概率论基础教程

第一章

P29 2.5 等可能结果

例 5d 一个坛子里共有 n 个球，其中一个做了标记. 如果依次从中随机取出 k 个球，那么做了标记的球被取出来的概率有多大?

解：从 n 个球中选取 k 个球，一共有 $\binom{n}{k}$ 种选取方法，每一种选取方法都是等可能的

与事件"做标记的球被取出"相关的选法共有 $\binom{1}{1}\binom{n-1}{k-1}$ 种，因此：

$$P\left(\{\text{做标记的球被取出}\}\right)=\frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}}=\frac{k}{n}$$

也可以这样求解：设 k 个球是顺序地被取出的，用 A_i 表示做标记的球在第 i 次被取出（i=1，…，k）。 既然所有的球在第 i 次被抽取的概率是一样的，那么 $P(A_i)=1/n$. 而这些事件是彼此互不相容的，因此，

$$P(\{\text{做标记的球被取出}\})=P\left(\bigcup_{i=1}^{k}A_i\right)=\sum_{i=1}^{k}P(A_i)=\frac{k}{n}$$

另外，$P(A_i)=1/n$ 可以这样推导：考虑到抽球的过程是有顺序的，一共有 $n(n-1)\cdots(n-k+1)=n!/(n-k)!$ 种等可能试验结果，其中有 $(n-1)(n-2)\cdots(n-i+1)(1)(n-i)\cdots(n-k+1)=(n-1)!/(n-k)!$ 种试验结果表示做标记的球被第 i 次取出，因此

$$P(A_i)=\frac{(n-1)!}{n!}=\frac{1}{n}.$$

P30

例 5e 假设有 $n+m$ 个球，其中 n 个红的，m 个蓝的，将他们随即排成一排，即所有 $(n+m)!$ 种排列都是等可能的。 如果只记录连续排列的球的颜色，证明各种可能的结果的概率是一样的。

解：我们将 $(n+m)$ 个球的次序排列称为一组球的排列，将 $n+m$ 个球的颜色次序排列称为一组球的颜色次序排列。球的排列共有 $(n+m)!$ 种，在红球之间作任何一个位置置换，在蓝球之间作任何一个位置置换，置换的结果并不影响球的颜色次序排列。 从而，一组球的颜色次序排列，对应于 $n!m!$ 个球的排列，这说明球的次序排列也是等可能的，并且每一种颜色次序出现的概率为 $n!m!/(n+m)!$.

例如，假设有 2 个红球，记为r_1，r_2，两个蓝球，记为b_1，b_2，这样，一共有 4! 种球的排列，对于每一种颜色次序排列，对应于$2!2!$个球的排列。例如，下面 4 个球的排列对应于相同的颜色次序排列：

$$r_1,b_1,r_2,b_2 \qquad r_1,b_2,r_2,b_1 \qquad r_2,b_1,r_1,b_2 \qquad r_2,b_2,r_1,b_1$$

因此，每一个颜色次序排列出现的概率都是 4/24=1/6.

例 5f 一手牌有 5 张，如果这 5 张牌是连续的，但又不是同一花色，那么称为顺子. 例如，“黑桃 5，黑桃 6，黑桃 7，黑桃 8，红桃 9”就是一个顺子. 那么一手牌是顺子的概率是多大？

解：假设所有$\binom{52}{5}$种组合都是等可能的。先看看由“A，2，3，4，5”这 5 张牌（花色不同）能组成多少个顺子。 因为，“A”有 4 中可能的花色，同样其他 4 张牌也分别有 4 种可能的花色，所以，一共有 4^5 种可能，但是，其中有 4 种可能是 5 张牌是同花色（这种情况称为同花顺），所以一共是 4^5-4 种顺子。 类似地，“10，J，Q，K，A”这种顺子也有 4^5—4 种，因此一共有$10\times(4^5-4)$种顺子，于是所求概率为：

$$\frac{10\times(4^5-4)}{\binom{52}{5}}\approx 0.0039$$

P33 2.5 等可能结果的样本空间

例 5l 一个俱乐部里有 36 人会打网球，28 人会打软式网球，18 人会打羽毛球; 22 人会打网球和软式网球，12 人会打网球和羽毛球，9 人会打软式网球和羽毛球; 4 人三种球都会打。至少会打一种球的有多少人？

解：记N为俱乐部总人数。设从俱乐部中随机地抽取一人，又假设C为它的任一子集，那么抽到一人刚好在C中的概率为：

$$P(C)=\frac{C\text{中的人数}}{N}$$

设T，S，B分别表示会打网球、软式网球和羽毛球的人的集合，那么利用上述公式以及命题 4.4 可知

$$P(T\cup S\cup B)=P(T)+P(S)+P(B)-P(TS)-P(TB)-P(SB)+P(TSB)$$
$$=\frac{36+28+18-22-12-9+4}{N}=\frac{43}{N}$$

因此，至少会打一种球的人数为 43 人.

P59 3.3 贝叶斯公式

例 3k 一架飞机失踪了，推测它等可能地坠落在 3 个区域。令$1-\beta_i (i=1,2,3)$表示飞机事实上坠落在第i个区域，且被发现的概率（β_i称为忽略概率，因为它表示忽略飞机的概率，通常由该区域的地理和环境条件决定）。已知对区域 1 的搜索没有发现飞机，求在此条件下，飞机坠落在第i $(i=1,2,3)$区域的条件概率。

解： 令$R_i (i=1,2,3)$表示“飞机坠落在第i个区域”这一事件，令E表示“对第 1 个区域的搜索没有发现飞机”这一事件，利用贝叶斯公式可得

$$P(R_1|E)=\frac{P(ER_1)}{P(E)}=\frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3}P(E|R_i)P(R_i)}=\frac{\beta_1\times\frac{1}{3}}{\beta_1\times\frac{1}{3}+1\times\frac{1}{3}+1\times\frac{1}{3}}=\frac{\beta_1}{\beta_1+2}$$

对于$j=2,3$，有

$$P(R_j|E)=\frac{P(E|R_j)P(R_j)}{P(E)}=\frac{1\times\frac{1}{3}}{\beta_1\times\frac{1}{3}+\frac{1}{3}+\frac{1}{3}}=\frac{1}{\beta_1+2}\qquad j=2,3$$

注意，当搜索了第 1 个区域没有发现飞机时，飞机坠落在第j $(j\neq 1)$个区域的更新（即条件）概率会增大，而坠落在第 1 个区域的概率会减小。这是一个常识问题：因为既然在第 1 个区域没有发现飞机，当然飞机坠落在该区域的概率会减小，而坠落在其它区域的概率会增大。而且飞机坠落在第 1 个区域的条件概率是忽略概率β_1的递增函数。当β_1增加时，增大了飞机坠落在第 1 个区域的条件概率。类似的，$P(R_j|E)$ $(j\neq 1)$是β_1的递减函数。

第二章

P166 第 5 章 连续型随机变量

例 4f VAR（Value at Risk）是财务核算中的一个核心概念，投资的 VAR 可以定义为一个值ν，满足投资的损失大于ν的概率只有 1%。如果投资收益X服从正态分布$N(\mu,\sigma^2)$，那么，因为损失是收益的相反数，所以我们有

$$0.01=P\{-X>\nu\}$$

由$-X$服从正态分布$N(-\mu,\sigma^2)$，可得

$$0.01=P\left\{\frac{-X+\mu}{\sigma}>\frac{\nu+\mu}{\sigma}\right\}=1-\Phi(\frac{\nu+\mu}{\sigma})$$

根据正态分布表，$\Phi(2.33)=0.99$，于是我们有

$$\frac{\nu+\mu}{\sigma}=2.33$$

也就是

$$v = VAR = 2.33\sigma - \mu$$

结论是，在所有收益服从正态分布的投资集合中，使$\mu - 2.33\sigma$达到最大值的投资风险最小.

5.6.1 Γ 分布

如果随机变量具有密度函数

$$f(x) = \begin{cases} \dfrac{\lambda \mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

其中，$\Gamma(\alpha)$称为Γ函数，则称该随机变量具有Γ分布，其参数为(α, λ)，$\alpha > 0,\ \lambda > 0$. Γ函数的定义如下：

$$\Gamma(\alpha) = \int_0^\infty \mathrm{e}^{-y} y^{\alpha-1} \mathrm{d}y$$

对$\Gamma(\alpha)$分部积分可得

$$\begin{aligned} \Gamma(\alpha) &= -\mathrm{e}^{-y} y^{\alpha-1} \Big|_0^\infty + \int_0^\infty \mathrm{e}^{-y}(\alpha-1) y^{\alpha-2} \mathrm{d}y \\ &= (\alpha-1)\int_0^\infty \mathrm{e}^{-y} y^{\alpha-2} \mathrm{d}y = (\alpha-1)\Gamma(\alpha-1) \end{aligned}$$

对应于α的积分值，比如说$\alpha = n$，重复利用式（6.1）得到

$$\begin{aligned} \Gamma(n) &= (n-1)\Gamma(n-1) = (n-1)(n-2)\Gamma(n-2) = \cdots \\ &= (n-1)\times(n-2)\times\cdots\times 3\times 2\times\Gamma(1) \end{aligned}$$

又因为$\Gamma(1) = \int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 1$，可得$n$的积分值为

$$\Gamma(n) = (n-1)!$$

当α为一正整数，比方说$\alpha = n$时，参数为 (α, λ)的Γ分布在实践中经常作为某个事件总共要发生n次的等待时间的分布。更具体地说，如果n个事件是随机发生的，且满足 4.7 节的三个公理，那么可以证明要等待某个事件发生共n次的时间是服从参数为(n, λ)的Γ分布的随机变量. 为了证明这点，令T_n表示第n个事件发生的时间，并注意到T_n小于或等于t的充要条件是在时刻t以前至少发生了n个事件，即时间区间$[0, t]$内发生的事件数$N(t) \geq n$。因此

$$P\{T_n \leq t\} = P\{N(t) \geq n\} = \sum_{j=n}^{\infty} P\{N(t) = j\} = \sum_{j=n}^{\infty} \frac{\mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}$$

最后一个等式成立是因为在$[0, t]$内发生的事件数服从参数为λt的泊松分布， 对上式求导得到T_n的密度函数如下

$$f(t)=\sum_{j=n}^{\infty}\frac{e^{-\lambda t}j(\lambda t)^{j-1}\lambda}{j!}-\sum_{j=n}^{\infty}\frac{\lambda e^{-\lambda t}(\lambda t)^{j}}{j!}=\sum_{j=n}^{\infty}\frac{\lambda e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!}-\sum_{j=n}^{\infty}\frac{\lambda e^{-\lambda t}(\lambda t)^{j}}{j!}=\frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$$

因此，T_n 服从参数为 (n,λ) 的 Γ 分布。（在文献中，这个分布也常称为 n-Erlang 分布.）注意，当 $n=1$ 时，该分布退化为指数分布。

参数为 $\lambda=1/2,\alpha=n/2$ 的 Γ 分布（n 为一个正整数）称为自由度为 n 的 χ^2 （读作"卡方"）分布。 实际中，卡方分布常出现在误差分布中，例如，在 n 维空间中试图击中某一靶子，其中各坐标的偏差相互独立且为标准正态分布，则偏差的平方和服从自由度为 n 的 χ^2 分布。第 6 章研究 χ^2 分布，并详细介绍 χ^2 分布与正态分布之间的关系。

第三章

P75 3.5 $P(\bullet|F)$ 是概率

P54 第 3 章条件概率和独立性

例 3a　第一部分　保险公司认为人可以分为两类，一类易出事故，另一类则不易出事故。统计表明，一个易出事故者在一年内发生事故的概率为 0.4，而对不易出事故者来说，这个概率则减少到 0.2，若假定第一类人占人口的比例为 30%，现有一个新人来投保，那么该人在购买保单后一年内将出事故的概率有多大？

解：以这个投保客户是不是易出事故者作为条件，我们将得到所求概率。令 A_1 表示"投保客户一年内将出事故"这一事件，而以 A 表示"投保人为容易出事故者"这一事件，则所求概率 $P(A_1)$ 为

$$P(A_1)=P(A_1|A)P(A)+P(A_1|\bar{A})P(\bar{A})=0.4\times0.3+0.2\times0.7=0.26$$

例 5a　考虑例 3a，保险公司认为人可以分为不同的两类，一类是易出事故的，另一类不易出事故。在任意给定的一年内，易出事故的人将发生事故的概率为 0.4,而对不易出事故的人来说，此概率为 0.2。若已知某新保险客户在第一年已经出过一次事故，问他在保险有效的第二年又出一次事故的条件概率是多大？

解： 如果令 A 表示"该保险客户是易出事故者"这一事件，而 $A_i(i=1,2)$ 表示"他在第 i 年出一次事故"。那么，以他是不是易出事故者为条件，可以算出所求概率 $P(A_2|A_1)$ 如下：

$$P(A_2|A_1)=P(A_2|AA_1)P(A|A_1)+P(A_2|\bar{A}A_1)P(\bar{A}|A_1)$$

而

$$P(A|A_1)=\frac{P(A_1A)}{P(A_1)}=\frac{P(A_1|A)P(A)}{P(A_1)}$$

但是，例 3a 中已经假设 $P(A)=3/10$，且算出了 $P(A_1)=0.26$，因此

$$P(A|A_1)=\frac{0.4\times 0.3}{0.26}=\frac{6}{13}$$

从而

$$P(\bar{A}|A_1)=1-P(A|A_1)=\frac{7}{13}$$

因为 $P(A_2|AA_1)$=0.4， $P(A_2|\bar{A}A_1)$=0.2，所以

$$P(A_2|A_1)=0.4\times\frac{6}{13}+0.2\times\frac{7}{13}\approx 0.29$$

P199 6.2 独立随机变量

例 2d（蒲丰投针问题）桌面上画着一些平行线，它们之间的距离都是 D，向此桌面上随意投掷一长度为 L 的针，其中 $L\le D$，问此针与桌面上的某一根平行线相交的概率是多大（另一种可能是此针正好在某两条平行线之间)?

解： 首先，我们需要确定针的位置：（1）从针的中点向距离该点最近的一条平行线引一条垂直线，设这条垂线的长度为 X；（2）设针与这条垂直线的夹角为 θ，垂直线、平行线以及针所在直线会形成一个直角三角形（见图 6.2). 如果直角三角形的斜边长小于 $L/2$ 时，针会与这一条直线相交. 即如果

$$\frac{X}{\cos\theta}<\frac{L}{2} \quad 或 \quad X<\frac{L}{2}\cos\theta$$

那么针与这一条平行线相交。又因为 X 是取值于 θ 到 $D/2$ 之间，θ 取值于 0 到 $\pi/2$ 之间。 假定 X 和 θ 相互独立且在各自取值范围内服从均匀分布（这个假定是很合理的）。 因此

$$P\left\{X<\frac{L}{2}\cos\theta\right\}=\iint_{X<(L\cos y)/2} f_X(x)f_\theta(y)\mathrm{d}x\mathrm{d}y=\frac{4}{\pi D}\int_0^{\pi/2}\int_0^{(L\cos y)/2}\mathrm{d}x\mathrm{d}y$$
$$=\frac{4}{\pi D}\int_0^{\pi/2}\frac{L}{2}\cos y\mathrm{d}y=\frac{2L}{\pi D}$$

6.7 随机变量函数的联合分布

设 X_1, X_2 是联合连续的随机变量，具有联合密度函数 f_{X_1,X_2}, Y_1，Y_2 为 X_1, X_2 的函数，有时我们需要求出 Y_1，Y_2 的联合分布。具体地说，设 $Y_1=g_1(X_1,X_2)$, $Y_1=g_2(X_1,X_2)$，函数 g_1, g_2 满足下列两个条件：

1.由下列方程组

$$y_1=g_1(x_1,x_2)$$
$$y_2=g_2(x_1,x_2)$$

可唯一地解出 x_1, x_2 来，即求出 $x_1 = h_1(y_1, y_2),\ x_2 = h_2(y_1, y_2)$.

2. 函数 g_1, g_2 对一切 (x_1, x_2) 具有连续偏导数，并且下面的 2 x 2 行列式对一切 (x_1, x_2) 有

$$J(x_1, x_2) = \begin{vmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} \\ \dfrac{\partial g_2}{\partial x_1} & \dfrac{\partial g_2}{\partial x_2} \end{vmatrix} \equiv \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \neq 0$$

在上述两条件之下，可以证明 Y_1，Y_2 的联合密度函数为

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)\left|J(x_1, x_2)\right|^{-1} \tag{7.1}$$

其中 $x_1 = h_1(y_1, y_2),\ x_2 = h_2(y_1, y_2)$.

式（7.1）的证明可从下式入手

$$P\{Y_1 \le y_1, Y_2 \le y_2\} = \iint\limits_{\substack{(x_1,x_2): \\ g_1(x_1,x_2)\le y_1 \\ g_2(x_1,x_2)\le y_2}} f_{X_1,X_2}(x_1, x_2)\,\mathrm{d}\,x_1\,\mathrm{d}\,x_2 \tag{7.2}$$

Y_1，Y_2 联合密度函数可通过对上式关于 y_1, y_2 的偏微商得到。 微商的结果刚好等于（7.1）式的右边。微商的过程作为高等微积分的一个练习不再赘述。

例 7a 设 X_1, X_2 为联合连续的随机变量，其联合密度函数为 f_{X_1,X_2}。令 $Y_1 = X_1 + X_2,\ Y_2 = X_1 - X_2$. 求出 Y_1，Y_2 的联合密度函数。

解： 设 $g_1(x_1, x_2) = x_1 + x_2$，$g_2(x_1, x_2) = x_1 - x_2$，经计算

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

由 $y_1 = x_1 + x_2,\ y_2 = x_1 - x_2$ 解得 $x_1 = (y_1 + y_2)/2,\ x_2 = (y_1 - y_2)/2$. 利用（7.1）可得所求的密度函数是

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1,X_2}\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right)$$

例如，如果 X_1， X_2 为独立同分布的（0,1）均匀随机变量，则

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{1}{2} & 0 \le y_1 + y_2 \le 2,\ 0 \le y_1 - y_2 \le 2 \\ 0 & 其他 \end{cases}$$

又或者，如果 X_1， X_2 为相互独立的指数随机变量，其相应的参数为 λ_1，λ_2，那么

$$f_{Y_1,Y_2}(y_1,y_2)=\begin{cases}\dfrac{\lambda_1\lambda_2}{2}\exp\left\{-\lambda_1(\dfrac{y_1+y_2}{2})-\lambda_2(\dfrac{y_1-y_2}{2})\right\} & y_1+y_2\ge 0,\ y_1-y_2\ge 0\\ 0 & 其他\end{cases}$$

最后，如果 X_1， X_2 为相互独立的标准正态随机变量，则

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{4\pi}\mathrm{e}^{-[(y_1+y_2)^2/8+(y_1-y_2)^2/8]}=\frac{1}{4\pi}\mathrm{e}^{-(y_1{}^2+y_2{}^2)/4}$$

$$=\frac{1}{\sqrt{4\pi}}\mathrm{e}^{-y_1{}^2/4}\frac{1}{\sqrt{4\pi}}\mathrm{e}^{-y_2{}^2/4}$$

因此,我们不仅得知 X_1+X_2 与 X_1-X_2 为正态随机变量,期望为 0,方差为 2,而且, X_1+X_2 与 X_1-X_2 还是相互独立的。(事实上，我们还可以得到如果随机变量 X_1 和 X_2 独立且同分布，其分布函数为 F ，则 X_1+X_2 与 X_1-X_2 相互独立当且仅当 F 是正态分布函数。)

第四章

P106 第 4 章 随机变量

例 4b 某种季节性销售的产品，如果每卖出一件商品，可获得纯利润 b 元，如果季节末仍未卖出，则每件商品将损失 l 元。设某百货商店在某个季节的销售量（即卖出商品的件数）为一随机变量，其分布列为 $p(i)$，$i\ge 0$。商店决定销售旺季前要囤货，请问它要囤多少件才能使得期望利润最大化。

解： 令 X 表示季节需求量，如果囤货数量为 s，记利润为 $\pi(s)$，$\pi(s)$ 可表示为

$$\begin{aligned}\pi(s)&=bX-(s-X)l \qquad 若X\le s\\ &=sb \qquad\qquad\qquad\quad 若X>s\end{aligned}$$

因此，期望利润为

$$\begin{aligned}\mathrm{E}\left[\pi(s)\right]&=\sum_{i=0}^{s}[bi-(s-i)l]p(i)+\sum_{i=s+1}^{\infty}sbp(i)\\ &=(b+l)\sum_{i=0}^{s}ip(i)-sl\sum_{i=0}^{s}p(i)+sb[1-\sum_{i=0}^{s}p(i)]\\ &=(b+l)\sum_{i=0}^{s}ip(i)-(b+l)s\sum_{i=0}^{s}p(i)+sb\\ &=sb+(b+l)\sum_{i=0}^{s}(i-s)p(i)\end{aligned}$$

为了得到最佳的 s 值，我们先考虑当 s 增加一个单位时利润的变化值。利用上述公式得到，当 s 增加一个单位时，期望利润为

$$E[\pi(s+1)] = b(s+1) + (b+l)\sum_{i=0}^{s+1}(i-s-1)p(i)$$
$$= b(s+1) + (b+l)\sum_{i=0}^{s}(i-s-1)p(i)$$

因此，

$$E[\pi(s+1)] - E[\pi(s)] = b - (b+l)\sum_{i=0}^{s}p(i)$$

只要下列条件满足，那么囤货数量为$s+1$得到的期望利润会大于囤货数量为s的情形：

$$\sum_{i=0}^{s}p(i) < \frac{b}{b+l}$$

由于公式（4.1）的左边随着s的增加而增加，而右边为一常数，因此不等式对所有的$s \le s^*$总是成立的，其中s^*为满足（4.1）式的最大值。 因为

$$E[\pi(0)] < \cdots E[\pi(s^*)] < E[\pi(s^*+1)] > E[\pi(s^*+2)] > \cdots$$

这样，囤货数量为s^*+1时将会使得期望利润达到最大。

7.5 条件期望

7.5.1 定义

当X 和Y 的联合分布为离散分布时，对于$P\{Y=y\}>0$ 的y值，给定$Y=y$之下，X 的条件分布由下式定义为

$$p_{X|Y}(x|y) = P\{X=x|Y=y\} = \frac{p(x,y)}{p_Y(y)}$$

因此，很自然地定义，对于所有满足$p_Y(y)>0$的y，X 在给定$Y=y$之下的条件期望为

$$\mathrm{E}[X|Y=y] = \sum_x xP\{X=x|Y=y\} = \sum_x xp_{X|Y}(x|y)$$

例 5a 设X 和Y是独立同分布的二项分布随机变量，其参数为(n,p).计算在$X+Y=m$的条件下X 的条件期望.

解： 首先计算在给定$X+Y=m$的条件下，X 的条件分布列，对于$k \le \min\{n,m\}$，

$$P\{X=k|X+Y=m\}=\frac{P\{X=k,X+Y=m\}}{P\{X+Y=m\}}$$

$$=\frac{P\{X=k,Y=m-k\}}{P\{X+Y=m\}}=\frac{P\{X=k\}P\{Y=m-k\}}{P\{X+Y=m\}}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-m+k}}{\binom{2n}{m}p^m(1-p)^{2n-m}}=\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}$$

其中，我们利用了 $X+Y$ 的分布列是参数为 $(2n,p)$ 的二项分布的事实（参见第 6 章的例 3f）。因此，在给定 $X+Y=m$ 的条件下，X 的条件分布为超几何分布。由例 2g，我们得到

$$\mathrm{E}[X|X+Y=m]=\mathrm{E}[Y|X+Y=m]=\frac{1}{2}\mathrm{E}[X+Y|X+Y=m]=\frac{m}{2}$$

类似地，设 X 和 Y 有连续型联合分布，其联合密度函数为 $f(x,y)$，对于给定的 $Y=y$,只要 $f_Y(y)>0$，X 的条件密度函数定义为

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

很自然地，给定 $Y=y$ 的条件下，X 的条件期望由下式给出

$$\mathrm{E}[X|Y=y]=\int_{-\infty}^{\infty}xf_{X|Y}(x|y)\,\mathrm{d}x$$

此处假定 $f_Y(y)>0$.

例 5b 设 X 和 Y 的联合密度函数为

$$f(x,y)=\frac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y}\qquad 0<x<\infty,0<y<\infty$$

计算 $\mathrm{E}[X|Y=y]$.

解： 先计算条件密度

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)\mathrm{d}x}=\frac{(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{e}^{-y}\mathrm{d}x}=\frac{(1/y)\mathrm{e}^{-x/y}}{\int_0^{\infty}(1/y)\mathrm{e}^{-x/y}\mathrm{d}x}=\frac{1}{y}\mathrm{e}^{-x/y}$$

因此，X 在给定 $Y=y$ 之下的条件分布刚好是均值为 y 的指数分布，所以

$$\mathrm{E}[X|Y=y]=\int_0^{\infty}\frac{x}{y}\mathrm{e}^{-x/y}\mathrm{d}x=y$$

注释 正如条件概率满足概率的所有性质，条件期望也满足通常期望的性质，例如公式

$$\mathrm{E}[g(X)|Y=y]=\begin{cases}\sum_x g(x)p_{X|Y}(x|y) & \text{离散情形}\\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)\,\mathrm{d}x & \text{连续情形}\end{cases}$$

$$\mathrm{E}[\sum_{i=1}^{n} X_i|Y=y]=\sum_{i=1}^{n}\mathrm{E}[X_i|Y=y]$$

仍然成立。 事实上，给定 $Y=y$ 条件下的条件期望可以看成是减小了的样本空间中的普通的期望，这个减小的样本空间由满足 $\{Y=y\}$ 条件的那些样本点组成。

7.5.2 利用条件计算期望

记 $\mathrm{E}[X|Y]$ 表示随机变量 Y 的函数，它在 $Y=y$ 处的值为 $\mathrm{E}[X|Y=y]$，注意 $\mathrm{E}[X|Y]$ 本身是一个随机变量。下面给出的命题是条件期望一个极其重要的性质：

命题 5.1

$$\mathrm{E}[X]=\mathrm{E}[\mathrm{E}[X|Y]] \qquad (5.1)$$

如果 Y 是离散型随机变量，则公式（5.1）变成

$$\mathrm{E}[X]=\sum_y \mathrm{E}[X|Y=y]P\{Y=y\} \qquad (5.1a)$$

如果 Y 是连续型随机变量，密度函数为 $f_Y(y)$，则公式（5.1）变成

$$\mathrm{E}[X]=\int_{-\infty}^{\infty}\mathrm{E}[X|Y=y]f_Y(y)\mathrm{d}y \qquad (5.1b)$$

现在我们给出式（5.1）在 X 和 Y 均为离散型随机变量情形时的证明。

X 和 Y 为离散情形下（5.1）式的证明 我们只需证明

$$\mathrm{E}[X]=\sum_y \mathrm{E}[X|Y=y]P\{Y=y\} \qquad (5.2)$$

等式（5.2）的右边可以写为

$$\begin{aligned}\sum_y \mathrm{E}[X|Y=y]P\{Y=y\} &= \sum_y\sum_x xP\{X=x|Y=y\}P\{Y=y\}\\ &=\sum_y\sum_x x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\}=\sum_y\sum_x xP\{X=x,Y=y\}\\ &=\sum_x x\sum_y P\{X=x,Y=y\}=\sum_x xP\{X=x\}=\mathrm{E}[X]\end{aligned}$$

命题证毕。

可以这样来解释（5.2）式，期望值 $\mathrm{E}[X]$ 可以看成条件期望 $\mathrm{E}[X|Y=y]$ 的加权平均，而权重刚好是事件 $\{Y=y\}$ 的概率。这个结果对计算随机变量的期望是极其重要的，它可以让我们首先很容易地计算某随机变量在给定条件之下的条件期望， 然后再对条件期望求平

均。下面的例子说明了这个公式的用处。

例 5c 一个矿工在井下迷了路，迷路的地方有三个门，从第一个门出来，经过 3 个小时后，可到达安全之处。若从第二个门出去，经过 5 个小时后，他会回到原地。若从第三个门出来，经过 7 个小时后才会到原地。假定工人在任何时候都是随机地选择一个门。问这个工人为了走到安全之处，平均需要多少时间。

解：设 X 表示该矿工为到达安全之处所需的时间（单位：小时），又设 Y 为他选择的门的号码，则

$$\begin{aligned}\mathrm{E}[X] &= \mathrm{E}[X\,|Y=1]P\{Y=1\}+\mathrm{E}[X\,|Y=2]P\{Y=2\}+\mathrm{E}[X\,|Y=3]P\{Y=3\}\\ &=\frac{1}{3}(\mathrm{E}[X\,|Y=1]+\mathrm{E}[X\,|Y=2]+\mathrm{E}[X\,|Y=3]).\end{aligned}$$

然而，

$$\begin{aligned}\mathrm{E}[X\,|Y=1]&=3\\ \mathrm{E}[X\,|Y=2]&=5+\mathrm{E}[X]\\ \mathrm{E}[X\,|Y=3]&=7+\mathrm{E}[X]\end{aligned} \qquad (5.3)$$

在此，我们对（5.3)式作一解释，例如 $\mathrm{E}[X\,|Y=2]$ 的公式，其理由如下：如果矿工选择第二个门，他花 5 个小时后又回到了原地，但回到原地，问题与刚开始时一样，他到达安全地点所需时间为 $\mathrm{E}[X]$。因此 $\mathrm{E}[X\,|Y=2]=5+\mathrm{E}[X]$。式（5.3)的其余各等式的解释是类似。 因此

$$\mathrm{E}[X]=\frac{1}{3}(3+5+\mathrm{E}[X]+7+\mathrm{E}[X])$$

从而

$$\mathrm{E}[X]=15$$

例 5d（随机个数随机变量和的期望）假设在某一天进入百货商店的人数是一个随机变量，其平均值为 50。进一步假定这些顾客在店里花费的钱数是独立且同分布的随机变量，均值为 8 元，并且假定顾客的花钱数与进入百货店的人数也是相互独立的。 试求在这一天百货店营业额的期望值是多少？

解：记 N 为进入百货店的顾客人数，X_i是顾客 i 在店内的消费额，则百货店内消费总量可以表示成$\sum_{i=1}^{N} X_i$，所以有

$$\mathrm{E}\left[\sum_{i=1}^{N} X_i\right]=\mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^{N} X_i\,|N\right]\right]$$

但

$$\mathrm{E}[\sum_{i=1}^{N} X_i \big| N=n] = \mathrm{E}[\sum_{i=1}^{n} X_i \big| N=n] = \mathrm{E}[\sum_{i=1}^{n} X_i] \quad 由X_i与N的独立性$$

$$= n\mathrm{E}[X] \quad 其中\mathrm{E}[X] = \mathrm{E}[X_i]$$

由此可得，

$$\mathrm{E}[\sum_{i=1}^{N} X_i \big| N] = N\mathrm{E}[X]$$

从而

$$\mathrm{E}[\sum_{i=1}^{N} X_i] = \mathrm{E}[N\mathrm{E}[X]] = \mathrm{E}[N]\mathrm{E}[X]$$

因此，当天百货店营业额的期望值为 50 x 8 = 400 元.

例 5e 掷骰子的“Craps”游戏是这样的，每次掷两枚骰子，开始时，如果得到的点数之和是 2, 3 或 12，则玩家输，若得到 7 或 11，则玩家赢；若得到的是其他点数 i，则需继续玩下去，一直到掷出 7 或 i 为止.若玩家最后得到的点数为 7,则玩家输，若最后得到的点数为 i，则玩家赢. 记 R 为掷骰子的次数，求：

（a）$\mathrm{E}[R]$;（b）$\mathrm{E}[R \big| 玩家赢]$;（c）$\mathrm{E}[R \big| 玩家输]$.

解：如果令 P_i 表示每次掷骰子得到两枚骰子点数之和为 i 的概率，则有

$$P_i = P_{14-i} = \frac{i-1}{36} \quad i = 2, 3, \cdots, 7$$

为求 $\mathrm{E}[R]$，记 S 为第一次掷出点数，则给定 S 的条件下，有

$$\mathrm{E}[R] = \sum_{i=2}^{12} \mathrm{E}[R \big| S=i] P_i$$

其中，

$$\mathrm{E}[R \big| S=i] = \begin{cases} 1 & i = 2,3,7,11,12 \\ 1 + \dfrac{1}{P_i + P_7} & 其他 \end{cases}$$

在上式中，若第一次得到 i，$i \neq 2,3,7,11或12$，则玩家必须继续进行直到出现 i 或 7 为止，此时所需掷骰子的次数服从几何分布，参数为 $P_i + P_7$。所以，掷骰子的期望次数为 $\dfrac{1}{P_i + P_7}$ +1,其中+1 表示加上第一次掷骰子，因此有

$$\mathrm{E}[R] = 1 + \sum_{i=4}^{6} \frac{P_i}{P_i + P_7} + \sum_{i=8}^{10} \frac{P_i}{P_i + P_7} = 1 + 2(3/9 + 4/10 + 5/11) = 3.376$$

为求 $\mathrm{E}[R \big| 赢]$，先来计算玩家赢的概率 p。 给定第一次掷骰子的结果 S 的条件下，有

$$p=\sum_{i=2}^{12}P\left\{赢\middle|S=i\right\}P_i=P_7+P_{11}+\sum_{i=4}^{6}\frac{P_i}{P_i+P_7}P_i+\sum_{i=8}^{10}\frac{P_i}{P_i+P_7}P_i=0.493$$

其中上式用到事实：i 在 7 之前出现的概率为 $P_i/(P_i+P_7)$. 现在需要确定在玩家赢的条件下 S 的条件概率，记 $Q_i=P\left\{S=i\middle|赢\right\}$，我们有

$$Q_2=Q_3=Q_{12}=0,\quad Q_7=P_7/p,\ Q_{11}=P_{11}/p$$

对于 $i=4,5,6,8,9,10$

$$Q_i=\frac{P\left\{S=i,赢\right\}}{P\left\{赢\right\}}=\frac{P_iP\left\{赢\middle|S=i\right\}}{p}=\frac{P_i^2}{p(P_i+P_7)}$$

对第一次掷出的点数和取条件可得

$$\mathrm{E}[R|赢]=\sum_i\mathrm{E}[R|S=i]Q_i$$

在第 6 章例 2j 已经指出，已知 $S=i$ 的条件之下，需要掷多少次骰子与最后的结果是赢或输是相互独立的.(可以这样来看这个事实，在需要掷的次数为 R 的条件下，是赢是输的概率与已经掷了几次是无关的，再利用事件独立性的对称特性，即 A 独立于事件 B, 则 B 也独立于事件 A, 可以推出在输赢已知的条件下，R 的分布与输赢也是无关的.) 因此有

$$\mathrm{E}[R|赢]=\sum_i\mathrm{E}[R|S=i]Q_i=1+\sum_{i=4}^{6}\frac{Q_i}{P_i+P_7}+\sum_{i=8}^{10}\frac{Q_i}{P_i+P_7}=2.938$$

尽管我们可以仿照 $\mathrm{E}[R|玩家赢]$ 的计算方法来求 $\mathrm{E}[R|玩家输]$，但是还有一个更简单的方法，就是利用

$$\mathrm{E}[R]=\mathrm{E}[R|赢]p+\mathrm{E}[R|输](1-p)$$

由此可得

$$\mathrm{E}[R|输]=\frac{\mathrm{E}[R]-\mathrm{E}[R|赢]p}{1-p}=3.801$$

例 5g 考虑 n 次独立重复试验，每次试验的结果为 $1,2,\cdots,k$，相应的概率为 $p_1,\cdots,p_k$，$\sum_{i=1}^{k}p_i=1$. 令 N_i 表示试验中结果 i 发生的次数，$i=1,2,\cdots,k$. 对任意 $i\neq j$，计算

(a) $\mathrm{E}[N_j\mid N_i>0]$；(b) $\mathrm{E}[N_j\mid N_i>1]$

解： 对于（a)，令

$$\mathrm{I}=\begin{cases}0, & N_i=0\\ 1, & N_i>0\end{cases}$$

那么 $\mathrm{E}[N_j]$ 可以写成

$$\mathrm{E}[N_j] = \mathrm{E}[N_j \mid I=0]P\{I=0\} + \mathrm{E}[N_j \mid I=1]P\{I=1\}$$

或等价地

$$\mathrm{E}[N_j] = \mathrm{E}[N_j \mid N_i=0]P\{N_i=0\} + \mathrm{E}[N_j \mid N_i>0]P\{N_i>0\}$$

因为 N_j 的无条件分布是参数为 (n, p_j) 的二项分布，设 $N_i = r$ 给定，则其余的 $n-r$ 次试验的结果不会是 i，并且相互独立，是 j 的概率为 $P(j \mid \text{不是} i) = \dfrac{p_j}{1-p_i}$。 因此 N_j 在给定 $N_i = r$ 的条件下的条件分布为二项分布，参数为 $(n-r, \dfrac{p_j}{(1-p_i)})$，关于推导的细节可参见第 6 章例 4c。又因为 $P\{N_i=0\} = (1-p_i)^n$，上面 $\mathrm{E}[N_j]$ 的等式变成

$$np_j = n\frac{p_j}{1-p_i}(1-p_i)^n + \mathrm{E}[N_j \mid N_i>0][1-(1-p_i)^n]$$

从而

$$\mathrm{E}[N_j \mid N_i>0] = np_j\frac{1-(1-p_i)^{n-1}}{1-(1-p_i)^n}$$

对于（b）的讨论，方法是类似的，令

$$J = \begin{cases} 0, & N_i = 0 \\ 1, & N_i = 1 \\ 2, & N_i > 1 \end{cases}$$

则有

$$\mathrm{E}[N_j] = \mathrm{E}[N_j \mid J=0]P\{J=0\} + \mathrm{E}[N_j \mid J=1]P\{J=1\} + \mathrm{E}[N_j \mid J=2]P\{J=2\}$$

或等价地

$$\mathrm{E}[N_j] = \mathrm{E}[N_j \mid N_i=0]P\{N_i=0\} + \mathrm{E}[N_j \mid N_i=1]P\{N_i=1\} + \mathrm{E}[N_j \mid N_i>1]P\{N_i>1\}$$

由这个公式可以导出

$$\begin{aligned} np_j &= n\frac{p_j}{1-p_i}(1-p_i)^n + (n-1)\frac{p_j}{1-p_i}np_i(1-p_i)^{n-1} \\ &\quad + \mathrm{E}[N_j \mid N_i>1][1-(1-p_i)^n - np_i(1-p_i)^{n-1}] \end{aligned}$$

最后得到

$$\mathrm{E}[N_j \mid N_i>1] = \frac{np_j[1-(1-p_i)^{n-1} - (n-1)p_i(1-p_i)^{n-2}]}{1-(1-p_i)^n - np_i(1-p_i)^{n-1}}$$

类似地，也可以利用取条件的方法计算随机变量的方差。我们通过下面一个例子来介绍这种方法

例 5i 假设有 r 个玩家在赌博，玩家 i 最初拥有 n_i 单位赌资，$n_i > 0,\ i = 1, \cdots, r$. 在每一个阶段，两名玩家来玩一局，赢家从输家那里赢得一单位赌资。当玩家的财富值变为 0 时该玩家就被淘汰，游戏继续直到只有一个玩家拥有所有赌资 $n \equiv \sum_{i=1}^{r} n_i$ 时，那名玩家就是胜

利者。假设每场对局是独立的并且每局两名玩家获胜的机会是相等的，那么只有一名玩家得到所有的n单位赌资时的平均赌博局数是多少？

解：要求对局的平均局数，首先假设起初只有 2 名玩家。 玩家 1 和玩家 2 的最初的赌资分别为j和$n-j$个单位。 记X_j表示将要进行的对局数，$m_j = \mathrm{E}[X_j]$，对$j=1,\cdots,n-1$，有

$$X_j = 1 + A_j$$

A_j是在第一局之后还需要附加的对局数，取期望后得

$$m_j = 1 + \mathrm{E}[A_j]$$

在给定第一局的结果为条件时，得到

$$m_j = 1 + \mathrm{E}[A_j \mid \text{玩家1赢了第一局}] \cdot 1/2 + \mathrm{E}[A_j \mid \text{玩家2赢了第一局}] \cdot 1/2$$

现在，如果玩家 1 赢了第 1 局，情况就与假设玩家 1 初始时拥有$j+1$单位赌资而玩家 2 初始时拥有$n-(j+1)$单位赌资的情形相同，所以，

$$\mathrm{E}[A_j \mid \text{玩家1赢了第一局}] = m_{j+1}$$

类似地

$$\mathrm{E}[A_j \mid \text{玩家2赢了第一局}] = m_{j-1}$$

所以，

$$m_j = 1 + \frac{1}{2} m_{j+1} + \frac{1}{2} m_{j-1}$$

或者等价地，

$$m_{j+1} = 2m_j - m_{j-1} - 2 \qquad j = 1,\cdots,n-1 \qquad (5.4)$$

利用$m_0 = 0$，由上式可得

$$\begin{aligned} m_2 &= 2m_1 - 2 \\ m_3 &= 2m_2 - m_1 - 2 = 3m_1 - 6 = 3(m_1 - 2) \\ m_4 &= 2m_3 - m_2 - 2 = 4m_1 - 12 = 4(m_1 - 3) \end{aligned}$$

因此，我们猜想下式可能成立

$$m_i = i(m_1 - i + 1) \qquad i = 1,\cdots,n \qquad (5.5)$$

下面，我们利用数学归纳法证明上式。 因为已经得到上式在$i=1,\ 2$的时候是正确的，我们归纳假设当$i \le j < n$的时候等式也是成立的。下面只需要验证在$j+1$的情况下，结论是正确的。 利用式（5.4）可得

$$\begin{aligned} m_{j+1} &= 2m_j - m_{j-1} - 2 = 2j(m_1 - j + 1) - (j-1)(m_1 - j + 2) - 2 \quad \text{(由归纳假设)} \\ &= (j+1)m_1 - 2j^2 + 2j + j^2 - 3j + 2 - 2 = (j+1)m_1 - j^2 - j = (j+1)(m_1 - j) \end{aligned}$$

这就完成了式（5.5）的归纳证明。在式（5.5）中令$i=n$，并利用$m_n=0$，可得

$$m_1=n-1$$

再次利用式（5.5)，可以得到

$$m_i=i(n-i)$$

所以，在只有两名玩家的情况下，平均对局数就是最初他们各自持有的赌资单位i和$n-i$的乘积。 因为两名玩家参与了所有的阶段，所以这也是所有玩家 1 参与的对局数的平均值。

现在让我们回到包含r个玩家的问题，他们的初始赌资为n_i, $i=1,\cdots,r$, $\sum_{i=1}^{r}n_i=n.$ 令X表示获得一次胜利所需要的对局数，令X_i表示包含玩家i的对局数。对玩家i来说，初始拥有n_i单位赌资后一直对局，每局胜出的机会都是独立且均等的，直到他的财富是n或者 0。 所以他的对局数和当他只有一个初始财富为$n-n_i$的对手时的对局数是一样的。 于是，由前面的结论可知

$$\mathrm{E}\left[X_i\right]=n_i\left(n-n_i\right)$$

所以

$$\mathrm{E}\left[\sum_{i=1}^{r}X_i\right]=\sum_{i=1}^{r}n_i\left(n-n_i\right)=n^2-\sum_{i=1}^{r}\left(n_i\right)^2$$

但是因为每个对局包含两名玩家，所以有

$$X=\frac{1}{2}\sum_{i=1}^{r}X_i$$

上式两边取期望后得到

$$\mathrm{E}\left(X\right)=\frac{1}{2}\left(n^2-\sum_{i=1}^{r}{n_i}^2\right)$$

有趣的是，注意到：所得的平均对局数的值并不依赖于是怎么选择每次对局的组合，但这并不是说它不依赖于对局数的分配。 举例说，假设$r=3,n_1=n_2=1,n_3=2$ 如果玩家 1 和玩家 2 是被选第 1 组对局，那么就会需要至少 3 局才能得到胜利者，而如果玩家 3 出现在第 1 次对局的话，只用 2 局就可以了。

下面一个例子中，我们用取条件的方法重新证明在 6.3.1 节中提到的一个结果：平均来说，e个独立的（0,1）均匀随机变量之和大于 1。

例 5j 设$U_1,U_2\cdots$为一列相互独立的（0,1）均匀随机变量序列，令

$$N=\min\left\{n:\sum_{i=1}^{n}U_i>1\right\}$$

计算$\mathrm{E}\left[N\right]$

解： 我们通过求解一个更一般的结果，来得到$\mathrm{E}[N]$的值，对于$x \in [0,1]$，令

$$N(x) = \min\left\{n : \sum_{i=1}^{n} U_i > x\right\}$$

并令

$$m(x) = \mathrm{E}[N(x)]$$

即$N(x)$是部分和$\sum_{i=1}^{n} U_i$超过x的最小指标n，$m(x)$是$N(X)$的期望值。将U_1作为条件，利用公式（5.1b），得到

$$m(x) = \int_0^1 \mathrm{E}[N(x) | U_1 = y] \mathrm{d}y \qquad (5.6)$$

对于条件期望$\mathrm{E}[N(x) | U_1 = y]$，我们有

$$\mathrm{E}[N(x) | U_1 = y] = \begin{cases} 1 & y > x \\ 1 + m(x-y) & y \le x \end{cases} \qquad (5.7)$$

上式中，当$y > x$时等式是很显然的。当$y \le x$时，此时需要继续取$U_2, \cdots,$这相当于序列从U_2开始，要求出刚好超过$x - y$的最小时刻。将式（5.7）代入式（5.6），得到

$$m(x) = 1 + \int_0^x m(x-y) \mathrm{d}y = 1 + \int_0^x m(u) \mathrm{d}u \qquad (\text{作变量代换} u = x - y)$$

上式求微商得到

$$m'(x) = m(x)$$

或等价地，

$$\frac{m'(x)}{m(x)} = 1$$

再对上式求积分，得

$$\ln[m(x)] = x + c$$

或

$$m(x) = k\mathrm{e}^x$$

由$m(0) = 1$，得$k = 1$，这样

$$m(x) = \mathrm{e}^x$$

因此，要满足使得（0,1）区间上的均匀随机变量的部分和大于 1，平均最少需要的个数$m(1)$等于 e。

7.5.3 通过取条件计算概率

取条件期望的方法不仅可以用于计算一个随机变量的期望，还可以用于计算概率。设 E 为一随机事件，令 X 为 E 的示性变量，即

$$X=\begin{cases}1 & 若E发生\\ 0 & 若E不发生\end{cases}$$

由 X 的定义可得，

$$\mathrm{E}(x)=P(E)$$

$$\mathrm{E}\left[X|Y=y\right]=P\left(E|Y=y\right) \qquad 对任意随机变量Y$$

因此利用 $(5.1a)$ 与 $(5.1b)$，可得

$$P(E)=\begin{cases}\sum_y P\left(E|Y=y\right)P\left(Y=y\right) & Y为离散型随机变量\\ \int_{-\infty}^{+\infty}P\left(E|Y=y\right)f_Y\left(y\right)\mathrm{d}y & Y为连续型随机变量\end{cases} \tag{5.8}$$

注意，如果 Y 是离散型随机变量，且取值为 $y_1,\cdots y_n$，定义事件 $F_i=\{Y=y_i\}$，则式（5.8）变成

$$P(E)=\sum_{i=1}^{n}P\left(E|F_i\right)P\left(F_i\right)$$

其中 $F_1,\cdots,F_n$ 为互不相容的事件，且这些事件的并集构成一个样本空间。

例 5k （最优奖问题） 设有 n 个不同的奖陆续出台，当一个奖出台时，你可以拒绝或接受。当然，你接受了这个刚出台的奖，你不能再领以后出台的奖。若你拒绝刚出台的奖，那么你还有机会领以后出台的奖。当一个奖出台时，唯一的信息是刚出台的奖与已经出台的奖进行比较。 例如，当第 5 个奖出台时，你只能与前 4 个已经公布的奖进行比较。我们的目标是希望得到最高奖，或找到一种策略使得得到最高奖的概率尽可能大。 假设出台的奖项的 $n!$ 种次序都是等可能的。

解： 令人惊讶的是，我们可以得到很好的结果。对于固定的 $k, 0\le k<n$，考虑 如下的策略：首先拒绝前面 k 个奖项，然后从第 $k+1$ 个奖项出台开始算起，只要发现新出台的奖项比前面已经发布的好就接受这个奖项，否则就拒绝这个奖项而观察出台的下一个奖项，记 $P_k\left(\{最优\}\right)$ 表示利用这个策略得到最优奖项的概率，记 X 为最优奖项出台的顺序号，在给定 X 的条件下，有

$$P_k\left(最优\right)=\sum_{i=1}^{n}P_k\left(最优|X=i\right)P\left(X=i\right)=\frac{1}{n}\sum_{i=1}^{n}P_k\left(最优|X=i\right)$$

一方面，若最优的奖项在前面的 k 次发布，按这个选奖的策略，每次都拒绝拿奖，因此，不可能拿到最优奖。 这样

$$P_k\left(\text{最优}\middle|X=i\right)=0 \quad i\le k$$

另一方面，若最优奖的位置i在k之后，即$i>k$，那么就有可能拿到最优奖。 如果前面$i-1$个奖项的最大值在前面k个奖中（那么，第$k+1,\cdots k+2,\cdots i-1$次出台的奖项都被拒绝，直到最优奖$i$发布时，按规则接受最优奖）。 现在假定最优奖位置在i，在前面$i-1$个奖项中，最高奖的位置在$1,\cdots,i-1$处是等可能的。因此

$$P_k\left(\text{最优}\middle|X=i\right)=P\left\{\text{前面}i-1\text{个奖中最优奖在}\{1,2,3\cdots,\ k\}\text{中}\middle|X=i\right\}$$

$$=\frac{k}{i-1},i>k$$

这样我们得到

$$P_k\left(\text{最优}\right)=\frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1}\approx\frac{k}{n}\int_{k+1}^{n}\frac{1}{x-1}\mathrm{d}x=\frac{k}{n}\ln\frac{(n-1)}{k}\approx\frac{k}{n}\ln\left(\frac{n}{k}\right)$$

若考虑函数

$$g(x)=\frac{x}{n}\ln\left(\frac{n}{x}\right)$$

那么

$$g'(x)=\frac{1}{n}\ln\left(\frac{n}{x}\right)-\frac{1}{n}$$

所以

$$g^{'}(x)=0\Rightarrow\ln\left(\frac{n}{x}\right)=1\Rightarrow x=n/\mathrm{e}$$

由于$P_k\left(\text{最优}\right)\approx g(k)$, 当取$k=n/\mathrm{e}$时，$P_k(\text{最优})\approx g(n/\mathrm{e})=1/\mathrm{e}$, 最优策略是首先拒绝前面$k=n/\mathrm{e}$个奖项，然后等待出现第一个比以前的奖项都大的奖项，并接受这个奖项。 按这个策略，拿到最优奖的概率近似地等于$1/\mathrm{e}\approx0.36788$.

注释 大部分人对于以这么大的概率拿到最优奖感到吃惊， 一般认为这个概率当n很大时会趋于 0。 然而，即使不经精确计算，稍微思考,就会发现拿到最优奖的概率会相当大。取$k=n/2$，考虑这n个奖中的最优奖与第二最优奖. 考虑一个随机事件: 第二最优奖出现在前面一半, 第一最高奖出现在后面一半。 这个事件的概率为 1/4。 当这个事件发生时，我们一定能选到奖，并且是最优奖。因此看出，n无论怎么大，总是能找到一种策略，使得得到最高奖的可能性超过 1/4。

例 5l 设U为（0,1）上均匀分布的随机变量，又设在给定$U=p$的条件下，随机变量X服从参数为$(n,\ p)$的二项分布。计算X的分布列。

解： 在给定U的值的条件下，有

$$P\{X=i\}=\int_0^1 P\{X=i|U=p\}f_U(p)\mathrm{d}p=\int_0^1 P\{X=i|U=p\}\mathrm{d}p=\frac{n!}{i!(n-i)!}\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p,$$

又因为

$$\int_0^1 p^i(1-p)^{n-i}\mathrm{d}p=\frac{i!(n-i)!}{(n+1)!}$$

（这个公式的概率证明可参见 6.6 节

$$F_{X(j)}(y)=\sum_{k=j}^{n}C_n^k y^k(1-y)^{n-k}=\frac{n!}{(n-j)!(j-1)!}\int_0^y x^{j-1}(1-x)^{n-j}\mathrm{d}x,\quad 0\le y\le 1）,$$

由此可得

$$P\{X=i\}=\frac{1}{n+1}\qquad i=0,\cdots,n$$

由这个公式，我们可以得到一个令人吃惊的事实。如果将一个硬币连续掷n次，假定硬币的正面朝上的概率 p 服从（0,1）上的均匀分布，则正面朝上的次数为$0, 1, \cdots, n$上的可能性是相同的。

因为条件分布具有很好的形式，所以有必要给出另一种论证来加强我们的直观认识，设$U, U_1, \cdots, U_n$ 为$n+1$个独立同分布的（0,1）均匀随机变量。令 X 为$U_1, \cdots, U_n$中小于 U 的变量个数，由于$U_1, \cdots, U_n$和U具有相同分布，在$n+1$个变量的排序过程中U为最小，第 2 小，…，或最大，这$n+1$种可能性是相同的。 又因为在给定$U=p$的条件下，$U_i\le U(i=1, \cdots, n)$的个数的分布为二项分布，其参数为(n, p)，因此，X 的分布具有很直观的解释。

例 5m 设 X 和Y 为两个相互独立的随机变量，其密度分别为 f_X 和 f_Y 。计算$P\{X<Y\}$

解：对 y 的值取条件可得

$$\begin{aligned}P\{X<Y\}&=\int_{-\infty}^{\infty}P\{X<Y|Y=y\}f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}P\{X<y|Y=y\}f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{\infty}P\{X<y\}f_Y(y)\mathrm{d}y\qquad \text{由独立性}\\&=\int_{-\infty}^{\infty}F_X(y)f_Y(y)\mathrm{d}y\end{aligned}$$

其中

$$F_X(y)=\int_{-\infty}^{y}f_X(x)\mathrm{d}x$$

例 5n 设 X 和Y 为相互独立的连续随机变量，求 $X+Y$ 的分布。

解： 对 y 的值取条件可得

$$\begin{aligned}P\{X+Y<a\}&=\int_{-\infty}^{\infty}P\{X+Y<a|Y=y\}f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}P\{X+y<a|Y=y\}f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{\infty}P\{X<a-y\}f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}F_X(a-y)f_Y(y)\mathrm{d}y\end{aligned}$$

P176 第 5 章 连续型随机变量

例 6a 设随机变量 X 服从参数为 α 和 λ 的 Γ 分布，试计算（a） $\mathrm{E}[X]$；（b）$\mathrm{Var}(X)$。

解： （a）

$$\mathrm{E}[X]=\frac{1}{\Gamma(\alpha)}\int_0^{\infty}\lambda x\mathrm{e}^{-\lambda x}(\lambda x)^{\alpha-1}\mathrm{d}x$$

$$=\frac{1}{\lambda\Gamma(\alpha)}\int_0^{\infty}\lambda\mathrm{e}^{-\lambda x}(\lambda x)^{\alpha}\mathrm{d}x=\frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}=\frac{\alpha}{\lambda}\qquad \text{利用公式（6.1）}$$

（b）首先计算 $\mathrm{E}[X^2]$，再由方差计算公式可得

$$\mathrm{Var}(X)=\frac{\alpha}{\lambda^2}$$

例 5h （几何分布的方差 ）设有一独立重复试验序列，每次试验成功的概率为 p。记 N 为取得第一次成功所需的试验次数。求 $\mathrm{Var}(N)$。

解： 若第一次试验成功，令 $Y=1$；否则，$Y=0$，利用公式

$$\mathrm{Var}(N)=\mathrm{E}[N^2]-(\mathrm{E}[N])^2$$

我们只需计算 $\mathrm{E}[N^2]$。 在给定 Y 的条件下，有

$$\mathrm{E}[N^2]=\mathrm{E}[\mathrm{E}[N^2\mid Y]]$$

然而

$$\mathrm{E}[N^2\mid Y=1]=1\qquad \mathrm{E}[N^2\mid Y=0]=\mathrm{E}[(1+N)^2]$$

上述两式成立是因为：一方面，如果第一次实验成功，则有 $N=1$,从而N^2=1; 另一方面，若 $Y=0$，即第一次试验失败，则试验相当于重新开始，因此第一次成功所需实验次数变成 N+1。因为后者与 N 同分布，得到 $\mathrm{E}[N^2\mid Y=0]=\mathrm{E}[(1+N)^2]$。因此有

$$\mathrm{E}[N^2]=\mathrm{E}[N^2\mid Y=1]P\{Y=1\}+\mathrm{E}[N^2\mid Y=0]P\{Y=0\}$$
$$=p+(1-p)\mathrm{E}[(1+N)^2]=1+(1-p)\mathrm{E}[2N+N^2]$$

然而，在第 4 章例 8b 中，已经证明 $\mathrm{E}[N]=1/p$，因此有

$$\mathrm{E}[N^2]=1+\frac{2(1-p)}{p}+(1-p)\mathrm{E}[N^2]$$

由上式解得

$$\mathrm{E}[N^2]=\frac{2-p}{p^2}$$

从而

$$\mathrm{Var}(N)=\mathrm{E}[N^2]-(\mathrm{E}[N])^2=\frac{2-p}{p^2}-(\frac{1}{p})^2=\frac{1-p}{p^2}$$

7.5.4 条件方差

正如我们定义 $Y=y$ 之下 X 的条件期望一样，也可以定义 $Y=y$ 之下 X 的条件方差为

$$\mathrm{Var}(X|Y)\equiv \mathsf{E}[(X-\mathsf{E}[X|Y])^2|Y]$$

即 $\mathrm{Var}(X|Y)$ 是 X 和它的条件期望之差的平方的（条件）期望值。 换句话说，$\mathrm{Var}(X|Y)$ 通常的方差的定义完全一样，不过求期望换成了求在 Y 已知的条件下的条件期望。

条件方差 $\mathrm{Var}(X|Y)$ 和无条件方差 $\mathrm{Var}(X)$ 之间具有某种很有用的关系，人们通常利用这种关系计算一个随机变量的方差。首先，与普通方差的公式 $\mathrm{Var}(X)=\mathsf{E}[X^2]-(\mathsf{E}[X])^2$ 一样，条件方差也有

$$\mathrm{Var}(X|Y)\equiv \mathsf{E}[X^2|Y]-(\mathsf{E}[X|Y])^2$$

由此得到

$$\mathsf{E}[\mathrm{Var}(X|Y)]=\mathsf{E}[\mathsf{E}[X^2|Y]]-\mathsf{E}[(\mathsf{E}[X|Y])^2]=\mathsf{E}[X^2]-\mathsf{E}[(\mathsf{E}[X|Y])^2] \qquad (5.9)$$

同时

$$\mathrm{Var}(E[X|Y])=\mathsf{E}[(\mathsf{E}[X|Y])^2]-(\mathsf{E}[\mathsf{E}[X|Y]])^2=\mathsf{E}[(\mathsf{E}[X|Y])^2]-(\mathsf{E}[X])^2 \quad (5.10)$$

将（5.9）与式（5.10）相加，我们得到如下命题.

命题 5.2 [条件方差公式]

$$\mathrm{Var}(X)=\mathsf{E}[\mathrm{Var}(X|Y)]+\mathrm{Var}(\mathsf{E}[X|Y])$$

例 5o 设对任意时间 t，在 $(0,t)$ 内到达某火车站的人数是一个泊松随机变量，均值为 λt。现设火车在 $(0,T)$ 这个区间内随机到达，即到达时间是 $(0,T)$ 上的均匀分布，并且与旅客到达火车站的时间独立。 求火车到达时，上火车的旅客人数的期望和方差。

解： 对任意 $t\ge 0$，令 $N(t)$ 表示 t 以前到达车站的人数，Y 表示火车到达时间，$N(Y)$ 表示上火车的人数。给定 Y 的条件下有

$$\begin{aligned}\mathsf{E}[N(Y)|Y=t]&=\mathsf{E}[N(t)|Y=t]=\mathsf{E}[N(t)] && \text{由}Y\text{与}N(t)\text{的独立性}\\ &=\lambda t && N(t)\text{是均值为}\lambda t\text{ 的泊松随机变量}\end{aligned}$$

因此，

$$\mathsf{E}[N(Y)|Y]=\lambda Y$$

两边取期望可得

$$\mathsf{E}[N(Y)]=\lambda \mathsf{E}[Y]=\frac{\lambda T}{2}$$

为了计算 $\mathrm{Var}(N(Y))$，我们利用条件方差公式

$$\mathrm{Var}(N(Y)|Y=t)=\mathrm{Var}(N(t)|Y=t)=\mathrm{Var}(N(t))=\lambda t$$

因此，有

$$\mathrm{Var}(N(Y)|Y)=\lambda Y, \qquad \mathsf{E}[N(Y)|Y]=\lambda Y$$

再由条件方差公式，得

$$\mathrm{Var}(N(Y))=\mathsf{E}[\lambda Y]+Var(\lambda Y)=\lambda\frac{T}{2}+\lambda^2\frac{T^2}{12}$$

上式利用了 $\mathrm{Var}(Y)=T^2/12$ 的事实。

例 5p （随机个数随机变量之和的方差） 设 X_1，X_2，$\cdots$是一系列独立同分布的随机变量，N 是一取非负整数的随机变量，并且独立于序列 X_i，$i\ge 1$。为计算 $\mathrm{Var}(\sum_{i=1}^{N}X_i)$， 先固定 N 的值作为条件

$$\mathsf{E}[\sum_{i=1}^{N}X_i|N]=N\mathsf{E}[X], \qquad \mathrm{Var}(\sum_{i=1}^{N}X_i|N)=N\,\mathrm{Var}(X)$$

由前面已经得到的结果可知，对于给定的 N，$\sum_{i=1}^{N}X_i$ 是固定个数的独立随机变量的和，故它的期望和方差刚好是相应的期望和方差之和，再利用条件方差公式可得

$$\mathrm{Var}(\sum_{i=1}^{N}X_i)=\mathsf{E}[N]\mathrm{Var}(X)+(\mathsf{E}[X])^2\,\mathrm{Var}(N)$$

7.6 条件期望及预测

在实际问题中，有时会遇到这种情况，即某人观察到随机变量 X 的值，然后基于 X 的观察值，要对第二个随机变量 Y 进行预测，令g(x)表示预测值，即当观测到 X 的值 x 以后，g(X)就是 Y 的值的预测值。显然，我们希望选择 g 使g(X)最接近 Y，选择 g 的一个准则是极小化 $\mathrm{E}[(Y-g(X))^2]$。下面我们指出在这个准则之下，y 的最优的预测值为 $g(X)=\mathsf{E}[Y|X]$。

命题 6.1

$$\mathsf{E}[(Y-g(X))^2]\ge \mathsf{E}[(Y-\mathsf{E}[Y|X])^2]$$

证明：

$$\begin{aligned}\mathsf{E}[(Y-g(X))^2|X]&=\mathsf{E}[(Y-\mathsf{E}[Y|X]+\mathsf{E}[Y|X]-g(X))^2|X]\\&=\mathsf{E}[(Y-\mathsf{E}[Y|X])^2|X]+\mathsf{E}[(\mathsf{E}[Y|X]-g(X))^2|X]\\&\quad+2\mathsf{E}[(Y-\mathsf{E}[Y|X])(\mathsf{E}[Y|X]-g(X))|X]\end{aligned}$$

然而，对于给定的 X 值，$\mathsf{E}[Y|X]-g(X)$ 就是一个常数，于是

$$\begin{aligned}&\mathsf{E}[(Y-\mathsf{E}[Y|X])(\mathsf{E}[Y|X]-g(X))|X]=(\mathsf{E}[Y|X]-g(X))\mathsf{E}[Y-\mathsf{E}[Y|X]|X]\\&=(\mathsf{E}[Y|X]-g(X))(\mathsf{E}[Y|X]-\mathsf{E}[Y|X])=0\end{aligned}$$

这样，由（6.1）式和（6.2）式可得

$$\mathrm{E}[(Y-g(X))^2 \mid X] \geq \mathrm{E}[(Y-E[Y \mid X])^2 \mid X]$$

上式两边再求期望即可得到命题的结论。

注释 此处可以给出命题 6.1 的一个更加直观的证明，当然，在证明的严格性上要差一点。很容易证明 $\mathrm{E}[(Y-c)^2]$ 在 $c=\mathrm{E}[Y]$ 达到极小值（见理论习题 7.1）。因此在我们没有任何数据可用时，在均方误差最小的意义下，Y 的最优预测就是 $\mathrm{E}[Y]$。现在设得到了 X 的观察值 x,此时预测问题与没有数据时的预测问题完全一样，只是原来 Y 的期望改为事件 $\{X=x\}$ 之下的条件期望。因此，y 的最优预测是 Y 在 $X=x$ 之下的条件期望，于是命题 6.1 得证。

例 6a 设父亲的身高为 x 英寸，儿子的身高服从均值为 x+1，方差 4 的正态分布。假设父亲的身高为 6 英尺，那么其儿子成年以后的身高的最优预测值是多少。

解： 设父亲身高为 X，儿子身高为 Y，两者关系可表示为

$$Y = X + 1 + e$$

其中 e 为正态随机变量，独立于 X，并且期望为 0，方差为 4。 对于 6 英尺高的父亲，其儿子身高的最优预测为 $\mathrm{E}[Y \mid X=72]$，

$$\mathrm{E}[Y \mid X=72] = \mathrm{E}[X+1+e \mid X=72] = 73 + \mathrm{E}[e \mid X=72] = 73 + \mathrm{E}(e) = 73$$

例 6b 假设在 A 处发射一个强度为 s 的信号，在 B 处会接收到一个强度为 R 的信号，R 是一个正态随机变量，参数为（s，1）。现在假设发射端发射的信号强度 S 服从正态分布，参数为（μ，σ^2）。当接收端收到的 R 的值为 r 时，求发送信号强度的最优估计？

解： 首先计算发射端发送信号强度 S 在给定 R 之下的条件密度

$$f_{S|R}(s \mid r) = \frac{f_{S,R}(s,r)}{f_R(r)} = \frac{f_S(s) f_{R|S}(r \mid s)}{f_R(r)} = K\mathrm{e}^{-(s-\mu)^2/(2\sigma^2)}\mathrm{e}^{-(r-s)^2/2}$$

其中 K 不依赖于 s。注意

$$\frac{(s-\mu)^2}{2\sigma^2} + \frac{(r-s)^2}{2} = s^2\left(\frac{1}{2\sigma^2}+\frac{1}{2}\right) - \left(\frac{\mu}{\sigma^2}+r\right)s + C_1 = \frac{1+\sigma^2}{2\sigma^2}\left[s^2 - 2\left(\frac{\mu+r\sigma^2}{1+\sigma^2}\right)s\right] + C_1$$
$$= \frac{1+\sigma^2}{2\sigma^2}\left(s - \frac{\mu+r\sigma^2}{1+\sigma^2}\right)^2 + C_2$$

其中C_1, C_2均不依赖于 s，因此条件密度为

$$f_{S|R}(s \mid r) = C\exp\left\{\frac{-\left(s-\dfrac{\mu+r\sigma^2}{1+\sigma^2}\right)^2}{2\left(\dfrac{\sigma^2}{1+\sigma^2}\right)}\right\}$$

其中 C 与 s 无关，由上式可知，在给定 R=r 之下，S 的条件分布为正态分布，其期望和方差分别为

$$\mathrm{E}[S|R=r]=\frac{\mu+r\sigma^2}{1+\sigma^2},\quad \mathrm{Var}(S|R=r)=\frac{\sigma^2}{1+\sigma^2}$$

再利用命题 6.1，在给定 $R=r$ 之下，在均方误差最小的意义下，S 的最优估计为

$$\mathrm{E}[S|R=r]=\frac{1}{1+\sigma^2}\mu+\frac{\sigma^2}{1+\sigma^2}r$$

由上式看出，条件期望提供了关于 S 的信息，它是μ （信号的先验期望值）和r（接受到信号的期望值）的加权平均。而两个权值之比为 1 比σ^2，其中 1 代表信号 s 发出后接收到的信号的条件方差，σ^2表示发送信号的方差。

例 5f 在第 6 章例 5d 中，定义二维正态随机向量$(X,\ Y)$的联合密度为

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_x}{\sigma_x})^2+(\frac{y-\mu_y}{\sigma_y})^2-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}]\right\}$$

下面我们证明 ρ 实际上是 X 与 Y 的相关系数。 同例 5d，设 $\mu_x=\mathrm{E}[X],\ \sigma_x^2=\mathrm{Var}(X)$，且 $\mu_y=\mathrm{E}[Y],\ \sigma_y^2=\mathrm{Var}(Y)$。 则有

$$\mathrm{Corr}(X,\ Y)=\frac{\mathrm{Cov}(X,\ Y)}{\sigma_x\sigma_y}=\frac{\mathrm{E}[XY]-\mu_x\mu_y}{\sigma_x\sigma_y}$$

我们需要计算$\mathrm{E}[XY]$，为此，我们对Y取条件，即利用恒等式

$$\mathrm{E}[XY]=\mathrm{E}[\mathrm{E}[XY|Y]]$$

和第 6 章例 5d 中给定$Y=y$的条件下，X 的条件分布是均值为$\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$的正态分布的结论，可得

$$\mathrm{E}[XY|Y=y]=\mathrm{E}[Xy|Y=y]=y\mathrm{E}[X|Y=y]=y[\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)]$$

$$=y\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y^2-\mu_y y)$$

从而

$$\mathrm{E}[XY|Y]=Y\mu_x+\rho\frac{\sigma_x}{\sigma_y}(Y^2-\mu_y Y)$$

上式表明

$$\mathrm{E}[XY]=\mathrm{E}[Y\mu_x+\rho\frac{\sigma_x}{\sigma_y}(Y^2-\mu_y Y)]=\mu_x\mathrm{E}[Y]+\rho\frac{\sigma_x}{\sigma_y}\mathrm{E}[Y^2-\mu_y Y]$$

$$=\mu_x\mu_y+\rho\frac{\sigma_x}{\sigma_y}(E[Y^2]-\mu_y^2)=\mu_x\mu_y+\rho\frac{\sigma_x}{\sigma_y}\mathrm{Var}(Y)=\mu_x\mu_y+\rho\sigma_x\sigma_y$$

由此可得

$$\mathrm{Corr}(X,\ Y)=\frac{\rho\sigma_x\sigma_y}{\sigma_x\sigma_y}=\rho$$

有时候，利用条件期望恒等式（即（5.1a）或（5.1b））来计算 $E[X]$.

例 6c 在数字信号处理过程中必须把连续数据离散化。其过程如下：取一组递增数列，a_i，i = 0,±1,±2,.. .，使得 $\lim_{i\to+\infty} a_i=\infty$, $\lim_{i\to-\infty} a_i=-\infty$ 当 $X\in(a_i,a_{i+1}]$ 时，选一个代表值 y_i，用 Y 表示离散化后的值，Y 与 X 之间有如下关系：

$$Y=y_i \quad a_i<X\le a_{i+1}$$

Y 的分布由下式给出

$$P\{Y=y_i\}=F_X(a_{i+1})-F_X(a_i)$$

现在我们的目标是要选择各区间的代表值 $y_i, i=0,\pm1,\pm2,\cdots$ ，使得 $E[(X-Y)^2]$ 达到极小。

（a） 找到最优值 $y_i, i=0,\pm1,\pm2,\cdots$

对于最优的 Y，证明：

（b） $E[Y]=E[X]$，即均方误差最小意义下的离散化保持均值不变。

（c） $\mathrm{Var}(Y)=\mathrm{Var}(X)-E[(X-Y)^2]$

解： （a）对于任意的离散化随机变量 Y，在给定 Y 值的条件下，有

$$E[(X-Y)^2]=\sum_i E[(X-y_i)^2 \mid a_i<X\le a_{i+1}]P\{a_i<X\le a_{i+1}\}$$

如果我们令

$$a_i<X\le a_{i+1} \text{ 为 } I=i$$

则有

$$E[(X-y_i)^2 \mid a_i<X\le a_{i+1}]=E[(X-y_i)^2 \mid I=i]$$

利用 4.5 节命题的结论，当

$$y_i=E[X \mid I=i]=E[X \mid a_i<X\le a_{i+1}]=\int_{a_i}^{a_{i+1}}\frac{xf_X(x)\mathrm{d}x}{F_X(a_{i+1})-F_X(a_i)}$$

时 $E[(X-y_i)^2 \mid a_i<X\le a_{i+1}]$ 达到极小值。因此 $Y=E[X\mid I]$ 是最优的离散化随机变量。在最优的选择之下有下述结论成立：

(b) $E[Y]=E[E[X\mid I]]=E[X]$

(c) $\mathrm{Var}(X)=E[\mathrm{Var}(X \mid I)]+\mathrm{Var}(E[X \mid I])=E[E[(X-Y)^2 \mid I]]+\mathrm{Var}(Y)$

$$=E[(X-Y)^2]+\mathrm{Var}(Y)$$

在某些情况下，X 和 Y 的联合分布不是完全已知的，或者，即使知道联合分布，$E[Y\mid X=x]$ 的计算也十分复杂。然而，如果我们知道 X 和 Y 的期望、方差

和相关系数，那么我们至少可以求出依赖于 X 的最优线性预测。

为求得 Y 的最优线性预测，我们需要选择线性预测 $a+bX$ 的系数 a 和 b，使得 $\mathrm{E}[(Y-(a+bX))^2]$ 达到极小值。为此，先将 $\mathrm{E}[(Y-(a+bX))^2]$ 展成一个 a, b 的多项式：

$$\begin{aligned}\mathsf{E}[(Y-(a+bX))^2] &= \mathsf{E}[Y^2-2aY-2bXY+a^2+2abX+b^2X^2] \\ &= \mathsf{E}[Y^2]-2a\mathsf{E}[Y]-2b\mathsf{E}[XY]+a^2+2ab\mathsf{E}[X]+b^2\mathsf{E}[X^2]\end{aligned}$$

将上式对 a 和 b 求偏导数，得到

$$\frac{\partial}{\partial a}\mathsf{E}[(Y-a-bX)^2] = -2\mathsf{E}[Y]+2a+2b\mathsf{E}[X]$$
$$\frac{\partial}{\partial b}\mathsf{E}[(Y-a-bX)^2] = -2\mathsf{E}[XY]+2a\mathsf{E}[X]+2b\mathsf{E}[X^2]$$

令偏导数为 0，求解关于（a,b）的方程组（6.3）,得到

$$b = \frac{\mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y]}{\mathsf{E}[X^2]-(\mathsf{E}[X])^2} = \frac{Cov(X,Y)}{\sigma_x^2} = \rho\frac{\sigma_y}{\sigma_x}$$
$$a = \mathsf{E}[Y]-b\mathsf{E}[X] = \mathsf{E}[Y]-\frac{\rho\sigma_y\mathsf{E}[X]}{\sigma_x}$$

其中ρ为 X，Y 的相关系数， $\sigma_x^2 = \mathsf{Var}(X)$ ， $\sigma_y^2 = \mathsf{Var}(Y)$ 。容易验证由（6.4）给出的 a, b 值使得 $\mathsf{E}[(Y-(a+bX))^2]$ 达到极小。 因此，在均方误差意义下，Y 关于 X 的最优线性预测为

$$\mu_y+\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x)$$

其中μ_y=$\mathsf{E}[Y]$, μ_x=$\mathsf{E}[X]$.

这个线性预测的均方误差为

$$\begin{aligned}&\mathsf{E}[(Y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x))^2] \\ &= \mathsf{E}[(Y-\mu_y)^2]+\rho^2\frac{\sigma_y^2}{\sigma_x^2}\mathsf{E}[(X-\mu_x)^2]-2\rho\frac{\sigma_y}{\sigma_x}\mathsf{E}[(Y-\mu_y)(X-\mu_x)] \\ &= \sigma_y^2+\rho^2\sigma_y^2-2\rho^2\sigma_y^2 = \sigma_y^2(1-\rho^2)\end{aligned}$$

由（6.5）可以看出，当 p 接近于+1 或-1 时，其最优线性预测的均方误差接近于 0。

例 6d 当 X，Y 的联合分布为二元正态分布时，因为在给定 X 的条件下 Y 的条件期望为 X 的线性函数，因此 Y 关于 X 的最优线性预测就是最优预测。在第 6 章例 5c 已经给出，在正态情况下，

$$\mathrm{E}[\mathrm{Y} \mid X=x] = \mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

第 6 章

马尔可夫不等式 设 X 为取非负值的随机变量，则对于任何常数 $a>0$，有

$$P\{X \ge a\} \le \mathrm{E}(X)/a$$

证明：对于 $a>0$，令 $I=\begin{cases}1, 若X \ge a \\ 0, 其他\end{cases}$。并且注意到，由于 $X \ge 0$，有

$I \le X/a$，对不等式两边求期望，得 $\mathrm{E}(I) \le \mathrm{E}(X)/a$。因为 $\mathrm{E}(I)=P\{X \ge a\}$，所以不等式成立。

命题 5.1［单边的切比雪夫不等式］

设 X 具有 0 均值和有限方差 σ^2，则对任意 a>0，

$$P\{X \ge a\} \le \frac{\sigma^2}{\sigma^2+a^2}$$

证明： 令 b>0，注意到

$$X \ge a \Leftrightarrow X+b \ge a+b$$

故

$$P\{X \ge a\}=P\{X+b \ge a+b\} \le P\{(X+b)^2 \ge (a+b)^2\}$$

上式中，由于 $a+b>0$，$X+b \ge a+b$ 可推知 $(X+b)^2 \ge (a+b)^2$ 故不等式成立。再利用马尔可夫不等式。可得

$$P\{X \ge a\} \le \frac{\mathrm{E}[(X+b)^2]}{(a+b)^2}=\frac{\sigma^2+b^2}{(a+b)^2}$$

上式中，b 可以取任意正常数，取 $b=\sigma^2/a$，便得到本命题的结论。实际上。当 $b=\sigma^2/a$ 时，$(\sigma^2+b^2)/(a+b)^2$ 达到极小值。

例 5a 设某工厂每周的产量是一个随机变量，其均值为 $\mu=100$，方差为 $\sigma^2=400$。计算一周产量至少为 120 的概率的上界。

解：利用单边切比雪夫不等式

$$P\{X \ge 120\}=P\{X-100 \ge 20\} \le \frac{400}{400+20^2}=1/2$$

这说明本周产量至少为 120 的概率不会超过 1/2。

如果直接利用马尔可夫不等式，可得

$$P\{X \ge 120\} \le \frac{\mathrm{E}[X]}{120}=5/6$$

这个上界就比较弱。（上界越小，结论越强，若上界为 1，这个结论就没有任何意义了。）

现在设 X 具有均值μ，方差σ^2，由于$X-\mu$与$\mu-X$都具有均值 0 和方差σ^2，利用单边切比雪夫不等式可知，对于$a>0$，

$$P\{X-\mu\ge a\}\le\frac{\sigma^2}{\sigma^2+a^2}$$
$$P\{\mu-X\ge a\}\le\frac{\sigma^2}{\sigma^2+a^2}$$

因此我们得到下面的推论。

推论 5.1 若 $\mathrm{E}[X]=\mu$， $\mathrm{Var}(X)=\sigma^2$，则对于$a>0$，下列不等式成立

$$P\{X\ge\mu+a\}\le\frac{\sigma^2}{\sigma^2+a^2}$$
$$P\{X\le\mu-a\}\le\frac{\sigma^2}{\sigma^2+a^2}$$

例 5b 一个有 100 个男人和 100 个女人组成的集合，被随机分成两两一组的 100 组。试给出最多 30 个组是由一男一女组成的概率的上界。

解：对所有男人任意地从 1 至 100 进行编号，对于 $i=1,2,\cdots,100$，令

$$X_i=\begin{cases}1 & \text{男人}i\text{所在的组内有女人}\\ 0 & \text{其他}\end{cases}$$

这样，男女组的数量 X 可以表示为

$$X=\sum_{i=1}^{100}X_i$$

由已知第 i 个男人和其他 199 个人配对的概率是相等的，而其中有 100 个是女人，我们有

$$\mathrm{E}[X_i]=P\{X_i=1\}=\frac{100}{199}$$

类似地，对于 $i\ne j$

$$\mathrm{E}[X_iX_j]=P\{X_i=1,X_j=1\}=P\{X_i=1\}P\{X_j=1|X_i=1\}=\frac{100}{199}\bullet\frac{99}{197}$$

其中 $P\{X_j=1|X_i=1\}=\dfrac{99}{197}$. 这是因为当第 i 个男人已经和一个女人配对时，男人 j 只可能跟剩余的 197 人配对，其中 99 人为女人。因此我们得到

$$\mathrm{E}[X]=\sum_{i=1}^{100}\mathrm{E}[X_i]=100\cdot\frac{100}{199}\approx 50.25$$

$$\begin{aligned}\mathrm{Var}(X)&=\sum_{i=1}^{100}\mathrm{Var}(X_i)+2\sum_{i<j}\sum\mathrm{Cov}(X_i,X_j)\\&=100\frac{100}{199}\frac{99}{199}+2\binom{100}{2}\left[\frac{100}{199}\frac{99}{197}-\left(\frac{100}{199}\right)^2\right]\approx 25.126\end{aligned}$$

由切比雪夫不等式可得

$$P\{X\le 30\}\le P\{|X-50.25|\ge 20.25\}\le\frac{25.126}{(20.25)^2}\approx 0.061$$

由此看出，最多 30 对一男一女的概率上界为 0.061。然而，我们可以利用单边切比雪夫不等式对该上界进行改进，得到

$$P\{X\le 30\}=P\{X\le 50.25-20.25\}\le\frac{25.126}{25.126+(20.25)^2}\approx 0.058$$

当随机变量 X 矩母函数已知时，我们可以得到更加有效的 $P\{X\ge a\}$ 的上界。令

$$M(t)=\mathrm{E}[\mathrm{e}^{tX}]$$

为随机变量 X 的矩母函数。则对于 $t>0$，有

$$P\{X\ge a\}=P\{\mathrm{e}^{tX}\ge\mathrm{e}^{ta}\}\le\mathrm{E}[\mathrm{e}^{tX}]\mathrm{e}^{-ta}\qquad\text{利用马尔科夫不等式}$$

类似得，对于 $t<0$，

$$P\{X\le a\}=P\{\mathrm{e}^{tX}\ge\mathrm{e}^{ta}\}\le E[\mathrm{e}^{tX}]\mathrm{e}^{-ta}$$

这样，我们得到下列结果，被称为*切尔诺夫界*。

命题 5.2 [*切尔诺夫界*]

$$\begin{aligned}P\{X\ge a\}&\le\mathrm{e}^{-ta}M(t)\qquad\text{对一切}t>0\\P\{X\le a\}&\le\mathrm{e}^{-ta}M(t)\qquad\text{对一切}t<0\end{aligned}$$

由于切尔诺夫界对 t 为正数或负数的情况都成立，我们通过找到使 $\mathrm{e}^{-ta}M(t)$ 达到最小的 t 值，来获得 $P\{X\ge a\}$ 的最佳上界。

第 7 章　模拟

10.1 引言

我们怎样确定在一场纸牌赌博中赢的概率？一种可能的方法是假设一副牌有 $(52)!$ 种可能的排列，各种排列是等可能的，然后观察其中多少种排列能使我们获胜。然而这种方法显然不现实，因为没有任何系统性的算法能够算出何种组合能够获胜，而且 $(52)!$ 是相当大的数。似乎唯一的办法确定在比赛结束之后真正得到胜利的组合，但这种方法对于我们显然没有任何用处。

看起来，确定一副纸牌胜出的概率是数学的难题。然而，并非没有一点希望，因为概率不仅属于数学领域，还属于应用科学领域。在所有应用科学中，试验是非常有价值的技术。对于单人纸牌游戏，试验就是玩很多次这样的纸牌游戏，或者可以编制一个计算机程序，让机器去玩牌。经过几次玩牌以后，比如 n 次，令

$$X_i=\begin{cases}1 & i\text{次玩牌胜出}\\ 0 & \text{其他}\end{cases}$$

此时 $X_i, i=1,2,\cdots,n$ 是独立的伯努利随机变量，且

$$\mathrm{E}[X_i]=P\{\text{第}i\text{次玩牌胜出}\}$$

由此，由强大数定律可得

$$\sum_{i=1}^{n}\frac{X_i}{n}=\frac{\text{赢的次数}}{\text{玩牌总次数}}$$

以概率为 1 地收敛到 $P\{\text{玩牌赢}\}$。也就是说，玩大量次数的纸牌游戏以后，赢牌的频率的就作为赢的概率。用试验的方法来确定概率值的方法称为模拟。

为了使计算机实现模拟，我们必须产生（0,1）上均匀分布的随机变量的值，这些值称为随机数。大部分计算机有一内置程序，称为随机数发生器，它产生一个伪随机数序列，就所有实用目的来说，这个伪随机数序列与来自（0,1）均匀分布的样本没有区别。通常随机数的生成是从一个初始值 X_0 开始的，这个初始值被称作种子，然后给定正整数 a,c,m，令

$$X_{n+1}=(aX_n+c)\bmod m \qquad n\geq 0$$

上述表明 X_{n+1} 的值是 aX_n+c 除以 m 的余数。这样每个 X_n 的取值范围都是 $0,1,\cdots,m-1$，且 X_n/m 近似地在（0,1）上均匀分布。 可以证明当取定合适的 a,c,m 值时，可以产生一个类似于独立的（0,1）上均匀随机变量序列的序列。

开始模拟时，我们假定能够模拟（0,1）上均匀分布的随机变量，并用随机数序列这一术语表示（0,1）上均匀布的随机变量的一组样本。

在纸牌游戏的例子中，首先需要编程产生一个给定的纸牌的排列顺序，然而初始顺序必须是从（52）!种顺序中等可能地取出。因此，必须产生一个随机的排列。下面的算法说明怎样利用随机数产生一个随机排列：先把 n 个元素放在 $0,1,\cdots,n$ 共 n 个位置中，然后利用一

个随机数选定一个随机的位置，将这个位置上的对象放在 n 这个位置上，再在剩下的 n-1 个对象中随机地选出一个对象放在 n-1 这个位置上，最后所有的对象都放在相应的位置上，一个随机排列就产生了。

例 la （产生一随机排列 ） 假设要产生整数 $1,2,\cdots,n$ 的一个排列，使得所有 $n!$ 种排列都是等可能的。从任何一个初始排列开始，将通过 n-1 步得到最终结果，在每一步交换排列中两个数的位置。在整个过程中，用 $X(\mathrm{i})$ 表示在位置 i 上的数。其算法如下：

1. 考虑一个初始排列，$X(\mathrm{i})$ 表示在位置 i 上的对象，$i=1,2,\cdots,n$ [例如，令 $X(i)=i, i=1,2,\cdots,n$]
2. 产生一个随机变量 N_n，N_n 在数集 $\{1,2,\cdots,n\}$ 上的均匀分布。
3. 将 $X(N_n)$ 与 $X(n)$ 交换位置，交换以后，$X(n)$ 就是原来的 $X(N_n)$，并且将这个对象固定在位置 n。[例如，n=4,初始状态 $X_i=i, i=1,2,3,4$,若 $N_4=3$,此时,新的排列成为 $X(1)=1, X(2)=2, X(3)=4, X(4)=3$ 。而 3 这个对象此后不改位置，永远放在位置 4 上。]
4. 产生随机变量 N_{n-1}，它在整数集 $\{1,2,\cdots,n-1\}$ 上均匀分布。
5. 交换 $X(N_{n-1})$ 与 $X(n-1)$ 的位置。 [若 $N_3=1$，新的排列变成 $X(1)=4, X(2)=2, X(3)=1, X(4)=3$ 。]
6. 产生 N_{n-2} ,它在 $\{1,2,\cdots,n-2\}$ 上均匀分布。
7. 交换 $X(N_{n-2})$ 和 $X(n-2)$ 。[若 $N_2=1$，此时新的排列成为 $X(1)=2, X(2)=4, X(3)=1, X(4)=3$。]
8. 产生 $N_{n-3},\ldots$，直到 N_2 产生，然后交换 $X(N_2)$ 与 X_2 的位置，得到最后的排列。

要实现这个算法需要产生在 $\{1,2,\cdots,k\}$ 上等可能地取值的随机变量，为此，令 U 是一个随机数，即 U 在（0，1）上均匀分布。注意到此时 kU 在区间（0，k）上均匀分布。则

$$P\{i-1<kU<i\}=\frac{1}{k} \quad i=1,\cdots,k$$

取 $N_k=[kU]+1$,其中记号 $[x]$ 表示 x 的整数部分（即不大于 x 的最大整数)，则 N_k 就会在{1,…,k}上均匀分布。

这个算法可以简明地写成下列几步：

第 1 步 令 $X(1),\cdots,X(n)$ 为 $1,2,\cdots,n$ 的任意排列 [例如，$X(\mathrm{i})=\mathrm{i}, \mathrm{i}=1,\cdots,\mathrm{n}$]

第 2 步 令 $I=n$。

第 3 步 产生一个随机数 U，令 $N=[IU]+1$。

第 4 步 交换 $X(N)$ 与 $X(I)$ 的位置。

第 5 步 将 I 的值减去 1，如果 $I>1$，则转向第 3 步。

第 6 步 $X(1),\cdots,X(n)$ 就是随机排列。

上述产生的随机排列的算法很有用。例如，假设一个统计学家想通过试验来比较对 n 个试验对象进行 m 种不同处理的效果。他把试验对象分成容量分别为 $n_1,n_2,\cdots,n_m$ 的 m 组，显然 $\sum_{i=1}^{m} n_i = n$。第 i 组试验对象接受第 i 种治疗。为了清除任何分组偏差（例如，若把最好的个体在同一组，就会将处理的效果和个体的“好坏”作用相混淆，造成偏差）我们必须将各个个体随机地分入各组。怎样做才能完成这个任务?

一个简单而有效的方法是随意将对象由 1 到 n 进行编号，并产生一个 $1,2,\cdots,n$ 的随机排列 $X(1),\cdots,X(n)$。将编号为 $X(1),\cdots,X(n_1)$ 的对象归入第一组，将编号 $X(n_1+1),\cdots,X(n_1+n_2)$ 编为第二组，一般地，将编号为 $X(n_1+\cdots+n_{j-1}+k),k=1,\cdots,n_j$ 的对象编成第 j 组，然后对于第 j 组施行第 j 种处理。

10．2 具有连续分布函数的随机变量的模拟技术

本节中，我们提供两种模拟随机变量的一般方法，而这些随机变量具有连续分布函数。

10.2.1 反变换方法

基于下列命题，我们可用反变换方法模拟具有连续分布函数的随机变量。

命题 2.1 设 U 为（0,1）上均匀随机变量，F 为任意一个连续分布函数，如果定义随机变量

$$Y=F^{-1}(U)$$

则 Y 具有分布函数 F。[$F^{-1}(x)$ 是方程 $F(y)=x$ 的解]

证明：

$$F_Y(a)=P\{Y\le a\}=P\{F^{-1}(U)\le a\} \qquad (2.1)$$

由于 $F(x)$ 是一个单调函数，$F^{-1}(U)\le a$ 成立的充要条件是 $U\le F(a)$。因此，由（2.1）式可得

$$F_Y(a)=P\{U\le F(a)\}=F(a)$$

由命题 2.1，我们可以通过产生一个随机数 U 并令 $X=F^{-1}(U)$ 来模拟具有连续分布函数 F 的随机变量 X。

例 2a（**模拟一个指数随机变量**） 设 $F(x)=1-e^{-x}$，则 $F^{-1}(U)$ 是下列方程的解 x：

$$1-e^{-x}=u$$

或

$$x = -\ln(1-u)$$

因此，若 U 为（0，1）均匀随机变量，则

$$F^{-1}(U) = -\ln(1-U)$$

的分布为指数分布，均值为 1。由于 $(1-U)$ 也是（0，1）均匀随机变量，所以 $-\ln U$ 也是指数随机变量，其均值为 1。若 x 具有指数分布，其均值为 1，则 cX 具有指数分布，其均值为 c。利用指数分布的这个特点知 $-c\ln U$ 具有指数分布，均值为 c。例 2a 的结果也可以用来模拟 Γ 随机变量。

例 2b（**模拟一个 $\Gamma(\mathrm{n},\lambda)$ 随机变量**）为了模拟参数为 (n,λ) 的 Γ 随机变量，其中 n 是整数，我们可以利用 Γ 随机变量与指数随机变量的关系，即 n 个独立同分布的参数为 λ 的指数随机变量的和具有此分布。因此，设 $U_1,\cdots,U_n$ 为独立同分布的（0,1）均匀随机变量，则

$$X = -\sum_{i=1}^{n}\frac{1}{\lambda}\ln U_i = -\frac{1}{\lambda}\ln(\prod_{i=1}^{n}U_i)$$

具有 $\Gamma(\mathrm{n},\lambda)$ 分布。

10.2.2 舍取法

假设我们有一种方法能够模拟密度函数为 $g(x)$ 的随机变量。我们可以首先模拟一个密度函数为 g 的随机变量 Y，然后以正比于 $f(Y)/g(Y)$ 的概率采用 Y 的值，这样以 Y 为基础就能模拟出一个密度为 $f(x)$ 的随机变量。
具体来说，令 c 为一常数，满足

$$\frac{f(\mathrm{y})}{g(\mathrm{y})} \le c \quad \text{对一切 } y \text{ 成立}$$

然后采用下列方法模拟具有密度 f 的随机变量。

舍取法

第 1 步 模拟具有密度 g 的 Y，同时产生—随机数 U。

第 2 步 若 $U \le f(Y)/[\mathrm{c}g(Y)]$，则 $X = Y$，否则回到第一步。

舍取法模拟流程见图 10.1。下面我们要证明舍取法的可行性。

图 10.1

命题 2.2 由上述舍取法产生的随机变量具有密度函数 f。

证明：设 X 为由舍取法产生的随机变量，记 N 为舍取法中循环的次数，则

$$P\{X \le x\} = P\{Y_N \le x\} = P\left\{Y \le x \middle| U \le \frac{f(Y)}{cg(Y)}\right\}$$

$$= \frac{P\left\{Y \le x, U \le \dfrac{f(Y)}{cg(Y)}\right\}}{K}$$

其中K=$P\{U \le f(Y)/(cg(Y))\}$。由于Y与U相互独立，所以Y与U的联合密度由下式给出：

$$f(y,u)=g(y) \qquad 0<u<1$$

利用上述的结论可得，

$$P\{X \le x\}=\frac{1}{K}\iint\limits_{\substack{y\le x\\ 0\le u\le f(y)/(cg(y))}} cg(y)\mathrm{d}u\mathrm{d}y$$

$$=\frac{1}{K}\int_{-\infty}^{x}\int_{0}^{f(y)/(cg(y))}\mathrm{d}u\, g(y)\mathrm{d}y=\frac{1}{cK}\int_{-\infty}^{x}f(y)\mathrm{d}y \qquad (2.2)$$

由于 $f(y)$ 为密度函数，上式两边令 $x\to+\infty$，得

$$1=\frac{1}{cK}\int_{-\infty}^{+\infty}f(y)\mathrm{d}y=\frac{1}{cK}$$

因此，由式（2.2）可得

$$P\{X \le x\}=\int_{-\infty}^{x}f(y)\mathrm{d}y$$

注释 (a) 注意，前面提到“以概率 $f(Y)/(cg(Y))$ 接受 Y”是指产生一个随机数 U，若$U \le f(Y)/(cg(Y))$，则令 X=Y。

(b) 在产生随机数的过程中，每次循环独立地接受 Y 的概率为 $P\{U \le f(Y)/(cg(Y))\}=K=1/c$。由此可知，循环次数 N 具有以c 为均值的几何分布。

例 2c (模拟正态随机变量）模拟一个标准正态随机变量 Z（即均值为 0，方差为 1 的正态分布)，首先注意 $X=|Z|$具有密度函数

$$f(x)=\frac{2}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \qquad 0<x<\infty \qquad (2.3)$$

我们首先通过舍取法，模拟 X 取密度函数 $g(x)$，它是均值为 1 的指数分布的密度函数，即

$$g(x)=\mathrm{e}^{-x} \qquad 0<x<\infty$$

注意到

$$\frac{f(x)}{g(x)}=\sqrt{\frac{2}{\pi}}\exp\left\{\frac{-(x^2-2x)}{2}\right\}=\sqrt{\frac{2}{\pi}}\exp\left\{\frac{-(x^2-2x+1)}{2}+\frac{1}{2}\right\}$$

$$=\sqrt{\frac{2\mathrm{e}}{\pi}}\exp\left\{\frac{-(x-1)^2}{2}\right\}\le\sqrt{\frac{2\mathrm{e}}{\pi}} \qquad (2.4)$$

因此，取 $c=\sqrt{2\mathrm{e}/\pi}$，由式（2.4）知

$$\frac{f(x)}{cg(x)}=\exp\left\{\frac{-(x-1)^2}{2}\right\}$$

所以利用取舍法，可以按照下列步骤模拟标准正态随机变量：

(a)产生独立随机变量 Y 和 U，其中 Y 具有均值为 1 的指数分布，U 在(0,1)上均匀分布。

(b)若$U\le\exp\{-(Y-1)^2/2\}$，则 $X=Y$，否则转向(a)。

当得到具有密度（2.3）X 之后，可令 $Z=+X$ 或 $-X$，以 1/2 的概率取正号，1/2 的概率取负号。在步骤（b）中，条件$U\le\exp\left\{-(Y-1)^2/2\right\}$等价于 $-\ln U\le(Y-1)^2/2$，而在例 2a 中，$-\ln U$ 是均值为 1 的指数随机变量， 因此，步骤（a)，(b)等价于：

(a')产生两个相互独立的均值为 1 的指数随机变量Y_1和Y_2。

(b')若 $Y_2\ge(Y_1-1)^2/2$，令 $X=Y_1$，否则转向(a')。

现在假设Y_1被接受, 此时我们知道 Y_2 比 $(Y_1-1)^2/2$ 大, 但是 Y_2 比 $(Y_1-1)^2/2$ 大多少？

由上文我们知道 Y_2 是均值为 1 的指数随机变量，利用指数分布的无记忆性可知，

$Y_2-(Y_1-1)^2/2$ 也是指数随机变量，其均值为 1。因此，若接受Y_1，则得到 X 的同时，我们也能得到另一个指数随机变量$Y_2-(Y_1-1)^2/2$，它与 X 相互独立，并且均值为 1。

概况来说，我们可以利用下列步骤产生一个指数随机变量（均值为 1)和与之独立的标准正态随机变量。

第 1 步 产生Y_1，它是均值为 1 的指数随机变量。

第 2 步 产生Y_2，它是均值为 1 的指数随机变量。

第 3 步 若$Y_2-(Y_1-1)^2/2>0$，令$Y=Y_2-(Y_1-1)^2/2$， 转向第 4 步，否则转向第一步。

第 4 步 产生一个随机数 U，令

$$Z=\begin{cases}Y_1 & U\le\dfrac{1}{2}\\ -Y_1 & U>\dfrac{1}{2}\end{cases}$$

上述算法产生的随机变量 Y 与 Z 独立的，且随机变量 Z 服从标准正态分布而随机变量 Y 服从参数为 1 的指分布， （如果想要得到均值为 μ，方差为σ^2 的正态随机变量，只需取 $\mu+\sigma Z$ 即可。）

注释 (a) 由于 $c=\sqrt{2\mathrm{e}/\pi}\approx1.32$，在前述的产生随机变量过程中，第 3 步中要求有 N 步的循环，其中 N 是均值$c\approx1.32$ 的几何随机变量。

(b) 如果我们希望产生一个标准正态的随机变量序列，则在第 3 步中生产的 Y 代入第 1 步产生下一个正态随机变量。那么，产生一个正态随机变量平均需要 $1.64(=2\times1.32-1)$ 个指数随机变量以及 1.32 次平方运算。

例 2d （**模拟正态随机变量：极坐标法**）在第 6 章例 7b 指出，若 X 和 Y 为相互独立的标准正态随机变量，则它们的极坐标 $R=\sqrt{X^2+Y^2}$ 和 $\Theta=\arctan(Y/X)$ 相互独立，R^2 是均值 2 的指数随机变量。Θ 在 $(0,2\pi)$ 上均匀分布，因此，设 U_1,U_2 是随机数，则利用例 2a 的方法有

$$R=(-2\ln U_1)^{1/2} \qquad \Theta=2\pi U_2$$

从而

$$\begin{aligned} X&=R\cos\Theta=(-2\ln U_1)^{1/2}\cos(2\pi U_2)\\ Y&=R\sins\Theta=(-2\ln U_1)^{1/2}\sin(2\pi U_2) \end{aligned} \qquad (2.5)$$

是独立的标准正态随机变量。

上述生成标准正态随机变量的方法称为 Box-Muller 方法。由于要计算正弦值与余弦值，算法的效率受到影响，有一种方法可以避免这种潜在的费时的困难。首先注意，若 U 是（0, 1）上的均匀随机变量，则 $2U$ 是（0,2）上的均匀随机变量, 所以 $2U$-1 是（-1，+1）上的均匀随机变量，因此，假设我们产生 U_1,U_2 两个随机数，令

$$V_1=2U_1-1 \quad V_2=2U_2-1$$

则 (V_1,V_2) 是在面积为 4，中心为（0，0）的一个方块上均匀分布的随机向量（见图 10-2）。

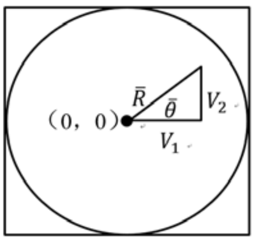

图 10-2

现在假设连续产生 (V_1,V_2)，直到产生一对 (V_1,V_2) 满足条件 $V_1^2+V_2^2\le 1$，即 (V_1,V_2) 处于以（0,0）为中心的单位圆内，显然这样得到的 (V_1,V_2) 是在单位圆内均匀分布的随机向量。取 $(\overline{R},\overline{\Theta})$ 为 V_1,V_2 的极坐标。容易验证 $\overline{R}$ 与 $\overline{\Theta}$ 相互独立，并且 $\overline{R}^2$ 是（0，1）上的均匀随机变量，$\overline{\Theta}$ 是 $(0,2\pi)$ 上的均匀随机变量（见习题 10.13.)。

由于

$$\sin\overline{\Theta}=\frac{V_2}{\overline{R}}=\frac{V_2}{\sqrt{V_1^2+V_2^2}} \qquad \cos\overline{\Theta}=\frac{V_1}{\overline{R}}=\frac{V_1}{\sqrt{V_1^2+V_2^2}}$$

由（2.5）式知，我们可以通过生成随机数 U 并令

$$X=(-2\ln U_1)^{1/2}V_1/\overline{R} \qquad Y=(-2\ln U_1)^{1/2}V_2/\overline{R}$$

生成独立标准正态分布随机变量 X 和 Y。事实上，因为（在 $V_1^2+V_2^2\le 1$ 的条件下）$\overline{R}^2$ 是（0，1）均匀随机变量，与 $\overline{\Theta}$ 独立，所以 $\overline{R}^2$ 可代替 U 而不必重新产生新的随机数，这表明

$$X=(-2\ln\overline{R}^2)^{1/2}V_1/\overline{R}=\sqrt{\frac{-2\ln S}{S}}V_1$$

$$Y=(-2\ln\overline{R}^2)^{1/2}V_2/\overline{R}=\sqrt{\frac{-2\ln S}{S}}V_2$$

是相互独立的标准正态随机变量，其中

$$S=\overline{R}^2=V_1^2+V_2^2$$

综上，对于生成一对独立的标准正态随机变量，我们有如下方法：

第 1 步 产生随机数 U_1,U_2；

第 2 步 令 $V_1=2U_1-1$， $V_2=2U_2-1$， $S=V_1^2+V_2^2$；

第 3 步 若 S>1,转向第 1 步；

第 4 步 得到独立的标准正态随机变量.

$$X=\sqrt{\frac{-2\ln S}{S}}V_1 \quad Y=\sqrt{\frac{-2\ln S}{S}}V_2$$

上述的方法称为*极坐标法*，因为正方形中的点同时也在圆中的概率为 $\pi/4$ (圆面积与正方形面积之比)，所以极坐标法平均要经过 $4/\pi\approx 1.273$ 次步骤 1。因此，要产生 2 个独立标准正态分布随机变量平均需要 2.546 个随机数，一次求对数，一次求平方根，一次除法和 4.546 次乘法。

例 2e（**模拟一个 χ^2 随机变量**）自由度为 n 的 χ^2 分布就是随机变量 $\chi_n^2=Z_1^2+\cdots+Z_n^2$ 的分布，其中 $Z_1,\cdots$， Z_n 是独立标准正态分布的随机变量，在第 6 章 6.3 节指出，$Z_1^2+Z_2^2$ 具有 χ^2 分布，均值为 2， 因此当 n 为偶数时（例如 $n=2k$），则 χ_{2k}^2 的分布为 Γ 分布，参数为（k，1/2)。这样，$-2\ln(\prod_{i=1}^{k}U_i)$ 具有自由度为 2k 的 χ^2 分布。当 $n=2k+1$ 为奇数时，我们只需先模拟一个标准正态随机变量 Z，再将 Z^2 加到 χ_{2k}^2 可得到 2k+ 1 个自由度的 χ^2 随机变量，即

$$\chi_{2k+1}^2=Z^2-2\ln(\prod_{i=1}^{k}U_i)$$

其中 $Z,U_1,\cdots,U_k$ 相互独立， Z 为标准正态随机变量，$U_1,\cdots,U_k$ 为（0，1）随机变量。

10.3 模拟离散分布

所有模拟具有连续分布函数的随机变量所使用的方法，都可适用于离散随机变量的模拟.例如，我们希望模拟具有下列分布列的随机变量 X:

$$P\{X = x_j\} = P_j \qquad j = 0,1,\cdots,\sum_j P_j = 1$$

可利用下面的方法，它是逆变换方法的离散版本.

设 U 为随机数，令

$$X = \begin{cases} x_1 & U \le P_1 \\ x_2 & P_1 < U \le P_1 + P_2 \\ \cdots & \\ x_j & \sum_{i=1}^{j-1} P_i < U \le \sum_{i=1}^{j} P_i \\ \cdots & \end{cases}$$

由于

$$P\{X = x_j\} = P\left\{\sum_{i=1}^{j-1} P_i < U \le \sum_{i=1}^{j} P_i\right\} = P_j$$

我们可以看出，所产生的随机变量 X 具有离散分布列 $\{P_j, j = 1,2,\cdots\}$。

例 3a（几何分布）假设有一独立重复试验，每次成功的概率为 p， $0 < p < 1$，试验一直进行到出现成功为止，记 X 为试验的次数，则

$$P\{X = i\} = (1-p)^{i-1} p \qquad i \ge 1$$

$X = i$ 表示前 i - 1 次试验的均为失败，而第 i 次成功。称随机变量 X 是参数为 p 的几何随机变量。 因为

$$\sum_{i=1}^{j-1} P\{X = i\} = 1 - P\{X > j-1\}$$

$$= 1 - P\{\text{前 j}-1\text{次试验均为失败}\} = 1 - (1-p)^{j-1} \qquad j \ge 1$$

所以 X 可以由下列方式产生： 产生一个随机数 U，当

$$1 - (1-p)^{j-1} < U \le 1 - (1-p)^j$$

时，X 取为 j，上式与

$$(1-p)^j \le 1 - U < (1-p)^{j-1}$$

是等价的，又由于 U 与 1-U 具有相同的分布，因此，我们可以定义 X 为

$$X=\min\left\{j:(1-p)^j\le U\right\}=\min\left\{j:j\ln(1-p)\le\ln U\right\}=\min\left\{j:j\ge\frac{\ln U}{\ln(1-p)}\right\}$$

上式中有一个不等号反向的过程，这是因为 $\ln(1-p)$ 是负的[原因是 $\ln(1-p)<\ln 1=0$]。利用记号 $[x]$（$[x]$为不超过 x 的最大整数），X 可以写成

$$X=1+\left[\frac{\ln U}{\ln(1-p)}\right]$$

和连续情形类似，对于常见的离散型随机变量，有一些特殊的模拟方法。 在这里我们例举两个。

例 3b（**模拟二项随机变量**）参数为(n,p)的二项随机变量可以表示成n个独立的伯努利随机变量之和，利用这一点，很容易进行模拟，设$U_1,\cdots,U_n$为一组随机数，令

$$X_i=\begin{cases}1 & U_i<p\\ 0 & 其他\end{cases}$$

易知 $X\equiv\sum_{i=1}^{n}X_i$是二项随机变量，其参数为(n,p).

例 3c（**模拟泊松随机变量**）要模拟一个均值为λ的泊松随机变量，首先产生一组随机数 $U_1,U_2,\cdots$，记

$$N=\min\left\{n:\prod_{i=1}^{n}U_i<\mathrm{e}^{-\lambda}\right\}$$

随机变量 $X\equiv N-1$即为所求。 也就是说，连续产生一些随机数直到他们的积小于 $\mathrm{e}^{-\lambda}$，此时所产生的随机数的个数减去 1 即为均值为 1 的泊松随机变量。注意，

$$X+1=\min\left\{n:\prod_{i=1}^{n}U_i<\mathrm{e}^{-\lambda}\right\}$$

与

$$X=\max\left\{n:\prod_{i=1}^{n}U_i\ge\mathrm{e}^{-\lambda}\right\}\qquad 其中\prod_{i=1}^{\infty}U_i\equiv 1$$

是等价的。两边取对数得

$$X=\max\left\{n:\sum_{i=1}^{n}\ln U_i\ge-\lambda\right\}$$

$$X=\max\left\{n:\sum_{i=1}^{n}-\ln U_i\le\lambda\right\}$$

然而，$-\ln U_i$是均值为 1 的指数随机变量。现在考虑一个泊松过程，其强度为 1，易知，这个过程在$(0,\lambda)$上事件的个数服从泊松分布，其均值为λ，而$(0,\lambda)$上每两个相邻事件之间的时间间隔刚好是参数为 1 的指数分布，而且这些时间间隔又相互独立。

由 X 的表达式知，其分布刚好与强度为 1 的泊松过程在$(0,\lambda)$上事件的个数的分布相同。 因此， X 的分布为泊松分布，其均值为λ。

10.4 方差缩减技术

令随机变量 $X_1,\cdots,X_n$ 具有给定的联合分布，现在我们希望计算

$$\theta \equiv \mathrm{E}[g(X_1,\cdots,X_n)]$$

其中 g 是一个已知的函数。有时会发现用解析的方法计算这个值是十分困难的， 在这种时候，可以利用模拟技术去估计 θ 的值。 方法如下：产生一组随机变量 $X_1^{(1)},\cdots,X_n^{(1)}$，使得 $X_1^{(1)},\cdots,X_n^{(1)}$ 与 $X_1,\cdots,X_n$ 具有相同的联合分布，令

$$Y_1 = g(X_1^{(1)},\cdots,X_n^{(1)})$$

同时，还可以模拟一组 $X_1^{(2)},\cdots,X_n^{(2)}$，使得它与第一组变量相互独立并且具有相同的联合分布，令

$$Y_2 = g(X_1^{(2)},\cdots,X_n^{(2)})$$

重复上述步骤直到得到 $Y_1,Y_2,\cdots,Y_k$（k 为事先确定），$Y_1,Y_2,\cdots,Y_k$ 相互独立，且与 $g(X_1,\cdots,X_n)$ 分布相同，记 $\overline{Y}$ 为 $Y_1,Y_2,\cdots,Y_k$ 的平均，即若

$$\overline{Y} = \frac{1}{k}\sum_{i=1}^{k} Y_i$$

则

$$\mathrm{E}[\overline{Y}] = \theta, \quad \mathrm{E}[(\overline{Y}-\theta)^2] = \mathrm{Var}(\overline{Y})$$

因此，我们可以用 $\overline{Y}$ 作为 θ 的一个估计。因为 $\overline{Y}$ 与 θ 之间的均方误差等于 $\overline{Y}$ 的方差，所以我们希望 $\mathrm{Var}(\overline{Y})$ 越小越好. [在我们所讨论的情况下， $\mathrm{Var}(\overline{Y}) = \frac{1}{k}\mathrm{Var}(Y_1)$，而通常这个量是不知道的，我们必须设法从 $Y_1,Y_2,\cdots,Y_k$ 的值去估计它。] 现在我们需要介绍几个方差缩减技术.

10.4.1 利用对偶变量

前面我们提到，用来估计 θ 所产生的随机变量 Y_1 与 Y_2 相互独立且同分布。具有均值 θ 现在讨论关于方差的一个公式

$$\begin{aligned}\mathrm{Var}\left(\frac{Y_1+Y_2}{2}\right) &= \frac{1}{4}[\mathrm{Var}(Y_1)+\mathrm{Var}(Y_2)+2\mathrm{Cov}(Y_1,Y_2)] \\ &= \frac{1}{2}\mathrm{Var}(Y_1)+\frac{1}{2}\mathrm{Cov}(Y_1,Y_2)\end{aligned}$$

因此（在方差减小的意义上），我们更希望 Y_1 与 Y_2 负相关而不是相互独立。现在看一

看如何实现方差缩减，先假定 $X_1,\cdots,X_n$ 是相互独立的，此外，每一个变量都是利用逆变换方法产生的，即 $X_i = F_i^{-1}(U_i)$，其中 U_i 是随机数，F_i 是 X_i 的分布函数，Y_1 可表示成

$$Y_1 = g(F_1^{-1}(U_1),\cdots,F_n^{-1}(U_n))$$

因为当 U 是随机数时，$1-U$ 也是（0，1）随机数，所以 $1-U$ 与 U 具有负相关，若定义

$$Y_2 = g(F_1^{-1}(1-U_1),\cdots,F_n^{-1}(1-U_n))$$

则 Y_2 与 Y_1 同分布。因此，若 Y_1，Y_2 负相关，那么 $(Y_1+Y_2)/2$ 的方差就会比 $\frac{1}{2}\text{Var}(Y_1)$ 小。（另外，从计算的角度，也节省了计算量，在产生 Y_2 的时候，就不必产生 n 个新的随机数，而是用 $1-U_i$（$i=1,\cdots,n$）代之。）虽然通常不能确定 Y_1 和 Y_2 为负相关，但一般而言这是正确的，尤其当 g 是单调函数时，确实可以证明它们是负相关的。

10.4.2 利用“条件”

首先回忆下列条件方差公式（7.5.4 节）

$$\text{Var}(Y) = \text{E}[\text{Var}(Y|Z)] + \text{Var}(E[Y|Z])$$

我们现在估计 $\text{E}[g(X_1,\cdots,X_n)]$，先模拟 $\boldsymbol{X}=(X_1,\cdots,X_n)$，再计算 $Y=g(\boldsymbol{X})$。若存在某随机变量 Z，并且能够计算 $\text{E}[Y|Z]$，那么，因为 $\text{Var}(Y|Z)\geq 0$，所以利用条件方差公式得

$$\text{Var}(\text{E}[Y|Z]) \leq \text{Var}(Y)$$

因此，由 $\text{E}[\text{E}[Y|Z]]=\text{E}[Y]$，可知 $\text{E}[Y|Z]$ 是一个比 Y 更好的估计。

例 4a（估计 π）令 U_1,U_2 为随机数，记 $V_i = 2U_i - 1, i=1,2,$，在例 2d 中已经指出 (V_1,V_2) 在面积为 4 中心为（0，0）的一个正方形内均匀分布，随机点 (V_1,V_2) 落在半径为 1 的以（0，0）为圆心的圆内的概率为 $\pi/4$（见图 10.2，$\pi/4$ 等于内接圆与正方形面积之比）。现在可模拟数组 (V_1,V_2) n 次，令

$$I_j = \begin{cases} 1 & \text{第 } j \text{ 次模拟落在单位圆内} \\ 0 & \text{其他} \end{cases}$$

故 $I_1,\cdots,I_n$ 是独立同分布的随机变量，$\text{E}[I_j]=\pi/4$，利用强大数定律，

$$\frac{I_1+\cdots+I_n}{n} \to \frac{\pi}{4} \quad n\to\infty$$

因此，通过模拟大量的数组 (V_1,V_2) 并将这个比例乘以 4，可得 π 的估计。

上述估计可以通过条件期望进行改善。对于示性变量 I，考虑条件概率

$$\mathrm{E}[I \mid V_1] = P\left\{V_1^2 + V_2^2 \le 1 \mid V_1\right\} = P\left\{V_2^2 \le 1 - V_1^2 \mid V_1\right\}$$

由于

$$\begin{aligned} &P\left\{V_2^2 \le 1 - V_1^2 \mid V_1 = v\right\} = P\left\{V_2^2 \le 1 - v^2\right\} \\ &= P\left\{-\sqrt{1-v^2} \le V_2 \le \sqrt{1-v^2}\right\} = \sqrt{1-v^2} \end{aligned}$$

故

$$\mathrm{E}[I \mid V_1] = \sqrt{1 - V_1^2}$$

因此，利用 $\sqrt{1-V_1^2}$ 的平均值为 $\pi/4$ 的估计是原来的估计的改进。 又由于

$$\mathrm{E}\left[\sqrt{1-V_1^2}\right] = \int_{-1}^{1} \frac{1}{2}\sqrt{1-v^2}\,\mathrm{d}v = \int_0^1 \sqrt{1-u^2}\,\mathrm{d}u = \mathrm{E}\left[\sqrt{1-U^2}\right]$$

其中 U 为随机数，因此，我们模拟产生 n 个随机数，利用 $\sqrt{1-U^2}$ 的平均值作为 $\pi/4$ 的估计具有更高的精度（习题 10.14 指出，利用 $\sqrt{1-V^2}$ 产生的估计与利用 $\sqrt{1-U^2}$ 产生的估计具有相同的精度。）.

这个估计可以进一步改进。注意到 $g(u) = \sqrt{1-u^2}, 0 \le u \le 1$ 是单调递减函数，利用对偶变量法可以减少 $\mathrm{E}\left[\sqrt{1-U^2}\right]$ 的估计量的方差。即我们不产生 n 个随机数，而是利用 $\sqrt{1-U^2}$ 的平均值作为 $\pi/4$ 的估计，因此产生 $n/2$ 个随机数 U，然后用 $n/2$ 个 $(\sqrt{1-U^2} + \sqrt{1-(1-U^2)})/2$ 的平均值作为 $\pi/4$ 的估计，就得到一个改进的估计值。表列 10-1 列出了当 $n = 10000$ 时，π 的估计值。

表 7-1

方法	π 的估计
落入单位圆内随机点的比例	3.1612
$\sqrt{1-U^2}$ 的平均值	3.128448
$\frac{1}{2}(\sqrt{1-U^2} + \sqrt{1-(1-U^2)})$ 的平均值	3.139578

利用最后一个方法，当 $n = 64000$ 时，π 的估值为 3.143288。

10.4.3 控制变量

假如我们希望通过模拟估计 $\mathrm{E}[g(\boldsymbol{X})]$，其中 $\boldsymbol{X} = (X_1, \cdots, X_n)$. 但是我们已知某 $f(\boldsymbol{X})$ 的期望，例如 $\mathrm{E}[f(\boldsymbol{X})] = \mu$。对于任何常数 a 我们用

$$W = g(\boldsymbol{X}) + a[f(\boldsymbol{X}) - \mu]$$

作为$\mathrm{E}[g(\boldsymbol{X})]$的估计量，此时

$$\operatorname{Var}(W)=\operatorname{Var}[g(\boldsymbol{X})]+a^2\operatorname{Var}[f(\boldsymbol{X})]+2a\operatorname{Cov}[g(\boldsymbol{X}),f(\boldsymbol{X})] \qquad (4.1)$$

通过简单运算可知，式（4.1）在

$$a=-\frac{\operatorname{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]}{\operatorname{Var}[f(\boldsymbol{X})]} \qquad (4.2)$$

时达到最小值，其最小值为

$$\operatorname{Var}(W)=\operatorname{Var}[g(\boldsymbol{X})]-\frac{\operatorname{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]^2}{\operatorname{Var}[f(\boldsymbol{X})]} \qquad (4.3)$$

但是，$\operatorname{Var}[f(\boldsymbol{X})]$和$\operatorname{Cov}[f(\boldsymbol{X}),g(\boldsymbol{X})]$通常是未知的，因此，得不到所需的方差缩减。实践中，我们可以利用模拟数据去估计这个值。 理论上我们可以利用这个方法对所有的模拟结果缩减相应的方差。

小结

设F是一个连续的分布函数，U是$(0,1)$上的均匀分布的随机变量（称为随机数）则$F^{-1}(U)$具有分布F，其中$F^{-1}(u)$是方程$F(x)=u$的解，这种由随机数构造其他随机变量的方法称为逆变换方法。

另一个产生随机变量的方法称为舍取法。 假定对于密度函数g，我们已经有一个产生随机变量的成熟的方法，现在希望模拟一个具有密度函数f的随机变量。 我们首先确定一个常数c，它满足

$$\max\frac{f(x)}{g(x)}\le c$$

然后经过下列步骤：

1. 产生Y，其密度为g。
2. 产生随机数U。
3. 若$U\le f(Y)/[cg(Y)]$，则令$X=Y$，过程中止。
4. 回到第 1 步。

此方法循环的次数具有几何分布，其平均值为c.

标准正态随机变量可通过舍取法产生（g为指数密度，均值为 1）或者利用极坐标方法产生。

为了估计某一个参数θ，首先模拟一个随机变量，使得它的期望值为θ。然后，利用统计方法缩减其相应的方差。 本文中介绍了三种缩减方差的方法：

1. 利用对偶变量。
2. 利用条件期望。

3. 利用控制变量。