第一章 概率论的若干结论

概率论的发展归功于精确地论述
越来越复杂的观测结果。
M. Loeve（1977，P. 7）

1. 引言

在本章中,我们将以定义、定理和例子的形式给出现代概率论的各种概念。这些概念取材于大量的材料，将为读者提供本书后续章节的背景材料。尽管本章的某些节内对于一些概念一带而过，但有些节将详细叙述诸如鞅和最优停时等概念，读者可以从几个例子中看到概率论的鞅论和最优停时理论在经济分析和金融中的应用。阅读本章至少可以从以下两个方面受益：首先，了解各种应用概率模型的理论框架；其次，熟悉应用研究中广泛使用的鞅和最优停时理论。

2. 概率空间

在一次试验中可能发生，也可能不发生，具有偶然性的事件称为试验的结果。试验所有的结果构成一个集合，由Ω来表示。集合或空间Ω的元素由ω来表示，称为样本点。Ω的一个子集称为一个事件。例如，考虑下个月美国失业人数占整个劳动力人数的比例，样本空间Ω由$[0,1]$区间上的所有有理数构成，而$\omega = 0.065$即为一个元素或样本点。失业率不超过0.07，或者失业率介于0.06和0.08之间均为事件。直观上讲，概率就是对在族上的某类事件的评价。

设Ω为由点ω组成的任意空间。在概率研究中，如果Ω的某一子集族是重要的，可定义Ω的该族子集为$\sigma-$域或$\sigma-$代数，本书中以F来表示。我们称F为一个$\sigma-$域，如果它满足以下三个条件：

（1）$\Omega \in F$,即F包含了空间Ω本身；

（2）$A \in F$意味着$A^c \in F$，即如果事件A（A为Ω的一个子集）属于F，则A的补集A^c也属于F；

（3）$A_1, A_2, A_3, ... \in F$意味着$A_1 \cup A_2 \cup A_3 \cup \cdots \in F$，即如果$\Omega$的可列子集属于$F$，则这些可列子集的并集也属于$F$。

我们称F的所有元素为可测集。值得一提的是，对于一个给定的集合Ω，其最小的$\sigma-$域是由空集ϕ和集合Ω自身所构成，而最大的$\sigma-$域则由Ω的幂集，即Ω的所有子集而构成。最大的$\sigma-$域由Ω的2^{Ω}个子集组成。这样一对(Ω, F)称为一个可测空间。

令ς为Ω一个子集族。所有包含ς的$\sigma-$域的交称为由ς生成的$\sigma-$域，记为$\sigma(\varsigma)$。作为例子，我们考虑由实直线R^1的子集—$(a,b]$区间族而生成的$\sigma-$域，这个$\sigma-$域用$\Re^1$来表示，其元素称为$Borel$集。注意，$\Re^1$是包含所有R^1区间族的最小的$\sigma-$域；特别地，它包含了R^1的所有开集和闭集。直观上来说，$\Re^1$是由R^1区间开始，经过一系列所有可能的有限和可列集合理论运算（并、交和补）而得到的。

如前所述，概率是对某类事件族上的评价。现在给出其严格的定义。令(Ω,F)为一可测空间。定义在F上的一个集合函数μ称为一个测度，如果它满足下面的条件：

（1）$\mu(\phi)=0$；

（2）$A\in F$意味着$0\le\mu(A)\le\infty$；

（3）$A_1,A_2,\cdots\in F$，如果A_n是两两不相容的，即$A_k\cap A_m=\phi$，$k\neq m$，则：

$$\mu(\bigcup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}\mu(A_n)$$

一个最重要的测度为$Lebesgue$测度，它定义在实直线R^1的$Borel$集合族上，由λ来表示。这个测度赋予每一个区间的是其区间的长度，即$\lambda(a,b]=b-a$。

$Lebesgue$测度可以直接推广到所有的$Borel$集。要了解详情的读者，请参阅$Ash(1972,Ch.1)$。值得注意的是，$\lambda(R^1)=\infty$以及可数集合$Lebesgue$的测度为0。定义在$\Re^1$上的$Lebesgue$测度λ可以推广到k维空间R^k的$Borel$集$\Re^k$上。

概率是一种特殊测度，记为P，它满足$P(\Omega)=1$。因此对于$A\in F$，有$0\le P(A)\le 1$。三元体(Ω,F,P)称为概率空间，其中Ω是由试验结果构成的非空的空间，F是由Ω的代表各种事件的子集所构成的$\sigma-$域，而P为定义在F上的概率测度。

作为例子，我们来考虑$[0,1]$区间上的所有有理数构成的集合Ω，它们表示未来某个月中，美国失业人数占总劳动力人数的比例。令F为可列空间Ω的所有子集构成的$\sigma-$域，$\mu(\omega)$为定义在Ω上的非负函数，满足$\sum_{\omega\in\Omega}\mu(\omega)=1$，定义$P(A)=\sum_{\omega\in A}\mu(\omega)$，则$(\Omega,F,P)$就是一个概率空间。

概率空间的一种特殊情形为完备概率空间。令(Ω, F, P)为一个概率空间，N为Ω的一个子集。如果存在一个集合$A \in F$使得$N \subset A$，且$P(A)=0$，则称N是可忽略的。如果F包含了Ω的所有对P可忽略的子集，则称概率空间是完备的。这个概念将在本章§7中用到。

考虑概率空间(Ω, F, P)的一列事件$\{A_n\}$，定义：

$$\begin{aligned}\limsup_n A_n &= \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k \\ &= \{\omega : \omega\text{属于无穷多个}A_n\}\end{aligned}$$

和

$$\begin{aligned}\liminf_n A_n &= \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \\ &= \{\omega : \omega\text{属于除有限个外所有的}A_n\}\end{aligned}$$

如果所有$A_n \in F$，则$\limsup\limits_n A_n$和$\liminf\limits_n A_n$都属于F。

在结束本节之前，我们需要给出独立性的定义，并给出一个重要引理。我们由初等概率论知，如果$P(A \cap B) = P(A)\cdot P(B)$，则称事件$A$和事件$B$独立。我们将这一概念推广为有限个事件的独立性。事件$A_1, A_2, ..., A_n$是相互独立的，如果它们满足对任意一组j个不同的下标$k_1, k_2, ..., k_j \in \{1,2,...,n\}$，有：

$$P(A_{k_1} \cap ... \cap A_{k_j}) = P(A_{k_1}) ... P(A_{k_j}) \tag{2.1}$$

对于无穷多个事件，如果它的任意有限个事件是独立的，则称这无穷多个事件是独立的。下面的引理非常重要，本章的后面将会用到它。

引理 2.1 (Borel-Cantelli) 令$\{A_n\}$为概率空间(Ω, F, P)上的一事件列。如果$\sum\limits_n P(A_n) < \infty$，则$P(\limsup\limits_n A_n) = 0$。如果事件列$\{A_n\}$是独立的，且$\sum\limits_n P(A_n) = \infty$，则：

$$P(\limsup_n A_n) = 1$$

证明：参见 Neveu(1965,128-129 页)。

3. 随机变量

设(Ω, F)和(Ω', F')为两个可测空间。称映射$X : \Omega \to \Omega'$是(F, F')可测的，如果对每一个$A' \in F'$，有：

$$X^{-1}(A') = \{\omega : X(\omega) \in A'\} \in F$$

直观上讲，一个映射的可测性指的是对于象空间Ω'的每一个有意义的事件，定义域空间Ω总有一个有意义的事件与其对应。这里，一个事件有意义指的是它属于一个适当的$\sigma-$域。由X生成的$\sigma-$域记为$\sigma(X)$，它是使得X可测的最小的$\sigma-$域。

一个重要的特殊情形是象空间Ω'为实直线R^1时。这时我们取$\sigma-$域为$\Re^1$，即 Borel 集合族。称一个函数$f:\Omega \to R^1$是可测的，如果对于每一个$A' \in \Re^1$，有$f^{-1}(A') = \{\omega : f(\omega) \in A'\} \in F$。在上述意义下可测的实函数就称为随机变量。

假设$f:\Omega \to R^k$是可测的。那么它就称为随机向量。注意f有如下形式：

$$f(\omega) = (f_1(\omega),..., f_k(\omega))$$

这里每一个分量$f_i(\omega)$都是实函数。映射f是可测的，当且仅当它的每一个分量都是可测的。在许多应用中，我们假设f是连续的，这意味着f也是可测的。

令(Ω, F, P)为一个概率空间，$(R,^1 \Re^1, \lambda)$为可测空间。其中R^1为实直线，$\Re^1$为 Borel 集构成的$\sigma-$域，λ为$Lebesgue$测度，并假设X为一个随机变量，$X:\Omega \to R^1$。对于所有的$A \in \Re^1$，我们定义X的分布或分布律P_X为：

$$\begin{aligned} P_X(A) &= P(X^{-1}(A)) \\ &= P(\omega : X(\omega) \in A) \end{aligned} \quad (3.1)$$

随机变量X的分布对于目标空间中的每一个集合$A \in \Re^1$赋予了一个概率测度，它依赖于X的原象。对于$x \in R^1$，定义X的分布函数F如下：

$$F(x) = P\{\omega : X(\omega) \le x\} \quad (3.2)$$

注意到F是一个递减、右连续的函数，且：

$$F(x) \to 0, \text{当} x \to -\infty \text{时},$$
$$F(x) \to 1, \text{当} x \to \infty \text{时}。$$

概率论中有一个非常重要的定理。该定理叙述如下：对于一个给定的具有上述性质的分布函数F，总可以构造一个概率空间(Ω, F, P)和一个随机变量X，使得F恰好是X的分布函数。参见 Billingsley(1979,159 页)。于是分布函数的概念可以推广到随机向量$X:\Omega \to R^k$上。对每一个$x \in R^k$

$$\begin{aligned} F(x) &= F(x_1, x_2, \cdots, x_k) \\ &= P\{\omega : X_1(\omega) \le x_1, \cdots, X_k(\omega) \le x_k\} \end{aligned} \tag{3.3}$$

这里，$X_1,...,X_k$ 为 X 的 k 个分量。我们称 F 为联合分布函数。令 F 为 K 维随机向量 $X=(X_1,...,X_k)$ 的联合分布函数。X 的 m 个分量 $(m \le k)$ 的边际分布函数可通过在 F 的未涉及到的变量中以 ∞ 代替而得到。作为示例，一个二维（$m=2$）边际分布，可写为：

$$\begin{aligned} &F(x_1, x_2, \infty, \cdots, \infty) \\ &= P\{\omega : X_1(\omega) \le x_1, X_2(\omega) \le \infty, \cdots, X_k(\omega) \le \infty\} \end{aligned} \tag{3.4}$$

称随机变量 X 和其分布 P_X 关于 $Lebesgue$ 测度具有密度 f，如果 f 为一个非负实函数，且使得对于每一个 $A \in \Re^1$，有：

$$P\{\omega : X(\omega) \in A\} = \int_A f(x)dx \tag{3.5}$$

注意，密度函数 f 仅由一个 $Lebesgue$ 零测度集确定。当随机变量 X 有密度 f 时，则密度 f 和分布函数 F 有如下关系：

$$F(x) = \int_{-\infty}^{x} f(s)ds \tag{3.6}$$

在许多应用中，分布函数 F 是连续可微的。此时 F 的导数就是密度 f。有几个分布函数的例子，在初等概率论中大家已经熟悉，如二项分布：

$$P_X(x) = \binom{n}{x} P^x (1-P)^{n-x} \tag{3.7}$$

$x = 0,1,2,...,n$。在此例中，我们取 $\Omega = \{0,1,2,...,n\}$，F 由 Ω 的所有子集构成，定义变量 X 为 $X(x) = x$。

另一个例子为带有正参数 m 的 $Poisson$ 分布：

$$P_X(x) = \frac{e^{-m} m^x}{x!} \tag{3.8}$$

$x = 0,1,2,...$。

最后，称随机变量 X 服从参数 μ 和 $\sigma > 0$ 的正态分布，如果对于 $x \in A$，且 $A \in \Re^1$，有：

$$P_X(A) = \frac{1}{\sqrt{2\pi\sigma}} \int_A \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] dx \tag{3.9}$$

假设 $X_1, X_2, \ldots, X_n$ 为定义在概率空间 (Ω, F, P) 上的随机变量，$X_i : \Omega \to R^1, i = 1,2,\ldots,n$。

如果对所有的 Borel 集 $A_1, A_2, \ldots, A_n$，有：

$$\begin{aligned} & P\{\omega : X_1(\omega) \in A_1, \cdots, X_n(\omega) \in A_n\} \\ & = P\{\omega : X_1(\omega) \in A_1\} \cdots P\{\omega : X_n(\omega) \in A_n\} \end{aligned} \tag{3.10}$$

则称 $X_1, X_2, \ldots, X_n$ 是相互独立的随机变量。另外还有两个等价性定义：$X_1, X_2, \ldots, X_n$ 是相互独立的随机变量，如果对任意的实数 $x_1, x_2, \ldots, x_n$，有：

$$P[X_1 \le x_1, \cdots, X_n \le x_n] = P[X_1 \le x_1] \cdots P[X_1 \le x_1] \cdots P[X_n \le x_n] \tag{3.11}$$

注意，在（3.11）中，为了记号的方便，将变量 ω 省略掉了。换言之，$X_1, X_2, \ldots, X_n$ 是相互独立的，如果对于任意的实数 $x_1, x_2, \ldots, x_n$，有：

$$F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n) \tag{3.12}$$

其中 F 是 $X_1, X_2, \ldots, X_n$ 的联合分布函数，$F_1, F_2, \ldots, F_n$ 为一维边际分布函数。

上述定义可以非常容易地推广到随机向量情形。在（3.10）中，我们将 X_i 视为随机向量，而 A_i 则为 k 维 Borel 集。此外，独立性的定义还可推广到无穷多个随机变量 $X_1, X_2, \ldots$ 的情形，根据（3.10）或其他等价性定义，只要要求任意有限子集独立即可。

对于定义在 (Ω, F, P) 上的随机变量列 $\{X_n\}$ 和定义在同一个空间上的随机变量 X，有三种常用的收敛性概念。首先，如果存在一个集合 $N \in F, P(N) = 0$，使得对所有的 $\omega \notin N$，实数列 $\{X_n(\omega)\}$ 在通常的意义下收敛于 $X(\omega)$，则称 $\{X_n\}$ 以概率 1 收敛于 X，并记为：

$$X_n \to X \;\; w.p.1 \;\; , n \to \infty \tag{3.13}$$

其次，我们称随机变量 $\{X_n\}$ 以概率收敛于 X，如果对于任给的 $\varepsilon > 0$，有：

$$P\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} = P[|X_n - X| > \varepsilon] \to 0 \, , n \to \infty \tag{3.14}$$

并记为：

$$X_n \xrightarrow{P} X \quad , n \to \infty \tag{3.15}$$

最后，设$\{F_n\}$和F 分别为$\{X_n\}$和X 的分布函数。若当$n \to \infty$时：

$$\int_R f(x)dF_n(x) \to \int_R f(x)dF(x) \tag{3.16}$$

对于任意的定义在R上的实值、连续有界的函数 f 都成立，则称$\{X_n\}$依分布收敛于X，并记为：

$$X_n \Rightarrow X \quad ,n \to \infty \tag{3.17}$$

注意：（3.13）成立表明（3.15）成立，而且（3.15）成立表明（3.17）成立。

4. 数学期望

如果X 为定义在概率空间(Ω, F, P)上的随机变量，若积分$\int_\Omega XdP$存在，即$\int_\Omega XdP < \infty$，则称其为X 的期望，记为：

$$E(x) = \int_\Omega XdP \tag{4.1}$$

问题随之而来，（4.1）右端的意义是什么？通过定义积分

$$\int_\Omega X(\omega)dP(\omega) = \int_\Omega X(\omega)P(d\omega) = \int_\Omega XdP \tag{4.2}$$

可以立即回答这个问题。

首先，假设X 为一个非负随机变量。对于空间Ω有限分割集合$\{A_i\}$，A_i满足$A_i \in F, A_i \cap A_j = \phi$，当$i \neq j$时，且$\bigcup_i A_i = \Omega$，考虑总和：

$$\sum_i [\inf_{\omega \in A_i} X(\omega)]P(A_i) \tag{4.3}$$

利用（4.3），定义（4.2）的意义如下：

$$\int_\Omega XdP = \sup \sum_i [\inf_{\omega \in A_i} X(\omega)]P(A_i) \tag{4.4}$$

在（4.4）中，上确界$\sup$是对Ω的定义在F上的所有有限分割$\{A_i\}$而取的。

其次，注意到任意的随机变量X 都可以表示为：

$$X = X^+ - X^- \tag{4.5}$$

其中X^+表示X 的正部，由

$$X^+(\omega) = \begin{cases} X(\omega), & \text{如果} \ 0 \le X(\omega) \le \infty \\ 0, & \text{其他} \end{cases} \tag{4.6}$$

给出，而X^-表示X 的负部，由

$$X^-(\omega)=\begin{cases}-X(\omega), & \text{如果 } -\infty\le X(\omega)\le 0\\ 0, & \text{其他}\end{cases} \tag{4.7}$$

给出。由于X^+和X^-都是非负可测的，从而（4.4）对于它们都成立。因此任意随机变量的积分可以定义为：

$$\int_\Omega XdP=\int_\Omega X^+dP-\int_\Omega X^-dP \tag{4.8}$$

如果（4.8）右端的每一项都是有限的，则我们称X是可积的。由于$|X|=X^++X^-$，所以X可积的充要条件为$\int_\Omega |X|dP<\infty$。

期望具有几个良好的性质。在下个定理介绍这些性质之前，我们先来解释测度论和概率论中的一个习惯术语。在概率空间(Ω,F,P)上，一个与Ω的元素ω有关的命题被认为是几乎必然成立的，记为$a.s.$，或等价地，以概率 1 成立，记为$w.p.1$，如果该命题对于可能除一个零测度$N[N\in F,P(N)=0]$以外的所有$\omega\in\Omega$都成立。如果我们以任意测度来代替P，则$w.p.1$由几乎处处来代替，记为$a.e$。

定理 4.1 设(Ω,F,P)为一个概率空间，f和h为定义在这个空间上的随机变量。

（1） 若f可积，$\alpha\in R$，则$\int_\Omega \alpha fdP<\infty$，且$\int_\Omega \alpha fdP=\alpha\int_\Omega fdP$。

（2） 若f和h可积，α和β为有限实数，则$\alpha f+\beta h$可积，且

$\int_\Omega(\alpha f+\beta h)dP=\alpha\int_\Omega fdP+\beta\int_\Omega hdP$。

（3） 若f和g可积，且$f\le h\ \ w.p.1$，则$\int_\Omega fdP\le\int_\Omega hdP$。

（4） 若f可积，则$\left|\int_\Omega fdP\right|\le\int_\Omega |f|dP$。

（5） 若f可积，则对任意的$A\in F$，$\int_A fdP<\infty$。

关于这个定理的证明，参阅 Ash(1972,41-42 页)。

定理 4.2 （单调收敛定理）设$\{X_n\}$为(Ω,F,P)上的非负单调递增的随机变量列，且$X_n(\omega)\to X(\omega), w.p.1$，则$\int_\Omega X_ndP\to\int_\Omega XdP$。

注意，这个定理保证了极限和期望运算的可交换性，即$\lim\limits_{n\to\infty}E(X_n)=E(\lim\limits_{n\to\infty}X_n)$。

对于实的随机变量列X_n，下面定义$\limsup_n X_n$和$\liminf_n X_n$：

$$(\liminf_n X_n)(\omega) = \sup_n \inf_{k\geq n} X_k(\omega)$$

$$(\limsup_n X_n)(\omega) = \inf_n \sup_{k\geq n} X_k(\omega)$$

定理 4.3 （Fatou 引理）对于(Ω, F, P)上的非负随机变量列$\{X_n\}$，有：

$$\int_\Omega \liminf_n X_n dP \leq \liminf_n \int_\Omega X_n dP$$

定理 4.4 （控制收敛定理）设$\{X_n\}$为一随机变量列，Y 为一可积的随机变量，所有变量都定义在(Ω, F, P)上，且$|X_n| \leq Y \ w.p.1$对所有n都成立。如果$X_n \to X \ w.p.1$，则X 和X_n 都可积，且$\int_\Omega X_n dP \to \int_\Omega X dP$。

这些定理的证明见 Ash(1972,44-50 页)。

利用上述积分的背景，我们现在来叙述一些基本定义。设X 为(Ω, F, P)上的随机变量，$k > 0$。我们称$E(X^k)$为X 的k阶矩，而$E((X - E(X))^k)$为k阶中心矩。当$k = 1$时，$E(x)$就是通常所称的X 的均值；当$k = 2$时，二阶中心矩称为X 的方差，记为$VarX$，用σ^2表示，即：

$$\sigma^2 = VarX = E((X - E(X))^2) \tag{4.9}$$

假定$E(X) < \infty$。正的平方根σ称为标准差。

注意到，如果$k > 0, E(X^k) < \infty$，则$E(X^l) < \infty$对$0 < l < k$成立。还有，如果$X_1,..., X_n$为定义在(Ω, F, P)上的相互独立的随机变量，$E(X_i) < \infty, i = 1,2,..., n$，则$E(X_1 \cdots X_n) < \infty$，且有：

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n) \tag{4.10}$$

而且，

$$Var(X_1 + ... + X_n) = \sum_{i=1}^n VarX_i \tag{4.11}$$

对于具有有限期望的两个随机变量X 和Y，X 和Y 的协方差定义为：

$$\begin{aligned} Cov(X,Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y) \end{aligned} \tag{4.12}$$

由（4.10），如果X 和Y 是独立的，则$Cov(X,Y) = 0$。但是反之则不成立。

考虑随机变量 X 和 Y 。假设它们的方差 σ_X^2 和 σ_Y^2 是非零且有限的。我们定义 X 和 Y 的相关系数，由 $\rho(X,Y)$ 表示，为：

$$\rho(X,Y) = Cov(X,Y)/\sigma_X\sigma_Y \tag{4.13}$$

我们通过叙述一个有用的事实来结束这一节。设 X 为定义在 (Ω,F,P) 上的随机变量，其分布函数为 F 。假设：$g: R \to R$ 为一可测函数，且 $Y = g(X)$ ，则：

$$E(Y) = \int_R g(x)dF(x) \tag{4.14}$$

在各种应用中，通过 R 上的积分而代替 Ω 上的积分来计算随机变量的期望会更容易些。基于此，我们可以在（4.14）中，令 $g(x) = x$ 。

5. 条件概率

由初等概率论知，给定一个集合 B ，集合 A 的条件概率 $P(A|B)$ 由

$$P(A|B) = P(A \cap B)/P(B) \tag{5.1}$$

给出，这里假定 $P(B) \neq 0$ 。条件概率与给定空间 Ω 的一个子集中的事件有关。直观上讲，条件概率表示的是在 B 已经发生这一信息的前提下，重新计算 A 发生的概率。本节，我们将研究样本空间 (Ω,F,P) 下具有一般意义的条件概率。这里,集合 A 的条件概率是针对 F 中的 $\sigma-$域 ζ 而定义的,所使用的记号为 $P[A|\zeta]$ 。直观的解释类似于（5.1）的解释：集合 A 的条件概率是根据包含在 F 中的 $\sigma-$域 ζ 所提供的信息而计算的。

在此，我们离开本题转而给出两个定义以及 Radon-Nikodym 定理。Radon-Nikodym 定理将建立条件概率的存在性。

考虑可测空间 (Ω,F) 上两个测度 v 和 P 。我们称测度 v 相对于测度 P 绝对连续，或称 v 被 P 控制，只要对每个 $A \in F$ ， $P(A) = 0$ 一定也有 $v(A) = 0$ 。若 Ω 可由 F 中至多可数多个无穷集合的并组成，每一个都有有限测度，则称测度 P 为 $\sigma-$有限。现在给出定理的陈述。

定理 5.1 （Radon-Nikodym） 设 (Ω,F) 是一可测空间。若对 (Ω,F) 上的两个 $\sigma-$有限测度 v 和 P ，其中 v 被 P 控制，则存在一个非负可测函数 f 使得

$$v(A)=\int_A fdP \qquad A\in F \tag{5.2}$$

证明参见 Ash（1972，PP. 63-65）。

（5. 2）中的函数 f 称为 v 关于 P 的 Radon-Nikodym 导数，记为 dv/dP。

基于上述简要的背景资料，我们将在任意一个概率空间 (Ω,F,P) 上，给出集合 A 在给定 F 中的 $\sigma-$ 域 ζ 下的条件概率的定义。我们将按照 Billingsley（1979）的方法进行分析。

对给定集合 $A\in F$，如下定义 $\sigma-$ 域 ζ 上的测度 v：

$$v(G)=P(A\cap G) \tag{5.3}$$

对任意 $G\in\zeta$。将概率测度 P 限制在 $\sigma-$ 域 ζ 上。注意到若 $P(G)=0$，则 $v(G)=0$。这表示（5.3）所定义的测度 v 是受 P 控制的。对于可测空间 (Ω,ζ)，由于 v 和 P 都是 σ 有限的，且 v 受控于 P，所以由 Radon-Nikodym 定理得知存在一个非负随机变量 f 使（5.2）成立。综合（5.2）、（5.3）我们得出：

$$v(G)=P(A\cap G)=\int_G fdP \tag{5.4}$$

对于任意 $G\in\zeta$。随机变量 f 就是条件概率 $P[A|\zeta]$，且满足以下两条性质：

（1）$P(A|\zeta)$ 是对 ζ 的测度且可表示为 P 的积分：

（2）对 $G\in\zeta$，$P[A|\zeta]$ 满足等式：

$$\int_G P[A|\zeta]dP=P(A\cap G) \tag{5.5}$$

存在许多 $w.p.1$ 等价意义下的随机变量 $P[A|\zeta]$。对任一给定的 $P[A|\zeta]$，我们称此随机变量为条件概率的一种表示。当 A 的条件概率由随机变量 X 生成的 σ 域表示时，我们记为 $P[A|\sigma(X)]$，或简记为 $P[A|X]$。由此我们也可以得到任一给定的随机变量序列 $X_1,X_2,\ldots$ 的条件下，A 的条件概率 $P[A|\sigma(X_1,X_2,\ldots)]$ 或 $P[A|X_1,X_2,\ldots]$。

下述定理综合了条件概率的一些基本性质。

定理 5.2 设 (Ω,F,P) 是一概率空间，ζ 是 F 中的一个 $\sigma-$ 域。

（1）若 $A\in F$，则 $0\le P[A|\zeta]\le 1$。特别地，若 $P(A)=1$，则 $P[A|\zeta]=1$，且若 $P(A)=0$，则 $P[A|\zeta]=0$。

（2）若$\{A_n\}, n=1,2,...$是一列互不相交的集合，则$P[\bigcup_n A_n \mid \zeta]=\sum_n P[A_n \mid \zeta]$。

（3）若 $A \in F, B \in F$ ，且 $A \subset B$ ，则 $P[B-A \mid \zeta]=P[B \mid \zeta]-P[A \mid \zeta]$ ，且 $P[A \mid \zeta] \leq P[B \mid \zeta]$。

（4）若$\{A_n\}$是一递增（减）序列，以A为极限，则当$n \to \infty$时，$P[A_n \mid \zeta]$递增（减），并以$P[A \mid \zeta]$为极限。

在简单讨论了条件概率后，下面给出条件期望的概念。

设X为(Ω, F, P)上的非负可积随机变量，定义ζ上的测度v如下：

$$v(G)=\int_G XdP \tag{5.6}$$

对任给的$G \in \zeta$成立。这里ζ是F上的$\sigma-$域。如前所述，若P限制在ζ上，则由（5.6）所定义的v受P控制，且应用 Radon-Nikodym 定理得知存在随机函数f满足：

$$v(G)=\int_G fdP \tag{5.7}$$

这个随机变量由$E[X \mid \zeta]$来表示，称为给定$\sigma-$域ζ的条件下随机变量X的条件期望。注意$E[X \mid \zeta]$满足以下两条性质：

（1）$E[X \mid \zeta]$可积且ζ可测；

（2）由（5.6）和（5.7）可得对于$G \in \zeta$：

$$\int_G E[X \mid \zeta]dP=\int_G XdP \tag{5.8}$$

对一般的随机变量X，X可以不必非负，它的条件期望可由（4.5）得到，定义如下：

$$E[X \mid \zeta]=E[X^+ \mid \zeta]-E[X^- \mid \zeta]。$$

由于存在许多仅在一零概率集上不同的$E[X \mid \zeta]$值，因而对每一个这样的值，我们都称其为随机变量X的条件期望的一种表示。值得一提的是$E[X \mid \zeta]$为一个依赖于ω的随机变量，即$E[X \mid \zeta](\omega)$，我们可将其解释为给定$\sigma-$域ζ的信息的条件下X的期望值。

条件期望满足以下定理中的各种性质。在定理 5.3 中我们假定X和Y是可积的。

定理 5.3 设(Ω, F, P)是一概率空间，ζ是F的$\sigma-$域。

（1）若 X 为常数，即 $X=c, c\in R$，则 $E[X\mid\zeta]=c$ 以概率 1 成立 $w.p.1$。

（2）若 X,Y 为随机变量，$a,b\in R$，则：

$E[aX\pm bY\mid\zeta]=aE[X\mid\zeta]\pm bE[Y\mid\zeta]$

（3）若 X,Y 为随机变量，且 $X\le Y$ 以概率 1 成立 $w.p.1$，则：

$$E[X\mid\zeta]\le E[Y\mid\zeta]$$

（4）对于任给的随机变量 X，有：

$$\left|E[X\mid\zeta]\right|\le E[\left|X\right|\mid\zeta]$$

证明参见 Billingsley（1979. P. 379）。

定理 5.4 设 X 为 (Ω,F,P) 上的随机变量且可积，若 ζ_1 和 ζ_2 为 F 上的 $\sigma-$域，且 $\zeta_1\subset\zeta_2$，则

$$E[E[X\mid\zeta_2]\mid\zeta_1]=E[X\mid\zeta_1] \text{ 以概率 1 成立。}$$

证明参见 Tucker（1967，p. 212）。

定理 5.5 设 X,Y 为 (Ω,F,P) 上的随机变量，且 Y 和 XY 均可积，X 为 ζ 可测。则

$$E[XY\mid\zeta]=XE[Y\mid\zeta] \text{ 以概率 1 成立 } w.p.1\text{。}$$

其中 ζ 为 F 的 $\sigma-$域。

证明参见 Tucker（1967，pp. 213-214）。

最后我们得出：

定理 5.6 （条件 Jensen 不等式） 设 h 为 R 上的凸函数，X 为 (Ω,F,P) 上的随机变量。ζ 是 F 的一个 $\sigma-$域。若 X 和 $h(X)$ 均可积，则：

$$h(E[X\mid\zeta])\le E[h(X)\mid\zeta] \text{ 以概率 1 成立 } w.p.1\text{。}$$

证明参见 Tucker（1967，p.217）。

最后我们阐述两个重要的事实来结束这一节。

假定 X 为 (Ω,F,P) 上的随机变量，ζ 是 F 的 $\sigma-$域。迄今为止的讨论中，我们已经解释了条件概率的意义。由于 $P[X\in A'\mid\zeta]=P[A\mid\zeta]$，这里 $A=\{\omega: X(\omega)\in A'\}$，$A'\in\Re^1$，因此我们可以将条件概率推广到随机变量情形。我们也曾解释了条件期望的意义，并记作

$E[X \mid \zeta]$。条件概率和条件期望的概念可以用概率分布函数统一起来。我们按照 Billingsley（1979. p. 390）的形式表述如下：

事实 1. 对(Ω, F, P)上的随机变量 X，其条件概率为：

$$P[A \mid \zeta] = P[X \in A' \mid \zeta] = P[\omega : X(\omega) \in A' \mid \zeta]$$

其中 $A' \in \Re^1$，存在一个定义在 $A' \in \Re^1$ 和 $\omega \in \Omega$ 上的函数 $P(A', \omega)$，满足两条性质：

（1）任给一 $\omega \in \Omega$，$P(A', \omega)$ 作为 A' 的函数，为 $\Re^1$ 上的概率测度；

（2）任给一 $A' \in \Re^1$，$P(A', \omega)$ 作为 ω 的函数，为 $P[X \in A \mid \zeta]$ 的一个版本。

概率测度 $P(\cdot, \omega)$ 称为 X 在给定 ζ 条件下的条件分布。

事实 2. 设 $P(\cdot, \omega)$ 为由事实 1 所描述的给定 ζ 条件下 X 的条件分布，且 X 可积。则

$\int_R xP(dx, \omega)$ 为 $E[X \mid \zeta]$ 的一种版本。

见 Billingsley（1979. p. 399）。

6. 鞅及其应用

在给出了条件概率的内容后我们现在来研究鞅论及其在经济与金融中的一些应用。由于鞅论广泛地利用了条件概率，因此上一节的作用很快就会显示出来。我们也将谈到鞅论如同概率论的其他领域一样，在赌博的思维方式中可以找到其根源：设 $X_1, X_2, \ldots$ 是定义在同一概率空间(Ω, F, P)上的一个随机变量序列。在赌博的情况中这些随机变量表示赌徒连续在 $1,2,\ldots,n$ 次赌博后的全部赢余。在赌徒已经完成了 n 次赌博的条件下，记其第 $n+1$ 次赌博后的期望财富为 $E[X_{n+1} \mid X_1, \ldots, X_n]$。由前节可知，$E[X_{n+1} \mid X_1, \ldots, X_n]$ 表示在随机变量 $X_1, \ldots, X_n$ 生成的 $\sigma-$域下，X_{n+1} 的条件期望。直观上讲，由 $X_1, \ldots, X_n$ 生成的 $\sigma-$域包含了赌徒前 n 次赌博的关于其财富的过去所有信息。假如赌博是公平的，那么平均而言，赌徒在第 $n+1$ 次赌博中的期望所得将既不多于也不少于这次赌博前的所得，即：

$$E\left[X_{n+1} \mid X_1, \cdots, X_n\right] = X_n \tag{6.1}$$

换言之，方程（6.1）称为鞅性质，它表示在一个公平的博奕中，赌徒在下一局中的财富从平均意义上来讲就是当前的财富，它与先前的历史无关。

将（6.1）中的等号换作≥或≤，这种博弈就分别成为有利或无利的。由此引出鞅的定义。

6.1 定义

设 $X_1, X_2, \ldots$ 为定义在同一概率空间 (Ω, F, P) 上的随机变量序列。令 $F_1, F_2, \ldots$ 为属于 F 的 $\sigma-$域。序列 $\{(X_n, F_n), n = 1,2,\ldots\}$ 称为鞅，如果对于任意的 n，它满足以下四个条件：

(1) $F_n \subset F_{n+1}$；

(2) X_n 为 F_n 可测；

(3) $E(|X_n|) < \infty$；

(4) $E[X_{n+1} \mid F_n] = X_n$，以概率 1 成立 $w.p.1$。

条件（1）表示 $\{F_n\}, n = 1,2,\ldots$ 为 F 中的一组递增的 $\sigma-$域序列。直观上讲，$\{F_n\}$ 递增的要求意味着 $\sigma-$域序列包含的信息量是递增的，这被称为 F 的 $\sigma-$域序列 $\{F_n\}$ 的单调性。它试图反映这样一个实际想法，即从过去到第 $n+1$ 时期所包含的事件、信息或历史比到第 n 时期要多。由单调递增序列 $\{F_n\}$ 所表现的总体信息结构反映了“学而不忘”的概念。在某些应用中，F_n 是由随机变量 $X_1, \ldots, X_n$ 生成的 $\sigma-$代数，当然也可以由其他随机变量，如 $Y_1, \ldots, Y_n$，所生成的 $\sigma-$代数。在这样的应用中，我们常用 $E[X_{n+1} \mid X_1, \ldots, X_n]$ 或 $E[X_{n+1} \mid Y_1, \ldots, Y_n]$ 等来替代 $E[X_{n+1} \mid F_n]$。

条件（2）中的可测性质指的是对每一个 n, X_n 以可测空间 (Ω, F_n) 作为定义域，$(R, \Re)$ 作为目标空间。有些学者如 Meyer（1966.p.77）用了如下术语：X_n 适应于 $\sigma-$域 F_n。如前所述，当 $\{F_n\}$ 取成由 $X_1, X_2, \ldots X_n$ 生成的 $\sigma-$域，即 $F_n = \sigma(X_1, X_2, \ldots X_n)$ 时，可测条件自动满足。注意，由 $X_1, \ldots, X_n$ 生成的 $\sigma-$域 $\sigma(X_1, \ldots, X_n)$ 是使 $X_1, \ldots X_n$ 可测的最小 $\sigma-$域。假定 Y 也是同一空间上的随机变量，考虑由 $Y, X_1, \ldots, X_n$ 生成的 $\sigma-$域，记为 $\sigma(Y, X_1, \ldots, X_n)$。应有 $\sigma(X_1, \ldots, X_n) \subset \sigma(Y, X_1, \ldots, X_n)$，且 $X_1, \ldots, X_n$ 关于新的 $\sigma-$域 $\sigma(Y, X_1, \ldots, X_n)$ 仍然可测。允许 $\sigma-$域 $\{F_n\}$ 比最小的 $\sigma-$域 $\sigma(X_1, \ldots, X_n)$ 大，这在理论上是显然的，在许多应用中也是恰当的。Fama（1970）利用 $\sigma-$域的大小程度给出了不同的信息级别。

条件(3)，称为可积性，它仅说明 X_n 是可积的，即 X_n 的期望有限。最后，由条件(4)表示的鞅性质指 X_n 为 $E[X_{n+1} | F_n]$ 的一种表示。对于赌博而言,此条件表示赌博是公平的。注意条件（4） 等价于

$$\int_A E[X_{n+1} | F_n]dP = \int_A X_n dP \qquad (6.2)$$

其中 $A \in F_n$， $n = 1,2,\ldots$。而由方程（5.8）可得：

$$\int_A E[X_{n+1} | F_n]dP = \int_A X_{n+1} dP \qquad (6.3)$$

对 $A \in F_n$ 成立。由（6.2）和（6.3）可得：

$$\int_A X_{n+1} dP = \int_A X_n dP \qquad (6.4)$$

对 $A \in F_n$ 成立。因此条件（4）和（6.4）是等价的。由归纳法，同理可得：

$$\int_A X_n dP = \int_A X_{n+1} dP = \cdots = \int_A X_{n+k} dP$$

对 $A \in F_n \subset F_{n+1} \subset \cdots \subset F_{n+k}, k \geq 1$ 成立。这表示 X_n 是 $E[X_{n+k} | F_n]$ 的一个版本。因此条件（4）中的鞅性质也可写为：$E[X_{n+k} | F_n] = X_n$

下面给出一些补充定义。对于不满足条件（3）的非负 X_n，条件（4） 仍然成立。对这样的满足（1）、（2）和（4）但不满足（3）的非负 X_n，我们称 $\{(X_n, F_n), n = 1,2,\ldots\}$ 为广义鞅。若序列 $\{(X_n, F_n), n = 1,2,\ldots\}$ 满足（1），（2），（3）和（4^*）$E[X_{n+1} | F_n] \geq X_n$，以概率 1 成立 $w.p.1$，我们称其为下鞅。若将（4^*）中 $\geq$ 号换为 $\leq$ 号，则 $\{(X_n, F_n), n = 1,2,\ldots\}$ 称为上鞅。利用（6.4）我们可知 $\{(X_n, F_n), n = 1,2,\ldots\}$ 为下鞅，当且仅当

$$\int_A X_{n+1} dP \geq \int_A X_n dP \qquad (6.5)$$

对 $A \in F_n$ 成立。$\{(X_n, F_n), n = 1,2,\ldots\}$ 为上鞅，当且仅当

$$\int_A X_{n+1} dP \geq \int_A X_n dP \qquad (6.6)$$

对 $A \in F_n$ 成立。

6.2 示例

在说明鞅论在经济与金融中的应用之前，利用上述概念并参照 Billingsley（1979. P. 408）和 Doob（1971），我们给出一些具有上述概率性质的例子。

例（1）:设 (Ω, F, P) 为一概率空间，v 为 F 的上的一有限测度，$F_1, F_2, \ldots$ 为 F 的非递减 $\sigma-$ 域序列。假定 v 和 P 限制在 F_n 上时，v 受控于 P，即假定若 $A \in F_n$ 且 $P(A)=0$，则有 $v(A)=0$。由定理 5.1 得知，当 v 和 P 都限制在 F_n 上时，存在 v 对 P 的 Radon-Nikodym 导数，记为 X_n。注意到 X_n 是 F_n 可测且对 P 可积，则由（5.2）得知，对于任意的 $A \in F_n$，X_n 满足：

$$v(A) = \int_A X_n dP \tag{6.7}$$

但是 $A \in F_n$ 意味着 $A \in F_{n+1}$，故 $v(A) = \int_A X_{n+1} dP$。因此（6.4）成立，证得 $\{(X_n, F_n), n = 1,2,\ldots\}$ 为鞅。

例（2）：设 Y 为 (Ω, F, P) 上的可积随机变量，$\{F_n\}$ 为 F 上的非递减 $\sigma-$ 域序列。定义 X_n 为 $X_n = E[Y \mid F_n]$，即 X_n 为 Y 在 $\sigma-$ 域 F_n 条件下的条件期望。由前一节可知 X_n 既是可测的，又是可积的。此外条件（4）成立，因为由定理 5.4 可得：

$$E[X_{n+1} \mid F_n] = E[E[Y \mid F_{n+1}] \mid F_n] = E[Y \mid F_n] = X_n\text{。}$$

因此 $\{(X_n, F_n), n = 1,2,\ldots\}$ 为鞅。这种鞅是在我们了解越来越多的条件时，对 Y 连续求条件数学期望而得。

例（3）：假定 $Y_0, Y_1, Y_2, \ldots$ 为独立随机变量，且 $Y_0 = 0, E(|Y_n|) < \infty, E(Y_n) = 0$，$n \geq 1$。定义 $X_0 = Y_0, X_n = Y_1 + \cdots + Y_n$，$n \geq 1$。则由 Y_n 生成的 $\sigma-$ 域 X_n 为鞅。条件（1）成立，这是因为 $Y_1, \ldots, Y_n$ 生成的 $\sigma-$ 域 $\sigma(Y_1, \ldots, Y_n)$ 包含于由 $Y_1, \ldots, Y_{n+1}$ 生成的 $\sigma-$ 域 $\sigma(Y_1, \ldots, Y_{n+1})$ 中，即 $\sigma(Y_1, \ldots, Y_n) \subset \sigma(Y_1, \ldots, Y_{n+1})$。由于加法运算保持可测性，故条件（2）成立；条件（3）由

$$E(|X_n|) \leq E(|Y_1|) + \cdots E(|Y_n|) < \infty \text{ 可得。}$$

最后，条件（4）也成立，因为

$$
\begin{aligned}
&E[X_{n+1} \mid Y_0,...,Y_n] = E[X_n + Y_{n+1} \mid Y_0,...,Y_n] \\
&= E[X_n \mid Y_0,...,Y_n] + E[Y_{n+1} \mid Y_0,...,Y_n] \\
&= X_n + E(Y_{n+1}) \\
&= X_n
\end{aligned}
$$

注意在证明条件（4）时，我们用到了定理 5.3。还有一点要注意的是由$Y_0, Y_1, Y_2, ...$是独立的随机变量，从而有$E[Y_{n+1} \mid Y_0,...,Y_n] = E[Y_{n+1}]$。

例（4）是下鞅的一个例子：假定$\{(X_n, F_n), n = 1,2,...\}$为鞅，则我们断言$\{(|X_n|, F_n), n = 1,2,...\}$必为下鞅。条件（1）满足是显然的，因为我们考虑的是同一族$\sigma -$域。条件（2）和（3）成立是因为由X_n的可测性和可积性推出$|X_n|$的可测性和可积性。条件（4^*）的成立可由定理 5.3 推得，因为有：

$$E[|X_{n+1}| \mid F_n] \geq |E[X_{n+1}] \mid F_n| = |X_n|$$

6．3 鞅在经济与金融中的应用

鞅的概念已经在经济与金融中得到多处应用，下面我们将给出一些应用示例。

例（1）：本例仿效 Samuelson（1965），我们稍加修改，指出在一定条件下期货定价是一个鞅。

设$..., X_t, X_{t+1}, ..., X_{t+T}, ...$为定义在概率空间$(\Omega, F, P)$上的有界随机变量序列。例如，这个序列可表示价格，如小麦或黄金现货价格的时间序列。X_t表示目前的现货价格，X_{t+T}表示从现在到T单位时间后的现货价格。随机变量有界的假定并不苛刻，因为商品价格总是有限的。假定经纪人至少像了解过去的价格那样了解今天的价格。用概率的语言来说，我们假设经纪人了解由在$\sigma -$域F_t中的过程产生的所有可供信息，其中：

$$F_t = \sigma(X_0, X_1, ..., X_t) \tag{6.8}$$

特别提醒注意的是F_t包含了过去的价格实现过程。这些价格可记做$x_0, x_1, ..., x_{t-1}, x_t$，它们是对特定$\omega \in \Omega$过程的实现值。这里：

$$X_0(\omega) = x_0, ..., X_{t-1}(\omega) = x_{t-1}, X_t(\omega) = x_t$$

经纪人无法确切知道明天的价格X_{t+1}或任意未来价格X_{t+T}。然而，随着时间的推移，他的信息不断增加，这就是因为他观测到了额外的价格现实。自然（6.8）中的F_t是单调递增序

列。我们插入提一下，Fama（1970）引进了术语弱、半强和强信息集来刻画信息集 F_t，它们分别表示包含过程所有的过去值，所有公开的过去信息以及所有的过去信息（公开的和内幕的）。本例中 F_t 至少是弱信息集以保证 X_t 可测。

下面考虑今天的针对后 T 期实际现货价的期货报价。我们用 $Y(T,t)$ 表示在 t 时刻所报，T 期后的期货价格。当一期过后，同一期货价格的新的报价就记为 $Y(T-1,t+1)$。这样我们就得到一个序列：

$$Y(T,t), Y(T-1,t+1),..., Y(T-n,t+n), Y(1,t+T-1) \tag{6.9}$$

Samuelson(1965)的基本假定是期货价格由期末价格的目前期望值竞价决定。这类似于 Muth(1961)的理性预期假设，可表示为：

$$Y(T,t) = E[X_{t+T} \mid F_t] \qquad T = 1,2,... \tag{6.10}$$

注意（6.10）是有道理的，因为 X_{t+T} 关于 F_t 的条件期望存在，这由 X_{t+T} 的可积性而保证。而 X_{t+T} 的可积性可由本例开头处 X_{t+T} 的有界而得。

现在我们来证明（6.9）是鞅。与（6.9）有关的 $\sigma-$ 域为 $F_t, F_{t+1},..., F_{t+n},..., F_{t+T-1}$。由（6.8）定义的这些 $\sigma-$ 域形成一个单调递增序列，使鞅的第一个条件满足。(6.9）的可测性和可积性条件由方程（6.10）和 X_{t+T} 的条件期望性质可得。确切地说，$E[X_{t+T} \mid F_t]$ 是 F_t 可测，而由（6.10）知 $Y(T,t)$ 也是 F_t 可测，此外，$E[X_{t+T} \mid F_t]$ 可积且：

$$\int_A E[X_{t+T} \mid F_t] dP = \int_A X_{t+T} dP < \infty \tag{6.11}$$

对 $A \in F_t$ 成立。由（6.10）知，这对 $Y(T,t)$ 亦成立。(6.11）右端可由 X_{t+T} 的有界假定得到。因此，我们仅需证一条鞅性质：

$$E[Y(T-1,t+1] \mid F_t] = Y(T,t) \tag{6.12}$$

利用（6.10）和定理 5.5 很容易证明方程（6.12）。具体如下：

$$\begin{aligned} E[Y(T-1,t+1] \mid F_t] &= E[E[X_{t+T} \mid F_{t+1}] \mid F_t] \\ &= E[X_{t+T} \mid F_t] \\ &= Y(T,t) \end{aligned}$$

因此，期货定价在所给的假定下为鞅。

本文概述的 Samuelson（1965）的理论文献和 Mandelbrot（1966）的著作激发了计量经济学界对股票价格特性的检验的浓厚兴趣。尽管 Samuelson 的文章所建立的鞅性质是针对期

货价格，而不是股票价格，但是，股票价格可以被看作一系列相继到期的期货报价。这样，鞅性质对股票价格同样适用，它可以作为资本市场有效的一种度量。如果证券价格完全包含了所有可提供的信息，则称资本市场是有效的。这里我们简单介绍 Jensen（1969）的几个发现，有兴趣的读者可参考 Fama（1970）关于有效资本市场的详细的理论调查和实证发现。Jensen 把信息区分开弱和强两种，进而他检验了强鞅假设，其中t时刻T期后的价格X_{t+T}的期望是建立在t时刻的所有可提供的信息条件下。如前所述，这种信息包含了过程的所有过去值以及公开的和厂商的内幕信息。Jensen（1969）将包含所有这样信息的$\sigma-$域记做Θ_t，于是强鞅假设可表示为：

$$E[X_{t+T} \mid \Theta_t] = f(T) X_t \tag{6.13}$$

按 Jensen(1969)的提法，这里的$f(T)$表示正常积累率，它依赖于时期长度T。注意（6.13）实际上表示序列$X_t, X_{t+1}, \ldots, X_{t+T}$为一个下鞅，因为$E[X_{t+T} \mid \Theta_t] \geq X_t$对$f(T) \geq 1$成立。Jensen(1969)的 115 种共同基金的实证分析表明证券的当前价格包含了所有可提供的信息，因此未来价格的最佳预测就是现价加上一个T期正常的预期收益。

作为结束，我们再来研读一下 Samuelson(1965)提到的序列$\ldots, X_t, \ldots, X_{t+T}, \ldots$，以及它们在（6.9）的无条件期望。注意这个序列假定为外生给出的。LeRoy（1973）在外生给出的假设放宽时，试图导出鞅性质。他分析了股票风险和投资者风险厌恶的关系以便形式推导收益率的外生概率分布。LeRoy（1973, P. 437）得出结论：“任意特殊的系统偏离鞅性质这种情形是罕见的，我们所面临的是在风险厌恶下，无法找到鞅性质的严格的理论证明。”在评论 LeRoy 的文章时，Ohlson(1977)指出当投资者具有相对固定的风险厌恶以及红利的百分比变化平稳时，鞅性质成立。

例（2）：作为鞅的期货定价应用可按照 Samuelson（1965）的方法进行推广。令$\alpha = (1+r)^{-1}$，其中r为放弃安全利率的度量，并假定以

$$Y(T,t) = \alpha^T E[X_{t+T} \mid F_t] \tag{6.14}$$

来代替（6.10），即我们允许从现在起T期后价格的条件期望作适当的折现。这样序列$Y(T,t), \ldots, Y(T-n, t+n), \ldots$满足当前（6.14）的折现公理，为关于$\sigma-$域$F_t, \ldots, F_{t+n}, \ldots$的下鞅。这里我们假定$\sigma-$域至少包含弱信息。

为了证明此序列为下鞅，我们只需验证下鞅性质。类似于前述应用，条件（1）、（2）和（3）很容易证明。注意，在经济与金融文献中很少有仔细验证条件（1）、（2）和（3）的，但是读者可以用前面的例子分析作为一个模型来验证单调、可测和可积条件。

下鞅条件可以利用定理 5.5 和方程（6.14）得出如下：

$$\begin{aligned}E[Y(T-1,t+1)\mid F_t] &= E[\alpha^{T-1}E[X_{t+T}\mid F_{t+1}]\mid F_t]\\ &= \alpha^{T-1}E[E[X_{t+T}\mid F_{t+1}]\mid F_t]\\ &= \alpha^{T-1}E[X_{t+T}\mid F_t]\\ &= \alpha^{T-1}\alpha^{-T}Y(T,t)\\ &= \alpha^{-1}Y(T,t)\\ &= (1+r)Y(T,t)\\ &\geq Y(T,t)\end{aligned}$$

在此例中，折现的期货价格每期均以一个等于r的百分比增加。Samuelson（1965）用这个例子给出了现货溢价的合理解释，概要是：期货定价的鞅性质说明，试图利用过去已知价格的信息来构造超常收益模型的所有方法注定是要失败的。

例（3）：在本例中，我们按 Samuelson（1973）的思路，给出在一定条件下，经过折现后的股票价格满足鞅性质。尽管我们的目的是证明鞅性质，但在这之前还是先给出大家所熟悉的资本贴现法则。

令$x_t,\ldots,x_{t+T}$为给定的股票在$t,\ldots,t+T$时刻所支付的红利序列，并将其在$t,\ldots,t+T$时刻再投资。假设折现率r保持恒定。允许r随每单位时间而变化并不增加分析的难度，只是记号复杂而已。开始我们先假定$x_t,\ldots,x_{t+T}$是非随机的，并写出熟知的方程：

$$V_t=\sum_{T=1}^{\infty}\frac{x_{t+T}}{(1+r)^T} \qquad (6.15)$$

这里V_t是由资本贴现法则所确定的t时刻的股票价值。利用（6.15）我们可得出下期股票价值V_{t+1}，如下：

$$V_{t+1}=\sum_{T=1}^{\infty}\frac{x_{t+T+1}}{(1+r)^T}=\sum_{T=2}^{\infty}\frac{x_{t+T}}{(1+r)^{T-1}} \qquad (6.16)$$

利用（6.15）和（6.16）可得：

$$V_t-\frac{V_{t+1}}{1+r}=\frac{x_{t+1}}{1+r}$$

也可表示为：

$$V_{t+1}=(1+r)V_t-x_{t+1} \qquad (6.17)$$

方程（6.17）的作用在于它将股票下期价值表示为本期价值V_t、折现率r和下期末股票红利x_{t+1}的函数。作为特例，若$rV_t=x_{t+1}$，则$V_{t+1}=V_t$，这意味着如果折现率等于收益率$\frac{x_{t+1}}{V_t}$，

则股票价格将保持不变。

利用上述讨论，我们将方程（6.15）和（6.17）推广到随机情形。令（Ω, F, P）为一概率空间，$x_t(\omega),\ldots,x_{t+T}(\omega)$为一股票红利的随机变量序列，以下我们将省略掉$\omega$以简化记号。对于每一个随机变量$x_{t+T}, T=1,2,\ldots$，我们假定其分布和条件期望存在，$F$中的$\sigma-$域序列记为$F_{t+T}$，它至少含有弱信息集，即$\sigma(x_t,\ldots,x_{t+T})\subset F_{t+T}$。这表示投资者至少知道此股票的历史红利。

至此，我们马上可以将（6.15）推广如下：

$$v_t = E[V_t \mid F_t] = E[\sum_{T=1}^{\infty}\frac{x_{t+T}}{(1+r)^T} \mid F_t] = \sum_{T=1}^{\infty}(E[\frac{x_{t+T}}{(1+r)^T} \mid F_t]) \tag{6.18}$$

（6.17）的随机化推广也非常容易。首先，利用（6.16）和（6.17）计算v_{t+1}：

$$v_{t+1} \equiv E[V_{t+1} \mid F_{t+1}] = E[\sum_{T=2}^{\infty}\frac{x_{t+T}}{(1+r)^{T-1}} \mid F_{t+1}] \tag{6.19}$$

其次，注意到由（6.18）、（6.19）和定理5.5我们可得：

$$\begin{aligned}E[v_{t+1} \mid F_t] &= E[E[\sum_{T=2}^{\infty}\frac{x_{t+T}}{(1+r)^{T-1}} \mid F_{t+1}] \mid F_t]\\ &= E[\sum_{T=2}^{\infty}\frac{x_{t+T}}{(1+r)^{T-1}} \mid F_t]\\ &= E[(-x_{t+1}+\sum_{T=1}^{\infty}\frac{x_{t+T}}{(1+r)^{T-1}}) \mid F_t]\end{aligned} \tag{6.20}$$

最后，在（6.20）的第二项中乘以$\frac{1}{1+r}$，并利用（6.18）中v_t的定义，我们得到：

$$\begin{aligned}E[v_{t+1} \mid F_t] &= (1+r)E[\sum_{T=1}^{\infty}\frac{x_{t+T}}{(1+r)^T} \mid F_t] - E[x_{t+1} \mid F_t]\\ &= (1+r)v_t - E[x_{t+1} \mid F_t]\end{aligned} \tag{6.21}$$

方程（6.21）是（6.17）随机化的推广。它可以用来帮助我们确定什么条件下序列$v_t,\ldots,v_{t+T}$为鞅。换句话说我们可以提问：什么条件下$E[v_{t+1} \mid F_t]=v_t$？答案是当 $rv_t = E[x_{t+1} \mid F_t]$时。这表示当折现率等于股票的条件预期收益率时，鞅条件成立。

因此我们得出结论：倘若

$$r = E[x_{t+1+T} \mid F_{t+T}]/v_{t+T} \tag{6.22}$$

则股票的折现期望值序列$v_t, v_{t+T}, T=1,2,\ldots$为鞅。

将最后这个方程中的“=”换作“≥”，我们立即可得$v_t,\cdots,v_{t+T},T=1,2,\ldots$为下鞅。这是由于（6.22）的“=”改为“≥”，并由（6.21）可得：

$$E[v_{t+1+T}\mid F_{t+T}]=v_{t+T}+(r-\frac{E[x_{t+1+T}\mid F_{t+T}]}{v_{t+T}})v_{t+T}\geq v_{t+T}$$

下鞅性质表明下期股票的条件期望值大于或等于它的现值，之所以如此是由于条件预期收益率小于或等于折现率。

我们将给出一个重要的注释以结束本例的讨论。回忆一下，在本例中，我们假定了序列$x_t,\ldots,x_{t+T}$的条件期望是存在的，但我们当时没有解释投资者个人作出他自己预期的方式。更完整的分析要求投资者个人将其主观预期$y_t,\ldots,y_{t+T}$作为一列随机变量，其中$y_{t+T},T=1,2,\ldots$表示从 t 时刻起，T期后的预期收益率。作此要求后，还需要建立y_{t+T}与x_{t+T}的关系。在这一点上，Samuelson（1965）的预期构成公理或 Muth（1961）的理性预期假设可以用来作为一种假设：

$$y_{t+T}=E[x_{t+T}\mid F_t] \tag{6.23}$$

方程（6.23）将投资者的主观预期与市场的客观条件预期联系在一起。这样，我们就可以从投资者的自身主观预期y_{t+T}出发，利用（6.23）将其转为客观预期，再利用x_{t+T}进行如前讨论。

例（4）：本例中，我们将按 Hall（1978）的思路来研究一种简单的瞬态随机优化模型，从而得到消费边际效用的鞅性质。

考虑具有单期严格凹效用函数$u(\cdot)$的一个人，其基于不确定性的生命周期消费问题由下式给出：

$$\max E[\sum_{T=0}^{N}\frac{u(t+T)}{(1+\delta)^T}\mid F_t] \tag{6.24}$$

约束条件为：

$$\sum_{T=0}^{N}\frac{c_{t+T}-w_{t+T}}{(1+r)^T}=A_t \tag{6.25}$$

（6.24）和（6.25）式中的记号分别为：$c_{t+T},T=0,1,\ldots,N$为消费，δ为主观时间偏好利率，r为实际利率，w_{t+T}为收入，A_t为t时刻时资产的价值。该人在时刻t考虑消费问题，其经济生命的长度就是$t+N$期。$\sigma-$域$F_t,F_{t+1},\ldots,F_{t+N}$至少相应地包含关于$t,\ldots,t+N$各期的消费序列和这种消费的边际效用序列的信息。

Hall（1978）给出了这一最大化问题解的必要条件：

$$E[u'(c_{t+1})\,|\,F_t]=\frac{1+\delta}{1+r}u'(c_t) \tag{6.26}$$

其中$u'(\cdot)$为边际效用。假设$\delta=r$，则（6.26）式表明边际效用序列满足鞅性质。若我们假设$r<\delta$，则（6.26）式即说明边际效用序列为下鞅。

Hall（1978）还证明了若一期到下期边际效用随机变动很小，则消费为下鞅，可表示为：

$$E[c_{t+1}\,|\,F_t]=\lambda_t c_t\ge c_t \tag{6.27}$$

其中λ_t由下式给出：

$$\lambda_t=\left\{\frac{1+\delta}{1+r}\right\}^{u'(c_t)/c_t u''(c_t)} \tag{6.28}$$

由（6.28）式，因为$u''<0$及假定$r>\delta$，所以可得$\lambda_t\ge 1$。

在结束本例之前，我们还要提一下，Foldes（1978）为随机最优储蓄的动态离散时间模型建立了鞅条件。

6.4 鞅的基本定理

我们将以定理的形式叙述一些有用的结论，从而结束本节关于鞅的讨论。这些鞅定理尚未广泛应用于经济与金融文献中，它们的提出主要是来自概率论方面。在一些非常优秀的文献中可以找到这些结论以及另外一些鞅论的结果，例如Ash（1972），Billingsley（1979），Doob（1953），Meyer（1966），Neveu（1975）和Tucker（1967）。这里我们用的是Ash（1972）和Billingsley（1979）的成果。

定理6.1

（1） 设$X_1,X_2,\ldots$为(Ω,F,P)上关于$F_1,F_2,\ldots$的鞅。若ϕ为一凸函数且$\phi(X_n)$可积，则$\phi(X_1),\phi(X_2),\ldots$为关于$F_1,F_2,\ldots$的下鞅。

（2） 设$X_1,X_2,\ldots$为(Ω,F,P)上关于$F_1,F_2,\ldots$的下鞅。若ϕ为一非减凸函数且$\phi(X_n)$可积，则$\phi(X_1),\phi(X_2),\ldots$为关于$F_1,F_2,\ldots$的下鞅。

证明：要证明（1）我们只需证明$E[\phi(X_{n+1})\,|\,F_n]\ge\phi(X_n)$。因为$X_n$为鞅，所以$E[X_{n+1}\,|\,F_n]=X_n$。于是$\phi\{E[X_{n+1}\,|\,F_n]\}=\phi(X_n)$。而$\phi$为凸函数，且$X_n$和$\phi(X_n)$可积，由定理5.6所述的条件期望Jensen不等式得：

$$E[\phi(X_{n+1})\,|\,F_n]\ge\phi\{E[X_{n+1}\,|\,F_n]\}=\phi(X_n)。$$

要证明（2）我们只需证明 $E[\phi(X_{n+1})\,|\,F_n] \ge \phi(X_n)$ 。由 X_n 为下鞅得知，$E[X_{n+1}\,|\,F_n] \ge X_n$。由假设 ϕ 为非递减，则 $\phi\{E[X_{n+1}\,|\,F_n]\} \ge \phi(X_n)$ 。再次利用 Jensen 不等式得：

$$E[\phi(X_{n+1})\,|\,F_n] \ge \phi\{E[X_{n+1}\,|\,F_n]\} \ge \phi(X_n)\text{。}$$

定理 6.2 （Kolmogorov 不等式）设 $X_1, X_2, ..., X_n$ 为 (Ω, F, P) 上的下鞅，且 $\lambda > 0$ 。则：

$$P[\max_{i \le n} X_i \ge \lambda] \le \frac{1}{\lambda} E[|X_n|]\text{。}$$

证明：给定 $\lambda > 0$ ，定义 $A_1, ..., A_n, A$ 如下：

$$\begin{aligned}
A_i &= [\omega : X_1(\omega) \ge \lambda] \\
&\vdots \\
A_k &= [\omega : \max_{i<k} X_i(\omega) < \lambda \le X_k(\omega)] \\
A &= \bigcup_{k=1}^{n} A_k = [\omega : \max_{i \le k} X_i(\omega) \ge \lambda]
\end{aligned}$$

注意，上述定义的 A_k 是互不相交的，且由于 $X_1, X_2, ..., X_n$ 为下鞅，则由归纳法可得 $E[X_{n+1}\,|\,F_n] \ge X_n$，对任意的 $k \ge 1$ 都成立。于是：

$$\begin{aligned}
\int_A X_n dP &= \sum_{k=1}^{n} \int_{A_k} X_n dP = \sum_{k=1}^{n} \int_{A_k} E[X_n \,|\, F_k] dP \\
&\ge \sum_{k=1}^{n} \int_{A_k} X_k dP \ge \lambda \sum_{k=1}^{n} P(A_k) = \lambda P(A)
\end{aligned}$$

其中 $A_k \in F_k = \sigma(X_1, ..., X_k)$。因此：

$$\begin{aligned}
\lambda P(A) &= \lambda P[\omega : \max_{i \le n} X_i(\omega) \ge \lambda] \\
&\le \int_A X_n dP \le \int_\Omega X_n^+ dP \le E(|X_n|)
\end{aligned}$$

定理得证。

下面我们给出上穿的概念，这一概念是上穿理论的基础。上穿理论被概率论学者用来证明鞅收敛定理这一重要结果。我们在此给出上穿的概念是由于它可能对金融和经济领域中的研究者有用。

令 $[\alpha, \beta]$ 为一区间，其中 $\alpha < \beta$ ，且 $X_1, X_2, ..., X_n$ 为随机变量。$X_1, X_2, ..., X_n$ 在 $[\alpha, \beta]$ 上的上穿数表示序列从低于 α 开始向上穿过 α 和 β 的次数（见图 6.1）。

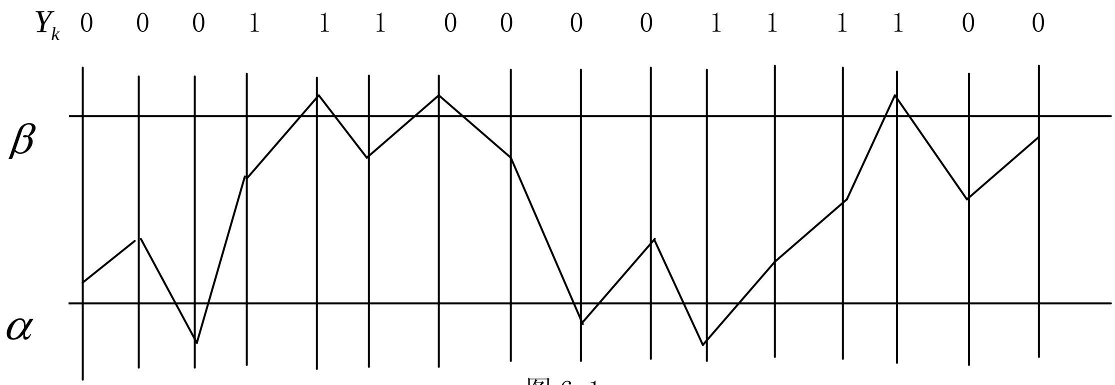

图 6-1

上图中，有两次上穿，其中 $n=16$。这对应于变量 Y 的图形上方相继为 1 的字符串，这里 Y 的定义如下：

$$Y_1 = 0,\ 2 \le k \le n+1$$

$$Y_k = \begin{cases} 0\text{，如果} Y_{k-1} = 1\text{，且} X_{K-1} \ge \beta\text{，} \\ 0\text{，如果} Y_{k-1} = 0\text{，且} X_{K-1} > \alpha\text{，} \\ 1\text{，如果} Y_{k-1} = 1\text{，且} X_{K-1} < \beta\text{，} \\ 1\text{，如果} Y_{k-1} = 0\text{，且} X_{K-1} \le \alpha\text{。} \end{cases}$$

按以上定义，上穿对 $Y_2, Y_3, ..., Y_n$ 而言对应于一串不间断的 1，且其两侧都是 0。现在定义

$$Z_k = \begin{cases} 1 & (Y_k = 1, Y_{k+1} = 0) \\ 0 & \text{其他} \end{cases}$$

则上穿次数为：$U = \sum_{k=2}^{n} Z_k$ 。

令 $F_0 = \{0, \Omega\}, F_k = \sigma(X_1, ..., X_k)$。则 Y_k 关于 F_{k-1} 可测，其中 $k = 1,2,...,n+1$，且 U 也可测。

定理 6.3 （鞅上穿）令 $X_1, X_2, ..., X_n$ 为 (Ω, F, P) 上的下鞅。则 $[\alpha, \beta]$ 上的上穿次数 U 满足 $E[U] \le \dfrac{E(|X_n|) + |\alpha|}{\beta - \alpha}$。

证明参见 Billingsley（1979，p.415）或 Ash（1972，p.291）。

最后给出鞅收敛定理。

定理 6.4 （鞅收敛）设 $X_1, X_2, ...$ 为 (Ω, F, P) 上的下鞅，$K = \sup_n E(|X_n|) < \infty$。则 $X_n \to X \quad w.p.1$，其中 X 是随机变量且满足 $E(|X|) \le K$。

证明参见 Billingsley（1979，p.416）。

7. 随机过程

随机过程是定义在同一概率空间 (Ω, F, P) 上的随机变量 $\{X_t, t \in T\}$ 的集合。注意 $X_t(\omega) \equiv X(t, \omega)$ 以乘积空间 $T \times \Omega$ 作为其定义域，其值域为 R 或 R^k 。其下标或参数集 T 可以认为表示时间。若 T 是可数的，特别是 $T = \{0,1,2,3,...\} \equiv N$ ，即非负整数集，则过程称为离散参数过程。若 $T = R$ 或 $T = [a,b]$，这里 a 和 b 为实数，或 $T = [0,\infty)$，即 T 为不可数的，则我们称其为连续参数过程。尽管下标集 T 可以是任意的，但在本节及本书其余部分，最常用的指标集合为 $T = [0,\infty)$。Ω 表示样本或随机空间。对给定 $\omega \in \Omega$， $X_t(\omega) = X(\cdot, \omega)$， $t \in T$ 称作相应于 ω 的样本路径或样本函数。各种书中所用来描述 ω 固定时 $X_t(\omega)$ 的其他术语有过程的实现或轨道。对固定的 $\omega \in \Omega$，过程通常的记号为 X_t 。然而本书与其他文献一样，有时也用 $X(t)$ 来表示。显然，对固定的 $t \in T$， $X_t(\omega) = X(t, \cdot)$ 为随机变量， X_t 的所有可能值所在的空间称为状态空间。通常状态空间为实直线 R，此时称 X_t 为实值随机过程或简称随机过程。状态空间也可以是 R^k ，此时称 X_t 为 k 维随机过程。前节中所讨论的各种鞅即为随机过程很好的例子。本节中将讨论另外一些例子。尤其地我们将主要讨论 Brownian 运动（或 Wiener 过程），Markov 过程和 Poisson 过程。在此之前，我们先以定义或定理的形式来介绍随机过程的一些基本概念。

7.1 基本概念

随机过程 $\{X_t, t \in T\}$ 的一个重要特征为过程中的随机变量，如 $X_{t_1},..., X_{t_n}$，$t_1,...t_n \in T$ 之间的关系。这种关系可以用这些变量的联合分布函数来确定：

$$P_{X_{t_1},...,X_{t_n}}(H) = P[\omega : (X_{t_1}(\omega),..., X_{t_n}(\omega)) \in H] \tag{7.1}$$

其中 $H \in R^n$ 。首先我们必须指出的是对任意的指标集 T，形如（7.1）的有限维分布族并不能完全确定过程的特征。然而，在随机过程的一般理论中，第一步就是对于给定的有限维分布族来构造随机过程。

假设给定一个有限维分布族如（7.1）的随机过程。注意（7.1）必需满足两个相容性性质。第一条性质是对称条件。令 p 为 $(1,2,...,n)$ 的一个排列，且定义 $f_p : R^n \to R^n$ 如下：

$$f_p(x_{p1},..., x_{pn}) = (x_1,..., x_n) \tag{7.2}$$

由（7.3）和（7.1）的左端可知，随机向量

$$(x_{t_1},...,x_{t_n}) = f_p(x_{t_{p1}},...,x_{t_{pn}}) \tag{7.3}$$

一定具有分布 $P_{X_{t_1},...,X_{t_n}}$ ，而由（7.3）的右端可得 $P_{X_{t_{p1}},...,X_{t_{pn}}} f_p^{-1}$ 。于是导出了对称性条件如下：

$$P_{X_{t_1},...,X_{t_n}} = P_{X_{t_{p1}},...,X_{t_{pn}}} f_p^{-1} \tag{7.4}$$

第二个相容性为一致性条件，记为：

$$P_{X_{t_1},...,X_{t_n}}(H) = P_{X_{t_1},...,X_{t_n},X_{t_{n+1}}}(H \times R^1) \tag{7.5}$$

对 $H \in R^n$ 成立。

上面分析得出的结论是：给定一个随机过程 $\{X_t, t \in T\}$，则它的有限维分布满足性质（7.4）和（7.5）。自然，数学问题就出来了：它的逆成立吗？也就是说，给定一个具有（7.4）和（7.5）性质的有限维分布族是否存在一个随机过程以其作为它的有限维分布族？著名的 Kolmogorov 定理对这个问题做出了肯定回答。

定理 7.1 （kolmogorov）给定一族满足一致性和对称性相容条件的有限维分布族，则存在一个概率空间 $(\Omega, F.P)$ 和一个定义在此空间上的随机过程 $\{X_t, t \in T\}$，使给定的有限维分布族恰为 X_t 的有限维分布族。

证明参见 Billingsley（1979，36 节）。

Kolmogorov 存在性定理为随机过程理论奠定了坚实的基础。接下来，继续给出一些有用的定义。

考虑两个定义在同一概率空间 (Ω, F, P) 上的随机过程 $\{X_t, t \in T\}$ 和 $\{Y_t, t \in T\}$。这两个过程称为随机等价的，如果对所有的 $t \in T$，$P[\omega : X_t(\omega) \neq Y_t(\omega)] = 0$。换句话说，如果对所有的 $t \in T$， $X_t(\omega) = Y_t(\omega)$ 以概率 1 成立，则这两个过程等价。若两个过程等价，则称一个是另一个的表示，且我们可知它们的有限维分布族是相同的。然而等价的过程并不总是具有相同性质的样本轨道。这可以用一个简单的例子来说明，它也证实前面所提及的有限维分布族不能完全确定过程的全部性质。此例是考虑两个过程 $\{X_t, t \geq 0\}$ 和 $\{Y_t, t \geq 0\}$，它们定义在同一概率空间 (Ω, F, P) 上。对 $t \geq 0$，首先定义 X_t 为：

$$X_t(\omega) = 0 \ , \qquad \text{对所有} \ \omega \in \Omega \tag{7.6}$$

再对 $t \geq 0$，定义 Y_t 为：

$$Y_t(\omega)=\begin{cases}1\text{，如果}V(\omega)=t\\0\text{，如果}V(\omega)\neq t\end{cases}\tag{7.7}$$

其中V为(Ω,F,P)上的正随机变量，且具有连续分布，由$P[\omega:V(\omega)=x]=0$，$x>0$给出。对于如上定义的X_t和Y_t，$P[\omega:X_t(\omega)\neq Y_t(\omega)]=0$，$t\geq 0$。因此它们是随机等价的。$X_t$和$Y_t$也具有相同的有限维分布族，如下给出：对 $t_1,...,t_n$，

$$P_{X_{t_1},...,X_{t_n}}(H)=P_{Y_{t_1},...,Y_{t_n}}(H)=\begin{cases}1\text{，如果}H\text{包含}R^n\text{的原点，}\\0\text{，其他，}\end{cases}$$

其中$H\in\Re^n$。然而，两个过程的等价并不能足以保证X_t和Y_t有相同的样本轨道。对$\omega\in\Omega$，由（7.6）式我们得知$X_t(\omega)=0$，而由（7.7）式我们提知$Y_t(\omega)=0$，$Y_t(\omega)$在$t=V(\omega)$处不连续，其值为 1。这样，两个过程的样本轨道是不同的。为修正这种形式的不规则性，概率论专家引进了可分过程的概念。

考虑定义在完备概率空间(Ω,F,P)上的过程$\{X_t,t\in[0,\infty)\}$。如果存在$T=[0,\infty)$的一个可数稠密子集，记做$S=\{t_1,t_2,...\}$，使得对每一个区间$(a,b)\subset[0,\infty)$和每一个闭集$A\subset R$，有：

$$\begin{aligned}&P[\omega:X_t(\omega)\in A,\text{对所有}t\in(a,b)\cap S]\\&=P[\omega:X_t(\omega)\in A,\text{对所有}t\in(a,b)]\end{aligned}\tag{7.8}$$

则称过程$\{X_t,\ t\in[0,\ \infty)\}$是可分的。

注意这个定义要求集合$[X_t\in A,\forall t\in(a,b)\cap S]$与集合$[X_t\in A,\forall t\in(a,b)]$不同的概率为零。此定义的动机是利用可数个时间点集来刻划过程的样本轨道。重要的问题是：对一个给定的具有相容性的有限维分布族的随机过程，是否存在具有相同分布族的该过程的可分表示？幸运的是回答是肯定的。我们从此就可以考虑给定过程的可分表示。关于这一存在性定理的详细表述和证明，可参见 Billingsley（1979，38 节）和 Tucker(1967，8.2 节)。

下面我们来介绍三个重要过程的几个简单结论。

7.2 维纳过程或布朗运动

维纳过程或布朗运动$\{z_t,t\in[0,\infty)\}$是定义在概率空间(Ω,F,P)上的一个随机过程，它满足以下性质：

（1）$z_0(\omega)=0\ \ w.p.1$，即通常我们假定过程从 0 开始。

（2）若$0 \le t_0 \le t_1 \le ... \le t_n$是时间点，则对$H \in \Re^1$，有：

$$P[z_{t_i} - z_{t_{i-1}} \in H_i, i \le n] = \prod_{i \le n} P[z_{t_i} - z_{t_{i-1}} \in H_i]$$

这表示增量过程$z_{t_i} - z_{t_{i-1}}, i \le k$为独立随机变量。

（3）对$0 \le s < t$，增量$z_t - z_s$具有分布：

$$P[z_t - z_s \in H] = \frac{1}{\sqrt{2\pi(t-s)}} \int_H \exp[-\frac{x^2}{2(t-s)} dx$$

这表明每一个增量$z_t - z_s$都服从期望为 0，方差为$\sigma^2(t-s)$的正态分布。这里假定$\sigma = 1$，即我们将其标准化。

（4）对每一个$\omega \in \Omega, z_t(\omega)$在$t \ge 0$时关于$t$连续。

注意，条件（2）反应的是无记忆性。它表示过程在区间$[t_0, t_1], ..., [t_{n-2}, t_{n-1}]$上的变化$z_{t_1} - z_{t_0}, ..., z_{t_{n-1}} - z_{t_{n-2}}$不影响过程在$[t_{n-1}, t_n]$上的变化。过程的历史不影响其未来的位置。过程未来的行为仅依赖于其现在的位置，而与其如何到达是无关的。确切地说，如果$0 \le t_0 < t_1 < ... < t_n < t$，则对于实数$x, x_0, ..., x_n$，有：

$$P[z_t \le x \mid z_{t_0} = x_0, ..., z_{t_n} = x_n] = P[z_t \le x \mid z_{t_n} = x_n] \tag{7.9}$$

等式（7.9）称为马尔可夫性，它将在定义 Markov 过程中起重要作用。这里我们需要强调一点，维纳过程的条件（2）要求独立增量，这实际上比马尔可夫性更强。

为了理解条件（3），我们假定z_t表示一粒子在时刻t位于一固定水平面上的高度。假定$z_t - z_s$期望为 0 这一事实表明粒子向上和向下运动的倾向是一样的，即没有漂移，方差随区间$[s,t]$的长度而增加这一假定则表明粒子有离开它在时刻s所处位置的趋势，而且一旦离开，不存在任何力可以使其恢复到原来的位置。

维纳过程的增量在$z_t - z_s$的分布只依赖于差$t-s$的意义下是平稳的。由性质（1）可知，$z_0 = 0$，从而我们可以这样描述增量的行为：z_t服从正态分布，其均值$E(z_t) = 0$，方差$E(z_t^2) = t$。为计算方差，注意对于$0 \le s < t$，有：

$$
\begin{aligned}
Cov(z_s, z_t) &= E(z_s z_t) = E(z_s z_t - z_s z_s + z_s z_s) \\
&= E[z_s(z_t - z_s) + z_s^2] \\
&= E(z_s[z_t - z_s]) + Ez_s^2 \\
&= Ez_s E(z_t - z_s) + E(z_s^2) = E(z_s^2) = s \\
&= \min\{t, s\}
\end{aligned}
$$

最后，加上条件（4）是因为在许多应用中，连续性是必要的。不过我们马上会给出关于维纳样本轨道微分性质的重要定理。

定理 7.2 （维纳过程的不可微性）设$\{z_t, t \geq 0\}$为(Ω, F, P)上的维纳过程，则对于某个零概率集以外的所有ω，样本轨道$z_t(\omega), t \geq 0$处处不可微。

证明参见 Billingsley（1979，第 37 节）。

直观上讲，处处不可微的样本轨道表示一个粒子的运动在任何时刻都没有速度。因此尽管样本轨道是连续的，但定理 7.2 表明它们非常奇怪，它们的导数处处不存在。有关布朗运动样本轨道结构的完整分析可以参见 *Itô* 和 *McKean* (1974)。

直观上，我们可以给出说明维纳过程不可微性的一种简单方法。由性质（3），对于$0 \leq s < t$，有：

$$
E\left(\frac{z_t - z_s}{t - s}\right)^2 = \frac{\sigma^2}{t - s} = \frac{1}{t - s}
$$

这里假定$\sigma = 1$；取极限：

$$
\lim_{t \to s} E\left(\frac{z_t - z_s}{t - s}\right)^2 = \infty \qquad (7.10)
$$

然而如果假定维纳过程是可微的，我们记其导数为z_s'，则：

$$
\lim_{t \to s} E\left(\frac{z_t - z_s}{t - s} - z_s'\right)^2 = 0
$$

此方程意味着：

$$
\lim_{t \to s} E\left(\frac{z_t - z_s}{t - s}\right)^2 = E(z_s')^2 \qquad (7.11)
$$

于是（7.11）与（7.10）矛盾。

在工程文献中，维纳过程的导数称为白噪声。下一章我们将作进一步讨论。

设$\{z_t, t \geq 0\}$为一维纳过程。我们用它来构造新的过程$\{w_t, t \geq 0\}$如下：

$$
w_t = z_t + \mu t \qquad (t \geq 0) \qquad (7.12)
$$

其中 μ 为常数。于是我们称 $\{w_t, t \geq 0\}$ 为带有漂移的维纳过程或布朗运动，μ 称为漂移参数。在这种情形下维纳过程定义中唯一修改的地方是性质（3），即改为 $w_t - w_s$ 服从均值为 $\mu(t-s)$，方差为 $\sigma^2(t-s)$ 的正态分布，且可设 $\sigma = 1$。

最后，设 w_t 为由（7.12）定义的维纳过程。考虑如下新过程：

$$y_t = \exp(w_t)\,,\ t \geq 0 \tag{7.13}$$

则 $\{y_t, t \geq 0\}$ 称为几何布朗运动或几何维纳过程。

7.3 马尔可夫链和马尔可夫过程

马尔可夫链是指定义在(Ω, F, P)上具有可数状态空间E的随机过程$\{X_t, t=0,1,2,...\}$，且满足性质：对任意的$j \in E$和$t \in T = N \equiv \{0,1,2,...\}$，

$$P[X_{t+1} = j \mid X_0,..., X_t] = P[X_{t+1} = j \mid X_t] \qquad (7.14)$$

方程（7.14）称为马尔可夫性。它表明在给定随机变量$X_0,..., X_t$的过去行为的条件下（即在给定由$X_0,..., X_t$生成的$\sigma-$域$\sigma(X_0,..., X_t) \equiv \sigma(X_n, n \le t)$的条件下），随机变量$X_{t+1}$位于状态$j$的概率等于仅给定由$X_t$所提供的目前信息条件下[即在给定由$X_t$生成的$\sigma-$域$\sigma(X_t)$] X_{t+1}位于状态j的概率。换句话说，马尔可夫性说明当现在已知时，过去和未来是相互独立的。决定过程的下一状态X_{t+1}仅仅是其当前状态X_t，而与其如何到达X_t无关。

在X_t位于状态i的条件下X_{t+1}位于状态j的概率称为一步转移概率，记为：

$$P_{ij}^{t,t+1} \equiv P[X_{t+1} = j \mid X_t = i] \quad ,i，j \in E \qquad (7.15)$$

注意到（7.15）表示概率依赖于初始状态、最终状态以及时间。如果一步转移概率与时间无关，即：

$$P_{ij} = P[X_{t+1} = j \mid X_t = i] \qquad (7.16)$$

与$t \in N$无关，则我们称马尔可夫链为时齐或称马尔可夫链具有平稳转移概率。这些概率可用一个矩阵来表示：

$$P \equiv \begin{bmatrix} P_{11} & P_{12} & P_{13} & \cdots \\ P_{21} & P_{22} & P_{23} & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix} \qquad (7.17)$$

（7.17）式的矩阵P称为马尔可夫链的转移矩阵。首先，（7.7）式的矩阵满足条件$P_{ij} \ge 0$，$i, j \in E$；其次，还满足条件$\sum_{j \in E} P_{ij} = 1$，　$i \in E$成立。

马尔可夫链从状态i经过s步转移到状态j的概率为P的s次幂的第(i, j)号元素。这可表示为：

$$P[X_{t+s}=j\,|\,X_t=i]=P_{ij}^s \tag{7.18}$$

在（7.18）式中，P_{ij}^s 记为 P 的 s 次幂的第 (i,j) 号元素。用矩阵代数的知识 $P^{s+r}=P^sP^r$，我们可知对 $i,j\in E$：

$$P_{ij}^{s+r}=\sum_{k\in E}P_{ik}^sP_{kj}^r \tag{7.19}$$

这称为 Chapman-Kolmogorov 方程。方程（7.19）表示如果马尔可夫链从状态 i 开始，为使其在 $s+r$ 步后到达状态 j，则它一定在 s 步后经过中间状态 k，然后在剩余的 r 步内从 k 到达状态 j。

对于 $i\in E$，定义函数 $\pi_0(i)$ 为：

$$\pi_0(i)=P[X_0=i] \tag{7.20}$$

称其为马尔可夫链的初始分布。首先，它满足：$\pi_0(i)\ge 0$，$i\in E$；其次，$\sum_{i\in E}\pi_0(i)=1$；最后，对于具有转移概率 P_{ij} 的马尔可夫链 X_t，$t\in N$，如果 $i\in E$，$\pi(i)\ge 0$，$\sum_{i\in E}\pi(i)=1$，且：

$$\sum_{i\in E}\pi(i)P_{ij}=\pi(j),\quad j\in E \tag{7.21}$$

则我们称 $\pi(i)$ 为平稳分布。假定平稳分布 π 存在，且：

$$\lim P^t(i,j)\to\pi(j),\ j\in E \tag{7.22}$$

当 $t\to\infty$ 时成立。对任意的初始分布 π_0，可以证明，X_t 的分布当 $t\to\infty$ 时都趋向于 π。这时我们称 π 是一不变状态分布。详细讨论可见 Hoel，Port 和 Stone（1972,第 2 章）。

在简单讨论了马尔可夫链的一些概念后，我们现在给出关于马尔可夫过程的各种定义。

一个定义在概率空间 (Ω,F,P) 上，状态空间为实直线 R 的随机过程称为马尔可夫过程，如果它对于 $0\le s\le t<\infty$ 和 $A\in\Re^1$，有：

$$P[X_t\in A\,|\,\sigma(X_u,u\le s)]=P[X_t\in A\,|\,\sigma(X_s)] \tag{7.23}$$

$w.p.1$成立。

方程（7.23）是方程（7.14）的推广，也称为马尔可夫性。在（7.23）式中左端表示以所有随机变量 $X_u,u\in[0,s]$ 生成的 $\sigma-$域为条件的条件概率，而右端表示仅以随机变量 X_s 生成的 $\sigma-$域为条件的条件概率。类似于前面，确定马尔可夫过程的未来行为的不是它的过去

而是它现在的状态。等价于（7.23）的一个条件是：对于 $n \geq 1, A \in \mathfrak{R}^1$ 和 $0 \leq t_0 < t_1 < ... < t_n < t < \infty$，

$$P[X_t \in A \mid X_{t_0},...,X_{t_n}] = P[X_t \in A \mid X_{t_n}] \tag{7.24}$$

$w.p.1$成立。

对于马尔可夫过程 $X_t, t \geq 0$，我们定义转移概率，记为 $P(s,x,t,A)$ 如下：对于 $0 \leq s < t$，$A \in \mathfrak{R}^1$ 和 $x \in R$，

$$P(s,x,t,A) = P[X_t \in A \mid X_s = x] \quad , \quad w.p.1 \tag{7.25}$$

转移概率的存在性可由本章第 5 节的事实 1 得出。它具有两条性质。首先，对于 $0 \leq s < t$，$A \in \mathfrak{R}^1$，$P(s,\cdot,t,A)$ 为 $\mathfrak{R}^1$ 可测的函数，这里 s,t 和 A 是固定的；其次，对于固定的 s,t 和 $x \in \mathfrak{R}^1$，$P(s,x,t,\cdot)$ 为 $\mathfrak{R}^1$ 上的概率测度。类似于（7.19）的 Chapman-Kolmogorov 方程，对于一个马尔可夫过程，当 $0 \leq s \leq u \leq t < \infty$，$A \in \mathfrak{R}^1$ 时，对于所有的 $x \in R^1$（$x \in N$ 的一个满足 $P[\omega : X_s(\omega) = x \in N] = 0$ 的零概率集可除外），我们有：

$$P(s,x,t,A) = \int_R P(u,y,t,A)P(s,x,u,dy) \tag{7.26}$$

马尔可夫过程称为时齐的，如果其转移概率 $P(s,x,t,A)$ 是时间无关或平稳的，即对任意 $u > 0$，

$$P(s+u,x,t+u,A) = P(s,x,t,A) \tag{7.27}$$

这意味着转移概率为一个仅依赖三个变量 $x, t-s$ 和 A 的函数。这样，Chapman-Kolmogorov 方程可表示为：

$$P(s+t,x,A) = \int_R P(s,y,A)P(t,x,dy) \tag{7.28}$$

齐次马尔可夫过程的标准例子是维纳过程 $z_t, t \geq 0$，其平稳转移概率为：

$$\begin{aligned} P(t,x,A) &= P[z_{s+t} \in A \mid z_s = x] \\ &= \frac{1}{\sqrt{2\pi t}} \int_A \exp\left(-\frac{(y-x)^2}{2t} \right) dy \text{，其中} t \geq 0\text{，} A \in \mathfrak{R}^1\text{。} \end{aligned}$$

本节最后我们讨论泊松过程。

7.4 泊松过程

具有参数λ的泊松过程是一族定义在 (Ω, F, P) 上，状态空间为$N=\{0,1,2,...\}$的随机变量$\{X_t, t\in[0,\infty)\}$，且满足下列三条性质：

（1）$X_0 = 0 \quad w.p.1$。

（2）对于任意的$0<t_1<t_2<...t_n$，增量$X_{t_1}, X_{t_3}-X_{t_2},..., X_{t_n}-X_{t_{n-1}}$相互独立。

（3）对于$0\le s<t<\infty$，增量$X_t - X_s$服从参数为$\lambda(t-s)$的泊松分布，即增量分布如下：

$$P[X_t - X_s = k] = \frac{[\lambda(t-s)]^k}{k!}\exp[-\lambda(t-s)] \text{ ， } k\in N \text{ 。}$$

注意性质（1）只是为方便起见，而性质（2）则可简述为在不同的时间区间内事件个数是独立的。性质（3）可由 Karlin 和 Taylor（1975,p23-26）的两条假设推出。这两条假设如下：

首先，在Δt时间段内至少有一个事件发生的概率为：

$$P(\Delta t) = \lambda\Delta t + o(\Delta t) \tag{7.29}$$

其中$\lambda>0, \Delta t\to 0$。记号$o(\Delta t)$表示当Δt趋于零时，$o(\Delta t)$也趋于 0，但比Δt趋于 0 的速度更快。

其次，在Δt时间区间内，有两个或两个以上的事件发生的概率为$o(\Delta t)$。这一概率非常小，它实质上说明在Δt时间区间内，多于一个事件同时发生的概率为 0。

在第二章的第 12 节，我们将利用（7.29）式。

8. 最优停时

假定连续抛一枚均匀的硬币。在每次抛完后，我们都必须决定是停止还是继续抛下去。令$Y_1, Y_2,...$为表示连续抛投结果的独立随机变量，它们具有相同的概率分布$P(Y_i=1)=P(Y_i=-1)=\frac{1}{2}$，其中$Y_i=1$表示第$i$次为正面，$Y_i=-1$表示第$i$次为反面。如果在第$n$次抛币后停止，那么我们将得到一种收益，记其为$X_n$。$X_n$是最初$n$次抛投结果的函数，即$X_n = f_n(Y_1, Y_2,..., Y_n)$。数学问题是为了使平均收益最大化，我们什么时候应该停止。停时（也称作停止规则或停止变量）是一个在$\{1,2,3,...\}$上取值的随机变量τ，且事件

$\{\tau = n\}$仅依赖于过去值$Y_1,...,Y_n$而与未来值$Y_{n+1},Y_{n+2},...$无关。若τ为一个停时，那么$E(X_\tau)$就可以度量相对于停时τ的平均报酬。记C为使$E(X_\tau)<\infty$的所有停时的集合，且：

$$V = \sup_{\tau\in C}\{E(X_\tau)\} \tag{8.1}$$

则称V为序列$\{X_n\}$的值。如果存在停止规则τ满足：

$$E(X_\tau) = V \tag{8.2}$$

则称τ为最优停时。

为说明最优停时的概念，我们来讨论 Robbins（1970， pp.334-336）的两个例子。在第一个例子中，我们将考虑如下收益函数：

$$X_n = \min\{1, Y_1+\cdots+Y_n\} - \frac{n}{n+1} \tag{8.3}$$

$n\geq 1$，这里 min 表示 1 与$Y_1+\cdots+Y_n$的最小值。作为一个停时，可考虑：

$$\tau = \text{第一个 使}Y_1+\cdots+Y_n = 1\text{的整数}n\text{，}n\geq 1 \tag{8.4}$$

对于如（8.4）的停时，精确计算$E(X_\tau)$是很困难的。但我们知道$E(X_\tau)>0$，因为：

$$E(X_\tau) = E(\min\{1, Y_1,...,Y_\tau\}) - E(\frac{\tau}{\tau+1}) = 1 - E(\frac{\tau}{\tau+1}) > 0 \tag{8.5}$$

为证明由（8.4）给出的停时是最优的，我们下面证明异于（8.4）的任一停时，其相应的$E(X_\tau)$均为负。为此，我们需要一个著名的结果。

引理 8.1（Wald）设$Y_1,Y_2,...$为独立同分布随机变量，且$E(Y_i)\equiv\mu<\infty$。又设τ为序列$Y_1,Y_2,...$的任一停时，且$E(\tau)<\infty$。则$E(\sum_{i=1}^{\tau}Y_i)$总存在，且有：

$$E(\sum_{i=1}^{\tau}Y_i) = \mu\cdot E(\tau) \tag{8.6}$$

证明参见 Wald（1944）或 Shiryayev（1978, p.175）。

为了在我们的例子中应用 Wald 引理，注意$Y_1,Y_2,...$表示一枚均匀硬币的连续抛投结果，是独立同分布的随机变量，其期望为：

$$\mu = E[Y_i] = 1\cdot\frac{1}{2} + (-1)\cdot\frac{1}{2} = 0$$

此外，设τ为满足$E(X_\tau)<\infty$的任一停时，特别地，若$\tau\in C$，则$E(Y_1+\cdots Y_\tau)=0$。这意味着：

$$E(X_{\tau}) \le E(Y_1 + \cdots Y_{\tau}) - E(\frac{\tau}{\tau+1}) \le -\frac{1}{2} \qquad 8.7）$$

因此，对异于（8.4）的所有停时，方程（8.7）说明$E(X_{\tau})<0$，从而对$\tau \in C$的所有$E(X_{\tau})$的上确界就是（8.5）。对于（8.4）给出的τ，由（8.1）得：

$$V = E(X_{\tau}) > 0$$

由此 Robbins（1970，p.335）得出（8.4）为最优停时。

举第二个例子，假设（8.3）中的收益函数改为：

$$X_n = \frac{n2^n}{n+1}\prod_{i=1}^{n}\left(\frac{Y_i+1}{2}\right) \qquad （8.8）$$

其中$n=1,2,...$。注意（8.8）表明如果在抛投n次全是正面后停止，则我们的收益为$n2^n/(n+1)$，否则收益为 0。假如我们已进行到第n次，且全是正面，收益为$n2^n/(n+1)$。在我们停止前，再多抛一次的条件期望收益会是多少呢？答案如下：

$$E\left[X_{n+1}\middle| X_n = \frac{n2^n}{n+1}\right] = \frac{1}{2}\frac{(n+1)2^{n+1}}{n+2} = \frac{2^n(n+1)}{n+2} > X_n \qquad （8.9）$$

方程（8.9）告诉我们不应该停止，因为在X_n给定条件下，多抛一次后的期望收益X_{n+1}比X_n的收益要大。但是，假设我们采用了聪明的决策，进行了另一次抛投，那么这样进行下去，反面终究会出现，这时我们的最终收益将变成 0。因此，每一步聪明的行动并不意味着一个最佳的长期策略。这表明对于（8.8）给出的收益函数，最优停时不存在。然而，停时确实存在。例如，考虑这样一类停时$\{\tau_k\}, k=1,2,...$，其中$\{\tau_k\}$表示无论反正面序列如何出现，都将在k次抛投后停止。对于这样的停止规则，我们有：

$$E(X_{\tau_k}) = \frac{1}{2^k}\frac{k2^k}{k+1} + (1-\frac{1}{2^k})\cdot 0 = \frac{k}{k+1} \qquad （8.10）$$

由（8.10）式知当$k \to \infty$时，利用（8.1）我们可得$V=1$。这样，停时存在，但最优停时不存在。

8.1 数学结论

通过启发给出最优停时的概念后，我们来建立一些基本的数学结果，然后给出源于求职理论和随机资本理论的一些示例。我们的分析主要参照 Chow，Robbins 和 Siegmund（1971）以及 DeGroot（1970）这两本经典著作。

设(Ω, F, P)为一概率空间，且$\{F_n, n=1,2,...\}$为一属于F的递增$\sigma-$域序列。令

$Y_1, Y_2, ...$ 为具有已知联合概率分布函数 F 并定义在 (Ω, F, P) 上的随机变量。假定我们可以依次观测 $Y_1, Y_2, ...$，并记 $X_1, X_2, ...$ 为收益序列。如果我们在第 n 步停止，$X_n = f_n(Y_1, Y_2, ..., Y_n)$。假设 $X_1, X_2, ...$ 关于 $F_1, F_2, ...$ 可测，即 X_n 关于 F_n 可测，$n = 1,2,...$。停时（也可称为停止规则或停时变量）是一个随机变量 $\tau = \tau(\omega)$，它定义在 (Ω, F, P)，目标空间为正整数 $1,2,...$，且满足两个条件。首先：

$$P[\omega : \tau(\omega) < \infty] = 1 \tag{8.11}$$

其次：

$$\{\omega : \tau(\omega) = n\} \in F_n \quad , \quad n = 1,2,\cdots \tag{8.12}$$

等式（8.11）说明停时以概率 1 取有限值，而（8.12）则表示在时刻 n 停时决策仅依赖于包含在 $\sigma -$ 域 F_n 中的过去信息。换句话说，（8.12）表明任何未来信息都不会影响在时刻 n 停时的决策。

尽管在一些应用中，我们可将 F_n 取为 $F_n = \sigma_n(Y_1, Y_2, ..., Y_n)$，但一般情况下，$F_n$ 具有相当的任意性。对任意整数 n 满足 $A \cap \{\tau = n\} \in F_n$ 的所有集合 $A \in F$ 的集合构成 F 中的 $\sigma -$ 域，记作 F_τ。注意 τ 和 X_τ 都是 F_τ 可测的。

对任意的停时 τ，τ 时刻的收益记为 X_τ，它是如下形式的随机变量：

$$X_\tau = \sum_{n=1}^{\infty} X_n I_{\{\tau = n\}} = \begin{cases} X_n \text{，在}\{\tau = n\}\text{上} \\ 0 \quad \text{，其他} \end{cases} \tag{8.13}$$

其中 $n = 1,2,...$。正如（8.1）式所定义，收益序列的值 V 为 $\tau \in C$ 时，$E(X_\tau)$ 的上确界。注意，$I_{\{\tau = n\}}$ 为指标函数，它在集合 $\{\omega : \tau(\omega) = n\}$ 上取值为 1，其他为 0。

对一个给定的停时 τ，收益 X_τ 的期望存在，即 $E(X_\tau) < \infty$，当且仅当

$$E(|X_\tau|) = \sum_{n=1}^{\infty} E[|X_n| \mid \{\tau = n\}] P[\tau = n] < \infty \tag{8.14}$$

利用以上概念，我们现在可以问这样一个问题：在什么条件下最优停时存在？答案由下述定理给出。

定理 8.1 [最优停时的存在性]假设 $X_1, X_2, ...$ 为如上描述的定义在 (Ω, F, P) 上的收益序列，

且满足：

$$E(\left|\sup_n X_n\right|) < \infty \text{，} w.p.1 \qquad (8.15)$$

且当$n \to \infty$时，

$$\lim X_n \to -\infty \text{ ，} \quad w.p.1 \qquad (8.16)$$

则最优停时存在。

证明参见 DeGroot（1970, pp.347-348）。

定理 8.1 的两个充分条件有直观的解释。条件（8.15）说明即使我们能观测到完整的随机变量序列$Y_1, Y_2, \ldots$，然后以最大化收益的原则来选择停时，平均收益仍会是有限的。换句话说，即使能够完全预测，收益也是有限的。但即使有如（8.15）那样的有限平均收益，不停止或许是有利的。条件（8.16）保证我们以概率 1 在某个有限时刻停止。

定理 8.1 十分有用，因为它给出了最优停时存在性的充分条件。尽管如此它并没有告诉我们最优停时的性质。数学上的研究已经获得了在某些特定条件下关于最优停时性质的一些结果。我们依照 DeGroot（1970）来讨论一种重要情形。考虑一个具有如下形式的收益函数：

$$X_n = \max\{Y_1, Y_2, \ldots, Y_n\} - nc \qquad (8.17)$$

其中$n = 1,2,\ldots$，$c > 0$表示每次观测的固定成本。等式（8.17）说明我们在n期的收益为随机变量$Y_1, Y_2, \ldots Y_n$中所观测到的最大值与进行这些观测或抽样所需成本的差，于是我们有下面的定理。

定理 8.2 假设$Y_1, Y_2, \ldots$为(Ω, F, P)上独立同分布且分布函数为F的随机变量序列。令$X_n, n = 1,2,\ldots$为由（8.17）给出的收益序列。如果：

$$E(Y_n^2) < \infty \qquad (8.18)$$

其中$n = 1,2,\ldots$，则存在一个停时，使得$E(X_\tau)$最大化，其形式为：一旦某观察值$Y \geq V$，则停止；如果$Y < V$，则继续，其中V为收益序列的值，它是如下方程的唯一解：

$$\int_V^\infty (Y - V)dF(Y) = c \qquad (8.19)$$

证明参见 DeGroot（1970, p.352）。

8.2 求职

在叙述另外的关于最优停时的定理前，我们先来把到目前为止已经讲过的结果应用于求

职理论中。在这个应用中，我们遵循 Lippman 和 McCall（1976a）。

考虑一个正在找工作的人，他每天都在找工作，直到找到一份工作为止。假设在找工作的过程中，每天恰好有一份工作提供于他。找到每份工作的成本为 $c>0$ 。有两种可能性：一种为恢复抽样，即找到的所有工作都保留；另一种为不可恢复抽样，即如果不接受，就失去该个工作机会。记号上，随机变量 Y_n 为时期 n 的工作薪酬， $n=1,2,\ldots$ 。我们假定求职者知道薪水分布函数 F 的参数。求职者的薪酬 Y_n 就是以该分布函数 F 而随机产生的。为简化分析，我们假定求职参与者为风险中性的，且试图使其期望净收益最大化。求职者要做的决策是何时停止求职而接受一份工作。注意他的收益序列具有如下简单形式：在恢复抽样情况下，如（8.17）所示：

$$X_n=\max\{Y_1,\ldots,Y_n\}-nc$$

$n=1,2,\ldots$ ，而在不可恢复样情形下，有

$$\overline{X}_n=Y_n-nc$$

最后这个方程仅仅表明在不可恢复抽样情形下，求职者的收益为当前提供的薪水与其求职总成本之差。下面我们来讨论在 $Y_1,\ldots,Y_n$ 独立同分布条件下，恢复抽样的情形。

这种求职问题可以利用最优停时来分析。我们继续来讨论最优停时的存在性和性质，以期从分析中可以得到一些新的视点。

为了建立求职问题的最优停时的存在性，由定理 8.1 知，我们只需建立（8.15）和（8.16）这两个条件。这可以由下面的引理完成。在该引理中，不需要 $Y_1,Y_2,\ldots$ 的独立性。

引理 8.2 设 $Y_1,Y_2,\ldots$ 为同分布随机变量序列，分布函数为 F 。又设 $c>0$ 为一个给定的数，并定义

$$Z=\sup_n X_n \tag{8.20}$$

这里 X_n 由方程（8.17）给出。

如果 F 的期望存在，则当 $n\to\infty$ 时， $\lim X_n\to-\infty$ ，以概率 1 成立。

如果 F 的方差有限，则 $E(|Z|)<\infty$ 。

证明参见 DeGroot（1970, pp.350-352）。

这个引理的作用在于帮助我们建立了求职问题最优停时的存在性。注意此引理是证明充分条件（8.15）、（8.16）成立的纯数学结果。特别地，若公共分布 F 的期望和方差存在，则（8.15）和（8.16）成立。因此，假设 $E(Y_n^2)<\infty,n=1,2,\ldots$ ，我们可以利用引理 8.2 和定理

8.1 得出最优停时的存在性。

求职模型中最优停时的结构可由定理 8.2 来刻划。特别地，对任意薪水Y，求职者的最优停止规则的结构或形式为：

$$\begin{array}{ll}\text{接受工作，如果} & Y \ge V; \\ \text{继续寻找，如果} & Y < V。\end{array} \tag{8.21}$$

在求职文献中，（8.21）中的临界值V称为预定薪水，形如（8.21）的任一决策称为具有预定薪水性质的决策。

考虑独立同分布随机变量$Y_1, Y_2, \ldots$序列中的第一个观测点Y_1。由（8.21）表示的最优决策期望收益为$E(\max\{V, Y_1\}) - c$。由V的定义，最优停时中的最优期望收益满足：

$$V = E\max(V, Y_1) - c \tag{8.22}$$

注意到：

$$\begin{aligned} E\max(V, Y_1) &= V\int_0^V dF(Y) + \int_V^\infty YdF(Y) \\ &= V\int_0^V dF(Y) \pm V\int_V^\infty dF(Y) + \int_V^\infty YdF(Y) \\ &= V\int_0^V dF(Y) + V\int_V^\infty dF(Y) + \int_V^\infty YdF(Y) - V\int_V^\infty dF(Y) \\ &= V\int_0^\infty dF(Y) + \int_V^\infty (Y - V)dF(Y) \\ &= V + \int_V^\infty (Y - V)dF(Y) \end{aligned} \tag{8.23}$$

将（8.23）的结果代入（8.22），我们得到定理 8.2 中的方程（8.19）。让我们进一步分析刚刚推导的方程（8.19）。定义：

$$H(V) \equiv \int_V^\infty (Y - V)dF(Y)$$

函数H为凸、非负、严格递减且满足下列性质：

$$\text{当}V \to \infty\text{时}, \lim H(V) \to 0\text{；当}V \to 0\text{时}, \lim H(V) \to E(Y_1)\text{ ；}$$

$$\frac{dH(V)}{dV} = -[1 - F(V)], \quad \frac{d^2H(V)}{dV^2} \ge 0$$

用图形表示，我们可以将 H 表示为如图 8.1

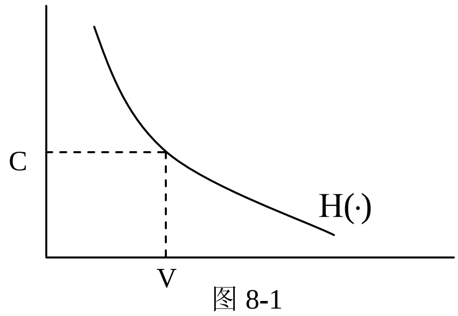

图 8-1

由图 8.1 可以看出，求职成本 $c>0$ 越低，则预留薪水 V 越高，求职的时间也将越长。研究（8.19）中的方程 $H(V)=c$，我们得到一个简单的经济解释，即选取 V 值使得多一次观测的期望边际收益 $H(V)$ 等于多获得一次工作机会的边际成本 c。在这一应用中，通过比较接受一份工作所得到的薪水与恰好再多一个工作机会而得到的期望薪水，求职者表现得缺乏远见。

我们最后再谈一点。在时间为无限以及 $F(\cdot)$ 已知的情形下，恢复抽样与不可恢复抽样的分析没有什么差异。对于后一种情况，求职一直继续，直到预定薪水被后面提供的工作薪酬超过为止。然而需要注意的是，如果时间是有限的或 $F(\cdot)$ 未知，则这两种假设将导致不同的结果，这已由 Lippman 和 McCall（1976a）所证明。

8.3 另外一些数学结果

我们通过说明鞅论对最优停时问题的作用来继续时最优停时的分析。

令 $X_1, X_2, ...$ 为 (Ω, F, P) 上的随机变量序列，并代表收益。令 $F_1, F_2, ...$ 为 F 中的递增 $\sigma-$ 域序列。如前假定 X_n 为 F_n 可测，$n=1,2,...$，$X_n = f_n(Y_1,...,Y_n)$。二元对 $\{(X_n, F_n), n=1,2,...\}$ 叫做随机序列。

如果我们将一个随机序列解释为鞅，那么自然要问，对于停时 τ 是否有 $E(X_1)=E(X_\tau)$。读者回忆一下，鞅性质可以看作代表公平博奕的概念。因此问是否有 $E(X_1)=E(X_\tau)$ 指的是在任意停时 τ 下，公平博奕的性质是否保持。下述的结果探讨了这个问题，它由 Chow,Robbins 和 Siegmund 和 DeGroot（1970）给出。

定理 8.3 设 $\{(X_n, F_n), n=1,2,...\}$ 为 (Ω, F, P) 上的下鞅，τ 为停时，n 是一个正整数。

（1）如果 $P[\tau \le n]=1$，则 $E[X_n \mid F_t] \ge X_\tau$

（8.24）

（2）如果 $P[\tau<\infty]=1$，$E[X_\tau]<\infty$，且：

$$\liminf_n \int_{\{\tau>n\}} X_n^+ dP = 0 \quad (8.25)$$

则对任意的 n，在 $\{\tau\geq n\}$ 上有：

$$E[X_\tau \mid F_n]\geq X_n \quad (8.26)$$

定理证明参见 Chow,Robbins 和 Siegmund（1971, p.21）。

由方程（8.24）我们可知无条件期望满足关系：

$$E(X_n)\geq E(X_\tau)\geq E(X_1) \quad (8.27)$$

若 $P[\tau\leq n]=1$。对上鞅形如（8.27）的结论中的不等号要颠倒。对鞅我们将（8.27）的不等号换为等号即可。

关于条件（8.25），我们指出对一致可积的随机变量序列 $X_1,X_2,\ldots$ 满足该条件。定义如下：称 (Ω,F,P) 上的随机变量序列 $\{X_n,n=1,2,\ldots\}$ 为一致可积的，若当 $\alpha\to\infty$ 时有：

$$\limsup_n \int_{\{|X_n|>\alpha\}} |X_n| dP \to 0 \quad (8.28)$$

注意（8.28）表明 $\sup_n E(|X_n|)<\infty$，特别地我们可以得出 $E(X_\tau)<\infty$ 对停时 τ 成立，故（8.26）成立。

这一小节的最后一个结果是单调性原理。对一随机序列 $\{(X_n,F_n),n=1,2,\ldots\}$，假设其可积，令：

$$A_n=\{E[X_{n+1}\mid F_n]\leq X_n\} \quad (8.29)$$

其中 $n=1,2,\ldots$

$$A_1\subset A_2\subset\ldots\subset A_n\subset\ldots \quad (8.30)$$

且：

$$\bigcup_{n=1}^{\infty} A_n=\Omega \quad (8.31)$$

均成立。则我们称单调性成立。在此情况下，下述定理讲述了最优停时的性质。

定理 8.4 （单调性情形）假定 $\{(X_n,F_n),n=1,2,\ldots\}$ 为满足单调性的随机序列。停止变量 s 如下定义：

$$s = \text{第一个}\ n \geq 1\ \text{使得}\ X_n \geq E[X_{n+1} \mid F_n],$$

只要$E(X_s^-) < \infty$。那么若：

$$\liminf_n \int_{\{s>n\}} X_n^+ = 0$$

成立。则我们可得：

$$E(X_s) \geq E(X_\tau)$$

对任给τ有$E(X_\tau) < \infty$且：

$$\liminf_n \int_{\{\tau \geq n\}} X_n^- = 0$$

证明参见 Chow,Robbins 和 Siegmund（1971,p.55）。

8.4 随机资本理论

我们将遵循 Brock,Rothschild 和 Stiglitz（1979）给出随机资本理论的主要分析，从而结束本节。

资本理论的介绍性演讲经常从下述问题的分析开始：你有一棵树，在t时刻砍倒它的价值为$X(t), t = 0,1,2,...$，若折现率为r，何时才应该砍树？也就是问这棵的树最大现值是多少？这一问题的回答是直接的，即我们选择砍期τ应能最大化$e^{-r\tau} X(\tau)$。指出在$t < \tau$时树的价值为$e^{rt} e^{-r\tau} X(\tau)$。资本理论的大部分内容都可以构筑在这一简单基础上。我们的目的是分析当随机生长时停时估值这一简单问题。假定若在t时刻砍树，树的价值将是$X(t,\omega)$，这里$X(t,\omega)$为一随机过程。何时才应该砍树呢？它的最大现值是多少？我们询问这些问题是因为可以把这种确定情况的分析方法用到许多其他的停时估值问题上。

在我们分析何时砍随机增长的树的问题前，我们先确定两个用于树的生长和估值原则的随机过程。这里我们分析离散时间模型，因为在这种模型中处理某些问题会更容易和更有普遍性。我们假设树的价值遵循离散时间的实值马尔可夫过程，记为X_1，$X_2,\ldots$，X_t。为完善问题的描述，我们一定要弄清楚树的拥有者试图最大化的东西。最简单的假设为他想最大化的是目前的期望折现值。这样，若我们令C为X_t的停时集，问题转化为找$\tau \in C$以最大化$EX(\tau)e^{-r\tau}$。显然更一般的方法应该包含最大化期望效用，$E(U[X(\tau)]e^{-r\tau})$。这仅仅是更一般而已，因为$U(X)$为严格增函数，$U(X(t,\omega))$为一与$X(t,\omega)$本质上有相同属性的马

尔可夫过程，在最大化$E(e^{-r\tau}X(\tau))$和最大化$E(U[X(\tau)]e^{-r\tau})$没有任何分析上的不同。当然，$X(t)$随机性质的解释依赖于$X(t)$是以美元度量还是以效用单位度量。

若折现率为：

$$\beta = \frac{1}{1+r} \tag{8.32}$$

则我们的问题就是选择停时τ以最大化$E(\beta^{\tau}X_{\tau})$，让我们选择一个最简单的过程来确定X_t，即假设：

$$X_{t+1} = X_t + \varepsilon_t \tag{8.33}$$

这里ε_t是期望值为μ的独立同分布随机变量。显然这种条件下最优法则有特别简单的形式：挑选一个树的大小$\hat{X}$，在$X_t \ge \hat{X}$的第一个时间就砍倒它。

若过程X_t为确定的，$X_{t+1} = X_t + \mu$，那么很容易找到最优砍伐尺寸，我们记其为$\overline{X}$，$\overline{X}$一定满足：

$$X = \beta(X + \mu) \tag{8.34}$$

由于（8.34）的左边为此刻砍伐一棵树的价值，所以右边为下期此树的折现值。若$\overline{X}$满足（8.34），则树的拥有者是现在卖掉此树还是继续保有一个时期没有任何差异。而对较小的X，右边的值（保持此树一期）超过左边的值（此刻砍伐的收获）。注意$\overline{X}$为（8.34）的解，则$\mu/\overline{X} = (1-\beta)/\beta = r$或$\overline{X} = \mu/r$，即增长率等于利率。若树的生长改为不确定的这种分析会怎样呢？答案依赖于随机过程X_t是否严格递增。令$\hat{X}$为最优砍伐尺寸，它是满足（8.33）的随机序列。那么，若ε为正，即表示X_t为递增过程，不确定性不影响砍伐尺寸。

定理 8.5 若：

$$P[X_{t+1} > X_t] = 1 \tag{8.35}$$

则 $\overline{X} = \hat{X}$

代替严格性证明我们给出一个启发性的思路。此结果可以由后面更一般的定理 8.7 表示。令$V(X)$为拥有一棵尺寸为X的树的价值，当达到最优尺寸时将会被砍伐。若$\hat{X}$为砍伐尺寸则一定有：

$$V(X) = X \quad 对于 \quad X \geq \hat{X} \tag{8.36}$$

且：

$$V(X) < X \quad 对于 \quad X < \hat{X} \tag{8.37}$$

也就是说在 $\hat{X}$ 处，树的主人此刻砍伐与允许其继续生长到下一期没有任何差异。这就是说，$\hat{X}$ 一定满足：

$$\hat{X} = \beta E(\max\{\hat{X} + \varepsilon, V(\hat{X} + \varepsilon)\}) \tag{8.38}$$

不过，由于 ε 非负，$V(\hat{X} + \varepsilon) = \hat{X} + \varepsilon$ 则（8.38）式变为：$\hat{X} = \beta E(\hat{X} + \varepsilon) = \beta(\hat{X} + \mu)$，即与（8.34）相同。而（8.34）的解唯一，故 $\hat{X} = \overline{X}$ 。

若 ε 为负，这种思路是无法走通的。

定理 8.6 若 $P[\varepsilon < 0] > 0$，则 $\hat{X} > \overline{X}$ 。

这一定理说明了树在不确定条件下比确定条件下大时才会被砍伐。替代严格证明，我们还是给出一个启发性思路。简单地假定 ε 有密度函数 $f(\cdot)$，且具有支撑集 $[-1,1]$。那么（8.38）式变为：

$$\begin{aligned}\hat{X} &= \beta(\int_{-1}^{0} V(\hat{X} + \varepsilon) f(\varepsilon) d\varepsilon + \int_{0}^{1} (\hat{X} + \varepsilon) f(\varepsilon) d\varepsilon) \\ &> \beta(\int_{-1}^{0} (\hat{X} + \varepsilon) f(\varepsilon) d\varepsilon + \int_{0}^{1} (\hat{X} + \varepsilon) f(\varepsilon) d\varepsilon) \\ &= \beta(\hat{X} + \mu)\end{aligned}$$

因此： $\hat{X} > \dfrac{\mu\beta}{1-\beta} = \dfrac{\mu}{r} = \overline{X}$

可以注意这些简单的命题有两个隐含的意义：不确定性能增大树的价值；严格增过程的表现不同于允许递减的过程。

定理 8.6 蕴含在某些情况下的不确定性增大了树的价值。假定我们有一棵尺寸为 $\overline{X}$ 的树。那么若未来增长是确定的，它的价值恰为 $\overline{X}$ 。若未来增长为不确定的，如果树的尺寸可以减小，那么你将不会砍伐它；它的价值超过 $\overline{X}$ 。记确定形式下树的价值为 $V^{d}(X)$，当树的增长增量为（非退化的）随机变量 ε 时记树的价值为 $V^{\varepsilon}(X)$，我们可以得到若 ε 为负时，

$$V^{\varepsilon}(X) > V^{d}(X) \tag{8.39}$$

对某些 X 成立。(连续性暗示(8.39)式对某个 δ， $X \in [\overline{X}-\delta, \overline{X}]$ 成立)。很自然地会问在更一般的环境中(8.39)是否成立。

定理 8.5 和 8.6 的另一层意义是严格递增过程不同于允许递减的过程。我们通过下述定理可知定理 8.5 对很大一类递增过程都是成立的。

定理 8.7 设 X_t 为马尔可夫过程满足：

$$P[X_{t+1} \ge X_t] = 1$$

且：

$$E[\beta X_{t+1} - X_t \mid X_t \ge \overline{X}_t] \le 0，$$

$$E[\beta X_{t+1} - X_t \mid X_t \le \overline{X}_t] \ge 0。$$

这里 $\overline{X}_t$ 为非递增序列。令 $Y_t = \beta^t X_t$。若 $X_t \ge \overline{X}_t$，那么有 $Y_t, Y_{t+1}, Y_{t+2}, \ldots$ 为上鞅。

这一命题表示不确定性不影响砍伐树的时间。由定理 8.3 我们得知，若 τ 为任一停时，那么 $E[Y_\tau \mid X_t \ge \overline{X}_t] \le Y_t$。这就是说当树高首次超过 $\overline{X}_t$ 时砍伐为最优。

证明这个定理我们要证对所有 t：

$$E[Y_{t+\tau+1} - Y_{t+\tau} \mid X_t \ge \overline{X}_t] \le 0。$$

不过由于 $Y_{t+\tau+1} - Y_{t+\tau} = \beta^{t+\tau}(\beta X_{t+\tau+1} - X_{t+\tau})$，注意到

$$\begin{aligned} E[\beta X_{t+\tau+1} - X_{t+\tau} \mid X_t \ge \overline{X}_t] &= E[E[\beta X_{t+\tau+1} - X_{t+\tau} \mid X_{t+\tau}] \mid X_t \ge \overline{X}_t] \\ &= E[E[\beta X_{t+\tau+1} - X_{t+\tau} \mid X_{t+\tau} \ge \overline{X}_{t+\tau}] \mid X_t \ge \overline{X}_t] \\ &\le 0 \end{aligned}$$

就足够了。

9. 各种应用和习题

(1) 设 $\{A_n\}$ 为 $\sigma-$域 F 上的可列集序列。证明 $\bigcap_n A_n \in F$。这一事实可知 $\sigma-$域对可列集的交封闭。又假定 $A \in F, B \in F$，这里 F 为 $\sigma-$域，证 $A - B \in F$。

(2) 令 Z 为正整数集，$\tilde{A}$ 为 Z 的子集 A 组成的集类，满足 A 或 A^c 有限，那么 $\tilde{A}$ 为 $\sigma-$域吗？解释之。

(3) 令 R^1 为实数集，$\Re^1$ 为 $Borel$ 集构成的 $\sigma-$域，集类 $\Re^1$ 包含所有实直线上的开

集、闭集。这表明 $Borel$ 集构成的集类 $\Re^1$ 是相当大的，不过，提醒读者：确实存在 R^1 上的集不属于 $\Re^1$。这种例子见 Billingsley（1979.第 3 节）。

（4） 假设 $\Omega=\{1,2,3,...\}$ 即 Ω 为正整数集，F 为包含 Ω 全部子集的 $\sigma-$ 域。定义 $\mu(A)$ 为 A 中点，即“，”的个数。如果 A 是一个无限集，$\mu(A)=\infty$。那么 μ 为 F 上的测度，它叫计数测度。下面，考虑如上所设的 (Ω,F)，令 $P_1,P_2,...$ 为相对于正整数集的非负数且满足 $\sum_i P_i=1,\ i=1,2,...$，定义 $\mu(A)=\sum_{x_i\in A}P_i$，则 μ 为概率测度。

（5） 概率作为一种特殊的测度，它满足许多条有用的性质。下面即为其中一些性质。令 (Ω,F,P) 为概率空间。

（a）若 $A\in F,B\in F$ 且 $A\subset B$，则 $P(A)\le P(B)$，称为单调性。

（b）若 $A\in F$，则 $P(A^c)=1-P(A)$。

（c）若 $\{A_n\}$ 为 F 上的可列个集序列，则 $P(\cup_n A_n)\le\sum_n P(A_n)$，称为可列次可加性，或叫 Boole 不等式。

（d）若 $\{A_n\}$ 为 F 上递增集序列，A 为其极限，则 $P(A_n)$ 递增至 $P(A)$。证明这一命题可见 Billingsley（1979,第 2 节）和 Tucker（1967, pp.6-7）。

（6） 设 X 和 Y 为定义在同一空间 (Ω,F,P) 上的随机变量，则：

（a）对 $c\in R$，cX 为随机变量。

（b）$X+Y$ 为随机变量，只要 $X(\omega)+Y(\omega)\ne\infty-\infty$ 对每一 ω 成立。

（c）$X\cdot Y$ 为随机变量，只要 $X(\omega)Y(\omega)\ne 0\cdot\infty$ 对每一 ω 成立。

（d）X/Y 为随机变量，只要 $X(\omega)/Y(\omega)\ne\infty/\infty$ 对每一 ω 成立。

（7） 令 X 为 (Ω,F,P) 上具有有限期望 EX 和有限方差 $VarX$ 的随机变量。则对满足 $0<k<\infty$ 的 k，我们有 $P[\omega:|\left|X(\omega)-E(X(\omega))\right|\ge k]\le\frac{VarX}{k^2}$，这叫做 Chebyshev 不等式。表示具有较小方差的随机变量更逼近其期望值。

（8） 令 X 为 (Ω,F,P) 上具有有限期望的随机变量，假定 ϕ 为凸实值函数，满足 $E(\phi(X))<\infty$，则 $\phi(E(X))\le E(\phi(X))$。这称为 Jensen 不等式。它在经济理论

中可以找到若干应用。若消费者喜好风险，则其效用函数为凸，Jensen 不等式说明他的随机效用函数期望值大于或等于随机变量 X 的期望值点的效用。这种情况下消费者将愿意参加公平博奕。假定 ϕ 为凹的实值函数且 $E(X)<\infty, E(\phi(X))<\infty$，则 $\phi(E(X))\geq E(\phi(X))$，这也称为 Jensen 不等式，在经济中它被用来刻划那些通过购买保险而避免参加公平赌博的风险厌恶消费者。Friedman 和 Savage（1948）曾建议效用函数应由凹凸两部分构成。Jensen 不等式在经济中的进一步应用见 Rothschild 和 Stiglitz（1970）。

（9） 令 (Ω, F, P) 为概率空间，$A\in F$，满足 $P(A)>0$。设 $\{B_n\}$ 为有限或可列不相交事件满足 $P(\bigcup_n B_n)=1$ 和 $P(B_n)>0$，对所有 n 成立。则对每一个 k，有：

$$P[B_k \mid A]=P[A\mid B_k]P[B_k]/\sum_n P[A\mid B_n]P[B_n]$$

这叫做 Bayes 定理。利用条件概率，该定理可如下表示：

$$P[B\mid A]=\int_B P[A\mid \widetilde{B}]dP/\int_\Omega P[A\mid \widetilde{B}]dP$$

其中 $B\in\widetilde{B}$，这里 $\widetilde{B}$ 为 F 中的 $\sigma-$域。

（10） 定理 4.2，4.3 和 4.4 可扩展到条件期望，表述如下：

（a）单调收敛定理的条件形式。

令 $\{X_n\}$ 为一递增非负随机变量序列，定义在 (Ω, F, P) 上，设 $X_n(\omega)\to X(\omega)\quad w.p.1$ 且满足 $E(X)<\infty$。则任给一个 F 中的 $\sigma-$域 ζ，当 $n\to\infty$ 时，有：

$$E[X_n\mid\zeta]\to E[X\mid\zeta]\quad, w.p.1$$

（b）Fatou 引理的条件形式。

若 $\{X_n\}$ 为一定义在 (Ω, F, P) 上具有有限期望的非负随机变量序列。当 $n\to\infty$ 时，$E(\liminf_n X_n)<\infty$，则当 $n\to\infty$ 时，有：

$$E[\liminf_n X_n\mid\zeta]\leq\liminf_n E[X_n\mid\zeta]\ \ w.p.1$$

其中 ζ 为 F 中的 $\sigma-$域。

（c）控制收敛定理的条件形式。

令 $\{X_n\}$ 为一随机变量序列，Y 为可积随机变量，均定义在 (Ω, F, P) 上且对所有 n 满足 $|X_n|\leq Y\quad w.p.1$。若 $X_n\to X\quad w.p.1$ 且 ζ 为 F 中的 $\sigma-$域。则当 $n\to\infty$ 时，有：

$$E[X_n \mid \zeta] \to E[X \mid \zeta] \ \ w.p.1。$$

三个定理的证明可以在 Tucker（1967, pp.215-216）找到。

（11） 考虑股票的折现期望值序列 $v_{t+T}, T = 1,2,...$，即如第 6 节的应用（3）所设。假定 $r \ge E[X_{t+T} \mid F_t]/v_t$，股票持有者自然希望了解 v_{t+T} 的最大值且超过 λ 美元的概率，这里 λ 为某正数。找出 v_{t+T} 上的条件使投资者可以计算 $P[\omega : \max_T v_{t+T} \ge \lambda]$。此外，寻找 v_{t+T} 上的条件使 v_{t+T} 在 $T \to \infty$ 收敛到 v。v 是某个表示股票折现期望值的随机变量。

（12） 考虑具有漂移的维纳过程 $\{W_t, t \ge 0\}$，其中 $W_t = z_t + \mu t, \mu \ne 0$，这里 $\{z_t, t \ge 0\}$ 为一标准维纳过程。证明增量 $W_t - W_s$，$0 \le s < t$ 具有期望 $\mu(t-s)$ 和方差 $t-s$。

（13） 设 $\{z_t, t \ge 0\}$ 为一维纳过程，证明 $\{(z_t, F_t), t = 0,1,2,...\}$ 为鞅。这里 $F_t = \sigma(z_0, z_1, ..., z_t)$。

（14） 第 8 节中的求职应用可以推广至允许考虑折现。讨论类似无折现情况，当求职者找到超过一件的工作时，他的期望收益为 $\beta[E(\max(V, Y_1) - c)]$，这表示预订工资 V 为下列方程

$$V = \beta[E(\max(V, Y_1) - c)]$$

的解，这里 $\beta = \dfrac{1}{1+r}$，r 为利率。注意，由最后一个方程所示，我们假定寻职费用发生在期末，工酬也是在期末收到。研究这种应用可知，当利率增加时，保留工资率将降低。详见 Lippman 和 McCall（1976a）。

10. 进一步的注释和参考

有几本非常好的概率论教材。本章中的材料，特别是第 2、3、4、5 节中的内容可以在 Ash（1972）、Billingsley（1979）、Chung（1974）、Loeve（1977）、Neveu（1965）、Papoulis（1965）和 Tucker（1967）的书中找到。Loeve 的是标准教材，现在已是第四版。Tucker 和 Neveu 的简明扼要，Neveu 相对于 Tucker 用了更高级的材料。Ash 在他书中的第一部分发展了测度和积分理论、泛函分析和拓扑学。他随后将这些概念应用到了概率论中。另一方面，Billingsley（1979）注重了分析和由概率问题所导引出的分析问题的概率的相互融合。关于各种收敛性讨论的标准的参考书，可见 Billingsley（1968）。对在微观经济学中应用的一些概率概念的简单介绍可见 McCall（1971）。

也许需要指出，概率论于十七世纪由 Pascal 和 Fermat 等人创建，目的是把机会博弈问题用数学术语公式化。大家今天所知的现代概率论则由 Kolmogorov 在 1933 年建立起稳固的数学基础，他开创性的工作是在 1950 年由苏联介绍到英国的。参见 Kolmogorov（1950）。

第 6 节鞅的定义、例子和定理来自 Ash（1972，第 7.3 节）、Billingsley（1979，第 35 节）、Doob（1953， 第 6 章）、Meyer（1966， B 部）、Neveu（1975）。鞅论两本经典的来源是 Doob（1953）和 Meyer（1966），随后的书基本是前面的拓广。关于这两本经典的书我们还应加上 Neveu（1975）和新版的 Meyer 和 Dellacherie（1978），主要介绍可见于 Doob 的综述文章，它发表在 *Amerian Mathematical Monthly*，可见于 Doob（1971）或 Feller（1971）。Karlin 和 Taylor（1975,第 6 章）关于鞅也有很好的讨论。注意 Doob（1953）用术语半鞅和下半鞅分别代替下鞅和上鞅。

第 6 节中经济和金融中的应用是基于 Samuelson（1965,1973）的两篇文章和 Hall（1978）的论文，也可见于 MaCurdy（1978）和 Malliaris（1981）。有兴趣的读者欲详细讨论可参考 Fama（1970）。Grossman 和 Stiglitz（1980）对有效市场问题提出了建设性的评述并重新定义了有效性概念。信息、鞅、价格等的概念的非技术陈述见 Alchian（1974），鞅性的经济检验和宏观经济鞅的最近综述见 O’Neill（1978）。

尽管在作为鞅的应用的期货定价中我们主要参考 Samuelson（1965），但还有另外值得一读的是 Mandelbrot（1966）的文章，也可见于 Houthakker（1961）的早期论文。Samuelson 和 Mandelbrot 的早期文章激发了近来对鞅概念的兴趣。Mandelbrot 认为 Bachelier（1900）数学上的博士论文“*Théorie de la Spéculation*”，是这一领域的第一篇文章。事实上，Bachelier 发现布朗运动理论比爱因斯坦早了五年，历史所记载的鞅的命名是归功于 Ville（1939）。

鞅性不仅在股票市场，也在别处得到了验证。关于利率可见于 Roll（1970）、Sargent（1972,1976）和 Modigliani 和 Shiller（1973）；而汇率可见于 Cornell（1977），价格期望可见 Mullineaux（1978）和 McNees（1978）。在连续交易理论中，Harrison 和 Pliska（1981）曾证明证券市场是完全的，当且仅当它的价格向量过程有某一鞅表现性质。

与鞅性有关的经济金融文献证明下列几个概念有重要的相互影响：鞅性、随机游动、市场有效性、理性预期、套利价格、稳态投机、价格统计相关。有几篇理论文章尝试清晰表示这些概念的精确关系。这些文章中，我们注意到的有 Fama（1970）、Mandelbrot（1971）、Danthine（1977,1978）、Shiller（1978）和 Lucas（1978）。

随机过程方面的基本参考书是 Doob（1953）的经典著作。关于随机过程的详尽分析也可见于 Gihman 和 Skorohod（1974,1975,1979）三卷书。对这一主题有兴趣的读者可参阅更高级的版本，我们推荐 Karlin 和 Taylor（1975）、Cox 和 Miller（1965）、Prabhu（1965）、Cinlar（1975）和 Hoel,Port 和 Stone（1972）等。Billingsley（1979）和 Tucker（1967）等人的课本中也有一章是关于随机过程的。

历史上，首先被较详尽地讨论的随机过程是布朗运动。英国植物学家 R.Brown 在 1827 年观察到悬浮在液体中的微小颗粒进行着不停止无规则运动。稍后的 1905 年，爱因斯坦用数学化的术语描述了这一过程，观测中的微粒由于与周围溶液中分子的不停碰撞而运动。布朗

运动理论的精确数学公式是由维纳在 1918 年给出的。本书及其他许多书中将布朗运动过程和维纳过程名称交互使用。过去，布朗运动与维纳过程的区别在于前者仅满足定义中的性质（1）、（2）、（3），而后者满足定义中的性质（1）、（2）、（3）、(4)。

马尔可夫序列应归功于俄国数学家 A.A.Markov，他为了努力解决由 D. Billingsley 在 1769 年首先提出的问题，于 1907 年发展了这种思想。另外对马尔可夫序列和马尔可夫过程做出贡献的是 A.Kolmogorov 和 W.Feller。读者可参考的一些基本的参考书有 Karlin 和 Taylor（1975）、Bharucha-Reid（1960）、Cox 和 Miller（1965）、Cinlar（1975）、Hoel,Port 和 Stone（1972）和 Feller（1968）。更高深和详细的书可参考 Dynkin（1965）。

关于最优停时介绍性说明，读者可以参考 Rabbins（1970）的描述性论文，它发表在 American Mathematical Monthly 上，也可见于 Breiman（1964）。正如第 8 节所提到的，我们广泛利用了 DeGroot（1970）和 Chow、Robbins 和 Siegmund（1971）这两本书。在这些书中，几个重要的结果是由 Chow 和 Robbins（1967）、DeGroot（1968）、Dvoretzy（1967）、Yahav（1966）及其他几位数学家首先发现的。第 8 节中我们没有给出关于马尔可夫链和马尔可夫过程的最优停时问题。这一重要问题可参见 Cinlar（1975）和 Dynkin 和 Yushkevich（1969）的经典书籍。更高深的书可参考 Shiryayev（1978）和 Ruiz（1968）。实际上，Shiryayev（1978, 第 2 章）可用来完整证明方程（8.33）之后提出的最优停时问题。Leonardz（1974）的著作是一本关于最优停时的有用书籍，它主要是为对商业应用有兴趣的学生而写的。对于求职模型中的许多最优停时的应用，标准参考书是 Lippman 和 McCall（1976a,1976b,1979）。

最后要提及的是我们本打算把本章的小部分内容作为附录给出。当资料积累到现在的篇幅，我们认为以一章的形式组织比一个长长的附录要更合适。作出这番说明是希望读者以第一章作为参考或基础。很显然，本书的主要问题将在二、三、四章予以讨论。