曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

曹乾　译

(东南大学　caoqianseu@163.com)

译者序

一切都应尽量简单，但不要太简单。

——爱因斯坦

范里安（H.R.Varian）的这本教科书，我断断续续翻译了一年多。中断的原因在于业余时间主要用于翻译 MWG 的《微观经济理论》（受人大出版社之托）。

这是一本经典教科书，二十多年时间过去了，仍然能在高级微观经济教科书领域占有重要位置。原因之一大概是因为范里安有很好的穿透力。出身于数学背景，使得范里安的文字也简单明确。MWG 的教科书在文字表达上要比 Varian, Myerson，Tirole 等相关教科书略逊一筹，尽管 MWG 已是最为重要的微观经济理论教科书。

严格意义上来说，这本教材已算不上高级教程，它的难度只比范里安的中级教程高那么一点点。适合于经管类专业本科生和一年级研究生学习。

翻译的目的当然是为了教学方便的需要，此书所有内容的知识产权归 Varian 及其出版社所有，请勿用于商业用途。

此版翻译基本无理解上的错误，但也许只是因为“身在庐山中”，若有错误包括笔误，欢迎指出。

曹乾　caoqianseu@163.com

东南大学　江苏.南京

2012 年 10 月

目录

序言

《微观经济分析》第一版出版于 1977 年。15 年过后，是时候来次大的修订了。对于第三版，我做了两类大的修改：结构上的修改和内容上的修改。

结构上的变化体现在我重新整理了内容，将它们按照“模块”分章。大部分章节的标题，对应着我写的本科生教材《中级微观经济学》。这使得读者在有需要时能够容易地参考本科教材。另一方面，如果本科生想进一步学习某一主题的高级内容，自然可以学习这本《微观经济分析》的相应章节。这种模块章节的安排还有另外两个好处：可以按照各种顺序学习；也更方便于查阅。

除了章节上的变化之外，内容上也进行了修订。首先，我重写了本书的大部分章节。现在的内容更简洁，当然我也希望它们更好懂了。其次，我更新了很多内容，尤其是关于垄断和寡头的内容——我根据 1980 年代产业组织理论的主要进展更新了这些内容。

第三，我增添了很多新内容。具体说，我增加了关于博弈论、资产市场、以及信息的章节。这些章节可以作为一年级研究生的引论性质的学习材料。对于这些主题，我没有提供深入的讨论，因为我发现最好在研究生二年级或三年级时再这么做，那时学生已经掌握了经济分析的标准工具。

第四，我新添了一些习题，并且提供了所有奇数习题的答案。我必须说我对提供答案这种做法感到矛盾——但我希望大多数研究生应该有足够的意志力不去翻阅答案，除非他们已经亲自努力解答习题。

本书的结构

正如我在前面指出的，本书的编排特征是每一章都比较短。我认为几乎每个人都应该系统地学习本书前半部分，因为它描述的微观经济学基本工具对于所有经济学家都是有用的。本书的后半部分包含一些微观经济学主题的介绍。我估计大多数会选学部分内容。有些教师想强调博弈论，有些教师喜欢讲授一般均衡。有些课程花费很多时间来学习动态模型；另外一些课程则强调福利经济学的学习。

我不可能对所有这些主题都提供详尽的介绍，因此我决定只提供引论性质的材料。我试图使用本书第一部分描述的符号和方法，从而使得这些章节能为教材或专业期刊文章的更详尽阐述铺平道路。幸运的是，目前已有几本教材详细讨论资产市场、博弈论、信息经济学和一般均衡理论。总之，认真的学生在学习这些主题时不会缺乏相应的学习材料。

本书的制作

在重写这本书时，我已将所有材料都搬到 Donald Knuth 的 TEX 系统上。我认为现在这本书好看多了；而且，交叉参考、公式编号、索引等等对于作者和读者来说都变得容易多了。由于现在作者修改教材的成本大幅减少了，读者应该能期望看到更频繁的更新。（也许我在下一版中会将该段落中的最后一句编成习题…）

本书的一部分是在 MS-DOS 系统下写成的，但是大部分是在 NeET 计算机上完成的…

致谢

这些年，很多人写信告诉我教材中出现的打印错误，或者提供评语与建议。他们是（略去）。

最后，我给读者提供一个建议。当你读这本书时，请牢记 Sir Richard Steele（1672-1729）的不朽名言：“It is to be noted that when any part of this paper appears dull there is a design in it.”

密歇根大学安娜堡

1991 年 11 月

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 1 章：技术

曹乾 译

(东南大学 caoqianseu@163.com)

1 技术

刻画厂商技术最简单和最常见方法是生产函数，中级微观经济学课程中一般已研究生产函数。然而，刻画厂商的技术还有另外一些方法，这些方法在某些场景下更一般而且更有用。本章我们将讨论几种表示厂商成产可能性的方法，以及用经济学的语言描述与厂商技术相关的问题。

1.1 投入与产出的衡量

厂商使用各种投入的组合生产产出。为了研究厂商的选择，我们需要一种概括厂商生产可能性的简便方法，即哪种投入和产出的组合在技术上是可行的。

用流量（flows）衡量投入和产出通常最令人满意：单位时段的一定数量的投入被用于生产单位时段的一定数量的产出。在描述投入和产出时，最好说明时间尺度。如果你这么做，你就不大可能犯诸如以下的低级错误：使用了不可比较的单位；混淆了存量和流量；等等。例如，如果我们用每周多少小时计量劳动时间，我们要保证资本运行时间也用每周多少小时衡量，产量用每周多少单位衡量。然而，当我们抽象地讨论技术选择时，比如本章，我们通常省略时间尺度。

我们可能还希望按以下几个方面区分投入和产出：何时得到它们的；何地得到它们的；甚至在什么样的环境下才能得到它们。使用何时和何地修饰投入和产出，我们可以描述生产的时间或空间性质。例如，某既定年份得到的混粘土可用于建筑某幢大楼，这幢大楼将于下一年完工。类似地，在某地点购买的混凝土可用于其它地点的建筑施工。

“混凝土”这种投入应被认为是某特定等级、在某特定地点和特定时间得到的混凝土。在有些情形下，我们甚至需要把上述限定条件再加上诸如“如果气候干燥”这样的条件，也就是是说，我们可能考虑混凝土是在什么样的环境或者自然状态下得到的。至于应该对投入和产出描述到什么样的细致程度，取决于我们具体研究的问题，但是我们应该意识到，如果需要，我们就能将某特定投入和产出描述得非常详尽。

1.2 对技术的说明

假设厂商有 n 种可能的商品，这 n 种商品可以被用作投入和/或产出。如果某厂商使用 y_j^i 单位的商品 j 作为投入，生产 y_j^o 单位的产出，则商品 j 的净产出（net out）为 $y_j = y_j^o - y_j^i$。如果商品 j 的净产出为正，则表明厂商正在生产的商品 j 的数量多于将其用作投入品的数量；如果净产出为负，则表明厂商正在使用商品 j 作为投入品的数量多于他生产的该商品数量。

一个生产方案（A production plan）就是一组不同商品的净产出数值。我们可以用 R^n 中的向量 y 表示某个生产方案，其中：y_j 为负若商品 j 是净投入，y_j 为正若商品 j 是净产出。所有技术上可行的生产方案组成的集合称为厂商的生产可能集（production possibilities set），我们用 R^n 中的子集 Y 表示。集合 Y 描述了投入和产出的所有技术上可行的模式。生

产可能集Y完整地描述了厂商面对的技术可行性。

当我们研究处于某特定经济环境下厂商的行为时，我们希望将生产方案区分为“立即可行”的方案和“最终可行”的方案。例如，在短期（short run），厂商的某些投入是固定不变的，因此，只有与这些固定投入要素相容的方案才是可行的。在长期（long run），这些固定投入是可变的，因此商场的技术可能性也可能改变。

我们一般假设上述限制可用R^n中的向量z表示。例如，向量z是某特定时期下的一组各种投入和产出的最大数值。受限的生产可能集或称短期生产可能集用$Y(z)$表示；该集合由所有满足约束条件z的可行净产出束（net output bundles）组成。例如，假设要素n在短期的数量固定为$\bar{y}_n$，则$Y(\bar{y}_n)=\{Y$中的$y: y_n=\bar{y}_n\}$。注意，$Y(z)$是Y的子集，这是因为Y包含了所有生产可行方案，这意味着$Y(z)$包含在Y中而且$Y(z)$还要满足某些额外条件。

例子：必要投入集

假设我们要研究的厂商只生产一种产品。在这种情形下，我们将净产出束写为$(y,-x)$，其中x为生产y单位产品的投入向量。这样，我们可以定义一种特殊的受限生产可能集，即必要投入集(input requirement set):

$$V(y)=\{R_+^n\text{中的}x:(y,-x)\text{在}Y\text{中}\}$$

必要投入集是能至少生产y单位产出的所有投入束组成的集合。

注意，此处定义的投入必要集中，投入用正数表示，而在生产可能集中我们用负数表示净投入。

例子：等产量集

根据前面的例子，我们可以定义一个等产量集：

$$Q(y)=\{R_+^n\text{中的}x:x\text{在}V(y)\text{中但不在}V(y')\text{中，其中}y'>y\}$$

等产量集由恰好生产y单位产出的所有投入束组成。

例子：短期生产可能集

假设某企业使用劳动以及某类机器（我们将其称为“资本”）生产某些产出。于是，生产方案为$(y,-1,-k)$，其中：y为产出水平，1 为劳动投入两，k 为资本投入量。假设在短期，劳动可以立即改变但资本固定在水平$\bar{k}$。则

$$Y(\bar{k})=\{Y\text{中的}(y,-1,-k):k=\bar{k}\}$$

就是短期生产可能的一个例子。

例子：生产函数

如果厂商只有一种产出，我们可以将生产函数定义为

$f(x)=\{R$中的$y:y$是与$-x$关联的最大产出，其中y和$-x$都在Y中$\}$.

例子：转换函数

类比一元生产函数，我们可以定义 n-维“生产函数”，这个函数在研究一般均衡理论时将非常有用。Y 中的某个生产方案 y 是有效率的（efficient），如果我们无法在 Y 中找到 y'（$y' \neq y$）使得 $y' \geq y$ 成立；也即是说如果对于某生产方案来说，已无法使用相同的投入生产更多的产出，或者已无法使用更少的投入生产相同的产出，则该生产方案是有效率的。（仔细注意投入的符号规则）。我们通常假设可将技术上有效率的生产方案用某个转换函数（transformation function）$T: R^n \to R$ 表示，其中 $T(y)=0$ 当且仅当 y 是有效率的。正如生产函数选取最大的标量产出(scalar output)作为投入的生产函数一样，转换函数选取的是最大的净产出向量。

例子：柯布-道格拉斯技术

令参数 a 满足 $0<a<1$，则柯布-道格拉斯（Cobb-Douglas）可用以下各种方式定义。请见图 1.1A。

$$Y=\{R^3 \text{ 中的 } (y,-x_1,-x_2)\text{：} y \leq x_1^a x_2^{1-a}\}$$

$$V(y)=\{R_+^2 \text{ 中的 } (x_1,x_2): y \leq x_1^a x_2^{1-a}\}$$

$$Q(y)=\{R_+^2 \text{ 中的 } (x_1,x_2): y = x_1^a x_2^{1-a}\}$$

$$Y(z)=\{R^3 \text{ 中的 } (y,-x_1,-x_2): y \leq x_1^a x_2^{1-a}, x_2=z\}$$

$$T(y,x_1,x_2)=y-x_1^a x_2^{1-a}$$

$$f(x_1,x_2)=x_1^a x_2^{1-a}$$

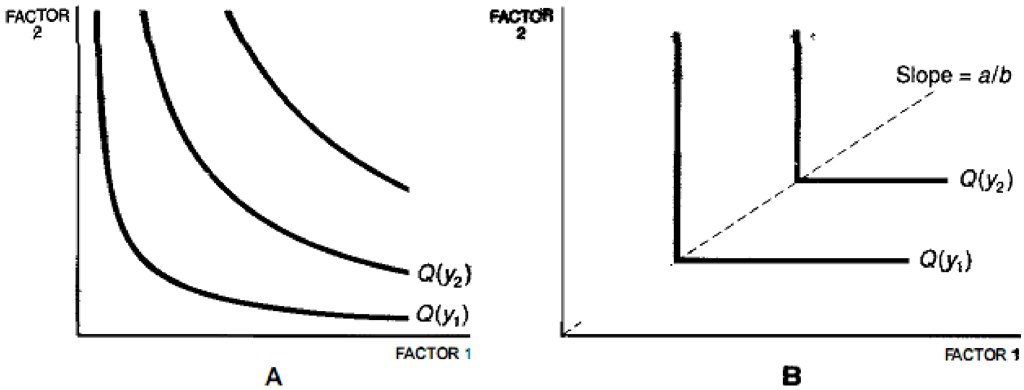


图 1.1:**柯布-道格拉斯技术和里昂惕夫技术**。A 图描述了柯布-道格拉斯技术的一般形状；B 图描述了里昂惕夫技术的一般形状。

例子：里昂惕夫技术

令参数 $a>0, b>0$，则里昂惕夫(Leontief)技术可用以下各种方式定义。请见图 1.1B。

$$Y=\{R^3 \text{ 中的 } (y,-x_1,-x_2)\text{：} y\le \min\{ax_1,bx_2\}\}$$

$$V(y)=\{R_+^2 \text{ 中的 } (x_1,x_2): y\le \min\{ax_1,bx_2\}\}$$

$$Q(y)=\{R_+^2 \text{ 中的 } (x_1,x_2): y= \min\{ax_1,bx_2\}\}$$

$$T(y,x_1,x_2)=y-\min\{ax_1,bx_2\}$$

$$f(x_1,x_2)=\min\{ax_1,bx_2\}$$

在本章，我们主要分析只生产一种产出的厂商；因此，我们通常用必要投入集或生产函数描述它们的生产技术。以后我们再使用生产集和转换函数。

1.3 生产技术分析

描述生产或必要生产集的最简单的方法，就是列举各种可行的生产方案。例如，假设我们可用要素 1 和 2 生产一种产品。生产方法有两种不同的**途径**（activities）或**技术**（techniques）：

技术 A：一单位要素 1 和两单位要素 2 能产出一单位产品。

技术 B：两单位要素 1 和一单位要素 2 能产出一单位产品。

令产出为商品 1，投入要素为商品 2 和 3，则我们可以将上述两种技术暗含的生产可能性用下列生产集表示

$$Y=\{(1,-1,-2),(1,-2,-1)\}$$

或用下列必要**投入**集表示

$$V(1)=\{(1,2),(2,1)\}.$$

该必要投入集请见图 1.2A。

在生产 y 单位产出时，有可能出现下面的情形，即每种要素的投入量恰好是生产一单位产出所需要素的 y 倍，其中 y=1,2…。在这种情形下，你可能会认为生产 y 单位产出的所有可行方法的集合为

$$V(y)=\{(y,2y),(2y,y)\}.$$

但是，上述集合漏掉了部分可行生产方法。的确，若使用技术 A，则 $(y,2y)$ 能生产 y 单位的产出；若使用技术 B，则 $(2y,y)$ 也能生产 y 单位产出。但是，你忘了我们还可以同时使用这两种方法进行生产。

在这种情形下，为避免混淆，我们使用 y_A 表示技术 A 的产量，y_B 表示技术 B 的产量。于是 $V(y)$ 可用下列集合表示

$$V(y)=\{(y_A+2y_B),(y_B+2y_A): y=y_A+y_B\}.$$

因此，例如$V(2)=\{(2,4),(4,2),(3,3)\}$，如图 1.2B 所示。注意投入束(3,3)也能生产两单位产出：技术 A 生产了一单位，技术 B 生产了一单位。

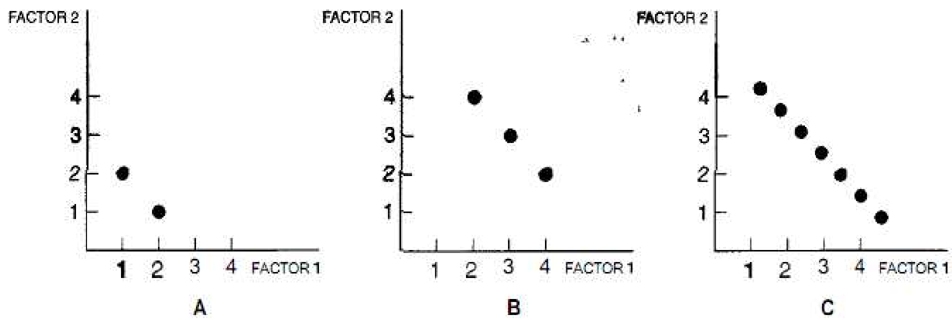


图 1.2：**必要投入集**。A 图表示 V（1）；B 图表示 V（2）；C 图表示 V（y），其中 y 为更大的产量。

1.4 单调的技术

我们继续分析上一节介绍的两种生产技术的例子。假设我们有个投入向量（3,2)。它足以生产一单位产出吗？我们可能认为，既然可以去掉 2 单位要素 1，这样只剩下（1,2)，因此技术 A 可确保（3,2）能成产一单位产出。所以，如果允许这样的**自由取舍**（free disposal），则当然可以认为：如果投入向量x是能生产 y 单位产出的一种可行方法，并且投入向量x'中的每种投入都至少和x一样多，则x'也是生产 y 单位产出的可行方法。因此，必要投入集在下列意义上是**单调的**（monotonic）：

单调性（monotonicity）。**若**x**在**$V(y)$**中且**$x'\geq x$**，则**x'**也在**$V(y)$**中**。

如果我们假设了单调性，则图 1.2 中的必要投入集现在变为图 1.3 中的相应集合。

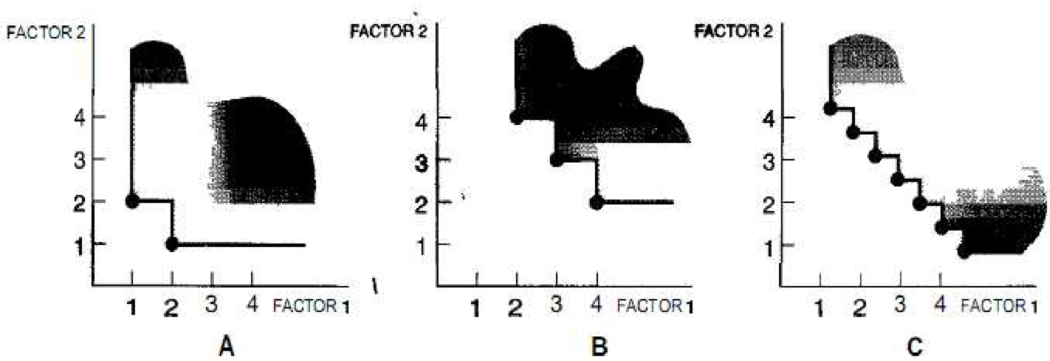


图 1.3：**单调性**。在单调性假设下，图 1.2 中的必要投入集就相应呈现上图的形状。

单调性对于生产集来说通常是个合适的假设。在生产集中，我们通常假设如果 y 在 Y 中而且若$y'\leq y$，则y'必定也在 Y 中。提醒你注意此处的符号规则。$y'\leq y$表示向量y'的每个分量都小于或等于向量y的相应分量。这意味着与y代表的生产方案相比，y'代表的生产方案使用至少一样多的所有各种投入，但生产出的产量相同或者更少。因此，自然可以

假设如果 y 可行，则 y' 也可行。

1.5 凸的技术

现在我们考虑如果生产 100 单位的产出，必要投入集将会呈现什么样的形状。作为第一步，我们可能认为若我们将向量（1,2）和（2,1）乘以标量 100，我们应该能**复制**（replicate）前面介绍的两种技术，因此可以生产 100 倍得产量。显然不是所有的生产过程都能允许这样的复制，但是在很多环境下，复制似乎是合理的。

如果上述复制可行，我们可以断言（100,200）和（200,100）都在 V（100）中。还有没其他可能方式来生产 100 单位产出？有，我们可以将技术 A 和 B 各复制 50 次，即使用 150 单位要素 1 和 150 单位要素 2 生产 100 单位产出；因此，（150,150）也应该在必要投入集之中。类似地，我们可以将技术 A 复制 25 次将技术 B 复制 75 次。这意味着

$$0.25(100,200)+0.75(200,100)=(175,125)$$

应该在 V（100）中。更一般地，

$$t(100,200)+(1-t)(200,100)=(100t+200(1-t),200t+100(1-t))$$

应该也在 V（100）中，其中 $t=0, 0.01, 0.02, ..., 1$.

我们不妨作出明显近似处理，令 t 取闭区间[0,1]内的任何小数，则可以得到图 1.4A 形状的生产集。下一个定义准确介绍了这个性质：

凸性（Convexity）。**若 x 和 x' 都在 $V(y)$ 中，则 $tx+(1-t)x'$ 也在 $V(y)$ 中，其中 $0\le t\le 1$。也就是说 $V(y)$ 是个凸集（convex set）。**

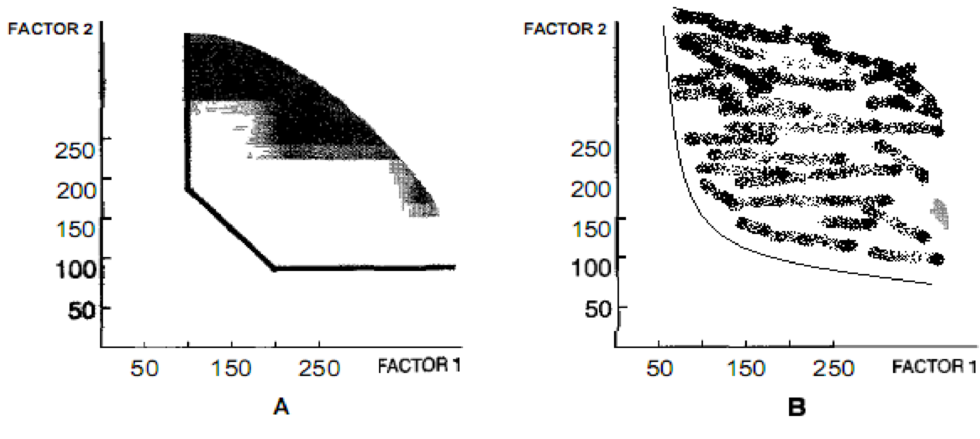


图 1.4: 凸的必要投入集。如果 x 和 x' 都可以生产 y 单位产出，则任何加权平均 $tx+(1-t)x'$ 也能生产 y 单位产出。图 A 中的凸必要投入集是两种生产技术下的情形；图 B 中的凸必要投入集是很多种生产技术下的情形。

生产技术可以复制驱使我们作出凸性的假设。如果我们想生产“大量的”产出，并且如果我们可以复制“小的”生产过程，那么假设技术为凸就比较合理。然而，如果生产技术的规模相对很大，而想要生产的产量却相对较小，那么这种情形下再假设技术为凸就不合理了。

然而，我们还有其他的理由说明为何凸性在某些环境下是合理的假设。例如，假设我

曹乾（东南大学 caoqianseu@163.com）

们考虑的是每月的产出。如果一个投入向量 x 生产 y 单位的产生每月，另外一个向量 x' 也可以生产 y 单位的产出每月，则我们可以用 x 和 x' 各生产半个月。如果转换生产方法不会产生任何问题，我们有理由认为产出为 y 单位。

我们曾将上述论断应用于必要投入集，但类似的论断也可以用于生产集。通常可以假设 y 和 y' 都在 Y 中，则 $ty+(1-t)y'$ 也在 Y 中，其中 $0 \le t \le 1$；换句话说，Y 是一个凸集。然而，需要指出的是，与假设必要投入集为凸相比，假设生产集为凸的问题更大些。例如，凸生产集排除了"启动成本"（start up costs）和其他形式的规模报酬。我们稍后会更详细地分析这个问题。现在我们只简单描述 V(y)的凸性、生产函数的曲率和 Y 的凸性之间的一些关系。

定理(凸生产集蕴涵着凸必要投入集)：若生产集 Y 是一个凸集，则与其相关的必要投入集 $V(y)$，也是一个凸集。

证明。若 Y 是个凸集，这意味着对于任何 x 和 x'，只要 $(y,-x)$ 和 $(y,-x')$ 都在 Y 中，则 $(ty+(1-t)y,-tx-(1-t)x')$ 必然也在 Y 中。这只是要求 $(y,-(tx+(1-t)x'))$ 也在 Y 中。由此可见，若 x 和 x' 都在 $V(y)$ 中，则 $tx+(1-t)x'$ 也在 $V(y)$ 中，这表明 $V(y)$ 是凸集。■

定理（凸必要投入集等价于拟凹生产函数）：$V(y)$ 是一个凸集当且仅当生产集 $f(x)$ 是一个拟凹函数。

证明。$V(y)=\{x: f(x)\ge y\}$，这正好是 $f(x)$ 的**上等高集**（upper contour set）。但一个函数是拟凹的当且仅当它又一个凸的上等高集；参见第 27 章拟凹函数和拟凸函数那一节。■

1.6 正则技术

最后我们考虑关于 $V(y)$ 的一种弱正则条件（a weak regularity condition）。

正则性。$V(y)$ 对于所有 $y \ge 0$ 来说是闭的非空集合。

$V(y)$ 是非空的这个假设，要求对于任何既定的产出水平都有法生产。这个要求只是想避免每次都需要强调"假设 y 能够被生产出。"

作出 $V(y)$ 是闭的这个假设是基于技术原因，在多数情形下这个假设是合理的。假设 $V(y)$ 为闭的一种解释是：假设我们的投入束序列 (x^n) 中每一个投入束都可以生产 y 单位产出，而且这个序列收敛于投入束 x^0。也就是说，这个序列中的投入束无限接近于 x^0。若 $V(y)$ 是一个闭集则该极限投入束 x^0 必然能成产出 y。大致来说，必要投入集必定"包含自身的边界。"

1.7 技术的参数表示

假设我们有很多种方法生产某既定产出水平。则我们自然可以将这个投入集用图 1.5 中的"平滑的"投入集表示。也就是说我们想用良好的曲线拟合穿过可能生产点的曲线。如果生产既定产量产出的方法的确有很多，则这样的平滑化的过程就不会带来比较大的问题。

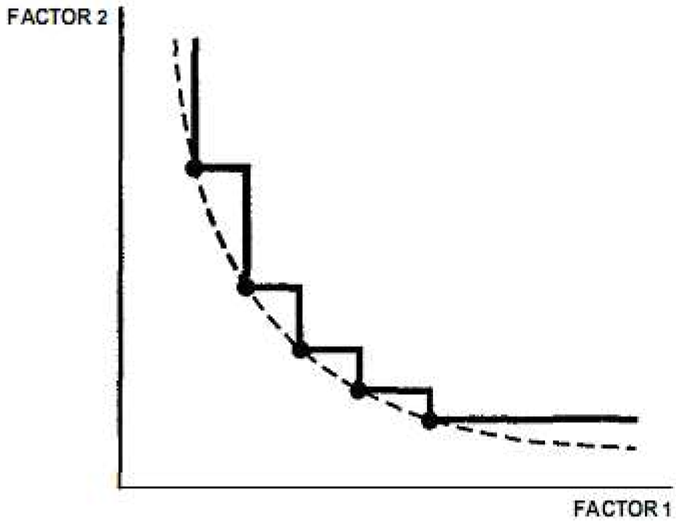


图 1.5：将等产量线平滑化。一个必要投入集及其“平滑”近似。

如果我们用这样的近似去“光滑”必要投入集，自然我们想寻求表示技术的一种简便方法，即带有若干未知参数的参数方程。例如，前面提到过的柯布-道格拉斯技术蕴涵着，满足 $x_1^a x_2^b \geq y$ 的任何投入束 (x_1, x_2) 都能生产至少 y 单位的产量。

你不应该将这些参数技术表示仅看作为生产可能性的文字描述。生产可能性是描述实际生产方案的工程数据。这些工程数据完全有可能用比较方便的函数形式比如柯布-道格拉斯函数表述。如果行得通，这样的参数描述将会非常有用。

在绝大多数应用中，我们只关心是在某些特定投入和产出水平范围内，能否找到对某技术的参数化描述，而且通常用相对简单形式的函数去做参数化近似。这些参数化表示是非常简便的教学工具，因此我们通常假设我们的技术能用这样的参数表示。这样，我们就可以使用微积分和代数工具区分析企业的生产选择决策。

1.8 技术替代率

假设我们的技术可以用一个平滑的生产函数表示，而且我们在某个特定点 $y^* = f(x_1^*, x_2^*)$ 进行生产。假设为维持产量水平不变，我们想增加要素 1 的投入量减少要素 2 的投入量。我们如何确定这两种要素之间的**技术替代率**（technical rate of substitution，TRS）？

在二维情形下，技术替代率就是等产量线的斜率：当 x_1 微小变动时，你应该怎样调整 x_2 才能使产量不变，如图 1.6 所示。在 n 维情形下，技术替代率是等产量面在特定方向上的斜率。

令 $x_2(x_1)$ 表示隐函数，这个函数告诉我们如果我们投入 x_1 单位要素 1，则还需要投入多少 x_2 才能生产 y 单位产量。于是根据定义可知，函数 $x_2(x_1)$ 必定满足下列不等式

$$f(x_1, x_2(x_1)) \equiv y$$

我们想要的是 $\partial x_2(x_1^*)/\partial x_1$ 的表达式。对上式微分可得

$$\frac{\partial x_2(x^*)}{\partial x_1}+\frac{\partial x_2(x^*)}{\partial x_1}\frac{\partial x_2(x_1^*)}{\partial x_1}$$

$$\frac{\partial x_2(x_1^*)}{\partial x_1}=-\frac{\partial f(x^*)/\partial x_1}{\partial f(x^*)/\partial x_2}.$$

其中 $x^*=(x_1^*,x_2(x_1^*))$ 。

这样我们就得到了技术替代率的显性表达式。

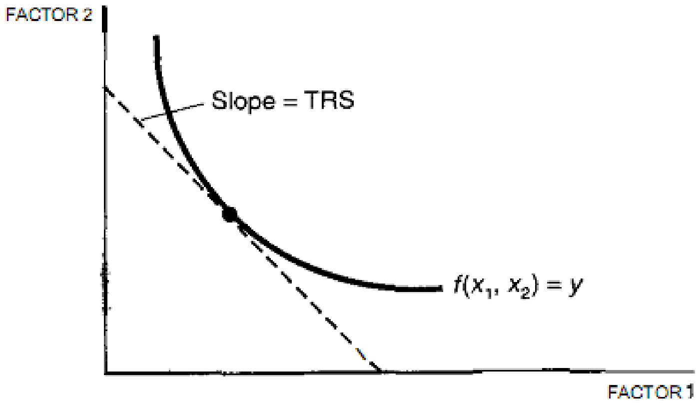


图 1.6：技术替代率。技术替代率衡量当一种投入变动时，我们必须如何调整另外一种投入才能使产量不变。

下面我们介绍推导技术替代率的另外一种方法。考虑投入向量的在投入水平上的（微小）变化，我们将其记为 $dx=(dx_1,dx_2)$ 。相应的产出变化近似为

$$dy=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2.$$

上述表达式称为函数 $f(x)$ 的全微分。考虑只有要素 1 和要素 2 变动的这种特殊情形，而且要素投入量的这种变动不会改变产量大小。也就是说， dx_1 和 dx_2 “沿着等产量线”调整。

由于产量维持不变，可得

$$0=\frac{\partial f}{\partial x_1}dx_1+\frac{\partial f}{\partial x_2}dx_2,$$

从该式可以解得

$$\frac{dx_2}{dx_1}=-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}.$$

在计算技术替代率时，我们可以使用隐函数的方法，也可以使用全微分的方法。隐函数的方法相对严格一些，但是全微分的方法可能相对直观。

例子：柯布-道格拉斯的技术替代率

给定函数 $f(x_1,x_2)=x_1^a x_2^{1-a}$，求导可得

$$\frac{\partial f(x)}{\partial x_1}=ax_1^{a-1}x_2^{1-a}=a[\frac{x_2}{x_1}]^{1-a}$$

$$\frac{\partial f(x)}{\partial x_2}=(1-a)x_1^a x_2^{-a}=(1-a)[\frac{x_1}{x_2}]^a .$$

由此可得

$$\frac{\partial x_2(x_1)}{\partial x_1}=-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}=-\frac{a}{1-a}\frac{x_2}{x_1} .$$

1.9 替代弹性

技术替代率衡量无差异曲线的斜率。替代弹性衡量等产量线的曲率。更具体地说，替代弹性衡量在产量维持不变的情形下，要素投入比率的变动百分比除以 TRS 变动百分比。如果令 $\Delta(x_2/x_1)$ 表示要素投入比率的变动，令 ΔTRS 表示技术替代率的变动，我们可以将替代弹性表达为

$$\sigma=\frac{\frac{\Delta(x_2/x_1)}{x_2/x_1}}{\frac{\Delta TRS}{TRS}}$$

这是曲率的相对自然的衡量方法：它问的是当等产量线的斜率变动时，要素投入比率如何变动。如果斜率的较小变动引起要素投入比率相对较大的变动，则等产量是相对平坦的——替代弹性较大。

在实践中，我们可以将变动百分比想象为非常小，并对上述表达式求当 ΔTRS 趋于 0 时的极限。因此，替代弹性 σ 的表达式变为

$$\sigma=\frac{TRS}{(x_2/x_1)}\frac{d(x_2/x_1)}{dTRS}.$$

用对数形式的导数计算替代弹性通常比较方便。一般来说，若 $y=g(x)$，则 y 对于 x 的弹性是指 x 的（微小）变动百分比引起的 y 变动的百分比。

$$\varepsilon=\frac{dy/y}{dx/x}=\frac{dy}{dx}\frac{x}{y}.$$

如果 x 和 y 均为正，则这个导数可以写为

$$\varepsilon=\frac{d\ln y}{d\ln x},$$

下面证明这个结论。由链式法则可知

$$\frac{d\ln y}{d\ln x}\frac{d\ln x}{dx}=\frac{d\ln y}{dx}.$$

将等式左右两端的项分别计算可得

$$\frac{d\ln y}{d\ln x}\frac{1}{x}=\frac{1}{y}\frac{dy}{dx},$$

$$\frac{d\ln y}{d\ln x}=\frac{x}{y}\frac{dy}{dx}。$$

另外一种证明方法是使用全微分：

$$d\ln y=\frac{1}{y}dy$$

$$d\ln x=\frac{1}{x}dx,$$

因此

$$\varepsilon=\frac{d\ln y}{d\ln x}\equiv\frac{dy}{dx}\frac{x}{y}.$$

同样，第一种方法比较严格，但第二种方法比较直观。

将上述表达式应用于替代弹性可得

$$\sigma=\frac{d\ln(x_2/x_1)}{d\ln|TRS|}.$$

(分母中的绝对值符号是将 TRS 转换为正数，从而使取对数有意义。)

例子：柯布-道格拉斯生产函数的替代弹性

我们已经知道

$$TRS_{12}=-\frac{a}{1-a}\frac{x_2}{x_1}$$

或

$$\frac{x_2}{x_1}=-\frac{a}{1-a}TRS_{12}$$

由此可得

$$\ln\frac{x_2}{x_1}=\ln\frac{1-a}{a}+\ln|TRS_{12}|.$$

这又意味着

$$\sigma = \frac{d\ln(x_2/x_1)}{d\ln|TRS|} = 1 .$$

1.10 规模报酬

假设我们使用某个投入向量 x 生产某既定的产量 y，现在我们决定将所有投入按比例同时增加或减少，使其变为原来的 t 倍，产量将怎样变化？

在前面的情形中，我们只希望产量等比例增加（scale up），我们通常假设只要将以前的生产模式复制，就能生产出 t 倍的产量。如果这种等比例增加总是可行，我们说生产技术呈现规模报酬不变(constant returns to scale)的现象。更正式地，

定义（规模报酬不变）：某生产技术呈现规模报酬不变的现象，若它满足下列条件：

（1） y **在 Y 中蕴涵着** ty **也在 Y 中，其中** $t \geq 0$**；**

（2） x **在** $V(y)$ **中蕴涵着** tx **在** $V(ty)$ **中，其中** $t \geq 0$**；**

（3） $f(tx) = tf(x)$**，其中** $t \geq 0$**；即生产函数是** $f(x)$ **是一阶齐次函数。**

上述技术可以复制的论断表明，规模报酬不变通常是对技术的一种合理假设。然而，在某些情形下，再假设规模报酬不变就不合理了。

例如，当我们试图“细分”（subdivide）生产过程时，规模报酬不变的假设就不再成立。即使我们能将生产过程以整数倍复制，但可能无法做到将生产过程细分 n 倍。例如，某种生产技术可能存在最小生产规模，如果产量低于该最小生产规模，可能需要使用不同的技术。一旦产量超过最小生产规模，我们就可以复制这种生产技术生产更多的产量。

规模报酬不变假设不再合理的另外一种情形是，当生产过程复制的倍数不是整数时。当然，前面所说的整数倍的复制很简单，如果我们要复制 1.5 倍，应该怎么做？

这两种情形下，只有当生产规模相对于最小产出规模比较小时，规模报酬不变的假设才不合理。

规模报酬不变假设不合理的第三种情形是，当所有投入都翻倍，但此时出现了更有效率的生产方法。复制是说将投入翻倍产出也翻倍，但也有可能存在生产更多产量的方法。例如，某企业在两点之间铺设油管，使用的投入为劳动、及其设备和钢材。我们可以使用油管线路的运输能力作为该企业的产出。显然如果我们将所有投入翻一番，则产量将不止翻一番，因为油管的表面积变为原来的 2 倍，则体积变为原来的 4 倍[一]。在这种情形下，产量增加的倍数超过投入增加的倍数，我们说这样的技术是规模报酬递增的（increasing returns to scale）。

定义（规模报酬递增）：若 $f(tx) > tf(x)$ **（其中** $t > 1$**），则该技术是规模报酬递增的。**

规模报酬假设不合理的第四种情形是有些产出是不能复制的。例如，考虑某 100 亩农田。

[一] 当然，体积更大的油管更难制造，因此我们不需要认为产量一定恰好变为原来的 4 倍。但是，产量很有可能大于原来的 2 倍。

如果我们希望产量变为原来的 2 倍，则每种要素投入量也要变为原来的 2 倍。但这意味着土地面积也应变为原来的 2 倍。在这种情形下，可能我们无法得到那么大的土地面积。尽管**所有**投入增加相同比例，技术可能会呈现规模报酬不变的特征，但是这种情形下，我们通常说，对于**受限制**的投入来说，技术呈现**规模报酬递减**（decreasing return to scale）的特征。更准确地说，我们有：

定义（规模报酬递减）：若**若 $f(tx) < tf(x)$（其中 $t > 1$），则该技术是规模报酬递减的**。

规模报酬递减的最自然情形是我们无法复制某些投入。因此，我们应该预期受限（restricted）生产集通常会具有规模报酬递减的特征。可以证明，规模报酬递减总是由于某些投入是固定不变而引起的。

为证明这个结论，假设 $f(x)$ 是 k 种要素的生产函数，并且具有规模报酬递减的性质。那么，我们可以引入一种新的"虚构的"投入，用 z 表示它的投入数量。定义一个新的生产函数

$$F(z,x) = zf(x/z).$$

注意，函数 F 具有规模报酬不变性质。如果我们将所有投入（x 和 z）同乘以 $t \geq 0$，则产量也变为原来的 t 倍。如果我们将 z 固定为 1，此时的技术和以前一样（即 $F(x) = f(x)$），因此，我们可以认为原来的规模报酬递减技术 $f(x)$，是由对规模报酬不变的技术 $F(z,x)$ 施加 z=1 的这个限制而得到的。

最后，需要指出，前面介绍的几种规模报酬在本质上是全局（global）性质的，比如在定义域内技术为规模报酬不变的，等等。但是很有可能出现下列现象：某种技术对于某些 x 值来说是规模报酬递增的，但对另外一些 x 值来说又是规模报酬递减的。因此，在很多情形下，刻画规模报酬的局部（local）性质就比较有用。**规模弹性**（elasticity of scale）衡量所有投入同时增加 1%（即由于**规模**增加）引起的产量增加百分比。

令 $y = f(x)$ 表示生产函数，t 为一个正的标量。考虑函数 $y(t) = f(tx)$。若 $t = 1$，则生产模就是我们当前的生产规模；若 $t > 1$，则我们将所有投入都增加为原来的 t 倍；若 $t < 1$，则所有投入都减少为原来的 t 倍。

规模弹性的表达式为

$$e(x) = \frac{\frac{dy(t)}{y(t)}}{\frac{dt}{t}}$$

将上式变形并计算 $t = 1$ 时的数值可得

$$e(x) = \frac{dy(t)}{y(t)} \frac{t}{y}\Big|_{t=1} = \frac{df(tx)}{dt} \frac{t}{f(tx)}\Big|_{t=1}.$$

注意，在计算点 x 的规模弹性时，我们必须计算该表达式在 $t = 1$ 时的数值。当 $e(x)$ 大于、等于或小于 1 时，我们说技术分别呈现规模报酬局部递增、局部不变或局部递减的性质。

例子：规模报酬和柯布-道格拉斯技术

假设 $y=x_1^a x_2^b$ 。则 $f(tx_1,tx_2)=(tx_1)^a(tx_2)^b=t^{a+b}x_1^a x_2^b=t^{a+b}f(x_1,x_2)$ 。因此 $f(tx_1,tx_2)=tf(x_1,x_2)$ 当且仅当 $a+b=1$ 时。类似地，$a+b>1$ 意味着规模报酬递增，$a+b<1$ 意味着规模报酬递减。

事实上，柯布-道格拉斯技术的规模弹性恰好为 $a+b$。为了证明这一点，我们应用规模弹性的定义：

$$\frac{d(tx_1)^a(tx_2)^b}{dt}=\frac{dt^{a+b}x_1^a x_2^{\,b}}{dt}=(a+b)t^{a+b-1}x_1^a x_2^{\,b}.$$

计算 $t=1$ 时的数值，再除以 $f(x_1,x_2)=x_1^a x_2^b$ 即可得到我们想要的结果。

1.11 齐次技术和位似技术

函数 $f(x)$ 称为 k 次齐次的，若 $f(tx)=t^k f(x)$ 对于所有 $t>0$ 成立。经济学中最重要的齐次函数是零次齐次和一次齐次的。零次齐次函数具有 $f(tx)=f(x)$ 的性质，而一次齐次函数则具有 $f(tx)=tf(x)$ 的性质。

将一次齐次函数的定义与规模报酬不变的定义相比较可知，某技术是规模报酬不变的当且仅当它的生产函数是一次齐次的。

函数 $g:R\to R$ 称为**正单调变换**（positive monotonic transformation）若 g 是一个严格递增的函数；也就是说，对于该函数来说 $x>y$ 蕴涵着 $g(x)>g(y)$。（正单调变换还是负单调变换可从上下文判断出）。**位似函数**（homothetic function）是一次齐次函数的正单调变换。换句话说，$f(x)$ 是位似的当且仅当它可以写为 $f(x)=g(h(x))$，其中 $h(\cdot)$ 是一次齐次函数，$g(\cdot)$ 是个单调函数。图 1.7 提供了几何解释。

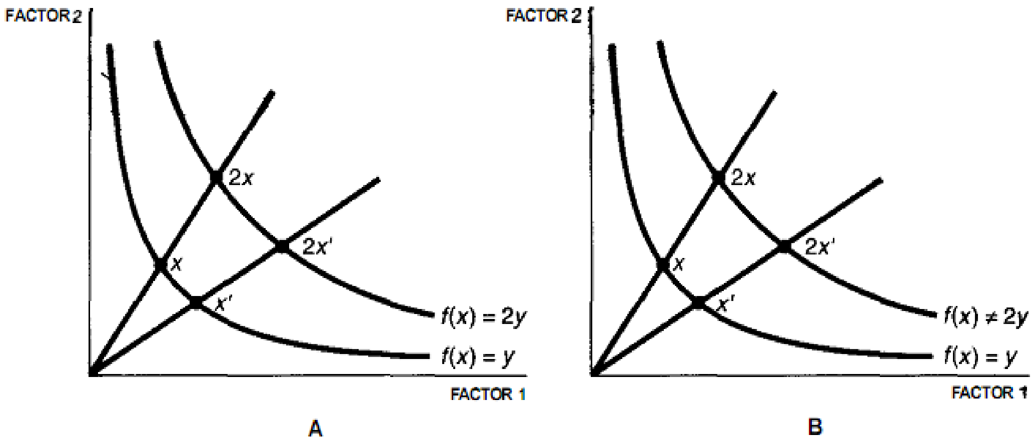


图 1.7：齐次函数和位似函数。A 图中的函数为一次齐次的。若 x 和 x' 都可以生产出 y 单位的产量，则 $2x$ 和 $2x'$ 都可以生产出 $2y$ 单位的产量；B 图中的函数为位似函数。若若 x 和 x' 都可以生产出**相同**水平的产量——y 单位的产量，则 $2x$ 和 $2x'$ 都可以生产出**相同**单位的产量，但未必是 $2y$ 单位的产量。

你可以将单调变换看成用不同单位来衡量产出的一种方法。例如，我们可以用升或毫

升衡量液体产量。在这种情形下，不同单位之间的转换非常简单——我们只要乘以 1000 或者处以 1000 即可。一种更奇怪的单调变换是我们用“升的平方”衡量产出。根据这种解释，位似技术是衡量产出的一种方法，使得这种技术“看上去类似”规模报酬不变。

我们对齐次函数和位似函数感兴趣，这是因为当产量变化时它们的等产量线的变化比较简单。在齐次函数的情形下，等产量线族就象是由一条等产量线相应“扩大”（blown up）而得到的。如果 $f(x)$ 是一次齐次的，则若 x 和 x' 都可以生产出 y 单位的产量，那么 tx 和 tx' 都可以生产出 ty 单位的产量，如图 1.7A 所示。位似函数具有几乎完全相同的性质：若 x 和 x' 都可以生产出相同单位的产量，则 tx 和 tx' 生产出的产量也相同——但未必是原来产量的 t 倍。位似技术的等产量线和齐次技术的等产量线非常类似，唯一区别是等产量线代表的产量水平是不同的。

齐次函数和位似函数比较有趣的原因是，它们对技术替代率施加了限制——当生产规模变化时，技术替代率是如何变化的。特别地，对于这两种函数来说，技术替代率和生产规模无关。

这个结论可以直接从第 26 章齐次函数的知识推导出，在那里我们指出若 $f(x)$ 是一次齐次的，则 $\partial f(x)/\partial x_i$ 是零次齐次的，这正是我们想要的结果。

例子：CES 生产函数

不变替代弹性（constant elasticity of substitution, CES）生产函数的表达式为

$$y = [a_1 x_1^{\rho} + a_2 x_2^{\rho}]^{\frac{1}{\rho}}.$$

容易验证，CES 函数具有规模报酬不变性质。对参数 ρ 取不同的数值，我们就可以得到几个著名的生产函数的特例。我们将这些函数介绍如下，并用图 1.8 描述这些函数。在我们的讨论中，令参数 $a_1 = a_2 = 1$ 将非常方便。

（1）线性生产函数（$\rho = 1$）。将 $\rho = 1$ 代入 CES 生产函数可得

$$y = x_1 + x_2 .$$

（2）柯布-道格拉斯生产函数（$\rho = 0$）。当 $\rho = 0$ 时 CES 生产函数无定义，因为除数为零。然而我们可以证明当 ρ 趋近于零时，CES 的等产量线看上去非常象柯布-道格拉斯生产函数的等产量线。

使用技术替代率很容易看清这一点。直接计算可得，

$$TRS_{12} = -(\frac{x_1}{x_2})^{\rho - 1}. \qquad (1.1)$$

当 ρ 趋近于零时，上式趋近于极限

$$TRS_{12} = -\frac{x_2}{x_1},$$

这正是柯布-道格拉斯生产函数的 TRS。

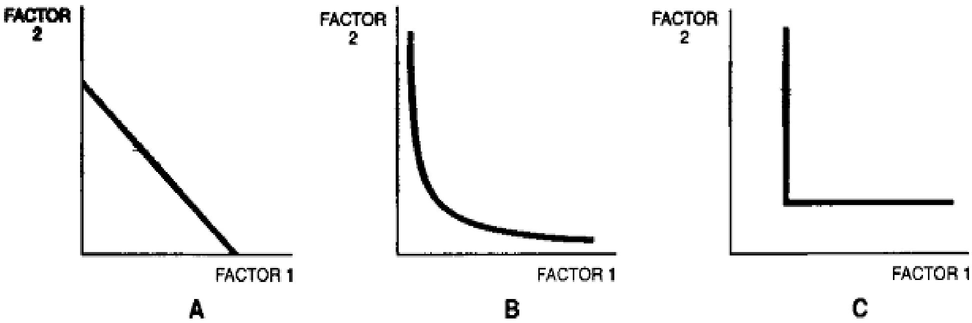


图 1.8：CES 生产函数。CES 生产函数的形状有多种，这取决于参数ρ的取值。A 图中的情形为$\rho=1$；B 图中的情形为$\rho=0$，C 图中的情形为$\rho=-\infty$。

（3）里昂惕夫生产函数（$\rho=-\infty$）。我们已经知道 CES 生产函数的 TRS 可由（1.1）式计算。当$\rho\to-\infty$时，（1.1）式趋近于

$$TRS_{12}=-(\frac{x_1}{x_2})^{-\infty}=-(\frac{x_2}{x_1})^{\infty}.$$

若$x_2>x_1$，则 TRS 是（负）无穷；若$x_2<x_1$，则 TRS 为零。这表示当$\rho\to-\infty$时，CES 等产量线看上去类似里昂惕夫技术的等产量线。■

CES 生产函数的替代弹性为常数，这一点你也许不会惊讶。为了验证这一点，注意 TRS 的计算公式

$$TRS_{12}=-(\frac{x_1}{x_2})^{\rho-1},$$

因此

$$\frac{x_2}{x_1}=\left|TRS_{12}\right|^{\frac{1}{1-\rho}}.$$

取对数可得

$$\ln\frac{x_2}{x_1}=\frac{1}{1-\rho}\ln\left|TRS_{12}\right|.$$

使用对数导数的弹性定义可得

$$\sigma=\frac{d\ln x_2/x_1}{d\ln\left|TRS\right|}=\frac{1}{1-\rho}.$$

注释

替代弹性的最初工作源于希克斯（Hicks,1932）。对于替代弹性的一般形式（n 种投入）请参考 Blackorby & Russell(1989)。规模弹性的工作源于 Frisch（1965）。

习题

1.1. 对还是错？若$V(y)$是一个凸集，则相应的生产集 Y 必定为凸。

1.2. 当$a_1 \neq a_2$时，一般 CES 技术$y=[a_1x_1^{\rho}+a_2x_2^{\rho}]^{\frac{1}{\rho}}$的替代弹性为多大？

1.3. 定义要素i的产出弹性为

$$\varepsilon_i(x)=\frac{\partial f(x)}{\partial x_i}\frac{x_i}{f(x)}.$$

如果$f(x)=x_1^a x_2^b$，每种要素的产出弹性为多大？

1.4 如果$\varepsilon(x)$是规模弹性，$\varepsilon_i(x)$是要素i的产出弹性，证明$\varepsilon(x)=\sum_{i=1}^{n}\varepsilon_i(x)$。

1.5CES 技术$f(x_1,x_2)=(x_1^{\rho}+x_2^{\rho})^{1/\rho}$的规模弹性为多大？

1.6 对还是错？ 可微函数$g(x)$为严格递增的当且仅当$g'(x)>0$。

1.7 教材中断言，如果$f(x)$是位似技术而且x和x'生产的产量水平相同，则tx和tx'生产的产量水平也相同。你能严格证明这个结论吗？

1.8 令$f(x_1,x_2)$是位似函数。证明它在(x_1,x_2)点的技术替代率等于在(tx_1,tx_2)点的技术替代率。

1.9 考虑 CES 技术$f(x_1,x_2)=[a_1x_1^{\rho}+a_2x_2^{\rho}]^{1/\rho}$。证明我们总可以将这个式子写为$f(x_1,x_2)=A(\rho)[bx_1^{\rho}+(1-b)x_2^{\rho}]^{1/\rho}$的形式。

1.10 令Y为一个生产集。如果y和y'都在Y中蕴涵$y+y'$也在Y中，那么我们说技术是可加的。如果y在Y中蕴涵ty也在Y中（其中$0\leq t\leq 1$），那么我们说该技术是可分的。证明如果某技术既是可加的又是可分的，则Y必定是凸的，而且它是规模报酬不变的。

1.11 对于每个必要投入集，确定它是正则的（regular）、单调的和/或凸的。假设参数a和b以及产量水平都是严格正的。

（a）$V(y)=\{x_1,x_2: ax_1\geq \log y, bx_2\geq \log y\}$

（b）$V(y)=\{x_1,x_2: ax_1+bx_2\geq y, x_1\geq 0\}$

（c）$V(y)=\{x_1,x_2: ax_1+\sqrt{x_1x_2}+bx_2\geq y\}$

（d）$V(y)=\{x_1,x_2: ax_1+bx_2\geq y\}$

（e）$V(y)=\{x_1,x_2: x_1(1-y)\geq a, x_2(1-y)\geq b\}$

（f）$V(y)=\{x_1,x_2: ax_1-\sqrt{x_1x_2}+bx_2\geq y\}$

（g）$V(y)=\{x_1,x_2: x_1+\min(x_1,x_2)\geq 3y\}$

1.12 为简单起见，线性齐次生产函数通常用“人均”（per capita）表示，方法很简单：将函

数表达式中的一种生产要素作为公因子提取出来（这种要素通常为劳动，因此说人均）。

例如，考虑函数 $y = f(x_1, x_2)$，其中 $f(x_1, x_2)$ 是一次齐次的。如果将 x_2 提取出来，可以定义如下的新函数 $\phi(X)$：

$$Y = x_2 f((x_1/x_2), 1) \equiv x_2\phi(X) \quad （1）$$

其中，$X \equiv x_1/x_2$，或者，$(Y/x_2) = \phi(X)$ 是以 x_2 表示的人均生产函数（per capita production function）。

（a）在产品 y 的价格为 1 的假设前提下，将利润最大化条件表达为含有 w_1, w_2 和 $\phi(X)$ 的式子。

（b）证明

$$\sigma = \frac{\phi'(X)[\phi(X) - X\phi'(X)]}{X\phi(X)\phi''(X)} \quad （2）$$

（c）证明如果 $f(x_1, x_2)$ 是一个标准的 CES 生产函数，即

$$y = f(x_1, x_2) = (ax_1^\rho + (1-a)x_2^\rho)^{1/\rho}$$

证明(b)中的表达式将为

$$\sigma = 1/(\rho - 1) \quad （2'）$$

1.13 假设有两种产品：投入品 z 和产出品 q。生产函数为 $q = f(z)$。假设生产函数 $f(\cdot)$ 是规模报酬递增的。

（a）假设 $f(\cdot)$ 是可微的，那么 $f(\cdot)$ 规模报酬递增是否意味着平均产量随着投入增加是非递减的？边际产量的情形又是如何的？

（b）假设存在一个代表性消费者，他的效用函数为 $u(q) - z$（负号表示从他那里拿走投入品）。假设 $q = f(z)$ 是个生产计划，它的目标是使代表性消费者的效用最大化。请以数学形式或经济学形式（不需要考虑边界解）说明这个最大化问题的必要条件为边际效用等于边际成本。

（c）在（b）中，边际效用等于边际成本也是该最大化问题的充分条件吗？

习题解答

1.1. 对还是错？若 $V(y)$ 是一个凸集，则相应的生产集 Y 必定为凸。

【知识回顾】生产集为凸意味着必要投入集为凸，但反过来说不成立。

定理(凸生产集蕴涵着凸必要投入集)：若生产集 Y 是一个凸集，则与其相关的必要投入集 $V(y)$，也是一个凸集。

证明。若 Y 是个凸集，这意味着对于任何 x 和 x'，只要 $(y, -x)$ 和 $(y, -x')$ 都在 Y 中，则 $(ty + (1-t)y, -tx - (1-t)x')$ 必然也在 Y 中。这只是要求 $(y, -(tx + (1-t)x'))$ 也在 Y 中。由此可见，若 x 和 x' 都在 $V(y)$ 中，则 $tx + (1-t)x'$ 也在 $V(y)$ 中，这表明 $V(y)$ 是凸集。

【参考答案】这种说法是错误的，因为生产集为凸意味着必要投入集为凸，但反过来说不成立。

我们举一个反例说明。考虑由生产函数 $f(x)=x^2$ 生成的技术。必要投入集 $V(y)=\{x:x\geq\sqrt{y}\}$ 是一个凸集，但生产集 $Y=\{(y,-x):y\leq x^2\}$ 显然不是凸的。

1.2. 当 $a_1\neq a_2$ 时，一般 CES 技术 $y=[a_1x_1^{\rho}+a_2x_2^{\rho}]^{\frac{1}{\rho}}$ 的替代弹性为多大？

【知识回顾】替代弹性的两种计算方法，一是 $\sigma=\dfrac{TRS}{(x_2/x_1)}\dfrac{d(x_2/x_1)}{dTRS}$；二是 $\sigma=\dfrac{d\ln(x_2/x_1)}{d\ln\left|TRS\right|}$。

【参考答案】我们使用第二种方法即 $\sigma=\dfrac{d\ln(x_2/x_1)}{d\ln\left|TRS\right|}$ 计算替代弹性。

$$TRS=-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}=-\frac{a_1x_1^{\rho-1}}{a_2x_2^{\rho-a}}\Rightarrow\ln\left|TRS\right|=\ln(a_1/a_2)+(1-\rho)\ln(x_2/x_1)$$

$$\sigma=\frac{d\ln(x_2/x_1)}{d\ln\left|TRS\right|}=\frac{d\ln(x_2/x_1)}{d[\ln(a_1/a_2)+(1-\rho)\ln(x_2/x_1)]}=\frac{d\ln(x_2/x_1)}{(1-\rho)d\ln(x_2/x_1)}=\frac{1}{1-\rho}.$$

1.3. 定义要素 i 的产出弹性为

$$\varepsilon_i(x)=\frac{\partial f(x)}{\partial x_i}\frac{x_i}{f(x)}.$$

如果 $f(x)=x_1^ax_2^b$，每种要素的产出弹性为多大？

【参考答案】

$f(x)=x_1^ax_2^b\Rightarrow\dfrac{\partial f(x)}{\partial x_1}=ax_1^{a-1}x_2^b$；$\dfrac{\partial f(x)}{\partial x_2}=bx_1^ax_2^{b-1}$

$\varepsilon_1(x)=\dfrac{\partial f(x)}{\partial x_1}\dfrac{x_1}{f(x)}=(ax_1^{a-1}x_2^b)\dfrac{x_1}{x_1^ax_2^b}=a$；类似地可计算出 $\varepsilon_2(x)=b.$

1.4 如果 $\varepsilon(x)$ 是规模弹性，$\varepsilon_i(x)$ 是要素 i 的产出弹性，证明 $\varepsilon(x)=\sum_{i=1}^{n}\varepsilon_i(x)$。

【知识回顾】令 $y=f(x)$ 表示生产函数，t 为一个正的标量。考虑函数 $y(t)=f(tx)$。规模弹性的表达式为 $\varepsilon(x)=\dfrac{dy(t)/y(t)}{dt/t}$。在计算点 x 的规模弹性时，我们必须计算该表达式在 $t=1$ 时的数值。$\varepsilon(x)=\dfrac{dy(t)}{y(t)}\dfrac{t}{y}\Big|_{t=1}=\dfrac{df(tx)}{dt}\dfrac{t}{f(tx)}\Big|_{t=1}$。

【参考答案】

令 $y=f(x)$ 表示生产函数，t 为一个正的标量。考虑函数 $y(t)=f(tx)$。

$$\varepsilon(x)=\frac{dy(t)}{y(t)}\frac{t}{y}\Big|_{t=1}=\frac{df(tx)}{dt}\frac{t}{f(tx)}\Big|_{t=1}=\frac{1}{f(x)}\sum_{i=1}^{n}\frac{\partial f(x)}{\partial x_i}x_i=\sum_{i=1}^{n}\frac{\partial f(x)}{\partial x_i}\frac{x_i}{f(x)}=\sum_{i=1}^{n}\varepsilon_i(x)$$

其中最后一个等式用到了要素 i 的产出弹性的定义 $\varepsilon_i(x)=\frac{\partial f(x)}{\partial x_i}\frac{x_i}{f(x)}$。

1.5CES 技术 $f(x_1,x_2)=(x_1^{\rho}+x_2^{\rho})^{1/\rho}$ 的规模弹性为多大？

【参考答案】

方法一：

$$f(tx_1,tx_2)=[(tx_1)^{\rho}+(tx_2)^{\rho}]^{1/\rho}=t(x_1^{\rho}+x_2^{\rho})^{1/\rho}=tf(x_1,x_2)$$

这表明 CES 生产函数的规模报酬是不变的，因此规模弹性为 1。

方法二：使用规模弹性的定义 $\varepsilon(x)=\frac{dy(t)}{y(t)}\frac{t}{y}\Big|_{t=1}=\frac{df(tx)}{dt}\frac{t}{f(tx)}\Big|_{t=1}$ 计算

$$\frac{df(tx)}{dt}=\frac{d[(tx_1)^{\rho}+(tx_2)^{\rho}]^{1/\rho}}{dt}=\frac{d[tf(x)]}{dt}=\frac{f(x)dt}{dt}=f(x)$$

$$\varepsilon(x)=\frac{dy(t)}{y(t)}\frac{t}{y}\Big|_{t=1}=\frac{df(tx)}{dt}\frac{t}{f(tx)}\Big|_{t=1}=f(x)\frac{1}{f(x)}=1\text{。}$$

1.6 对还是错？ 可微函数 $g(x)$ 为严格递增的当且仅当 $g'(x)>0$。

【参考答案】这种说法是错误的。这是因为，如果 $g'(x)>0$ 则可微函数 $g(x)$ 为严格递增的，但反过来未必成立。例如，$g(x)=x^3$ 是严格递增的，但 $g'(0)=0$。

1.7 教材中断言，如果 $f(x)$ 是位似技术而且 x 和 x' 生产的产量水平相同，则 tx 和 tx' 生产的产量水平也相同。你能严格证明这个结论吗？

【知识回顾】

函数 $g:R\to R$ 称为正单调变换（positive monotonic transformation）若 g 是一个严格递增的函数；也就是说，对于该函数来说 $x>y$ 蕴涵着 $g(x)>g(y)$。位似函数（homothetic function）是一次齐次函数的正单调变换。换句话说，$f(x)$ 是位似的当且仅当它可以写为 $f(x)=g(h(x))$，其中 $h(\cdot)$ 是一次齐次函数，$g(\cdot)$ 是个单调函数。

【参考答案】

令 $f(x)=g(h(x))$ 并假设 $g(h(x))=g(h(x'))$。由于 $g(\cdot)$ 是个单调函数，必有 $h(x)=h(x')$。

由于 $h(\cdot)$ 是一次齐次的即 $h(tx)=th(x)$，从而：$g(h(tx))=g(th(x))$；$g(h(tx'))=g(th(x'))$。

所以有 $g(h(tx))=g(h(tx'))$，这表明 tx 和 tx' 生产的产量水平是相同的。

1.8 令 $f(x_1,x_2)$ 是位似函数。证明它在 (x_1,x_2) 点的技术替代率等于在 (tx_1,tx_2) 点的技术替代率。

【知识回顾】位似函数；技术替代率；

【参考答案】

$f(x)$ 是位似的当且仅当它可以写为 $f(x)=g(h(x))$，其中 $h(\cdot)$ 是一次齐次函数，$g(\cdot)$ 是个单调函数。因此位似函数 $f(x_1,x_2)$ 的技术替代率表达式为

$$-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}=-\frac{\frac{\partial f}{\partial h}\frac{\partial h}{\partial x_1}}{\frac{\partial f}{\partial h}\frac{\partial h}{\partial x_2}}=-\frac{\frac{\partial h}{\partial x_1}}{\frac{\partial h}{\partial x_2}}$$

注意在上式中我们用到了链式法则。从上式可以看出，位似函数 f 的替代率与相应的一次齐次函数 h 的替代率是相等的。而我们已经知道，一次齐次函数在 (x_1,x_2) 点的技术替代率等于在 (tx_1,tx_2) 点的技术替代率是相等的，因此就得到了题目中的结论。

1.9 考虑 CES 技术 $f(x_1,x_2)=[a_1x_1^\rho+a_2x_2^\rho]^{1/\rho}$。证明我们总可以将这个式子写为 $f(x_1,x_2)=A(\rho)[bx_1^\rho+(1-b)x_2^\rho]^{1/\rho}$ 的形式。

【参考答案】

注意，我们可以将 $[a_1x_1^\rho+a_2x_2^\rho]^{1/\rho}$ 写为 $(a_1+a_2)^{1/\rho}[\frac{a_1}{a_1+a_2}x_1^\rho+\frac{a_2}{a_1+a_2}x_2^\rho]^{1/\rho}$。

现在令 $A(\rho)=(a_1+a_2)^{1/\rho}$，令 $b=\frac{a_1}{a_1+a_2}$（从而 $1-b=\frac{a_2}{a_1+a_2}$），这样 $f(x_1,x_2)$ 就可以写为 $A(\rho)[bx_1^\rho+(1-b)x_2^\rho]^{1/\rho}$ 的形式。

1.10 令 Y 为一个生产集。如果 y 和 y' 都在 Y 中蕴涵 $y+y'$ 也在 Y 中，那么我们说技术是可加的。如果 y 在 Y 中蕴涵 ty 也在 Y 中（其中 $0\le t\le 1$），那么我们说该技术是可分的。证明如果某技术既是可加的又是可分的，则 Y 必定是凸的，而且它是规模报酬不变的。

【参考答案】

要证明凸性，我们必须证明：对于 Y 中的所有 y 和 y' 以及所有 $0\le t\le 1$，$ty+(1-t)y'$ 必定在 Y 中。

但可分性意味着 ty 和 $(1-t)y'$ 都在 Y 中，使用可加性可知 $ty+(1-t)y'$ 在 Y 中，这样就证明了凸性。

为了证明规模报酬不变，我们必须证明如果 y 在 Y 中，则 sy（其中 $s>0$）也在 Y 中。

给定任何 $s>0$，令 n 为满足 $n\ge s\ge n-1$ 的非负整数。由可加性可知，ny 在 Y 中；由于 $s/n\le 1$，根据可分性可知，$(s/n)ny=sy$ 在 Y 中。这样就证明了规模报酬不变。

1.11 对于每个必要投入集，确定它是正则的（regular）、单调的和/或凸的。假设参数 a 和 b 以及产量水平都是严格正的。

（a）$V(y)=\{x_1,x_2:ax_1\ge\log y,bx_2\ge\log y\}$

（b）$V(y)=\{x_1,x_2:ax_1+bx_2\geq y,x_1\geq 0\}$

（c）$V(y)=\{x_1,x_2:ax_1+\sqrt{x_1x_2}+bx_2\geq y\}$

（d）$V(y)=\{x_1,x_2:ax_1+bx_2\geq y\}$

（e）$V(y)=\{x_1,x_2:x_1(1-y)\geq a,x_2(1-y)\geq b\}$

（f）$V(y)=\{x_1,x_2:ax_1-\sqrt{x_1x_2}+bx_2\geq y\}$

（g）$V(y)=\{x_1,x_2:x_1+\min(x_1,x_2)\geq 3y\}$

【知识回顾】

单调性：若x在$V(y)$中且$x'\geq x$，则x'也在$V(y)$中。

凸性：若x和x'都在$V(y)$中，则$tx+(1-t)x'$也在$V(y)$中，其中$0\leq t\leq 1$。也就是说$V(y)$是个凸集（convex set）。

正则性：对于所有$y\geq 0$，$V(y)$是闭的非空集合。

【参考答案】

（a）这个集合对于所有$y>0$是闭的和非空的（如果要素投入可为负），因此是正则的。等产量线类似里昂惕夫技术，唯一的区别是我们用$\log y$而不是y衡量单位产量。由此可知该技术是单调的和凸的。

（b）这个必要投入集是非空的但是开的，因此不是正则的；它是单调的和凸的。

（c）该必要投入集为非空且为闭的，因此是正则的。函数$f(x_1,x_2)$的偏导数都是正的，因此，该技术是单调的。为了使等产量线凸向原点，只要生产函数为凹函数即可（注意凹函数是等产量线凸向原点的充分但非必要条件）。为了检验这一点，使用生产函数的二阶导数构建一个矩阵，看看它是不是负半定的。Hessian 矩阵的第一个主子阵的行列式必定为负，第二个主子式的行列式必定是非负的。检验可知，结果的确是这样的，因此题目中的必要投入集为凸的。

（d）是正则的，单调的和凸的；

（e）它是非空的，但它无法生产任何$y>1$的产量。它是单调的和弱凸的。

（f）它是正则的。为了检验单调性，将生产函数写出来$f(x_1,x_2)=ax_1-\sqrt{x_1x_2}+bx_2$，则$\partial f/\partial x_1=a-\frac{1}{2}x_1^{-1/2}x_2^{1/2}$。只有当$a>\frac{1}{2}\sqrt{x_2/x_1}$式，这个偏导数才为正，因此题目中的这个必要投入集并不总是单调的。

现在来看$f(x_1,x_2)$的 Hessian 矩阵，它的行列式为零，第一个主子式为正，因此f不是凹的。但这不能说明必要投入集不是凸的。但是由于f是凸的，因此所有具有下列形式的集合

$$\{x_1,x_2:ax_1-\sqrt{x_1x_2}+bx_2\leq y\}$$

都是凸的。注意到上述凸集除了边界上的点之外，它是我们题目中必要投入集的补集。作为

凸集的补集，题目中的必要投入集不可能是凸的。

（g）这个函数是一个线性函数加上一个里昂惕夫函数，因此它具有这两类函数的性质，当然包括正则性、单调性和凸性。

1.12 为简单起见，线性齐次生产函数通常用“人均”（per capita）表示，方法很简单：将函数表达式中的一种生产要素作为公因子提取出来（这种要素通常为劳动，因此说人均）。

例如，考虑函数 $y = f(x_1, x_2)$，其中 $f(x_1, x_2)$ 是一次齐次的。如果将 x_2 提取出来，可以定义如下的新函数 $\phi(X)$：

$$Y = x_2 f((x_1 / x_2), 1) \equiv x_2 \phi(X) \qquad (1)$$

其中，$X \equiv x_1 / x_2$，或者，$(Y / x_2) = \phi(X)$ 是以 x_2 表示的人均生产函数（per capita production function）。

（a）在产品 y 的价格为 1 的假设前提下，将利润最大化条件表达为含有 w_1, w_2 和 $\phi(X)$ 的式子。

（b）证明

$$\sigma = \frac{\phi'(X)[\phi(X) - X\phi'(X)]}{X\phi(X)\phi''(X)} \qquad (2)$$

（c）证明如果 $f(x_1, x_2)$ 是一个标准的 CES 生产函数，即

$$y = f(x_1, x_2) = (ax_1^{\rho} + (1-a)x_2^{\rho})^{1/\rho}$$

证明(b)中的表达式将为

$$\sigma = 1/(\rho - 1) \qquad (2')$$

【参考答案】

（a）利润最大化问题为 $\max \quad pf(x_1, x_2) - w_1 x_1 - w_2 x_2$， $p = 1$，利润最大化的一阶条件为

$$w_1 = \frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{\partial [x_2 \phi(X)]}{\partial x_1} = x_2 \phi'(X) x_2^{-1} = \phi'(X) \qquad (3)$$

$$w_2 = \frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{\partial [x_2 \phi(X)]}{\partial x_2} = \phi(x) - \phi'(X)(x_1 / x_2) = \phi(X) - X\phi'(x) \qquad (4)$$

（b）鉴于（a）中的 w_1 和 w_2 的表达式形式，先求 σ^{-1} 是方便的

$$\sigma^{-1} = \frac{\partial(w_1/w_2)}{\partial x}\frac{x}{w_1/w_2}$$
$$= \frac{(\phi - X\phi')\phi'' + X\phi'\phi''}{(\phi - X\phi')^2}\frac{X(\phi - X\phi')}{\phi'}$$
$$= \frac{X\phi\phi''}{\phi'(\phi - X\phi')}$$

所以，$\sigma = \dfrac{\phi'(X)[\phi(X) - X\phi'(X)]}{X\phi(X)\phi''(X)}$。

（c）对于标准的 CES 函数

$$\phi(X) = [(aX^{\rho} + (1-a)]^{1/\rho} \tag{6}$$

从（3）和（4）可知

$$w_1 = \phi'(X) = a(\phi(X)/X)^{1-\rho}$$
$$w_2 = \phi(X) - X\phi'(x) = \phi(X) - aX(\phi(X)/X)^{1-\rho}$$

$$\text{因此，}\quad \frac{w_2}{w_1} = (\frac{X^{1-\rho}}{a})\phi^{\rho} - X \tag{7}$$

最后，使用（6）式，可得

$$\frac{w_2}{w_1} = (\frac{X^{1-\rho}}{a})aX^{\rho} + (\frac{X^{1-\rho}}{a})(1-a) - X = (\frac{1-a}{a})X^{1-\rho}$$

而且，由于

$$\phi''(X) = a(\frac{\phi}{X})^{-\rho}(\frac{X\phi' - \phi}{X^2})$$

注意到 $\phi\phi''X = a(\phi/X)^{-\rho}(\dfrac{X\phi' - \phi}{X^2})(1-\rho)$

由（2）式可知，

$$\sigma = \frac{\phi'(\phi) - X\phi'}{\phi\phi''X} = \frac{1}{\rho - 1}$$

1.13 假设有两种产品：投入品 z 和产出品 q。生产函数为 $q = f(z)$。假设生产函数 $f(\cdot)$ 是规模报酬递增的。

（a）假设 $f(\cdot)$ 是可微的，那么 $f(\cdot)$ 规模报酬递增是否意味着平均产量随着投入增加是非递减的？边际产量的情形又是如何的？

（b）假设存在一个代表性消费者，他的效用函数为 $u(q) - z$（负号表示从他那里拿走投入品）。假设 $q = f(z)$ 是个生产计划，它的目标是使代表性消费者的效用最大化。请以数学形

式或经济学形式（不需要考虑边界解）说明这个最大化问题的必要条件为边际效用等于边际成本。

（c）在（b）中，边际效用等于边际成本也是该最大化问题的充分条件吗？

【参考答案】

（a）生产函数 $f(\cdot)$ 规模报酬递增当且仅当 $f(\lambda z)\geq \lambda f(z)$ 对于所有 z 和所有 $\lambda \geq 1$ 成立。因此，如果 $z' \geq z > 0$，那么

$$(1/z')f(z'') = (1/z')f[(z'/z)z] \geq (1/z')(z'/z)f(z) = (1/z)f(z)$$

因此平均产量是非减的。然而，边际产量可能在某个产量区间递减，如图 1 所示。

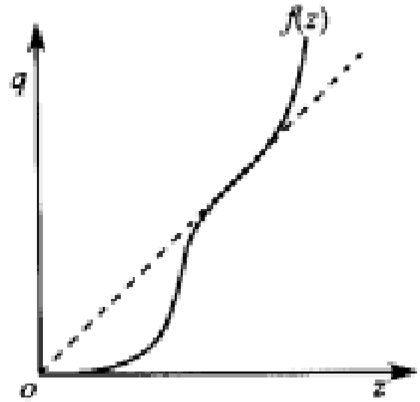


图 1:规模报酬递增不意味着边际产量递增

（b）消费者的最大化问题为

$$\max_{z\geq 0}\ u(f(z)-z$$

一阶必要条件为 $u'[(f(z)]f'(z)=1$，变形可得 $u'[(f(z)]=[f'(z)]^{-1}$。由于成本函数为 $z=f^{-1}(q)$，边际成本等于 $[f'(z)]^{-1}$，因此边际成本等于边际效用是这个最大化问题的必要条件。

（c）不是。即使在某个投入水平上边际成本和边际效用相等，也可能存在着使得消费者效用更高的另外一个投入水平。原因在于当生产函数为规模报酬递增（它产生的可行集是非凸的）时，一阶必要条件不是充分条件。请自行画下图。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 2 章：利润最大化

曹乾　译

(东南大学　caoqianseu@163.com)

2 利润最大化

经济利润的定义为企业的收入减去成本。在计算利润时应确保所有的成本都已被包含进来。如果某人开了个小卖部，而且他也在这个小卖部工作，他的工资应该算作成本。如果一帮人借钱给企业使用，企业承诺按月偿还，则支付的利息应视作生产的成本。

企业的收入和成本都取决于企业自身的行为。这些行为有多种形式：实际生产活动；购买生产要素；购买广告等。在比较抽象的水平，我们可认为企业参与种类繁多的活动。我们将收入写为 n 种活动水平的函数，即 $R(a_1,...,a_n)$，类似地，成本函数为 $C(a_1,...,a_n)$。

对企业行为的大多数经济分析都要做一个基本假设，即假设企业的目标是利润最大化；也即是说，企业选择行为 $(a_1,...,a_n)$ 使得 $R(a_1,...,a_n)-C(a_1,...,a_n)$ 最大。这个行为假设将贯穿本书。

即使在如此广泛的意义上，利润最大化仍有两个基本原理。第一条原理可运用微积分直接得出，企业的利润最大化问题可以写为

$$\max_{a_1,...,a_n} R(a_1,...,a_n)-C(a_1,...,a_n)$$

简单运用一下微积分可知，最优活动集 $\mathbf{a}^*=(a_1^*,...,a_n^*)$ 由下列条件决定

$$\frac{\partial R(\mathbf{a}^*)}{\partial a_i}=\frac{\partial C(\mathbf{a}^*)}{\partial a_i} \qquad i=1,...,n.$$

这些条件符合直觉：若边际收入大于边际成本，则应该增加活动水平；若边际收入小于边际成本，则应该降低活动水平。

这个刻画利润最大化的基本条件有几种具体的解释。例如，企业的一种决策是选择产出水平。这个条件告诉我们，企业选择的产量应使得再额外多生产一单位产量的边际收入和边际生产成本相等。企业的另外一种决策是决定具体要素的投入量，比如劳动雇佣量。这个条件告诉我们，企业选择雇佣的劳动量应使得再额外雇佣一单位劳动的边际产出等于雇佣这个额外一单位劳动的边际成本。

利润最大化的第二个基本条件是，长期利润相等的条件。假设两家企业的收入函数和成本函数都相同。那么显然在长期这两家企业的利润应该相等，因为每家企业可以模仿另外一家企业的行为。这个条件非常简单，但是它的应用意义却非常强大。

为了更具体地应用这些条件，我们需要将收入函数和成本函数分解为更基本的部分。收入由两部分组成：企业每种产品的销量乘以相应产品的价格。成本也由两部分组成：企业每种要素的使用量乘以相应要素的价格。

企业利润最大化问题因此可以分解为以下两个问题：一是价格决定问题，即如何决定产品的价格和要素的价格；二是数量决定问题，即产量和要素使用量的决策。当然，它不能

单方面地设定价格以及设定活动水平。在决定最优决策时，企业受到两类约束：技术约束和市场约束。

技术约束（technological constraints）是和生产方案可行性有关的约束。我们在上一章已经介绍了描述技术约束的方法。另外，企业在市场中会受到其他行为人的活动的影响，由此产生的对该企业的约束称为市场约束（market constraints）。例如，购买该企业产品的消费者对某既定数量的产品只愿意支付一定的价格；类似地，企业的供应商对生产要素的价格也会提出要求，若价格过低，他们不会供给生产要素。

当企业决定它的最优行为时，它必须考虑这两类约束。然而，最好将这两类约束分别研究。因此，在下面几节研究的企业，我们假设企业将表现出最简单的市场行为，即价格接受者的行为。我们假设每个企业都把价格视为它的利润最大化问题的外生的、给定的变量。因此，企业只需要关心决定利润最大化的产出和投入水平。这样的价格接受企业通常称为竞争性企业。

为什么把这样的企业称为竞争性企业？我们稍后再来分析这个问题。然而，此处我们可以简要指出在哪些情形下价格接受行为才是合理的模型。假设消费者都是信息灵通的（well-informed），再假设很多这样的消费者购买某种同质产品，而且这种产品的生产企业的数量也很多。现在你就清楚了：所有企业对这种产品必须索要相同的价格——索要价格高于现行市价的任何企业将会立即失去所有的消费者。因此，每个企业在制定它的最优决策时，必须认为市场价格是外生给定的。我们在本章将分析，在市场价格给定的情形下，企业对生产方案的最优选择决策。

2.1 利润最大化

假设企业在产品市场和它使用要素的市场都是价格接受者。令 $\mathbf{p}$ 表示企业的投入和产出的价格向量[一]。企业的利润最大化问题可以表示为

$$\pi(\mathbf{p}) = \max \ \mathbf{py}$$

$$\text{使得 } \mathbf{y} \text{ 在 } Y \text{ 中}$$

由于产出是用正数衡量，投入是用负数衡量，该问题的目标函数就是利润：收入减去成本。函数 $\pi(\mathbf{p})$ 将最大利润作为价格向量的函数，这个函数称为企业的利润函数。

利润函数有几种有用的变形(variants)。例如，如果我们分析的是短期最大化问题，我们可定义短期利润函数，有时也称为受约束的利润函数（restricted profit function）：

$$\pi(\mathbf{p}, \mathbf{z}) = \max \ \mathbf{py}$$

[一] 通常，我们将价格向量视为行向量，而把产品或要素数量向量视为列向量。

$$\text{使得}\,\mathbf{y}\,\text{在}\,Y(\mathbf{z})\,\text{中}$$

如果企业只生产一种产品，利润函数可以写为

$$\pi(p,\mathbf{w}) = \max \quad pf(\mathbf{x}) - \mathbf{wx}$$

其中 p 现在为产品的（标量）价格，$\mathbf{w}$ 为要素的价格向量，投入用（非负）向量 $\mathbf{x} = (x_1,...,x_n)$ 衡量。在这种情形下，我们可以定义受约束利润函数的一种变体，即成本函数

$$c(w,y) = \min \ wx$$

$$\text{使得}\,x\,\text{在}\,V(y)\,\text{中}.$$

在短期，我们可能想要分析受限或者短期成本函数

$$c(w,y,z) = \min \ wx$$

$$\text{使得}\,(y,-x)\,\text{在}\,Y(z)\,\text{中}.$$

成本函数表示要素价格为 w 时，生产 y 单位产品的最小成本。由于在这个问题中，只有要素价格被视为外生变量，因此我们可用成本函数分析下列这种企业的行为：企业在要素市场为价格接受者，但在产品市场不是价格接受者。我们在分析垄断行为时将用到这个工具。

利润最大化行为的特征可用微积分进行刻画。例如，企业产品只有一种时，它的利润最大化问题为

$$p\frac{\partial f(x^*)}{\partial x_i} = w_i \quad i = 1,...,n.$$

这个条件是说每种要素的边际产量的价值必定等于该种要素的价格。使用向量符号可以把上述条件写得更紧凑些

$$pDf(x^*) = w.$$

其中

$$Df(x^*) = (\frac{\partial f(x^*)}{\partial x_1},...,\frac{\partial f(x^*)}{\partial x_n})$$

是 f 的梯度（gradient）：f 对它的各个变量的一阶偏导数。

一阶条件表明“每种要素的边际产量的价值必定等于该要素的价格。”这只是我们前面说过的最优法则的一种特殊情形：每种行为的边际收入必定等于它的边际成本。

这个一阶条件也可以同图形表示。考虑图 2.1 画出的生产可能集。在这个二维情形下，利润表达式为 $\prod = py - wx$。若 p 和 w 固定不变，则该函数的水平集（level sets）是一条直线，该直线的表达式为 $y = \prod / p + (w/p)x$。这条直线称为等利润线(isoprofit line)。该等利润线的斜率为以单位产品价格衡量的工资；纵截距为以单位产品价格衡量的利润。

追求利润最大化的企业希望在生产集中找到一点，使得该点的利润水平最大——这一点所在的等利润线的纵截距最大。由图可知，这样的最优点可用相切条件刻画：

$$\frac{df(x^*)}{dx} = \frac{w}{p}.$$

在二维情形下，容易看出利润最大化的二阶条件，也就是说生产函数对投入的二阶导数必定是非正的：

$$\frac{d^2 f(x^*)}{dx^2} \le 0.$$

在图形上，这表示在利润最大化的那一点 x^*，生产函数必定位于它的切线以下；也就是说，它必定是“局部凹的”（locally concave）。我们通常假设二阶导数严格为负。

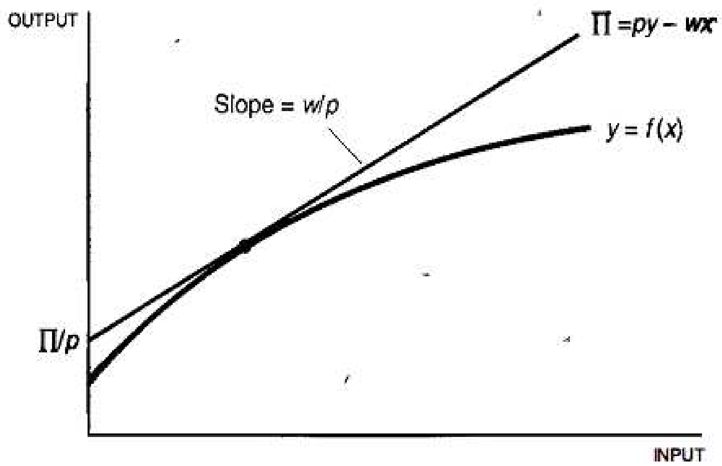


图 2.1：**利润最大化**。利润最大化的投入量发生在等利润线的斜率等于生产函数的斜率之处。

类似的二阶条件在多投入的情形下也是成立的。在这种情形下，利润最大化的二阶条件是生产函数二阶导数的矩阵在最优点必定为负半定（negative semidefinite）；也就是说，二阶条件要求海赛矩阵（Hessian matrix）

$$D^2 f(x^*) = (\frac{\partial^2 f(x^*)}{\partial x_i \partial x_j})$$

必须满足下列条件：对所有向量 h 都有 $hD^2 f(x^*)h^t \le 0$。（上标 t 表示矩阵的转置。）注意，若投入只有一种，则海赛矩阵是个标量，这个条件简化为我们前面说过的单一投入情形下的二阶条件。

在图形上，要求海赛矩阵为负半定意味着生产函数在最优点的领域内必定为局部凹的，也就是说，生产函数在必定位于它的切超平面（tangent hyperplane）之下。

在很多应用中，我们关心的是正常的最大化（a regular maximum），因此需要检验的条件为海赛矩阵是否为负半定。在第 26 章，我们指出检验海赛矩阵为负半定的必要和充分条件是海赛矩阵的前主子式（leading principal minors）的正负号必须为交替改变。这个代数条件在检验二阶条件时有时比较有用，下面我们将看到这一点。

2.麻烦

对于每个价格向量(p,w)，通常会存在要素的最优选择x^*。要素最优选择是价格向量的函数，这个函数称为企业的要素需求函数。我们将该函数记为$x(p,w)$。类似地，函数$y(p,w)=f(x(p,w))$称为企业的供给函数。我们通常假设这些函数是定义清晰和良好性状的（well-behaved），但是有必要考虑它们不满足这两个条件时的问题。

首先，生产技术有可能不能用可微分的生产函数表示，因此就不能再使用导数工具。里昂惕夫技术就是这样的例子。

其次，上面推导出的微积分条件，只有当选择变量在最优选择的一个开邻域内变动才合理。在很多经济问题中，变量自然为非负的；如果某些变量在最优选择处的值为零，上述微积分条件可能是不合适的。上述微积分条件只对内部解（interior solutions）有效，内部解是指每种要素的使用量是正的。

如果想要最优解包含边界解（boundary solutions），那么必须对上述微积分条件进行修改，但是这并不难做到。例如，如果我们在利润最大化问题中限定 x 为非负，则可以证明，相关的一阶条件为

$$p\frac{\partial f(x)}{\partial x_i}-w_i\le 0 \qquad \text{若 } x_i=0$$

$$p\frac{\partial f(x)}{\partial x_i}-w_i=0 \qquad \text{若 } x_i>0.$$

因此，增加x_i使用量带来的边际利润必为非正，否则企业就会增加x_i的使用量。若$x_i=0$，则增加x_i使用量带来的边际利润可能为负——这就是说，企业会减少x_i的使用量。但是由于x_i已经为零，继续减少显然是不可能的。最后，若$x_i>0$，即非负约束这个条件没有起到约束作用，这种情形就是通常的内部解的条件。

涉及非负约束或其他类型的不等式约束的最优化问题，可用库恩-塔克(Kuhn-Tucker)定理求解。请参考第 27 章。在成本最小化那一章，我们将举例说明如何运用这一定理。

第三个问题是可能不存在利润最大化的生产方案。例如，考虑生产函数为$f(x)=x$的情形，因此一单位x可以生产一单位产出。不难看出，对于$p>w$，不存在利润最大化的方案。当$p>w$时，如果你想要最大化$px-wx$，你会选择无限大的x。只有当$p\le w$时，利润最大化的生产方案才会存在，在这种情形下，最大化利润为零。

事实上，任何规模报酬不变的技术都存在与上述相同的现象。为了证明这一点，假设我们找到了某个(p,w)，其中最大利润为严格正：

$$pf(x^*)-wx^*=x^*>0.$$

假设我们按比例增加产出，乘子为$t>1$；我们的利润为

$$pf(tx^*)-wtx^*=t[pf(x^*)-wx^*]=tx^*>x^*.$$

这意味着，如果利润为正，那么利润可以做得更大。因此，利润是无边界的，在这种情形下不存在利润最大化的生产方案。

从这个例子明显可以看出，规模报酬不变企业唯一的非平凡（nontrivial）利润最大化之处是零利润之处。如果企业的产量为正但利润为零，那么它在不同产量之间是无差异的。

这带来了第四个难题：即使存在利润最大化的方案，这样的方案也不是唯一的。如果对于规模报酬不变的技术，(y,x)产生的最大利润为零，那么(ty,tx)也会产生零利润，因此也是利润最大化的。在规模报酬不变的情形下，如果对于(p,w)存在利润最大化的选择，则通常存在着一系列能使利润最大化的生产方案。

例子：柯布-道格拉斯技术的利润函数

考虑$f(x)=x^a$，其中$a>0$类型的生产函数的利润最大化问题。一阶条件为

$$pax^{a-1}=w,$$

二阶条件可以化简为

$$pa(a-1)x^{a-2}\le 0.$$

二阶条件只有当$a\le 1$时才能满足，这意味着为使竞争利润最大化有意义，生产函数必定有规模报酬不变或递减的时候。

如果$a=1$，一阶条件简化为$p=w$，因此当$w=p$时，x的任何值都是利润最大化的选择。当$a<1$时，我们使用一阶条件求解要素需求函数

$$x(p,w)=\left(\frac{w}{ap}\right)^{\frac{1}{a-1}}.$$

供给函数为

$$y(p,w)=f(x(p,w))=\left(\frac{w}{ap}\right)^{\frac{a}{a-1}}.$$

利润函数为

$$\pi(p,w)=py(p,w)-wx(p,w)=w\left(\frac{1-a}{a}\right)\left(\frac{w}{ap}\right)^{\frac{1}{a-1}}.$$

2.3 需求函数和供给函数的性质

将投入和产出的最优选择作为价格的函数，分别称为要素需求函数和产品供给函数。这些函数是既定形式最大化问题（利润最大化问题）的解的事实，意味着需求函数和供给函数的行为受到某些限制。

例如，容易看出，如果将所有价格同乘以一个正数t，使利润最大化的要素投入向量不会变动。（你能严格证明这个结论吗？）因此，要素需求函数$x_i(p,w)$，其中$i=1,...,n$，必然满足下列限制

$$x_i(tp,tw)=x_i(p,w).$$

换句话说，要素需求函数必定是零次齐次函数。这个性质是利润最大化行为的一个重要应用：某个行为是否来自利润最大化行为的快速检查方法，是判断需求函数是否为零次齐次的。如果不是，那么该企业不可能是在最大化自己的利润。

我们还想找出需求函数的其他类似限制。事实上，我们想找到这类限制的一个完整名单。这样的名单有两个用途。首先，我们可以将它用于分析关于利润最大化企业对经济环境变化的反应的理论论断。例如，有这样一个论断："如果所有价格都变为原来的二倍，追求利润最大化企业的要素需求和商品供给水平都不会变动。"根据前面的分析可知，这是正确的。其次，我们可用这些限制实证地分析企业的行为是否符合利润最大化模型。如果我们观测到某个企业在所有价格变为原来的二倍时，它的需求和供给改变但其他事项都未变，我们可以推断（可能不情愿地）该企业不是一个利润最大化者。

因此，无论是出于理论还是实证角度的考虑，都有必要确定需求函数和供给函数的性质。我们使用三种方法解决这个问题。第一种方法是分析刻画最优选择的一阶条件。第二种方法是直接分析需求函数和供给函数的最大化的性质。第三种方法是分析利润函数和成本函数的性质并且将这些性质和需求函数关联起来。这种方法有时称为"对偶方法"（dual approach）。上述每一种分析最优化行为的方法对于经济学中其他类型的问题的分析都是有益的，相关的操作方法需要认真学习。

经济学家将一个经济变量如何对经济环境变化作出反应的这种分析，称为比较静态分析（comparative statics）。例如，我们可以提问：追求利润最大化的企业的产量供给如何随着产品价格的变化而变化的？这是供给函数静态分析的一部分内容。

术语 "比较"是指比较经济环境变化"之前"和"之后"的情形。术语"静态"是指在所有的调整都进行完之后才进行比较，也就是说，我们是将一种均衡状态和另外一个均衡状态进行比较。

其实，术语“比较静态”的说法不够好，似乎只有经济学家这么说。这类分析的一个更好的称呼是**敏感性分析**（sensitivity analysis）。这种叫法的额外好处在于，这个术语在其他研究领域也实用。然而，由于比较静态的叫法在经济学中已称为一个传统，在经济分析中已根深蒂固，任何试图改变这种称呼的做法可能都是徒劳的。

2.4 使用一阶条件进行比较静态分析

我们首先分析一种简单的情形：追求利润最大化的企业使用一种要素生产一种产品。该企业面对的最大化问题为

$$\max_{x} \; pf(x) - wx.$$

如果 $f(x)$ 是可微的，需求函数 $x(p,w)$ 必定满足一阶条件和二阶条件

$$\begin{aligned} pf'(x(p,w)) - w &\equiv 0 \\ pf''(x(p,w)) &\leq 0. \end{aligned}$$

注意，这些条件关于 p 和 w 恒成立。根据定义可知 $x(p,w)$ 是在企业在 (p,w) 时的利润最大化选择，因此 $x(p,w)$ 必定满足利润最大化的一阶条件。由于一阶条件是个恒等式，我们可以将其对 w 微分，从而得到

$$pf''(x(p,w))\frac{dx(p,w)}{dw} - 1 \equiv 0.$$

假设这个最大化问题是正则的（regular），因此 $f''(x)$ 不等于零，将上式两侧同除以 $f''(x)$ 可得

$$\frac{dx(p,w)}{dw} \equiv \frac{1}{pf''(x(p,w))}. \tag{2.1}$$

这个式子告诉我们，要素需求函数 $x(p,w)$ 对 w 的变动是如何反应的。首先，它明确地将 dx/dw 表达为生产函数的函数。如果生产函数在最优点的领域内非常弯曲，从而二阶导数的数值比较大，那么要素需求随要素价格变动而变动的程度就较小。（你可以模仿图 2.1 画图验证一下这个结论。）

其次，它让我们知道导数的符号：由于最大化问题的二阶条件意味着，生产函数的二阶导数 $f''(x(p,w))$ 是负的，（2.1）式意味着 $dx(p,w)/dw$ 也为负。换句话说：要素需求曲线向下倾斜。

对一阶条件求微分的程序同样适用于当投入要素有多种的情形。为简单起见，我们以两种要素为例进行分析。为了简化符号，我们将产品价格标准化为 1 即 $p=1$，重点分析要素需求行为如何随要素价格的变动而变动。要素需求函数必定满足一阶条件

$$\frac{\partial f(x_1(w_1,w_2),x_2(w_1,w_2))}{\partial x_1}\equiv w_1$$
$$\frac{\partial f(x_1(w_1,w_2),x_2(w_1,w_2))}{\partial x_2}\equiv w_2.$$

上面两个式子对 w_1 求导可得

$$f_{11}\frac{\partial x_1}{\partial w_1}+f_{12}\frac{\partial x_2}{\partial w_1}=1$$
$$f_{21}\frac{\partial x_1}{\partial w_1}+f_{22}\frac{\partial x_2}{\partial w_1}=0.$$

类似地，对对 w_1 求导可得

$$f_{11}\frac{\partial x_1}{\partial w_2}+f_{12}\frac{\partial x_2}{\partial w_2}=0$$
$$f_{21}\frac{\partial x_1}{\partial w_2}+f_{22}\frac{\partial x_2}{\partial w_2}=1.$$

将上面四个式子用矩阵的形式表示可得

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}\begin{pmatrix} \dfrac{\partial x_1}{\partial w_1} & \dfrac{\partial x_1}{\partial w_2} \\ \dfrac{\partial x_2}{\partial w_2} & \dfrac{\partial x_2}{\partial w_2} \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

假设这个最大化问题的解是正则的。这意味着 Hessian 矩阵为严格负定的，因此是非奇异的（nonsingular)。(这个假设类似于一维情形下的假设 $f''(x)<0$ 。) 求解上述矩阵表达式中的一阶导数，可得

$$\begin{pmatrix} \dfrac{\partial x_1}{\partial w_1} & \dfrac{\partial x_1}{\partial w_2} \\ \dfrac{\partial x_2}{\partial w_2} & \dfrac{\partial x_2}{\partial w_2} \end{pmatrix}=\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}^{-1}.$$

上式左边的矩阵称为替代矩阵（substitution matrix)，这是因为它刻画了当要素价格变动时，企业如何使用一种要素替代另一种要素的。根据我们的计算可知，替代矩阵只是 Hessian 矩阵的逆阵。这个结论有几个重要的应用。

我们已经知道，(严格）利润最大化的二阶条件是 Hessian 矩阵是一个对称的负定矩阵。从线性代数的知识可知，一个对称的负定矩阵的逆矩阵仍然是一个对称的负定矩阵。这表示替代矩阵本身必定是一个对称的负定矩阵。特别地：

1）$\partial x_i/\partial w_i<0$ ，其中 $i=1,2$ ，这是因为一个负定矩阵的主对角线上的元素必定是负的；

2）$\partial x_i / \partial w_j = \partial x_j / \partial w_i$，这是由于矩阵是对称的。

尽管直觉上要素需求曲线的斜率应为负，但是替代矩阵是对称这一事实却不那么明显。为什么当要素 j 的价格变动时企业对要素 i 的需求，必定等于当要素 i 的价格变动时企业对要素 j 的需求？没有明显的原因…但利润最大化行为暗示这一点必须成立。

类似的计算适用于投入要素为任意多种的情形。将产品价格标准化为 1 即 $p=1$，利润最大化的一阶条件为

$$Df(x(w)) - w \equiv 0.$$

如果对 w 求导，可得

$$D^2 f(x(w))Dx(w) - 1 \equiv 0.$$

从上式求解替代矩阵，可得

$$Dx(w) \equiv [D^2 f(x(w))]^{-1}.$$

由于 $D^2 f(x(w))$ 是一个对称的负定矩阵，替代矩阵 $D^2 x(w)$ 也是一个对称的负定矩阵。这个表达式显然类似于前文描述的一种要素和两种要素的情形。

如果说替代矩阵是负半定的，它的实证内容是什么？我们提供以下解释。假设要素价格矩阵从 w 变为 $w + dw$，则要素需求相应的变动为

$$dx = Dx(w)dw^t.$$

上式两侧同乘以 dw 可得

$$dwdx = dwDx(w)dw^t \leq 0.$$

从负半定矩阵的定义可知，上式的符号为非正。我们看到，替代矩阵的负半定性意味着要素价格变动和要素需求变动的内积必定是非正的，至少对于要素价格的微小变化是这样的。例如，如果第 i 种要素的价格上升，其他价格都不变，则该种要素的需求必定下降。一般来说，需求量的变动 dx 与价格变动 dw 的夹角是钝角。大致来说，需求量变动的方向多少和价格变动的方向“相反”。

2.5 使用代数进行比较静态分析

在这一节我们直接使用利润最大化的定义分析最大化行为的结果。我们使用的分析环境和上一节稍微有些不同。我们不再使用需求函数和供给函数分析企业的行为，而是假设能得到企业行为的一些观察值，根据这些观察值进行分析。这种方法能避免微分的冗长计算，而且这种方法更符合实际。（谁也不曾得到企业行为的无限数据，因此微分方法是有缺陷的。）

因此，假设我们观测到企业的一组价格向量 p^t 和相应的净产出向量 y^t，其中 $t = 1,...,T$. 我们将这个集合称为**数据**（data）。如果使用前面介绍的净供给函数进行表达，数据就是

$(p^t, y(p^t))$，其中 $t = 1,...,T$.

我们问的第一个问题是，利润最大化模型能传达出关于数据集的什么信息。如果企业是追求利润最大化的，那么价格为 p^t 时我们观测到的企业对净产出的选择产生的利润，必定大于或等于其他任何选择产生的利润。我们不知道这种情形下企业的**所有**其他可行选择，但是我们的确知道一些可行选择，也就是说我们可观测到企业的其他选择 y^s，其中 $s = 1,...,T$.因此，利润最大的**必要**条件为

$$\text{对于所有 } t \text{ 和 } s = 1,...,T \text{，都有 } p^t y^t \geq p^t y^s.$$

我们将这个条件称为利润最大化弱公理（Weak Axiom of Profit Maximization, WAPM）。

在图 2.2A 我们画出了违背利润最大化弱公理的两个观察值，而在图 2.2B 则画出了符合该弱公理的情形。

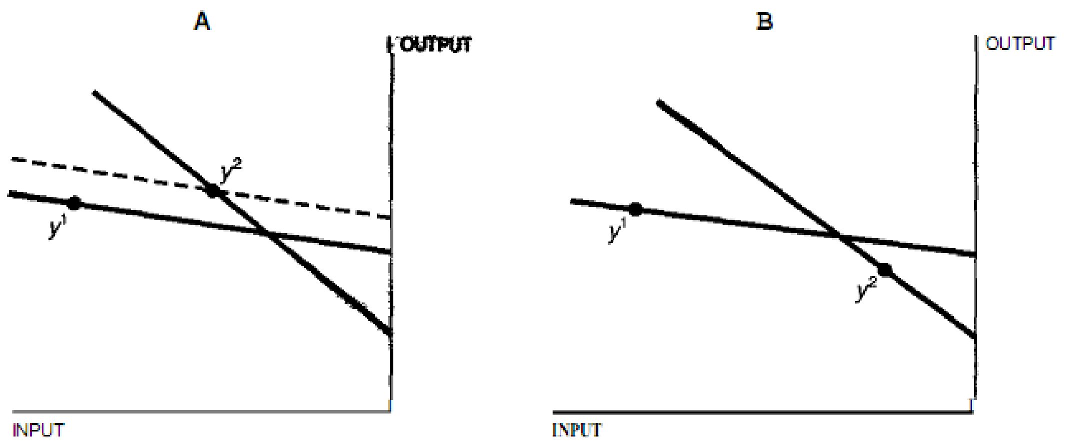


图 2.2：利润最大化弱公理。A 图违背了弱公理，因为 $p^1 y^2 \geq p^1 y^1$。B 图符合弱公理。

利润最大化弱公理是一个简单但非常有用的条件；下面我们推导出一些结果。固定两个观察期 t 和 s，对每一期应用利润最大化弱公理，可得

$$p^t (y^t - y^s) \geq 0$$

$$-p^s (y^t - y^s) \geq 0$$

将这两个不等式相加可得

$$(p^t - p^s)(y^t - y^s) \geq 0.$$

令 $\Delta p = (p^t - p^s)$ 以及 $\Delta y = (y^t - y^s)$，上式可以简化为

$$\Delta p \Delta y \geq 0. \qquad (2.2)$$

也就是说，**价格向量改变与相应的净产出改变的内积必定是非负的**。

例如，如果 Δp 是向量 $(1,0,...,0)$，则（2.2）式意味着 Δy_1 必然是非负的。如果第一种商品对于企业来说是个产出品，而且因此它的数量为正，则当该商品价格上升时，该商品的供

给量不可能下降。另一方面，如果第一种商品对于企业来说是投入品，它的数量因此为负，那么当该商品的价格上升时，该商品的需求量不可能增加。

当然，（2.2）式只是上一节推导出的 $dpdy \geq 0$ 的“德尔塔（Δ）版本。但是这个版本的应用性更强，因为它可以应用于所有价格变动，而不只是价格的微小变动。注意，（2.2）式是直接从利润最大化的定义推导出的，不需要假设生产技术是正则的。

2.6 可还原性

利润最大化弱公理穷尽了利润最大化的所有意义吗？或者说，利润最大化还蕴涵着一些其他有用的条件？回答这个问题的一种方法是试图构建一种生产技术，使得该技术能够使得观察到的企业行为 (p^t, y^t) 是利润最大化行为。如果我们能对满足利润最大化弱公理的任何数据集找到这样的一种技术，那么利润最大化弱公理的确穷尽了利润最大化行为的所有意义。我们将构建与可观测到的选择相一致的生产技术的这种做法称为还原。

我们将证明，如果一组数据满足利润最大化弱公理，那么总是可以找到某种生产技术使得可观测到的选择是利润最大化的选择。事实上，总能找到某个闭且凸的生产集 Y。本节剩下的部分将介绍这个论断的大致证明过程。

我们的任务是构建一个生产集，使得这个生产集产生的可观测到的选择 (p^t, y^t) 为利润最大化的选择。事实上我们将构建两个这样的生产集，一个充当实际技术的“内部边界”，另外一个作为实际技术的“外部边界”。我们先构建内部边界。

假设真正的生产集 Y 为凸且单调的。由于 Y 必然包含 y^t，其中 $t = 1,...,T$，自然可以将内部边界取为包含 $y^1,...,y^T$ 的最小的凸且单调的集合。这个集合称为点 $y^1,...,y^T$ 的凸单调包，我们将其表示为

$$YI = \{y^t : t = 1,...,T\} \text{ 的凸单调包}$$

集 YI 如图 2.3A 所示。

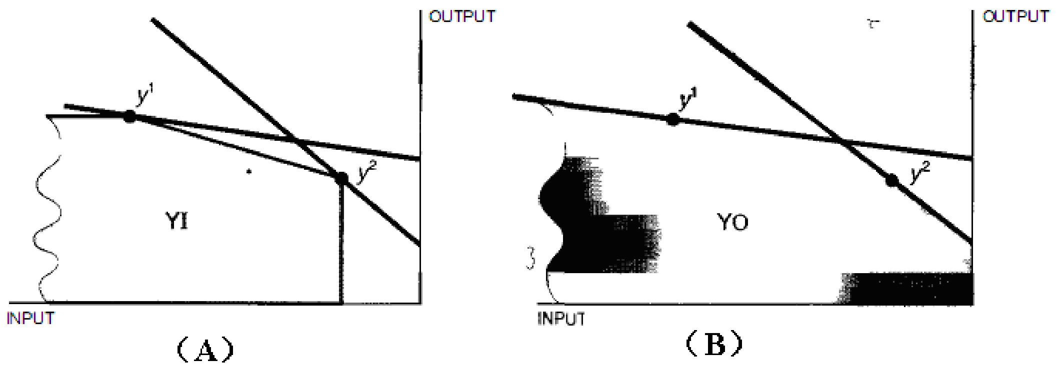


图 2.3：集 YI 和集 YO。集 YI 是可能成为与数据相符的生产集的最小凸单调包。集 YO 是可能成为与数据相符的生产集的最大凸单调包。

容易证明对于技术YI， y^t是价格为p^t时的利润最大化的选择。我们所做的工作就是检验对于所有t，

$$p^t y^t \geq p^t y \text{ 对于 } YI \text{ 中的所有 } y \text{ 都成立.}$$

假设事实并非如此。那么对于某个观测t，YI中存在某个y满足$p^t y^t < p^t y$。但是仔细看下图你就知道必定存在某个观测s使得$p^t y^t < p^t y^s$。但这个不等式违背了利润最大化弱公理。

因此集YI合理解释了（rationalize）观测到的行为，因为它是产生这种行为的一种可能技术。不难看出YI必然包含于产生该观测行为的任何凸技术之内：如果Y产生了上述观测到的行为而且还是凸的，那么它必定包含观测到的选择y^t，这些点的凸包是最小的这样的集合。在这种意义上，YI该处了能产生这些观测到选择的实际技术的"内部边界"。

我们自然会问能否找到这个"实际"技术的外部边界？也就是说，我们能否找到某个集合YO，使得这个集合能包含与观测到的数据相一致的任何技术？

回答这个问题的技巧是将不可能在实际技术中的所有点排除，然后将剩下的一切作为这个集合。更准确地说，我们把$NOTY$（即不是 Y）定义为

$$NOTY = \{y : p^t y > p^t y^t \text{ 对于某个 } t\}.$$

$NOTY$包含所产生的利润高于观测到选择的所有净产出束。如果企业是追求利润最大化的，这样的产出束在技术上是不可行的，否则企业必定已经选择了这些产出束。因此，我们取$NOTY$集的补集就得到了 Y 的外部边界。

$$YO = \{y : p^t y \leq p^t y^t \text{ 对于所有 } t = 1,...,T\}.$$

集 YO 如图 2.3B 所示。

为了证明 YO 合理解释了观测到的行为，我们必须证明这些观测到选择的利润至少和 YO 中的任何其他y一样大。假设不是。那么存在某个y^t使得$p^t y^t < p^t y$对于 YO 中的某个y成立。但是这违背了 YO 的上述定义。

从我们对 YO 的构建方法可知，它必定包含于数据(y^t)相一致的任何生产集。因此，YO 和 YI 构成了产生这些数据的真正生产集的最紧贴的内部和外部边界。

注释

如果你想学习更多的比较静态方法，请参考 Silberberg（1974）和 Silberberg（1990）。本章的代数方法是受 Afriat（1967）和 Samuelson（1947）；更进一步的探讨请参阅 Varian（1982b)。

习题

2.1 使用 Kuhn-Tucker 定理推导利润最大化条件和成本最小化的条件，要求这些条件对边界解（即某要素使用量为零）也适用。

2.2 证明对于规模报酬递增的技术如果存在产生正利润的某个点，那么利润最大化的点通常是不存在的。

2.3 计算技术 $y = x^a$（其中 $0 < a < 1$）的利润函数，验证它关于 (p, w) 是齐次的和凸的。

2.4 令 $f(w_1, w_2)$ 表示含有两种要素的生产函数，w_1, w_2 分别为这两种要素的价格。证明要素份额 $(w_2 x_2 / w_1 x_1)$ 关于 (x_1 / x_2) 的弹性为 $1/\sigma - 1$。

2.5 证明要素份额关于 (w_2 / w_1) 的弹性为 $1-a$.

2.6 令 (p^t / y^t) 表示满足利润最大弱公理的一组观察到的选择，其中 $t = 1,\ldots,T$，令 YI 和 YO 分别为真实生产集 Y 的内部边界和外部边界。令 $\pi^+(p)$ 是与 YO 相伴的利润函数，$\pi^-(p)$ 是与 YI 相伴的利润函数，$\pi(p)$ 是和 Y 相伴的利润函数。证明对于所有的 p，都有 $\pi^+(p) \ge \pi(p) \ge \pi^{-1}(p)$。

2.7 生产函数 $f(x) = 20x - x^2$，产品价格已标准化为 1。令 w 表示要素 x 的价格。$x \ge 0$。

（a）如果 $x > 0$，利润最大化的一阶条件是什么？

（b）w 取何值时，x 的最优数量为零？

（c）w 取何值时，x 的最优数量为 10？

（d）求要素需求函数。

（e）求利润函数。

（f）求利润函数关于 w 的导数。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 3 章：利润函数

曹乾　译

(东南大学　caoqianseu@163.com)

3 利润函数

给定任何生产集 Y，我们已经知道如何计算利润函数$\pi(p)$，该利润函数告诉我们企业在价格 p 时能达到的最大利润。从利润函数的定义可以直接推导出它的若干重要的性质。这些性质对于分析利润最大化行为非常有用。

我们已经知道，利润函数是净产出价格向量的函数，在价格一定的情形下，它是企业能够实现的最大利润：

$$\pi(p) = \max_y \; py$$

从数学推导的角度来说，重要的是该问题中的目标函数是价格的线性函数。

3.1 利润函数的性质

我们开始分析利润函数的性质。要记住这些性质仅是通过利润最大化假设就能推导出的。不需要凸性、单调性或其他类型的正规性假设。

利润函数的性质

1）**对产出价格的非递减性以及对投入价格的非递增性**。如果对于所有产出都有 $p_i' \geq p_i$ 并且对于所有投入都有 $p_j' \leq p_j$，则$\pi(p') \geq \pi(p)$.

2）**对于价格 p 是一次齐次的**。对于所有$t > 0$，都有$\pi(tp) = t\pi(p)$.

3）**对于价格 p 是凸的**。令 $p'' = tp + (1-t)p'$，其中$0 \leq t \leq 1$。

则$\pi(p'') \leq t\pi(p) + (1-t)\pi(p')$.

4）**对于价格 p 是连续的**。函数$\pi(p)$是连续的，至少当$\pi(p)$为良好定义以及 $p_i > 0$（其中$i = 1,...,n$）时是这样的。

证明。我们再次强调，这些性质直接可以从利润最大化函数的定义推导出，而不需要借助生产技术的任何其他假设。

1）令 y 是当价格为 p 时的企业利润最大化的净产出向量，因此$\pi(p) = py$；令 y' 是当价格为 p' 时的企业利润最大化的净产出向量，因此$\pi(p') = p'y'$。于是根据利润最大化的定义可知，$p'y' \geq p'y$。根据假设，当 $y_i \geq 0$ 时，$p_i' \geq p_i$ 以及当 $y_j \leq 0$ 时，$p_j' \leq p_j$，我们有 $p'y \geq py$。将这两个式子放在一起可得，$\pi(p') = p'y' \geq py = \pi(p)$，这正是我们想证明的。

2）令 y 是当价格为 p 时的企业利润最大化的净产出向量，因此 $py \geq py'$ 对于 Y 中的所有 y' 都成立。由此可知，对于任意$t \geq 0$，$tpy \geq tpy'$ 对于 Y 中的所有 y' 都成立。因此，

当价格为tp时，y也能使得企业的利润最大化。因此$\pi(tp)=tpy=t\pi(p)$.

3）令y当价格为p时使企业利润最大；y'当价格为p'时使企业利润最大；y''当价格为p''时使企业利润最大。于是我们有

$$\pi(p'')=p''y''=(tp+(1-t)p')y''=tpy''+(1-t)p'y''. \tag{3.1}$$

根据利润最大化的定义，可知

$$tpy''\le tpy=t\pi(p)$$

$$(1-t)py''\le(1-t)p'y'=(1-t)\pi(p').$$

将这两个不等式相加并且使用（3.1）式，可得

$$\pi(p'')\le t\pi(p)+(1-t)\pi(p').$$

4）$\pi(p)$的连续性可以从第27章的最大化定理推导出。■

利润函数对于价格是一次齐次函数以及对产出价格是递增的，这一事实并不特别令人奇怪。另一方面，凸性的性质倒不是那么明显。尽管这一事实不明显，但凸性的结果有充足的经济学理由，凸性有着重要的应用。

考虑利润对（versus）单一产品价格的图形，其中要素价格既定不变，如图3.1所示。在价格向量为(p^*,w^*)时，利润最大化生产方案(y^*,x^*)产生的利润为$p^*y^*-w^*x^*$。假设p上升，但是企业继续使用与前述相同的生产方案(y^*,x^*)。将这种被动行为产生的利润成为"被动利润函数"（passive profit function），并且用$\Pi(p)=py^*-w^*x^*$表示。容易看出这个被动利润函数是一条直线。最优生产方案产生的利润必定至少和被动生产方案一样大，因此，$\pi(p)$的图形必然位于$\Pi(p)$上方。类似地结论对于任何价格p都成立，因此，利润函数必然在每一点都位于它的切线上方，由此可知$\pi(p)$必定是凸函数。

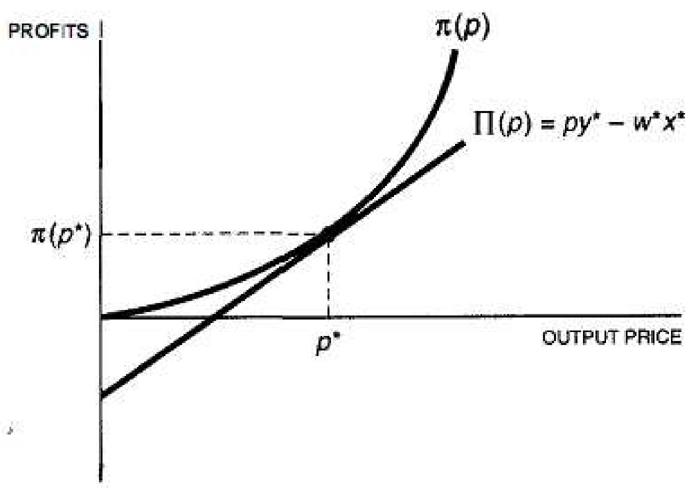


图3.1：利润函数。随着产品价格上升，利润函数以递增的速率上升。

利润函数的性质有几个用途。这些性质提供了利润最大化行为可以观测到的应用，我

们对此感到满意。例如，假设我们可以得到某个企业的会计数据，而且可以看到当所有的价格变为原来的t（$t>0$）倍时，该企业的利润并不是变为原来的t倍。继续假设其他的坏境没有明显的改变，则我们怀疑这样的企业不是追求利润最大化的企业。

例子：价格稳定的效应

假设某个竞争行业的产品价格是随机波动的。为简单起见，假设它的产品价格为p_1的概率为q，价格为p_2的概率为$(1-q)$。有人建议该行业应该将价格稳定在平均价格$\overline{p}=qp_1+(1-q)p_2$。这对行业中的企业利润有何影响？

我们必须比较价格p波动时的平均利润以及价格为平均价格时的利润。由于利润函数为凸函数，

$$q\pi(p_1)+(1-q)\pi(p_2)\geq\pi(qp_1+(1-q)p_2)=\pi(\overline{p}).$$

因此，价格波动时的平均利润至少和价格稳定(stabilization)时的利润一样大。

乍看起来，这个结论不符合我们的直觉，但是当我们思考一下利润函数为凸函数的经济学原因，我们就会明白。当价格升高时，每个企业会增加产量，当价格下降时，每个企业都会减少产量。这样做产生的利润，大于企业在平均价格生产数量固定产品所得到利润。

3.2 从利润函数推导供给函数和需求函数

如果给定一个净供给函数$y(p)$，则不难计算出利润函数。我们只要将净供给函数代入利润函数，就可以得到

$$\pi(p)=py(p).$$

现在假设给定的是利润函数，让我们求净供给函数。怎么求？有一种简单的求法：对利润函数微分。这种做法的证明是下一个命题的内容。

Hotelling 引理。（利润函数的可导性）令$y_i(p)$是企业对商品i的净供给函数。则

$$y_i(p)=\frac{\partial\pi(p)}{\partial p_i}\qquad i=1,\ldots,n$$

假设该导数存在而且$p_i>0$。

证明。假设(y^*)是价格为(p^*)时的净产出向量，则定义函数

$$g(p)=\pi(p)-py^*.$$

显然，价格为p时的利润最大化生产方案y产生的利润，至少和生产方案y^*产生的利润一

样大。然而，生产方案 y^* 是在价格为 p^* 时的利润最大化生产方案，因此函数 $g(p)$ 在 p^* 达到最小值 0。 $p_i > 0$ 的假设意味着这是个内部最小值。

最小值的一阶条件意味着

$$\frac{\partial g(p^*)}{\partial p_i} = \frac{\partial \pi(p^*)}{\partial p_i} - y_i^* = 0 \qquad i = 1,\ldots,n$$

由于该式对所有价格 p^* 都成立，证毕。■

上面的证明只是图 3.1 表示的关系的代数版本。由于“被动”利润线位于利润函数图形的下方，而且二者在某一点重合，所以这两条线必定在该点相切。但这意味着利润函数在 p^* 点的导数，必定等于价格为 p^* 时的利润最大化的要素供给： $y(p^*) = \partial\pi(p^*)/\partial p$ 。

上述对可导性（the derivative property）的论证是让人信服的（我希望是!），但是却没多少启发性。下面的论证可以帮助你理解这个性质是怎么一回事。

我们考虑产出和投入均只有一种的情形。在这种情形下，利润最大化的一阶条件为

$$p\frac{df(x)}{dx} - w = 0. \tag{3.2}$$

要素需求函数 $x(p,w)$ 必定满足这个一阶条件。

利润函数为

$$\pi(p,w) \equiv pf(x(p,w)) - wx(p,w).$$

将利润函数对 w 求导，可得

$$\begin{aligned}\frac{\partial \pi}{\partial w} &= p\frac{\partial f(x(p,w))}{\partial x}\frac{\partial x}{\partial w} - w\frac{\partial x}{\partial w} - x(p,w)\\ &= \left[p\frac{\partial f(x(p,w))}{\partial x} - w\right]\frac{\partial x}{\partial w} - x(p,w).\end{aligned}$$

将（3.2）代入上式可得

$$\frac{\partial \pi}{\partial w} = -x(p,w).$$

上式为负源于以下事实：要素的价格上升，利润必定下降。

这个论证展现了 Hotelling 引理背后的经济合理性。当产品的价格稍微上升时，会产生两种效应。首先是直接效应：由于价格上升，企业挣得的利润更多，即使它的产量维持在原来的水平。

其次是间接效应：产品价格稍微上升会使企业相应稍微增加一些产量。然而，由于产

量微小增加导致的利润变化必定为零，这是由于我们已经在利润最大化的生产方案上。因此，间接效应的影响为零，剩下的只是直接效应。你可以使用上述利润函数对 p 求导验证一下。

3.3 包络定理

利润函数的可导性质是包络定理的一种特殊情形，请参见第 27 章。考虑一个任意的最大化问题，其中目标函数取决于某个参数 a：

$$M(a)=\max\ f(x,a).$$

函数 $M(a)$ 表明目标函数的最大值是参数 a 的函数。在利润函数的情形下，a 是某个价格；x 为某种要素的需求；$M(a)$ 为利润的最大值，它是价格的函数。

令 $x(a)$ 是该利润最大化问题的解 x 的值。于是我们也可以将 $M(a)$ 写为 $M(a)=f(x(a),a)$。这个式子的意思是说，函数的最优值等于函数在最优选择点的值。

我们通常对 $M(a)$ 如何随着 a 的变动而变动感兴趣。包络定理告诉了我们这个问题的答案：

$$\frac{dM(a)}{da}=\frac{\partial f(x(a),a)}{\partial a}\bigg|_{x=x(a)}$$

这个式子是说：M 关于 a 的导数，等于 f 关于 a 的偏导数在 $x=x(a)$ 时的值。这正是上式右侧竖线的意思。包络定理的证明比较直接，请参见第 27 章。（在参见答案之前，你应该自己证明一下这个结论。）

我们看看包络定理在一种投入-一种产出的利润最大化问题中的应用。利润最大化问题为

$$\pi(p,w)=\max_x\ pf(x)-wx.$$

包络定理中的参数 a 在这种情形下为 p 或 w，$M(a)$ 为 $\pi(p,w)$。由包络定理可知，$\pi(p,w)$ 关于 p 的导数，等于目标函数关于 p 的导数在最优选择之处的值：

$$\frac{\partial\pi(p,w)}{\partial p}=f(x)\Big|_{x=x(p,w)}=f(x(p,w)).$$

上式是企业在价格 (p,w) 时的产品最优供给量。

类似地，

$$\frac{\partial\pi(p,w)}{\partial w}=x\Big|_{x=x(p,w)}=-x(p,w)$$

是企业在价格 (p,w) 时的要素最优净供给量。

3.4 使用利润函数进行比较静态分析

本章一开始，我们就证明了利润函数必定满足某些性质。我们已经知道，净供给函数是利润函数的导数。我们对从利润函数的性质推导净供给函数的性质感兴趣。下面我们逐一分析这些性质。

首先，利润函数是价格的单调函数。因此，如果商品i是投入品，则$\pi(p)$关于p_i的偏导数为负；如果商品i是产出品，则$\pi(p)$关于p_i的偏导数为正。这只是我们对净供给符号采用的一种惯例。

其次，利润函数是价格的一次齐次函数。我们已经知道，这意味着利润函数的偏导数必定是零次齐次的。所有的价格变为原来的t（其中$t>0$）倍，不会改变企业的最优选择，因此利润也变为原来的t倍。

第三，利润函数是价格的凸函数。因此，利润函数π关于价格p的二阶偏导数，即Hessian矩阵，必定是正半定的。但是利润函数的二阶导数构成的矩阵就是净供给函数一阶导数构成的矩阵。例如，在商品只有两种的情形下，我们有

$$\begin{pmatrix} \dfrac{\partial^2\pi}{\partial p_1^2} & \dfrac{\partial^2\pi}{\partial p_1\partial p_2} \\ \dfrac{\partial^2\pi}{\partial p_1\partial p_2} & \dfrac{\partial^2\pi}{\partial p_1^2} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial y_1}{\partial p_1} & \dfrac{\partial y_1}{\partial p_2} \\ \dfrac{\partial y_2}{\partial p_1} & \dfrac{\partial y_2}{\partial p_2} \end{pmatrix}$$

右侧的矩阵就是替代矩阵：当商品j的价格变化时，商品i的净供给如何变化。从利润函数的性质可知，它必定是一个对称的、正半定的矩阵。

净供给函数是利润函数的导数的事实，在利润函数的性质和净供给函数的性质之间搭建了一个桥梁。使用这种关系很容易推导出利润最大化行为的有关命题。

例子：LeChatelier 原理

下面我们将企业供给行为的短期反应与长期反应进行比较。根据定义，长期与短期相比，企业可以调整更多的生产因素，因此，在长期企业似乎对价格变化的反应更大。这个根据直觉得出的命题可以进行严格证明。

为简单起见，假设产出只有一种，并且假设所有要素的价格是固定不变的。因此，利润函数仅取决于产出的（标量）价格。定义短期利润函数为$\pi_S(p,z)$，其中z是短期中的固定要素。令$z(p)$表示长期中使利润最大化的要素z的需求，因此长期利润函数为$\pi_L(p)=\pi_S(p,z(p))$。最后，令p^*表示产出的某个既定价格，令$z^*=z(p^*)$表示产出价格为p^*时的要素z的最优长期需求。

长期利润总是至少和短期利润一样大，这是因为短期可以调整的要素集是长期可以调整的要素集的子集。由此可知，对于所有价格p都有

$$h(p)=\pi_L(p)-\pi_S(p,z^*)=\pi_S(p,z(p))-\pi_S(p,z^*)\geq 0.$$

当价格为 p^* 时，长期利润和短期利润的差为零，因此当 $p=p^*$ 时 $h(p)$ 达到最小值。所以，在 p^* 处一阶导数必定等于零。根据 Hotelling 引理可知，在 p^* 处，每种商品的长期净供给必定等于短期净供给。

但是事实不止这些。由于在 p^* 处 $h(p)$ 达到最小值，因此在此处 $h(p)$ 的二阶导数必定是非负的。这意味着

$$\frac{\partial^2 \pi_L(p^*)}{\partial p^2}-\frac{\partial^2 \pi_S(p^*,z^*)}{\partial p^2}\geq 0.$$

再次使用 Hotelling 引理，可知

$$\frac{dy_L(p^*)}{dp}-\frac{\partial^2 y_S(p^*,z^*)}{\partial p^2}=\frac{\partial^2 \pi_L(p^*)}{\partial p^2}-\frac{\partial^2 \pi_S(p^*,z^*)}{\partial p^2}\geq 0.$$

这个式子表明，与短期相比，在 $z^*=z(p^*)$ 处，企业长期供给对价格变动的反应至少和短期供给对价格变动反应一样大。

注释

利润函数的性质证明归功于 Hotelling(1932)，Hicks（1946）和 Samuelson（1947）。

习题

3.1.一个竞争性的利润最大化企业的利润函数为 $\pi(w_1,w_2)=\phi_1(w_1)+\phi_2(w_2)$。产品的价格已标准化为 1。

（a）对于函数 $\phi_i(w_i)$ 的一阶导数和二阶导数，我们知道些什么？

（b）如果 $x_i(w_1,w_2)$ 是要素 i 的需求函数，$\partial x_i/\partial w_j$ 的符号是什么样的？

（c）令 $f(w_1,w_2)$ 是产生题目中的这种形式利润函数的生产函数。这个生产函数的形式什么样的？（提示：注意一阶条件。）

3.2 考虑由下列这样的技术：对于 $x\leq 1$， $y=0$；对于 $x>1$， $y=\ln x$。计算这种技术的利润函数。

3.3 给定生产函数 $f(x_1,x_2)=a_1\ln x_1+a_2\ln x_2$，计算利润最大化的需求函数和供给函数，计算利润函数，为简单起见假设存在内部解。假设 $a_i>0$。

3.4 给定生产函数 $f(x_1,x_2)=x_1^{a_1}x_2^{a_2}$，计算利润最大化的需求函数和供给函数，计算利润函数。假设 $a_i>0$。a_1 和 a_2 应该满足什么样的限制？

3.5 给定生产函数 $f(x_1,x_2)=\min\{x_1,x_2\}^a$，计算利润最大化的需求函数和供给函数，计算利润函数。a 必须满足什么样的限制？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 4 章：成本最小化

曹乾　译

(东南大学　caoqianseu@163.com)

4 成本最小化

在本章我们将研究企业的成本最小化行为。我们对该问题感兴趣的原因有两个：一是成本最小化提供了对竞争企业的供给行为的另外一种研究方法；二是成本函数可以让我们为非竞争企业建立模型。而且，成本最小化的分析可让我们了解约束最优化的求解方法。

4.1 成本最小化的微积分分析

下面我们分析企业在生产既定产量时，如何找到成本最小化生产方法的问题：

$$\min_{x} wx$$

$$\text{使得 } f(x) = y.$$

我们使用拉格朗日乘数的方法分析这个越睡最小化问题。我们首先写出拉格朗日函数

$$L(\lambda, x) = wx - \lambda(f(x) - y)\,,$$

将这个式子对每个选择变量 x_i 以及拉格朗日乘数 λ 求导。内部解 x^* 的一阶条件为：

$$w_i - \lambda \frac{\partial f(x^*)}{\partial x_i} = 0 \qquad i = 1,\ldots,n$$

$$f(x^*) = y.$$

这些条件也可用向量符号表示。令 $Df(x)$ 表示梯度向量，即由 $f(x)$ 一阶偏导数组成的向量。我们可以将 $w_i - \lambda\partial f(x^*)/\partial x_i = 0$ 这个条件写为

$$w = \lambda Df(x^*).$$

将这些一阶条件中的第 i 个条件除以第 j 个条件，可得

$$\frac{w_i}{w_j} = \frac{\dfrac{\partial f(x^*)}{\partial x_i}}{\dfrac{\partial f(x^*)}{\partial x_j}} \qquad i, j = 1,\ldots,n. \tag{4.1}$$

上式右侧是技术替代率，即维持产量不变的情形下要素 j 替换为要素 i 的比率。上式左侧是经济替代率，即维持成本不变的情形下要素 j 替换为要素 i 的比率。（4.1）式这个条件要求技术替代率等于经济替代率。如果二者不等，则存在着某种调整方法，使得生产相同产量的成本更小。

例如，假设

$$\frac{w_i}{w_j}=\frac{2}{1}\neq\frac{1}{1}=\frac{\dfrac{\partial f(x^*)}{\partial x_i}}{\dfrac{\partial f(x^*)}{\partial x_j}}.$$

则如果我们少用一单位要素 i 多用一单位要素 j，产出不变但是成本下降了。因为少用一单位要素 i 节省了 2 元，多用一单位要素 j 多花了 1 元，综合效果是成本减少了 1 元。

（4.1）式这个一阶条件也可用图形表示。在图 4.1 中，曲线代表等产量线，直线表示固等成本曲线。当 y 固定不变时，企业的问题是在既定的等产量线上找到成本最小化的点。等成本曲线 $C=w_1x_1+w_2x_2$，可以写为 $x_2=C/w_2-(w_1/w_2)x_1$。对于固定的 w_1 和 w_2，企业想在既定的等产量线上找到一点，使得相应的等成本曲线的纵截距最小。显然这样的点可用相切条件进行刻画，因此等成本曲线的斜率必定等于等产量线的斜率。这一条件如用代数式表达就是（4.1）式。

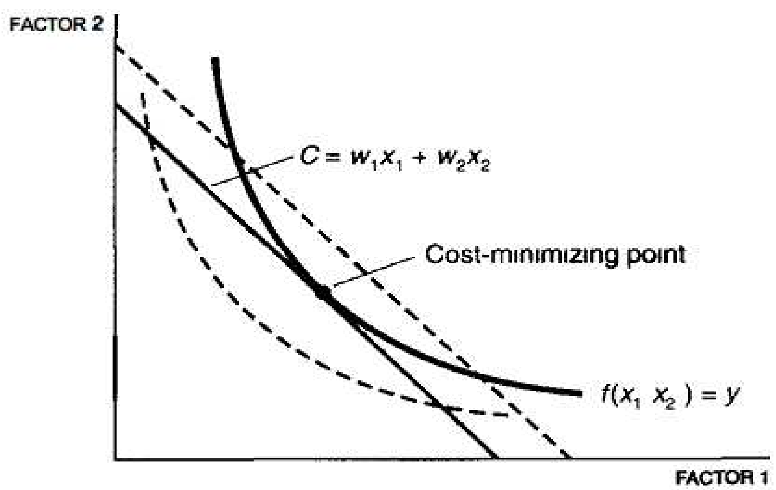


图 4.1：成本最小化。在使成本最小化的点，等产量线必定与等成本线相切。

审视图 4.1 表明，在成本最小化的选择之处，还必须满足二阶条件，也就是说，等产量线必定位于等成本线上方。换一句说，维持成本不变的任何要素投入量的变动，即沿着等产量线的运动，必定导致产量减少或不变。

这个条件意味着什么？令 (h_1,h_2) 表示要素 1 和 2 的微小变动，考虑产量的相应变化。假设函数 f 满足必要的可导条件，我们可以写出二阶泰勒级数展开式

$$f(x_1+h_1,x_2+h_2)\approx f(x_1,x_2)+\frac{\partial f}{\partial x_1}h_1+\frac{\partial f}{\partial x_2}h_2+\frac{1}{2}\left[\frac{\partial^2 f}{\partial x_1^2}h_1^2+2\frac{\partial^2 f}{\partial x_1\partial x_2}h_1h_2+\frac{\partial^2 f}{\partial x_2^2}h_2^2\right].$$

写成矩阵的形式更方便些

$$f(x_1+h_1,x_2+h_2)\approx f(x_1,x_2)+\begin{pmatrix}f_1 & f_2\end{pmatrix}\begin{pmatrix}h_1\\ h_2\end{pmatrix}+\frac{1}{2}\begin{pmatrix}h_1 & h_2\end{pmatrix}\begin{pmatrix}f_{11} & f_{12}\\ f_{21} & f_{22}\end{pmatrix}\begin{pmatrix}h_1\\ h_2\end{pmatrix}.$$

如果(h_1, h_2)微小变动而又使成本不变，则必定满足$w_1h_1 + w_2h_2 = 0$。从成本最小化的一阶条件中替代w_i，我们可以将它写为

$$w_1h_1 + w_2h_2 = \lambda f_1h_1 + \lambda f_2h_2 = \lambda[f_1h_1 + f_2h_2] = 0.$$

因此，沿着等产量线，这个泰勒展开式中的一阶条件必定等于零。所以，沿着等成本线的任何移动，如果产量下降，则：

$$\begin{pmatrix} h_1 & h_2 \end{pmatrix}\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \le 0 \text{ 对于满足} \begin{pmatrix} f_1 & f_2 \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = 0 \text{的所有} \begin{pmatrix} h_1 & h_2 \end{pmatrix}\text{都成立。} \quad (4.2)$$

直觉上，在成本最小化的点上，一阶条件意味着如果离开该点则产量不变，而二阶条件意味着离开该点产量下降。

这种二阶条件的表达方法可以推广到要素为n种的情形；二阶条件要求在下列线性约束条件下

$$h^t D^2 f(x^*) h \le 0 \qquad \text{其中 } h \text{ 满足 } wh = 0,$$

生产函数的 Hessian 矩阵是负半定的。

4.2 二阶条件的另外分析方法

在第 27 章，我们说明可以用拉格朗日函数的 Hessian 矩阵分析二阶条件。下面我们使用这种分析方法分析上一节的问题。

在这种情形下，拉格朗日函数为

$$L(\lambda, x_1, x_2) = w_1x_1 + w_2x_2 - \lambda[f(x_1, x_2) - y].$$

成本最小化的一阶条件为：拉格朗日函数关于λ, x_1和x_2的导数等于零。二阶条件涉及到拉格朗日函数的 Hessian 矩阵

$$D^2L(\lambda^*, x_1^*, x_2^*) = \begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda \partial x_1} & \frac{\partial^2 L}{\partial \lambda \partial x_2} \\ \frac{\partial^2 L}{\partial x_1 \partial \lambda} & \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_2 \partial \lambda} & \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{pmatrix}.$$

使用f_{ij}表示$\partial^2 f / \partial x_i \partial x_j$是方便的。计算上面的各个二阶导数，并使用$f_{ij}$进行表示，可得

$$D^2L(\lambda^*, x_1^*, x_2^*) = \begin{pmatrix} 0 & -f_1 & -f_2 \\ -f_1 & -\lambda f_{11} & -\lambda f_{12} \\ -f_2 & -\lambda f_{21} & -\lambda f_{22} \end{pmatrix}. \quad (4.3)$$

上式右侧称为加边 Hessian 矩阵。从第 27 章的知识可知，当且仅当加边 Hessian 矩阵的行列

式为负时，（4.2）式表述的二阶条件变为严格不等式。这样，给定某种情形，我们就有了一个判断该情形是否满足二阶条件的相对简便的方法。

在要素为 n 种的一般情形下，二阶条件变得相对复杂。在这种情形下，我们必须检查加边 Hessian 矩阵子矩阵的行列式的符号。请参见第 27 章。

例如，假设生产要素为 3 种。加边 Hessian 矩阵为

$$D^2L(\lambda^*, x_1^*, x_2^*, x_3^*) = \begin{pmatrix} 0 & -f_1 & -f_2 & -f_3 \\ -f_1 & -\lambda f_{11} & -\lambda f_{12} & -\lambda f_{13} \\ -f_2 & -\lambda f_{21} & -\lambda f_{22} & -\lambda f_{23} \\ -f_3 & -\lambda f_{31} & -\lambda f_{32} & -\lambda f_{33} \end{pmatrix}. \qquad (4.4)$$

当要素为 3 种时，二阶条件要求（4.3）和（4.4）的行列式在最优选择处的值都为负。如果有 n 种要素，则二阶条件为严格不等式的条件是，所有上述形式的加边 Hessian 矩阵都必须为负。

4.3 麻烦

对于 w 和 y 的每一选择，都存在着某个 x^*，使得生产 y 单位产品的成本最小。这个函数给出了要素的最优选择，我们将其称为条件要素需求函数(conditional factor demand function)，并将其记为 $x(w,y)$。注意，条件要素需求函数不仅取决于产量 y，还取决于要素价格 w。成本函数是要素价格为 w 产量为 y 时的最小成本，即 $c(w,y)=wx(w,y)$。

一阶条件符合直觉，但是机械地套用一阶条件可能导致错误，利润最大化问题也存在着这样的麻烦问题。下面我们分析一下利润最大化问题中四种可能的麻烦，并探讨它们如何与成本最小化问题相关联。

首先，我们所分析的生产技术可能无法使用可微的生产函数进行描述，因此无法运用微积分技术。例如，里昂惕夫生产技术就是这样的。稍后我们将计算它的成本函数。

其次，一阶条件仅适用于内部（interior）解。如果成本最小化的点发生在边界之处，则一阶条件需要修改。因此完整的一阶条件需要区分以下两种情况：

$$\lambda \frac{\partial f(x^*)}{\partial x_i} - w_i \le 0 \qquad \text{若 } x_i^* = 0$$

$$\lambda \frac{\partial f(x^*)}{\partial x_i} - w_i = 0 \qquad \text{若 } x_i^* > 0$$

稍后我们将用具体的例子进行说明。

第三，我们讨论利润最大化问题时必须分析是否存在着利润最大化的商品束。因为在某些情形下，利润在理论上可以无限大。然而，在成本最小化问题的分析中，成本最小化的要素束一般是存在的。论证如下。我们知道，连续函数在有界闭集上能取得最小值和最大值。目标函数 wx 显然是连续函数，而且根据假设 $V(y)$ 是闭集。剩下的任务是找到 $V(y)$ 的有界子集。但这很容易做到。任意选取 x 的一个值，比如 x'。显然成本最小的要素束的成本小于 wx'。因此，我们将注意力集中在子集 $\{V(y)$ 中的 $x: wx \le wx'\}$，只要 $w >> 0$，这个子

集显然是个闭集。

第四，一阶条件确定的解可能不是唯一的。一阶条件毕竟只是必要条件。尽管对于局部最优来说，一阶条件通常也是充分条件，但是只有在某些凸性约束条件下（对于大多数成本最小化问题，要求$V(y)$是凸集），一阶条件才能确定唯一的全局最优解。

例子：柯布-道格拉斯技术的生产函数

考虑以下成本最小化问题

$$c(w, y) = \min_{x_1, x_2} w_1 x_1 + w_2 x_2$$

$$\text{使得 } A x_1^a x_2^b = y.$$

使用约束条件解出x_2，并将其代入目标函数，我们看到这个最小化问题等价于

$$\min_{x_1} w_1 x_1 + w_2 A^{-\frac{1}{b}} y^{\frac{1}{b}} x_1^{-\frac{a}{b}}.$$

一阶条件为

$$w_1 - \frac{a}{b} w_2 A^{-\frac{1}{b}} y^{\frac{1}{b}} x_1^{-\frac{a+b}{b}} = 0,$$

由此可以解出要素 1 的条件需求函数：

$$x_1(w_1, w_2, y) = A^{-\frac{1}{a+b}} \left[\frac{a w_2}{b w_1}\right]^{\frac{b}{a+b}} y^{\frac{1}{a+b}}$$

要素 2 的条件需求函数为：

$$x_2(w_1, w_2, y) = A^{-\frac{1}{a+b}} \left[\frac{a w_2}{b w_1}\right]^{-\frac{a}{a+b}} y^{\frac{1}{a+b}}.$$

成本函数为

$$\begin{aligned} c(w_1, w_2, y) &= w_1 x_1(w_1, w_2, y) + w_2 x_2(w_1, w_2, y) \\ &= A^{-\frac{1}{a+b}} \left[\left(\frac{a}{b}\right)^{\frac{b}{a+b}} + \left(\frac{a}{b}\right)^{-\frac{a}{a+b}}\right] w_1^{\frac{a}{a+b}} w_2^{\frac{b}{a+b}} y^{\frac{1}{a+b}}. \end{aligned}$$

当我们使用柯布-道格拉斯技术为例时，我们通常假设$A = 1$以及使用规模报酬不变的假设$a + b = 1$。在这种情形下，成本函数简化为

$$c(w_1, w_2, y) = k w_1^a w_2^{1-a} y,$$

其中$k = a^{-a}(1-a)^{a-1}$。

例子：CES 技术的成本函数

假设 $f(x_1,x_2)=(x_1^\rho+x_2^\rho)^{\frac{1}{\rho}}$。与其相伴的成本函数是什么样的？成本最小化问题为

$$c(w,y)=\min_{x_1,x_2} w_1x_1+w_2x_2$$

$$\text{使得 } x_1^\rho+x_2^\rho=y^\rho.$$

一阶条件为

$$w_1-\lambda\rho x_1^{\rho-1}=0$$

$$w_2-\lambda\rho x_2^{\rho-1}=0$$

$$x_1^\rho+x_2^\rho=y^\rho.$$

从前两个式子解出 x_1^ρ 和 x_2^ρ，我们有

$$\begin{aligned} x_1^\rho &= w_1^{\frac{\rho}{\rho-1}}(\lambda\rho)^{-\frac{\rho}{\rho-1}} \\ x_2^\rho &= w_2^{\frac{\rho}{\rho-1}}(\lambda\rho)^{-\frac{\rho}{\rho-1}}. \end{aligned} \tag{4.5}$$

将这两个式子代入生产函数可得

$$(\lambda\rho)^{\frac{-\rho}{\rho-1}}\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]=y^\rho.$$

解出 $(\lambda\rho)^{\frac{-\rho}{\rho-1}}$ 并代入（4.5）式。由此就得到了条件要素需求函数

$$x_1(w_1,w_2,y)=w_1^{\frac{1}{\rho-1}}\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]^{-\frac{1}{\rho}}y$$

$$x_2(w_1,w_2,y)=w_2^{\frac{1}{\rho-1}}\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]^{-\frac{1}{\rho}}y.$$

将这些函数代入成本函数的定义可得

$$\begin{aligned} c(w_1,w_2,y) &= w_1x_1(w_1,w_2,y)+w_2x_2(w_1,w_2,y) \\ &= y\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]^{-\frac{1}{\rho}}=y\left[w_1^{\frac{\rho}{\rho-1}}+w_2^{\frac{\rho}{\rho-1}}\right]^{\frac{\rho-1}{\rho}}. \end{aligned}$$

令 $r=\rho/(\rho-1)$ 可将上式写得更简洁些：

$$c(w_1, w_2, y) = y\left[w_1^r + w_2^r\right]^{\frac{1}{r}}.$$

注意，这个成本函数的形式和它的原始 CES 函数的形式是相同的，只不过此处用 r 替代了原 CES 函数中的 ρ 。对于下列的一般情形

$$f(x_1, x_2) = \left[(a_1 x)^\rho + (a_2 x_2)^\rho\right]^{\frac{1}{\rho}},$$

计算成本函数的过程类似，可得

$$c(w_1, w_2, y) = \left[(w_1 / a_1)^r + (w_2 / a_2)^r\right]^{\frac{1}{r}} y.$$

例子：里昂惕夫技术的生产函数

假设 $f(x_1, x_2) = \min\{ax_1, bx_2\}$ 。与此相伴的成本函数是什么样的？由于我们知道企业不会浪费价格为正的任何投入，企业必定在 $y = ax_1 = bx_2$ 之处生产。因此，如果企业想生产 y 单位产品，不管这两种要素的价格为多大，它必须使用 y/a 单位要素 1 和 y/b 单位要素 2。所以，成本函数为

$$c(w_1, w_2, y) = \frac{w_1 y}{a} + \frac{w_2 y}{b} = y(\frac{w_1}{a} + \frac{w_2}{b}).$$

例子：线性技术的成本函数

假设 $f(x_1, x_2) = ax_1 + bx_2$，因此要素 1 和 2 是完全替代的。相应的成本函数是什么样的？由于这两种要素是完全替代的，哪种要素便宜企业就会使用哪种。所以，成本函数的形式为 $c(w_1, w_2, y) = \min\{w_1 / a, w_2 / b\} y.$

在这种情形下，成本最小化问题的答案通常涉及边界解：其中一种要素的使用量为零。尽管对于这个特殊的问题，答案容易看出，我们还是要提供一个更为正式的解法，目的是展示 Kuhn-Tucker 定理的用法。由于线性技术下，我们几乎得不到内部解，使用 Kuhn-Tucker 定理是合适的。

为方便起见假设 $a = b = 1$。最小化问题为

$$\min_{x_1, x_2} w_1 x_1 + w_2 x_2$$

$$\text{使得} \quad \begin{aligned} x_1 + x_2 &= y \\ x_1 &\geq 0 \\ x_2 &\geq 0. \end{aligned}$$

这个问题的拉格朗日函数为

$$L(\lambda, \mu_1, \mu_2, x_1, x_2) = w_1 x_1 + w_2 x_2 - \lambda(x_1 + x_2 - y) - \mu_1 x_1 - \mu_2 x_2.$$

Kuhn-Tucker 一阶条件为

$$
\begin{aligned}
w_1 - \lambda - \mu_1 &= 0 \\
w_2 - \lambda - \mu_2 &= 0 \\
x_1 + x_2 &= y \\
x_1 &\geq 0 \\
x_2 &\geq 0.
\end{aligned}
$$

补充的松弛条件（slackness conditions）为

$$\mu_1 = 0, \mu_2 \geq 0 \quad 若\, x_1 > 0$$

$$\mu_1 \geq 0, \mu_2 = 0 \quad 若\, x_2 > 0.$$

为了确定这个最小化问题的解，我们必须检查各种可能的情形，其中不等式约束可能为束缚的（binding）也可能是非束缚的。由于不等式约束条件有两个，每个都可能是束缚的或非束缚的，我们需要考虑四种情形。

1） $x_1 = 0$， $x_2 = 0$。在这种情形下，我们不可能满足条件 $x_1 + x_2 = y$，除非 $y = 0$。

2） $x_1 = 0$， $x_2 > 0$。在这种情形下，我们知道 $u_2 = 0$。因此，根据前两个一阶条件可知，

$$
\begin{aligned}
w_1 &= \lambda + \mu_1 \\
w_2 &= \lambda
\end{aligned}
$$

由于 $\mu_1 \geq 0$，这种情形只有在 $w_1 \geq w_2$ 时出现。由于 $x_1 = 0$，所以 $x_2 = y$。

3）$x_2 = 0$，$x_1 > 0$。推理和 2）类似，结果为 $x_1 = y$，这种情形下只有 $w_2 \geq w_1$ 时才能出现。

4）$x_1 > 0$，$x_2 > 0$。在这种情形下，补充松弛条件意味着 $\mu_1 = 0$ 和 $\mu_2 = 0$。因此一阶条件意味着 $w_1 = w_2$。

上面的问题，尽管多少有些平凡，但它是运用 Kuhn-Tucker 定理的典型方法。如果有 k 个约束条件，这些条件可能是束缚的也可能不是束缚的，则最优解存在 2^k 种情形。每种情形都需要分析，以判断它是否与所有约束条件相容，若是它就可能是个最优解。

4.4 条件要素需求函数

下面我们转而分析成本最小化问题以及条件要素需求。根据通常的论证可知，条件要素需求函数 $x(w, y)$ 必定满足下列一阶条件

$$
\begin{aligned}
f(x(w, y)) &\equiv y \\
w - \lambda Df(x(w, y)) &\equiv 0.
\end{aligned}
$$

在接下来的矩阵运算中很容易出错，因此我们以两种商品为例进行分析。在这种情形下，一阶条件为

$$f(x_1(w_1, w_2, y), x_2(w_1, w_2, y)) \equiv y$$

$$w_1 - \lambda \frac{\partial f(x_1(w_1, w_2, y), x_2(w_1, w_2, y))}{\partial x_1} \equiv 0$$

$$w_2-\lambda\frac{\partial f(x_1(w_1,w_2,y),x_2(w_1,w_2,y))}{\partial x_2}\equiv 0$$

就象上一章一样，这些一阶条件是从条件要素需求函数的定义得出的，它们对于 w_1, w_2 和 y 的任何值都成立。因此，我们可以对这些恒等式求导，比如关于 w_1 求导，可得

$$\frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial w_1}+\frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial w_1}\equiv 0$$

$$1-\lambda\left[\frac{\partial^2 f}{\partial x_1^2}\frac{\partial x_1}{\partial w_1}+\frac{\partial^2 f}{\partial x_1\partial x_2}\frac{\partial x_2}{\partial w_1}\right]-\frac{\partial f}{\partial x_1}\frac{\partial \lambda}{\partial w_1}\equiv 0$$

$$1-\lambda\left[\frac{\partial^2 f}{\partial x_2\partial x_1}\frac{\partial x_1}{\partial w_1}+\frac{\partial^2 f}{\partial x_2^2}\frac{\partial x_2}{\partial w_1}\right]-\frac{\partial f}{\partial x_2}\frac{\partial \lambda}{\partial w_1}\equiv 0.$$

这些式子可以写成矩阵形式：

$$\begin{pmatrix}0 & -f_1 & -f_2\\ -f_1 & -\lambda f_{11} & -\lambda f_{21}\\ -f_2 & -\lambda f_{12} & -\lambda f_{22}\end{pmatrix}\begin{pmatrix}\frac{\partial \lambda}{\partial w_1}\\ \frac{\partial x_1}{\partial w_1}\\ \frac{\partial x_2}{\partial w_1}\end{pmatrix}\equiv\begin{pmatrix}0\\ -1\\ 0\end{pmatrix}.$$

注意下列重要事实：上式左侧的矩阵正是最大化问题二阶条件涉及的加边 Hessian 矩阵。请参考第 27 章。我们可以使用线性代数中的一个标准技术，即 Cramer 法则（请参见第 26 章），以求解 $\partial x_1/\partial w_1$：

$$\frac{\partial x_1}{\partial w_1}=\frac{\begin{pmatrix}0 & 0 & -f_2\\ -f_1 & -1 & -\lambda f_{21}\\ -f_2 & 0 & -\lambda f_{22}\end{pmatrix}}{\begin{pmatrix}0 & -f_1 & -f_2\\ -f_1 & -\lambda f_{11} & -\lambda f_{21}\\ -f_2 & -\lambda f_{12} & -\lambda f_{22}\end{pmatrix}}.$$

令 H 表示上式右侧分母中的矩阵。根据最大化问题的二阶条件，我们知道 $H<0$。将分子展开进行计算可得

$$\frac{\partial x_1}{\partial w_1}=\frac{f_2^2}{H}<0.$$

因此，条件要素需求函数是向下倾斜的。

类似地，我们可以推出 $\partial x_2/\partial w_1$ 的表达式。再次运用 Cramer 法则，可得

$$\frac{\partial x_2}{\partial w_1}=\frac{\begin{vmatrix}0 & -f_1 & 0\\ -f_1 & -\lambda f_{11} & -1\\ -f_2 & -\lambda f_{12} & 0\end{vmatrix}}{\begin{vmatrix}0 & -f_1 & -f_2\\ -f_1 & -\lambda f_{11} & -\lambda f_{21}\\ -f_2 & -\lambda f_{12} & -\lambda f_{22}\end{vmatrix}}.$$

由上式可得

$$\frac{\partial x_2}{\partial w_1}=-\frac{f_2 f_1}{H}>0. \tag{4.6}$$

类似地，我们可以计算$\partial x_1/\partial w_2$，可得

$$\frac{\partial x_2}{\partial w_1}=\frac{\begin{vmatrix}0 & 0 & -f_2\\ -f_1 & 0 & -\lambda f_{21}\\ -f_2 & -1 & -\lambda f_{22}\end{vmatrix}}{\begin{vmatrix}0 & -f_1 & -f_2\\ -f_1 & -\lambda f_{11} & -\lambda f_{21}\\ -f_2 & -\lambda f_{12} & -\lambda f_{22}\end{vmatrix}}.$$

这意味着

$$\frac{\partial x_1}{\partial w_2}=-\frac{f_1 f_2}{H}>0. \tag{4.7}$$

比较（4.6）和（4.7）可知，它们是相同的。因此，$\partial x_1/\partial w_2$等于$\partial x_2/\partial w_1$。正如利润最大化的情形一样，我们找到了一个对称条件：成本最小化模型的一个结果是“交叉价格效应必定相等”。

在投入要素为两种的情况下，交叉价格效应必定为正的。也就是说，这两种要素必定互为替代品（substitutes）。要注意，这个结论只对投入要素为两种的情形适用。如果要素多于两种，交叉价格效应可能为正也可能为负。

下面我们使用矩阵代数的术语重新表达前面的计算过程。由于在所有的计算中，y是维持不变的。为简单起见，我们将y从条件要素需求函数的变量中舍去。成本最小化的一阶条件为

$$f(x(w))\equiv y$$

$$w-\lambda Df(x(w))\equiv 0.$$

将上面的恒等式对w求导，可得

$$Df(x(w))Dx(w)\equiv 0$$

$$1-\lambda D^2 f(x(w))Dx(w)-Df(x(w))D\lambda(w)\equiv 0.$$

将上面的式子稍微变形可得

$$\begin{pmatrix} 0 & -Df(x) \\ -Df(x)^t & -\lambda D^2 f(x) \end{pmatrix} \begin{pmatrix} D\lambda(w) \\ Dx(w) \end{pmatrix} = -\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

注意，上式左侧的第一个矩阵是加边 Hessian 矩阵，即拉格朗日函数二阶导数组成的矩阵。假设最优解为正则的，所以Hessian矩阵为非退化的（nondegenerate）。我们可以通过求Hessian矩阵的逆矩阵的方法求解替代矩阵 $Dx(w)$：

$$\begin{pmatrix} D\lambda(w) \\ Dx(w) \end{pmatrix} = \begin{pmatrix} 0 & Df(x) \\ Df(x)^t & \lambda D^2 f(x) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

（我们已经将原等式左右两侧同乘以 -1 以消除两侧的负号。）由于加边 Hessian 矩阵是对称的，它的逆也是对称的，这意味着交叉价格效应是对称的。因为下面我们将用另外的方法证明这个结论，所以在此处我们就不再证明了。

4.5 成本最小化的代数分析方法

和利润最大化的情形一样，我们也可以使用代数方法分析成本最小化问题。假设我们观察到企业的产量选择为 y^t，要素价格为 w^t，要素使用水平为 x^t，其中 $t=1,...,T$。这些数据何时与成本最小化模型相符？

一个明显的必要条件是，观察到的投入数量产生的成本，不大于任何其他要素投入水平（至少生产同样产量）的成本。用数学语言表达，

对于满足 $y^s \geq y^t$ 的所有 s 和 t，都有 $w^t y^t \leq w^t y^s$。

我们将这个条件称为成本最小化弱公理（Weak Axiom of Cost Minimization, WACM）。

和利润最大化情形一样，可以使用成本最小化弱公理推导出向下倾斜需求的“德尔塔”（Δ）版本。维持产量不变的情形下，取两套不同的观察值，成本最小化意味着

$$w^t x^t \leq w^t x^s$$

$$w^s x^s \leq w^s x^t$$

第一个式子是说，在 t 期价格下，t 期的成本必定更小；第二个式子是说，在在 s 期价格下，s 期的成本必定更小。

将第二个不等式变形

$$-w^s x^t \leq -w^s x^s,$$

将其与第一个式子相加可得

$$(w^t - w^s)(x^t - x^s) \leq 0,$$

或

$$\Delta w \Delta x \leq 0.$$

大致来说，要素需求向量和要素价格向量的运动方向必定“相反”。

你可以构建产生上述数据的真实必要投入集的内界（inner bound）和外界（outer bound）。我们在此处将描述一下这两个边界，但验证的工作留给读者完成。论证过程类似于利润最大化的情形。

内界为：

$$VI(y)=\{x^t : y^t \geq y\}\text{的凸单调包。}$$

也就是说，内界是能至少生产 y 单位产品的所有投入的凸单调包（convex monotonic hull）。外界为：

$$VO(y)=\{x : w^t x \geq w^t x^t \text{对于所有满足}\ y^t \leq y\ \text{的}\ t\ \text{成立}\}.$$

这些构造类似于前几章的 YO 和 YI 的构造。图 4.2 给出了 VO 和 VI 的图形。

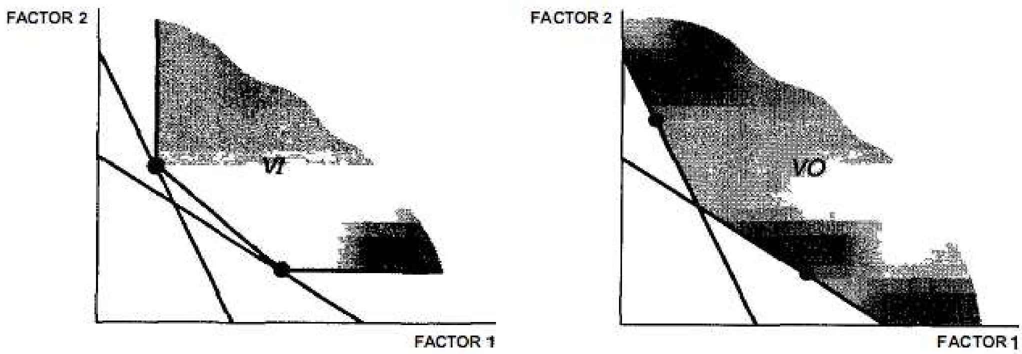


图 4.2：内界和外界。集合 VI 和 VO 分别给出了真实必要投入集的内界和外界。

容易看出，$VI(y)$ 包含于 $V(y)$，至少只要 $V(y)$ 是凸且单调时是这样的。$V(y)$ 包含于 $VO(y)$，尽管这一事实不是那么明显。所以我们要证明一下。

反证。假设存在某个 x，x 在 $V(y)$ 中但不在 $VO(y)$ 中。由于 x 不在 $VO(y)$ 中，必定存在某个观察 t 使得 $y^t \leq y$ 以及

$$w^t x < w^t x^t. \tag{4.8}$$

但是由于 x 在 $V(y)$ 中，它至少可以生产 y^t 单位的产品，并且（4.8）式表明它的成本小于 x^t。但这与 x^t 是成本最小化的要素束矛盾。

注释

成本最小化的代数分析方法进一步的研究可以参见 Varian(1982b)。

习题

4.1 严格证明利润最大化蕴涵成本最小化。

4.2 使用 Kuhn-Tucker 定理推导成本最小化的条件,要使得这个条件对于边界最优解也适用。

4.3 某个企业有两个工厂，这两个工厂的成本函数分别为$c_1(y_1)=y_1^2/2$和$c_2(y_2)=y_2$。求该企业的成本函数。

4.4 某个企业有两个工厂，这两个工厂的生产函数分别为$x_1^a x_2^{1-a}$和$x_1^b x_2^{1-b}$。求该企业的成本函数。

4.5 假设某企业生产某产品的技术有两种，一种技术使用a_1单位要素 1 和a_2单位要素 2 生产一单位产品；另外一种技术使用b_1单位要素 1 和b_2单位要素 2 生产一单位产品。这两种要素只能按上述固定比例搭配使用。如果要素的价格为(w_1,w_2)。求着两种要素的需求；求该企业的成本函数；要素价格为多大时成本函数是不可微的？

4.6 某企业有两个工厂，成本函数分别为$c_1(y_1)=4\sqrt{y_1}$和$c_2(y_2)=2\sqrt{y_2}$。求该企业产量为y时的生产成本。

4.7 下表给出了某个企业的要素需求（x_1,x_2）、要素价格（w_1,w_2）和相应产量（y）的观察数据。该企业的行为符合成本最小化行为吗？

Obs	y	w_1	w_2	x_1	x_2
A	100	2	1	10	20
B	110	1	2	14	10

4.8 某企业的成本函数为$y=x_1x_2$。如果要素价格为$w_1=w_2=1$时的最小生产成本等于 4，求产量y。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 5 章：成本函数

曹乾　译

(东南大学　caoqianseu@163.com)

5 成本函数

成本函数衡量在要素价格固定不变的情形下，生产既定产量的最小成本。成本函数刻画了企业可以使用的生产技术的信息。事实上，成本函数行为可以告诉我们企业生产技术性质的很多信息。

正如生产函数是描述生产的技术可行性的主要方法，成本函数是描述企业生产的经济可行性的主要方法。在下面两节，我们主要分析成本函数 $c(w,y)$ 关于价格和产量的行为。但在分析之前，我们需要定义一些与成本函数相关的函数，即平均成本函数和边际成本函数。

5.1 平均成本和边际成本

下面我们考虑一下成本函数的结构。一般来说，成本函数总是可以表达为条件要素需求的价值：

$$c(w,y) \equiv wx(w,y)$$

这个式子只是说，生产 y 单位产品的最小成本就是生产 y 单位产品的最便宜方法的成本。

在短期，某些生产要素的数量是既定不变的。令 x_f 表示固定要素的向量，x_v 表示可变要素的向量，这样要素价格向量 w 分解为可变要素的价格向量和固定要素的价格向量，即 $w=(w_v,w_f)$。短期条件要素需求函数一般取决于 x_f，所以我们将其写为 $x_v(w,y,x_f)$。于是，短期成本函数可以写为

$$c(w,y,x_f) = w_v x_v(w,y,x_f) + w_f x_f .$$

其中，$w_v x_v(w,y,x_f)$ 称为短期可变成本（short-run variable cost, SVC）；$w_f x_f$ 是固定成本（fixed cost, FC）。使用这些基本的单元，我们可以定义各种衍生成本概念：

$$短期总成本= STC = w_v x_v(w,y,x_f) + w_f x_f$$

$$短期平均成本= SAC = \frac{c(w,y,x_f)}{y}$$

$$短期平均可变成本= SAVC = \frac{w_v x_v(w,y,x_f)}{y}$$

$$短期平均固定成本= SATC = \frac{w_f x_f}{y}$$

$$短期边际成本= SMC = \frac{\partial c(w,y,x_f)}{\partial y}$$

当所有要素是可变的，企业会最优化对 x_f 的选择。因此，长期成本函数仅取决于要素的价格和产量水平，前面已经指出过这一点。

我们可以使用短期成本函数表达长期成本函数，方法如下。令 $x_f(w,y)$ 表示对固定要素的最优选择，令 $x_v(w,y)=x_v(w,y,x_f(w,y))$ 表示对可变要素的长期最优选择。于是，长期成本函数可以写为

$$c(w,y)=w_v x_v(w,y)+w_f x_f(w,y)=c(w,y,x_f(w,y)).$$

我们可以使用长期成本函数定义一些与上述短期成本类似的概念：

$$\text{长期平均成本}=LAC=\frac{c(w,y)}{y}$$

$$\text{长期边际成本}=LMC=\frac{\partial c(w,y)}{\partial y}$$

注意，长期平均成本等于长期可变平均成本，这是由于在长期，所有成本都是可变的。出于相同的原因，长期固定成本为零。

长期和短期当然是相关的概念。哪些要素被视为可变的、哪些被视为固定的取决于你分析的具体问题。你应该首先考虑分析企业行为的时间段，然后判断企业在这个时间段内可以调整哪些要素。

例子：短期柯布-道格拉斯函数

假设柯布-道格拉斯技术中的要素 2 的使用量被限制在水平 k。于是成本最小化问题为

$$\min\ w_1x_1+w_2x_2$$

$$\text{使得 } y=x_1^a k^{1-a}.$$

从约束条件中解出 x_1，它是 y 和 k 的函数：

$$x_1=(yk^{a-1})^{\frac{1}{a}}$$

因此，$c(w_1,w_2,y,k)=w_1(yk^{a-1})^{\frac{1}{a}}+w_2k.$

下列各个成本函数也可以求出来：

$$\text{短期平均成本}=w_1(\frac{y}{k})^{\frac{1-a}{a}}+\frac{w_2k}{y}$$

$$\text{短期平均可变成本}=w_1(\frac{y}{k})^{\frac{1-a}{a}}$$

$$\text{短期平均固定成本}=\frac{w_2k}{y}$$

$$\text{短期边际成本}=\frac{w_1}{a}(\frac{y}{k})^{\frac{1-a}{a}}$$

例子：规模报酬不变和成本函数

如果生产函数是规模报酬不变的，则成本函数显然是产量的线性函数：如果产量变为原来的 2 倍，成本也变为原来的 2 倍。这个结论的证明是下列命题的内容：

规模报酬不变。如果生产函数是规模报酬不变的，则成本函数可以写为 $c(w,y) = yc(w,1)$ 。

证明。令 x^* 为在要素价格为 w 时生产一单位产品的成本最小的方法，因此 $c(w,1) = wx^*$ 。于是可以断言 $c(w,y) = wyx^* = yc(w,1)$ 。首先可以注意到，yx^* 是生产 y 单位产品的一种可行方法，因为生产技术是规模报酬不变的。假设它未能使成本最小；令 x' 为在要素价格为 w 时生产 y 单位产品的成本最小的方法，因此 $wx' < wyx^*$ 。于是 $wx'/y < wx^*$ ，并且注意到 x'/y 可以生产一单位产品，这是因为生产技术是规模报酬不变的。这与 x^* 的定义矛盾。■

如果生产技术是规模报酬不变的，则平均成本函数、平均可变成本函数和边际成本函数都是相同的。

5.2 成本的图形

成本函数是研究企业的经济行为的最为重要的工具。稍后我们就会明白，成本函数总结了关于企业生产技术的所有经济相关信息。在下一节我们将分析成本函数的某些性质。分析这些性质时，最方便的方法是分为两步进行研究：第一步，在要素价格固定不变的假设下研究成本函数的性质；在这种情形下，我们可将成本函数写为 $c(y)$ 。第二步，在要素价格自由可变的情形下研究成本函数的性质。

由于我们将要素价格视为固定不变的，成本仅取决于企业的产量水平，这样我们就可以画出成本和产量之间的关系图。通常假设总成本曲线是产量的单调函数：你生产得越多，花费的成本越大。然而，平均成本曲线可能随着产量的增加而上升或者下降，这取决于总成本是比线性增长的速度更快还是更慢。通常人们认为最现实的情形是，平均成本先下降然后再上升，至少在短期内是这样的。原因如下。

在短期，成本函数由两部分组成：固定成本和可变成本。因此，我们可将短期平均成本写为

$$SAC = \frac{c(w,y,x_f)}{y} = \frac{w_f x_f}{y} + \frac{w_v x_v(w,y,x_f)}{y} = SAFC + SAVC.$$

在多数情形下，短期固定要素是机械、建筑物以及其他资本设备，而可变要素是劳动和原材料。下面我们分析当产量变化时，由于这些要素的变化带来的成本变动。

当我们增加产量时，平均可变成本最初可能是下降的，这是因为最初生产区域存在着规模经济的现象。然而，由于受固定要素数量的限制，当我们达到某个产量水平时，可变要素多少会和产量呈现线性增长关系。当我们接近这个产量水平时，进一步增加产量需要投入更多的可变要素（即产量增加一单位要素投入增加多于一单位）。因此，随着产量增加，平均可变成本函数最终会上升。如图 5.1A 所示。显然，平均固定成本必定随着产量的增加而下降，如图 5.1B 所示。将平均可变成本曲线和平均固定成本曲线相加就得到了平均成本曲线，它是 U 形曲线，如图 5.1C 所示。平均成本最初下降，这是由于平均固定成本下降；平均成

本最终会上升，这是因为平均可变成本的增加。平均可变成本达到最小值时的产量水平，有时称为**最小效率规模**（minimal efficient scale）。

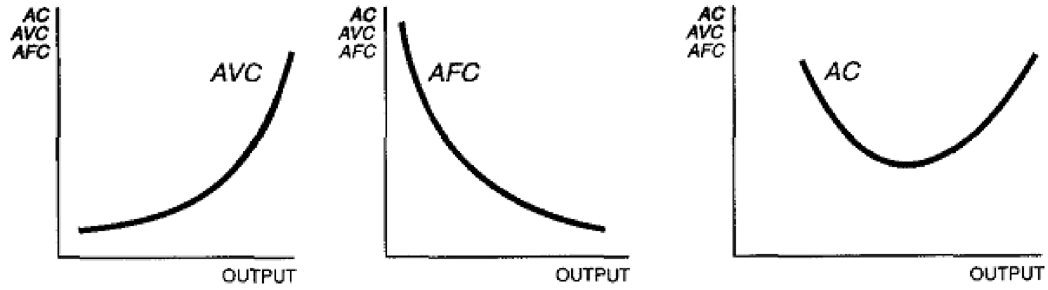


图 5.1：平均成本曲线。平均可变成本曲线最终会随着产量的增加而上升，而平均固定成本曲线总是随着产量的增加而下降。这两种效应的综合结果是平均成本曲线呈 U 形。

在长期，所有成本都是可变成本；在这样的情形下，平均成本增加似乎是不合理的，因为企业总可以复制它的生产过程。因此，长期平均成本的合理情形是平均成本是不变或递减的。另一方面，正如我们曾经指出的，某些类型的企业可能不会呈现长期规模报酬不变的现象，这是有意长期固定要素的存在。如果某些要素在长期的确是固定要素，长期平均成本曲线就会呈现 U 形，原因类似于短期平均成本曲线呈 U 形。

下面我们分析边际成本曲线。边际成本曲线和平均成本曲线有何关系？令 y^* 表示平均成本最小值之处的产量水平，则在 y^* 的左侧，平均成本曲线是下降的，即对于 $y \le y^*$ 有

$$\frac{d}{dy}\left(\frac{c(y)}{y}\right) \le 0 .$$

上式对 y 求导可得

$$\text{对于 } y \le y^*\text{，有 } \frac{yc'(y)-c(y)}{y^2} \le 0 .$$

这意味着

$$\text{对于 } y \le y^*\text{，有 } c'(y) \le \frac{c(y)}{y} .$$

这个不等式是说在最小平均成本点的左侧，边际成本小于平均成本。类似地分析可以表明

$$\text{对于 } y \ge y^*\text{，有 } c'(y) \ge \frac{c(y)}{y} .$$

由于在 y^* 点，上述两个不等式都成立，我们有

$$c'(y^*) = \frac{c(y^*)}{y^*} ;$$

也就是说，在最小平均成本点，边际成本等于平均成本。

边际成本曲线和平均可变成本曲线有何关系？将上面论证过程中的总成本替换为总可

变成本即可得到答案：当平均可变成本曲线下降时，边际成本曲线位于平均可变成本曲线的下方；当平均可变成本曲线上升时，边际成本曲线位于平均可变成本曲线的上方。由此可知，边际成本曲线必然穿过平均可变成本曲线的最低点。

不难证明，对于第一单位产品来说，边际成本必定等于平均可变成本。毕竟，第一单位产品的边际成本，等于第一单位产品的平均可变成本，原因在于这两种成本的数值此时都等于$c_v(1)-c_v(0)$。更正式的证明如下。平均可变成本的定义为

$$AVC(y)=\frac{c_v(y)}{y}.$$

如果$y=0$，上式变为$0/0$，这是未定型（indeterminate）。然而，根据 L'Hopital 法则（请参见第 26 章）可知，$c_v(y)/y$的极限可用下面的方法计算：

$$\lim_{y\to 0}\frac{c_v(y)}{y}=\frac{c_v'(0)}{1}.$$

由此可知，产量为零时的平均可变成本正是边际成本。

以上的所有分析对长期和短期都适用。然而，如果生产在长期是规模报酬不变的，即成本是产量水平的线性函数，那么平均成本、平均可变成本和边际成本都相等，这使得前面分析的这些关系显得非常平凡。

例子：柯布-道格拉斯成本曲线

从前面几章我们已经知道，一般的柯布-道格拉斯技术的成本函数为

$$c(y)=Ky^{\frac{1}{a+b}}\quad a+b\le 1$$

其中，K是要素价格和参数的函数。因此，

$$AC(y)=\frac{c(y)}{y}=Ky^{\frac{1-a-b}{a+b}}$$

$$MC(y)=c'(y)=\frac{K}{a+b}y^{\frac{1-a-b}{a+b}}.$$

如果 $a+b<1$，则平均成本曲线是上升的；如果 $a+b=1$，则平均成本固定不变。

我们在前面已经知道，柯布-道格拉斯技术的短期成本函数为

$$c(y)=Ky^{\frac{1}{a}}+F.$$

因此，

$$AC(y)=\frac{c(y)}{y}=Ky^{\frac{1-a}{a}}+\frac{F}{y}.$$

5.3 长期成本曲线和短期成本曲线

下面我们分析长期成本曲线和短期成本曲线的关系。显然，长期成本曲线绝不会位于任何短期成本曲线的上方，这是因为短期成本最小化问题只是长期成本最小化问题带有约束条件的版本。

我们将长期成本函数写为 $c(y)=c(y,z(y))$. 此处我们删去了要素价格变量，因为根据假设它们是固定不变的，令 $z(y)$ 表示对于某个固定要素的成本最小化的需求。令 y^* 表示某个既定的产量水平，$z^*=z(y^*)$ 表示在该产量水平下对该固定要素的长期需求。对于所有产量 y 来说，短期成本 $c(y,z^*)$ 必定不会小于长期成本 $c(y,z(y))$；对于产量 y^*，短期成本和长期成本是相等的，因此 $c(y^*,z^*)=c(y^*,z(y^*))$ 。因此，长期成本曲线和短期成本曲线必定在 y^* 点相切。

上述相切的结论只是包络定理在几何图形上的表现。长期成本曲线在 y^* 点的斜率为

$$\frac{dc(y^*,z(y^*))}{dy}=\frac{\partial c(y^*,z^*)}{\partial y}+\frac{\partial c(y^*,z^*)}{\partial z}\frac{\partial z(y^*)}{\partial y}.$$

但是由于 z^* 点是产量为 y^* 时对固定要素的最优需求，我们必然有

$$\frac{\partial c(y^*,z^*)}{\partial z}=0.$$

所以，长期边际成本在 y^* 之处的值等于短期边际成本在 (y^*,z^*) 处的值。

最后，我们注意到，如果长期成本曲线和短期成本曲线相切，则长期平均成本曲线和短期平均成本曲线必然也是相切的。如图 5.2 所示。

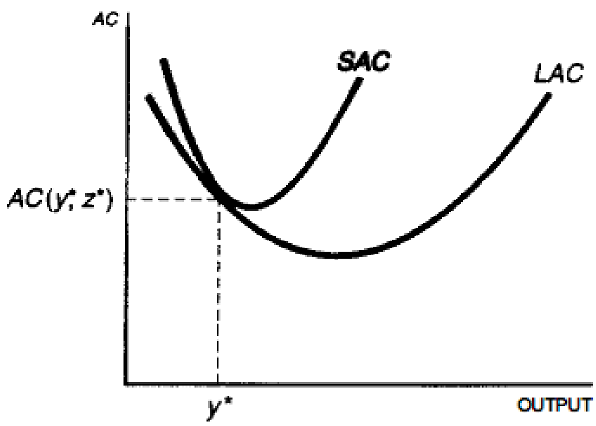


图 5.2：长期和短期平均成本曲线。注意，长期平均成本曲线和短期平均成本曲线必定相切，这意味着长期边际成本和短期边际成本必定相等。

长期平均成本曲线和短期平均成本曲线关系的另外一种分析方法，是从短期平均成本曲线族开始分析。例如，假设某固定要素的数量水平只有三种：z_1, z_2 和 z_3 。我们将短期平均成本曲线族画在图 5.3 中。长期成本曲线是什么样的？长期成本曲线就是这些短期成本曲线的下包络线（lower envelope），因为：生产 y 单位产品的最优选择 z ，就是生产 y 单位产品成本最小的选择。这个包络操作产生的长期平均成本曲线是锯齿形状的。如果固定要素的

数量水平有很多种，那么这些锯齿就变成了一条平滑的曲线。

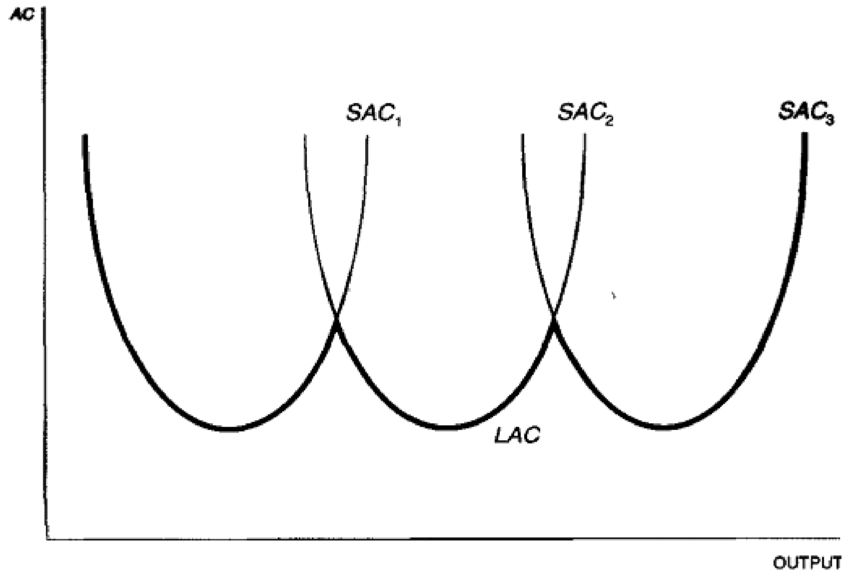


图 5.3：长期平均成本曲线。长期平均成本曲线 LAC，是短期成本曲线 SAC_1, SAC_2 和 SAC_3 的下包络线。

5.4 要素价格和成本函数

我们转而分析成本函数的价格行为。从成本函数的定义可以直接推出若干有趣的性质。总结如下。注意这些性质和利润函数性质的类似性。

成本函数的性质。

1）对 w 是非减的。如果 $w' \geq w$，那么 $c(w', y) \geq c(w, y)$.

2）关于 w 是一次齐次的。$c(tw, y) \geq tc(w, y)$，其中 $t > 0$.

3）关于 w 是凹的。$c(tw + (1-t)w', y) \geq tc(w, y) + (1-t)c(w', y)$，其中 $0 \leq t \leq 1$.

4）关于 w 是连续的。$c(w, y)$ 是 w 的连续函数，其中 $w >> 0$.

证明。

1）这个结论很明显，但是正式的证明具有启发性。令 x 和 x' 分别是与 w 和 w' 相伴的成本最小化要素束，而且 $w \leq w'$。那么，根据最小化可知 $wx \leq wx'$；并且 $wx' \leq w'x'$，这是由于 $w \leq w'$。将这些不等式放在一起可知 $wx \leq w'x'$，这正是我们想要的。

2）我们需要证明：若 x 是价格为 w 时的成本最小的要素束，则 x 也是价格为 tw 时的成本最小的要素束。反证法。假设不是，令 x' 是价格为 tw 时的成本最小的要素束，于是 $twx' < twx$。但这意味着 $wx' < wx$，这和 x 的定义矛盾。因此，将要素价格都乘以一个正数 t 不会改变成本最小要素束的构成，所以，成本也恰好变为原来的 t 倍：$c(tw, y) = twx = tc(w, y)$.

3）令 (w, x) 和 (w', x') 是两个成本最小化的要素价格-数量组合，令 $w'' = tw + (1-t)w'$，其

中 $0 \le t \le 1$.现在

$$c(w'', y) = w''x'' = twx'' + (1-t)w'x''.$$

由于 x'' 未必是要素价格为 w' 或 w 时生产 y 单位产品的成本最小方法，我们有 $wx'' \ge c(w, y)$ 以及 $w'x'' \ge c(w', y)$ 。于是，

$$c(w'', y) \ge tc(w, y) + (1-t)c(w', y).$$

4）$c(w, y)$ 的连续性可从最大化定理推出，参见第 27 章。■

以上性质中唯一让人稍有惊讶的是凹性。然而，我们可以提供一个直觉解释，这个解释类似于对利润函数的解释。假设成本函数是某种要素价格的函数，维持其他价格不变。如果该要素的价格上升，成本绝不会下降（性质 1），但是成本上升的速率是递减的（性质 3）。为什么？因为如果这种要素的价格变得更高但其他要素价格不变时，追求成本最小化的企业会转而使用其他要素。

借助图形 5.4 可以说得更清楚。令 x^* 是价格为 w^* 时的成本最小化要素束。假设要素 1 的价格从 w_1^* 变为 w_1 。如果我们对上述价格的变动的反应是被动的，从而继续使用 x^* ，那么我们的成本为 $C = w_1 x_1^* + \sum_{i=2}^n w_i^* x_i^*$ 。最小成本 $c(w, y)$ 必定小于这个“被动的”成本函数；因此，$c(w, y)$ 的图形必定位于被动成本函数的下方，两条曲线在 w_1^* 点重合。不难看出，这意味着 $c(w, y)$ 关于 w_1 是凹的。

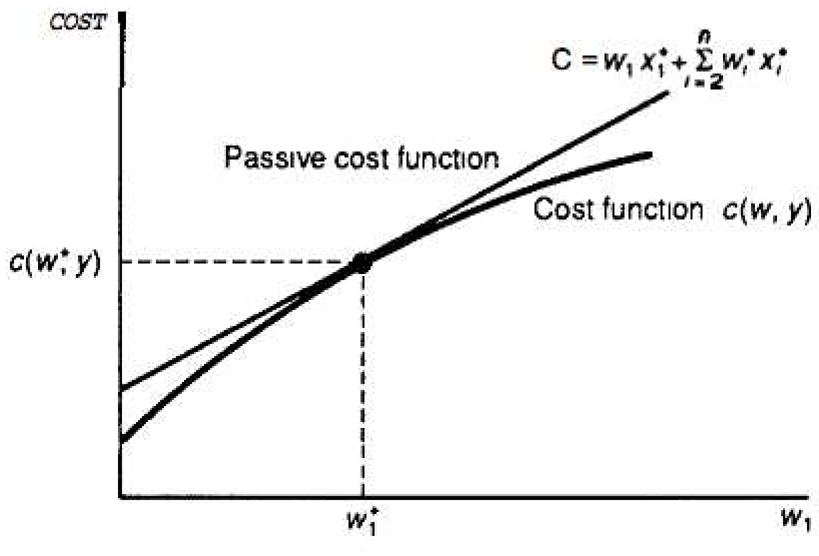


图 5.4：成本函数的凹性。成本函数是要素价格的凹函数，这是因为它必然位于“被动”成本函数的下方。

使用图 5.4 还可以发现条件要素需求函数的另外一种有用的表达方法。我们首先正式给出结论：

Shephard 引理。（可导性）令 $x_i(w, y)$ 是企业对要素 i 的条件要素需求。则如果成本函数在 (w, y) 点是可微的，并且 $w_i > 0$ ，其中 $i = 1, \ldots, n$ ，那么

$$x_i(w,y)=\frac{\partial c(w,y)}{\partial w_i} \qquad i=1,\ldots,n.$$

证明。这个证明非常类似 Hotelling 引理的证明。令 x^* 是价格为 w^* 时生产 y 单位产品的成本最小化要素束。定义函数

$$g(w)=c(w,y)-wx^*.$$

由于 $c(w,y)$ 是生产 y 单位产品的成本最小的方法，上述函数总是非正的。在点 $w=w^*$，$g(w^*)=0$。由于这是 $g(w)$ 的最大值，它的导数必定为零：

$$\frac{\partial g(w^*)}{\partial w_i}=\frac{\partial c(w,y)}{\partial w_i}-x_i^*=0 \quad i=1,\ldots,n$$

所以，成本最小化的要素投入向量，就是由成本函数关于要素价格的导数组成的向量。■

由于这个命题的重要性，我们再给出四种不同的证明方法。第一种方法，根据成本函数的定义可知 $c(w,y)\equiv wx(w,y)$。将这个式子对 w_i 求导，并且使用一阶条件就可得到结果。（提示：$x(w,y)$ 也满足恒等式 $f(x(w,y))\equiv y$。你需要将这个式子对 w_i 求导。）

第二种方法，上述计算过程正好是包络定理的推导过程，我们将在下一节介绍这一定理。运用包络定理可以直接得到我们想要的结果。

第三种方法，使用图形 5.4 进行论证。我们知道在图 5.4 中，曲线 $C=w_1x_1^*+\sum_{i=2}^{n}w_i^*x_i^*$ 位于 $c(w,y)$ 的上方，而且这两条曲线在 $w_1=w_1^*$ 处重合。因此，这两条曲线必定相切，所以 $x_i(w,y)=\partial c(w,y)/\partial w_i$.

最后一种方法，我们分析这个命题背后的基本经济学直觉。如果我们在某个成本最小化之处生产，价格 w_1 上升后将产生一个直接效应，即花费在第一种要素身上的成本会增加。还有一个间接效应，即我们希望改变要素组合。但是既然我们已经在成本最小化之处生产，任何微小的变动必定导致额外利润为零。

5.5 约束最优化的包络定理

Shephard 引理是包络定理的另外一个例子。然而，在这种情形下，我们运用的包络定理的版本是适用于约束最优化问题的版本。这种情形的证明请参见第 27 章。

考虑一个一般的约束最优化问题

$$M(a)=\max_{x_1,x_2} g(x_1,x_2,a)$$

$$\text{使得 } h(x_1,x_2,a)=0.$$

在成本函数的情形下，$g(x_1,x_2,a)=w_1x_1+w_2x_2$，$h(x_1,x_2,a)=f(x_1,x_2)-y$，参数 a 可以为一种要素的价格。在这种情形下，需要将目标函数前面的 max 换成 min，因为这是个最小化问题。

曹乾（东南大学 caoqianseu@163.com）

这个问题的拉格朗日函数为

$$L = g(x_1, x_2, a) - \lambda h(x_1, x_2, a),$$

一阶条件为

$$\begin{aligned} \frac{\partial g}{\partial x_1} - \lambda \frac{\partial h}{\partial x_1} &= 0 \\ \frac{\partial g}{\partial x_2} - \lambda \frac{\partial h}{\partial x_2} &= 0 \\ h(x_1, x_2, a) &= 0. \end{aligned} \tag{5.1}$$

这些一阶条件确定了最优选择函数 $(x_1(a), x_2(a))$，而这个函数反过来又确定了最大值函数

$$M(a) \equiv g(x_1(a), x_2(a), a). \tag{5.2}$$

包络定理为我们提供了最大化问题中的值函数关于参数导数的表达式。具体地说，这个表达式为

$$\begin{aligned} \frac{dM(a)}{da} &= \frac{\partial L(x,a)}{\partial a}\Big|_{x=x(a)} \\ &= \frac{\partial g(x_1, x_2, a)}{\partial a}\Big|_{x_i = x_i(a)} - \lambda \frac{\partial h(x_1, x_2, a)}{\partial a}\Big|_{x_i = x_i(a)} \end{aligned}$$

和以前一样，对偏导数进行解释时需要特别小心：它们是 g 和 h 关于 a 的导数，但要在 x_1 和 x_2 的最优选择处计算。包络定理的证明请参见第 27 章。此处我们只是将它用于分析成本最小化问题。

在这个问题中，参数 a 可以为某种要素的价格 w_i，最优值函数 $M(a)$ 为成本函数 $c(w, y)$。包络定理断言

$$\frac{\partial c(w,y)}{\partial w_i} = \frac{\partial L}{\partial w_i} = x_i\Big|_{x_i = x_i(w,y)} = x_i(w, y),$$

这正是 Shephard 引理。

例子：边际成本再分析

包络定理的另外一个应用是，考虑成本函数关于 y 的导数。根据包络定理可知，这个导数等于拉格朗日函数关于 y 的导数。成本最小化的拉格朗日函数为

$$L = w_1 x_1 + w_2 x_2 - \lambda[f(x_1, x_2) - y].$$

因此，

$$\frac{\partial c(w_1, w_2, y)}{\partial y} = \lambda.$$

换句话说，成本最小化问题中的拉格朗日乘子就是边际成本。

5.6 使用成本函数进行比较静态分析

在前面我们已经证明，从成本最小化问题的定义可以推导出成本函数的若干性质；在上面我们已经证明，条件要素需求函数是成本函数的导数。因此，我们发现的成本函数的那些性质将体现在对它的导数（即条件要素需求函数）的限制上。这些限制我们已经使用其他方法证明过，但是使用成本函数证明这些限制特别漂亮。

我们逐一审查这些限制。

1）成本函数对于要素价格是非增的。这一结论可直接从 $\partial c(w,y)/\partial w_i = x_i(w,y) \geq 0$ 推导出。

2）成本函数关于 w 是一次齐次的。因此，成本函数的导数，即要素需求函数，关于 w 是零次齐次的。（参见第 26 章）。

3）成本函数关于 w 是凹的。因此，成本函数的二阶导数，即要素需求函数的一阶导数组成的矩阵，是对称的负半定矩阵。这个结论是从成本最小化行为推出的，但它不是那么明显。它有几个应用。

a)交叉价格效应是对称的。即

$$\frac{\partial x_i(w,y)}{\partial w_j} = \frac{\partial^2 c(w,y)}{\partial w_j \partial w_i} = \frac{\partial^2 c(w,y)}{\partial w_i \partial w_j} = \frac{\partial x_j(w,y)}{\partial w_i}.$$

b)自价格效应是非正的。大致来说，条件要素需求曲线是向下倾斜的。这个结论可从 $\partial x_i(w,y)/\partial w_i = \partial^2 c(w,y)/\partial w_i^2 \leq 0$ 推导出，这个式子中的最后一个不等式成立的原因是，负半定矩阵主对角线上的元素必定是非正的。

c) 要素需求变动向量和要素价格变动向量的运动方向“相反”，即 $dwdx \leq 0$。

注意，既然成本函数为凹的结论仅使用成本最小化的假设就可推出，因此要素需求函数一阶导数矩阵的负半定性仅使用成本最小化的假设即可推出，不需要对技术的结构施加任何限制。

注释

成本函数的性质是由几个学者发展出的，但是最系统的研究见 Shephard（1953）和 Shephard（1970）。比较全面的研究也可参见 Diewert（1974）。本章的处理方法主要借鉴了 McFadden（1978）

习题

5.1 某企业有两个工厂。这两个工厂的成本函数分别为$c_1(y_1)=y_1^2$和$c_2(y_2)=y_2^2$。要素的价格都是固定不变的，因此分析时可不用考虑。计算该企业的成本函数。

5.2 某企业有两个工厂。这两个工厂的成本函数分别为$c_1(y_1)=3y_1^2$和$c_2(y_2)=y_2^2$。计算该企业的成本函数。

5.3 某企业的生产函数为$f(x_1,x_2,x_3,x_4)=\min\{2x_1+x_2,x_3+2x_4\}$。求该技术的成本函数。求要素 1 和 2 的条件需求函数，注意要将它们表达为要素价格(w_1,w_2,w_3,w_4)和产量y的函数。

5.4 某企业的生产函数为$f(x_1,x_2)=\min\{2x_1+x_2,x_1+2x_2\}$。求该技术的成本函数。求要素 1 和 2 的条件需求函数，注意要将它们表达为要素价格(w_1,w_2)和产量y的函数。

5.5 某企业的生产函数为$f(x_1,x_2)=\max\{x_1,x_2\}$。该企业的必要投入集为凸还是非凸的？求要素 1 的条件要素需求函数。求成本函数。

5.6 某企业的条件要素需求函数为

$$x_1=1+3w_1^{-1/2}w_2^{a}$$

$$x_2=1+bw_1^{1/2}w_2^{c}$$

为简单起见，已将产量设定为 1。求参数a,b,c的值并解释原因。

5.7 某企业的生产函数为$y=x_1x_2$。如果$w_1=w_2=1$时的最小生产成本等于 4，求产量y。

5.8 某企业的成本函数为

$$c(y)=\begin{cases}y^2+1, & y>0\\ 0 & y=0\end{cases}$$

令p为产品的价格，并且令要素的价格都是固定不变的。计算$p=2$时企业的产量。计算$p=1$时的企业产量。求该企业的利润函数。（提示：务必小心！）

5.9 某个典型的硅谷芯片生产企业的成本函数为$c(y)$，边际成本是递增的。在该企业生产出来的芯片中，有$(1-a)$比例的芯片是有缺陷的无法出售。芯片的市场价格为p，该市场为完全竞争的。

（a）计算利润关于参数a的导数并指出它的符号为正还是负。

（b）计算产量关于参数a的导数并指出它的符号为正还是负。

（c）假设有n个相同生产企业，令$D(p)$为需求函数，令$p(a)$为竞争均衡价格。计算(dp/da)并指出它的符号为正还是负。

5.10 假设某企业在产品市场和要素市场上都是完全竞争的。假设每种要素要素的价格随需求量增加而上升，令dw_i表示要素i的价格上升量。在什么样的条件下，利润最大化的产量是下降的？

5.11 企业 A 使用 4 种要素生产 1 种产品，生产函数为

$$f(x_1, x_2, x_3, x_4) = \min\{x_1, x_2\} + \min\{x_3, x_4\}.$$

（a）求要素价格向量 $w = (1,2,3,4)$ 时，生产 1 单位产品的条件要素需求向量。

（b）计算成本函数。

（c）该技术的规模报酬情形是什么样的（递减、不变还是递增）？

（d）另外一个企业 B 的生产函数为 $f(x_1, x_2, x_3, x_4) = \min\{x_1 + x_2, x_3 + x_4\}$。求要素价格向量 $w = (1,2,3,4)$ 时，生产 1 单位产品的条件要素需求向量。

（e）计算 B 企业的成本函数。

（f）B 企业的技术的规模报酬情形是怎样的（递减、不变还是递增）？

5.12 生产要素 i 称为劣等的（inferior），如果该要素的条件需求随着产量的增加而下降，即 $\partial x_i(w, y) / \partial y < 0$。

（a）画图说明劣等要素是可能存在的。

（b）证明如果某技术是规模报酬不变的，则所有要素都不是劣等的。

（c）证明如果边际成本随着某种要素的价格上升而下降，则该要素必定是劣等的。

5.13 某个追求利润最大化的企业，它生产的产品在完全竞争市场上出售。我们观测到当产品价格上升时，企业雇佣更多熟练工和更少的熟练工。现在非熟练工联合起来迫使企业提高他们的工资，他们的努力成功了。假设所有其他价格维持不变。

（a）企业对非熟练工的需求有什么样的变化？

（b）企业产品供给有什么变化？

5.14 假设你有下列一些时间序列观测数据：产量的变动 Δy，成本的变动 Δc，要素价格的变动 Δw，要素 i 的需求水平 x_i（其中 $i = 1, ..., n$）。请问你应该如何估计出每个时期的边际成本 $\partial c(w, y) / \partial y$？

5.15 计算下列技术的成本函数

$$V(y) = \{(x_1, x_2, x_3) : x_1 + \min(x_2, x_3) \geq 3y\}.$$

5.16 对于下列每个成本函数，确定它是否为一次齐次的、单调的、凹的和（或）连续的。如果是，推导出相应的生产函数。

（a）$c(w, y) = y^{1/2}(w_1 w_2)^{3/4}$

（b）$c(w, y) = y(w_1 + \sqrt{w_1 w_2} + w_2)$

（c）$c(w, y) = y(w_1 e^{-w_1} + w_2)$

（d）$c(w, y) = y(w_1 - \sqrt{w_1 w_2} + w_2)$

（e） $c(w,y)=(y+\frac{1}{y})\sqrt{w_1w_2}$

5.17 某个企业的必要投入集为$V(y)=\{x>0: ax_1+bx_2\geq y^2\}$。

（a）计算生产函数。

（b）计算条件要素需求。

（c）计算成本函数。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 6 章：对偶

曹乾　译

(东南大学　caoqianseu@163.com)

6 对偶

在上一章，我们分析了成本函数的性质。成本函数衡量生产既定产量的最小成本。给定任何技术，可以直接推导出成本函数，至少在理论上是这样的：我们只要求解利润最小化问题即可。

在本章，我们将证明这个过程可以反转过来。给定一个成本函数，我们可以“解出”产生这个成本函数的技术。这意味着成本函数和生产函数在本质上包含着同样的信息。用生产函数性质定义的任何概念，都对应着使用成本函数性质定义的一个“对偶”概念，反过来也是这样。这个一般的结论称为**对偶**（duality）原理。对偶原理有着若干重要的结果，我们将在本章进行分析。

同一种经济行为有着不同的表达方法，这些表达方法之间的对偶，在研究消费者行为理论、福利经济学和经济学其他领域中的问题非常有用。很多看上去难以理解的关系，如果使用对偶的工具进行分析，则会变得非常简单甚至平凡（trivial）。

6.1 对偶

在第 4 章，我们刻画了集合$VO(y)$，我们说它是真实必要投入集$V(y)$的“外边界”。给定数据(w^t, x^t, y^t)，$VO(y)$的定义为

$$VO(y) = \{x : w^t x \geq w^t x^t \text{ 对于所有满足 } y^t \leq y \text{ 的 } t \text{ 成立}\}.$$

容易验证$VO(y)$是一个闭的、单调的而且凸的技术。更进一步地说，正如我们在第 4 章指出的，$VO(y)$包含可以产生数据(w^t, x^t, y^t)的任何技术，其中$t = 1,...,T$.

如果我们观测在不同要素价格下的企业选择，当这些观察值越来越多时，似乎$VO(y)$在某种意义上应该“逼近”真实的必要投入集。为了更准确说明，令要素价格在所有可能的价格向量$w \geq 0$上变动，则VO的一般化形式为

$$V^*(y) = \{x : wx \geq wx(w, y) = c(w, y) \text{ 对于所有 } w \geq 0 \text{ 成立}\}.$$

$V^*(y)$和真实必要投入集$V(y)$有什么关系？当然，正如第 4 章指出的，$V^*(y)$包含$V(y)$。一般来说，$V^*(y)$严格包含$V(y)$。例如，在图 6.1A 中，我们看到阴影区域不能从$V^*(y)$中排除出去，这是因为这个区域中的点满足条件$wx \geq c(w, y)$。

图 6.1B 的情形也类似。成本函数仅包含与$V(y)$在经济上相关区域的信息，即，那些真正为成本最小化问题的解的要素束，也就是指真正的条件要素需求。

然而，假设我们最初的生产技术是凸的而且是单调的。在这种情形下，$V^*(y)$等于

$V(y)$。这是因为，在凸且单调的情形下，$V(y)$ 边界上的每个点在某个价格向量 $w \geq 0$ 下，都是成本最小化的素需求。因此，满足 $wx \geq c(w,y)$ 的点集，其中 $w \geq 0$，正是必要投入集。更正式地：

$V(y)$ 何时等于 $V^*(y)$。假设 $V(y)$ 是一个正则、凸且单调的技术，则 $V^*(y)=V(y)$.

证明。（只给出证明的梗概）我们已经知道 $V^*(y)$ 包含 $V(y)$，所以我们只需要证明 $V(y)$ 包含 $V^*(y)$ 即可，也就是说我们要证明如果 x 在 $V^*(y)$ 中，则 x 必在 $V(y)$ 中。假设 x 不是 $V(y)$ 的元素。于是由于 $V(y)$ 是一个闭且凸的集合，而且它满足单调性假设，我们可以运用分离超平面定理（参加第 26 章），从而找到一个向量 $w^* \geq 0$ 使得 $w^*x < w^*z$ 对于 $V(y)$ 中的所有 z 成立。令 z^* 为 $V(y)$ 中的一点，该点使得当价格为 w^* 时的生产成本最小。于是，特别地，我们有 $w^*x < w^*z^* = c(w^*, y)$。但是由 $V^*(y)$ 的定义可知，x 不可能在 $V^*(y)$ 中。

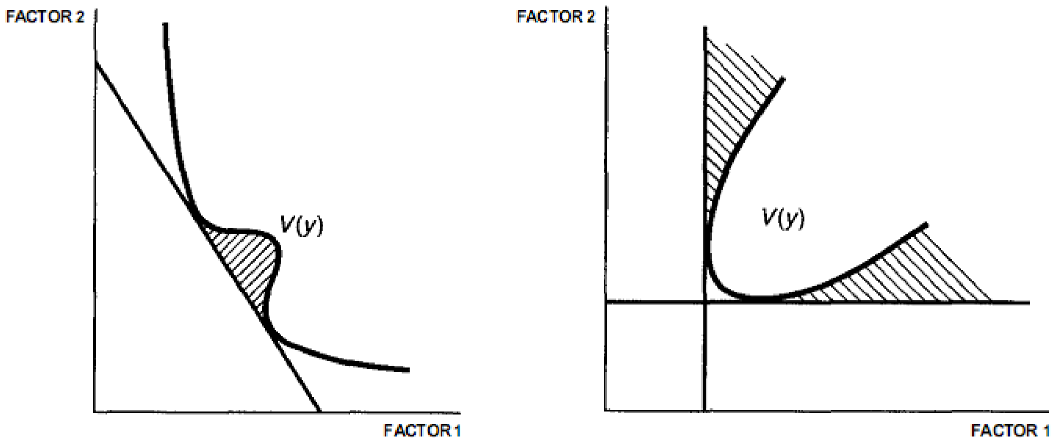


图 6.1：$V(y)$ 与 $V^*(y)$ 的关系。一般来说，$V^*(y)$ 严格包含 $V(y)$。

这个命题表明，如果原来的技术是凸而且是单调的，则与生产技术相伴的成本函数可被用来完全重构原来的生产技术。如果我们知道在每个可能的价格向量 w 下，最小成本的生产之处，那么我们就可以知道企业可使用的全部技术选择集。

在技术为凸且单调的情形下，上述结果是令人满意的。但是对于表现不怎么良好的情形，结果会怎样？假设我们的生产技术 $V(y)$ 可能是非凸的。我们找到它的成本函数 $c(w,y)$ 并且生成 $V^*(y)$。从上面的结果我们知道 $V^*(y)$ 未必等于 $V(y)$，除非 $V(y)$ 正好是凸的而且是单调的。然而，假设我们定义

$$c^*(w,y) = \min\ wx$$

$$\text{使得 } x \text{ 在 } V^*(y) \text{ 中.}$$

$c^*(w,y)$ 和 $c(w,y)$ 有什么关系？

$c(w,y)$ 何时等于 $c^*(w,y)$。从 $c^*(w,y)$ 的定义可知，$c^*(w,y)=c(w,y)$.

证明。容易看出$c^*(w,y)\le c(w,y)$；由于$V^*(y)$总是包含$V(y)$，$V^*(y)$中的最小成本束必定至少与$V(y)$中的最小成本束一样小。假设当价格为w'时，$V^*(y)$中的最小成本束具有$w'x'=c^*(w',x)<c(w',x)$的性质。但是这是不可能发生的，因为根据$V^*(y)$的定义可知，$w'x'\ge c(w',x)$。■

这个命题表明，生产技术$V(y)$的成本函数与它的凸版本的技术$V^*(y)$的成本函数是相同的。在这个意义上，从经济的观点来看，必要投入集为凸的假设不具有很强的限制性。

我们总结一下我们的讨论：

（1）给定一个成本函数，我们可以定义一个必要投入集$V^*(y)$。

（2）如果原来的技术是凸而且单调的，构建出的技术将和原来的技术相同。

（3）如果原来的技术是非凸的或非单调的，构建出的必要投入集是原来必要投入集的凸化和单调化的版本，更为重要的是，构建出的技术和原来技术的成本函数是相同的。

使用生产的对偶原理，我们可以将以上的三点简洁地总结为：一个企业的成本函数总结了它的生产技术的所有经济相关的信息。

6.2 成本函数的充分条件

在上一节我们已经看到，成本函数总结了企业生产技术的所有经济相关信息。在以前的章节我们已经知道，所有的成本函数都是价格的非减、齐次、凹的和连续的函数。问题产生了：假设给你一个关于价格的非减、齐次、凹的和连续的函数，它一定是某个生产技术的成本函数吗？

这个问题的另外一种表达方法是：上一章中列举的成本函数的性质是否是成本最小化行为的全部性质？给定一个具有这些性质的函数，它一定是由某个技术产生的吗？答案是肯定的，下列命题表明如何构建这样的技术。

$\phi(w,y)$ **何时是成本函数**。令$\phi(w,y)$是一个满足下列性质的可微函数：

1）$\phi(tw,y)=t\phi(w,y)$对于所有$t\ge 0$成立；

2）$\phi(w,y)\ge 0$，其中$w\ge 0$和$y\ge 0$；

3）$\phi(w',y)\ge\phi(w,y)$，其中$w'\ge w$；

4）$\phi(w,y)$是w的凹函数。

那么，$\phi(w,y)$是技术$V^*(y)=\{x\ge 0: wx\ge\phi(w,y)$（其中$w\ge 0$）的成本函数。

证明。给定某个$w\ge 0$，我们定义

$$x(w,y)=(\frac{\partial\phi(w,y)}{\partial w_1},...,\frac{\partial\phi(w,y)}{\partial w_n})$$

注意到$\phi(w,y)$是w的一次齐次函数，欧拉法则（Euler’s law）意味着$\phi(w,y)$可以写为

$$\phi(w,y)=\sum_{i=1}^{n}w_i\frac{\partial\phi(w,y)}{\partial w_1}=wx(w,y).$$

（欧拉法则请参见第 26 章）。注意，$\phi(w,y)$的单调性意味着$x(w,y)\geq 0$。

我们需要证明的是，对于任何给定的$w'\geq 0$，$x(w',y)$对于$V^*(y)$中的所有x来说，都使得$w'x$最小：

$$\phi(w',y)=w'x(w',y)\leq w'x \text{ 对于 } V^*(y) \text{ 中的所有 } x \text{ 成立.}$$

首先，我们证明$x(w',y)$是可行的，也就是说证明$x(w',y)$在$V^*(y)$中。由于$\phi(w,y)$是w的凹函数，我们有

$$\phi(w',y)\leq\phi(w,y)+D\phi(w,y)(w'-w)$$

对于所有$w\geq 0$成立。（请参考第 27 章。）

使用前面的欧拉法则，上式化简为

$$\phi(w',y)\leq w'x(w,y) \text{ 对于所有 } w\geq 0 \text{ 成立.}$$

根据$V^*(y)$的定义可知，$x(w',y)$在$V^*(y)$中。

接下来我们将证明$x(w',y)$对于$V^*(y)$中的所有x来说使得wx最小。如果x在$V^*(y)$中，则根据定义可知，它必须满足

$$wx\geq\phi(w,y).$$

但是根据欧拉法则可知，

$$\phi(w,y)=wx(w,y).$$

上面两个式子意味着

$$wx\geq wx(w,y)$$

对于$V^*(y)$中的所有x成立。证毕。 ■

6.3 需求函数

上一节证明的命题提出了一个有趣的问题，假设给你一组函数，这组函数满足上一章描述的条件要素需求函数的性质，也就是说，它们是价格的零次齐次函数，而且

$$\left(\frac{\partial g_i(w,y)}{\partial w_j}\right)$$

是一个对称的负半定矩阵。这些函数一定是某个生产技术的要素需求函数吗？

下面我们试图应用上述命题。首先，我们构建一个候选的成本函数：

$$\phi(w,y)=\sum_{i=1}^{n}w_i g_i(w,y).$$

接下来我们检验这个函数是否满足我们刚证明的那个命题的性质。

1）$\phi(w,y)$ 是 w 的一次齐次函数吗？为了检验这个问题，我们分析 $\phi(tw,y)=\sum_i tw_i g_i(tw,y)$. 由于根据假设，$g_i(w,y)$ 是零次齐次的，$g_i(tw,y)=g_i(w,y)$。因此

$$\phi(tw,y)=t\sum_{i=1}^{n}w_i g_i(w,y)=t\phi(w,y).$$

2）对于 $w\geq 0$，$\phi(w,y)\geq 0$ 吗？由于 $g_i(w,y)\geq 0$，答案是肯定的。

3）$\phi(w,y)$ 是 w_i 的非减函数吗？使用求导的乘积法则，我们计算

$$\frac{\partial\phi(w,y)}{\partial w_i}=g_i(w,y)+\sum_{j=1}^{n}w_j\frac{\partial g_j(w,y)}{\partial w_i}=g_i(w,y)+\sum_{j=1}^{n}w_j\frac{\partial g_i(w,y)}{\partial w_j}.$$

由于 $g_i(w,y)$ 是零次齐次的，最后一项为零，$g_i(w,y)$ 显然大于或等于零。

4）最后，$\phi(w,y)$ 是 w 的凹函数吗？为了检验这个问题，我们对 $\phi(w,y)$ 微分两次，得到

$$\left(\frac{\partial^2\phi}{\partial w_i\partial w_j}\right)=\left(\frac{\partial g_i(w,y)}{\partial w_j}\right).$$

为了证明凹形，我们希望这些矩阵是对称的和负半定的，根据假设可知，这些都是对的。

因此，可以应用上一节证明过的命题，由此可知，存在技术 $V^*(y)$ 使得它生成的 $g_i(w,y)$ 是它的条件要素需求。这意味着齐次性和负半定构成了成本最小化行为对需求函数施加的所有限制。

当然，类似的结果对于利润函数和（非条件）需求函数和供给函数也成立。如果六年函数满足第 3 章描述的限制，或者等价地，如果需求函数和供给函数满足第 3 章描述的限制，那么存在一种技术使得它生成这个利润函数或这些需求函数以及供给函数。

例子：对偶映射的应用

假设我们的成本函数为 $c(w,y)=yw_1^a w_2^{1-a}$。我们如何解出与该函数相伴的技术？根据

可导性质

$$x_1(w,y) = ayw_1^{a-1}w_2^{1-a} = ay(\frac{w_2}{w_1})^{1-a}$$

$$x_2(w,y) = (1-a)yw_1^{a}w_2^{-a} = (1-a)y(\frac{w_2}{w_1})^{-a}.$$

我们想从这两个式子中消去 w_2/w_1，从而将 y 表示成 x_1 和 x_2 的函数。将这两个式子变形可得

$$\frac{w_2}{w_1} = \left(\frac{x_1}{ay}\right)^{\frac{1}{1-a}}$$

$$\frac{w_2}{w_1} = \left(\frac{x_2}{(1-a)y}\right)^{-\frac{1}{a}}$$

令这两个式子相等，并且左右两端各取 $-a(1-a)$ 次方，可得

$$\frac{x_1^{-a}}{a^{-a}y^{-a}} = \frac{x_2^{1-a}}{(1-a)^{1-a}y^{1-a}},$$

或

$$\left[a^a(1-a)^{1-a}\right]y = x_1^a x_2^{1-a}.$$

这正是柯布-道格拉斯技术。

例子：规模报酬不变和成本函数

由于成本函数告诉我们与技术有关的所有经济相关信息，我们可以使用施加在技术身上的限制，来解释施加在成本身上的各种限制。在第 5 章，我们已表明，如果技术是规模报酬不变的，则成本函数的形式为 $c(w)y$ 。此处我们将证明反过来说也是成立的。

规模报酬不变。令 $V(y)$ 是凸且单调的；如果 $c(w,y)$ 可以写成 $yc(w)$，则 $V(y)$ 必定是规模报酬不变的。

证明。使用凸性和单调性以及假设的成本函数形式，我们知道

$$V(y) = V^*(y) = \{x : wx \geq yc(w) \text{ 对于所有 } w \geq 0 \text{ 成立}\}.$$

我们想证明，如果 x 在 $V^*(y)$ 中，则 tx 在 $V^*(ty)$ 中。如果 x 在 $V^*(y)$ 中，我们知道 $wx \geq yc(w)$ 对于所有 $w \geq 0$ 成立。将这个式子的两侧同乘以 t 可得：$wtx \geq tyc(w)$ 对于所有 $w \geq 0$ 成立。但这意味着 tx 在 $V^*(ty)$ 中。■

例子：规模弹性和成本函数

给定一个生产函数 $f(x)$，我们可以考虑规模报酬的局部测量，这称为规模弹性（elasticity of scale）：

$$e(x)=\frac{df(tx)}{dt}\frac{t}{f(x)}\Big|_{t=1}$$

这个定义我们曾经在第 1 章给出。当$e(x)$ 小于、等于或大于 1 时，技术分别呈现局部规模报酬递减、不变或递增的特征。

给定要素价格向量，我们可以计算企业的成本函数$c(w,y)$。令x^*是在(w,y)时的成本最小的要素束。于是我们可以根据下列式子计算$e(x^*)$：

$$e(x^*)=\frac{c(w,y)/y}{\partial c(w,y)/\partial y}=\frac{AC(y)}{MC(y)}.$$

为了看清这一点，我们对前面定义的$e(x)$ 微分：

$$e(x^*)=\frac{\sum_{i=1}^{n}\frac{\partial f(x^*)}{\partial x_i}x_i^*}{f(x^*)}.$$

由于 x^* 使得成本最小，它满足一阶条件 $w_i=\lambda\frac{\partial f(x^*)}{\partial x_i}$。而且，根据包络定理可知，$\lambda=\partial c(w,y)/\partial y$（参见第 5 章），因此

$$e(x^*)=\frac{\sum_{i=1}^{n}w_i x_i^*}{\lambda f(x^*)}=\frac{c(w,y)/f(x^*)}{\partial c(w,y)/\partial y}=\frac{AC(y)}{MC(y)}.$$

6.4 对偶的几何图形

在这一章我们将分析企业的技术（由生产函数刻画）和它的经济行为（由成本函数刻画）之间的几何关系。

在图 6.2 中，我们画出了企业的等产量线和在该产量水平 y 时的等成本曲线。这条等成本曲线在点(w_1^*,w_2^*)的斜率为

$$\frac{dw_2(w_1^*)}{dw_1}=-\frac{\frac{\partial c(w^*,y)}{\partial w_1}}{\frac{\partial c(w^*,y)}{\partial w_2}}=-\frac{x_1(w^*,y)}{x_2(w^*,y)}$$

另一方面，等产量曲线的定义为

$$f(x)\equiv y.$$

等产量曲线在点 x^* 的斜率为

$$\frac{dx_2(x_1^*)}{dx_1} = -\frac{\dfrac{\partial f(x^*)}{\partial x_1}}{\dfrac{\partial f(x^*)}{\partial x_2}}.$$

现在，如果 (x_1^*, x_2^*) 是价格为 (w_1^*, w_2^*) 时的一个成本最小化的点，那么它必定满足一阶条件

$$\frac{w_1^*}{w_2^*} = \frac{\dfrac{\partial f(x^*)}{\partial x_1}}{\dfrac{\partial f(x^*)}{\partial x_2}}.$$

注意这个美妙的**对偶**：等产量曲线的斜率给出了要素价格之比，而等成本曲线给出的是要素使用数量之比。

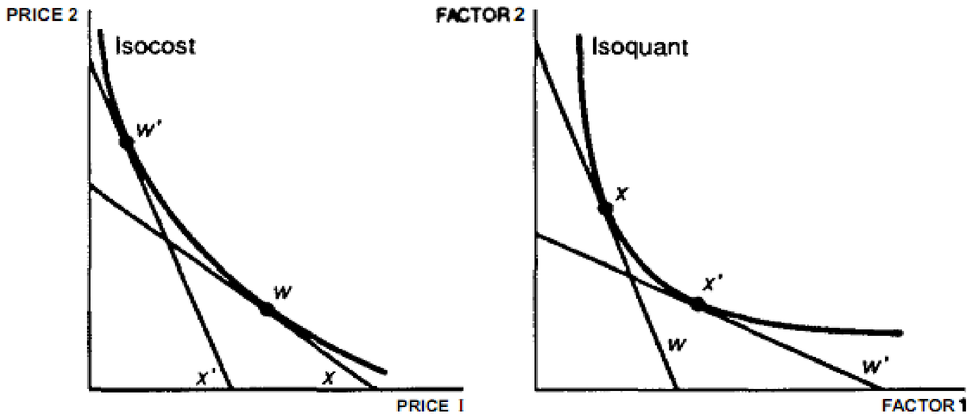


图 6.2：等产量曲线的曲率和等成本曲线的曲率。等产量曲线的曲率越大，等成本曲线的曲率越小。

等产量曲线的**曲率**（curvature）和等成本曲线的曲率有什么关系？可以证明它们的曲率是反相关的：如果等成本曲线非常弯曲，则等产量曲线将非常平坦，反之亦然。为了看清这一点，考虑等产量曲线的某个具体的点 (w_1, w_2)，然后将这个点沿着这条曲线移动到较远的距离，到达另外一个点 (w_1', w_2')。假设我们发现等成本曲线的斜率变化并不大，也就是说等成本曲线的曲率较小。由于等成本曲线的斜率给出的是要素需求量之比，这意味着成本最小化要素束之间是非常相似的。借助图 6.2 我们可以看出这表示等产量曲线是急剧弯曲的。在极端情形下，我们发现里昂惕夫技术的成本函数是一个线性函数，而且 L 形成本函数都对应着线性技术。

例子：生产函数，成本函数和条件要素需求

假设我们有一个平滑且凸的等产量曲线。那么等成本曲线也是凸且平滑的，条件要素需求曲线也是良好表现的，如图 6.3 所示。

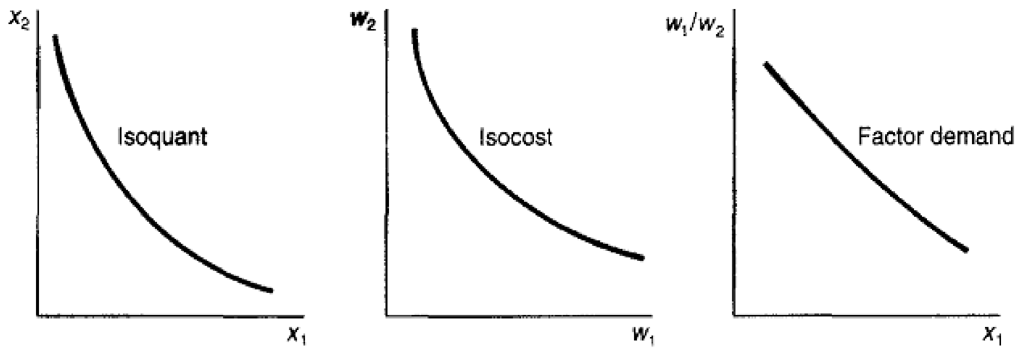


图 6.3：技术，成本和需求。等产量曲线为平滑且凸的情形。

假设等产量曲线有水平段（flat spot），因此在某组要素价格水平下，不存在唯一的要素需求束，也就是说要素需求束有很多个。于是在这样的要素价格水平下，等成本曲线必定是不可微分的，条件要素需求函数是多值（multivalued）函数，如图 6.4 所示。

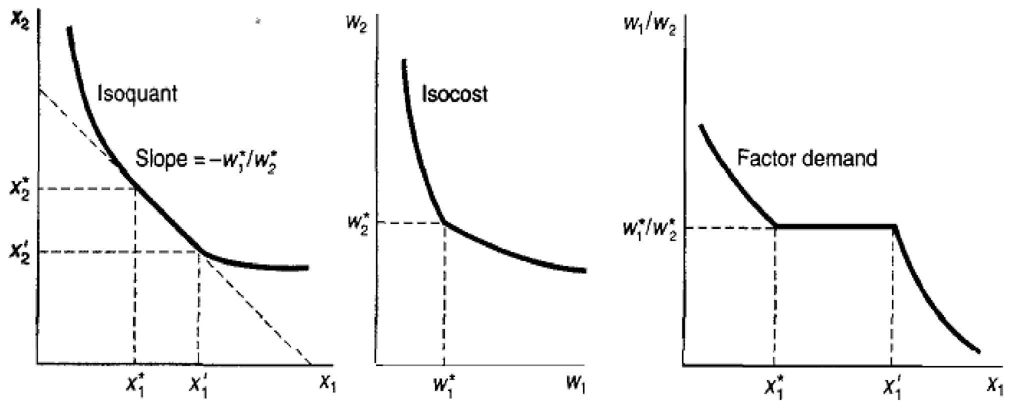


图 6.4：技术，成本和需求。等产量线有水平段的情形。等成本曲线在价格等于水平段斜率之处存在着弯折（kink）。在这样的要素价格水平下，成本最小束有好多个。

假设等产量曲线在某个点有个弯折，那么对于某个价格区域，要素需求束是固定一点。这意味着等成本曲线必定有水平段，如图 6.5 所示。

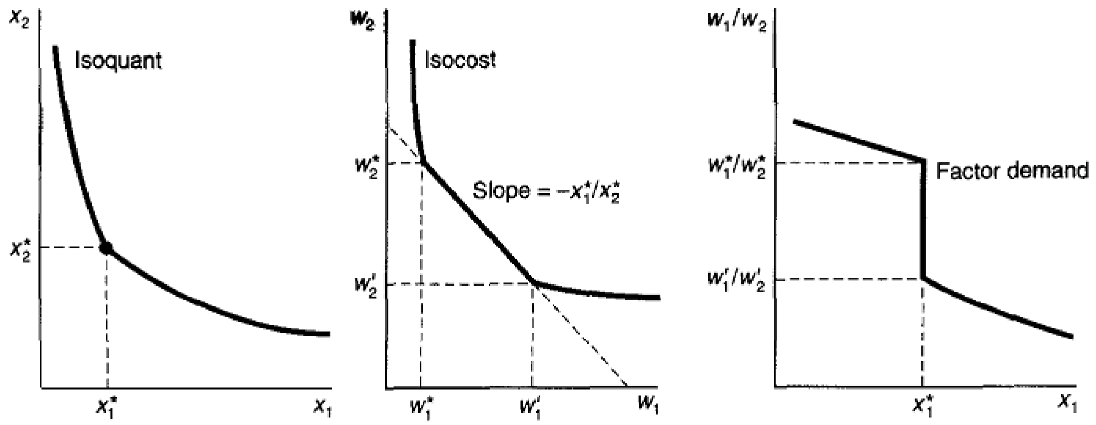


图 6.5：技术，成本和需求。等产量线有弯折的情形。等成本曲线有水平段，对于同一个要素束有多组价格都可以使得成本最小化。

假设等产量曲线在某个区域是拟凸的。那么等成本曲线在某点出现了弯折，条件要素需求是不连续的和多值的，如图 6.6 所示。比较图 6.4 和 6.6，注意这个技术的成本函数与这个技术在凸化后的成本函数的区别。

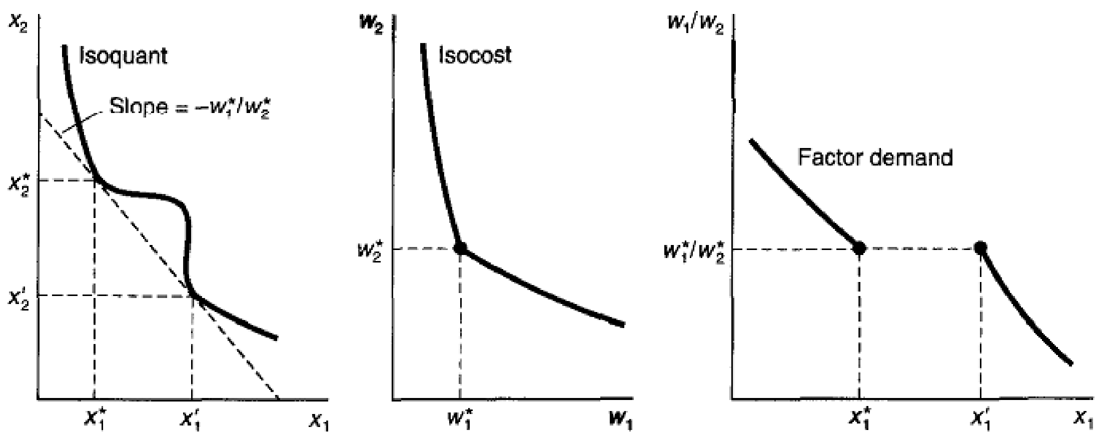


图 6.6：技术、成本与需求。等产量线非凸的情形。它的等成本线看上去和有水平段的等产量线的等成本线一样，但是要素需求函数现在是不连续的。

6.5 对偶的用途

生产技术与它相伴的成本函数之间存在着对偶关系，这个事实对于生产经济学有几个重要的影响。我们曾经顺带着涉及到了这些影响中的一部分，此处需要将它们再总结一下。

首先，拥有描述技术性质的两种不同方法在理论上非常方便，因为某些类型的论断如果借助生产函数或利润函数将非常容易证明，而如果使用技术的直接表示方法进行证明则比

较困难。例如，考虑前面的那个例子，在该例中，价格波动比价格固定在期望值上能带来更高的利润。这只是利润函数凸性的微不足道的一个推论。然而，如果我们使用技术直接表示方法进行证明的话，就比较困难一些。

第二，在均衡分析中，行为的对偶（例如成本函数和利润函数之间的对偶）非常有用，因为它们已将行为假设包含在方程说明中。例如，如果我们想分析某种特别税收政策对企业利润的影响方式，我们可以先分析这种税如何影响企业面的的价格，然后看看这些特定的价格变动如何影响利润函数的。我们不需要求解任何最大化问题，因为它们在利润函数的界定中已经 “被解决了”。

第三，成本函数和利润函数的全部性质就是齐次性、单调性和曲率性（curvature），这个事实使我们更容易验证企业行为的某些类别的命题。我们只要检验一下某个特定性质是不是成本或利润函数的齐次性、单调性或曲率性的结果即可。如果不是，那么这个性质就不能从最大化行为中推导出来。

第四，利润和成本函数可用三个相对简单的数学条件刻画，这样我们就可以比较容易地将技术表达为参数形式。例如，为了完全描述某个技术，我们所要做的全部工作，就是规定函数关于要素价格是连续的、单调的和凹的即可。这要比规定生产函数简单得多。在实践计算或计量经济工作中，这样的参数表达形式非常有用。

第五，对偶表示方法在经济计量工作中的表现通常更让人满意。原因在于进入对偶关系的变量——价格变量——通常视为企业选择问题的外生变量。如果要素市场是完全竞争的，那么经济学家认为企业将要素价格作为外生给定的，它要做的就是选择投入水平，因此在统计学上，要素价格因素可能和生产关系的误差项不相关。从统计学的观点看，这个性质非常让人满意。我们将在 12 章进一步分析这个问题。

注释

Shephard（1953）第一个首先严格证明了成本函数和生产函数的基本对偶关系。Diewert（1954）则综述了这一主题的历史进展并且提供了更一般的现代处理方法。

习题

6.1 成本函数为 $c(w_1, w_2, y) = \min\{w_1, w_2\}y$ 。求生产函数和条件要素需求。

6.2 成本函数为 $c(w_1, w_2, y) = y\{w_1 + w_2\}$ 。求条件要素需求和生产函数。

6.3 成本函数为 $c(w_1, w_2, y) = w_1^a w_2^b y$ 。我们对 a 和 b 知道些什么？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 7 章：效用最大化

曹乾　译

(东南大学　caoqianseu@163.com)

7 效用最大化

在本章我们开始分析消费者行为。在竞争性厂商理论中，供给和需求函数都是从利润最大化行为模型和基本的技术约束条件中推出。在消费者理论中，我们将根据效用最大化行为和基本经济约束条件推出需求函数。

7.1 消费者偏好

考虑某消费者面对集合 X 中某些可能的消费束的情形，集合 X 称为他的**消费集**（consumption set）。在本书中，我们通常假定 X 是 $\mathbb{R}^K$ 中的非负象限，但也可以使用更具体的消费集。例如，为维持消费者生存所必需的那些消费束组成的集合。我们总是假定 X 是闭且凸的集合。

消费者对于 X 中的消费束存在着偏好。当我们写出 $x \succeq y$ 时，我们的意思是说“消费者认为消费束 x 至少和消费束 y 一样好”。我们希望偏好能对消费束排序。因此，我们需要假定偏好满足某些标准的属性。

完备性（complete）。对于 X 中所有的 x 和 y，或者 $x \succeq y$ 或者 $y \succeq x$（或二者都成立）。

反身性（reflective）。对于 X 中所有的 x，$x \succeq x$。

传递性（transitive）。对于 X 中所有的 x，y 和 z，若 $x \succeq y$ 且 $y \succeq z$，则 $x \succeq z$。

第一个假设只是说任意两个消费束之间都可以进行比较；第二个假设是微不足道的；第三个假设是分析偏好最大化（maximization）时所必须的。如果偏好不是传递的，则可能存在这样的消费束集，它们不含有最优元素。

给定描述“弱偏好”的排序 $\succeq$，我们可以定义严格偏好 $\succ$：$x \succ y$ 的意思是不是 $y \succeq x$。我们将 $x \succ y$ 读为“ x 被严格偏好于 y ”。类似地，我们可以定义无差异~：$x \sim y$ 当且仅当 $x \succeq y$ 和 $y \succeq x$ 同时成立。

我们通常还希望对消费者的偏好作出其他的假设；例如。

连续性。对于 X 中的所有 y，集合 $\{x : x \succeq y\}$ 和 $\{x : x \preceq y\}$ 都是闭集。由此可知，$\{x : x \succ y\}$ 和 $\{x : x \prec y\}$ 是开集。

这个假设将某些不连续的行为排除了；它的意思是说，如果 (x') 是消费束的一个序列，而且该序列中的每个消费束都至少和消费束 y 一样好，如果该序列收敛于某个消费束 x^*，则 x^* 至少和 y 一样好。

连续性最重要的后果是：如果 y 严格比 z 好，并且如果 x 是非常接近于 y 的一个消费束，则 x 严格比 z 好。这种说法只是严格偏好消费束集合为开集的另外一种表达方法。对于开集和闭集的简明讨论请见第 26 章。

在经济分析中，通常用**效用函数**（utility function）总结某个消费者的行为；也就是说，函数 $u: X \to R$ 使得 $x \succ y$ 当且仅当 $u(x) > u(y)$。可以证明，如果偏好排序是完备的、反身的、传递的和连续的，则它可以用一个连续的效用函数表示。下面我们将提供一个相对较弱版本的证明。效用函数通常是描述偏好的便利方法，但是不应该对此进行任何心理学上的解释。效用函数唯一重要的特征是它的序数性质。如果 $u(x)$ 代表某个偏好 $\succsim$ 而且 $f: R \to R$ 是一个单调函数，则 $f(u(x))$ 将和 $u(x)$ 代表同一偏好，这是由于 $f(u(x)) \geq f(u(y))$，当且仅当 $u(x) \geq u(y)$。

对于偏好还有一些有用的假设。例如：

弱单调性。若 $x \geq y$，则 $x \succsim y$。

强单调性。若 $x \geq y$ 且 $x \neq y$，则 $x \succ y$.

弱公理是说，如果消费束 x 中的每种商品的数量都不小于消费束 y 中的，则 x 至少和 y 一样好。强单调性是说如果消费束 x 中的每种商品的数量都不小于消费束 y 中的，并且某些商品的数量严格大于消费束 y 中的，则 x 严格比 y 好。这只是假设商品是好的（good）。

如果某种商品是坏商品（bad），例如垃圾或污染，那么强单调性就不再合理。但是在这些情形下，重新将这些商品定义为类似“缺乏垃圾”或“缺乏污染”，那么在这些重新定义过的商品上，偏好是满足强单调性假设的。

比上述两种单调性都弱的另外一个假设为：

局部非饱和性（local nonsatiation）。给定 X 中的任何 x 以及任何 $\varepsilon > 0$，则在 X 中存在着满足 $|x - y| < \varepsilon$ 的某个消费束 y，使得 $y \succ x$ [一]。

局部非饱和性是说，即使你被限制只能对消费束稍微进行改变，你也总能做得稍微更好。读者需要验证一下，强单调性蕴涵着局部非饱和性，但反过来则不成立。局部非饱和性排除了“厚的”无差异曲线。

我们还需要两个假设，这两个假设通常用来保证消费者需求函数表现良好：

凸性。给定 X 中 x, y 和 z 并且 $x \succsim z$ 和 $y \succsim z$，则 $tx + (1-t)y \succsim z$，其中 $0 \leq t \leq 1$。

严格凸性。给定 X 中 x, y（$x \neq y$）和 z 并且 $x \succsim z$ 和 $y \succsim z$，则 $tx + (1-t)y \succ z$，其中 $0 < t < 1$。

给定一个偏好排序，我们通常用图形表示。由彼此无差异的所有消费束组成的集合称为**无差异曲线**（indifference curve）。你可以将无差异曲线想象成效用函数的**水平集**(level

[一] 符号 $|x - y|$ 表示欧式空间中 x 和 y 的距离。

sets)；无差异曲线类似于生产理论中的等产量线。在一条无差异曲线上面或上方（on or above）的所有消费束$\{X$ 中的 $x : x \succeq y\}$，称为**上轮廓集**（upper contour set）[一]。上轮廓集类似于生产理论的必要投入集。

凸性意味着消费者偏好平均消费束胜于极端消费束，但是除了这个意义之外，它不具有其他经济意义。在凸偏好的情形下，无差异曲线可能含有水平段，然而在严格凸偏好的情形下，无差异曲线是严格圆滑的。凸性是新古典经济学中"边际替代率递减"假设的推广。

例子：效用函数的存在性。

效用函数的存在性。假设偏好是完备的、反身的、传递的、连续的而且是强单调的，则存在着可以代表这些偏好的一个连续效用函数$u : R_+^k \to R$。

证明。令e表示R_+^k中的单位向量，即$e = \underbrace{(1,...,1)}_{k个1}$。给定任何向量$x$，令$u(x)$是使得$x \sim u(x)e$的那个数。我们需要证明这样的数存在而且是唯一的。

令$B = \{R$ 中的 $t : te \succeq x\}$以及$W = \{R$ 中的 $t : x \succeq te\}$。强单调性意味着B是非空的；W显然是非空的，因为它包含 0。连续性意味着上述两个集合都是闭的。由于实直线是连续的，存在某个t，使得$t_x e \sim x$。我们必须证明这个效用函数真正代表了消费者隐含的偏好。令

$$u(x) = t_x \quad 其中 t_x e \sim x$$

$$u(y) = t_y \quad 其中 t_y e \sim y.$$

于是如果$t_x < t_y$，根据强单调性可知$t_x e \prec t_y e$，根据传递性可知，

$$x \sim t_x e \prec t_y e \sim y.$$

类似地，如果$x \succ y$，则$t_x e \succ t_y e$，因此$t_x > t_y$。

证明$u(x)$是连续函数需要额外技术，此处省略。■

例子：边际替代率

令$u(x_1,...,x_n)$是一个效用函数。假设我们增加商品i的数量，为了使效用不变，消费者应该怎样改变商品j的数量？

根据第 1 章的做法，令dx_i和dx_j分别表示x_i和x_j的变化量。根据假设，效用的变化量必定为零。因此

$$\frac{\partial u(x)}{\partial x_i} dx_i + \frac{\partial u(x)}{\partial x_j} dx_j = 0.$$

[一] 有时翻译为"上水平集"等，译者注。

因此，

$$\frac{dx_i}{dx_j} = -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial xj}}.$$

这个式子称为商品 i 和商品 j 之间的**边际替代率**（marginal rate of substitution）。

边际替代率不取决于我们选取哪个效用函数来代表偏好，也就是说效用函数的单调变换不会改变边际替代率。为了证明这一点，令 $v(u)$ 是效用函数 u 的单调变换。这个效用函数的边际替代率为

$$\frac{dx_i}{dx_j} = -\frac{v'(u)\frac{\partial u(x)}{\partial x_i}}{v'(u)\frac{\partial u(x)}{\partial xj}} = -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial xj}}.$$

7.2 消费者行为

现在我们有了偏好的描述方法——效用函数，可以分析消费者行为了。我们的基本假设是理性的消费者总是从他买得起的消费束集中选择最偏好的消费束。

在偏好最大化的基本问题中，消费者能够买得起的消费束集合就是满足消费者预算约束的所有消费束。令 m 表示消费者的既定收入，令 $p = (p_1,..., p_k)$ 表示商品 $1,...,k$ 的价格向量。消费者能够买得起的消费束集合，即消费者的预算集为

$$B = \{X \text{ 中的 } x : px \le m\}.$$

效用最大化的问题因此可以写为：

$$\max_x \ u(x)$$

$$\text{使得 } px \le m$$

$$x \text{ 在 } X \text{ 中}.$$

我们先来分析一下这个问题的一些基本特征。第一个问题为解的存在性问题。根据第 27 章可知，我们需要检验目标函数是连续的，并且约束集是闭且有界的。根据假设可知，效用函数是连续的，约束集明显是闭的。如果 $p_i > 0$（其中 $i = 1,...,k$ ）并且 $m \ge 0$，不难证明约束集是有界的。如果某种商品的价格为零，消费者对于该商品的需求可能是无限的。我们一般忽略这样的边界问题。

第二个问题是偏好的代表性问题。这里我们注意到最大化的选择 x^* 和我们选择用哪个效用函数代表这些偏好无关。这是因为最优选择 x^* 必定满足下列性质：对于 B 中的所有 x，都有 $x^* \succeq x$，因此代表偏好 $\succeq$ 的任何效用函数必然都会选择 x^* 作为约束最大化的解。

第三，如果我们将所有商品的价格以及消费者的收入同乘以一个正常数，我们不会改

曹乾（东南大学 caoqianseu@163.com）

变预算集，因此我们不会改变最优选择集。也就是说，如果 x^* 具有下列性质：$x^* \succeq x$ 对于所有满足 $px \le m$ 的 x 成立，那么 $x^* \succeq y$ 对于所有满足 $tpy \le tm$ 的所有 y 成立。大致来说，最优选择集关于价格和收入是"零次齐次"的。

如果我们对偏好再作出一些正规性的假设，我们对消费者效用最大化行为描述得会更详细。例如，假设偏好满足局部非饱和性；我们能否得到某个 x^* 使得 $px^* < m$？假设我们能够得到；于是，由于 x^* 的花费严格小于 m，X 中每个充分接近 x^* 的消费束也是可行的。但是根据局部非饱和性假设，存在无限接近于 x^* 的某个消费束 x，x 比 x^* 更受偏好。但是这意味着 x^* 在预算集上并未使得效用最大。

因此，在局部非饱和性的假设下，效用最大化的消费束 x^* 必定在预算线上，即 $px^* = m$。这样，我们可以将消费者的最大化问题重新表述为

$$v(p,m) = \max_x \ u(x)$$

$$使得\ px = m.$$

函数 $v(p,m)$ 称为**间接效用函数**（indirect utility function），它是在既定价格和收入条件下能实现的最大效用值。这个最大化问题的解 x 称为消费者的**需求束**（demanded bundle）。需求束给出了在价格和收入既定条件下消费者想要的每种商品的数量。我们假设每个预算集都有唯一的需求束；作出这个假设的目的是出于方便，对于分析来说，这个假设不是必需的。

需求束是 p 和 m 的函数，这个函数称为消费者的**需求函数**（demand function）。我们用 $x(p,m)$ 表示需求函数。就象企业的情形一样，我们需要作出一些假设确保需求函数是定义清晰的。特别地，我们假设使效用最大的消费束是唯一的。稍后我们将看到，偏好的严格凸性将满足上述唯一性的行为。

正象企业的情形一样，消费者的需求函数是 (p,m) 的零次齐次函数。我们在前面已经说过这个结论，将价格和收入同乘以一个正数，丝毫不会改变预算集，因此也不会改变效用最大化问题的解。

和生产的情形一样，我们可以使用微积分刻画最优化行为，只要效用函数是可微的。效用最大化问题的拉格朗日函数可以写为

$$L = u(x) - \lambda(px - m),$$

其中 λ 是拉格朗日乘子。将这个拉格朗日函数对 x_i 求导，可得到一阶条件

$$\frac{\partial u(x)}{\partial x_i} - \lambda p_i = 0 \qquad i = 1,\dots,k.$$

为了解释这些条件，我们将第 i 个一阶条件除以第 j 个一阶条件可得

$$\frac{\dfrac{\partial u(x^*)}{\partial x_i}}{\dfrac{\partial u(x^*)}{\partial x_j}} = \frac{p_i}{p_j} \qquad i, j = 1,\dots,k.$$

上式的左侧是商品i和商品j之间的边际替代率，右侧可以称为商品i和商品j之间的经济替代率（economic rate of substitution）。最大化意味着这两个替代率必定相等。如果不相等，假设

$$\frac{\frac{\partial u(x^*)}{\partial x_i}}{\frac{\partial u(x^*)}{\partial x_j}}=\frac{1}{1}\neq\frac{2}{1}=\frac{p_i}{p_j}.$$

于是，如果消费者放弃一单位商品商品i并且购买一单位商品j，那么他将位于同一条无差异曲线上，并且还余下一元钱。所以，总效用会增加，这与最大化矛盾。

这个结论也可以用几何方法论证，如图 7.1 所示。消费者的预算线为$\{x: p_1x_1+p_2x_2=m\}$。这个式子也可以写成隐函数的图形形式：$x_2=m/p_2-(p_1/p_2)x_1$。因此，预算线的斜率为$-p_1/p_2$，纵截距为m/p_2。消费者希望在预算线上找到一点使得他的效用最大。这显然必须满足相切条件，即无差异曲线的斜率等于预算线的斜率。将这些语句翻译成代数语言，就得到了前面的条件。

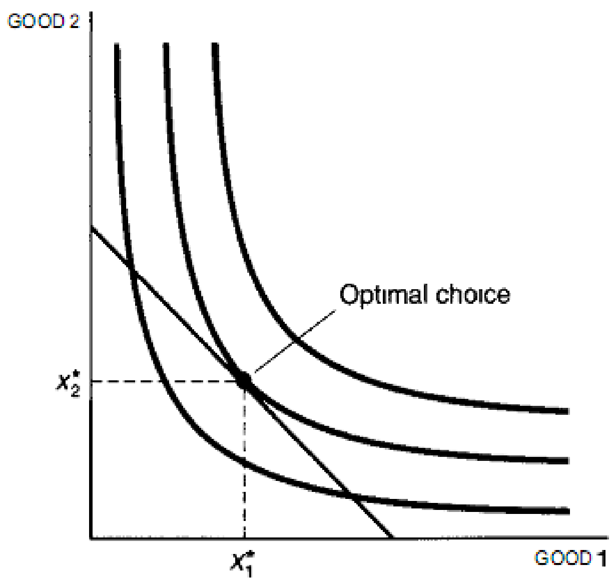


图 7.1：效用最大化。最优消费束位于无差异曲线和预算线的切点之处。

最后，我们可以使用向量术语表达这个条件。令x^*为一个最优选择，令dx表示满足预算约束的x^*的扰动。因此，我们必然有

$$p(x^*\pm dx)=m.$$

由于$px=m$，这个式子意味着$pdx=0$，这反过来意味着dx必定与p正交。

对于dx的这类任何扰动，效用不会发生变化，否则x^*不是最优的。因此，我们也有

$$Du(x^*)dx = 0$$

这表明 $Du(x^*)$ 与 dx 也是正交的。既然这个结论对于满足 $pdx = 0$ 的所有扰动 dx 成立，$Du(x^*)$ 必然和 p 成比例，这就是我们已经发现的一阶条件。

效用最大化的二阶条件可用第 27 章的知识进行分析。拉格朗日函数对商品 i 和商品 j 的二阶导数为 $\partial^2 u(x)/\partial x_i \partial x_j$。因此，二阶条件可以写为

$$h^t D^2 u(x^*) h \leq 0 \text{ 对于所有满足 } ph = 0 \text{ 的 } h \text{ 成立.} \tag{7.1}$$

这个条件要求效用函数的海赛矩阵对于与价格向量正交的 h 是负半定的。这等价于要求 $u(x)$ 是局部拟凹的。从几何上来说，这个条件表示上轮廓集在最优选择 x^* 之处必然位于预算超平面的上方。

和往常一样，二阶条件可用涉及加边的海赛矩阵表示。通过第 27 章的知识可知，（7.1）式为严格不等式的充分必要条件是加边海赛行列式的自然序主子式交替改变符号。因此，

$$\begin{vmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{vmatrix} > 0$$

$$\begin{vmatrix} 0 & -p_1 & -p_2 & -p_3 \\ -p_1 & u_{11} & u_{12} & u_{13} \\ -p_2 & u_{21} & u_{22} & u_{23} \\ -p_3 & u_{31} & u_{32} & u_{33} \end{vmatrix} < 0$$

等等。

7.3 间接效用

我们在前面定义了间接效用函数的概念。这个函数 $v(p,m)$ 是指最大效用是价格 p 和收入 m 的函数。

间接效用函数的性质。

（1）$v(p,m)$ 是价格 p 的非增函数；也就是说，如果 $p' \geq p$, 则 $v(p',m) \leq v(p,m)$。类似地，$v(p,m)$ 是收入 m 的非减函数。

（2）$v(p,m)$ 是 (p,m) 的零次齐次函数。

（3）$v(p,m)$ 是价格 p 的拟凹函数；也就是说，$\{p : v(p,m) \leq k\}$ 对于所有 k 来说都是凸集。

（4）$v(p,m)$ 是连续的，若 $p >> 0, m > 0.$

证明。

（1）令 $B=\{x: px\le m\}$ 以及 $B'=\{x: p'x\le m\}$，其中 $p'\ge p.$ 则 B' 包含于 B。因此，$u(x)$ 在 B 上的极大值，至少和 $u(x)$ 在 B' 上的极大值一样大。类似地，可以证明关于 m 的结论。

（2）如果价格和收入同乘以一个正数，预算线根本不会改变。因此，$v(tp,tm)=v(p,m)$，其中 $t>0.$

（3）假设 p 和 p' 分别满足 $v(p,m)\le k$ 和 $v(p',m)\le k$。令 $p''=tp+(1-t)p'$。我们想证明 $v(p'',m)\le k$。定义以下预算集：

$$B=\{x: px\le m\}$$

$$B'=\{x: p'x\le m\}$$

$$B''=\{x: p''x\le m\}$$

我们将证明，任何 x 若在 B'' 中，则 x 必然在 B 中或在 B' 中；也就是说，$B\cup B'\supset B''$。假设不是如此；则 x 满足 $tpx+(1-t)p'x\le m$ 但 $px>m$ 且 $p'x>m$。这两个不等式可以写为

$$tpx>tm$$

$$(1-t)p'x>(1-t)m.$$

将上面两个式子相加，可得

$$tpx+(1-t)p'x>m.$$

这与原来的假设矛盾。

现在注意到

$v(p'',m)=\max u(x)$ 使得 x 在 B'' 中

$\le \max u(x)$ 使得 x 在 $B\cup B'$ 中　　（因为 $B\cup B'\supset B''$）

$\le k$　　（因为 $v(p,m)\le k$ 且 $v(p',m)\le k$）。

（4）从第 27 章的最大化定理可以推知。■

在图 7.2 中，我们画出了一种特别的集合，称为“价格无差异曲线”，这些曲线就是间接效用函数的水平集。根据上面定理中的性质（1）可知，当我们向原点移动时，效用是非减的；根据性质（3）可知，下轮廓集是凸的。注意，下轮廓集位于价格无差异曲线的东北方，这是因为间接效用随着价格增高而下降。

我们注意到如果偏好满足局部非饱和性假设，则 $v(p,m)$ 是 m 的严格增函数。在图 7.3 中，我们已经画出了在价格既定不变的情形下，$v(p,m)$ 和 m 之间的关系。因为 $v(p,m)$ 是 m 的严格增函数，所以它是可逆的，我们可以从该函数中解出 m:它是效用水平的函数；也就是说，给定任何效用水平 u，我们可以找到当价格为 p 时实现效用 u 所必需的最小收入额。如图 7.3 所示。收入和效用的这种函数关系，即间接效用函数的反函数，称为支出函数（expenditure function），并用 $e(p,u)$ 表示它。

支出函数的一个等价定义是：

$$e(p,u) = \min \ px$$

$$\text{使得}\, u(x) \geq u.$$

支出函数给出了实现既定效用水平的所必需的最小成本。

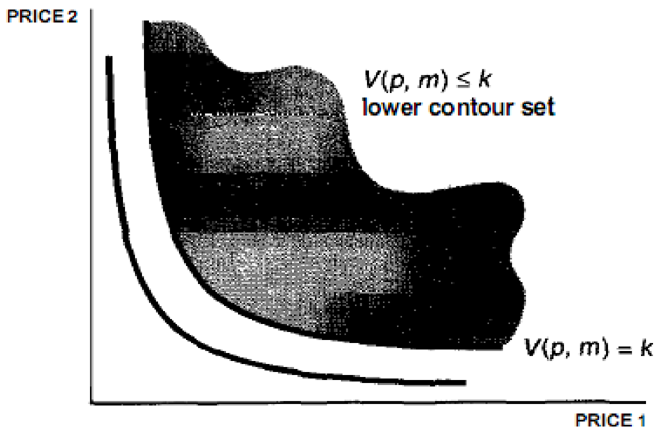


图 7.2：价格无差异曲线。价格无差异曲线是指满足$v(p,m)=k$（其中k为常数）的所有价格组合。下轮廓集包含所有满足$v(p,m)\leq k$的价格组合。

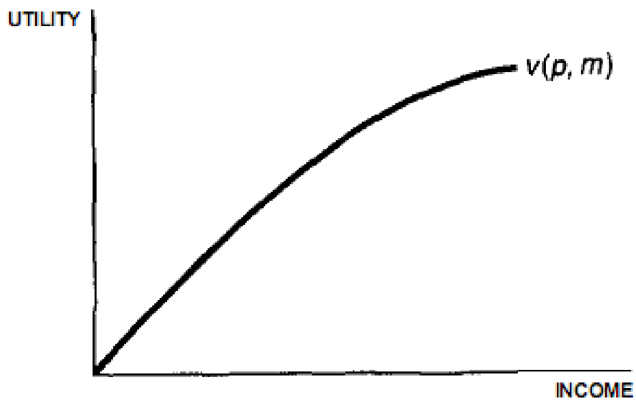


图 7.3：效用作为收入的函数。当收入增加时，间接效用必然增加。

支出函数完全类似于我们曾研究过的企业行为中的成本函数，因此，它具有成本函数的全部性质（参见第 5 章）。为方便起见，我们将这些性质重复表述如下。

支出函数的性质。

（1） $e(p,u)$是p的非减函数。

（2） $e(p,u)$是p的一次齐次函数。

（3） $e(p,u)$是p的凹函数。

（4） $e(p,u)$ 是 p 的连续函数，若 $p>>0$ 。

（5）若 $h(p,u)$ 是价格为 p 时实现效用水平 u 的**最小支出**束，则 $h_i(p,u)=\dfrac{\partial e(p,u)}{\partial p_i}$ ，其中 $i=1,...,k$ ，假设导数存在而且 $p_i>0$.

证明。这些性质正是成本函数的性质。请参见第 5 章的证明。■

函数 $h(p,u)$ 称为**希克斯**（Hicksian）**需求函数**。希克斯需求函数类似于前几章中的条件要素需求函数。希克斯需求函数告诉我们实现既定效用水平所必需的最小支出。

希克斯需求函数有时又称为**补偿需求函数**（compensated demand function）。这个名字来源于下列对需求函数的构造方法：变动价格和收入是的消费者保持在既定的效用水平上。因此，我们调整收入的目的是“补偿”价格的变化。

希克斯需求函数不是可以直接观测到的，因为它依赖于效用，但效用是不可直接观测到的。作为价格和收入函数的需求函数是可以观测到的；当我们想强调希克斯需求函数和通常的需求函数的区别时，我们通常将后者称为**马歇尔**（Marshallian）**需求函数** $x(p,m)$ 。马歇尔需求函数就是我们一直讨论的普通需求函数。

7.4 一些重要的恒等式

有一些重要的恒等式，它们将支出函数、间接效用函数、马歇尔需求函数和希克斯需求函数连接在一起。

我们考虑以下的效用最大化问题

$$v(p,m^*)=\max u(x)$$

$$\text{使得 } px\le m^*.$$

令 x^* 是上述最大化问题的解并且令 $u^*=u(x^*)$ 。考虑支出函数问题

$$e(p,u^*)=\min\ px$$

$$\text{使得 } u(x)\ge m^*.$$

分析图 7.4 可知，在非反常的情形下，这两个问题的解都应该为 x^* 。（更严格的论证参见本章的附录。）这个简单的结论使我们得到四个重要的恒等式：

（1） $e(p,v(p,m))\equiv m$ 。实现效用 $v(p,m)$ 的必要最小支出为 m 。

（2） $v(p,e(p,u))\equiv u$ 。收入 $e(p,u)$ 能实现的最大效用为 u 。

（3） $x_i(p,m)\equiv h_i(p,v(p,m))$ 。收入为 m 时的马歇尔需求函数与效用为 $v(p,m)$ 时的希克斯需求函数是相同的。

（4） $h_i(p,u)\equiv x_i(p,e(p,u))$ 。效用为 u 时的希克斯需求函数与收入为 $e(p,u)$ 时的马歇

尔需求函数是相同的。

最后一个恒等式也许是最重要的，因为它将“可观测的”马歇尔需求函数和“不可观测的”希克斯需求函数联系起来。这个恒等式表明，希克斯需求函数——支出最小化的解——等于在相应收入水平时的马歇尔需求函数——即，在给定价格水平下，为实现合意的效用水平所必需的最小收入。因此，任何需求束都可以表达为效用最大化问题的解，也可以表达为支出最小化问题的解。在本章的附录部分，我们给出了二者相等的条件。目前，我们只是分析这个对偶性的结果。

正是这种联系产生了“补偿需求函数”的概念。在价格变动的情形下，我们调整消费者的收入，即对他进行“补偿”以让他维持在既定的效用水平上，这种情形下他对商品的马歇尔需求就是希克斯需求。

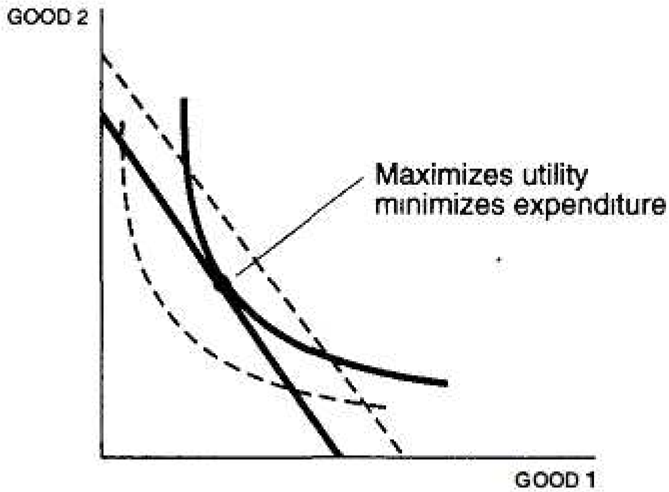


图 7.4：最大化效用和最小化支出。一般来说，使效用最大化的消费束也使支出最小化；反之亦然。

下列定理给出了这些恒等式的一个应用：

罗伊(Roy)恒等式。如果 $x(p,m)$ 是马歇尔需求函数，那么

$$x_i(p,m) = -\frac{\dfrac{\partial v(p,m)}{\partial p_i}}{\dfrac{\partial v(p,m)}{\partial m}} \quad i = 1,\ldots,k$$

当然前提条件是上式右侧是良好定义的，而且 $p_i > 0$ 以及 $m > 0$ 。

证明。假设 x^* 在 (p^*, m^*) 处使得效用最大，最大值为 u^*，从前面的恒等式我们知道

$$x(p^*, m^*) \equiv h(p^*, u^*). \quad (7.2)$$

从另外一个恒等式，我们还知道

$$u^* \equiv v(p, e(p, u^*)).$$

这个恒等式是说，不管价格在什么水平上，如果在这些价格下为使消费者达到效用u^*，你给他一笔最小的收入，那么他能达到的最大效用就是u^*。

由于这是一个恒等式，该式两侧同时对 p_i 求导可得

$$0 = \frac{\partial v(p^*, m^*)}{\partial p_i} + \frac{\partial v(p^*, m^*)}{\partial m} \frac{\partial e(p^*, u^*)}{\partial p_i}$$

将上式变形并结合（7.2）式可知

$$x_i(p^*, m^*) \equiv h_i(p^*, u^*) \equiv \frac{\partial e(p^*, u^*)}{\partial p_i} \equiv -\frac{\dfrac{\partial v(p^*, m^*)}{\partial p_i}}{\dfrac{\partial v(p^*, m^*)}{\partial m}}.$$

由于这个恒等式对于所有(p^*, m^*)成立，而且由于$x^* = x(p^*, m^*)$，就证明了最终的结果。■

以上的证明，尽管简练，但它的启发性不够强。下面我们再给出另外一种直接的证明方法。间接效用函数为

$$v(p, m) \equiv u(x(p, m)) \tag{7.3}$$

如果我们将上式对 p_j 求导，可得

$$\frac{\partial v(p, m)}{\partial p_j} \equiv \sum_{i=1}^{n} \frac{\partial u(x)}{\partial p_i} \frac{\partial x_i}{\partial p_j}. \tag{7.4}$$

由于$x(p, m)$为需求函数，它满足利润最大化的一阶条件。将这些一阶条件代入（7.4）式可得

$$\frac{\partial v(p, m)}{\partial p_j} = \lambda \sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial p_j}. \tag{7.5}$$

需求函数还满足预算约束 $px(p, m) \equiv m$，将该式两侧对 p_j 求导可得

$$x_j(p, m) + \lambda \sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial p_j} = 0. \tag{7.6}$$

将（7.6）代入（7.5）可得

$$\frac{\partial v(p, m)}{\partial p_j} = -\lambda x_j(p, m). \tag{7.7}$$

现在将（7.3）式两侧对m求导可得

$$\frac{\partial v(p, m)}{\partial m} = \lambda \sum_{i=1}^{k} p_i \frac{\partial x_i}{\partial m}. \tag{7.8}$$

将预算约束对 m 求导可得

$$\sum_{i=1}^{k} p_i \frac{\partial x_i}{\partial m} = 1 \tag{7.9}$$

将（7.9）代入（7.8）可得

$$\frac{\partial v(p,m)}{\partial m} = \lambda. \tag{7.10}$$

这个式子只是说，一阶条件中的拉格朗日乘子就是收入的边际效用。将（7.7）和（7.10）合在一起就得到了罗伊恒等式。

最后，罗伊恒等式还有一种证明方法，这就是直接使用包络定理进行证明（参见第 27 章）。上述的论证过程就是将包络定理的证明过程重述了一遍。

7.5 以货币度量的效用函数

关于支出函数有一个漂亮的构造，福利经济学不少地方都要用到它。考虑某些价格 p 和某个给定的商品束 x。我们提出以下问题：为了和消费商品束 x 的状况一样好，在价格 p 下某个既定的消费者至少需要多少钱？

图 7.5 告诉我们如何使用图形构建这个问题的答案，前提是我们知道消费者偏好。我们只要看看消费者为了达到经过 x 的无差异曲线需要多少钱即可。在数学上，我们要求解下列问题：

$$\min_z \; pz$$

$$\text{使得} \; u(z) \geq u(x)$$

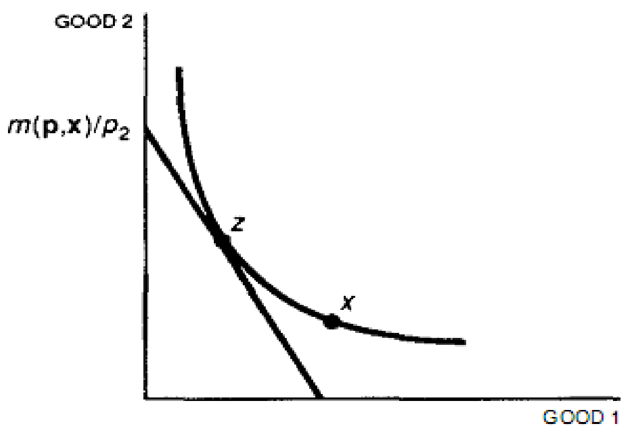


图 7.5：以货币度量的直接效用函数。货币制的效用函数，给出了在价格 p 下为使消费者的效用至少和商品束 x 一样好，他应该拥有的最小钱数。

这类函数经常出现，因此需要给它起个名字；遵循萨缪尔森（Samuelson, 1974）的叫法，

我们将其称为**以货币度量的效用函数**（money metric utility function）。有时也称为“最低收入函数”，或“直接补偿函数”等其他名字。另外的一种定义方法是

$$m(p,x) \equiv e(p,u(x)).$$

容易看出，如果 x 固定不变，则 $u(x)$ 也是固定不变的，因此 $m(p,x)$ 的行为和支出函数的行为是一样的：它对于价格 p 是单调的、齐次的和凹的，等等。但下列事实不怎么明显：当 p 固定不变时，$m(p,x)$ 实际上就是一个效用函数。证明很简单：价格固定不变时，支出函数是效用水平的增函数，如果你想得到更高的效用水平，你必须花费更多的钱。事实上，如果偏好是连续的、局部非饱和的，则支出函数是 u 的严格增函数。

因此，对于固定不变的 p，$m(p,x)$ 只是效用函数的单调变换，因此它本身也是个效用函数。

由图 7.5 不难看出这一点。对于通过 x 的无差异曲线上的所有点，我们赋予相同的数值；对于更高无差异曲线上的所有点，我们赋予更大的数值。这正是效用函数要求的全部条件，因此 $m(p,x)$ 是一个效用函数。

对于间接效用函数也有一个类似的构造，称为**以货币度量的间接效用函数**（money metric indirect utility function）。它的表达式为

$$\mu(p;q,m) \equiv e(p,v(q,m)).$$

也就是说，$\mu(p;q,m)$ 衡量当价格为 p 时某人需要多少钱才能使他的状况，和当价格为 q 且他的收入为 m 时的状况一样好。和前面直接效用情形一样，$\mu(p;q,m)$ 的行为类似支出函数关于 p 的行为，但是现在它的行为类似于间接效用函数关于 q 和 m 的行为。这是因为，它毕竟是间接效用函数的单调变换。图 7.6 给出了一个例子。

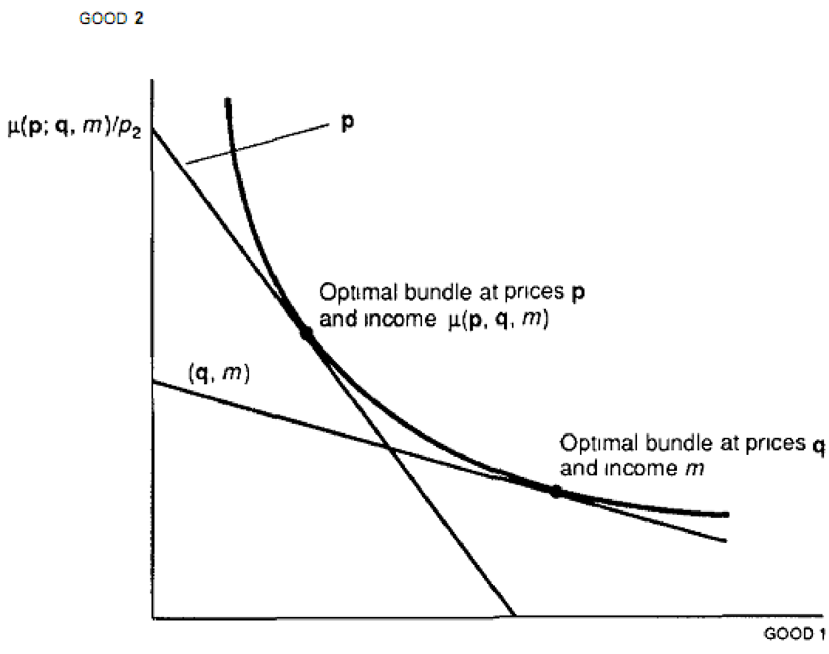


图 7.6:以货币单位衡量的间接效用函数。这个函数是说当价格为 p 时，消费者至少应拥有多少钱，才能使他的状况和当价格为 q 且他的收入为 m 时的状况一样好。

直接和间接补偿函数的一个没好特征是他们仅包含**可以观测到的**变量。它们是一类**特别的**直接和间接效用函数，使用它们可以衡量我们感兴趣的一些东西，而且关于单调变换也是很明确的。在讨论可积理论（integrability theory）和福利经济学时，我们将发现这一特征非常有用。

例子：柯布-道格拉斯效用函数

柯布-道格拉斯效用函数形式为$u(x_1,x_2)=x_1^a x_2^{1-a}$. 由于该函数的任何单调变换仍然代表着相同的偏好，我们也可以将它写为$u(x_1,x_2)=a\ln x_1+(1-a)\ln x_2$.

支出函数和希克斯需求函数是相同的，只要相应将符号改变一下即可；成本函数和条件要素需求函数（请参见第 4 章）也是相同的。马歇尔需求函数和间接效用函数可用求解下列问题的方法推导出来：

$$\max \ \ a\ln x_1+(1-a)\ln x_2$$

$$\text{使得 } p_1x_1+p_2x_2=m.$$

构造拉格朗日函数

$$L=a\ln x_1+(1-a)\ln x_2-\lambda(p_1x_1+p_2x_2-m).$$

一阶条件为

$$\frac{a}{x_1}-\lambda p_1=0$$

$$\frac{1-a}{x_2}-\lambda p_2=0$$

$$p_1x_1+p_2x_2-m=0.$$

由前两个式子可得

$$\frac{a}{p_1x_1}=\frac{1-a}{p_2x_2}.$$

将上式交叉相乘并使用预算约束可得

$$ap_2x_2=p_1x_1-ap_1x_1$$

$$am=p_1x_1$$

$$x_1(p_1,p_2,m)=\frac{am}{p_1}$$

将x_1代入预算约束可得商品 2 的马歇尔需求函数

$$x_2(p_1,p_2,m)=\frac{(1-a)m}{p_2}$$

将商品 1 和 2 的马歇尔需求函数代入目标函数并整理可得到间接效用函数：

$$v(p_1,p_2,m)=\ln m-a\ln p_1-(1-a)\ln p_2. \tag{7.11}$$

得到间接效用函数的一种更快方法是求柯布-道格拉斯成本函数或支出函数（参见第 4 章）的逆函数。我们已知道这样的支出函数为

$$e(p_1,p_2,u)=Kp_1^a p_2^{1-a}u,$$

其中 K 是取决于参数 a 的常数。求上述函数的逆时，用 m 替换上式中的 $e(p_1,p_2,u)$，用 $v(p_1,p_2,m)$ 替代 u 可得

$$v(p_1,p_2,m)=\frac{m}{Kp_1^a p_2^{1-a}}.$$

这个式子只是（7.11）式的一个单调变换，对上式两侧同时取对数就可以看清这一点。

以货币度量的效用函数可用替代的方法得到。我们有

$$\begin{aligned} m(p,x)&=Kp_1^a p_2^{1-a}u(x_1,x_2)\\ &=Kp_1^a p_2^{1-a}x_1^a x_2^{1-a} \end{aligned}$$

以及

$$\begin{aligned} \mu(p;q,m)&=Kp_1^a p_2^{1-a}v(q_1,q_2,m)\\ &=Kp_1^a p_2^{1-a}q_1^{-a}q_2^{a-1}m. \end{aligned}$$

例子：CES 效用函数

CES 效用函数的形式为 $u(x_1,x_2)=(x_1^\rho+x_2^\rho)^{1/\rho}$. 由于效用函数的单调变换不会改变原来的偏好，我们可以选择 $u(x_1,x_2)=\dfrac{1}{\rho}\ln(x_1^\rho+x_2^\rho)$.

在以前我们已经知道 CES 技术的成本函数的形式为 $c(w,y)=(x_1^r+x_2^r)^{1/r}y$，其中 $r=\rho/(\rho-1)$。因此，CES 效用函数的支出函数必然具有下列形式

$$e(p,u)=(p_1^r+p_2^r)^{1/r}u.$$

需求函数可通过罗伊法则求出：

$$x_1(p,m)=\frac{-\partial v(p,m)/\partial p_1}{-\partial v(p,m)/\partial m_1}=\frac{\frac{1}{r}(p_1^r+p_2^r)^{-(1+\frac{1}{r})}mrp_1^{r-1}}{(p_1^r+p_2^r)^{-1/r}}=\frac{p_1^{r-1}m}{(p_1^r+p_2^r)}.$$

CES 效用函数的以货币度量的效用函数也可以使用替代的方法得到：

$$m(p,x)=(p_1^r+p_2^r)^{1/r}(x_1^\rho+x_2^\rho)^{1/\rho}$$

$$\mu(p;q,m)=(p_1^r+p_2^r)^{1/r}(q_1^r+a_2^r)^{-1/r}m.$$

附录

考虑下列两个问题：

$$\begin{aligned} &\max u(x) \\ &s.t. \ \ px \le m. \end{aligned} \qquad (7.12)$$

$$\begin{aligned} &\min \ px \\ &s.t. \ \ u(x) \ge u. \end{aligned} \qquad (7.13)$$

假设：

（1）效用函数是连续的；（2）偏好满足局部非饱和性；（3）上述两个最优化问题的解存在。

效用最大化蕴涵支出最小化。假设上述假设都得到满足。令 x^* 是（7.12）的解，令 $u = u(x^*)$。则 x^* 也是（7.13）的解。

证明。假设不是，令 x' 是（7.13）的解，因此 $px' < px^*$ 而且 $u(x') \ge u(x^*)$。根据局部非饱和性可知，存在充分接近于 x' 的消费束 x''，使得 $px'' < px^* = m$ 而且 $u(x'') > u(x^*)$。但这样一来，x^* 就不可能是（7.12）的解。■

支出最小化蕴涵效用最大化。假设上述假设都得到满足。令 x^* 是（7.13）的解，令 $m = px^*$，并假设 $m > 0$。则 x^* 也是（7.12）的解。

证明。假设不是，令令 x' 是（7.12）的解，因此 $u(x') \ge u(x^*)$ 并且 $px' = px^* = m.$ 由于 $px^* > 0$ 且效用函数是连续的，我们可以找到 $0 < t < 1$ 使得 $ptx' < px^* = m$ 且 $u(tx') > u(x^*)$。因此，x^* 不可能是（7.13）的解。■

注释

文中对效用函数存在性的证明参考了 Wold（1943）的文献。效用函数存在性一般定理可在 Debru(1964)中找到。

间接效用函数的重要性首先由 Roy（1942，1947）提出。支出函数似乎源于 Hicks（1946）。使用对偶方法研究消费者理论，参考的是 McFadden & Winter(1968)。货币度量的效用函数要归功于 McKenzie（1957）和 Samuelson（1974）。

习题

7.1 考虑定义在非负象限的偏好$(x_1,x_2)\succ(y_1,y_2)$若$x_1+x_2<y_1+y_2$。这样的偏好是否是局部非饱和的？如果消费品只有上面两种，而且它们的价格均为正，消费者会将他的收入全部花完吗？请解释。

7.2 某消费者的效用函数为$u(x_1,x_2)=\max\{x_1,x_2\}$。求该消费者对于商品 1 的需求函数。求他的间接效用函数和支出函数。

7.3 某个消费者的间接效用函数具有下列形式

$$v(p_1,p_2,m)=\frac{m}{\min\{x_1,x_2\}}.$$

求这个消费者的支出函数、（拟凹）效用函数和他对商品 1 的需求函数。

7.4 考虑下列间接效用函数

$$v(p_1,p_2,m)=\frac{m}{p_1+p_2}.$$

（a）求需求函数；（b）求支出函数；（c）求直接效用函数

7.5 某个消费者的直接效用函数为下列形式

$$u(x_1,x_2)=v(x_1)+x_2.$$

商品 1 是离散商品；商品 1 的可能消费水平只有两种即$x_1=0$和$x_1=1$。为简单起见，假设$u(0)=0$和$p_2=1$。

（a）这个消费者有什么样的偏好？

（b）如果p_1严格小于多少时，消费者将肯定选择$x_1=1$？

（c）与这个直接效用函数相伴的间接效用函数的代数表达式是什么？

7.6 某个消费者的间接效用函数为$v(\mathbf{p},m)=A(\mathbf{p})m.$

（a）这个消费者有什么样的偏好？

（b）他的支出函数$e(\mathbf{p},u)$是什么样的？

（c）他的以货币度量的间接效用函数$\mu(\mathbf{p};q,m)$是什么样的？

（d）现在假设消费者的间接效用函数为$v(\mathbf{p},m)=A(\mathbf{p})m^b$，其中$b>1$。在这种情形下，该消费者的以货币度量的间接效用函数是什么样的？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 8 章：选择

曹乾　译

(东南大学　caoqianseu@163.com)

8 选择

在本章，我们将对消费者的需求行为进行比较静态分析：消费者的需求如何随价格和收入的变动而变动。和企业的情形一样，我们使用三种方法进行分析：对一阶条件进行微分；使用支出函数和间接效用函数的性质；使用最优化模型蕴涵的代数不等式。

8.1 比较静态分析

下面我们对两种商品的消费者效用最大化问题分析得更详细一些。我们对消费者的需求如何随着问题中的参数变化而变化感兴趣。我们维持价格不变允许收入变动；效用最大化的商品束的运动轨迹称为**收入扩展路径**（income expansion path）。从收入扩展路径，我们可以推导出一个函数，它将收入和消费者对每种商品的需求（价格不变）联系起来。这样的函数称为**恩格尔曲线**（Engel curves）。收入扩展路径和恩格尔曲线的形状有以下几种：

（1）收入扩展路径（因此恩格尔曲线）是一条经过原点的直线。在这种情形下，我们说消费者的需求曲线是单位弹性的。消费者消费每种商品的比例在不同收入水平下是相同的。

（2）收入扩展路径向其中一种商品弯曲靠近。也就是说，当消费者的收入增加时，两种商品的消费量都增加，但是其中一种商品的增加比例更大一些（奢侈品），另外一种商品增加比例更小一些（必需品）。

（3）收入扩展路径可能向后弯曲。在这种情形下，收入增加后消费者对其中一种商品的消费反而更少。例如，某个消费者认为当收入增加后，他会减少土豆的消费。这样的商品称为低档商品（inferior goods）；收入增加，需求也增加的商品称为正常商品（normal goods）。如图 8.1 所示。

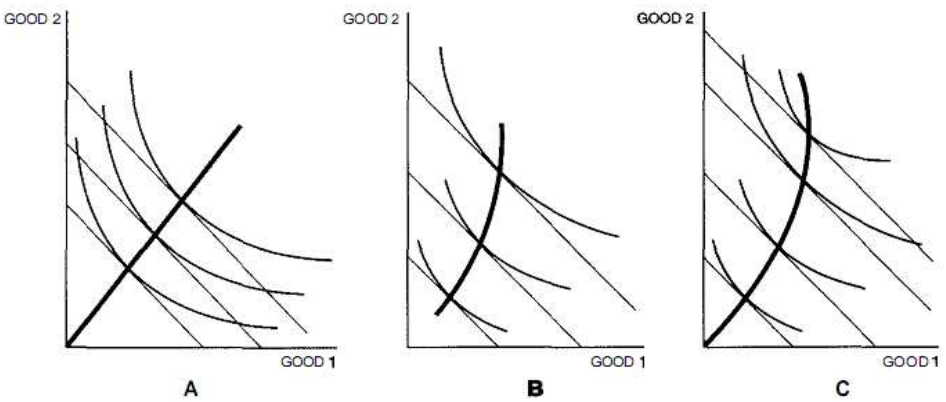


图 8.1：收入扩展路径。A 图表示单位弹性需求；B 图表示商品 2 为奢侈品；C 图表示商品 1 为低档商品。

我们可以维持收入固定不变但允许价格变动。如果令 p_1 变动而维持 p_2 和 m 不变，则预算线将会转动，预算线和无差异曲线的切点的运动轨迹称为**价格提供曲线**（price offer curve）。图 8.2 中的第一种情形是一种普通情形，因为商品 1 的价格降低后会导致商品 1 的需求增加；第二种情形是，商品 1 的价格下降导致商品 1 的需求下降。这样的商品称为吉芬商品（Giffen good）。土豆在某些情形下可能是吉芬商品；如果土豆价格下降，消费者能够买得起原来消费量的土豆而且还有余钱。消费者可用这部分多出的钱购买更多的通心面条。但是由于消费者消费更多的通心面条，他可能想减少土豆的消费。

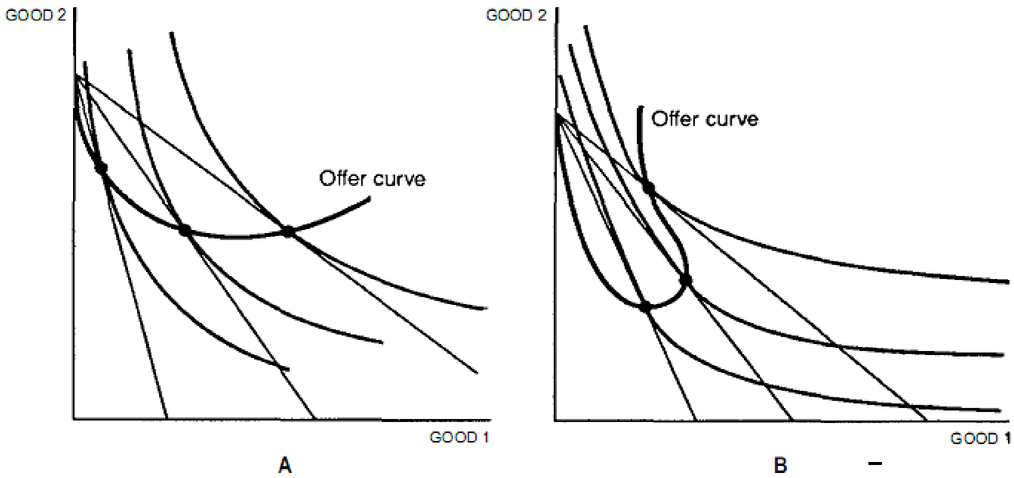


图 8.2：价格提供曲线。A 图中，商品 1 的价格下降后，商品 1 的需求增加，因此它是一种普通商品；在 B 图中，商品 1 价格下降但它的需求也下降，因此它是一种吉芬商品。

在以上的例子中，我们看到当某种商品价格下降后，可能产生两种效应：一是该商品相对另外一种商品更便宜；二是总“购买力”增加了。消费者理论的一个基本结果是斯勒茨基方程（Slutsky equation）。稍后我们将使用多种方法推导这个方程。

例子：商品税和收入税

假设我们想对某个追求效用最大化的消费者征税，以获得某些税收收入。最初，消费者的预算约束为 $p_1x_1+p_2x_2=m$，但是在对商品 1 征税后，消费者的预算约束变为 $(p_1+t)x_1+p_2x_2=m$。这种商品税（excise tax）的效应如图 8.3 所示。如果我们用 (x_1^*,x_2^*) 表示税后的消费水平，则税收收入为 tx_1^*.

假设现在我们决定征收收入税，但税收总额与上述商品税总额相同。在这种情形下，消费者的预算线变为 $p_1x_1+p_2x_2=m-tx_1^*$。这条预算线的斜率为 $-p_1/p_2$，而且它经过点

(x_1^*, x_2^*)。如图 8.3 所示。注意，由于这条预算线和无差异曲线相交于点(x_1^*, x_2^*)，在征税总额相同的情形下，征收收入税比征收商品税对消费者的影响小，因为消费者在前者情形下效用更高。

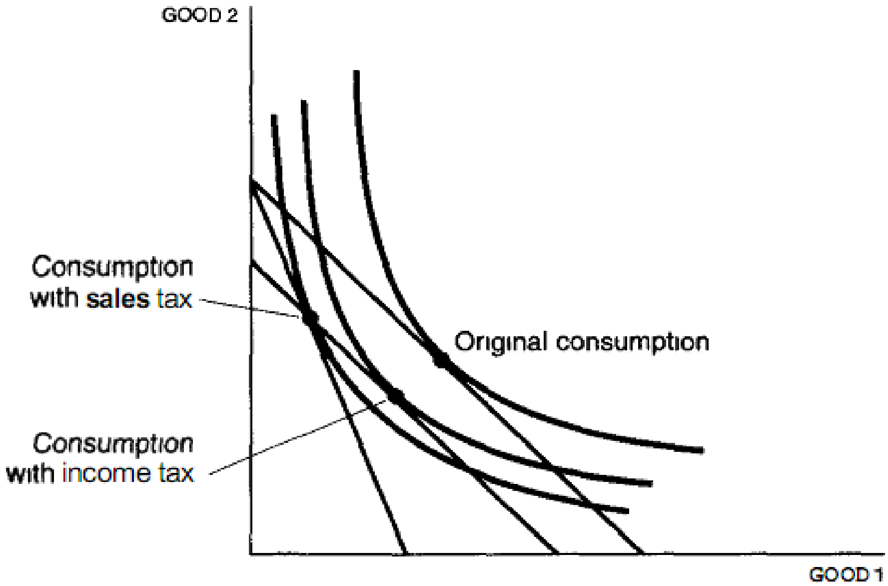


图 8.3：商品税和收入税。在征税总额相同的情形下，征收商品税与征收收入税相比，会使消费者的状况更差。（注：图中标注的 sales tax 为商品税）

8.2 斯勒茨基方程

我们已经知道希克斯需求曲线或称为补偿需求曲线，和企业理论中的条件要素需求是相同的。因此，它们具有完全相同的性质；特别地，它们都具有对称的负半定的替代矩阵。

在企业的情形下，这类限制通常是可以观测到的对企业行为的限制，因为企业的产出是可以观测到的变量。在消费者的情形下，这类限制似乎没有多大用处，因为效用是不可以直接观察到的。

然而，可以证明这种表象是不对的。尽管补偿需求函数不可以直接观察到，我们将看到，它的导数可以从可观察到的事物计算出来，即马歇尔需求关于价格和收入的导数。这个关系称为**斯勒茨基方程**（Slutsky equation）。

斯勒茨基方程。

$$\frac{\partial x_j(p,m)}{\partial p_i} = \frac{\partial h_j(p,v(p,m))}{\partial p_i} - \frac{\partial x_j(p,m)}{\partial m} x_i(p,m)$$

证明。令 x^* 在 (p^*,m^*) 之处使效用最大，令 $u^* = u(x^*)$ 。下列等式是恒成立的

$$h_j(p,u^*) \equiv x_j(p,e(p,u^*)).$$

我们可以将上述恒等式对 p_i 求导，并且在 $p = p^*$ 处求（导数）值，可得

$$\frac{\partial h_j(p^*,u^*)}{\partial p_i} = \frac{\partial x_j(p^*,m^*)}{\partial p_i} - \frac{\partial x_j(p^*,m^*)}{\partial m}\frac{\partial e(p^*,u^*)}{\partial p_i}$$

注意上式的意思。上式的左侧是说当 p_i 变动时，补偿需求如何变动。上式右侧是说，补偿需求的变动等于下列两项**之和**：一是维持收入为 m^* 不变时的需求变动；二是收入变动时的需求变动**乘以**为维持效用不变而必需的最小收入的变动。但是右侧的最后一项 $\partial e(p^*,u^*)/\partial p_i$ 就是 x^*。将上式变形可得

$$\frac{\partial x_j(p^*,u^*)}{\partial p_i} = \frac{\partial h_j(p^*,u^*)}{\partial p_i} - \frac{\partial x_j(p^*,m^*)}{\partial m} x_i^*$$

这就是斯勒茨基方程。■

斯勒茨基方程将由价格变动 Δp_i 引起的需求变动分解为两种独立的效应：替代效应和收入效应：

$$\Delta x_j \approx \frac{\partial x_j(p^*,u^*)}{\partial p_i}\Delta p_i = \frac{\partial h_j(p^*,u^*)}{\partial p_i}\Delta p_i - \frac{\partial x_j(p^*,m^*)}{\partial m} x_i^* \Delta p_i$$

我们也可以考虑所有价格同时变动的效应；在这种情形下，我们将导数解释为 n 维导数而不是偏导数。在两种商品情形下，斯勒茨基方程的形式为：

$$D_p x(p,m) = D_p h(p,u) - D_m x(p,m) x$$

$$\begin{bmatrix} \frac{\partial x_1(p,m)}{\partial p_1} & \frac{\partial x_1(p,m)}{\partial p_2} \\ \frac{\partial x_2(p,m)}{\partial p_1} & \frac{\partial x_2(p,m)}{\partial p_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1(p,u)}{\partial p_1} & \frac{\partial h_1(p,u)}{\partial p_2} \\ \frac{\partial h_2(p,u)}{\partial p_1} & \frac{\partial h_2(p,u)}{\partial p_2} \end{bmatrix} - \begin{bmatrix} \frac{\partial x_1(p,m)}{\partial m} \\ \frac{\partial x_2(p,u)}{\partial m} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix}$$

其中 $u = v(p,m)$.

将最后一项展开可得

$$\begin{bmatrix} \frac{\partial x_1(p,m)}{\partial m} \\ \frac{\partial x_2(p,u)}{\partial m} \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial x_1(p,m)}{\partial m} x_1 & \frac{\partial x_1(p,m)}{\partial m} x_2 \\ \frac{\partial x_2(p,u)}{\partial m} x_1 & \frac{\partial x_2(p,m)}{\partial m} x_2 \end{bmatrix}$$

假设我们考虑价格变动 $\Delta p = (\Delta p_1, \Delta p_2)$ ，我们对近似的需求变动 $\Delta x = (\Delta x_1, \Delta x_2)$ 感兴趣。根据斯勒茨基方程，我们可以使用下列式子计算这个变动

$$\begin{bmatrix}\Delta x_1\\ \Delta x_2\end{bmatrix}\approx\begin{bmatrix}\dfrac{\partial h_1}{\partial p_1} & \dfrac{\partial h_1}{\partial p_2}\\ \dfrac{\partial h_2}{\partial p_1} & \dfrac{\partial h_2}{\partial p_2}\end{bmatrix}\begin{bmatrix}\Delta p_1\\ \Delta p_2\end{bmatrix}-\begin{bmatrix}\dfrac{\partial x_1}{\partial m}x_1 & \dfrac{\partial x_1}{\partial m}x_2\\ \dfrac{\partial x_2}{\partial m}x_1 & \dfrac{\partial x_2}{\partial m}x_2\end{bmatrix}\begin{bmatrix}\Delta p_1\\ \Delta p_2\end{bmatrix}$$

$$=\begin{bmatrix}\Delta x_1^s\\ \Delta x_2^s\end{bmatrix}-\begin{bmatrix}\Delta x_1^m\\ \Delta x_2^m\end{bmatrix}$$

第一个向量是替代效应。它表明希克斯需求如何变动。由于希克斯需求变动不会改变效用，$(\Delta x_1^s,\Delta x_2^s)$将和无差异曲线相切。第二个向量是收入效应。价格变动导致“购买力”的变动量为$x_1\Delta p_1+x_2\Delta p_2$，向量$(\Delta x_1^m,\Delta x_2^m)$衡量这种变动对需求的影响，当然价格要固定在原来的水平。这个向量因此位于收入扩展路径上。

不仅对于需求的微小变动，对于需求的有限变动我们也可以做出类似的分解，如图 8.4 所示。此处，价格从p^0变为p'，需求从x变为x'。为了构建希克斯分解，我们首先绕着无差异曲线转动预算线，这样我们就可以找到价格为p'但效用维持在原来水平时的最优消费束。接下来我们将预算线平移到x'的位置，这样就找到了收入效应。总效应是这两种移动之和。

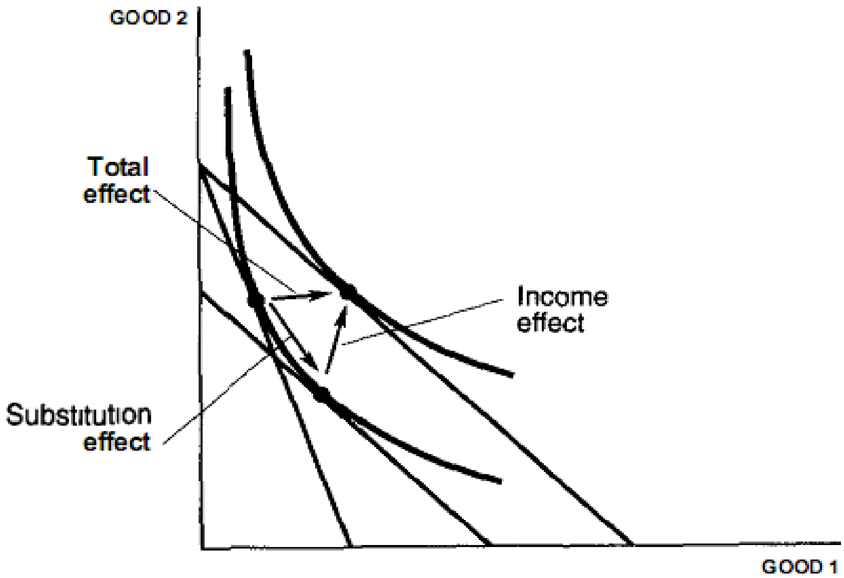


图 8.4：需求变动的希克斯分解。我们可以将需求的变动分解为两个部分：替代效应和收入效应。

例子：柯布-道格拉斯型斯勒茨基方程

我们使用柯布-道格拉斯效用函数检验斯勒茨基方程。我们已经知道，在这种情形下，我们有

$$v(p_1,p_2,m)=mp_1^{-a}p_2^{a-1}$$
$$e(p_1,p_2,u)=up_1^{a}p_2^{1-a}$$
$$x_1(p_1,p_2,m)=\frac{am}{p_1}$$
$$h_1(p_1,p_2,u)=ap_1^{a-1}p_2^{1-a}u.$$

因此，

$$\frac{\partial x_1(p,m)}{\partial p_1}=-\frac{am}{p_1^2}$$
$$\frac{\partial x_1(p,m)}{\partial m}=\frac{a}{p_1}$$
$$\frac{\partial h_1(p,u)}{\partial p_1}=a(a-1)p_1^{a-2}p_2^{1-a}u$$
$$\frac{\partial h_1(p,v(p,m))}{\partial p_1}=a(a-1)p_1^{a-2}p_2^{1-a}mp_1^{-a}p_2^{a-1}=a(a-1)p_1^{-2}m.$$

现在代入斯勒茨基方程，可得

$$\frac{\partial h_1}{\partial p_1}-\frac{\partial x_1}{\partial m}x_1=\frac{a(a-1)m}{p_1^2}-\frac{a}{p_1}\frac{am}{p_1}=\frac{[a(a-1)-a^2]m}{p_1^2}=\frac{-am}{p_1^2}=\frac{\partial x_1}{\partial p_1}.$$

8.3 需求函数的性质

支出函数的性质让我们比较容易地探讨新古典经济学中消费者行为的主要命题：

（1）替代矩阵 $(\partial h_j(p,u)/\partial p_i)$ 是负半定的。这是因为

$$(\partial h_j(p,u)/\partial p_i)=(\partial^2 e(p,u)/\partial p_i\partial p_j)$$

是负半定的，这又是因为支出函数是凹的（参看第 27 章）。

（2）替代矩阵是对称的，因为

$$\frac{\partial h_j(p,u)}{\partial p_i}=\frac{\partial^2 e(p,u)}{\partial p_j\partial p_i}=\frac{\partial^2 e(p,u)}{\partial p_i\partial p_j}=\frac{\partial h_i(p,u)}{\partial p_j}.$$

（3）特别地，“补偿性自身价格效应（the compensated own-price effect）是非正的”，也就是说，希克斯需求曲线是向下倾斜的：

$$\frac{\partial h_i(p,u)}{\partial p_i}=\frac{\partial^2 e(p,u)}{\partial p_i^2}\le 0.$$

这是因为替代矩阵是负半定的，因此其主对角线的元素是非正的。

这些限制都是关于希克斯需求函数的，这个函数是不可以直接观测到的。然而，我们在前面曾经指出过，斯勒茨基方程可以让我们将 h 关于 p 的导数，表达为 x 关于 p 和 m 的导数，这些都是可以观测到的。例如，由斯勒茨基方程和上面的评论可知，

（4）替代矩阵 $\left(\frac{\partial x_j(p,m)}{\partial p_i}+\frac{\partial x_j(p,m)}{\partial m}x_i\right)$ 是对称的、负半定的矩阵。

这个结论从直觉上不易看出：需求函数对价格的导数以及对收入的导数的特殊形式的和，必然导致负半定的矩阵。然而这个结论却是由最大化行为的逻辑严谨地推导出的。

8.4 使用一阶条件进行比较静态分析

斯勒茨基方程也可以从一阶条件的微分推导出。由于计算稍微有些冗长，我们仅分析商品为两种的情形，只给出论证的大致轮廓。

在这种情形下，一阶条件具有以下形式

$$
\begin{aligned}
p_1x_1(p_1,p_2,m)+p_2x_2(p_1,p_2,m)-m &\equiv 0\\
\frac{\partial u(x_1(p_1,p_2,m),x_2(p_1,p_2,m))}{\partial x_1}-\lambda p_1 &\equiv 0\\
\frac{\partial u(x_1(p_1,p_2,m),x_2(p_1,p_2,m))}{\partial x_2}-\lambda p_2 &\equiv 0.
\end{aligned}
$$

将上述一阶条件分别对 p_1 求导，并使用矩阵形式进行表达可得

$$
\begin{bmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{bmatrix}
\begin{bmatrix} \frac{\partial \lambda}{\partial p_1} \\ \frac{\partial x_1}{\partial p_1} \\ \frac{\partial x_2}{\partial p_1} \end{bmatrix}
=\begin{bmatrix} x_1 \\ \lambda \\ 0 \end{bmatrix}
$$

使用克莱姆（Cramer）法则求解 $\partial x_1/\partial p_1$ 可得

$$
\frac{\partial x_1}{\partial p_1}=\frac{\begin{vmatrix} 0 & x_1 & -p_2 \\ -p_1 & \lambda & u_{12} \\ -p_2 & 0 & u_{22} \end{vmatrix}}{H},
$$

其中 $H>0$，是加边海赛（Hessian）矩阵的行列式。

上式右侧分子中的行列式按照第二列的余子式展开，可得

$$
\frac{\partial x_1}{\partial p_1}=\lambda\frac{\begin{vmatrix} 0 & -p_2 \\ -p_2 & u_{22} \end{vmatrix}}{H}-x_1\frac{\begin{vmatrix} -p_1 & u_{12} \\ -p_2 & u_{22} \end{vmatrix}}{H}.
$$

这个式子看上去已有点象斯勒茨基方程了。注意上式右侧中的第一项，可以证明它是替代效应，因此是负的。现在回到一阶条件，将这些一阶条件对 m 求导，可得：

$$
\begin{bmatrix} 0 & -p_1 & -p_2 \\ -p_1 & u_{11} & u_{12} \\ -p_2 & u_{21} & u_{22} \end{bmatrix}
\begin{bmatrix} \dfrac{\partial \lambda}{\partial m} \\ \dfrac{\partial x_1}{\partial m} \\ \dfrac{\partial x_2}{\partial m} \end{bmatrix}
= \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}.
$$

因此，根据克莱姆法则

$$
\frac{\partial x_1}{\partial m} = \frac{\begin{vmatrix} -p_1 & u_{12} \\ -p_2 & u_{22} \end{vmatrix}}{H}.
$$

将其代入前面推导出的$\partial x_1 / \partial p_1$，我们就得到了斯勒茨基方程的收入效应部分。为了得到替代效应，我们必须建立支出最小化问题并据此解出$\partial h_1 / \partial p_1$。这个计算过程类似于条件要素需求函数的计算（参见第 4 章）。可以证明这个结果就是上面方程中的替代效应，这样就建立了斯勒茨基方程。

8.5 可积性问题

我们已经知道，效用最大化假设对消费者行为施加了某些观测性的限制。特别地，我们知道替代矩阵

$$
\left(\frac{\partial h_i(p,m)}{\partial p_j}\right) = \left(\frac{\partial x_i(p,m)}{\partial p_j} + \frac{\partial x_i(p,m)}{\partial m} x_j(p,m)\right),
$$

必定是一个对称的、负半定的矩阵。

假设我们有某个需求函数组，它有一个对称的、负半定的替代矩阵。是否一定存在一个效用函数，从这个函数我们可以推导出那些需求函数？这个问题称为可积性问题（the integrability problem）。

例如，罗伊（Roy）法则告诉我们

$$
x_i(p,m) = -\frac{\partial v(p,m)/\partial p_i}{\partial v(p,m)/\partial m}. \qquad (8.1)
$$

一般来说，如果给我们一个间接效用函数，我们就可以使用上式计算需求函数。然而，可积性问题问的是相反的问题：给定需求函数，以及（8.1）式中$i = 1,\ldots,k$个关系，我们如何根据这些方程找到$v(p,m)$？或者，更基本地，我们如何知道这个解是否存在？

（8.1）式给出的方程组，是一个由偏导数方程组成的方程组。可积性问题让我们确定这个方程组的解。

使用支出函数比使用间接效用函数更容易提出上述问题。假定我们给定某个需求方程组$(x_i(p,m))$，其中$i = 1,\ldots,k$。我们选择某个点$x^0 = x(p^0, m)$并且任意赋予它一个效用值u^0。我们如何构造支出函数$e(p,u^0)$？一旦我们发现一个与需求函数相符的支出函数，我们可以使用这个支出函数解出潜在的直接和间接效用函数。

如果这样的支出函数的确存在，它必然满足下列偏导数方程组

$$\frac{\partial e(p,u^0)}{\partial p_i}=h_i(p,u^0)=x_i(p,e(p,u^0)) \quad i=1,...,k \tag{8.2}$$

并且它还要满足初始调价

$$e(p^0,u^0)=p^0x(p^0,m^0).$$

这些式子只是说，，每种商品在效用为u的情形下的希克斯需求函数，就是收入为$e(p,u)$时的马歇尔需求函数。现在根据可积条件（integrability condition，请参见第 26 章）可知，下列形式的偏导数方程组

$$\frac{\partial f(p)}{\partial p_i}=g_i(p) \qquad i=1,...,k$$

具有一个（局部）解的充分必要条件是

$$\frac{\partial g_i(p)}{\partial p_j}=\frac{\partial g_j(p)}{\partial p_i} \qquad i,j=1,...,k.$$

将这个条件应用于上述问题，我们看到这个问题化简为要求矩阵

$$\left(\frac{\partial x_i(p,m)}{\partial p_j}+\frac{\partial x_i(p,m)}{\partial m}\frac{\partial e(p,u)}{\partial p_j}\right)$$

是对称的。但这正是斯勒茨基的限制！因此，斯勒茨基限制意味着我们可以将这些需求函数“积分”，从而找到与观测到的选择行为相符的一个支出函数。

这个对称条件足以保证存在函数$e(p,u^0)$，使得它至少在某个区域满足（8.2）式。（确保存在全局解的条件涉及的限制更多。）然而，为了保证这个函数是一个真正的支出函数，它必须还是价格的凹函数。也就是说，$e(p,u)$的二阶导数的矩阵必须为负半定的。但是，我们已经知道，$e(p,u)$的二阶导数矩阵正是斯勒茨基替代矩阵。如果它是负半定的，那么上述偏导方程组的解必定是凹的。

这些观察让我们得到了可积问题的一个解。给定一组需求函数$(x_i(p,m))$，我们只要验证它们是否拥有一个对称的、负半定的替代矩阵即可。如果的确如此，原则上我们可以解出（8.2）的解，这样就可以找到与那些需求函数相符的一个支出函数。

在从需求函数复原支出函数的同时，有一种方法可以同时复原出间接效用函数。（8.2）式对于所有的效用水平u^0都是有效的，因此我们可以选择一些基础价格q和基础收入水平m，令$u^0=v(q,m)$。使用这个替代，我们可以将（8.2）式写为

$$\frac{\partial e(p,v(q,m))}{\partial p_i}=x_i(p,e(p,v(q,m))) \ ,$$

其中边界条件现在变为

$$e(p,v(q,m))=m.$$

回忆以货币度量的（间接）效用函数（参见第 7 章，$\mu(p;q,m)\equiv e(p,v(q,m)).$）使用这个定义，我们可以将这组式子写为：

$$\frac{\partial\mu(p;q,m)}{\partial p_i} = x_i(p,\mu(p;q,m)) \qquad i=1,\dots,k$$
$$\mu(p;q,m) = m.$$

我们将这个方程组称为可积方程（integrability equations）。这个问题的解$\mu(p;q,m)$就是一个间接效用函数——一个特别的间接效用函数——它描述了可以观测到的需求行为$x(p,m)$。这个以货币度量的效用函数可以方便地用于福利分析。

例子：两种商品情形下的可积性

如果只消费两种商品，可积方程的形式非常简单，因为自变量只有一个，即两种商品的相对价格。类似地，也只有一个独立方程，因为如果我们知道其中一种商品的需求，就可以通过预算约束找到另一种商品的需求。

我们将商品 2 的价格标准化为 1，令 p 表示商品 1 的价格，$x(p,m)$ 表示商品 1 的需求函数。那么可积方程组变为一个单个的方程加上一个边界条件：

$$\frac{d\mu(p;q,m)}{dp} = x(p,\mu(p;q,m))$$
$$\mu(p;q,m) = m.$$

这是一个普通的带有边界条件的微分方程，可以使用标准方法解出这个方程的解。

例如，假设需求函数是对数线性的：

$$\ln x = a\ln p + b\ln m + c$$
$$x = p^a m^b e^c$$

可积方程为

$$\frac{d\mu(p;q,m)}{dp} = p^a m^b e^c.$$

变形可得

$$\mu^{-b}\frac{d\mu(p;q,m)}{dp} = p^a e^c.$$

将上式积分，

$$\int_p^q \mu^{-b}\frac{\partial\mu}{\partial t}dt = e^c\int_p^q t^a dt.$$

$$\left.\frac{\mu^{1-b}}{1-b}\right|_p^q = \frac{q^{a+1}-p^{a+1}}{a+1}e^c$$

其中$b \neq 1$。解上式可得

$$\frac{m^{1-b}-\mu(p;q,m)^{1-b}}{1-b}=\frac{q^{a+1}-p^{a+1}}{a+1}e^{c},$$

或，

$$\mu(p;q,m)=\left[m^{1-b}+\frac{b-1}{1+a}e^{c}[q^{a+1}-p^{a+1}]\right]^{\frac{1}{1-b}}$$

例子：多种商品情形下的可积性

我们现在考虑商品为三种的情形，这种情形下，独立的需求方程有两个。为具体起见，我们以柯布-道格拉斯方程组说明：

$$x_1=\frac{a_1 m}{p_1}$$

$$x_2=\frac{a_2 m}{p_2}$$

我们在前面曾经验证过这个方程组满足斯勒茨基的对称性，因此我们知道可积方程组有解。我们只要求出下列偏微分方程组的解即可：

$$\frac{\partial\mu}{\partial p_1}=\frac{a_1\mu}{p_1}$$

$$\frac{\partial\mu}{\partial p_2}=\frac{a_2\mu}{p_2}$$

$$\mu(p_1,p_2;q_1,q_2,m)=m$$

第一个方程意味着

$$\ln\mu=a_1\ln p_1+C_1$$

其中C_1是常数。第二个方程意味着

$$\ln\mu=a_2\ln p_2+C_2.$$

因此，我们自然想到要寻找具有下列形式的解

$$\ln\mu=a_1\ln p_1+a_2\ln p_2+C_3,$$

其中C_3是常数，和 p_1,p_2 无关。

将上式代入边界条件，可得

$$\ln\mu(p;q,m)=a_1\ln p_1+a_2\ln p_2+C_3.$$

解出C_3，代入前面想出的那个解可得

$$\ln\mu(p;q,m)=a_1\ln p_1+a_2\ln p_2-a_1\ln q_1-a_2\ln q_2+\ln m.$$

这个式子的确是柯布-道格拉斯效用函数的以货币度量的间接效用函数。第 7 章，我们提供了这个函数的另一个推导方法。

8.6 消费的对偶性

我们已经知道如何从观测到的需求函数还原间接效用函数，方法是使用可积方程组。下面我们将探讨如何求解直接效用函数。

答案表明直接效用函数和间接效用函数之间存在着漂亮的对偶性。最方便的方法是使用标准化过的间接效用函数进行计算，其中我们将收入和所有价格都除以收入，因此支出恒等于 1。所以，标准化的间接效用函数由下列问题给出

$$v(p) = \max_x u(x)$$

$$\text{使得 } px = 1.$$

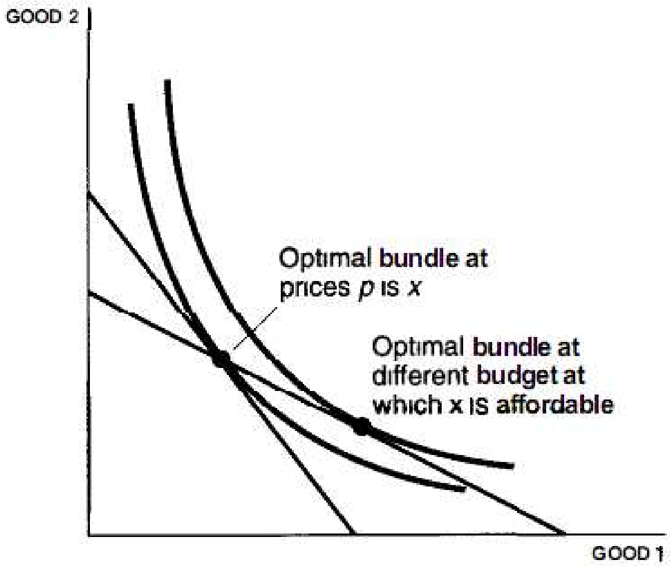


图 8.5：求解直接效用函数。需求束 x 产生的效用必定不会大于在任何价格下能买得起 x 时的效用（the utility that can be achieved at any prices p at which x is affordable）。

可以证明，如果给我们一个间接效用函数 $v(p)$，我们可以通过求解下列问题的方式找到直接效用函数

$$u(x) = \min_p v(p)$$

$$\text{使得 } px = 1.$$

在看到我们下面的阐述后，你就会知道上述结论的证明并不难。令 x 表示价格为 p 时的需求束。于是根据定义可知 $v(p) = u(x)$。令 p' 表示满足预算约束 $p'x = 1$ 的任何价格向量。于是 x 在价格为 p' 时总是一个可行的选择，根据预算集的形状可知，效用最大化的选择产生的效用不会小于 x 产生的效用；也就是说，$v(p') \geq u(x) = v(p)$。因此，这个间接效用函数在所有满足预算约束的价格上的最小值，就给出了 x 的效用。

上述论证过程请见图 8.5。任何满足预算约束 $px = 1$ 的价格向量产生的效用，都大于 $u(x)$，这意味着 $u(x)$ 是上述最小化问题的解。

例子：求解直接效用函数

假设我们的间接效用函数为 $v(p_1, p_2) = -a\ln p_1 - b\ln p_2$ 。相应的直接效用函数是什么？我们建立下列最小化问题：

$$\min_{p_1, p_2} = -a\ln p_1 - b\ln p_2$$

$$\text{使得 } p_1x_1 + p_2x_2 = 1.$$

一阶条件为

$$-a/p_1 = \lambda x_1$$
$$-b/p_2 = \lambda x_2,$$

或

$$-a = \lambda p_1x_1$$
$$-b = \lambda p_2x_2,$$

将上面两个式子加在一起，并使用预算约束可得

$$\lambda = -a - b.$$

代入一阶条件可得

$$p_1 = \frac{a}{(a+b)x_1}$$
$$p_2 = \frac{b}{(a+b)x_2}.$$

这个价格选择 (p_1, p_2) 使得间接效用函数最小。现在将其代入间接效用函数可得：

$$\begin{aligned} u(x_1, x_2) &= -a\ln\frac{a}{(a+b)x_1} - b\ln\frac{b}{(a+b)x_2} \\ &= a\ln x_1 + b\ln x_2 + C \end{aligned}$$

其中 C 为常数，这就是我们熟悉的柯布-道格拉斯效用函数。

8.7 显示偏好

在研究消费者行为时我们将偏好作为原生概念，并由此推导出效用自大化模型施加在可观测到的需求函数的限制。这些限制基本上是斯勒茨基限制，即替代矩阵是对称的和负半定的。

这些限制在原则上是可以观测到的，但在实践中还存在着一些问题。毕竟谁真正看到过需求函数？在现实中我们最多能看到的是不同环境下的一组选择。例如，对于消费者的行为我们可能有一些观测数据，即给定一组价格 p^t 和消费者相应选择的消费束 x^t ，其中 $t = 1,...,T$ 。我们如何知道这些数据来源于追求利润最大化的消费者？

如果对于满足 $p^tx^t \geq p^tx$ 的所有 x 都有 $u(x^t) \geq u(x)$ ，其中 $t = 1,...,T$ ，我们就说效用函数 $u(x)$ 使得这个观测到的行为合理化。也就是说，如果 $u(x)$ 在被选中的消费束上达到了最大值，它就使得观测到的行为合理化。假设这些数据是由这样的最大化过程产生的。那么，观测到的行为应该满足什么样的可观测到的限制？

如果我们对$u(x)$不加任何假设限制，那么上述问题的答案是平凡的，即答案为无任何限制。例如，假设$u(x)$是一个常数值函数，因此消费者对于所有消费束都是无差异的。于是不存在对观测到的消费选择模式的限制：一切都有可能。

为使这个问题有趣，我们必须排除这样的平凡情形。最容易的做法是要求潜在的效用函数满足局部非饱和性。我们的问题现在变为：一个满足局部非饱和性的效用函数，它对选择行为施加了什么样的可以观测到的限制？

首先，我们注意到如果$p^t x^t \geq p^t x$，则必然有$u(x^t) \geq u(x)$。由于消费者在本来可以选择x的情形下却选择了x^t，x^t的效用不会小于x的效用。在这种情形下，我们说x^t**被直接显示偏好于**x（x^t is directly revealed preferred to x）。根据这个定义以及根据数据是由效用最大化行为产生的假设，我们可以断言"$x^t R^D x$意味着$u(x^t) \geq u(x)$"。

假设$p^t x^t > p^t x$。是否由此可推知$u(x^t) > u(x)$？不难看出，局部非饱和性意味着上述问题的答案是肯定的。因为我们从上一段可知，在这种情形下$u(x^t) \geq u(x)$；如果$u(x^t) = u(x)$，那么根据局部非饱和性可知，存在着充分接近于x的x'使得$p^t x^t > p^t x'$并且$u(x') > u(x^t) = u(x)$。这和效用最大化的假设矛盾。

如果$p^t x^t > p^t x$，我们说x^t被直接显示**严格**偏好于x，我们将其记为$x^t P^D x$。

现在假设有一系列这样的显示偏好之间的比较，比如$x^t R^D x^j$，$x^j R^D x^k$，…，$x^n R^D x$。在这种情形下，我们说x'**被显示偏好于**x，记为$x^t R x$。关系R有时称为关系R^D的**传递闭包**（transitive closure）。如果我们假设数据是由效用最大化行为产生的，则由此可推知"$x^t R x$意味着$u(x^t) \geq u(x)$"。

现在考虑两个观测值x^t和x^s。我们现在有办法确定$u(x^t) \geq u(x^s)$是否成立，也有观测性条件来确定$u(x^s) > u(x^t)$是否成立。显然，这两个条件不应该同时得到满足。这个条件可以表述为：

显示偏好的一般性公理（**G**eneralized **A**xiom of **R**evealed **P**reference，GARP）。若x^t被显示偏好于x^s，则x^s不可能被直接显示严格偏好于x^t。

使用我们前面定义的符号，可以将这个公理写为：

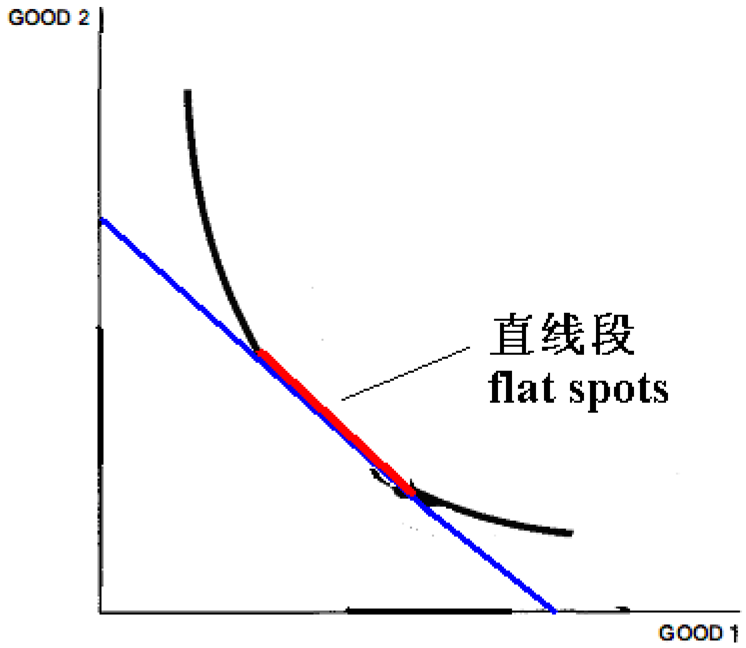
GARP。$x^t R x^s$意味着$x^s R^D x^t$不成立。换句话说，$x^t R x^s$意味着$p^s x^s \leq p^s x^t$.

正如 GAPR 的名字所意味的，它是各种其他显示偏好检验的一般化。下面是两个标准的条件。

显示偏好弱公理（**W**eak **A**xiom of **R**evealed **P**reference, WARP）。若$x^t R^D x^s$且x^t不等于x^s，则$x^s R^D x^t$不成立。

显示偏好强公理（**Strong Axiom of Revealed Preference, SARP**）。若 $x^t R x^s$ 且 x^t 不等于 x^s，则 $x^s R x^t$ 不成立。

上面这两个公理，每个都要求在一个预算中只能有唯一一个需求束，而 GARP 允许多个需求束。因此，GAPR 允许产生观测到的选择的无差异曲线存在直线段（flat spots）。如下图所示（此图为译者所加）。



图：GAPR 允许产生观测到的选择的无差异曲线存在直线段。

8.8 最大化的充分条件

如果数据 (p^t, x^t) 是由拥有非饱和偏好的追求效用最大化的消费者产生的，这些数据必定满足 GARP。因此，GARP 是效用最大化的可观测的结果。但是，它揭示了这个模型的所有蕴意了吗？如果某些数据满足这个公理，那么它必定来自效用最大化行为吗或者至少应该按照这个思路思考吗？GARP 是效用最大化的充分条件吗？

答案是肯定的。如果某个有限数据集与 GAPR 一致，则存在能使得观测到的行为理性化的效用函数。也就是说，存在能产生该行为的效用函数。因此，GARP 穷尽了效用最大化模型施加的种种限制。

下列定理是表达这个结论的最好方式。

阿弗里亚特定理（Afriat's theorem）。令 (p^t, x^t)（其中 $t = 1, \ldots, T$）是个关于价格向量和消费束的有限数量的观测值。则下列条件是等价的：

（1）存在能理性化这些数据的局部非饱和效用函数；

（2）这些数据满足 GARP；

（3）存在满足下列阿弗里亚特不等式的正数(u^t,λ^t)（其中$t=1,...,T$）：

$$u^s \le u^t+\lambda^t p^t(x^s-x^t) \text{ 对于所有 } t,s \text{ 成立。}$$

（4）存在能理性化这些数据的局部非饱和的、连续的、凹的且单调的效用函数。

证明。我们已经看到（1）意味着（2）。$(2)\Rightarrow(3)$的证明请参见 Varian（1982a），在此不再赘述。$(4)\Rightarrow(1)$的证明是平凡的。因此，我们只需证明$(3)\Rightarrow(4)$。

我们将说明下列效用函数能够完成证明$(3)\Rightarrow(4)$的任务。定义

$$u(x)=\min_t\{u^t+\lambda^t p^t(x-x^t)\}.$$

注意这个函数是连续的。只要$p^t\ge 0$而且不存在$p^t=0$，这个函数将是局部非饱和的且单调的。不难证明，这个函数也是凹的。在几何图形上，该函数就是有限个超平面的下包络。

我们需要证明该函数理想化了这些数据，也就是说，当价格为p^t时，该效用函数在x^t达到了约束最大化。首先我们证明$u(x^t)=u^t$。反正法。假设不是这样，则我们有：

$$u(x^t)=u^m+\lambda^m p^m(x^t-x^m)<u^t.$$

但这违背了阿弗里亚特不等式，因此$u(x^t)=u^t$。

现在假设$p^s x^s\ge p^s x$。由此可以推得

$$u(x)=\min_t\{u^t+\lambda^t p^t(x-x^t)\}\le u^s+\lambda^s p^s(x-x^s)\le u^s=u(x^s).$$

这表明对于满足$p^s x^s\ge p^s x$的所有x都有$u(x^s)\ge u(x)$。换句话说，$u(x)$理性化了观测到的选择。■

我们在证明阿弗里亚特定理过程中定义的效用函数，有着自然而然的解释。假设$u(x)$是理性化观测到的选择行为的一个凹且可微的效用函数。$u(x)$是可微的事实意味着它必定满足一阶条件

$$\mathrm{D}u(x^t)=\lambda^t p^t. \tag{8.3}$$

$u(x)$是凹函数的事实意味着它必定满足凹性条件

$$u(x^t)\le u(x^s)+\mathrm{D}u(x^s)(x^t-x^s). \tag{8.4}$$

将（8.3）式代入（8.4）式，可得

$$u(x^t)\le u(x^s)+\lambda^s p^s(x^t-x^s).$$

因此，阿弗里亚特数（Afriat numbers）u^t和λ^t可以解释为与观测到的选择相一致的效用水平和边际效用水平。

阿弗里亚特定理的最重要的含义在于（1）意味着（4）：如果存在能理性化这些数据的任何局部非饱和的效用函数，则必定存在能理性化这些数据的连续、单调且凹的效用函数。

这和第 6 章中的观测到的结论是一样的，在那里我们证明了，如果必要投入集存在非凸的部分，则生产者不会选择在非凸部分生产，因为成本非最小。

类似的结论对于效用最大化也成立。如果潜在的效用函数在某些点上存在“错误的”曲率（curvature），我们不会观测到消费者选择这些点，因为它们无法满足正确的二阶条件。因此，市场数据不允许我们拒绝偏好的凸性和单调性假设。

8.9 使用显示性偏好进行比较静态分析

由于 GARP 是效用最大化的必要且充分条件，它必然蕴涵着某些条件，这些条件与前面得到的比较静态结果类似。这包括将价格变化的效应进行斯勒茨基分解（分解为收入效应和替代效应），还包括自身替代效应为负的事实。

我们首先分析自身替代效应为负这个结论。当我们考虑价格的有限变化而非无穷小变化时，存在两种补偿性需求的概念。第一种补偿性需求是我们先前定义的自然而然的扩展，即如果我们变动收入水平来恢复原来的效用水平，那么商品的需求将会发生什么样的变化。也就是说，当价格从 p 变为 $p+\Delta p$ 时，商品 i 的补偿性需求数量为 $x_i(p+\Delta p, m+\Delta m) \equiv x_i(p+\Delta p, e(p+\Delta p, u))$，其中 u 是在初始状态 (p,m) 实现的（初始）效用水平。这个补偿需求概念称为希克斯补偿（Hicksian compensation）。

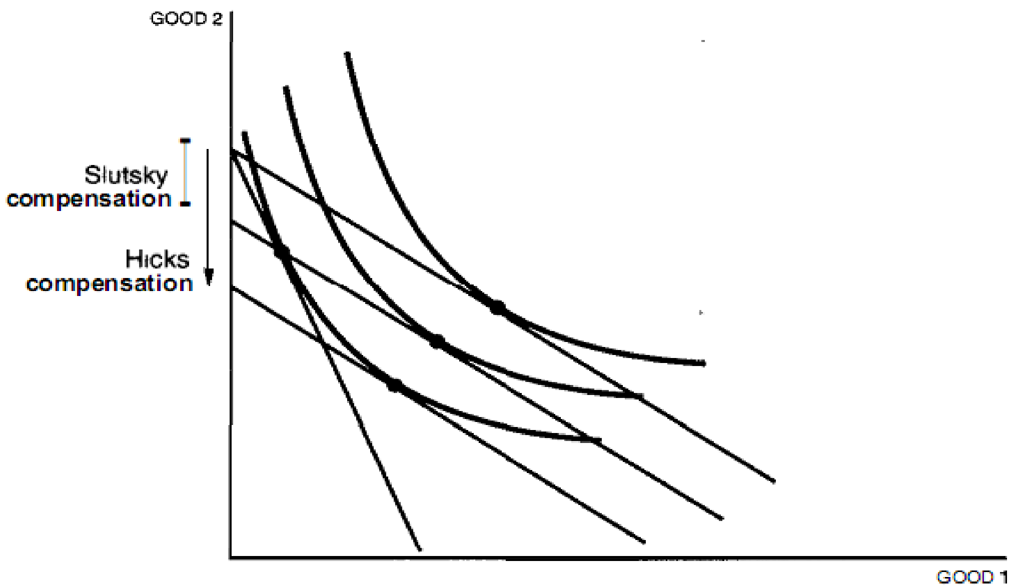


图 8.6：希克斯补偿和斯勒茨基补偿。希克斯补偿是调整收入使得消费者的效用水平维持在原来的水平上；斯勒茨基补偿是调整收入使得消费者恰好能买得起原来的消费束。

第二种补偿需求的概念称为斯勒茨基补偿（Slutsky compensation）。当价格从 p 变为 $p+\Delta p$ 时，相应调整收入使得恰好能买得起原来的**消费水平**，这就是斯勒茨基补偿。这种补偿可用下式描述。当价格从 p 变为 $p+\Delta p$ 时，我们变动收入 Δm，使得原来的消费水平

$x(p,m)$ 在新价格 $p+\Delta p$ 下仍是可行的。也就是说，

$$(p+\Delta p)x(p,m)=m+\Delta m\ 。$$

由于 $px(p,m)=m$ ，上式简化为 $\Delta p\ x(p,m)=\Delta m$ 。

图 8.6 说明了这两种补偿概念的区别。斯勒茨基补偿概念可以直接计算，无需知道偏好信息；但在分析性的研究工作中，使用希克斯补偿概念则更为简便。

对于无穷小的价格变动来说，没有必要区分这两个概念，因为在这种情形下，它们是相同的。只需要考察支出函数就能证明这个结论。如果商品 j 的价格变化为 dp_j ，为了维持效用水平不变，我们需要调整支出，调整量为 $(\partial e(p,u)/\partial p_j)dp_j$ 。

如果我们想让消费者能买得起原来的消费束，我们需要调整收入，调整量为 $x_j dp_j$ 。根据支出函数的微分性质可知，这两个调整量是相同的，即 $x_j dp_j=(\partial e(p,u)/\partial p_j)dp_j$ 。

不论你喜欢使用哪个补偿概念，你都可以使用显示性偏好来证明“补偿性自身价格效应是负的”。假设我们使用的是希克斯补偿概念。价格向量从 p 变为 $p+\Delta p$ ，因此补偿性需求变为 $x(p+\Delta p,m+\Delta m)$ ，其中 Δm 是使得 $x(p+\Delta p,m+\Delta m)$ 与 $x(p,m)$ 无差异的收入调整。

由于 $x(p,m)$ 和 $x(p+\Delta p,m+\Delta m)$ 彼此是无差异的，谁都无法显示偏好于另一个。也就是说，我们必定有

$$px(p,m)\le px(p+\Delta p,m+\Delta m)\ ,$$

$$(p+\Delta p)\ x(p+\Delta p,m+\Delta m)\le (p+\Delta p)\ x(p,m)\ 。$$

将这两个不等式相加，我们有

$$\Delta p\,[x(p+\Delta p,m+\Delta m)-x(p,m)]\le 0.$$

令 $\Delta x=x(p+\Delta p,m+\Delta m)-x(p,m)$ ，上式变为

$$\Delta p\ \Delta x\le 0.$$

假设只有一种商品（商品 i ）的价格发生了变化，因此 $\Delta p=(0,\dots,\Delta p_i,\dots,0)$ 。那么上式意味着 x_i 的变化与自身的价格变化方向相反。

现在假设我们使用的是斯勒茨基的补偿概念。我们使用的记号和前面相同，唯一不同是在这里，我们将 Δm 解释为：为了让消费者恰好能买得起原来的消费束，收入应变动多少。由于（根据假设）原来的消费束 $x(p,m)$ 在新价格水平 $p+\Delta p$ 下仍是可行的，那么消费者在 $p+\Delta p$ 下实际选择的消费束不可能被显示差于 $x(p,m)$ 。也就是说，

$$px(p,m)\le px(p+\Delta p,m+\Delta m)\ 。$$

根据我们对 Δm 构造的可知，

$$(p+\Delta p)\ x(p+\Delta p,m+\Delta m)=(p+\Delta p)x(p,m)\ 。$$

将前面的不等式减去上面的等式，可得

$$\Delta p\ \Delta x\le 0.$$

这个结果与使用希克斯补偿概念得到的结果是相同的。

8.10 离散形式的斯勒茨基方程

在前面，我们对一个涉及希克斯和马歇尔需求的恒等式进行微分，得到了连续形式的斯勒茨基方程。现在我们的任务是推导出离散形式的斯勒茨基方程。我们从下列算式恒等式入手：

$$\begin{aligned}&x_i(p+\Delta p,m)-x_i(p,m)\\&=x_i(p+\Delta p,m+\Delta m)-x_i(p,m)-[x_i(p+\Delta p,m+\Delta m)-x_i(p+\Delta p,m)]\end{aligned}$$

你可以验证一下，这个式子恒成立。

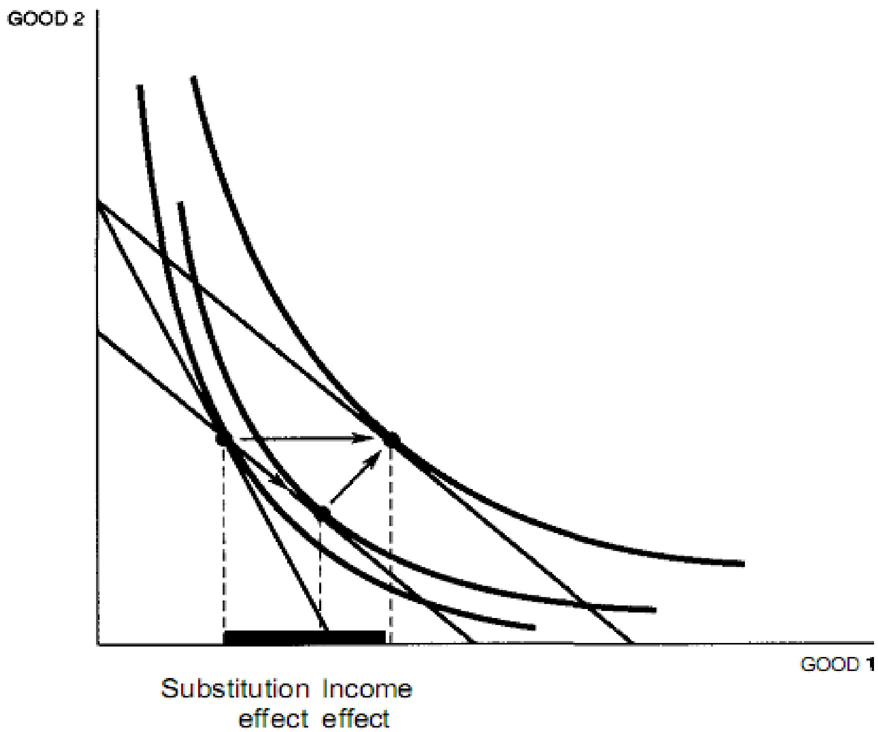


图 8.7：价格变动的斯勒茨基分解。首先将预算线绕着原来的消费束转动，然后向外平移到最终选择。

假设只有一种商品（商品 j ）的价格发生了变化，因此 $\Delta p=(0,...,\Delta p_j,...,0)$ 。于是斯勒茨基补偿变动 $\Delta m=x_j(p,m)\Delta p_j$ 。如果我们将前面的那个恒等式左右两侧同除以 Δp_j ，并使用 $\Delta p_j=\Delta m/x_j(p,m)$ 这个事实，可得

$$\begin{aligned}&\frac{x_i(p+\Delta p,m)-x_i(p,m)}{\Delta p_j}\\&=\frac{x_i(p+\Delta p,m+\Delta m)-x_i(p,m)}{\Delta p_j}-x_j(p,m)\frac{[x_i(p+\Delta p,m+\Delta m)-x_i(p+\Delta p,m)]}{\Delta m}\end{aligned}$$

使用 Δ 记号，可将上式简写为

$$\frac{\Delta x_i}{\Delta p_j} = \frac{\Delta x_i}{\Delta p_j}\Big|_{补偿} - x_j \frac{\Delta x_i}{\Delta m}\text{。}$$

注意，上式只不过是离散版本的斯勒茨基方程。上式左侧是说当商品 j 的价格变化时，商品 i 的需求将会如何变化。它被分解为替代效应和收入效应。替代效应是指，当商品 j 的价格变化，并相应调整收入使得消费者能够买得起原来的消费束时，商品 i 的需求将会如何变化。收入效应是指，当维持价格不变但收入变化引起的商品 i 的需求变化乘以商品 j 的需求。图 8.7 画出了价格变动的斯勒茨基分解。

8.11 复原偏好

由于显示性偏好条件是由效用最大化行为施加的一个完备约束集，它们必定包含潜在偏好的所有可得到的信息。如何使用显示性偏好关系来确定观测到的选择 x^t （其中 $t = 1,..,T$ ）之间的偏好关系，这个问题的答案在某种程度上是显而易见的。然而，如何使用显示性偏好关系来判断从未被观测到的选择之间的偏好关系，这个问题的答案就不那么明显了。

看清这个问题的最简单方法是举例说明。图 8.8 画出了我们观测到的消费者的一个选择行为 (p^1, x^1) 。这个选择意味着通过消费束 x^0 的无差异曲线是什么样子的？注意， x^0 以前从未被观测到；具体地说，我们没有直接的价格数据来判断 x^0 何时可能是最优选择。

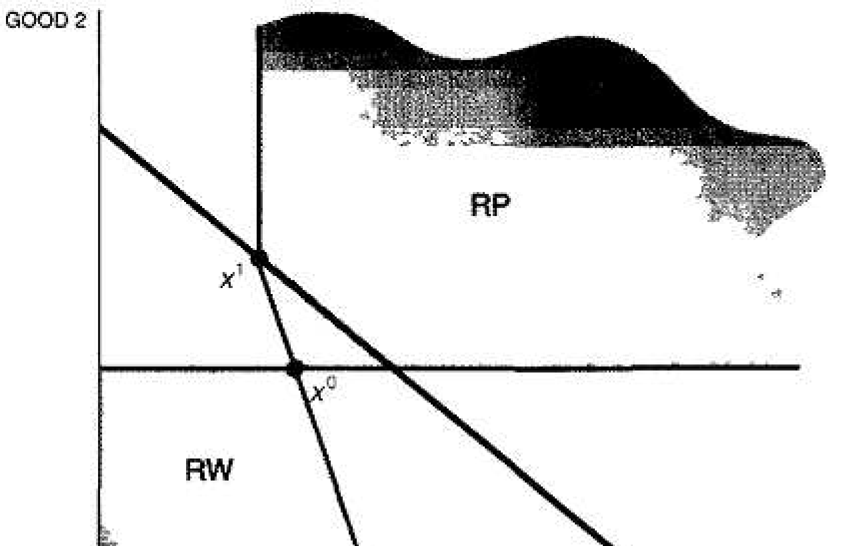


图 8.8：内界和外界。RP 是通过 x^0 点的无差异曲线的内界（inner bounds）；RW 的补集是其外界（outer bounds）。

下面我们尝试用显示性偏好来确定通过 x^0 点的无差异曲线的“界限”。首先，我们注意到 x^1 被显示偏好于 x^0 。假设偏好是凸的和单调的。那么所有位于连接 x^0 和 x^1 形成的线段上的消费束，都至少与 x^0 一样好，所有位于 x^0 东北方的消费束都至少和 x^0 一样好。将这个

消费束集合称为“被显示偏好于 x^0 的消费束组成的集合”，记为 $RP(x^0)$ 。不难证明这个集合是通过点 x^0 的上轮廓集（upper contour）的最优“内界”。为了推导出最优外界，我们必须考虑通过 x^0 的所有可能的预算线。令 $RW(x^0)$ 表示对于所有这些预算线来说，那些被显示劣于 x^0 的所有消费束组成的集合。无论使用哪条预算线，$RW(x^0)$ 内的消费束都比 x^0 差。

我们把通过点 x^0 的上轮廓集的外界定义为集合 $RW(x^0)$ 的补：$NRW(x^0)=$所有不在 $RW(x^0)$ 中的集合。这是最优外界，因为如果一个消费束不在 $NRW(x^0)$ 之中，那么该消费束不可能被显示偏好于 x^0 。为什么？因为根据我们对这两个集合的构造可知，一个消费束若不在 $NRW(x^0)$ 之中，则它必定在 $RW(x^0)$ 中，这样它就会显示劣于 x^0 。

在只有一个观测选择数据的情形下，边界不是很紧密（tight）。如果有很多关于选择的数据，这些边界将变得非常紧凑，有效地描述了位于它们之间的真实的无差异曲线。我们再举一个例子说明，见图 8.9。我们建议读者仔细考察这些边界是如何构建的，以便知道它们是怎么来的。一旦我们构建了上轮廓集的内界和外界，我们就复原了包含在观测到的需求行为中的偏好的所有信息。因此，集合 RP 和 RW 的构建类似于求解可积方程。

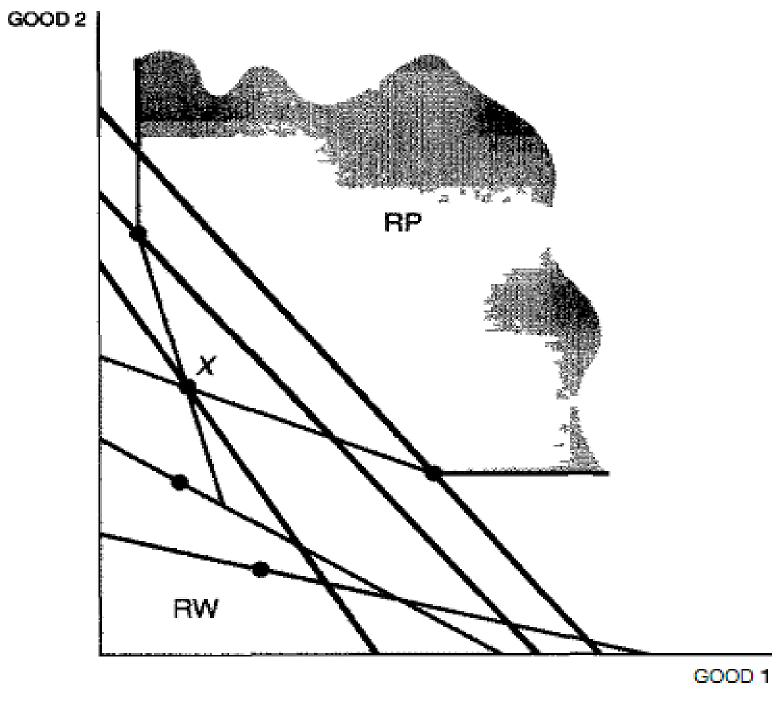


图 8.9：当观测数据较多时，内界和外界会非常紧凑。

直到现在，我们对 RP 和 RW 的构建都是通过图形进行的。然而，将这种分析推广到多种商品也是可行的。可以证明，决定一个消费束是否被显示偏好于或劣于另外一个消费束，涉及检验一组特定线性不等式是否有解。（本章结束）■

注释

本章中的斯勒茨基方程的证明使用了 McKenzie（1957）和 Cook（1972）的方法。可积性的详细讨论请参考 Hurwicz and Uzawa（1971）。显示性偏好的思想源于 Samuelson（1948），本章采用了 Afriat（1967）和 Varian（1982）的方法。至于使用显示性偏好推导斯勒茨基方程，我们采用了 Yokoyama（1968）的方法。

习题

8.1 弗兰克.费雪的支出函数为 $e(p,u)$。他对“开玩笑”（jokes）这种商品的需求为 $x_j(p,u)$，其中 p 为价格向量，$m \gg 0$ 为他的收入。证明开玩笑对弗兰克来说是正常商品当且仅当 $\partial^2 e / \partial p_j \partial u > 0$。

8.2 计算柯布-道格拉斯需求函数（假设商品有两种）的替代矩阵。证明该矩阵的对角元素为负的，交叉价格效应是对称的。

8.3 假设某个消费者的需求函数为线性的 $x = ap + bm + c$。为了找到用货币度量的效用函数，你需要使用微分方程，将这个微分方程写出来。如果可能求出这个方程的解。

8.4 假设某个消费者的需求函数是半对数的（semi-log），即 $\ln x = ap + bm + c$。为了找到用货币度量的效用函数，你需要使用微分方程，将这个微分方程写出来。如果可能求出这个方程的解。

8.5 某个消费者的效用函数为 $u(x_1, x_2) = x_1^{3/2} x_2$，他的预算约束为 $3x_1 + 4x_2 = 100$，求该消费者的需求束。

8.6 如果某个消费者的效用函数为 $u(x_1, x_2) = x_1^{1/2} x_2^{1/3}$，他的预算约束为 $m = p_1 x_1 + p_2 x_2$，求 $x(p,m), v(p,m), h(p,u)$ 和 $e(p,u)$。

8.7 如果某个消费者的效用函数为 $u(x_1, x_2) = (x_1 - \alpha_1)^{\beta_1}(x_2 - \alpha_2)^{\beta_2}$，他的预算约束为 $m = p_1 x_1 + p_2 x_2$，求 $x(p,m), v(p,m), h(p,u)$ 和 $e(p,u)$。验证替代矩阵的项 $\left(\dfrac{\partial h_j(p,u)}{\partial p_i}\right)$ 的对称性。

8.8 假设习题 8.6 中的效用函数现在为 $u^*(x_1, x_2) = \dfrac{1}{2}\ln x_1 + \dfrac{1}{3}\ln x_2$，求 $x(p,m), v(p,m), h(p,u)$ 和 $e(p,u)$，现在这些结果与习题 8.6 中的结果一样吗？为什么？证明如果我们用 e^{u^*} 替换 u^*，这些结果仍然成立。

8.9 假设某消费者的偏好可用效用函数 $u = \phi(x)$ 表示，并且假设我们已计算出了他的支出函数、间接效用函数和马歇尔需求函数。如果他的那些偏好也可以用效用函数 $u^* = \psi(\phi(x))$ 表

示，其中$\psi(\cdot)$是一个单调递增的函数。证明现在$e(p,u)$变为$e(p,\psi^{-1}(u^*))$，$v(p,m)$变为$\psi(v(p,m))$，$h(p,u)$变为$h(p,\psi^{-1}(u^*))$。你还要验证马歇尔需求函数$x(p,m)$没发生变化。

8.10 消费者迪夫的二期消费模型可用效用函数$u(x_1,x_2)$表示，其中x_1表示他在第 1 期的消费，x_2表示他在第 2 期的消费。该消费者的禀赋为$(\overline{x}_1,\overline{x}_2)$，他可以在这两个时期消费这些禀赋，也可以在当前消费和未来消费之间进行交易。因此，他的预算约束为

$$p_1x_1+p_2x_2=p_1\overline{x}_1+p_2\overline{x}_2,$$

其中p_1和p_2分别为第 1 期和第 2 期商品的价格。

（a） 推导出这个模型的斯勒茨基方程。（注意现在消费者迪夫的收入取决于他的禀赋的价值，而他的禀赋的价值又取决于价格：$m=p_1\overline{x}_1+p_2\overline{x}_2$。）

（b） 假设他的最优选择是使得$x_1<\overline{x}_1$。如果p_1下降，他的状况会变好还是变坏？如果p_2下降呢？

（c） 消费品的报酬率为多少？

8.11 某个消费者消费两种商品（商品 1 和 2）。当商品的价格为（2,4）时，他的需求为（1,2）；当价格为（6,3）时，他的需求为（2,1）。除了这个变化之外，其他都不变。这个消费者是追求效用最大化的吗？

8.12 假设间接效用函数为$v(p,y)=f(p)y$。支出函数是什么样的？间接补偿函数函数$\mu(p;q,y)$是什么样的（以函数$f(\cdot)$和y表示）？

8.13 假设效用函数为$u(x_1,x_2)=\min\{x_2+2x_1, x_1+2x_2\}$。

（a）画出$u(x_1,x_2)=20$的无差异曲线。将$u(x_1,x_2)\geq 20$的区域涂上阴影。

（b）如果唯一的最优选择是$x_1=0$，求p_1/p_2。

（c）如果唯一的最优选择是$x_2=0$，求p_1/p_2。

（d）如果x_1和x_2都不等于 0，而且最优选择是唯一的，求p_1/p_2。

8.14 在当前的税收法律下，人们每年最高可向自己的个人退休账户储蓄 2000 元，这种储蓄享有税收优惠。如果某个人在某个时点的收入为Y，他想将收入用于以下三项：消费C、个人退休账户储蓄S_1、以及普通储蓄S_2。假设他的“简化型”（reduced form）效用函数为

$$U(C,S_1,S_2)=S_1^{\alpha}S_2^{\beta}C^{\gamma}.$$

（这是个简化型的效用函数，因为参数不是真正的外生偏好参数，它还包括财产的税收待遇等。）该消费者的预算约束为：

$$C+S_1+S_2=Y,$$

再假设他向个人退休账户存入的钱数上限以L表示。

（a） 如果L对他的约束是不紧的（即他实际存入个人退休账户的钱数达不到L），求

他对 S_1 和 S_2 的需求函数。

（b） 如果 L 对他他的约束是紧的（binding），求他对 S_1 和 S_2 的需求函数。

8.15 如果闲暇是一种低档商品，劳动供给函数的斜率是什么样的（是正、负还是无法确定）？

8.16 某个追求效用最大化的消费者的偏好是严格凸的、严格单调的；他消费两种商品：x_1 和 x_2；这两种商品的价格都为 1。他对这两种商品的消费量都是非负的。该消费者的每年收入都为 m。他的当前消费水平为 (x_1^*, x_2^*)，其中 $x_1^* > 0$，$x_2^* > 0$。假设下一年他得到了一笔奖金 $g_1 \le x_1^*$，但规定他只能将这笔奖金用于消费商品 1（如果他愿意，他可以拒绝接受这笔奖金）。

（a）（判断正误）"如果商品 1 是正常商品，那么这笔奖金对他的消费的影响，必定和不带附加条件的等额奖金对他消费的影响是相同的。" 如果这个结论是正确的，证明它为真。若是错误的，请证明它为假。

（b）（判断正误）"如果商品 1 在该消费者所有收入 $m > x_1^* + x_2^*$ 上，对于来说都是低档商品，那么如果他得到了一笔规定只能用于购买商品 1 的奖金 g_1，那么这笔奖金对他的消费影响，必定和不带附加条件的等额奖金对他消费的影响是相同的。" 如果这个结论是正确的，证明它为真。若是错误的，请说明如果给他这笔奖金他会怎么做。

（c）假设该消费者的偏好是位似的（homothetic），他当前的消费为 $x_1^* = 12, x_2^* = 36$。先画一个图，将 g_1 画在横轴上，将 x_1 画在纵轴上。使用这个图说明：如果该消费者的正常收入为 $m = 48$，并且假设该消费者得到了一笔只能用于购买商品 1 的奖金 g_1，那么他对商品 1 的需求是什么样的？g_1 为多大时，他的需求曲线将会出现一个拐折（kink）?[在回答这个问题之前，请先思考一下。给出数值解。]

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 9 章：需求

曹乾　译

(东南大学　caoqianseu@163.com)

9 需求

在本章，我们将分析需求行为的几个主题。这些主题中的大部分涉及到特别形式的预算约束，或者涉及到导致特别需求行为的偏好。在很多环境中，这样的特殊情形非常便于分析，我们也需要理解它们是如何运行的。

9.1 含有禀赋的预算约束

在前几章的消费者行为理论中，我们假设消费者的收入是外生的。但在更详细的消费者行为模型中，需要考虑收入是怎么产生的。比较常规的方法是认为消费者拥有各种商品的禀赋（endowment）$\omega = (\omega_1, ..., \omega_k)$，这些禀赋可以按照当前的市场价格 p 出售。这样，消费者的收入为 $m = p\omega$，他可以使用这些收入购买其他商品。

效用最大化问题变为

$$\max_x u(x)$$

$$\text{使得 } px = p\omega$$

这个问题可用标准方法求解，从而得到需求函数 $x(p, p\omega)$。商品 i 的净需求为 $x_i - \omega_i$。消费者对某种商品的净需求可能为正也可能为负，这取决于他对该商品的需求是大于还是小于他的该商品禀赋的数量。

在这个模型中，价格影响到消费者所拥有的商品（禀赋）的价值，也影响到消费者想要出售的商品价值。使用斯勒茨基方程可清楚地看到这一点。下面我们进行推导。首先，将需求对价格进行微分：

$$\frac{dx_i(p, p\omega)}{dp_j} = \frac{\partial x_i(p, pw)}{\partial p_j}\Big|_{p\omega=\text{常数}} + \frac{\partial x_i(p, p\omega)}{\partial \text{m}}\omega_j$$

上式右侧第一项是维持收入不变时将需求对价格求导。右侧第二项是需求对收入的导数乘以收入的变动量。右侧第一项可用斯勒茨基方程展开。合并同类项可得

$$\frac{dx_i(p, p\omega)}{dp_j} = \frac{\partial h_i(p, u)}{\partial p_j} + \frac{\partial x_i(p, p\omega)}{\partial \text{m}}(\omega_j - x_j).$$

现在收入效应取决于商品 j 的净需求而不是总需求。考虑正常商品的情形，当商品价格上升时，替代效应和收入效应都使需求减少。但是假设某消费者是该商品的净出售者。那么，他的实际收入会增加，这种额外的禀赋收入效应可能会导致他增加该种商品的消费。

劳动供给

假设消费者选择两种商品：消费和劳动。他的非劳动收入为 m 。令 $v(c,l)$ 表示消费和劳动的效用，将这个效用最大化问题写为

$$\max_{c,l} v(c,l)$$

$$使得\ pc = wl + m.$$

这个问题和我们以前研究的问题看上去稍有不同：劳动可能是一种“厌恶品”，而不是消费者喜欢的商品，并且劳动出现在预算约束的右侧。

然而，将这个问题转换为标准形式并不难。令 $\overline{L}$ 表示消费者能够工作的最长时间，则 $L = \overline{L} - l$ 就是闲暇时间。消费和闲暇的效用函数为 $u(c,\overline{L}-l) = v(c,L)$ 。相应地，我们可以将效用最大化问题改写为

$$\max_{c,l} u(c,\overline{L}-l)$$

$$使得\ pc + w(\overline{L}-l) = w\overline{L} + m.$$

或者，使用 $L = \overline{L} - l$ 的定义，我们写成

$$\max_{c,L} u(c,L)$$

$$使得\ pc + wL = w\overline{L} + m.$$

上述问题的形式和我们以前研究的形式基本是一样的。此处消费者按照价格 w“出售”他的劳动禀赋并且买回一些闲暇。

根据斯勒茨基方程可以计算出当工资率变动时闲暇的需求是如何变动的。我们有

$$\frac{dL(p,w,m)}{dw} = \frac{\partial L(p,w,u)}{\partial w} + \frac{\partial L(p,w,u)}{\partial m}[\overline{L} - L].$$

根据定义可知方括号内的项是非负的，在现实中几乎肯定为正。因此，上式是说闲暇需求的导数，等于一个负数和一个正数的加和，所以它的符号在本质上是不明确的（即可能为正也可能为负）。换句话说，工资率的增加可能导致劳动供给的增加也可能减少。

本质上，工资率的增加倾向于使得劳动的供给增加，这是因为闲暇此时变得更昂贵——多工作一些会得到更多的消费。但是，与此同时，工资率的上升也使得你更有钱，而这又会增加你对闲暇的需求。

9.2 位似效用函数

函数 $f: R^n \to R$ 称为一次齐次的，若 $f(tx) = tf(x)$ 对于所有 $t > 0$ 成立。方程 $f(x)$ 是**位似的**（homothetic），若 $f(x) = g(h(x))$，其中 g 是一个严格增函数 h 是一个一次齐次函数。第 26 章进一步讨论了这种函数的数学性质。

经济学家通常发现假设效用函数是齐次的或位似的，将给分析带来方便。事实上，效用理论中的这两个概念存在着稍许区别。位似函数是齐次函数的单调变换，但效用函数的单调变化仍代表着原来的偏好。因此假设偏好可用位似函数代表，等价于假设该偏好可用一次齐次的函数代表。如果消费者的偏好可用位似效用函数表示，经济学家说该消费者的偏好是位似的。

我们在生产理论中已经知道，如果生产函数是一次齐次的，则成本函数可以写成 $c(w, y) = c(w)y$。从这个结论可知，如果效用函数是一次齐次的，则支出函数可以写成 $e(p, u) = e(p)u$。

这反过来意味着间接效用函数可以写为 $v(p, m) = v(p)m$。于是，罗伊恒等式意味着需求函数的形式为 $x_i(p, m) = x_i(p)m$，即需求函数是收入的线性函数。这种特殊形式的“收入效应”在需求分析中比较有用，下面我们将看到这一点。

9.3 商品之间的加总

在很多环境下，使用某些“部分”（partial）最大化问题对消费者选择建模是合理的。例如，我们在为消费者对“肉类”消费选择建模时，不区分牛肉、猪肉和羊肉的数量。在大多数实证研究中，对这类商品的某些类型的加总是有必要的。

为了描述和此类消费决策可分离性有关的某些有用结果，我们必须引进一些新的符号。我们将消费束分割为两个“亚消费束”，因此消费束的形式为 (x, z)。例如，x 为不同肉的消费数量组成的向量，z 为所有其他商品消费数量组成的向量。

类似地，我们将价格向量分割为 (p, q)。其中 p 为各种肉类的价格组成的向量，q 为其他商品价格组成的向量。使用这些符号可将标准的效用最大化问题写为

$$\max_{x,z} u(x, z)$$

$$\text{使得 } px + qz = m. \tag{9.1}$$

我们感兴趣的问题是，在什么样的条件下，我们可以将 x 商品作为一个整体看待，从而研究它的需求问题，而不必担心这个需求在 x 商品（比如肉类）的组成部分之间如何分配（即分成猪肉、牛肉等）的问题。

一种方法是将这个问题使用下列数学方法阐述。我们想构建某个标量数量指标 X，以及某个标量价格指标 P，这两个指标分别为数量向量和价格向量的函数：

$$P = f(p)$$
$$X = g(x) \tag{9.2}$$

在上面的表达式中，P 是某个“价格指数”，它给出了商品的“平均价格”，而 X 是一个数量指数，它给出了各种肉的平均消费“数量”。我们希望找到构建这些价格和数量指数的方法，从而使得它们的行为就象普通价格和数量一样。

也就是说，我们希望找到一个新的效用函数$U(X,z)$，这个函数仅取决于商品 x 的数量指数，这个函数给与我们的答案正如我们求解(9.1)的整个最大化问题一样。更正式地，考虑问题

$$\max_{X,z} U(X,z)$$

$$\text{使得 } PX + qz = m.$$

数量指数 X 的需求函数为某个函数 $X(P,q,m)$。我们想知道何时下式才成立

$$X(P,q,m) \equiv X(f(p),q,m) = g(x(p,q,m)).$$

这要求我们使用两条不同的路线得到的 X 值是相同的：

1）首先使用 $P = f(p)$ 加总价格，然后在预算约束为 $PX + qz = m$ 的情形下最大化 $U(X,z)$。

2）首先在预算约束为 $px + qz = m$ 的情形下最大化$u(x,z)$，然后加总数量得到 $X = f(x)$。

在两种情形下，这类加总是可行的。第一种情形对价格运动施加了约束，这种情形称为希克斯可分性（Hicksian separability）。第二种情形对偏好的结构施加了约束，这种情形称为函数分割（functional separability）。

希克斯可分性

假设价格向量总是与某个固定不变的基价格向量 p^0 成比例，因此 $p = tp^0$，其中t为标量。如果商品 x 是各种肉，这个条件要求各种肉的相对价格保持不变——（绝对）价格以相同的比例增加或减少。

根据前面的描述，我们定义商品 x 的价格指数和数量指数如下

$$P = t$$
$$X = p^0 x.$$

我们将与这些指数相伴的**间接**效应函数定义如下：

$$V(P,q,m) = \max_{x,z} u(x,z)$$

$$\text{使得 } Pp^0 x + qz = m.$$

容易验证这个间接效用函数具有所有通常的性质：它是拟凸的，它关于价格和收入是齐次的，等等。特别地，直接运用包络定理可知，根据罗伊恒等式，我们可以还原商品 x 的需求函数：

$$X(P,q,m) = -\frac{\partial V(P,q,m)/\partial P}{\partial V(P,q,m)/\partial m} = p^0 x(p,q,m).$$

这个计算表明 $X(P,q,m)$ 对于消费商品 x 来说，是一个合适的数量指数：如果我们首先加总价格然后最大化 $U(X,z)$，我们得到的结果和首先最大化 $u(x,z)$ 然后再加总数量得到的结果是一样的。

我们可以求出与 $V(P,q,m)$ 对偶的直接效用函数，计算方法如下：

$$U(X,z) = \min_{P,q} V(P,q,m)$$

$$\text{使得 } PX + qz = m.$$

根据该直接效用函数的构造可知，它具有下列性质

$$V(P,q,m) = \max_{X,z} U(X,z)$$

$$\text{使得 } PX + qz = m.$$

因此，使用这种方法构造出来的价格指数和数量指数，能保证它们的行为与普通价格、普通数量一样。

两种商品情形下的模型

当我们研究一种商品的需求时，我们经常使用希克斯加总。在这种情形下，将商品 z 看成一种商品，而将商品 x 看成“所有其他商品”。最大化问题因此为

$$\max_{x,z} u(x,z)$$

$$\text{使得 } px + qz = m.$$

假设商品 x 的相对价格维持不变，因此 $p = Pp^0$ 。也就是说，价格向量是某个基价格向量 p^0 乘以某个价格指数 P 。于是希克斯加总表明我们可以将商品 z 的需求写为

$$z = z(P,q,m).$$

由于这个需求函数是零次齐次的，我们也可以将其写为

$$z = z(q/P, m/P).$$

这个式子是说商品 z 的需求取决于商品 z 相对于“所有其他商品”的价格，以及取决于收入除以“所有其他商品”的价格。在实践中，所有其他商品的价格指数通常选用某些标准的消

费者价格指数（CPI）。商品 z 的需求变成只有两个变量的函数：商品 z 相对于 CPI 的价格以及相对于 CPI 的（相对）收入。

函数可分性

分解消费者的消费决策的第二种情形称为函数可分性（functional separability）。假设潜在的偏好排序具有下列性质

$$(x,z) \succ (x',z) \text{ 当且仅当 } (x,z') \succ (x',z')$$

对于所有消费束 x, x', z 和 z' 成立。这个条件是说如果对于另外商品的某些选择来说，x 比 x' 更受偏好，那么对于另外商品的所有选择来说，x 比 x' 更受偏好。或者，更简洁地说，商品 x 的偏好独立于商品 z 的偏好。

如果这个“独立性”得以满足，而且偏好是局部非饱和的，则可以证明商品 x 和商品 z 的效用函数可以写成 $u(x,z) = U(v(x),z)$，其中 $U(v,z)$ 是 v 的增函数。也就是说，商品 x 和商品 z 产生的总效用，可以写成商品 x 的次效用（subutility）$v(x)$ 和商品 z 消费水平的函数。

如果效用函数可以写成这种形式，我们说这种效用函数是弱可分的（weakly separable）。可分性意味着效用最大化问题有什么样的结构？和往常一样，我们把需求函数写为 $x(p,q,m)$ 和 $z(p,q,m)$。令 $m_x = px(p,q,m)$ 表示商品 x 的最小支出。

可以证明，如果总效用函数是弱可分的，则商品 x 的最优选择可以通过求解下列的次效用最大化问题而获得：

$$\max_x v(x)$$

$$\text{使得 } px = m_x. \qquad (9.3)$$

这就是说如果我们知道花费在商品 x 上的支出 $m_x = px(p,q,m)$，我们可以求解次效用最大化问题从而确定商品 x 上的最优选择。换句话说，商品 x 的需求只是商品 x 的价格以及花费在商品 x 上的支出 m 的函数。其他商品的价格只有在确定花费在商品 x 上的支出时才和这个问题有关。

上述结论不难证明。假设 $x(p,q,m)$ 不是上述最大化问题的解。于是令 x' 是 x 的另一个值，它满足预算约束并且产生严格更大的次效用。那么 (x',z) 产生的总效用高于 $(x(p,q,m), z(p,q,m))$，这和需求函数的定义矛盾。

需求函数 $x(p,m)$ 有时称为条件需求函数（conditional demand functions），这是因为它们给出商品 x 的需求——这个需求取决于这些商品支出水平。因此，例如，牛肉的需求是牛肉价格、猪肉价格和羊肉价格以及这些肉的总支出的函数。

令 $e(p,v)$ 表示（9.3）式中的次效用最大化问题的支出函数。这个函数告诉我们在价格

为 p 时花费在商品 x 上的支出为多大时才能实现次效用 v 。

不难看出我们可以将消费者的总效用最大化问题写为

$$\max_{v,z} u(v,z)$$

$$\text{使得} e(p,v)+qz=m.$$

这种形式几乎就是我们想要的形式了：v 是商品 x 的一个合适的数量指数，但是价格指数却不怎么正确。我们想要的是 P 乘以 X ，但是我们现在拥有的是价格的非线性函数以及 $X=v$ 。

为了得到一个与数量指数呈线性关系的预算约束，我们需要假设次效用函数具有特别的结构。例如，假设次效用函数是位似的。于是从第5章我们知道，可以将 $e(p,v)$ 写成 $e(p)v$ 。因此，我们可以选择数量指数为 $X=v(x)$ ，价格指数为 $P=e(p)$ ，效用函数为 $U(X,z)$ 。我们通过求解下列问题

$$\max_{X,z} U(X,z)$$

$$\text{使得} PX+qz=m.$$

得到的 X ，同通过下列方法得到的 X 是完全一样的：

首先求解

$$\max_{x,z} u(v(x),z)$$

$$\text{使得} px+qz=m.$$

然后使用 $X=v(x)$ 将 x 加总。

在上述表达方法中，我们可以将消费决策想象成分为两个阶段：首先，消费者通过求解总效用最大化问题得到复合商品（例如肉类）的消费量，它是肉类价格指数的函数；其次，在其他肉的价格以及肉类总支出给定的情况下，消费者选择牛肉的消费量，它是次效用最大化问题的解。使用这种两阶段预算过程的思想可以非常方便地分析需求问题。

9.4 消费者之间的加总

我们已经研究过消费者需求函数 $x(p,m)$ 的加总。现在我们考虑消费者的一个集合，该集合包含 $i=1,...,n$ 个消费者，每个消费者对于 k 种商品拥有一个需求函数，因此消费者 i 的需求函数是一个向量 $x_i(p,m_i)=(x_i^1(p,m_i),...,x_i^k(p,m_i))$ 。注意，此处我们已经稍微修改了一下我们使用的符号：商品种类现在用上标表示而消费者用下标表示。加总需求函数（aggregate demand function）定义为 $X(p,m_1,...,m_n)=\sum_{i=1}^n x_i(p,m_i)$ 。商品 j 的加总需求记为 $x^j(p,m)$ ，其中 m 表示收入向量 $(m_1,...,m_n)$ 。

加总需求函数继承了个人需求函数的某些性质。例如，如果个人需求函数是连续的，则加总需求函数当然是连续的。

个人需求函数的连续性是加总需求函数连续性的充分条件，但不是必要条件。例如，考虑洗衣机的需求。似乎下列假设是合理的：大部分消费者想要且只想要一台洗衣机。因此，消费者 i 的需求函数的形状将如图 9.1 所示的那样。

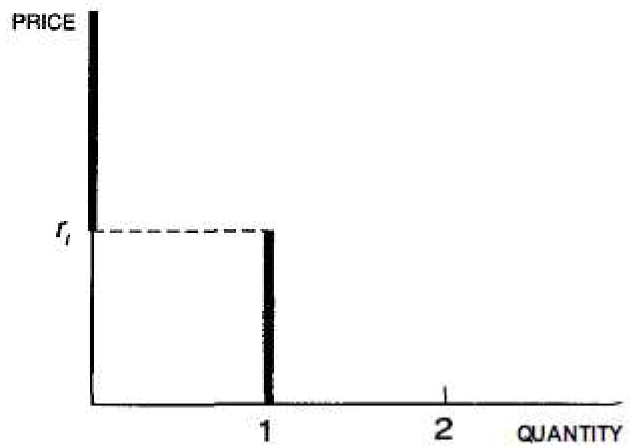


图 9.1：离散商品的需求。在任何大于 r_i 的价格水平上，消费者对洗衣机的需求为零。如果价格小于或等于 r_i，消费者将会需求一台洗衣机。

价格 r_i 称为消费者 i 的保留价格。如果消费者的收入和偏好改变，我们可以预期保留价格将有好几个。洗衣机的加总需求 $X(p)=$ 保留价格至少为 p 的那些消费者的个数。如果消费者的数量很多而且保留价格也比较分散，那么自然可以将这个需求函数看成是连续函数：如果价格稍微上升，只有少数几个消费者——“边际”消费者——将决定不购买该商品。即使这些边际消费者的需求变动是不连续的，加总需求的变动量也只会很小。

加总需求函数还从个人需求函数继承了哪些性质？是否存在加总版本的斯勒茨基方程或者加总版本的显示偏好强公理？不幸的是，上述问题的答案都是否定的。实际上，加总函数从个人需求函数继承的性质只有齐次性和连续性。因此，总体上来说，消费者理论未对总需求行为施加限制。

然而，在某些情形下，加总需求函数的行为可能就让人认为它是由一个“具有代表性的”消费者产生的。下面我们就来看一个这样的情形。

假设所有单个消费者的间接效用函数都为高曼（Gorman）形式：

$$v_i(p,m_i)=a_i(p)+b(p)m_i.$$

注意，$a_i(p)$ 可能随着消费者的不同而不同，但是 $b(p)$ 却被假设为对于所有消费者都

是相同的。根据罗伊恒等式可知，消费者i对商品 j 的需求将具有下列形式

$$v_i^j(p,m_i)=\alpha_i^j(p)+\beta^j(p)m_i. \tag{9.4}$$

其中，

$$\alpha_i^j(p)=-\frac{\dfrac{\partial a_i(p)}{\partial p_j}}{b(p)}$$

$$\beta^j(p)=-\frac{\dfrac{\partial b(p)}{\partial p_j}}{b(p)}.$$

注意，商品 j 的边际消费倾向$\partial x_i^j(p,m_i)/\partial m_i$，和任何消费者的收入水平无关，而且对于每个消费者来说它是固定不变的，因为$b(p)$对于每个消费者来说都是相同的。商品 j 的加总需求将具有下列形式

$$X^j(p,m^1,...,m^n)=-\left[\sum_{i=1}^n\frac{\dfrac{\partial a_i}{\partial p_j}}{b(p)}+\frac{\dfrac{\partial b(p)}{\partial p_j}}{b(p)}\sum_{i=1}^n m_i\right].$$

这个需求函数事实上可由一个具有代表性的消费者产生。他的具有代表性的间接效用函数为

$$V(p,M)=\sum_{i=1}^n a_i(p)+b(p)M=A(p)+B(p)M\ ,$$

其中$M=\sum_{i=1}^n m_i$ 。

证明方法是将罗伊恒等式应用到这个间接效用函数，并且注意它产生了（9.4）式中的需求函数。事实上可以证明，高曼形式的间接效用函数是能够对代表性的消费者模型进行加总的最为一般的函数形式。因此，高曼形式的函数不仅是代表性消费者模型成立的充分条件，还是必要条件。

完整证明这个事实比较复杂。我们提供一个简单的论证思路。为简单起见，假设有两个消费者。根据假设，商品 j 的加总需求可以写成

$$X^j=(p,m_1+m_2)=x_1^j(p,m_1)+x_2^j(p,m_2).$$

如果我们将上式先关于m_1微分，然后再关于m_2微分，可得到下列恒等式：

$$\frac{\partial X^j(p,M)}{\partial M}\equiv\frac{\partial x_1^j(p,m_1)}{\partial m_1}\equiv\frac{\partial x_2^j(p,m_2)}{\partial m_2}.$$

所以，所有消费者对商品 j 的边际消费倾向都是相等的。如果我们将上面的恒等式关于

m_1 再微分一次，可得

$$\frac{\partial^2 X^j(p,M)}{\partial M^2} \equiv \frac{\partial^2 x_1^j(p,m_1)}{\partial m_1^2} \equiv 0.$$

因此，消费者 1 对商品 j 的需求，从而消费者 2 对商品 j 的需求，关于收入是仿射的。所以，商品 j 的需求函数的形式为 $x_i^j(p,m_i)=\alpha_i^j(p)+\beta^j(p)m_i$。如果所有商品的情形都是这样的，那么每个消费者的间接效用函数必定都是高曼形式的。

具有高曼形式的一类特别效用函数，是位似效用函数。在这种情形下，间接效用函数的形式为 $v(p,m)=v(p)m$，这显然是高曼形式的。另外一类具有高曼形式的特殊效用函数，是拟线性效用函数。在这种性情下，$v(p,m)=v(p)+m$，这显然也是高曼形式的。位似效用函数与（或）拟线性效用函数的很多性质，也为高曼形式的效用函数所拥有。

9.5 反需求函数

在很多情形下，在表达消费者的需求行为时，我们需要即将商品价格描述为商品数量的函数。也就是说，给定某个商品向量 x，我们希望找到一个价格向量 p 和收入 m 使得 x 是需求束。

由于需求函数是零次齐次的，我们可以将收入固定在某个既定的水平，并且确定相对于这个收入水平来说的商品的（相对）价格。最方便的做法是将收入固定为 1，即 $m=1$。

在这种情形下，效用最大化问题的一阶条件为

$$\frac{\partial u(x)}{\partial x_i}-\lambda p_i=0 \quad \text{对于 } i=1,\dots,k$$

$$\sum_{i=1}^{k} p_i x_i = 1 \text{。}$$

我们希望消去第一组等式中的 λ。

为了做到这一点，将第一组中的每个等式乘以 x_i，然后加总可得

$$\sum_{i=1}^{k}\frac{\partial u(x)}{\partial x_i}x_i=\lambda\sum_{i=1}^{k}p_i x_i=\lambda.$$

将 $\lambda=\sum_{i=1}^{k}\frac{\partial u(x)}{\partial x_i}x_i$ 代入到第一组等式中，并将 p 表示为 x 的函数可得

$$p_i(x)=\frac{\dfrac{\partial u(x)}{\partial x_i}}{\sum_{i=1}^{k}\dfrac{\partial u(x)}{\partial x_i}x_i}. \tag{9.5}$$

给定任何需求向量 x，我们可以使用上式找到满足最大化一阶条件的价格向量 $p(x)$。如果效用函数是拟凹的，因此这些必要条件也是最大化的充分条件，于是我们就得到了反需求函数。

如果效用函数不是处处拟凹的，将会出现什么样的情形？那么，将会存在下列这样的一些商品束，这些商品束在任何价格下都不会成为需求束；位于无差异曲线非凸部分上的任何商品束都是上述这样的商品束。

上面这个反需求函数存在着对偶形式，我们可以从第 8 章的讨论推导出这个对偶形式的反需求函数。我们在第 8 章已经知道，对于满足预算约束的所有价格向量来说，需求束 x 必定**最小化**间接效用。因此，x 必定满足一阶条件

$$\frac{\partial v(p)}{\partial p_i} - \mu x_i = 0 \quad \text{对于 } i = 1, \ldots, k$$

$$\sum_{i=1}^{k} p_i x_i = 1.$$

现在将第一组式子中的每个式子乘以 p_i，然后相加可得 $\mu = \sum_{i=1}^{k} \frac{\partial v(p)}{\partial p_i} p_i$。将这个式子代入一阶条件，我们就得到了需求函数，它是以经过标准化的间接效用函数作为变量的函数：

$$x_i(p) = \sum_{i=1}^{k} \frac{\frac{\partial v(p)}{\partial p_i}}{\sum_{j=1}^{k} \frac{\partial v(p)}{\partial p_j} p_j} \qquad (9.6)$$

注意本节的漂亮的对偶：直接需求函数的表达式（9.6），和间接需求函数的表达式（9.5）具有相同形式。这个表达式也可以从经过标准化的间接效用函数的定义和罗伊恒等式推导出。

9.6 需求函数的连续性

直到现在我们一直假设我们的需求函数都具有良好行为；也就是说，假设它们是连续的甚至是可微的。这些假设是合理的吗？

借助于第 27 章的最大值定理，我们知道，只要需求函数是良好定义的，它们就是连续的，至少当 $p \gg 0$ 和 $m > 0$ 时是连续的；也就是说，只要 $x(p,m)$ 在价格收入组合为 (p,m) 时是**唯一的**最大化商品束，那么需求关于 p 和 m 就是连续的。

如果我们想保证对于所有 $p \gg 0$ 和 $m > 0$，需求都是连续的，那么我们需要保证需求束总是唯一的。为了做到这一点，我们所需要严格凸性这个条件。

唯一需求束（unique demanded bundle）。**如果偏好是严格凸的，那么对于每个** $p \gg 0$**，**

存在为一个商品束 x 使得 u 在消费者的预算集 $B(p,m)$ 上达到最大。

证明：假设 x' 和 x'' 都能使得 u 在 $B(p,m)$ 上最大。那么 $\frac{1}{2}x'+\frac{1}{2}x''$ 也在 $B(p,m)$ 中，但 $\frac{1}{2}x'+\frac{1}{2}x''$ 严格偏好于 x' 和 x''，矛盾。■

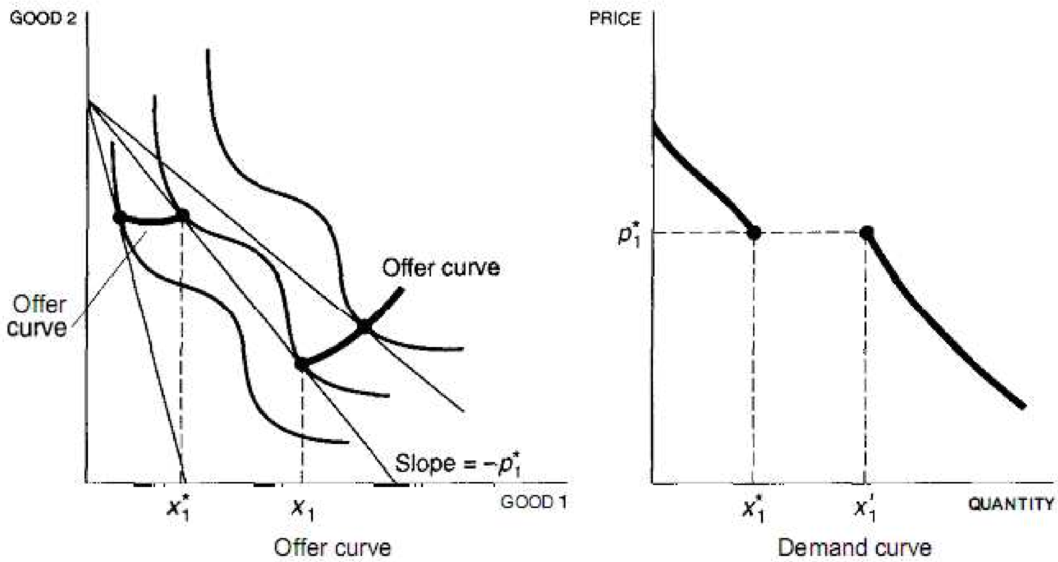


图 9.2：不连续的函数。由于偏好非凸，需求不连续。

粗略地说，如果需求函数是良好定义的、处处连续的，并且是从偏好最大化推导出的，那么潜在的偏好必定是严格凸的。如若不然，那么对于某个价格向量来说，将存在若干个最优商品束，如图 9.2 所示。注意，在图 9.2 描绘的情形下，价格的微小变动就能导致需求束的巨大变动：需求“函数”不连续。（本章结束）■

注释

条件需求请参考 Pollak（1969）。可分性请参考 Blackorby, Primont and Russell（1979）。可分性在消费者需求估计中的进一步发展和运用，可参考 Deaton and Muellbauer（1980）。商品加总的讨论基于 Gorman（1953）。加总中正的结果和负的结果可参见 Shafer and Sonnenschein（1982）。

习题

9.1 假设偏好是位似的。证明 $\frac{\partial x_i(p,m)}{\partial p_j}=\frac{\partial x_j(p,m)}{\partial p_i}$。

9.2 某种特殊商品的需求函数为 $x=a+bp$。求与该需求函数相伴的直接效用函数和间接效

用函数。

9.3 某种特殊商品的需求函数为 $x = a + bp + cm$，求与该需求函数相伴的直接效用函数和间接效用函数。（提示：如果想完全解答这个问题，你必须知道如何求解线性的非齐次的微分方程。如果你不会解，只需要写出方程就可以了。）

9.4 两种商品的需求函数为

$$x_1 = a_1 + b_{11} p_1 + b_{12} p_2$$

$$x_2 = a_2 + b_{21} p_1 + b_{22} p_2$$

需求理论对这些参数施加了什么样的限制？求与需求函数相伴的用货币衡量的效用函数。

9.5 在上个问题中，求与需求函数相伴的间接效用函数。

9.6 令 (q, m) 分别表示价格和收入，令 $p = q / m$。使用罗伊恒等式证明

$$x_i(p) = \frac{\dfrac{\partial v(p)}{\partial p_i}}{\sum_{j=1}^{k} \dfrac{\partial v(p)}{\partial p_j} p_j}。$$

9.7 考虑效用函数 $u(x_1, z_2, z_3) = x_1^a z_2^b z_3^c$。这个效用函数关于 (z_2, z_3) 是（弱）可分的吗？求商品 z 的次效用函数。给定商品 z 的支出 m，求商品 z 的条件需求。

9.8 假设有两种商品 x 和 y。某消费者对商品 x 的需求函数为 $\ln x = a - bp + cm$，其中 p 为商品 x 相对商品 y 的价格，m 为货币收入除以商品 y 的价格。

（a）为了确定产生这种需求行为的间接效用函数，你应该求解什么样的方程？

（b）这个微分方程的边界条件是什么？

9.9 某个消费者的效用函数为 $u(x, y, z) = \min\{x, y\} + z$。这三种商品的价格为 (p_x, p_y, p_z)，消费者必须用掉的钱数为 m。

（a）可以证明这个效用函数可以写为 $U(V(x, y), z)$ 的形式。分别求函数 $V(x, y)$ 和函数 $U(V, z)$ 的表达式。

（b）求这三种商品的需求函数。

（c）求间接效用函数。

9.10 假设有两种商品 x_1 和 x_2。令商品 1 和 2 的价格分别为 p_1 和 1。令收入为 y。某个消费者对商品 1 的需求为 $x_1 = 10 - p_1$。

（a）求商品 2 的需求函数。

（b）为了计算产生这些需求函数的收入补偿函数，你应该求解什么方程？

（c）求与这些需求函数相伴的收入补偿函数。

9.11 假设有两个消费者。消费者 1 的支出函数为$e_1(p_1, p_2, u_1) = u_1\sqrt{p_1 p_2}$，消费者 2 的效用函数为$u_2(x_1, x_2) = 43x_1^3 x_2^a$。

（a）求每个消费者对于每种商品的马歇尔（市场）需求函数。（两个消费者的收入分别用m_1和m_2表示）

（b）参数a取什么值时，存在和收入分配无关的加总需求函数？

曹乾（东南大学 caoqianseu@163.com）

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 10 章：消费者剩余

曹乾　译

(东南大学　caoqianseu@163.com)

10 消费者剩余

当经济环境改变时，消费者的状况可能会更好也可能更差。经济学家通常希望衡量消费者的状况如何受经济环境改变的影响，他们已经发明出几种方法来做此事。

福利变化的经典衡量方法是使用消费者剩余，初级微观经济学使用的就是这个工具。然而，消费者剩余只有在特定的环境中才能准确衡量福利变化。在本章，我们将介绍几个衡量福利变化的更一般方法。在这些更一般的方法中，消费者剩余只是一种特殊情形。

10.1 补偿变化和等价变化

我们首先考虑福利变化的"理想"衡量方法是什么样的。在最基本的层面上，我们想要的衡量方法是，它能衡量某些政策变化导致的效用变化。假设有两个预算 (p^0, m^0) 和 (p', m')，它们代表某个既定的消费者在不同政策框架下面对的价格和收入。通常我们将 (p^0, m^0) 视为现状，将 (p', m') 视为变化后的情形，尽管这不是唯一解释。

从 (p^0, m^0) 到 (p', m') 变化导致的福利变化，很自然地就可以用间接效用之差进行衡量：

$$v(p', m') - v(p^0, m^0).$$

如果这个效用之差为正，则政策的改变就是合理的，至少站在消费者的角度上是这样的；如果为负，则不需要改变现行的政策。

一般来说，做到这一步我们已经尽力了；效用理论在本质上纯粹是序数性质的，因此不存在明确的量化效用变化的方法。然而，在某些情形下，我们可以使用货币衡量消费者的福利变化。也许政策分析专家希望大致了解福利变化程度，以便以重要性为序安排各项工作。或者他们想比较不同消费者得到的利益或承担的成本。在上述这些情形下，有必要选择一种"标准"的效用差值衡量方法。一种合理的方法是使用以货币度量的（间接）效用函数（参见第 7 章）。

我们已经知道 $\mu(q; p, m)$ 衡量为了使得消费者为了和他在价格 p 和收入 m 的状一样好，在价格为 q 时他应该至少需要多少收入。也就是说，$\mu(q; p, m)$ 的定义为 $e(q; v(p, m))$。如果我们接受效用的这种衡量方法，我们发现上述效用的差值变为

$$\mu(q; p', m') - \mu(q; p^0, m^0).$$

剩下的问题是选择基价 q。选择明显有两个：我们可以令 q 为 p^0 或 p'。这样就得到以下两种衡量效用差值的方法：

$$\begin{aligned} EV &= \mu(p^0; p', m') - \mu(p^0; p^0, m^0) = \mu(p^0; p', m') - m^0 \\ CV &= \mu(p'; p', m') - \mu(p'; p^0, m^0) = m' - \mu(p'; p^0, m^0). \end{aligned} \tag{10.1}$$

第一种衡量方法称为等价变化（equivalent variation）。这种方法使用当前价格作为基价，它求解的是在当前价格水平下收入应该怎样变动，才能等价于政策变化对效用的影响。第二种衡量方法称为补偿变化（compensating variation）。这种方法使用新价格作为基价，它求解的是收入应该怎样变动才能补偿价格的变动对消费者的影响。（补偿发生在政策变化之后，因此补偿变化使用变化之后的价格。）

这两种方法都可以衡量价格变动引起的福利效应。它们的大小通常不同，因为钱的价值取决于相关价格是多少。但是它们的符号总是相同的，因为它们衡量的是相同的效用差值，只不过使用不同的效用函数而已。图 10.1 画出了两种商品情形下的等价变化和补偿变化。

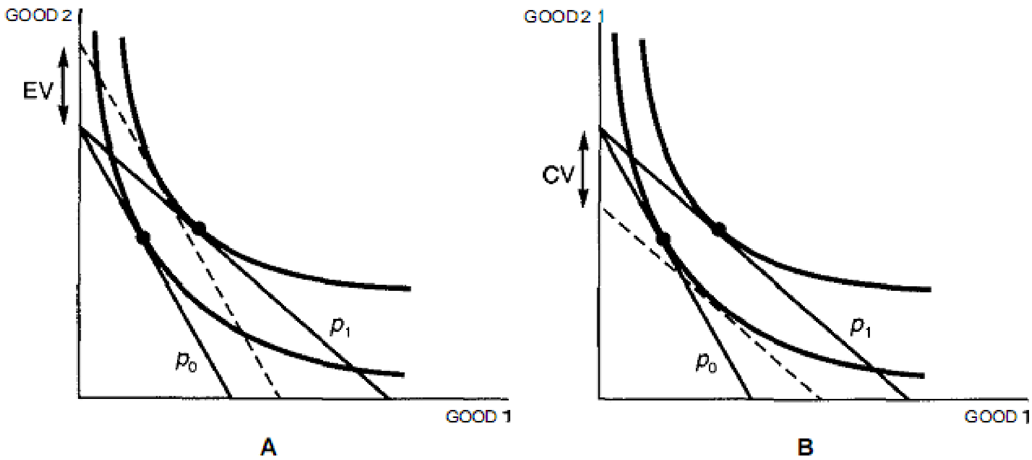


图 10.1：等价变化和补偿变化。在图 10.1 中，$p_2=1$，商品 1 的价格从 p_0 下降为 p_1。A 图画出的是收入的等价变化——在价格 p_0 下消费者应该额外需要多少钱，才能使他的状况和他在价格 p_1 下的状况一样好。B 图画出的是收入的补偿变动——应该从消费者手中拿走多少钱才能使他的状况和他在 p_0 下的状况一样好。

哪一种衡量方法更合适？这要取决于具体的环境以及你试图解决的问题。如果你想解决的是在新价格下的补偿方案的问题，那么使用补偿变化似乎更合理。然而，如果你只是想合理衡量“支付意愿”，那么等价变化的方法更好。原因有两个。首先，等价变化衡量的是在当前价格下的收入变动，而且对于决策者来说，使用货币的现值（当前价格下的价值）进行判断要比使用假想的价格进行判断容易。其次，在比较多个可能政策的优劣时，如果使用补偿变化方法，则对于每个政策来说，基价都不同，而等价变化则总是使用当前的价格，因此对于不同政策来说，基价是统一的。因此，如果评价的项目比较多，使用等价变化的方法更合适。

如果我们认为补偿变化和等价变化可以作为效用变化的指示器，在实践中我们如何计算它们？也就是说：在实践中我们如何衡量 $\mu(q;p,m)$？

在研究可积理论（第 8 章）时，我们已经回答了上述问题。在第 8 章，我们分析了如何通过观察需求行为为 $x(p,m)$ 复原 $\mu(q;p,m)$ 代表的偏好问题。给定任何观测到的需求行为，

你可以解出可积分方程组的解，至少理论上如此。从而可以推出相伴的以货币度量的效用函数。

在第 8 章我们已经知道如何根据需求函数（包括线性、非线性以及半对数函数）求出以货币度量的效用函数。在原理上，对于满足可积条件的任何函数，我们都可以进行类似的计算。

然而，在实践中通常在**另外**一个方向上进行参数刻画，因为这种方法更容易一些：首先确定间接效用函数的函数形式，然后根据罗伊恒等式推导出需求函数的函数形式。毕竟，对一个函数求导要比解一个偏微分方程组要容易得多！

如果我们确定了间接效用函数的参数形式，那么估计出与需求方程组相伴的参数，立即就可得到潜在效用函数的参数。一旦我们有了这些参数，我们就可以容易地推导出以货币度量的效用函数，以及补偿变化和等价变化。第 12 章详细介绍了这种方法。

当然这种方法只有在估计出的参数满足最优化模型的各种限制时，它才有意义。我们需要检验这些限制，目的在于判断它们对于我们特定的实证案例来说是否是合理的。如果是合理的，接下来就要估计出满足这些限制条件的相关参数。

总结一下：如果需求函数是可以观测到的，而且需求函数满足效用最大化的条件，那么补偿变化和等价变化也是可观测到的。可以观测到的需求行为可被用于构建福利变动的衡量方法，然后使用这种方法比较各个备选的政策。

10.2 消费者的剩余

衡量福利变化的经典工具就是**消费者的剩余**（consumer’s surplus）。如果 $x(p)$ 是某商品的需求函数，它是它自身价格的函数，那么与价格从 p^0 变化到 p^1 相伴的消费者的剩余为

$$CS = \int_{p^0}^{p'} x(t)dt .$$

上式就是需求曲线左侧位于价格线 p^0 和 p^1 之间的面积。可以证明当消费者的偏好可用拟线性效用函数表示时，消费者的剩余是一种精确衡量福利变动的方法。更准确地说，当效用是拟线性的，补偿变化等于等价变化，而且这两种变化都等于消费者剩余的积分。对于一般形式的效用函数来说，补偿变化不等于等价变化，消费者剩余就不再是福利变动的精确衡量方法。然而，即使效用不是拟线性的，消费者的剩余也是更准确衡量方法的合理近似。下面我们进一步分析这些思想。

10.3 拟线性效用

假设某个效用函数在单调变换后具有以下形式

$$U(x_0, x_1, ..., x_k) = x_0 + u(x_1, ..., x_k).$$

注意，该效用函数对于其中一种商品来说是线性的，但对于其他商品来说（可能）不是线性的。正因为此，我们将它称为**拟线性效用函数**（quasilinear utility function）。

在本节我们主要分析 $k=1$ 的情形，因此效用函数的形式为 $x_0 + u(x_1)$，本节所得的结论可以推广到任意数量的商品种类上。我们假定 $u(x_1)$ 为严格凹函数。

我们考虑下列形式效用函数的效用最大化问题：

$$\max_{x_0, x_1} x_0 + u(x_1)$$

$$\text{使得 } x_0 + p_1 x_1 = m.$$

将约束条件代入目标函数，可将上述问题化简为一个无约束的最优化问题

$$\max_{x_1} \ u(x_1) + m - p_1 x_1.$$

这个问题的一阶条件为

$$u'(x_1) = p_1$$

这个式子要求商品 1 的边际效用等于它的价格。

由一阶条件可知，商品 1 的需求只是它自身价格的函数，因此我们可以将需求函数写为 $x_1(p_1)$。于是商品 0 的需求函数可从预算线求出，$x_0 = m - p_1 x_1(p_1)$。将这些需求函数代入效用函数就可得到间接效用函数

$$v(p_1, m) = u(x_1(p_1)) + m - p_1 x_1(p_1) = v(p_1) + m,$$

其中 $v(p_1) = u(x_1(p_1)) - p_1 x_1(p_1)$。

这种方法很好，但它隐藏了一个问题。稍加思索，你就会明白商品 1 的需求不可能独立于所有价格和收入水平的收入。如果收入足够小，商品 1 的需求必然受到需求的约束。

假设我们将上述效用最大化问题写成下列形式，即明确列出对于 x_0 的非负约束：

$$\max_{x_0, x_1} x_0 + u(x_1)$$

$$\text{使得} \begin{array}{l} x_0 + p_1 x_1 = m \\ x_0 \geq 0. \end{array}$$

现在我们知道我们得到的解有两类，取决于 $x_0 > 0$ 还是 $x_0 = 0$。如果 $x_0 > 0$，那么我们得到的解和前面的那个最大化问题的解是一样的——商品 1 的需求仅取决于它自身的价格，与收入无关。如果 $x_0 = 0$，则间接效用由 $u(m/p_1)$ 给出。

想象消费者的最初收入 $m=0$，逐渐小额增加他的收入。效用的增额于是为

$u'(m/p_1)/p_1$。如果这个比值大于 1，则消费者将最初的一元钱花费在商品 1 比花费在商品 0 上更好。我们继续将钱花费在商品 1 上，直到额外一元钱的边际效用恰好等于 1；也就是说直到消费的边际效用等于商品的价格。然后，所有额外的收入都花费在商品 0 上。

拟线性效用函数经常用于福利经济学的分析，因为它的需求结构比较简单。需求仅取决于价格——至少收入足够高时是这样的——因此一点也不必担心收入效应。可以证明，这种效用函数简化了市场均衡的分析。你应该将这种情形认为我们模拟的是对收入不怎么敏感的商品的需求。想象一下你对纸或笔的需求：当你的收入变化时，你对它的需求变化多大？非常可能的是，任何收入的增加都被你用于其它商品的消费。

而且，使用拟线性效用的另外一个好处是可积性问题变得很简单。由于反需求函数为 $p_1(x_1) = u'(x_1)$，这意味着通过积分，我们就可以从反需求函数找到与商品 1 的特定消费水平相伴的效用：

$$u(x_1) - u(0) = \int_0^{x_1} u'(t)dt = \int_0^{x_1} p_1(t)dt.$$

选择消费 x_1 的总效用包含消费商品 1 的效用，加上消费商品 0 的效用：

$$u(x_1(p_1)) + m - p_1 x_1(p_1) = \int_0^{x_1} p_1(t)dt + m - p_1 x_1(p_1).$$

如果我们忽略常数 m，上式右侧正是商品 1 需求曲线下方的面积减去花费在商品 1 上的支出。或者说，这是商品 1 需求曲线左侧的面积。

得到上述结论的另外一种方法是从间接效用函数 $v(p_1) + m$ 开始分析。根据罗伊法则可知，$x_1(p_1) = -v(p_1)$。将这个式子积分可得

$$v(p_1) + m = \int_{p_1}^{\infty} x_1(t)dt + m.$$

这是需求曲线左侧、价格线 p_1 以上的面积。这正好是上一段落那个面积的另外一种表达方式。

10.4 拟线性效用和以货币度量的效用

假设效用函数的形式为拟线性的 $u(x_1) + x_0$。我们已经知道，对于这样的效用函数来说，需求函数 $x_1(p_1)$ 和收入无关。由前面的内容可知，我们可以通过对变量 p_1 积分的方法，可以得到与这个需求函数一致的间接效用函数。

当然，这个间接效用函数的任何单调变换还是一个间接效用函数，它们描述是消费者的同一行为。如果消费者的选择使得他的消费者剩余最大，那么当然也会使得消费者剩余的平方最大。

在前面我们已经知道，以货币度量的效用函数对于多种目的来说，都是个非常方便的效用函数。可以证明对于拟线性效用函数来说，需求的积分在本质上就是以货币度量的效用函数。

上述结论的证明方法是，写出可可积分方程，然后验证消费者剩余是这些方程的解即可。如果 $x_1(p_1)$ 是需求函数，则可积分方程为

$$\frac{d\mu(t;q,m)}{dt} = x_1(t)$$
$$\mu(q;q,m) = m.$$

直接计算可以验证这些方程的解为

$$\mu(p;q,m) = \int_p^q x_1(t)dt + m.$$

上式右侧正是与价格从 p 变为 q 相伴的消费者剩余，加上收入。

对于这种以货币度量的效用函数来说，补偿变化和等价变化的形式为

$$EV = \mu(p^0;p',m') - \mu(p^0;p^0,m^0) = A(p^0,p') + m' - m^0$$
$$CV = \mu(p';p',m') - \mu(p';p^0,m^0) = A(p^0,p') + m' - m^0.$$

在这种特别的情形下，补偿变化等于等价变化。不难看出这一结果背后的直觉。由于补偿函数是收入的线性函数，额外一元钱的效用，即收入的边际效用，和价格是无关的。因此，收入的补偿变化或等价变化的值和价格无关。

10.5 作为近似值的消费者剩余

我们已经知道，消费者剩余只有当效用函数为拟线性时，才是补偿变化和等价变化的精确衡量。然而，对于一般情形来说，它可能是一个合理的近似。

例如，考虑下面这样的情形：只有商品 1 的价格发生变动，从 p^0 变为 p'；收入固定不不变，$m = p^0 = m'$。在这种情形下，我们可以使用（10.1）式以及 $\mu(p;p,m) \equiv m$ 的事实，将等价变化和补偿变化写为

$$EV = \mu(p^0;p',m) - \mu(p^0;p^0,m) = \mu(p^0;p',m) - \mu(p';p',m)$$
$$CV = \mu(p';p',m) - \mu(p';p^0,m) = \mu(p^0;p^0,m) - \mu(p';p^0,m).$$

我们已经将这些式子写为唯一一个变量 p 的函数，因为所有其他价格都被假设为固定不变的。令 $u^0 = v(p^0,m)$ 以及 $u' = v(p',m)$ 并且使用以货币度量的效用函数的定义（参见第 7 章），我们有

$$EV = e(p^0,u') - e(p',u')$$
$$CV = e(p^0,u^0) - e(p',u^0).$$

最后，使用希克斯需求函数是支出函数的导数的事实，因此有 $h(p,u) \equiv \partial e / \partial p$，我们可以将这些式子写为

$$\begin{aligned} EV &= e(p^0,u') - e(p',u') = \int_{p'}^{p^0} h(p,u')dp \\ CV &= e(p^0,u^0) - e(p',u^0) = \int_{p'}^{p^0} h(p,u^0)dp. \end{aligned} \tag{10.2}$$

从这些式子可知，补偿变化是与初始效用水平相伴的希克斯需求函数的积分，等价变化是与最终效用水平相伴的希克斯需求函数的积分。福利的正确衡量是对需求曲线积分——但是你必须使用希克斯需求曲线而不是马歇尔需求曲线。

然而，我们可以使用（10.2）式推导出一个有用的界限。斯勒茨基方程告诉我们

$$\frac{\partial h(p,u)}{\partial p} = \frac{\partial x(p,m)}{\partial p} + \frac{\partial x(p,m)}{\partial m} x(p,m).$$

如果我们研究的商品是正常商品，希克斯需求曲线的导数就会大于马歇尔需求曲线的导数，如图 10.2 所示。

由此可知马歇尔需求曲线左侧的面积以希克斯需求曲线左侧的面积为界。在我们描述的情形中 $p^0 > p'$，因此所有的面积都是正的。由此可知，$EV > CS > CV$，其中 CS 表示消费者剩余。

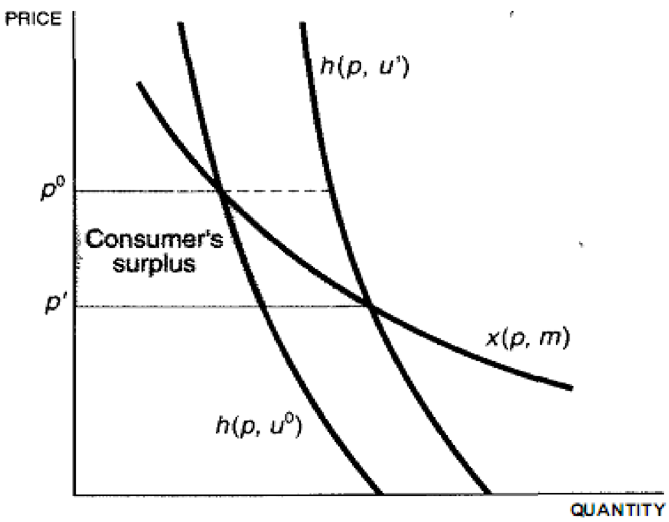


图 10.2：消费者剩余的界。对于正常商品来说，希克斯需求曲线比马歇尔需求曲线陡峭，因此，马歇尔需求曲线左侧的面积以希克斯需求曲线左侧的面积为界。

10.6 加总

补偿变化、等价变化和消费者的剩余的上述关系对于单个消费者来说是成立的。此处我们分析涉及许多消费者的一些问题。

在第 9 章我们已经知道，只有当个人 i 对某商品的间接效用函数具有高曼形式

$$v_i(p,m_i)=a_i(p)+b(p)m_i$$

时，该商品的加总需求函数才是价格以及加总收入的函数。

在这种情形下，每种商品的加总需求函数可从具有下列形式的加总间接效用函数推出

$$V(p,M)=\sum_{i=1}^{n}a_i(p)+b(p)M\ ,$$

其中 $M=\sum_{i=1}^{n}m_i$ 。

在前面我们已经知道与拟线性偏好相伴的间接效用函数的形式为

$$v_i(p)+m_i.$$

这显然是高曼形式的一种特殊情形，其中 $b(p)\equiv 1$。因此，产生加总需求函的加总间接效用函数为 $v(p)+m=\sum_{i=1}^{n}v_i(p)+\sum_{i=1}^{n}m_i.$

这个效用函数和加总的消费者剩余如何关联起来？为简单起见，我们回到单一价格的情形，罗伊法则表明函数 $v_i(p)$ 由下式给出

$$v_i(p)=\int_p^{\infty}x_i(t)dt\,.$$

由此可得

$$V(p)=\sum_{i=1}^{n}v_i(p)=\sum_{i=1}^{n}\int_p^{\infty}x_i(t)dt=\int_p^{\infty}\sum_{i=1}^{n}v_i(p)dt.$$

也就是说，产生加总需求函数的间接效用函数就是加总需求函数的积分。

如果所有消费者的效用函数都是拟线性的，则加总需求函数似乎可以使得加总的消费者剩余最大化。然而，加总消费者剩余是福利比较的合适工具这一事实并非特别明显。为什么某个特殊形式的效用的非加权（unweighted）之和是福利的有用衡量方法？我们在第 13 章分析这一问题。在那里我们将看到，在拟线性效用的情形下，加总消费者剩余可用于福利衡量，但是拟线性效用是一种非常特殊的情形。一般来说，加总消费者剩余不是福利的精确衡量方法。然而，在应用性研究中，通常使用加总消费者剩余作为消费者福利的近似衡量。

10.7 非参数边界

我们已经知道在给定参数形式的间接效用情形下，如何使用罗伊恒等式计算需求函数。如果给定需求函数的参数表达式，我们可以使用可积性理论计算用货币度量的效用函数的参数表达式。然而，这些计算都要求我们首先设定需求函数或间接效用函数的参数表达式。

我们自然会问，如果不设定参数表达式，我们能走多远？结果表明，即使完全使用非参数方法，我们也有可能推导出用货币度量的效用函数的紧致非参数边界。

我们在第 8 章讨论还原性时已经知道，给定任何一个消费束，我们可以构建比这个消费束“显示更好”（revealed preferred）或“显示更差”的消费束集合。这些集合可以视为消费者偏好集的内界和外界。

令 $NRW(x_0)$ 表示“不会显示比 x_0 更差”的消费束集合。这个集合正好是集合 $RW(x_0)$ 的补。我们从第 8 章知道，与 x_0 相伴的真正偏好集 $P(x_0)$ 必定包含集合 $RP(x_0)$，而且 $P(x_0)$ 必定包含于集合 $NRW(x_0)$ 之中，即 $RP(x_0)\subset P(x_0)\subset NRW(x_0)$。

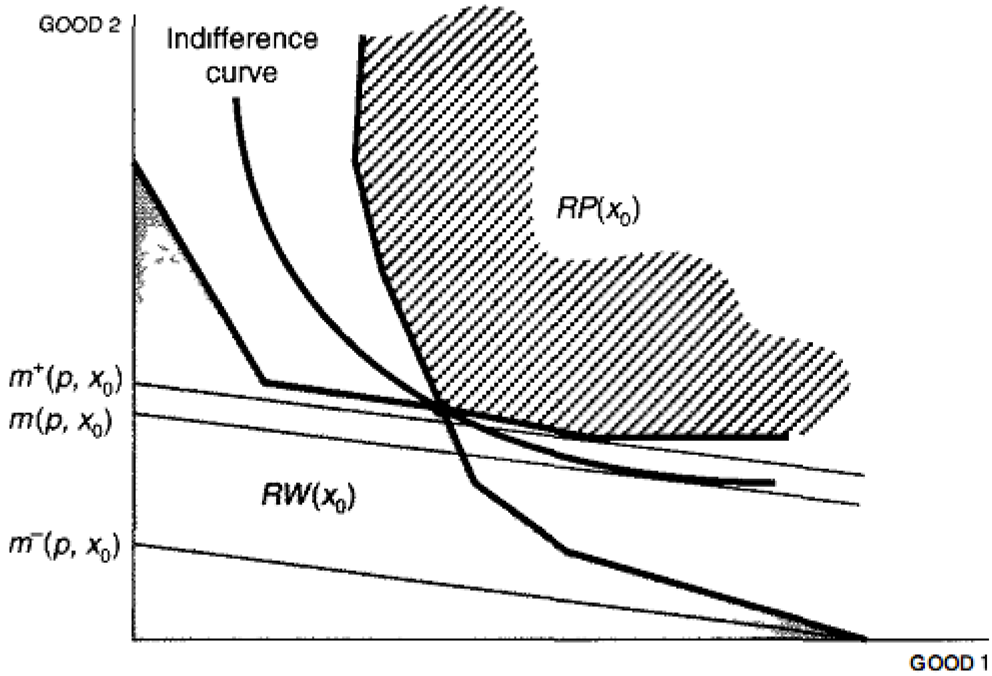


图 10.3：以货币度量的效用的界限。真正偏好集 $P(x_0)$ 必定包含集合 $RP(x_0)$，而且 $P(x_0)$ 必定包含于集合 $NRW(x_0)$ 之中，因此，$P(x_0)$ 上的最小支出必然位于我们画出的两个界限中，如图所示。

我们用图 10.3 说明这种情形。为了让图更清晰一些，我们省略了很多预算线和观测到的选择，只画出了 $RP(x_0)$ 和 $RW(x_0)$。我们还画出了通过 x_0 的“真正”无差异曲线。根据定义，x_0 用货币度量的效用为

$$m(p,x_0)=\min_x\ px$$
$$\text{s.t.}\ u(x)\geq u(x_0)$$

这个问题等价于

$$m(p,x_0)=\min_x \ px$$
$$\text{s.t. } x\in P(x_0)$$

定义$m^+(p,x_0)$和$m^-(p,x_0)$为

$$m^-(p,x_0)=\min_x \ px$$
$$\text{s.t. } x\in NRW(x_0)$$

和

$$m^+(p,x_0)=\min_x \ px$$
$$\text{s.t. } x\in RP(x_0)$$

由于$NRW(x_0)\supset P(x_0)\supset RP(x_0)$，由此可知$m^+(p,x_0)\geq m(p,x_0)\geq m^-(p,x_0)$。因此，补偿过度（overcompensation）函数$m^+(p,x_0)$，和补偿不足（undercompensation）函数$m^-(p,x_0)$构成了真正补偿函数$m(p,x_0)$的上下界。（本章结束）■

注释

Hicks（1956）首先讨论了补偿变化、等价变化以及它们与消费者剩余的关系。Willig（1976）讨论了消费者剩余的更紧致边界。Varian（1982a）分析了用货币度量的效用函数的非参数边界。

习题

10.1 假设效用是拟线性的。证明间接效用函数关于价格是凸的。

10.2 消费者埃尔斯沃思（Ellsworth）的效用函数为$U(x,y)=\min\{x,y\}$。该消费者有 150 元，商品x和y的价格都为 1。他的老板打算派他到另外一个镇上去。在这个镇上，商品x和y的价格分别为 1 和 2。老板不给他加工资。由于他很了解补偿变化和等价变化，他抱怨得很厉害。他说，尽管他不在意工作地点变动，因为新地点和原先地点一样让人愉快，但搬到新镇等同于他的收入降低了 A 元。他又说，如果给他增加 B 元的工资，他就不会抱怨搬到新镇去工作了。求 A 和 B 的数值。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 11 章：不确定性

曹乾　译

(东南大学　caoqianseu@163.com)

11 不确定性

直到现在，我们关注的都是消费者在确定条件下的行为。然而消费者的很多选择是在不确定条件下作出的。在本章我们将探讨如何使用消费者选择理论描述这样的行为。

11.1 彩票

第一个任务就是描述消费者面临的选择。我们可以将消费者面临的选择想象为彩票（lotteries）的形式。一张彩票可用 $p\circ x\oplus(1-p)\circ y$ 表示。这个符号表示："消费者得到奖品 x 的概率为 p，得到奖品 y 的概率为 $(1-p)$。"奖品可以为钱、商品或者甚至彩票。风险条件下的绝大多数行为都可以使用彩票架构进行解释。

对于消费者对彩票的认知，我们作出以下一些假设。

L1. $1\circ x\oplus(1-1)\circ y\sim x$. 以一的概率得奖和肯定得奖是一回事。

L2. $p\circ x\oplus(1-p)\circ y\sim(1-p)\circ y\oplus p\circ x$. 消费者不关心彩票描述的次序。

L3. $q\circ(p\circ x\oplus(1-p)\circ y)\oplus(1-p)\circ y\sim(qp)\circ x\oplus(1-qp)\circ y$. 消费者对彩票的认知仅取决于得到各种奖品的相应净概率。

假设（L1）和（L2）似乎是无害的。假设（L3）有时称为"复合彩票的化简"，这个假设多少让人怀疑它的合理性，因为有证据表明消费者认为复合彩票和一次性彩票是不同的。然而此处我们不打算深究这一点。

在这些假设条件下，我们可以定义消费者可以得到的彩票空间 L。我们假设消费者在这个彩票空间上存在着偏好：给定任何两张彩票，他可以做出选择。和往常一样，我们假设偏好是完备的、反身的和传递的。

我们假设彩票只有两个结果是合理的，因为我们可以允许它的结果仍是彩票。这样，通过不断复合只有两个结果的彩票，我们可以构造含有任何数量奖品种类的彩票。例如，假设我们想描述奖品有 x,y 和 z 的这样三种结果彩票，每种奖品的概率都为 1/3。根据复合彩票的化简可知，这个彩票等价于下列彩票

$$\frac{2}{3}\circ\left[\frac{1}{2}\circ x\oplus\frac{1}{2}y\right]\oplus\frac{1}{3}\circ z\,.$$

根据上面的假设 L3，消费者只关心奖品涉及的净概率，因此，这个彩票的确等价于原来的彩票。

11.2 期望效用

如果我们再增加一些小的假设，则效用函数存在性的相关定理（参见第 7 章）可用于证明：存在一个连续的效用函数 u，它可以描述消费者的偏好。也就是说，

$$p \circ x \oplus (1-p) \circ y \succ q \circ w \oplus (1-q) \circ z$$

当且仅当

$$u(p \circ x \oplus (1-p) \circ y) > u(q \circ w \oplus (1-q) \circ z).$$

当然，这个效用函数不是唯一的；它的任何单调转换一样合适。在某些额外的假设下，我们可以找到这个效用函数的某个特别的单调转换，它具有非常方便的性质，即**期望效用性质**（expected utility property）：

$$u(p \circ x \oplus (1-p) \circ y) = pu(x) + (1-p)u(y).$$

期望效用性质是说，彩票的效用等于它的各个奖品的期望效用。我们可以通过下列方式计算任何彩票的效用：对彩票的每个结果都相应取一个效用值，然后分别乘以各自的概率，最后将它们相加。彩票的效用在加法上对于彩票的结果来说是可分离的，而且彩票效用是概率的线性函数。

需要强调的是，效用函数的存在性问题是不存在争议的；任何良好表现（well-behaved）的偏好顺序都可以用效用函数来表示。我们此处关注的，是具有期望效用性质的效用函数的存在性问题。为了回答这个问题，我们还需要以下的一些公理：

U1. $\{p$ 在$[0,1]$中: $p \circ x \oplus (1-p) \circ y \succeq z\}$ 和 $\{p$ 在$[0,1]$中: $z \succeq p \circ x \oplus (1-p) \circ y\}$ 对于 L 中的所有 x, y 和 z 来说都是闭集。

U2. 如果 $x \sim y$，那么 $p \circ x \oplus (1-p) \circ z \sim p \circ y \oplus (1-p) \circ z.$

假设（U1）是关于连续性的假设；这个假设基本无害。假设（U2）是说奖品不同的彩票在某些条件下是无差异的。也就是说，如果给我们一种彩票 $p \circ x \oplus (1-p) \circ z$，而且我们知道 $x \sim y$，那么我们可以用 y 替代 x，从而构建另一种彩票 $p \circ y \oplus (1-p) \circ z$，消费者认为这两张彩票是等价的。这个假设似乎很合理。

为了避免一些技术上的细节，我们再做出两个假设。

U3. 存在最好的彩票 b 和最差的彩票 w。对于 L 中的任何 x 都有 $b \succeq x \succeq w.$

U4. 彩票 $p \circ b \oplus (1-p) \circ w$ 比彩票 $q \circ b \oplus (1-q) \circ w$ 更受偏好，当且仅当 $p > q$。

假设（U3）纯粹是出于方便的目的而做出的。假设（U4）可从其他公理推导出。这个公理是说，如果两种彩票的奖品都介于最好的彩票和最差的彩票之间，而且一种彩票比另外一种彩票更受偏好，那么原因必然是前者得到最好奖品的概率更高。

在这些假设之下，我们可以陈述主要的定理了。

期望效用定理。如果 $(C, \succeq)$ 满足上述公理，则存在定义在 C 上的效用函数 u，这个函数满足期望效用性质：

$$u(p \circ x \oplus (1-p) \circ y) = pu(x) + (1-p)u(y)。$$

证明。定义 $u(b) = 1$ 和 $u(w) = 0$。为了找到任意彩票 z 的效用，我们令 $u(z) = p_z$，其中 p_z 的定义为

$$p_z \circ b \oplus (1-p_z) \circ w \sim z. \qquad (11.1)$$

在这个构造中，消费者对于下列两个选择无差异：一是 z；二是在最好结果（概率为 p_z）与最差结果（概率为 $1-p_z$）之间的博彩。

为了保证上述定义是定义清楚的，我们需要验证两件事情。

（1） p_z 存在吗？集合 $\{p$ 在 $[0,1]$ 中: $p \circ x \oplus (1-p) \circ y \succeq z\}$ 和集合 $\{p$ 在 $[0,1]$ 中: $z \succeq p \circ x \oplus (1-p) \circ y\}$ 都是闭且非空的，并且 $[0,1]$ 中的每个点必然（至少）位于这两个集合中的一个。由于单位区间是连通的，因此必然存在某个 p 位于这两个集合之中——但这正是我们想要的 p_z。

（2） p_z 是唯一的吗？假设 p_z 和 p_z' 是两个不同的数，而且都满足（11.1）式。这两个数既然不同，则必然一个比另一个大。根据假设（U4）可知，得到最好奖品的概率较高的彩票，不可能和得到最好奖品概率较低的彩票一样好。因此 p_z 是唯一的。所以 u 是定义清晰的。

下面我们证明 u 具有期望效用性质。作一些简单的替代就可以看清这一点：

$$\begin{aligned}
&p \circ x \oplus (1-p) \circ y \\
&\quad \sim p \circ [p_x \circ b \oplus (1-p_x) \circ w] \oplus (1-p) \circ [p_y \circ b \oplus (1-p_y) \circ w] && (1) \\
&\quad \sim [pp_x + (1-p)p_y] \circ b \oplus [1 - pp_x - (1-p)p_y] \circ w && (2) \\
&\quad \sim [pu(x) + (1-p)u(y)] \circ b \oplus [1 - pu(x) - (1-p)u(y)] \circ w. && (3)
\end{aligned}$$

使用（U2）以及 p_x 和 p_y 的定义替代（1）。使用（L3）替代 2，这就是说真正要紧的是得到 b 或 w 的净概率。使用效用函数的构造来替代（3）。

根据效用函数的构造可知，

$$u(p \circ x \oplus (1-p) \circ y) = pu(x) + (1-p)u(y).$$

最后，我们证明 u 是一个效用函数。假设 $x \succ y$。则

$$u(x) = p_x \text{ 使得 } x \sim p_x \circ b \oplus (1-p_x) \circ w$$

$$u(y) = p_y \text{ 使得 } y \sim p_y \circ b \oplus (1-p_y) \circ w$$

根据公理（U4）可知，我们必然有 $u(x) > u(y)$。■

11.3 期望效用函数的唯一性

我们已经证明了期望效用函数 $u: L \to R$ 的存在性。当然，u 的任何单调转换都和 u 代表消费者的同样的选择行为。但是这样的单调转换能否保留期望效用的性质？上面介绍的效用函数的构造方法是否能够刻画期望效用函数的特征？

不难看出，如果 $u(\cdot)$ 是刻画某个消费者的期望效用函数，则 $v(\cdot) = au(\cdot) + c$ 也是，其中 $a > 0$；也就是说，期望效用函数的任何仿射转换（affine transformation）仍然是期望效用函数。这一点是很清楚的，因为

$$
\begin{aligned}
v(p\circ x\oplus(1-p)\circ y) = au(p\circ x\oplus(1-p)\circ y)+c \\
&= a[pu(x)+(1-p)u(y)]+c \\
&= p[au(x)+c]+(1-p)[au(y)+c] \\
&= pv(x)+(1-p)v(y).
\end{aligned}
$$

不难看出其逆命题为：u 的任何单调转换如果仍具有期望效用性质，那么这个单调转换必然是仿射转换。换一种表达方法：

期望效用函数的唯一性。期望效用函数唯一适用于仿射转换。

证明。根据前面的评价，我们只需要证明，如果某种单调转换保留了期望效用性质，它必然是仿射转换。令 $f:R\to R$ 为 u 的具有期望效用性质的单调转换。则

$$f(u(p\circ x\oplus(1-p)\circ y)) = pf(u(x))+(1-p)f(u(y)),$$

或

$$f(pu(x)+(1-p)u(y)) = pf(u(x))+(1-p)f(u(y)).$$

但这正等价于仿射转换的定义（参见第 26 章）。■

11.4 和期望效用有关的其他符号

我们已经证明了只有两种结果的彩票情形下的期望效用定理。正如以前指出的，很容易将这个证明扩展到结果为有限多种的彩票的情形，方法是使用复合彩票。如果得到结果 x_i 的概率为 p_i，其中 $i=1,...,n$，则这种彩票的期望效用为

$$\sum_{i=1}^{n} p_i u(x_i). \tag{11.2}$$

如果增加某些技术上的细节，就可以证明期望效用定理对于连续概率分布也是成立的。如果 $p(x)$ 是定义在结果 x 上的概率密度函数，则这个博彩的期望效用函数可以写为

$$\int u(x)p(x)dx \tag{11.3}$$

我们可以使用期望算符将这两种情形归纳在一起。令 X 表示取值为 x 的**随机变量**（random variable）。则 X 的效用也是一个随机变量 $u(X)$。这个随机变量的**期望**（expectation）$Eu(X)$ 是和彩票 X 相伴的期望效用。在离散随机变量的情形下，$Eu(X)$ 由（11.2）给出，而在连续随机变量的情形下，$Eu(X)$ 由（11.3）给出。

11.5 风险厌恶

下面我们考虑彩票空间仅由奖品为钱的博彩组成。我们知道，如果消费者的选择行为满足各种必要的公理，我们可以找到具有期望效用性质的效用函数来表示他的行为。这意味着如果我们只要知道这种特别形式的效用函数，我们就可以描述消费者对于一切金钱博彩的行为。例如，为了计算消费者对于博彩 $p\circ x\oplus(1-p)\circ y$ 的期望效用，我们只要使用

$pu(x)+(1-p)u(y)$ 计算即可。

这种构造可用图 11.1 表示，其中 $p=1/2$。注意在这个例子中，消费者更喜欢得到彩票的期望价值。也就是说，彩票的效用 $u(p\circ x\oplus(1-p)\circ y)$ 小于彩票期望价值的效用 $pu(x)+(1-p)u(y)$。这样的行为叫做**风险厌恶**（risk aversion）。消费者也可能是**风险喜好的**（risk loving），在这种情形下，消费者喜欢彩票胜过喜欢彩票的期望价值。

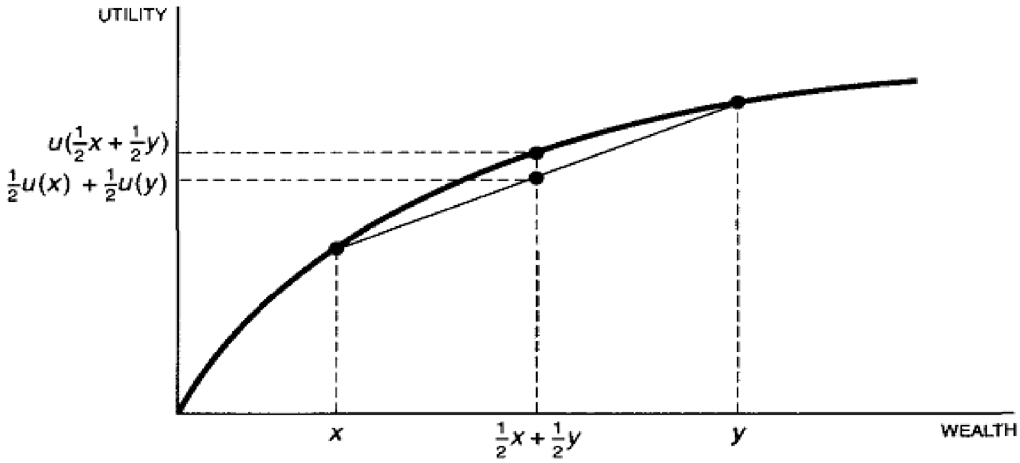


图 11.1：某个博彩的期望效用。该博彩的期望效用为 $\frac{1}{2}u(x)+\frac{1}{2}u(y)$。该博彩期望价值的效用为 $u(\frac{1}{2}x+\frac{1}{2}y)$。在这个例子中，期望价值的效用大于期望效用，因此消费者是风险厌恶的。

如果消费者在某个区间是风险厌恶的，那么他的效用函数曲线上的任何两点的连线，必然位于该曲线下方。这等价于凹函数的数学定义。因此，期望效用函数的凹性等价于风险厌恶。

通常我们想对风险厌恶的程度进行衡量。在直觉上，期望效用函数越凹，消费者越厌恶风险。因此，也许你会认为用期望效用函数的二阶导数衡量风险厌恶程度。然而，这种衡量方法存在着缺陷，因为它会随着期望效用函数的变化而变化：如果我们将期望效用函数乘以 2，消费者的行为没有改变，但是这种方法衡量出来的厌恶程度却变化了。

然而，如果我们将二阶导数标准化，即用二阶导数除以一阶导数，我们得到了一种合理的衡量风险厌恶程度的方法，这种方法称为**阿罗–普拉特的（绝对）风险厌恶的衡量方法**(Arrow-Pratt measure of (absolute) risk aversion):

$$r(w)=-\frac{u''(w)}{u'(w)}.$$

下列分析可以进一步说明上述衡量方法是合理的。现在我们用一对数字(x_1, x_2)表示一个博彩，其中消费者得到x_1如果事件E发生，得到x_2如果不是E发生。于是我们将消费者的**可接受集**（acceptance set）定义为由消费者在初始财富水平为w时可以接受的所有博彩组成的集合。如果消费者是风险厌恶的，可接受集将是个凸集。这个集合的边界——由那些无差异博彩组成的集合——可由隐函数$x_2(x_1)$给出，如图 11.2 所示。

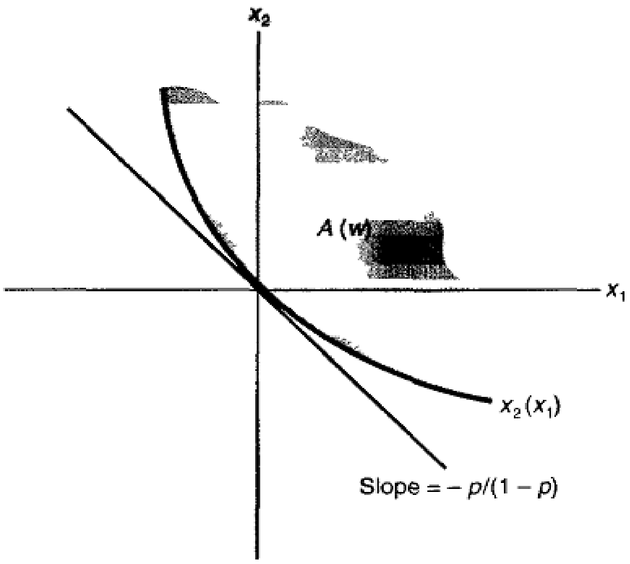


图 11.2：可接受集。这个集合刻画了消费者在初始财富水平情形下可以接受的所有博彩。如果消费者是风险厌恶的，可接受集将是凸集。

假设消费者的行为可由期望效用最大化刻画。则$x_2(x_1)$必然满足下列恒等式：

$$pu(w+x_1)+(1-p)u(w+x_2(x_1)) \equiv u(w).$$

可接受集边界在点(0,0)的斜率可用下列方法找到，将上式对x_1求导并计算该导函数在$x_1=0$的数值：

$$pu'(w)+(1-p)u'(w)x_2'(0)=0. \qquad (11.4)$$

由此可以解得该斜率为

$$x_2'(0)=-\frac{p}{1-p}.$$

也就是说，可接受集在点(0,0)的斜率给出了博彩的赔率。这种方法比较妙，因为我们可以得到概率——找到消费者恰好愿意为某事件押注一小笔钱的赔率。

现在假设我们有两个消费者，他们在事件E上的概率相同。我们自然可以说，消费者i比消费者j更厌恶风险，如果他的可接受集包含于消费者j的可接受集之中。由于上述说法意味着消费者j可以接受消费者i接受的任何博彩，这种对于风险厌恶的陈述是总体性的

（global）。如果我们将注意力限定在较小的博彩上，我们可以得到更有用的衡量方法。

自然可以说，消费者 i 比消费者 j **在局部上**（locally）更厌恶风险，如果他的可接受集包含于消费者 j 的可接受集中的点 $(0,0)$ 的某个领域之中。这意味着消费者 j 可以接受消费者 i 接受的任何较小的博彩。如果这种包含关系是严格的，那么消费者 i 可以接受的较小博彩数量比消费者 j 可接受的少。

不难看出，消费者 i 比消费者 j 在局部上更厌恶风险，如果他的可接受集在点 $(0,0)$ 附近比消费者 j 的可接收集"更弯曲"。这是一个有用的结论，因为我们可以通过计算 $x_2(x_1)$ 二阶导数的方法来找到可接受集的曲率。将恒等式（11.4）再一次对 x_1 求导，并计算该导函数在零点的值，可得

$$pu''(w)+(1-p)u''(w)x_2'(0)x_2'(0)+(1-p)u'(w)x_2''(0)=0.$$

将 $x_2'(0)=-p/(1-p)$ 代入上式，整理可得

$$x_2''(0)=-\frac{p}{(1-p)^2}\left[-\frac{u''(w)}{u'(w)}\right].$$

这个表达式与前面定义过的阿罗-普拉特的局部风险厌恶程度成比例。我们可以断言消费者 j 接受的较小博彩的数量比消费者 i 多，当且仅当消费者 i 的阿罗-普拉特局部风险厌恶程度比消费者 j 的大。

例子：保险的需求

假设某消费者初始拥有的钱数为 W 元。他损失 L 元的概率为 p，比如他家有可能失火。消费者可以购买保险，如果购买保险，那么在上述损失发生时，他可以从保险公司得到 q 元的赔偿。如果保险金额为 q 元，他必须向保险公司缴纳 πq 元的保险费；其中 π 是每一元钱保险金额的保险费，即保险费率。

消费者会购买多大的保险金额？我们看看这个效用最大化问题

$$\max\ pu(W-L-\pi q+q)+(1-p)u(W-\pi q).$$

将目标函数对 q 求导，并令该导函数等于零，可得

$$pu'(W-L+q^*(1-\pi))(1-\pi)+(1-p)u'(W-\pi q^*)\pi=0.$$

$$\frac{u'(W-L+(1-\pi)q^*)}{u'(W-\pi q^*)}=\frac{1-p}{p}\frac{\pi}{1-\pi}.$$

如果保险事故发生，保险公司的净收入为 $(\pi q-q)$ 元。如果保险事故不发生，保险公司的净收入为 πq 元。因此，保险公司的期望利润为

$$(1-p)\pi q-p(1-\pi)q.$$

假设竞争使得保险行业的利润为零。这意味着

$$-p(1-\pi)q+(1-p)\pi q=0,$$

由此可得 $\pi = p.$

在零利润的假设下，保险公司索要的是**精算公平费率**（actuarially fair premium）：保险单的成本恰好等于它的期望价值，即 $p = \pi$。将其代入效用最大化的一阶条件，可得

$$u'(W - L + (1-\pi)q^*) = u'(W - \pi q^*).$$

如果消费者是严格厌恶风险的，因此 $u''(W) < 0$，于是上式意味着

$$W - L + (1-\pi)q^* = W - \pi q^*.$$

由此可得 $L = q^*$。因此，消费者会购买足额保险以抵御损失 L。

这个结果主要取决于消费者无法影响损失发生的概率这个假设。如果消费者的行为的确会影响到损失发生的概率，保险公司可能只希望提供部分保险（partial insurance），因此消费者有动机进行风险的预防。在第 25 章我们将分析这样的一个模型。

11.6 总体的风险厌恶

阿罗-普拉特度量似乎是**局部**（local）风险厌恶的合理解释：一个人比另一个人更厌恶风险，如果他愿意接受的较小博彩比后者愿意接受的少。然而，在很多情形下，我们需要风险厌恶的总体（global）度量——也就是说，我们想表达的是，在所有财富水平上，一个人都比另一个人更厌恶风险。那么很自然地表达这种条件的方法是什么呢？

第一种可行的方法是，如果 A 的效用函数为 $A(w)$，B 的为 $B(w)$，那么若 A 比 B 更厌恶风险，则必然有

$$-\frac{A''(w)}{A'(w)} > -\frac{B''(w)}{B'(w)}$$

对所有财富水平 w 成立。这只是说 A 的风险厌恶程度处处都比 B 高。

第二种可行的方法是，如果 A 比 B 更厌恶风险，那么 A 的效用函数比 B 的“更凹”。更准确地说，我们说 A 的效用函数是 B 的效用函数的一个凹转换（concave transformation）；也就是说，存在递增、严格凹的函数 $G(\cdot)$ 使得

$$A(w) = G(B(w)).$$

第三种方法是，如果 A 比 B 更厌恶风险，则为了规避一定的风险，A 愿意比 B 支付更多的钱。为了将这种想法表达清楚，令 $\tilde{\varepsilon}$ 表示一个随机变量，它的期望为零：$E\tilde{\varepsilon} = 0$。然后令 $\pi_A(\tilde{\varepsilon})$ 表示消费者为了规避随机变量 $\tilde{\varepsilon}$ 而愿意放弃的最大财富数量。用符号可将这个**风险贴水**（risk premium）表示为

$$A(w - \pi_A(\tilde{\varepsilon})) = EA(w + \tilde{\varepsilon}).$$

上式左侧表示的是初始财富 w 减去 $\pi_A(\tilde{\varepsilon})$ 之后的效用；上式右侧表示的是博彩 $\tilde{\varepsilon}$ 的期望效用。自然可以说，A 比 B（总体上）更厌恶风险，如果 $\pi_A(\tilde{\varepsilon}) > \pi_B(\tilde{\varepsilon})$ 对于所有财富水平 w 都成立。

如果我们想表达一个人比另一个人总体上更厌恶风险，我们应该选择上述三种方法中的哪一种？似乎很难作出取舍。幸运地是，我们不需要作出选择：这三种方法是等价的！为了证明这个结论，我们需要下列结果，这个结果在处理期望效用函数时非常有用。

詹森不等式（Jensen's inequality）。令 X 表示一个非退化的（nondegenerate）随机变量，$f(X)$ 表示该随机变量的严格凹函数，则 $Ef(X) < f(EX)$。

证明。这个结论具有一般性，但我们将在可微的凹函数情形下证明这个结论，因为这么做最容易。这样的函数具有下列性质：对于任何点 $\bar{x}$ 都有 $f(x) < f(\bar{x}) + f'(\bar{x})(x-\bar{x})$。令 $\overline{X}$ 表示 X 的期望值，将前面的那个式子的两边取期望值，可得

$$Ef(X) < f(\overline{X}) + f'(\overline{X})E(X-\overline{X}) = f(\overline{X}),$$

由此可得

$$Ef(X) < f(\overline{X}) = f(EX). \quad \blacksquare$$

普拉特定理（Pratt's theorem）。令 $A(w)$ 和 $B(w)$ 表示财富 w 的两个期望效用函数，它们都是可微的、递增的而且凹的函数，则下列性质是等价的。

1） $-\dfrac{A''(w)}{A'(w)} > -\dfrac{B''(w)}{B'(w)}$ 对于所有 w 成立。

2） $A(w) = G(B(w))$ 对于某些递增的严格凹函数 G 成立。

3） $\pi_A(\tilde{\varepsilon}) > \pi_B(\tilde{\varepsilon})$ 对于 $E\tilde{\varepsilon} = 0$ 的所有 $\tilde{\varepsilon}$ 成立。

证明。

（1）蕴涵（2）。将 $G(B)$ 隐含地定义为 $A(w) = G(B(w))$。该效用函数为单调函数的事实意味着 G 是定义清晰的——即，对于每个 B，$G(B)$ 是唯一的。现在对该定义微分两次可得

$$A'(w) = G'(B)B'(w)$$

$$A''(w) = G''(B)B'(w)^2 + G'(B)B''(w).$$

由于 $A'(w) > 0$ 和 $B'(w) > 0$，上面第一个式子表明 $G'(B) > 0$。将上面的第二个式子除以第一个式子可得

$$\frac{A''(w)}{A'(w)} = \frac{G''(B)}{G'(B)}B'(w) + \frac{B''(w)}{B'(w)}.$$

重新整理可得

$$\frac{G''(B)}{G'(B)}B'(w) = \frac{A''(w)}{A'(w)} - \frac{B''(w)}{B'(w)} < 0,$$

其中不等式的符号由（1）推知。这表明 $G''(B) < 0$，这正是我们想要的。

（2）蕴涵（3）。这可由下列的一系列式子推出：

$$\begin{aligned} A(w-\pi_A) &= EA(w+i) = EG(B(w+\tilde{\varepsilon})) \\ &< G(EB(w+i) = G(B(w-\pi_B)) \\ &= A(w-\pi_B). \end{aligned}$$

上面的式子中，除了不等式之外，其余均根据风险贴水的定义推出；不等式是根据詹森不等式推出。比较第一项和最后一项可知，$\pi_A > \pi_B$.

（3）蕴涵（1）。由于（3）对于所有均值为零的随机变量 $\tilde{\varepsilon}$ 都成立，它必然对于任意小的随机变量成立。固定一个 i，考虑 ti 定义的随机变量族，其中 $t \in [0,1]$。令 $\pi(t)$ 表示风险贴水，它是 t 的函数。在 $t=0$ 附近的 $\pi(t)$ 的二阶泰勒序列展开式为

$$\pi(t) \approx \pi(0) + \pi'(0)t + \frac{1}{2}\pi''(0)t^2 \tag{11.5}$$

我们需要计算这个泰勒序列中的每一项，目的在于分析对于较小的 t，$\pi(t)$ 的行为是怎样的。$\pi(t)$ 的定义为

$$A(w-\pi(t)) \equiv EA(w+ti).$$

根据这个定义可知 $\pi(0)=0$。将上式对 t 求导两次可得

$$-A'(w-\pi(t))\pi'(t) = E[A'(w+t\tilde{\varepsilon})\tilde{\varepsilon}]$$

$$A''(w-\pi(t))\pi'(t)^2 - A'(w-\pi(t))\pi''(t) = E[A''(w+t\tilde{\varepsilon})\tilde{\varepsilon}^2]$$

（有些读者可能对期望的微分不熟悉。但是求期望值只是求和或者求积分，因此它们的法则相同：期望的导数等于导数的期望。）

求上面第一个式子在 $t=0$ 时的值，可得 $\pi'(0)=0$。求上面第二个式子在 $t=0$ 时的值，可得

$$\pi''(0) = -\frac{EA''(w)i^2}{A'(\sim)} = -\frac{A''(w)}{A'(w)}\sigma^2$$

其中 σ^2 是 $\tilde{\varepsilon}$ 的方差。将这些导数值代入（11.5）可求得 $\pi(t)$：

$$\pi(t) \approx 0 + 0 - \frac{A''(w)}{A'(w)}\frac{\sigma^2}{2}t^2.$$

这意味着对于任意小的 t 值，风险贴水单调地取决于风险厌恶程度，这正是我们想要证明的结果。■

例子：一个简单的投资组合问题的比较静态分析

我们使用所学过的知识分析一个简单的两阶段投资组合问题，这个组合涉及两种资产，一种是风险资产，另一种是无风险资产。由于风险资产的收益率是不确定的，我们用随机变量 $\tilde{R}$ 表示该收益率。

令 w 表示初始财富，令 $a \geq 0$ 表示投资在风险资产上的钱数，预算约束表明投资在无风险资产上的钱数为 $w-a$ 。为方便起见，假设无风险资产的收益率为零。

在这种情形下，第二阶段的财富为

$$\tilde{W} = a(1+\tilde{R}) + w - a = a\tilde{R} + w.$$

注意，第二阶段的财富是个随机变量，这是因为 $\tilde{R}$ 是个随机变量。投资 a 元在风险资产上的期望效用可以写为

$$v(a) = Eu(w + a\tilde{R}),$$

该期望效用关于 a 的前两阶导数分别为

$$v'(a) = Eu'(w + a\tilde{R})\tilde{R},$$
$$v''(a) = Eu''(w + a\tilde{R})\tilde{R}^2.$$

注意到风险厌恶意味着 $v''(a)$ 处处为负，因此二阶条件自动满足。

我们首先考虑边界解。求一阶导数在 $a=0$ 的值，可得 $v'(0) = Eu'(w)\tilde{R} = u'(w)E\tilde{R}$ 。由此可知，如果 $E\tilde{R} \leq 0$ ， $v'(0) \leq 0$ ；而且在严格风险厌恶的情形下， $v'(a) < 0$ 对于所有 $a > 0$ 成立。因此 $a = 0$ 是最优的当且仅当 $ER \leq 0$ 。也就是说，一个厌恶风险的人对风险资产的投资钱数为零，当且仅当该风险资产的期望收益是非正的。

相反，如果 $ER > 0$ ，由此可知 $v'(0) = u'(w)E\tilde{R} > 0$ ，因此投资者一般对风险资产的投资钱数为正。最优投资钱数满足下列一阶条件

$$Eu'(w + a\tilde{R})\tilde{R} = 0, \tag{11.6}$$

上式要求财富的期望边际效用等于零。

下面我们对该选择问题进行比较静态分析。首先我们分析当 w 变化时 a 如何变化。令 $a(w)$ 表示 a 的最优值，它是 w 的函数；这等价于要求满足下列一阶条件

$$Eu'(w + a(w)\tilde{R})\tilde{R} \equiv 0.$$

上式对 w 求导可得

$$Eu''(w + a\tilde{R})\tilde{R}[1 + a'(w)\tilde{R}] \equiv 0,$$

$$a'(w) = -\frac{Eu''(w + a\tilde{R})\tilde{R}}{Eu''(w + a\tilde{R})\tilde{R}^2}.$$

和往常一样，分母是负的，这由二阶条件可知，因此我们知道

$$a'(w) \text{ 的符号与 } Eu''(w + a\tilde{R})\tilde{R} \text{ 的符号相同.}$$

$Eu''(w + a\tilde{R})\tilde{R}$ 为正还是负并不明显。然而，可以证明它由绝对风险厌恶 $r(w)$ 的行为决定。

风险厌恶。$Eu''(w+a\tilde{R})\tilde{R}$ 为正、负或零，当 $r(w)$ 为 w 的递增、递减或不变的函数。

证明。我们选择证明 $r'(w)<0$ 蕴涵 $Eu''(w+a\tilde{R})\tilde{R}>0$，因为这是最合理的情形。其他情形的证明是类似的。

首先考虑 $\tilde{R}>0$ 的情形。在这种情形下我们有

$$r(w+a\tilde{R})=-\frac{u''(w+a\tilde{R})}{u'(w+a\tilde{R})}<r(w),$$

将上式变形可得

$$u''(w+a\tilde{R})>-r(w)u'(w+a\tilde{R}). \qquad (11.7)$$

由于 $\tilde{R}>0$，

$$u''(w+a\tilde{R})\tilde{R}>-r(w)u'(w+a\tilde{R})\tilde{R}. \qquad (11.8)$$

现在考虑 $\tilde{R}<0$ 的情形。检查（11.7）式，我们可知绝对风险厌恶递减意味着

$$u''(w+a\tilde{R})<-r(w)u'(w+a\tilde{R}).$$

由于 $\tilde{R}<0$，

$$u''(w+a\tilde{R})\tilde{R}>-r(w)u'(w+a\tilde{R})\tilde{R}.$$

将该式与（11.8）比较可知，（11.8）对于 $\tilde{R}>0$ 和 $\tilde{R}<0$ 必然都成立。因此，对于所有的 $\tilde{R}$ 取期望可得

$$u''(w+a\tilde{R})\tilde{R}>-r(w)u'(w+a\tilde{R})\tilde{R}=0,$$

其中最后一个等式是根据一阶条件推导出。■

这个引理给出了我们想要的结果：当风险厌恶函数为财富的递减、不变或递增函数时，对风险资产的投资额是增加、不变或减少的。

下面我们转而分析当收益的概率分布变化时，对风险资产的需求如何变化。随机收益率移动可用 $(1+h)\tilde{R}$ 进行参数化表示，其中 h 是一个移动变量。当 $h=0$ 时，就是原来的随机变量；如果 $h>0$，这表示每个已实现的收益比以前相应的收益高了 h 个百分点。

在（11.6）式中用 $(1+h)\tilde{R}$ 代替 $\tilde{R}$，并将该式的两侧同时除以 $(1+h)$ 可得

$$Eu'(w+a(1+h)\tilde{R})\tilde{R}=0. \qquad (11.9)$$

我们可以对这个式子再对 h 微分一次，并分析出它的符号。但是还有一种更简单的方法能看清当 h 变化时 a 是如何变化的。令 $a(h)$ 表示风险资产的需求，它是 h 的函数。我们可以断言

$$a(h)=\frac{a(0)}{1+h}.$$

将这个表达式代入一阶条件（11.9）式就可知道上述结论是成立的。

直觉上，如果随机变量变为原来的 $(1+h)$ 倍，消费者就会相应减少持有的投资资产，减少比例为 $1/(1+h)$，这样就可以恰好恢复到随机变量移动前的收益模式。随机变量的这种线性移动，消费者可相应调整自己的投资组合，从而将其完全抵消。

随机变量的一个更有趣的移动，称为保留均值的伸展（mean-preserving spread）。这种移动使得 R 的方差变大，但不会导致均值变动。这种变动的其中一种参数化表达方式是将其写成 $R+h(\tilde{R}-\overline{R})$ 。这个随机变量的期望值为 $\overline{R}$ ，但是方差为 $(1+h)^2\delta_R^2$，因此 h 的增加不会使均值变动但却使方差变大。

我们也可以将这个表达式写为 $(1+h)R-h\overline{R}$ 。这表明这样的保留均值的伸展，可以看成将随机变量乘以 $(1+h)$，然后再减去 $h\overline{R}$ 。根据前面的分析结果可知，随机变量乘以 $(1+h)$ 使得需求按 $(1+h)$ 的比例下降，进一步减去财富之后使得需求下降得更多，当然前提是绝对风险厌恶函数是减函数。因此，保留均值的扩展使得风险资产的投资额减少幅度超过了线性减少幅度。

例子：资产定价

假设风险资产有很多种，但无风险资产只有一种。每种风险资产的随机总收益（random total return）为 $\tilde{R}_i$，其中 $i=1,\ldots,n$；无风险资产的总收益为 R_0 。（总收益 R 等于 1 加上收益率；在上一节我们使用 R 表示收益率，而在本节我们用它表示总收益。）消费者最初拥有财富 w，对于资产 i 他选择的投资金额占财富的比例为 x_i，其中 $i=0,1,\ldots,n$ 。因此，消费者在第二阶段的财富——当随机收益实现时——将由下式给出

$$\tilde{W}=w\sum_{i=0}^{n}x_i\tilde{R}_i \tag{11.10}$$

我们假设消费者希望选择 (x_i) 来使随机财富 $\tilde{W}$ 的期望效用最大化。

这个最大化问题的预算约束为 $\sum_{i=0}^{n}x_i=1$ 。由于 x_i 是消费者在第 i 种资产投资额占财富 w 的比重，因此这些比重的和必然为 1.我们也可以将这个预算约束写为

$$x_0+\sum_{i=1}^{n}x_i=1$$

因此，$x_0=1-\sum_{i=1}^{n}x_i$ 。将这个式子代入（11.10）并且经整理可得

$$\tilde{W}=w\left[x_0R_0+\sum_{i=1}^{n}x_i\tilde{R}_i\right]=w\left[(1-\sum_{i=1}^{n}x_i)R_0+\sum_{i=1}^{n}x_i\tilde{R}_i\right]=w\left[R_0+\sum_{i=1}^{n}x_i(\tilde{R}_i-R_0)\right].$$

将预算约束这样变形之后，我们就得到了一个无约束的最大化问题

$$\max_{x_1,\dots,x_n} Eu\left(w\left[R_0+\sum_{i=1}^{n}x_i(\tilde{R}_i-R_0)\right]\right).$$

将目标函数对 x_i 求导可得到一阶条件

$$Eu'(\tilde{W})(\tilde{R}_i-R_0)=0,$$

其中 $i=1,\dots,n$。注意到这个式子在本质上和前一节推导出的式子是相同的。

这个式子也可以写为

$$Eu'(\tilde{W})\tilde{R}_i=R_0Eu'(\tilde{W}).$$

使用随机变量协方差的恒等式 $\operatorname{cov}(X,Y)=EXY-EXEY$，我们可以将这个式子写为

$$\operatorname{cov}(u'(\tilde{W}),\tilde{R}_i)+E\tilde{R}_iEu'(\tilde{W})=R_0Eu'(\tilde{W}),$$

整理可得

$$E\tilde{R}_i=R_0-\frac{1}{Eu'(\tilde{W})}\operatorname{cov}(u'(\tilde{W}),\tilde{R}_i). \tag{11.11}$$

这个式子是说，任何资产的期望收益可以写成两部分之和：一是无风险收益；二是风险贴水。风险贴水取决于财富的边际效用与资产收益之间的协方差。（注意这里的风险贴水概念和我们证明普拉特定理时用到的风险贴水概念是不一样的。不幸的是，它们的名字是一样的。）

考虑某种资产，它的收益与财富正相关。由于风险厌恶意味着财富的边际效用随着财富的增加而递减，由此可推知，这样的资产将和边际效用负相关。因此，为了补偿风险，这样的资产的期望收益率必然高于无风险资产的收益率。

另一方面，某资产如果和财富是负相关的，则它的期望收益率小于无风险收率率。在直觉上，某资产和财富负相关，则该资产对于减少风险来说具有特别重要的价值，因此人们为了持有这样的资产，愿意牺牲期望收益。

11.7 相对风险厌恶

考虑一个财富为 w 的消费者，假设我们提供给消费者下列的博彩形式：在 p 概率下他可得到他当前财富的百分之 x；在 $1-p$ 概率下他可得到他当前财富的百分之 y。如果消费者使用期望效用评价彩票，上述彩票的效用将为

$$pu(xw)+(1-p)u(yw).$$

注意，这个多重的博彩的结构和前面分析过的可加性博彩的结构不同。然而，这类相对博彩经常在经济问题中出现。例如，投资的收益通常表达为相对于投资水平来说。

和以前一样，我们也可以问：给定一定的财富水平，一个消费者在什么条件下才愿意比另一个消费者接受更多的较小相对博彩。回顾前面使用过的类似分析，我们发现合适的度量是**阿罗-普拉特的相对风险厌恶度量**（Arrow-Pratt measure of relative risk aversion）:

$$\rho = -\frac{u''(w)w}{u'(w)}.$$

我们对绝对和相对风险厌恶如何随财富的变化而变化感兴趣。自然我们可以假设绝对风险厌恶是随着财富的增加而下降的：当你变得更富有时，你可能愿意接受赌博额更大的赌博。相对风险厌恶的行为有些问题；随着你的财富增加，你是更愿意还是更不愿意失去某个既定份额的财富呢？假设相对风险厌恶程度不变，这可能是个不错的假设，至少当财富小额变化时，这样的假设是合理的。

例子：均值-方差效用

一般来说，一个博彩的期望效用取决于所有结果的整个的概率分布。然而，在某些情形下，博彩的期望效用仅取决于分布的某些概述性质的统计量（summary statistics）。最常见的例子是**均值-方差效用函数**（mean-variance utility function）。

例如，假设期望效用函数是二次型的，因此$u(w) = w - bw^2$。于是期望效用为

$$Eu(w) = Ew - bEw^2 = \overline{w} - b\overline{w}^2 - b\sigma_w^2.$$

因此，这种博彩的期望效用只是财富的均值和方差的函数。

不幸的是，二次型效用函数具有人们不喜欢的性质：在某些区间，它是财富的减函数，而且它表现出递增的绝对风险厌恶。

在一个更有用的情形下，均值-方差分析也是合理的，这就是财富是正态分布的情形。我们都知道正态随机变量可完全由均值和方差刻画；因此，在正态分布的随机变量中的选择问题简化为对它们均值和方差的比较。

我们也对一种特别的情形感兴趣，这就是消费者的效用函数具有$u(w) = -e^{-rw}$的形式。可以证明这种效用函数具有绝对风险厌恶程度不变的性质。而且，当财富是正态分布时

$$Eu(w) = -\int e^{-rw} f(w)dw = -e^{-r[\overline{w} - r\sigma_w^2/2]}$$

（积分时，可以进行配方，或者注意到这个积分就是找到正态分布的矩母函数（moment generating function)。）注意，期望效用关于$\overline{w} - r\sigma_w^2/2$是递增的。这意味着我们可以对该期望效用函数进行单调转换，并且使用效用函数$u(\overline{w}, \sigma_w^2) = \overline{w} - \frac{r}{2}\sigma_w^2$计算财富的分布。这个效用函数非常方便，因为它对财富的均值和方差是线性的。

11.8 依赖于状态的效用

我们分析消费者在不确定环境下的选择决策时，一开始我们假设奖品只是抽象的商品束；后来我们专注于只具有货币结果的彩票的分析。然而，这样做有一定的缺陷。毕竟，一元钱的价值取决于当前商品的价格；因此金钱博彩结果的完整描述，不仅应该包括每种结果下能得到的钱数，还应该包括每种结果下的当前价格水平。

一般来说，某商品的有用性取决于它所处的自然状态（state of nature）。对于消费者来说时，雨天的雨伞和晴天的雨伞似乎是大不相同的。这样的例子说明，在某些选择问题中，有必要区分商品所处的自然状态。

例如，假设有两种自然状态，炎热和寒冷，我们分别用 h 和 c 表示这两种状态。令 x_h 表示天气炎热状态下分发的冰淇淋数量， x_c 表示寒冷状态下分发的冰淇淋数量。如果天气炎热的概率为 p，我们可以将这种特别的彩票写为 $pu(h,x_h)+(1-p)u(c,x_c)$。此处分发的商品束在一种状态下为“天热和 x_h 单位冰淇淋”,在另外一种状态下为“天冷和 x_c 单位冰淇淋”。

更正式的例子是健康保险。钱的价值取决于你的健康状态——如果你处于昏迷状态，一百万元对你来说价值多少？在这种情形下，我们可将效用函数写为 $u(h,m_h)$，其中 h 是健康状况的指标， m 表示钱数。这些情形下的效用函数都是取决于状态的效用函数（state-dependent utility functions）。这样的效用函数只是说，消费者对于商品的偏好取决于商品所处的自然状态。

11.9 主观概率理论

在期望效用理论的讨论中，我们对进入期望效用函数的“概率”的本质有些含糊其辞。最简单的解释是他们是“主观”概率——比如基于观测到的频率基础之上的概率计算。不幸的是，大多数选择问题涉及的是主观概率（subjective probabilities）:某个既定的人对某事件发生可能性的预期。

在期望效用理论情形下，我们曾关注消费者选择行为的什么样的公理假设，将蕴涵代表行为的期望效用函数的存在。类似地，我们想知道消费者选择行为的什么样公理，可以推断主观概率的存在性问题；也就是说，可以将消费者的选择行为看成如下的情形：他使用某些主观期望测度方法测度期望效用，并根据期望效用评估博彩。

幸运的是，这样的公理是存在的而且是相当合理的。主观概率的构造方法类似期望效用函数的构造方法。我们已经知道某个博彩 x 的效用可按下列方法构造：选择数 $u(x)$ 使得

$$x \sim u(x)\circ b\oplus(1-u(x))\circ w.$$

假设我们想确定某个消费者对于某天将下雨的主观概率预期。于是我们可以询问在什么样的概率下，他在下列两个事件之间无差异：一是博彩 $p\circ b\oplus(1-p)\circ w$；二是“若下雨则得到 b 元，否则得到 w 元”。

更正式地，令 E 为某个事件，令 $p(E)$ 表示该事件发生的（主观）概率。我们将事件 E 发生的主观概率定义为:数 $p(E)$，它需要满足

$$p(E)\circ b\oplus(1-p(E))\circ w \sim \text{若 } E \text{ 发生则得到 } b \text{ 元否则得到 } w \text{ 元}.$$

可以证明，在某些正则性（regularity）假设下，上述方法定义出的主观概率具有普通客观概率的所有性质。特别地，它们遵循条件概率计算的常用法则。这对经济行为有着一些有用的应用。

我们将简要分析其中一种应用。令 $p(H)$ 表示某消费者对于某个特定假设为真的主观概

率，E 是表明 H 为真的某个事件。一个理性人如何根据 E 调整他的概率预期？也就是说，在事件 E 发生的条件下，H 为真的概率为多大？

我们可以写出事件 E 发生和 H 为真的联合概率

$$p(H,E) = p(H|E)\,p(E) = p(E|H)\,p(H).$$

将上式右侧整理可得

$$p(H|E) = \frac{p(E|H)\,p(H)}{p(E)}.$$

这是贝叶斯法则的一种形式，它将先验概率 $p(H)$ 和后验概率联系起来。先验概率是在观察到证据之前假设为真的概率；后验概率是在观测到证据之后假设为真的概率。

贝叶斯法则是直接从条件概率的简单运算中推导出的。如果某个人的行为满足的限制足以保证主观概率的存在，这些概率必定满足贝叶斯法则。贝叶斯法则非常重要，因为它表明了一个理性人在看到证据的情形下应该如何更新自己的概率，因此这个法则是多大多理性学习行为（rational learning behavior）模型的基础。

因此，效用函数和主观概率都可以从观测到的选择行为中构建，只要观测到的选择行为满足某些在直觉上合理的公理。然而，我们应该强调，尽管这些公理在直觉上是合理的，但这并不意味着它们就是个人实际行为的准确描述。这个决定必须依据于实证证据。

例子：阿莱悖论（Allais paradox）和埃尔斯伯格悖论（Ellsberg paradox）

期望效用理论和主观概率理论是受理性行为启发而构建出的。期望效用理论背后的公理似乎是合理的，我们用于构建主观概率理论的公理似乎也是合理的。

不幸的是，现实生活中的个人行为似乎违背了某些公理。此处我们介绍两个著名的例子。

阿莱悖论

你要在下面两个赌博中做出选择：

赌博 A：赢得 100 万元的机率为 100%。

赌博 B：赢得 500 万元的机率为 10%，赢得 100 万元的机率为 89%，赢得零万元的机率为 1%。

在进一步阅读下面的内容之前，请先选择一个赌博（A 或 B），把它写在纸上。现在考虑下面两个赌博：

赌博 C：赢得 100 万元的机率为 11%，赢得零万元的机率为 89%。

赌博 D：赢得 500 万元的机率为 10%，赢得零万元的机率为 90%。

再次选择一个赌博（C 或 D），并将它记在纸上。

很多人偏好 A 胜于 B，偏好 D 胜于 C。然而，这些选择违背了期望效用公理！为了看清这一点，只要将 $A \succsim B$ 蕴涵的期望效用关系写出即可：

$$u(1) > 0.1u(5) + 0.89(1) + 0.01u(0)$$

将上式变形可得

$$0.11u(1) > 0.1u(5) + 0.01u(0)$$

再在上式两侧同时加上 $0.89u(0)$ 可得

$$0.11u(1) + 0.89u(0) > 0.1u(5) + 0.90u(0).$$

这意味着，根据期望效用最大化，博弈 C 必定好于 D。但在现实中，大多人偏好 D 胜于 C，所以违背了期望效用公理。

埃尔斯伯格悖论

埃尔斯伯格悖论和主观概率理论有关。你要在一个壶中摸出一个球。已知该壶一共装有 300 个小球，其中 100 个为红色的，另外 200 个要么是蓝色的要么是绿色的。

赌博 A：如果你摸出的球是红的，你得到 1000 元。

赌博 B：如果你摸出的求是蓝的，你得到 1000 元。

在纸上记下你的选择（A 或 B）。现在考虑下面的赌博：

赌博 C：如果你摸出的球不是红的，你得到 1000 元。

赌博 D：如果你摸出的球不是蓝色的，你得到 1000 元。

再次在纸上记下你的选择（C 或 D）。

人们通常严格偏好 A 胜于 B，严格偏好 C 胜于 D。但是这些偏好违背了标准的主观概率理论。为了看清这一点，令 R 表示球是红色的这个事件，令 $\neg R$ 表示球为非红的事件，令 B 表示球是蓝色的这个事件，令 $\neg B$ 表示球为非蓝的事件。根据概率常识，

$$\begin{aligned} p(R) &= 1 - p(\neg R) \\ p(B) &= 1 - p(\neg B) \end{aligned} \qquad (11.12)$$

为简单起见，将 $u(0)$ 标准化为 0，即 $u(0) = 0$。于是，如果 A 比 B 更受偏好，我们必定有 $p(R)u(1000) > p(B)u(1000)$，由此可知

$$p(R) > p(B). \qquad (11.13)$$

如果 C 比 D 更受偏好，我们必定有 $p(\neg R)u(1000) > p(B)u(\neg 1000)$，由此可知

$$p(\neg R) > p(\neg B). \qquad (11.14)$$

然而，容易看出（11.2）式、（11.13）式和（11.4）式是不一致的。

埃尔斯伯格悖论出现的原因似乎是人们认为赌红球比赌篮球更“安全”一些。

阿莱悖论和埃尔斯伯格悖论重要吗？经济学家对此看法不一。有些经济学家认为需要为这些反常现象构建新的模型来描述人们的行为。另外一些经济学家认为这些悖论类似于“视觉错觉”。尽管人们在某些情形下不能很好地判断距离，但这并不意味着我们必须发明新的距离概念。（本章结束）■

注释

期望效用理论源于 Neuman and Morgenstern(1944)。我们此处对期望效用理论的处理采用的是 Herstein and Milnor(1953)的方法。风险厌恶的衡量归功于 Arrow(1970) 和 Pratt（1964）。我们此处对风险厌恶衡量的处理采用的是 Yaari（1969）的方法。期望效用理论一般化的近期研究综述，可参见 Machina（1982）。我们对主观概率理论的简短介绍是基于 Ansombe and Aumann(1963)。

习题

11.1 证明为了避免参加方差为 v 的小赌博，一个人愿意付出的钱数大约为 $r(w)v/2$ 。

11.2 如果风险厌恶度为常数，期望效用函数是什么样的？如果相对风险厌恶度是常数呢？

11.3 如果对风险资产的投资和财富数量无关，那么期望效用函数是什么样的？

11.4 考虑二次（quadratic）期望效用函数的情形。证明在某个财富水平上，边际效用是递减的。更重要的是，证明绝对风险厌恶度在任何财富水平上都是递增的。

11.5 一枚硬币落地时正面向上的概率为 p 。有人邀请你参加下面这样的赌博：如果第 j 次投掷硬币时，硬币落地后才首次出现正面，那么你将得到 2^j 元。

（a）当 $p=1/2$ 时，这个赌博的期望价值为多少？

（b）假设你的期望效用函数为 $u(x)=\ln x$ 。求这个赌博带给你的效用表达式，这是一个求和式。

（c）求（b）中表达式的值。（这要求你掌握一定的求和公式。）

（d）令 w_0 表示一笔钱数，这笔钱带给你的效用和你参加这个赌博的效用是一样的。求 w_0 。

11.6 西野加奈（Esperanza）从他五岁起就是个期望效用最大化者。由于他在一所保守的英国寄宿学校上学，他的效用函数 u 是严格递增的和严格凹的。现在，已经 30 岁左右的他正打算评估某个风险资产，该资产的随机结果 R 是正态分布的（其中均值为 μ ，方差为 σ^2 ）。因此，它的密度函数为

$$f(r)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2\right\}$$

（a）证明西野加奈从 R 得到的期望效用是一个仅关于μ和σ^2的函数。因此，证明$E[u(R)]=\phi(\mu,\sigma^2)$。

（b）证明$\phi(\cdot)$关于μ递增。

（c）证明$\phi(\cdot)$关于σ^2递减。

11.7 令R_1和R_2分别表示两种风险资产的随机报酬。假设R_1和R_2是独立的而且具有相同的分布。证明一个期望效用最大者如果是厌恶风险的，那么他将会把他的财产分配在这两种风险资产上；如果他喜欢风险，他会将他的所有财产全部用于投资其中一种风险资产。

11.8 假设某个消费者面对两种风险，在这两种风险中只有一种可以消除。令$\tilde{w}=w_1$和$\tilde{w}=w_2$的概率分别为p和$1-p$。令$\tilde{\varepsilon}=0$若$\tilde{w}=w_2$；如果$\tilde{w}=w_1$，则$\tilde{\varepsilon}=\varepsilon_1$和$\tilde{\varepsilon}=\varepsilon_2$的概率分别为 1/2。现在定义$\tilde{\varepsilon}$的风险贴水$\pi_u$满足

$$E\left[u(\tilde{w}-\pi_u)\right]=E\left[u(\tilde{w}+\tilde{\varepsilon})\right] \qquad (*)$$

（a）证明如果$\tilde{\varepsilon}$足够小，那么

$$\pi_u \approx \frac{-\frac{1}{2}pu''(w_1)\varepsilon^2}{pu'(w_1)+(1-p)u'(w_2)}.$$

[提示：对（*）式两侧取泰勒适当阶数的展开式——左侧取一阶，右侧取二阶。]

（b）令$u(w)=-e^{-aw}$和$v(w)=-e^{-bw}$。计算u和v的阿罗-普拉特度量。

（c）假设$a>b$。证明如果$p<1$那么存在足够大的(w_1-w_2)使得$\pi_v>\pi_u$。对于风险只能部分降低的问题，上述结果意味着阿罗-普拉特度量还有用吗？

11.9 某个人的期望效用函数为$u(w)=\sqrt{w}$。他的初始财富为 4 元。他有一张彩票，这张彩票值 12 元和 0 元的概率分别为 1/2。求他的期望效用。求他卖掉彩票的最低价格p。

11.10 某个消费者的期望效用函数为$u(w)=\ln w$。他受邀参加投掷硬币的赌博，已知硬币正面向上的概率为π。如果他赌注x元，那么若正面向上他将得到$w+x$元，如果反面向上他将得到$w-x$元。请将x的最优值表示为π的函数。当$\pi=1/2$时，x的最优值为多少？

11.11 某个消费者的期望效用函数为$u(w)=-1/w$。他受邀参加一个赌博，在该赌博之下，他的财富为w_1和w_2的概率分别为p和$1-p$。他现在的财富水平为多少时正好能使得他在下列两个选择之间无差异：一是保留当前财产（不参加赌博）；二是参加赌博。

11.12 某个人关心他在下一期的货币收益，已知下一期可能发生的自然状态为$s=1,\ldots,S$。用x_s表示他在状态s下的货币收益，状态s发生的概率为p_s。假设此人为了使得收益的贴现期望值最大化，他将选择$x=(x_1,\ldots,x_s)$。贴现因子记为α，即$\alpha=1/(1+r)$，其中r为贴现率。可行收益集用X表示，我们假设这个集合非空。

（a）写出此人的最大化问题。

（b）定义$v(p,\alpha)$为此人在概率$p=(p_1,\ldots,p_S)$和贴现因子α之下能实现的最大贴现期

望值。证明 $v(p,\alpha)$ 关于 α 是一次齐次的。（提示：$v(p,\alpha)$ 和你见过的什么函数看起来相似？）

（c）证明 $v(p,\alpha)$ 是 p 的凸函数。

（d）假设对于各种 p 和 α 值，你能观察到任意大次数的关于 x 的选择。集合 X 应该具有什么性质，才能从上述观测到的选择行为将其还原？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 12 章：计量经济

曹乾　译

(东南大学　caoqianseu@163.com)

12 计量经济

在前面几章，我们分析了各种最优化行为的模型。在本章，我们将阐述如何使用在前面几章得出的理论思想，以帮助我们评估可能由这些最优化行为产生的各种关系。

理论分析和计量经济分析之间的互动表现在几个方面。首先，我们可使用理论推导出在经济上可进行检验的假设。其次，理论可以指导我们对模型参数进行更合理的估计。第三，理论可以帮助我们确定模型中的结构关系，从而可以导致更合适的估计。最后，理论可以帮助我们确定用于估计的合理方程形式。

12.1 最优化假设

我们已经知道，最优化选择模型对可观测到的行为施加了某些限制。这些限制可用几种方法进行表达：1）代数关系，例如利润最大化弱公理（WAPM），成本最小化弱公理（WACM），显示偏好的广义公理（GARP）等；2）导数关系，例如某些替代矩阵的条件是，这些矩阵必须是对称的而且是正半定或负半定的；3）对偶关系，例如利润必定是价格的凹函数这个事实。

最大化模型蕴涵的条件比较重要，原因有两个。一是，它们可以允许我们检验最大化行为的模型。如果数据不满足我们想使用的模型蕴涵的假设，那么一般来说这样的模型不适合用来描述可观测到的行为。

二是，这些条件允许我们更准确估计我们的模型。如果我们发现最优化施加的理论限制与某一特定数据集不矛盾，我们会试图重新估计我们的模型，以便要求这样的估计满足最优化蕴涵的限制。

例如，假设我们的最优化模型意味着某个参数a等于零。首先，我们想要验证这个限制，看看估计值α是否显著不为零。如果这个参数不是显著不等于零，我们可以接受$a=0$的假设，并且根据这个假设重新估计我们的模型。如果这个假设为真，对该系统内的其他参数的第二个估计集的估计就较准确。

当然，如果这个假设是错误的。重新估计的程序就不再是合适的。我们对估计的信心在某种程度上取决于我们对于最优化限制初始测试施加的信心。

12.2 最优化行为的非参数估计

如果我们获得关于企业选择的一组观测数据，我们可以直接检验前几章介绍的利润最大化弱公理和/或成本最小化弱公理的不等式。如果我们获得的是消费者选择的数据，类似显示偏好广义公理的条件的检验要稍微困难些。这些条件可以让我们判断我们获得的数据是

否是最优化行为产生的。

这些不等式条件容易检验；我们只要看看我们得到的数据是否满足某些不等式关系。如果我们观测到某个不等式被违背，我们就可以拒绝最优化模型。例如，假设我们得到一个企业选择的几套数据，这些数据是不同价格向量下企业的净产出：(p^t, y^t)，其中$t=1,...,T.$ 假设我们对该企业是否在竞争环境下最大化其利润感兴趣。我们知道利润最大化蕴涵利润最大化弱公理：$p^t y^t \geq p^s y^s$ 对于所有s和t成立。检验这个弱公理只要检验这些T^2个不等式是否成立即可。

在这个框架下，只要有一组观测数据使得$p^t y^t < p^s y^s$成立，我们就可以拒绝利润最大化模型。但是这个要求可能过于严格。一般来说我们真正关注的不是某个企业是否正好使得它的利润最大化，而是关注它的行为是否能由利润最大化模型合理刻画。通常，我们不仅想知道这个企业是否实现了利润最大化，还想知道如果未能实现，它离最大化的差距是多大。如果这个差距很小，我们仍然可以接受该企业“几乎”是利润最大化者的理论。

有种方法可以很自然地衡量利润最大化弱公理的被违背程度，这个方法称为“残差（residuals）”$R_t = \max_s\{p^t y^s - p^t y^t\}$。残差$R_t$衡量如果企业在观测期$t$作出与原来不同的选择将能多赚取多少利润。这个残差可以合理地衡量企业行为偏离利润最大化行为的程度。如果R_t的平均值较小，那么该企业的行为“几乎”为最优化行为。

12.3 最优化行为的参数估计

上面介绍的非参数检验是最优化的“精确”检验，它们是数据与最优化模型相一致的必要和充分条件。然而，经济学家经常对下列问题感兴趣：某种特定的参数形式是否是潜在的生产函数或效用函数的合理近似？

回答上述问题的一种方法是使用回归分析，或者使用更精细的统计技巧，来估计某个函数形式的参数，并且检验我们是否满足最优化模型施加的限制。例如，假设我们观测到k种商品的价格和选择。柯布-道格拉斯效用函数意味着商品i的需求函数是收入与其价格之比的线性函数：$x_i = a_i m / p_i$，其中$i=1,...,k.$

需求数据不可能恰好是m/p_i的线性函数，因此我们要增添一个误差项来代表衡量误差、错定(misspecification)和遗漏的变量等等。使用ε_i表示第i个式子的误差项，我们就得到了回归方程

$$x_i = a_i \frac{m}{p_i} + \varepsilon_i \qquad i=1,...,k. \tag{12.1}$$

从这个最大化模型可知$\sum_{i=1}^{k} a_i = 1.$ 我们可以估计（12.1）描述的模型中的参数，看看它们是否满足这个限制。如果满足，说明函数可能是柯布-道格拉斯形式；如果估计出的参数不满足这个限制，那么说明函数可能不是柯布-道格拉斯形式的。

如果我们使用更精细的函数形式，我们就得到一组更为复杂的可检验的限制。我们在学习消费者行为理论时已经知道，最大化行为施加的基本的限制是替代矩阵必定是负半定的。这个条件施加了一些交叉方程限制，这些限制可用标准假设检验程序进行检验。

12.4 施加最优化限制

如果我们的统计检验没有拒绝某种特定的参数限制，我们希望重新估计对估计程序施加这些限制的模型。仍使用前面的例子，（12.1）描述的柯布-道格拉斯方程组意味着 $\sum_{i=1}^{k} a_i = 1.$ 我们希望检验施加这个限制的参数组 (a_i) 是否满足这个假设。如果这个假设为真，那么我们进行的估计一般比无约束的估计的结果要好。

最优化模型通常不仅对误差项施加限制，也对参数施加限制。例如，这个理论模型施加的另外一个限制是 $\sum_{i=1}^{k} p_i x_i(p,m) = m.$ 一般来说，我们观测到的选择满足 $\sum_{i=1}^{k} p_i x_i = m$ 这个限制。如果的确如此，（12.1）意味着

$$\sum_{i=1}^{k} p_i x_i = m = \sum_{i=1}^{k} p_i m + \sum_{i=1}^{k} p_i \varepsilon_i .$$

如果我们想在 $\sum_{i=1}^{k} a_i = 1$ 的限制下估计我们的方程组，我们也希望它们满足 $\sum_{i=1}^{k} p_i \varepsilon_i = 0.$ 也就是说，这 k 个误差项必然与价格向量是正交的。

12.5 最优化模型的拟合度

上一节简要介绍的参数检验，描述了如何在统计上检验观测到的选择是否由某个特定参数形式的最优化模型产生的。这样的检验是非常“敏感的”（sharp），因为我们要么拒绝最优行为的假设，要么接受。但在很多情形下，进行拟合度（goodness-of-fit）测度更为合适：观测到的选择与最优化选择的接近程度。

为了回答这个问题，我们需要对“接近”这个词进行准确定义。在利润最大化的非参数分析中，我们已经知道，一种合理的测度方法是如果企业另行选择，它能额外赚取多少利润。这个思想的应用性比较广泛：拟合度的一种测度方法是看看经济人离最优化规定的目标函数的距离有多远。

在分析企业行为的情形下，可以直接计算这种测度。如果我们的假设是利润最大化或成本最小化我们可将最佳拟合最优化模型与实际选择进行比较，计算利润损失或成本的增加额。如果这种方法应用于效用最大化，则结果更为微妙。

假设我们使用柯布-道格拉斯类型的函数来考察消费者的选择行为。如果最优拟合柯布-道格拉斯效用函数可用参数 $(\hat{a}_i)$ 描述，我们可以使用估计出的效用函数计算最优选择的效用，并将其与实际选择的效用进行比较。

这种衡量方法的问题在于，效用函数的单位是任意的。到底哪个函数最“接近”实际情形，我们并不是很清楚。这个问题的解决之道是在计算拟合度时使用特殊的效用函数。此处的一个自然而然的选择是使用货币制效用函数或称为用货币度量的效用函数（详见第 7 章）。货币制效用函数用货币单位衡量能够：在某个既定的价格向量下，为了使消费者的效用和他消费商品束 x 的效用相同，他需要多少钱。

下面我们看看如何使用货币制的效用函数构造优秀的拟合。假设我们观测到某组数据 (p^t, x^t)，其中 $t = 1, ..., T$ 。我们假设消费者打算最大化的是效用函数 $u(x,\beta)$，其中 β 是一个未知参数（或一组未知参数）。给定 $u(x,\beta)$，我们就可以使用标准的最优化技术来构建货币制效用函数 $m(p,x,\beta)$ 。

我们使用选择数据来估计能最准确描述观测到的选择行为的效用函数 $u(x,\hat{\beta})$ 。判断这个效用函数拟合程度的一种方法是计算 t “残差”（residuals）

$$G^t = \frac{m(p^t, x^t, \hat{\beta})}{p^t x^t}.$$

这里的 G^t 衡量的是与消费者实际花费的钱数相比，他为了得到效用 $u(x^t,\hat{\beta})$ 而必需的最小钱数。这自然可用效率术语解释：如果 G^t 的平均值为 $\bar{G}$ ，那么我们可以说，在平均意义上，消费者的选择行为具有 $(100\bar{G})\%$ 的效率。

如果消费者完全实现了效用函数 $u(x,\hat{\beta})$ 的最大化，那么 $\bar{G}$ 将等于 1，也就是说消费者的选择行为具有 100%的效率。如果 $\bar{G}$ 等于 0.95，则消费者的选择行为具有 95%的效率，以此类推。

12.6 结构模型与简化形模型

假设我们的某个理论表明，一些变量之间存在着某种关系。一般来说，我们的模型中存在着两类变量。一类称为内生变量（endogenous variables），外生变量的值由我们的模型决定；另一类称为外生变量（exogenous variables），外生变量的值是由模型以外的因素事先决定。例如，在我们的利润最大化行为模型中，价格和技术是外生变量，要素选择是内生变量。

例如，考虑一个简单的需求和供给系统：

$$\begin{aligned} D &= a_0 - a_1 p + a_2 z_1 + \varepsilon_1 \\ S &= b_0 - b_1 p + b_2 z_2 + \varepsilon_2 \\ D &= S \end{aligned} \tag{12.2}$$

这里的 D 和 S 分别表示某种商品的需求和供给，它们都是内生变量；P 是商品的价格，也是内生变量；(a_i) 和 (b_i) 是参数；z_1 和 z_2 是影响需求和供给的其他外生变量。变量 ε_1 和 ε_2 是误差项。（12.2）是个结构性系统。

我们可以求解结构系统，使得内生变量 p 可以表示为外生变量的函数：

$$p=\frac{a_0-b_0}{a_1+b_1}+\frac{a_2}{a_1+b_1}z_1-\frac{b_2}{a_1+b_1}z_2+\frac{\varepsilon_2-\varepsilon_1}{a_1+b_1} \qquad (12.3)$$

这称为该系统的**简化形**（reduced form）。

简化形的模型通常容易估计。在上述需求供给模型的例子中，我们估计下列回归方程即可：

$$p=p_0+\beta_1 z_1+\beta_2 z_2+\varepsilon_3.$$

参数(β_i)是关于参数(a_i,b_i)的函数，但一般不可能从简化形方程的参数β_i还原出原来结构方程的参数(a_i,b_i)。简化形参数可用于预测当外生变量变化时，均衡价格将如何变化。这在某些情形下比较有用。

但对于另外一些情形，我们有必要估计结构参数。例如，假设我们想估计的是下列情形：政府对商品征税后，均衡市场价格将如何变化？结构模型（12.2）表明，商品供给者得到的均衡价格p_s应该为税收的线性函数：

$$p=\frac{a_0-b_0}{a_1+b_1}+\frac{a_2}{a_1+b_1}z_1-\frac{b_2}{a_1+b_1}z_2-\frac{b_1}{a_1+b_1}t. \qquad (12.4)$$

如果我们有很多数据，这些数据是关于税的不同选择和供给者得到的相应价格。那么我们就可以估计如（12.4）描述的简化形方程。但是，如果我们没有这样的数据，就无法估计这样的简化方程。为了预测均衡价格对税收的反应，我们应该知道结构参数$b_1/(a_1+b_1)$。类似（12.3）式中的简化形参数无法提供足够的信息来回答这个问题。

这意味着我们必须想出估计类似（12.2）的结构方程系统的方法。最简单的方法似乎就是使用普通最小二乘法（ordinary least squares, OLS）回归技术，分别估计需求方程和供给方程。这种方法能提供可接受的参数估计吗？

我们从统计学知识知道，如果某些假设条件得以满足，OLS 方法具有让人满意的性质。其中一个特别的假设是，回归方程右侧的变量不能和误差项相关。

然而，我们的问题显然不满足上述这个假设。从（12.3）式很容易看出，变量p取决于误差项ε_1和ε_2。可以证明，这种依赖性通常将导致参数的有偏估计。

为了估计结构方程系统，我们一般需求更复杂的估计技术，例如量阶段最小二乘法或者各种最大似然技术。在估计涉及方程组的问题时，这些方法通常比 OLS 法具有更好的统计性质。

在上面介绍的需求供给例子中，变量之间的理论上的关系意味着某些估计方法比另外的估计方法更合适。事实上，情形通常如此；计量经济学的艺术之一就是用理论指导统计方法的选择。我们将在下一节继续分析这一问题。

12.7 估计技术关系

假设我们想估计一个简单的柯布-道格拉斯生产函数的参数。为了更具体一些，假设我们有农场的样本，农场i的玉米产量C_i，取决于种植的玉米数量K_i和玉米生长期的晴天天数S_i。我们暂时假设影响玉米产量的变量只有这两个。

我们假设生产关系由柯布-道格拉斯生产函数$C_i = K_i^a S_i^{1-a}$所给出。取对数，我们可以将产出和投入的关系写为

$$\log C_i = a \log K_i + (1-a) \log S_i \tag{12.5}$$

假设农场主在制定种植决策时无法观测到晴天天数。而且，计量经济学家没有每个农场的晴天天数数据。因此，计量经济学家将（12.5）视为下列形式的回归模型

$$\log C_i = a \log K_i + \varepsilon_i \tag{12.6}$$

其中ε_i是“误差项” $(1-a)\log S_i$。

计量经济理论告诉我们，如果$\log K_i$和ε_i是不相关的，则 OLS 能比较准确地估计出参数a。如果农场主们在选择K_i时没有观测到$\log S_i = \varepsilon_i$，那么他们的选择不可能受其影响。因此，在这种情形下，这是个合理假设，从而 OLS 是合适的估计方法。

下面我们再举个例子，在个例子中，OLS 不再是个好的估计方法。现在假设生产关系也取决于每个农场土地的质量，因此$C_i = Q_i K_i^a S_i^{1-a}$，或者

$$\log C_i = \log Q_i + a \log K_i + (1-a) \log S_i.$$

和前面的情形一样，我们假设计量经济学家和农场主都观测不到S_i。然而，我们假设农场主能观测到Q_i而计量经济学家不能。现在如果我们估计（12.6）的回归方程，我们能得到比较准确的a吗？

答案是否定的。由于每个农场主能观测到Q_i，他对K_i的选择将取决于Q_i。因此，K_i将和误差项相关，可能导致有偏估计。

如果我们假设农场主的行为是利润最大化的，我们将清楚地知道农场主如何使用K_i的信息。农场主i的（短期）利润最大化问题为

$$\max \quad p_i Q_i K_i^a S_i^{1-a} - q_i K_i,$$

其中p_i是产出的价格，q_i是农场主i面对的种子价格。对K_i求导，求要素需求函数可得

$$K_i = \left[a Q_i \frac{p_i}{q_i} \right]^{\frac{1}{1-a}} S_i \tag{12.7}$$

显然，农场主掌握的Q_i信息直接影响他的种植量选择，从而影响产量。

考虑图 12.1 中$\log K_i$和$\log C_i$散点图。我们还画出了函数$\log C_i = a \log K_i + \bar{Q}$的图形，

其中$\overline{Q}$是平均产量。

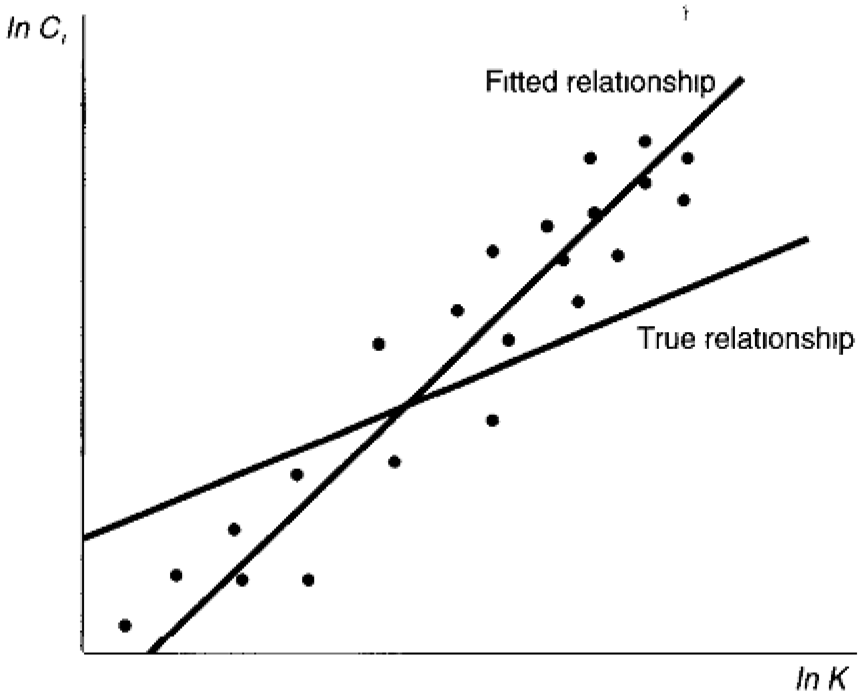


图 12.1：散点图。这是$\log K_i$和$\log C_i$的散点图。注意，K_i大的农场的土地质量通常高于平均水平，因此它的产量也高于平均质量土地的产量。因此，这样的点将位于拥有平均质量土地的农场的生产关系曲线的上方。

从（12.7）式明显可以看出，土地质量较高的农场主将希望多种玉米。这意味着如果我们看到某个农场投入的K_i越大，该农场的土地质量Q_i可能越高。因此，这样的农场的产量比拥有平均质量土地的农场的产量高，因此K_i较大的农场的数据点将位于拥有平均土地质量的农场的实际生产关系曲线的上方。类似地，某个农场投入的K_i越小，该农场的土地质量越有可能低于平均质量。

结果是，拟合这些数据的回归线估计出的a比a的实际值大。潜在的问题是高产量并不完全是由高投入引起的。还存在第三个变量 [遗漏变量（omitted variable）]，即土地质量，这个变量不仅会影响产量也会影响农场主对投入的选择。

这种偏差在计量经济工作中非常普遍：通常情形是有些影响选择的回归变量本身是由决策者选择的。例如，假设我们想估计教育的回报。通常，收入较高的人拥有教育的数量较多，但是教育不是先决变量（predetermined variable）:人们选择教育数量。如果人们选择不同的教育数量，他们在其他未被观察到的方面也许也会不同。但这些不同也许也会影响他们的收入。

例如，假设智商较高的人挣得工资较高，不管他们的教育数量如何。但是智商较高的人会发现他们容易获得较多的教育。这意味着对于拥有较多教育数量的人来说，他们工资较高的原因有两个：首先，他们的平均 IQ 较高；其次，因为他们获得的教育数量较多。如果简单地用工资对教育回归，将会高估教育对收入的影响。

或者，你可能认为父母富有的人的收入也较高。但是富有的父母能够买得起更多的教育，也可以分给子女更多的财富。再一次地，较高收入和较高的教育水平相关，但这两个变量之间可能不存在直接的因果关系。

简单的回归分析适合于对照实验（controlled experiments），但是它们通常不足以处理解释变量是由当事人选择的情形。在这种情形下，我们需要使用结构模型，从而将所有相关变量表达为真正的外生变量的函数。

12.8 估计要素需求

在生产关系的情形下，间接估计生产关系的参数有时是有用的。例如考虑（12.7）式，取对数，我们可以将其写为

$$\log K_i = \frac{1}{1-a}\log a + \frac{1}{1-a}\log p_i - \frac{1}{1-a}\log q_i + \frac{1}{1-a}\log Q_i + \log S_i$$

这个式子的合适回归方程为

$$\log K_i = \beta_0 + \beta_1 \log p_i + \beta_2 \log q_i + \varepsilon_i\,,$$

其中：常数项 β_0 是关于 a， $\log Q_i$ 与 $\log S_i$ 的均值的某个函数。注意这个式子意味着 $\beta_2 = -\beta_1$。

这个方程能用 OLS 方法估计吗？如果农场主们面对的产出品和投入物市场都是竞争性的，答案是肯定的，这是因为在竞争性的市场上，价格不受农场主控制。如果价格和误差项不相关，那么 OLS 是合适的估计方法。

另外，在最优化模型中 $\beta_1 = -\beta_2$ 这个事实，也为我们检验柯布-道格拉斯生产函数是否为最优化模型的一种方法。如果我们发现 β_1 明显与 $-\beta_2$ 不同，我们可以拒绝上述最优化模型的假说。另一方面，如果我们不能拒绝 $\beta_1 = -\beta_2$ 这个假说，我们可能倾向于接受这个假说，并且估计下列模型

$$\log K_i = \beta_0 + \beta_1 \log(p_i / q_i) + \varepsilon_i\,.$$

在这种情形下，需求函数是个结构方程：它将消费者的选择表达为外生变量的函数。我们可以使用这个方程的估计式推断生产技术的其他性质。

12.9 更复杂的生产技术

考虑下列情形：在生产函数中，产出是若干种投入物的函数。为简单起见，我们考虑投入物为两种的柯布-道格拉斯类型的生产函数： $f(x_1,x_2)=Ax_1^a x_2^b$.

我们在第 4 章（详见第 4 章 4.3 节中的例子）知道，要素需求函数的形式为

$$x_1(w_1,w_2,y)=A^{-\frac{1}{a+b}}\left[\frac{aw_2}{bw_1}\right]^{\frac{b}{a+b}} y^{\frac{1}{a+b}}$$

$$x_2(w_1,w_2,y)=A^{-\frac{1}{a+b}}\left[\frac{aw_2}{bw_1}\right]^{-\frac{a}{a+b}} y^{\frac{1}{a+b}}$$

这些需求函数具有对数线性性质，所以我们可以将回归模型写为

$$\log x_1=\beta_{01}+\beta_{11}\log(w_2/w_1)+\beta_{21}y+\varepsilon_1$$

$$\log x_2=\beta_{02}+\beta_{12}\log(w_2/w_1)+\beta_{22}y+\varepsilon_2.$$

在这里，生产技术的参数是回归系数的函数。然而，重要的是可观察到参数 a 和 b 进入了回归系数的定义。这表明这两个方程的参数不是不受约束的，而是相关的。例如，容易看出 $\beta_{01}=\beta_{02}$。在估计这个方程组时，应该考虑**交叉方程约束**（cross-equation restrictions）。

或者，我们可以将这两个方程组成成本函数 $c(w,y)$：

$$c(w_1,w_2,y)=A^{\frac{-1}{a+b}}\left[\left(\frac{a}{b}\right)^{\frac{b}{a+b}}+\left(\frac{a}{b}\right)^{\frac{-a}{a+b}}\right]w_1^{\frac{a}{a+b}}w_2^{\frac{a}{a+b}}y^{\frac{1}{a+b}}.$$

这个式子也有对数线性形式

$$\log c=\log\gamma_0+\gamma_1\log w_1+\gamma_2\log w_2+\gamma_3\log y.$$

这样，要素需求函数的交叉方程限制就很方便地整合到一个式子即成本函数之中。而且，从我们的成本函数的理论研究可知，它应该是个递增、齐次的凹函数。我们可以检验这些限制，在有需要的时候也可以做出这样的限制性假设。

实际上，成本函数可以视为要素需求函数组的简化形式。与我们前面研究的需求和供给例子不同，成本函数包含了结构模型的所有相关信息。因为我们已经知道对成本函数求导即可得到条件要素需求。因此，成本函数参数的估计自动地给出了条件要素需求函数的参数估计。

然而，应该强调的是，上述结论只有在成本最小化的假设下才成立。如果我们考察的企业真得是追求成本最小化或利润最大化，我们可以使用若干间接方法估计生产技术参数。如果最优化假设成立，这些间接方法通常比直接方法好。

12.10 函数形式的选择

我们在前面几节所举的例子中使用的都是柯布-道格拉斯函数形式。这是出于简化分析的目的，未必符合现实。事实上，我们应该使用范围更宽的参数形式来表示生产技术之间的权衡。

你可以将任意一个函数作为生产函数，但接下来你必须计算该生产函数蕴涵的要素需求和（或）成本函数。通常，直接从参数形式的成本函数入手更为简单，因为这样做之后，剩下的工作只是通过微分方法找到合适的要素需求。

我们从第 6 章已经知道，任何关于价格的单调、齐次的凹函数都是某个良好表现生产技术的成本函数。因此，我们的全部工作仅是找到满足这些性质的一个函数形式。

一般来说，我们希望选择下面这样的参数形式，对于某些值，这些参数满足最优化施加的限制，对另外一些值，则不满足。于是我们可以估计这些参数，检验所估计出的参数满足理论施加的相关限制的假说。下面举例说明。

例：戴沃特成本函数

戴沃特成本函数（Diewert cost function）的形式为

$$c(w,y)=y\sum_{i=1}^{k}\sum_{j=1}^{k}b_{ij}\sqrt{w_i w_j}.$$

对于这个函数形式，我们要求 $b_{ij}=b_{ji}$。注意，我们也可以将上式写为

$$c(w,y)=y\left[\sum_{i=1}^{k}b_{ii}w_i+\sum_{i\neq j}^{k}\sum_{j\neq i}^{k}b_{ij}\sqrt{w_i w_j}\right].$$

由于上式的第一部分是里昂惕夫成本函数形式，这个函数也称为**广义的里昂惕夫成本函数**（generalized Leontief cost function）。

要素需求的形式为

$$x_i(w,y)=y\sum_{j=1}^{k}b_{ij}\sqrt{w_j/w_i}.$$

这些需求关于参数 b_{ij} 是线性的。如果 $b_{ij}\geq 0$ 和某些 $b_{ij}>0$，容易验证这种形式满足该函数是成本函数的必要条件。

参数 b_{ij} 与不同要素之间的替代弹性有关；b_{ij} 越大，要素 i 和要素 j 之间的替代弹性越大。函数形式对各种弹性未施加任何限制；戴沃特成本函数可以作为任何成本函数的局部二阶近似。

例：超越对数成本函数

超越对数成本函数（translog cost function）的形式为

$$\log c(w,y)=a_0+\sum_{i=1}^{k}a_i\log w_i+\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}b_{ij}\log w_i\log w_j+\log y.$$

对于这个函数，我们要求

$$\sum_{i=1}^{k}a_i=1$$

$$b_{ij}=b_{ji}$$

$$\sum_{j=1}^{k}b_{ij}=0\text{。}$$

在这些限制之下，超越对数成本函数关于价格是齐次的。如果对于所有 i 和 j 都有 $a_i>0$ 和 $b_{ij}=0$，那么成本函数变为柯布-道格拉斯函数。

条件要素需求关于参数不是线性的，但是要素份额 $s_i(w,y)=w_ix_i(w,y)/c(w,y)$ 关于参数是线性的，而且要素份额由下式给出

$$s_i(w,y)=a_i+\sum_{j=1}^{k}b_{ij}\ln w_i\ .$$

12.11 估计消费者需求

我们前面的例子主要是生产关系的估计。它们有着方便的特征，即目标函数（利润函数或成本函数）是可观测到的。在消费者需求行为的情形下，目标函数不是可直接观测到的。这在概念上使得事情变得稍微复杂一些，但这并不意味着事情变得非常棘手。

假设我们已获得消费数据 (p^t,y^t)（其中 $t=1,...,T$），因此想估计某个参数形式的需求函数。我们首先分析单一商品的需求情形，然后分析多种商品的需求情形。

单一商品的需求函数

首先需要注意，即使我们仅关注单一商品的需求，实际上仍然涉及到两种商品：一是我们想研究的那种商品；二是“所有其他商品”。我们通常将这样的选择问题视为消费者在我们想研究的那种商品和他花费在所有其他商品上的货币上的选择。请参见第 9 章关于希克斯可分性的讨论。

假设我们用 x 表示 x 商品的购买量，用 y 表示花费在所有其他商品上的钱数。如果 x 商品和 y 商品的价格分别为 p 和 q，那么消费者的效用最大化问题可以表述为

$$\max_{x,y} u(x,y)$$
$$\text{s.t.} \quad px+qy=m.$$

我们用$x(p,q,m)$表示需求函数。由于需求函数是零次齐次的，我们可以用q将其他两个变量标准化，因此需求函数变为x的相对价格和消费者相对收入的函数$x(p/q,m/q)$。在实践中，p是我们想要研究的那种商品的名义价格，而q通常取成消费者价格指数。于是，需求关系是说观测到的需求量是“实际价格”p/q和“实际收入”m/q的函数。

两种商品情形下的问题具有一种非常方便的性质：几乎任何函数形式都与效用最大化一致。我们从第 8 章 8.5 节中的例子可知，两种商品情形下的积分方程可以表述为一个常微分方程。因此，通常存在能够产生需求方程的间接效用函数（使用罗伊法则）。在本质上，最优化为两种商品情形施加的唯一限制是补偿性的自身价格效应（compensated own price effect）应该为负。

这意味着你在选择符合最优化的函数形式方面有很大的自由。三种常见的函数形式为

1）线性需求：$x=a+bp+cm$。

2）对数需求：$\ln x=\ln a+b\ln p+c\ln m$。

3）半对数需求：$\ln x=a+bp+cm$。

上述每个式子都对应着一个间接效用函数。我们在第 26 章推导出了对数需求的间接效用函数，而将线性需求和半对数需求的情形留作习题。对需求函数参数的估计可自动得到间接效用函数的参数估计。

一旦我们有了间接效用函数，我们可以使用它进行各种预测。例如，我们可以使用估计式计算与价格变化对应的补偿变化或等价变化。具体细节请参见第 10 章。

多个方程

假设我们想估计一个含有两种商品以上的需求系统。在这种情形下，我们可以从需求方程的某个函数形式入手，然后将它们积分来找到效用函数。然而，下面的方法通常更简单一些：先找到直接效用或间接效用的函数形式，然后微分即可找到需求函数。

例子：线性支出系统

假设效用函数形式为

$$u(x)=\sum_{i=1}^{k} a_i \ln(x_i-\gamma_i)$$

其中$x_i>\gamma_i$。效用最大化问题为

$$\underset{x_i}{\text{Max}} \sum_{i=1}^{k} a_i \ln(x_i - \gamma_i)$$
$$\text{s.t. } \sum_{i=1}^{k} p_i x_i = m.$$

如果我们令 $z_i = x_i - \gamma_i$，我们可以将上面的效用最大化问题写为

$$\underset{z_i}{\text{Max}} \sum_{i=1}^{k} a_i \ln z_i$$
$$\text{s.t. } \sum_{i=1}^{k} p_i z_i = m - \sum_{i=1}^{k} p_i \gamma_i.$$

这是关于 z_i 的柯布-道格拉斯最大化问题。容易看出 x_i 的需求函数具有下列形式

$$x_i = \gamma_i + a_i \frac{m - \sum_{i=1}^{k} p_i \gamma_i}{p_i} \text{。}$$

例子：几乎理想的需求系统

几乎理想的需求系统（Almost Ideal Demand System, AIDS）的支出函数具有下列形式

$$e(p,u) = a(p) + b(p)u \qquad (12.8)$$

其中

$$a(p) = \alpha_0 + \sum_i a_i \log p_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij}^* \log p_i \log p_m$$
$$b(p) = \beta_0 \prod p_i \beta_i.$$

由于 $e(p,u)$ 关于 p 必定是齐次的，参数必定满足

$$\sum_{i=1}^{k} a_i = 1$$
$$\sum_{i=1}^{k} \gamma_{ij}^* = \sum_{j=1}^{k} \gamma_{ij}^* = \sum_{i=1}^{k} \beta_i = 0.$$

对（12.8）式微分可得到需求函数。然而，估计支出份额通常更方便一些

$$s_i = \alpha_i + \sum_{j=1}^{k} \gamma_{ij} \log p_j + \beta_i \log \frac{m}{P} \qquad (12.9)$$

其中 P 是由下式给出的价格指数

$$\log P = \alpha_0 + \sum_{i=1}^{k} \log p_i + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} \gamma_{ij} \log p_i \log p_j,$$

和

$$\gamma_{ij} = \frac{1}{2}(r_{ij}^* + r_{ji}^*).$$

除了价格指数这一项之外，AIDS 系统接近于线性。在实践中，计量经济学家通常使用任意的价格指数计算 m/P 项，然后使用（12.9）估计系统的其他参数。

12.12 总结

我们已经看到，最优化模型的理论分析能够在几个方面对计量经济学家给与指导。首先，它提供了使用非参数或参数形式来检验理论的方法。其次，理论上的一些假设限制能帮助经济学家构造更有效率的估计。第三，理论可以说明模型中的结构关系，指导经济学家选择什么样的估计方法。最后，理论能帮助经济学家决定选择什么样的函数形式。（本章结束）■

注释

用消费者理论估计需求系统方面的内容可参见 Deaton and Muelbauer（1980）。Varian（1990）更加详细地讨论了拟合度并且给出了一些实证例子。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 13 章：竞争性市场

曹乾　译

(东南大学　caoqianseu@163.com)

13 竞争性市场

到目前，我们已研究了个体（企业和消费者）的最优化行为。我们将经济环境视为给定的，完全由市场价格向量刻画。在本章，我们开始分析市场价格是如何由个体的行为决定的。我们从最简单的模型入手分析：单个的竞争性市场。

13.1 竞争性企业

竞争性企业（competitive firm）是指将产品的市场价格视为给定的、不受自己控制的企业。在竞争性市场中，每个企业都认为价格和自己的行为无关，尽管市场价格是由所有企业的行为共同决定的。

令 $\overline{p}$ 为市场价格。于是某个标准竞争性企业面临的需求曲线为

$$D(p)=\begin{cases}0 & p>\overline{p}\\ \text{任何数量} & p=\overline{p}\\ \infty & p<\overline{p}.\end{cases}$$

一个竞争性企业可以自由定价也可以自由定产。然而，如果该竞争性企业制定的价格高于当前的市场价格，没有人会购买它的产品。如果它制定的价格低于市场价格，那么它想要有多少消费者就有多少消费者；但它没有必要放弃利润，因为如果它制定的价格等于市场价格，它也能做到想要多少消费者就有多少消费者。有时我们也说竞争性企业面临的需求曲线具有完全弹性。

如果竞争性企业想出售产品，它必须按照市场价格出售。当然，现实世界很少能做到这么理想。问题不是在于是否任何特定的市场是完全竞争的——几乎没有市场是完全竞争的。合理的问题是什么样程度的完全竞争模型可以产生洞穿现实市场的思想。正如物理学中无摩擦的模型可以描述现实世界中的一些重要现象一样，完全竞争模型可以产生让我们理解经济世界的思想。

13.2 利润最大化问题

由于竞争性企业将市场价格视为给定的，它的利润最大化问题非常简单。它选择的产量 y 必然是下列问题的解

$$\max_{y} \; py-c(y)$$

这个问题的内部解的一阶条件和二阶条件分别为

$$p=c'(y^*)$$

$$c''(y^*)\geq 0.$$

我们通常假设二阶条件为严格不等式。这么做不是必需的，但是它可以简化计算。我们将这种情形称为常规情形。

反供给函数，以 $p(y)$ 表示，它衡量为使企业供给既定数量的产品市场价格必须为多少。根据一阶条件可知，反供给函数为

$$p(y) = c'(y),$$

只要 $c''(y) > 0.$

供给函数给出了在每个价格水平下的利润最大化的产量。因此，供给函数 $y(p)$ 必定恒等地满足一阶条件

$$p = c'(y(p)), \tag{13.1}$$

也必须满足二阶条件

$$c''(y(p)) > 0.$$

直接供给曲线和反供给曲线衡量的关系是相同的，它们都衡量价格和利润最大化产量水平之间的关系。区别在于它们描述的方法不同。

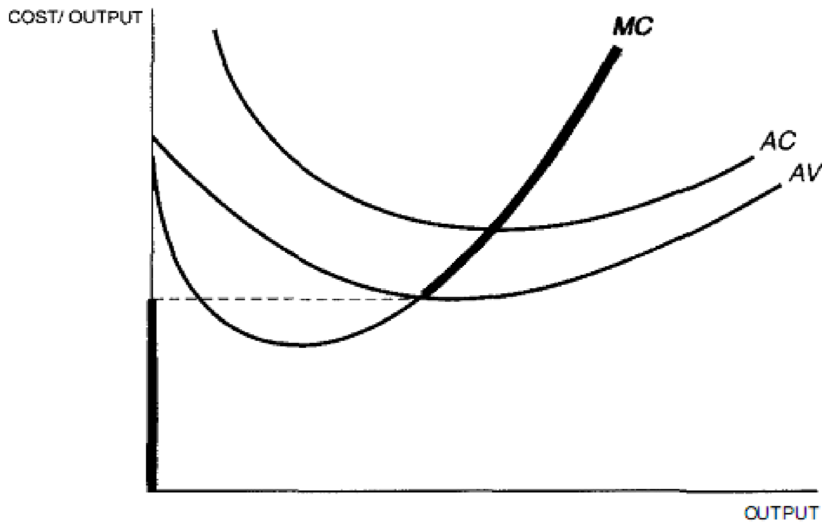


图 13.1：供给函数和成本曲线。在常规情形，竞争性企业的供给函数是位于平均可变成本曲线之上的向上倾斜的那部分边际成本曲线。

竞争性企业的供给如何随产品的价格变化而变化？我们将（13.1）式对 p 求导可得

$$1 = c''(y(p))y'(p).$$

由于通常 $c''(y) > 0$，由此可知 $y'(p) > 0$。因此，竞争性企业的供给曲线斜率为正，至少在常规情形是这样的。我们在第 2 章已经得到了这一结果，不过在那里我们使用的是另外一种方法。

我们的注意力一直放在利润最大化问题的内部解上，但是何时选择内部解也是个有趣的问题。我们将成本函数写为 $c(y)=c_v(y)+F$ ，即总成本等于可变成本和固定成本之和。我们将固定成本解释为真正固定的——即使产量为零也必须支付这些成本。在这种情形下，企业会发现当产量为正时的利润大于产量为零时的利润（损失）时，企业会进行生产：

$$py(p)-c_v(y(p))-F\geq -F.$$

整理可知，企业在

$$p\geq\frac{c_v(y(p))}{y(p)}$$

时，即当价格大于平均可变成本时，企业会进行生产。如图 13.1 所示。

13.3 行业供给函数

行业供给函数（industry supply function）是行业内所有企业供给函数之和。如果 $y_i(p)$ 表示第 i 个企业的供给函数，企业总数为 m 个，则行业供给函数为

$$Y(p)=\sum_{i=1}^{m}y_i(p).$$

行业的反供给函数就是行业供给函数的反函数：它给出了行业愿意供给某给定数量的最低价格。由于每个企业选择的产量水平都要满足价格等于边际成本这个条件，产量相同的企业的边际成本必定也是相同的。行业供给函数衡量行业产量和生产该产量的共同边际成本之间的关系。

例子：不同成本函数

假设某行业由两个企业组成，一个企业的成本函数为 $c_1(y)=y^2$，另外一个企业的成本函数为 $c_2(y)=2y^2$。供给函数为

$$y_1=p/2$$
$$y_2=p/4.$$

行业供给曲线因此为 $Y(p)=3p/4.$ 对于行业的任何产量水平 Y，每个企业的边际生产成本为 $4Y/3.$

例子：相同成本函数

假设行业中共有 m 个企业，每个企业的成本函数都为 $c(y)=y^2+1$。边际成本函数为 $MC(y)=2y$，平均可变成本函数为 $AVC(y)=y.$ 由于在这个例子中，边际成本总是大于平均可变成本，企业的反供给函数为 $p=MC(y)=2y.$

由此刻知，企业的供给函数为 $y(p) = p/2$，行业的供给函数为 $Y(p,m) = mp/2$。行业的反供给函数因此为 $p = 2Y/m$. 注意，行业反供给曲线的斜率随着行业中企业数量增加而变小。

13.4 市场均衡

行业供给曲线衡量在任何价格水平下行业的总产量供给。行业需求函数衡量在任何价格水平下对该行业产品的总需求。均衡价格是需求量等于供给量时的价格。

为什么要将这样的价格称为均衡价格？通常认为，如果在某个价格水平下需求不等于供给，某个经济人就会单方面改变自己的行为。例如，如果在某价格水平下，供给大于需求。在这种情形下，某些企业的产品无法全部销售出去。这些企业会减少产量，因为这样会降低生产成本而又不会导致销售收入的减少，因此增加了利润。所以这样的价格不可能是均衡价格。

如果令 $x_i(p)$ 表示个人 i 的需求函数，其中 $i = 1,...,n$； $y_j(p)$ 表示企业 j 的供给函数，其中 $j = 1,...,m$，则均衡价格是指下列方程的解

$$\sum_{i=1}^{n} x_i(p) = \sum_{j=1}^{m} y_j(p).$$

例子：相同企业

假设行业需求曲线是线性的， $X(p) = a - bp$, 行业供给曲线是上一节推导出的，$Y(p,m) = mp/2$。均衡价格是下列方程的解

$$a - bp = mp/2$$

由此可解得

$$p^* = \frac{a}{b + m/2}$$

注意在这个例子中，均衡价格随着企业数量增加而下降。

对于任意的行业需求曲线，均衡由下式决定

$$X(p) = my(p).$$

当 m 变化时均衡价格如何变化？我们将 p 看成 m 的隐函数，微分可得

$$X'(p)p'(m) = my'(p)p'(m) + y(p),$$

由此可得

$$p'(m) = \frac{y(p)}{X'(p) - my'(p)}.$$

假设行业需求曲线的斜率为负，则均衡价格必然随着企业数量的增加而下降。

13.5 进入

上一节计算行业供给曲线时我们假定企业数量是外生给定的。然而，在长期，行业中的企业数量是可变的。如果某个企业认为它生产某种产品可以挣钱，我们可以预期它会生产。类似地，如果行业中已有的企业发现它持续亏损，我们可以预期它会退出该行业。

根据对进入和退出成本以及潜在进入者的预见能力的假设不同，可以做出几种不同的进入和退出模型。在本节我们介绍一种特别简单的模型，该模型假设进入和退出成本为零，以及潜在进入者具有完全的预见能力。

假设行业中的企业数量任意多，但它们的成本函数 $c(y)$ 是相同的。我们可以计算出盈亏平衡时的价格 p^*，此时最优供给数量产生的利润为零。这种情形也就是平均成本等于边际成本。

现在我们可以画出行业中企业数量为 1,2,…时的行业供给曲线，看看企业盈亏平衡时行业内能容纳的最大企业数量。如图 13.2 所示。如果均衡时企业数量很大，那么相关的供给曲线将非常平坦，均衡价格将接近于 p^*。因此，经常假设完全竞争行业（可自由进出）的供给曲线基本是一条水平线，此时价格等于平均成本的最小值。

在这个进入模型中，均衡价格可以大于盈亏平衡时的价格。尽管行业中的企业可以赚取正的利润，潜在进入者也不会再进入，因为它们可以正确地预见到如果它们进入将会导致利润为负。

和往常一样，正的利润可以视为经济租金。在这种情形下，我们可以将利润视为“捷足先登者的租金（rent to being first）”。也就是说，投资者愿意支付一定资金将行业中的现有企业买下，这笔资金等于该企业利润流的现值，投资者这么做的原因在于获取这个利润流。这个租金可以算为继续呆在行业中的（机会）成本。如果我们按照这种会计传统计算，均衡时企业的利润为零。

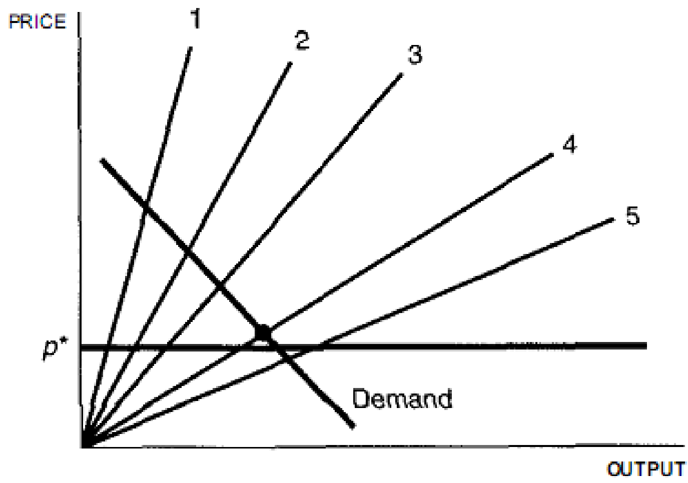


图 13.2：**企业的均衡数量**。在我们的进入模型中，企业的均衡数量是指盈亏平衡时行业内能容纳的企业的最大数量。如果这个数量很大，那么均衡价格必然接近于平均成本的最小值。

例子：进入和长期均衡

如果 $c(y)=y^2+1$，则令平均成本等于边际成本，就可以找到盈亏平衡时的产量：

$$y+1/y=2y,$$

由此可得 $y=1$。在这个产量水平下，边际成本为 2，因此这是盈亏平衡时的价格。根据我们的进入模型，只要企业认为它们的进入不会迫使均衡价格下降到 2 以下，它们就会进入该行业。

假设需求是线性的，我们继续使用上例中的那个线性需求。于是均衡价格是满足下列条件的最小 p^*

$$p^*=\frac{a}{b+m/2}$$
$$p^*\geq 2.$$

随着 m 的增加，均衡价格必定越来越接近于 2。

13.6 福利经济学

我们已经知道如何计算竞争均衡：价格使得供给等于需求。在本节，我们分析均衡的福利性质。分析方法有多种，此处我们使用的是**代表性消费者**（representative consumer）方法，这种方法也许是最简单的。稍后，在讨论一般均衡理论时，我们再使用另外一种不同的而且更为一般的方法。

假设市场需求曲线 $x(p)$ 是由一个代表性消费者的效用最大化产生的，这个消费者的效

用函数为$u(x)+y$。在这个特定市场中我们关注的是x商品。y商品代表“所有其他商品”。y商品最简单的处理方法是，将它认为是消费者在购买x商品最优数量之后剩下的花费在其他商品的钱数。

在第 10 章我们已经知道，这种效用函数的反需求曲线为

$$p=u'(x).$$

直接需求函数$x(p)$是上述函数的反函数，因此它满足一阶条件：

$$u'(x(p))=p.$$

注意这种函数的独特性质：在拟线性的情形下，需求函数和收入无关。这个性质使得均衡分析和福利分析特别简单。

我们已假定了一个代表性的消费者，类似地我们可以假定一个代表性的企业，令它的成本函数为$c(x)$。我们将此解释为：生产x单位的产品需要$c(x)$单位的y产品，并且假设$c(0)=0$。我们还假设$c''(\cdot)>0$，因此只使用一阶条件就可解出代表性企业的利润最大化的产量[一]。

代表性企业的利润最大化的（反）供给函数为$p=c'(x)$。因此，x商品的均衡产量水平就是下列方程的解

$$u'(x)=c'(x). \tag{13.2}$$

在这个产量水平上，消费者对x商品的边际支付意愿恰好等于生产的边际成本。

13.7 福利分析

假设我们不再使用市场机制确定产量，而是根据代表性消费者的效用最大化直接确定产量。这个最大化问题可以表述为

$$\max_{x,y}\ u(x)+y$$

$$\text{使得 } y=w-c(x).$$

其中w为消费者拥有的y商品的初始禀赋。

将约束函数代入目标函数，我们可将这个最大化问题重新写为

$$\max_{x}\ u(x)+w-c(x).$$

一阶条件为

$$u'(x)=c'(x). \tag{13.3}$$

[一] 当然，在字面上说只有一个企业时，谈及竞争行为就非常不合理。最好将这个企业认为是竞争行业中的“平均的”（average）或“代表性的”行为。

根据我们前面对曲率的假设即$c''(\cdot) > 0$可知，二阶条件自动满足。注意（13.2）和（13.3）决定的产量水平是相同的：在这个例子中，竞争市场的结果和效用最大化，恰好导致了相同的产量水平和相同的消费水平。

福利最大化问题就是使得总效用最大：消费者消费x商品的效用加上消费y商品的效用。由于x单位x商品意味着放弃了$c(x)$单位y产品，我们的社会目标函数为$u(x)+w-c(x)$。初始禀赋w只是个常数，因此我们完全可以将我们的目标函数写为$u(x)-c(x)$。

我们已经知道$u(x)$是（反）需求曲线下方直到x的面积。类似地，$c(x)$是边际成本线下方直到x的面积，这是因为

$$c(x)-c(0)=\int_0^x c'(x)dx$$

而且我们假设$c(0)=0.$

因此，选择x使得效用减去成本的差值最大，等价于选择x使得需求曲线以下供给曲线以上的面积最大，如图 13.3 所示。

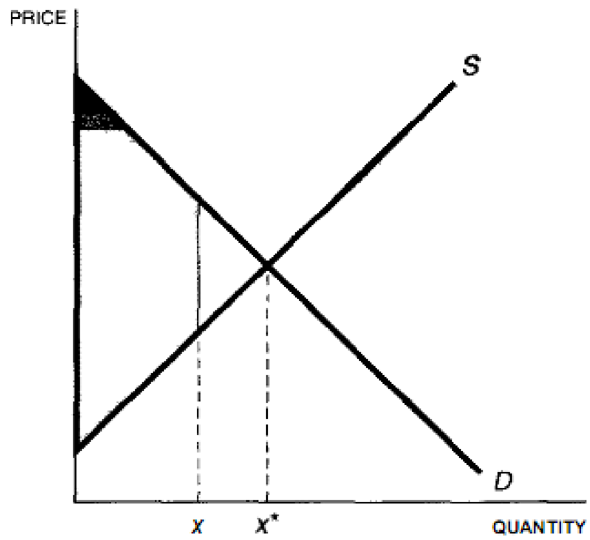


图 13.3：**直接效用**。均衡数量使得需求曲线和供给曲线之间的垂直面积最大。

我们使用例外一种方法重复上述计算。令$CS(x)=u(x)-px$表示与给定产出相伴的**消费者剩余**：它衡量消费x商品获得的“总收益”与在x商品上的支出之差。类似地，令$PS(x)=px-c(x)$表示利润，或代表性企业赚取的**生产者剩余**。

于是**总剩余**最大使得

$$\max_x CS(x)+PS(x)=[u(x)-px]+[px-c(x)],$$

或

$$\max_x u(x) - c(x).$$

因此，我们也可以说竞争均衡的产量水平使得总剩余最大。

13.8 多个消费者

上一节的分析只涉及到一个消费者和一个企业。然而很容易推广到多个消费者和多个企业的情形。假设有 $i = 1,..., n$ 个消费者和 $j = 1,..., m$ 个企业。每个消费者 i 的效用函数都为拟线性的，即 $u_i(x_i) + y_i$，每个企业 j 的成本函数为 $c_j(x_j)$ 。

在这样的情形下，我们将**配置**（allocation）定义为每个消费者消费 x 商品和 y 商品的数量 (x_i, y_i) 以及每个企业生产多少 x 商品 z_j，其中 $i = 1,..., n$； $j = 1,..., m$ 。我们假设每个消费者的**初始禀赋**（initial endowment）为他们最初拥有 y 商品的数量 w_i 以及 x 商品的数量 0，即 $(0, w_i)$ 。

在上述情形下，福利最大化的一种合理备选的配置是在产出可行的情形下使得效用之和最大的配置。效用之和为

$$\sum_{i=1}^{n} u_i(x_i) + \sum_{i=1}^{n} y_i.$$

y 商品的总数等于初始禀赋之和减去生产过程中用掉的数量：

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} w_i - \sum_{j=1}^{m} c_j(z_j).$$

将上式代入目标函数，并且注意到可行性约束意味着 x 产品的总产量必定等于它的总消费量，我们有以下的最大化问题

$$\max_{x_j, z_j} \sum_{i=1}^{n} u_i(x_i) + \sum_{i=1}^{n} w_i - \sum_{j=1}^{m} c_j(z_j)$$

$$\text{使得} \sum_{i=1}^{n} x_i = \sum_{j=1}^{m} z_j.$$

令 λ 表示约束方程的拉格朗日乘子，这个最大化问题的答案必然满足

$$u_i'(x_i^*) = \lambda$$
$$c_j'(z_j^*) = \lambda,$$

以及可行性约束。

但是注意这些条件正好是均衡价格 $p^* = \lambda$ 必须满足的条件。这样的均衡价格使得边际效用等于边际成本并且同时使得需求等于供给。因此，市场均衡必然使得福利最大化，至少在用效用之和衡量福利的情形下是这样的。

当然，这个问题一点也没涉及到总效用的分配问题，因为分配问题取决于初始禀赋(w_i)的配置模式。在拟线性的情形下，均衡价格不取决于财富的分配状况，而且初始配置的任何分配状况都与上述均衡条件是一致的。

13.9 帕累托效率

我们已经看到，竞争均衡使得效用之和最大，至少在拟线性效用的情形下是这样的。但是下列事实却非常不明显：尽管这种效用的局限性很大，效用之和仍然是一个合理的目标函数。

一个更一般的目标是**帕累托效率**（Pareto efficiency）。帕累托有效率的配置是指无法使所有人的状况都变好的配置。换一句话说，帕累托有效率配置是指，在给定其它人效用的前提下，每个人的状况已达到最佳的那种配置。

我们在拟线性效用函数情形下分析帕累托有效率的条件。为简单起见，假设只有两种商品而且它们的数量是固定的，即$(\bar{x},\bar{y})$。另外，假设只有两个人。在这种情形下，帕累托有效率的配置是在维持一个人（比如第 2 人）效用不变的前提下，使得另一个人（比如第 1 人）的效用最大:

$$\max_{x_1,y_1}\ u_1(x_1)+y_1$$

$$\text{使得}\,u_2(\bar{x}-x_1)+\bar{y}-y_1=\bar{u}_2.$$

将约束函数代入目标函数，我们得到了无约束的最大化问题

$$\max_{x_1}\ u_1(x_1)+u_2(\bar{x}-x_1)+\bar{y}-\bar{u}_2,$$

一阶条件为

$$u_1'(x_1)=u_2'(x_2). \tag{13.4}$$

对于任何给定的x_1，上述条件将唯一确定x_2的一个有效率的水平。然而，y_1和y_2的分配是任意的。将y商品在这两个消费者之间分配来分配去，会使一个消费者的状况变好而另外一个消费者的状况变差，但是却一点也不会影响上述效率条件。

最后，思考一下（13.4）式和竞争均衡之间的关系。在均衡价格p^*处，每个消费者调整他对x商品的消费，从而使得

$$u_1'(x_1^*)=u_2'(x_2^*)=p^*.$$

因此，帕累托效率的必要条件得以满足。而且，任何配置若是帕累托有效率的，则必定满足（13.4）式，它在本质上决定了价格p^*，在这个价格水平下，该帕累托有效率配置可以通过竞争均衡实现。

巧合的是，一般情形下在本质上也必须有和上述相同的结果，即使效用函数不是拟线

性的。然而，一般来说，均衡价格取决于商品 y 的分配。我们在讨论一般均衡时，将进一步分析这种决定性。

13.10 效率和福利

我们已经知道，我们求解福利之和最大化问题和求解帕累托有效率问题时，得到的答案是相同的。如果你是初次遇到这个问题，可能你会感到有些惊讶。我们再多分析一下。为简单起见，我们仍假设只有两个消费者和两种商品，当然可以容易地将其推广到多个消费者和多种商品的情形。

假设两种商品 x 和 y 的初始数量分别为 $\bar{x}$ 和 $\bar{y}$ 。一个有效率的配置能做到在一人的效率水平既定的前提下使得另外一人的效用最大：

$$\max_{x_1, y_1} \ u_1(x_1) + y_1 \tag{13.5}$$

$$\text{使得} u_2(\bar{x} - x_1) + \bar{y} - y_1 = \bar{u}_2.$$

两个消费者的效用之和为

$$[u_1(x_1) + y_1] + [u_2(\bar{x} - x_1) + (\bar{y} - y_1)] = u_1(x_1) + u_2(\bar{x} - x_1) + y_1 + \bar{y} - y$$

因此，使得效用之和最大化的配置必然是下列问题的解：

$$\max_{x_1, y_1} \ u_1(x_1) + u_2(\bar{x} - x_1) + y_1 + \bar{y} - y_1 \tag{13.6}$$

我们已经观察到，同一个 x_1^* 既是（13.5）的解又是（13.6）的解。然而，这两个问题的解中的 y 是不同的。任何数对 (y_1, y_2) 都可以使效用之和最大，但是满足（13.5）中约束条件的 y_1 只有一个。（13.5）的解只是（13.6）众多解中的一个。

拟线性效用的特殊结构意味着通过解（13.6）可以找到所有的帕累托有效率的配置：在所有帕累托有效率的配置中，(x_1^*, x_2^*) 是相同的，但是 (y_1^*, y_2^*) 是不同的。这就是效用之和最大化的解和帕累托有效率配置的解是相同的原因[一]。

13.11 离散商品模型

离散商品模型是分析市场行为的另外一个有用的特殊情形。在这种情形下，我们仍假设有两种商品，x 和 y 。但是，x 商品只能以离散数量消费。特别地，我们假设消费者总是要么购买一单位要么购买零单位商品 x 。

因此，一个收入为 m 的消费者购买价格为 p 的商品，可以实现的效用为 $u(1, m-p)$ ；

[一] 对这些结论要附加一下说明：它们要求我们的解 (y_1, y_2) 是内部解。如果消费者 2 的目标效用水平非常低，因此只有令 $y_2 = 0$ 时才能实现他的效用水平，那么上述两个问题的等价性就不再成立。

如果他不购买，则他的效用为$u(0,m)$。这个消费者对该商品的保留价格（reservation price）r是他在买和不买该商品之间恰好无差异时的价格。也就是说，保留价格r满足方程

$$u(1,m-r)=u(0,m).$$

单个消费者的需求曲线的形状如图 13.4A 所示；如果有很多消费者而且他们的保留价格是不同的，那么他们的加总需求曲线是阶梯状得，如图 13.4B 所示。

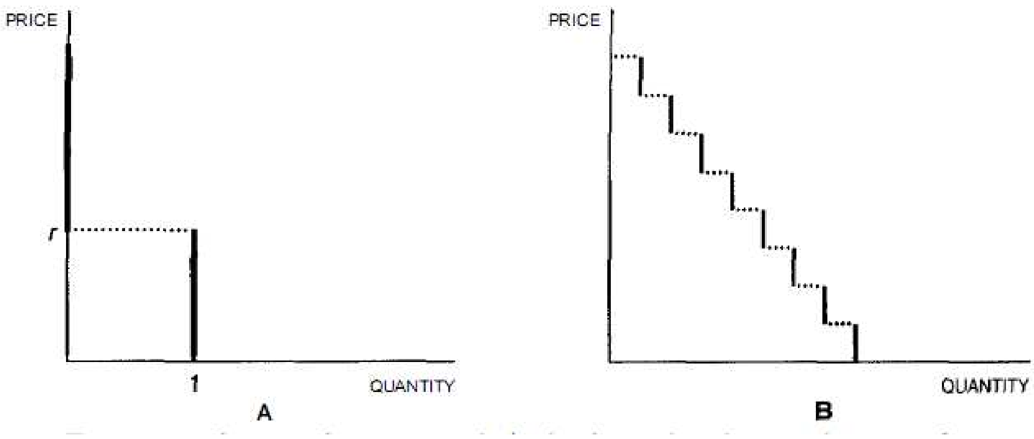


图 13.4：**保留价格**。A 图画出的是单个消费者的需求曲线；B 图画出的是保留价格不同的多个消费者的加总需求曲线。

在离散商品的情形下，如果偏好是拟线性的，则结果非常简单。在这样的情形下，消费者如果购买，则他的效用为$u(1)+m-p$；如果他不买则他的效用为$u(0)+m.$保留价格r是下列方程的解

$$u(1)+m-r=u(0)+m,$$

容易得到$r=u(1)-u(0).$ 使用将$u(0)$标准化为 0 的惯例，即$u(0)=0$，那么$r=u(1)$。由此可知，保留价格等于消费x商品的效用。

如果x商品的价格为p，那么选择消费该商品的消费者得到的效用为

$$u(1)+m-p=r+m-p=m+(r-p)$$

因此，消费者剩余$(r-p)$是一种衡量消费者面对商品价格p所能实现的效用。

这种特别的结构使得均衡分析和福利分析非常简单。市场均衡价格衡量的是**边际消费者**（marginal consumer）的保留价格，边际消费者是指在买和不买该商品之间恰好无差异的消费者。边际消费者得到的消费者剩余（近似）为零；**边际以内的消费者**（inframarginal consumers）通常得到正的消费者剩余。

13.12 税收和补贴

我们已经知道术语比较静态指的是，分析当经济环境变动时经济结果如何变动。在竞争市场的情形下，我们通常关注当某些政策变量变动时，均衡价格和（或）均衡数量如何变动。税收和补贴就是这方面的例子。

在征税的情形下，最重要的事情是要记住此时价格有两种：**需求价格**（demand price）和**供给价格**（supply price）。需求价格 p_d 是某商品需求者支付的价格；供给价格 p_s 是该商品供给者得到的价格；它们之间的差额就是税收或补贴。

例如，数量税（quantity tax）是按某商品的消费数量征收的税。这意味着需求者支付的价格大于供给者得到的价格，二者之差等于数量税：

$$p_d = p_s + t.$$

价值税（value tax）是按商品支出额征收的税。通常用百分数表示，例如 10%的销售税。税率为 τ 的价值税将会导致

$$p_d = (1+\tau) p_s.$$

补贴具有类似的结构；金额为 s 元的数量补贴（quantity subsidy）表示对于每单位商品来说，卖方得到的价格比买方支付的价格都多 s 元，即 $p_d = p_s - s.$

需求者的行为取决于他面对的价格；供给者的行为取决于他面对的价格。因此我们分别用 $D(p_d)$ 和 $S(p_s)$ 表示他们各自的行为。均衡条件是需求等于供给，这样就得到了下面两个式子:

$$D(p_d) = S(p_s)$$

$$p_d = p_s + t.$$

将第二个式子代入第一个式子，我们可以求解

$$D(p_s+t)=S(p_s),$$

或

$$D(p_d)=S(p_d-t).$$

显然，求解 p_d 和 p_s 与使用上面哪一个式子无关。也就是说，使用上面两个式子中的任何一个得到的解必定是相同的。

求解此类税收问题的另外一种方法是使用反需求函数和反供给函数。在这种情形下，方程变为

$$P_d(q)=P_s(q)+t,$$

或

$$P_s(q)=P_d(q)-t.$$

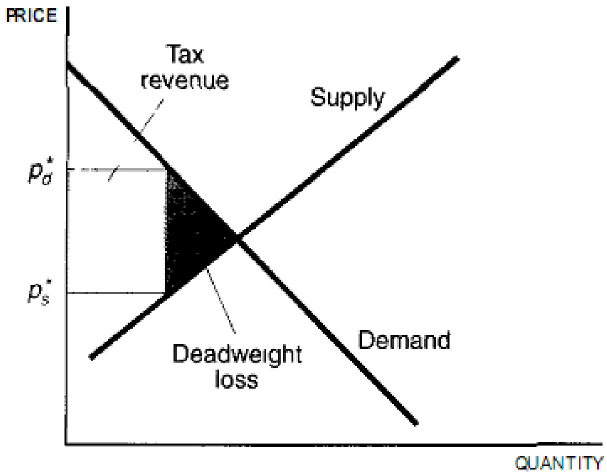


图 13.5：**征税造成的净损失**。矩形区域表示的是税收总收入。黑色三角形区域表示的是征税造成的净损失。

一旦我们解出均衡价格和均衡数量，进行福利分析就非常简单。消费者消费均衡数量 x^* 得到的效用为 $u(x^*)-p_d x^*$。生产者得到的利润为 $p_s x^*-c(x^*)$。最后，政府得到的税收总收入为 $tx^*=(p_d-p_s)x^*$. 最简单的情形是企业的利润和税收收入都为代表性消费者得到，由此产生的净福利为

$$W(x^*)=u(x^*)-c(x^*).$$

这正是需求曲线下方的面积减去边际成本曲线下方的面积，如图 13.5 所示。征税后得到的剩余与原来均衡时实现的福利之间的差值，称为**净损失**（deadweight loss）；如图 13.5 中的

三角形阴影区域所示。净损失衡量因产量损失而给消费者带来的价值损失。

注释

本章的内容是单个市场的标准的新古典主义分析。本章采用的形式可能最早见于 Marshall(1920).

习题

13.1 令 $v(p)+m$ 表示代表性消费者的间接效用，令 $\pi(p)$ 表示代表性企业的利润。令福利是价格的函数，即用 $v(p)+\pi(p)$ 表示福利函数。证明竞争价格使得该函数值最小。你能解释一下为什么均衡价格使得上述福利衡量方式的值最小而不是最大吗？

13.2 证明当价格由 p_0 变为 p_1 导致的利润变动可用供给函数的积分（积分下限和上限分别为 p_0 和 p_1）表示。

13.3 某行业由很多企业组成，每个企业的成本函数都具有下列形式

$$c(w_1,w_2,y)=(y^2+1)w_1+(y^2+2)w_2.$$

（a）求单个企业的平均成本曲线并描述该曲线如何随着要素价格 w_1/w_2 的变动而变动。

（b）求单个企业的短期供给曲线。

（c）求长期行业供给曲线。

（d）写出单个企业的必要投入集。

13.4 农民用土地和劳动生产玉米。生产 y 单位玉米所需劳动的成本为 $c(y)=y^2$。玉米行业内有 100 家企业且为完全竞争的。

（a）求单个农民的玉米供给曲线。

（b）求玉米的市场供给。

（c）假设玉米的需求曲线为 $D(p)=200-50p$。求均衡价格和均衡数量。

（d）求土地的均衡租金。

13.5 考虑关于美国和英国雨伞贸易的一个模型。英国代表性企业根据生产函数 $f(K,L)$ 生产出口雨伞，其中 K 和 L 分别代表资本和劳动。令 r 和 w 分别表示英国国内资本和劳动的价格，令与生产函数 $f(K,L)$ 相伴的成本函数为 $c(w,r,y)$。假设雨伞的初始均衡价格和均衡数量分别为 p^* 和 y^*。为简单起见，假设英国生产的雨伞全部用于出口，美国不生产雨伞，而且假设所有市场都是竞争性的。

（a）英国决定对生产和出口的雨伞进行补贴，每把出口雨伞补贴 s 元，因此英国企业每把雨伞的销售收入为 $p+s$。为了完全抵消这种补贴，也即是说，为了使英国生产和出口的雨伞数量稳定在 y^*，美国应该对进口的每把雨伞征收多大的进口税 $t(s)$？（提示：这个问题比较简单，别想得太复杂了。）

（b）由于美国轻易就能将英国的出口补贴抵消，英国决定使用资本补贴。特别地，英国决定对企业购买资本进行补贴，每单位资本的补贴为 s 元，因此英国雨伞生产企业购买资本的价格相当于 $r-s$。美国决定实施进口税进行报复，为了使英国雨伞的生产数量稳定在 y^*，

美国决定对每把进口雨伞征收 $t(s)$ 元。那么消费者支付的价格 p 、税收 $t(s)$ 和成本函数 $c(w,r,y)$ 之间必定有何种关系？

（c）计算 $t'(s)$ 的表达式，该表达式要涉及到资本的条件要素需求函数 $K(w,r,y)$ 。

（d）假设生产函数是规模报酬不变的。在这种情形下，你能对 $t'(s)$ 的表达式进行怎样简化处理？

（e）假设资本在雨伞的生产中是一种劣等要素（inferior factor）。那么抵消英国资本补贴的美国关税 $t(s)$ 有什么独特之处吗？

13.6 在一个热带小岛上有 100 个造船人，编号从 1 到 100。每个造船人每年最多可以造船 12 艘，而且在给定市场价格的情形下，每个造船人都追求利润最大化。令 y 表示某个特定造船人每年的造船数量，假设造船人 1 的成本函数为 $c(y)=11+y$ ，造船人 2 的成本函数为 $c(y)=11+2y$ ，以此类推。也就是说，造船人 i 的成本函数为 $c(y)=11+iy$ ，其中 $i=1,...,100.$ 假设固定成本 11 元是准固定成本；也就是说，造船人只有在产量为正的情形下才支付这些固定成本。如果船得价格为 25，有多少造船人会选择生产正的产量？如果船的价格为 25，每年一共造多少艘船？

13.7 考虑具有下列结构的某个行业。行业内有 50 个企业的行为是竞争性的，它们的成本函数相同，均为 $c(y)=y^2/2$ 。行业内还有一个垄断企业，它的边际成本为 0。该产品的需求曲线为 $D(p)=1000-50p.$

（a）垄断企业的利润最大化产量是多少？

（b）垄断企业的利润最大化价格是多少？

（c）竞争性部门在上述价格下的供给为多少？

13.8 美国消费者对雨伞的需求函数为 $D(p)=90-p$ 。雨伞由美国企业和英国企业供给。为简单起见，假设每个国家都有一个代表性企业，它们的行为都是竞争性的。每个国家生产雨伞的成本都为 $c(y)=y^2/2$ 。

（a）雨伞的加总供给函数是什么？

（b）求均衡价格和均衡数量。

（c）现在美国国内企业游说国会同意对进口雨伞征税收 3 元/把的关税。美国消费者对雨伞支付的新价格是多少？

（d）有多少雨伞是进口的（即由英国提供的），有多少雨伞是美国国内生产的？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 14 章：垄断

曹乾　译

(东南大学　caoqianseu@163.com)

14 垄断

垄断（monopoly）这个词最初是指专卖权。后来它用于描述某个企业或若干企业组成的小集团，这样的企业或企业集团在某个既定的市场上对某商品拥有排外的控制权。这个定义的模糊之处在于“既定的市场”指得是什么意思。软饮料市场上的企业很多，但在可乐市场上只有少数几家企业。

从经济分析的角度看，垄断企业的关键特征是它拥有市场力量（market power），因为它能销售的产量是它索要价格的连续函数。但在竞争企业的情形下，如果企业索要的价格高于当前的市场价格，它的销量就会锐减到零。竞争性企业是**价格接受者**；垄断企业是**价格制定者**。

垄断企业在选择价格和产量水平时受到两类约束。首先，它面临我们以前阐述过的标准技术约束——技术可行的投入和产出方案是既定的。我们将会发现用成本函数 $c(y)$ 描述技术约束很方便。（注意在成本函数中我们省略了要素价格这类变量，因为我们假定要素价格都是不变的。）

垄断企业面临的第二类约束是由消费者的行为施加的。消费者在不同的商品价格水平下愿意购买不同的数量，我们使用需求函数 $D(p)$ 表示这种关系。

垄断企业的利润最大化问题可以写为

$$\max_{p,y} \ py - c(y)$$

$$\text{使得} D(p) \geq y$$

在我们感兴趣的大多数情形下，垄断企业生产的产量等于消费者的需求，因此约束条件可以写成等式 $y = D(p)$ 。将 y 代入目标函数，可得下列的无约束的最大化问题

$$\max_{p} \ pD(p) - c(D(p)).$$

尽管这可能是提出垄断企业利润最大化问题的最自然方法，然而在绝大多数情形下使用反需求函数比使用直接需求函数更方便。

令 $p(y)$ 表示反需求函数——为了销售出 y 单位产品必须索要的价格。那么垄断企业销售 y 单位产品，它期望得到的销售收入为 $r(y) = p(y)y$ 。我们可以这样提出垄断企业的利润最大化问题：

$$\max_{y} \ p(y)y - c(y).$$

这个问题的一阶条件和二阶条件分别为

$$p(y) + p'(y)y = c'(y) \qquad (14.1)$$

$$2p'(y) + p''(y)y - c''(y) \le 0. \tag{14.2}$$

一阶条件是说产量的利润最大化选择是边际收入必定等于边际成本。我们进一步考虑这个条件。当垄断企业考虑是否多销售 dy 单位产品时，它必须考虑两方面的效应。首先，若多销售 dy 单位产品，则它的销售收入增加了 pdy，因为它是按当前价格销售这些额外的产品的；但是另外一方面，为了将这些额外的产品销售出去，它必须将价格降低 $dp = \frac{dp}{dy}dy$ 那么多，降低后的价格适用于所有销售的产品而不仅限于额外的产品。销售额外产品导致的额外收入因此为

$$pdy + dpy = [p + \frac{dp}{dy}y]dy,$$

这个额外收入必须与边际成本进行比较权衡。

二阶条件要求边际收入的导数必定小于边际成本的导数；也就是说，边际收入曲线从上方穿过边际成本曲线。

一阶条件变形后可得如下形式

$$r'(y) = p(y)[1 + \frac{dp}{dy}\frac{y}{p}] = c'(y),$$

$$p(y)[1 + \frac{1}{\varepsilon(y)}] = c'(y), \tag{14.3}$$

其中 $\varepsilon(y) = \frac{dy}{dp}\frac{p}{y}$ 是垄断企业面对的需求（价格）弹性。注意，只要消费者的需求曲线斜率为负（这当然是常规情形），那么需求价格弹性是负的。

从一阶条件可知，在最优产量水平上，需求价格弹性的绝对值必定大于 1。如果不是如此，边际收入将为负，因此不可能等于非负的边际成本。

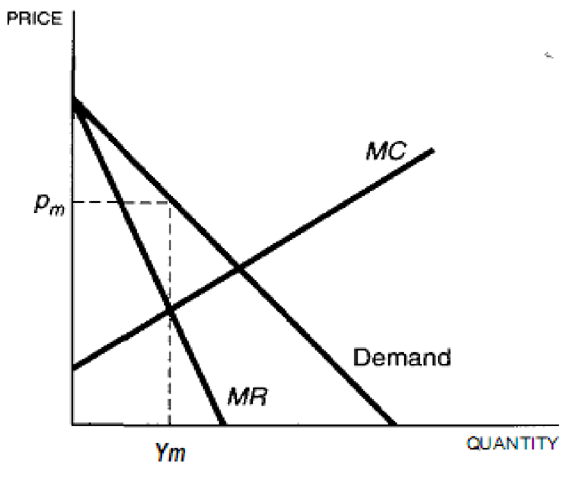


图 14.1：**垄断产量的确定**。垄断企业在边际收入等于边际成本之处进行生产。

垄断企业的最优产量可用图 14.1 表示。边际收入曲线为 $r'(y) = p(y) + p'(y)y$ 。由于根据假设可知 $p'(y) < 0$ ，边际收入曲线必定位于反需求曲线的下方。

当 $y = 0$ 时，销售额外一单位产品带来的边际收入恰好就是 $p(0)$ 。然而，当 $y > 0$ 时，销售额外一单位产品带来的边际收入必定小于产品价格，这是因为增加销量的唯一方法是降低价格，这种价格的降低会影响边际以内的所有销售数量的收入。

垄断企业的最优产量必定位于边际收入曲线与边际成本曲线相交之处。为了满足二阶条件，边际收入曲线必定从上方穿过边际成本曲线。我们通常假定利润最大化的产量水平是唯一的。给定产量水平，比如 y^* ，企业索要的价格将为 $p(y^*)$.

14.1 特殊情形

垄断行为有两种特殊的情形值得一提。第一种情形是线性需求。如果反需求曲线为 $p(y) = a - by,$ 那么收入函数为 $r(y) = p(y)y = ay - by^2$ ，边际收入为 $r'(y) = a - 2by$ 。因此，边际收入曲线的陡峭程度是需求曲线陡峭程度的 2 倍。如果企业的成本函数为 $c(y) = cy$ ，则其边际成本是常数。这样以来，我们就可以令边际收入等于边际成本，从而直接求解出垄断价格和垄断产量：

$$y^* = \frac{a-c}{2b}$$

$$p^* = \frac{a+c}{2}.$$

另外一种特殊情形是需求价格弹性为常数的需求函数 $y = Ap^{-b}$ 。在前面我们已经知道，这个需求函数的需求价格弹性 $\varepsilon(y) = -b$ 。在这种情形下，我们可以运用（14.3）式，可得

$$p(y) = \frac{c}{1-1/b}.$$

因此，对于需求价格弹性为常数的需求函数来说，价格等于边际成本的某个固定加成（markup），加成数量取决于需求价格弹性。

14.2 比较静态

我们经常对垄断企业的产量和价格如何随成本的变动而变动感兴趣。为简单起见，假设边际成本为常数。那么利润最大化问题为

$$\max_y \; p(y)y - cy$$

它的一阶条件为

$$p(y) + p'(y)y - c = 0.$$

我们从标准的比较静态计算可知，dy/dc 的符号与一阶条件关于 c 的导数的符号相同。容易看出，这个符号为负，因此我们可以断言：当边际成本增加时，追求利润最大化的企业总是会减少产量。

我们更关注成本变化对价格的效应。使用链式法则可知

$$\frac{dp}{dc}=\frac{dp}{dy}\frac{dy}{dc},$$

从上述表达式可知 $dp/dc>0$，但是我们还想知道 dp/dc 的大小。

标准的静态比较计算告诉我们

$$\frac{dy}{dc}=-\frac{\partial^2\pi/\partial y\partial c}{\partial^2\pi/\partial y^2}.$$

对利润函数求二阶导数可得

$$\frac{dy}{dc}=\frac{1}{2p'(y)+yp''(y)}.$$

由此可知，

$$\frac{dp}{dc}=\frac{p'(y)}{2p'(y)+yp''(y)}.$$

这个式子也可以写成

$$\frac{dp}{dc}=\frac{1}{2+yp''(y)/p'(y)}.$$

使用这个表达式，可以容易看出上述两种特殊情形意味着什么。如果需求是线性的，那么 $p''(y)=0$ 而且 $dp/dc=1/2.$ 如果需求函数的价格弹性是常数 ε，则 $dp/dc=\varepsilon/(1+\varepsilon).$ 在线性需求曲线的情形下，成本增加之后有一半的成本转嫁到价格之上。在需求函数的价格弹性为常数的情形，价格上升数额大于成本上升数额——需求越是缺乏弹性，转嫁的成本越多。

14.3 福利和产量

我们在第 13 章已经知道，在一定条件下，使得价格等于边际成本的产量水平是帕累托有效率的。由于边际收入曲线总是位于反需求曲线的下方，显然垄断企业生产的产量必定小于该帕累托有效率的产量。在本节我们将稍微详细地分析垄断造成的无效率问题。

为简单起见，假设某经济体内只有一个消费者，该消费者的效用函数为拟线性的：$u(x)+y$。我们在第 13 章已经知道，这种效用函数的反需求函数为 $p(x)=u'(x).$ 令 $c(x)$ 表示生产 x 单位的 x 产品所需要的 y 产品的最小数量。则一个合理的社会目标是选择 x 使得效用最大：

$$W(x) = u(x) - c(x).$$

这意味着社会最优的产量水平 x_o 由下式给出

$$u'(x_o) = p(x_o) = c'(x_o).$$

另外一方面，垄断产量水平满足条件

$$p(x_m) + p'(x_m)x_m = c'(x_m).$$

因此，福利函数的导数在垄断产量处的值为

$$W'(x_m) = u'(x_m) - c'(x_m) = -p'(x_m)x_m = -u''(x_m)x_m > 0.$$

从 $u(x)$ 为凹函数可知，产量增加会使效用增加。

我们可以使用另外一种稍微不同的方法得到上述同样的结论。我们也可以将社会目标函数写成消费者剩余与利润的和：

$$W(x) = [u(x) - p(x)x] + [p(x)x - c(x)].$$

在垄断产量之处，利润关于 x 的导数等于零，这是因为垄断企业选择能使利润最大的产量水平。消费者剩余在 x_m 之处的导数为

$$u'(x_m) - p(x_m) - p'(x_m)x_m = -p'(x_m)x_m,$$

上式当然是正的。

14.4 质量选择

垄断企业不仅选择产量水平，还选择产品的其它维度。例如，产品质量。假设我们可以将产品质量用某个数值水平 q 表示。我们假设效用和成本都取决于质量，我们采用的社会目标函数为

$$W(x,q) = u(x,q) - c(x,q).$$

（和往常一样，为简单起见，我们假设效用为拟线性的。）我们假设质量是人们喜欢的东西，因此 $\partial u / \partial q > 0$，生产这样的东西要花费成本，因此 $\partial c / \partial q > 0$。

垄断企业的利润最大化问题为：

$$\max_{x,q} \; p(x,q)x - c(x,q).$$

这个问题的一阶条件为

$$p(x_m,q_m) + \frac{\partial p(x_m,q_m)}{\partial x} x_m = \frac{\partial c(x_m,q_m)}{\partial x}$$

$$\frac{\partial p(x_m,q_m)}{\partial q}x_m=\frac{\partial c(x_m,q_m)}{\partial q}$$

下面我们计算福利函数在(x_m,q_m)之处的导数值。我们有

$$\frac{\partial W(x_m,q_m)}{\partial x}=\frac{\partial u(x_m,q_m)}{\partial x}-\frac{\partial c(x_m,q_m)}{\partial x}$$
$$\frac{\partial W(x_m,q_m)}{\partial q}=\frac{\partial u(x_m,q_m)}{\partial q}-\frac{\partial c(x_m,q_m)}{\partial q}$$

将一阶条件代入上面两个式子可得

$$\frac{\partial W(x_m,q_m)}{\partial x}=-\frac{\partial p(x_m,q_m)}{\partial x}x_m>0 \qquad (14.4)$$

$$\frac{\partial W(x_m,q_m)}{\partial q}=\frac{\partial u(x_m,q_m)}{\partial q}-\frac{\partial p(x_m,q_m)}{\partial q}x_m \qquad (14.5)$$

第一个式子表明，维持质量不变，垄断企业生产的产品数量与社会最优水平相比显得太少。第二个式子的解释不是那么容易。由于$\partial p/\partial q$等于生产更多质量的边际成本（注意可把质量理解为一种产品），它必定为正，因此福利关于质量的导数等于两个正数之差（见 14.5 式），因此它的符号为正还是负在表面上不容易看出。

问题在于我们能否找到需求行为的某个可行条件，使得我们可以容易地确定上面这个表达式的符号？如果我们将社会目标函数写为消费者剩余加上利润，而不是写成效用减去成本，那么答案就非常容易看出。社会目标函数的形式为

$$W(x,q)=[u(x,q)-p(x,q)x]+[p(x,q)x-c(x,q)]$$

=消费者剩余+利润

现在对上式关于x和q求导，并且求其在垄断产量之处的导数值。由于垄断企业是追求利润最大化的，垄断利润关于产量和质量的导数必定等于零，这表明福利关于产量和质量的导数，正好是消费者剩余关于产量和质量的导数。

消费者剩余关于产量的导数总是正的，这是垄断企业产量太少的另一种说法。消费者剩余关于质量的导数的符号是不明确的，它可以为正也可以为负。它的符号取决于$\partial^2 p(x,q)/\partial x\partial q$.

为了看清这一点，请看图 14.2。当质量增加时，需求曲线向上移动而且（可能）变得平缓或者陡峭。将这个移动分解为平移和转动，如图所示。消费者剩余不受需求曲线平移的影响，因此总变化取决于反需求函数是变得更平坦还是更陡峭。如果反需求曲线变得更平缓，则消费者的剩余下降，反之则反是。

解释（14.5）式的另外一种方法是基于保留价格模型的考虑。将$p(x,q)$认为是消费x的保留价格，因此$u(x,q)$正好是保留价格之和。在这种解释方法下，$u(x,q)/x$是平均支付意愿，$p(x,q)$是边际支付意愿。我们可以将（14.5）重新写为

$$\frac{1}{x_m}\frac{\partial W(x_m,q_m)}{\partial q}=\frac{\partial}{\partial q}[\frac{u(x_m,q_m)}{x_m}-p(x_m,q_m)].$$

现在我们可以看到福利关于q的导数和如下数量成比例，即与质量变动的平均支付意愿的导数与质量变动的边际支付意愿的导数之差成比例。

社会福利取决于消费者的效用之和或支付意愿之和；但是垄断企业只关注边际消费者的支付意愿。如果上述两个值不相等，垄断企业的质量选择从社会角度看不是最优的。

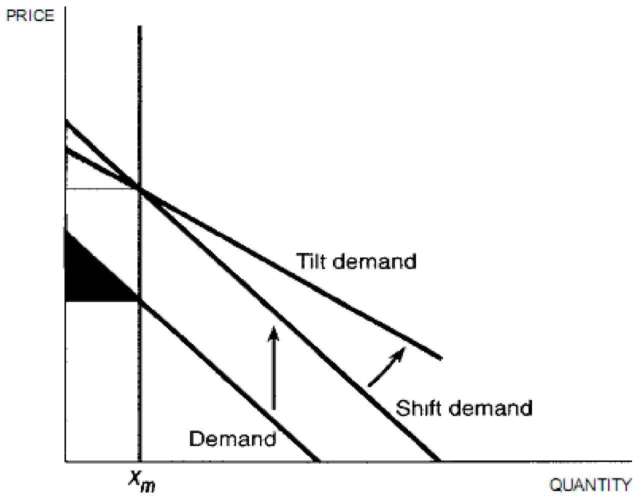


图 14.2：**质量变动对消费者剩余的影响**。当需求曲线向上移动并倾斜时，质量变动对消费者剩余的影响仅取决于需求曲线倾斜的方向（即需求曲线更平缓还是更陡峭）。

14.5 价格歧视

粗略地讲，价格歧视就是向消费者销售某种商品时，不同数量段制定不同的价格，这里的消费者可以是同一个消费者群体也可以是不同的消费者群体。垄断行为自然地涉及到价格歧视问题，这是因为垄断企业如果能找到某种方法多销售一些产品，而且这种方法不会降低当前售价时，它肯定就会多销售一些。

企业为了使价格歧视策略可行，它必须具有区分消费者的能力以及阻止消费者转售的能力。阻止转售通常问题不大，价格歧视最困难的部分在于区分消费者。最容易的情形是企业可根据消费者的外在特征比如年龄进行区分。当企业必须在某些外在特征（例如消费者购买的数量或购买的时间段）进行价格歧视时，有必要进行更复杂的分析。在这样的情形下，垄断企业必须制定不同价格从而使得消费者根据自己所在的类型进行“自我选择”。

价格歧视的传统分类要归功于庇古（Pigou,1920）。

第一级价格歧视（First-degree price discrimination），是卖方对每单位商品都索要不同

的价格，使得每单位商品的价格等于消费者对该单位商品的最大支付意愿。这种价格歧视也称为**完全价格歧视**（perfect price discrimination）。

第二级价格歧视（Second-degree price discrimination），是指价格根据购买的商品数量不同而不同，但不是对人的歧视——同一人购买不同的数量也要支付不同的价格。这种价格歧视也称为**非线性定价**（nonlinear pricing）。每个消费者面对的是相同的价目表，但是这个价目表中规定购买的数量不同价格也不同。数量折扣或奖励是最常见的例子。

第三级价格歧视（Third-degree price discrimination），是指不同的消费者群体支付的价格不同，但是同一个消费者群体中的消费者对每单位商品支付的价格相同。这也许是最常见的价格歧视的情形；常见的例子有对学生打折，或者一周内按照日期实施不同的价格。

我们用一个非常简单的模型分析这些价格歧视的情形。假设有两个潜在的消费者，他们的效用函数为 $u_i(x)+y$，其中 $i=1,2$。为简单起见，将效用标准化，因此 $u_i(0)=0$。消费者 i 对消费水平 x 的支付意愿为 $r_i(x)$。它是下列方程的解

$$u_i(0)+y=u_i(x)-r_i(x)+y.$$

上式的左侧是消费零单位商品 x 的效用；右侧是支付 $r_i(x)$ 消费 x 单位商品的效用。由 $u_i(0)=0$ 可知， $r_i(x)\equiv u_i(x).$

与效用函数相伴的另外一个有用函数是边际支付意愿函数，也就是（反）需求函数。这个函数衡量的是为了诱使消费者需求 x 单位商品，每单位商品的价格应该为多大。如果消费者在面对的每单位商品的价格 p 的情形下选择最优消费量，他必须求解下列效用最大化问题：

$$\max_{x,y} u_i(x)+y$$

$$\text{使得 } px+y=m.$$

我们已经知道这个问题的一阶条件为

$$p=u_i'(x). \tag{14.6}$$

因此，（14.6）式给出了反需求函数：诱使消费者 i 选择消费 x 单位商品的价格为 $p=p_i(x)=u_i'(x).$

我们假设消费者 2 对商品的最大支付意愿，总是大于消费者 1 对商品的最大支付意愿，也就是

$$u_2(x)>u_1(x)\text{ 对于所有 } x>0 \text{ 都成立} \tag{14.7}$$

我们还假设消费者 2 的边际支付意愿，总是大于消费者 1 的边际支付意愿，也就是

$$u_2'(x)>u_1'(x)\text{ 对于所有 } x>0 \text{ 都成立} \tag{14.8}$$

因此自然可以将消费者 2 称为**高需求**（high demand）的消费者，将消费者 1 称为**低需求**（low demand）的消费者。

我们假设商品的卖方只有一个，它的边际成本为常数 c，因此该垄断企业的成本函数为 $c(x)=cx.$

14.6 第一级价格歧视

我们暂时假设只有一个消费者，因此可以将区分消费者的下标 i 去掉。垄断企业希望提供给消费者能使它自己利润最大的价格和产量组合 (r^*, x^*)。价格 r^* 是要么接受要么走人的价格（take-it-or-leave-it price），也就是说消费者可以支付 r^* 消费 x^* 单位商品，或者消费零单位商品。

垄断企业的利润最大化问题为

$$\max_{r,x} r-cx$$

$$\text{使得} u(x)\geq r.$$

约束条件表明消费者从消费 x 商品中获得非负的剩余，否则他不会消费。由于垄断企业希望 r 越高越好，因此约束条件变成了等式。

将等式约束 $u(x)=r$ 代入目标函数，可得到无约束的最大化问题

$$\max_{x} u(x)-cx$$

一阶条件决定了最优产量水平为

$$u'(x^*)=c. \tag{14.9}$$

给定这个产量水平，则消费者的要么接受要么走人的价格为

$$r^*=u'(x^*).$$

这个解有几点值得注意。首先，垄断企业会选择帕累托有效率的产量水平——在该产量水平上边际支付意愿等于边际成本。然而，垄断企业也会设法攫取该产量水平的所有利益——它的利润实现了最大而消费者正好在消费和不消费该商品之间无差异。

第二，这个市场中的垄断企业生产的产量和竞争行业生产的产量是相同的。竞争行业会在价格等于边际成本而且供给等于需求之处进行生产。这两个条件意味着 $p(x)=c$，这正好是（14.9）式外加（14.6）式中反需求函数的定义。当然，完全竞争情形下，对交易好处的分配与完全价格歧视的垄断企业情形对交易好处的分配是不同的。在完全竞争情形下，消费者得到的效用为 $u(x^*)-cx^*$，企业的利润为零。

第三，如果垄断企业将每单位产品按不同的价格卖给消费者，则也可以实现与上述相

同的结果。例如，假设企业将产量 x 分解为 n 单位，每单位的数量为 Δx ，即 $x = n\Delta x$ 。那么消费者对第一单位产品的支付意愿为

$$u(0) + m = u(\Delta x) + m - p_1,$$

或

$$u(0) = u(\Delta x) - p_1.$$

类似地，消费者对第二单位产品的边际支付意愿为

$$u(\Delta x) = u(2\Delta x) - p_2.$$

依次类推，直到找到消费者对第 n 单位产品的边际支付意愿，这样我们就得到了下列一系列等式，

$$\begin{gathered} u(0) = u(\Delta x) - p_1 \\ u(\Delta x) = u(2\Delta x) - p_2 \\ \cdots \\ u((n-1)\Delta x) = u(x) - p_n \end{gathered}$$

将这 n 个式子相加并且使用标准化的效用 $u(0) = 0$ ，可得 $\sum_{i=1}^{n} p_n = u(x)$ 。也就是说边际支付意愿之和必定等于总支付意愿。因此，垄断企业采用下列哪种价格歧视方法，结果是一样的：一是制定的价格让消费者要么接受要么走人；二是对每单位商品都按照消费者最高的边际支付意愿进行销售。

14.7 第二级价格歧视

第二级价格歧视有时称为非线性定价。例如数量折扣，在这种情形下，企业的销售收入是销量的非线性函数。在这一节我们分析第二级价格歧视的一个简单问题。

回顾前文引入的符号。有两个消费者，他们的效用函数分别为 $u_1(x_1) + y_1$ 和 $u_2(x_2) + y_2$ ，其中我们假设 $u_2(x) > u_1(x)$ 而且 $u_2'(x) > u_1'(x)$ 。我们将消费者 2 称为高需求的消费者，将消费者 1 称为低需求的消费者。总支付意愿较大意味着边际支付意愿也较大的这个假设有时称为**单次相交的性质**（single crossing property），因为这个假设意味着这两个人的任何两条无差异曲线最多相交一次。

假设垄断企业选择的是某个（非线性）函数 $p(x)$ ，这个函数表明如果消费者需求 x 单位商品它应该索要的价格为多少。假设消费者 i 需求 x_i 单位，该消费者支付的钱数为 $r_i = p(x_i)x_i$ 元。无论是从消费者还是从垄断企业的观点来看，最关键的事情是消费者支付了 r_i 元购买了 x_i 单位的商品。因此，函数 $p(x)$ 的选择问题化简为 (r_i, x_i) 的选择问题。消费者 1 选择 (r_1, x_1) ，消费者 2 选择 (r_2, x_2) 。

垄断企业面临的约束是这样的。首先，每个消费者必定希望消费 x_i 单位的商品而且愿

意支付价格 r_i：

$$u_1(x_1)-r_1\ge 0$$
$$u_2(x_2)-r_2\ge 0.$$

上面两个式子是说，每个消费者消费 x 商品的状况至少要和不消费一样好。其次，每个消费者必须偏好他自己的消费胜于别人的消费，即：

$$u_1(x_1)-r_1\ge u_1(x_2)-r_2$$
$$u_2(x_2)-r_2\ge u_2(x_1)-r_1.$$

这样的条件称为**自我选择的约束**（self-selection constraints）。如果方案 (x_1,x_2) 可行，即消费者会自动地量体裁衣地进行选择，那么与消费其他人的商品束相比，每个消费者必定偏好消费为他量身定做的商品束。

将上一段落中的不等式整理可得

$$r_1\le u_1(x_1) \quad （14.10）$$

$$r_1\le u_1(x_1)-u_1(x_2)+r_2 \quad （14.11）$$

$$r_2\le u_2(x_2) \quad （14.12）$$

$$r_2\le u_2(x_2)-u_2(x_1)+r_1. \quad （14.13）$$

当然，垄断企业希望选择尽可能大的 r_1 和 r_2。由此可知一般来说前两个不等式中有一个是具有约束力的（binding），后面两个不等式中有一个是有约束力的。可以证明我们前面的假设 $u_2(x)>u_1(x)$ 和 $u_2'(x)>u_1'(x)$ 足以保证我们确定哪一个不等式是具有约束力的，下面我们将证明这一点。

不妨假设(14.12)式是具有约束力的。那么（14.13）式意味着

$$r_2\le r_2-u_2(x_1)+r_1,$$

或

$$u_2(x_1)\le r_1.$$

使用（14.7）式，我们可以写为

$$u_1(x_1)<u_2(x_1)\le r_1,$$

这与（14.10）式矛盾。由此可知，（14.12）是不具有约束力的，（14.13）是具有约束力的。我们指出一个事实，这个事实将来我们会用到：

$$r_2=u_2(x_2)-u_2(x_1)+r_1. \quad （14.14）$$

现在考虑（14.10）式和（14.11）式。如果（14.11）是具有约束力的，我们有

$$r_1 = u_1(x_1) - u_1(x_2) + r_2.$$

将（14.4）式代入可得

$$r_1 = u_1(x_1) - u_1(x_2) + u_2(x_2) - u_2(x_1) + r_1,$$

这意味着

$$u_2(x_2) - u_2(x_1) = u_1(x_2) - u_1(x_1).$$

我们可以将上式重新写为

$$\int_{x_1}^{x_2} u_1'(t)dt = \int_{x_1}^{x_2} u_2'(t)dt.$$

然而，这违背了 $u_2'(x) > u_1'(x)$ 这个假设。由此可知，（14.11）是不具有约束力的，（14.10）是具有约束力的，因此

$$r_1 = u_1(x_1). \tag{14.15}$$

（14.4）式和（14.15）式是说企业会向低需求消费者索要他们的最高支付意愿，企业会向高需求消费者索要能诱使他们消费 x_2 而不是 x_1 的最高价格。

垄断企业的利润函数为

$$\pi = [r_1 - cx_1] + [r_2 - cx_2],$$

将 r_1 和 r_2 代入上式可得

$$\pi = [u_1(x_1) - cx_1] + [u_2(x_2) - u_2(x_1) + u_1(x_1) - cx_2].$$

因此，我们得到的最大化问题为

$$\max_{x_1, x_2} [u_1(x_1) - cx_1] + [u_2(x_2) - u_2(x_1) + u_1(x_1) - cx_2]$$

它的一阶条件为

$$u_1'(x_1) - c + u_1'(x_1) - u_2'(x_1) = 0 \tag{14.16}$$

$$u_2'(x_2) - c = 0. \tag{14.17}$$

整理（14.6）式可得

$$u_1'(x_1) = c + [u_2'(x_1) - u_1'(x_1)] > c, \tag{14.18}$$

这意味着低需求消费者对商品的边际评价大于商品的边际成本。因此他购买的数量是无效率的较小数量。（14.17）式是说在最优的非线性价格上，高需求消费者的边际支付意愿等于边际成本。因此高需求消费者消费的数量从社会角度看是最优数量。

注意如果一次相交的性质得不到满足，那么（14.18）式括号内的项将是负的，这意味着低需求消费者消费的数量将大于在有效率点上的消费量。这种情形可能发生，但它的确比

较特殊。

高需求消费者支付边际成本是一种普遍的现象。如果高需求消费者支付的价格大于边际成本，垄断企业会向最大的消费者稍微降低一些价格，从而诱使他购买更多数量。由于稍微降价后，价格仍然大于边际成本，垄断企业在这些额外数量的销售上仍有利润可赚。而且，这样的策略不会影响垄断企业从其他消费者身上得到的利润，因为垄断企业在在这些消费者的低价值的消费上已将利润榨干。

例子：使用图形进行分析

价格歧视中的自我选择问题可以借助图形进行分析。图 14.3 画出了两个消费者的需求曲线；为简单起见，假设边际成本为零。图 14.3A 画出的是不存在自我选择问题的价格歧视。企业将分别向高需求消费者和低需求消费者出售 x_h^o 单位和 x_l^o 单位产品，售价分别等于这两个消费者各自的消费者剩余，也就是各自需求曲线下方的面积。因此，高需求的消费者支付 $A+B+C$，消费 x_h^o 单位产品；低需求消费者支付 A，消费 x_l^o 单位产品。

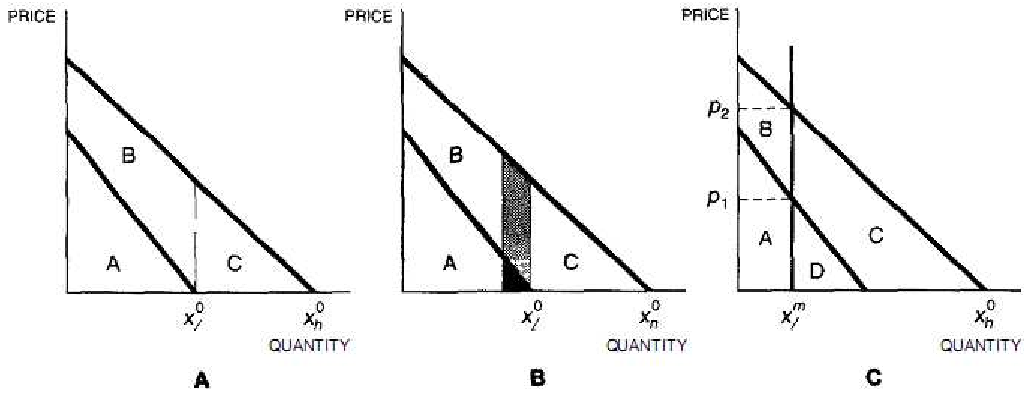


图 14.3：**第二级价格歧视**。A 图画出的是不存在自我选择问题时的解；B 图表明减少低需求消费者的消费束能够增加利润；C 图给出了低需求消费者的利润最大化的消费水平。

然而，这样的策略违背了自我选择的约束。高需求的消费者将偏好和选择低需求消费者的消费束，因为这样做他可以得到消费者剩余 B。为了满足自我选择的约束，垄断企业必须以价格 $A+C$ 供给 x_h^o，这样一来无论高需求消费者怎么选择都可以得到消费者剩余 B。

这种策略是可行的，但它是最优的吗？答案是否定的：如果垄断企业提供给低需求消费者稍微小一点的消费束，该企业损失的利润相当于图 14.3B 中黑色三角形区域那么大，但是却获得了相当于阴影梯形面积那么大的利润。稍微减少提供给低需求消费者的消费束不会对利润的一阶条件产生影响，这是因为在 x_l^o 处边际支付意愿等于零。然而，它在非边际上增加了利润，因为高需求消费者的支付意愿在这一点上大于零。

在低需求消费者的利润最大化的消费水平 x_l^m 处（图 14.3C），进一步降低价格使得从低

需求消费者身上获得**利润的边际下降**（marginal decrease） p_1，正好等于从高需求消费者身上获得利润的边际增加 $p_2 - p_1$。（注意，这一结论也可从（14.18）式推知。）最终的结果是低需求消费者支付 A 从而消费 x_l^m，因此，他的消费者剩余为零；高需求消费者消费量为 x_h^o，这正是社会最优数量。高需求消费者为此支付的钱数等于 $A+C+D$，这使得他得到了消费者剩余 B.

14.8 第三级价格歧视

第三级价格歧视是指企业对不同消费者索要不同的价格，但是每个消费者购买任何单位商品的价格都是相同的。这也许是价格歧视最常见的形式。

以教材为例，通常有两个分离的市场，企业可以很容易地将这两个市场隔离。另外一个例子是根据年龄进行歧视定价，例如电影对年青人打折。如果我们令 $p_i(x_i)$ 表示消费者群体 i 的反需求函数，那么垄断企业的利润最大化问题为

$$\max_{x_1,x_2} p_1(x_1)x_1 + p_2(x_2)x_2 - cx_1 - cx_2$$

这个最大化问题的一阶条件为

$$p_1(x_1) + p_1'(x_1)x_1 = c$$

$$p_2(x_2) + p_2'(x_2)x_2 = c.$$

令 ε_i 表示市场 i （即消费者群体 i ）的需求价格弹性，我们可以将上述两个式子改写为

$$p_1(x_1)[1-\frac{1}{|\varepsilon_1|}] = c$$

$$p_2(x_2)[1-\frac{1}{|\varepsilon_2|}] = c.$$

由此可知，当且仅当 $|\varepsilon_1| < |\varepsilon_2|$ 时， $p_1(x_1) > p_2(x_2)$。因此，市场的需求价格弹性绝对值越大，即市场对价格越敏感，则企业索要的价格越低。

现在假设垄断企业不能将市场分割得一清二白，因此它在一个市场索要的价格会影响另一个市场的需求。例如，某个剧院在星期一晚上打折；那么星期一较低的价格会在一定程度上影响星期二的需求。

在这种情形下，该垄断企业的利润最大化问题为

$$\max_{x_1,x_2} p_1(x_1,x_2)x_1 + p_2(x_1,x_2)x_2 - cx_1 - cx_2,$$

一阶条件为

$$p_1 + \frac{\partial p_1}{\partial x_1} x_1 + \frac{\partial p_2}{\partial x_1} x_2 = c$$

$$p_2 + \frac{\partial p_2}{\partial x_2} x_2 + \frac{\partial p_1}{\partial x_2} x_1 = c.$$

将上述条件整理变形可得

$$p_1 [1 - \frac{1}{|\varepsilon_1|}] + \frac{\partial p_2}{\partial x_1} x_2 = c$$

$$p_2 [1 - \frac{1}{|\varepsilon_2|}] + \frac{\partial p_1}{\partial x_2} x_1 = c.$$

由于我们假设效用是拟线性的，由此可知$\partial p_1 / \partial x_2 = \partial p_2 / \partial x_1$；也就是说交叉价格效应是对称的。上面第一式减去第二式可得

$$p_1 [1 - \frac{1}{|\varepsilon_1|}] - p_2 [1 - \frac{1}{|\varepsilon_2|}] = [x_1 - x_2] \frac{\partial p_2}{\partial x_1}.$$

自然可以假设这两种商品是互相替代的——毕竟它们是同种商品只不过销往不同的消费者群体而已——因此$\partial p_2 / \partial x_1 > 0.$

不失一般性，不妨假设$x_1 > x_2$，那么根据上面的式子立即可知

$$p_1 [1 - \frac{1}{|\varepsilon_1|}] - p_2 [1 - \frac{1}{|\varepsilon_2|}] > 0.$$

重新整理可得

$$\frac{p_1}{p_2} > \frac{1 - 1/|\varepsilon_2|}{1 - 1/|\varepsilon_1|}.$$

根据上式可知，如果$|\varepsilon_2| > |\varepsilon_1|$，则必有 $p_1 > p_2$. 也就是说，如果较小的市场的需求更富有弹性，该市场的价格必定较低。因此，分离市场的这种直觉在某些额外假设条件下，适用于更广泛的情形。

福利效应

对第三级价格歧视的很多讨论都要涉及到允许这种歧视存在的福利效应问题。第三级价格歧视存在与不存在相比，消费者剩余与生产者剩余之和是更大还是更小？

下面我们分析这个问题。首先构造一个福利改进的一般性检验。为简单起见，假设只有两种消费者群体，他们的总效用函数为$u(x_1, x_2) + y$。其中，x_1, x_2分别表示两个群体的消费数量，y是其他所有商品的消费钱数。这两种商品的反需求函数为

$$p_1(x_1,x_2)=\frac{\partial u(x_1,x_2)}{\partial x_1}$$
$$p_2(x_1,x_2)=\frac{\partial u(x_1,x_2)}{\partial x_2}.$$

我们假设$u(x_1,x_2)$是凹而且可微的函数，但是需要指出这个假设不是必需的。

令$c(x_1,x_2)$表示供给x_1和x_2的成本，因此社会福利为

$$W(x_1,x_2)=u(x_1,x_2)-c(x_1,x_2).$$

现在考虑产量的两个组合(x_1^0,x_2^0)和(x_1',x_2')，与此相伴的价格分别为(p_1^0,p_2^0)和(p_1',p_2')。由$u(x_1,x_2)$为凹函数的假设可知，

$$u(x_1',x_2')\le u(x_1^0,x_2^0)+\frac{\partial u(x_1^0,x_2^0)}{\partial x_1}(x_1'-x_1^0)+\frac{\partial u(x_1^0,x_2^0)}{\partial x_2}(x_2'-x_2^0)$$

重新整理并且使用反需求函数的定义，可得

$$\Delta u\le p_1^0\Delta x_1+p_2^0\Delta x_2.$$

类似地，可得

$$\Delta u\ge p_1'\Delta x_1+p_2'\Delta x_2.$$

由于$\Delta w=\Delta u-\Delta c$，我们得到最终的结果：

$$p_1^0\Delta x_1+p_2^0\Delta x_2-\Delta c\ge\Delta W\ge p_1'\Delta x_1+p_2'\Delta x_2-\Delta c. \tag{14.9}$$

在边际成本为常数的情形，$\Delta c=c\Delta x_1+c\Delta x_2$，因此上式变为

$$(p_1^0-c)\Delta x_1+(p_2^0-c)\Delta x_2\ge\Delta W\ge(p_1'-c)\Delta x_1+(p_2'-c)\Delta x_2. \tag{14.20}$$

注意，这些福利的上下限具有很好的一般性，因为我们只假定了效用函数的凹形，这个假设反过来，只是要求需求曲线向下倾斜。Varian（1985）使用间接效用函数推导出的不等式，一般性稍微更高。

为了将这些不等式应用于价格歧视的问题，令初始价格为不变的垄断价格，因此$p_1^0=p_2^0=p^0$，令(p_1',p_2')为歧视价格。则（14.20）的上下界变为

$$(p_1^0-c)(\Delta x_1+\Delta x_2)\ge\Delta W\ge(p_1'-c)\Delta x_1+(p_2'-c)\Delta x_2. \tag{14.21}$$

上界意味着福利增加的必要条件是总产量增加。假设与此相反即总产量下降$\Delta x_1+\Delta x_2<0$。由于$p^0-c>0$，（14.21）意味着$\Delta W<0$。下界给出了在价格歧视情形下福利增加的充分条件，这个条件为加权产量变动之和为正，其中权重等于价格减去边际成本。

图 14.4 给出了上下界的简单几何表示。福利增加ΔW用梯形表示。这个梯形的面积显然是有上下界的，这些界就是两个矩形的面积。

作为福利上下界的一个简单应用，我们分析两个市场的需求曲线均为线性的情形，

$$x_1 = a_1 - b_1 p_1$$

$$x_2 = a_2 - b_2 p_2.$$

为简单起见，假设边际成本等于零。那么如果垄断企业实施价格歧视，他的销售收入最大时有 $x_1 = a_1 / 2$ 和 $x_2 = a_2 / 2$ 。

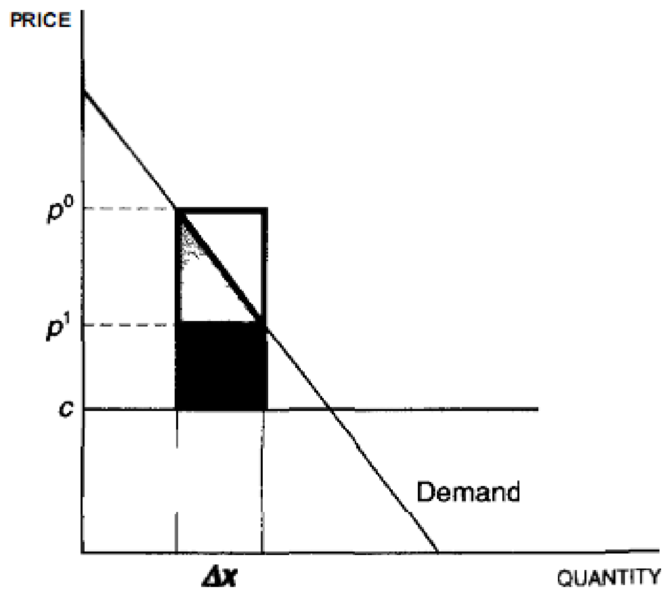


图 14.4：**福利上下界的图示**。梯形区域是消费者剩余的真正变动。

现在假设该企业的产品在两个市场上的售价相同，则总需求曲线为

$$x_1 + x_2 = a_1 + a_2 - (b_1 + b_2) p.$$

这种情形下，为使销售收入最大，垄断企业的产量选择为

$$x_1 + x_2 = \frac{a_1 + a_2}{2}$$

因此，在线性需求的情形下，价格歧视下的总产量和一般垄断情形下的总产量是相同的。于是（14.21）给出的上下界意味着在价格歧视的情形下，福利必定降低。

然而，上述结论依赖于垄断企业在这两个市场上都进行销售的假设。否则上述结论就不成立。例如假设市场 2 很小，如果不允许价格歧视，追求利润最大化的企业将会选择在市场 2 的销量零单位产品（即不销售）。如图 14.5 所示。

在这种情形下，允许价格歧视将导致 $\Delta x_1 = 0$ 以及 $\Delta x_2 > 0$ ，这使得（14.21）中福利变动符号不明朗。当然，这不仅是福利变动，事实上它是一种帕累托改进。

这个例子很稳健。如果由于价格歧视而新开了一个市场，这个市场在普通的垄断情形下是不存在的，那么我们将有帕累托改进，福利提高了。另一方面，如果线性需求是个不错的一阶近似，产出不会因为价格歧视而剧烈变化，那么我们预期福利的净影响是负的。（本章结束）■

注释

我们对质量选择的讨论是基于 Spence（1975）。若想了解价格歧视的更多内容，请参考 Varian（1989a）。

习题

14.1 反需求函数为 $p(y)=10-y$，某垄断者的供给固定为 4 单位商品。求他的销量和售价。在竞争性的市场下，如果需求和供给仍具有这些特征，求响应的产量和售价。如果该垄断者可以供给 6 单位商品，将会出现什么样的情形？[假设自由取舍（free disposal）]

14.2 假设某垄断者面对的需求曲线为 $D(p)=10-p$，他可以供给 7 单位商品。他的利润最大化价格为多少？他的最大利润为多少？

14.3 某垄断者面临的需求曲线为 $x=10/p$，他的边际成本固定为 1。求他的利润最大化产量。

14.4 什么样的需求曲线有 $dp/dc=1$？

14.5 假设某垄断者面临的反需求曲线为 $p(y,t)$，其中 t 是使得需求曲线移动的参数。为简单起见，假设该垄断者的技术具有边际成本不变性质。推导出表明产量如何随 t 的变化而变化的公式。如果 $p(y,t)=ay+bt$，化简你推导出的上面的那个式子。

14.6 某垄断者面临的需求曲线为 $D(p)=10/p$，他的边际成本为正数 c。求他的利润最大化产量。

14.7 假设边际成本固定为 $c>0$，需求函数为

$$D(p)=\begin{cases}10/p & \text{若}p\le 20\\ 0 & \text{若}p>20\end{cases}$$

求利润最大化价格。

14.8 给定垄断者的数量选择，对于什么样的效用函数和需求曲线来说，该垄断者生产的质量水平是最优的？

14.9 在教材中，我们用图形论证了如果 $\partial^2 p/\partial x\partial q>0$，那么 $\partial u/\partial q - x\partial p/\partial q<0$。现在请用代数方法证明。证明步骤如下：

1）证明题目中的假设意味着如果 $z<x$，则

$$\frac{\partial p(z,q)}{\partial q}<\frac{\partial p(x,q)}{\partial q}\text{。}$$

2）将上式左侧用效用函数进行表达。

3）不等式两侧对 z 积分，积分区间为 $[0,x]$。

14.10 价格歧视的一种常见方法是：消费者缴纳一笔“入场费”后才有权购买商品，然后再按每单位商品成本收取消费费用。标准的例子是游乐园，垄断者收取门票后，再对游乐项目收费。这样的定价策略称为两部收费制（two part tariff）。假设所有消费者的效用函数相同，都为 $u(x)$，垄断者的供给成本为 $c(x)$。如果采取两部收费制，垄断者生产的产品数量是大于还是小于有效率的产量？

14.11 考虑第二级价格歧视问题的图形。仔细查看教材图 14.3C，回答下列问题：在什么样的条件下，垄断者只会把商品卖给高需求的消费者？

14.12 如果垄断者选择把商品既卖给高需求的消费者，又卖给低需求的消费者。证明面积 B 必定小于面积 A。

14.13 假设有两个消费者，每个消费者可能购买一单位商品。如果商品的质量为 q，那么消费者 t 达到的效用为 $u(q,t)$。假设垄断者生产质量的成本为零。令消费者 t 愿意为质量 q 支付的价格为 w_t。垄断者无法区分这两个消费者，因此必须提供至多两种不同的质量让消费者自由选择。构建垄断者的利润最大化问题，并详细分析。提示：这个问题和你以前见过的什么样的问题类似？

14.14 我们也可以认为垄断者选择价格而把销量交给市场决定。写出垄断者的利润最大化问题，证明在利润最大化价格上有 $p(1+1/\varepsilon)=c'(y)$。

14.15 某个垄断者的生产技术具有边际成本不变的性质，即 $c(y)=cy$。市场需求曲线具有不变的弹性 ε。政府对该垄断者出售的商品征收从价税 τ，因此当消费者支付价格 P_D 时，该垄断者得到的价格为 $P_S=(1-\tau)P_D$。（此处 P_D 是消费者的需求价格，P_S 是垄断者的供给价格。）

税收部门打算将从价税改为从量税 t，因此 $P_D=P_S+t$。t 为多大时（用 τ 表示），消费者面对的最终价格在这两种税的情形下都是一样的？

14.16【与 14.5 是同一个题目，重复，略去】

14.17 考虑一个简单经济，这个经济运行起来就象一个消费者的行为，该消费者的效用函数为 $u_1(x_1)+u_2(x_2)+y$，其中 x_1 和 x_2 分别为商品 1 和 2 的数量，y 为花费在所有其他商品上的钱。假设商品 1 的供给者是竞争性的，商品 2 的供给者是垄断性的。商品 i 的成本函数记为 $c_i(x_i)$，政府对行业 i 的产品征收数额为 t_i 的从量税。假设 $c_i''>0$，$p_i''<0$ 和 $p_i'<0$。

（a）分别推导出 $i=1,2$ 的 dx_i/dt_i 的表达式，并确定它们的符号。

（b）给定产出变化(dx_1, dx_2)，推导出福利变化的表达式。

（c）假设我们考虑对其中一个行业征税，并用征得的税收收入补贴另外一个行业。我们应该对竞争性行业还是垄断性行业征税？

14.18 有两个消费者，他们的效用函数分别为$u_1(x_1, y_1) = a_1x_1 + y_1$和$u_2(x_2, y_2) = a_2x_2 + y_2$。商品$y$的价格为 1，而且每个消费者的初始财富都“较大”。已知$a_2 > a_1$。两种商品的消费量都不能为负。

商品x是由某个垄断者供给的。该垄断者的边际成本为零，但有产能约束：他至多只能供给 10 单位的商品x。该垄断者至多提供两个价格数量组合(r_1, x_1)和(r_2, x_2)。这里的r_i是消费者购买x_i单位商品支付的钱数。

（a）写出该垄断者的利润最大化问题。约束条件应该有 4 个，另外还有产能约束$x_1 + x_2 \le 10$。

（b）在最优解的情形下，哪个约束是等式约束？

（c）将那些约束条件代入目标函数。最终得到的表达式是什么样的？

（d）(r_1, x_1)和(r_2, x_2)的最优值分别为多少？

14.19 某垄断企业在两个市场销售产品。该垄断者的产品的需求曲线：在市场 1 为$x_1 = a_1 - b_1p_1$；在市场 2 为$x_2 = a_2 - b_2p_2$。这里的x_1和x_2分别为两个市场的销量；p_1和p_2分别为两个市场的售价。该垄断企业的边际成本为零。注意，尽管它在两个市场上能索要不同的价格，但对同一个市场销售的所有单位产品只能索要相同的单位价格。

（a）参数(a_1, b_1, a_2, b_2)满足什么条件时，该垄断企业的最优策略是不进行价格歧视？（假设解为内部解，忽略边界解。）

（b）假设需求函数的形式为$x_i = A_i p_i^{-b_i}$（其中$i = 1, 2$），该垄断企业的边际成本为常数$c > 0$。在什么样的条件下，它会选择不进行价格歧视？（假设解为内部解，忽略边界解。）

14.20 某垄断企业想使得$p(x)x - c(x)$最大化。为了分享一些垄断利润，政府对企业的收入征收数量为t的税，从而使得该垄断企业的目标函数变为$p(x)x - c(x) - tp(x)x$。最初，政府保存着这笔税收收入。

（a）这个税增加还是降低了垄断企业的产量？

（b）现在政府决定将税收收入返还一部分给购买该垄断企业产品的消费者。每个这样的消费者都能得到一定的退款。具体地说，代表性消费者若花费了px元购买该垄断企业的产品，他得到的退款为tpx元。假设该消费者的偏好是拟线性的，推导出该消费者的反需求函数，它是x和t的函数。

（c）该退税计划对垄断企业的产量有何影响？

14.21【与 14.15 相同，重复，略去】

14.22 某垄断企业的成本函数为$c(y)=y$，因此它的边际成本固定为 1。它面对的需求曲线如下：

$$D(p)=\begin{cases}0 & \text{若}p>20;\\ 100/p & \text{若}p\le 20.\end{cases}$$

（a）利润最大化的产量为多少?

（b）如果政府对该垄断企业的产品规定了价格上限，迫使它象竞争性企业一样，政府设定的价格为多少?

（c）如果该企业被迫使象竞争性企业一样，它的产量为多少?

14.23 某个经济体有两类消费者和两种商品。类型 A 的消费者的效用函数为$U(x_1,x_2)=4x_1-(x_1^2/2)+x_2$；类型 B 的消费者的效用函数为$U(x_1,x_2)=2x_1-(x_1^2/2)+x_2$。消费者消费的数量不能为负。商品 2 的价格为 1，每个消费者的收入都为 100 元。类型 A 和 B 的消费者各有 N 个。

（a）假设商品 1 是由某个垄断企业生产，它生产每单位商品 1 的成本固定为c，它不能实施任何形式的价格歧视。计算它的价格和数量。c为多少时，它将选择将商品 1 卖给类型 A 的消费者也卖给类型 B 的消费者?

（b）假设该垄断企业使用“两部收费制”。在这种收费方法下，消费者需要支付一笔入场费k之后才有权购买它的产品。在支付了入场费之后，消费者能以p的单位价格购买任意数量。消费者不能转售商品 1。如果$p<4$，类型 A 的消费者为了有权以价格p购买产品，他对入场费的支付意愿为多大（即k的最大值）? 如果类型 A 的消费者的确支付了入场费，他将会购买多少商品 1? 将类型 A 消费者对商品 1 的需求表达为p和k的函数。类型 B 的消费者对商品 1 的需求函数是什么? 现在将所有消费者对商品 1 的总需求表达为p和k的函数。

（c）如果经济体只包含 N 个类型 A 的消费者，不存在类型 B 的消费者。利润最大化时p和k各为多少?

（d）如果$c<1$，计算在两类消费者都购买产品这个约束条件下，使得垄断企业利润最大的p值和k值。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 15 章：博弈论

曹乾　译

(东南大学　caoqianseu@163.com)

15 博弈论

博弈论（Game theory）研究的是决策制定者的互动。在以前的章节中我们研究的是单个行为人（一个企业或一个消费者）在非常简单环境中的最优决策。在这样的情形下，行为人之间的互动不是非常复杂。在本章我们将为更复杂环境中经济人的行为分析奠定基础。

研究决策制定者的互动行为可从多个方向入手。你可以从社会学、心理学、生物学等角度分析他们的行为。这些方法在相应的环境中各有用处。博弈论强调的是冷血般的“理性”决策分析，因为人们认为这是绝大多数经济行为最适宜的分析方法。

在过去的几十年里，博弈论已被广泛应用于经济分析，人们在经济模型的策略性互动的本质解释方面取得了很多进展。的确，很多经济行为可以看成博弈论的特殊情形，博弈论已称为经济学家必不可少的分析工具。

15.1 博弈的描述

博弈的描述方法有若干种。对于我们的目的来说，策略形和展开形已经足够用了。大致来说，展开形是对博弈“展开”进行描述，而策略形则提供了博弈信息的“缩减版本的”概述。我们首先描述策略形，而将展开形留待介绍序贯博弈时才开始使用。

博弈的策略形提供了下列信息：一组选手；一组策略；每个选手可以作出的选择；一组收益数据，表明每个选手若选择某个策略组合时的收益情况。为了方便介绍，在本章我们仅介绍两人博弈。下面介绍的所有的概念可以容易地推广到多人博弈的情形。

我们假设博弈的描述——每个选手可得到的收益以及可使用的策略——是选手们所共知的。也就是说，每个选手不仅知道他自己的收益和策略，而且知道别的选手的收益和策略。更进一步，每个选手知道其他选手也知道这些信息，等等。我们还假设每个选手都是“完全理性的”，也就是说给定选手的主观信念，每个选手可以选择自己的行为最大化自己的效用，而且当新信息出现时，这些选手可以根据贝叶斯法则（Bayes’ law）修改自己的主观信念。

博弈论，是标准的单人决策理论的一般化形式。如果两个追求效用最大化的消费者的效用取决于对方的选择，每个消费者将如何作出选择？显然，每个消费者在作出合理的选择之前必须考虑其他选手面对的问题。下面我们看这样的一个例子。

例：硬币配对

在这个博弈中，有两个选手，行选手和列选手（以下简称 R 和 C）。每个选手都有一枚硬币，每个人可以决定自己的硬币是正面向上还是反面向上。因此，每个选手有两个策略——我们简称为正面和反面策略（heads and tails）。一旦每个人选定了策略，就会产生相应的

收益。每个选手的收益都取决于双方作出的选择。

每个选手独立作出选择，他们在做决策时都不知道对方的选择。假设如果两个选手的硬币都是正面向上或者反面向上，那么 R 赢得一元 C 输掉一元钱。与此相反，如果一个选手的硬币正面向上而另一个选手的反面向上，那么 C 赢得一元 R 输掉一元钱。

		Column	
		Heads	Tails
Row	Heads	1, −1	−1, 1
	Tails	−1, 1	1, −1

表 15.1:硬币配对博弈矩阵

我们可以使用**博弈矩阵**（game matrix）刻画这个策略性互动。盒内元素（正面，反面）表示如果这个策略组合被选中的话，选手 R 的收益为-1且 C 的收益为$+1$。注意，在盒内每个元素中，R 的收益正好是 C 的收益的相反数。换句话说，两个选手的收益之和为零，因此这样的博弈称为**零和博弈**（zero-sum game）。在零和博弈中，选手的收益是截然相反的而且特别容易分析。然而，经济中的博弈大都不是零和博弈。

例子：囚犯的困境

这个博弈中仍然有两个选手 R 和 C，但是现在他们的利益只是部分冲突。每个选手都有两个策略：合作和背叛。在最初版本的故事中，R 和 C 是同案犯。他们可以选择合作拒绝承认罪行，或者选择背叛从而供出另一方。

在其他情形下，合作和背叛有其他的解释。例如，在两个寡头的情形下，合作表示“继续索要高价”，背叛表示“降价从而抢夺竞争对手的市场”。

Aumann(1987)介绍了一个极其简单的囚犯的困境。在这个博弈中，每个选手向仲裁人宣称：“给我 1000 元，”或者“给对方 3000 元”。注意，这里涉及的金钱来自第三方而不是就行博弈的选手；囚犯的困境是个**变和博弈**（variable-sum game）。

选手在博弈之前可以就博弈事项就行讨论，但是在实际决策时必须是独立决策。每个选手的合作策略是宣布赠送对方 3000 元，而背叛策略是自己拿走 1000 元（并且逃跑!）。表 15.2 给出了 Aumann 版本的囚犯困境的收益矩阵，收益的单位为 1000 元。

稍后我们将详细分析这个博弈，但在此之前我们先要指出这里的“困境”所在。问题在于每个选手都有背叛的动机，无论他是否相信对方合作。如果我相信对方会选择合作的策略给我 3000 元，那么我如果选择背叛就可以一共得到 4000 元。另一方面，如果我相信对方

将会背叛从而拿走 1000 元，那么我最好也是拿走 1000 元。

		Column Cooperate	Column Defect
Row	Cooperate	3,3	0,4
	Defect	4,0	1,1

表 15.2：囚犯的困境

例子：古诺的双寡头模型

考虑一个简单的双寡头之间的博弈，该问题首先由古诺[Cournot（1838）]进行了分析。我们假设有两家企业生产相同的产品，而且生产成本均为零。每个企业都不知道对方的产量决策，但每个企业都需要作出产量决策。如果这两家企业生产的产品数量一共为 x，则市场价格为 $p(x)$，也就是说，$p(x)$ 是这两个企业面临的反需求曲线。

如果企业 i 的产量水平为 x_i，市场价格将为 $p(x_1+x_2)$，企业 i 的利润为 $\pi_i = p(x_1+x_2)x_i$。在这个博弈中，企业 i 的策略是选择产量水平，企业 i 的收益为它自己的利润。

例子：伯特兰的双寡头模型

在这个模型中，基本构造同上例中的古诺博弈，唯一不同的是每个选手的策略现在变为，给定任何产品数量，他要报出自己愿意供给该数量的价格。在这种情形下，收益函数的形式发生了很大的变化。显然消费者只会从报价最低的企业购买产品；如果这两个企业的报价相同，消费者从这两个企业购买的数量相同。令 $x(p)$ 表示需求函数，这样我们就得到了企业 1 的收益：

$$\pi_1(p_1, p_2) = \begin{cases} p_1 x(p_1) & p_1 < p_2 \\ p_1 x(p_1)/2 & p_1 = p_2 \\ 0 & p_1 > p_2. \end{cases}$$

这个博弈的结构类似于囚徒困境的结果。如果两个选手合作，他们可以索要的价格为垄断价格，每个选手都得到垄断利润的一半。但是每个选手稍微降价从而占有整个市场的诱惑总是存在。然而如果两个选手都降低价格，那么他们的状况都变坏了。

15.2 策略性选择的经济模型

注意古诺博弈和伯特兰博弈的结构截然不同，尽管它们看起来都是模拟同一种经济现象——双寡头。在古诺博弈中，每个企业的收益是它自身的策略性选择的连续函数；在伯特兰博弈中，收益是策略的非连续函数。正如我们的预期，它们导致截然不同的均衡结果。这些模型中的哪一个是“正确的”？

抽象地问哪个模型是“正确的”毫无意义。答案取决于你想模拟的情形是什么样的。因此更有益的问题是：在模拟行为人的策略集时应该考虑哪些事项。

一种方法是根据现实证据进行判断。如果我们注意到石油输出国组织(OPEC)在报告中指出，它们试图确定每个成员国的生产份额，但油价交由世界石油市场确定。那么，对该博弈策略的合理模拟应从产量水平而不是价格入手。

另外一个应该考虑的事项是，策略是承诺性质的或者一旦观测到对手的行为策略很难被改变。上面介绍的那些博弈是“一次性”博弈，但是这些博弈模拟的事实却是实时发生的。假设我会我的产量选择了一个价格，然后发现我的竞争对手稍微降低了一些价格。在这种情形下，我会立即修改我的价格。既然一旦知道了对手的策略，我可以很快修改我自己的策略，那么使用一次型博弈模拟这种互动是不合理的。因此，在分析这类价格制定的博弈中，为了描述详细的策略互动，我们应该使用多阶段的博弈。

另一方面，假设我们将古诺博弈中的产量解释为“产能”，产能是很难被撤销和修改的。在这种情形下，如果我发现了竞争对手的产量水平，我要想改变我自己的产量将要花费很大的代价。因此产能或产量似乎是合理的策略性变量，即使在一次性博弈中也是比较合理的。

和大多数其他经济模型一样，如何表示博弈的策略性选择（能够描述现实策略互动的细节）而又让该博弈容易分析，是一门艺术。

15.3 解的概念

在很多博弈中，策略性互动的本质意味着选手在选择策略时不想让对方事先猜测到他的选择。例如，在前面讨论的硬币配对博弈中，两个选手显然都不希望对方能够准确猜测到他自己的选择。因此，自然可以考虑选择随机策略，比如使正面向上的概率为 p_h，使反面向上的概率为 p_t。这样的策略称为混合策略（mixed strategy）。如果某个策略被选择的概率为 1，那么这个策略称为纯策略（pure strategy）。

令 R 表示行选手可以使用的纯策略集，则行选手可以使用的混合策略是在 R 上所有概率分布组成的集合，其中在 R 中选择策略 r 的概率为 p_r。类似地， p_c 表示列选手在其纯策略集 C 中选择策略 c 的概率。为了找到该博弈的解，我们要找到一组混合策略 (p_r, p_c)，这组策略在某种意义上是均衡的。有时，均衡混合策略解中某些选择的概率为 1，这样的情形我们称为纯策略。

寻找解概念的自然出发点是标准的决策理论：我们假设每个选手对其他选手的策略选择具有概率信念；而且假设每个选手选择使得他自己期望收益最大的策略。

例如假设行选手的收益为$u_r(r,c)$如果行选手选择r而且列选手选择c。我们假设行选手对于列选手的选择具有主观上的概率分布，以π_c表示；主观概率的基本思想请参见第11章。此处假设π_c表示行选手猜想的列选手选择c的概率。类似地，列选手对行选手的行为也有某些新年，我们用π_r表示。

我们假设每个选手选择的是混合策略，以p_r表示行选手的实际混合策略选择，以p_c表示列选手的实际混合策略选择。由于行选手在做选择时不知道列选手的选择，对于行选手来说，结果(r,c)出现的概率为$p_r\pi_c$。这个概率就是行选手选择r的（客观）概率乘以行选手认为列选手选择c的（主观）概率。因此，行选手的目标是选择概率分布π_r使得下式最大化：

$$\text{行选手的期望收益}=\sum_r\sum_c p_r\pi_c u_r(r,c).$$

另一方面，列选手希望最大化

$$\text{列选手的期望收益}=\sum_c\sum_r p_c\pi_r u_c(r,c).$$

到目前为止，我们仅仅将标准的决策理论上的模型应用到这个博弈——每个选手在给定他自己信念的情形下使得他的期望效用最大化。给定我对其他选手行为的信念，我会选择使我的期望效用最大的策略。

在这个模型中，我对其他选手行为的信念是外生变量。然而，现在我们对这个标准决策模型增加一个新变化，问：对其他选手行为持有何种信念是合理的？毕竟博弈中的每个选手都知道对方试图使自己的收益最大化，而且每个人应该使用这个信息来确定对对方的行为持有何种信念是合理的。

15.4 纳什均衡

在博弈论中，我们接受下列假设：每个选手都努力使得自己的收益最大化，而且每个选手知道对方的目标也是这样的。因此，为了确定我对其他选手的行为持有何种信念是合理的，我必须知道他们对我的信念。在上一节给出的期望收益表达式中，行选手的行为——他选择每种策略的可能性——是用概率分布p_r代表的，而列选手对行选手行为的信念是用列选手的（主观）概率分布π_r代表的。

一个自然的一致性要求是，每个选手对于对方行为的信念和对方实际行为是一样的。与实际频率一致的预期有时称为**合理预期**（rational expectations）。纳什均衡就是一种合理预期的均衡。更正式地说：

纳什均衡（Nash equilibrium）。一个纳什均衡包含对于策略(r,c)的信念概率(π_r,π_c)，以及选手实际选择策略的概率(p_r,p_c)，使得：

1）选手们的信念是正确的： $p_r = \pi_r$ 并且 $p_c = \pi_c$ 对于所有 r 和 c 成立；而且

2）每个选手选择 p_r, p_c 以便在给定他的信念情形下使得他自身的期望效用最大化。

在这个定义中，显然可以看出纳什均衡是**行为**和**信念**之间的均衡。在均衡时，每个选手正确地预测出对方的行为，而且这两个选手之间的信念是互相一致的。

纳什均衡的一个更常规的定义是：它指的是一对混合策略 (p_r, p_c)，这样的策略能保证在给定对方策略的情形下使得每个选手的选择能最大化自己的期望效用。这个定义和我们的定义是等价的。但是这个定义存在着缺陷，因为它模糊了选手们的信念和选手们行为之间的区别。我们的定义在区别这两个概念上比较小心。

纳什均衡的一种特别有趣的情形，是纯策略纳什均衡。也就是说，在纳什均衡时，每个选手选择某特定策略的概率都为 1。即：

纯策略（pure strategies）。纯策略纳什均衡是一对 (r^*, c^*) 使得：对于行选手的所有策略 r 都有 $u_r(r^*, c^*) \geq u_r(r, c^*)$；对于列选手的所有策略 c 都有 $u_c(r^*, c^*) \geq u_c(r^*, c)$。

纳什均衡是对一组策略施加的最低的一致性限制：如果行选手认为列选手会选择 c^*，那么行选手的最优反应是 r^*；列选手的最优反应可以类似分析。没有选手会认为单方面偏离纳什均衡策略符合他自己的利益。

如果一组策略不是纳什均衡，那么至少有一个选手对对方的信念与对方的行为不符。也就是说，其中一个选手必然预期对方的行为不符合他自己的利益——这与我们最初的假设是矛盾的。

人们通常将均衡概念视为某个调整过程中的“静止点”。纳什均衡的一种解释是，对对方动机“认知”的调整过程。行选手可能认为：“如果我认为列选手会选择策略 c_1，那么我的最优反应是选择 r_1。但是如果列选手想到我会选择 r_1，那么他的最优反应是选择策略 c_2。但是如果列选手选择 c_2，我的最优反应是选择 r_2 …”等等。于是，纳什均衡是一组信念和策略，它们要能使得每个选手关于对方的信念正好和对方的实际选择是一致的。

有时，人们将上一段介绍的“认知”的调整过程解释为实际调整过程，在这个调整过程中，每个选手实验性地选择不同策略，目的是理解对方的选择。尽管这样的实验和了解过程是在现实生活中策略性互动中发生的，但是严格来说，这种观点不能很好地解释纳什均衡的概念。原因在于如果每个选手知道博弈将进行若干次，那么每个选手都试图将他自己在 t 期的行为，建立在观测到对方 $1 \sim t$ 期的选择的基础之上。在这种情形下，纳什均衡的正确概念是指一系列的选择，这些选择都是针对对方一系列选择的（在某种程度上的）最优反应。

例子：计算纳什均衡

下面的博弈称为“性别大战（Battle of the sexes）”。该博弈背后的故事是这样的。Rhonda 行选手和 Calvin 列选手（以下简称 R 和 C）在讨论这学期是选择微观经济学还是宏观经济学课程。如果都选择微观经济学，那么 R 的效用为 2，C 的效用为 1；如果都选择宏观经济学，那么 R 的效用为 1，C 的效用为 2；如果他们选择不同的课程，那么他们的效用都为零。

		Calvin	
		Left (micro)	Right (macro)
Rhonda	Top (micro)	2, 1	0, 0
	Bottom (macro)	0, 0	1, 2

表 15.3：性别大战

下面我们计算出这个博弈的所有纳什均衡解。首先，我们寻找纯策略纳什均衡。方法是系统地检验针对各种策略选择的最优反应。假设 C 认为 R 会选择上。C 选择左的收益为 1，选择右的收益为 0，因此 C 会选择左，也就是说 C 对 R 选择上的最优反应是选择左。另一方面，如果 C 选择左，容易看出 R 的最优反应是选择上。这个推理过程表明（上，左）是一个纳什均衡。类似地，（下，右）也是一个纳什均衡。

我们也可以按照如下方法系统地求出均衡解：写出每个选手面对的最大化问题并且分析一阶条件。令(p_t, p_b)表示 R 选择上（Top）和下(Bottom)的概率，类似地(p_l, p_r)表示 C 选择左和右的概率。那么 R 的最大化问题为

$$\max_{p_t, p_b} \; p_t[p_l 2 + p_r 0] + p_b[p_l 0 + p_r 1]$$

$$\text{使得 } p_t + p_b = 1； \quad p_t \geq 0； \quad p_b \geq 0.$$

令λ， μ_t 和 μ_b 分别表示约束的 Kuhn-Tucker 乘子，因此拉格朗日函数的形式为

$$L = 2p_t p_l + p_b p_r - \lambda(p_t + p_b - 1) - \mu_t p_t - \mu_b p_b.$$

上式分别关于 p_t 和 p_b 求导可得 R 的 Kuhn-Tucker 条件:

$$\begin{aligned} 2p_l &= \lambda + \mu_t \\ p_r &= \lambda + \mu_b. \end{aligned} \tag{15.1}$$

由于在前面我们已经知道了纯策略均衡解，下面我们只考虑 $p_t > 0$ 和 $p_b > 0$ 的情形。松弛条件意味着$u_t = u_b = 0.$使用 $p_t + p_b = 1$的事实可知，当 $p_l = 1/3$ 和 $p_r = 2/3$ 时，R 的最优反应是选择混合策略。

对于 C 的问题可以类似地进行分析，我们发现 $p_t = 2/3$ 和 $p_b = 1/3$ 。每个选手从混合策略中得到的收益很容易计算，将上述这些概率数字代入目标函数。在这个例子中，每个选

手的期望收益是 $2/3$。注意，与这个混合策略相比，每个选手都偏爱两个纯策略中的任何一个（前面已求解出），因为纯策略的收益比较高。

15.5 对混合策略的解释

有时很难对混合策略给出行为解释。对于有些博弈，例如前面的介绍的硬币配对博弈，显然混合策略是唯一的合理均衡。但是对于其他博弈，例如双寡头博弈，混合策略似乎是不现实的。

混合策略在某些情形下是不现实的，除此之外，还存在着纯逻辑基础的困难。我们再次考虑性别大战中的混合策略。这个博弈中的混合策略具有以下性质：如果行选手选择的是他的均衡混合策略，列选手纯策略（两个纯策略中的任何一个）得到的收益必然等于他选择均衡混合策略得到的收益。看清上述结论的最简单方法是审视（15.1）的一阶条件。由于 $2p_l = p_r$，选择上的期望收益与选择下的期望收益是相同的。

这个现象不是偶然的。事实必然总是下列这样的：对于任何混合策略均衡，如果一个选手认为对方将选择混合策略，那么他在选择均衡混合策略还是选择任何纯策略（纯策略是混合策略的组成部分）之间必然是无差异的。其中的逻辑很简单：如果某个纯策略（纯策略是他的均衡混合策略的组成部分）的期望收益高于混合策略的其他组成部分，那么应该增加选择这样策略的概率。但是如果在选择的混合策略中所有纯策略的概率都为正，而且它们有相同的期望收益，那么这必然也是混合策略的期望收益。这反过来意味着选手在选择纯策略还是选择混合策略之间是无差异的。出现这种“退化”的原因是期望效用函数是概率的线性函数。人们想要更有说服力的理由来“迫使”混合策略结果的出现。

在某些情形下这不会造成严重的问题。假设你是一大群人的成员，成员之间随机配对进行硬币配对游戏。假设一开始每个人选择的是唯一的混合策略纳什均衡 $(1/2,1/2)$。最终有些人厌倦了选择混合策略，从而决定自此以后都选择正面向上或反面向上。如果这样决定的人群中选择正面向上和选择反面向上的人数一样多，那么每个选手的选择问题没有发生大的变化：每个选手仍然理性地认为他的对手选择正面向上与选择反面向上的概率为 50:50。

按照这种方式这个人群中的每个成员都可以选择纯策略，但是具体到某一局博弈中，选手们无法知道对方选择的是哪种纯策略（正面向上还是反面向上）。这种将混合策略概率解释为群体频率在动物行为的模型中是比较常见的。

混合策略的另外一种解释方法是考虑一个既定的选手在一次性博弈中到底是选择正面向上还是反面向上。这样的选择被认为是由对方无法确定的特质因素决定的。例如如果你正处于“正面情绪”中，你会选择正面向上，反之则反是。你可以察觉到你的心情，但是你的对手不能。因此，从每个选手的角度看，对方的策略是随机的，即使每个选手的策略是已确定好了的。一个选手的混合策略真正要紧的事情是它造成了其他选手的不确定性。

15.6 重复博弈

我们在前面指出过，同样的选手参与的重复博弈和这些选手参与的一次性博弈的重复，是不同的。这是因为重复博弈的策略空间要大得多：每个选手在某个时点上的选择是到这个时点止的整个博弈**历史**的函数。由于我的对手会根据我的历史选择情况改变自己的行为，在制定我自己的选择时，我必须将这个因素考虑进去。

我们以前面介绍过的囚犯的困境为例说明。在这个博弈中，选手们选择（合作，合作）符合他们的“长期”利益。因此一个选手有可能向另一个选手“发出信号”表示他愿意在第一轮博弈中选择合作。在短期，对方选择背叛符合他自己的利益，但是在长期他会这么选择吗？他也许会这么推理：如果他背叛，那么对方就会失去耐心，因此对方从此以后选择的都是背叛。因此，第二个选手如果选择短期的最优策略，那么他在长期就会输。这个推理背后的事实是我在第一轮的选择将对将来的选择产生后果——其他选手将来的选择可能会依据我当前作出的选择。

下面我们分析一下策略（合作，合作）是否是囚犯困境重复博弈的纳什均衡解。首先我们考虑的情形是，每个选手知道博弈进行的既定次数。考虑选手在最后一轮博弈进行之前的推理。每个选手都认为，在这个时点上，他们即将进行的是一次性博弈。既然大家以后没有机会博弈了，纳什均衡的标准逻辑可以用上了，因此双方的选择都是背叛。

现在开始考虑倒数第二轮的博弈。在这一轮博弈中，似乎每个选手都应该选择合作，以便向对方发出他是个“好人”的信号，从而期望在下一轮仍合作。但是我们已经看到，当下一轮（最后一轮）到来时，每个选手都想选择背叛。因此，在最后一轮选择合作没有什么好处——只要两个选手都相信对方在最后一轮会选择背叛，那么在倒数第二轮试图以向对方示好的方式来影响对方在最后一轮的行为，是不明智的。同样的逻辑也适用于倒数第三轮、倒数第四轮、以此类推。因此，在重复进行的囚犯困境博弈中，若重复的具体次数（有限次）是事先知道的，则纳什均衡是在每一轮都选择背叛。

如果重复博弈是无限次的，情形就与有限次大不相同。在这种情形下，在每一轮，双方都知道博弈还将继续下去，因此合作行为存在（潜在的）收益。我们以囚犯的困境为例看看无限次的重复博弈是如何运行的。

考虑无限次重复进行的囚犯困境博弈。这个重复博弈的策略是个函数序列，这些函数指明了在某个特定阶段（轮）每个选手选择合作还是背叛，这些函数是以前各轮博弈历史信息的函数。重复博弈的收益是每一轮博弈的收益贴现之和；也就是说，如果某个选手在时期t得到的收益为u_t，他在重复博弈中的收益为$\sum_{t=0}^{\infty} u_t/(1+r)^t$，其中$r$是贴现率。

我们断言，只要贴现率不是很高，就存在纳什均衡的策略组合，使得每个选在在每一轮选择合作对自己是有利的。事实上，容易举出这样策略的例子。考虑下面的策略：“在当前轮次的博弈上选择合作，除非对方在最后一轮选择背叛。如果对方在最后一轮选择背叛，则一直背叛下去。[一]”这样的策略有时称为惩罚策略（punishment strategy），原因很明显：如

[一] 范里安（作者）的意思在这里有些含糊，因为在无限次重复博弈中不存在最后一轮的说法。这里的惩罚策略其实就是“以牙还牙”策略——双方一开始都选择合作策略，但一旦有人选择背叛，则双方一直背叛下去。译者注。

果一个选手背叛，则他会一直受到处罚，因为自此以后他的收益降低了（变成了背叛的贴现收益）。

为了说明惩罚策略组合构成了一个纳什均衡，我们只要证明一个选手选择了惩罚策略，另外一个选手最好也选择惩罚策略。假设选手一直合作到 T 轮，如果某个选手在这一轮选择背叛，将会出现什么样的情形？我们使用 15.1 节囚犯困境例子中的数值，背叛者立即得到了收益 4，但是从此以后他的收益都是 1，这个无穷收益流的贴现值为$1/r$，因此选择背叛得到点的总期望收益为$4+(1/r)$。

另一方面，如果他继续选择合作，他的期望收益将为$3+(3/r)$。只要$3+(3/r)>4+(1+r)$即$r<2$，选择继续合作就是明智的。只要这个条件得到满足，惩罚策略就构成了一个纳什均衡：如果有一方选择惩罚策略，另外一个选手也会选择惩罚策略，任何一方偏离这个选择都不是明智的。

这个结构非常稳健。在本质上，类似的逻辑适用于任何策略，只要这样的策略的收益大于（背叛，背叛）的收益。无名氏定理（Folk Theorem）指出了一个重要的结论[二]：在重复进行的囚犯困正中，只要某种策略组合的收益大于（背叛，背叛）的收益，那么这种策略组合就是个纳什均衡。这个定理的证明多少沿着我们上面构建的思路。

例：维持卡特尔

考虑一个简单的关于双寡头的重复博弈的例子。如果两个企业选择古诺（Cournot）博弈，那么它们的利润为(π_c,π_c)；如果这两个企业选择能使得它们联合（joint）利润最大的产量，那么它们的利润为(π_j,π_j)。我们都知道，能使联合利润最大的产量水平通常不是只有一期的博弈的纳什均衡解——如果每个选手都相信对方维持约定的产量不变，它有激励多生产。然而，只要贴现率不是很高，联合利润最大化这个解将是重复博弈的一个纳什均衡。适当的惩罚措施是，双方维持卡特尔产量，但只要有人背叛，大家自此以后都生产古诺产量。可以证明，这是个纳什均衡。证明方法类似于囚犯困境情形。

15.7 纳什均衡的凝练

纳什均衡概念似乎是博弈均衡解的合理定义。和其他任何均衡概念一样，这里立即有两个我们感兴趣的问题：1）纳什均衡一般存在吗？（存在性）；2）纳什均衡是唯一的吗？（唯一性）。

幸运的是，存在性不是个问题。纳什（1950）证明了只要选手数量是有限的、纯策略数量也是有限的，总存在博弈均衡。当然，这里可能的均衡可能涉及混合策略。

[二] 无名氏定理之所以叫这个名字，是因为它是博弈论中一个比较重要的定理，但不知道是谁（无名氏）首先提出了它。译者注。

然而，唯一性一般来说是不成立的。我们已经看到，一个博弈可能存在若干个纳什均衡。博弈理论学家已投入大量精力，来寻找可以用于在若干个纳什均衡之间进一步进行选择的标准。这些标准成为纳什均衡的**凝练**（refinements），下面我们将考察一些凝练。

15.8 优势策略

令 r_1 和 r_2 为行选手的策略。如果无论列选手怎么选择，行选手的 r_1 的收益都严格大于 r_2 的，那么我们说：与 r_2 相比，r_1 是**优势策略**（dominant strategy）。如果行选手的 r_1 的收益都至少与列选手的所有选择的收益一样大，而且严格大于列选手的某些选择的收益，那么我们说，与 r_2 相比，r_1 是**弱优势策略**（weakly dominant strategy）。

优势策略均衡（dominant strategy equilibrium）是指选手们的一个策略选择组合比如 (r_1, c_1)，该策略组合要能使得：与行选手的任何其他策略相比，r_1 是（弱）优势策略；与列选手的任何其他策略相比，c_1 是（弱）优势策略。

囚犯困境博弈拥有优势策略均衡，这个优势策略均衡是（背叛，背叛）。如果我认为你合作，那么我选择背叛对我有利；如果我认为你背叛，我选择背叛仍然对我有利。

显然，优势策略均衡是纳什均衡，但是并非所有的纳什均衡都是优势策略均衡。优势策略均衡（如果存在的话），是个让人信服的博弈解，因为每个选手都有唯一的最优选择。

15.9 删除劣势策略

当不存在优势策略均衡时，我们必须求助于纳什均衡的思想。但通常存在着多于一个的纳什均衡。于是我们的问题是删除某些“不合理的”纳什均衡。

一个关于选手们行为的合理信念是，选手们不会选择劣势策略（比其他策略差的策略）。这意味着如果给定一个博弈，我们应该首先删除所有的劣势策略，然后计算剩下的博弈的纳什均衡。这个程序称为**删除劣势策略**（elimination of dominated strategies）；这种方法有时能大大减少纳什均衡解的数量。

例如考虑表 15.4 中的博弈，该博弈具有劣势策略。

注意，这个博弈有两个纯策略纳什均衡：（上，左）和（下，右）。然而对于列选手来说，右弱优于左（左是劣势策略）。如果行选手认为列选手绝不会选择劣势策略，那么该博弈的唯一均衡是（下，右）。

经济学家普遍认为删除严格劣势策略（strictly dominated strategies）能够简化博弈的分析。然而，删除弱劣势策略（weakly dominated strategies）则存在着不少问题。有例子表明，删除弱劣势策略可能会显著改变博弈的策略性性质。

		Column	
		Left	Right
Row	Top	2,2	0,2
	Bottom	2,0	1,1

表 15.4：具有劣势策略的博弈

15.10 序贯博弈

到目前为止，我们在本章介绍的博弈都有着非常简单的动态结构：它们要么是一次性博弈，要么是一次性博弈的重复序列。它们的信息结构也非常简单：每个选手知道其他选手的收益和可用的策略，但是事前不知道每个选手实际选择哪个策略。这样的博弈也称为**同时行动**（simultaneous moves）博弈。

但是经济学家感兴趣的很多博弈都不具有这种结构。在很多情形下，选手至少在某个选择上有先后顺序之分，一个选手可能在它决定自己的选择**之前**知道其他选手的选择。经济学家对这种博弈很感兴趣，因为很多经济博弈就有这样的结构：例如，垄断企业在决定具体产量时，事先已知道消费者的选择；或者，在双头垄断的情形下，一个寡头在决定自己的产量决策前已知道对手的资本投资情况；等等。这类博弈的分析需要一些新的概念。

例如，考虑表 15.5 描述的简单博弈。容易看出，这个博弈有两个纯策略纳什均衡，即（上，左）和（下，右）。隐含在这种博弈描述中的思想是，两个选手同时进行选择，而不知道对方的选择。但是现在我们假设行选手先选择，列选手在看到行选手的选择之后才进行选择。

		Column	
		Left	Right
Row	Top	1,9	1,9
	Bottom	0,0	2,1

表 15.5：同时行动博弈的收益矩阵

为了描述这样的序贯博弈，有必要引入一个新的工具，即**博弈树**（game tree）。这是个树状图，它表明了每个时点上的每个选手的选择。每个选手的收益是用博弈树的“树叶”表示，如图 15.1 所示。这个博弈树是**展开形**（extensive form）的一部分，展开形是描述博弈的另外一种方法。

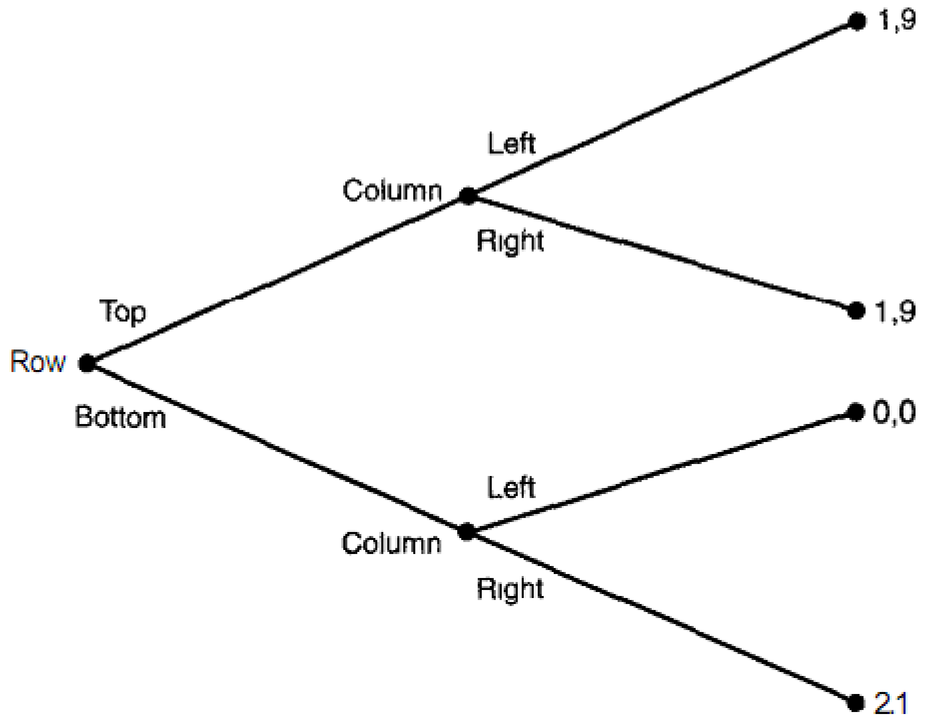


图 15.1：博弈树。这个图描述了表 15.5 中博弈的收益，但在这里行选手首先行动。

博弈树的有点在于它描述了博弈的动态结构：选择有先后之分。博弈中的一个选择对应着博弈树的分枝的选择。一旦做出了某个选择，选手们就处于**子博弈**（subgame）中，子博弈由自此点起的选手们的策略和收益组成的博弈，它是原博弈的一部分。

对于每个可能的子博弈，可以直接计算纳什均衡，尤其对于我们所举的这个例子来说，因为该博弈非常简单。如果行选手选择上，他实际上选择了一个非常简单的子博弈，在这个子博弈中，列选手只剩下一次选择。在这个子博弈中，列选手在他的两个策略中是无差异的，因此，如果行选手选择上，那么他的收益必定为 1。

如果行选手选择下，那么列选手的最优选择是选择右，这样行选手的收益为 2。由于 2 大于 1，行选手选择下显然好于选择上。因此，这个博弈的合理均衡是（下，右）。这个解当然只是表 15.5 中那个同时行动博弈的其中一个纳什均衡。如果列选手宣称他将选择右，那么行选手的最优反应是选择下；如果行选手宣称他讲选择下，那么列选手的最优反应是选择右。

但是，另外一个均衡即（上，左）为何是不合理的？如果行选手认为列选手将选择左，那么行选手的最优选择当然是上。但是为什么行选手认为列选手将会选择左？一旦行选手选择了下，在由此产生的子博弈中，列选手的最优选择是选择右。在这样的时点上，选择左不

是该子博弈的均衡选择。

在这个例子的两个纳什均衡中，只有一个满足下列条件：它不仅是整体博弈的均衡，**而且是每个子博弈的均衡**。具有这种性质的纳什均衡称为**子博弈完美均衡**（subgame perfect equilibrium）。

子博弈完美均衡容易计算，至少在我们讨论的这类博弈中是这样的。你只要从博弈的最后一步向前推导即可，这种方法称为逆向归纳（backwards induction）。最后进行选择（博弈最后一步）的那个选手在子博弈中有着很简单的最优化问题，因为此时已不存在策略分枝，这个最优化问题容易解出。拥有倒数第二步的选手事先可以观察上述选手对他的选择的反应，以此类推。这种分析模式类似于动态规划（详见第 19 章）。一旦选手们通过逆向归纳的方法看清了如何进行博弈，选手们就可以正向行动（博弈）了[一]。

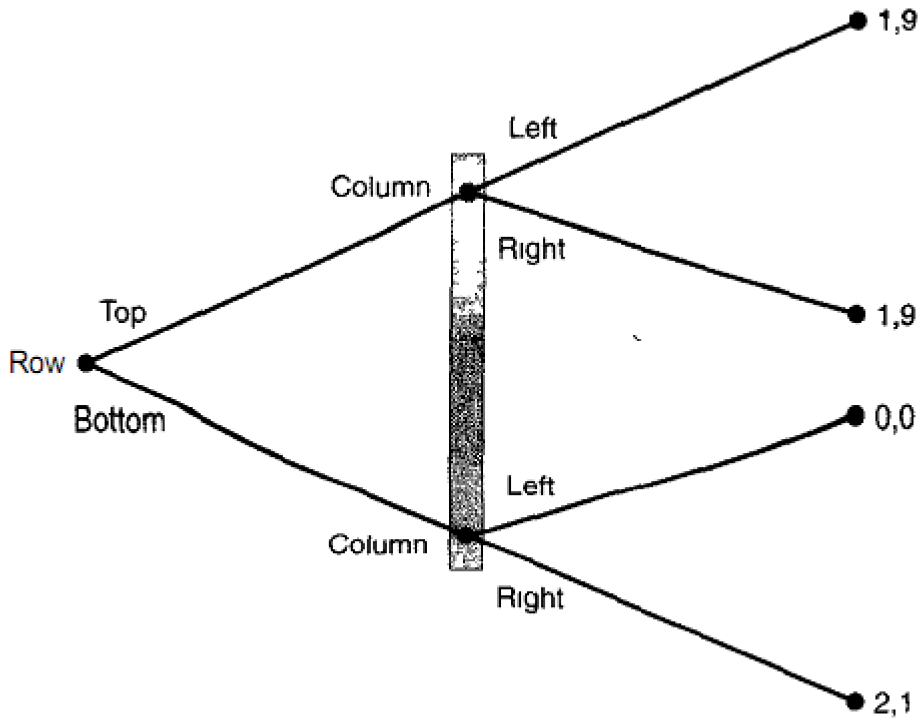


图 15.2：信息集。这是原来同时行动博弈的展开形。信息集（阴影区域）表示列选手在进行选择时不知道行选手究竟选择了哪个策略。

展开形博弈也能够描述某些行动是序贯的而有些行动是同时进行的情形。此处需要引入的概念是**信息集**（information set）。某个选手的信息集是该选手无法进行区分的博弈树的所有节点组成的集合。例如，本节一开始介绍的同时行动的博弈，可用图 15.2 表示。在这个图中，阴影区域表示当列选手在决策时不能区分行选手究竟选择了哪个策略。因此，这类似于选手们同时进行选择。

[一] 类比 Kierkegaard（1938）的话："哲学家说过，生活必须从终点向起点看（逆向地看），才能看得更明白。这当然非常正确。但是他们忘记了另外一个事实，生活必须从起点开始（正向进行）。"

因此，展开形博弈并不仅可以描述策略形博弈的所有内容，还可以描述序贯选择的信息和信息集。在这个意义上，与策略形相比，展开形是个更强大的概念，因为它包含了选手们策略互动的更详细信息。正是这些额外信息使得人们可以删除某些“不合理的”纳什均衡。

例子：一个简单的讨价还价模型

两个选手 A 和 B 要分 1 元钱。他们同意最多花费三天时间协商如何分配。第一天，A 提出分配方案，B 可以选择接受或不接受，如果 B 不接受，B 需要在第二天提出分配方案。如果 A 不接受，那么在第三天 A 需要提出最终分配方案。如果他们在三天之内无法达成协议，这两个选手都只能得到零元钱。

A 和 B 的耐心程度不同：对于未来的收入，A 和 B 分别用α和β的日贴现率贴现。比如第二天的 1 元钱，在 A 和 B 看来分别只相当于现在的α元和β元。最后，我们假设如果某个选手在两个分配方案之间无差异，他将会选择对手更喜欢的那个方案。其中的思想在于，他的对手可能提出只分配给他任意小的钱，从而使他更偏好于某个选择，这个假设能让我们把这样的“任意小的数字”近似为零。可以证明，这个讨价还价博弈存在着唯一一个子博弈完美均衡[一]。

使用逆向归纳的思想，我们从博弈的最后阶段（第三天的博弈）开始分析。在这个时点上，A 提出的分配方案是“若不接受就滚远点”。显然，A 此时的最优策略是分配给 B 能接受的尽量小的数额（根据假设，这个数额可以视为零）。因此，如果博弈实际进行了三天，A 将会得到 1 元，B 得到零元（即任意小的数额）。

现在回到上一步即第二天的博弈，当轮到 B 提出分配方案时，此时 B 应该认识到 A 的选择将是拒绝 B 提出的方案，这样 A 能确保自己得到 1 元钱。第三天的 1 元钱对 A 来说，只相当于第二天的α元，因此如果 B 提出的分配方案中，分配给 A 的钱数小于α元，那么 A 必定会拒绝这个方案。对于 B 来说，他当然偏好现在的$1-\alpha$元胜于第三天的 0 元，所以他会提出分配α元给 A，于是 A 会接受。因此，如果博弈在第二天终止，那么 A 得到α元，B 得到$1-\alpha$元。

现在回到第一天。在这个时点上，A 提出分配方案。A 认识到只要 B 拒绝，那么 B 在第二天就能得到$1-\alpha$元。因此，为了避免将博弈拖到第二天，A 提出的方案中，分配给 B 的现值不能小于$1-\alpha$元。因此，他承诺分配$\beta(1-\alpha)$元给 B。B 发现这个方案他刚好能接受，博弈结束。最终的结果是博弈在第一步结束，其中 A 得到$1-\beta(1-\alpha)$元，B 得到$\beta(1-\alpha)$元。

图 15.3A 说明了$\alpha=\beta<1$时的博弈过程。最外面的对角线表明第一天的可能的收益，即所有满足$x_A+x_B=1$的收益。第二条线（从外向内数）表示，如果博弈在第二期结束，那么各选手的收益现值为多大：$x_A+x_B=\alpha$。第三条线（从外向内数）表示，如果博弈在

[一] 这是鲁宾斯坦-斯坦伯格讨价还价模型的简化版本；更具体的内容请参考本章后的附录部分。

第三期结束，那么各选手的收益现值为多大：$x_A + x_B = \alpha^2$。图 15.3A 中的直角路径描述的是，在通向最终子博弈完美均衡过程中，每一期的最小可接受的钱数。图 15.3B 描述的是讨价还价博弈有更多期的情形，这个过程类似于图 15.3A。

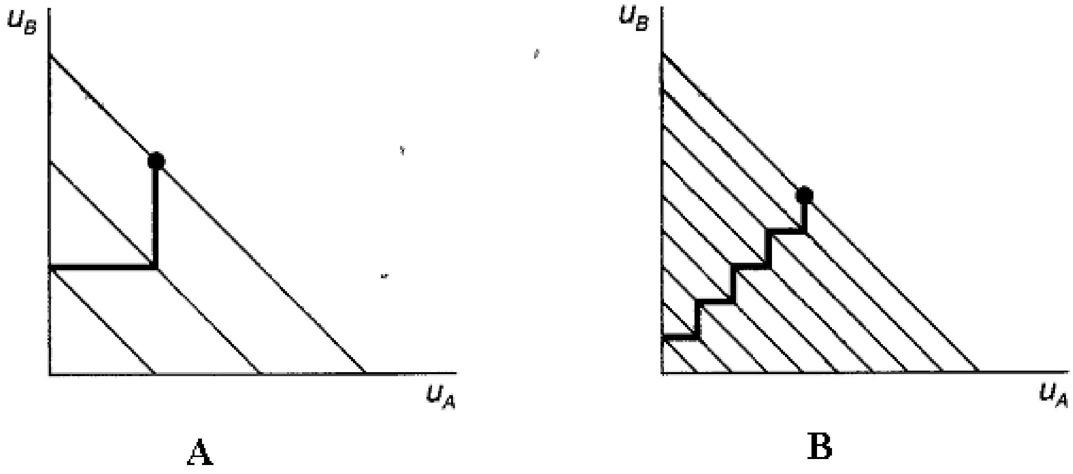


图 15.3：讨价还价博弈。粗线将各个子博弈的均衡结果连接起来。最外面的那条线上的点表示子博弈完美均衡。

自然地，我们令博弈期数无穷多，在这个无穷期的博弈中，结果将是怎样的？可以证明子博弈完美均衡分配为

$$\text{A 的收益}=\frac{1-\beta}{1-\alpha\beta}$$

$$\text{B 的收益}=\frac{\beta(1-\beta)}{1-\alpha\beta}。$$

注意，如果$\alpha = 1$且$\beta < 1$，那么选手 A 得到了全部收益，这与福音书（Gospels）所说的下列准则是一样的：“让耐心得到它的[子博弈]完美结果吧。”（James 1:4）。

15.11 重复博弈和子博弈完美

子博弈完美的思想能够删除下列这样的纳什均衡，在该纳什均衡中选手们的威胁行为是不可信的，即执行威胁计划不符合选手的利益。例如，我们在前面介绍过的惩罚策略不是子博弈完美均衡。如果选手 A 背离了（合作，合作）路径，那么选手 B 选择永远背叛未必符合 B 的利益。合理的做法是 B 对 A 的背叛行为进行某种程度的惩罚，但是永远惩罚似乎过于极端了。

一个不那么严厉的惩罚策略称为以牙还牙（tit-for-tat）：在第一步选择合作，在以后各

步选择你的对手在上一步选择的策略。在这种情形下，背叛者得到了惩罚，但只受到一次惩罚，只要他能“改邪归正”。在这个意义上，以牙还牙策略是个“宽容性的”策略。

尽管对于重复进行的囚犯困境博弈来说，惩罚性的策略不是子博弈完美的，但是存在着能使得合作解成为子博弈完美的策略。这样的策略不大容易进行描述，但是它们有着西点军校的准则特征：每个选手一致同意惩罚背叛者，而且也惩罚那些不惩罚背叛者的人，等等。你不惩罚背叛者你就会受到惩罚这个事实，使得执行惩罚成为子博弈完美。

不幸的是，类似策略也支持重复囚犯困境博弈的很多其它结果。**无名氏定理**（Folk Theorem）断言：在重复一次性博弈（one-shot game）中，几乎所有的效用分配都可能是重复博弈的均衡。

均衡解数量过多是个麻烦事。一般来说，策略空间越大，均衡解越多，这是因为选手对背叛者的“威胁”报复方法更多。为了删除“不合意的”均衡，我们需要找到删除策略的准则依据。一个自然而然的准则是删除“过于复杂的”策略。尽管在这个方面已取得了一些进展，但复杂性是个难以描述的概念，很难提出令人十分满意的定义。

15.12 不完全信息博弈

直到现在，我们考察的都是完全信息博弈。特别地，我们假设每个选手都知道其他选手的收益，每个选手都知道其他选手知道他知道，等等。在很多情形下，这不是个合理的假设。如果某个选手不知道其他选手的收益，那么纳什均衡就没有多少意义。然而，Harsanyi（1967）提出了一种考察不完全信息博弈的方法，这种方法能够系统地分析博弈的性质。

这种方法的关键是将一个选手对其他选手的不确定性归纳为一个变量，称为选手的**类型**（type）。例如，某个选手可能不知道另外一个选手对某种商品的评价，或者不知道另外一个选手的风险态度等等。每一类型的选手被视为一个不同的选手，每个选手在不同选手类型上有着先验概率分布。

于是，这种博弈的一个**贝叶斯-纳什均衡**（Bayes-Nash equilibrium）是指每个选手类型的一个策略集，该策略集要能使得在给定其他选手选择的策略条件下，每个选手类型的期望收益达到最大。这个定义在本质上与纳什均衡的定义相同，唯一的区别是这里增加了其他选手类型的不确定性。每个选手知道其他选手从一组可能的类型中进行选择，但是不知道他具体选择哪种类型。注意，为了完整描述均衡，我们必须有关于所有选手类型的一组策略，而不能只有选手在具体情形下实际选择的类型，这是因为每个单个选手不知道其他选手实际选择的类型，只能考虑所有可能性。

在同时行动的博弈中，这个均衡定义已经足够了。但是在序贯博弈中，我们应该允许选手根据他观测到的行动更新他对其他选手类型的信念。通常，我们假设这个更新过程是与

贝叶斯法则相一致的[一]。因此，如果某个选手比如 A 观察到选手 B 选择了某个策略 s，选手， A 要确定策略 s 被各种选手类型选中的概率，从而更新他对选手 B 选择哪个类型的信念。

例：密封拍卖

考虑某商品的一个简单密封拍卖，假设有两个竞价人（博弈选手）。每个选手都不知道对手的报价，因此报价是独立进行的，商品将由报价最高的选手获得。每个选手知道他对商品的评价 v，但不知道对手对商品的评价。然而，每个选手知道对手对商品的评价在 0 和 1 之间均匀分布。（而且每个选手知道每个选手相信这一点，等等。）

在这个博弈中，选手类型就是他对商品的评价。因此，该博弈的贝叶斯-纳什均衡是个函数 $b(v)$， $b(v)$ 的意思是说对于选手类型 v 来说，最优报价为 b 。给定这个博弈的对称性质，我们寻找下面这样的均衡，在这个均衡中每个选手选择同样的策略。

自然可以想到函数 $b(v)$ 是严格递增的，也就是说对商品的评价越高，报价越高。因此，我们可以令 $V(b)$ 为 $b(v)$ 的反函数，这样 $V(b)$ 就给出了报价为 b 的选手对商品的评价 V 。当某个选手报出某个具体的 b，他中标的概率为对手报价小于 b 的概率。但这正好是对手对商品的评价小于 $V(b)$ 的概率。由于 v 在 0 和 1 之间均匀分布，对手对商品的评价小于 $V(b)$ 的概率正好为 $V(b)$ 。

因此，如果某个选手对商品的评价为 v 报价为 b，他的期望收益为

$$(v-b)V(b)+0[1-V(b)].$$

上式第一项是如果该选手报价最高（从而中标）而获得的期望消费者剩余；第二项表示未中标时他的消费者剩余为零。最优报价必定要能使得上式最大，因此

$$(v-b)V(b)-V(b)=0.$$

对于每个 v 值，这个式子确定了选手的最优报价，它是 v 的函数。由于根据假设，$V(b)$ 是描述最优报价和选手对商品评价之间关系的函数，我们必定有

$$(V(b)-b)V(b)\equiv V(b)$$

对于所有 b 恒成立。

这个微分方程的解为

$$V(b)=b+\sqrt{b^2+2C}\ ,$$

其中 C 为积分常数。（请验证一下。）为了确定这个积分常数，我们注意到当 $v=0$ 时，我们必定有 $b=0$，这是因为如果对商品的评价为 0，最优报价自然为 0。将这个关系代入上式

[一] 贝叶斯法则的讨论请参见第 11 章。

可得

$$0 = 0 + \sqrt{2C} \text{，}$$

因此，$C = 0$。由此可知，$V(b) = 2b$，或$b = v/2$是这个拍卖的一个贝叶斯-纳什均衡。也就是说，每个选手的报价若是他对商品评价的一半，则这就是个贝叶斯-纳什均衡。

我们求解的方法相当标准。在本质上，我们猜测最优报价函数是可逆的，然后推导它必定满足的微分方程。可以证明，这个报价函数拥有合意的性质。这种方法的缺陷在于它只能找到贝叶斯博弈的一个特解，然而，在理论上可能存在着很多其它解。

需要指出的是，对于我们所举的这个博弈例子来说，我们计算的解碰巧是唯一的，但这个解一般不会实际发生。特别地，在不完全信息博弈下，某些选手会试图隐藏自己的类型，因为这样做对他自己有利。例如，某选手类型可能伪装成其他类型选择同样的策略。在这种情形下，将类型和策略关联起来的函数不再是可逆的，它的分析将变得复杂的多。

15.13 贝叶斯-纳什均衡的讨论

贝叶斯-纳什均衡思想是天才之作，但是也许它太天才（精巧）了。问题在于贝叶斯-纳什均衡计算过程中涉及的推理通常太复杂了。尽管有人说完全理性的选手会根据贝叶斯-纳什理论进行博弈，这种看法也许并非不合理，但是现实生活中的选手们有能力完成这么复杂的计算吗？很多人认为不能。

此外，模型在预测方面也存在着问题。每个选手的选择主要取决于他对各种选手类型分布的信念。对于选手类型分布的不同信念导致了不同的最优行为。由于我们通常不能观测到选手关于这个分布的信念，模型的预测能力通常难以检验。Ledyard（1986）证明，对于某种模式的信念来说，几乎任何模式的博弈玩法都是贝叶斯纳什均衡。

在最原始的纳什均衡概念中，存在着对选手们的信念的一致性要求：只允许那些与最优化行为相容的信念。但是一旦我们允许不同选手类型拥有不同效用函数，一致性的思想的效力丧失了大半。几乎任何模型的行为都能与某个模式的信念相容。

注释

纳什均衡的概念来自 Nash（1951）。贝叶斯均衡的概念归功于 Harsanyi（1967）。关于讨价还价模型的更详细处理请参见 Binmore and Dasgupta（1987）。

这一章只是博弈论梗概性的介绍；大多数学生想更详细地学习博弈论。幸运的是，最近有更严谨和详细的材料可供学习。综述性的文章可参考 Aumann（1987），Myerson（1986）和 Tirole（1988）。教材可以学习 Kreps（1990），Binmore（1991），Myerson（1991），Rasmusen（1989）和 Fudenberg and Tirole（1991）。

习题

15.1 计算硬币配对博弈中所有的纳什均衡。

15.2 在有限轮重复进行的囚犯困境博弈中，我们证明了选手在每一轮都选择背叛是个纳什均衡。请证明，事实上这是个优势策略均衡。

15.3 对于下列博弈，在删除了劣势策略之后的纳什均衡是什么？

		Column Left	Column Middle	Column Right
Row	Top	3, 3	0, 3	0, 0
	Middle	3, 0	2, 2	0, 2
	Bottom	0, 0	2, 0	1, 1

15.4 在教材 15.12 节的密封拍卖例子中，如果每个选手根据自己对商品的评价 v 选择贝叶斯-纳什均衡策略，计算每个选手的期望收益。

15.5 考虑下面的博弈矩阵。

		Column Left	Column Right
Row	Top	a, b	c, d
	Bottom	e, f	g, h

（a）如果（上，左）是个优势策略均衡，那么 $a,\ldots,h$ 必须满足什么样的不等式？

（b）如果（上，左）是个纳什均衡，上面哪个不等式必须得到满足？

（c）如果（上，左）是个优势策略均衡，它一定是个纳什均衡吗？

15.6 两个加利福尼亚州的年青人在玩懦夫博弈（play Chicken）。比尔和泰德各自驾车迎面猛开。每个人都有两个策略：不转向（相撞）或转向（不相撞）。如果一个选手选择了转向，他就是懦夫；如果两个选手都选择了转向，他们都是懦夫。然而，如果他们都选择不转向，他们都会身亡。该博弈的收益矩阵如下：

		Column Left	Column Right
Row	Top	-3,-3	2,0
	Bottom	0,2	1,1

（a）找到所有的纯策略均衡。

（b）找到所有的混合策略均衡。

（c）两个选手都活命的概率为多大？

15.7 在一个重复进行的对称的双头垄断博弈中，如果它们都选择生产能使得他们联合利润最大的产量水平，那么每个企业得到的收益为 π_j；如果它们选择生产古诺产量水平，每个企业得到的收益为 π_c。如果一方选择联合利润最大的产量水平，另一方能得到的最大收益为 π_d。贴现率为 r。如果任何一方偏离了联合利润最大化策略，它们采取惩罚策略即生产古诺产量。r 最多能为多大？

15.8 考虑以下博弈。

		Column Left	Column Middle	Column Right
Row	Top	1, 0	1, 2	2,-1
Row	Middle	1, 1	1, 0	0,-1
Row	Bottom	-3,-3	-3,-3	-3,-3

（a）行选手的哪个策略是严格劣势的（无论列选手选择什么样的策略）？

（b）行选手的哪个策略是弱劣势的？

（c）列选手的哪个策略是严格劣势的？

（d）如果我们删除了列选手的劣势策略，行选手的策略中存在着弱劣势策略吗？

15.9 考虑以下的合作博弈。

		Column Left	Column Right
Row	Top	2'2	-1,-1
Row	Bottom	-1,-1	1,1

（a）计算该博弈的所有纯策略均衡。

（b）存在比任何其他策略都占优势的纯策略吗？

（c）假设行选手首先行动，他承诺选择上或下。你在上面描述的策略还是纳什均衡吗？

（d）这个博弈的子博弈完美均衡是什么？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 16 章：寡头

曹乾　译

(东南大学　caoqianseu@163.com)

16 寡头

寡头市场由少数几家企业组成的市场，对寡头的研究主要是关注市场上企业之间的相互作用。寡头的现代研究几乎全部建立在博弈论基础之上的。当然，这样的发展相当自然。很多早期的关于策略性市场互动的非正式描述，若使用博弈论工具，可以说得很清楚。在本章，我们主要分析寡头理论，尽管我们关注的问题不限于这一方面。

16.1 古诺均衡

我们先分析古诺均衡这个经典模型，我们在上一章已提到过该模型。考虑两个生产某同质产品的企业，它们的产量分别为y_1和y_2，因此总产量为$Y=y_1+y_2$。和该总产量相伴的（反需求函数）为$p(Y)\equiv p(y_1+y_2)$。企业i的成本函数为$c_i(y_i)$，其中$i=1,2$。

企业 1 的利润最大化问题为

$$\max_{y_1}\pi_1(y_1,y_2)=p(y_1+y_2)y_1-c_1(y_1)$$

显然，企业 1 的利润取决于企业 2 选择的产量y_2，为了制定合理决策企业 1 必须预测企业 2 的产量决策。这正好是抽象博弈要考虑的因素——每个选手必须猜测对方的选择。因此，自然可将古诺模型看作一次性博弈：企业i的收益就是它自己的利润，企业i的策略空间就是它可以生产的可能的产量水平。于是，一个（纯策略）纳什均衡就是一组产量(y_1^*,y_2^*)，这组产量能使得：给定每个企业对其他企业选择的信念，它选择的是使其自身利润最大化的产量水平；而且每个企业关于其他企业选择的信念事实上也是正确的。

假设每个企业都有内部最优解，这意味着古诺均衡必定满足下列一阶条件：

$$\frac{\partial\pi_1(y_1,y_2)}{\partial y_1}=p(y_1+y_2)+p'(y_1+y_2)y_1-c_1'(y_1)=0$$

$$\frac{\partial\pi_2(y_1,y_2)}{\partial y_2}=p(y_1+y_2)+p'(y_1+y_2)y_2-c_2'(y_2)=0$$

每个企业的二阶条件为

$$\frac{\partial^2\pi_i}{\partial y_i^2}=2p'(Y)+p''(Y)y_i-c_i''(y_i)\le 0,\qquad i=1,2$$

其中$Y=y_1+y_2$。

企业 1 的一阶条件决定了企业 1 的产量最优选择，它是企业 1 关于企业 2 产量选择的信念的函数；这种关系称为企业 1 的**反应曲线**（reaction curve）:它描述的是，给定企业 1 关于企业 2 产量选择的各种信念，企业 1 如何反应。

假设充分的正则性，企业 1 的反应曲线 $f_1(y_2)$ 隐性地由下式定义

$$\frac{\partial \pi_1(f_1(y_2), y_2)}{\partial y_1} \equiv 0.$$

为了确定企业 1 的最优产量选择如何随它对企业 2 产量的信念变化而变化的，我们对上面的等式微分求出 $f_1'(y_2)$：

$$f_1'(y_2) = -\frac{\partial^2 \pi_1 / \partial y_1 \partial y_2}{\partial^2 \pi_1 / \partial y_1^2}.$$

根据二阶条件可知，和往常一样，上式分母为负，因此反应曲线的斜率由混合偏导的符号决定。容易看出（$\partial \pi_1 / \partial y_1$ 对 y_2 求导），这个混合偏导的表达式为

$$\partial^2 \pi_1 / \partial y_1 \partial y_2 = p'(Y) + p''(Y) y_1$$

如果反需求曲线是凹的，或者至少不是“非常”凸，这个表达式将是负的，这表明企业的古诺反应曲线的斜率一般为负。图 16.1 就是一个典型情形。

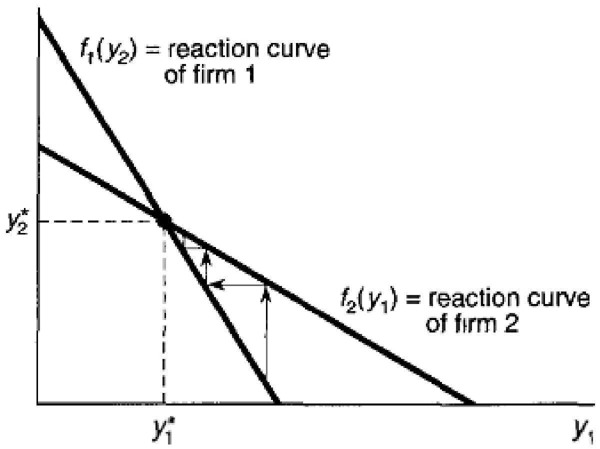


图 16.1：反应曲线。两条反应曲线的交点就是一个古诺-纳什均衡。

在下面我们将看到，寡头互动的很多重要性质取决于反应曲线的斜率，而这个斜率由取决于利润关于两个选择变量的混合偏导。如果选择变量是产量，这个混合偏导的符号“自然”为负。在这种情形下，我们说 y_1 和 y_2 是策略性替代的。如果这个混合偏导为正，这就是一个策略性互补的情形。下面我们看看它们之间的区别。

系统的稳定性

尽管我们已经认真地强调了古诺博弈的一次性博弈的性质，古诺自己却认为它是一个动态情形。这个模型的确存在着一种自然的（尽管多少有些让人怀疑）的动态解释。假设我们考虑一个学习过程，在这个过程中，每个企业在观察对方的实际产量选择后完善自己的信

念。

给定时期 0 的任意一对产量(y_1^0, y_2^0)，企业 1 猜测企业 2 在时期 1 将继续选择产量y_2^0，因此他会选择与其预测相符的利润最大化的产量即$y_1^1 = f_1(y_2^0)$。企业 2 观察到这个产量y_1^1并推测企业 1 将维持在y_1^1产量水平，因此企业 2 在时期 2 会选择$y_2^2 = f_1(y_1^1)$。一般来说，企业i在时期t的产量选择为$y_i^t = f_i(y_j^{t-1})$。

这个关于产量的差分方程勾勒出了两个企业的产量选择的变化轨迹，它是“蛛网”状的，如图 16.1 所示。在这个图中，企业的反应曲线比企业 2 的陡峭，蛛网向古诺-纳什均衡收敛。我们因此说这种均衡是**稳定的**（stable）。如果企业 1 的反应曲线比企业 2 的平坦，那么均衡将是**不稳定的**（unstable）。

我们还可以这么想，企业假定对方的产量固定不变，然后相应沿着利润增加的方向调整自己的产量。这样的想法让我们得到了一个与上述模型稍微不同的动态模型。这就是具有下列形式的**动态系统**（dynamical system）：

$$\frac{dy_1}{dt} = \alpha_1\left[\frac{\partial \pi_1(y_1, y_2)}{\partial y_1}\right]$$

$$\frac{dy_2}{dt} = \alpha_2\left[\frac{\partial \pi_2(y_1, y_2)}{\partial y_2}\right]$$

这里的参数$\alpha_1 > 0$和$\alpha_2 > 0$表示调整速度。

这个动态系统的局部稳定的充分条件为

$$\begin{vmatrix} \dfrac{\partial^2 \pi_1}{\partial y_1^2} & \dfrac{\partial^2 \pi_1}{\partial y_1 \partial y_2} \\ \dfrac{\partial^2 \pi_2}{\partial y_1 \partial y_2} & \dfrac{\partial^2 \pi_2}{\partial y_2^2} \end{vmatrix} > 0. \qquad (16.1)$$

这个条件也“几乎”是稳定性的必要条件；问题在于这个行列式可能为零即使该动态系统是局部稳定的。为了不将问题复杂化，我们不打算考虑这些边界情形。

这个行列式条件在推导比较静态结果时很有用。然而，应该强调这个假定的调整过程是相当特别的。每个企业期望对方维持产量不变，尽管它自己可以调增产量。这种不相容（inconsistent）的信念是博弈论所厌恶的。问题在于一次性的古诺博弈不能用动态方法解释；为了分析多时期博弈的动态特征，你应该使用重复博弈分析。尽管存在着这种异议，有人认为这种原始的动态模型多少都得到了经验的验证。企业为了获知市场对它们决策的反应，它们可能需要进行实验，因此可将这种特定动态调整过程视为对学习过程的描述。

16.2 比较静态

假设a是使企业 1 利润函数移动的某个参数。古诺均衡可用下列条件描述

$$\frac{\partial \pi_1(y_1(a), y_2(a), a)}{\partial y_1} = 0$$

$$\frac{\partial \pi_2(y_1(a), y_2(a))}{\partial y_2} = 0.$$

上面这两个式子关于 a 微分可得

$$\begin{pmatrix} \frac{\partial^2 \pi_1}{\partial y_1^2} & \frac{\partial^2 \pi_1}{\partial y_1 \partial y_2} \\ \frac{\partial^2 \pi_2}{\partial y_1 \partial y_2} & \frac{\partial^2 \pi_2}{\partial y_2^2} \end{pmatrix} \begin{pmatrix} \frac{\partial y_1}{\partial a} \\ \frac{\partial y_2}{\partial a} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 \pi_1}{\partial y_1 \partial a} \\ 0 \end{pmatrix}$$

应用克莱姆法则可得

$$\frac{\partial y_1}{\partial a} = \frac{\begin{vmatrix} -\frac{\partial^2 \pi_1}{\partial y_1 \partial a} & \frac{\partial^2 \pi_1}{\partial y_1 \partial y_2} \\ 0 & \frac{\partial^2 \pi_2}{\partial y_2^2} \end{vmatrix}}{\begin{vmatrix} \frac{\partial^2 \pi_1}{\partial y_1^2} & \frac{\partial^2 \pi_1}{\partial y_1 \partial y_2} \\ \frac{\partial^2 \pi_2}{\partial y_1 \partial y_2} & \frac{\partial^2 \pi_2}{\partial y_2^2} \end{vmatrix}}$$

上式分母的符号是由（16.1）中的稳定性条件决定的；我们假设它是正的。分子的符号取决于

$$(-\frac{\partial^2 \pi_1}{\partial y_1 \partial a})(\frac{\partial^2 \pi_2}{\partial y_2^2})$$

的符号。上式中的第二项是负的，这由利润最大化的二阶条件可知。因此，

$$\frac{\partial y_1}{\partial a} \text{的符号和} \frac{\partial^2 \pi_1}{\partial y_1 \partial a} \text{的符号相同}$$

这个条件是说，为了确定利润变动对均衡产量的影响，我们只要计算出混合偏导 $\partial^2 \pi_i / \partial y_i \partial a$ 即可。

我们将这个结果应用于双头垄断模型。假设 a 等于边际成本（常数），利润函数为

$$\pi_1(y_1, y_2, a) = p(y_1 + y_2) y_1 - a y_1.$$

则 $\partial^2 \pi_1 / \partial y_1 \partial a = -1$。这表明增加企业 1 的边际成本将降低它的古诺均衡产量。

16.3 几个企业

即使有 n 个企业，古诺模型的味道还是一样的。在这种情形下，企业 i 的一阶条件变为

$$p(Y)+p'(Y)y_i=c_i'(y_i) \quad (16.2)$$

其中$Y=\sum_i y_i$。将（16.2）变形可得

$$p(Y)\left[1+\frac{dp}{dY}\frac{y_i}{p}\right]=c_i'(y_i).$$

令$s_i=y_i/Y$表示企业i的产量占行业产量的份额，我们可写成

$$p(Y)\left[1+\frac{dp}{dY}\frac{Y}{p}s_i\right]=c_i'(y_i),\text{或}$$

$$p(Y)\left[1+\frac{dp}{dY}\frac{s_i}{\varepsilon}\right]=c_i'(y_i), \quad (16.3)$$

其中ε为市场需求弹性。

最后这个等式说明古诺模型在某种意义上“介于”寡头垄断和完全竞争情形之间。如果$s_i=1$，我们就正好得到了垄断条件；当s_i趋于零，从而每个企业的市场份额无穷小时，古诺均衡接近于完全竞争均衡[一]。

（16.2）和（16.3）有一些特殊情形，我们可以使用这些特殊情形构建一些例子。首先，假设每个企业的边际成本都是不变的，都为c。于是将（16.2）的左右两边对于所有n个企业相加，可得

$$np(Y)+p'(Y)Y=\sum_{i=1}^{n}c_i.$$

因此，行业总产量仅取决于边际成本之和，而不是取决于边际成本在企业之间是怎样分布的[二]。如果所有企业的边际成本都为常数c，则在对称的均衡中$s_i=1/n$，我们可以将这个式子写为

$$p(Y)\left[1+\frac{1}{n\varepsilon}\right]=c, \quad (16.4)$$

如果ε也为常数，这个式子表明价格是边际成本的固定加成（markup）。在这种情形下，显然当$n\to\infty$时，价格必定接近边际成本。

福利

我们已经知道垄断行业生产的产量水平是无效率的，因为在这个产量上价格大于边际成

[一] 实际上，你多少应该对这样不严密的表达保持警惕，因为这依赖于每个企业的份额是怎样趋于零的。Novshek（1980）提供了一个一致性的说明。

[二] 当然，我们前提是存在内部解。如果企业的边际成本差异很大，均衡时某些企业会不愿意生产。

本。古诺行业也是这样的。借助图形说明古诺行业对产量的扭曲，实际上就是问古诺行业最大化的是什么。

我们已经知道，需求曲线下方的面积$U(Y)=\int_0^Y p(x)dx$，在某些情形下可以合理衡量总收益。使用这个定义可以证明，边际成本不变的对称古诺均衡的产量水平可使下式最大：

$$W(Y)=[p(Y)-c]Y+(n-1)[U(Y)-cY].$$

证明很简单，只要将这个式子关于Y微分，并且注意到微分后的结果满足（16.4）即可。（我们假设相关的二阶条件是满足的。）

一般来说，我们希望一个行业能使得效用与成本之差达到最大。完全竞争行业的确做到了这一点，而垄断行业最大化的是利润。古诺行业最大化的是这两个目标的加权平均，其中权数取决于企业的数量。当n增加时，给与效用与成本之差这个社会目标的权重越来越大，而给与利润这个私人目标的权重越来越小。

16.4 伯特兰均衡

上一节介绍的古诺模型在很多方面都是个让人喜欢的模型，但是它只是寡头垄断行为的一种可能模型。古诺模型将产量作为企业的策略空间，但是自然地我们也可以考虑如果寡头垄断企业选择价格作为相关策略变量，将会出现什么样的情形？这种模型称为伯特兰（Bertrand）寡头垄断模型。

现在我们假设有两个企业，它们的边际成本分别为常数c_1和c_2，它们面对的市场需求曲线为$D(p)$。为了更具体一些，假设$c_2>c_1$。我们假设这两个企业生产的产品是同质的，因此企业 1 的需求曲线为

$$d_1(p_1,p_2)=\begin{cases}D(p_1) & 若p_1<p_2\\ D(p_1)/2 & 若p_1=p_2\\ 0 & 若p_1>p_2\end{cases}$$

也就是说，企业 1 相信如果它的产品定价小于企业 2 的，它能占有整个市场。当然，企业 2 也会这么想。

这个博弈的纳什均衡是什么？假设企业 1 的定价p_1大于c_2。这不可能是个均衡。为什么？如果企业 2 预期到企业 1 将会做出这样的选择，企业 2 会选择位于p_1和c_2之间的p_2，即$c_2<p_2<p_1$。这将使得企业 1 的利润为零而企业 2 的利润为正。类似地，在任何低于c_2的价格上，企业 1 都没有做到最好（“leaving money on the table”）。在任何这样的价格上，企业 2 会选择生产零单位产品，但企业 1 能增加自己的利润，方法是稍微提高一下自己产品的价格。

因此，这个博弈的一个纳什均衡是，企业 1 将价格定在$p_1=c_2$水平，生产$D(c_2)$单位

产品，而企业 2 将价格定在 $p_2 \geq c_2$ 水平，生产零单位产品。

在双头垄断行业中，企业的定价等于边际成本。这个结论多少有些不符合我们的直觉。问题部分出在伯特兰博弈是个一次性博弈：选手们选择自己的价格，然后博弈结束。这通常不是现实生活中的市场做法。

理解伯特兰模型的一种方法是将其视为一个竞争性的（拍卖）竞价模型。每个企业提交自己的密封报价，给出它愿意满足所有消费者的价格；拍卖人公开各个企业的报价，报价最低者胜出。按照这种方式理解，伯特兰结果不再那么让人费解。我们都知道密封报价是一种诱使企业激烈竞争的一种好方法，即使企业数量很少。

到目前为止，我们一直假设每个企业的固定成本都为零。现在我们放松这个假设，考虑：如果每个企业的固定成本 $K > 0$，将会出现什么样的情形。我们假设每个企业总是有权关门停业，即生产零单位产品，从而成本为零。在这种情形下，上面介绍的逻辑立即产生了伯特兰均衡：只要企业 1 的利润非负，均衡价格等于企业 2 的边际成本（企业 2 的边际成本比企业 1 高）。如果企业 1 的利润为负，那么不存在纯策略均衡。

然而，博弈通常存在混合策略均衡，事实上这样的解通常能够明确计算出。在混合策略均衡中，每个企业在其他企业可能选择的价格上有着一个概率分布，每个企业选择自己的概率分布以使得自己的期望利润最大。在伯特兰模型中，混合策略似乎是不合理的；然而，我们已说过问题所在：这是由于我们假设该博弈是一次性的。即使对于重复博弈，我们也可以将混合策略解释为一种“销售”策略：小镇零售市场上的每个商店将自己的价格随机化，从而在任何给定的一个星期内，某家商店的价格为小镇最低，从而捕获所有消费者。然而，每个星期胜出的商店通常都是不同的。

例：一个销售模型

我们计算双头垄断销售模型中的混合策略均衡。为简单起见，假设每个商店的边际成本都为零，固定成本都为 k。消费者有两类：知情消费者和不知情消费者。**知情消费者**（informed consumers）知道商店索要的最低价格，不知情消费者（uniformed consumers）随机选择一家商店。假设有 I 个知情的消费者和 $2U$ 个不知情的消费者。因此，每个商店在每个阶段铁定能得到 U 个不知情的消费者，如果它想得到知情的消费者，它的价格必须为最低价格。每个消费者的保留价格为 r。

我们仅考虑对称性的均衡，其中每个企业使用相同的混合策略。假设 $F(p)$ 为均衡策略的累积分布函数；也就是说，$F(p)$ 是商店选定的价格小于或等于 p 的概率。令 $f(p)$ 表示相应的概率密度函数，我们假设它是连续函数，因为这样我们可以忽略两家商店价格相等的概率的情形[一]。

[一] 可以证明在均衡时，零概率就是这样的。

给定这个假设，当某个企业设定了价格 p ，与其相关的事件恰好有两个：要么它的价格最低，这个事件的概率为 $1-F(p)$ ；要么它不是价格最低，这个事件的概率为 $F(p)$ 。如果它的价格最低，它得到的收入为 $p(U+I)$ ；如果它的价格不是最低，它得到的收入为 pU 。在这两种情形下，它的固定成本都为 k 。因此，商店的期望利润 $\bar{\pi}$ 可以写为

$$\bar{\pi}=\int_0^\infty [p(1-F(p))(I+U)+pF(p)U-k]f(p)dp.$$

现在注意下面的简单结论：**在均衡混合策略中，商店实际索要的每个价格必定产生相同的期望利润**。如若不然，必定有些价格的期望利润大于另外一些价格的期望利润，这样商店可以增加选择前者的概率，从而增加期望利润。

这意味着我们必定有

$$p(I+U)(1-F(p))+pF(p)U-k=\bar{\pi},$$

或者解出 $F(p)$ ，

$$F(p)=\frac{p(I+U)-k-\bar{\pi}}{pI} \tag{16.5}$$

剩下的任务是确定 $\bar{\pi}$ 。商店索要的价格小于或等于 r 的概率为 1，因此我们必定有 $F(r)=1$ 。由此可得 $\bar{\pi}=rU-k$ ，将其代入（16.5）式可得

$$F(p)=\frac{p(I+U)-rU}{pI}$$

令 $u=U/I$ ，我们可以将上式写为

$$F(p)=1+u-\frac{ru}{p}.$$

当 $\bar{p}=ru/(1+u)$ 时，这个式子等于零，因此当 $p\le\bar{p}$ 时，$F(p)=0$ ；当 $p\ge r$ 时，$F(p)=1$ 。

16.5 互补与替代

在我们的寡头垄断的两个模型（即古诺模型和伯特兰模型）中，我们一直假设两个企业生产的产品是完全替代的。然而，我们可以放松这个假设，这样做的好处是我们可以指出古诺均衡和伯特兰均衡之间存在着漂亮的对偶关系。这一点在线性需求函数下最容易看清楚，尽管它的成立不依赖于线性需求函数，是普遍成立的。令消费者的反需求函数为

$$\begin{aligned}p_1&=\alpha_1-\beta_1 y_1-\gamma y_2\\ p_2&=\alpha_2-\gamma y_1-\beta_2 y_2\end{aligned}$$

注意，在这里，“交叉价格效应是对称的”，这是因为良好表现的消费者需求函数要求它是这样的。

相应的直接需求函数为

$$y_1 = a_1 - b_1 p_1 - c p_2$$
$$y_2 = a_2 + c p_1 - b_2 p_2$$

其中参数 $a_1, a_2, \ldots$ 是参数 $\alpha_1, \alpha_2, \ldots$ 的函数。

当 $\alpha_1 = \alpha_2$ 和 $\beta_1 = \beta_2 = \gamma$ 时，商品是完全替代的。当 $\gamma = 0$，不同企业的商品市场是独立的。一般来说，$\gamma^2 / \beta_1\beta_2$ 可以用作描述产品差异程度的指数。当它为零时，市场是独立的；当它为 1 时，商品是完全替代的。

为简单起见，假设边际成本为零。那么如果企业 1 是古诺竞争者，它最大化的是

$$(\alpha_1 - \beta_1 y_1 - \gamma y_2) y_1,$$

如果它是伯特兰竞争者，它最大化的是

$$(a_1 - b_1 p_1 - c p_2) p_1.$$

注意，上面两个式子在结构上非常相似：将第一个式子中的 $\alpha_1, \beta_1, \gamma$ 分别替换为 $a_1, b_1, -c$ 就得到了第二个式子。因此，产品可替代情形（此时 $\gamma > 0$）下的古诺均衡和产品可互补情形（此时 $c < 0$）下的伯特兰均衡的数学结构是相同的。

这个“对偶性”能让我们一次证明两个定理：当我们计算出了涉及古诺竞争的结果时，只要将希腊字母替换为罗马字母，我们就能得到伯特兰竞争的结果。

例如，我们在前面已经知道，反应曲线的斜率在确定古诺模型中的比较静态结果时有着重要作用。在这里讨论的异质性商品的情形下，企业 1 的反应曲线是下列最大化问题的解：

$$\max_{y_1} [\alpha_1 - \beta_1 y_1 - \gamma y_2] y_1$$

容易得到

$$y_1 = \frac{\alpha_1 - \gamma y_2}{2\beta_1}.$$

将希腊字母变为罗马字母，伯特兰模型中的反应曲线为

$$p_1 = \frac{a_1 + c p_2}{2 b_1}.$$

注意，古诺模型中的反应曲线的斜率和伯特兰模型中的反应曲线斜率的符号是相反的。我们已经知道，古诺模型中的反应曲线通常是向下倾斜的，这意味着伯特兰模型中的反应曲线通常是**向上倾斜的**。这非常符合我们的直觉。如果企业 2 增加它的产量，那么企业 1 通常想要降低自己的产量，目的在于迫使价格上升。然而，如果企业 2 提高了自己的价格，企业 1 通常发现：为了与价格上升相匹配，提高自己的价格对自己有利。

另外一种解释方法是使用策略性互补和策略性替代的概念。企业们的产量是策略性替代

的，这是因为企业 2 增加产量 y_2 之后，如果企业 1 也增加产量 y_1，那么企业 1 的利润就会降低。然而，企业 2 提高价格 p_2 之后，如果企业 1 也提高价格，企业 1 的利润就会增加。由于交叉偏导数的符号不同，反应曲线的斜率符号也不同。

16.6 产量领导

双头垄断的另外一个模型是**产量领导模型**（quantity leadership model），也称为**斯坦科尔伯格模型**（Stackelberg model）。这个模型在本质上是个两阶段模型，其中充当领导者的企业先行动。另外一个企业即追随者观察领导者的产量选择，从而选择自己的最优产量水平。使用上一章的博弈术语来说，数量领导模型是个序贯博弈。

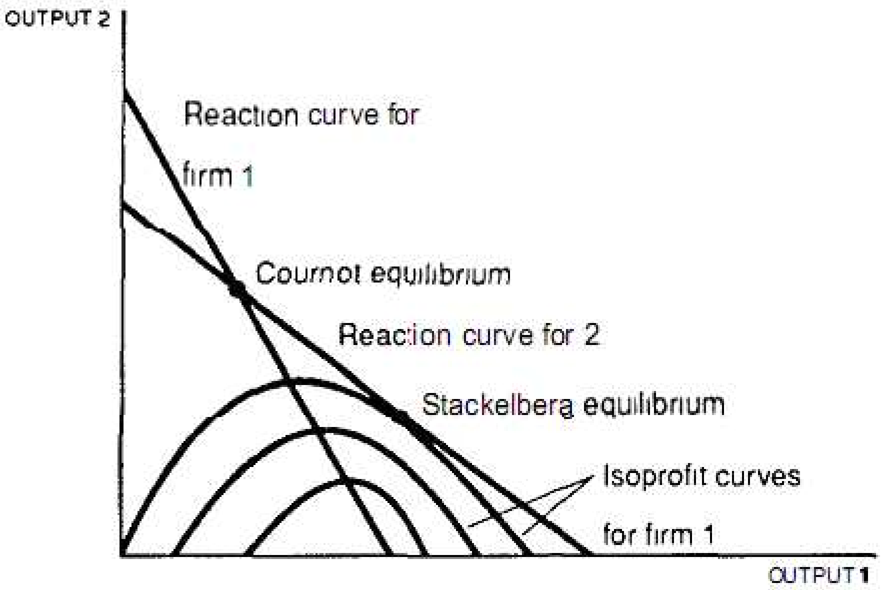


图 16.2：古诺均衡和斯坦科尔伯格均衡的比较。纳什均衡发生在两个企业的反应曲线的交点处。斯坦科尔伯格均衡发生在一个企业的反应曲线与另外一个企业的等利润线相切之处。

和以前一样，我们“逆向”求解。假设企业 1 是领导者，企业 2 是追随者。那么企业 2 的问题是直观的：给定企业 1 的产量，企业 2 希望使得它自身的利润 $p(y_1+y_2)y_2-c_2(y_2)$ 最大化。注意 $p(y_1+y_2)$ 是需求函数。这个最大化问题的一阶条件为

$$p(Y)+p'(Y)y_2=c_2'(y_2). \tag{16.6}$$

这正是我们在前面描述的古诺条件，和以前一样，我们可以使用这个式子推导出企业 2 的反应函数 $f_2(y_1)$。

回到该博弈的第一阶段。现在企业 1 想选择它自己的产量水平，并且它知道企业 2 如何对企业 1 的选择做出反应。因此，企业 1 希望使得它的利润最大

$$p(y_1+f_2(y_1))y_1-c_1(y_1).$$

这个最大化问题的一阶条件为

$$p(Y)+p'(Y)[1+f_2'(y_1)]=c_1'(y_1). \tag{16.7}$$

联立（16.6）和（16.7）式，我们就可以求得企业 1 和 2 的最优产量水平。

图 16.2 描述了斯坦科尔伯格均衡。在此图中，企业 1 的等利润线表示在这条线上的任何产量都能产生相等的利润。位置较低的等利润线对应着较高的利润水平。企业 1 想在企业 2 的反应曲线上能使得企业 1 的利润最大的那点上生产，如图所示。

我们如何比较斯坦科尔伯格均衡和古诺均衡？从显示偏好原理立即可知：由于在斯坦科尔伯格均衡中，企业 1（领导者）在企业 2（追随者）的反应曲线上选择它（企业 1）的最优点，而在古诺均衡情形下，企业 1 在企业 2 的反应曲线上选择的点是“任意的”，因此，斯坦科尔伯格均衡下企业 1 的利润通常比古诺均衡中企业 1 的利润大。

作为领导者和作为追随者的利润有何不同？企业愿意做领导者还是追随者？这个问题有着漂亮的、一般性的结论，但需要证明。我们将在下列假设条件下分析异质性产品这样的一般情形，假设企业 1 和 2 的产品分别为 y_1 和 y_2。（当然，这些假设包含产品同质的情形，即 y_1 和 y_2 是完全替代的。）

假设 1：替代性产品。$\pi_1(y_1,y_2)$ 是 y_2 的严格递减函数，$\pi_2(y_1,y_2)$ 是 y_1 的严格递减函数。

假设 2：向下倾斜的反应曲线。反应曲线 $f_i(y_i)$ 是严格递减的函数。

<u>命题【喜欢做领导者】</u>。在假设 1 和 2 之下，企业总是弱偏好于做领导者而不是追随者。

证明。令 $(y_1^*,y_2^*)=(y_1^*,f_2(y_1^*))$ 为企业 1 为领导者时的斯坦科尔伯格均衡。首先，我们需要证明

$$f_1(y_2^*)\le y_1^*. \tag{16.8}$$

反证。假设不是，因此有

$$f_1(y_2^*)>y_1^*. \tag{16.9}$$

将上式两侧应用函数 $f_2(\cdot)$ 可得

$$f_2(f_1(y_2^*))<_1 f_2(y_1^*)=_2 y_2^* \tag{16.10}$$

不等式（1）可从假设 2 推出；等式（2）可从斯坦科尔伯格均衡的定义推知。

现在我们有下列一连串不等式

$$\pi_1(y_1^*,y_2^*)\le_1 \pi_1(f_1(y_2^*),y_2^*)<_2 \pi_1(f_1(y_2^*),f_2(f_1(y_2^*))). \tag{16.11}$$

不等式（1）可从反应函数的定义中推出；不等式（2）可从（16.10）式和假设 1 推出。根据（16.11）式，点 $(f_1(y_2^*), f_2(f_1(y_2^*)))$ 产生的利润高于 $(y_1^*, f_2(y_1^*))$ 产生的。这与我们已开始假设 $(y_1^*, f_2(y_1^*))$ 是斯坦科尔伯格均衡相矛盾，从而（16.9）不成立而（16.8）成立。

现在，我们想要的结果可立即从下列不等式推出

$$\max_{y_2} \pi_2(f_1(y_2), y_2) \geq_1 \pi_2(f_1(y_2^*), y_2^*) \geq_2 \pi_2(y_1^*, y_2^*).$$

不等式（1）可从利润最大化推出，不等式（2）可从（16.8）和假设 1 推出。上面这个不等式的左侧和右侧的项表明，对于企业 2 来说，它做领导者的利润不会小于企业 1 为领导者时的它（企业 2）的利润。■

由于人们通常认为反应曲线向下倾斜，已经企业间的商品是替代的，我们这个命题表明每个企业都偏好做领导者。然而，哪个企业真正做领导者通常要取决于历史因素，例如哪个企业先进入市场等。

16.7 价格领导

价格领导（price leadership）是一个企业制定价格而其他企业将这个价格作为给定的情形。价格领导模型和斯坦科尔伯格模型的求解方法是一样的：首先我们推导出追随者的行为，然后再推导出领导者的行为。

在异质性商品的模型中，令 $x_i(p_1, p_2)$ 表示消费者对企业 i 产品的需求。追随者（企业 2）将 p_1 视为给定的，它的策略是选择 p_2。也就是说，追随者（企业 2）的最大化问题是

$$\max_{p_2} p_2 x_2(p_1, p_2) - c_2(x_2(p_1, p_2)). \tag{16.12}$$

我们令 $p_2 = g_2(p_1)$ 表示企业 2 的反应函数，它将企业 2 的最优价格 p_2 表示为 p_1 的函数。

然后，领导者（企业 1）通过求解自己的最大化问题

$$\max_{p_1} p_1 x_1(p_1, g_2(p_1)) - c_1(x_1(p_1, g_2(p_1))).$$

找到自己的最优 p_1 值。

一种有趣的特殊情形是当这两个企业销售相同的产品。在这种情形下，如果企业 2 的销量为正，它的售价 p_2 必定等于 p_1，即 $p_2 = p_1$。对于每个价格 p_1，追随者将 p_1 视为给定的，并在此情形下选择使得利润最大化的产量 $S_2(p_1)$。因此，这种情形下的反应函数就是竞争性的供给曲线。

如果企业 1 索要价格 p_1，企业 1 将销售 $r(p_1) = x_1(p_1) - S_2(p_1)$ 单位产品。函数 $r(p_1)$ 称为企业 1 面对的**剩余需求曲线**（residual demand curve）。企业 1 选择能满足下列问题的 p_1：

$$\max_{p_1} \; p_1 r(p_1) - c_1(r(p_1)).$$

这正好是面对着剩余需求曲线 $r(p_1)$ 的那个垄断企业的利润最大化问题。

图 16.3 描述了这个问题的解。我们从市场需求曲线减去企业 2 的供给曲线，从而得到剩余需求曲线。然后我们使用 MR=MC 条件来求领导者的产量。

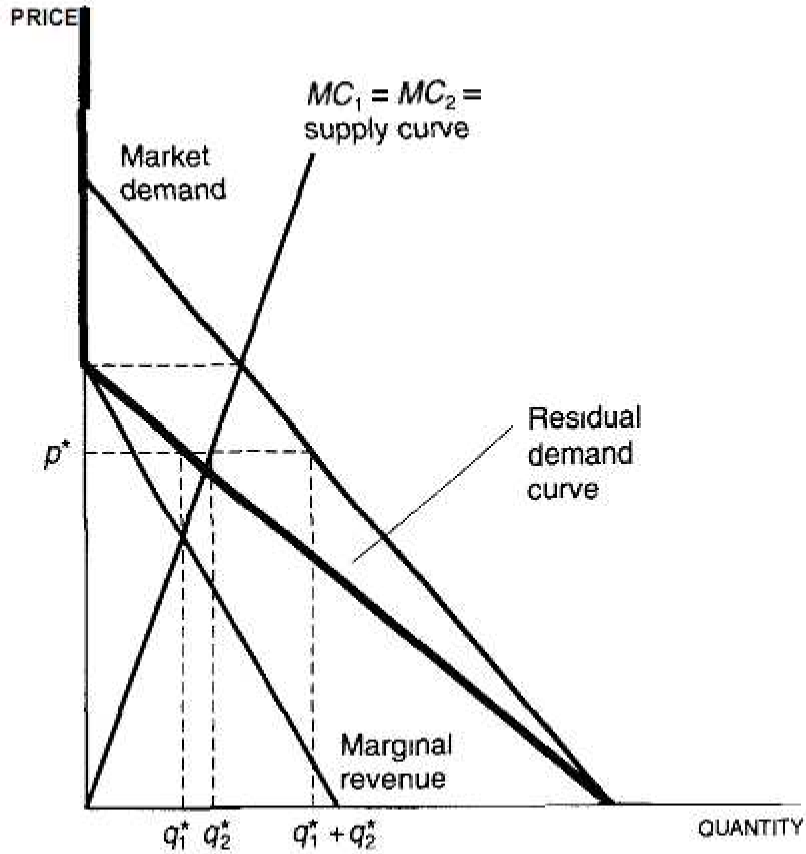


图 16.3：价格领导。企业 1 从市场需求曲线减去企业 2 的供给曲线得到剩余需求曲线。然后在这条曲线上选择利润最大的产量水平。

我们回到异质性产品情形，在这里我们对企业是喜欢做追随者还是领导者感兴趣。我们注意到：只要将 y_i 替换为 p_i，我们在上面证明的结果立即可以扩展到价格领导模型。然而，这种扩展存在着两个问题。首先，一个企业的利润未必是另外一个企业产品价格的减函数。企业 1 的利润关于企业 2 产品价格的导数为

$$\frac{\partial \pi_1(p_1, p_2)}{\partial p_2} = [p_1 - c_1'(x_1(p_1, p_2)]\frac{\partial x_1(p_1, p_2)}{\partial p_2}$$

这个导数值的符号取决于价格是否大于边际成本。可以证明，这个问题能够得到解决；在价

格领导模型中，只要反应函数是向下倾斜的，企业仍然喜欢做领导者。

然而，在价格领导模型中，反应曲线不可能是向下倾斜的。为简单起见，假设边际成本为零。那么企业 2 的反应函数必定满足一阶条件

$$p_2 \frac{\partial x_2(p_1, g_2(p_1))}{\partial p_2} + x_2(p_1, g_2(p_1)) \equiv 0.$$

根据比较静态计算可知，

$$g_2(p_1)\text{的符号与}\left[p_2 \frac{\partial^2 x_2}{\partial p_1 \partial p_2} + \frac{\partial x_2}{\partial p_1}\right]\text{的符号相同。}$$

上面方括号内的第一项可能为正也可能为负，但如果商品 1 和 2 是替代品，那么第二项的符号为正。因此，正如我们在前面指出的，在价格领导模型中，我们期望反应曲线是**向上倾斜的**。

命题 **【意见一致（Consensus）】。如果两个企业的反应函数都是向上倾斜的，那么如果一个企业偏好做领导者，则另外一个企业必定偏好做追随者**。

证明。参见 Dowrick（1986）。■

从上面的命题立即可以推知下列命题。

命题 **【喜欢做追随者（following preferred）】如果两个企业的成本函数和需求函数都相同，而且它们的反应曲线都是向上倾斜的，那么每个企业必定偏好做追随者而不是领导者。**

证明。如果一个企业偏好做领导者，根据对称性可知，另外一个企业也必定偏好做领导者。但这立即与前面的命题矛盾。■

在产品为同质的这种特殊情形下，我们还有一种证明方法。这个证明需要使用图 16.3。在这个图中，企业 1 选择 p^* 和产量水平 q_1^*。在价格为 p^* 时，企业 2 本来可以选择与企业 1 相同的产量 q_1^*，但是它没有这么做，而是选择位于企业 2 供给曲线上的产量。因此，均衡时企业 2 的利润比企业 1 的高。

在定价博弈中，为何企业偏好做追随者而不是领导者？在直觉上，原因在于在这种情形下，领导者为了维持价格必须降低产量，而追随者可以将领导者制定的价格作为既定的，并且它愿意生产多少就生产多少；也就是说，追随者可以搭乘追随者限制产量的便车。

16.8 模型分类与模型选择

我们已经讨论了双头垄断的四种模型：古诺模型、伯特兰模型、产量领导模型与价格领导模型。从博弈论的观点来看，这些模型的区别在于两个方面：首先，定义的策略空间（价格或产量）不同；其次，信息集不同：一个选手在行动时是否知道其他选手的选择。

在这四种模型中，哪一种是正确的？一般来说，这个问题没有答案；我们只能具体环

境具体分析，也就是说在选择模型时，我们需要考察特定的经济形势或产业。下面是一些有用的建议。

需要记住这些模型都是“一次性博弈”，但在现实应用中，我们通常试图模拟实时互动，即能持续很多期的产业结构。因此，我们自然要求用于模拟企业策略性变量的变量，不能立刻得到调整——这些变量一旦被选定，将持续某个时期，从而使得一次性博弈分析能在某种程度上代表现实生活中发生的经济现象。

例如，考虑伯特兰均衡。正式地说，伯特兰博弈是个一次性的博弈：两个寡头垄断企业同时定价而无需观察对方的选择。但是如果在你看到对手的价格后（在消费者看到你和你对手的价格之前），你能迅速且无成本地调整你的价格，那么伯特兰模型就不是个有吸引力的模型：只要竞争企业观察到另外一个企业的价格，它能立即作出反应，从而很有可能导致非伯特兰结果。

如果企业的产量调整起来比较缓慢，选择古诺模型似乎比较合适。在把“产量”解释为“产能”的情形下，古诺模型尤其具有吸引力。其中的思想在于，每个企业秘密地选择自己的产能，因为它们认识到一旦产能选定，它们将展开价格竞争即进行伯特兰博弈。Kreps and Scheinkman（1983）分析了这个两阶段的博弈，结果发现该博弈的结果通常是古诺均衡。我们将大致说说这两位经济学家提出的模型的简化版本。

假设在第一期，每个企业同时选择某个产量水平 y_i。在第二期，每个企业选择自己的售价。我们对这个两阶段的博弈的均衡比较感兴趣。

和往常一样，我们从第二期开始分析。在第二期，企业 i 的销量若小于 y_i 则它的边际成本为零，若它的销量大于 y_i，则它的边际成本无穷大。在均衡时，每个企业必定索要相同的价格；如若不然，一个企业索要的价格只要稍微比对手低，就对自己很有利。另外，每个企业索要的价格不会高于 $p(y_1+y_2)$。如果高于，那么只要其中一个企业稍微降低自己的价格，就能占据整个市场。最后，每个企业索要的价格不会低于 $p(y_1+y_2)$，这是因为当每个企业的销量达到产能上限时，提高价格对每个企业都有利。（这个论证的看起来相当直观，但要严格证明却非常困难。）

此处重要的结论是，当每个企业的销量达到产能上限时，那么每个企业都不会降低价格。的确，如果一个企业降价那么它将能夺走对手的所有消费者，但是由于它已将所有产品都卖光了，那些额外的消费者对它毫无用处，所以它不会降价[一]。

在知道第二期的均衡价格就是产量达到产能上限时的反需求，第一期的均衡就迎刃而解：它就是古诺均衡。因此，古诺产能竞争之后若是伯特兰价格竞争，那么其结果就是标准的古诺均衡。

[一] 然而，它的确带来了一个微妙的问题：如果一个企业得到额外的消费者，从而使它的产品供不应求后，我们必须建立一个分配规则来指明如何应付这些额外的消费者。Davidson and Deneckere（1986）证明，指明分配规则后，均衡的本质可能受到影响。

16.9 推测变量

上面介绍的价格领导博弈和产量领导博弈能进行有趣的推广。我们已经知道描述斯坦科尔伯格模型中最优产量选择的一阶条件为

$$p(Y)+p'(Y)[1+f_2'(y_1)]y_1=c_1'(y_1). \tag{16.13}$$

$f_2'(y_1)$ 这一项表示的是企业 1 关于企业 2 的最优行为如何随着 y_1 变化而变化的这种信念。

在斯坦科尔伯格模型中，这个信念等于企业 2 的实际反应函数的斜率。然而我们可以将 $f_2'(y_1)$ 这一项看作关于企业 2 如何对企业 1 的选择作出反应的任意“推测”。不妨将它称为企业 1 关于企业 2 的**推测变量**（conjectural variation），记为 v_{12}。现在一阶条件变为

$$p(Y)+p'(Y)[1+v_{12}]y_1=c_1'(y_1) \tag{16.14}$$

这种参数化的好处是不同的参数直接对应着我们在前面讨论的各种模型的一阶条件。

1）$v_{12}=0$。这是古诺模型，在这个模型中每个企业相信其他企业的选择和它的选择无关；

2）$v_{12}=-1$。这是竞争性模型，因为此时一阶条件简化为价格等于边际成本；

3）$v_{12}=$ 企业 2 的反应函数的斜率。这当然是斯坦科尔伯格模型；

4）$v_{12}=y_2/y_1$。这是合谋模型。在这种条件下，一阶条件变为行业利润最大，所以结果是合谋均衡。

这个表表明，我们在前面讨论的每个主要模型只是推测变量模型的一种特殊情形。在这个意义上，推测变量思想的作用是可以作为划分寡头垄断模型的依据。

然而，推测变量模型不是令人满意的行为模型。问题在于它把本质为静态的模型粉饰为动态的，因此是一种伪动态模型。我们在前面讨论的每个模型都是一次性模型：在古诺模型中，每个企业独立地选择产量；在斯坦科尔伯格模型中，领导者先选择产量，期望对方能作出最优反应，等等。推测变量模型表明一个企业选择某个产量，是因为它期望对方能以某种特定方式作出反应：但在一次性博弈中，对方如何做出反应？如果你想模拟动态情形，其中每个企业都能对对手的产量选择作出反应，那么你应该考察重复博弈。

16.10 合谋

截止到目前，我们描述的模型都是**非合作博弈**（non-cooperative games）。每个企业都竭力使得自己利润最大化，每个企业都独立地作出自己的决策。如果我们放松这个假设，考虑可能的合作，结果将会怎样？在一个行业中，如果该行业的企业在价格和产量的制定上能够实现某种程度的合谋，那么该行业称为**卡特尔**（cartel）。

我们自然会考虑如果两个企业选择它们的产量来实现联合利润最大化，将会出现什么样的情形。在这种情形下，这两个企业同时选择产量 y_1 和 y_2，以便使得行

业利润最大：

$$\max_{y_1,y_2} p(y_1+y_2)[y_1+y_2]-c_1(y_1)-c_2(y_2).$$

一阶条件为

$$\begin{aligned} p(y_1^*+y_2^*)+p'(y_1^*+y_2^*)[y_1^*+y_2^*]&=c_1'(y_1^*) \\ p(y_1^*+y_2^*)+p'(y_1^*+y_2^*)[y_1^*+y_2^*]&=c_2'(y_2^*). \end{aligned} \tag{16.15}$$

由于这两个式子的左侧是相同的，右侧必定相等——每个企业必定使得它们的生产边际成本相等，即 $c_1'(y_1^*)=c_2'(y_2^*)$。

卡特尔解的问题在于它不是“稳定的”。企业总有背叛的动力：生产的产量超过约定产量。为了看清这一点，考虑如果企业 1 稍微增加一些产量 dy_1 但假设企业 2 仍然生产卡特尔约定的产量 y_2^*，将会出现什么样的结果。企业 1 的利润随着 y_1 的变化而变化，企业 1 的利润在卡特尔解之处的导数为

$$\frac{\partial\pi(y_1^*,y_2^*)}{\partial y_1}=p(y_1^*+y_2^*)+p'(y_1^*+y_2^*)y_1^*-c_1'(y_1^*)=-p'(y_1^*+y_2^*)y_2^*>0$$

其中：等式的符号来自于一阶条件（16.15），而不等式的符号来自需求曲线向下倾斜这个事实。

如果一个企业认为其他企业将按照约定的产量生产，那么它自己增加产量就可以增加自己的利润，因为它卖得产品更多了。但是如果它不认为企业将按照约定的产量生产，它生产约定产量这个决策一般来说不是最优的。在这两种情形下，它都有增加产量的动机，尽可能地获得更高的利润。

这样的策略形势类似于囚犯的困境：如果你认为其他企业守约，背叛对你来说就是值得的——生产的产量超过约定产量。如果你认为其他企业不会守约，那么你生产的产量超过约定的产量，通常也是有利可图的。

为了使得卡特尔结果可行，卡特尔必须找到某种方法来稳定市场。这样的方法通常是对违背卡特尔协议的企业进行惩罚。例如，其中一个企业，比如企业 1，可能宣称：如果它发现有企业不按约定的产量生产，它自己也会增加自己的产量。这样就产生了一个有趣的问题：如果存在企业不守约，企业 1 应该增加多少产量？

我们在上一节已经知道，支持卡特尔解的推测变量 $v_{12}=y_1/y_2$。这意味着什么？假设企业 1 宣称：若果企业 2 增加产量 dy_2，那么企业 1 的反应是增加产量 $dy_1=(y_1/y_2)dy_2$。如果企业 2 相信这是个威胁，那么增加的产量 dy_2 带来的预期利润变化为

$$
\begin{aligned}
d\pi_2 &= p(y_1^*+y_2^*)dy_2 + p'(y_1^*+y_2^*)\left[dy_2+\frac{y_1^*}{y_2^*}\right]y_2^* - c_2'(y_2^*)dy_2 \\
&= \left[p(y_1^*+y_2^*)+p'(y_1^*+y_2^*)[y_1^*+y_2^*]-c_2'(y_2^*)\right]dy_2 \\
&= 0
\end{aligned}
$$

因此，如果企业 2 认为企业 1 将如此反应，那么企业 2 就不会期望背叛能带来更多的利润。

企业 1 的惩罚策略的本质在市场份额为不对称的情形下能看得更清楚。假设在卡特尔均衡解中，企业 1 的产量是企业 2 产量的 2 倍。那么为了防止企业 2 背叛，企业 1 必须威胁称它将生产的产量是企业 2 背叛后产量的 2 倍。另一方面，为了防止企业 1 背叛，企业 2 只需要宣称它将生产的产量是企业 1 背叛后产量的 1/2 即可。

尽管很形象，这类分析存在着和推测变量同样的问题：它硬生生将动态互动情形套进静态的模型之中。然而，我们将看到一个严格的动态分析涉及到的思想和这里有很多相同之处。让卡特尔结果成为可行结果的关键问题，是对背叛行为构建合适的惩罚措施。

16.11 重复寡头博弈

到目前为止我们讨论的所有博弈都是一次性博弈。但是现实市场的互动通常是实时的，因此，我们自然应该考虑重复博弈。最简单的方法是将古诺博弈想象成一个重复博弈。

处理方法类似于重复囚犯困境的分析方法（详见第 15 章）。在这种情形下，合作的结果就是卡特尔解。企业可以选择古诺产量作为惩罚措施。于是，支持卡特尔解的策略是下面这样的：选择卡特尔约定的产量，除非对方背叛；如果对方背叛，你就选择古诺产量。与囚犯困境的情形一样，这是个纳什策略集，只要贴现率不是很高。

不幸的是，这样的博弈有很多其它均衡策略：正如重复囚犯困境博弈的情形一样，几乎任何策略都可能成为纳什均衡。更为严重的是，对于**有限次**重复产量决策博弈，结果也是这样的，这与重复囚犯困境博弈不同（无限次重复囚犯困境博弈才会出现这样的结果）。

为了看清这一点，我们考虑两个相同企业之间的一个两期博弈，这两个企业的边际成本都为零。考虑企业 1 的下列策略：在第一期生产产量 y_1。如果你的对手在第一期生产的产量为 y_2，你的策略是在第二期生产古诺产量 y_1^c。如果你的对手生产的产量不是 y_2，那么你的策略是生产足够大的产量，将市场价格压低为零。

企业 2 对这个威胁的最优反应是什么？如果企业 2 在第一期生产 y_2，在第二期生产 y_2^c，那么它得到的收益为 $\pi_2(y_1,y_2)+\delta\pi_2(y_1^c,y_2^c)$。如果企业 2 在第一期生产的产量不是 y_2，比如说为 x，那么企业 2 得到的收益为 $\pi_2(y_1,x)$。因此，当

$$\pi_2(y_1,y_2)+\delta\pi_2(y_1^c,y_2^c) > \max_x \pi_2(y_1,x)$$

时，企业 2 会选择与企业 1 合作，因为利润更大。这个条件通常对一系列的 (y_1,y_2) 成立。

问题在于实际执行惩罚策略的这种威胁是不可信的：一旦第一期结束，超量生产通常不符合企业 1 的利益。换句话说，超量生产不是仅包含第二期的子博弈的均衡策略。使用第 15 章的术语来说，这个策略不是子博弈完美。

不难证明，在有限次重复进行的产量制定博弈中，唯一的子博弈完美均衡是重复一次性的古诺均衡，至少当一次性古诺均衡是唯一时是这样的。证明方法是常见的逆向归纳法：因为在最后一期，古诺均衡是唯一的结果，所以选手在第一期宣称将干预最后一期结果的威胁是不可信的，因此，“缺乏远见的”古诺均衡是唯一的选择。

自然我们会问：在有限次重复博弈中，卡特尔产量是个可持续的子博弈完美均衡吗？弗里德曼（Friedman, 1971）证明了答案是肯定的。可行的策略类似于上一章讨论的惩罚策略。令 π_i^c 表示企业 i 从一期古诺均衡得到的利润，令 π_i^* 表示企业 i 从一期卡特尔结果得到的利润。考虑企业 1 的下列策略：生产卡特尔产量水平，除非企业 2 背叛。如果企业 2 背叛，则永远生产古诺产量。

如果企业 2 相信企业 1 在某个给定时期将生产古诺产量水平，那么它的最优反应也是生产古诺产量水平。（这是古诺均衡的定义！）因此，重复古诺产量无疑是重复博弈的一个均衡。

为了看清企业 2 生产卡特尔产量水平是否为有利可图的，我们必须比较它选择背叛得到的利润现值和它选择合作得到的利润现值。令 π_2^d 表示企业 2 选择背叛得到的利润，上述条件变为

$$\pi_2^d + \delta \frac{\pi_2^c}{1-\delta} < \frac{\pi_2^*}{1-\delta}.$$

上式左侧的意思是，企业 2 在背叛当期得到的利润加上它在以后各期生产古诺产量得到的利润现值。将上式变形可得下列条件，

$$\delta > \frac{\pi_2^d - \pi_2^*}{\pi_2^d - \pi_2^c}. \qquad (16.16)$$

只要 δ 足够大即贴现率足够小，那么这个条件就能得到满足。与重复囚犯困境博弈一样，这个博弈也存在着很多其它解。

如果我们允许选手采取其它不同的惩罚措施，而不是仅局限于古诺版本的惩罚措施，那么我们就可以扩展（子博弈完美）惩罚策略的思想。例如，阿布鲁（Abreu, 1986）证明，如果在某一期选手采取惩罚措施，然后在下一期又回到古诺产量，那么这样的做法足以支持卡特尔均衡。这让我们想起了推测变量模型中的最优惩罚结果——只要一个企业能够非常迅速地惩罚另外一个企业，它就能保证对方不能从背叛中获得好处。夏普洛（Shapiro，1989）综述了重复寡头博弈的结果。感兴趣的读者可以参考这篇文献。

16.12 序贯博弈

上一节介绍的重复博弈只是一次性市场博弈的简单重复进行版本。一个企业比如企业 1 在某一期做出的选择会间接影响它的竞争对手的行为，但不会影响企业 1 自身在其他期的利润。然而，在现实中，某个时点的决策影响以后各期的产量。在某些博弈中，这类投资决策起着重要的策略性作用。

在分析这类行为的模型时，有必要区分纳什均衡和子博弈完美均衡。为了说明这种区别的重要性，我们举一个非常简单的例子。以企业进入市场为例。

考虑某个行业，该行业有两个企业准备在条件成熟时进入市场。假设进入不需要成本，而且存在某种外生的生产技术进步使得生产成本逐渐降低。令 $\pi_1(t)$ 表示如果在 t 时市场上只有一家企业情形下，该企业在 t 时赚取利润的现值。令 $\pi_2(t)$ 表示如果在在 t 时市场上有良家企业情形下，每家企业在 t 时赚取利润的现值。这种简化的利润函数掩盖了行业内的确切竞争形式；我们需要的仅是 $\pi_1(t) > \pi_2(t)$ 对于所有 t 成立，这意味着完全垄断企业的利润大于每个双头垄断企业的。

我们用图 16.4 表示这些利润流。我们假设一开始利润增长速度比利率快，从而导致贴现利润随着时间推移而增加。但是最终该行业中的技术进步速度减缓，使得利润增长速度低于利率增长速度，从而使得利润现值下降。

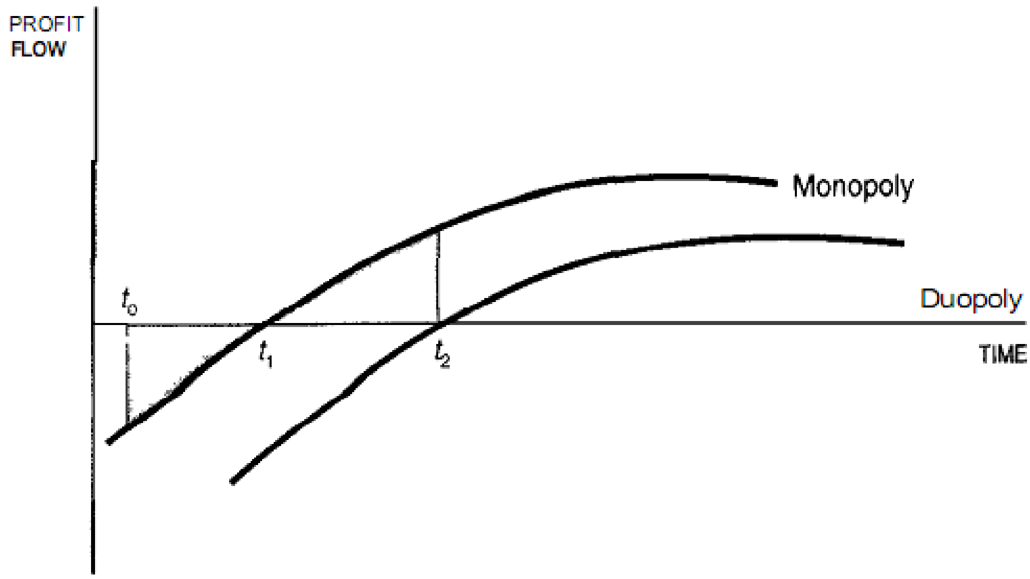


图 16.4：利润流与进入。在子博弈完美均衡中，第一家企业在 t_0 时进入，此时它的贴现利润为零。在 t_1 时进入是个纳什均衡，但不是子博弈完美均衡。

我们感兴趣的问题是进入模式问题。最明显的方案是 (t_1, t_2)，即当完全垄断利润比双头垄断利润大时，其中一个企业比如企业 1 进入；当双头垄断利润大于完全垄断利润时，企业 2 进入。这是常见的进入条件——利润为正。的确，容易验证，这是个纳什均衡。然而，多少有些令人意外的是，它不是子博弈完美均衡。

为了看清这一点，考虑如果企业 2（第二个进入者）决定进入时刻稍微比 t_1 早一点，将会发生什么样的结果。企业 2 无疑会在一段较短时期内赔钱，但现在企业 1 宣称在 t_1 时进入的威胁就不再可信。由于企业 2 进入市场，企业 1 在 t_1 时进入市场不再有利可图。因此，在区间 $[t_1, t_2]$，企业 2 得到的正垄断利润，在 t_2 之后得到双头垄断利润。

当然如果企业 2 决定进入时点比 t_1 时稍微早一些，企业 1 也会这么做。这个模型的唯一子博弈完美均衡是其中一个企业在 t_0 时进入，在这个时点上，初始垄断阶段的利润被压低为零；也就是说，位于 $\pi_1(t)$ 上方且在 t_0 与 t_1 之间的面积等于位于 $\pi_1(t)$ 下方且在 t_1 与 t_2 之间的面积，但符号相反。进入威胁使得垄断利润全部消失了！

稍微思考一下就知道这意味着什么。由于这两个企业的地位是相等的，如果它们最终得到的利润不同，那么这样的结果多少令人吃惊。在子博弈完美均衡中，先进入者的垄断利润被竞争消散，剩下的只是博弈的双头垄断阶段的利润。

16.13 限制性定价

经济学家通常把进入威胁看成寡头垄断的约束力。即使行业只有少数几家企业，也可能存在很多潜在的进入者，因此“实际”竞争程度非常高。即使对于垄断企业来说，当它面对进入威胁时，也可能进行价格竞争。这种旨在阻止进入的定价策略称为**限制性定价**（limit pricing）。

这个观点在表面上有很大的吸引力，但它存在着一些严重的问题。我们用一个正式模型来考察其中几个问题。假设有两个企业，一个为在位者（incumbent），该企业已进入市场并已生产；另外一个为潜在进入者，这个企业正考虑进入市场。市场需求函数和两个企业的成本函数为共同知识。博弈有两期：在第一期，垄断者制定价格和产量，潜在进入者能观测到这些变量，在这个时点上它决定是否进入市场。如果进入，在第二期这两个企业就构成了某种双头垄断博弈。如果不进入，在位者在第二期索要垄断价格。

在这个模型中，限制性定价特征是什么？几乎没什么特征。如果进入，必定是双头垄断均衡。潜在进入者唯一关注的事情是预测它能从双头垄断均衡中得到的利润。由于潜在进入者完全知道成本和需求函数，第一期的价格没有传递任何信息。因此，在位者在第一期能够并且索要垄断价格，从而获得垄断利润。

你可能会认为，为了向潜在进入者发出“如果进入，定当痛击”的信号，在位者也许会在第一期索要较低的价格。但这是个无效的威胁；如果其他企业进入了，那么在位者在这个时点上会理性地选择利润最大化的策略。由于潜在进入者知道所有相关信息，它会事先预测在位者的利润最大化策略，从而相应制定自己的计划。

在这样的架构中，限制性定价没有起到任何作用，这是因为第一期的定价不能传达第二期博弈的信息。然而，如果我们在模型中置入一些不确定因素，我们将发现限制性定价是个均衡策略。

考虑下面简单的模型。如果市场价格不大于 3，某商品的需求为 1 单位。在位者和潜在进入者分别只有一个。在位者的边际成本为常数 0 或 2 元（也就是说在位者可能是低成本的也可能是高成本的，因此存在着不确定性），潜在进入者的边际成本为常数 1 元。为了进入市场，进入者必须支付 1/4 元进入费。如果进入者进入市场，我们假设这两个企业展开伯特兰竞争。

由于这两个企业的成本不同，这意味着其中一个企业被驱逐出市场。如果在位者的成本更低，那么它制定的价格将稍微小于进入者的边际成本（1 元），从而将进入者赶出市场。在这种情形下，在位者的利润为 1，进入者损失了进入费（1/4 元）。如果在位者的成本更高，那么进入者制定的价格将稍微小于 2 元，从而赚取了 3/4 元（=2-1-1/4）的利润，并且把在位者赶出市场。

如果在位者的成本更高，而且潜在进入者不进入，那么在位者将索要垄断价格 3 元，赚取的利润为 1 元。问题在于，在位者在第一期制定的价格为多少？在本质上，低成本在位者倾向于制定下列这样的价格即该价格对于高成本在位者是不可行的，因为这种做法是在向潜在进入者示威：我的成本很低。假设低成本在位者在第一期制定的价格稍微小于 1 元，在第二期制定垄断价格 3 元。这对于低成本在位者仍然是有利可图的，因为它的成本为 0。但是这个政策对于高成本在位者不是有利可图的——在第一期，它损失的利润稍微大于 1 元，在第二期它的利润仅为 1 元。在整体上，这个政策导致了损失。由于只有低成本企业才有能力制定较低的价格（1 元），这是个可信的威胁信号。这个例子表明，在不完全信息模型中，限制性定价的确能发挥作用：它向潜在进入者发出了在位者成本类型的信号，从而阻止了进入，至少在某些情形下能阻止进入。（本章结束）■

注释

Shapiro（1989）很好地综述了寡头垄断理论。我沿用了他处理重复博弈的方法。寡头垄断的比较静态分析内容，借鉴了 Dixit（1986）。销售模型来自 Varian（1980）。古诺模型中的产能选择的讨论，是根据 Kreps and Scheinkman（1983）的材料改编而成。领导者-追随者博弈的利润分析借鉴了 Dowrick（1986）。Sonnenschein（1968）首先注意到了古诺均衡和伯特兰均衡的对偶性，后来 Singh and Vives（1984）扩展了这一分析。我在教材中给出的限制性定价模型是受到了 Milgrom and Roberts（1982）的启发。

习题

16.1 假设我们有四个企业，其中两个的边际成本为常数 c_1，另外两个企业的边际成本为常数 c_2，并且有 $c_2 > c_1$。求该模型的伯特兰均衡和竞争性均衡。

16.2 考虑教材 16.5 节一开始描述的销售模型。随着 U/I 增加，$F(p)$ 将会发生什么变化？

解释这个结果。

16.3 给定教材 16.5 节一开始描述的线性反需求函数，推导出直接需求函数的参数表达式。

16.4 使用上一个问题中的线性需求函数，证明与伯特兰竞争相比，古诺竞争中的产量总是更低、价格总是更高。

16.5 证明如果两个企业的反应函数曲线都是向上倾斜的即 $f_i'(y_j)>0$，而且假设 (y_1^*, y_2^*) 是斯坦科尔伯格均衡，那么 $f_2(f_1(y_2^*)) > f_2(y_1^*) = y_2^*$。

16.6 与竞争性模型相伴的推测变量 $v_{12}=-1$。这意味着当一个企业增加一单位产量时，另外一个企业将减少一单位产量。在直觉上，很难想象这竟然是竞争性行为。问题出在哪里？

16.7 证明如果对于所有 $x>0$ 都有 $c_1'(x)<c_2'(x)$，那么在卡特尔解中必然有 $y_1>y_2$。

16.8 假设有两个相同的企业，它们按照卡特尔解生产，而且每个企业都相信，如果它调整产量，另外一个企业也会调整产量，从而使得每个企业的市场份额都为 1/2。这意味着推测变量等于多少？这意味着该行业结构是什么样的？

16.9 为什么有限次重复古诺博弈有很多均衡，而在有限次重复囚犯困境博弈中只有一个均衡？

16.10 假设某个行业由 2 个企业组成，每个企业的边际成本都为 0。该行业面对的（反）需求曲线为

$$p(Y)=100-Y,$$

其中 $Y=y_1+y_2$ 是总产量。

（a）求该行业竞争性均衡的产量水平。

（b）如果每个企业都是古诺竞争者，给定企业 2 的产量，企业 1 的最优产量为多少？

（c）计算古诺均衡时每个企业的的产量。

（d）计算该行业的卡特尔产量。

（e）如果企业 1 是追随者，企业 2 是领导者。计算斯坦科尔伯格均衡时每个企业的产量。

16.11 考虑某个古诺行业，在该行业中，企业的产量分别记为 $y_1,\ldots,y_n$，总产量记为 $Y=\sum_{i=1}^n y_i$，行业需求曲线记为 $P(Y)$，每个企业的成本函数为 $c_i(y_i)=cy_i$。为简单起见，假设 $P''(Y)<0$。假设每个企业都需要缴税，税率为 t_i。

（a）写出企业 i 的一阶条件。

（b）证明该行业的产量和价格仅取决于税率之和 $\sum_{i=1}^n t_i$。

（c）考虑每个企业的税率都发生变化但行业税收负担不变的情形。令 Δt_i 表示企业 i 的税率变化，我们要求 $\sum_{i=1}^n \Delta t_i = 0$。假设所有的企业都没有退出市场，计算企业 i 的均衡产量变化即 Δy_i。提示：不需要求导数；这个问题可以通过考察问题（a）和（b）得到解答。

16.12 某个行业的结构如下。该行业有 50 家竞争性企业，每个企业的成本函数都相同即都为 $c(y)=y^2/2$。该行业还有一个垄断企业，它的边际成本为 0。该行业的需求曲线为

$D(p)=1000-50p.$

（a）求其中一个竞争性企业的供给曲线。

（b）竞争性部门的总供给为多少？

（c）如果垄断企业的定价为 p，求它的销量。

（d）求垄断企业的利润最大化产量。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 17 章：交换

曹乾　译

(东南大学　caoqianseu@163.com)

17 交换

在第 13 章我们探讨了单个市场的经济理论。我们已经知道如果市场里有很多人，可以合理地认为每个人都是市场价格的接受者，价格是不受他控制的。给定这些外生的价格，每个人于是决定他对商品的需求或供给。价格调整到使市场出清，在这样的均衡价格下，没有人会想改变他的行为。

单个市场的均衡是一种局部均衡模型，在该模型中，除了我们关注的某商品外，所有其他商品的价格都是假定不变的。在一般均衡模型中，所有价格都是变量，均衡要求所有的市场都出清。因此，一般均衡理论不仅需要考虑单个市场的运行，还需要考虑市场间的所有互动关系。

为了方便阐述，我们首先研究一般均衡模型的一种特别情形，在这种情形下，所有人都是消费者，不存在生产者。这种情形，称为纯粹交换，它阐述的很多现象同样适用于涉及生产者的这种更广泛的情形。

在一个纯粹交换的经济中，我们有几个消费者，每个消费者都完全由他的偏好和他拥有的商品进行刻画。这些人根据一定的规则进行交易，目的是让自己的状况变得更好。

这样过程的结果是怎样的？这样过程的理想结果是什么样的？什么样的分配机制才能实现上述理想结果？这些问题既涉及到实证的（positive）议题也涉及到规范的（normative）议题。正是这两类议题的相互作用才使得资源配置问题格外有趣。

17.1 消费者和商品

这里所说的商品范围非常广泛。商品可以按时间、位置和状态（state of world）进行区分。服务，例如劳动服务，是另外一类商品。假设每种商品都有各自的市场，每种商品价格由它所在的市场决定。

在纯粹交换模型中，只有消费者没有生产者。每个消费者 i 都完全由下列两个特征刻画：他的偏好 $\succsim i$ （或他的效用函数 u_i ）；他拥有的 k 种商品的初始禀赋 w_i 。假设每个消费者的行为都是竞争性的，也就是说，每个消费者都认为价格是给定的，和他的行为无关。我们假设每个消费者试图选择他能买得起的最优消费束。

一般均衡理论基本的关注点是商品如何在消费者之间进行分配的。消费者 i 拥有的商品 j 的数量用 x_i^j 表示。消费者 i 的消费束用 $x_i=(x_i^1,...,x_i^k)$ 表示，这是一个有 k 个分量的向量，表示消费者 i 消费每种商品的数量。配置 $x=(x_1,...,x_n)$ ，是 n 个消费束的集合，它表示 n 个消费者中每个人拥有的商品。可行配置是指在实物上可行的配置；在纯粹交换的情形下，可行配置就是指这个配置恰好将所有商品用光，也就是说， $\sum_{i=1}^n x_i=\sum_{i=1}^n w_i$ 。（在某些情形

下，如果以$\sum_{i=1}^{n} x_i \le \sum_{i=1}^{n} w_i$表示可行配置将更为方便。）

当只存在两个消费者和两种商品的情形下，我们可以使用一种方便的方法将配置、偏好和禀赋用二维形式表示，这就是埃奇沃思盒（Edgeworth box）。图 17.1 就是埃奇沃思盒的一个例子。

假设商品 1 的总数为$w^1 = w_1^1 + w_2^1$，商品 2 的总数为$w^2 = w_1^2 + w_2^2$。埃奇沃思盒的长为w^1，宽为w^2。盒内的一点(x_1^1, x_1^2)表示消费者 1 拥有的两种商品数量。同时，它也指明了消费者 2 拥有的两种商品的数量：$(x_2^1, x_2^2) = (w^1 - x_1^1, w^2 - x_1^2)$. 从几何图形上，我们从盒子的左下角衡量消费者 1 的消费束。消费者 2 的消费束则从盒子的右上角衡量。使用这种方法，可将每个可行的配置表示为盒子内的一点。

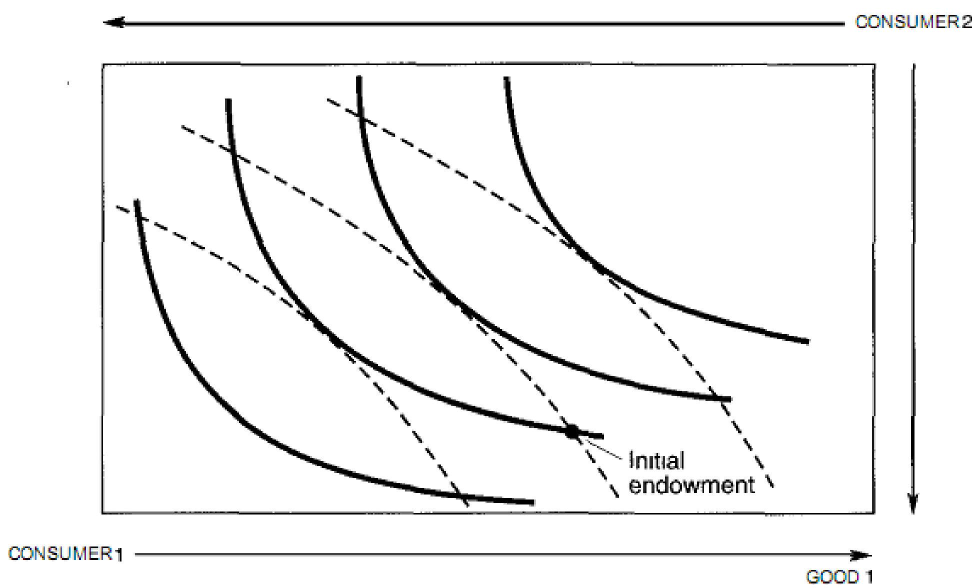


图 17.1：埃奇沃思盒。横轴的长度表示商品 1 的总数量；纵轴的高度表示商品 2 的总数量。盒内的每一点（包括边界）都是一个可行配置。

我们还可以在盒子内画出每个消费者的无差异曲线。这样盒内将有两组无差异曲线，每个消费者拥有一组。因此，两个消费者-两种商品的纯粹交换经济的所有信息，都可以用埃奇沃思盒这样简单的图形来表示。

17.2 瓦尔拉斯均衡

我们曾经指出，当市场上有很多消费者时，可以将每个消费者视为价格接受者，即市场价格和他的行为无关。考虑此处描述的纯粹交换的情形。假设市场价格为 $p=(p_1,...,p_n)$，即商品 i 的价格为 p_i。每个消费者将价格视为给定的，他们根据这些价格选择最喜欢的消费束；也就是说，每个消费者都试图求解下列问题：

$$\max_{x_i} \ u_i(x_i)$$

$$\text{使得 } px_i = pw_i.$$

这个问题的答案为 $x_i(p, pw_i)$，也就是消费者的需求函数，我们已在第 9 章研究过。在那一章，消费者的收入或财富（m）是外生的。此处，我们将消费者的收入视为他的初始禀赋的市场价值，因此 $m_i = pw_i$。我们在第 9 章已经知道，在偏好严格凸的假设条件下，需求函数是表现良好（well-behaved）连续函数。

当然，对于一个任意价格向量 p，未必能发生人们想要的交易，这是因为在任意价格下，总需求 $\sum_i x_i(p, pw_i)$ 可能不等于总供给 $\sum_i w_i$.

自然可以认为存在某个价格向量使得所有市场都出清；也就是说，在这组价格下，每个市场的需求都等于供给。然而，对于我们的目的来说，这组价格可能过于严格了。例如，如果我们不想要某些种类的商品。在这种情形下，均衡时这些商品可能是超额供给了。

正因为此，我们通常将**瓦尔拉斯均衡**（Walrasian equilibrium）定义为一对 (p^*, x^*)，使得

$$\sum_i x_i(p^*, p^* w_i) \le \sum_i w_i.$$

也就是说，如果所有商品都不存在正的超额需求，则 p^* 就是一个瓦尔拉斯均衡。稍后我们将看到，如果所有商品都是人们想要的，在准确意义上说，则实际上所有市场的需求都等于供给。

17.3 图形分析

瓦尔拉斯均衡可用埃奇沃思盒进行图形分析。给定任何价格向量，我们可以确定每个消费者的预算线，并且可以使用无差异曲线找到每个消费者的需求束。于是我们可以找到一组价格向量使得两个人的需求束是相容的。

在图 17.2 中，我们画出了这样的均衡配置。每个消费者都在他的预算线上最大化自己的效用，这些需求和总供给是相容的。注意，瓦尔拉斯均衡发生在两条无差异曲线的切点上。这个结论很明显，因为效用最大化要求每个消费者的边际替代率等于共同的价格比率

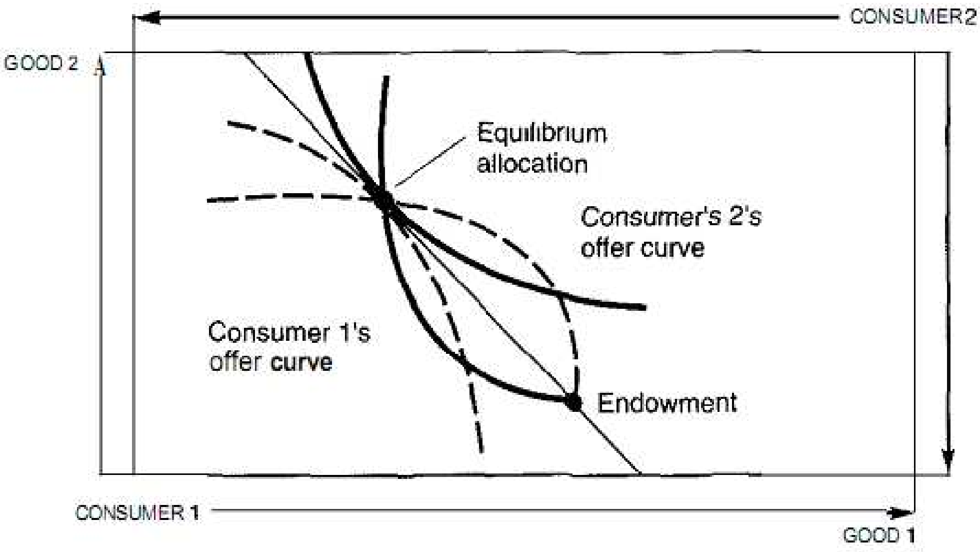


图 17.2：埃奇沃思盒中的瓦尔拉斯均衡。每个消费者在自己的预算线上最大化他的效用。

另外一种描述均衡的方法是使用**提供曲线**（offer curves）。我们已经知道，某个消费者的提供曲线是指当相对价格变动时，他的无差异曲线和预算线的切点的运动轨迹，也就是他的需求束的集合。因此，在埃奇沃思盒内的某个均衡点上，两个消费者的提供曲线相交于该点。在这样的交点上，每个消费者的需求束和供给是相容的。

17.4 瓦尔拉斯均衡的存在性

总是存在能让所有市场都出清的价格向量吗？在这一节我们将分析瓦尔拉斯均衡的存在性问题。

我们先回忆关于这个存在性问题的一些事实。首先，如果我们将所有商品的价格同乘以任何一个正的常数，则每个消费者的预算集保持不变；因此，每个消费者的需求函数具有以下性质：对于所有 $k>0$，都有 $x_i(p,pw_i)=x_i(kp,kpw_i)$；也就是说，需求函数对于价格是零次齐次的（homogeneous of degree zero）。由于齐次函数之和还是齐次的，总超额需求函数

$$z(p)=\sum_{i=1}^{n}[x_i(p,pw_i)-w_i],$$

对于价格也是零次齐次的。注意我们忽略了 z 依赖于初始禀赋向量 (w_i) 这个事实，这是因为初始禀赋在我们分析过程中保持不变。

如果所有个人需求函数都是连续的，则 z 也是连续函数，因为连续函数之和还是连续函数。而且，总超额需求函数必然满足一个条件，这个条件就是**瓦尔拉斯法则**（Walras's law）。

瓦尔拉斯法则。对于任何价格向量 p，我们有 $pz(p)\equiv 0$；也就是说，超额需

求的价值恒等于零。

证明。我们只要写出总超额需求的定义，并将其乘以价格向量 p 即可：

$$pz(p)=p\left[\sum_{i=1}^{n}x_i(p,pw_i)-\sum_{i=1}^{n}w_i\right]=\sum_{i=1}^{n}\left[x_i(p,pw_i)-pw_i\right]=0,$$

这是因为 $x_i(p,pw_i)$ 必定满足预算约束 $px_i=pw_i$，其中 $i=1,\ldots,n.$ ■

瓦尔拉斯的内容非常清楚：如果每个人满足预算约束，所以他的超额需求的价值等于零，于是超额需求总和的价值也必定为零。认识到瓦尔拉斯法则是个恒等式很重要，即超额需求的价值恒等于零，注意这里的恒等于零是对于**所有**价格向量都成立，而不是仅对于某些特殊价格向量成立。

结合瓦尔拉斯法则以及均衡的定义，我们得到两个有用的命题。

市场出清。**如果一共有 k 个市场，其中 $k-1$ 个的市场的需求都等于供给，并且 $p_k>0$，那么第 k 个市场的需求必定等于供给**。

证明。如若不然，则违背了瓦尔拉斯法则。■

免费商品。**如果 p^* 是一个瓦尔拉斯均衡而且 $z_j(p^*)<0$，则 $p_j^*=0$. 也就是说，如果某商品在瓦尔拉斯均衡时是超额供给的，那么它必定是一种免费商品**。

证明。由于 p^* 是一个瓦尔拉斯均衡，$z(p^*)\le 0$。由于价格都是非负的，

$$p^*z(p^*)=\sum_{i=1}^{k}p_i^*z_i(p^*)\le 0。$$

如果 $z_j(p^*)<0$ 并且 $p_j^*>0$，则会有 $p^*z(p^*)<0$，违背了瓦尔拉斯法则。■

这个命题表明了所有市场都出清的所要求的条件。假设所有商品在下列意义上是合意的：

合意性假设（Desirability）。**如果 $p_i=0$，则 $z_i(p)>0$，其中 $i=1,\ldots,k.$**

合意性假设是说如果某种商品的价格为零，则该商品的总超额需求为严格正。于是我们有下列命题：

供需相等。**如果所有商品都是合意的，而且 p^* 是一个瓦尔拉斯均衡，则 $z(p^*)=0.$**

证明。假设 $z_i(p^*)<0$。则根据免费商品命题可知 $p_i^*=0$。但是根据合意性假设知 $z_i(p^*)>0$，矛盾。■

总结一下：一般来说，我们对均衡的所有要求是任何商品都不存在着超额需求。但是，上述命题指出，如果均衡时，某商品实际上是超额供给的，则它的价格必定为零。因此，如

果每种商品都是合意的，即零价格意味着它是超额需求的，那么均衡实际上可由所有市场的供需相等状态进行刻画。

17.5 均衡解的存在性

由于总超额需求函数是零次齐次的，我们可以将价格标准化，从而用**相对价格**（relative prices）表达需求。标准化的方法有多种，然而对于我们的目的来说，比较方便的做法是将每个绝对价格 $\hat{p}_i$ 用标准化价格替换

$$p_i = \frac{\hat{p}_i}{\sum_{j=1}^{k} \hat{p}_i}.$$

这种做法的结果是，标准化价格之和必定等于 1。因此，我们可以将我们的注意力集中于，属于 $k-1$ 维的单位单纯形（unit simplex）的价格向量：

$$S^{k-1} = \{R_+^k \text{中的 } p : \sum_{i=1}^{k} p_i = 1\}.$$

S^1 和 S^2 的图形如图 17.3 所示。

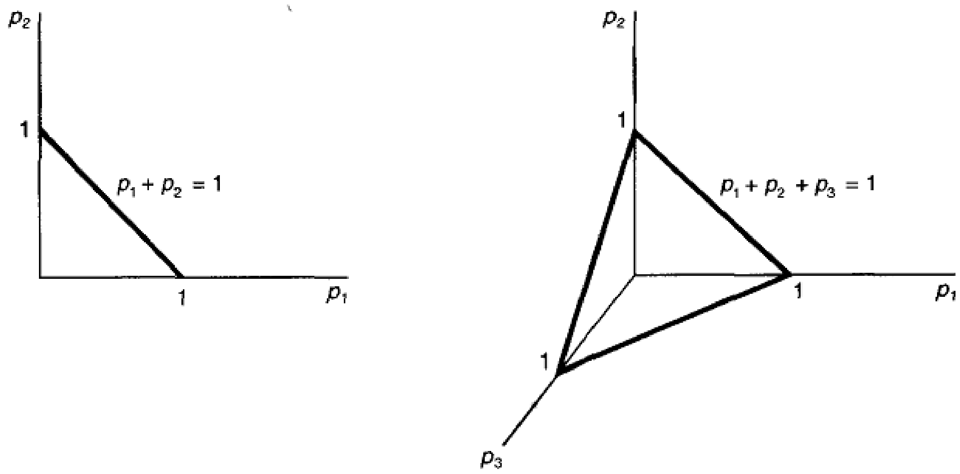


图 17.3：价格单纯形。第一个图是一维的价格单纯形 S^1；第二份图是二维的价格单纯形 S^2。

现在我们转向分析瓦尔拉斯均衡的存在性问题：是否存在能使所有市场出清的 p^*？我们使用 Brouwer 不动点定理证明这个存在性问题。

Brouwer 不动点定理。当且仅当 $S^{k-1} \to S^{k-1}$ 是一个从单位单纯形到其自身的连续函数时，存在 S^{k-1} 中的 x 使得 $x = f(x)$。

证明。一般情形的证明超出了本书的范围。Scarf(1973)提供的证明比较好。然而，我们证明 $k = 2$ 时的情形。

在这种情形下，我们可将一维单纯形 S^1 等同于单位区间。根据这个不动点定理的条件可知，我们有一个连续函数 $f:[0,1]\to[0,1]$，我们需要证明，$[0,1]$ 中存在某个 x 使得 $x=f(x)$.

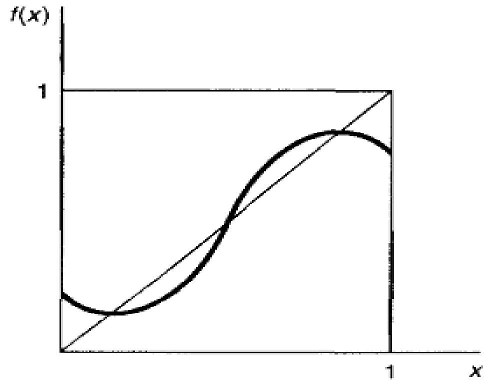


图 17.4：一维单纯形情形下的不动点

定义函数 g 为 $g(x)=f(x)-x$。在几何图形上，如图 17.4 所示，g 表示的是 $f(x)$ 与对角线之间的距离。映射 f 的不动点是指满足 $g(x^*)=0$ 的点 x^*。

由于 $f(0)$ 在 $[0,1]$ 中，$g(0)=f(0)-0\geq 0$ 且 $g(1)=f(1)-1\leq 0$。因为 f 是连续的，根据中值定理可知，$[0,1]$ 中存在某个 x 使得 $g(x)=f(x)-x=0$，这就证明了不动点定理。■

现在我们证明存在性定理。

瓦尔拉斯均衡的存在性。如果 $z:S^{k-1}\to R^k$ 是一个满足瓦尔拉斯法则 $pz(p)\equiv 0$ 的连续函数，则 S^{k-1} 中存在着某个 p^* 使得 $z(p^*)\leq 0$。

证明。定义映射 $g:S^{k-1}\to S^{k-1}$ 为

$$g_i(p)=\frac{p_i+\max(0,z_i(p))}{1+\sum_{j=1}^k\max(0,z_j(p))}\text{，其中 } i=1,\dots,k.$$

注意这个映射是连续的，这是因为 z 和 max 函数都是连续的。而且，$g(p)$ 是单纯形 S^{k-1} 中的一点，这是由于 $\sum_i g_i(p)=1$。这个映射也有合理的经济学解释：若某个市场是超额需求的，则 $z_i(p)\geq 0$，于是该商品的相对价格上升。

根据 Brouwer 不动点定理，存在 p^* 使得 $p^*=g(p^*)$；也就是说，

$$p_i^*=\frac{p_i^*+\max(0,z_i(p^*))}{1+\sum_{j=1}^k\max(0,z_j(p^*))}\text{，其中 } i=1,\dots,k. \tag{17.1}$$

我们需要证明 p^* 是一个瓦尔拉斯均衡。将（17.1）式交叉相乘并整理可得

$$p_i^*\sum_{j=1}^{k}\max(0,z_j(p^*))=\max(0,z_i(p^*))\text{，其中 } i=1,\ldots,k.$$

现在将上述 k 个式子中的每个式子乘以 $z_i(p^*)$：

$$z_i(p^*)p_i^*\sum_{j=1}^{k}\max(0,z_j(p^*))=z_i(p^*)\max(0,z_i(p^*))\text{，其中 } i=1,\ldots,k.$$

将这 k 个式子相加可得

$$\left[\sum_{j=1}^{k}\max(0,z_j(p^*))\right]\sum_{i=1}^{k}p_i^*z_i(p^*)=\sum_{i=1}^{k}z_i(p^*)\max(0,z_i(p^*)).$$

根据瓦尔拉斯法则可知 $\sum_{i=1}^{k}p_i^*z_i(p^*)=0$，我们有

$$\sum_{i=1}^{k}z_i(p^*)\max(0,z_i(p^*))=0.$$

这个加和中的每一项都大于或等于 0，这是由于每一项要么是 0，要么是 $z_i(p^*)^2$。但是，如果任何一项严格大于 0，则上式不可能成立。因此，每一项必须等于 0，这就是说

$$z_i(p^*)\le 0\text{，其中 } i=1,\ldots,k.\blacksquare$$

上述定理的一般性质值得强调。所有条件只需要假设超额需求函数是连续的而且满足瓦尔拉斯法则。瓦尔拉斯法则可直接从消费者必须满足某种预算约束的假设推导出；这样的行为在认可类型的经济模型中都是必要的。连续性的假设限制性稍微强些，但并非不合理。在前面我们已经知道，如果所有消费者的偏好都是严格凸的，则他们的需求函数将是良好定义而且是连续的。总超额需求函数因此也是连续的。但是，即使个人需求函数不是连续的，如果消费者的数量众多，总需求函数仍然是连续的。因此，总需求的连续性似乎是一个相对较弱的假设。

然而，上述对存在性问题的证明过程存在着一个小问题。的确，如果价格为正，总需求函数可能是连续的，但是如果某种商品的价格为零，假设总需求函数为连续就不再合理。例如，如果偏好是单调的，而且某种商品的价格为零，我们可以预期这样商品的需求是无限的。因此，超额需求函数在价格单纯形的边界上（即价格向量中某些价格为零）甚至可能不是良好定义的。然而，这类非连续性可用稍微复杂一些的数学方法进行处理。

例子：柯布-道格拉斯经济

令消费者 1 的效用函数为 $u_1(x_1^1,x_1^2)=(x_1^1)^a(x_1^2)^{1-a}$，他的禀赋为 $w_1=(1,0)$。令消费者 2 的效用函数为 $u_2(x_2^1,x_2^2)=(x_2^1)^b(x_2^2)^{1-b}$，他的禀赋为 $w_2=(0,1)$。则消费者 1 对商品 1 的需求函数为

$$x_1^1(p_1,p_2,m_1)=\frac{am_1}{p_1}$$

价格为(p_1, p_2)时，收入为$m_1 = p_1 \times 1 + p_2 = 0 = p_1$，将其代入上式可得

$$x_1^1(p_1, p_2) = \frac{ap_1}{p_1} = a.$$

类似地，消费者 2 对商品 1 的需求为

$$x_2^1(p_1, p_2) = \frac{bp_2}{p_1}.$$

均衡价格是每种商品的总需求等于总供给时的价格。根据瓦尔拉斯法则，我们只需要找到商品 1 的总需求等于商品 1 的总供给时的价格即可：

$$x_1^1(p_1, p_2) + x_2^1(p_1, p_2) = 1$$
$$a + \frac{bp_2}{p_1} = 1$$
$$\frac{p_2^*}{p_1^*} = \frac{1-a}{b}.$$

注意，和往常一样，均衡时确定的是相对价格。

17.6 福利经济学第一定理

瓦尔拉斯均衡的存在性作为一个实证结果是有趣的，当然前提是我们要相信该模型基于的行为假设。然而，即使这些假设在很多情形下不是特别合理，我们仍对瓦尔拉斯均衡的规范内容（normative content）感兴趣。我们考虑以下定义。

帕累托效率的定义。一个可行配置x称为一个弱帕累托有效率配置（weakly Pareto efficient allocation），如果不存在可行配置x'使得所有消费者都严格偏好x'胜于x。一个可行配置x称为一个强帕累托有效率配置，如果不存在可行配置x'使得所有消费者都弱偏好x'胜于x并且某个消费者严格偏好x'胜于x。

容易看出，强帕累托有效率的配置也是弱帕累托有效率的。一般来说，弱帕累托有效率的配置未必是强帕累托有效率的。然而，如果增加某些关于偏好的一些较弱的假设，弱帕累托有效率的配置也是强帕累托有效率的。因此，这两个概念可以替换使用。

弱帕累托效率和强帕累托效率的等价性。假设偏好是连续且单调的，则一个配置为弱帕累托有效率的当且仅当它是强帕累托有效率的。

证明。如果一个配置为强帕累托有效率的，则它当然是弱帕累托有效率的：如果在不损害其他人利益的情况下，你已无法使某个人的状况变好，这意味着你当然无法使每个人的状况变好。

我们需要证明如果某个配置是弱帕累托有效率的，则它是强帕累托有效率的。我们采

取的策略是证明与上述论断在逻辑上等价的论断：如果某个配置不是强有效率的，则它不是弱有效率的。

于是我们可以假设，存在可以让某个人i的状况变好而又不损害其他人利益的方法。这样，我们必须找到让每个人状况变好的方法。我们的做法是从个人i的消费束中稍微拿走一部分，并将这部分数量平均分配给其他人。更准确地说，我们用θx_i替换个人i的消费束x_i，用$x_j+\dfrac{(1-\theta)x_i}{n-1}$替换其余每个人$j$的消费束$x_j$。根据偏好的连续性可知，我们可以选择充分接近于 1 的θ使得个人i的状况仍然变好。根据单调性，所有其他人由于额外得到了一些商品，他们的状况都严格变好了。■

弱帕累托效率比强帕累托效率在数学表达上稍微方便一些，所以我们一般使用弱帕累托有效率的定义：当我们说“帕累托有效率”时，我们总是指“弱帕累托有效率”。然而，由于我们今后总是假设偏好是连续且单调的，因此这两个概念是等价的。

注意，帕累托效率概念在规范（normative）意义上很弱；在某个配置下，如果其中一个人拥有一切，其余人一无所有，那么这个配置仍然是帕累托有效率的，当然前提是拥有一切的这个人尚未达到饱和点。

帕累托有效率的配置可在前面介绍的埃奇沃思盒内画出。在两个人的情形下，帕累托有效率的配置可通过下列方法找到：将其中一个人的效用固定在某既定水平，并在这个约束条件下求另外一个人最大的效用水平。正式地，我们只需要求解下列的最大化问题：

$$\max_{x_1,x_2} u_1(x_1)$$

$$\text{使得}\ \ u_2(x_2)\geq \bar{u}_2$$

$$x_1+x_2=w_1+w_2.$$

这个问题可用埃奇沃思盒求解。我们只要在其中一个人的无差异曲线上找到一点，使得另外一个人的效用达到最大即可。到现在你应该明白最终的帕累托有效率点可用相切条件刻画：每个人的边际替代率必定相同。

对于消费者 2 的每个既定的效用水平，我们可以找到某个配置使得消费者 1 的效用最大，因此相切条件就会得到满足。帕累托有效率点的结合，即帕累托集，于是就是这两个人无差异曲线切点的轨迹，如图 17.5 所示。帕累托集有时也称为合同曲线（contract curve），这是由于它给出了有效率的“合同”或配置的集合。

比较图 17.5 和图 17.2，可以发现一个惊人的事实：似乎瓦尔拉斯均衡集和帕累托有效率配置集之间存在着一一映射。每个瓦尔拉斯均衡满足效用最大化的一阶条件，即每个人的两种商品的边际替代率都等于这两种商品的价格之比。由于所有人在瓦尔拉斯均衡时面对的商品价格比率是相同的，所有人的边际替代率必然也相同。

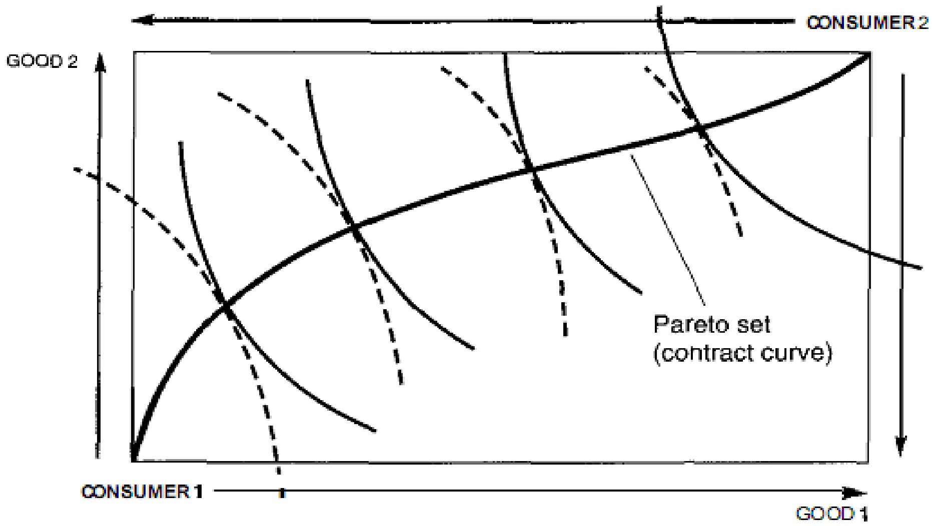


图 17.5：埃奇沃思盒中的帕累托有效率配置。帕累托集，或称为合同曲线，是所有帕累托有效率配置组成的集合。

而且，如果我们任意选择一个帕累托配置，我们知道每个人的边际替代率必定相等，而且我们可以选择某个价格比率使得它等于这两个人共同的边际替代率。从图形上说，给定任意一个帕累托有效率点，我们画出一条公切线将这两个人的无差异曲线分开。于是这条线上的任何一点都可以被当作初始禀赋。如果每个人都在自己的预算线上最大化自己的效用，他们最终正好位于帕累托有效率的配置上。

下面两个定理准确地给出了这个一一对应关系。首先，我们用更方便的方式重述瓦尔拉斯均衡的定义。

瓦尔拉斯均衡的定义。一个配置价格组合 (x, p) 是一个瓦尔拉斯均衡，如果（1）该配置是可行的，而且（2）每个人在自己的预算线上作出最优选择。用式子表达：

(1) $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} w_i.$

(2)如果个人 i 偏好 x_i' 胜于 x_1，则 $px_i' > pw_i$。

只要满足合意性条件，这个定义和瓦尔拉斯均衡原来的定义是等价的。这个定义可以让我们忽略免费商品的可能性，因为免费商品对于下面的论断来说，有些让人讨厌。

福利经济学第一定理。如果 (x, p) 是一个瓦尔拉斯均衡，则 x 是帕累托有效率的。

证明。如若不然，令 x' 是所有消费者都更偏好的另一个可行配置，即 $x' \succ x$。根据瓦尔拉斯均衡定义的性质 2 可知

$$px_i' > pw_i \text{，其中 } i = 1, ..., n.$$

将上面的n个式子相加，并利用x'是个可行配置的事实，我们有

$$p\sum_{i=1}^{n} w_i = p\sum_{i=1}^{n} x_i' > \sum_{i=1}^{n} pw_i,$$

这样我们就得到了一个矛盾。■

这个定理是说如果我们模型的行为假设得到满足，则市场均衡是有效率的。市场均衡在任何伦理道德意义上未必是“最优的”，这是因为市场均衡可能是非常“不公平的”。配置结果完全取决于禀赋的初始分配情况。我们需要的是更进一步的道德标准，据此再在有效率的配置中进行选择。这样的概念，即福利函数的概念，稍后我们将探讨。

17.7 福利经济学第二定理

我们已经证明每个瓦尔拉斯均衡都是帕累托有效率的。下面我们将证明每个帕累托有效率的配置都是瓦尔拉斯均衡。

福利经济学第二定理。假设x^*是一个帕累托有效率的配置，其中每个人持有的每种商品的数量都为正。假设偏好是凸的、连续且单调的，则x^*是与初始禀赋$w_i = x_i^*$相伴的瓦尔拉斯均衡，其中$i = 1,...,n.$

证明。令

$$P_i = \{R^k \text{ 中的 } x_i : x_i \succ_i x_i^*\}.$$

这是由对个人i来说，比x_i^*更好的所有消费束组成的集合。于是定义

$$P = \sum_{i=1}^{n} P_i = \{z : z = \sum_{i=1}^{n} x_i \text{ ，其中 } x_i \text{ 在 } P_i \text{ 中}\}.$$

P表示由k种商品构成的所有消费束，这些消费束可在n个人中分配从而使每个人的状况都变好。由假设条件可知，P_i是一个凸集，又由于凸集之和还是凸集，因此P是一个凸集。

令$w = \sum_{i=1}^{n} x_i^*$表示当前的**总消费束**（aggregate bundle）。由于x^*是帕累托有效率的，因此x^*的任何再分配都不可能使所有人的状况都变好。这意味着w不在集合P之中。

因此，根据分离超平面定理（请参考第 26 章）可知，存在$p \neq 0$使得

$$pz \geq p\sum_{i=1}^{n} x_i^* \text{ 对于 } P \text{ 中的所有 } z \text{ 都成立。}$$

将上式变形可得

$$p(z - \sum_{i=1}^{n} x_i^*) \geq 0 \text{ 对于 } P \text{ 中的所有 } z \text{ 都成立} \qquad (17.2)$$

我们想证明 p 实际上是一个均衡价格向量。证明分为三步。

（1） p 是非负的；即 $p \geq 0.$

为了看清这一点，令 $e_i = (0,...,1,...,0)$，即第 i 个分量为 1。由于每个人的偏好是单调的，$w + e_i$ 必定在 P 中；由于任何商品的数量都额外增加了一单位，因此再分配可以让每个人的状况变好。不等式（17.2）于是蕴含着

$$p(w + e_i - w) \geq 0，其中 i = 1,...,k .$$

由上式可得

$$pe_i \geq 0，其中 i = 1,...,k .$$

这个式子意味着 $p_i \geq 0$，其中 $i = 1,...,k$.

（2）如果 $y_j \succ_j x_j^*$，那么 $py_j \geq px_j^*$，其中 $i = 1,...,n$..

我们已经知道，如果每个人 i 偏好 y_j 胜于 x_j^*，则

$$p\sum_{i=1}^{n} y_i \geq p\sum_{i=1}^{n} x_i^*.$$

现在假设只有某个特定的消费者 j 偏好 y_j 胜于 x_j。将每种商品都从消费者 j 手里拿走一些并且分配给其他消费者，从而构建一个配置 z。正式地，令 θ 是一个很小的正数，定义配置 z 为

$$z_j = (1-\theta)y_j$$

$$z_i = x_i^* + \frac{\theta y_j}{n-1} \quad i \neq j.$$

对于足够小的 θ，强单调性意味着配置 z 比 x^* 更受帕累托偏好，因此 $\sum_{i=1}^{n} z_i$ 在 P 中。运用不等式（17.2），我们有

$$p\sum_{i=1}^{n} z_i \geq p\sum_{i=1}^{n} x_i^*$$

$$p\left[y_j(1-\theta) + \sum_{i\neq j} x_i^* + y_j\theta \right] \geq p\left[x_j^* + \sum_{i\neq j} x_i^* \right]$$

$$py_j \geq px_j^*.$$

这个论断表明，如果消费者 j 偏好 y_j 胜于 x_j^*，则 y_j 的花费不会比 x_j^* 小。剩下的证明是我们要将这个不等式变为严格不等式。

（3）如果 $y_j \succ_j x_j^*$，我们必然有 $py_j > px_j^*$.

我们已经知道 $py_j \geq px_j^*$. 我们想排除等号成立的可能性。因此，我们假设 $py_j = px_j^*$ 并且试图得到矛盾。

根据偏好的连续性假设，我们可以找到某个 θ，$0 < \theta < 1$，使得 θy_j 严格比 x_j^* 更受偏好。根据第（2）部分的论证，我们知道 θy_j 的花费至少和 x_j^* 的花费一样大：

$$p\theta y_j \geq px_j^*. \qquad (17.3)$$

这个定理的一个假设是 x_j^* 的每个分量都是严格正的；由此可知 $px_j^* > 0$.

因此，如果 $py_j - px_j^* = 0$，由此可得到 $\theta py_j < px_j^*$. 但这和（17.3）式矛盾，因此我们就证明了这个定理。■

这个命题的假设值得思考。偏好的凸性和连续性当然非常重要，但是强单调性的条件可以放松。你也可以放松 $x_j^* >> 0$ 的假设。

显示偏好的论证

福利经济学第二定理有个更为简单但有些间接的证明方法，这种证明方法要使用显示偏好理论和本章前面几节证明的存在性定理。

福利经济学第二定理。假设 x^* 是一个帕累托有效率的配置，而且假设偏好是非饱和的。进一步假设对于初始禀赋 $w_i = x_i^*$ 来说存在着竞争均衡，令 (p', x') 表示这个竞争均衡。则，事实上，(p', x^*) 是一个竞争均衡。

证明。根据构造，x_i^* 在消费者 i 的预算集中，我们必然有 $x_i' \succsim_i x_i^*$。由于 x^* 是帕累托有效率的，这意味着 $x_i^* \sim_i x_i'$。因此，如果 x_i' 是最优的，x_i^* 必然也是最优的。所以，(p', x^*) 是一个瓦尔拉斯均衡。■

这个论证表明，如果对于某个帕累托有效率的配置来说**存在着**竞争均衡，则该帕累托有效率**本身**就是一个竞争均衡。本章前面对于存在性定理的评价表明，存在性定理的唯一核心假设是总需求函数的连续性。连续性可从个人偏好的凸性推导出或者从经济很“大”的假设推导出。因此，第二定理成立的条件也与之相同。

17.8 帕累托效率和微积分

在上一节我们已经知道，每个竞争均衡都是帕累托有效率的，而且在本质上，每个帕累托有效率的配置也是与某个禀赋配置相伴的一个竞争均衡。在这一节，我们将使用微积分更仔细地分析这种关系。在本质上，我们要推导出市场均衡和帕累托有效率各自的一阶条件，然后将这两组条件进行比较。

市场均衡的一阶条件非常简单。

市场均衡的微分特征。如果 (x^*, p^*) 是一个市场均衡，其中每个消费者拥有的每种商品的数量都为正，则存在着一组数 $(\lambda_1,...,\lambda_n)$ 使得：

$$Du_i(x^*) = \lambda_i p^* \qquad i = 1,...,n.$$

证明。如果我们有一个市场均衡，则每个消费者都在自己的预算线上最大化其自身的效用，而这正是他们效用最大化的一阶条件。λ_i 是消费者 i 的收入的边际效用。■

帕累托有效率的一阶条件相对难以表达。然而，使用下列的技巧就可以解决这个问题。

帕累托有效率的微分特征。一个可行配置 x^* 是帕累托有效率的，当且仅当 x^* 是下列 n 个最大化问题的解，其中 $i = 1,...,n$.

$$\max_{(x_i^g, x_j^g)} u_i(x_i)$$

$$\text{使得} \sum_{h=1}^{n} x_h^g \le w^g \qquad g = 1,...,k$$

$$u_j(x_j^*) \le u_j(x_j) \qquad j \ne i.$$

证明。假设 x^* 是所有 n 个最大化问题的解但它不是帕累托有效率的。这意味着存在着另外一个配置 x' 使得每个人的状况更好。如果这样 x^* 不可能是上述任何一个最大化问题的解，矛盾。

相反地，假设 x^* 是帕累托有效率的，但它不是其中一个最大化问题的解。令 x' 是这个最大化问题的解。于是这意味着 x' 使得这个人的状况变好而又不损害其他人的利益，这和 x^* 是帕累托有效率的假设矛盾。■

在研究其中一个最大化问题的拉格朗日表达式之前，我们来做个简单的计算。这 n 个最大化问题种的每一个都有 $k+n-1$ 个约束条件。前 k 个条件是资源约束，后面的 $n-1$ 个条件是效用约束。在每个最大化问题中，有 kn 个选择变量：n 个消费者中的每个人拥有 k 种商品中的每种商品的数量。

令 q^g，其中 $g = 1,...,k$ 是资源约束的库恩-塔克乘子，令 a_j，其中 $j \ne i$，是效用约束的乘子。写出其中一个最大化问题的拉格朗日函数:

$$L = u_i(x_i) - \sum_{g=1}^{k} q^g \left[\sum_{i=1}^{n} x_i^g - w^g \right] - \sum_{j \ne i} a_j [u_j(x_j^*) - u_j(x_j)].$$

现在 L 对 x_j^g 求导，其中 $g = 1,...,k$，$j = 1,...,n$。我们得到下列形式的一阶条件

$$\frac{\partial u_i(x_i^*)}{\partial x_i^g} - q^g = 0 \qquad g = 1,...,k$$

$$a_j \frac{\partial u_j(x_j^*)}{\partial x_j^g} - q^g = 0 \qquad j \neq i;\ g = 1,\ldots,k.$$

乍看起来，这些条件可能有些奇怪，因为它们似乎是不对称的。对于消费者 i 的每个选择，我们可以得到乘子 (q^g) 和 (a_i) 的不同数值。然而，当我们注意到这些 q 的相对值和消费者 i 的选择无关后，这个不对称的难题解决了。q 的相对值和消费者 i 的选择无关，这是因为上面的条件意味着

$$\frac{\dfrac{\partial u_i(x_i^*)}{\partial x_i^g}}{\dfrac{\partial u_j(x_j^*)}{\partial x_i^h}} = \frac{q^g}{q^h} \qquad i = 1,\ldots,n; \quad g, h = 1,\ldots,k.$$

由于 x^* 是给定的，q^g / q^h 必然和我们求解的最大化问题无关。类似地，a_i / a_j 必然和我们求解的最大化问题无关。那个不对称问题的解现在变得清楚了：如果我们用其他人的效用作为约束，求解消费者 i 的最大化效用，则就好像我们将消费者 i 的 Kukn-Tucker 乘子设定为 $a_i = 1$.

使用第一定理，我们可以推导出权重 (a_i) 和 (q^g) 的美妙解释：如果 x^* 是一个市场均衡，则

$$Du_i(x_i^*) = \lambda_i p^* \qquad i = 1,\ldots,n.$$

然而，所有的市场均衡都是帕累托有效率的，因此必定满足

$$a_i Du_i(x_i^*) = q \qquad i = 1,\ldots,n.$$

由此明显可以看出，我们可以选择 $p^* = q$ 以及 $a_i = 1/\lambda_i$。用语言表达，资源约束的库恩-塔克乘子就是竞争价格；消费者效用的库恩-塔克乘子就是他们各自收入边际效用的倒数。

如果我们消除一阶条件中的库恩-塔克乘子，就可以得到描述有效率配置的下列条件：

$$\frac{\dfrac{\partial u_i(x_i^*)}{\partial x_i^g}}{\dfrac{\partial u_j(x_j^*)}{\partial x_i^h}} = \frac{p_g^*}{p_h^*} = \frac{q^g}{q^h} \qquad i = 1,\ldots,n; \quad g, h = 1,\ldots,k.$$

这个式子是说，每个帕累托有效率的配置必定满足下列条件：每对商品的边际替代率对于任何消费者都是相同的。这个边际替代率就是竞争价格的比率。

注意到下列事实也是有用的：帕累托有效率的一阶条件和效用加权和的一阶条件是相同的。为了看清这一点，考虑下列问题

$$\max \sum_{i=1}^{n} a_i u_i(x_i)$$

$$\text{使得} \sum_{i=1}^{n} x_i^g \le w^g \qquad g = 1,\ldots,k.$$

这个最大化问题的一阶条件为

$$a_i Du_i(x_i^*) = q\text{，} \tag{17.4}$$

这个条件和帕累托有效率的一阶条件是相同的。

当“福利权重”集 $(a_1,\ldots,a_n)$ 变动时，我们可以画出帕累托有效率配置集的变动轨迹。如果我们所有帕累托有效率配置感兴趣，我们需要对这些式子进行操作，消去福利权重。一般来说，这就是用边际替代率表示一阶条件。

另外一种分析方法是，将福利权重纳入效用函数的定义。如果消费者 i 原来的效用函数为 $u_i(x_i)$，作单调变换得到新的效用函数 $v_i = a_i u_i(x_i)$。由此得到的一阶条件刻画了一种特殊的帕累托有效率配置，这个配置最大化的是某种既定效用函数的和。但是，如果我们对一阶条件进行操作，从而使得它们用边际替代率表示，我们通常能够找到描述所有有效率配置的条件。

眼下我们注意到这个描述帕累托效率的微积分条件，能让我们比较容易地证明第二定理。让我们假设所有消费者的效用函数都是凹的，尽管这一要求不是必需的。于是如果 x^* 是一个帕累托有效率的配置，我们从一阶条件知道，

$$Du_i(x_i^*) = \frac{1}{a_i} q \qquad i = 1,\ldots,n.$$

因此，每个消费者效用函数的梯度（gradient）与某个既定的向量 q 成正比。我们选择竞争价格向量作为 q。我们需要验证每个消费者在他自己的预算集上 $\{x_i : qx_i \le qx_i^*\}$ 已达到效用最大。但这可由效用函数为凹立即推知；根据凹函数的数学性质：

$$u(x_i) \le u(x_i^*) + Du(x_i^*)(x_i - x_i^*)\text{，}$$

因此，

$$u(x_i) \le u(x_i^*) + \frac{1}{a_i}(x_i - x_i^*).$$

所以，如果 x_i 在消费者的预算集中，则 $u(x) \le u(x_i^*)$.

17.9 福利最大化

帕累托效率作为一个规范性的标准，它有一个缺陷，即它并不是很明确的（specific）。帕累托效率只和效率问题有关，它和福利分配标准无关。即使我们同意，我们想要的是一个帕累托有效率的配置，我们仍然不知道到底要哪一个。

这些问题的一种解决之道是假设存在着**社会福利函数**（social welfare function）。这个函数是将个人效用函数加总从而得到“社会效用”。社会福利函数最合理的解释是，它表示社会决策制定者的对不同个人效用权衡的偏好。此处我们不作出哲学意义上的评价，而只是假定这样的函数是存在的；也就是说，假设我们有

$$W: R^n \to R,$$

这样，给定任何个人效用分布$(u_1,...,u_n)$，我们都能从$W(u_1,...,u_n)$中得到“社会效用”。为了使这个构造有意义，我们必须选择每个人的效用的特定表达形式，并在讨论过程中保持不变。

我们假设W关于它的每个变量都是递增的——如果你增加任何一个人的效用但不减少其他人的效用，社会福利就会增加。我们假设社会应该在社会福利最大化的点上运行；也就是说，我们应该选择某个配置$\mathbf{x}^*$，使得$\mathbf{x}^*$是下列约束最大化问题的解

$$\max W(u_1(x_1),...,u_n(x_n))$$
$$\text{s.t.} \sum_{i=1}^{n} x_i^g \le w^g \quad g = 1,...,k.$$

我们如何比较下列两个配置：一个是使得上述福利函数最大化的配置，另一个是帕累托有效率配置。下列命题是单调性假设的平凡结果：

福利最大化和帕累托效率。如果$\mathbf{x}^*$使得社会福利函数最大化，那么$\mathbf{x}^*$是帕累托有效率的。

证明。如果$\mathbf{x}^*$不是帕累托有效率的，那么必定存在着另外的可行配置$\mathbf{x}'$，使得对于$i=1,...,n$我们有$u_i(x_i') > u_i(x_i^*)$。但是，这样一来我们就有

$$W(u_1(x_1'),...,u_n(x_n')) > W(u_1(x_1^*),...,u_n(x_n^*))。$$

矛盾。■

由于福利最大化点是帕累托有效率的，它们必定满足与帕累托有效率配置相同的一阶条件；而且，在凸性假设条件下，每个帕累托有效率配置是一个竞争性的均衡，因此，这也适用于社会福利最大化配置：每个福利最大化配置都是某个禀赋分配的竞争性均衡。

最后这个结论使我们得到了竞争性价格的一种进一步的解释：它们也是福利最大化问题的库恩-塔克乘子。运用包络定理，我们可以看到竞争性价格衡量的是一种商品的（边际）社会价值：如果这种商品数量额外增加一点，社会福利增加多大。然而，这个结论仅适用于我们考察的配置能够最大化我们选择的那个福利函数的情形。

我们已经看到每个福利最大化配置是帕累托有效率的，但是它的逆命题一定成立吗？我们在上一节已经知道，每个帕累托有效率配置满足与效用加权加总最大化问题相同的一阶条件，因此，似乎在凸性和凹性假设下，上述命题的逆应该成立。事实的确如此。

帕累托效率和福利最大化。**令$\mathbf{x}^*$是一个帕累托有效率的配置，其中$x_i^* \gg 0$，$i=1,...,n$。令效用函数u_i是凹的、连续的和单调的函数。那么，存在权重a_i^*使得x^*在资源约束下能最大化$\sum a_i^* u_i(x_i)$。而且权重满足$a_i^* = 1/\lambda_i^*$，其中λ_i^*是第i个人收入的边际效用；也就是说，如果m_i是第i个人在均衡价格为$\mathbf{p}^*$时的禀赋，那么**

$$\lambda_i^* = \frac{\partial v_i(\mathbf{p}^*, m_i)}{\partial m_i}。$$

证明。由于$\mathbf{x}^*$是帕累托有效率的，它是个瓦尔拉斯均衡。因此，存在着价格$\mathbf{p}^*$使得每个人都能最大化他自己的预算集；这又意味着

$$Du_i(x_i^*) = \lambda_i \mathbf{p}^*，\ i=1,...,n。$$

现在考虑社会福利最大化问题

$$\begin{aligned}
&\max \sum_{i=1}^n a_i u_i(x_i) \\
&\text{s.t.} \sum_{i=1}^n x_i^1 \le \sum_{i=1}^n x_i^{1*} \\
&\qquad \vdots \\
&\quad \sum_{i=1}^n x_i^k \le \sum_{i=1}^n x_i^{k*}
\end{aligned}$$

根据凹性约束的最大化问题的充分性定理（参见第 27 章 27.5 节）可知，如果存在非负数$(q_1,...,q_k) = \mathbf{q}$使得

$$a_i Du_i(x_i^*) = \mathbf{q}，$$

那么$\mathbf{x}^*$是上述最大化问题的解。

如果我们选择$a_i = 1/\lambda_i$，那么价格$\mathbf{p}$就充当了这个合适的非负数，证毕。■

将权重解释为收入的边际效用的倒数，这有着很好的经济含义。如果某个人在某个帕累托有效率配置上的收入较多，那么他收入的边际效用将比较小，从而在社会福利函数中他的权重较大。

上述两个命题完整描述了市场均衡、帕累托有效率配置与福利最大化之间的关系。我们简要重述如下：

（1）竞争性均衡总是帕累托有效率的；

（2）在凸性假设和禀赋重新分配条件下，帕累托有效率的配置是竞争性均衡。

（3）社会福利最大化的配置总是帕累托有效率的。

（4）在福利权重的凹性假设条件下，帕累托有效率配置是福利最大化的。

考察上面的关系，我们可以看到其中的寓意：竞争性市场体系能够给出有效率的配置，

但这一点也没涉及分配问题。收入分配的选择与禀赋重新分配的选择一样，这又等价于选择某个特定的福利函数。（本章结束）■

注释

Walras（1954）首先提出了一般均衡模型。Wald（1951）第一个完成了一般均衡存在性的证明；McKeinzie（1954）和 Arrow and Debreu（1954）证明了一般情形下的一般均衡的存在性。Debreu（1954）和 Arrow and Hahn（1971）给出了最可靠的证明。Arrow and Hahn（1971）这份文献包含了数不清的历史性注释。

基本福利定理源远流长。我们在教材中对第一福利定理的证明采用了 Koopmans（1957）的方法。Arrow（1951）和 Debreu（1953）强调了凸性在第二定理中的重要性。效率的微分处理方法是由 Samuelson（1947）首先严格发展出。我们对福利最大化和帕累托效率之间关系的介绍沿用了 Negisihi（1960）的方法。

第二定理的显示偏好证明方法是由 Maskin and Roberts（1980）提出的。

习题

17.1 考虑第二福利定理的显示偏好证明方法。请证明如果偏好为严格凸，则对于 $i = 1, \ldots, n$ 有 $\mathbf{x}_i' = \mathbf{x}_i^*$。

17.2 如果价格为有限个，画出瓦尔拉斯均衡时的埃奇沃思盒状图。

17.3 考虑图 17.6。在此图中 $\mathbf{x}^*$ 是个帕累托有效率配置，但 $\mathbf{x}^*$ 无法得到竞争性价格的支持。这违背了第二福利定理中的哪个假设？

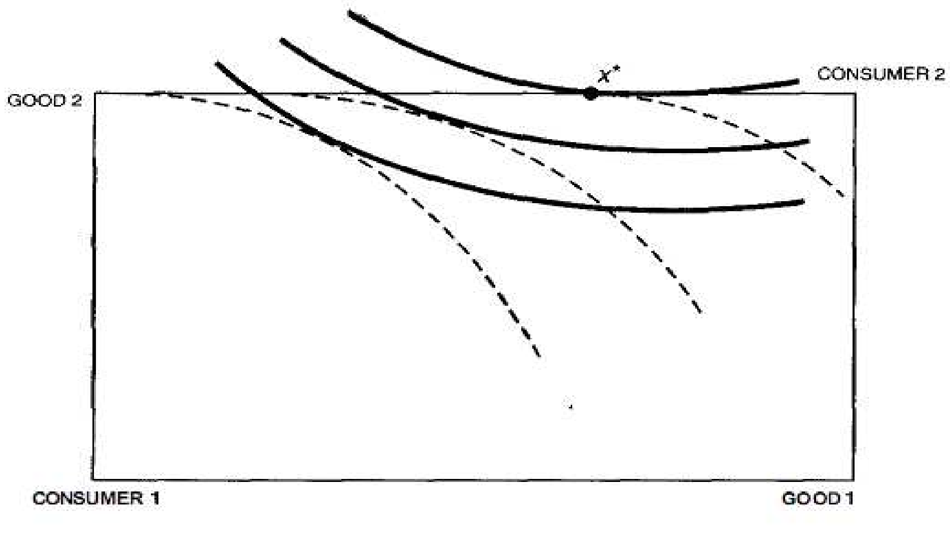


图 17.6：阿罗的例外情形。配置 $\mathbf{x}^*$ 是帕累托有销量的，但不存在能使得 $\mathbf{x}^*$ 为瓦尔拉斯均衡的价格。

17.4 有两个消费者 A 和 B，他们的效用函数和禀赋分别为

$$u_A(x_A^1, x_A^2) = a\ln x_A^1 + (1-a)\ln x_A^2 \quad \omega_A = (0,1)$$
$$u_B(x_B^1, x_B^2) = \min\{x_B^1, x_B^2\} \quad \omega_B = (1,0)$$

计算市场出清价格和均衡配置。

17.5 有 n 个人，他们的效用函数是相同的，且为严格凸函数。初始商品束为 ω。证明均分该商品束是个帕累托有效率的配置。

17.6 有两个人，他们的间接效用函数分别为：

$$v_1(p_1, p_2, y) = \ln y - a\ln p_1 - (1-a)\ln p_2$$
$$v_2(p_1, p_2, y) = \ln y - b\ln p_1 - (1-b)\ln p_2$$

他们的初始禀赋分别为

$$\omega_1 = (1,1) \quad \omega_2 = (1,1).$$

计算市场出清价格。

17.7 假设所有消费者的效用函数都是拟线性的，即 $v_i(p, m_i) = v_i(p) + m_i$。令 $\mathbf{p}^*$ 为一个瓦尔拉斯均衡。证明每种商品的总需求曲线在 $\mathbf{p}^*$ 处必定都是向下倾斜的。更一般地，证明总替代矩阵必定是负半定的。

17.8 假设有两个消费者 A 和 B，他们的效用函数相同的：

$$u_A(x_1, x_2) = u_B(x_1, x_2) = \max\{x_1, x_2\}\text{。}$$

商品 1 和 2 的数量分别为 1 单位和 2 单位。用埃奇沃思盒说明强帕累托有效率的集合与（弱）帕累托有效率的集合。

17.9 某个经济体有 15 个消费者和 2 种商品。消费者 3 的效用函数为柯布-道格拉斯类型 $u_3(x_3^1, x_3^2) = \ln x_3^1 + \ln x_3^2$。在某个帕累托有效率配置 $\mathbf{x}^*$ 上，消费者 3 持有的商品束为（10, 5）。求支持配置 $\mathbf{x}^*$ 的竞争性价格。

17.10 如果我们允许饱和的可能性，消费者的预算约束采取的形式为 $\mathbf{p}\mathbf{x}_i \le \mathbf{p}\omega_i$。于是瓦尔拉斯法则变为：$\mathbf{p}z(\mathbf{p}) \le 0$ 对于所有 $\mathbf{p} \ge 0$ 成立。证明我们在教材中对瓦尔拉斯均衡的证明仍然适用于上述一般形式的瓦尔拉斯法则。

17.11 消费者 A 的效用函数为 $u_A(x_1, x_2) = x_1 + x_2$，B 的效用函数为

$$u_B(x_1, x_2) = \max\{x_1, x_2\}\text{。}$$

这两个人的禀赋是相同的，都为（1/2, 1/2）。

（a）在埃奇沃思盒中画出这种情形。

（b） p_1 与 p_2 有何均衡关系？

（c）求均衡配置。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 18 章：生产

曹乾　译

(东南大学　caoqianseu@163.com)

18 生产

上一章只分析了纯交换的经济。在这一章，我们将描述如何将这种一般均衡模型推广到含有生产的经济中。首先，我们将讨论如何模型化企业的行为，接下来讨论如何模型化消费者的行为，最后讨论如何修改基本的存在和效率定理。

18.1 企业行为

我们将使用第 1 章介绍的生产技术表示方法。如果有 k 种产品，则企业 j 的净产出向量为含有 k 个分量的向量 y_j；企业 j 的可行净产出向量集，即生产可能集为 Y_j。我们已经知道，在净产出向量中，负的元素表示净投入而正的元素表示净产出。第 1 章已经给出过若干生产可能集的例子，此处不再赘述。

在这一章中，我们只分析完全竞争的企业。如果 p 为商品的价格向量， py_j 为与生产计划 y_j 相伴的利润。假设企业 j 选择的利润最大化方案为 y_j^*。

在第 2 章，我们分析过这个行为模型的结果。在那一章我们介绍了完全竞争企业净供给函数 $y_j(p)$ 的概念。这个函数是说：任意给定一个价格向量 p，企业相应的利润最大产出向量是什么。在某些假设条件下，单个企业的净供给函数是良好定义和表现良好的。如果我们有 m 个企业，**总净供给函数**（aggregate net supply function）为 $y(p)=\sum_{j=1}^{m} y_j(p)$。如果单个企业的净供给函数是良好定义且连续的函数，那么总的净供给函数也是良好定义且连续的函数。

我们也可以考虑**总生产可能集**（aggregate production possibilities set）Y。这个集合表示的是经济整体的所有可行净产出向量。它是单个生产可行集的加和，因此可将其写为

$$Y=\sum_{j=1}^{m} Y_j.$$

注意理解上式的意思。生产方案 y 在 Y 之中的充分必要条件是 y 可以写为

$$y=\sum_{j=1}^{m} y_j.$$

其中每个生产方案 y_j 都在 Y_j 中。因此，Y 代表的是若将生产在企业 $j=1,\ldots,m$ 分配后可实现的所有生产方案。

总利润最大化（aggregate profit maximization）。**某个总生产方案 y 使得总利润最大化，当且仅当每个企业的生产方案 y_j 使得自己利润最大化。**

证明。假设 $y=\sum_{j=1}^{m} y_j$ 使总利润最大但是某个企业 k 若选择 y_k' 可使利润更高。于是让

企业k选择方案y_k'但其他企业的方案不变，那么总利润会更高。

相反，令(y_j)表示单个企业j（其中$j=1,\ldots,m$）的一组利润最大化的生产方案。假设$y=\sum_{j=1}^{m}y_j$在价格向量为p时没有使得利润最大。这意味着另一个生产方案$y'=\sum_{j=1}^{m}y_j'$（其中y_j'在Y_j中）的利润更高：

$$\sum_{j=1}^{m}py_j'=p\sum_{j=1}^{m}y_j'>p\sum_{j=1}^{m}y_j=\sum_{j=1}^{m}py_j.$$

但是审视上述不等式的两侧可知，这意味着某个企业使用y_j'方案的利润比使用y_j的利润更高。 ■

这个命题是说，如果每个企业的利润都实现了最大化，则总利润必定实现了最大化；反过来，若总利润最大，则每个企业的利润必定实现了最大化。我们是根据以下假设推导出这个结论的：总生产可能性是单个企业生产可能性的加总。

由此可知，总净供给函数的构建方法有两种：一种是将单个企业的净供给函数加总；二是先将单个企业的生产集加总，然后在这个总生产集上找到能使利润最大的净供给函数。这两种方法的结果是一样的：都能得到同一个函数。

18.2 麻烦的情形

为简单起见，我们通常假设总净供给函数是良好表现的。但是更一般的分析要求从生产集的潜在性质推导出这个假设。如果生产集是严格凸而且是合理有界的，则不难证明净供给函数是良好表现的。相反，如果生产集具有非凸区域，那么净供给"函数"将是不连续的。这里的引号是想强调在非凸情形下，需求函数不是良好定义的。在某组价格水平下，利润最大化的消费束可能有若干个。如果不连续性程度"较小"，问题不大，但难以作出一般性的结论。

位于连续和不连续之间的情形是规模报酬不变。在第 2 章我们已经知道，在这种情形下，净供给行为让人非常讨厌：它可能为零、无穷大或者一系列产量，具体是哪种情况要取决于价格。尽管这种行为让人讨厌，但是，与规模报酬不变相伴的净供给"函数"关于价格具有一定连续性。

首先要说明的是，净供给"函数"可能并不是函数。根据函数的定义可知，函数要求对于定义域的每一点，值域都有**唯一**一点与之对应。如果生产集是规模报酬不变的，若某个净产出向量y使得利润最大，且最大利润为零，那么任何向量ty（其中$t\geq 0$）也能做到这一点。因此最优净供给解有无穷多个。

这种情形在数学上的处理方法是定义一类称为**映射**（correspondence）的广义函数。在一个映射中，定义域的每个点对应着值域内的若干个点组成的**点集**。如果这个点集是凸的，那么我们说这个映射是一个**凸映射**（convex correspondence）。当然，函数只是凸映射的一

种特殊情形。

不难证明，如果生产集是凸的，则净供给映射是凸映射。而且可以证明，当价格变动时，净供给映射近似连续变动。这一章我们使用的净供给函数的几乎所有结果都可以推广到映射的情形。感兴趣的读者可以参考本章结尾中的注释。然而，为简单起见，我们只分析净供给函数的情形。

18.3 消费者行为

生产行为为我们的消费者模型引入了两个新因素：劳动供给和利润分配。

劳动供给

在纯交换模型中，我们假设消费者拥有由各种商品组成的一个禀赋 ω_i 。如果该消费者将这个商品向量出售，他得到的收入为 $p\omega_i$ 。至于消费者是只消费部分禀赋，还是销售全部禀赋并买回一些商品，并不重要。这是因为，虽然消费者的货币收入可能不同，但他的经济收入是相同的。

如果我们向模型中引入劳动，那么我们引入了一种新的可能性：消费者根据工资率的不同供给不同的劳动数量。

我们回顾一下第 9 章介绍过的那个简单劳动供给模型。在这个模型中，消费者拥有的"时间"禀赋为 $\overline{L}$ ，他必须将这些时间在劳动（ l ）和闲暇（ $L=\overline{L}-l$ ）之间进行分配。消费者关心闲暇 L 和消费品 c 的数量。劳动的价格即工资率，用 w 表示；消费品的价格用 p 表示。消费者可能已经拥有一定数量的消费品禀赋 $\overline{c}$ ，这构成了他的非劳动收入。

我们可以将消费者的最大化问题写为

$$\begin{aligned}&\max u(c,L)\\&s.t.\quad pc=p\overline{c}+w(\overline{L}-L).\end{aligned}$$

上式中的预算约束可以写成我们熟悉的形式

$$pc+wL=p\overline{c}+w\overline{L}.$$

在预算约束的后面这种表达方法中，我们将闲暇看为另外一种商品：消费者拥有闲暇的禀赋为 $\overline{L}$ ，他将这些禀赋（时间）以价格 w 卖给企业，并且在该价格下"买回"某些闲暇。

对于某个消费者拥有很多种劳动的这种更为复杂的情形，上面的这种处理策略同样适用。给定商品和劳动的任何价格向量，消费者可以考虑出售他的禀赋然后买回他想要的商品和闲暇。当以这种方法看待劳动供给问题时，容易看出它正好适用于前面介绍的消费者行为模型。给定任何一个禀赋向量 ω 和任何一个价格向量 p ，消费者求解的问题是

$$\begin{aligned}&\max u(x)\\&s.t.\quad px=p\omega.\end{aligned}$$

这个模型与前面的消费者行为模型唯一的区别是，现在的约束条件变多了；例如，消费者消费的闲暇数量不可能超过 24 小时/天。正式地，这样的约束可以并入第 7 章定义的消费集之中。

利润分配

现在我们开始分析利润分配的问题。在一个资本主义经济中，消费者们拥有企业，有权享受部分利润。我们用一组数字 (T_{ij}) 表示这种所有权关系，其中 T_{ij} 表示消费者 i 在企业 j 利润中的份额。对于任何企业 j，我们要求 $\sum_{i=1}^{n} T_{ij} = 1$，因此它全部为消费者们所拥有。我们将所有权关系视为历史给定的，尽管更复杂的模型可以将股票市场包括进来。

在价格向量 p 下，每个企业 j 选择的生产方案产生的利润为 $py_j(p)$。消费者收到的利润收入等于他从每个企业收到的利润之和：$\sum_{j=1}^{m} T_{ij} py_j(p)$. 消费者的预算约束现在变为

$$px_i(p) = p\omega_i + \sum_{j=1}^{m} T_{ij} py_j(p).$$

我们假设消费者会选择满足他的预算约束的效用最大的商品束。因此，消费者 i 的需求函数可以写为价格向量 p 的函数。我们再一次有必要假设偏好为严格凸的，目的在于保证 $x_i(p)$ 是一个（单值）函数。然而，我们在第 9 章已经知道，在这样的假设下，$x_i(p)$ 是连续的，至少当价格和收入为严格正时，是连续的。

18.4 总需求

将所有的消费者的需求函数加总在一起就得到了总消费者需求函数（aggregate consumer demand function）$X(p) = \sum_{i=1}^{n} x_i(p)$。总供给向量是下列两项的加和：一是来自消费者的供给 $\sum_{i=1}^{n} \omega_i$；二是来自企业的总净供给 $Y(p)$。最后，我们将总超额需求函数（aggregate excess demand function）定义为

$$z(p) = X(p) - Y(p) - \omega.$$

上式是站在需求角度看。若站在供给角度看，也是可行的，但需注意 $z(p)$ 分量的正号和负号的含义：$z(p)$ 的某个分量为负，若相应商品是净超额供给的；$z(p)$ 的某个分量为正，若相应商品是净超额需求的。

我们在纯交换经济中阐述了瓦尔拉斯法则。现在来看含有生产的经济中的瓦尔拉斯法则。

瓦尔拉斯法则。如果 $z(p)$ 的定义如上，则对于所有 p 有 $pz(p) = 0$。

证明。根据 $z(p)$ 的定义将 $pz(p)$ 展开。

$$
\begin{aligned}
pz(p) &= p[X(p)-Y(p)-\omega] \\
&= p\left[\sum_{i=1}^{n} x_i(p) - \sum_{i=1}^{m} y_j(p) - \sum_{i=1}^{n} w_i\right] \\
&= \sum_{i=1}^{n} px_i(p) - \sum_{i=1}^{m} py_j(p) - \sum_{i=1}^{n} pw_i.
\end{aligned}
$$

消费者的预算约束为 $px_i(p)=p\omega_i+\sum_{j=1}^{m}T_{ij}py_j(p)$ 将其代入上式最后一行

$$
\begin{aligned}
pz(p) &= \sum_{i=1}^{n} pw_i(p) + \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}py_j(p) - \sum_{i=1}^{m} py_j(p) - \sum_{i=1}^{n} pw_i \\
&= \sum_{i=1}^{m} py_j(p)\sum_{i=1}^{n} T_{ij} - \sum_{i=1}^{m} py_j(p) = 0,
\end{aligned}
$$

上式最后一行等于 0，这是因为对于每个 j 有 $\sum_{i=1}^{n}T_{ij}=1$。 ∎

此处瓦尔拉斯法则成立的原因和纯经济情形中的原因是一样的：每个消费者满足了他的预算约束，因此经济作为一个整体必须满足总预算约束。

18.5 均衡的存在性

如果 $z(p)$ 是定义在满足瓦尔拉斯法则的价格单纯形上的连续函数，则我们可以使用第 17 章的论断，证明存在一个 p^* 使得 $z(p^*)\le 0$。我们已经知道若每个企业的生产集是严格凸的，则连续性成立。不难看出我们需要总生产可能集是凸的。即使单个企业的技术都是稍微非凸的，由于规模报酬递增的区间很小，由此导致的非连续性在加总后会被抹平。

注意，我们在上面给出的存在性证明方法只适用于我们处理的是需求函数的情形（即单值函数而非一般意义上的映射）。这就要求我们将规模报酬不变技术排除在外，而这种技术又是一种非常重要的情形。因此，我们将给出一个一般性的定理，并讨论这个定理的假设条件的经济含义。

均衡的存在性。某个经济存在着均衡，如果下列假设条件得到满足

（1）每个消费者的消费集是闭的、凸的和有下界的。

（2）任何消费者没有饱和消费束。

（3）对于每个消费者 $i=1,\dots,n$，集合 $\{x_i : x_i \succeq_i x_i'\}$ 和 $\{x_i : x_i' \succeq_i x_i\}$ 都是闭集。

（4）每个消费者都在他的消费集内部拥有一个初始禀赋向量。

（5）对于每个消费者 i，如果 x_i 和 x_i' 是两个消费束，则 $x_i \succ_i x_i'$ 意味着对于任何 $0<t<1$ 有 $tx_i+(1-t)x_i' \succ_i x_i'$；

（6）对于每个企业 j，0 是 Y_j 的元素；

（7） $Y=\sum_{j=1}^{m}Y_j$ **是闭且凸的；**

（8） $Y\cap(-Y)\subset\{0\}$

（9） $Y\supset(-R_+)$。

证明：参见 Debreu(1959)。 ■

尽管这个定理的证明超出了本书的范围，我们至少可以确保我们理解了每个假设的意图。假设（1）和（3）是用来构建存在效用最大化的消费束。假设（1）~（5）是用来构建消费者需求映射的连续性。

假设（6）要求企业总能退出经营，这保证了均衡利润是非负的。假设（7）用于保证每个企业（多值的）净供给函数的连续性；假设（8）意味着生产是不可逆的，因为你不能做到：生产净产出向量 y 然后又用它作为投入生产出所有要素。它的目的是保证可行配置集是有界的。最后，假设（9）说明任何使用所有商品作为投入品的生产方案都是可行的；这个假设在本质上就是自由处置（free disposal）假设；它意味均衡价格是非负的。

18.6 均衡的福利性质

配置 (x,y) 是可行的，若总财产等于总供给：

$$\sum_{i=1}^{n}x_i-\sum_{j=1}^{m}y_i-\sum_{i=1}^{n}w_i=0\,.$$

和以前一样，可行配置 (x,y) 是帕累托有效率的，如果不存在其他的可行集 (x',y') 使得 $x_i'\succ_i x_i$，其中 $i=1,\ldots,n$。

福利经济学第一定理。**如果** (x,y,p) **是一个瓦尔拉斯均衡，则** (x,y) **是帕累托有效率的**。

证明。假设不是，令 (x',y') 是帕累托占优配置（pareto dominating allocation）。于是由于消费者要最大化其效用，我们必然有

$$px_i'>p\omega_i+\sum_{j=1}^{m}T_{ij}py_j\text{，其中 }i=1,\ldots,n\text{。}$$

将所有消费者 $i=1,\ldots,n$ 的这个式子加总，我们得到

$$p\sum_{i=1}^{n}x_i'>p\sum_{i=1}^{n}\omega_i+\sum_{j=1}^{m}py_j.$$

此处我们使用了 $\sum_{i=1}^{n}T_{ij}=1$ 这个事实。

现在我们使用 x' 可行性的定义，用 $\sum_{j=1}^{m}y_j'+\sum_{i=1}^{n}\omega_i$ 代替 $\sum_{i=1}^{n}x_i'$ 可得：

$$p[\sum_{j=1}^{m} y'_j + \sum_{i=1}^{n} \omega_i] > \sum_{i=1}^{n} p\omega_i + \sum_{j=1}^{m} y_j$$

$$\sum_{j=1}^{m} py'_j > \sum_{j=1}^{m} py_j.$$

但是这意味着生产方案(y'_j)的总利润大于生产方案(y_j)的总利润，这违背了企业已实现利润最大化的假设。 ■

其他的基本福利定理证明也不难。我们只给出证明的大概。

福利经济学第二定理。**假设(x^*, y^*)是一个帕累托有效率的配置，其中每个消费者拥有的每种商品数量都为严格正，并且偏好是凸的、连续的和强单调的。假设每个企业的生产可能集Y_j是凸的，其中 $j=1,\ldots,m$。那么存在一个价格向量$p \geq 0$使得：**

（1）若$x'_i \succ_i x_i$，则$px'_i > px_i$，其中$i=1,\ldots,n$；

（2）若y'_j在Y_j中，则$py^*_j \geq py'_j$，其中 $j=1,\ldots,m$。

证明。(概要)和以前一样，令P表示由所有更受偏好的总消费束（aggregate preferred bundles）组成的集合。令F表示所有可行的总消费束组成的集合，即

$$F = \left\{ \omega + \sum_{j=1}^{m} y_j : y_j \in Y_j \right\}$$

则F和P都是凸集，而且由于(x^*, y^*)是帕累托有效率的，F和P是不相交的，我们可使用分离超平面定理（详见第 26 章），找到一个价格向量使得

$$pz' \geq pz'' \qquad \text{对于 } P \text{ 中的所有 } z' \text{ 和 } F \text{ 中的所有 } z'' \text{ 均成立。}$$

偏好的单调性意味着$p \geq 0$。我们可以使用纯交换经济情形下给出的构造证明，在这些价格下每个消费者的效用已达到最大而且每个企业的利润已达到最大。 ■

上面的命题表明，每个帕累托有效率的配置都可以通过对“财富”的合理再分配而实现。我们首先确定我们想要的配置(x^*, y^*)，然后确定相关的价格p。如果我们给予消费者i相当于px_i^*那么大的收入，那么他不会想要改变自己的消费束。

对于这个结果我们的解释方法有好几种：首先，我们可以把它想象为下面的情形：先将所有消费者的原来禀赋（包括商品和闲暇）没收，然后重新分配这些禀赋使得它与人们想要的收入再分配相符。注意这种再分配可能涉及商品的再分配、利润的再分配以及闲暇的再分配。

另一方面，我们可以这么认为：消费者保留各自的原来禀赋，但是现在要让他们缴纳一次性税收。这种税和一般的税收不同，这种税是对“潜在的”的收入而不是“已实现的”收入征收；也就是说，是对劳动的禀赋而不是对出售的劳动征税。消费者无论是否出售劳动都要纳税。在纯交换经济的意义上，对某个人 A 征收一次性税收然后将这笔钱给与另一个人 B，相当于将 A 的一部分劳动禀赋给与 B，然后允许 B 按照市场的工资率出售这些劳动。

当然，个体之间的能力可能不同，换句话说即他们的各种潜在的劳动的禀赋是不同的。在实践中，如果想征收这种一次税，必须观察人们的能力区别，但这很难做到。当个人能力不同时，收入再分配就存在着严重的效率问题。

福利经济学第二定理的显示偏好证明方法

下面我们介绍一种简单但多少不那么直接的第二定理的证明方法，这个方法是基于显示偏好定理基础之上的（详见第 17 章）。

福利经济学第二定理。**假设 (x^*, y^*) 是一个帕累托有效率的配置，并且偏好是局部非饱和的。再假设对于所有 i, j 在初始禀赋 $\omega_i = x_i^*$ 和利润份额 $T_{ij} = 0$ 情形下，存在一个竞争均衡，令其为 (p', x', y')。则事实上 (p', x^*, y^*) 是一个竞争均衡**。

证明。根据命题中关于 x_i^* 的设定可知 x_i^* 满足每个消费者的预算约束，因此必然有 $x_i' \succsim_i x_i$。由于 x^* 是帕累托有效率的，这意味着 $x_i' \sim_i x_i^*$。所以，如果 x_i' 在预算集上能使效用最大，则 x_i^* 也能。

由于偏好是非饱和的，每个消费者的预算约束变成等式，因此

$$p'x_i' = p'x_i^* \qquad i = 1, \ldots, n.$$

将所有消费者 $i = 1, \ldots, n$ 的上式相加，并使用可行配置的定义，我们可得到

$$p'\left(\sum_{j=1}^{m} y_j' + \sum_{i=1}^{n} \omega_i\right) = p'\left(\sum_{j=1}^{m} y_j^* + \sum_{i=1}^{n} \omega_i\right),$$

$$p'\sum_{j=1}^{m} y_j' = p'\sum_{j=1}^{m} y_j^*.$$

因此，如果 y' 使得总利润最大化，则 y^* 也可使得总利润最大化。根据本章第 1 节的那个定理可知，每个企业都已实现了利润最大化。 ■

这个命题表明如果某个帕累托有效率配置 (x^*, y^*) 存在竞争均衡，则 (x^*, y^*) 本身就是一个竞争均衡。你也许会问均衡存在的条件是什么？根据前面关于存在性的讨论可知，有下面两个假设就足够了：（1）所有的需求函数是连续的；（2）满足瓦尔拉斯法则。需求的连续性可从偏好和生产集的凸性推知。瓦尔拉斯法则可通过下列计算进行验证：

$$\begin{aligned} pz(p) &= pX(p) - p\omega - pY(p) \\ &= pX(p) - pX^* - pY(p) \\ &= 0 - pY(p) \le 0. \end{aligned}$$

在上面的模型中我们看到，超额需求的值总是非正的。这是因为我们没有将消费者在企业利润中的份额给与这些消费者。由于这些利润被“扔掉”了，超额需求的值当然为负。然而，仔细审

视第 17 章均衡存在性的证明，就会知道我们其实不需要 $pz(p)\equiv 0$ 的假设；只要有 $pz(p)\le 0$ 就足够了。

这个结果表明，福利经济学第二定理的关键假设条件，其实就是完全竞争均衡存在的条件，即凸性条件。

18.7 生产经济中的福利分析

生产经济中的福利分析方法和交换经济中的福利分析方法大致相同，你不应该为此感到惊讶。唯一真正的问题是在生产经济情形下如何描述可行配置集。

最简单的方法是使用转换函数（参见第 1 章）。我们已经知道，转换函数（transformation function）在下列意义上选择处了有效率的生产方案：当且仅当 $T(y)=0$ 。可以证明几乎任何合理的技术都可以使用转换函数刻画[一]。

于是，福利最大化的问题可以写为

$$\begin{aligned} &\max \quad W(u_1(x_1),\ldots,u_n(x_n)) \\ &s.t. \quad T(X^1,\ldots,X^k)=0. \end{aligned}$$

其中 $X^g=\sum_{i=1}^n x_i^g$ ， $g=1,\ldots,k$ 。这个问题的拉格朗日函数为

$$L=W(u_1(x_1),\ldots,u_n(x_n))-\lambda T(X)=0,$$

它的一阶条件为

$$\frac{\partial W}{\partial u_i}\frac{\partial u_i(x_i^*)}{\partial x_i^g}-\lambda\frac{\partial T(X^*)}{\partial X^g}=0 \qquad i=1,\ldots,n;\ \ g=1,\ldots,k.$$

这些条件可以变形为

$$\frac{\dfrac{\partial u_i(x_i^*)}{\partial x_i^g}}{\dfrac{\partial u_i(x_i^*)}{\partial x_i^h}}=\frac{\dfrac{\partial T(X^*)}{\partial X^g}}{\dfrac{\partial T(X^*)}{\partial X^h}} \qquad i=1,\ldots,n; \qquad g,h=1,\ldots,k$$

福利最大化的一阶条件要求，每对商品之间的边际替代率等于这对商品之间的边际转换率。

18.8 图形分析

我们可以借助类似于埃奇沃思盒的图形分析生产和一般均衡。假设我们考虑的由一个消费者构成的经济。这种情形下消费者的生活非常奇怪：一方面他是一个追求利润最大化的生产者（企

[一] 我们可以将资源禀赋融入转换函数的定义之中。

业）——使用劳动要素生产某种消费品；另一方面他又是一个拥有上述企业、追求效用最大化的消费者。这种情形有时称为鲁宾逊.克鲁索经济（Robinson Crusoe economy）。

在图 18.1 中，我们画出了企业的生产集。注意：劳动的数量用负数表示，因为它是生产过程的投入品；生产技术是规模报酬不变的。

劳动的供给量有上限，我们把劳动的这个最大供给量用 $\overline{L}$ 表示。为简单起见，假设消费者的初始消费品禀赋为零。消费者的消费束由消费品和闲暇组成，他的偏好由图中的无差异曲线给出。均衡工资是多大？

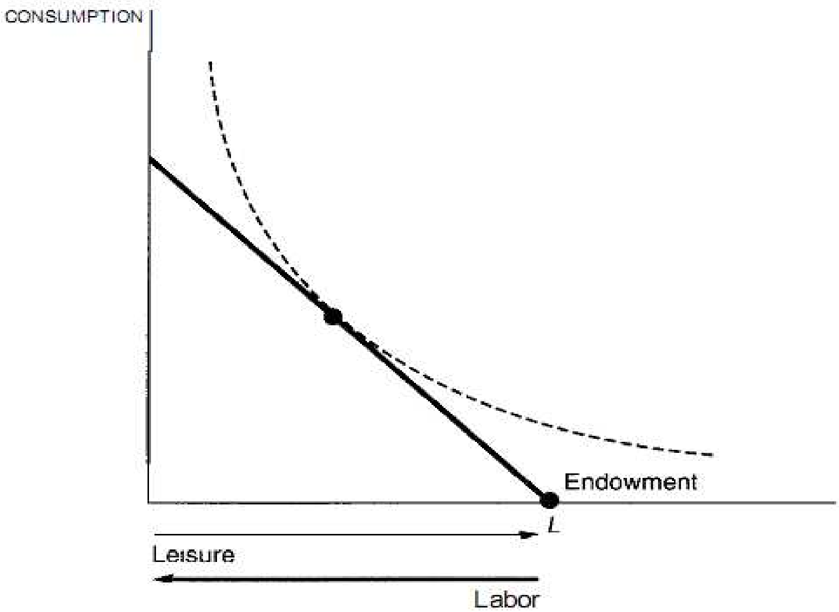


图 18.1：规模报酬不变的鲁滨逊克鲁索经济。劳动的数量用负数表示，生产技术是规模报酬不变的。

如果实际工资（real wage）由生产集的斜率给出，那么消费者的预算集与生产集重合。他的需求束是能给他最大效用的消费束。生产者也愿意生产这个需求束，因为他能得到零利润。因此，消费品市场和劳动市场都出清。

注意：实际工资完全由生产技术决定，而最终产量和消费束都由消费者需求决定。这个结论可以扩展为**非替代性定理**（Non-substitution Theorem）。这个定理是说如果生产过程只有一种天然性投入（nonproduced input）而且技术是规模报酬不变的，则均衡价格和偏好无关——它们完全由生产技术决定。稍后我们将证明这个定理。

规模报酬递减的情形请见图 18.2。在这个图中我们可以找到均衡配置点，因为在这一点上边际替代率等于边际转换率。这一点的斜率给出了均衡实际工资。

当然，在这一实际工资率下消费者的预算线不经过禀赋点 $(0, \overline{L})$。原因在于消费者从企业那里得到了部分利润。企业赚取的利润数额（以消费品的单位数衡量）用纵截距表示。由于消费者

拥有这个企业，全部利润当然归他所有，但要注意这些利润是他的“非劳动”收入。因此，他的预算集正如图中所示，仔细审视一下图即可知道，这两个市场的确都是出清的。

这使得一般均衡模型中的利润有些特殊。在上面的分析中，我们假设技术对劳动投入是规模报酬递减的，但没有解释具体原因。一种可能的原因是存在着固定要素——例如，土地。在这种解释中，罗宾逊的消费品的生产函数取决于劳动投入 L 和（固定的）土地投入 T. 如果我们同时增加这两种要素的投入数量，生产函数可能是规模报酬不变的，但是如果我们固定劳动投入从而将产量只看成劳动这个变量的函数，我们就可能得到规模报酬递减的结果。我们在第 1 章已经知道，对于每个规模报酬递减的技术，如果我们规定它们含有固定要素，那么可以将它们看为规模报酬不变的技术。

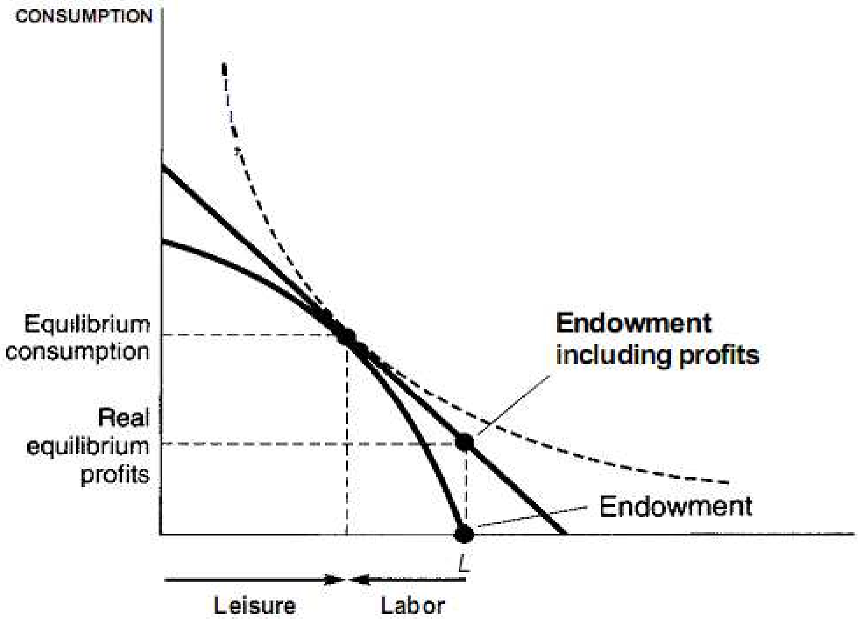


图 18.2：规模报酬递减的罗宾逊克鲁索经济。消费者的预算线不经过禀赋点 $(0,\overline{L})$，因为消费者从企业那里得到了部分利润。

从这个角度来说，“利润”或说非劳动收入可以解释为固定要素的**租金**（rent）。如果我们使用这种解释，那么一般来说利润为零：产品的价值必定等于要素的价值。这一点是从利润的定义得知的。产品价值减去可变要素价值剩下的部分，自动视为固定要素的报酬或租金。

例子：柯布-道格拉斯规模报酬不变的经济

假设某个消费者关于消费品 x 和闲暇 R 的效用函数是柯布-道格拉斯型的：

$$u(x,R)=a\ln x+(1-a)\ln R.$$

该消费者的禀赋为一单位的劳动和一单位的闲暇。另外，还存在着一个企业，该企业的技术是规模报酬不变的：$x=aL$。

稍微分析一下可知均衡实际工资必定等于劳动的边际产量，因此 $w^*/p^*=a$ 。消费者的最大化问题为

$$\begin{aligned}&\max\ a\ln x+(1-a)\ln R\\&s.t.\quad px+wR=w.\end{aligned}$$

在写上述问题中的预算约束时，我们用到了均衡利润为零的事实。我们已经知道柯布-道格拉斯效用函数的需求函数的形式为 $x(p)=am/p$ ，其中 m 为货币收入，我们发现

$$\begin{aligned}x(p,w)&=a\frac{w}{p}\\R(p,w)&=(1-a)\frac{w}{w}=1-a.\end{aligned}$$

因此，劳动的均衡供给量为 a ，均衡产量为 a^2 。

例子：规模报酬递减的经济

假设消费者的效用函数如上例所示，但现在生产者的生产函数为 $x=\sqrt{L}$ 。我们武断地将产品价格标准化为 1。利润最大化问题为

$$\max\ L^{1/2}-wL.$$

这个问题的一阶条件为：

$$\frac{1}{2}L^{-1/2}-w=0.$$

求出企业的需求函数和供给函数：

$$\begin{aligned}L&=(2w)^{-2}\\x&=(2w)^{-1}.\end{aligned}$$

将它们代入目标函数可得到利润函数的表达式,

$$\pi(w)=(2w)^{-1}-w(2w)^{-2}=(4w)^{-1}.$$

现在消费者的收入中需要加入利润收入，因此他对闲暇的需求为

$$R(w)=\frac{(1-a)}{w}(w+\frac{1}{4w})=(1-a)(1+\frac{1}{4w^2}).$$

根据瓦尔拉斯法则可知，我们只需要找到能使市场出清的实际工资即可：

$$\frac{1}{4w^2}=1-(1-a)(1+\frac{1}{4w^2}).$$

求解该方程，我们得到

$$w^* = \left(\frac{2-a}{4a}\right)^{1/2}.$$

因此，均衡利润水平为

$$\pi^* = \frac{1}{4}\left(\frac{2-a}{4a}\right)^{-1/2}.$$

下面我们介绍这个问题的另外一种求解方法。正如前面指出的，我们对技术呈现规模报酬递减的原因解释是存在着固定要素。令这种固定要素为“土地”，我们以单位数衡量它的大小，因此土地的总量为$\overline{T}=1$。令生产函数为$L^{1/2}T^{1/2}$。注意这个生产函数是规模报酬不变的，而且当$T=1$时它与原来的技术相同。令土地的价格为q。

该企业的利润最大化问题为

$$\max L^{1/2}T^{1/2} - wL - qT,$$

它的一阶条件为

$$\frac{1}{2}L^{-1/2}T^{1/2} - w = 0$$
$$\frac{1}{2}L^{1/2}T^{-1/2} - q = 0.$$

在均衡时土地市场将出清，因此$T=1$。将其代入上面的方程可得

$$L = (2w)^{-2}$$
$$L = (2q)^2.$$

联立这两个式子可得$q = 1/4w.$

现在消费者的收入由两部分构成：一是来自他的劳动禀赋的收入$w\overline{L}=w$；二是来自他的土地禀赋的收入$q\overline{T}=q$。因此，他的闲暇需求为

$$R = (1-a)\frac{m}{w} = (1-a)\frac{(w+q)}{w}.$$

令劳动的需求等于劳动的供给可得到均衡工资

$$w^* = \left(\frac{2-a}{4a}\right)^{1/2}.$$

土地的均衡租金为

$$q^* = \frac{1}{4}\left(\frac{2-a}{4a}\right)^{-1/2}.$$

注意这个结果和前面那种方法求得的结果是相同的。

18.9 非替代性定理

本节将要证明前面提及的非替代性定理。假设有 n 个行业，它们的产品分别为 $y_i, i=1,...,n$ 。每个行业只生产一种产品；不允许联合生产。生产的天然性投入只有一种，用 y_0 表示。我们一般认为这种天然性商品为劳动。假设这 $n+1$ 种商品的价格为 $w=(w_0, w_1,..., w_n)$ 。

和往常一样，所能确定的均衡价格是以相对价格形式表现的。我们假设劳动是每个行业的必要投入要素。因此，均衡时 $w_0>0$ ，我们可以将劳动选为计价物（numeraire）；也就是说我们可以武断地设定 $w_0=1$ 。

假设生产技术是规模报酬不变的。在第 5 章我们已经知道，这种技术意味着每个行业的成本函数可以写为 $c_i(w, y_i)=c_i(w)y_i$ ，其中 $i=1,...,n$ 。函数 $c_i(w)$ 是单位成本函数，即在价格为 w 时（注意，这是相对价格，计价标准为 $w_0=1$ ），该行业生产一单位产品所花费的成本。

我们还假设劳动对于生产是必不可缺的，因此劳动的单位要素需求是严格正的。用 x_i^0 表示当 $y=1$ 时企业 i 对要素 0 的需求，我们可以利用成本函数的可微性质将 x_i^0 写为

$$x_i^0(w)=\frac{\partial c_i(w)}{\partial w_0}>0.$$

注意，这意味着成本函数关于 w_0 严格增。由于成本函数关于至少其中一种产品的价格是严格增的，对于 $t>1$ 有 $c_i(tw)=tc_i(w)>c_i(w).$

非替代性定理。假设生产过程只有一种天然性要素，这种要素对于生产必不可缺；假设不允许联合生产，而且生产技术是规模报酬不变的。令 (x, y, w) 为一个瓦尔拉斯均衡，其中 $y_i>0, i=1,...,n$ 。则 w 是 $w_i=c_i(w), i=1,...,n$ 的唯一解。

证明。如果 w 是规模报酬不变经济中的一个均衡价格向量，那么每个行业的利润必定为零；即

$$w_i y_i - c_i(w) y_i = 0 \qquad i=1,...,n.$$

由于 $y_i>0$ （其中 $i=1,...,n$ ），这个条件可以写为

$$w_i - c_i(w) = 0 \qquad i=1,...,n.$$

上式是说任何均衡价格必须满足价格等于平均成本这个条件。由于 $w_0>0$ 而且劳动是生产的不可缺少的要素，必有 $c_i(w)>0$ 。而这又意味着 $w_i>0$ ，其中 $i=0,1,...,n$ 。换句话说，所有均衡价格向量都是严格正的。

我们将证明这样的均衡价格向量是唯一的。假设 w' 和 w 是上述方程组的两个不同解。定义

$$t=\frac{w'_m}{w_m}=\max_i \frac{w'_i}{w_i}.$$

这里，两个向量的分量的最大比值出现在分量 m 之处——此处 w'_m 是 w_m 的 t 倍。

假设$t>1$。我们可得到下面一串不等式：

$$w'_m =_1 tw_m =_2 tc_m(w) =_3 c_m(tw) >_4 c_m(w') =_5 w'_m.$$

这些等式和不等式成立的理由分别为：

（1）t的定义；（本句前面的序号（1）对应着上式中的第一个等式，即标有下标 1 的那个式子；以下类推。）（2）w是解的假设；（3）成本函数的线性齐次性；（4）t的定义，假设$t>1$；以及成本函数关于要素价格向量是严格单调的；（5）w'解的假设。

因此，假设$t>1$使我们得到了一个矛盾的结果，所以$t\le 1$，从而$w\ge w'$。在上面的论证中，w和w'的地位是对称的，因此我们也有$w'\ge w$。联立这两个不等式可知，$w'=w$。 ■

这个定理是说，如果某经济体存在着均衡价格向量，那么这个向量必定是$w_i\ge c_i(w)$的解，其中$i=1,...,n$。令人惊讶的是，w一点也不取决于需求条件，即w完全独立于消费者的偏好和禀赋。

我们用术语**技术**（technique）表示生产一单位产品所必需的要素需求。令w^*为满足零利润条件的价格向量。那么我们可以确定企业的均衡技术，方法是成本函数关于每种要素j价格微分：

$$x_i^j(w^*)=\frac{\partial c_i(w^*)}{\partial w_j}.$$

由于均衡价格独立于需求条件，均衡技术的选择将和需求条件无关。无论消费者的需求如何变动，企业也不会用其它技术替代均衡技术；这就是非替代定理名称的由来。

18.10 一般均衡中的行业结构

在瓦尔拉斯模型中，企业的数量是给定的。在第 13 章我们认为行业中的企业数量是个变量。我们该如何调和这两个模型？

我们首先考虑规模报酬不变的情形。在这种情形下，我们知道与均衡相容的唯一利润最大水平就是零利润。而且，在与零利润相容的价格水平上，企业愿意生产任意产量。因此，在这种情形下，经济的行业结构是不确定的——企业根本不关心它占有多少市场份额。如果企业数量是个变量，它也是不确定的。

现在考虑规模报酬递减的情形。如果所有的技术都是规模报酬递减的，我们知道存在着均衡利润。在直到目前我们讨论的一般均衡模型中，企业没有理由获得相同的利润。人们通常认为企业利润相同的原因是假设企业会进入利润最高的行业；但是如果企业的数量是固定的，这种情形不能发生。

如果企业的数量是可变的，实际情况是怎样的呢？企业很有可能会进入。如果技术真是规模报酬不变的，那么企业的最优规模是无穷小，这是因为拥有两个小企业总比拥有一个大企业好。因此，我们预期企业会持续进入，不断压低利润水平。在长期均衡时，企业的数量是无穷大的，每个企业的规模都是无穷小。

这样的情形似乎非常不切合实际。一种解释方法是使用第 13 章提及的结论：如果我们总是可以复制企业，那么唯一合理的长期技术就是规模报酬不变的技术。因此，技术呈现规模报酬递减必定是由于存在固定要素。在这种解释中，均衡“利润”真得应该看作是固定要素的报酬。

注释

非替代性定理参见 Samuelson（1951）。本章对该定理的分析采纳的是 von Weizsaker(1971) 的方法。

习题

18.1 某个经济体中有两种天然性（nonproduced）生产要素——劳动和资本，还有两种产品——苹果和手帕。苹果和手帕的生产技术都是规模报酬不变的。手帕生产只使用劳动，而苹果生产则要使用劳动和土地。有 N 个相同的人，每个人的初始禀赋为 15 单位劳动和 10 单位土地。每个人的效用函数都具有下列形式$U(A,B)=c\ln A+(1-c)\ln B$，其中$0<c<1$，A 和 B 分别为每个人消费的苹果和手帕数量。苹果是用固定系数的（fixed-coefficients）技术生产的，每生产一单位苹果需要使用一单位劳动和一单位土地。而手帕的生产则只使用劳动，生产一单位手帕需要一单位劳动。令劳动为这个经济体的计价物。

（a）计算该经济体得竞争均衡价格和消费量。

（b）参数c取何值时，土地禀赋的微小变动不会导致竞争均衡价格的变动？

（c）参数c取何值时，土地禀赋的微小变动不会导致竞争均衡消费量的变动？

18.2 某个经济体有两个企业和两个消费者。企业 1 完全为消费者 1 所拥有。它使用石油（x）生产枪支（g），生产函数为$g=2x$。企业 2 完全为消费者 2 所拥有，它使用石油（x）生产黄油（b），生产函数为$b=2x$。每个消费者原有 10 单位石油。消费者 1 的效用函数为$u(g,b)=g^{0.4}b^{0.6}$，消费者 2 的效用函数为$u(g,b)=10+0.5\ln(g)+0.5\ln(b)$。

（a）计算枪支、黄油和企业的市场出清价格。

（b）计算每个消费者消费的枪支和黄油数量。

（c）计算每个企业使用的石油数量。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 19 章：时间

曹乾　译

(东南大学　caoqianseu@163.com)

19 时间

在本章我们分析跨时期的消费者和经济体的行为。我们将看到，在某些情形下，跨期行为可以视为前面讨论过的静态模型的简单扩展。然而，时间也对偏好和市场施加了有趣而特殊的结构。由于未来具有不确定性，讨论涉及不确定性的问题就显得自然合理。

19.1 跨期偏好

我们的标准消费者选择理论足以描述跨期选择（intertemporal choice）。选择的对象——消费束——现在变为一定时期的消费流。我们假设消费者在这些消费流上的偏好满足通常的正则条件。按照标准的思路，这些偏好通常用效用函数代表。然而，和期望效用最大化的情形一样，我们考虑的是一类特殊的选择问题。这意味着偏好具有特别的结构，从而它产生的效用函数也具有特别的形式。一种非常普遍的选择问题涉及在时间上满足可加性的效用函数，即

$$U(c_1,...,c_T)=\sum_{t=1}^{T}u_t(c_t).$$

其中$u_t(c_t)$是时期t消费的效用。这个函数也可进一步具体为时间静态形式

$$U(c_1,...,c_T)=\sum_{t=1}^{T}\alpha^t u(c_t).$$

在这个情形下，每个时期的效用函数是相同的，但是时期t的效用要乘以一个折扣因子α^t。

注意这个函数的结构与期望效用结构的相似性。在期望效用模型中，消费者在每种自然状态下的效用是相同的，但每一种自然状态下的效用要乘以该状态发生的概率。事实上，仿照期望效用公理，从对潜在偏好的限制角度，我们可以给出这类具有时间可加性效用函数的合理性理由。

假设未来的消费可能性是不确定的。正如我们前面看到的，给定一组公理，我们可以选择出某个效用函数，使得不同自然状态的效用是可加的。然而，情形更有可能是这样的：效用函数的一种单调变换满足不同状态的效用的可加性，而另外一种单调变换则使得不同时期的效用是可加的。没有理由硬要规定效用函数同时满足跨期选择和不确定选择情形下的可加性。尽管如此，最常见的做法还是假设跨期瞎用函数同时满足这两种情形的可加性。这样的假设不怎么符合实际，但好处是计算简便。

19.2 跨期最优化：两个时期

在第 11 章我们已研究过一个简单的两时期的投资组合模型。此处我们分析如何将这个模型扩展为多时期模型。这个例子的目的是说明如何使用动态规划（dynamic programming）方法，这种方法将多时期的最优化问题分解为两期最优化问题来求解。

我们首先回顾一下两时期模型。用(c_1, c_2)表示这两个时期中的消费。消费者在时期 1 的禀赋为w_1元，他可以将他的财富投资于两种资产身上。一种资产的收益率R_0是固定的，另外一种资产的收益率$\tilde{R}_1$是随机的。我们通常将收益率视为总收益率（total returns），因为这比较方便，总收益率等于 1 加上收益率。

假设消费者决定在时期 1 消费c_1元，并将剩下财富投资于风险资产和无风险资产，投资比例分别为x和$1-x$。在这个投资组合中，$(w_1-c_1)x$元的收益率为R_0，$(w_1-c_1)(1-x)$元的收益率为$\tilde{R}_1$。因此他在时期 2 的财富——这等于他在时期 2 的消费——为

$$\tilde{w}_2 = \tilde{c}_2 = (w_1 - c_1)\left[\tilde{R}_1 x + R_0(1-x)\right] = (w_1 - c_1)\tilde{R}.$$

其中$\tilde{R} = \tilde{R}_1 x + R_0(1-x)$为消费者的**投资组合的收益**（portfolio return）。注意$\tilde{R}$一般是一个随机变量，这是因为$\tilde{R}_1$是个随机变量。

由于投资组合的收益是不确定的，消费者在时期 2 的消费是不确定的。我们假设消费者的效用函数具有下列形式

$$U(c_1, \tilde{c}_2) = u(c_1) + \alpha E u(\tilde{c}_2),$$

其中$\alpha < 1$是折扣因子。

令$V_1(w_1)$为消费者在时期 1 拥有财富w_1可以实现的最大效用：

$$V_1(w_1) = \max_{c_1, x} u(c_1) + \alpha E u[(w_1 - c_1)\tilde{R}]. \tag{19.1}$$

函数$V_1(w_1)$本质上是间接效用函数：它给出了财富效用函数的最大值。

将（19.1）中的目标函数分别对c_1和x微分，可得到一阶条件

$$u'(c_1) = \alpha E u'(\tilde{c}_2)\tilde{R} \tag{19.2}$$

$$E u'(\tilde{c}_2)(\tilde{R}_1 - R_0) = 0. \tag{19.3}$$

（19.2）式是跨期最优化的条件：它是说时期 1 的消费的边际效用必须等于时期 2 消费的期望边际效用的折现。（19.3）式是投资组合最优化的条件：它是说将一小笔钱从无风险资产移动到风险资产的期望边际效用为零。我们在第 11 章已分析过类似的一阶条件。

给定上述含有两个未知数（x和c_1）的两个方程，我们一般可以解出最有消费和投资组合选择。在下一节我们将举例说明如何求解。

19.3 跨期最优化：多个时期

假设有$T>2$个时期。如果$(\tilde{c}_1,\ldots,\tilde{c}_T)$是某个（可能随机的）消费流，我们假设消费者根据下列效用函数评价这个消费流

$$U(\tilde{c}_1,\ldots,\tilde{c}_T)=\sum_{t=0}^{T}\alpha^T Eu(\tilde{c}_t).$$

如果消费者在时期t的财富为w_t，他投资于风险资产的钱数占投资总钱数（总财富减去消费）的比例为x_t，那么他在时期$t+1$的财富为

$$\tilde{w}_{t+1}=[w_t-c_t]\tilde{R},$$

其中$\tilde{R}=\tilde{R}_1x+R_0(1-x)$是时期$t$和时期$t+1$之间的（随机）投资组合收益率。

为了求解跨期最优化问题，我们使用动态规划的方法将其分解为一系列的两期最优化问题。考虑时期$T-1$。如果消费者在此期的财富为w_{T-1}，他能得到的最大效用为

$$V_{T-1}(w_{T-1})=\max_{c_{T-1},x_{T-1}} u(c_{T-1})+\alpha Eu\left[(w_{T-1}-c_{T-1})\tilde{R}\right]\tag{19.4}$$

上式和（19.1）式是一样的，只不过此处用下标$T-1$替换了（19.1）的下标 1。这个最大化问题的一阶条件为

$$u'(c_{T-1})=\alpha Eu'(\tilde{c}_T)\tilde{R}\tag{19.5}$$

$$Eu'(\tilde{c}_T)(\tilde{R}_1-R_0)=0.\tag{19.6}$$

我们已经知道如何求解这个问题，至少在原理上知道怎么做；也知道如何确定间接效用函数$V_{T-1}(w_{T-1})$。

现在我们返回到$T-2$期。如果消费者选择(c_{T-2},x_{T-2})，那么在$T-1$期，他的（随机）财富为

$$\tilde{w}_{T-1}=\left[w_{T-2}-c_{T-2}\right]\tilde{R}.$$

在这个财富上他可以实现的期望效用为$V_{T-1}(w_{T-1})$。因此，该消费者在$T-2$期的最大化问题为

$$V_{T-2}(w_{T-2})=\max_{c_{T-2},x_{T-2}} u(c_{T-2})+\alpha Eu\left[(w_{T-2}-c_{T-2})\tilde{R}\right]$$

这个问题和（19.4）的问题类似，但是“时期 2”的效用由间接效用函数$V_{T-1}(w_{T-1})$而不是直接效用函数给出。

时期$T-2$的最大化问题的一阶条件为

$$u'(c_{T-2})-\alpha EV'(\tilde{c}_{T-1})\tilde{R}=0\tag{19.7}$$

$$EV'(\tilde{w}_{T-1})(\tilde{R}_1-R_0)=0.\tag{19.8}$$

和以前一样，（19.7）式是跨期最优化条件：当前消费的边际效用必须等于未来消费的**间接**

边际效用的折现值。(19.8)式是投资组合的最优化条件。

我们可以使用这些条件求解$V_{T-2}(w_{T-2}),...,V_1(w_1)$。给定间接效用函数$V_t(w_t)$，$T$期的跨期最优化问题就变成了一系列两期最优化问题。

例子：对数效用

假设$u(c)=\ln(c)$。那么（19.5）和（19.6）中的一阶条件变为

$$\frac{1}{c_{T-1}}=\alpha\frac{\tilde{R}}{[w_{T-1}-c_{T-1}]\tilde{R}}=\frac{\alpha}{[w_{T-1}-c_{T-1}]} \quad (19.9)$$

$$0=E\left[\frac{\tilde{R}_1-R_0}{[w_{T-1}-c_{T-1}]\tilde{R}}\right]. \quad (19.10)$$

注意在（19.9）中收益率抵消掉了，这是对数效用的非常方便的性质。

从（19.9）中解出c_{T-1}，可得

$$c_{T-1}=\frac{w_{T-1}}{1+\alpha}.$$

将上式代入目标函数可得到间接效用函数：

$$V_{T-1}(w_{T-1})=\ln\frac{w_{T-1}}{1+\alpha}+\alpha E\ln\frac{\alpha w_{T-1}\tilde{R}}{1+\alpha}$$

使用对数的性质可得，

$$V_{T-1}(w_{T-1})=(1+\alpha)\ln w_{T-1}+\alpha E\ln\tilde{R}+\alpha\ln\alpha-(1+\alpha)\ln(1+\alpha).$$

注意下面这个重要性质：间接效用函数V_{T-1}是财富的对数。随机收益率影响V_{T-1}的可加性；但它不影响财富的**边际**效用，因此它不进入相应的一阶条件。

由此可推知，时期$T-2$最大化问题的一阶条件为

$$\frac{1}{c_{T-2}}=\frac{\alpha(1+\alpha)}{[w_{T-2}-c_{T-2}]} \quad (19.11)$$

$$0=E\left[\frac{\tilde{R}_1-R_0}{[w_{T-2}-c_{T-2}]\tilde{R}}\right]. \quad (19.12)$$

这些条件非常类似时期$T-1$最大化问题的条件；与（19.9）相比，（19.11）在右侧多出了一个因子即$(1+\alpha)$；而（19.2）的形式则与（19.10）相同。由此可知，消费者在每一时期选择的投资组合都是相同的——和他求两期最优化的解是相同的；消费者在$T-1$时期的

消费选择总是与该期的财富成正比。

19.4 跨期情形下的一般均衡

我们以前说过，阿罗-德布鲁一般均衡模型中的商品是非常一般的。因此，商品可以按人们关心的任何特征划分。例如，如果他们关心的是何时得到商品，那么在**不同**时间得到的“同种”商品也应该视为不同商品。如果他们关心的是在何种环境下得到商品，那么商品应该按照被提供时的自然状态进行区分。

当我们按照这些方法区分商品时，我们就可以做到以全新的而且更深入的方法理解均衡价格的作用。例如，我们若考虑的是一个简单的一般均衡模型，其中只有一种商品。消费者可以获得该消费品的时期分别为$t=1,...,T$。由上面的评价可知，我们可以把这种商品视为T种不同的商品，令c_t表示时期t的消费量。

在纯交换经济模型中，消费者i在时期t将有一些消费禀赋c_{it}。在含有生产的经济模型中，将存在某种技术使得时期t的消费可以转换为未来时期的消费。放弃某一时期的消费，消费者可以在未来享受（更多的）消费。

消费者在消费流上具有偏好，而且存在可让不同时点商品进行交易的市场。这种市场的一种组织方式是使用**阿罗-德布鲁证券**（Arrow-Debreu securities）。这些证券形式特别：证券t在时期t支付一元报酬，在其他时期支付零元。这样的证券在现实世界中是存在的；它们就是纯折扣债券（pure discount bonds）。纯折扣债券在某个特定日期支付一定金额的收益（例如 10,000 元）。

这个模型具有标准阿罗-德布鲁模型的所有要素：偏好、禀赋和市场。我们可以运用标准的存在性结论证明，使所有市场同时出清的阿罗-德布鲁证券的均衡价格（p_t）必然存在。注意，p_t是在时期零对时期t交付的商品支付的价格。在这个模型中，所有金融交易都在零期进行而消费则是跨期的。

在现实生活的跨期市场中，我们通常用另外一种方法衡量远期价格（future prices），这就是利率。想象有一家银行它经营以下生意：消费者在时期零每存入1元，在时期t银行向消费者支付$(1+r_t)$元。我们说银行的存款利率为r_t。利率r_t和阿罗-德布鲁的价格p_t有什么样的关系？

假设某个人在时期零持有 1 元钱。他可以将钱存入银行，这样在时期t他可以得到$(1+r_t)$元。或者他可以购买阿罗-德布鲁证券t。如果该证券的价格为p_t，他可以购买$1/p_t$单位的证券。

由于每单位证券在时期t的价值为 1 元，他在时期t可以得到$1/p_t$元。显然，不管他采用哪种投资方案，投资收益必定相等，因此

$$1 + r_t = \frac{1}{p_t}.$$

这意味着利率等于阿罗-德布鲁价格的倒数减去 1。

我们可以使用阿罗-德布鲁价格评估消费流的价值，评估方法和往常一样。例如，下飞着的预算约束的形式为

$$\sum_{t=1}^{T} p_t c_t = \sum_{t=1}^{T} p_t \overline{c}_t.$$

使用价格和利率的关系，我们可以将上式写为

$$\sum_{t=1}^{T} \frac{c_t}{1+r_t} = \sum_{t=1}^{T} \frac{\overline{c}_t}{1+r_t}.$$

因此，预算约束采取的形式是消费折现现值必定等于禀赋折现现值。

由于它的结构和前面说过的阿罗-德布鲁模型是相同的，相同定理必定成立：在各种凸假设下均衡存在且有效率。

无限情形

在很多情形下，用有限的时间衡量活动的期限是不合适的，因为人们会合理预期经济将“无限长”地持续下去。然而，如果我们使用的是无限长的时期，均衡的存在性和福利定理将存在某些麻烦之处。

第一组问题是技术性问题：含有无限多个自变量的函数的合理定义是什么样的？相应的不动点定理或分离超平面定理是什么样的？我们可以借助各种数学工具解决这些问题；大多数问题在本质上和处理技术有关。

然而，无限期模型也存在着某些根本性的特殊之处。也许最著名的例子是**叠代模型**（overlapping generations model），这个模型也称为**纯消费借贷模型**（pure consumption-loan model）。考虑具有如下结构的经济体。在每一期，有 n 个人出生，每个人的寿命都是两期。因此，时期 1 之后的每个时点上都有 $2n$ 个人：n 个年青人和 n 个老年人。每个人在出生时拥有 2 单位的消费禀赋，而且他在年青时消费还是年老时消费这两个选择之间是无差异的。

这个简单的模型毫无疑问存在着均衡。显然每个人都消费自己的禀赋就是一个均衡。这个均衡的支撑价格是对于所有 t， $p_t = 1$。然而这个均衡不是帕累托有效率的！

为了看清这一点，假设第 $t+1$ 代的每个人将他的一单位禀赋转移给第 t 单位的人。现在第一代的人的状况变好了，因为这一代的每个人在生存起见都消费了 3 单位。哪一代人的状况都没有变坏，因为他们在年青时的付出到了年老时得到了补偿。这意味着我们在原来的均衡基础上找到了一个帕累托改进。

现在思考一下，为什么在这个模型中第一福利定理“不成立”？症结在于假设商品数量无限多。如果商品在各期的均衡价格都是 1，那么总消费流和总禀赋的价值都是无穷大的。这样在我们使用反证法证明第一福利定理时，最后一步得到的矛盾就不再成立，证明失败。

这个例子很简单，但是由此的结果却是可靠的。在将有限维度的模型推广到无限维度的模型时你必须非常小心。

19.5 不同自然状态情形下的一般均衡

在前面我们说过人们可能关注商品得到的具体环境或者说在什么样的自然状态下得到这些商品的。毕竟下雨时的雨伞和不下雨时的雨伞是不同的商品。

假设市场在时刻零开始交易，但是在时刻 1 存在着某些不确定性，尽管交易仍会发生。具体地说，假设在时刻 1 自然状态有两种：一是下雨；二是不下雨。

假设人们签订的是**条件合同**（contingent contracts）：“消费者 i 向该合同的持有者交付一单位商品 j 当且仅当天下雨。”时刻 0 的交易是合同交易，即承诺如果某种自然状态发生则在未来提供某种商品或者服务。

我们可以想象存在进行这种合同交易的市场，在合同的任何价格向量下，消费者可以根据他们的偏好和生产技术决定各种合同的需求量和供给量。注意，合同是在时刻 0 签订但却在时刻 1 执行，而且只有在时刻 1 发生某种既定的自然状态的条件下，合同才执行。和往常一样，均衡价格向量是使得任何合同都不存在超额需求的价格向量。从抽象理论的角度看，合同和其他商品没有什么区别。因此，标准的均衡存在性和效率结果仍然适用。

正确理解这个效率结果是重要的。偏好的定义域是彩票空间。如果冯.诺依曼-摩根斯坦公理得以满足，定义在随机事件上的偏好可用期望效用函数刻画。因此，我们说不存在可行配置使得所有消费者状况变好，就是说不存在任何形式的条件合同使得每个人的期望效用增加。

现实生活中存在着类似条件合同的合同。可能最常见的例子就是保险合同。保险合同给付一定资金的充分必要条件是某些既定事件发生。然而，必然承认严格意义上的条件在实践中非常罕见。

注释

Ingersoll（1987）介绍了几个动态投资组合优化模型及其解；Geanakoplos（1987）很好地综述了叠代模型。

习题

19.1 考虑本章 19.4 使用过的对数效用的例子。证明任意时期t的消费量为

$$c_t = w_t/[1+\alpha+\alpha^2+...+\alpha^{T-t}] = \frac{1-\alpha}{1-\alpha^{T-t+1}} w_t.$$

19.2 考虑下列“租金稳定”（rent stabilization）计划。允许房东每年提高租金，每年租金提高额为通货膨胀率的 3/4。新建公寓的房东可以任意设定自己公寓的初始租金。租金稳定计划的倡议者认为既然新公寓的初始租金可以设定为任意水平，这样的措施不会打击新公寓的供给。我们用一个简单的模型分析这种说法。

假设公寓持续 2 期。令r为名义利率，π为通货膨胀率。假设没有租金稳定计划时，时期 1 的租金为p，时期 2 的租金为$(1+\pi)r$。令c表示建造新公寓的边际成本，它是个常数。令每个时期的公寓需求函数为$D(p)$。最后，令K表示租金受到控制的公寓的供给。

（a）在没有租金稳定计划时，时期 1 的租金p和建造新公寓的边际成本之间必定有什么样的关系？

（b）如果租金稳定计划实施，上述关系变为什么样的？

（c）画出简单的需求-供给图并说明在没有租金稳定计划时新公寓的数量。

（d）租金稳定计划会导致更多还是更少的新公寓？

（e）在租金稳定计划的情形下，新公寓的均衡租金是变高了还是变低了？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 20 章：资产市场

曹乾 译

(东南大学 caoqianseu@163.com)

20 资产市场

研究资产市场需求需要使用一般均衡方法。下面我们将看到，某种给定资产的均衡价格严重地依赖于它的价值与其他资产价值如何相关。因此，多元资产（multi-asset）的定价内在地涉及到一般均衡思想。

20.1 确定性下的均衡

在资产市场的研究中，核心问题是什么因素决定了资产价格的差异。在确定性的世界中，资产市场的分析很简单：资产的价格就是它的收入流现期折现价值。如果不是如此，将存在无风险**套利**(arbitrage)的可能性。

例如，在一个简单的两期模型中，我们假设某种资产 0 的**总收益率**（total return）是确定的（无风险的），假设为 R_0，即在资产 0 上投资 1 元在下一期肯定能得到 R_0 元。如果 R_0 是资产 0 的总收益率，则 $r_0 = R_0 - 1$ 为（净）**收益率**。

还存在另外一种资产 a，它在下一期的价值为 V_a。资产 a 的现期均衡价格是多少？

在确定性的世界中，这个问题的答案很简单：资产 a 的现期价格必定等于它的现值，

$$p_a = \frac{V_a}{R_0} = \frac{V_a}{r_0 + 1}. \qquad (20.1)$$

如果答案不是如此，那么有人肯定会套利赚钱。如果 $p_a > V_a / R_0$，那么拥有资产 a 的人会出售资产 a 然后再投资于无风险的资产。下一期他拥有的钱数 $p_a R_0 > V_a$。既然至少有一人想卖掉资产 a，那么 p_a 不可能是均衡价格。

我们还有一种方法写出（20.1）式的均衡条件，这就是使用资产 a 的总收益率 $R_a = V_a / p_a$。如果我们将（20.1）的两侧同除以 p_a，整理可得

$$p_a = \frac{V_a}{p_a} = R_0. \qquad (20.2)$$

这个式子是说均衡时，所有具有确定收益率的资产的收益率必定是**相同的**，因为没有人愿意持有收益率较低的资产。

20.2 不确定性下的均衡

在一个具有不确定性的世界中，资产的期望收益将取决于它的风险性。通常，我们认为在其他条件相同的情形下，资产的风险性越大，它的价格就越低。换句话说，资产的风险

越大，为了诱使人们购买该资产，它的期望收益应越大。

类比（20.2）式，我们可以将资产 a 的期望收益（expected return）$\overline{R}_a$ 写为

$$\overline{R}_a = R_0 + \text{资产}\, a\, \text{的风险溢价}$$

上式右侧是无风险收益率加上资产 a 的风险溢价（risk premium）。我们也可以将这个条件写为

$$\overline{R}_a - R_0 = \text{资产}\, a\, \text{的风险溢价}$$

上式左侧称为资产 a 的超额收益（excess return）。这个式子断言：均衡时，每种资产的超额收益等于它的风险溢价。

当然，这些式子都只是相关术语的定义。资产市场理论试图使用诸如消费者偏好、资产收益模式等这些“基本概念”推导出风险溢价的表达式。

这一分析涉及一般均衡的思想，这是因为某种风险资产的价格内在地取决于其他资产存在与否，这些其他资产是该风险资产的补充品或替代品。因此，在资产定价的绝大多数模型中，某种资产的价值最终取决于它如何与其他资产协（同）变（动）（covary）。

令人惊讶的是，尽管资产定价模型之间的差异似乎很大，但都基本体现了上述思想。在本章我们将推导和比较几个资产定价模型。

20.3 符号

此处列出了本章用到的所有符号。这种做法的好处是在有需要时你可以容易地查看这些符号的定义。其中一些术语的定义将在相关章节进一步详细介绍。

我们一般分析两期模型，当前时期视为时期 0。各种资产在时期 1 的价值是不确定的。我们使用自然状态（states of nature）的概念刻画这种不确定性。也就是说，我们假设存在着各种可能的结果 $s = 1, \ldots, S$ ，而且每种资产在时期 1 的价值取决于时期 0 结果实际发生的结果。

i　个人投资者，$i = 1, \ldots, I$

W_i　投资者 i 在时期 0 的财富

c_i　投资者 i 在时期 0 的财富

$W_i - c_i$　投资者 i 在时期 0 的投资额

s　第二期（即时期 1）的自然状态，$s = 1, \ldots, S$

π_s　自然状态 s 的发生概率。我们假设所有消费者具有相同的概率信念；这种情形称

为同种期望（homogeneous expectations）。

C_{is}　投资者 i 在第二期的状态 s 下的消费

$\tilde{C}_i$　投资者 i 在第二期的消费（视为随机变量）

注意，我们对消费有两种不同的看法：一是作为每种状态下一系列可能的消费(C_{is})；二是作为随机变量 $\tilde{C}_i$，它以概率 π_s 取值 C_{is}。

$u_i(c_i)+\delta Eu_i(\tilde{C}_i)$　投资者 i 冯.诺依曼-摩根斯坦效用函数。注意，我们假设这个函数在时间上满足加法可分离性（additively separable），而且注意这个函数含有折现因子 δ。

p_a　资产 a 的价格，其中 $a=0,1,...,A$

X_{ia}　投资者 i 对资产 a 的购买量

x_{ia}　投资者 i 对资产 a 的投资额占其投资总额的比重。如果我们用 W_i 表示他对所有资产的投资总额，则 $x_{ia}=p_a X_{ia}/W_i$，因此有 $\sum_{a=0}^{A} x_{ia}=1.$

$(x_{i0},...,x_{iA})$　投资者 i 持有的投资组合。注意，投资组合是用每种资产投资额占投资总额的比重表示的。

V_{as}　资产 a 在第二期的自然状态 s 下的价值

$\tilde{V}_a$　资产 a 在第二期的价值，视为随机变量

R_{as}　资产 a 在自然状态 s 下的（总）收益率。根据定义，$R_{as}=V_{as}/p_a$。

$\tilde{R}_a$　资产 a 的总收益率，视为随机变量。随机变量 $\tilde{R}_a$ 取值 $\tilde{R}_{as}$ 的概率为 π_s。

R_0　无风险资产的总收益率

$\sigma_{ab}=\mathrm{cov}(\tilde{R}_a,\tilde{R}_b)$　资产 a 总收益率和资产 b 总收益率之间的协方差。

20.4 资本资产定价模型

我们大致按照资产市场模型出现的历史先后顺序对这些模型进行分析。因此，我们首先分析这些模型的老前辈——著名的资本资产定价模型（Capital Asset Pricing Model, CAPM）。CAPM 首先界定了一种特别的效用函数，即财富随机分布的效用仅取决于概率分布的前两个矩（moments）——均数和方差。

这个函数只在一定条件下才与期望效用模型一致。例如，当所有的资产是正态分布的，或者当期望效用函数是二次函数(quadratic)时。然而，在很多情形下，均值-方差模型可以大致近似很多类型的效用函数。在使用均值-方差模型时，“风险厌恶”意味着消费者喜欢期望消费的增加，而厌恶消费的方差的增加。

我们首先推导出预算约束。在其他几个模型中，你将看到它们的预算约束和 CAPM 的预算约束类似。为简单起见，我们将用于表示哪个投资者的下标 i 删去。第二期的消费为

$$\tilde{C}=(W-c)\sum_{a=0}^{A}x_a\tilde{R}_a=(W-c)\left[x_0R_0+\sum_{a=1}^{A}x_a\tilde{R}_a\right]$$

因为投资组合的权重之和必定等于 1，所以 $x_0=1-\sum_{a=1}^{A}x_a$，我们可以将上述预算约束写为

$$\tilde{C}=(W-c)\left[R_0+\sum_{a=1}^{A}x_a(\tilde{R}_a-R_0)\right].\tag{20.3}$$

上式方括号内的项为**投资组合的收益率**（portfolio return）。给定关于均值-方差效用函数的假设，无论投资水平如何，对于既定的期望价值，消费者都喜欢方差最小的投资组合。也就是说，投资者愿意购买**均衡-方差有效率的**（mean-variance efficient）投资组合。至于投资者实际选择哪个投资组合，要取决于他的效用函数；但是，无论他的效用函数是什么样的，实际选择的投资组合必须做到：对于给定的期望收益水平使得方差最小。

在继续介绍之前，我们先看看这个最小化问题的一阶条件。我们想在以下约束条件下使得投资组合收益的方差最小：一是投资组合要能达到既定的期望收益 $\overline{R}$；二是满足预算约束 $\sum_{a=0}^{A}x_a=1$。这个最小化问题用式子表示为

$$\begin{aligned}&\min_{x_0,\ldots,x_A}\sum_{a=0}^{A}\sum_{b=0}^{A}x_ax_b\sigma_{ab}\\ &s.t.\ \ \sum_{a=0}^{A}x_a\overline{R}_a=\overline{R}\\ &\qquad\sum_{a=0}^{A}x_a=1.\end{aligned}$$

在这个问题中，我们允许 x_a 可为正也可为负。这比为这消费者可以长期或短期持有任何资产，包括风险资产。

令 λ 为第一个约束条件的拉格朗日乘子，μ 为第二个约束条件的拉格朗日乘子，一阶条件的形式为

$$2\sum_{b=0}^{A}x_b\sigma_{ab}-\lambda\overline{R}_a-\mu=0\qquad \text{其中 } a=0,1,\ldots,A.\tag{20.4}$$

由于目标函数是凸的而且约束是线性的，二阶条件自动得以满足。

使用这些一阶条件可以得到期望收益率模式的表达式。我们的推导方法比较优美但有些拐弯抹角。令 $(x_1^e,\ldots,x_A^e)$ 表示某个投资组合，这个投资组合由均值-方差有效率的所有风险资产组成。假设投资者可以购买到的其中一种风险资产（比如称为资产 e），是持有该有效率组合 (x_a^e) 的某个“共同基金”。那么下列投资组合是均值-方差有效率的：投资 1 元于资

产e但投资 0 元于所有其他资产。这意味着这样的投资组合必定满足（20.4）的条件。

注意，对于这个投资组合来说$x_b = 0, b \neq e$，我们看到第a个一阶条件变为

$$2\sigma_{ae} - \lambda \overline{R}_a - \mu = 0. \qquad (20.5)$$

$a = 0$和$a = e$是两种特殊情形：

$$-\lambda R_0 - \mu = 0$$
$$2\sigma_{ee} - \lambda \overline{R}_e - \mu = 0.$$

当$a = 0$时，$\sigma_{ae} = 0$，这是因为资产 0 是没有风险的资产。当$a = e$，$\sigma_{ee} = \sigma_{ee}$，因为某个变量和其自身的协方差就是这个变量的方差。

从这两个方程中解出λ和μ可得

$$\lambda = \frac{2\sigma_{ee}}{\overline{R}_e - R_0}$$

$$\mu = -\lambda R_0 = \frac{2\sigma_{ee} R_0}{\overline{R}_e - R_0}.$$

将它们的值代入（20.5），整理可得

$$\overline{R}_a = R_0 + \frac{\sigma_{ae}}{\sigma_{ee}}(\overline{R}_e - R_0). \qquad (20.6)$$

这个式子是说，任何资产（a）的期望收益率，等于无风险的利率（R_0）加上“风险溢价”（$\frac{\sigma_{ae}}{\sigma_{ee}}(\overline{R}_e - R_0)$），风险溢价取决于该资产的收益率与风险资产某个有效率投资组合收益率的协方差。

为了给这个式子增添经验性（empirical）内容，我们需要能够识别出某个特定的有效率的投资组合。为了做此事，我们使用图形分析有效率投资组合的结构。在图 20.1 中，我们画出了风险资产某个特定组合产生的期望收益率和标准差（standard deviations）[一]。可以证明，

全部由风险资产组成的均值-标准差有效率投资组合的集合，必定具有图 20.1 所示的双曲线形状。但是需要指出，这种特殊形状并不是下列论断所必需的。

我们想使用风险资产和无风险资产构建一个有效率的投资组合的集合。为了做此事，画出一条经过无风险收益率R_0的直线，并且要使得该直线恰好与双曲线接触，如图 20.1 所示。将这一接触点记为$(\overline{R}_m, \sigma_m)$。这一点是某个投资组合$m$的期望收益率和标准差。我们断言：该条直线上的每个点（期望收益率和标准差的每个组合），都可以通过对下列两种投资组合取凸组合（convex combination）而得到：一是无风险的投资组合；二是投资组合m。

[一] 某随机变量的标准差就是它的方差的平方根。

例如，为了构建期望收益率为$(\overline{R}_m - R_0)/2$标准差为$\sigma_m/2$的一个投资组合，我们只要将我们的一半资金投资于资产组合m、一半资金投资于无风险资产即可。这表明了如何实现$(\overline{R}_m, \sigma_m)$左方的均值-标准差组合。为了产生$(\overline{R}_m, \sigma_m)$右方的组合，我们必须按照利率$R_0$借入资金并将其投资于资产组合$m$。

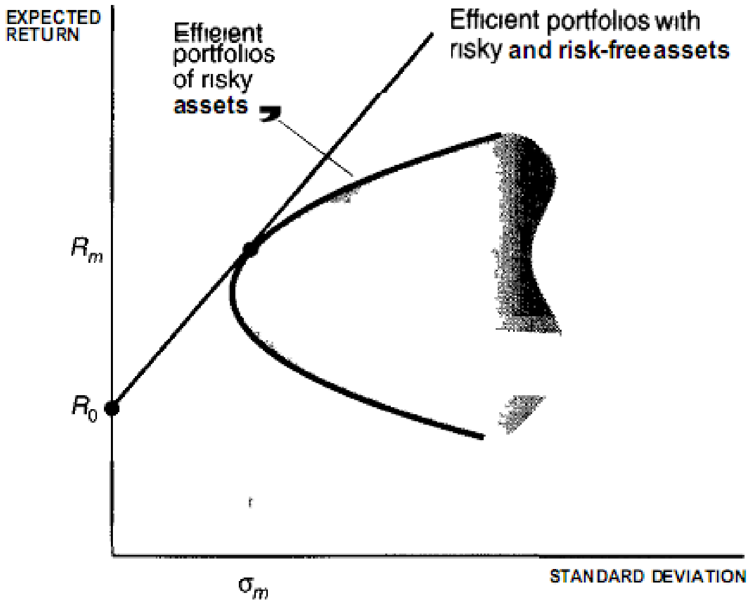


图 20.1：风险收益率和标准差的组合。所有均值-方差有效率的投资组合都可以通过对收益为R_0和R_m的资产（组合）取凸组合而得到。

从上面的论证过程可知，有效率投资组合的集合的结构的确非常简单：该集合仅通过两个资产组合就可完全构造出，其中一个组合是无风险资产构成的组合，另一个是资产组合m。

在资产组合m中令投资于资产a的资金比重为x_a^m，当然我们必有$\sum_{a=1}^{A} x_a^m = 1$。令W_i表示个人i投资于风险组合的资金额，令X_{ia}表示个人i投资于风险资产a的资金比重，令p_a表示资产a的价格。由于每个投资者持有相同的风险资产组合，我们必有

$$x_a^m = \frac{p_a X_{ia}}{W_i} \qquad i = 1, \ldots, I.$$

将上式两侧同乘以W_i，然后在i上遍加（即将I个式子相加）可得

$$x_a^m = \frac{p_a \sum_{i=1}^{I} X_{ia}}{\sum_{i=1}^{I} W_i}.$$

上式的分子是资产a的总市场价值，分母是所有风险资产的总市场价值。因此，x_a^m表示的

是投资于资产 a 的钱数占投资于所有风险资产钱数的比重。这个组合称为**风险资产的市场组合**（market portfolio of risky assets）。这个组合可能观测的到——只要我们能衡量出持有的风险资产的总市值。

由于风险资产的市场组合是一种特别的均值-标准差有效率的投资组合，我们可以将（20.6）式改写为

$$\overline{R}_a = R_0 + \frac{\sigma_{am}}{\sigma_{mm}}(\overline{R}_m - R_0).$$

这是 CAPM 的基本结果。它向风险溢价注入了可验证的内容：风险溢价是资产 a 与市场组合的协方差 σ_{am} 除以市场组合的方差 σ_{mm} 然后再乘以市场组合的超额收益率 $(\overline{R}_m - R_0)$ 。

σ_{am}/σ_{mm} 这一项可以视为 $\tilde{R}_a$ 对 $\tilde{R}_m$ 回归得到的理论回归系数。正是由于这个原因，这一项通常简写为 β_a 。进行替换之后我们就得到了 CAPM 的最终形式：

$$\overline{R}_a = R_0 + \beta_a(\overline{R}_m - R_0). \qquad (20.7)$$

CAPM 是说为了确定任何资产的期望收益率，我们只要知道该资产的“贝塔值” β——它与市场组合的协方差。注意，资产收益率的方差和这个问题无关；要紧的不是该资产的“自身风险”，而是该资产的收益率如何影响投资者的整个投资组合的风险。由于在 CAPM 模型中，每个投资者持有相同的风险资产组合，关键的风险是某种资产如何影响市场组合的风险。

CAPM 的迷人之处在于它涉及的因素似乎可以进行经验验证：风险资产的市场组合的期望收益率；某种特定资产收益率和市场组合收益率的回归系数。然而，你必须记住相关的理论构建是基于所有风险资产构成的投资组合基础之上的；这一点可能不容易观测到。

20.5 套利定价理论

CAPM 从对消费者的偏好界定开始；而套利定价理论（Arbitrage Pricing Theory, APT）则从对资产收益产生过程的说明开始。在这个意义上，CAPM 是个需求方模型，而 APT 是个供给方模型。

人们发现绝大多数资产的价格是一起运动的；也就是说资产价格之间存在较高程度的协方差。自然可以想到，可以将资产的收益率写成某些共同因素和某些与具体资产有关的特定风险的函数。例如，如果只有两个共同因素，我们可以写为

$$\tilde{R}_a = b_{0a} + b_{1a}\tilde{f}_1 + b_{2a}\tilde{f}_2 + \tilde{\varepsilon}_a. \qquad a = 1,\ldots,A.$$

在上式中，我们将 $(\tilde{f}_1, \tilde{f}_2)$ 想象为“宏观经济方面的因素”，它们是影响所有资产收益的宏观经济因素。每种资产 a 对于因素 i 有个特别的“敏感度” b_{ia} 。资产-特定风险（asset-specific risk），根据定义，是独立于宏观经济因素 $\tilde{f}_1$ 和 $\tilde{f}_2$ 的。

由于存在常数项 b_{0a}，可以假设因素 $\tilde{f}_i$（其中 $i=1,2$）和资产-特定风险 $\tilde{\varepsilon}_a$（其中 $a=1,...,A$）的期望值为零。（如果期望值不为零，将其并入 b_{0a} 即可。）我们还假设 $E\tilde{f}_1\tilde{f}_2=0$，也就是说这两个因素是真正互相独立的因素。

我们首先分析不存在资产-特定风险的 APT，这是 APT 的特殊情形。我们从只有一种风险因素的情形开始分析，因此

$$\tilde{R}_a = b_{0a} + b_{1a}\tilde{f}_1 \qquad a=0,...,A.$$

象以前一样，我们想将资产的期望收益以风险溢价进行表达。根据 APT 的构造可知 $\overline{R}_a = b_{0a}$，因此我们只要分析 b_{0a} 的行为即可，其中 $a=1,...,A.$

假设我们构造出某个投资组合，这个投资组合由资产 a 和资产 b 组成，我们持有资产 a 和 b 的比例分别为 x 和 $1-x$。该投资组合的收益率为

$$x\tilde{R}_a + (1-x)\tilde{R}_b = [xb_{0a} + (1-x)b_{0b}] + [xb_{1a} + (1-x)b_{1b}]\tilde{f}_1.$$

我们选择 x^* 使得上式右侧第二个方括号内的项等于零。这意味着

$$x^* = \frac{b_{1b}}{b_{1b} - b_{1a}}. \qquad (20.8)$$

注意此处必须假设 $b_{1b} \neq b_{1a}$，即资产 a 和 b 的敏感系数是不同的。

因为这种方法构造的是无风险组合，所以它的收益率必定等于无风险的收益率，这意味着

$$x^* b_{0a} + (1-x^*)b_{0b} = R_0,$$

或

$$x^*(b_{0a} - b_{0b}) = R_0 - b_{0b}.$$

将上式代入（20.8），整理可得

$$\frac{b_{0b} - R_0}{b_{1b}} = \frac{b_{0b} - b_{0a}}{b_{1b} - b_{1a}} \qquad (20.9)$$

在上式中交换 a 和 b 的角色可得

$$\frac{b_{0a} - R_0}{b_{1a}} = \frac{b_{0a} - b_{0b}}{b_{1a} - b_{1a}} \qquad (20.10)$$

可以看到（20.9）和（20.10）的右侧是相同的。由于这适用于所有资产 a 和 b，可知对于某个常数 λ_1 和所有资产 a，有

$$\frac{b_{0a} - R_0}{b_{1a}} = \lambda_1$$

使用 $\overline{R}_a = b_{0a}$ 这个事实，将上式变形可得到单因素 APT 的最终形式

$$\overline{R}_a = R_0 + b_{1a}\lambda_1. \qquad (20.11)$$

（20.11）是说任何资产 a 的期望收益率等于无风险收益率 R_0 加上风险溢价 $b_{1a}\lambda_1$，其中风险溢价等于资产 a 对共同风险因素的敏感程度 b_{1a} 乘以一个常数 λ_1。常数 λ_1 可以解释为某投资组合的风险溢价，这个投资组合对于因素 1 代表的风险的敏感度为 1。

两个因素

现在我们考虑一个两因素模型：

$$\tilde{R}_a = b_{0a} + b_{1a}\tilde{f}_1 + b_{2a}\tilde{f}_2.$$

现在我们构建一个投资组合 (x_a, x_b, x_c)，它有三种资产：a、b 和 c，它们满足下列方程组：

$$\begin{aligned} x_a b_{1a} + x_b b_{1b} + x_c b_{1c} &= 0 \\ x_a b_{2a} + x_b b_{2b} + x_c b_{2c} &= 0 \\ x_a + x_b + x_c &= 1. \end{aligned}$$

第一个方程是说该投资组合消除了因素 1 导致的风险，第二个方程的意思类似，第三个式子是说资产份额之和等于 1，即我们真正有一个投资组合。

由上面的分析可知，这个投资组合的风险为零。因此，它的收益率等于无风险的收益率，所以 $x_a b_{0a} + x_b b_{0b} + x_c b_{0c} = R_0$。将这些条件写为矩阵形式可得

$$\begin{pmatrix} b_{0a} - R_0 & b_{0b} - R_0 & b_{0c} - R_0 \\ b_{1a} & b_{1b} & b_{1c} \\ b_{2a} & b_{2b} & b_{2c} \end{pmatrix} \begin{pmatrix} x_a \\ x_b \\ x_c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

向量 (x_a, x_b, x_c) 的分量不能全部是 0，这是因为 $x_a + x_b + x_c = 1$。由此可知上式左侧矩阵必定是奇异的。如果最后两行不是共线性的（collinear），事实上我们假设它们不是，那么第一行必定是最后两行的线性组合。也就是说，第一行的每个项是后面两行相应项的线性组合。这意味着

$$\overline{R}_a - R_0 = b_{1a}\lambda_1 + b_{2a}\lambda_2 \qquad a = 1, \ldots, A. \qquad (20.12)$$

这些 λ 的解释方法和前面相同，以 λ_2 为例：它是投资组合的超额收益率，这个投资组合对因素 2 的的风险敏感程度为 1。两因素模型是单因素模型的自然推广：资产 a 的超额收益率取决于它对这两个风险因素的敏感程度。

资产-特定风险

我们已经看到如果风险因素的数量相对资产种类数量较小，并且如果不存在资产-特定风险，那么可以构建出无风险的投资组合。这些无风险的投资组合的收益率必定等于无风险的收益率，这实际上对期望收益率集合构成了某些约束。

但是，使用风险资产构造这些无风险投资组合的前提条件是**所有**风险都是由宏观经济因素引起的。如果除了宏观经济风险之外还存在着资产-特定风险，结果又将如何？

根据定义，资产-特定风险不仅独立于宏观经济风险因素，而且互相之间也是独立的。大数法则意味着持有高度分散投资组合的风险中，资产-特定风险非常小。这个结论是说我们可以忽略资产-特定风险，而且期望收益率和风险因素的线性关系仍然成立，至少近似是线性的。感兴趣的读者如果想了解细节，可参考本章结尾列出的参考文献。

20.6 期望效用

下面我们考虑一个基于跨期期望效用最大化的资产定价模型。考虑下列两期问题：

$$\max_{c,x_1,\ldots,x_A} \quad u(c)+\alpha E\left[u\left(W-c\left(R_0+\sum_{a=1}^{A}x_a(\tilde{R}_a-R_0)\right)\right)\right].$$

注意，在上式中，为简化符号，我们省略掉了投资者 i 的下标。

这个最大化问题要求我们确定第一期的储蓄额 $W-c$ 和投资组合方案 $(x_1,\ldots,x_A)$，以使得折现期望效用最大化。

令 $\tilde{R}=(R_0+\sum_{a=1}^{A}x_a(\tilde{R}_a-R_0))$ 为投资组合的收益率，令 $\tilde{C}=(W-c)\tilde{R}$，我们可以将这个问题的一阶条件写为

$$\begin{aligned} u'(c)&=\alpha Eu'(\tilde{C})\tilde{R} \\ 0&=Eu'(\tilde{C})(\tilde{R}_a-R_0) \qquad a=1,\ldots,A. \end{aligned}$$

第一个条件是说，第一期消费的边际效用应该等于第二期消费的折现期望效用。第二组条件是说，投资组合从安全资产移动至资产 a 的期望边际效用必定等于零，其中 $a=1,\ldots,A.$

我们重点关注第二组条件，看看它们对于资产定价的含义。使用协方差恒等式（详见 26 章），我们可以将这些条件写为

$$Eu'(\tilde{C})(\tilde{R}_a-R_0)=\mathrm{cov}(u'(\tilde{C}),\tilde{R}_a)+Eu'(\tilde{C})(\overline{R}_a-R_0)=0 \qquad a=1,\ldots,A.$$

整理可得

$$\overline{R}_a=R_0-\frac{1}{Eu'(\tilde{C})}\mathrm{cov}(\tilde{R}_a,u'(\tilde{C})). \tag{20.13}$$

这个式子让我们想起 CAPM 中的资产定价形式即（20.7）式，但有一些区别。现在风险溢价取决于某资产收益率$\tilde{R}_a$与边际效用的协方差，而不是取决于它与风险资产市场组合的协方差。

如果某种资产和消费正相关，那么它与消费的边际效用是负相关的，这是由于$u'' < 0$。这意味着它的风险溢价是正的——为了让投资者持有它必须提供较高的期望收益率。而于消费负相关的资产的情形正好相反。

事实上，这个式子只对于单个投资者i成立。然而，在某些条件下，这个条件是可以加总的。例如，假设所有资产是正态分布的。那么消费也将是正态分布的，这样我们可以运用 Rubinstein（1976）发现的定理

$$\mathrm{cov}(u'(\tilde{C}), \tilde{R}_a) = Eu''(\tilde{C})\,\mathrm{cov}(\tilde{C}, \tilde{R}_a).$$

对（20.13）应用该定理，为区分投资者再加入下标，我们有

$$\bar{R}_a = R_0 + \frac{Eu''(\tilde{C}_i)}{Eu'(\tilde{C}_i)}\mathrm{cov}(\tilde{C}_i, \tilde{R}_a). \qquad (20.14)$$

上式右侧中与协方差相乘的项（$Eu''(\tilde{C}_i)/Eu'(\tilde{C}_i)$），有时称为个人$i$的**总体**风险厌恶程度（global risk aversion）。我们将其记为r_i。将（20.14）变形可得

$$\frac{1}{r_i}(\bar{R}_a - R_0) = \mathrm{cov}(\tilde{C}_i, \tilde{R}_a).$$

将上式对投资者$i = 1,\ldots, I$加总，并使用$\tilde{C} = \sum_{i=1}^{I}\tilde{C}_i$表示总消费，可得

$$(\bar{R}_a - R_0)\sum_{i=1}^{I}\frac{1}{r_i} = \mathrm{cov}(\tilde{C}, \tilde{R}_a).$$

上式也可以写为

$$\bar{R}_a = R_0 + \left[\sum_{i=1}^{I}\frac{1}{r_i}\right]^{-1}\mathrm{cov}(\tilde{C}, \tilde{R}_a). \qquad (20.15)$$

现在风险溢价取决于总消费和资产收益率之间的协方差，它与该协方差成比例。这个比例因子$\left[\sum_{i=1}^{I}\frac{1}{r_i}\right]^{-1}$衡量了风险厌恶程度的大小。

我们也可以使用某特定资产的超额收益率推导出这个比例因子。假设某种资产c和总消费完美相关（这种资产本身可以为其他资产的组合。）那么该资产的收益率R_c必定满足（20.15）式：

$$\bar{R}_c = R_0 + \left[\sum_{i=1}^{I}\frac{1}{r_i}\right]^{-1}\mathrm{cov}(\tilde{C}, \tilde{R}_a) = R_0 + \left[\sum_{i=1}^{I}\frac{1}{r_i}\right]^{-1}\mathrm{var}(\tilde{C})$$

从这个式子可以解除平均风险厌恶程度

$$\left[\sum_{i=1}^{I}\frac{1}{r_i}\right]^{-1}=\frac{\overline{R}_c-R_0}{\sigma_{cc}}.$$

这样我们可以把资产定价公式（20.15）改写为

$$\overline{R}_a=R_0+\frac{\sigma_{ca}}{\sigma_{cc}}\mathrm{cov}(\tilde{C},\tilde{R}_a). \tag{20.16}$$

上式右侧中的协方差比值有时称为某资产（这里为资产a）的**消费贝塔值**（consumption beta）。它是资产a的收益率与资产c的收益率的理论回归系数，其中资产c与总消费完美相关。在 CAPM 模型中，我们对于“市场贝塔值”的解释与这里的消费贝塔值的解释是相同的。事实上，（20.16）和（20.17）是如此相似，以至于读者可能怀疑这二者之间是否真正存在着区别。

在两期模型中，它们不存在着区别。如果只有两期，那么第二期的总财富（也就是市场组合）等于总消费。然而，在多期模型中，财富可能不等于消费。

虽然我们使用两期情形推导出了（20.16）这个公式，实际上它适用于多期情形。为了看清这一点，考虑下面的实验。在时期t，如果从无风险资产那儿移出一元到风险资产a，那么在时期$t+1$，你的消费的变化量等于$x_a(\tilde{R}_a-R_0)$。如果你有最优消费方案，这种移动的期望效用必定为零。但是这个条件和收益率分布的正态性是我们推导出（20.16）的全部前提条件！

例子：期望效用和 APT

由于 APT 对收益率的特征施加了限制，而期望效用模型只对偏好施加了限制，我们可以结合这两个模型的结果，对因素-特定风险（factor-specific risks）进行解释。

为方便参考，我们将（20.13）式列在此处作为（20.17）式

$$\overline{R}_a=R_0-\frac{1}{Eu'(\tilde{C})}\mathrm{cov}(\tilde{R}_a,u'(\tilde{C})). \tag{20.17}$$

两因素 APT 模型规定

$$\tilde{R}_a=b_{0a}+b_{1a}\tilde{f}_1+b_{2a}\tilde{f}_2.$$

将上式代入（20.17）可得

$$\overline{R}_a=R_0-\frac{1}{Eu'(\tilde{C})}\left[b_{1a}\mathrm{cov}(u'(\tilde{C}),\tilde{f}_1)+b_{2a}\mathrm{cov}(u'(\tilde{C}),\tilde{f}_2)\right]$$

将这个式子与（20.12）比较，我们看到λ_1和λ_2同下列因素成比例：消费的边际效用与相应

风险因素的协方差。

20.7 完全市场（complete market）

我们现在考虑资产定价的另外一个模型。假设有 S 种不同的自然状态，而且对于每种状态 s 都相应存在着一种资产，该资产当状态 s 发生时支付 1 元，如果 s 不发生则支付 0 元。这样的资产称为阿罗-德布鲁证券。令 p_s 表示阿罗-德布鲁证券的均衡价格。

现在考虑状态 s 下的价值为 V_{as} 的任意一种资产。这样的资产在时期 0 的价值是多大？考虑下面的推理：构造一个持有 V_{as} 单位的阿罗-德布鲁证券 s 的投资组合。由于阿罗-德布鲁证券 s 在状态 s 下的价值为 1 元，这个投资组合的价值在状态 s 下的价值将为 V_{as}。因此，该投资组合的收益模式和资产 a 的收益模式是相同的。从套利方面考虑，资产 a 的价值必定和这个投资组合的价值是相同的。因此，

$$p_a = \sum_{s=1}^{S} p_s V_{as}.$$

这个论证表明任何资产的价值都可以通过阿罗-德布鲁证券的价值确定。

令 π_s 表示状态 s 的发生概率，我们可以将上式写为

$$p_a = \sum_{s=1}^{S} \frac{p_s}{\pi_s} V_{as} \pi_s = E \frac{\tilde{p}}{\tilde{\pi}} \tilde{V}_a,$$

其中 E 为期望算子。上式是说资产 a 的价值是资产 a 的价值与随机变量（ $\tilde{p}/\tilde{\pi}$ ）乘积的期望值。使用协方差恒等式（详见第 26 章），我们可以将上式写为

$$p_a = \mathrm{cov}(\frac{\tilde{p}}{\tilde{\pi}}, \tilde{V}_a) + E \frac{\tilde{p}}{\tilde{\pi}} \tilde{V}_a. \qquad (20.18)$$

根据定义

$$E \frac{\tilde{p}}{\tilde{\pi}} = \sum_{s=1}^{S} \frac{p_s}{\pi_s} \pi_s = \sum_{s=1}^{S} p_s.$$

因此， $E(\tilde{p}/\tilde{\pi})$ 是下一期支付 1 元的那种投资组合的价值。令 R_0 表示这种投资组合的无风险收益率，我们有

$$E \frac{\tilde{p}}{\tilde{\pi}} = \frac{1}{R_0}.$$

将上式代入（20.18）并稍加整理可得

$$p_a = \frac{\bar{V}_a}{R_0} + \mathrm{cov}(\frac{\tilde{p}}{\tilde{\pi}}, \tilde{V}_a). \qquad (20.19)$$

因此资产 a 的价值必定等于它的折现期望价值加上风险溢价。

所有这些只是操作上的定义；现在我们增加一个行为假设。如果个人 i 购买 c_{is} 单位阿罗-德布鲁证券 s，那么他必定满足一阶条件

$$\pi_s u_i'(c_{is}) = \lambda p_s$$
$$\frac{u_i'(c_{is})}{\lambda_i} = \frac{p_s}{\pi_s}.$$

由此可知，p_s / π_s 必定和投资者 i 的消费边际效用成比例。这个式子的左侧是消费的严格递减函数，这是因为风险厌恶。令 f_i 表示 u_i' / λ_i 的反函数；这也是一个递减函数。于是我们可以写为

$$c_{is} = f_i(p_s / \pi_s).$$

对 i 加总，并使用 C_s 表示状态 s 下的总消费，可得

$$C_s = \sum_{i=1}^{I} f_i(p_s / \pi_s).$$

由于每个 f_i 都是递减函数，上式右侧也是递减函数。因此，它有反函数 F，所以我们可以写为

$$\frac{p_s}{\pi_s} = F(C_s),$$

其中 $F(C_s)$ 是总消费的递减函数。

将上式代入（20.19）可得

$$p_a = \frac{\overline{V}_a}{R_0} + \operatorname{cov}(F(\tilde{C}), \tilde{V}_a). \qquad (20.20)$$

因此，资产 a 的价值等于期望价值的折现加上风险溢价，风险溢价取决于资产价值和总消费的递减函数的协方差。如果资产和总消费正相关，则需要对它的折现期望价值进行负的调整，反之则反是，这一点和其他模型是相同的。

20.8 纯套利

我们最后考虑的资产定价模型，与上述几种模型相比，假设条件最少：只要求不存在纯套利（pure arbitrage）机会即可。

将一组资产排列成 $A \times S$ 矩阵，矩阵元素 V_{sa} 衡量资产 a 在状态 s 下的价值。将这个矩阵记为 V。令 $X = (X_1, ..., X_A)$ 表示持有资产 A 的模式。于是这个投资模式在第二期的价值将是一个 S-向量，它由矩阵乘积 VX 得到。

假设 X 在每种自然状态下的收益都是非负的：$VX \geq 0$。那么假设这个投资模式的价值为非负似乎是合理的：$pX \geq 0$。否则就存在着明显的套利机会。我们将这个条件表达为如

下原理：

无套利原理。若$VX \geq 0$，则 $pX \geq 0$。

这再本质上是要求“不存在免费的午餐。”可以证明，无套利原理意味着存在着一组“状态价格” $\rho_s \geq 0$（其中$s = 1, \ldots, S$）使得任何资产a的价值

$$p_a = \sum_{s=1}^{S} \rho_s V_{as}. \tag{20.21}$$

由于这个结论的证明不是我们的兴趣所在，所以证明过程在附录部分给出。在这里我们将分析这个条件对于资产定价的意义。

我们知道π_s衡量的是自然状态s的发生概率，我们可以将（20.21）写为

$$p_a = \sum_{s=1}^{S} \frac{\rho_s}{\pi_s} V_{as} \pi_s.$$

上式右侧是两个随机变量乘积的期望。令$\tilde{Z}$表示取值为ρ_s / π_s的随机变量，令$\tilde{V}_a$表示取值为V_{as}的随机变量。

于是，应用协方差恒等式可得

$$p_a = E\tilde{Z}\tilde{V}_a = \mathrm{cov}(\tilde{Z}, \tilde{V}_a) + \overline{Z}\overline{V}_a.$$

根据定义

$$\overline{Z} = \sum_{s=1}^{S} \frac{\rho_s}{\pi_s} \pi_s = \sum_{s=1}^{S} \rho_s.$$

上式右侧是每种状态都支付 1 元的这种证券的价值，也就是说，它是个无风险的债券。根据定义可知，这个价值为$1/R_0$。

进行替换并整理后，我们有

$$\overline{V}_a = p_a R_0 - R_0 \mathrm{cov}(\tilde{Z}, \tilde{V}_a).$$

将上式两侧同除以 p_a 可将其转化为资产收益率的表达式

$$\overline{R}_a = R_0 - R_0 \mathrm{cov}(\tilde{Z}, \tilde{R}_a).$$

从这个式子，我们看到在非常一般的条件下，每种资产的风险溢价取决于资产收益和单个随机变量的协方差——这个随机变量对于所有资产都是相同的。

在我们已经考察的不同模型中，$\tilde{Z}$ 的表达式都是不同的。在资本资产定价模型（CAPM）中，$\tilde{Z}$ 是风险资产投资组合的收益。在消费-贝塔模型中，$\tilde{Z}$ 是消费的边际效用。在阿罗-德布鲁模型中，$\tilde{Z}$ 是特殊形式的总消费函数。

附录

我们想证明：教材中的无套利原理意味着存在着非负状态的价格$(\rho_1,...,\rho_S)$。为了解决这个问题，我们考虑下列线性规划问题：

$$\begin{aligned}&\min\ \ \mathbf{px}\\&\text{s.t.}\ \ \mathbf{vx}\geq 0\end{aligned}$$

这个线性规划问题要求我们找到满足非负报酬条件的最便宜的投资组合。当然$\mathbf{x}=\mathbf{0}$是这个问题的可行选择，而且无套利原理意味着它的确能使得目标函数最小化。因此，线性规划问题有确切解。

这个线性规划的对偶问题为

$$\begin{aligned}&\min\ \ \mathbf{op}\\&\text{s.t.}\ \ \rho\mathbf{v}=\mathbf{p}\end{aligned}$$

其中ρ是 S 维的对偶变量非负向量。由于原问题有确切可行解，因此对偶问题也有确切解。因此，无套利条件意味着必定存在能满足$\mathbf{p}=\rho\mathbf{v}$的 S 维向量。

注释

我们对资本资产定价模型（CAPM）的处理参照了 Ross（1977）。套利定价模型（APT）归功于 Ross（1976）；我们对 APT 的处理方法参照了 Ingersoll（1987）。阿罗-德布鲁模型的资产定价公式参照了 Rubinstein（1976）。纯套利的分析参照了 Ross（1978），但我们使用了 Ingersoll（1987）的证明。

习题

20.1 在期望效用最优投资组合的一阶条件为：对于所有资产a，都有$Eu'(\tilde{C})(\tilde{R}_a-R_0)=0$。证明，这个条件也可以写为：对于任何资产$a$和$b$，$Eu'(\tilde{C})(\tilde{R}_a-\tilde{R}_b)=0$。

20.2 将（20.2）式表达为资产a报酬率的函数。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 21 章：均衡分析

曹乾　译

(东南大学　caoqianseu@163.com)

21 均衡分析

在这一章我们讨论一些一般均衡议题，这些议题放在其他章节都不太合适。我们第一个主题是关于核（核是帕累托集的推广）以及核与瓦尔拉斯均衡的关系。接下来我们简要讨论凸性和经济体大小之间的关系。然后，我们分析在什么样的条件下，瓦尔拉斯均衡是唯一的。最后讨论一般均衡的稳定性。

21.1 交换经济的核

我们已经看到瓦尔拉斯均衡通常存在而且一般是帕累托有效率的。但是竞争市场体系只是配置资源的一种方法。如果我们使用其它社会制度帮助交易达成，情形又会如何？结果会不会“非常接近”瓦尔拉斯均衡？

为了分析这个问题，我们考虑一个“市场博弈”，其中每个消费者 i 在交易前的初始禀赋为 ω_i。假设消费者不使用价格机制，而是在市场里东走西逛、进行试探性的交易。当所有消费者找到最好的约定时，开始交易。

介绍到现在我们还基本没涉及到这个博弈的结构。我们不打算给出计算均衡的必要细节，而是问一个一般性的问题。这个博弈的“合理”结果是什么样的？下列一组定义对于回答这个问题有帮助。

某个配置的改进。**一组人 S 在某个给定的配置 x 上存在着改进余地，如果存在另外一个配置 x' 使得**

$$\sum_{i\in S} x_i' = \sum_{i\in S} \omega_i \text{，并且}$$

$$x_i' \succ_i x_i \text{ 对于所有 } i \in S \text{ 成立。}$$

如果能改进某个配置 x，那么就存在着一小伙人，他们不需要进行市场交易就能使得他们的状况变得更好；他们只需要在他们彼此之间进行交易就可以了。比如，一伙消费者建立合作性的商店以和高价格的杂货店对抗。任何配置如果存在着改进的余地，那么这个配置似乎不应该是均衡——某些小团伙总是有动机从经济体分离出来。

经济的核（core of an economy）。**某个可行配置 x 在经济的核之中，如果不能通过任何联盟（coalition）对它进行改进。**

注意，如果配置 x 在核中，那么 x 必定是帕累托有效率的。因为如果 x 不是帕累托有效率的，那么由全体当事人组成的最大联盟的状况在 x 上存在着改进的余地。在这种意义上，核是帕累托集这个概念的推广。如果某个配置在核中，每个小团伙都从交易中得到好处——没有哪个小团伙有动机偏离这个配置。

核这种思想的缺陷在于它对当事人的信息要求很高——分散于各个联盟的人为了形成新的联盟，必须能找到志同道合的人。而且，它还假设形成联盟是没有成本的，即使组成联盟只能实现非常小的利益，他们也会组建联盟。

借助埃奇沃思盒，我们可以对核进行图形分析。我们以两人两种商品的情形为例。见图 21.1。在这种情形下，核是帕累托集的子集，在核中每个当事人的状况比进行交易要好。

经济的核一般是非空的吗？如果我们进一步作出假设保证市场均衡的存在，那么它一般是非空的，这是因为市场均衡总是包含于核之中。

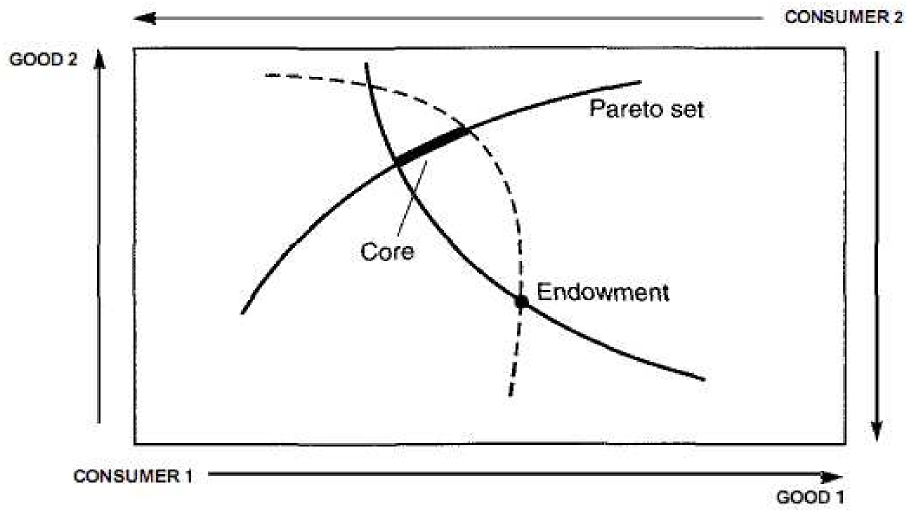


图 21.1：埃奇沃思盒中的核。在埃奇沃思盒中，核就是通过初始禀赋的两条无差异曲线之间的帕累托集的子集。

核中的瓦尔拉斯均衡。**如果 (x^*, p) 是由初始禀赋 ω_i 出发达到的一个瓦尔拉斯均衡，那么 x^* 在核之中**。

证明。假设 x^* 不在核之中；那么将存在某个联盟 S 以及某个可行集 x'，使得 S 中的每个人 i 都严格偏好 x_i' 胜于 x_i^*，而且

$$\sum_{i \in S} x_i' = \sum_{i \in S} \omega_i .$$

但是，瓦尔拉斯均衡的定义意味着

$$p x_i' > p \omega_i \qquad \text{对于 } S \text{ 中的所有 } i \text{ 成立}$$

$$p \sum_{i \in S} x_i' > p \sum_{i \in S} \omega_i$$

这和第一个式子矛盾。 ■

从埃奇沃思盒可以看出，核之中除了市场均衡点之外还存在着其他点。然而，如果我们允许我们的两人经济壮大，即变成多人经济，那么可能的联盟会更多，因此任何既定配置的

改进机会更多。所以，你可能怀疑，随着经济增长，核可能会收缩。表达这种思想遇到的困难是，核是配置空间的子集，因此核会随着经济增长而改变维度。所以我们只分析特别简单的增长类型。

当两个人的偏好和禀赋都相同时，我们说两个人是相同**类型**的。我们说一个经济体是另一个经济体的**复制品**（replica），如果前者中每类人的数量是后者的 r 倍。这意味着如果一个大经济体从小经济体复制，那么它是小经济的“放大”版本。为简单起见，我们仅限于分析两类人，A 类和 B 类。考虑一个固定两人的经济体，该经济体的 r **核**（**r-core**）是指原来经济第 r 次复制后得到经济体的核。

可以证明在任何核配置中，所有相同类型的人得到的商品束必定是相同的。这个结果有助于进一步简化分析。

核中的公平待遇（Equal treatment in the core）。**假设人们的偏好是严格凸、强单调且连续的。如果 x 是某给定经济体的 r 核，那么相同类型的任何两个人必定得到相同的商品束。**

证明。令 x 为核中的配置，令 $2r$ 个人分别用 $A_1,...,A_r$ 和 $B_1,...,B_r$ 表示。如果相同类型的所有人得不到相同的商品束，那么每种类型中都有一人的待遇最悲惨。我们将这两个人称为“A 类弱者（underdog）”和“B 类弱者”。如果 A 类弱者或 B 类弱者不止一人，则任选一人即可。

令 $\bar{x}_A = \frac{1}{r}\sum_{j=1}^{1} x_{A_j}$ 和 $\bar{x}_B = \frac{1}{r}\sum_{j=1}^{1} x_{B_j}$ 分别为 A 类人和 B 类人的平均商品束。由于配置 x 是可行的，

$$
\begin{aligned}
\frac{1}{r}\sum_{j=1}^{1} x_{A_j} + \frac{1}{r}\sum_{j=1}^{1} x_{B_j} &= \frac{1}{r}\sum_{j=1}^{1} \omega_{A_j} + \frac{1}{r}\sum_{j=1}^{1} \omega_{B_j} \\
&= \frac{1}{r} r\omega_A + \frac{1}{r} r\omega_B = \omega_A + \omega_B .
\end{aligned}
$$

由此可知，

$$\bar{x}_A + \bar{x}_B = \omega_A + \omega_B$$

因此 $(\bar{x}_A, \bar{x}_B)$ 对于 A 类弱者和 B 类弱者这两个人组成的联合是可行的，也就是说经济体完全可以做到让这两个人得到他所属类型的平均数。我们假设**至少**有一种类型的两个人，比如 A 类型中的两个人得到的商品束是不同的。因此，A 类弱者必定严格偏好 $\bar{x}_A$ 胜于他实际得到的商品束，理由在于偏好是严格凸的（因为 $\bar{x}_A$ 是至少和 x_A 一样好的商品束的加权平局商品束）。B 类弱者认为 $\bar{x}_B$ 至少和他实际得到的商品束一样好。强单调性和连续性允许 A 稍微从 $\bar{x}_A$ 拿出一些贿赂 B 类弱者，这样就形成了可以改进配置的一个联盟。 ■

由于核中的任何配置必定给相同类型的每个人分配相同的商品束，我们可以使用埃奇沃思盒分析两人经济，因为大型经济都是从两人经济复制出的。我们不再将核中的点 x 看成 A 得到多少和 B 得到多少，而是看成 A 类人中的**每个人**得到多少和 B 类人中的每个人得到

多少。上面的引理告诉我们r核中的所有点都可以用这种方式表示。

下面的命题表明任何配置若不是市场均衡配置，则必然最终不会存在于经济体的r核中。这意味着大经济体的核配置类似瓦尔拉斯均衡。

缩小的核（Shrinking core）。**假设偏好是严格凸、强单调而且从初始禀赋ω开始只能得到一个唯一的市场均衡x^*。于是如果y不是市场均衡，那么存在某个复制经济体r使得y不在r核之中。**

证明。参考图 21.2。我们想证明象y这样的点最终都可以得到改进。由于y不是瓦尔拉斯均衡，连接y点和ω点的线段必然与至少一个人经过点y的无差异曲线相交。因此，可以选择另外一个点比如g点使得其中一人比如 A 的效用比y点更高。具体情形有好几种，这取决于g点的位置；然而，无论哪种情形，论证过程在本质上是相同的，因此我们只分析我们画出的这种情形。

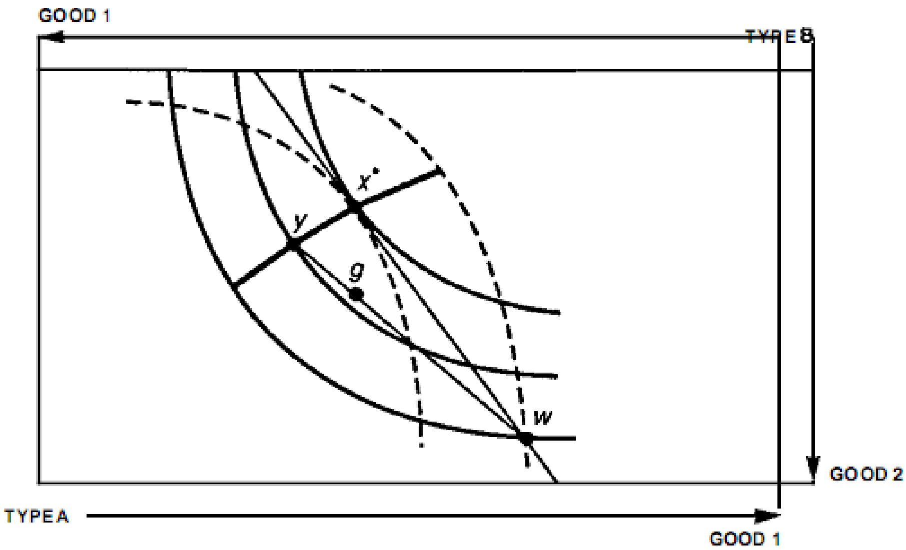


图 22.2：缩小的核。随着经济不停复制扩大，象y这样的点最终不可能位于核之中。

由于g在连接y点和ω点的线段上，我们可以将其表示为

$$g=\theta\omega_A+(1-\theta)y_A \qquad \text{其中}\ \theta>0$$

由于偏好是连续的，我们也可以假设$\theta=T/V$，其中T和V都是正整数。因此

$$g=\frac{T}{V}\omega_A+(1-\frac{T}{V})y_A.$$

假设经济体已经复制了V次。然后，我们构造一个联盟，这个联盟由V个 A 类型的消费者和$(V-T)$个 B 类型的消费者组成。考虑配置z，这个配置分配给 A 类消费者的商品束为g_A、分配给 B 类消费者的为y_B。与y相比，联盟中的所有消费者更偏好z（为了得到严格偏好，我们可以从 A 类消费者中拿走一部分商品束给与 B 类消费者）。我们要证明这个配置是可行的。从下列计算可看清这一点：

$$
\begin{aligned}
Vg_A+(V-T)y_B &= V\left[\frac{T}{V}\omega_A+(1-\frac{T}{V})y_A\right]+(V-T)y_B \\
&= T\omega_A+(V-T)y_A+(V-T)y_B \\
&= T\omega_A+(V-T)[y_A+y_B] \\
&= T\omega_A+(V-T)[\omega_A+\omega_B] \\
&= T\omega_A+V\omega_A-T\omega_A+(V-T)\omega_B \\
&= V\omega_A+(V-T)\omega_B.
\end{aligned}
$$

这正好是我们构造的联盟的禀赋，因为这个联盟由V个 A 类人和$(V-T)$个 B 类人组成。因此，这个联盟可以改进y，命题得证。 ■

这个命题中的很多限制性假设是可以放松的。特别地，我们可以容易地去掉强单调性和唯一市场均衡的假设。凸性似乎对这个命题至关重要，但是和存在性定理一样，这个解释对大的经济体不是必需的。当然，我们也可以允许当事人的类型超过两类。

在研究瓦尔拉斯均衡时，我们已发现市场机制可导致良好定义的均衡。在帕累托有效率配置中，我们发现几乎所有帕累托有效率配置可以通过对禀赋的合理再分配以及价格机制实现，在一般纯交换经济的研究中，价格以第三种角色出现：大型经济体的核中的配置是市场均衡配置。缩小的核这个定理表明瓦尔拉斯均衡是稳健的：即使非常弱的均衡概念，例如核，对于大型经济体来说，也接近于瓦尔拉斯均衡。

21.2 凸性和经济体的大小

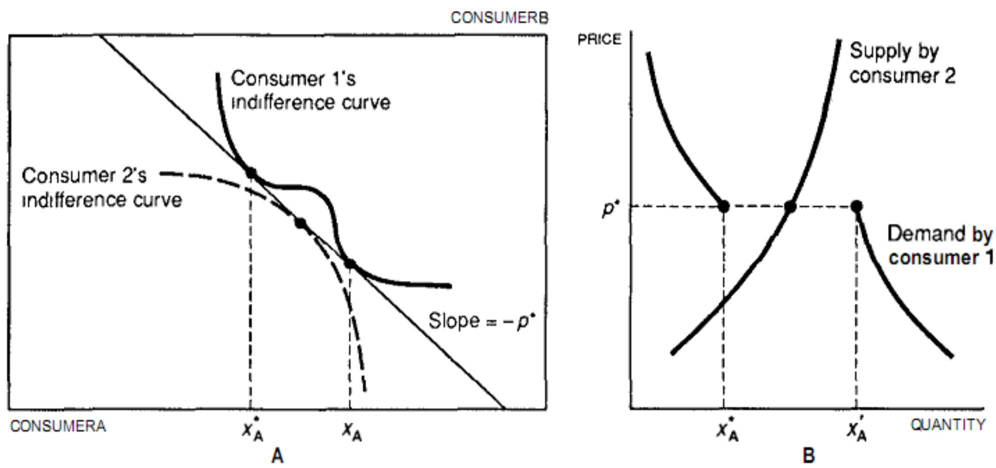


图 21.3：非凸偏好情形下不存在均衡。在 A 图的埃奇沃思盒中，消费者的偏好是非凸的。B 图画出的是相应的总需求曲线，它是不连续的。

偏好的凸性假设在几个均衡模型中都出现过。通常，假设偏好为凸的目的是保证需求函数是良好定义的——即在每个价格水平下只有唯一一个需求束，而且保证需求函数是连续的

——价格的微小变化只会导致需求的微小变化。凸性假设似乎是均衡配置存在的必要条件，因为我们容易构造出实例说明非凸会导致需求不连续，因此导致均衡价格不存在。

例如考虑图 21.3 中的埃奇沃思盒。此处消费者 A 的偏好是非凸的而 B 的偏好是凸的。在价格为 p^* 时，有两个点使得效用最大；但是在这两个点上，供给都不等于需求。

然而，也许均衡并不象本例那样难以达到。下面我们考虑一个具体的例子。假设商品的供给位于价格为 p^* 时的两个需求量的中点，如图 21.3B 所示。现在考虑如果经济体复制一次情形将会怎样。在这种情形下，A 类和 B 类消费者各有两个人。于是在价格为 p^* 时，一个 A 类消费者的需求为 x_A^*，另外一个 A 类消费者的需求为 x_A'。在这种情形下，总需求的确等于总供给。这个复制扩大后的经济体存在着瓦尔拉斯均衡。

不难看出无论供给曲线的位置在什么地方，我们都可以进行类似地构造：如果供给位于 x_A^* 和 x_A' 之间距离的 2/3 处，我们可以将原来的经济体复制三次，以此类推。只要将经济体复制足够的次数，我们就能使纵需求任意接近总供给。

这个论证表明，在大型经济体中，非凸性的规模相对于市场规模比较小，因此通常存在着某个价格向量使得需求接近于供给。对于大型经济体来说，小的非凸性不会造成严重的问题。

这个结论和我们讨论竞争企业行为时的复制观点密切相关。考虑具有固定成本和 U 形平均成本函数的经典企业模型。单个企业的供给函数一般是不连续的，但如果市场规模足够大，这些不连续性是微不足道的。

21.3 均衡的唯一性

在一般均衡的存在性那一节，我们已经知道，在适当条件下，存在使得所有市场都同时出清的价格向量，即存在 p^* 使得 $z(p^*) \leq 0$。我们本节关注的是唯一性问题：什么样的条件下使得所有市场出清的价格向量是唯一的？

我们对免费物品不怎么感兴趣，因此我们通过合意性（desirability）假设将这类物品排除在外：假设当某物品的相对价格为零时，该商品的超额需求是严格正的。这个假设的经济学意义是，当某物品的价格接近于零时，每个人的需求量都很大，这一点似乎非常合理。在这种情形下，结果非常明显：在任何均衡价格向量中每种商品的价格必须为严格正。

和以前一样，我们假设 z 是连续的，但是现在我们需要进一步加强这个假设——假设 z 是连续可微的。理由比较明显。如果无差异曲线存在着拐折（kinks），那么存在一系列价格能使市场均衡。在这种情形下，均衡不仅不是唯一的，甚至不是**局部**唯一的。

在这些假设下，我们遇到一个纯粹数学上的问题：给定从价格单纯形到 R^k 的一个平滑映射 z，何时存在一个唯一的点，它的映射值为零？不要奢望它在一般情形下都成立，因为我们可以构造一个简单的反例，即使在两维情形下它也不成立。因此，我们希望能够找到保证唯一性的对超额需求函数的需求条件。然后我们分析这些限制是强限制还是弱限制，它们的经济意义是什么，等等。

此处我们考虑 z 的两个限制条件，这两个条件分别都能保证前面提到的唯一性。第一种

情形是总替代，我们对该情形感兴趣的原因在于，它有明显的经济意义而且唯一性的证明简单直接。第二种情形是指数分析，我们对它感兴趣是因为它具有一般性。实际上，所有其他的唯一性都是这种情形的特例。不幸的是，这种情形的证明需要使用比较高深的微分拓扑学中的定理。

总替代

大致来说，两种商品是总替代的（gross substitutes），如果一种商品价格上升导致另外一种商品需求增加。在初级经济学课程中，这样的情形通常定义为替代。在高级经济学课程中，有必要区分净替代和总替代。净替代是指当一种商品价格上升时，另外一种商品的希克斯需求增加；而总替代则是将净替代定义中的“希克斯需求”替换为“马歇尔需求”。

总替代。**两种商品 i 和 j 在价格向量为 p 时是总替代的，若** $\dfrac{\partial z_j(p)}{\partial p_i} > 0$**，其中** $i \neq j$。

这个定义是说，如果一种商品 i 的价格上升导致另外一种商品 j 的超额需求增加，则这两种商品是总替代的。如果所有商品都是总替代的，z 的雅可比矩阵 $Dz(p)$ 的非主对角线元素都为正。

总替代性蕴均衡的唯一性。**如果在所有价格水平下所有商品都是总替代的，那么如果 p^* 是一个均衡价格向量，那么它是唯一的均衡价格向量**。

证明。假设 p' 是某个其它均衡价格向量。由于 $p^* >> 0$，我们可以定义 $m = \max p_i' / p_i^* \neq 0$。根据齐次性和 p^* 是一个均衡的事实，我们知道 $z(p^*) = z(mp^*) = 0$。根据 m 的定义，可知对于某个价格 p_k，我们有 $mp_k^* = p_k'$。我们把除了 p_k 之外的每个价格 mp_i^* 依次降低为 p_i'。由于除了商品 k 之外，其它商品的价格在从 mp^* 向 p' 的移动中都降低了，商品 k 的需求必然下降。因此 $z_k(p') < 0$，这意味着 p' 不可能是一个均衡。 ■

指数分析

考虑一个只有两种商品的经济体。选择商品 2 作为计价物，将商品 1 的超额需求作为它自身价格的函数，画出这一需求曲线。瓦尔拉斯法则意味着，当商品 1 的超额需求为零时，我们就得到了一个均衡。合意性假设意味着当商品 1 的相对价格比较大时，商品 1 的超额需求是负的；而当商品 1 的相对价格比较小时，商品 1 的超额需求是正的。

参照图 21.4，在该图中我们画出了几种可能的情形。注意，在图（1）中均衡通常是分离的；图（2）和（3）中，均衡不是分离的；另外，均衡也不是“稳定的”，因为较小的扰动就能让其偏离均衡；图（4）通常有奇数个均衡；在图（5）中如果超额需求曲线在所有均衡点都是向下倾斜的，那么只可能存在唯一一个均衡，如果只有唯一一个均衡，超额需求曲线在该均衡点必然向下倾斜。

在上述一维情形下，注意：如果在所有均衡点$dz(p)/dp<0$，那么只可能存在唯一的一个均衡。指数分析（index analysis）是将这个结果推广到k维的一种方法，该方法给出了均衡唯一性的一个简单的充分必要条件。

给定一个均衡p^*，按照下列方法定义p^*的**指数**（index）：写出超额供给函数的雅可比矩阵的负矩阵$-Dz(p^*)$，舍弃最后一行和最后一列，取剩余矩阵的行列式。若该行列式为正，则对p^*点赋予指数$+1$；若该行列式为负，则赋予p^*点指数-1。（舍弃最后一行和最后一列，等价于选择最后一种商品作为计价物，这和我们一维情形是一样的。）

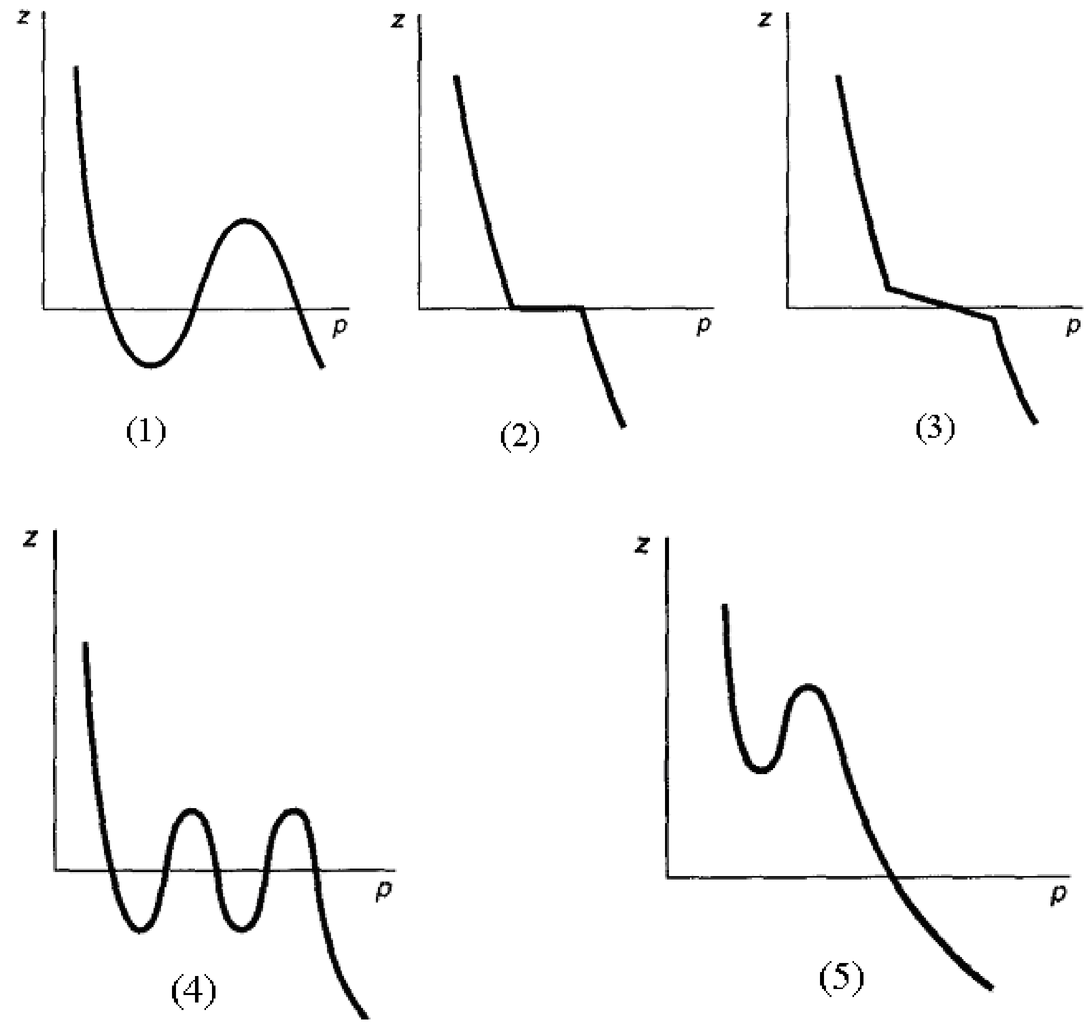


图 21.4：均衡的唯一性和均衡的局部唯一性。这些图画出了我们在讨论均衡唯一性时提及的几种情形。

我们也需要一个边界条件；边界条件存在着几种一般的可能性，但是最简单的情形是假设当$p_i=0$时$z_i(p)>0$。在这种情形下，微分几何中的一个基本定理表明，如果所有均衡都有正的指数，那么均衡将是唯一的。这样，我们立即得到了唯一性定理。

均衡的唯一性。假设z是定义在价格单纯形上的连续可微的总超额需求函数，且当

当 $p_i = 0$ 时 $z_i(p) > 0$。如果 $(k-1)\times(k-1)$ 矩阵 $(-Dz(p^*))$ 的行列式在所有均衡点上的值为正时，则只存在唯一的一个均衡。

这个唯一性定理是一个纯粹数学上的结论。它的优势在于该定理可以应用于很多不同的均衡问题。如果均衡存在性定理可表述为不动点问题，那么我们通常可以使用指数定理来找到在什么样的条件下均衡是唯一的。然而，这个定理的缺陷在于，我们难以使用经济学术语解释它的含义。

在我们分析的这个情形中，我们感兴趣的是总超额供给函数的行列式。我们可以使用斯勒茨基方程将总超额供给函数的导数写为

$$-Dz(p) = -\sum_{i=1}^{n} D_p h_i(p, u_i) - \sum_{i=1}^{n} D_m x_i(p, p\omega_i)[\omega_i - x_i].$$

上式左侧矩阵的行列式何时为正？我们来看看上式右侧的表达式。右侧第一项的符号容易确定；替代矩阵是负半定的矩阵，因此该矩阵的 $(k-1)\times(k-1)$ 主子阵的负矩阵通常是正定矩阵。正定矩阵的和仍是正定的，因此行列式为正。

第二项的符号是不确定的。这一项在本质上是商品超额供给与这些商品边际消费倾向的协方差。没有理由认为这一项具有一般性的特定结构。我们只能说如果收入效应相对于替代效应较小，那么上式右侧第一项的符号起主导作用，这样认为均衡是唯一的就比较合理。

21.4 一般均衡动态

我们已经证明，在对经济人的行为适当的假设条件下，总存在使得需求等于供给的价格向量。但是我们从未保证现实经济会在这个“均衡”点上运行。什么力量可以使得价格向市场出清的价格向量移动？在试图模拟竞争经济的价格调整时我们会遇到一些问题，本节的任务就是分析这些问题。

最大的问题也是最根本的问题是，竞争和价格调整之间的悖论关系：如果所有人将市场价格作为给定的、不受他们个人控制的，那么价格如何移动呢？谁来调整价格？

为了解决这个难题，经济学家杜撰了一个神话人物——“瓦尔拉斯拍卖者”，他的唯一职能是寻找市场出清价格。根据这种解释，竞争市场运行如下：

在时刻零，瓦尔拉斯拍卖者喊出某个价格向量。所有人确定他自己在这样的价格下的现期和远期的商品需求和供给。拍卖者计算总超额需求向量并根据某个规则调整价格，他会提高处于超额需求状态的商品的价格，降低处于超额供给状态的商品的价格。这个过程一直继续下去，直到找到均衡价格向量。在这一点上，所有交易（包括远期交易合同的交换）都被确定下来。经济体继续运行，每个人按照签订的合同执行。

这当然是一个非常不现实的模型。然而，价格沿着超额需求方向移动这个基本思想似乎是合理的。在什么样的条件下，这样的调整过程才能到达均衡？

21.5 反复试探以达到均衡的过程（tatonnement processes）

我们考虑某个时期内的经济体。每天市场都开放，人们在市场上进行买和卖。任意给定一个价格向量 p，某些市场通常存在超额需求和供给。我们假设价格根据下列规则进行调整，我们将这一规则称为供给和需求规则。

价格调整规则。**对于** $i=1,\ldots,k$**，** $\dot{p}_i = G_i(z_i(p))$**。其中，**G_i **是平滑且保号的超额需求函数**。

我们想作出假设将均衡在零价格向量发生这种情形排除，因为这么做便于分析。因此，我们通常假设：当 $p_i = 0$ 时，$z_i(p) > 0$。

画出这个价格调整规则定义的动态系统的图形是有益的。我们考虑一种特殊情形，其中：对于 $i=1,\ldots,k$，$G_i(z_i)$ 等于恒等函数。再加上边界假设，我们有了一个由下式定义的 R^k 中的系统：

$$\dot{\mathbf{p}} = z(\mathbf{p})$$

从通常的考虑可知，这个系统遵循瓦尔拉斯法则 $\mathbf{p}z(\mathbf{p}) \equiv \mathbf{0}$。在几何图形上，这意味着 $z(\mathbf{p})$ 与价格向量 $\mathbf{p}$ 正交。

瓦尔拉斯法则蕴涵着一个非常方便的性质。下面我们看看价格向量的欧几里得范数（norm）如何随时间变化而变化，根据瓦尔拉斯法则可知：

$$\frac{d}{dt}\left(\sum_{i=1}^{k} p_i^2(t)\right) = \sum_{i=1}^{k} 2p_i(t)\dot{p}_i(t) = 2\sum_{i=1}^{k} 2p_i(t)z_i(\mathbf{p}(t)) = 0。$$

因此，瓦尔拉斯法则要求当价格变化时，价格的平方和维持不变。这意味着价格路径被限制在 K 维球体的表面。而且，由于当 $p_i = 0$ 时 $z_i(\mathbf{p}) \geq 0$，我们知道价格运动总是指向 $p_i = 0$ 附近的点。在图 21.5 中，我们画出了 k=2 和 k=3 的情形。

第三张图特别让人不舒服。它画出了只有一个均衡的情形，但这个均衡完全不稳定。我们所描述的调整过程几乎永远不会向均衡收敛。这似乎是个反常的情形，但它非常容易发生。Debreu（1974）已经证明：任何满足瓦尔拉斯法则的连续函数都是某个经济的超额需求函数；因此，效用最大化假设对**总需求**行为没有施加任何限制，价格球体上的任何动态系统都可从我们的经济模型中产生。显然，为了得到全局稳定结果，必须对需求函数施加特别的假设条件。于是，结果的值取决于假设条件的经济自然性。

我们将给出一个特殊调整过程下的一个这样特殊假设下的全局稳定性的证明梗概。这个假设就是总需求行为满足第 8 章 8.7 节描述的显示偏好弱公理。这个弱公理是说如果 $\mathbf{px}(\mathbf{p}) > \mathbf{px}(\mathbf{p}^*)$，我们必定有 $\mathbf{p}^*\mathbf{x}(\mathbf{p}) > \mathbf{p}^*\mathbf{x}(\mathbf{p}^*)$ 对于所有的 $\mathbf{p}$ 和 $\mathbf{p}^*$ 成立。由于这个条件对所有 $\mathbf{p}$ 和 $\mathbf{p}^*$ 成立，它对于均衡价格 $\mathbf{p}^*$ 也必定成立。下面我们将推导出这个条件对于超额需求函数的含义。

将 $\mathbf{px}(\mathbf{p}) > \mathbf{px}(\mathbf{p}^*)$ 的左右两侧同时减去 $\mathbf{p}\omega$，将 $\mathbf{p}^*\mathbf{x}(\mathbf{p}) > \mathbf{p}^*\mathbf{x}(\mathbf{p}^*)$ 的两侧同时减去 $\mathbf{p}^*\omega$，可得到下列蕴含关系

$$\mathbf{px}(\mathbf{p}) - \mathbf{p}\omega > \mathbf{px}(\mathbf{p}^*) - \mathbf{p}\omega \text{蕴含着} \mathbf{p}^*\mathbf{x}(\mathbf{p}) - \mathbf{p}^*\omega > \mathbf{p}^*\mathbf{x}(\mathbf{p}^*) - \mathbf{p}^*\omega.$$

使用超额需求的定义，我们可以将上面的关系写为

$$\mathbf{p}z(\mathbf{p}) \geq \mathbf{p}z(\mathbf{p}^*) \text{蕴含着} \mathbf{p}^*z(\mathbf{p}) \geq \mathbf{p}^*z(\mathbf{p}^*). \qquad (21.1)$$

现在可以注意到，任何均衡价格 $\mathbf{p}^*$ 必定满足（21.1）左侧的条件。为了看清这一点，注意到瓦尔拉斯法则意味着 $\mathbf{p}z(\mathbf{p}) \equiv 0$，以及均衡的定义意味着 $\mathbf{p}z(\mathbf{p}^*) = 0$ 。因此，任何均衡价格 $\mathbf{p}^*$ 必定满足（21.1）右侧的条件。所以，我们必然有：$\mathbf{p}^*z(\mathbf{p}) > 0$ 对于所有 $\mathbf{p} \neq \mathbf{p}^*$ 成立。

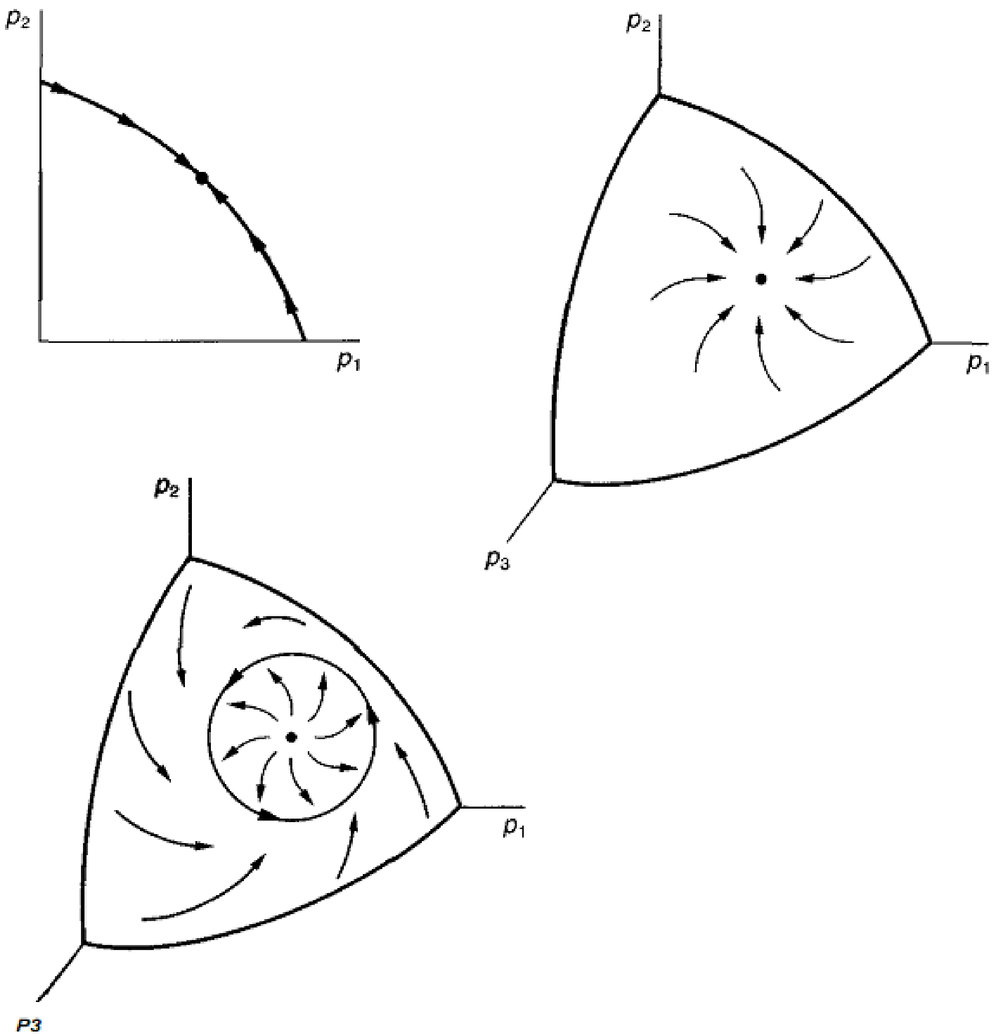


图 21.5：价格动态。前两个例子是稳定均衡；第三个例子有唯一的不稳定均衡。

WARP 蕴含着稳定性（WARP implies stability）。**假设调整规则为** $\dot{\mathbf{p}}_i = z_i(\mathbf{p})$ **（其中** $i = 1, \ldots, k$ **）；假设超额需求函数满足显示偏好弱公理（WARP），也就是说，如果** $\mathbf{p}^*$ **是经济的一个均衡，那么对于所有** $\mathbf{p} \neq \mathbf{p}^*$ **，有** $\mathbf{p}^* z(\mathbf{p}) > 0$ **。那么，遵循上述调整规则的所有价格路径都收敛于** $\mathbf{p}^*$ **。**

证明。（梗概）我们需要构造经济的李雅普诺夫（Liaponov）函数（参见第 26 章 26.13 节）。令 $V(\mathbf{p}) = \sum_{i=1}^{k}\left[(p_i - p_i^*)^2\right]$。那么，

$$\frac{dV(\mathbf{p})}{dt} = \sum_{i=1}^{k} 2(p_i - p_i^*)\dot{p}_i(t) = 2\sum_{i=1}^{k}(p_i - p_i^*)z_i(\mathbf{p})$$
$$= 2\sum_{i=1}^{k}[p_i z_i(\mathbf{p}) - p_i^* z_i(\mathbf{p})] = 0 - 2\mathbf{p}^* z(\mathbf{p}) < 0.$$

这意味着 $\mathbf{p} \neq \mathbf{p}^*$ 时，$V(\mathbf{p})$ 沿着解路径是单调递减的。根据李雅普诺夫（Liaponov）定理可知，只要我们能够证明 $\mathbf{p}$ 是有界的，我们就能证明 $V(\mathbf{p})$ 是个李雅普诺夫函数，而且经济是全局稳定的。我们省略了这部分的证明。■

21.6 无试探过程

试探过程在两种情形下有意义：一是直到均衡之前没有交易发生；二是商品不可储存，因此消费者在每个阶段的禀赋是相同的。如果商品可以积累，那么消费者的禀赋将随着时间的推移而发生变化，这又会影响需求行为。能够考虑禀赋的这种变化的模型称为**无试探模型**（nontatonnement models）。

在这样的模型中，我们必须用当前的价格向量 $\mathbf{p}(t)$ 和当前的禀赋 $\omega_i(t)$ 刻画时点 t 的经济状态。和以前一样，我们通常假设价格根据超额需求的符号进行调整。但禀赋应该如何变化？

我们考虑两种规格（specifications）。第一种规格称为**埃奇沃思过程**（Edgeworth process），在这种过程下，消费者之间的交易技术具有下列性质：每个消费者的效用必定连续递增。这个结论背后的思想是消费者不会自愿交易除非交易能让他们的状况变好。这个规格有一个方便的性质，它能快速导致稳定性；我们只要将李雅普诺夫函数定义为 $\sum_{i=1}^{n} u_i(\omega_i(t))$ 即可。根据假设，效用之和必定随着时间的推移而增加，因此只要证明了有界性，即可证明收敛性。

第二种规格称为**哈恩过程**（Hahn process）。对于这个过程，我们假设交易规则具有下列性质：不存在下列这样的商品——对于该商品有些人存在着超额需求而有些人存在超额供给。也就是说，在任何时点上，如果某个人对某种商品存在超额需求，那么该商品存在**总超额需求**（aggregate excess demand）。

这个假设具有重要含义。我们已经假设当某个商品存在着超额需求时，它的价格将会上升。这会导致该商品需求者的间接效用值降低。但已经承诺以当前价格供给这种商品人的效用，不会受到这种价格变化的影响。因此，**总间接效用**随着时间的推移而下降。

为了严格证明这个结论，我们需要进一步对禀赋的变化做出假设。消费者 i 在时点 t 的禀赋价值为 $m_i(t) = \sum_{j=1}^{k} p_j(t)\omega_i^j(t)$. 将这个式子关于 t 求导可得

$$\frac{dm_i(t)}{dt} = \sum_{j=1}^{k} p_j(t)\frac{d\omega_i^j(t)}{dt} + \sum_{j=1}^{k} \frac{dp_j(t)}{dt}\omega_i^j(t).$$

可以合理假设上式右侧第一项为零，这意味着禀赋在任何时间上的瞬间变化，若用当前的价格来计算价值，其值为零。这就是说每个人用价值一元的商品交换价值一元的另外商品。禀赋的价值随着时间的推移会发生变化，但其原因在于价格发生了变化，而不是因为消费者在当前的价格下想进行有利可图的交易。

给定上述结论，容易证明间接效用之和随着时间的推移而下降。消费者 i 的间接效用函数的导数为

$$\frac{dv_i(\mathbf{p}(t), \mathbf{p}(t)\omega_i(t))}{dt} = \sum_{j=1}^{k} \frac{\partial v_i}{\partial p_j}\frac{dp_j}{dt} + \frac{\partial v_i}{\partial m_i}\left[\sum_{j=1}^{k} p_j \frac{d\omega_i^j}{dt} + \frac{dp_j}{dt}\omega_i^j\right].$$

使用罗伊法则（Roy's law）以及在当前价格下禀赋变化的价值为零这个事实，可得

$$\frac{dv_i(\mathbf{p}(t), \mathbf{p}(t)\omega_i(t))}{dt} = -\frac{\partial v_i}{\partial m_i}\sum_{j=1}^{k}\left[x_i^j(\mathbf{p}, \mathbf{p}\omega_i) - \omega_i^j\right]\frac{dp_j(t)}{dt}.$$

根据假设，如果消费者 i 对商品 j 存在超额需求，则 $dp_j / dt > 0$；反之则反是。由于收入的边际效用为正，所以只要总需求不等于总供给，整个式子的符号为负。因此，当经济不是处于均衡状态时，每个消费者 i 的间接效用必定减少。

注释

这些主题的更详细讨论请参见 Arrow and Hahn（1971）。Dierker（1972）首先认识到了拓扑指数对于唯一性的重要性。Debreu and Scarf（1963）严格证明了核收敛的结论。

习题

21.1 有两个消费者：他们的偏好相同，且都为严格凸；他们的禀赋也是相同的。描述这个经济的核，并在埃奇沃思盒中画图说明。

21.2 在某个纯交换的经济中，所有消费者都具有可微的拟线性效用函数$u(x_1,...,x_n)+x_0$。假设$u(x_1,...,x_n)$为严格凸。证明均衡必定是唯一的。

21.3 假设瓦尔拉斯主义的拍卖者遵循下列下列价格调整规则$\mathbf{p}=[\mathrm{D}z(\mathbf{p})]^{-1}z(\mathbf{p})$。证明$V(\mathbf{p})=-z(\mathbf{p})z(\mathbf{p})$是此动态系统的李雅普诺夫函数。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 22 章：福利

曹乾 译

(东南大学 caoqianseu@163.com)

22 福利

在本章，我们将分析一些福利经济学概念，这些概念不怎么适合于放在本书其他章节介绍。第一个概念是补偿标准，这个标准常用于收益-成本分析。接下来我们讨论在计算产量或价格变动引致的福利效应时常用到的一个技巧。最后，我们分析最优商品税的问题。

22.1 补偿标准

人们通常想知道政府政策何时能够改善社会福利。例如，建设水坝具有降低水电价格等方面的经济利益。然而，为了正确评价这样的项目，我们必须将这些收益与成本进行权衡，这些成本包括可能的环境损坏造成的成本和建坝本身的成本等。一般来说，项目的收益和成本对不同人的影响方式也不同——建坝导致的水供给增加可能会降低某些地区的水费但可能会提高另外一些地区的水费。我们应该怎样比较这些不同的收益和成本？

以前我们曾分析过下列情形：由于某种消费品的价格或数量变化对**某人**造成的收益和成本变动的测度问题。本节我们试图使用**帕累托标准**（Pareto criterion）和**补偿标准**（compensation criterion）将上述这类分析推广到**社区**情形。

考虑两个配置$\mathbf{x}$和$\mathbf{x}'$。如果每个人偏好$\mathbf{x}'$胜于$\mathbf{x}$，我们就说配置$\mathbf{x}'$**帕累托优于**（Pareto dominate）$\mathbf{x}$[一]。如果每个人都偏好$\mathbf{x}'$胜于x，那么必然可以断言$\mathbf{x}'$比$\mathbf{x}$“更好”，因此可将人们从$\mathbf{x}$移动到$\mathbf{x}'$的任何项目都应该被采纳。这就是**帕累托标准**。然而，项目受到一致偏好的情形是很罕见的。一般情形下，有些人偏好$\mathbf{x}'$胜于$\mathbf{x}$，有些人则偏好$\mathbf{x}$胜于$\mathbf{x}'$。在这样的情形下，我们该如何决策？

补偿标准提议的检验方法是：如果存在对$\mathbf{x}'$的再分配方法（达到配置$\mathbf{x}''$）使得每个人偏好x''胜于原来的配置$\mathbf{x}$，则$\mathbf{x}'$是**潜在帕累托优于**（potentially Pareto preferred to）x的。下面我们更正式地进行表述：$\mathbf{x}'$潜在帕累托优于$\mathbf{x}$，若存在某个配置$\mathbf{x}''$，$\sum_{i=1}^{n}\mathbf{x}_i''=\sum_{i=1}^{n}\mathbf{x}_i'$（即$\mathbf{x}''$是$\mathbf{x}'$再分配后达到的配置）使得对于所有人$i$都有$\mathbf{x}_i'' \succ_i \mathbf{x}_i'$。

因此，补偿标准仅要求$\mathbf{x}'$是$\mathbf{x}$的帕累托改进（而不是要求$\mathbf{x}'$是帕累托最优的）。为便于表达，称某个人为“赢家”如果他偏好$\mathbf{x}'$胜于$\mathbf{x}$，称某个人为“输家”如果他偏好$\mathbf{x}$胜于$\mathbf{x}'$。于是，在补偿检验的意义上我们说$\mathbf{x}'$比$\mathbf{x}$好，如果赢家们能够补偿输家们——也就是说，赢家们能让渡一部分收益给输家们从而使得每个人的状况都比以前更好。

现在如果赢家们真正补偿了输家们，则每个人将接受政府提议的项目，这似乎是合理的。但是你也许不明白为什么赢家们可能补偿输家们就意味着$\mathbf{x}'$比$\mathbf{x}$好。

[一] 为简单起见，此处我们使用了严格偏好的概念；这个思想可以容易地推广到弱偏好。

经济学家认为补偿标准可行的原因在于，补偿是否真正执行事实上是一个关于收入分配的问题，而基本福利定理表明收入分配问题和配置效率问题是可以分离开的。补偿标准关注的仅仅是配置效率，而合理的收入分配问题可通过其他方法比如再分配税进行处理。我们在第 22 章将进一步分析这一点。

我们借助图形重新表述这个问题。假设只有两个个人，他们正在考虑两个配置 $\mathbf{x}$ 和 $\mathbf{x}'$。对于每个配置我们赋予相应的效用可能集：

$$U=\{u_1(y_1),u_2(y_2):y_1+y_2=x_1+x_2\}$$
$$U'=\{u_1(y_1),u_2(y_2):y_1+y_2=x_1'+x_2'\}.$$

这些集合的右上边界称为**效用可能性边界**（utility possibility frontiers）。效用可能性边界给出了与所有帕累托有效率的配置 $\mathbf{x}$ 和 $\mathbf{x}'$ 相伴的效用分布。图 22.1 给出了效用可能性集的一些例子。

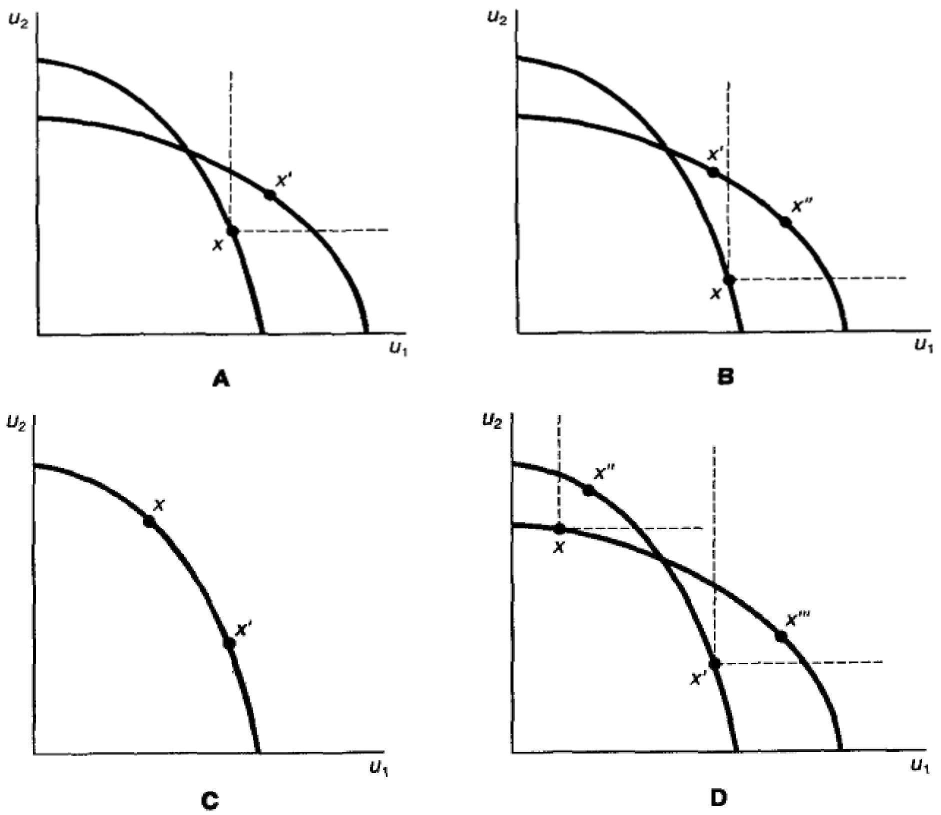


图 22.1：补偿检验。在 A 图中，配置 $\mathbf{x}'$ 帕累托优于 $\mathbf{x}$。在 B 图中，$\mathbf{x}'$ 在补偿检验意义上优于 $\mathbf{x}$。在 C 图中，$\mathbf{x}$ 和 $\mathbf{x}'$ 是不可比较的。在 D 图中，$\mathbf{x}'$ 优于 $\mathbf{x}$ 而且 $\mathbf{x}$ 优于 $\mathbf{x}'$。

在图 22.1A 中，配置 $\mathbf{x}'$ 是帕累托优于 $\mathbf{x}$ 的，这是因为 $u_1(x_1')>u_1(x_1)$ 且 $u_2(x_2')>u_1(x_2)$。在图 22.1B 中，$\mathbf{x}'$ 潜在帕累托优于 $\mathbf{x}$：存在 $\mathbf{x}'$ 的再分配 $\mathbf{x}''$ 使得 $\mathbf{x}''$ 是帕累托优于 $\mathbf{x}$ 的，尽管 $\mathbf{x}'$ 本身可能不是帕累托优于 $\mathbf{x}$。因此，$\mathbf{x}'$ 满足补偿标准：若从 $\mathbf{x}$ 移动到 $\mathbf{x}'$，赢家们就会对输家们进行补偿。在图 22.1C 中，$\mathbf{x}'$ 和 $\mathbf{x}$ 是不可以比较的——无法从补偿检验或帕累托检验说出哪个配置更好。在图 221.D 中，我们有最矛盾的情形：$\mathbf{x}'$ 潜在帕累托优于 $\mathbf{x}$，因为 $\mathbf{x}''$

帕累托优于$\mathbf{x}$；但是$\mathbf{x}$也潜在帕累托优于$\mathbf{x}'$，因为$\mathbf{x}''$帕累托优于$\mathbf{x}'$！

情形 C 和 D 说明了补偿标准的主要缺陷：它对于比较帕累托有效率配置不具有知道作用，而且它可能导致不相容的比较。尽管如此，补偿检验仍普遍应用于应用福利经济学中。

正如我们说过的那样，补偿检验要求我们考虑项目对所有相关消费者的效用影响。表面上，这似乎要求对人群进行广泛而细致地调研。然而，在某些情形下，不必这么做。

如果拟实施的项目是公共产品，在作出社会决策时必然要进行人群调查研究。我们在第 23 章分析这个问题。如果拟实施的项目是私人产品，事情就好办一些，因为私人产品的当前价格在某种意义上可以反映它们对于个人的边际价值。

假设我们目前处在某个市场均衡$(\mathbf{x},\mathbf{p})$处，我们考虑是否将配置$\mathbf{x}$移动到$\mathbf{x}'$。那么

国民收入检验（national income test）。**如果$\mathbf{x}'$是潜在帕累托优于$\mathbf{x}$的，我们必定有**

$$\sum_{i=1}^{n}\mathbf{p}\mathbf{x}_i' > \sum_{i=1}^{n}\mathbf{p}\mathbf{x}_i.$$

也就是说，若以当前价格衡量，$\mathbf{x}'$点的国民收入要大于$\mathbf{x}$点的国民收入。

证明。如果在补偿标准意义上$\mathbf{x}'$优于$\mathbf{x}$，那么必定存在某个配置$\mathbf{x}''$使得$\sum_{i=1}^{n}\mathbf{x}_i''=\sum_{i=1}^{n}\mathbf{x}_i'$且所有$i$都有$\mathbf{x}_i''\succ_i \mathbf{x}_i'$。由于$x$是一个市场均衡，这意味着对于所有$i$都有$\mathbf{p}\mathbf{x}_i''>p\mathbf{x}_i$。相加即可得到$\sum_{i=1}^{n}\mathbf{p}\mathbf{x}_i''=\sum_{i=1}^{n}\mathbf{p}\mathbf{x}_i$。但是，

$$\sum_{i=1}^{n}\mathbf{p}\mathbf{x}_i''=\mathbf{p}\sum_{i=1}^{n}\mathbf{x}_i''=\mathbf{p}\sum_{i=1}^{n}\mathbf{x}_i'.$$

这就得到了我们想要的结果。 ■

这个结果比较有用，因为它为拟实施的项目提供了一种单向检验：如果国民收入（以当前价格衡量）下降，那么这个项目不可能潜在地帕累托优于当前的配置。

图 22.2 直观说明了这个结论。两个坐标轴分别衡量两种商品的总量。当前的配置可用一个总消费束$\mathbf{X}=(X^1,X^2)$表示，其中$X^1=\sum_{i=1}^{n}x_i^1$，X^2的定义类似（注意我们用下标表示消费者，用上标表示商品。）

对于一个总消费束$\mathbf{X}'$，如果$\mathbf{X}'$能在消费者之间分配构建出一个配置$\mathbf{x}'$，而且$\mathbf{x}'$帕累托优于$\mathbf{x}$，那么我们说$\mathbf{X}'$潜在地帕累托优于$\mathbf{x}$。换句话说，潜在的帕累托更受偏好的总消费束为

$$\mathbf{P}=\{\sum_{i=1}^{n}\mathbf{x}_i' : \mathbf{x}_i'\succ_i \mathbf{x}_i \text{对于所有} i \text{成立}\}$$

图 22.2 说明了一种典型情形。集合$\mathbf{P}$是性质良好的和凸的，而且总消费束$\mathbf{X}$在它的边界上。竞争性价格将$\mathbf{X}$和$\mathbf{P}$分离开。从这个图容易看出上述命题的内容：如果在补偿标准

的意义上，$\mathbf{x}'$比$\mathbf{x}$更受偏好，那么$\mathbf{X}'$必定在$\mathbf{P}$之中，因此$\mathbf{pX}' > \mathbf{pX}$。

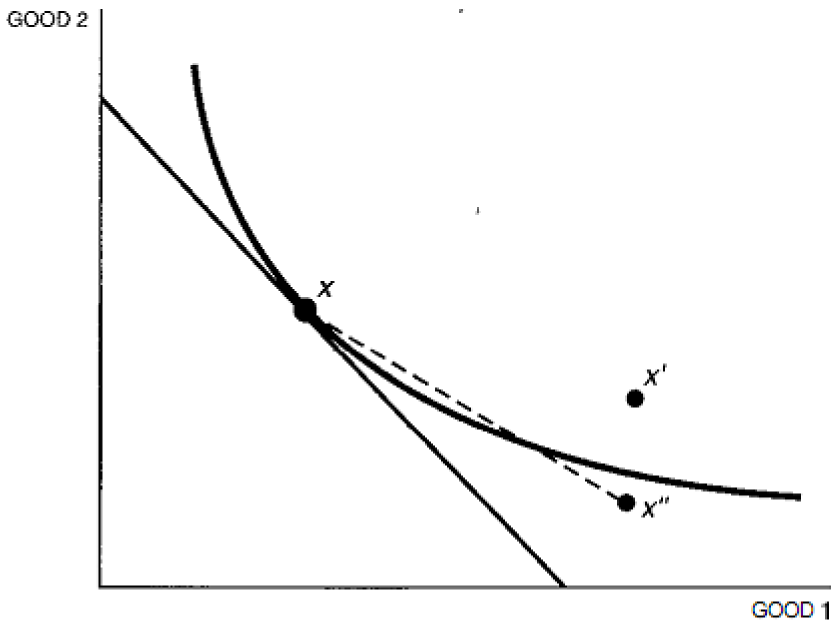


图 22.2：国民收入检验。如果国民收入下降，这个变化不可能是潜在帕累托受偏好的。如果国民收入增加，这个变化可能是潜在帕累托受偏好的，也可能不是。然而，如果较小变动就能导致国民收入增加，那么这个变化可能是潜在帕累托受偏好的。

我们还可以看到逆命题非真。消费束$\mathbf{X}''$有$\mathbf{pX}'' > \mathbf{pX}$，但它不潜在地帕累托优于$\mathbf{x}$。然而，从这个图中我们可以猜想：如果$\mathbf{pX}'' > \mathbf{pX}$而且$\mathbf{X}''$充分接近于$\mathbf{X}$，那么$\mathbf{X}''$必定潜在地帕累托优于$\mathbf{x}$。更准确地说，考察连接$\mathbf{X}''$与$\mathbf{X}$的虚线。这条虚线上的所有点的值都大于$\mathbf{X}$，但是这条虚线上的所有点并非都位于穿过$\mathbf{X}$的无差异曲线的上方。然而，这条虚线上充分接近于$\mathbf{X}$的点都位于这条无差异曲线的上方。下面我们正式地用代数表达这个思想。

论证基于以下事实。对于一阶近似来说，个人效用的变化与收入变化成正比。这可从泰勒展开式推知：

$$u_i(\mathbf{x}_i') - u_i(\mathbf{x}_i) \approx \mathrm{D}u_i(\mathbf{x}_i)[\mathbf{x}_i' - \mathbf{x}_i] = \lambda_i[\mathbf{x}_i' - \mathbf{x}_i].$$

根据这个表达式可知，若消费束的效用值变化为正，则消费束$\mathbf{x}_i$的微小变化是受到偏好的；反之若为负，则不受到偏好。

我们使用这个思想证明，如果$\mathbf{p}\sum_i \mathbf{x}_i' > \mathbf{p}\sum_i \mathbf{x}_i$而且$\mathbf{x}_i'$充分接近于$\mathbf{x}_i$，那么能够找到$\mathbf{x}'$的一个再分配——称为$\mathbf{x}''$——使得每个人偏好$\mathbf{x}''$胜于$\mathbf{x}$。为了证明这一点，令$\mathbf{X} = \sum_i \mathbf{x}_i$和$\mathbf{X}' = \sum_i \mathbf{x}_i'$，并且定义$\mathbf{x}''$为

$$\mathbf{x}_i'' = \mathbf{x}_i + \frac{\mathbf{X}' - \mathbf{X}}{n}$$

此处每个消费者得到了从 $\mathbf{x}$ 移动到 $\mathbf{x}'$ 带来的总收益的 $1/n$ 。根据上面的泰勒序列展开式可得，

$$u_i(\mathbf{x}'') - u_i(\mathbf{x}) \approx \lambda_i \mathbf{p}[\mathbf{x}_i + \frac{\mathbf{X}' - \mathbf{X}}{n} - \mathbf{x}_i] \approx \lambda_i \mathbf{p}[\frac{\mathbf{X}' - \mathbf{X}}{n}].$$

因此，如果上式右侧为正——国民收入在最初的价格下上升——那么它必定增加每个人的效用。当然，只有变化足够小从而保证泰勒近似有效的情形下，它才成立。国民收入检验通常用于评价边际政策变动对消费者福利的影响。

22.2 福利函数

我们在本章前面说过，补偿方法的缺陷在于它忽略了分布上的考虑。一个配置若潜在地帕累托优于当前的配置，那么该配置的福利更高。但有人认为真正重要的是实际福利。

如果你愿意假定存在着某个福利函数，那么你可以将分布上的考虑纳入成本收益分析。我们假设我们有线性效用福利函数

$$W(u_1,...,u_n) = \sum_{i=1}^{n} a_i u_i.$$

正如我们在第 17 章 17.8 节看到的，参数 (a_i) 与个人的“福利权重”有关。这些权重可以视为“社会计划者”的价值判断。我们假设我们在某个市场均衡 $(\mathbf{x},\mathbf{p})$ 处，并且正考虑移向配置 $\mathbf{x}'$ 。这个移动能增加福利吗？如果 $\mathbf{x}'$ 充分接近于 $\mathbf{x}$ ，我们可以使用泰勒展开式得到

$$W(u_1(\mathbf{x}_1'),...,u_n(\mathbf{x}_n')) - W(u_1(\mathbf{x}_1),...,u_n(\mathbf{x}_n)) \approx \sum_{i=1}^{n} a_i \mathrm{D}u_i(\mathbf{x}_i)(\mathbf{x}_i' - \mathbf{x}_i).$$

由于 $(\mathbf{x},\mathbf{p})$ 是个市场均衡，我们可以将上式写为

$$W(u_1(\mathbf{x}_1'),...,u_n(\mathbf{x}_n')) - W(u_1(\mathbf{x}_1),...,u_n(\mathbf{x}_n)) \approx \sum_{i=1}^{n} a_i \lambda_i \mathbf{p}(\mathbf{x}_i' - \mathbf{x}_i).$$

我们看到福利检验简化为考察支出加权的变化。权重和我们最初纳入效用函数的价值判断有关。

作为一个特例，假设最初的配置 $\mathbf{x}$ 是帕累托最优的。那么第 17 章 17.8 节的结果告诉我们 $\lambda_i = 1/a_i$ 。在这种情形下，我们发现

$$W(u_1(\mathbf{x}_1'),...,u_n(\mathbf{x}_n')) - W(u_1(\mathbf{x}_1),...,u_n(\mathbf{x}_n)) \approx \sum_{i=1}^{n} \mathbf{p}(\mathbf{x}_i' - \mathbf{x}_i).$$

分布项被消掉了——因为分布已是帕累托最优的——我们得到了一个简单的准则：如果某个

项目使得国民收入（在原来的价格水平上）增加，那么该项目增加了福利。这正好是与补偿检验相关的标准。

这意味着如果社会计划者始终遵循使得福利不仅关于总额收入分布最大化而且关于音响配置的其它政策选择最大化，那么影响配置的政策的评价与收入分配效应无关。

22.3 最优税收

我们在第 8 章 8.1 节看到，定额收入税（lump-sum income tax）总是比商品税更能得到偏好。然而，在很多情形下，定额收入税是不可行的。如果我们不能使用定额税，什么样的税才是最优的？

我们用只有一个消费者的经济来考察这个问题。令 $u(\mathbf{x})$ 表示消费者的间接效用函数，令 $v(\mathbf{p},m)$ 表示他的间接效用函数。我们将 $\mathbf{p}$ 解释为生产者的价格。如果 $\mathbf{t}$ 是税收向量，那么消费者面对的价格向量为 $\mathbf{p}+\mathbf{t}$ 。这样，消费者得到的效用为 $v(\mathbf{p}+\mathbf{t},m)$ ，政府的税收收入为 $R(\mathbf{t})=\sum_{i=1}^{k}t_i x_i(\mathbf{p}+\mathbf{t},m)$ 。

最优税收问题为消费者的效用关于税率最大化，约束条件为税收系统筹集的钱数为既定的，即等于 R 。

$$\max_{t_1,\dots,t_k} \quad v(\mathbf{p}+\mathbf{t},m)$$
$$\text{s.t.} \quad \sum_{i=1}^{k} t_i x_i(\mathbf{p}+\mathbf{t},m)=R$$

这个问题的拉格朗日函数为

$$\mathcal{L}=v(\mathbf{p}+\mathbf{t},m)-\mu\left[\sum_{i=1}^{k} t_i x_i(\mathbf{p}+\mathbf{t},m)-R\right]。$$

将其关于 t_i 微分可得

$$\frac{\partial v(\mathbf{p}+\mathbf{t},m)}{\partial p_i}-\mu\left[x_i+\sum_{j=1}^{k}t_j\frac{\partial x_j(\mathbf{p}+\mathbf{t},m)}{\partial p_i}\right]=0 \qquad i=1,\dots,k.$$

运用罗伊法则，我们可以写成

$$-\lambda x_i-\mu\left[x_i+\sum_{j=1}^{k}t_j\frac{\partial x_j(\mathbf{p}+\mathbf{t},m)}{\partial p_i}\right]=0 \qquad i=1,\dots,k.$$

解 x_i 可得：

$$x_i=-\frac{\mu}{\mu+\lambda}\sum_{j=1}^{k}t_j\frac{\partial x_j(\mathbf{p}+\mathbf{t},m)}{\partial p_i}$$

现在使用上式右侧的斯勒茨基方程可得

$$x_i = -\frac{\mu}{\mu+\lambda}\sum_{j=1}^{k} t_j \left[\frac{\partial h_j}{\partial p_i} - \frac{\partial x_j}{\partial m} x_i\right].$$

在经过一些变换后，上式可以写为

$$\theta x_i = \sum_{j=1}^{k} t_j \frac{\partial h_j}{\partial p_i},$$

其中θ为关于μ、λ和$\sum_j t_j \partial x_j / \partial m$的函数。

运用斯勒茨基矩阵的对称性，我们可以将前面的那个式子写为

$$\theta x_i = \sum_{j=1}^{k} t_j \frac{\partial h_i}{\partial p_j}. \tag{22.1}$$

将（22.1）变换成弹性形式，可得

$$\theta = \sum_{j=1}^{k} \frac{\partial h_i}{\partial p_j} \frac{p_j}{x_i} \frac{t_j}{p_j} = \sum_{j=1}^{k} \varepsilon_{ij} \frac{t_j}{p_j}.$$

这个式子是说，我们对税率的选择必须使得希克斯交叉价格弹性的加权和对于所有商品来说都是相同的。

在$\varepsilon_{ij} = 0$（其中$i \neq j$）这个极端情形下，上述条件变为

$$\frac{t_i}{p_i} = \frac{\theta}{\varepsilon_{ij}} \tag{22.2}$$

因此，商品i的税率与价格之比值，与需求价格弹性成反比。这称为**反弹性法则**（inverse elasticity rule）。这符合直觉：你应该对需求相对缺乏弹性的商品征收较高的税率，对需求相对富有弹性的商品征收较低的税率。这种做法对消费者决策的扭曲最少。

在另外一种情形下，即当税率t_i很小时，我们可以对前面的那个条件进行简化。在这种情形下

$$dh_i \approx \sum_{j=1}^{k} t_j \frac{\partial h_i}{\partial p_j}.$$

将上式代入（22.1）可得

$$\frac{dh_i}{h_i} \approx \theta.$$

这个式子是说，由较小税率组成的最优集使得所有补偿需求按相同的比例减少。■

注释

本节的内容是相等标准的；更为详细的内容可参考任何一本关于成本收益分析的教科书。最优税收理论的文献综述可参见 Mirrlees（1982）或 Atkinson and Stiglitz（1980）。

习题

22.1 在教材推导出的最优税收表达式（22.1）中，证明如果所需税收收入为正，那么θ是非负的。

22.2 一个公用事业生产的产品产量为$x_1,...,x_k$。某个代表性的消费者消费这些商品，他的效用函数为$u_1(x_1)+\cdots+u_k(x_k)+y$，其中$y$为充当计量物（numeraire）的商品。该公用事业生产商品i的边际成本为c_i，而且它的固定成本为F。请推导出最优定价法则的表达式，在该式中你要将(p_i-c_i)与商品i的弹性联系起来。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 23 章：公共产品

曹乾　译

(东南大学　caoqianseu@163.com)

23 公共产品

直到现在，我们对资源配置的讨论还仅限于私人产品，私人产品的消费仅影响单个的消费者。例如，面包。你和我可以消费不同数量的面包，而且如果我吃一块面包，那么这块面包你就无法吃了。

如果人们能被阻止消费某种商品，我们说该商品是可排外的（excludable）。我们说一种商品是非竞用性的（nonrival），如果一个人消费该商品并不会减少其他人消费该商品的数量。竞用的商品有时称为可缩小的（diminishable）商品。普通的私人商品既是可排外的也是竞用的。

有些商品不具有这样的性质。路灯就是一个很好的例子。某个给定区域的路灯数量是一定的——你和我拥有相同的潜在消费，而且我的“消费”不会影响你的消费数量。因此，路灯是非竞用的。更进一步地说，我对路灯的消费没有限制你对路灯的消费。非排外而且非竞用的商品称为**公共产品**（public goods）；公共产品的例子还有警察服务、消防、高速公路、国防、灯塔、电视和无线电广播、空气等等。

有些商品介于私人商品和公共产品之间。例如加密的电视节目。这种产品是非竞用性的——因为一个人的消费不会减少另外一个人的消费——但是它是排他性的，这是因为只有拥有解码器的人才能看这种电视节目。具有这种性质的商品通常称为**俱乐部商品**（club goods）。

还有一类商品，它们是非排他性的，但却是竞用性的。例如，拥挤的街道就具有这样的性质：每个人都可以在街道上走（非排他性），但是你在街道上走却减少了其他人的空间（竞用性）。

最后，某些商品在本质上是私人商品，但是却被当成公共产品看待。例如，教育本质上是私人产品——它具有可排他性，而且在某种程度上，具有竞用性。然而，大多数国家的政治决策都是提供公共教育服务。通常政治决策也要求对每个公民提供相同数量的教育经费。这种限制要求我们将教育看成一种公共产品。

公共产品的资源配置问题和私人产品的资源配置问题大不相同。我们在前面已经知道，竞争市场是一种有效的（effective）社会机制，它可以有效率的配置私人商品。然而，人们发现竞争市场并**不是**配置公共产品的好方法。一般来说，必须使用其他社会机制，例如投票等来进行公共产品的配置。

23.1 离散公共产品的有效率的提供

我们首先研究两个人-两种商品的简单例子。一种商品，x，是私人商品，可以看成花费

在私人消费的金钱。另外一种商品，G，是公共产品，可以看成花费在公共产品（例如路灯）上的金钱。每个人最初拥有一些私人商品禀赋 w，并且需要决定要为公共产品花费多少钱。如果个人 i 决定为公共产品出资 g_i，那么他的私人商品的消费为 $x_i = w_i - g_i$。假设每个人的效用函数都是公共产品和私人商品消费的严格增函数。我们用 $u_i(G, x_i)$ 表示个人 i 的效用函数。

我们先分析公共产品的数量为离散的（discrete）情形，即为整数的情形；公共产品的数量要么提供既定的整数数量，要么为零。假设每单位公共产品的成本为 c，则它的技术如下：

$$G = \begin{cases} 1 & \text{若} g_1 + g_2 \geq c \\ 0 & \text{若} g_1 + g_2 < c. \end{cases}$$

稍后我们将考虑更一般的技术。

我们首先关心的问题是提供公共产品是否为帕累托有效率的。如果存在某些出资方式 (g_1, g_2) 使得 $g_1 + g_2 \geq c$ 并且使得

$$\begin{aligned} u_1(1, w_1 - g_1) > u_1(0, w_1) \\ u_2(1, w_2 - g_2) > u_2(0, w_2) \end{aligned} \tag{23.1}$$

则提供公共产品就是帕累托有效率的。

令 r_i 表示个人 i 为了得到一单位公共产品，愿意放弃私人消费的最大数量。我们将其称为个人 i 的最大支付意愿或者个人 i 的保留价格（参考第 9 章）。

根据保留价格 r_i 的定义可知， r_i 必定满足下式

$$u_i(1, w_i - r_i) = u_i(0, w_i). \tag{23.2}$$

将上式代入（23.1）可得

$$u_i(1, w_i - g_i) > u_i(0, w_i) = u_i(1, w_i - r_i),$$

其中 $i = 1,2.$ 由于效用函数是个人消费的严格增函数，

$$w_i - g_i > w_i - r_i$$

其中 $i = 1,2.$

将这两个人的上述不等式相加，可得

$$r_1 + r_2 > g_1 + g_2 \geq c.$$

因此，如果提供公共产品是帕累托有效率的，我们必然有 $r_1 + r_2 > c.$ 也就是说，对公共产品的支付意愿之和，大于提供公共产品的成本。注意这个效率条件和提供私人产品效率条件的区别。在私人产品的情形下，如果个人 i 的支付意愿等于私人产品的生产成本，则提供私人产品是有效率的。而对于公共产品，这个条件变弱，即只要求支付意愿**之和**大于提供公

共产品的成本即可。

证明上述定理的逆定理并不难。假设我们有 $r_1 + r_2 > c$. 于是我们可以选择 g_i 使得 g_i 比 r_i 稍小，从而使得不等式 $g_1 + g_2 > c$ 以及

$$u_i(1, w_i - g_i) > u_i(0, w_i)$$

都成立，其中 $i = 1,2$。这表明当 $r_1 + r_2 > c$ 时，提供公共物品不仅是可行的，而且是帕累托改进的（pareto improving）。我们将上述讨论过程总结如下：当且仅当个人支付意愿之和大于提供公共物品的成本，提供离散公共物品是一种帕累托改进（pareto improvement）。

23.2 离散公共物品的私人提供

私人市场提供公共物品可行吗？假设 $r_i = 100$，其中 $i = 1,2$；$c = 150$。因此，支付意愿之和大于提供公共物品的成本。每个人独立进行是否购买公共产品的决策。然而，由于公共物品的性质是公共的，每个人都无法限制对方消费它。

我们可用一个简单的博弈矩阵表示各自的策略和收益，如表 23.2 所示。

		Consumer 2	
		Buy	Don't buy
Consumer 1	Buy	-50,-50	-50,100
	Don't buy	100,-50	0,0

表 23.2：离散公共物品的私人提供

如果消费者 1 购买，他的收益为 100 元，但是需要支付 150 元。如果消费者 1 购买，但是消费者 2 不购买，消费者 2 就免费得到了 100 元的收益。在这种情形下，我们说消费者 2 对消费者 1 **搭便车**（free riding）。

注意这个博弈和囚犯的两难（Prisoner’s Dilemma）博弈(参考第 15 章)具有类似的结构。在这个博弈中，占优策略是（不买，不买）。每个消费者都不愿意购买公共物品，因为每个人都想搭对方的便车。这个博弈的均衡解因此意味着公共产品不会提供，尽管提供公共产品是有效率的。

这表明，我们**不能**期望人们完全独立的决策能导致帕累托有效率的公共物品的提供数量。一般来说，有必要使用更为复杂的机制。

23.3 离散公共产品的投票表决

公共产品的数量通常使用投票表决的方式决定。这种方法能达到有效率的结果吗？假设有三个消费者决定对是否购买公共物品进行投票表决，该公共物品的价格为 99 元。如果多数票赞成购买，他们需要均摊费用，即每个人支付 33 元。这三个人的保留价格分别为 $r_1 = 90, r_2 = 30$ 以及 $r_3 = 30$。

显然，三人的保留价格之和超过了公共物品的价格，因此购买是帕累托有效率的。然而，在这种情形下，只有消费者 1 会投票赞成购买，因为只有他能从购买公共产品后获得正的收益（90>33）。多数人投票方法的缺陷在于它只衡量了公共物品的序数效用，然而效率条件要求比较支付意愿。消费者 1 也许会对其他消费者进行补偿，如果其他消费者投票赞成购买的话，但只是也许会而不是一定会补偿。

另外一类投票方法是个人报出自己的支付意愿，规定当他们各自报出的支付意愿之和大于公共产品的成本时，才提供公共产品。如果费用平摊，那么通常该博弈不存在均衡解。例如前面三个消费者进行投票的例子。在这种情形下，提供公共产品会使消费者 1 的状况变好，因此他可能故意报出很高的支付意愿。类似地，消费者 2 和 3 会报出很小的支付意愿。

还有一类投票方法是，每个人报出自己的支付意愿。如果他们报出的支付意愿之和大于或等于公共产品的成本，则提供公共产品，而且每人都必须按照他报出的支付意愿出钱（即 $g_i = r_i$）。在这种情形下，如果提供公共产品是帕累托有效率的，那么这个博弈存在着均衡解。例如，每个人报出的支付意愿不大于他的保留价格，而且他们报出的支付意愿之和大于公共产品的成本，就是一个均衡解。然而，这个博弈也存在**无效率**的均衡解。例如，所有人报出的支付意愿都为零，这是一个典型的均衡解。

23.4 连续公共产品有效率的供给

现在假设公共物品的数量是连续的；为简单起见，仍考虑只有两个人的情形。若 $g_1 + g_2$ 表示这两个人对公共产品的出资之和，则公共产品的数量用 $G = f(g_1 + g_2)$ 表示，个人 i 的效用用 $u_i(f(g_1 + g_2), w_i - g_i)$ 表示。我们也可以将生产函数纳入效用函数，此时可将效用函数写为 $u_i(g_1 + g_2, w_i - g_i)$，其中 $u_i(G, x_i)$ 定义为 $U_i(f(G), x_i)$。将生产技术纳入效用函数不会导致一般性的损失，因为最终效用取决于这两个人对公共产品的出资总额。

我们知道，效率的一阶条件可通过效用函数的加权平均和的最大化问题获得：

$$\max_{g_1, g_2} a_1 u_1(g_1 + g_2, w_1 - g_1) + a_2 u_2(g_1 + g_2, w_2 - g_2).$$

g_1 和 g_2 的一阶条件可以写为

$$
\begin{aligned}
a_1\frac{\partial u_1(G,x_1)}{\partial G}+a_2\frac{u_2(G_2,x_2)}{\partial G}&=a_1\frac{\partial u_1(G,x_1)}{\partial x_1}\\
a_1\frac{\partial u_1(G,x_1)}{\partial G}+a_2\frac{u_2(G_2,x_2)}{\partial G}&=a_2\frac{\partial u_1(G,x_1)}{\partial x_2}.
\end{aligned}
\tag{23.3}
$$

由此可得$a_1\partial u_1/\partial x_1=a_2\partial u_2/\partial x_2$. 根据这个式子，并将（23.3）式的左边除以右边，可知

$$
\frac{\dfrac{\partial u_1(G,x_1)}{\partial G}}{\dfrac{\partial u_1(G,x_1)}{\partial x_1}}+\frac{\dfrac{\partial u_2(G,x_2)}{\partial G}}{\dfrac{\partial u_2(G,x_2)}{\partial x_2}}=1,
\tag{23.4}
$$

或

$$MRS_1+MRS_2=1.$$

在公共产品的数量为连续的情形下，效率的条件是边际支付意愿之和等于公共产品的边际提供成本，在这个例子中，边际成本为 1，这是由于公共产品的数量等于各人出资之和。

通常，满足效率条件的配置(G,x_1,x_2)是一系列的。一般来说，由于对公共产品的边际支付意愿取决于个人消费的数量，G 的有效率水平通常取决于x_1和x_2。

然而，在一种特别的情形下，即在拟线性效用情形下，公共产品有效率的提供数量和个人消费无关。为了看清这一点，假设效用函数的形式为$u_i(G)+x_i$. 则效率条件（23.4）可以写为$u_1'(G)+u_2'(G)=1$，这个式子决定了公共产品的数量是唯一的[一]。

例子：求解公共产品有效率的提供数量

假设效用函数为柯布-道格拉斯形式，即$u_i(G,x_i)=a_i\ln G+\ln x_i$. 在这种情形下，

$$MRS_i=\frac{a_ix_i}{G},$$

因此，效率条件为

$$\frac{a_1x_1}{G}+\frac{a_2x_2}{G}=1,$$

或，

$$G=a_1x_1+a_2x_2. \tag{23.5}$$

若私人产品禀赋总和为w，则我们有下列条件

$$x_1+x_2+G=w. \tag{23.6}$$

[一] 这个结论假设提供数量为正的公共产品是有效率的。但是在收入很低的情况下，这个结论可能并不成立。

（23.5）式和（23.6）式刻画了帕累托有效率配置的集合。

现在考虑效用函数为拟线性的情形，比如$u_i(G,x_i)=b_i\ln G+x_i$.一阶效率条件为

$$\frac{b_1}{G}+\frac{b_2}{G}=1,$$

或

$$G=b_1+b_2. \tag{23.7}$$

显然该配置是可行的，因此帕累托有效率的配置可用（23.6）式和（23.7）式描述。注意，在拟线性效用的条件下，公共产品有效率的提供数量是唯一的，而在一般情形下，公共产品有效率的提供数量有很多种。

23.5 连续公共产品的私人提供

假设每个人独立地决定他对公共产品的出资额。如果个人 1 认为个人 2 出资额为g_2，则个人 1 的效用最大化问题为

$$\max_{g_1} u_1(g_1+g_2,w_1-g_1)$$

$$\text{使得 } g_1\geq 0.$$

在这种情形下，约束条件$g_1\geq 0$是一个很自然的约束；它表明个人 1 可以自愿地增加公共物品的数量，但他不能单方面减少公共物品的数量。下面我们将会看到，这个不等式约束条件比较重要。

这个问题的库恩-塔克一阶条件为

$$\frac{\partial u_1(g_1+g_2,x_1)}{\partial G}-\frac{\partial u_1(g_1+g_2,x_1)}{\partial x_1}\leq 0, \tag{23.8}$$

其中，当$g_1>0$时，上式取等号。我们可以将上述条件写为

$$\frac{\dfrac{\partial u_1(g_1+g_2,x_1)}{\partial G}}{\dfrac{\partial u_1(g_1+g_2,x_1)}{\partial x_1}}\leq 1.$$

如果个人i对公共产品的出资额为正，他在公共产品和私人产品之间的边际替代率必定等于边际成本 1。如果他的边际替代率小于边际成本，则他不会愿意出资。

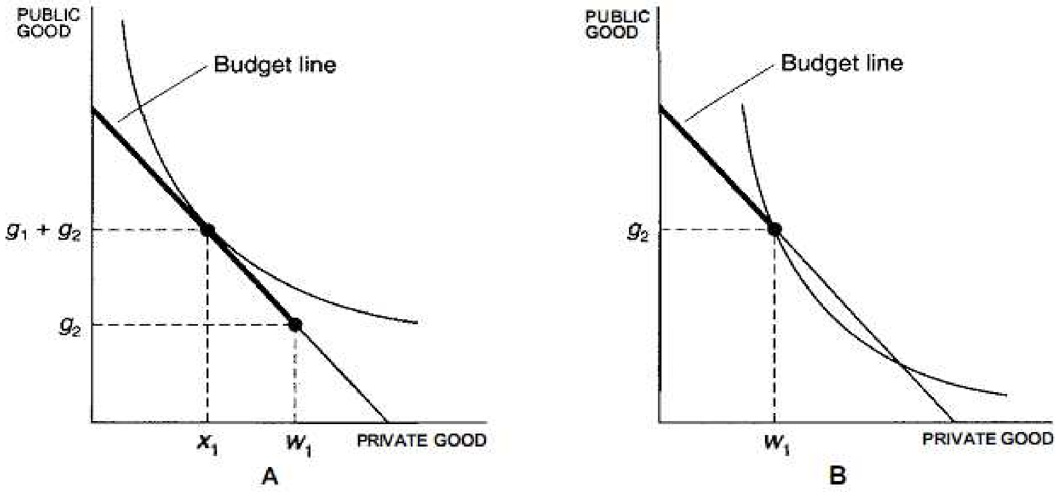


图 23.1：公共产品的私人提供。在 A 图中，个人 1 的出资额为正。在 B 图中，个人 1 发现最优选择是搭个人 2 的便车。

这个条件可用图 23.1 表示。图中，个人 1 的“禀赋”是点(w_1,g_2)，这是因为如果他不出资，那么他的私人消费为w_1，他消费的公共产品数量为g_2。“预算”线是经过原点斜率为-1的直线。预算线上的可行点是满足$g_1=w_1-x_1\geq 0$的那些点。我们画出了两种情形：一种情形是个人 1 愿意出资，另外一种情形是他希望搭便车。

这个博弈的一个纳什均衡解，是一组出资额(g_1^*,g_2^*)，该组出资额能使得在给定对方出资额的情形下，每个人的出资额都是最优的。因此（23.8）式必须对这两个人同时成立。我们可以该博弈的将纳什均衡的条件写为

$$\begin{aligned}&\frac{\dfrac{\partial u_1(G^*,x_1^*)}{\partial G^*}}{\dfrac{\partial u_1(G^*,x_1^*)}{\partial x_1^*}}\leq 1\\&\frac{\dfrac{\partial u_2(G^*,x_2^*)}{\partial G^*}}{\dfrac{\partial u_2(G^*,x_2^*)}{\partial x_2^*}}\leq 1.\end{aligned}\qquad（23.9）$$

如果公共产品的数量 G 为正，那么上面的两个不等式中至少有一个取等号。我们还可以继续分析，找出在什么样的条件下只有一个人出资，以及在什么样的条件下两个人都会出资等等。

然而，在这种情形下，我们还有一种更有用的表达纳什均衡的方法。为了做此事，我们需要解出个人i的反应函数（reaction function）。反应函数将一个人的意愿出资额表示为另一个人出资额的函数。

我们可以将个人 1 的最大化问题写为

$$\max_{g_1,x_1} u_1(g_1+g_2,x_1) \tag{23.10}$$

$$使得\ g_1+x_1=w_1;\quad g_1\geq 0.$$

利用$G=g_1+g_2$，我们可以将上述最大化问题写为

$$\max_{g_1,x_1} u_1(G,x_1) \tag{23.11}$$

$$使得\ G+x_1=w_1+g_2;\quad G\geq g_2.$$

将（23.11）式看仔细点。这个式子是说个人 1 在他自己的预算约束下选择最优的公共产品总量，约束条件表明他选择的总量必定至少和另外一个人提供的公共产品的数量一样大。预算约束表明个人 1 的消费总价值必定等于他的“禀赋”价值w_1+g_2。

（23.11）式表示的最大化问题和消费者一般的最大化问题类似，只有一点不同，即此处多了一个不等式约束。令$f_1(w)$表示个人 1 对公共产品的需求，它是他的财富的函数，暂时忽略不等式的约束。则满足（23.10）的公共产品的数量为

$$G=\max\{f_1(w_1+g_2),g_2\}.$$

将上式两端同时减去g_2可得

$$g_1=\max\{f_1(w_1+g_2)-g_2,0\}.$$

这个式子就是个人 1 的反应函数；它将个人 1 的最优出资额表示为另一个人出资额的函数。纳什均衡解是一组出资额(g_1^*,g_2^*)，使得

$$\begin{aligned}g_1^*&=\max\{f_1(w_1+g_2^*)-g_2^*,0\}\\ g_2^*&=\max\{f_2(w_2+g_1^*)-g_1^*,0\}.\end{aligned} \tag{23.12}$$

这个表达式通常比（23.9）那个表达式重要，因为这个表达式让我们知道需求函数f_1和f_2可能是什么样子。在下面的例子中我们会说明这一点。

当效用函数为拟线性时，有必要检查均衡条件的形式。在这种情形下，我们可将（23.9）式写为

$$\begin{aligned}u_1'(g_1^*+g_2^*)&\leq 1\\ u_2'(g_1^*+g_2^*)&\leq 1.\end{aligned}$$

注意，一般来说，上述两个不等式中只有一个不等式真正有约束力。假设个人 1 对公共产品的边际评价高于个人 2，即对于所有 G，都有$u_1'(G)>u_2'(G)$.那么，只有个人 1 会出资——个人 2 会永远搭便车。只有在这两个人对公共产品的边际偏好相同的情形下，他们才都会出资。

另外一种方法是，我们注意到当效用为拟线性时，对公共产品的需求将和收入无关，因此$f_i(w)\equiv\overline{g}_i$。则（23.12）变为

$$g_1^* = \max\{\bar{g}_1 - g_2^*, 0\}$$
$$g_2^* = \max\{\bar{g}_2 - g_1^*, 0\}.$$

由这两个式子可知，若 $\bar{g}_1 > \bar{g}_2$，则 $g_1^* = \bar{g}_1$ 和 $g_2^* = 0$。

例子：求公共产品提供的纳什均衡解

考虑前文的柯布-道格拉斯效用函数的例子。运用柯布-道格拉斯需求函数的标准形式，可得

$$f_i(w) = \frac{a_i}{1 + a_i} w.$$

由此可推知，（23.12）的解必定满足

$$g_1 = \max\{\frac{a_1}{1 + a_1}(w_1 + g_2) - g_2, 0\}$$
$$g_2^* = \max\{\frac{a_2}{1 + a_2}(w_2 + g_1) - g_2, 0\}. \qquad (23.13)$$

对于拟线性效用的例子，我们得到的一阶条件为

$$\frac{b_1}{G} \le 1$$
$$\frac{b_2}{G} \le 1.$$

因此，$G = \max\{b_1, b_2\}$。如果 $b_1 > b_2$，个人 1 对公共产品出全资，个人 2 搭便车。

23.6 投票

假设有一群人要对公共产品的提供数量进行投票表决。如果公共产品的当前数量为 G，则他们投票表决的是是增加还是减少公共产品的数量。如果多数人赞成增加（或减少）公共产品的数量，这个问题就解决了。**投票均衡**（voting equilibrium）是指某数量，该数量能使得不存在多数人偏好更多或更少的数量情形。

如果没有额外的限制条件，在投票模型中很可能不存在均衡解。例如，假设有三个人，A,B 和 C。假设公共产品的提供数量也有三种，1,2 或 3 单位。A 的偏好顺序为 1 好于 2,2 好于 3；B 的偏好顺序为 2 好于 3,3 好于 1；C 的偏好顺序为 3 好于 1,1 好于 2。在这种情形下，多数人偏好 1 胜于 2，多数人偏好 2 胜于 3，多数人偏好 3 胜于 1。因此，不管公共产品的提供数量为多少，多数人总想改变这个数量。这个例子就是著名的**投票悖论**（paradox of voting）。

然而，如果我们增加一些额外的限制，我们就能消除这个悖论。假设所有人都同意下列做法：如果多数人赞成增加公共产品的数量，那么由此增加的成本中，个人 i 的出资份额为 s_i。再假设所有人的效用函数都为拟线性的。如果公共产品的提供数量为 G，则个人 i 的效用为 $u_i(G)-s_iG$。因此，如果 $u_i'(G)>s_i$，他会投票赞成增加公共产品的数量。

我们说个人 i 拥有**单峰偏好**（single-peaked preferences），若 $u_i(G)-s_iG$ 的最大值只有一个。假设每个人的偏好都是单峰的，令 G_i 表示个人 i 的效用达到最大值的点。那么，我们可以断言，唯一的投票均衡是各个 G_i 的**中值**（median value）。为简单起见，假设每个人的 G_i 是不同的，而且投票人的人数为奇数。假设有 $n+1$ 个投票人，则**中值投票人**（median voter）是指除了这个人之外，有 $n/2$ 的人偏好更多的公共产品、有 $n/2$ 的人偏好更少的公共产品。如果个人 m 是中间投票人，则公共产品的投票均衡的数量 G 由下式给出

$$u_m'(G_v)=s_m.$$

这样的均衡称为**鲍恩**（Bowen）**均衡**。显然这是一个均衡，因为不存在多数人想减少或增加公共产品的情形。不难证明，这个均衡解是唯一的。

一个有趣的问题是将这个数量水平与帕累托有效率的水平进行比较。我们已经知道，公共产品帕累托有效率的水平满足

$$\sum_{i=1}^{n}u_i'(G_e)=1.$$

我们可以将这个式子写为

$$\frac{1}{n}\sum_{i=1}^{n}u_i'(G_e)=\frac{1}{n}.$$

这个式子的左端是“平均”效用函数的导数，右侧是平均成本份额。因此，公共产品有效率的水平由下列条件决定：**平均**支付意愿必定等于平均成本。而在投票均衡条件下，决定公共产品均衡数量的是支付意愿的**中值**（median）。如果中值消费者想要的公共产品数量，和平均消费者想要的公共产品的数量相等，那么投票决定的公共产品数量就是有效率的。然而，一般来说，投票决定的公共产品数量要么过多要么过少，这是因为在多数人投票机制中，最终结果取决于中值投票人的决定，而中值投票人想要的公共产品数量，与平均投票人相比，可能过多也可能过少。

例子：拟线性效用和投票

假设效用函数为 $b_i\ln G+x_i$，而且假设每个人必须平摊公共产品的费用，这样每个人所出的资金占总费用的 $1/n$。公共产品有效率的的数量为 $G_e=\sum_i b_i$。投票均衡数量是对中值投票人来说最佳的数量。令 b_m 表示中值投票人偏好的参数，我们有

$$\frac{b_m}{G_v} = \frac{1}{n},$$

或者 $G_v = nb_m$. 因此

$$G_e > G_v \text{ 当且仅当 } \frac{1}{n}\sum_i b_i > b_m.$$

也就是说，如果平均消费者对公共产品的评价高于中值消费者，那么公共产品有效率的数量就会大于由多数人投票决定的公共产品的数量。

23.7 林达尔（Lindahl）配置

假设我们试图用价格体系支撑公共产品有效率的配置。我们赋予每个消费者 i 按照价格 p_i 想“买”多少公共产品就买多少的权利。消费者 i 的最大化问题因此为

$$\max_{x_i, G} u_i(G, x_i)$$

$$\text{使得 } x_i + p_i G = w_i.$$

这个问题的一阶条件为

$$\frac{\frac{\partial u_i}{\partial G}}{\frac{\partial u_i}{\partial x_i}} = p_i .$$

公共产品的最优数量 G 是 p_i 和 w_i 的函数，这就是消费者对公共产品的需求函数，我们将其写为 $G_i(p_i, w_i)$.

是否存在一组价格使得消费者自然会选择有效率的公共产品数量？在标准凸性条件下，答案是肯定的。我们从前面的效率分析中知道，公共产品有效率的产量必定满足

$$\frac{\frac{\partial u_1(G^*, x_1^*)}{\partial G}}{\frac{\partial u_1(G^*, x_1^*)}{\partial x_1}} + \frac{\frac{\partial u_2(G^*, x_2^*)}{\partial G}}{\frac{\partial u_2(G^*, x_2^*)}{\partial x_2}} = 1.$$

因此选择

$$p_i^* = \frac{\frac{\partial u_i(G^*, x_i^*)}{\partial G}}{\frac{\partial u_i(G^*, x_i^*)}{\partial x_i}}$$

就能取得成功。这些价格——支持有效率公共产品配置的价格——称为**林达尔价格**。

我们也可以将这些价格解释为税率。如果公共产品的提供数量为 G，则个人 i 支付的税收为 $p_i G$ 。正因为此，有时人们把林达尔价格称为**林达尔税**。

23.8 需求显示机制

在本章前面几节，我们已经知道分权化的资源配置方式不适合于公共产品的配置。私人提供公共产品的方式，通常导致公共产品的数量小于有效率的数量。投票方式则可能导致过多或过少的公共产品。是否存在某种机制，它导致提供的公共产品数量不多不少刚刚好？

为了分析这个问题，我们回到离散公共产品模型。假设 G 要么为 0 要么为 1。令 r_i 表示个人 i 的保留价格，令 s_i 表示个人 i 出资份额。由于提供公共产品花费的成本为 c ，因此如果想要提供公共产品，则个人 i 必须出资 $s_i c$ 。令 $v_i = r_i - s_i c$ 表示个人 i 对公共产品的**净值**（net value）。根据我们前面的讨论，如果 $\sum_i v_i = \sum_i (r_i - s_i c r_i) > 0$ ，则提供公共产品就是有效率的。

我们可以使用的一种机制是让每个人报告自己的净值，然后将这些净值相加，如果净值之和大于或等于 0，则提供公共产品。这种方案的缺陷是，它无法让个人有动机显示自己真实的支付意愿。例如，如果个人 1 的净值大于零而且真实净值未必大，但他可能报出一个任意大的数值。由于他报告出的净值不影响他的实际出资额，但会影响公共产品是否能提供，因此，他可能报出一个非常大的净值。

我们如何诱导每个人**诚实地**显示自己对公共产品的真实评价？下面的方案就可以做到这一点：

格罗夫-克拉克（Groves-Clarke）机制

（1）每个人对公共产品报出一个价格 b_i 。 b_i 未必是自己对公共产品的真实评价。

（2）若 $\sum_i b_i \geq 0$，则提供公共产品；反之，若 $\sum_i b_i < 0$，则不提供。

（3）如果提供公共产品，那么每个人 i 会收到一笔单方面支付（sidepayments），数额等于其他人的报价之和 $\sum_{j \neq i} b_j$ 。（如果该数额为正，个人 i 接收它；如果该数额为负，则他必须支付该数额。）

下面我们证明每个人的最优选择是如实报出真实的评价。有 n 个人，每个人的真实评价为 v_i ，报出的评价为 b_i 。我们需要证明每个人的最优选择是报价 $b_i = v_i$ ，而不管其他人如何报价。也就是说，我们想证明说实话是**占优策略**（dominant strategy）。

$$\text{个人 } i \text{ 的收益} = \begin{cases} v_i + \sum_{j \neq i} b_j & \text{若} \mathrm{b}_i + \sum_{j \neq i} b_j \geq 0 \\ 0 & \text{若} \mathrm{b}_i + \sum_{j \neq i} b_j < 0. \end{cases}$$

假设 $v_i+\sum_{j\neq i}b_j>0$，则个人 i 若报价 $b_i=v_i$，可确保公共产品被提供。另一方面，假设 $v_i+\sum_{j\neq i}b_j<0$，则个人若报价 $b_i=v_i$，可确保公共产品不被提供。在上述两种情形下，个人 i 的最优选择都是说实话。不管其他人怎么报价，个人 i 没有激励说谎。实际上，在这种方式下，信息收集机制已经改变了，因此每个人面对的是社会决策问题而不是个人决策问题，也因此每个人都有激励如实报告自己的偏好。

不幸的是，上述偏好显示方案有个很大的缺陷。单方面支付的总额可能非常大：这个总额等于除了个人 i 之外的所有其他人的报价总和。因此，诱导个人说真话的代价非常大。

理想情形是，我们希望能找到某种机制使得单方面支付的总和为零。然而可以证明这样的愿望一般不可能实现。但是，我们可以设计一种机制使得单方面支付总和总是非正的。因此，个人可能需要“缴税”但永远不可能得到正的单方面支付。但是由于税收对于投票人来说是一种“浪费”，这种方式下的公共产品和私人商品的配置不可能是帕累托有效率的。然而，当且仅当提供公共产品是有效率时，公共产品才会被提供。

下面我们说说如何能做到这一点。基本的思想如下：我们要求每个人 i 还需要额外支付一笔钱，这笔钱仅取决于其他人的行为而且这些行为不会影响到个人 i 的激励。

令 b_{-i} 表示不包含个人 i 的所有其他人报价的向量，令 $h_i(b_{-i})$ 表示个人 i 额外支付的钱数。个人 i 的收益如下：

$$\text{个人 } i \text{ 的收益}=\begin{cases} v_i+\sum_{j\neq i}b_j-h_i(b_{-i}) & \text{若}\mathrm{b}_i+\sum_{j\neq i}b_j\geq 0 \\ -h_i(b_{-i}) & \text{若}\mathrm{b}_i+\sum_{j\neq i}b_j<0. \end{cases}$$

显然，这样的机制可以诱导每个人真实报告自己的偏好，原因和前面一样。如果我们能够明智地选择 h_i 函数，就能大幅度减少单方面支付的金额。对 h_i 函数的一种漂亮选择是选择如下形式的函数：

$$h_i(b_{-i})=\begin{cases} \sum_{j\neq i}b_j & \text{若}\sum_{j\neq i}b_j\geq 0. \\ 0 & \text{若}\sum_{j\neq i}b_j<0. \end{cases}$$

这样的选择产生了主角机制或关键人机制（pivotal mechanism），也称为克拉克税（Clarke tax）。个人 i 的收益为：

$$\text{个人 } i \text{ 的收益}=\begin{cases} v_i & \text{若}\sum_i \mathrm{b_i}\geq 0\text{且}\sum_{j\neq i}b_j\geq 0 \\ \mathrm{v_i}+\sum_{j\neq i}b_j & \text{若}\sum_i \mathrm{b_i}>0\text{且}\sum_{j\neq i}b_j<0 \\ -\sum_{j\neq i}b_j & \text{若}\sum_i \mathrm{b_i}<0\text{且}\sum_{j\neq i}b_j\geq 0 \\ 0 & \text{若}\sum_i \mathrm{b_i}<0\text{且}\sum_{j\neq i}b_j<0. \end{cases} \qquad (23.14)$$

注意个人 i 绝不会得到正的单方面支付；他可能需要缴税，但永远不可能得到补贴。在方案中加入单方面支付因素后，它的影响是只有当个人 i 改变了社会决策时，他才需要缴税。

例如，请看（23.14）式的第二行和第三行。个人 i 只有当他的报价将报价总和由正改变为负或者由负该为正时，才需要缴税。个人 i 缴纳的税额必须等于他的报价对其他人造成的损害（根据他们的报价计算）。个人 i 如果想改变公共产品的提供数量，那么他需要承担的代价等于他对其他人造成的损害。注意每个人都会发现使用这种决策过程对其自身是有利的，这是因为他缴纳的税收永远不会大于这个决策对他的价值。

23.9 连续公共产品的需求显示机制

现在假设我们关心的是连续公共产品的提供问题。如果公共产品的提供数量为 G，则消费者 i 的效用为

$$v_i(G) = u_i(G) - s_i G,$$

其中 $u_i(G)$ 是他对公共产品的（拟线性）效用函数， s_i 是他的费用份额。假设要求每个人 i 报告出的他自己的函数 $v_i(G)$ 。

令 $b_i(G)$ 表示他报告出的函数。政府宣布提供的公共产品数量为 G^* ，该数量使得个人报告函数的总和达到最大值。每个人 i 得到的单方面支付等于 $\sum_{i\neq j} b_j(G^*)$.

在这种机制下，如实报出自己的效用函数符合每个人的利益。为了看清这一点，只要注意到个人 i 的最大化问题为

$$v_i(G) + \sum_{i\neq j} b_j(G),$$

而政府的最大化问题为

$$b_i(G) + \sum_{i\neq j} b_j(G).$$

个人 i 若报告 $b_i(G) = v_i(G)$ ，则他能确保政府选择的公共产品提供数量为 G^* ，而这个数量恰好能使他的效用达到最大。

和离散情形一样，单方面支付的总额可能非常大。然而，正如前面指出的，如果合理选择单方面支付，那么这个总额可以大幅度降低。在这种情形下，最好的选择是选择单方面支付的数额为 $-\max_G \sum_{j\neq i} b_j(G)$ 。这样，个人 i 的净效用为

$$v_i(G) + \sum_{i\neq j} b_j(G) - \max_G \sum_{j\neq i} b_j(G).$$

注意，上式中的最后两项之和的符号必定为负。和前面一样，个人 i 缴纳的税额等于他对社会福利的改变额。

注释

公共产品的效率问题是由 Samuelson(1954)首先系统进行研究的。Bergstrom, Blume &

Varian(1986)广泛研究了公共产品的私人提供问题。Lindahl（1919）引入了 Lindahl 价格的概念。需求显示机制则是由 Clarke(1971)和 Groves(1973)引入的。

习题

23.1. 假设下面的博弈是两个人进行离散公共产品提供问题的解。每个人 i 的报价为 b_i。如果 $b_1+b_2\geq c$，则提供公共产品，每个人按照自己的报价 b_i 出资；若 $b_1+b_2<c$，则不提供公共产品，每个人都不需要出资。帕累托有效率的结果是这个博弈的一个均衡解吗？还有没有其他的均衡解？

23.2. 假设 u_1 和 u_2 对于 (x_i,G) 都是位似得。推导这两个人为公共产品出资水平的纳什均衡条件。

23.3. 现在假设两个人的财富不同，但是有着相同的柯布-道格拉斯型效用函数 $u_i(G,x_i)=G^{\alpha}x_i^{1-\alpha}$。这两个人的财富差值为多大时才能使得第 2 个人在均衡时的贡献为零？

23.4 假设有 n 个人，他们有着相同的柯布-道格拉斯效用函数 $u_i(G,x_i)=G^{\alpha}x_i^{1-\alpha}$。财富总额 w 要在 $k\leq n$ 个人中分配。计算应该提供的公共产品的数量。当 k 增大时，公共产品的数量是如何变化的？

23.5 克拉克税能导致帕累托有效率的配置吗？克拉克税能导致帕累托有效率的公共产品数量吗？

23.6 南海中有个土著部落称为 Grads，他们的消费品只有椰子。椰子对他们来说有两种用途：一是当饭吃；二是在公共宗教祭祀时烧掉椰子。（他们相信这种祭祀对他们有帮助。）

假设部落中的每个人 i 的拥有椰子的初始禀赋为 $w_i>0$，令 $x_i\geq 0$ 表示此人把椰子当饭吃的数量，$g_i\geq 0$ 表示他上交给部落供祭祀用的数量。祭祀用的椰子总量为 $\sum_{i=1}^n g_i$。每个人 i 的效用函数为

$$u_i(G,x_i)=x_i+a_i\ln G$$

其中 $a_i>1$。

（a）每个人 i 在确定自己的祭礼（即上交的椰子数量）时，都假设其他人的祭礼数量是固定不变的，然后他在这个基础上再确定自己的祭礼数量。令

$$G_{-i}=\sum_{j\neq i}g_j$$

表示除了 i 之外的所有其他人的祭礼数量。请写出决定 i 祭礼数量的效用最大化问题。

（b）由于所有人的祭礼之和为 $G=g_i+G_{-i}$，求公共物品（祭礼）的均衡数量。（提示：不是每个人的祭礼数量都为正）

（c）在这个问题中谁将搭便车？

（d）在这个经济体中，帕累托有效率的公共产品的提供数量为多少？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 **3** 版）

完美中文翻译版

第 24 章：外部性

曹乾 译

(东南大学 caoqianseu@163.com)

24 外部性

当一个人的行为直接影响另外一个人的处境时，我们说存在着**外部性**（externality）。在**消费外部性**（consumption externality）中，某个消费者的效用直接受到其他消费者行为的影响。例如某个消费者可能受到其他人下列行为的影响：抽烟、喝酒、大声播放音乐等等。消费者也有可能受到企业的负影响，比如企业制造污染或噪音。

在**生产外部性**（production externality）中，某个企业的生产集受到其他企业行为的影响。例如，钢厂制造的烟尘直接影响洗衣店晾晒的干净衣物，养蜂人直接影响临近苹果园的产量等。

在本章我们进行外部性的经济学分析。我们发现，当存在外部性时一般市场均衡是无效率的。自然地，我们将分析各种可能导致帕累托有效率结果的资源配置的其他方法。

当存在着外部性时，福利经济学第一定理不再成立。原因在于存在着某些人们关注但却没有价格的东西。在这种情形下，如果要想实现有效率的配置就要确保当事人面对他们行为的正确代价。

24.1 生产外部性的一个例子

假设有两个企业。企业 1 生产的产量为 x，在竞争市场上销售。然而，这些产量也为企业 2 造成了成本 $e(x)$。例如，假设生产技术为 x 单位的产品伴随着 x 单位的污染，这些污染危害了企业 2。

令 p 表示产品的价格，两个企业的利润分别为

$$\pi_1 = \max_x \; px - c(x)$$
$$\pi_2 = -e(x).$$

假设两个企业的成本函数都是递增且凸的。（企业 2 可能会从自身的生产活动中获利，但是为简单起见，我们不考虑它自身的生产。）

均衡产量 x 由 $p = c'(x_q)$。然而，从社会的观点看，这个产量太大了。企业 1 考虑的只是私人成本，即它生产活动给自己造成的成本，但没有考虑社会成本，即私人成本加上它对其他企业造成的成本。

为了确定有效率的产量，我们可以这么想：如果这两个企业合并，情形将是什么样的？企业合并使得外部性内部化。在这种情形下，合并后的企业将最大化总利润

$$\pi = \max_x \; px - c(x) - e(x),$$

这个问题的一阶条件为

$$p = c'(x_e) + e'(x_e) \tag{24.1}$$

产量 x_e 是有效率的产量；它由价格等于边际社会成本这个条件确定。

24.2 外部性问题的解决方法

经济学家提出过若干种解决外部性的无效率问题的方法。

庇古税

这种方法认为，企业 1 面对的是自身行为的错误定价，因此可以征收矫正税，从而达到有效率的资源配置。这类矫正税称为庇古税（Pigovian taxes）。

例如，假设政府对企业 1 每单位产品征收 t 元税收，那么企业 1 的利润最大化问题的一阶条件变为

$$p = c'(x) + t.$$

由于我们假设成本函数为凸，我们可以令 $t = e'(x_e)$，这会引导企业 1 选择产量 $x = x_e$，正如（24.1）确定的那样。即使成本函数不是凸的，我们只要对企业征收非线性税 $e(x)$，也能引导它将外部性的成本内部化。

这种方法的问题是，它要求征税当局知道外部性成本函数 $e(x)$。但是如果征税当局知道这个成本函数，那么它可以直接告诉企业 1 生产多少产量就够了，何必再多此一举征税呢。

缺少外部性的市场

这种方法认为，问题在于企业 2 关注企业 1 制造的污染数量但却无可奈何。如果设置污染这种产品的市场，让企业 2 能表达它对污染的需求，即污染每减少一单位企业 2 愿意支付 r 元，从而借助市场机制达到有效率的配置。

在我们的模型中，x 单位的产品必然伴随着 x 单位的污染。如果污染的市场价格为 r，那么企业 2 可以决定出售多少单位的污染（x_1），企业 2 可以决定购买多少单位的污染（x_2）。这两个企业各自的利润最大化问题变为

$$\pi_1 = \max_{x_1}\ px_1 + rx_1 - c(x_1)$$
$$\pi_2 = \max_{x_2} - rx_2 - e(x).$$

一阶条件为

$$\begin{aligned} p + r &= c'(x_1) \\ -r &= e'(x_2). \end{aligned}$$

当污染的供给等于污染的需求时，我们有 $x_1 = x_2$，此时这些一阶条件等价于（24.1）中的条件。注意，污染的均衡价格 r 是负的。这是当然的，因为污染是种厌恶品而不是人们喜欢的商品。

更一般地，假设污染和产品不是 1:1 生产出的。如果企业 1 生产 x 单位产品和 y 单位污染，那么企业 1 支付的成本为 $c(x, y)$。假设 y 从零增加后企业 1 生产 x 的成本会降低；这个假设是自然的，因为如果污染增加企业 1 自身的生产成本反而增加，那么企业 1 自己就会想法解决污染问题。

在缺乏控制问题的任何机制时，企业 1 的利润最大化问题为

$$\max_{x,y} \; px - c(x, y)$$

一阶条件为

$$\begin{aligned} p &= \frac{\partial c(x, y)}{\partial x} \\ 0 &= \frac{\partial c(x, y)}{\partial y}. \end{aligned}$$

企业 1 将使得污染的价格等于自己的边际成本。在这种情形下，污染的价格为零，因此企业 1 制造的污染数量会一直增加直至生产成本达到最小时为止。

现在我们设置污染这种产品的市场。令每单位污染的价格为 r，企业 1 和 2 的污染供给和污染需求分别为 y_1 和 y_2。最大化问题为

$$\begin{aligned} \pi_1 &= \max_{x, y_1} \; px + ry_1 - c(x, y_1) \\ \pi_2 &= \max_{y_2} \; -ry_2 - e(y_2). \end{aligned}$$

一阶条件为

$$\begin{aligned} p &= \frac{\partial c(x, y_1)}{\partial x} \\ r &= \frac{\partial c(x, y_1)}{\partial y_1} \\ -r &= \frac{\partial e(y_2)}{\partial y_2}. \end{aligned}$$

令污染供给和污染需求相等，即 $y_1 = y_2$，我们就得到了 x 和 y_1, y_2 的有效率水平的一阶条件。

这种方法的问题是污染市场上的买方和卖方数量太少，在我们所举的这个例子中只有两个企业。没有什么特别的原因能让我们相信这样的市场是竞争性的市场。

产权

这种方法认为，基本的问题在于产权没有帮助资源配置实现充分效率。在上面这个例子中，如果两种生产都是由一个企业负责，不会存在问题。然而，我们将看到市场信号将激励当事人确定产权的有效率方式。

如果一个企业的外部性对另外一个企业的运行造成了不利影响，那么一个企业买下另外一个企业总是有利可图的。协调两个企业的生产行为产生的利润，总是大于企业独自生产时的利润之和。因此，一个企业愿意按照另外一个企业的市场价值（存在外部性时的市值）将其买下，因为当外部性在合理调整之后该企业的市值会超过当前的市值。这个论断表明市场机制本身提供了信号——调整产权将外部性内部化的信号。

我们在证明福利经济学第一定理时（详见第 18 章），已经说明了上述这样的思想。那里的论断表明如果某个配置不是帕累托有效率的，那么必定存在着增加总利润的方法。仔细审查第一定理的证明可知，该定理的全部必要条件是消费者关注的所有商品都要有定价，或者等价地，消费者的偏好仅取决于自己的消费束。在存在着生产外部性的情形下，我们对于第一定理的证明过程（详见第 18 章）也是成立的，在那里我们证明了帕累托有效率配置处的总利润大于初始配置的纵利润。如果存在着生产外部性，那么第 18 章的这个论证过程表明存在着可以使总利润增加的其他生产方案——因此，某个企业有激励买下另外一个企业，协调它们二者的生产，将外部性内部化。

在本质上，企业将所有相关的生产外部性内部化之前会一直增长壮大。这适用于某些类型的外部性，但不是所有外部性。例如，它对于消费外部性或者公共产品引起的外部性几乎是无能为力的。

24.3 补偿机制

我们在前面指出过，庇古税一般不足以解决外部性问题，原因在于信息不对称：征税当局一般不知道外部性造成的成本有多大。然而，导致外部性的当事人可能知道。如果是这样，将外部性内部外就好办了。方案如下：这种方案要设立外部性这种产品的市场，它的机理在于激励企业正确显示它为别人造成的成本。方法如下。

宣布阶段。企业 $i=1,2$ 分别报出各自的庇古税税率 t_i，不必要求这样的税率是有效率的税率水平。

选择阶段。如果企业 1 生产 x 单位的产品，它必须缴纳 $t_2 x$ 元的税，企业 2 得到 $t_1 x$ 元的补偿。除此之外，每个企业还需要支付罚金，罚金的大小取决于它们报出的税率之差。

罚金的具体形式和我们的目的没有任何关系；只要能做到当 $t_1=t_2$ 时罚金为零且 $t_1 \neq t_2$ 是罚金为正即可。为了便于说明，我们将罚金表达为 $(t_1-t_2)^2$ 这种平方形式。在这种情形下，

企业 1 和 2 的最终收益为

$$\pi_1 = \max_x \quad px - c(x) - t_2 x - (t_1 - t_2)^2$$
$$\pi_2 = t_1 x - e(x) - (t_1 - t_2)^2.$$

我们想证明：该博弈的均衡结果涉及外部性生产的有效率水平。为了做此事，我们先思考一下这个博弈的均衡是由什么构成的。由于该博弈有两个阶段，自然需要一个**子博弈完美均衡**（subgame perfect equilibrium）——也就是说，在这个均衡中，每个企业考虑了第一阶段选择后果对第二段结果的影响。详细内容请参考第 15 章。

和往常一样，在解这样的博弈时，我们先分析第二阶段。考虑企业在第二阶段的产量选择。企业 1 选择的 x 满足条件

$$p = c'(x) + t_2 \tag{24.2}$$

t_2 的每个选择，都对应着每个最优选择 $x(t_2)$。如果 $c''(x) > 0$，那么容易证明 $x'(t_2) < 0$。

在第一阶段，每个企业选择选择税率以使得各自的利润最大化。对于企业 1 来说，作出选择很简单：如果企业 2 选择 t_2，那么企业 1 希望选择

$$t_1 = t_2. \tag{24.3}$$

为了验证这一点，只要将企业 1 的利润函数关于 t_1 微分即可。

企业 2 的选择有些棘手，因为它必须认识到它的选择 t_2 通过函数 $x(t_2)$ 影响企业 1 的产量。对企业 2 的利润函数微分，在微分时要考虑到上述影响，我们有

$$\pi_2'(t_2) = (t_1 - e'(x))x'(t_2) - 2(t_2 - t_1) = 0. \tag{24.4}$$

将（24.3）式代入（24.4）式可得 $t_2 = e'(x)$，然后将其代入（24.2）可得

$$p = c'(x) + e'(x),$$

这个式子就是效率条件。

这种方法的机制在于为两个企业设置相互对立的激励。由（24.3）可以看出，企业 1 总是有激励选择与企业 2 宣称的税率相等的税率。但是考虑一下企业 2 的激励。如果企业 2 知道 1 会为 2 提出较大的赔偿率（税率）t_1，那么他希望政府对企业 1 征收的税收越少越好，因为这样企业 1 会尽可能地多生产。另一方面，如果企业 2 认为企业 1 提出的赔偿率较低，那么企业 2 希望政府对企业 1 征收的税收越多越好。企业 2 对于企业 1 产量水平无差异的点，正是企业 2 的外部性成本在边际上刚好被补偿的那个点。

24.4 外部性存在时的效率条件

下面我们推导在存在外部性时的效率条件。假设有两种商品，商品 x 和商品 y；再假

设有两个人。每个人关新对方消费的商品 x 的数量，但没人关心对方对商品 y 的消费。商品 x 和 y 的初始数量分别为 $\bar{x}$ 和 $\bar{y}$ 。

根据第 17 章的内容可知，帕累托有效率的配置能在资源既定约束下使得效用之和最大

$$\max_{x_i, y_i} \ au(x_1, x_2, y_1) + au(x_1, x_2, y_2)$$

$$\text{使得 } x_1 + x_2 = \bar{x}\ ,\ y_1 + y_2 = \bar{y}.$$

一阶条件为

$$a_1 \frac{\partial u_1}{\partial x_1} + a_2 \frac{\partial u_2}{\partial x_1} = \lambda$$

$$a_1 \frac{\partial u_1}{\partial x_2} + a_2 \frac{\partial u_2}{\partial x_2} = \lambda$$

$$a_1 \frac{\partial u_1}{\partial y_1} = \mu$$

$$a_2 \frac{\partial u_2}{\partial y_2} = \mu.$$

经过整理，这些条件可以写为

$$\frac{\frac{\partial u_1}{\partial x_1}}{\frac{\partial u_1}{\partial y_1}} + \frac{\frac{\partial u_2}{\partial x_1}}{\frac{\partial u_2}{\partial y_1}} = \frac{\lambda}{u}$$

$$\frac{\frac{\partial u_1}{\partial x_2}}{\frac{\partial u_1}{\partial y_1}} + \frac{\frac{\partial u_2}{\partial x_2}}{\frac{\partial u_2}{\partial y_1}} = \frac{\lambda}{u}.$$

效率条件为边际替代率之和等于一个常数。当确定消费者 1 增加商品 1 的消费是否值得时，我们必须不仅考虑他愿意为额外一单位商品 1 支付多少钱，还必须考虑消费者 2 愿意支付多少钱。这些条件在本质上和公共产品的效率条件是相同的。

从这些条件可以清楚地看到如何将外部性内部化。只要将 x_1 和 x_2 看成不同的商品即可。x_1 的价格是 $p_1 = \partial u_2 / \partial x_1$， x_2 的价格是 $p_2 = \partial u_1 / \partial x_2$ 。如果每个消费者面对他的行为的合适价格，市场均衡将会导致有效率的产出水平。

注释

Pigou(1920)和 Coase(1960)的文献是外部性的经典文献。补偿机制的进一步分析请参见 Varian(1989b)。

习题

24.1 假设两个消费者正在考虑应将自己的车开得多快。消费者 i 选择的车速为 x_i，他因此得到的效用为 $u_i(x_i)$；我们假设 $u_i'(x_i) > 0$。然而，车开得越快，这两个人越有可能相撞。令 $p(x_1, x_2)$ 表示相撞的概率，假设这个函数关于 x_1 和 x_2 都是递增的，令 $c_i > 0$ 表示车祸给消费者 i 造成的成本。假设每个消费者的效用关于钱数是线性的。

（a）证明：从社会的观点看来，每个消费者都有激励开得过快。

（b）假设若发生车祸消费者 i 需要缴纳罚款 t_i，为了将外部性内部化，t_i 应为多大？

（c）如果罚款额是最优的，每个消费者支付的总成本（包括罚款）是多少？请将它们与车祸的总成本进行比较。

（d）现在假设消费者 i 只有在没有车祸时才能得到效用 $u_i(x)$，求这种情形下的最优罚款额。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 25 章：信息

曹乾　译

(东南大学　caoqianseu@163.com)

25 信息

在过去的几十年里，经济理论发展最快的一个领域就是信息经济学。在本章，我们将分析信息经济学的一些基本议题。

我们将要分析的绝大多数内容都涉及信息不对称的情形，即某个人知道某事情而另外一个人不知道。例如，雇员比雇主更了解自己能够生产多少产品，再比如生产厂家比潜在的消费者更了解自己生产的产品质量。

然而，通过仔细观察雇员的行为，雇主也可能推断出员工生产能力的部分信息。类似地，消费者可以根据某企业的产品是怎么销售的来判断其质量。好的员工希望雇主能知道他的好，也有可能希望雇主不知道，因为这取决于雇主怎么支付他们工资。善于生产高质量产品的员工一般希望雇主知道这一点，但只能生产低质量产品的员工却希望能打扮成高质量的。因此，信息不对称情形下的行为研究必然涉及当事人之间的策略互动。

25.1 委托人-代理人问题

很多激励问题都可以使用下列框架进行建模。委托人希望诱使代理人从事某种活动，这种活动对于代理人来说当然是有代价的。委托人可能不能够直接观察到代理人的行为，因此他观察产量 x，产量自然是由代理人的行为决定的（至少部分决定）。委托人的问题是设计有激励的报酬支付方案 $s(x)$，从而诱导代理人能代表委托人的利益从事最佳活动。

委托人-代理人问题的最简单例子是经理和员工之间的例子。经理希望员工能尽最大的努力来生产尽可能多的产品，而员工在努力程度和激励方案一定的情形下会理性地进行能使自己效用最大的选择。

零售企业和消费者之间的例子就不那么明显了。企业希望消费者购买它的产品——这对于消费者来说是有代价的活动。企业希望按照每个消费者的保留价格来要价——消费者最高支付意愿。企业不能直接观测到消费者的保留价格，但是它可以观测到不同偏好的消费者在不同价格下购买的数量。该企业的问题是设计一种定价方案使得自己的利润最大。这是个垄断企业在实施价格歧视策略时面对的问题，请参考第 14 章。

我们将这类问题称为委托人-代理人问题。在下面各节，我们将分析经理-员工问题，但不难将这种情形推广到非线性定价等其他情形。

令 x 表示委托人收到的产量，令 a 和 b 表示代理人可能从他的可行行为集 A 中选择出的行为。在下面的内容中若假设可行的行为只有两种，论证起来将比较方便，但我们暂时不需要使用这个假设。假设最初没有不确定性，因此结果完全由代理人的行为所决定，我们将这种关系表示为 $x = x(a)$。令 $c(a)$ 表示行为 a 的函数，$s(x)$ 表示委托人支付给代理人的激励性报酬。

委托人的效用函数为 $x-s(x)$，即产出减去他支付给代理人的激励性报酬；代理人的效用函数为 $s(x)-c(a)$，即激励性报酬减去他的行为的成本。委托人希望选择函数 $s(\cdot)$，使得 $s(\cdot)$ 是下列最大化问题的解：在代理人最优行为的约束条件下，委托人的效用最大。

涉及代理人的约束条件通常有两类：第一类约束是代理人可能还有其他的工作选择，这样的工作能使得他得到保留效用水平（reservation level of utility），因此委托人若想确保代理人能够愿意为他效劳，就必须确保给与代理人的报酬不能低于上述保留水平。我们将这个条件称为参与约束（participation constraint）。（这个条件有时称为个人理性约束。）

第二类约束是激励相容（incentive compatibility）约束：给定委托人选择的激励方案，代理人选择自己的最优行为。委托人不能直接选择代理人的行为：他能做到的仅是通过选择激励方案影响代理人的行为。

我们仅关注两类委托人-代理人问题情形。第一类情形是委托人只有一个，他是垄断者：委托人设定激励方案，对于该方案来说，只要它大于代理人的保留效用水平，代理人就愿意接受。在这类情形中，我们希望确定激励方案的性质使得委托人的效用最大。第二类情形是有很多竞争性的委托人，每个委托人设定自己的激励方案。在这类情形中，我们希望确定均衡激励方案体系的性质。

在第一类情形，即委托人是垄断者的情形中，代理人的保留效用水平是外生的：它通常是与委托人提出的激励方案无关的活动的效用。在第二类情形，即委托人是竞争的情形中，代理人的保留效用水平是内生的：它是与其他委托人提供的激励方案相关的效用。类似地，在垄断情形下，目标是使得利润最大化。而在竞争情形下，我们通常假设利润在竞争中消散。因此，这种情形下，零利润就成为了一个重要的均衡条件。

25.2 完全信息：垄断情形的解

我们首先分析一个简单的例子，在这个例子中，委托人完全了解代理人的成本和行为。在这种情形下，委托人的目标就是确定他希望代理人选择什么样的行为，并且设计能够诱导代理人选择该行为的激励方案。由于委托人只有一个，我们将这种情形称为垄断情形[一]。

令 a 表示代理人可以采取的各种行为，令代理人的产出为他的行为的函数 $x(a)$。令 b 表示委托人想要诱导代理人选择的行为。（可将 b 理解为对于委托人来说，代理人“最好的”的那个行为；将 a 理解为代理人的“可以选择的”那些行为。）

最优激励方案 $s(\cdot)$ 设计问题可以写为

$$\max_{b,s(\cdot)} \quad x(b)-s(x(b))$$

$$\text{使得：}\ s(x(b))-c(b)\geq \overline{u}\,, \qquad (25.1)$$

[一] 我们也可以将这种情形称为买方垄断情形，由于我们分析的是买方只有一个而不是卖方只有一个的情形，但我习惯于使用一般意义上的垄断（包含买方垄断和卖方垄断），因为根据上下文读者自然明白说得是哪种情形。

$$s(x(b))-c(b)\geq s(x(a))-c(a)\quad \text{对于 A 中的所有}\ a\ \text{成立。}\qquad (25.2)$$

条件（25.1）是参与约束，它的意思是说，代理人得到的报酬必须不能小于他的保留效用水平，否则他可以选择不参与；条件（25.2）是激励相容约束，它是指代理人会发现选择 b 是最优的，而这正好是委托人想要的。注意，在这种激励方案中，委托人选择的是代理人的行为 b，虽然他不是直接选择而是诱导代理人自行选择的。委托人面对的约束是，确保代理人想采取的行为正好是委托人希望的那个行为。

尽管这个最大化问题看上去有些特别，但容易看到它的解是平凡的。下面我们就说明这一点。暂时不用考虑激励相容约束，而将注意力放在目标函数和参与约束，可以看到：对于任何 x，委托人希望 $s(x)$ 尽可能地小。由（25.1）中的参与约束可知，这意味着 $s(x(b))$ 应该等于 $\bar{u}+c(b)$；也就是说，委托人支付给代理人的报酬扣除代理人的成本之后，剩下的只是代理人的保留效用水平。

因此，从委托人的角度看，最优行为是使得 $x(b)-\bar{u}-c(b)$ 最大的那个行为。将这个行为称为 b^*，相应的产出水平为 $x^*=x(b^*)$。我们的问题是：委托人能否设定某个激励方案 $s(x)$ 使得 b^* 称为代理人的最优选择？这容易做到：委托人只要选取能满足 $s(x^*)-c(b^*)\geq s(x(a))-c(a)$ 的任何函数 $s(x)$ 即可。例如，令

$$s(x^*)=\begin{cases}\bar{u}+c(b^*) & x=x(b^*)\\ -\infty & x\neq x(b^*)\end{cases}$$

这个激励方案称为**目标产出方案**（target output scheme）：委托人为代理人设定既定的目标产出 x^*，代理人若实现该目标就可以得到与他的保留价格相等的报酬，否则就会受到狠狠的惩罚。（实际上，惩罚性报酬只要小于代理人能达到设定的目标而得到的报酬，就可实现这一目的。）

解决激励问题的方案有很多，我们上面给出的只是其中一种。下面再举一个例子。在这个例子中，我们选择的是**线性激励报酬**（linear incentive payment）$s(x(a))=x(a)-F$。在这种情形下，代理人向委托人支付一笔固定租金 F 后就可以得到他自己的全部产出。这种方案可行，因为代理人有激励选择使 $x(a)-c(a)$ 最大的那个行为。委托人对于 F 的选取要能使得代理人正好满足参与约束，即 $F=x(b^*)-c(b^*)-\bar{u}$。在这种情形下，代理人是产出的**剩余索取者**（residual claimant）。一旦代理人向委托人支付了租金 F，代理人就可以得到所有剩下的利润。

关于完全信息情形下委托代理问题的解，有几点需要注意。首先，激励相容约束实际上不是“束紧的”（binding）。一旦选定了最优产出水平，总可以选择某个激励方案使得该产出水平是代理人的最优选择。其次，由于激励相容约束不是束紧的，帕累托有效率的产量水平总会被生产出来。也就是说，不存在委托人和代理人都偏爱的另外一种产量水平。这一结论可从下列事实推导出来——无激励约束情形下的最大化问题是帕累托最优化的标准形式：在维持其他人效用不变的情形下，最大化其中一人的效用。

这些激励方案存在的问题是它们对信息的轻微扰动非常敏感。例如，假设投入和产出之间的关系不是完全确定的。也许系统中存在着某些“噪音”，从而低产量是由运气差而不是不努力导致的。在这种情形下，我们在上面描述的激励方案类型可能不再合适。如果代理人在完成目标产量水平时只能得到工资，那么他的期望效用——随机产量的均值——可能低于他的保留效用水平。因此他可能拒绝参与。

为了满足参与约束，委托人提供给代理人的报酬支付方案必须能使代理人得到他的保留效用水平。这样的方案通常涉及若干个产量水平上的正的报酬，因为这些不同的产量水平可能都与目标努力水平相一致。这种问题称为**隐藏行动**（hidden action）的激励问题，这是因为委托人通常不能完全观察到代理人的行动。

我们感兴趣的第二种不完全信息是，委托人不能完全观测到代理人的目标函数。代理人的种类可能有很多种，即他们的效用函数或成本函数是不同的。委托人设计的激励方案，在平均意义上，必须能够很好地应付任何类型的代理人。这种激励问题称为**隐藏信息**（hidden information）问题，因为问题出在代理人的种类信息对于委托人来说是隐藏的。下面我们将分析这两种激励问题。

25.3 完全信息：竞争解

在讨论上述两种激励问题之前，我们先考察竞争性环境中的完全信息的委托代理问题。正如上面指出的，完成模型的一种方法是增加一个条件：竞争迫使利润为零。

为了便于说明，假设有一组工厂老板和一组相同的工人。每个工厂老板设定一个激励方案，试图将工人吸引到自己的工厂。为了吸引工人，工厂老板们必须彼此竞争；另外一方面，为了得到工作，工人们也必须彼此竞争。

一个既定工厂老板面临的问题和垄断情形下的问题是一样的：他知道诱使各种努力水平所需的成本，也知道为了吸引工人到他自己工厂而花费的成本，从而选择合适的组合来使得收入与成本之差最大化。

我们已经看到，在这种情形下，一种最优的激励方案为：报酬是产量的线性函数，即 $s(x)=x-F$ 。在垄断模型中， F 由参与约束决定

$$x(b^*)-F-c(b^*)=\bar{u}\ ,$$

其中 $\bar{u}$ 是个外生变量，是从其他活动得到的效用水平。

在竞争模型中，这通常不合适。在竞争性框架下，确定 F 的方法是假设参与约束不是束紧的，但是行业中的竞争迫使利润为零。在这种情形下， F 由下列条件决定

$$x-(x-F)=0,$$

这意味着 $F=0$ 。工人们攫取了所有边际产品，而且“垄断租金”被压低为零。

均衡租金为零这个事实，是由规模报酬不变技术人为造成的结果。如果工厂老板的固定成本为 K，那么均衡条件要求 F=K。

从正式的观点来看，垄断解和竞争解的区别在于租金 F 是如何决定的。在垄断模型中，正是这个租金数额使得工人在为委托人工作与从事其他活动之间无差异。在竞争模型中，租金是由利润为零这个条件所决定。

25.4 隐藏行动：垄断解

在本节，我们将考察一个简单的委托代理关系模型，在这个模型中行动不可被直接观测到。为便于分析，我们将作出一些假设。特别地，我们将假设可能的产量水平 $(x_1,...,x_n)$ 为有限个。代理人的行动选择有两种：a 或 b，他的行动将影响各种产量水平出现的概率。因此，我们令 π_{ia} 表示代理人选择行动 a 时产量水平 x 能被观测到的概率，π_{ib} 的意思类似。令 $s_i = s(x_i)$ 表示如果委托人观测到产量水平 x 而支付给代理的人报酬。于是，代理人选择比如行动 b 时委托人的期望利润为

$$\sum_{i=1}^{n}(x_i - s_i)\pi_{ib} \qquad (25.3)$$

至于代理人，我们假设他是厌恶风险的，而且追求某个关于报酬的冯.诺依曼-摩根斯坦效用函数 $u(s_i)$ 的最大化。假设代理人的行动成本 c_a 线性地进入他的效用函数。因此，代理人会选择行动 b 如果

$$\sum_{i=1}^{n}u(s_i)\pi_{ib} - c_b \ge \sum_{i=1}^{n}u(s_i)\pi_{ia} - c_a \qquad (25.4)$$

他将选择行动 a，如果在上面不等式中出现的是≤。这是激励相容约束。

我们假设代理人还有一个行动选择，即他可以不参与委托人的生产。假设代理人不参与，他得到的效用为 $\bar{u}$。因此，代理人通过参与委托人生产而得到的期望效用必定满足：

$$\sum_{i=1}^{n}u(s_i)\pi_{ib} - c_b \ge \bar{u} \qquad (25.5)$$

这是参与约束。

委托人希望在约束条件（25.4）和（25.5）之下使得（25.3）最大化。最大化发生在行动 b 和报酬 (s_i) 上。注意，在这个问题中，委托人和代理人都试图使自己的收益最大。代理人将会选择行动 b，在委托人设定激励系统 (s_i) 情形下，这是代理人的最优选择。理解了这一点之后，委托人希望他提供的激励报酬对于委托人是最优的。因此，委托人在制定激励报酬时必须将代理人随后的行动作为一个约束条件。在本质上，委托人是在考虑相关成本的情形下，为代理人选择委托人想要的行动，也就是说，他设计的激励报酬必须能使得委托人想要代理人选择的行动，也正好是代理人自身想要选择的行动。

委托人可以观测到代理人的行动

在上一节讨论的完全信息问题中，报酬方案是建立在行动还是产量上是无关紧要的。这是因为行动和产出之间是一一对应的关系。在现在的不完全信息的问题中，这种区分非常重要。如果报酬可以建立在行动之上，那么委托人就有可能实施最优（first-best）的激励方案，即使产出是随机的。在这种情形下，委托人全部的工作就是：首先确定代理人各种可能行动带给委托人的（期望）利润；然后诱导代理人选择使得委托人期望利润最大的行动。

为了用数学表示，假设委托人支付的报酬是委托人选择的行动（而不是产量）的函数。于是，代理人得到的报酬为 $s(b)$ 。注意，这个报酬是确定性的，因此，代理人的效用为 $\sum_{i=1}^{n} u(s_i)\pi_{ib} - c_b = u(s(b)) - c_b$ 。于是，上面描述的激励问题简化为

$$\begin{aligned}
&\max_{s(b),b} \sum_{i=1}^{n} x_i \pi_{ib} - s(b) \\
&\text{s.t.} \quad u(s(b)) - c_b \geq \bar{u} \\
&\qquad\quad u(s(b)) - c_b \geq u(s(a)) - c_a.
\end{aligned}$$

这与前面考察的完全信息问题是一样的：激励相容约束是不重要的。

委托代理问题的有趣版本只有在下列情形下才会发生：代理人的行动是隐藏的，因此激励报酬只能建立在产量水平上。在这种情形下，委托人支付给代理人的报酬必须是随机的，而且最优激励方案涉及委托人和代理人的某种程度的风险共摊。当产量较小时，委托人想支付的报酬也小，但问题是委托人无法判断产量小的原因——是由于代理人偷懒导致的，还是由于运气不好导致的。如果委托人对低产量的处罚过重，那么他就对代理人施加了过多的风险，为了补偿这种风险，委托人必须支付更高水平的报酬。这是委托人在设计最优激励方案时必须面对的权衡。

假设不存在激励问题，唯一的问题是风险共摊问题。在这种情形下，委托人的最大化问题为

$$\begin{aligned}
&\max_{(s_i)} \sum_{i=1}^{n} (x_i - s_i)\pi_{ib} \\
&\text{s.t.} \quad \sum_{i=1}^{n} u(s_i)\pi_{ib} - c_b \geq \bar{u}
\end{aligned}$$

令 λ 表示约束条件的拉格朗日乘子，那么一阶条件为

$$-\pi_{ib} - \lambda u'(s_i)\pi_{ib} = 0,$$

这意味着 $u'(s_i)$ 是个常数，从而 s_i 是个常数。在本质上，委托人承保了代理人的全部风险。这是自然而然的，因为委托人是风险中性的而代理人是风险厌恶的。

当存在某个激励约束时，上面这个解通常不是合理的。如果委托人提供了足额保险，代理人不会关心实际产出水平，因此代理人没有动机选择委托人想要的行动：如果代理人得到某个确定性的报酬，而不管他的努力水平如何，那么代理人为何要辛苦工作呢？最优激励合同的设计涉及到下面二者之间的权衡：一是委托人承保代理人带来的收益；二是这种保险产生的激励成本。

最优激励方案的分析

我们使用下面策略分析最优激励方案的设计。首先，我们将确定能必然诱导出各个可能行动的最优激励方案。然后我们将比较致谢方案带给委托人的效用，判断哪个方案对他来说成本最小。为简单起见，假设行动只有两种即 a 和 b，并且我们的问题是如何设计方案才能诱导出行动 b。令 $V(b)$ 表示委托人设计的能诱导委托人选择 b 的方案带给委托人的最大可能的效用。委托人面对的最大化问题为

$$V(b) = \max_{(s_i)} \sum_{i=1}^{n} (x_i - s_i)\pi_{ib}$$

$$\text{s.t.} \sum_{i=1}^{n} u_i(s_i)\pi_{ib} - c_b \geq \bar{u} \tag{25.6}$$

$$\sum_{i=1}^{n} u_i(s_i)\pi_{ib} - c_b \geq \sum_{i=1}^{n} u_i(s_i)\pi_{ia} - c_a \tag{25.7}$$

此处，条件（25.6）为参与约束，条件（25.7）为激励相容约束。

在这个最大化问题中，目标函数是线性的，约束条件是非线性的。尽管我们可以直接分析这个问题，但是为了图形处理的方便，我们有必要将这个问题重新表述为约束条件为线性、目标函数为非线性的形式。令 u_i 为结果 i 带来的效用，因此 $u(s_i) = u_i$。令 f 表示效用函数的反函数，将其写为 $s_i = f(u_i)$。函数 f 表明了委托人提供给代理人效用 u_i 而花费的成本。容易证明 f 是递增的凸函数。使用函数 f 重写（25.6）和（25.7）可得

$$V(b) = \max_{(u_i)} \sum_{i=1}^{n} (x_i - f(u_i))\pi_{ib}$$

$$\text{s.t.} \sum_{i=1}^{n} u_i\pi_{ib} - c_b \geq \bar{u} \tag{25.8}$$

$$\sum_{i=1}^{n} u_i\pi_{ib} - c_b \geq \sum_{i=1}^{n} u_i\pi_{ia} - c_a \tag{25.9}$$

在这里，我们将最大化问题看成为代理人选择一个效用分布，其中委托人提供效用 u_i 的成本为 $s_i = f(u_i)$。

当 $n = 2$ 时，我们可用图形分析这个问题。在这种情形下，产量水平只有两个：x_1 和 x_2。

委托人知需要设定两个效用水平：一个为u_1，它表示当产量为x_1时代理人得到的效用；另外一个为u_2，它表示产量为x_2时代理人得到的效用。

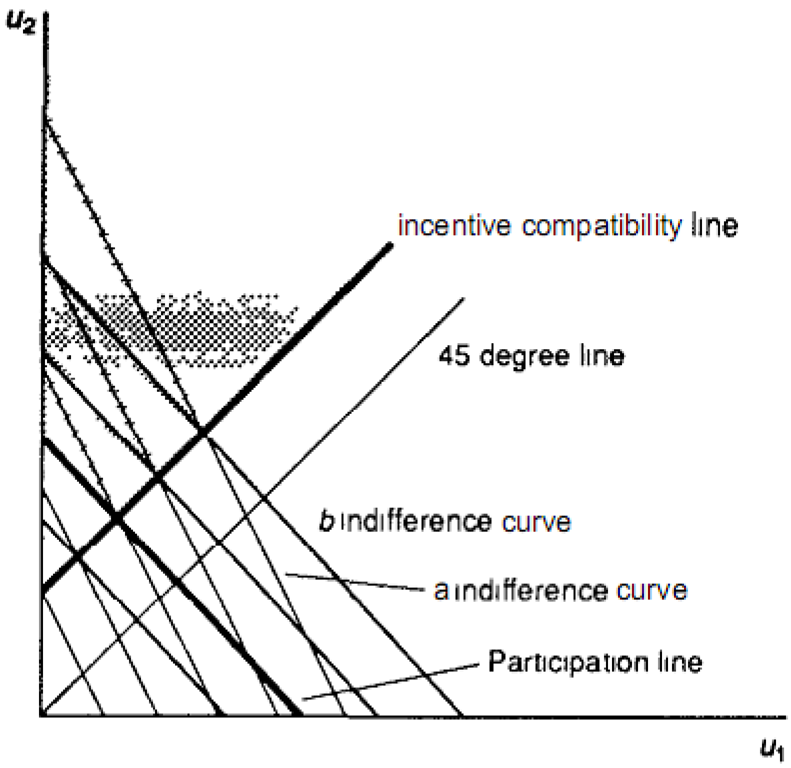


图 25.1：隐藏行动情形下的委托-代理问题的可行集。参与线的东北区域满足参与约束。激励相容线西北的区域满足激励相容约束。这两个区域的交为阴影区域。

图 25.1 给出了（25.8）-（25.9）确定的约束集。代理人选择行动a或b时的无差异曲线是具有下列形式的直线

$$\pi_{1b}u_1+\pi_{2b}u_2-c_b=\text{常数，}$$

$$\pi_{1a}u_1+\pi_{2a}u_2-c_a=\text{常数。}$$

考察激励相容约束（25.9），并且考虑效用组合(u_1,u_2)——代理人正好在行动b和a之间无差异的那些点。这些点是行动a的无差异曲线与代表同等效用水平的行动b的无差异曲线的交点。这样的效用组合(u_1,u_2)满足方程

$$\pi_{1b}u_1+\pi_{2b}u_2-c_b=\pi_{1a}u_1+\pi_{2a}u_2-c_a.$$

将u_2视为u_1的函数，解出，可得

$$u_2=\frac{\pi_{1a}-\pi_{1b}}{\pi_{2a}-\pi_{2b}}u_1+\frac{c_b-c_a}{\pi_{2b}-\pi_{2a}}=u_1+\frac{c_b-c_a}{\pi_{2b}-\pi_{2a}}. \tag{25.10}$$

u_1 的系数为 1，这是因为

$$\pi_{1a}+\pi_{2a}=\pi_{1b}+\pi_{2b}=1.$$

由此可知，（25.10）式中确定的激励相容线的斜率为+1。行动b比行动a更受代理人偏好的区域是位于这条直线上方的区域。

参与约束要求

$$\pi_{1b}u_1+\pi_{2b}u_2-c_b\geq\bar{u}.$$

上述条件以等式成立时的集合(u_1,u_2)就是代理人的一条b无差异曲线。相交区域满足激励相容约束，满足参与约束的区域请见图 25.1。

这个图也画出了 45 度线。这条 45 度线比较重要，因为它画出了满足$u_1=u_2$的效用组合(u_1,u_2)。我们已经看到如果不存在激励相容约束，委托人只要承保代理人即可，而且最优解满足条件$u_1=u_2=\bar{u}$。

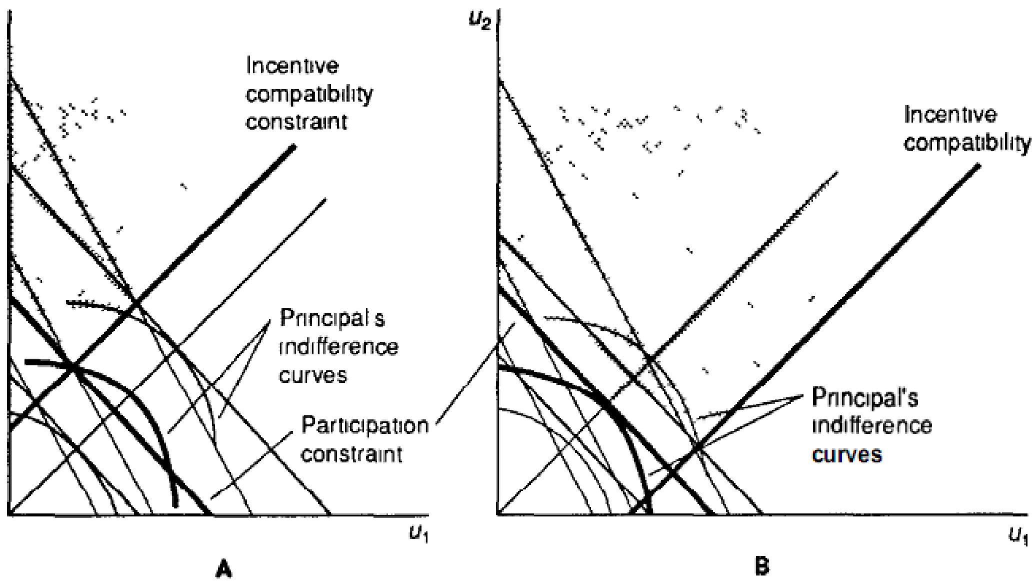


图 25.2：委托代理问题的两种解。在 A 图中，我们画出的是在最优解中代理人需要承担一定风险的情形；在 B 图中，足额保险是最优的。

由于存在激励相容约束，足额保险的点可能不是可行的。委托代理问题的解的性质，取决于激励相容线是否与纵轴或者横轴相交。我们已经在图 25.2 中画出了这些情形。为了找到最优解，我们只要画出委托人的无差异曲线即可。这些无差异曲线的形式为

$$\pi_{1b}(x_1-f(u_1))+\pi_{2b}(x_2-f(u_2))=常数.$$

委托人的效用随着s_1和s_2的下降而增加。我们如何知道委托人无差异曲线的斜率？这些斜率由下式给出

$$MRS = -\frac{\pi_{1b} f'(u_1)}{\pi_{1b} f'(u_2)}.$$

当$u_1 = u_2$，我们必定有$MRS = -\pi_{1b} / \pi_{2b}$。由于代理人的无差异曲线是由$\pi_{1b}u_1 + \pi_{2b}u_2 =$常数这个条件确定的，当$u_1 = u_2$时代理人的无差异曲线也为$-\pi_{1b} / \pi_{2b}$。因此，委托人的无差异曲线必定与代理人的无差异曲线交在 45 度线上。这只不过是下列事实在几何图形上的反映：如果不存在激励问题，委托人将足额承保代理人。

因此，如果足额保险解是可行的，如图 25.2B 所示，那么它将是最优解。如果足额保险解是不可行的，我们发现在最优解中，代理人将承担一定风险。

为了从几何图形上考察最优激励方案的性质，我们回到 n 个结的情形，并且对（25.6-25.7）描述的最大化问题建立拉格朗日函数。

$$\mathcal{L} = \sum_{i=1}^{n} (x_i - s_i)\pi_{ib} - \lambda\left[c_b + \overline{u} - \sum_{i=1}^{n} u(s_i)\pi_{ib}\right] - u\left[c_b - c_a - \sum_{i=1}^{n} u(s_i)(\pi_{ib} - \pi_{ia})\right].$$

为得到库恩-塔克一阶条件，我们将这个式子对s_i微分：

$$-\pi_{ib} + \lambda u'(s_i)\pi_{ib} + \mu u'(s_i)[\pi_{ib} - \pi_{ia}] = 0.$$

将这个式子两侧同除以$\pi_{ib}u'(s_i)$并整理，即可得到确定激励方案形状的基本方程：

$$\frac{1}{u'(s_i)} = \lambda + \mu\left[1 - \frac{\pi_{ia}}{\pi_{ib}}\right]. \tag{25.11}$$

我们通常期望保留效用受到的约束是束紧的（binding），从而$\lambda > 0$。

第二个约束的问题更大；我们已从图形分析中知道，这个约束可能是束紧的，也可能不是。假设$\mu = 0$。那么（25.11）式意味着$u'(s_i)$等于某个常数$1/\lambda$；即，代理人得到的报酬和结果无关。由此可知s_i等于某个常数$\overline{s}$，将其代入激励相容约束，可知

$$u(\overline{s})\sum_{i=1}^{n} \pi_{ib} - c_b > u(\overline{s})\sum_{i=1}^{n} \pi_{ia} - c_a.$$

由于每个概率分布的和等于 1，这意味着

$$c_a > c_b.$$

因此，这个情形只有当委托人偏好的行动同时也是代理人的低成本行动时才会发生。图 25.2B 画出了这种情形，在这种情形下，委托人和代理人之间不存在利益冲突，而且委托人为代理人提供保险。

下面我们分析第二个约束为束紧约束的情形，在这种情形下$\mu > 0$，我们看到代理人得到的报酬s_i将随着结果x_i的变动而变动。在这种情形下，委托人想要的行动是代理人的高成本行动，因此代理人得到的报酬将取决于比值π_{ia} / π_{ib}的行为。

在统计文献中，π_{ia}/π_{ib} 通常使用或然率比值表示的。它衡量的是给定代理人选择行动 a 时观测到 x_i 的可能性与给定代理人选择行动 b 时观测到 x_i 的可能性的比值。这个比值越高，意味着代理人越有可能选择行动 a；这个比值越低，意味着代理人越有可能选择行动 b。

公式中含有或然率比值强烈地意味着最优激励方案的设计与统计推断问题密切相关。这表明我们可以将统计文献中的正则（regularity）条件与最优方案的行为分析联系起来。例如，一个常用的条件，即单调或然率比值性质（Monotone Likelihood Ratio Property），要求比值 π_{ia}/π_{ib} 关于 x_i 单调递减。如果这个条件得以满足，那么由此可知 $s(x_i)$ 将是关于 x_i 的单调递增函数。详细内容可参考 Milgrom（1981）。（25.11）式的显著特征是最优激励方案是如此简单：在本质上，它是或然率比值的线性函数。

例子：比较静态

和往常一样，我们可以通过考察这个问题的拉格朗日函数来了解最优激励方案的一些信息。包络定理告诉我们：委托人最优值函数关于这个问题的某个参数的导数，正好等于拉格朗日函数关于这个参数的导数。

例如，拉格朗日函数关于 c_a 和 c_b 的导数分别为

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial c_a} &= \mu \\ \frac{\partial \mathcal{L}}{\partial c_b} &= -(\lambda+\mu).\end{aligned} \tag{25.12}$$

我们可以使用这些导数回答下列古老问题：胡萝卜和大棒哪个更好？将胡萝卜想象为选择行动 b 降低的成本，将大棒想象为选择行动 a 增加的等量成本。根据（25.12）式可知，一个行动所降低的微小成本，与另外一个行动的等量的成本增加相比较，前者总是带给委托人更大的效用。实际上，胡萝卜放松了两个假设，而大棒只放松了一个。

接下来，我们考察概率分布的微小变动（$d\pi_{ia}$）。这个变动对委托人效用的效应为

$$d\mathcal{L} = -\mu\sum_{i=1} u(s_i)d\pi_{ia}.$$

这表明当激励相容方案是束紧（binding）从而 $\mu>0$ 使，委托人和代理人的利益在关于可选行动的概率分布变动上是截然相反的：任何使得代理人状况变好的变化，必定使得委托人的状况变差。

例子：均值-方差效用下的委托代理模型

下面我们介绍一个简单的激励方案问题，这个方案来自 Holmstrom and Milgrom（1987）。令行动 a 表示代理人的努力水平，令 $\tilde{x}=a+\tilde{\varepsilon}$ 表示委托人观察到的产量水平。随机变量 $\tilde{\varepsilon}$ 服

从均衡为零、方差为σ^2的正态分布。

假设委托人选择的激励方案是线性的，因此$s(\tilde{x})=\delta+\gamma\tilde{x}=\delta+\gamma a+\gamma\tilde{\varepsilon}$。其中$\delta$与$\gamma$是待确定的参数。由于委托人是风险中性的，他的效用为

$$E[\tilde{x}-s(\tilde{x})]=E[a+\tilde{\varepsilon}-\delta-\gamma a-\gamma\tilde{\varepsilon}]=(1-\gamma)a-\delta.$$

假设代理人的效用函数是不变绝对厌恶风险的（constant absolute risk averse）：$u(w)=-e^{-rw}$，其中r是绝对厌恶风险系数，w为财富水平。代理人的财富就是$s(\tilde{x})=\delta+\gamma\tilde{x}$。因为$\tilde{x}$是正态分布的，所以财富也是正态分布的。我们在第 11 章 11.7 节（均值方差效用例子）已经知道，在这种情形下，代理人的效用线性地取决于财富的均值和方差。由此可知，与激励方案$s(\tilde{x})=\delta+\gamma\tilde{x}$相伴的代理人的效用为

$$\delta+\gamma a-\frac{\gamma^2 r}{2}\sigma^2.$$

代理人的目标函数是效用与努力所耗费的成本$c(a)$之差，他的问题是使该目标函数最大：

$$\max_a\ \delta+\gamma a-\frac{\gamma^2 r}{2}\sigma^2-c(a).$$

这个问题的一阶条件为

$$\gamma=c'(a) \tag{25.13}$$

委托人的最大化问题是在下列两个约束条件下确定最优的δ和γ：一个约束是代理人得到的效用不能小于他的保留效用水平$\bar{u}$；另外一个约束是激励约束（25.13）。这个约束最优化问题可以写为

$$\begin{aligned}&\max_{\delta,\gamma,a}\ (1-\gamma)a-\delta\\&\text{s.t. } \delta+\gamma a-\frac{\gamma^2 r}{2}\sigma^2-c(a)\geq\bar{u}\\&\qquad c'(a)=\gamma.\end{aligned}$$

在第一个约束条件中求出δ，在第二个约束条件中求出γ，将它们代入目标函数。经过化简，

可得

$$\max_a \ a - \frac{c'(a)^2 r}{2}\sigma^2 - c(a).$$

这个问题的一阶条件为

$$1 - rc'(a)c''(a)\sigma^2 - c'(a) = 0.$$

将$c'(a) = \gamma$代入，可得

$$\gamma = \frac{1}{1 + rc''(a)\sigma^2}.$$

这个式子表明了解的本质特征。如果$\sigma^2 = 0$，即不存在风险时，我们有$\gamma = 1$：最优激励方案的形式为$s = \delta + \tilde{x}$。如果$\sigma^2 > 0$，我们有$\gamma < 1$，因此代理人承担部分风险。不确定性越大，或代理人越厌恶风险，γ越小。

25.5 隐藏行动：竞争性市场

如果有很多委托人在他们提供的激励合同上展开竞争，结果将是怎样的？在这种情形下，我们希望假设：竞争迫使委托人的利润为零，从而均衡合同必定正好盈亏平衡。在这种情形下，图 25.2 仍然适用，但是我们需要重新解释等利润线和无差异曲线的水平。

在竞争性市场中，参与约束不是束紧的（binding），而且零利润条件确定了委托人的一条特定等利润线。和垄断情形一样，这种情形下存在两种可能的均衡格局：足额保险（full insurance）或不足额保险（partial insurance）。

在足额保险合同上，所有工人都得到固定数额的报酬，不管他们生产多少产量。工人们的反应是付出最低的努力水平。在不足额保险合同均衡上，工人的工资取决于他的产量水平。由于工人承担的风险越大，为了增加工资，他付出的努力水平越大，从而使得产量水平可能越高。

考虑图 25.3 中的不足额保险情形。为了使得它成为一个均衡，需要保证不存在能为代理人带来更高效用和为委托人带来更高利润的其它合同。根据这个构造，不存在能诱使具有这些性质的行动b；然而，可能存在某个合同，它能诱导出更受当事人帕累托偏好的行动a。所谓更受帕累托偏好是指这个合同使得委托人的利润为正，并且使得代理人更偏好这个合同。

为了看清是否存在这样的合同，我们画出行动a的无差异曲线，使得它穿过不足额保险合同和行动a的零利润线。如果令利润线不与更受工人偏好的区域相交，如图 25.3a 所示，那么不足额保险合同是个均衡。如果零利润线与更受工人偏好的区域相交，如图 25.3b 所示，那么这个合同不可能是个均衡，这是因为某个企业可以提供足额保险合同，使得该合同能带来正的利润而且使得工人更偏好这个足额保险合同。在这种情形下，不存在均衡。

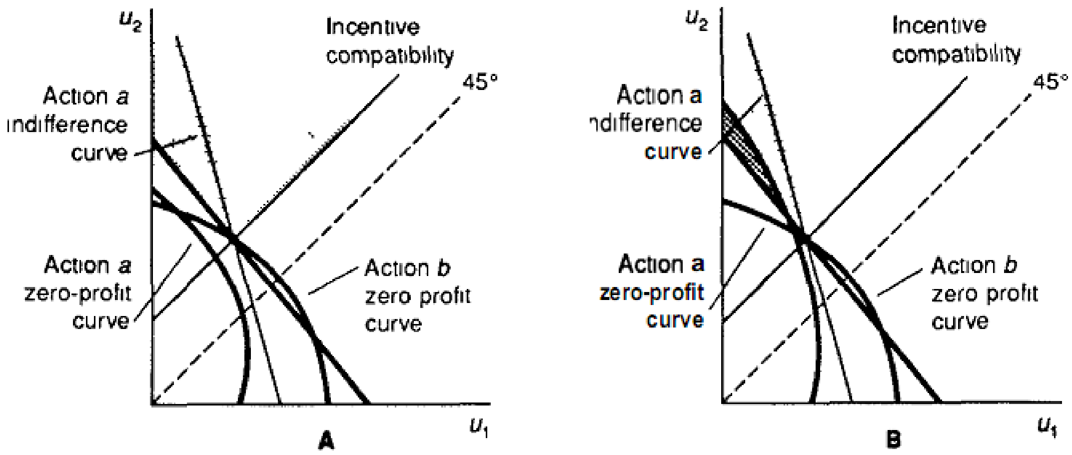


图 25.3：均衡合同。在 A 图中，不足额保险合同是个均衡。在 B 图中，不足额保险合同不是个均衡，这是因为行动 a 的零利润线与代理人更偏好的集合相交。

例子：保险市场上的道德风险

在保险市场上，带有隐藏行动的委托代理问题称为**道德风险问题**（moral hazard problem）。“道德风险”是指购买保险合同的人不会采取合理的预防措施水平。我们在下列背景下考察这个问题。这个背景就是我们在第 11 章 11.5 节（保险需求的例子）分析的例子。

假设有很多相同的消费者，他们正在考虑购买汽车偷窃险。如果某个消费者的汽车被偷，他承担的损失为 L。令状态 1 为消费者的汽车被偷的自然状态，状态 2 为未被偷的自然状态。汽车被偷的概率取决于消费者（车主）的行动，比如他是否锁好了车。令 π_{1b} 表示消费者为车上锁情形下汽车被偷的概率，π_{1a} 表示消费者忘记锁车情形下汽车被偷的概率。令 c 表示锁车需要花费的成本，令 s_i 表示消费者在状态 i 下从保险公司得到的净钱数（赔偿）。最后，令 w 表示消费者的财富。

假设保险公司希望消费者锁车，激励问题为

$$\max_{s_1,s_2} \ \pi_{1b}s_1+\pi_{2b}s_2$$

$$\text{s.t. } \pi_{1b}u(w-s_1-L)+\pi_{2b}u(w-s_2)\geq \bar{u}$$

$$\pi_{1b}u(w-s_1-L)+\pi_{2b}u(w-s_2)-c\geq \pi_{1a}u(w-s_1-L)+\pi_{2a}u(w-s_2).$$

如果不存在激励问题[因此汽车被偷的概率与消费者（代理人）的行为无关]，而且如果保险市场上的竞争迫使保险公司的期望利润为零，我们在第 11 章 11.5 节（保险需求的例子）中已经知道，最优解将是 $s_2=s_1+L$。也就是说，保险公司将向消费者提供足额保险，因此消费者的财富在汽车被偷和未被偷的情形下是完全相等的。

当损失概率取决于消费者（代理人）的行动，足额保险将不再是最优的。一般来说，

委托人希望使得代理人的消费取决于代理人自身的选择，从而使得代理人有预防损失的激励。在这种情形下，消费者对保险合同的需求是理性的。消费者希望在精算公平费率下购买更多的保险，但是保险公司不会提供这样的保险合同，因为这种做法将导致消费者对损失的预防不足。

在竞争情形下，参与约束不是束紧的，均衡由零利润条件和激励相容约束确定：

$$\begin{aligned}&\pi_{1b}s_1^* + \pi_{2b}s_2^* = 0\\&\pi_{1b}u(w - s_1^* - L) + \pi_{2b}u(w - s_2^*) - c = \pi_{1a}u(w - s_1^* - L) + \pi_{2a}u(w - s_2^*).\end{aligned} \tag{25.14}$$

这两个式子确定了均衡 (s_1^*, s_2^*)。和往常一样，我们必须验证从而保证不存在能打破这个均衡的足额保险合同。如果不作出额外的假设条件，这样的合同是可能存在的，因此这个模型不存在均衡。

25.6 隐藏信息：垄断

我们现在考察另一类委托代理问题，在这种情形下，代理人的效用或成本函数信息是不可观测到的。为简单起见，假设代理人的种类只有两种，可按照成本函数对他们进行区分；另外令代理人的行为是他生产的产量水平。我们仍然沿用前面的委托代理模型，但是现在假设委托人能够完全观测到代理人的产量，但是某些代理人发现他们的成本高于另外一些代理人。委托人能够完全看到代理人的行动，但他无法知道这些代理人的行动所花费的成本。

令 x_t 和 $c_t(x)$ 分别表示 t 类代理人的产量和成本函数。更具体一些，令类型 2 代理人是高成本的代理人，因此对于所有 x 都有 $c_2(x) > c_1(x)$。令 $s(x)$ 表示代理人得到的报酬，它是产量的函数，并且假设 t 类代理人的效用函数为 $s(x) - c_t(x)$。委托人不能肯定他面对的代理人是哪一类的，但他为 t 类代理人赋予概率 π_t。和往常一样，我们要求每个代理人得到的效用不能小于他的保留效用水平，为简单起见，假设保留效用水平为零。

为简单起见，我们进一步假设成本函数具有下列性质：高成本代理人的边际成本也高，即对于所有 x 都有 $c_2'(x) > c_1'(x)$。这个性质有时称为一次相交性质（single-crossing property），这是因为它意味着类型 1 代理人的任何给定的无差异曲线与类型 2 的任何给定的无差异曲线最多相交一次。我们给出下列简单事实，我们要求你在习题中证明它：

一次相交性质。**假设对于所有 x 都有 $c_2'(x) > c_1'(x)$，那么对于任何两个不同的产量水平 $x_2 > x_1$，我们必定有 $c_2(x_2) - c_2(x_1) > c_1(x_2) - c_1(x_1)$。**

为了更好地理解激励方案问题，我们首先考虑在委托人能够观测到代理人的成本函数的情形下，最优激励方案是什么样的。在这种情形下，委托人具有充分信息，因此解在本质上就是我们前面讨论过的目标产量问题。委托人只要使得总产量减去总成本 $x_1 + x_2 - c_1(x_1) - c_2(x_2)$ 最大即可。这个最大化问题的解要求 $c_t'(x_t^*) = 1$（其中 $t = 1, 2$）。因

此，委托人支付给代理人的报酬恰好等于代理人的保留效用水平，$s_t - c_t'(x_t^*) = 0$。

我们将上述情形画在图 25.4 中。在此图中，我们用纵轴表示边际成本，用横轴表示产量。t 类代理人生产的产量为 x_t^*，此时 $c_t'(x_t^*) = 1$。如果委托人能够明确明确区分这两类代理人，他只需要求 t 类代理人生产产量 x_t^* 即可，和前面的例子一样，这是一种目标产量方案。即，t 类代理人得到的报酬为：在 x^* 的产量上，他得到的报酬为 $s_t(x^*) = c_t(x_t^*)$；在所有其它产量 x 上，他得到的报酬 $s_t(x) < c_t(x)$。

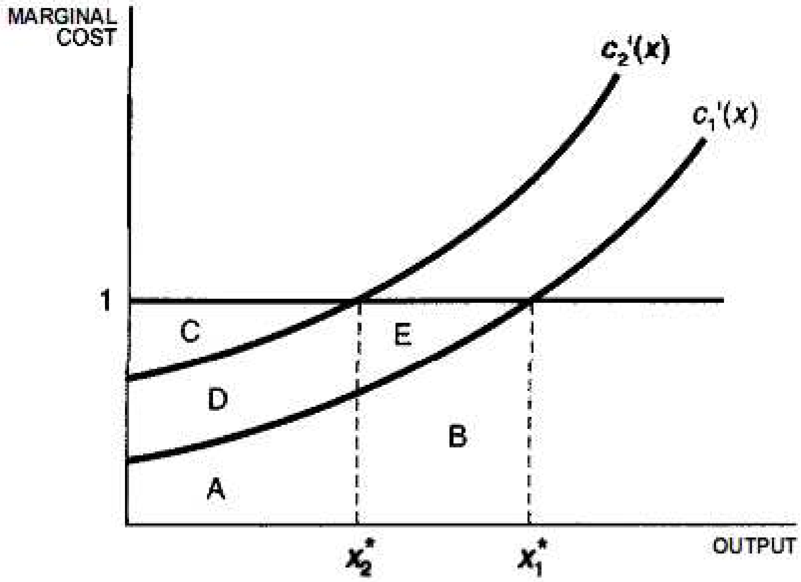


图 25.4：隐藏信息情形下的委托代理问题。在最优（first-best）方案中，1 类代理人生产产量 x_1^*，2 类代理人生产产量 x_2^*。

这意味着每个代理人的总剩余被委托人攫取了。反映在图形上，1 类代理人得到的报酬为（$A+B$），这正好等于他的总生产成本；类似地，类型 2 的代理人得到的报酬（$A+D$）也等于类型 2 代理人的总成本。

这种方案的问题是它不满足激励相容约束。如果高成本代理人正好满足他的参与约束，低成本代理人必定偏好 (s_2, x_2^*) 而不是 (s_1, x_1^*)。用符号表示，

$$s_2 - c_1(x_2^*) > s_2 - c_2(x_2^*) = 0 = s_1 - c_1(x_1^*)$$

这是因为对于所有 x 都有 $c_2(x) > c_1(x)$。反映在图形上，低成本代理人偏爱冒充成高成本代理人，从而只生产产量 x_2^*，这样低成本代理人得到了面积为 D 的剩余。

这种道德风险问题的一种解决方法是改变报酬。假设如果产量为 x_2^*，支付给代理人的报酬为 A，但是如果产量为 x_1^*，支付报酬 $A+D$。这使得低成本代理人得到了净剩余 D，从而使得这类代理人在生产 x_1^* 和 x_2^* 之间是无差异的。

这种方案当然是可行的，但对于委托人来说，它是最优的吗？答案为否定的，这存在着一种有趣的解释。假设我们稍微降低高成本代理人的目标产量。由于代理人的最优产量（站在代理人的角度）是使得价格等于边际成本，这只会导致利润的一阶减少：产量的减少量正好等于 2 类代理人报酬的减少量。

但是由于现在 x_2 和面积 D 都变小了，低成本代理人生产 x_2 得到的报酬也变小了。这样，高成本代理人的产量和得到的报酬都减少了，从而使得这样的产量水平对于低成本代理人缺乏吸引力。这种做法的效应大于一阶效应，这是因为现在低成本代理人在边际成本小于 1 之处生产。

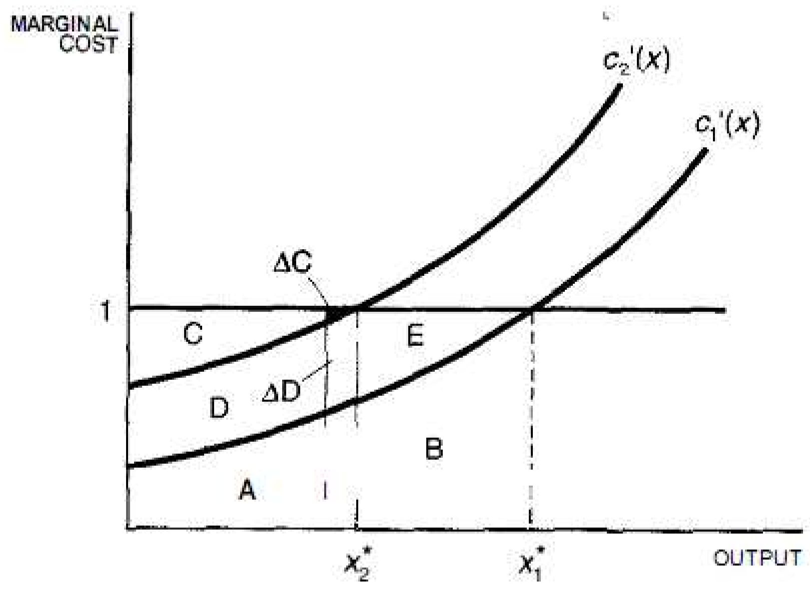


图 25.5：利润增加。如果稍微降低高成本代理人的目标产量，委托人的利润就会增加。

我们将上述情形画在图 25.5 中。高成本代理人目标产量的下降，使得委托人从高成本代理人得到的利润下降了 ΔC，但是从低成本代理人得到的利润增加了 ΔD。因此，委托人发现，如果将高成本代理人的目标产量减少到小于有效率产量的某个产量水平，那么这对委托人有利可图。通过减少高成本代理人报酬的方法，委托人减少了他必须支付给低成本代理人的报酬。

为了进一步考察激励方案的结构，我们需要从代数角度表达这个问题。

正如几何图形分析指出的，基本的激励问题是低成本代理人试图冒充成高成本的代理人。如果 x_1 是 1 类代理人选择的产量，那么委托人的激励方案必须使得：1 类代理人从选择

x_1得到的效用，大于他选择x_2得到的效用。类似的逻辑也适用于 2 类代理人。这只是一种特殊形式的激励相容条件，在这种情形下，我们将其称为**自我选择约束**（self-selection constraints）。

给定这些结论，我们可以将委托人的最优化问题写为：

$$\max_{x_1,x_2,s_1,s_2} \quad \pi_1(x_1-s_1)+\pi_2(x_2-s_2)$$

s.t.

$$s_1-c_1(x_1)\ge 0 \tag{25.15}$$

$$s_2-c_2(x_2)\ge 0 \tag{25.16}$$

$$s_1-c_1(x_1)\ge s_2-c_1(x_2) \tag{25.17}$$

$$s_2-c_2(x_2)\ge s_1-c_2(x_1) \tag{25.18}$$

前两个约束是参与约束。后面两个月是是激励相容约束或称为自我选择约束。最优激励方案$(x_1^*,s_1^*,x_2^*,s_2^*)$是这个最大化问题的解。

下面两个式子来自自我选择约束的变形：

$$s_2\le s_1+c_1(x_2)-c_1(x_1) \tag{25.19}$$

$$s_2\ge s_1+c_2(x_2)-c_2(x_1) \tag{25.20}$$

这两个不等式表明，如果自我选择得以满足，那么

$$c_1(x_2)-c_1(x_1)\ge c_2(x_2)-c_2(x_1) \tag{25.21}$$

一次相交条件意味着 2 类代理人的边际成本总是高于 1 类代理人的。如果$x_2>x_1$，这将与（25.21）矛盾。因此，在最优解中必定有$x_2\le x_1$，它的意思是说低成本代理人生产的产量不会小于高成本代理人的。

现在考察约束（25.15）和（25.17）。我们可以将这两个约束重新写为

$$s_1\ge c_1(x_1) \tag{25.15$'$}$$

$$s_1\ge c_1(x_1)+[s_2-c_1(x_2)] \tag{25.17$'$}$$

由于委托人希望s_1尽可能小，上述两个约束最多有一个是束紧的。从约束（25.16）和成本函数的性质可知，

$$s_2-c_1(x_2)>s_2-c_2(x_2)=0.$$

因此，（$25.17'$）中的方括号内的项为正，（$25.15'$）不可能是束紧的。由此可知，

$$s_1=c_1(x_1)+[s_2-c_1(x_2)]. \tag{25.22}$$

类似地，在约束（25.16）和（25.18）中，也正好只有一个约束是束紧的。（25.18）是束紧

的吗？如果是，我们可以将（25.22）代入（25.18）可得

$$s_2 = s_1 + c_2(x_2) - c_2(x_1) = s_2 + c_1(x_1) - c_1(x_2) + c_2(x_2) - c_2(x_1).$$

整理可得

$$c_1(x_2) - c_1(x_1) = c_2(x_2) - c_2(x_1),$$

这违背了一次相交条件。由此可知最优的激励合同必定满足

$$s_2 = c_2(x_2) \tag{25.23}$$

到目前为止，尽管我们尚未真正考察这个最优化问题，但我们已经看到约束条件和目标函数本身蕴涵着两个重要性质：高成本代理人得到的报酬恰好能使得他在参与和不参与之间无差异，低成本代理人得到了一部分剩余。低成本代理人得到的剩余正好能使得他在伪装和不伪装成高成本代理人之间无差异。

为了确定最优行动，我们用（25.22）式（即 $s_1 = c_1(x_1) + [s_2 - c_1(x_2)]$ ）和（25.3）式（即 $s_2 = c_2(x_2)$ ）分别替换委托人最大化问题中的 s_1 和 s_2，从而可将委托人的最大化问题写为

$$\max_{x_1, x_2} \pi_1[x_1 - c_1(x_1) - c_2(x_2) + c_1(x_2)] + \pi_2[x_2 - c_2(x_1)]$$

这个问题的一阶条件为

$$\pi_1[1 - c_1'(x_1)] = 0$$

$$\pi_1[c_1'(x_2) - c_2'(x_2)] + \pi_2[1 - c_2'(x_2)] = 0.$$

我们可以将这两个条件写为

$$\begin{aligned} & c_1'(x_1^*) = 1 \\ & c_2'(x_2^*) = 1 + \frac{\pi_1}{\pi_2}[c_1'(x_2^*) - c_2'(x_2^*)]. \end{aligned} \tag{25.24}$$

第一个式子意味着低成本代理人的生产的产量水平，恰好等于只有他这一类代理人时他生产的产量，即帕累托有效率的产量水平。给定一次相交性质，高成本的代理人生产的产量，小于只有他这类代理人时他生产的产量，这是因为 $c_2'(x_2^*) - c_1'(x_2^*) > 0.$

为了从图形上描述这些条件，也为简单起见，假设 $\pi_1 = \pi_2 = 1/2$ 。于是（25.24）中的第二个式子意味着 $2c_2'(x_2^*) = 1 + c_1'(x_2^*)$ 。在这个点上，稍微降低 x_2 得到的边际收益正好等于边际成本。图 25.6 画出了最优解。低成本代理人在他的边际收益等于他的边际成本处生产；高成本代理人在他的边际收益大于他的边际成本处生产。高成本代理人得到的报酬为 $A + D$，正好等于他的成本；低成本代理人得到的报酬为 $A + B + D$，这个报酬使得他正好在伪装和不伪装成高成本代理人之间无差异。

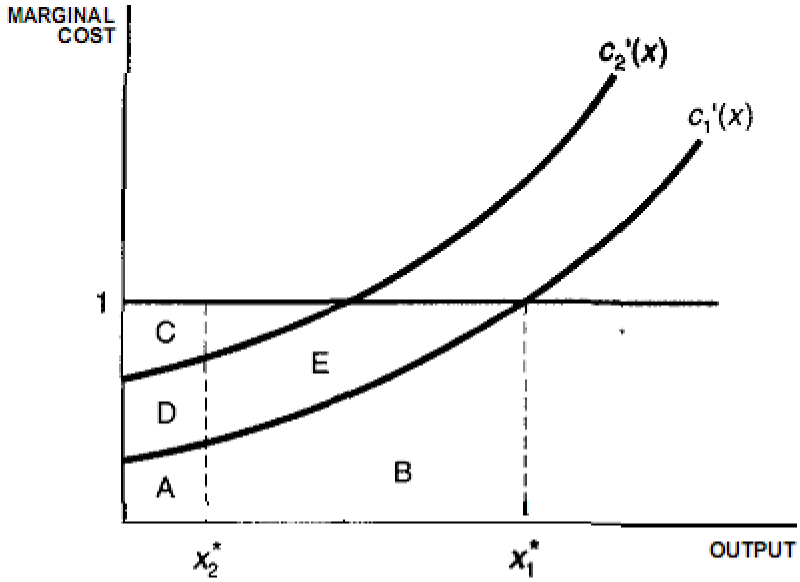


图 25.6：最优合同。高成本代理人的产量为x_2^*，低成本代理人的产量为x_1^*。高成本代理人得到的报酬为$A+D$，低成本代理人得到的报酬为$A+B+D$。

图 25.7 画出了最优激励合同的另外一个图。在这个图中，我们在空间(s,x)中描述合同。t类代理人的效用函数的形式为$u_t = s_t - c_t(x_t)$。因此，他的无差异曲线的形式为$s_t = u_t + c_t(x_t)$。根据一次相交性质可知，高成本代理人的无差异曲线总是比低成本代理人的无差异曲线陡峭。

我们已经知道在均衡时，高成本代理人得到了他的保留效用水平（等于零）。这决定了高成本代理人的无差异曲线和所有激励合同(s_2,x_2)必定位于效用为零的无差异曲线上。企业（委托人）从t类代理人身上赚取的利润为$P_t = x_t - s_t$。因此，等利润线的形式为$s_t = x_t - P_t$。这些等利润线是相互平行的直线，它们的斜率为+1，纵截距为$-P_t$。企业的总利润为$\pi_1 P_1 + \pi_2 P_2$。注意，当等利润线向东南方向移动，企业的利润增加；当无差异曲线向西北方向移动，代理人的效用增加。

从条件（25.24）可知，低成本代理人必定满足条件$c_1'(x_1^*) = 1$。这意味着等利润曲线必定与低成本代理人的无差异曲线相切。我们还知道$c_2'(x_2^*) < 1$，因此等利润线穿过高成本代理人的无差异曲线。

如果不存在低成本代理人，委托人将希望高成本代理人生产更多产量，而且高成本代理人也希望这么做。图 25.7 中的阴影区域表示的是高成本代理人和委托人的状况都能变好的区域。但是由于低成本代理人的存在，如果委托人增加高成本代理人的产量，那么委托人支付给低成本代理人的报酬必须增加。在均衡时，委托人通过降低x_2和s_2从而增大P_2带来的好处正好被P_1的下降所抵消。

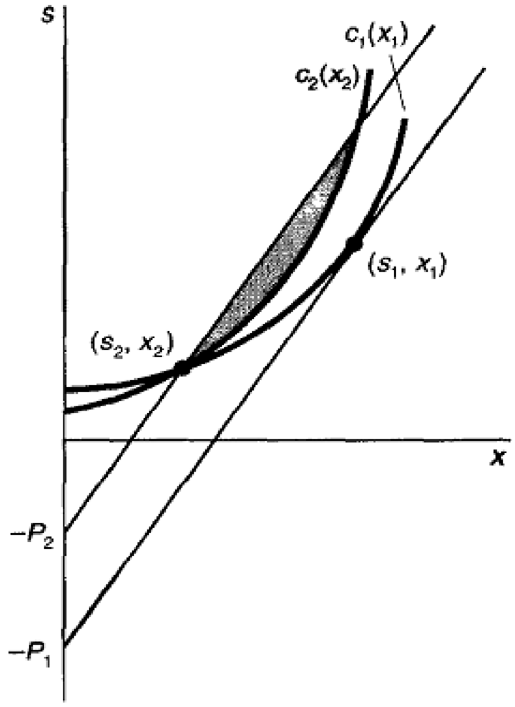


图 25.7：最优激励合同。企业（委托人）的利润为$\pi_1 P_1+\pi_2 P_2$。阴影区域表示的是由自我选择约束引致的对高成本工人（代理人）的无效率使用。

正是高成本代理人和低成本代理人之间的负外部性，才导致了无效率的均衡。如果垄断企业能够区分工人类型，从而提供不同的工资，那么结果将是有效率的。这与第 14 章讨论的第二级价格歧视非常类似。在这个模型中，如果只有一类消费者，那么垄断企业将实施完全价格歧视：对消费者提供的选择是“要么购买要么走人”。但是若垄断企业面对的消费者类型有多种，那么若它实施价格歧视，通常只能得到无效率的结果。

25.7 市场均衡：隐藏信息

和往常一样，我们可以通过对模型增加零利润条件以及重新解释保留效用来分析竞争性均衡。随着更多的企业进入市场，它们抬高了工人的工资，降低了代表性企业的利润。在垄断问题中，保留价格决定了利润水平；在竞争均衡问题中，零利润条件决定了工人们的效用。

为了看清这一点，可考察图 25.7。在垄断情形下，高成本代理人的无差异曲线决定了企业的利润$\pi_1 P_1+\pi_2 P_2$。在竞争情形下，企业的利润被压低为零，代理人移动到更高的无差异曲线上。

我们只考察对称性均衡，即所有企业提供的合同集是相同的。乍看起来，似乎有多种均衡情形：

（a）代表性企业提供单一合同，该合同对两类工人都有吸引力。

（b）代表性企业提供单一合同，该合同只对一类工人具有吸引力。

（c）代表性企业提供两种合同，每一种对相应类型的工人具有吸引力。

两类工人接受同一种合同的情形称为**一体性均衡**（pooling equilibrium）。另外一种情形，即不同类型的工人接受不同的合同称为**分离性均衡**（separating equilibrium）。这样，（b）和（c）都是一体性均衡，而（a）是分离性均衡。下面我们考察这些均衡的可能性。

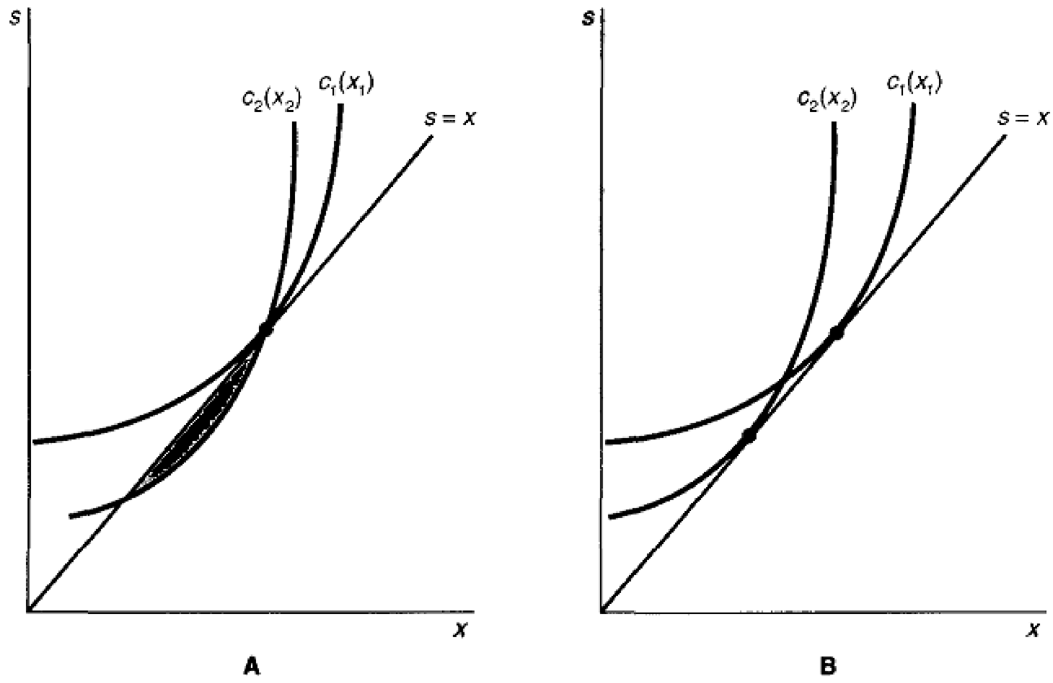


图 25.8：可能的均衡。A 图不可能是个均衡，原因见正文分析。唯一可能的均衡是 B 图所示的情形，其中每种工人得到自己的边际产品。

我们在图 25.8 中画出了可能的均衡图。不难看出，如果企业只提供一种合同，那么就不可能达到均衡，因此我们断言（a）和（c）不可能是均衡。如果代表性企业的利润为零，那么它必定在图 25.8A 中的 45 度线上经营。如果它只提供一种合同，比如(s^*, x^*)，那么这个合同必定只能对一种工人来说是最优的。假设它对于低成本工人来说是最优的。但是，企业会发现如果它偏离上述选择而是提供阴影区域中的合同（高成本工人偏好这样的合同），它就能赚取正的利润。类似地论证适用于如果合同对于高成本工人是最优的情形。

由上面的分析可知，只要每类工人至少得到自己的保留效用水平，那么唯一可能的均衡是图 25.8 所示的分离性均衡。企业支付给每类工人的报酬等于该类工人的产品价值，从而利润为零。

例：一个代数例子

下面我们使用代数例子说明隐藏信息的垄断模型和隐藏信息的竞争模型之间的区别。假设$c_t(x_t)=tx_t^2/2$，$\pi_1=\pi_2=1/2$。那么垄断者的最优解可由（25.22）、（25.23）和（25.24）式确定。你应该验证一下这些方程的解为

$$x_1^*=1,\quad x_2^*=1/3,\quad s_1^*=5/9,\quad s_2^*=1/9.$$

垄断者的利润为

$$\frac{1}{2}(x_1^*-s_1^*)+\frac{1}{2}(x_2^*-s_2^*)=\frac{1}{2}\times\frac{4}{9}+\frac{1}{2}\times\frac{2}{9}=\frac{1}{3}.$$

在垄断模型中，高成本工人正好得到他的保留效用水平（等于零）。在竞争模型中，由于企业竞争抬高了价格，工人得到的效用增加了。

我们已经知道在竞争均衡中，工资是线性的，因此t类工人希望使得$x_t-c_t(x_t)$最大化。由此可知$x_1=1$，$x_2=1/2$。企业的利润为零，因此我们必定有$s_1=x_1=1$，$s_2=x_2=1/2$。低成本工人的剩余为 1/2，高成本工人的剩余为 1/4。

25.8 逆向选择

考虑上一节模型的一个变种。假设工人除了成本函数不同之外，他们的生产率也不同。高成本工人生产v_2x_2单位产品，而低成本工人生产v_1x_1单位。我们假设$v_1>v_2$，因此现在企业喜欢低成本的工人的原因有二：更高的生产率和更低的成本。

在这种情形下，均衡工资合同是什么样的？和上一章一样对于对称性均衡来说，存在这两种符合逻辑的均衡可能：一是企业 1 类和 2 类工人提供同一种合同(s^*,x^*)；二企业是对 1 类和 2 类工人分别提供合同(s_1^*,x_1^*)和(s_2^*,x_2^*)。我们已经知道，如果企业对所有类型的工人都提供同一种合同，我们将其称为一体性均衡；如果对不同类的工人提供不同的合同，则称为分离性均衡。

首先考虑一体性均衡。在这种情形下，即使某类工人比另一类具有更高的生产率，所有工人也只能得到同样的报酬。由于总利润为零，低成本工人带给企业的利润必定为负，高成本工人带给企业的利润必定为正。两类工人的产品总价值$(\pi_1v_1+\pi_2v_2)x^*$等于总成本$\pi_1s^*+\pi_2s^*=s^*$。因此，(s^*,x^*)必定位于直线$s=(\pi_1v_1+\pi_2v_2)x$上，它的斜率为这两类工人生产率的加权平均，如图 25.9 所示。

可能的一体性均衡位于这条线的某个点上。在任何这样的点，画出穿过这个点的每类工人的无差异曲线。根据假设高生产率工人的无差异曲线比低生产率工人的平缓。这意味着阴影区域存在着下列这样的合同：该合同对高生产率工人有利，对低生产率工人不利。企业会发现如果它选择这样的合同，那么它吸引得只是高生产率工人，因此利润为正。由于这样的构造（configuration）在零利润线上的任何点都能发生，因此不存在一体性均衡。

下面我们考察分离性均衡的可能性。图 25.10 描述了有效率且均衡的合同。合同 (s_1^*, x_1^*) 和 (s_2^*, x_2^*) 是（完全信息的）有效率合同，但是它们不满足自我选择约束：低生产率工人更喜欢企业为高生产率工人制定的合同。如果一个企业提供合同 (s_2^*, x_2^*)，那么它希望只吸引低成本且高生产率的工人。但是这个企业会经历逆向选择——这两类工人都更喜欢这个合同。

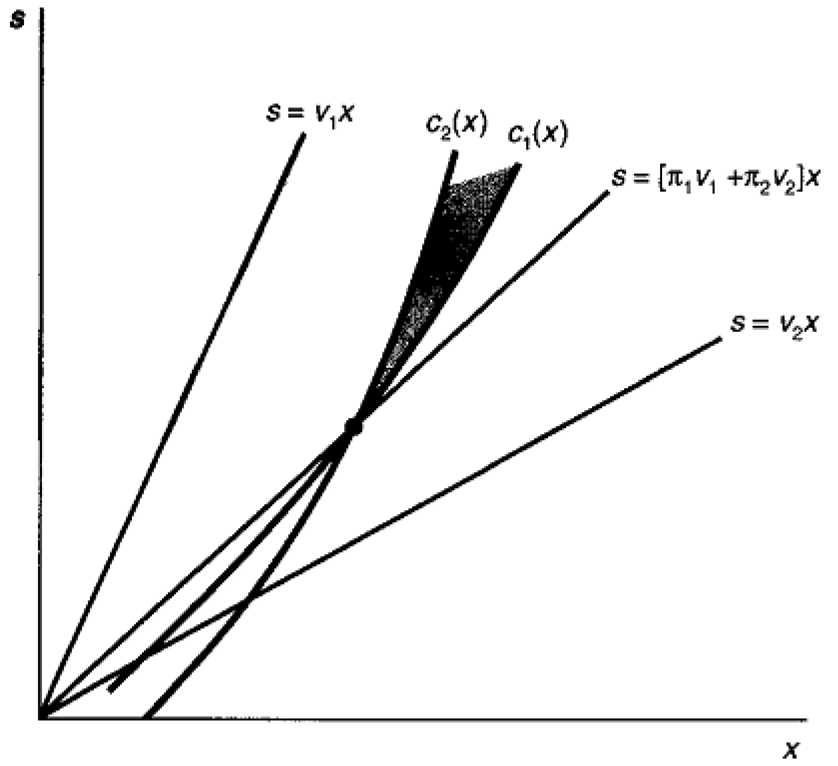


图 25.9：一体性均衡不可能发生。如果企业对不同类工人提供同一种合同，它必定位于零利润线上。画出通过这个合同的无差异曲线，并且注意到由于低生产工人的无差异曲线更陡峭，企业总有可能在阴影区域找到只吸引高生产率的工人的合同，从而赚取正利润。

这个逆向选择问题的解决方法是把高生产率工人的零利润线移动到象 (s_1', x_1') 点这样的点。现在 (s_1', x_1') 和 (s_2^*, x_2^*) 是合同的一个均衡构造：低生产率的工人正好在他的合同与高生产率工人合同之间无差异。任何一类工人的无差异曲线上方的点对企业不是有利可图的，这样我们得到了一个均衡。如图 25.10 所示。

然而，也可能不存在均衡。注意到，通过 (s_1', x_1') 的无差异曲线必定穿过零利润线。由此可知，存在诸如图 25.10 中的阴影区域，使得企业和高生产率工人都喜欢这个区域。然而，企业不会提供位于这个区域的合同，这是因为它也会吸引低生产率的工人，从而使企业无利可图。这样的合同之所以是无利可图的，是因为在图 25.10 中一体性（pooled）工人的零利

润线在阴影区域的下方。

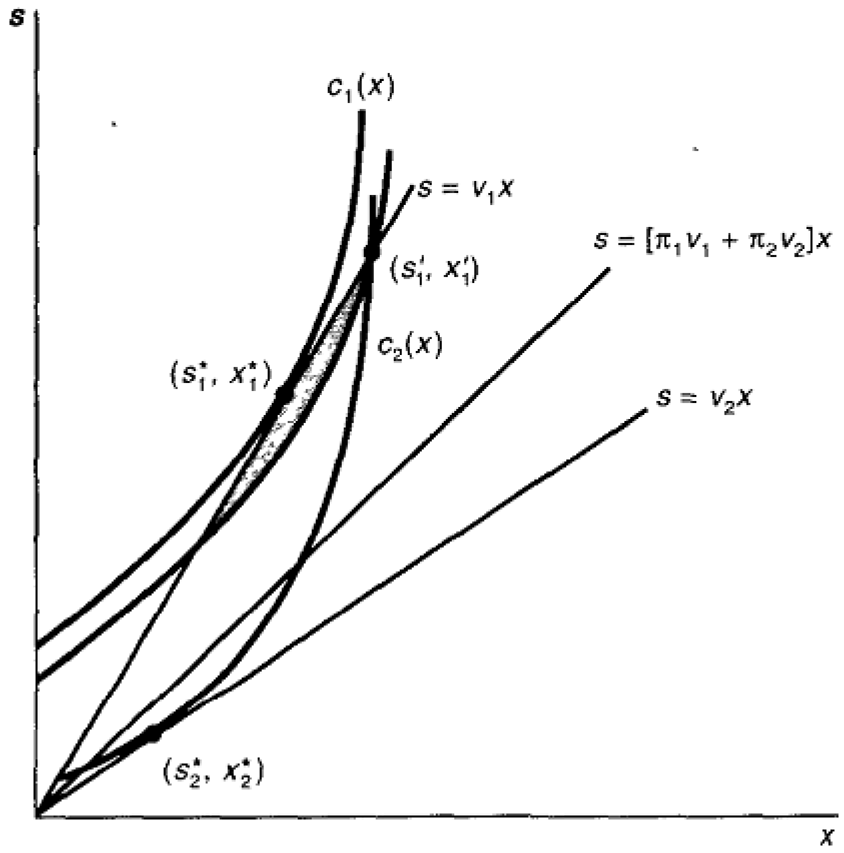


图 25.10：分离性均衡。合同 (s_1^*, x_1^*) 和 (s_2^*, x_2^*) 是有效率的，但是它们不满足自我选择约束。合同 (s_2^*, x_2^*) 和 (s_1', x_1') 满足自我选择约束。

但是假设高生产率的工人有很多，从而使得线 $s = \pi_1 v_1 + \pi_2 v_2$ 与阴影区域相交。在这种情形下，企业提供位于这个区域的一体性合同就是有利可图的。因此，可能的分离性均衡就被破坏，不存在纯策略均衡。

25.9 次品市场和逆向选择

下面我们再举一个由于存在逆向选择而导致模型不存在均衡的例子。考虑二手车市场。假设当前的车主比潜在买者更了解汽车的质量。一旦买者认识到这一点，他们就可能不愿意买二手车，这是因为他们担心可能买到次品（这种担心显然是正确的）。买者的疑问是：如果车的质量很好，为何车主要卖掉呢？尽管二手车市场上存在很多的潜在买者和卖者，但它的规模可能仍然很小（成交量有限）。

上述现象符合我们的直觉，但一直没有人将其模型化，直到 Akerlof（1970）论文的出

现。在他所谓的**次品市场**（lemons market）中，假设我们可用数字q表示某辆二手车的质量，q在区间$[0,1]$均与分布。后面将使用下列事实：如果q在区间$[0,b]$之间均匀分布，那么q的平均值为$b/2$，因此二手车市场上汽车的平均质量为$b/2$。

二手车市场上有大量的买者和卖者，买者愿意最多花$\frac{3}{2}q$的钱来买质量为q的车，卖者愿意把质量为q的车最少卖q元。因此，如果车的质量是可观测到的，那么每辆质量为q的车的成交价必定位于区间$[q,\frac{3}{2}q]$之中。

然而，假设车的质量是不可观测到的。那么买者将会用市场上车的平均质量来估计他买到车的质量。我们假设车的平均质量是可观测到的，尽管每辆车的质量不可知。因此，买者对每辆车最多愿意支付$\frac{3}{2}\overline{q}$。

在这个市场上，均衡价格为多少？假设均衡价格是某个数$p>0$。那么车的质量小于p的车主将希望以这个价格把车卖掉，因为对于这些车主来说，p大于他们的保留价格。由于车的质量在区间$[0,p]$均匀分布，准备卖掉的车的平均质量为$\overline{q}=p/2$。将它代入买者的保留价格表达式$\frac{3}{2}\overline{q}$，可知买者最多愿意支付$\frac{3}{2}\overline{q}=\frac{3}{2}\times\frac{p}{2}=\frac{3}{4}p$。由于买者的保留价格（$\frac{3}{4}p$）小于卖者的保留价格（$p$），因此在价格$p$的水平上，车的成交量为零。由于$p>0$是任意的，我们已经证明了二手车市场的成交量为零。唯一的均衡价格是$p=0$。在$p=0$这个价格水平上，需求为零，供给也为零：买卖双方之间的信息不对称彻底摧毁了二手车市场。

任何吸引好车车主的价格水平，对差车车主的吸引力都更大。卖者对车的自我选择是偏态的——偏向于差车。车的质量越差的车主卖车的激励越大，这是逆向选择的另外一个例子。

25.10 发送信号

在本节也是最后一节，我们将说明隐藏信息问题如何导致带有逆向选择的均衡。在次品市场上，成交量微乎其微，这是因为好车和差车（质量好坏）不能轻易被区分开。在劳动市场上，有效率的合同集是不可行的，这是因为低生产率的工人希望选择企业原本为高生产率工人设计的合同。

在二手车市场上，好车车主希望**发送信号**（signal）：他卖的是好车而不是差车。一种可能是他提供保证书——如果在一定时间内车出了问题，他保证承担由此产生的费用。比如，他向你承诺如果你买他的车，他为你购买汽车保险。

为了与均衡一致，信号必定是下列这样的：好车车主有能力提供而差车车主不能。这样的信号使得好车车主向潜在的买者“证明”他的车的确是好车。提供保证需要花钱，这对于差车车主不合算对好车车主合算。因此，这种信号使得买者能区分好车与差车。在这种情形下，发送信号能让市场更有效率的运行。然而，结果并非总是如此，下面我们将看到这一点。

25.11 教育信号

我们回到劳动市场。假设劳动市场有两类工人，他们的生产率分别为v_1和v_2。假设每类工人的工作时间是固定不变的。如果企业无法区分这两类工人，那么竞争均衡时，工人只能得到等于他们平均生产率的工资，即

$$\overline{s} = \pi_1 v_1 + \pi_2 v_2 .$$

注意，高生产率工人得到的工资$\overline{s}$小于他的边际产品价值，而低生产率工人得到的工资$\overline{s}$则大于他的边际产品价值。高生产率工人有发送信号的激励，以此表明他的生产率高。

假设存在某种信号，使得高生产率工人比低生产率工人更容易得到它。教育就是这样的信号：能力高的人学习不费劲，花费的成本低，也就是说高生产率工人更容易获得教育。为具体起见，假设低生产率和高生产率工人接受e年教育需要花费的成本分别为$c_1 e$和$c_2 e$，其中$c_1 > c_2$。

现在假设教育对生产率没有任何影响。然而，企业仍然发现将工资与教育年限挂钩是有利可图的，因为这么做能吸引高生产率的工人。假设工人们相信企业支付的工资为$s(e)$，其中s是e的增函数。发送信号均衡将是下面这样的：工人们猜想出关于工资的一个组合，而且企业的行为证明它是正确的。

令e_1和e_2分别表示低生产率和高生产率工人实际选择的教育年限。那么分离性发送信号均衡必须满足零利润条件

$$s(e_1) = v_1, \quad s(e_2) = v_2 ,$$

以及自我选择约束

$$s(e_1) - c_1 e_1 \geq s(e_2) - c_1 e_2,$$
$$s(e_2) - c_2 e_2 \geq s(e_1) - c_2 e_1.$$

一般来说，满足这些条件的函数$s(e)$可能有很多。我们能找到其中一个就足够了。

令e^*表示满足下列条件的一个数

$$\frac{v_2 - v_1}{c_2} > e^* > \frac{v_2 - v_1}{c_1}.$$

假设工人们猜想出的工资函数为

$$s(e)=\begin{cases} v_2, & 若 e>e^* \\ v_1 & 若 e\le e^* \end{cases}$$

容易证明这个函数满足自我选择约束，因此这个函数给出的工资组合与均衡是一致的。

注意，站在社会角度看，这个发送信号的均衡存在着浪费（即不是有效率的）。在我们的例子中，教育对于社会生产没有贡献，因为它不影响生产率。教育的唯一作用是将高生产率工人和低生产率工人区分开。■

注释

在正文中我们围绕着两种行为的例子进行了讨论，这样的例子虽然简单，但它包含的很多思想却有着广泛的适用性。关于委托代理的文献可参考 Hart and Holmstrom（1987）。Spence（1974）首先将发送信号的情形引入了经济学。Alkerlof（1970）第一个分析了次品市场。带有逆向选择的市场均衡模型可参考 Rothschild and Stiglitz（1976）。Kreps（1990）更为详细地讨论了涉及不对称信息的模型的均衡问题。

习题

25.1 考虑教材中描述的隐藏信息的委托代理问题。令 $f=u^{-1}$。假设 $u(s)$ 是递增且凹的，证明 f 是个递增且凸的函数。

25.2 令 $V(c_a,c_b)$ 表示委托人使用最优激励方案而得到的效用，其中 c_a,c_b 分别表示行动 a 和 b 的成本。推导 $\partial V/\partial c_a$ 和 $\partial V/\partial c_b$，将它们用出现在基本条件中的参数进行表示；使用这些表达式解释这些参数的意思。

25.3 假设 $c_a=c_b$，那么最优激励方案是什么样的？

25.4 假设在委托代理问题中，c_b 降低且其它所有参数维持不变。证明代理人的状况至少与以前一样好。

25.5 假设在隐藏行动的委托代理问题中，代理人是风险中性的。证明最优结果（first-best outcome）能够实现。

25.6 考虑垄断版本的隐藏信息问题。假设两类代理人的成本函数相同，但保留效用水平不同。分析结果将会发生什么样的变化？

25.7 证明一次相交性质的下列含义：如果对于所有 x 都有 $c_2'(x)>c_1'(x)$，那么对于任何两个不同的产出水平 $x_1<x_2$，我们必定有 $c_2(x_2)-c_2(x_1)>c_1(x_2)-c_1(x_1)$。

25.8 在教材中我们断言如果 $c_2(x)>c_1(x)$ 且 $c_2'(x)>c_1'(x)$，那么类型 1 代理人的任何一条无差异曲线和类型 2 代理人的任何一条无差异曲线最多相交一次。证明之。

25.9 考虑教材中描述的隐藏信息的竞争均衡。如果高成本代理人的保留效用水平足够高，那么就可能存在只是用低成本代理人的均衡。那么$\bar{u}_2$硬为多大？

25.10 P 教授雇佣了一个教学助理 A。P 教授关心 A 的教学时间和他支付给 A 的报酬。P 教授希望使得他自己的收益$x-s$最大化，其中x表示 A 的教学时间，s表示他支付给 A 的报酬。如果 A 教学x小时得到的报酬为s，那么 A 的效用为$s-c(x)$，其中$c(x)=x^2/2$。A 的保留效用为零。

（a）如果 P 教授在 A 愿意为他工作的约束下，通过选择x和s来使得自己的效用最大化，求 A 的教学时间。

（b）为了让 A 教那么长的时间，求 P 教授必须支付给 A 的报酬。

（c）假设教授 P 为 A 制定了以下激励方案。教授 P 制定的工资形式为$s(x)=ax+b$，并且让 A 选择教学时间。为了使 P 教授的收益最大，求a和b的值。如果 P 教授使用更一般的函数形式来设计工资方案，他（教授 P）能得到更高的收益吗？

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 26 章：数学

曹乾　译

(东南大学　caoqianseu@163.com)

26 数学

本章简要说明了教材中使用的绝大多数数学工具。如果你忘记了某些定义或者某些重要性质，你可以在本章找到它们。本章不能代替系统的数学教科书，如果你想学习这些教科书，请参见本章结尾处提供的参考文献。

26.1 线性代数

我们用 R^n 表示所有实数 n 元组的集合。非负实数的 n 元组用 R_+^n 表示。这些集合的要素称为点(points)或向量（vectors）。向量通常用粗体字表示。若 $x=(x_1,...,x_n)$ 表示一个向量，我们将它的第 i 个分量表示为 x_i。

两个向量的加法是将量向量各分量相应相加：$x+y=(x_1+y_1,...,x_n+y_n)$。向量的标量乘法（scalar multiplication）是用各分量分别与一个既定的实数 t 相乘：$tx=(tx_1,...,tx_n)$。在几何图形上，向量加法可以按下列办法处理：画出向量 x 和 y，然后再将 y 平移到 x 的尾部；标量乘法可用原向量的 t 倍表示。

向量 x 是包含 n 个向量的集合 A 的**线性组合**（linear combination），若 $x=\sum_{i=1}^n t_i y_i$，其中 $y_i \in A$，t_i 为标量。包含 n 个向量的集合 A 称为**线性无关的**（linearly independent），若不存在集合 (t_i, x_i) 使得 $\sum_{i=1}^n t_i x_i = 0$，其中至少存在一个 $t_i \neq 0$，$x_i \in A$。一个等价的定义是 A 中的向量都不能表示为 A 中向量的线性组合。

给定两个向量，它们的**内积**（inner product）为 $xy=\sum_i x_i y_i$。向量 x 的范数用 $|x|$ 表示，定义为 $|x|=\sqrt{xx}$。注意根据毕达哥拉斯定理，x 的范数是点 x 到原点的距离；也就是说，是向量 x 的长度。

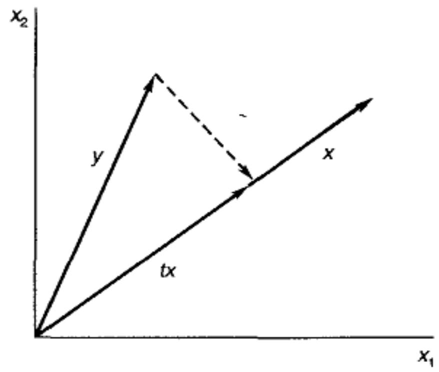


图 26.1：向量 y 在向量 x 上的投影。
该图表明了如何计算投影。

内积的几何解释非常重要，如图 26.1 所示。我们有两个向量 x 和 y，虚线是从 y 的头部画向 x 而且垂直于 x。从原点到虚线与 x 交点的那个向量，称为向量 y 在向量 x 上的投影（projection）。y 在 x 上的投影当然是一个形式为 tx 的向量。我们使用毕达哥拉斯公式计算 t：

曹乾（东南大学 caoqianseu@163.com）

$$|tx|^2+|y-tx|^2=|y|^2$$

$$t^2xx+(y-tx)(y-tx)=yy$$
$$t^2xx+yy-2txy+t^2xx=yy$$
$$txx=xy$$
$$t=\frac{xy}{xx}$$

因此，若我们将 y 投影于 x 上，我们得到了一个向量：该向量的方向和 x 的方向一样，但常度是 x 长度的 xy/xx。若 $xy=0$，我们说 x 和 y 是正交的（orthogonal）。

令 θ 表示向量 x 和 y 之间的夹角。从初等三角学知识可知 $t|x|=|y|\cos\theta$. 如果我们将该表达式和上述 t 的表达式结合起来，可得 $xy=|x||y|\cos\theta$. 因此，若 $\theta=90^o$， $xy=0$；若 $\theta>90^o$，则 $xy<0$；若 $\theta<90^o$，则 $xy>0$。

我们考虑从 R^n 中的向量到 R^m 中向量的映射，将该映射表示为 $f:R^n\to R^m$. 一个映射称为线性函数（linear function），若 $f(tx+sy)=tf(x)+sf(y)$ 对于所有标量 s,t 和所有向量 x,y 均成立。若 $f:R^n\to R^1$ 且其为线性函数，我们将其称为线性函数的(linear functional)。如果 p 是一个线性函数的，我们可以用一个向量 $p=(p_1,...,p_n)$ 代表它，并将其写成 $p(x)=px$。具有 $H(p,a)=\{x:px=a\}$ 形式的点集称为一个超平面。

超平面 $H(p,0)$ 包含了正交于向量 p 的所有向量 x。不难看出，这是一个 $n-1$ 维的集合。超平面 $H(p,a)$ 是 $H(p,0)$ 这个基本的超平面平移后得到的。超平面在经济学中比较重要，因为 $H(p,a)$ 包含了价格为 p 时值为 a 的所有向量 x。

26.2 矩阵的定性（definite）与半定(semidefinite)

令 A 为对称的方阵。于是在 A 的右边乘以某向量 x，并且再在 A 的左边乘以该向量 x 的转置向量，就得到了一个二次型，例如

$$(x_1\quad x_2)\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=a_{11}x_1^2+(a_{21}+a_{12})x_1x_2+a_{22}x_2^2.$$

假设 A 为单位矩阵。在这种情形下，不难看出 x_1 和 x_2 的值，二次型必定是非负的。事实上，若 x_1 和 x_2 不全为零，则 xAx 为严格正。单位矩阵是正定矩阵的一个特例。

定性矩阵（Definite matrices）. 一个方阵 A 为：

（a）正定的，若 $x^tAx>0$ 对于所有 $x\neq 0$ 均成立；

（b）负定的，若 $x^tAx<0$ 对于所有 $x\neq 0$ 均成立；

（c）正半定的，若 $x^tAx\geq 0$ 对于所有 x 均成立；

（d）负半定的，若 $x^tAx\leq 0$ 对于所有 x 均成立。

曹乾（东南大学 caoqianseu@163.com）

在某些情形下，我们不想要求 $x^t Ax$ 对于所有 x 都有确定性的符号，只要求对某些受限的 x 的集有确定性的符号即可。我们说 A 是受约束于 bx=0 的正半定矩阵，若 $x^t Ax > 0$ 对于所有 $x \neq 0$ 且 bx=0 均成立。上述矩阵的其他定义可以自然地推广到受约束的情形。

定性矩阵的检验

若能辨别矩阵何时为负半定或何时正半定，对研究将比较方便。一种有用的必要条件如下：若某个矩阵是正半定的，那么该矩阵的主对角线上的元素不能全为零。证明此事很简单，只要注意到例如若 $x=(1,0,...,0)$，则 $x^t Ax = a_{11}$。

必要条件和充分条件则比较复杂。某个矩阵 A 的**子矩阵**（minor matrices）是指从矩阵 A 中删除 k 行和 k 列形成的矩阵。矩阵 A 的**自然序主子矩阵**（naturally ordered principal minor matrices）或称为**嵌套的主子矩阵**(nested principal minor matrices)是下列形式的矩阵

$$a_{11} \qquad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

等等。某矩阵的子行列式（minor determinants，minors）是指它的子矩阵的行列式。令 $\det A$ 或 $|A|$ 表示矩阵 A 的行列式。

给定一个方阵 A 和一个向量 b，我们可**用 b 给 A 加边**（border A by b）：

$$\begin{pmatrix} 0 & b_1 & \cdots & b_n \\ b_1 & a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots & \\ b_n & a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

这个矩阵称为**加边（的）矩阵**（bordered matrix）。可以将这个概念推广到子矩阵，加边后的子矩阵称为**保留加边的主子矩阵**（border-preserving principal minor matrices）。以矩阵 A 为例，A 的保留加边的主子矩阵是从左上角（包括加边元素）开始的子矩阵。可以证明，使用这些子矩阵的行列式可以方便地确定某个矩阵是正定的还是负定的。

矩阵定性的测试。方阵 A 为：

（a）**正定**，当且仅当它的所有主子矩阵的行列式（以下简称为主子式）均为正；

（b）**负定**，当且仅当它的 k 阶主子式的符号为 $(-1)^k$，其中:k=1,…,n.

（c）bx=0 时为**正半定**，当且仅当它的所有保留加边主子式均为负；

（d）bx=0 时为**负半定**，当且仅当它的所有保留加边主子式的符号为 $(-1)^k$，其

中:k=2,…,n.

注意下列特别的事实：正定矩阵的所有主子式均为正，而有约束条件（bx=0）矩阵的所有保留加边主子式均为负。第 27 章给出了一些例子。

推论。在上面描述的情形中，若某矩阵的自然序主子式满足条件(a)-(d)，则该矩阵的所有主子式满足相应的条件。因此，检验自然序主子式的符号就足以判断矩阵的定性。

26.3 克莱姆法则

下列线性形式的方程有一种简便的求解法则

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

我们可以把这个方程写得更方便一些，即 Ax=b。

克莱姆法则。为了找到找到这个线性方程组解 x 的分量 x_i，可将矩阵 A 的第 i 列换成向量 b，这样就可以形成一个新的矩阵 A_i，则 x_i 等于 A_i 的行列式除以 A 的行列式：

$$x_i = \frac{|A_i|}{A}$$

26.4 实分析的一些内容

给定 R^n 中的一个向量 x 以及一个正实数 e，则圆心在 x、半径为 e 的**开球**（open ball）可定义为 $B_e(x)=\{y\in R^n:|y-x|<e\}$。点集 A 为**开集**（open set），若 A 中的每个 x 都存在 $B_e(x)$ 包含于 A。若 x 为任意集合中的一点，且存在一个 $e>0$ 使得 $B_e(x)$ 包含于 A，则称 x 为 A 的**内点**。

R^n 中的集合 A 的**补集**（complement）包含 R^n 中所有不在 A 中的点；可用 $R^n\setminus A$ 表示 A 的这一补集。

集合 A 为**闭集**（closed set），若 $R^n\setminus A$ 为开集。集合 A 为有界的，若存在 A 中的点 x，并且存在 $e>0$ 使得 A 包含于 $B_e(x)$ 之中。如果 R^n 的一个非空集合既是闭的又是有界的，则称其为**紧集**（compact set）。

R^n 中的某个无穷**序列**（sequence）$(x^i)=(x^1,x^2,...)$，就是一个无穷的点集，该序列中的任何一个正整数都可用上述点集中的一个点表示。称序列 (x^i) **收敛**(converge)到点 x^*，若对于任何 $e>0$，都存在一个整数 m 使得对于所有 $i>m$，x^i 都在 $B_e(x)$ 中。有时我们将上述情形称为 x^i 无限接近于 x^*。我们也将上述情形说为 x^* 是序列 (x^i) 的**极限**（limit），并将

其写为 $\lim_{i\to\infty} x^i = x^*$。若某个序列收敛到一点，我们将其称为**收敛序列**（convergent sequence）。

闭集。集合 A 为闭集，若 A 中的每个序列都收敛到 A 中的点。

紧集。若 A 为紧集，则 A 中的每个序列都有一个收敛子序列。

方程 $f(x)$ 在点 x^* 是**连续的**，若对于收敛到 x^* 的每个序列，都有 $f((x^i))$ 收敛到 $f(x^*)$。若某个函数在其定义域内的每个点上都是连续的，则该函数是一个**连续函数**。

26.5 微积分

微积分综合运用线性代数和数学分析的知识，使用线性函数去近似某些函数。给定一个函数 $f: R \to R$，如果它在某点 x^* 存在下列极限，则将它在该点的导数定义为

$$\frac{df(x^*)}{dx} = \lim_{t\to 0} \frac{f(x^*+t) - f(x^*)}{t}$$

导数 $df(x^*)/dx$ 也可以用 $f'(x^*)$ 表示。若果 f 在点 x^* 存在导数，则称 f 在点 x^* **可微**（differentiable）。

考虑以下线性函数 $F(t)$

$$F(t) = f(x^*) + f'(x^*)t.$$

这个线性函数很好地近似了 f 在点 x^* 的值，因为

$$\lim_{t\to 0} \frac{f(x^*+t) - F(t)}{t} = \lim_{t\to 0} \frac{f(x^*+t) - f(x^*) - f'(x^*)t}{t} = 0.$$

类似地，任意给定一个函数 $f: R^n \to R^m$，则它在 x^* 点的导数 $Df(x^*)$ 可以定义为 R^n 到 R^m 的一个线性映射，该映射很好地近似了 f 在 x^* 的值，因为

$$\lim_{|t|\to 0} \frac{\left| f(x^*+t) - f(x^*) - Df(x^*)t \right|}{|t|} = 0,$$

当然，前提是存在这样的映射。注意，我们使用范数（norm）的符号进行表达，因为上式的分子和分母都是向量。映射 $f(x^*) + Df(x^*)t$ 是 f 在 x^* 点的一个很好地近似，这是由于对于向量 $|t| \to 0$，

$$f(x^*+t) \approx f(x^*) + Df(x^*)t.$$

给定函数 $f: R^n \to R$，我们可以定义 f 关于 x_i 在点 x^* 的偏导数。为做此事，保持除了 x_i 之外的所有其他分量不变，这样 f 只是 x_i 的函数，然后可以计算出这个普通的一维导数。令 $\partial f(x^*)/\partial x_i$ 表示 f 关于 x_i 在点 x^* 的偏导数。

由于 $Df(x^*)$ 是一个线性转换，我们可以用矩阵表示它，这个矩阵是

$$Df(x^*) = \begin{pmatrix} \frac{\partial f_1(x^*)}{\partial x_1} & \cdots & \frac{\partial f_1(x^*)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x^*)}{\partial x_1} & \cdots & \frac{\partial f_m(x^*)}{\partial x_n} \end{pmatrix}$$

这个矩阵称为 $x = x^*$ 时的 **Jacobian 矩阵**。我们通常处理的方程是 $f: R^n \to R$，在这种情形下，$Df(x^*)$ 是一个 $n \times 1$ 矩阵，因此它就是一个向量。

高阶导数

若我们的方程为 $f: R^n \to R$，则该方程的 **Hessian 矩阵**是它的混合偏导数的矩阵

$$D^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right).$$

注意，$D^2 f(x)$ 是一个对称矩阵。

令 $f: R^n \to R$ 是一个可微函数，令 x 和 y 为 R^n 中的向量，则可以证明

$$f(y) = f(x) + Df(z)(y - x)$$

$$f(y) = f(x) + Df(z)(y - x) + \frac{1}{2}(y - x)^t D^2 f(w)(y - x)$$

其中 z 和 w 是 x 和 y 连成线段中的点。这些表达式称为 f 在 x 的**泰勒序列展开式**（Taylor series expansions）。

若 x 和 y 非常接近，而且 f 的导函数是连续的，则 $Df(z)$ 和 $D^2 f(w)$ 分别近似等于 $Df(x)$ 和 $D^2 f(x)$。因此，我们可以将泰勒序列展开式写为

$$f(y) \approx f(x) + Df(x)(y - x)$$

$$f(y) \approx f(x) + Df(x)(y - x) + \frac{1}{2}(y - x)^t D^2 f(x)(y - x).$$

26.6 梯度和切平面

考虑函数 $f: R^n \to R$。f 在点 x^* 的**梯度**（gradient）是一个向量，它的分量是 f 在点 x^* 的偏导数：

$$Df(x^*)=\left(\frac{\partial f(x^*)}{\partial x_1},\dots,\frac{\partial f(x^*)}{\partial x_n}\right).$$

f 在点 x^* 的梯度和 f 在点 x^* 的导数代表的是同一个东西，但在概念上有些区别。导数是 R^n 上的线性函数，而梯度是 R^n 中的向量。由于线性函数可用向量表示，因此它们的外表很相似，尽管它们是不同的东西。在以后的表述中，我们对它们不加区别，用相同的符号表示。

梯度的概念有着重要的集合解释：它指向函数 f 的值增加最快的那个方向。为了看清这一点，令 h 表示范数为 1 的一个向量。f 在沿着方向 h 上的点 x^* 的导数为 $Df(x^*)h$。用内积表示，

$$Df(x^*)h=|Df(x^*)|\cos\theta,$$

这个式子的值显然当 $\theta=0$ 时达到最大，也就是说当向量 $Df(x^*)$ 和 h 共线（collinear）时。

某个函数的**水平集**（level set）是使得该函数的值为常数的所有 x 组成的集合：$Q(a)=\{x:f(x)=a\}$。可微函数 $f:R^n\to R$ 的水平集通常为 $n-1$ 维的曲面。

函数 $f:R^n\to R$ 的**上轮廓集**（upper contour set）是使得该函数的值至少和某个数一样大的所有 x 组成的集合：$U(a)=\{x:f(x)\geq a\}$.

下面我们希望找到，水平集在某点 x^* 的**切超平面**（tangent hyperplane）的表达式。我们知道线性映射 $f(x^*)+Df(x^*)(x-x^*)$ 可以很好地近似 f 在点 x^* 的值。因此，$\{x:f(x)=a\}$ 的最佳线性近似应为 $H(a)=\{x:f(x^*)+Df(x^*)(x-x^*)=a\}$。由于 $f(x^*)=a$，切超平面的表达式如下：

$$H(a)=\{x:Df(x^*)(x-x^*)=0\} \qquad (26.1)$$

超平面（Hyperplane）。若 x 是切超平面内的一个向量，则向量 $x-x^*$ 在点 x^* 正交于（垂直于）f 的梯度。

这个结论可以直接由（26.1）式推出，它也非常符合直觉。沿着曲面 $Q(a)$，函数 f 的值是常数。因此，$f(x)$ 在这些方向上的导数应为零。

26.7 极限

在教材中，我们需要使用 L'Hopital 法则计算某些点的极限。假设我们试图计算分式 $f(x)/g(x)$ 在 $x\to 0$ 时的极限，但是 $f(0)=g(0)=0$，因此该分式在 $x=0$ 处无定义。然而，如果 f 和 g 都是可微的，而且若 $g'(0)\neq 0$，L'Hopital 法则表明

$$\lim_{x\to 0}\frac{f(x)}{g(x)}=\frac{f'(x)}{g'(x)}$$

因此，比值的极限等于它们导数的比值。

26.8 齐次函数

函数 $f:R_+^n \to R$ 称为 k 次**齐次的**（homogeneous），若对于所有 $t>0$ 都有 $f(tx)=t^k f(x)$。两个最重要的情形是当 $k=0$ 和 $k=1$ 时。如果我们将零次齐次函数的所有自变量都变为原来的 2 倍，那么该函数的函数值不会改变。如果我们将一次齐次函数的所有自变量都变为原来的 2 倍，那么该函数的函数值也变为原来的 2 倍。

欧拉定理（Euler's law）。若 f 为一次齐次函数且可微，则

$$f(x)=\sum_{i=1}^{n}\frac{\partial f(x)}{\partial x_i}x_i .$$

证明时需要注意到根据定义有 $f(tx)\equiv tf(x)$ 即可。$f(tx)\equiv tf(x)$ 的两边对 t 微分可得

$$\sum_{i=1}^{n}\frac{\partial f(tx)}{\partial(tx_i)}x_i = f(x)$$

在上式中令 $t=1$ 即可得到最终结果。

齐次性（Homogeneity）。如果 $f(x)$ 是 $k\ge 1$ 次齐次的，则 $\partial f(x)/\partial x_i$ 是 $k-1$ 次齐次的。

为了看清这一点，恒等式 $f(tx)=t^k f(x)$ 两边对 x_i 微分：

$$\frac{\partial f(tx)}{\partial(tx_i)}t=t^k\frac{\partial f(x)}{\partial x_i}.$$

两边同除以 t 即可得到想要的结果。

上述结论的一个重要应用是：沿着从原点出发的射线上，齐次函数的水平集的斜率的斜率为常数：对于所有 $t>0$，都有

$$\frac{\dfrac{\partial f(tx)}{\partial x_i}}{\dfrac{\partial f(tx)}{\partial x_j}}=\frac{\dfrac{\partial f(x)}{\partial x_i}}{\dfrac{\partial f(x)}{\partial x_j}} \qquad (26.2)$$

然而，有一类函数，尽管它们不是齐次函数，但仍具有上述性质。某函数称为**位似**（homothetic）**函数**，若它是由一次齐次函数的正单调变换得到的。也就是说，位似函数可以写为 $f(x)=g(h(x))$ 的形式，其中 $h(x)$ 是一次齐次函数。不难证明，位似函数满足条件（26.2）。

26.9 仿射函数

如果某个函数可以写成 $f(x)=a+bx$ 的形式，则称其为**仿射函数**（affine function）。仿射函数有时称为线性函数，但是，严格来说，只有当 $a=0$ 时，才可以这么说。

显然，可微函数为仿射函数当且仅当 $f''(x)=0$。下面是个应用：函数满足条件 $f(pu+(1-p)v)\equiv pf(u)+(1-p)f(v)$（其中 $0\le p\le 1$）当且仅当 $f(u)$ 是仿射函数。证明时，只要将上述等式两端对 p 求导即可，这样可得到

$$f'(pu+(1-p)v)(u-v)\equiv f(u)-f(v)$$

上式再对 p 求导可得 $f''(pu+(1-p)v)(u-v)\equiv 0$。

26.10 凸集

R^n 中的点集 A 为**凸集**(convex set)，若 x 和 y 均在 A 中蕴涵着 $tx+(1-t)y$ 也在 A 中，其中 $0\le t\le 1$。点集 A 称为严格凸（strictly convex），若 $tx+(1-t)y$ 在 A 的内部，其中 $0<t<1$。

我们经常要用到凸集之和仍为凸集的事实，现在证明这个结论。令 A_1 和 A_2 是两个凸集，令 $A=A_1+A_2$。再令 x 和 y 是 A 中的两个点。根据定义 $x=x_1+x_2$，其中 x_1 在 A_1 中，x_2 在 A_2 中；y 的情形类似。由此可得

$$tx+(1-t)y=t(x_1+x_2)+(1-t)(y_1+y_2)=[tx_1+(1-t)y_1]+[tx_2+(1-t)y_2]$$

上式方括号内的项分别在 A_1 和 A_2 之中，因为每项都是一个凸集。因此 $tx+(1-t)y$ 也在 A 中，这就证明了 A 是一个凸集。

26.11 分离超平面

令 A 和 B 是 R^n 中两个不相交的凸集，即它们的交集为空集。显然我们可以找到一个超平面将这两个集合“分离”开。也就是说，A 和 B 分别位于这个超平面的两侧。正式地，我们给出以下定理：

分离超平面定理（Separating hyperplane theorem）。若 A 和 B 是 R^n 中两个非空的、不相交的凸集，则存在一个线性函数 p 使得对于所有 A 中的 x 和 B 中的 y 都有 $px\ge py$。

26.12 偏微分方程

偏微分方程组（a system of partial differential equations）是下列形式的方程组

$$\frac{\partial f(p)}{\partial p_i}=g_i(f(p),p) \qquad i=1,\dots,n$$
$$f(q)=0.$$

后面一个方程称为**边界条件**（boundary condition）。一般的偏微分方程组远比上述方程组复杂，但这个方程组对于我们的目的来说已经够用了。

偏微分方程组的一个解是同时满足各个方程的一个函数 $f(p)$。偏微分方程组解存在的

必要条件来自交叉偏导数（cross-partial derivatives）的对称性：

$$\frac{\partial g_i}{\partial f}\frac{\partial f}{\partial p_j}+\frac{\partial g_i}{\partial p_j}=\frac{\partial^2 f(p)}{\partial p_i \partial p_j}=\frac{\partial^2 f(p)}{\partial p_j \partial p_i}=\frac{\partial g_i}{\partial f}\frac{\partial f}{\partial p_i}+\frac{\partial g_i}{\partial p_i}$$

由此可知

$$\frac{\partial g_i}{\partial f}\frac{\partial f}{\partial p_j}+\frac{\partial g_i}{\partial p_j}=\frac{\partial g_i}{\partial f}\frac{\partial f}{\partial p_i}+\frac{\partial g_i}{\partial p_i}$$

是这个偏微分方程组存在**局部解**（local solution）的必要条件。这个条件称为**可积条件**（integrability condition）。方程组存在全局解的必要条件比较复杂，通常取决于定义域的拓扑性质。

偏微分方程组的显性解通常难以求出，但有一种特殊的情形例外。这个情形就是 $f(p)$ 不是显性地存在于方程组的右侧。这类方程组求解方法是积分。

例如，当方程组为下列情形时，请求解。

$$\begin{aligned}&\frac{\partial f(p_1,p_2)}{\partial p_1}=g_1(p_1,p_2)\\&\frac{\partial f(p_1,p_2)}{\partial p_2}=g_2(p_1,p_2)\\&f(q_1,q_2)=0.\end{aligned}$$

如果满足下列可积条件

$$\frac{\partial g_1(p_1,p_2)}{\partial p_2}=\frac{\partial g_2(p_1,p_2)}{\partial p_1}$$

则可以证明，这个方程组的解为

$$f(p_1,p_2)=\int_{q_1}^{p_1} g_1(t,q_2)dt+\int_{q_2}^{p_2} g_2(p_1,t)dt\,.$$

显然，这个方程满足边界条件，而且通过微分可以发现$\partial f/\partial p_2=g_2(p_1,p_2)$。我们只需要验证$\partial f/\partial p_1=g_1(p_1,p_2)$。

求导可得

$$\frac{\partial f(p_1,p_2)}{\partial p_1}=g_1(p_1,q_2)+\int_{q_2}^{p_2}\frac{\partial g_2(p_1,t)}{\partial p_1}dt$$

使用可积条件

$$\begin{aligned}\frac{\partial f(p_1,p_2)}{\partial p_1}&=g_1(p_1,q_2)+\int_{q_2}^{p_2}\frac{\partial g_2(p_1,t)}{\partial p_1}dt\\&=g_1(p_1,q_2)+g_1(p_1,p_2)-g_1(p_1,q_2)\\&=g_1(p_1,p_2).\end{aligned}$$

26.13 动态系统

系统的状态包括对影响该系统行为所有变量的描述。一个系统的状态空间包含所有可行的状态。例如，某个特定经济系统的状态空间包含所有可能的价格组合，或者所有可能的价格和消费束组合。

令 S 表示状态空间，则我们可用函数 $F: S \times R \to S$ 描述动态系统。实直线 R 可以解释为时间；如果某系统在时间 0 的状态为 x，则它在时间 t 的状态可用 $F(x,t)$ 表示。

通常状态函数 F 不是以显性形式给出，而是用一组微分方程隐性地给出。例如，$x = f(x(t))$ 是个微分方程，它告诉我们当系统在状态 $x(t)$ 时的 x 的变动率。如果 $f(x)$ 满足一定的正则条件，可以证明由 f 定义的动态系统是唯一的。

微分方程组 $\mathbf{x} = f(\mathbf{x})$ 的解是个函数 $\mathbf{x}: \mathbb{R} \to \mathbb{R}^n$，它使得对于所有 t 都有 $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$。解有时也称为解曲线、轨线（trajectory）、轨道（orbit）等。

动态系统的一个均衡是满足 $f(\mathbf{x}^*) = 0$ 的一种状态 $\mathbf{x}^*$。粗略地说，如果动态系统一旦进入稳定状态，它就永远维持在这种状态上。

假设给定某个动态系统 $\mathbf{x} = f(\mathbf{x})$，而且在零时刻我们在某个任意状态 $\mathbf{x}_0$ 上。我们通常对系统状态何时移动到某个均衡 $\mathbf{x}^*$ 感兴趣。我们说一个系统是全局稳定的（globally stable），如果 $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$ 对于所有初始值 $\mathbf{x}_0$ 成立。我们需要指出，全局稳定均衡必定是唯一的。

有个非常方便的标准能确定动态系统何时是全局稳定的。我们只研究状态空间 S 为紧的（compact）情形，因此我们知道这个系统总是处于有界区域。假设我们能够找到某个微分方程 $V: S \to \mathbb{R}$，它具有下列两个性质：

（1）V 在 $\mathbf{x}^*$ 处达到最小值。

（2）对于所有 $\mathbf{x}(t) \neq \mathbf{x}^*$，有 $\dot{V}(\mathbf{x}(t)) < 0$。也就是对于所有 $\mathbf{x}(t) \neq \mathbf{x}^*$，有 $DV(\mathbf{x}) f(\mathbf{x}(t)) < 0$。

这样的函数称为李雅普诺夫函数（Liaponov function）。

李雅普诺夫定理。如果对某个动态系统我们能找到一个李雅普诺夫函数，那么唯一的均衡 $\mathbf{x}^*$ 是全局稳定的。

26.14 随机变量

给定一个随机变量 $\tilde{R}_a$，它取值 R_{as} 的概率为 π_s（其中 $s = 1,...,S$），我们定义这个随机变量的期望为

$$E\tilde{R}_a = \bar{R}_a = \sum_{s=1}^{S} R_{as}\pi_s .$$

我们把两个随机变量 $\tilde{R}_a$ 和 $\tilde{R}_b$ 的**协方差**（covariance）定义为

$$\operatorname{cov}(\tilde{R}_a, \tilde{R}_b) = \sigma_{ab} = E(R_a - \bar{R}_a)(R_b - \bar{R}_b).$$

随机变量 $\tilde{R}_a$ 的**方差**（variance）为

$$\sigma_a^2 = \sigma_{aa} = \operatorname{cov}(\tilde{R}_a, \tilde{R}_a).$$

令 x_a、x_b 和 c 为非随机的。那么，给定上述定义，下列事实可由直接计算推导出：

和的期望。$E(x_a\tilde{R}_a + x_b\tilde{R}_b) = x_a E\tilde{R}_a + x_b E\tilde{R}_b = x_a\bar{R}_a + x_b\bar{R}_b$. 更一般地，

$$E\sum_{a=1}^{A} x_a\bar{R}_a = \sum_{a=1}^{A} x_a\bar{R}_a.$$

协方差恒等式。$\operatorname{cov}(\tilde{R}_a, \tilde{R}_a) = E\tilde{R}_a\tilde{R}_a - \bar{R}_b\bar{R}_b$.

和的协方差。$\operatorname{cov}(x_a\tilde{R}_a + c\tilde{R}_b) = x_a(\tilde{R}_a, \tilde{R}_b)$.

和的方差。

$$\begin{aligned}\operatorname{var}(x_a\tilde{R}_a + x_b\tilde{R}_b + c) &= x_a^2\operatorname{var}(\tilde{R}_a) + 2x_a x_b\operatorname{cov}(\tilde{R}_a, \tilde{R}_b) + x_b^2\operatorname{var}(\tilde{R}_b)\\ &= x_a^2\sigma_{aa} + 2x_a x_b\sigma_{ab} + x_b^2\sigma_{bb}.\end{aligned}$$

更一般地，

$$\operatorname{var}\left(\sum_{a=1}^{A} x_a\tilde{R}_a\right) = \sum_{a=1}^{A}\sum_{b=1}^{A} x_a x_b\sigma_{ab}.$$

注释

有很多经济数学的教材更为详细地考察了本章的内容。比如，Binmore（1982）、Binmore（1983）和 Blume and Simon（1991）。

曹乾●经济学译丛精品系列

Microeconomics Analysis

（3th edition）

Hal. R. Varian

（University of Michigan Ann Arbor）

范里安

微观经济分析（第 3 版）

完美中文翻译版

第 27 章：最优化

曹乾　译

(东南大学　caoqianseu@163.com)

27 最优化

本章的目的是复习一下最优化的相关内容。一般来说，学生在学习本教材之前，已花费几个星期学习最优化的方法。

27.1 单变量的最优化

令 $f: R \to R$ 为一个单变量函数。如果对于所有 x 都有 $f(x^*) \geq f(x)$ 我们说这个函数在 x^* 达到最大值。若对于所有 $x \neq x^*$ 都有 $f(x^*) > f(x)$，则我们说 x^* 是一个严格极大值点。类似地，如果对于所有 x 都有 $f(x^*) \leq f(x)$ 我们说这个函数在 x^* 达到最小值。若对于所有 $x \neq x^*$ 都有 $f(x^*) < f(x)$，则我们说 x^* 是一个严格极小值点。

注意，$f(x)$ 关于 x 的最大化问题，等价于 $-f(x)$ 关于 x 的最小化问题。

一阶条件和二阶条件

假设一个可微[一]函数 f 在在 x^* 达到最大值，则根据初级微积分可知，f 的一阶导数在 x^* 的值必定等于零，f 的二阶导数在 x^* 的值必定小于或等于零。这两个条件分别称为一阶条件和二阶条件，可用数学表达式表示：

$$f'(x^*) = 0$$
$$f''(x^*) \leq 0.$$

注意，这些条件是必要条件而非充分条件；若 R 中的一点是最大化问题的解，则它必定满足这些条件，但是满足这些条件的点未必是利润最大化问题的解。

对于最小化问题，一阶条件和最大化问题的一阶条件相同，但是二阶条件变为 $f''(x^*) \geq 0.$

例子：一阶条件和二阶条件

考虑函数 $f(x,a) = \ln x - ax$。关于 x 的最大化一阶条件为 $1/x - a = 0$，二阶条件为 $-1/x^2 \leq 0$。我们可以看到，二阶条件自动满足，因此可知最大值点 x 为 $x^* = 1/a$。

现在令 $g(x,b) = u(x) - bx$。在这个例子中，我们无法解出最大值点 x^* 的显性表达式，但是我们知道它必定满足以下两个条件：$u'(x) = b$；$u''(x) \leq 0$。

[一] 在本教材中，我们采用“索罗惯例（Solow’s conventions）:“每个函数都是可微的，而且其可微分的阶数总是比我们需要的高一阶。”

凹性

单变量方程为**凹的**（concave），若对于所有 x 和 y 以及所有 $0 \le t \le 1$ 都有

$$f(tx+(1-t)y) \ge tf(x)+(1-t)f(y).$$

如图 27.1 所示。凹函数的一个性质是 x 和 y 加权平均数的函数值即 $f(tx+(1-t)y)$，大于或等于函数值 $f(x)$ 和 $f(y)$ 的加权平均数即 $tf(x)+(1-t)f(y)$。

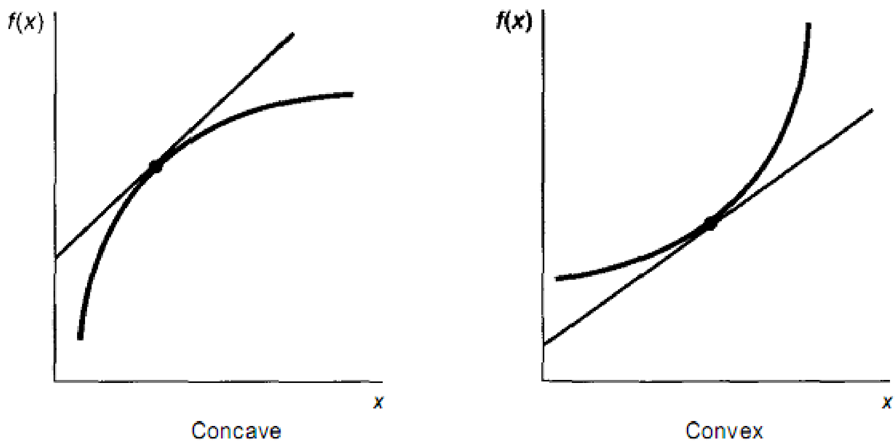


图 27.1：凹函数和凸函数

如果 f 是可微的，则 f 为凹当且仅当 $f''(x) \le 0$ 对于所有 x 成立。例如，$\ln x$ 是个凹韩式，因为它的二阶导数总是小于或等于零。

一个方程为严格凹的，若对于所有 x 和 y 以及所有 $0 \le t \le 1$ 都有

$$f(tx+(1-t)y) > tf(x)+(1-t)f(y).$$

若对于所有 x 都有 $f''(x) < 0$，则 f 为严格凹的，但是反过来说（即逆命题）非真。例如 $f(x) = -x^4$ 为严格凹，但 $f''(0) = 0$.

凹函数的另外一个性质是对于所有 x 和 y，都有

$$f(x) \le f(y) + f'(y)[x-y] \qquad (27.1)$$

线性函数增长的速率是不变的。大致来说，凹函数在每一点的增长速率，均小于与该点相切的线性函数的增长速率。

若 f 为凹，则它的二阶导数在每一点的值都小于或者等于零。由此可知，若 $f'(x^*) = 0$，则 x^* 是该函数的最大值点。最简单的证明方法是将 $y = x^*$ 和 $f'(x^*) = 0$ 代入（27.1）式，这样可发现对于所有 x 都有

$$f(x) \le f(x^*) + f'(x^*)[x - x^*] = f(x^*).$$

但这个不等式表明 x^* 是函数 f 的最大值点。因此，对于凹函数来说，一阶条件既是必要条件也是充分条件。

凸函数（convex function）满足下列性质

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y).$$

注意，若 $f(x)$ 为凸，则 $-f(x)$ 为凹。由这个结论可得到下列事实：

凹函数和凸函数。凹函数增长速度，小于和它在某点相切的线性函数的增长速度；而凸函数的增长速度则比线性函数快。

1）若 f 为凸函数，则对于所有 x 都有 $f''(x) \ge 0$。

2）若 f 为凸函数，则 $f(x) \ge f(y) + f'(y)[x - y]$。

3）若 f 为凸函数，且 $f'(x^*) = 0$，则 f 在 x^* 点达到最小值。

包络定理

假设 $f(x,a)$ 是变量 x 和 a 的函数。我们一般将 a 解释为模型以外的参数（外生变量），而将 x 视为我们想要研究的变量（内生变量）。假设我们选择 x 使得该函数的值达到最大，则对于 a 的每个不同值，使得函数值最大的 x 的值一般也不同。在充分规则的情形下，我们能够写出函数 $x(a)$，这个函数给出了在不同的 a 值下能使函数值达到最大的 x 的值。例如，在某些经济问题中，选择变量可能为商品的消费量或生产量，而参数 a 为商品价格。

我们还可以定义（最优）值函数（value function），$M(a) = f(x(a),a)$. 这个函数告诉我们对于不同的 a 值，函数 f 的最优值是多少。

例子：值函数

我们在上面已经知道，对于函数 $f(x,a) = \ln x - ax$，x 的最优值为 $x(a) = 1/a$。因此，该问题的值函数为 $M(a) = \ln(1/a) - a/a = -\ln a - 1$.

再例如，对于函数 $g(x,b) = u(x) - bx$，我们有 $M(b) = u(x(b) - bx(b)$.

在经济学中，我们通常对下列问题感兴趣：当参数 a 改变时，最优值如何改变。可以证明，有种简单的办法可以计算这种改变。根据定义可知

$$M(a) \equiv f(x(a),a).$$

这个恒等式的两端对 a 求导，可得

$$\frac{dM(a)}{da}=\frac{\partial f(x(a),a)}{\partial x}\frac{\partial x(a)}{\partial a}+\frac{\partial f(x(a),a)}{\partial a}$$

由于 $x(a)$ 是 f 最大值问题的解，我们知道

$$\frac{\partial f(x(a),a)}{\partial x}=0.$$

将其代入前面的表达式，我们有

$$\frac{dM(a)}{da}=\frac{\partial f(x(a),a)}{\partial a} \tag{27.2}$$

上式的一种更好的写法是

$$\frac{dM(a)}{da}=\left.\frac{\partial f(x(a),a)}{\partial a}\right|_{x=x(a)}$$

上述表示方法的好处是，可以提醒我们注意，求导时要将 x 固定在 $x(a)$，也就是说求 $x=x(a)$ 时的导数值。

换句话说：值函数关于参数的全微分 $dM(a)/da$，等于在最优选择时（即 $x=x(a)$）的 f 关于参数的偏导数 $\partial f(x(a),a)/\partial a$。这种陈述是包络定理的最简单表达方式。需要考虑一下为什么会出现这样的事情。当参数 a 变化时产生两种效应：一是 a 的变化直接引起 f 值的变化；二是 a 的变化引起 x 值的变化，而 x 的变化又会引起 f 值的变化。但是如果我们选择的 x 是最优的，则 x 的微小变化对 f 值的影响为零（$\partial f/\partial x=0$），因此间接效应就消失了，只剩下直接效应。

例子：包络定理

继续使用前面用过的例子 $f(x,a)=\ln x-ax$，在前面我们已知道 $M(a)=-\ln a-1$。因此，$M'(a)=-1/a$。我们也可以使用包络定理得到这个结论；直接计算 $\partial f(x,a)/\partial a$ 可知，$\partial f(x,a)/\partial a=-x$。令 x 等于它的最优值 $x(a)$，可得 $df(x,a)/da=-1/a=M'(a)$。

对于另外一个例子即 $M(b)=g(x(b),b)=u(x(b))-bx(b)$，我们有 $M'(b)=-x(b)$。

比较静态

在经济学中，我们还关注当某参数变化时，最优选择如何变化。这类分析称为比较静态分析（comparative statics analysis）或敏感性分析。基本计算如下。我们知道最优选择函数 $x(a)$ 必然满足条件

$$\frac{\partial f(x(a),a)}{\partial x}\equiv 0.$$

对上述恒等式两端进行微分可得

$$\frac{\partial^2 f(x(a),a)}{\partial x^2}\frac{dx(a)}{da}-\frac{\partial^2 f(x(a),a)}{\partial x\partial a}\equiv 0.$$

解出 $dx(a)/da$ 可得

$$\frac{dx(a)}{da}=-\frac{\partial^2 f(x(a),a)/\partial x\partial a}{\partial^2 f(x(a),a)/\partial x^2}.$$

根据最大化的二阶条件可知，上式右侧**分母**的符号为负。由于这个事实并且注意上式右侧前面的符号，因此可以断言：

$$\frac{dx(a)}{da}\text{的符号与}\frac{\partial^2 f(x(a),a)}{\partial x\partial a}\text{的符号相同}$$

因此，最优选择关于参数 a 的导数的符号，仅取决于目标函数对 x 和 a 的二阶交叉偏导数的符号。

这个结论的妙处在于我们不需要每次都重复计算这个结果；我们只要使用交叉偏导数传达的信息即可。

例子：比较静态分析

若 $f(x,a)=\ln x-ax$，在前面我们已知道 $x(a)=1/a$。该式直接对 a 求导可知 $x'(a)=-1/a^2<0$。但是我们原本不需要求解这个利润最大化问题，因为在比较静态分析时，我们有时只想知道 $x'(a)$ 的符号。事实上，如果注意到

$$\frac{\partial^2 f(x,a)}{\partial x\partial a}=-1<0.$$

那么利用 $\frac{dx(a)}{da}$ 的符号与 $\frac{\partial^2 f(x(a),a)}{\partial x\partial a}$ 的符号相同这个结论，我们立即可知 $x'(a)<0$。

类似地，对于目标函数为 $g(x,b)=u(x)-bx$，我们立即可知，$x'(b)$ 的符号与 $\frac{\partial^2 g(x,b)}{\partial x\partial b}=-1$ 的符号相同，因此 $x'(b)<0$。

上面这个例子太妙了：我们对函数 $u(x)$ 的形状几乎一无所知，但是我们仍然能确定当参数变动时最优选择如何变动。你已经注意到了，我们的方法是使用了上述目标函数为拟线性函数的性质。在教材中，我们将研究更多的此类例子。

对于最小化问题，分析的方法类似，只不过在最小化问题中下式右侧分母的符号改变了：

$$\frac{dx(a)}{da}=-\frac{\partial^2 f(x(a),a)/\partial x\partial a}{\partial^2 f(x(a),a)/\partial x^2}.$$

因为最小化的二阶条件意味着目标函数关于选择变量的导数为正，所以我们可以得出以下结论：选择变量对参数的导数的符号与交叉偏导数的符号相反，即

$$\frac{dx(a)}{da}\text{的符号与}\frac{\partial^2 f(x(a),a)}{\partial x\partial a}\text{的符号相反。}$$

27.2 多变量的最大化

下面我们分析更为复杂的最大化问题。假设选择变量有两个：x_1 和 x_2。通常我们将这两个变量写成一个向量的形式，即 $x=(x_1,x_2)$.

在这种情形下，我们将最大化问题写为

$$\max_{x_1,x_2} f(x_1,x_2),$$

或者，更一般地

$$\max_{x} f(x).$$

一阶条件和二阶条件

多变量最大化问题的一阶条件是，目标函数对每个选择变量的偏导数必定都为零。如果只有两个选择变量，则一阶条件（即必要条件）为

$$\frac{\partial f(x_1,x_2)}{\partial x_1}=0$$

$$\frac{\partial f(x_1,x_2)}{\partial x_2}=0.$$

如果有 n 个选择变量，那么使用梯度向量进行表示更简便些，即将 Df(x)定义为

$$Df(x)=\left(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_n}\right).$$

使用这种表示方法，可将 n 个变量的最大化问题的一阶条件写为

$$Df(x^*)=0.$$

这个式子是说，在最优选择 x^* 处，偏导数向量必定等于零向量。

两个选择变量最大化问题的二阶条件，通常用目标函数关于选择变量的二阶偏导数组

成的矩阵表示。这个矩阵称为海赛矩阵，它的形式如下

$$H=\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

其中，f_{ij} 代表 $\partial^2 f/\partial x_i \partial x_j$.

由微积分的知识可知，在最优选择 x^* 处，海赛矩阵必须为负半定。这意味着对于任何向量 (h_1,h_2)，我们必定有

$$(h_1,h_2)\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}\leq 0.$$

更一般地，若令 h 为列向量，令 h^t 表示 h 的转置（向量）。那么矩阵为负半定的条件就可以写为

$$h^t Hh \leq 0.$$

如果我们分析的问题是最小化问题，而不是最大化问题，那么它们的一阶条件是相同的，但是最小化问题的二阶条件变为要求海赛矩阵为正半定。

比较静态分析

假设我们想确定最优选择函数对参数 a 的变动是如何反应的。我们知道，最优选择必须满足下列一阶条件：

$$\frac{\partial f(x_1(a),x_2(a),a)}{\partial x_1}=0$$

$$\frac{\partial f(x_1(a),x_2(a),a)}{\partial x_2}=0.$$

将上述两个式子分别对参数 a 求导可得

$$f_{11}\frac{\partial x_1}{\partial a}+f_{12}\frac{\partial x_2}{\partial a}+f_{13}=0$$

$$f_{21}\frac{\partial x_1}{\partial a}+f_{22}\frac{\partial x_2}{\partial a}+f_{23}=0.$$

上面两个式子可以一并使用矩阵形式表示

$$\begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}\begin{pmatrix} x_1'(a) \\ x_2'(a) \end{pmatrix}=\begin{pmatrix} -f_{13} \\ -f_{23} \end{pmatrix}.$$

如果上式左侧的矩阵是可逆的，则这个方程组的解为

$$\begin{pmatrix} x_1'(a) \\ x_2'(a) \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}^{-1} \begin{pmatrix} -f_{13} \\ -f_{23} \end{pmatrix}.$$

事实上，我们一般使用克莱姆法则（Cramer’s rule）求解而不是使用矩阵的逆求解。

例如，如果我们想解出$\partial x_1/\partial a$，我们可以运用克莱姆法则，将这个解表示为两个行列式的比值：

$$\frac{\partial x_1}{\partial a} = \frac{\begin{vmatrix} -f_{13} & f_{12} \\ -f_{23} & f_{22} \end{vmatrix}}{\begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix}}$$

根据最大化问题的二阶条件可知，上式右侧分母中的矩阵为负半定矩阵。基本线性代数知识告诉我们这个矩阵的行列式必定为正。因此，$\partial x_1/\partial a$的符号取决于上式分子中的行列式的符号。

例子：比较静态分析

令$f(x_1, x_2, a_1, a_2) = u_1(x_1) + u_2(x_2) - a_1x_1 - a_2x_2$. 这个函数的最大化问题的一阶条件为：

$$u_1'(x_1^*) - a_1 = 0$$

$$u_2'(x_2^*) - a_2 = 0.$$

二阶条件是矩阵

$$H = \begin{pmatrix} u_1''(x_1^*) & 0 \\ 0 & u_2''(x_2^*) \end{pmatrix}$$

为负半定的。由于负半定的矩阵的对角线上的项必定小于或等于零，由此可知$u_1''(x_1^*) \le 0$和$u_2''(x_2^*) \le 0$。

最大化的值函数为

$$M(a_1, a_2) = \max_{x_1, x_2} \; u_1(x_1) + u_2(x_2) - a_1x_1 - a_2x_2.$$

使用包络定理简单计算一下即可知道

$$\frac{\partial M}{\partial a_1} = -x_1^*$$

$$\frac{\partial M}{\partial a_2} = -x_2^*.$$

由前面的比较静态计算结果可知:

$$\frac{\partial x_1}{\partial a_1}\text{的符号与}\begin{pmatrix}1 & 0 \\ 0 & u_2''(x_2^*)\end{pmatrix}\text{的符号相同.}$$

计算上述行列式的值可知

$$\frac{\partial x_1}{\partial a_1} \le 0.$$

注意，即使我们一点也不知道效用函数 u_1 或 u_2 的具体表达式，我们也能确定选择变量如何随参数的变动而变动；我们只需要知道目标函数的**结构**(structure)，例如，在上例中，目标函数是**加法可分离的**（additively separable）。

凸性和凹性

函数 $f : R^n \to R$ 为**凸的**（concave），若

$$f(tx + (1-t)y) \ge tf(x) + (1-t)f(y)$$

对于所有 x 和 y 以及所有 $0 \le t \le 1$ 都成立。上式的解释和单变量的解释是相同的；也就是说，函数 f 在 x 和 y 加权平均数之处的函数值，至少与 $f(x)$ 和 $f(y)$ 的加权平均和一样大。

可以证明凹函数必定满足下列不等式

$$f(x) \le f(y) + Df(y)[x-y].$$

在二维情形下，我们可以将上式写为

$$f(x_1,x_2) \le f(y_1,y_2) + \begin{pmatrix}\frac{\partial f(y_1,y_2)}{\partial x_1} & \frac{\partial f(y_1,y_2)}{\partial x_1}\end{pmatrix}\begin{pmatrix}x_1-y_1 \\ x_2-y_2\end{pmatrix},$$

或者，将上述不等式右侧第二项相乘展开

$$f(x_1,x_2) \le f(y_1,y_2) + \frac{\partial f(y_1,y_2)}{\partial x_1}[x_1-y_1] + \frac{\partial f(y_1,y_2)}{\partial x_2}[x_2-y_2].$$

上式是一维条件的自然推广。

对于凹性的二阶导数条件也有一个美妙的推广。我们已经知道，在一维情形下，凹函数的二阶导数必定小于或等于零。在多维情形下，凹性的条件是由二阶导数组成的矩阵在每一点都是负半定的。从几何上来说，这意味着凹函数的图形必定在每个方向上都“弯曲地离开”（curve away）它的切平面。也就是说，凹函数自动满足最大化问题的二阶条件。类似地，凹函数的海赛矩阵必定是正半定的。

如果某函数的海赛矩阵在每一点都是负半定的，则该函数必定是严格凹的。然而，它

的逆并不成立：严格凹函数的海赛矩阵在某些点可能是奇异的(singular)。例如，在一维情形下，考虑函数$-x^4$在点$x=0$的情形。

拟凹函数和拟凸函数

函数$f:R^n \to R$是拟凹的（quasiconcave）的，若该函数的上轮廓集（upper contour sets）是凸集。换句话说，集合$\{x\in R^n: f(x)\ge a\}$对于任意a都是凸集。函数$f(x)$是拟凸的（quasiconvex），若$-f(x)$是拟凹的。

27.3 约束最大化

考虑下列形式的约束最大化（constrained maximization）问题

$$\max_{x_1,x_2} \quad f(x_1,x_2)$$

$$\text{使得 } g(x_1,x_2)=0.$$

为了阐述这个问题的一阶条件和二阶条件，最好使用拉格朗日函数：

$$L(\lambda,x_1,x_2)=f(x_1,x_2)-\lambda g(x_1,x_2).$$

一阶条件要求拉格朗日函数对它的各个自变量的导数等于零：

$$\frac{\partial L}{\partial x_1}=\frac{\partial f}{\partial x_1}-\lambda\frac{\partial g}{\partial x_1}=0$$

$$\frac{\partial L}{\partial x_2}=\frac{\partial f}{\partial x_2}-\lambda\frac{\partial g}{\partial x_2}=0$$

$$\frac{\partial L}{\partial \lambda}=-g(x_1,x_2)=0.$$

未知变量有三个x_1,x_2和λ，方程数量也有三个；通常你可以解出这个最优化问题的方程组。

n维最优化问题的结构和上面的最优化问题相同。这个最大化问题是

$$\max_{x} \quad f(x)$$

$$\text{使得 } g(x)=0.$$

因此，拉格朗日函数变为

$$L(\lambda,x)=f(x)-\lambda g(x).$$

一阶条件有$n+1$个，它们的形式为

$$\frac{\partial L}{\partial x_i}=\frac{\partial f(x)}{\partial x_i}-\lambda\frac{\partial g(x)}{\partial x_i}=0\text{，其中 } i=1,\ldots,n.$$

$$\frac{\partial L}{\partial \lambda}=-g(x_1,x_2)=0.$$

二阶条件要使用拉格朗日函数的海赛矩阵进行表达。在二维问题中，这就是

$$D^2L(x_1,x_2)=\begin{pmatrix}\frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1\partial x_2}\\ \frac{\partial^2 L}{\partial x_2\partial x_1} & \frac{\partial^2 L}{\partial x_2^2}\end{pmatrix}=\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2}-\lambda\frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2}-\lambda\frac{\partial^2 g}{\partial x_1\partial x_2}\\ \frac{\partial^2 f}{\partial x_2\partial x_1}-\lambda\frac{\partial^2 g}{\partial x_2\partial x_1} & \frac{\partial^2 f}{\partial x_2^2}-\lambda\frac{\partial^2 g}{\partial x_2^2}\end{pmatrix}.$$

二阶条件要求上述矩阵在线性约束的情形下是负半定的，即

$$h^tD^2L(x)h\le 0\text{ 对于满足 }Dg(x)h=0\text{ 的所有 }h\text{ 都成立.}$$

直觉上，这个条件要求海赛矩阵在与约束平面相切方向上的任何变动都是负半定的。

如果海赛矩阵在有约束条件限制下是负半定的，我们就说我们得到的是一个**常规最大值**（regular maximum）。常规最大值必然是严格局部最大值，但反过来说未必正确。

27.4 二阶条件的另外一种表示方法

在常规局部最大化下，二阶条件有另外一种表达方法。在这种情形下，判断一个特定的矩阵是否在线性约束下是负半定的，可以化简为检验某个矩阵的各个行列式的符号。

考虑朗格拉日函数的二阶导数组成的矩阵，其中包含各个变量关于拉格朗日乘子 λ 的导数。如果有两个选择变量和一个约束，这个矩阵看起来就是下面这样的样子：

$$D^2L(\lambda,x_1,x_2)=\begin{pmatrix}\frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \lambda\partial x_1} & \frac{\partial^2 L}{\partial \lambda\partial x_2}\\ \frac{\partial^2 L}{\partial x_1\partial \lambda} & \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1\partial x_2}\\ \frac{\partial^2 L}{\partial x_2\partial \lambda} & \frac{\partial^2 L}{\partial x_2\partial x_1} & \frac{\partial^2 L}{\partial x_2^2}\end{pmatrix}. \qquad (27.3)$$

使用拉格朗日函数的定义和一阶条件很容易计算出下列各个导数：

$$\frac{\partial^2 L}{\partial \lambda^2}=0$$

$$\frac{\partial^2 L}{\partial \lambda\partial x_1}=\frac{\partial^2 L}{\partial x_1\lambda}=-\frac{\partial g(x)}{\partial x_1}$$

$$\frac{\partial^2 L}{\partial\lambda\partial x_2} = \frac{\partial^2 L}{\partial x_2 \lambda} = -\frac{\partial g(x)}{\partial x_2}.$$

将这些表达式代入（27.3）可得到**加边海赛矩阵**（bordered Hessian matrix）:

$$\begin{pmatrix} 0 & -\frac{\partial g}{\partial x_1} & -\frac{\partial g}{\partial x_2} \\ -\frac{\partial g}{\partial x_1} & \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ -\frac{\partial g}{\partial x_2} & \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} \end{pmatrix}.$$

注意这个矩阵具有下列形式

$$\begin{pmatrix} 0 & b_1 & b_2 \\ b_1 & h_{11} & h_{12} \\ b_2 & h_{21} & h_{22} \end{pmatrix},$$

其中，“加边项”是约束条件一阶导数的负数，h_{ij} 项是拉格朗日函数关于选择变量的二阶导数。如果有 n 个选择变量和 1 个约束条件，加边海赛矩阵将是 $n+1$ 维方阵。下面我们举例说明，在这个例子中，选择变量有四个。

可以证明，在常规最大化且 $b_1 \neq 0$ 的情形下，上面给出的二阶条件意味着加边海赛矩阵有一个正的行列式。这个事实的证明并不难，但比较繁琐，所以我们省略了。

需要注意，如果我们将加边海赛矩阵的行列式的第一行和第一列都乘以 -1，它的符号不会改变。这种运算的好处是消去加边项的负号，从而看起来更舒服一些。有些作者一开始就给出这种形式的加边海赛行列式。然而，这种做法的缺陷是掩盖了下列事实：加边海赛矩阵就是拉格朗日函数的海赛矩阵，其中包含各个变量关于 λ 的导数。

如果有 n 个选择变量和 1 个约束条件，加边海赛矩阵将是 $(n+1)\times(n+1)$ 的矩阵。在这种情形下，我们必须考察加边海赛矩阵的各个子矩阵的行列式。我们以 4×4 的加边海赛矩阵说明如何计算。和上面一样，我们以 b_i 表示加边项，以 h_{ij} 表示海赛项，因此加边海赛矩阵为

$$\begin{pmatrix} 0 & b_1 & b_2 & b_3 & b_4 \\ b_1 & h_{11} & h_{12} & h_{13} & h_{14} \\ b_2 & h_{21} & h_{22} & h_{23} & h_{24} \\ b_3 & h_{31} & h_{32} & h_{33} & h_{34} \\ b_4 & h_{41} & h_{42} & h_{43} & h_{44} \end{pmatrix}.$$

考虑常规最大值的情形，在这种情形下，海赛矩阵在约束条件下是负定的，并且假设 $b_1 \neq 0$。那么二阶条件等价于下列行列式必定成立：

$$\det\begin{pmatrix} 0 & b_1 & b_2 \\ b_1 & h_{11} & h_{12} \\ b_2 & h_{21} & h_{22} \end{pmatrix} > 0$$

$$\det\begin{pmatrix} 0 & b_1 & b_2 & b_3 \\ b_1 & h_{11} & h_{12} & h_{13} \\ b_2 & h_{21} & h_{22} & h_{23} \\ b_3 & h_{31} & h_{32} & h_{33} \end{pmatrix} < 0$$

$$\det\begin{pmatrix} 0 & b_1 & b_2 & b_3 & b_4 \\ b_1 & h_{11} & h_{12} & h_{13} & h_{14} \\ b_2 & h_{21} & h_{22} & h_{23} & h_{24} \\ b_3 & h_{31} & h_{32} & h_{33} & h_{34} \\ b_4 & h_{41} & h_{42} & h_{43} & h_{44} \end{pmatrix} > 0.$$

这种模式适用于任何数量的因子。我们可以将这个条件表示为：**加边海赛矩阵的自然序主子行列式的符号必然是交替变化**。

常规局部最小值有类似的二阶条件，只不过此时要求所有的行列式都为负。

如何记住二阶条件

你可能记不住二阶条件涉及的所有行列式条件，说实话至少我记不住。下面这种简单的方法可以帮助你记住它们。

正定矩阵的最简单例子是单位矩阵。计算单位矩阵的主子行列式，容易看到所有行列式的符号都是正的。

负定矩阵的最简单例子是负的单位矩阵。不难看出这个矩阵的所有主子行列式的符号是交替变化的：

$$\begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} > 0$$

$$\begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -1\begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} < 0.$$

假设某个矩阵在线性约束下是正定的。最简单的例子是约束条件为 $(h_1, h_2)(1,1) = 0$ 下的单位矩阵。这样，我们就得到了下列加边海赛矩阵

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

左上角的主子式为

$$\begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 < 0.$$

上述单位矩阵的行列式沿着第一列展开，可得

$$-1\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} + 1\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -2.$$

因此，约束条件下的矩阵为正定的条件是，加边海赛矩阵的所有所有主子式都是负的。

下面我们考察约束条件下的矩阵为负半定的情形。再一次地，我们举最简单的例子：约束条件$(h_1, h_2)(1,1) = 0$下的负的单位矩阵。我们得到的加边海赛矩阵为

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

两个主子式分别为

$$\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1 < 0.$$

$$\begin{vmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & 1 \\ 0 & -1 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ -1 & 0 \end{vmatrix} = 2 > 0.$$

因此，我们希望主子式的符号是交替变化的。

包络定理

考虑下列参数化的最大化问题

$$M(a) = \max_{x_1, x_2} g(x_1, x_2, a)$$
$$\text{s.t. } h(x_1, x_2, a) = 0.$$

这个问题的拉格朗日函数为

$$L = g(x_1, x_2, a) - \lambda h(x_1, x_2, a),$$

一阶条件为

$$\begin{aligned}
&\frac{\partial g}{\partial x_1}-\lambda\frac{\partial h}{\partial x_1}=0\\
&\frac{\partial g}{\partial x_2}-\lambda\frac{\partial h}{\partial x_2}=0\\
&h(x_1,x_2,a)=0.
\end{aligned}\tag{27.4}$$

这些条件确定了最优选择函数为 $(x_1(a),x_2(a))$，它又确定了最大值函数

$$M(a)=g(x_1(a),x_2(a),a).\tag{27.5}$$

包络定理（envelope theorem）告诉我们如何计算最优值函数关于最优化问题中变量的导数，具体地说，计算公式为

$$\frac{dM(a)}{da}=\frac{\partial L(\mathbf{x},a)}{\partial a}\Big|_{\mathbf{x}=\mathbf{x}(a)}=\frac{\partial g(x_1,x_2,a)}{\partial a}\Big|_{\mathbf{x}=\mathbf{x}(a)}-\lambda\frac{\partial h(x_1,x_2,a)}{\partial a}\Big|_{\mathbf{x}=\mathbf{x}(a)}$$

和以前一样，需要注意这里的偏导数的意思，它们是 g 和 h 在下列条件下的关于 a 的导数：维持 x_1 和 x_2 固定在最优值时。

包络定理的证明比较简单，直接计算即可。对（27.5）微分可得

$$\frac{dM}{da}=\frac{\partial g}{\partial x_1}\frac{dx_1}{da}+\frac{\partial g}{\partial x_2}\frac{dx_2}{da}+\frac{\partial g}{\partial a},$$

将一阶条件（27.4）代入，可得

$$\frac{dM}{da}=\lambda\left[\frac{\partial h}{\partial x_1}\frac{dx_1}{da}+\frac{\partial h}{\partial x_2}\frac{dx_2}{da}\right]+\frac{\partial g}{\partial a}.\tag{27.6}$$

现在可以看到，最优选择函数必定恒满足约束 $h(x_1(a),x_2(a),a)\equiv 0$。将这个恒等式关于 a 微分，可得

$$\frac{\partial h}{\partial x_1}\frac{dx_1}{da}+\frac{\partial h}{\partial x_2}\frac{dx_2}{da}+\frac{\partial h}{\partial a}\equiv 0.\tag{27.7}$$

将（27.7）代入（27.6），可得

$$\frac{dM}{da}=-\lambda\frac{\partial h}{\partial a}+\frac{\partial g}{\partial a},$$

这正是我们想要的结果。

27.5 不等式约束下的最大化问题

很多经济问题自然会用到不等式约束。下面我们考察这样问题的一阶条件的合理形式。

令 $f:R^n\to R$ 和 $g_i:R^n\to R$，其中 $i=1,\ldots,k$；考虑下面的最优化问题

$$\begin{aligned}&\max\quad f(\mathbf{x})\\&s.t.\ \ g_i(\mathbf{x})\le 0.\quad i=1,\ldots,k\end{aligned}\tag{27.8}$$

点集 $\{\mathbf{x}:g_i(\mathbf{x}^*)\le 0\quad i=1,\ldots k\}$ 称为**可行集**（feasible set）。如果在某个特定的 $\mathbf{x}^*$ 上我们有 $g_i(\mathbf{x}^*)=0$，我们说第 i 个约束是个**束紧约束**（binding constraint）；否则，我们说第 i 个约束不是紧的，或是个**松弛约束**（slack constraint）。

令 $G(\mathbf{x}^*)$ 为紧约束在 $\mathbf{x}^*$ 点上的梯度集：

$$G(\mathbf{x}^*)=\{Dg_i(\mathbf{x}^*):\text{对于所有}\,i\,\text{满足}\,g_i(\mathbf{x}^*)=0\}$$

如果向量集 $G(\mathbf{x}^*)$ 是线性无关的，那么我们说**约束规格**（constraint qualification）成立。

库恩–塔克定理（Kuhn-Tucker Theorem）。**如果 $\mathbf{x}^*$ 是（27.8）的解，而且约束规格在 $\mathbf{x}^*$ 点成立，那么存在着一组库恩-塔克乘子 $\lambda_i\ge 0$，其中 $i=1,\ldots,k$，使得**

$$\mathrm{D}f(\mathbf{x}^*)=\sum_{i=1}^{k}\lambda_i\mathrm{D}g_i(\mathbf{x}^*).$$

而且，我们有下列互补的松弛条件（complementary slackness conditions）：

$$\lambda_i\ge 0\qquad\text{对于所有}\ i\text{，}$$

$$\lambda_i=0\qquad\text{如果}\ g_i(\mathbf{x}^*)<0\text{。}$$

比较库恩-塔克定理与拉格朗日乘子定理可知，二者主要的区别在于库恩-塔克乘子是非负的，而拉格朗日乘子可正可负。这个额外的信息有时非常有用。

当然，库恩-塔克条件只是最大值的必要条件。然而，在一种重要的情形下，它既是必要条件也是充分条件。

库恩–塔克充分性。**假设 f 是凹函数，g_i 是凸函数（其中 $i=1,\ldots,k$）。令 $\mathbf{x}^*$ 是个可行点，假设我们可以找到与互补松弛条件一致的非负实数 λ_i，使得 $\mathrm{D}f(\mathbf{x}^*)=\sum_{i=1}^{k}\lambda_i\mathrm{D}g_i(\mathbf{x}^*)$. 那么 $\mathbf{x}^*$ 是最大化问题（27.8）的解。**

证明。令 $\mathbf{x}$ 是个可行点。由于 f 是凹的，我们可以将其写为

$$f(\mathbf{x})\le f(\mathbf{x}^*)+\mathrm{D}f(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*).$$

曹乾（东南大学 caoqianseu@163.com）

使用定理中的假设，我们可以将上式写为

$$f(\mathbf{x}) \le f(\mathbf{x}^*) + \sum_{i=1}^{k} \lambda_i \mathrm{D}g_i(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*). \tag{27.9}$$

由于 $g_i(\mathbf{x})$ 是凸的，我们有

$$g_i(\mathbf{x}) \ge g_i(\mathbf{x}^*) + \mathrm{D}g_i(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*). \tag{27.10}$$

如果约束 i 是束紧的，那么 $g_i(\mathbf{x}) \le g_i(\mathbf{x}^*) = 0$；由于 $\lambda_i \ge 0$，对于所有束紧约束，我们有 $\lambda_i \mathrm{D}g_i(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \le 0$。如果约束 i 是松弛的，$\lambda_i = 0$。将这些结论应用于（27.9），我们看到（27.9）中 $\sum_{i=1}^{k} \lambda_i \mathrm{D}g_i(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)$ 的每一项都是小于或等于零，因此 $f(\mathbf{x}) \le f(\mathbf{x}^*)$，这正是我们想要的。■

27.6 构建库恩-塔克问题

约束最大化问题的拉格朗日条件和约束最小化问题的拉格朗日条件是相同的，因为它们处理的只是一阶条件。库恩-塔克条件也是这样的，当然前提是我们构建的问题要正确。一般来说，这要求在运用库恩-塔克定理之前先要进行一些处理。

我们已经知道，我们研究的是下列最大化问题形式的库恩-塔克条件：

$$\begin{aligned} &\max_{\mathbf{x}} \quad f(\mathbf{x}) \\ &\quad \text{s.t.} \quad g_i(\mathbf{x}) \le 0 \quad i = 1, \dots, k. \end{aligned}$$

这个问题的拉格朗日函数为

$$L = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}).$$

如果最大化问题以这种方式构建而成，那么就可以保证库恩-塔克乘子是非负的。

在有些最优化问题中，有些约束条件要求函数大于或等于零。在这种情形下，为了应用库恩-塔克条件，我们必须先将这样的约束乘以 -1。例如，假设我们的最大化问题为

$$\begin{aligned} &\max_{\mathbf{x}} \quad f(\mathbf{x}) \\ &\quad \text{s.t.} \quad h_i(\mathbf{x}) \ge 0 \quad i = 1, \dots, k. \end{aligned}$$

这个问题等价于

$$\begin{aligned} &\max_{\mathbf{x}} \quad f(\mathbf{x}) \\ &\quad \text{s.t.} \quad -h_i(\mathbf{x}) \le 0 \quad i = 1, \dots, k. \end{aligned}$$

但现在这种形式可以运用库恩-塔克条件。

对于这个最大化问题，拉格朗日函数变为

$$L = f(\mathbf{x}) - \sum_{i=1}^{k} \lambda_i(-h_i(\mathbf{x})) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i(h_i(\mathbf{x})).$$

由于变换后的最大化问题具有合适的形式，可以保证乘子 λ_i 是非负的。

现在假设我们研究的是最小化问题。那么，为了让库恩-塔克乘子有正确的符号（非负），我们构建的最小化问题必须为下列形式

$$\begin{aligned} &\min_{\mathbf{x}} \quad f(\mathbf{x}) \\ &\text{s.t.} \ \ g_i(\mathbf{x}) \geq 0 \quad i = 1, ..., k. \end{aligned}$$

注意，在最小化问题中，约束条件为 $g_i(\mathbf{x}) \geq 0$，而在最大化问题中，约束条件为 $g_i(\mathbf{x}) \leq 0$。

27.7 最大值的存在性和连续性

考虑下列形式的参数化的最大化问题

$$\begin{aligned} M(a) = \max \quad & f(\mathbf{x}, a) \\ \text{s.t.} \quad & \mathbf{x} \in G(a) \end{aligned} \tag{27.11}$$

最大值的存在性。**如果约束集 $G(a)$ 是非空的和紧致的（compact），而且函数 f 是连续的，那么这个最大化问题存在着解 $\mathbf{x}^*$。**

最大值的唯一性。**如果函数 f 是严格凸的，而且约束集是凸的，那么若解存在，则它是唯一的。**

令 $\mathbf{x}(a)$ 是问题（27.11）的解。有时，我们希望知道在什么样的条件下 $\mathbf{x}(a)$ 具有良好的性状。我们必须面对的一的哥问题是 $\mathbf{x}(a)$ 未必为函数：最大化问题可能存在着若干个解。也就是说，对于每个 a， $\mathbf{x}(a)$ 可能是个点集。这种情形下，我们用对应（correspondence）而不是函数描述这样的关系，对应是将每个 a 与集合 $\mathbf{x}(a)$ 联系起来的一种规则。我们想知道当 a 变化时，点集 $\mathbf{x}(a)$ 如何变化；特别地，我们想知道在什么样的条件下，当 a 变化时，$\mathbf{x}(a)$ 的变化是“连续的”。

用于描述对应（correspondences）的连续性的定义有两个。如果当 a 稍微变化时，集合 $\mathbf{x}(a)$ 不会“急剧膨胀”（explode），我们说这个对应是**上半连续的**（upper-semicontinuous）。如果当 a 稍微变化时，集合 $\mathbf{x}(a)$ 不会“急剧收缩”（implode），我们说这个对应是**下半连续的**（lower-semicontinuous）。如果一个对应既是上半连续的，又是下半连续的，那么它是**连续的**（continuous）。

最大值定理（theorem of the maximum）。**令 $f(\mathbf{x},a)$ 是个连续函数，它的值域是紧致的，假设约束集 $G(a)$ 是 a 的非空、紧值**（compacted-value）**、连续的对应。那么：（1）$M(a)$ 是个连续函数；（2）$\mathbf{x}(a)$ 是个上半连续的对应。**

如果对应 $\mathbf{x}(a)$ 碰巧是单值的，即 $\mathbf{x}(a)$ 是个函数，那么它是个连续函数。

注释

Dixit（1990）提供了经济学中最优化的直观讨论和例子。更详细的讨论可参见 Blume and Simon（1991）。如果想了解对应（correspondences）的拓扑性质，可参考 Berge（1963）和 Hildenbrand and Kirman（1988）。

参考文献

Abreu, D.(1986). Extremal equilibria of oligopolistic supergames. Journal of Economic Theory, 39, 191-225.

Afriat, S. (1967). The construction of a utility function from expenditure data. International Economic Review, 8, 67-77.

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 89, 488-500.

Anscombe, F. & Aumann, R, (1963). A definition of subjective probability. Annals of Mathematical Statistics, 34, 199-205.

Arrow, K. (1951). An extension of the basic theorems of classical welfare economics. In P. Newman (Ed.), Readings in Mathematical Economics. Baltimore: Johns Hopkins Press.

Arrow, K. (1970). Essays in the Theory of Risk Bearing. Chicago: Markham.

Arrow, K. & Debreu, G. (1954). Existence of equilibrium for a competitive economy. Econometrica, 22, 265-290.

Arrow, K. & Hahn, F. (1971). General Competitive Analysis. San Francisco: Holden-Day.

Atkinson, T.& Stiglitz, J. (1980). Lectures on Public Economics. New York: McGraw-Hill.

Aumann, R. (1987). Game theory. In J. Eatwell, M. Milgate, & P. Newman (Eds.), The New Palgrave. London: MacMillan Press.

Berge, C. (1963). Topological Spaces. New York: Macmillan.

Bergstrom, T., Blume, L., & Varian, H. (1986). On the private provision of public goods. Journal of Public Economics, 29(1), 25-49.

Binmore, K. (1982). Mathematical Analysis (2 ed.). Cambridge: Cambridge University Press.

Binmore, K. (1983). Calculus. Cambridge: Cambridge University Press.

Binmore, K. (1991). Fun and Games. San F'rancisco: Heath.

Binmore, K. & Dasgupta, P. (1987). The Economics of Bargaining. Oxford: Basil Blackwell.

Blackorby, C., Primont, D., & Russell, R. (1979). Duality, Separability and Functional Structure: Thwry and Economic Applications. Amsterdam: North-Holland.

Blackorby, C. & Russell, R. R. (1989). Will the real elasticity of substitution please stand up. American Economic Review, 79 (4), 882-888.

Blume, L. & Simon, C. (1991). Mathematics for Economists. New York: W. W. Norton & Co.

Clarke, E. (1971). Multipart pricing of public goods. Public Choice, 11, 17-33.

Coase, R. (1960). The problem of social cost. Journal of Law and Economics, 3, 1-44.

Cook, P. (1972). A one-line proof of the Slutsky equation. American Economic Review, 42, 139.

Davidson, C. & Deneckere, R. (1986). Long-run competition in capacity, short-run competition in price, and the Cournot model. RAND Journal of Economics, 17, 404-415.

Deaton, A. & Muellbauer, J. (1980). Economics and Consumer Behavior. Cambridge: Cambridge University Press.

Debreu, G. (1953). Valuation equilibrium and Pareto optimum. In K. Arrow & T. Scitovsky (Eds.), Readings in Welfare Economics. Homewood, Ill.: Irwin.

Debreu, G. (1959). Theory of Value. New York: Wiley.

Debreu, G. (1964). Continuity properties of Paretian utility. International Economic Review, 5,

285-293.
Debreu, G. (1974). Excess demand functions. Journal of Mathematical Economics, 1, 15-22.
Debreu,G. & Scarf, H. (1963). A limit theorem on the core of an economy. International Economic Review, 4, 235-246.
Dierker, E. (1972). Two remarks on the number of equilibria of an economy. Econometrica, 40, 951-953.
Diewert, E. (1974). Applications of duality theory. In M. Intriligator & D. Kendrick (Eds.), Frontiers of Quantitative Economics. Amsterdam: North-Holland.
Dixit, A. (1986). Comparative statics for oligopoly. International Economic Review, 27, 107-122.
Dixit, A. (1990). Optimization in Economic Theory (2ed.). Oxford: Oxford University Press.
Dowrick, S. (1986). von Stackelberg and Cournot duopoly: Choosing roles. Rand Journal of Economics, 17(l), 251-260.
Friedman, J. (1971). A noncooperative equilibrium for supergames. Review of Economic Studies, 38, 1-12.
Frisch, R. (1965). Theory of Production. Chicago: Rand McNally.
Fudenberg, D. & Tirole, J. (1991). Game Theory. Cambridge: MIT Press.
Geanakoplos, J. (1987). Overlapping generations model of general equilibrium. In J. Eatwell, M. Milgate, & P. Newman (Eds.), The New Palgrave. London: MacMillan Press.
Gorman, T. (1953). Community preference fields. Econornetrica, 21, 63-80.
Groves, T. (1973). Incentives in teams. Econornetrica, 41, 617-631.
Harsanyi, J.(1967).Games of incomplete information played by Bayesian players. Management Science, 14, 159-182, 320-334, 486-502.
Hart, O. & Holmstrom, B. (1987). The theory of contracts. In T. Bewley (Ed.), Advances in Economic Theory. Cambridge: Cambridge University Press.
Herstein, I. & Milnor, J. (1953). An axiomatic approach to measurable utility. Econometrica, 21, 291-297.
Hicks, J. (1932). Theory of Wages. London: Macmillan.
Hicks, J. (1946). Value and Capital. Oxford, England: Clarendon Press.
Hicks, J. (1956). A Revision of Demand Theory. London: Oxford University Press.
Hildenbrand, W. & Kirman, A. (1988). Equilibrium Analysis. Amsterdam: North-Holland.
Holmstrom,B. & Milgrom,P. (1987). Aggregation and linearity in the provision of intertemporal incentives. Econornetrica, 55, 303-328.
Hotelling, H. (1932). Edgeworth's taxation paradox and the nature of demand and supply function. Political Economy, 40, 577-616.
Hurwicz, L. & Uzawa, H. (1971). On the integrability of demand functions. In J. Chipman, L. Hurwicz, M. Richter, & H. Sonnenschein (Eds.), Preferences, Utility, and Demand. New York: Harcourt, Brace, Jovanovich.
Ingersoll, J. (1987). Theory of Financial Decision Making. Totowa, New Jersey: Rowman & Littlefield.
Kierkegaard, S. (1938). The Journals of Soren Kzerkegaard. Oxford: Oxford University Press.
Koopmans, T. (1957). Three Essays on the State of Economic Science. New Haven: Yale University Press.
Kreps, D. (1990). A Course in Microeconomics Theory. Princeton University Press.

Kreps, D. & Scheinkman, J. (1983). Quantity pre-commitment and Bertrand competition yield Cournot outcomes. Bell Journal of Economics, 14, 326-337.
Ledyard, J. (1986). The scope of the hypothesis of Bayesian equilibrium. Journal of Economic Theory, 39, 59-82.
Lindahl, E. (1919). Just taxation-a positive solution. In R. Musgrave & A. Peacock (Eds.), Classzcs zn the Theory of Public Finance. London: Macmillan.
Machina, M (1982) 'Expected utility' analysis without the independence axiom. Econometnca, 50, 277-323.
Marshall, A. (1920). Principles of Economics. London: Macmillan.
Maskin, E. & Roberts, K. (1980). On the fundamental theorems of general equilibrium. Technical Report 43, Cambridge University, Cambridge, England.
McFadden, D. (1978). Cost, revenue, and profit functions. In M. Fuss & D. McFadden (Eds.), Production Economics: A Dual Approach to Theory and Applications. Amsterdam: North-Holland.
McFadden, D. & Winter, S. (1968). Lecture Notes on Consumer Theory. University of California at Berkeley: Unpublished.
McKenzie, L. (1954). On equilibrium in Graham's model of world trade and other competitive systems. Econometrica, 22, 147-161.
McKenzie, L. (1957). Demand theory without a utility index. Review of Economic Studies, 24, 183-189.
Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. Bell Journal of Economics, 13, 380-391.
Milgrom, P. & Roberts, J. (1982). Limit pricing and entry under incomplete information: An equilibrium analysis. Econometrica, 50, 443-459.
Mirrlees, J. (1982). The theory of optimal taxation. In K. Arrow & M. Intriligator (Eds.), Handbook of Mathematical Economics, volume 11. Amsterdam: North-Holland.
Myenon, R. (1986). An introduction to game theory. In S. Reiter (Ed.), Studies in Mathematical Economics. Mathematical Association of America.
Myerson, R. (1991). Game Theory. Cambridge: Harvard University Press. Nash, J. (1950). Equilibrium points in n-person games. Proceedings of the National Academy of Sciences, 36, 48-49.
Nash, J. (1951). Non-cooperative games. Annals of Mathematics, 54, 286-295.
Negisihi, T. (1960). Welfare economics and the existence of an equilibrium for a competitive economy. Econometrica, 12, 92-97.
Neumann, J. & Morgenstern, 0. (1944). Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press.
Novshek, W. (1980). Cournot equilibrium with free entry. Review of Economic Studies, 4 7, 473-486.
Pigou, A. (1920). The Economics of Welfare. London: Macmillan.
Pollak, R. (1969). Conditional demand functions and consumption theory. Quarterly Journal of Economics, 83, 60-78.
Pratt, J. (1964). Risk aversion in the small and in the large. Econometrica, 32, 122-136.
Rasmusen, E. (1989). Games and Information. Oxford: Basil Blackwell.
Ross, S. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory,

13, 341-360.
Ross, S. (1977). The capital asset pricing model (CAPM), short sales restrictions and related issues. Journal of Finance, 32, 177-183.
Ross, S. (1978). A simple approach to the valuation of risky streams. Journal of Business, 51, 453-475.
Rothschild, M. & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. Quarterly Journal of Economics, 80, 62M49.
Roy, R. (1942). De l'utilitd. Paris: Hermann.
Roy, R. (1947). La distribution de revenu entre les divers biens. Econometrica, 15, 205-225.
Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. Bell Journal of Economics, 7, 407-25.
Samuelson, P. (1947). Foundations of Economic Analysis. Cambridge, Mass.: Harvard University Press.
Samuelson, P. (1948). Consumption theory in terms of revealed preference. Econometnca, 15, 243-253.
Samuelson, P. (1951). Abstract of a theorem concerning substitutability in an open Leontief model. In T. Koopmans (Ed.), Activity Analysis of Production and Consumption. New York: Wiley.
Samuelson, P. (1954). The pure theory of public expenditure. The Review of Economics and Statistics, 64, 387-389.
Samuelson, P. (1974). Complimentarily: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory. Journal of Economic Literature, 64 (4), 1255-1289.
Scarf, H. (1973). The Computation of Economic Equilibrium. New Haven: Yale University Press.
Shafer, W. & Sonnenschein, H. (1982). Market demand and excess demand functions. In K Arrow & M. Intriligator (Eds.), Handbook of Mathematical Economics. Amsterdam: North-Holland.
Shapiro, C. (1989). Theories of oligopoly behavior. In R. Schmalensee & R. Willig (Eds.), Handbook of Industrial Organization, volume 1. Amsterdam: North-Holland
Shephard, R. (1953). Cost and Production functions. Princeton, NJ: Princeton University Press.
Shephard, R. (1970). Cost and Production Functions. Princeton, NJ: Princeton University Press.
Silberberg, E (1974). A revision of comparative statics methodology in economics. Journal of Economic Theory, 7, 159-172.
Silberberg, E. (1990). The Structure of Economics. New York: McGraw- Hill.
Singh, N. & Vives, X. (1984). Price and quantity competition in a differentiated duopoly. Rand Journal of Economics, 15, 546-554.
Sonnenschein, H. (1968). The dual of duopoly is complementary monopoly: or, two of Cournot's theories are one. Journal of Political Economy, 36, 316-318.
Spence, M. (1974). Market Signaling. Cambridge, Mass.: Harvard University Press.
Spence, M (1975). Monopoly, quality and regulation. Bell Journal of Economics, 6(2), 417-429.

Tirole, J. (1988). The Theory of Industrial Organization. Cambridge: MIT Press.
Varian, H. (1980). A model of sales. American Economic Review, 70, 651-659.
Varian, H. (1982a). The nonparametric approach to demand analysis. Econometrica, 50, 945-973.
Varian, H. (1982b). The nonparametric approach to production analysis. Econometrica, 52, 579-597
Varian, H. (1985). Price discrimination and social welfare. American Economic Review, 75 (4), 87Ck875.
Varian, H. (1989a). Price discrimination. In Handbook of Industrial Organization. Amsterdam: North-Holland.
Varian, H. (1989b). A solution to the problem of externalities when agents are well-informed. Technical report, University of Michigan, Ann Arbor.
Varian, H. (1990). Goodness-of-fit in optimizing models. Journal of Economics, 46, 125-140.
Wald, A. (1951). On some systems of equations in mathematical economics. Econometrica, 19, 368-403.
Walras, L. (1954). Elements of Pure Economics. London: Allen and Unwin.
Weizsacker, C. V. (1971). Steady State Capztal Theory. New York: Springer-Verlag.
Willig, R. (1976). Consumer's surplus without apology. American Economic Review, 66, 589-597.
Wold, H. (1943). A synthesis of pure demand analysis, i-iii. Skandznavzsk Aktuarzetzdsknft, 26, 27.
Yaari, M. (1969). Some remarks on measures of risk aversion and their uses. Journal of Economic Theory, 1, 315-329.
Yokoyama, T. (1968). A logical foundation of the theory of consumer's demand. In P. Newman (Ed.), Readings in Mathematical Economics. Baltimore: Johns Hopkins Press.