第四章　随机变量的数字特征

4.1 随机变量的数学期望

例1. 甲、乙两人进行打靶, 所射中环数分别记为 X_1、X_2, 它们的频率分别为:

X_1	8	9	10	X_2	8	9	10
p_k	0.3	0.1	0.6	p_k	0.2	0.5	0.3

试评定他们射击技术的好坏.

若使两个射手各射N枪，则他们打中的环数大约是：

甲： $8\times0.3N+9\times0.1N+10\times0.6N=9.3N;$

乙： $8\times0.2N+9\times0.5N+10\times0.3N=9.1N.$

他们平均射中的环数约为

$$\text{甲：}\ \bar{x}\approx8\times\frac{0.3N}{N}+9\times\frac{0.1N}{N}+10\times\frac{0.6N}{N}=9.3$$

$$\text{乙：}\ \bar{y}\approx8\times\frac{0.2N}{N}+9\times\frac{0.5N}{N}+10\times\frac{0.3N}{N}=9.1$$

平均起来甲每枪射中 9.3环，乙射中9.1环，因此，甲的技术要好些。

- 受此问题启发在上式中用概率代替频率引入如下定义：

定义: 设离散型 r.v. X 的分布律为: $P\{X = x_k\} = p_k, k = 1, 2, \cdots$

若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 为随机变量的数学期望(也叫做平均值), 记作E(X), 即 $E(X) = \sum_{k=1}^{\infty} x_k p_k$.

当$\sum_{i=1}^{\infty} |x_i| p_i$发散时，则说$X$的数学期望不存在 。

➢ 例2. 设 X 为投掷一颗骰子时出现的点数，则 X 的分布律为

$$P\{X=i\}=1/6,\ i=1,\ 2,\ \ldots,\ 6;$$

于是，X 的数学期望为：

$$\mathrm{E}(X)=\frac{1}{6}\times 1+\frac{1}{6}\times 2+\cdots+\frac{1}{6}\times 6=\frac{7}{2}$$

➢ 离散型随机变量的期望值：

1) (0-1) 分布

设 X 服从(0-1)分布，分布律为：

$$P\{X=1\}=p,\ P\{X=0\}=q,\ 0<p<1,\ q=1-p$$

X 的数学期望为 $E(X)=1\cdot p+0\cdot q=p$

2) 二项分布: 设 $X\sim b(n,p)$, $P\{X=k\}=C_n^k\ p^k q^{n\text{-}k}$, $k=0,\ 1,\ \cdots,n.$

解: $\mathrm{E}(X)=\sum_{k=0}^{n}k\cdot P\{X=k\}=\sum_{k=0}^{n}k\cdot C_n^k\ p^k q^{n\text{-}k}$

$$=\sum_{k=0}^{n}k\cdot\frac{n!}{k!(n-k)!}p^k q^{n\text{-}k}$$

$k=0$, 对均值的贡献为0;

$$=\sum_{k=1}^{n}\frac{np(n-1)!}{(k-1)![(n-1)-(k-1)]!}p^{k\text{-}1}q^{(n\text{-}1)\text{-}(k\text{-}1)}$$

$$=np\sum_{k'=0}^{n-1}\underbrace{\frac{(n-1)!}{k'![(n-1)-k']!}}_{C_{n-1}^{k'}}p^{k'}q^{(n\text{-}1)\text{-}k'}\ (\text{令}k'=k-1)$$

$$=np(p+q)^{n-1}=np.$$

3) 泊松分布: 设$X\sim\pi(\lambda)$, 即 $P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, k=0, 1, 2, \cdots \quad \lambda>0.$

解： $\mathrm{E}(X)=\sum\limits_{k=0}^{+\infty} k\cdot\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\sum\limits_{k=1}^{+\infty}\dfrac{\lambda^{k-1}}{(k-1)!}\cdot\lambda$

$$=\mathrm{e}^{-\lambda}\cdot\lambda\sum_{k'=0}^{+\infty}\frac{\lambda^{k'}}{k'!}$$

$$=\mathrm{e}^{-\lambda}\times\lambda\times\mathrm{e}^{\lambda}=\lambda$$

4) **几何分布** $p_k = q^{k-1}p, k = 1,2,\cdots$ 第一次成功发生在第 k 次试验的概率

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} kp_k = \sum_{k=1}^{\infty} kq^{k-1}p = p(1+2q+3q^2+\cdots)$$

$$= p(q+q^2+q^3+\cdots)'$$

$$= p(\frac{q}{1-q})' = p\frac{1}{(1-q)^2} = \frac{1}{p}$$

例3：随机变量X的分布率为$P\{X=x_k\}=\dfrac{1}{2^k},k=1,2,\cdots,$ 其中$x_k=(-1)^k\dfrac{2^k}{k}$.

显然级数 $\sum_{k=1}^{\infty}x_k p_k=\sum_{k=1}^{\infty}(-1)^k\dfrac{1}{k}=-\ln 2,$

但由于$\sum_{k=1}^{\infty}|x_k|p_k=\sum_{k=1}^{\infty}\dfrac{1}{k}=\infty$，因此$X$的期望不存在

➢ 连续型随机变量的数学期望：

设 $f(x)$ 为连续型随机变量 X 的概率密度，对 X 的取值区间作一分割，有：

$P\{x_i < X \leq x_{i+1}\} \approx f(x_i)\Delta x_i$，当$\Delta x_i \to 0^+$时，近似地有

$$E(X) \approx \sum_i x_i f(x_i)\Delta x_i \to \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x.$$

定义: 设连续型$r.v.X$的概率密度为$f(x)$,若积分$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ 绝对收敛,

称积分$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$的值为r.v.$X$的**数学期望**, 记为$\mathrm{E}(X)$. 即

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} xf(x)\mathrm{d}x.$$

数学期望简称期望, 又称为**均值**.

- 下面计算常用连续型变量的数学期望：

1^0 均匀分布: 设r.v. X服从区间$[a, b]$ 上的均匀分布，即 X 的密度函数为:

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & otherwise.\end{cases}$$

则$E(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x=\int_a^b x\cdot\dfrac{1}{b-a}\mathrm{d}x=\dfrac{a+b}{2}$, 是区间 $[a, b]$ 的中点；

2⁰ 指数分布: 设r.v.X服从参数为λ的指数分布，则其密度函数为:

$$f(x)=\begin{cases}\lambda \mathrm{e}^{-\lambda x}, x\ge 0,\\ 0,\quad x<0.\end{cases}$$

$$\mathrm{E}(X)=\int_0^{+\infty} x\cdot\lambda \mathrm{e}^{-\lambda x}\mathrm{d}t \overset{\text{令}t=\lambda x}{=} \frac{1}{\lambda}\int_0^{+\infty} t\cdot \mathrm{e}^{-t}\mathrm{d}t$$

$$=\frac{1}{\lambda}\left[-t\mathrm{e}^{-t}-\mathrm{e}^{-t}\right]\Big|_0^{+\infty}=\frac{1}{\lambda},$$

✓ 指数分布是最常用的“寿命分布”之一，期望表明 λ 值越小，产品平均寿命越长。

3^0 正态分布: 设$X \sim N(\mu,\sigma^2)$, 则

$$\mathrm{E}(X)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} x\cdot \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \ (\text{令} t=x-\mu)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(t+\mu)\cdot \mathrm{e}^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} t\cdot \mathrm{e}^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t+\mu\cdot\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2\sigma^2}}\mathrm{d}t$$

$$\therefore \mathrm{E}(X)=\mu.$$

第一项被积函数为奇函数,因而积分为0, 而第二项后一部分为1

例4. 设X服从柯西分布，其密度为$f(x)=\dfrac{1}{\pi\cdot(1+x^2)}, -\infty<x<+\infty.$ 由于

$$\int_{-\infty}^{+\infty}|x|\frac{1}{\pi(1+x^2)}\mathrm{d}x=\infty$$

因此，柯西分布的数学期望不存在.

$$\left(\frac{1}{2}\log(1+x^2)\right)'=\frac{x}{(1+x^2)}$$

例5. 设X概率密度为$f(x)=\dfrac{1}{2}\mathrm{e}^{-|x|},-\infty<x<+\infty.$

$$\text{则}\mathrm{E}(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\int_{-\infty}^{+\infty}x\cdot\frac{1}{2}\mathrm{e}^{-|x|}\mathrm{d}x$$

$$=\frac{1}{2}[\int_{-\infty}^{0}x\mathrm{e}^{x}\mathrm{d}x+\int_{0}^{+\infty}x\mathrm{e}^{-x}\mathrm{d}x]$$

$$=\frac{1}{2}[-1+1]=0$$

➢ 随机变量函数的数学期望公式:

定理: 设Y是r.v.X的函数, $Y=g(X)$ (g是连续函数)

(i) X是离散型r.v., 它的分布律为 $p_k=P\{X=x_k\}, k=1,2,\cdots,$

若$\sum_{k=1}^{+\infty} g(x_k)p_k$绝对收敛,则 $E(Y)=E[g(X)]=\sum_{k=1}^{+\infty} g(x_k)p_k$.

(ii) X是连续型r.v., 它的概率密度为$f(x)$, 若 $\int_{-\infty}^{+\infty} g(x)f(x)dx$

绝对收敛, 则$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)dx$.

➢ 说明:

- 1. 在已知 Y 是 X 的连续函数前提下, 当我们求 E(Y) 时不必知道 Y 的分布, 只需知道 X 的分布就可以了.

- 2. 上述定理可以推广到多维 r.v.函数.

若(X,Y)为离散型$r.v.$ 其分布律为$P\{X=x_i,Y=y_j\}=p_{ij}$,

$i,j=1,2,3,\cdots$, 则有 (假设级数绝对收敛):

$$\mathrm{E}(Z)=\mathrm{E}[g(X,Y)]=\sum_{j=1}^{+\infty}\sum_{i=1}^{+\infty}g(x_i,y_j)p_{ij} \qquad (4.1)$$

(X,Y)为连续型$r.v.$， $Z=g(X,Y)$(g是连续函数) 是 $r.v.\ X,Y$ 的函数, 若(X,Y)的概率密度为 $f(x,y)$，则 $r.v.Z$ 的期望为(假设积分绝对收敛):

$$\mathrm{E}(Z)=\mathrm{E}\left[g(X,Y)\right]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y, \quad (4.2)$$

例6．设$X\sim N(0,1)$，求$\mathrm{E}(X^2)$.

解法1：求得$Y=X^2$的密度 $f_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}y^{1/2}}\mathrm{e}^{-\frac{y}{2}}, y>0\\ 0, y\le 0\end{cases}$

$$\text{则}\mathrm{E}(X^2)=\mathrm{E}(Y)=\int_{-\infty}^{+\infty} yf_Y(y)\mathrm{d}y$$

$$=\int_0^{+\infty}\frac{2}{\sqrt{\pi}}\left(\frac{y}{2}\right)^{3/2-1}\mathrm{e}^{-\frac{y}{2}}\mathrm{d}\left(\frac{1}{2}y\right)$$

$$=\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right)=1$$

例6. 设$X \sim N(0,1)$，求$\mathrm{E}(X^2)$.

解法2：由定理 $\mathrm{E}(X^2)=\int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{d}x=\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x$

$$=-\int_{-\infty}^{+\infty} x\mathrm{d}\left(\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\right)$$

$$=-\left[x\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\right]\Bigg|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x$$

$$=1$$

例7. 某商品的市场需求量 X 服从[2000,4000]上的均匀分布，每售出一吨挣 3 万元，售不出则每吨需保养费1万元，问应组织多少货源才能使收益最大。

解：设y为进货量，$y \in [2000, 4000]$, 收益为Z.则

$$Z = H(x) = \begin{cases} 3y, & \text{当}x \geq y\text{时} \\ 3x-(y-x), x < y \end{cases}$$

$$\text{于是}\mathrm{E}(Z) = \int_{-\infty}^{+\infty} H(x) f(x)\mathrm{d}x = \frac{1}{2000}\int_{2000}^{4000} H(x)\mathrm{d}x$$

$$= \frac{1}{2000}\int_{2000}^{y} (4x-y)\mathrm{d}x + \frac{1}{2000}\int_{y}^{4000} 3y\mathrm{d}x$$

$$= \frac{1}{1000}[-y^2 + 7000y - 4\times 10^6]$$

当 y=3500 时达到最大值，因此组织3500吨货源是最好的决策

例8. 设二维随机变量(X,Y)的概率密度如下，求：XY 的数学期望.

$$f(x,y)=\begin{cases} x+y, & 0\le x\le 1,\ 0\le y\le 1, \\ 0, & \text{其它}, \end{cases}$$

解：由式(4.2)可得

$$\mathrm{E}(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)\mathrm{d}x\mathrm{d}y$$

$$=\int_0^1\int_0^1 xy(x+y)\mathrm{d}x\mathrm{d}y=\frac{1}{3}.$$

$$\mathrm{E}(Z)=\mathrm{E}\left[g(X,Y)\right]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y, \qquad (4.2)$$

例9. 已知(X,Y)的联合分布律为$P\{X=i,Y=j\}=p^2q^{j-2}, i=1,2,\cdots j-1,$

$j=2,3\cdots,\quad 0<p<1, q=1-p,$ 求：(1) $\mathrm{E}(XY)$；(2) $\mathrm{E}(\frac{X}{Y})$.

解：$$\mathrm{E}(XY)=\sum_{j=2}^{\infty}\sum_{i=1}^{j-1} ijP\{X=i,Y=j\}=\sum_{j=2}^{\infty} j\cdot\frac{j(j-1)}{2}p^2q^{j-2}$$

$$=\frac{1}{2}p^2[(\sum_{j=2}^{\infty}q^{j+1})'''-(\sum_{j=2}^{\infty}q^{j})'']=\frac{1}{2}p^2(\frac{6}{p^4}-\frac{2}{p^3})=\frac{2+q}{p^2}$$

$$\mathrm{E}(X/Y)=\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\frac{i}{j}P\{X=i,Y=j\}=\sum_{j=2}^{\infty}\frac{(j-1)}{2}p^2q^{j-2}$$

$$=\frac{1}{2}p^2(\sum_{j=2}^{\infty}q^{j-1})'$$

$$=\frac{1}{2}p^2(\frac{q}{1-q})'=\frac{1}{2}$$

➢均值的性质（能证明吗？）：

(1) $E(c)=c$; (c为常数)

(2) $E(cX)=cE(X)$;(c为常数)

(3) $E(X+Y)=E(X)+E(Y)$;

(4) 设 X、Y 相互独立, 则 $E(XY)=E(X)E(Y)$;

(5) $|E(XY)|^2 \le E(X^2)E(Y^2)$. (许瓦尔兹不等式)

说明: i. 性质 (3) 和 (4) 可以推广到有限个 r.v. $(X_1, X_2, ..., X_n)$的情况

.

ii. 对于“和”，不要求$X_1, X_2, ..., X_n$ 相互独立; 对于“积”要求$X_1, X_2, ..., X_n$相互独立.

例10 二项分布的均值的计算:

设 $X\sim b(n, p)$,引入 r.v. X_i $(i=1, 2, \ldots, n)$, 它们是相互独立的且都服从0--1分布: $P\{X_i=1\}=p$, $P\{X_i=0\}=q$; X 表示 n 次独立重复试验中A发生的次数, X_i 表示第 i 次试验的结果: $X_i=1$表示 A 发生, $X_i=0$表示A不发生, 所以

$$X=\sum_{i=1}^{n}X_i \quad 故\ E(X)=\sum_{i=1}^{n}E(X_i)=np.$$

➢ 说明: 将 X 分解成数个 r.v.之和, 然后利用 r.v.和的数学期望等于r.v. 的数学期望之和来求解. 这个方法具有一定的普遍意义.

例11 一民航送客车载有20位旅客自机场开出，旅客有10个车站可以下车，如到达一个车站没有旅客下车就不停车。以X 表示停车的次数，求E(X)（设每位旅客在各个车站下车是等可能的，并设各旅客是否下车相互独立）

解：引入随机变量$X_i = \begin{cases} 0, \text{在第 } i \text{ 站没有人下车,} \\ 1, \text{在第 } i \text{ 站有人下车,} \end{cases}$， $i = 1, 2, \cdots, 10$

易知$X = X_1 + X_2 + \cdots + X_n$,现在来求E($X$);

按题意，任一旅客在第i 站不下车的概率为0.9 ，因此,20 位旅客都不在第i 站下车的概率为0.9^{20} ,在第i 站有人下车的概率为$1-0.9^{20}$ ，也就是：

$$P\{X_i=0\}=0.9^{20},.\quad P\{X_i=1\}=1-0.9^{20}, i=1,2,\cdots,10$$

由此$\mathrm{E}(X_i)=1-0.9^{20}, i=1,2,\cdots 10.$

$$\mathrm{E}(X)=\mathrm{E}(X_1+X_2+\cdots+X_{10})=\mathrm{E}(X_1)+\mathrm{E}(X_2)+\cdots+\mathrm{E}(X_{10})$$

$$=10(1-0.9^{20})=8.784(\text{次})$$

例 12 –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为一随机变量,其分布律为 $p(i),\ i \geq 0$.

–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：

$$\pi(s) = bX - (s - X)l \qquad 若 X \leq s$$

$$= sb \qquad 若 X > s$$

-期望利润为：

$$\mathrm{E}\left[\pi(s)\right]=\sum_{i=0}^{s}[bi-(s-i)l]p(i)+\sum_{i=s+1}^{\infty}sbp(i)$$

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

$$=(b+l)\sum_{i=0}^{s}ip(i)-(b+l)s\sum_{i=0}^{s}p(i)+\mathrm{sb}$$

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

- 为得到最佳的s值，考虑当s增加一个单位时期望利润的变化值；
- 利用上述公式得到，当 s 增加一个单位时，期望利润为

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

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

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

$$\sum_{i=0}^{s} p(i) < \frac{b}{b+l} \qquad (4.3)$$

- 由于(4.3)式的左边随着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(4.3)式的最大值.
- 因为：$\mathrm{E}[\pi(0)] < \cdots \mathrm{E}[\pi(s^*)] < \mathrm{E}[\pi(s^*+1)] > \mathrm{E}[\pi(s^*+2)] > \cdots$
- 这样, 囤货数量为s^*+1时将会使得期望利润达到最大.

回 顾

➢ 离散型随机变量的数学期望：

定义: 设离散型 r.v. X 的分布律为: $P\{X=x_k\}=p_k, k=1,2,\cdots$

若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 为随机变量的数学期望(也叫做平均值), 记作E(X), 即 $\mathrm{E}(X)=\sum_{k=1}^{\infty} x_k p_k$.

当$\sum_{i=1}^{\infty}|x_i|p_i$发散时，则说$X$的数学期望不存在；

➢ 连续型随机变量的数学期望：

定义: 设连续型$r.v.X$的概率密度为$f(x)$,若积分$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$ 绝对收敛,称积分$\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x$的值为r.v.$X$的数学期望, 记为$\mathrm{E}(X)$. 即

$$\mathrm{E}(X)=\int_{-\infty}^{+\infty} xf(x)\mathrm{d}x.$$

数学期望简称期望, 又称为均值.

- 随机变量函数的数学期望公式:

定理: 设Y是r.v.X的函数, $Y = g(X)$ (g是连续函数)

(i) X是离散型r.v., 它的分布律为 $p_k = P\{X = x_k\}, k = 1,2,\cdots,$

若$\sum_{k=1}^{+\infty} g(x_k)p_k$绝对收敛,则 $\mathrm{E}(Y) = \mathrm{E}[g(X)] = \sum_{k=1}^{+\infty} g(x_k)p_k.$

(ii) X是连续型r.v., 它的概率密度为$f(x)$, 若 $\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$

绝对收敛, 则$\mathrm{E}(Y) = \mathrm{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x.$

4.2 条件期望

- 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 1 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 $k \le \min\{n, m\}$

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

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

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

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

$$=\frac{C_n^k C_n^{m-k}}{C_{2n}^m}$$

- 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 2，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的联合分布连续，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 $f_Y(y)>0$

例 3 设 X 和 Y 的联合密度函数为 $f(x,y)=\dfrac{\mathrm{e}^{-x/y}\mathrm{e}^{-y}}{y}, 0<x<\infty,$ $0<y<\infty$，计算 $E(X|Y=y)$；

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

- **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，例如以下公式仍然成立：

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

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

- 事实上，给定 $Y=y$ 条件下的条件期望可以看成是减小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.

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**

$$\mathrm{E}(X)=\mathrm{E}(\mathrm{E}(X|Y)) \tag{1}$$

$$\mathrm{E}(X)=\sum_y \mathrm{E}(X|Y=y)P\{Y=y\} \tag{2a}$$

– 如果 Y 是连续型随机变量，密度函数为 $f_Y(y)$，则公式(1)变成

$$\mathrm{E}(X)=\int_{-\infty}^{\infty}\mathrm{E}(X|Y=y)f_Y(y)\mathrm{d}y \tag{2b}$$

➢ 式（1）在 X 和 Y 均为离散型随机变量情形时的证明.

$$\mathrm{E}(X)=\mathrm{E}(\mathrm{E}(X|Y)) \qquad (1)$$

– 我们只需证明：$$\mathrm{E}(X)=\sum_{y}\mathrm{E}(X|Y=y)P\{Y=y\} \qquad (2)$$

– 等式(2)的右边可以写为：

$$\begin{aligned}\sum_{y}\mathrm{E}(X|Y=y)P\{Y=y\}&=\sum_{y}\sum_{x}xP\{X=x|Y=y\}P\{Y=y\}\\&=\sum_{y}\sum_{x}x\frac{P\{X=x,Y=y\}}{P\{Y=y\}}P\{Y=y\}\\&=\sum_{y}\sum_{x}xP\{X=x,Y=y\}=\sum_{x}x\sum_{y}P\{X=x,Y=y\}\\&=\sum_{x}xP\{X=x\}=\mathrm{E}(X)\end{aligned}$$

$$\mathrm{E}(X)=\sum_{y}\mathrm{E}(X|Y=y)P\{Y=y\} \qquad (2)$$

- (2)式的解释:
  - 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 3

- 一个矿工在井下迷了路，迷路的地方有三个门，
- 若从第一个门出来，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

$$E(X)=E(X|Y=1)P\{Y=1\}+E(X|Y=2)P\{Y=2\}+E(X|Y=3)P\{Y=3\}$$

$$=\frac{1}{3}(E(X|Y=1)+E(X|Y=2)+E(X|Y=3)).$$

然而，
$$E(X|Y=1)=3$$
$$E(X|Y=2)=5+E(X) \qquad (3)$$
$$E(X|Y=3)=7+E(X)$$

$$E(X|Y=1)=3$$
$$E(X|Y=2)=5+E(X) \qquad (3)$$
$$E(X|Y=3)=7+E(X)$$

- (3)式解释：
- 例如E(X|Y=2)：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E(X)，因此E(X|Y=2)=5+ E(X)；
- 其余各等式的解释类似；
  - 因此 $E(X)=\frac{1}{3}(3+5+E(X)+7+E(X))$
  - 从而 E(X)=15.

例4 （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

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

但， $\mathrm{E}(\sum_{i=1}^{N} X_i \middle| N = n) = \mathrm{E}(\sum_{i=1}^{n} X_i)$ 由X_i与N的独立性

$=n\mathrm{E}(X)$ 其中$\mathrm{E}(X) = \mathrm{E}(X_i)$

由此可得，$\mathrm{E}(\sum_{i=1}^{N} X_i | N) = N\mathrm{E}(X)$

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$50 \times 8 = 400$元。

例5 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) $\mathrm{E}(R\,|\,\text{玩家赢})$； (c) $\mathrm{E}(R\,|\,\text{玩家输})$.

解：如果令P_j表示每次掷骰子得到两枚骰子点数之和为 j 的概率，则有：

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

– 为求 $E[R]$，记s 为第一次掷出点数，则给定 s的条件下，有：

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

- 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 $\frac{1}{P_i + P_7}+1$，其中+1表示加上第一次掷骰子，因此有：

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

- 为求 $\mathrm{E}[R|\text{赢}]$，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

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

- 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 $Q_i = P\{S=i|赢\}$ 我们有：

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

- 对于 $i = 4,5,6,8,9,10$

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

- 对第一次掷出的点数和取条件可得：

$$\mathrm{E}(R|赢) = \sum_i \mathrm{E}(R|赢，\ S=i)Q_i = \sum_i \mathrm{E}(R|S=i)Q_i$$

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

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

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

- 由此可得：

$$\mathrm{E}(R|\text{输}) = \frac{\mathrm{E}(R) - \mathrm{E}(R|\text{赢})p}{1-p} = 3.801$$

$$\mathrm{E}(R|S=i) = \begin{cases} 1 & i = 2,3,7,11,12 \\ 1 + \dfrac{1}{P_i + P_7} & \text{其他} \end{cases} \quad Q_i = \frac{P\{S=i,\text{赢}\}}{P\{\text{赢}\}} = \frac{P_i P\{\text{赢}|S=i\}}{p} = \frac{P_i^2}{p(P_i + P_7)}$$

例 6 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 出现的次数，$i = 1, 2, \cdots, k$. 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 $I = \begin{cases} 0, & N_i = 0 \\ 1, & N_i > 0 \end{cases}$；

那么 $E[\mathrm{N}_j]$ 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。
- 且相互独立，是 j 的概率为 $P(j \mid 不是 i) = \dfrac{p_j}{1-p_i}$ ；
- 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 $(n-r,\ \dfrac{p_j}{(1-p_i)})$ 。
- 又因为 $P\{N_i = 0\} = (1-p_i)^n$ ，上面E(N_j)的等式变成

$$E(N_j) = E(N_j \mid N_i = 0)P\{N_i = 0\} + E(N_j \mid N_i > 0)P\{N_i > 0\}$$

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

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

- 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

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

- 最后得到

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

例 7 假设有 r 个玩家在赌博，玩家 i 最初拥有 n_i 单位赌资，$n_i>0,\ i=1,\ldots,r$,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 $n\text{-}j$ 个单位.

记 X_j 表示将要进行的对局数，$m_j = \mathrm{E}(X_j)$，对 $j=1,\ldots,n\text{-}1$ 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 $n\text{-}(j+1)$ 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， $m_{j+1} = 2m_j - m_{j-1} - 2 \qquad j = 1, \cdots, n-1 \qquad (4.4)$

- 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

$$m_i = i(m_1 - i + 1) \qquad i = 1, \cdots, n \qquad (4.5)$$

- 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.

$$m_{j+1}=2m_j-m_{j-1}-2 \qquad j=1,\cdots,n-1 \qquad (4.4)$$

✓ 利用(4.4)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（4.5）的归纳证明.

➢ 在式（4.5）中令$i=n$，并利用 $m_n=0$，可得 $m_1=n-1$

- 再次利用式（4.5），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

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

- 所以

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

- 但是因为每个对局包含两名玩家，所以

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

- 上式两边取期望后得到

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

- 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.

例8 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$\mathrm{E}[N]$；

解： 我们通过求解一个更一般的结果，来得到 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$U_1=y$作为条件, 得到

$$m(x)=\int_0^1 \mathrm{E}\left[N(x)\middle|U_1=y\right]\mathrm{d}y \qquad (4.6)$$

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

$$\mathrm{E}\left[N(x)\middle|U_1=y\right]=\begin{cases}1 & y>x\\ 1+m(x-y) & y\le x\end{cases} \qquad (4.7)$$

当$y \leq x$时，需要继续取$U_2, \cdots,$这相当于序列从U_2开始，要求刚好超过$x-y$的最小时刻。

将式（4.7）代入式（4.6），得到

$$m(x) = 1 + \int_0^x m(x-y)\mathrm{d}y$$

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

上式求微商得到 $m'(x) = m(x)$

或等价地， $\dfrac{m'(x)}{m(x)} = 1$

再对上式求积分，得 $\ln\left[m\left(x\right)\right]=x+c$

或 $m\left(x\right)=k\mathrm{e}^{x}$

由$m\ (0) = 1$,得$k = 1$，这样 $m\left(x\right)=\mathrm{e}^{x}$

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.

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$\mathrm{E}(X)=P(E)$
- 由X的定义可得，$X=\begin{cases}1 & 若E发生\\ 0 & 若E不发生\end{cases}$

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

因此利用式(4.1)与（4.2)可得

$$P(E)=\begin{cases}\sum_{y}P(E|Y=y)P(Y=y) & Y\text{为离散型随机变量}\\ \int_{-\infty}^{+\infty}P(E|Y=y)f_Y(y)\mathrm{d}y & Y\text{为连续型随机变量}\end{cases} \qquad (4.8)$$

- 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（4.8）变成

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

- 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9 最优奖问题 设有n个不同的奖陆续出台，当一个奖出来时，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个己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，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 $P_k(\{最优\})$ 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：

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

- 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

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

- 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 $i-1$ 个奖项的最大值奖的位置在前面 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：

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

- 这样我们得到

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

若考虑函数： $g(x) = \dfrac{x}{n}\ln\left(\dfrac{n}{x}\right)$

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

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

- 由于 $P_k(\text{最优}) \approx g(k)$, 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

  $1/\mathrm{e} \approx 0.36788$ ；

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 10 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：

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

又因为 $\int_0^1 p^i(1-p)^{n-i}dp=\dfrac{i!(n-i)!}{(n+1)!}$

由此可得：$P\{X=i\}=\dfrac{1}{n+1} \qquad i=0,\cdots,n$

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11 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

$$P\{X<Y\}=\int_{-\infty}^{\infty}P\{X<Y|Y=y\}f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}P\{X<y|Y=y\}f_Y(y)\mathrm{d}y$$

$$=\int_{-\infty}^{\infty}P\{X<y\}f_Y(y)\mathrm{d}y \qquad \text{由独立性}$$

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

$$F_X(y)=\int_{-\infty}^{y}f_X(x)\,dx$$

例12 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

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

$$=\int_{-\infty}^{\infty}P\{X<a-y\}f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}F_X(a-y)f_Y(y)\mathrm{d}y$$

4.3 方差

方差描述了r.v.对其数学期望的离散程度，在概率论和数理统计中十分重要.

➢ 定义

设X为一r.v., 若$\mathrm{E}\left[X-\mathrm{E}(X)\right]^2$ 存在, 则称它为 X 的方差, 记作 D(X)或Var(X) ,即

$$\mathrm{D}(X)=\mathrm{Var}(X)=\mathrm{E}\left[X-\mathrm{E}(X)\right]^2$$

称$\sqrt{\mathrm{D}(X)}$为X的均方差或标准差.

若 X 为离散型 $r.v.$ 其分布律为 $P\{X=x_k\}=p_k$, k=1,2,…, 则

$$\mathrm{D}(X)=\sum_{k}\left[x_k-\mathrm{E}(X)\right]^2\cdot p_k,$$

若X为连续型$r.v.$,其密度函数为$f(x)$, 则

$$\mathrm{D}(X)=\int_{-\infty}^{+\infty}[x-\mathrm{E}(X)]^2 f(x)\mathrm{d}x.$$

例1．甲、乙两人进行打靶，所射中环数分别记为 X_1、X_2, 它们的频律分别为:

X_1	8	9	10	X_2	8	9	10
p_k	0.3	0.1	0.6	p_k	0.2	0.5	0.3

试评定他们射击技术的好坏.（均值：甲9.3，乙9.1）

$$\mathrm{D}(X_1)=\mathrm{E}(X_1-\mathrm{E}(X_1))^2=0.3\times(8-9.3)^2+0.1\times(9-9.3)^2+0.6\times(10-9.3)^2=1.11$$

$$\mathrm{D}(X_2)=\mathrm{E}(X_2-\mathrm{E}(X_2))^2=0.2\times(8-9.1)^2+0.5\times(9-9.1)^2+0.3\times(10-9.1)^2=0.49$$

可见甲的技术不够“稳定”，乙方差小较“稳定”．

➢ 方差的计算公式:

$$D(X)=E(X-E(X))^2=E(X^2-2X\cdot E(X)+(E(X))^2)$$

$$=E(X^2)-[E(X)]^2$$

例2. 设r.v.X具有概率密度 $f(x)=\begin{cases}1+x, & -1\le x<0,\\ 1-x, & 0\le x<1,\\ 0, & otherwise.\end{cases}$ 求: $D(X)$.

解: $E(X)=\int_{-1}^{0} x(1+x)dx+\int_{0}^{1} x(1-x)dx=0,$

$$E(X^2)=\int_{-1}^{0} x^2(1+x)dx+\int_{0}^{1} x^2(1-x)dx=\frac{1}{6},$$

于是, $D(X)=E(X^2)-[E(X)]^2=\frac{1}{6}$.

➢ 下面计算一些常见分布的方差

1. 设随机变量 X 具有 (0-1) 分布，其分布律为 $P\{X=0\}=1-p$, $P\{X=1\}=p$, 则

$$E(X)=0\bullet(1-p)+1\bullet p=p$$

$$E(X^2)=0^2\bullet(1-p)+1^2\bullet p=p$$

故 $D(X)=E(X^2)-[E(X)]^2=p-p^2=p(1-p)$.

2. 二项分布: 设$X \sim b(n,p)$, $P\{X=k\}=C_n^k\, p^k q^{n-k}, k=0, 1, \cdots n.$

$$\mathrm{E}(X^2)=\sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k}=\sum_{k=2}^{n} k(k-1) C_n^k p^k q^{n-k}+\sum_{k=1}^{n} k C_n^k p^k q^{n-k}$$

$$=\sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k q^{n-k}+\mathrm{E}(X) \quad (\text{令} k'=k-2)$$

$$=n(n-1) p^2 \sum_{k'=0}^{n-2} \frac{(n-2)!}{k'!(n-2-k')!} p^{k'} q^{n-2-k'}+\mathrm{E}(X)$$

$$=n(n-1) p^2+\mathrm{E}(X)=n(n-1) p^2+np$$

故$\mathrm{D}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2=npq.$

3. 泊松分布: 设$X\sim\pi(\lambda)$, 即 $P\{X=k\}=\dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, k=0,1,2,\cdots\ \lambda>0.$

解： $$\mathrm{E}(X)=\sum_{k=0}^{+\infty}k\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\mathrm{e}^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}\cdot\lambda=\lambda.$$

$$\mathrm{E}(X^2)=\sum_{k=0}^{+\infty}k^2\cdot\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}=\sum_{k=1}^{+\infty}(k-1+1)\cdot\frac{\lambda^k}{(k-1)!}\mathrm{e}^{-\lambda}$$

$$=\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}\cdot\lambda^2}{(k-2)!}\mathrm{e}^{-\lambda}+\sum_{k=1}^{+\infty}\frac{\lambda^k}{(k-1)!}\mathrm{e}^{-\lambda}=\lambda^2+\lambda,$$

$$\therefore \mathrm{D}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2=\lambda.$$

1. 均匀分布: 设r.v. X服从区间$[a, b]$ 上的均匀分布, 即X 的密度函数为:

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a \le x \le b, \\ 0, & otherwise.\end{cases}$$

$$\mathrm{E}(X)=\int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \int_a^b x\cdot\frac{1}{b-a}\mathrm{d}x=\frac{a+b}{2},$$

$$\mathrm{E}(X^2)=\int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_a^b x^2\cdot\frac{1}{b-a}\mathrm{d}x=\frac{b^3-a^3}{3(b-a)}=\tfrac{1}{3}(b^2+ab+a^2),$$

$$\therefore \mathrm{D}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2=\frac{(b-a)^2}{12}.$$

2. 指数分布: 设r.v.X服从参数为λ的指数分布, 则其密度函数为: $f(x)=\begin{cases}\lambda e^{-\lambda x}, x\geq 0,\\ 0, \quad x<0.\end{cases}$ $\Gamma(p)=\int_0^{+\infty}x^{p-1}e^{-x}dx$ $\Gamma(n)=(n-1)!$

解: $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_0^{+\infty}x\cdot\lambda e^{-\lambda x}dx=\frac{1}{\lambda}.$

$$E(X^2)=\int_0^{+\infty}x^2\lambda e^{-\lambda x}dx\overset{t=\lambda x}{=}\frac{1}{\lambda^2}\int_0^{+\infty}t^2\cdot e^{-t}dt=\frac{2}{\lambda^2},$$

$$\therefore D(X)=E(X^2)-[E(X)]^2=\frac{1}{\lambda^2}$$

3. 正态分布: 设$X\sim N(\mu,\sigma^2)$, 则

$$D(X)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^2\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\ (令t=x-\mu)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}t^2\cdot e^{-\frac{t^2}{2\sigma^2}}dt\quad (令u=\frac{t^2}{2\sigma^2})$$

$$=\frac{2}{\sqrt{2\pi}\sigma}\int_{0}^{+\infty}2\sigma^2u\cdot e^{-u}\cdot\sigma\sqrt{2}\cdot\frac{1}{2}u^{-\frac{1}{2}}du$$

$t=\sigma\sqrt{2u}$

$dt=\sigma\sqrt{2}\cdot\frac{1}{2}\cdot u^{-\frac{1}{2}}du$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\int_{0}^{+\infty}u^{\frac{1}{2}}e^{-u}du\quad=\frac{2\sigma^2}{\sqrt{\pi}}\cdot\Gamma\left(\frac{3}{2}\right)$$

$$=\frac{2\sigma^2}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{2}=\sigma^2.$$

例 3（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 $\mathrm{Var}(N)=\mathrm{E}(N^2)-(\mathrm{E}(N))^2$，只需计算 $E[N^2]$.

在给定 Y 的条件下，有 $\mathrm{E}(N^2)=\mathrm{E}(\mathrm{E}(N^2 \mid Y))$

然而，$\mathrm{E}(N^2 \mid Y=1)=1$, $\mathrm{E}(N^2 \mid Y=0)=\mathrm{E}((1+N)^2)$

- 上述两式成立是因为：
  - ✓ 一方面，若第一次实验成功，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，我们得到 $\mathrm{E}(N^2 \mid Y=0)=\mathrm{E}((1+N)^2)$；
- 因此有 $$\begin{aligned}\mathrm{E}(N^2) &= \mathrm{E}(N^2 \mid Y=1)P\{Y=1\}+\mathrm{E}(N^2 \mid Y=0)P\{Y=0\} \\ &= p+(1-p)\mathrm{E}((1+N)^2) \\ &= 1+(1-p)\mathrm{E}(2N+N^2)\end{aligned}$$

➢ 然而，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 $\mathrm{E}(N^2)=\dfrac{2-p}{p^2}$

从而 $Var(N)=E(N^2)-(E(N))^2=\frac{2-p}{p^2}-\left(\frac{1}{p}\right)^2=\frac{1-p}{p^2}$

例4. X有密度函数$f(x)=\begin{cases}\dfrac{1}{x\sigma\sqrt{2\pi}}\exp\{-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\}, x>0\\ 0, x\le 0\end{cases}$

求$\mathrm{E}(X),\mathrm{D}(X)$.

解：$\mathrm{E}(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x$

$$=\int_0^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\}\mathrm{d}x$$ 令$y=\ln x$，$\mathrm{d}x=\mathrm{e}^y\mathrm{d}y$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}\cdot\mathrm{e}^y\mathrm{d}y$$

$$=\mathrm{e}^{\mu+\frac{\sigma^2}{2}}\cdot\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(y-\mu-\sigma^2)^2}{2\sigma^2}\}\cdot\mathrm{d}y$$

$$=\mathrm{e}^{\mu+\frac{\sigma^2}{2}}$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2 f(x)\mathrm{d}x$$ 令$y=\ln x$，$x=\mathrm{e}^y$，$\mathrm{d}x=\mathrm{e}^y\mathrm{d}y$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}\cdot\mathrm{e}^{2y}\mathrm{d}y$$

$$=e^{2(\mu+\sigma^2)}\int_{-\infty}^{\infty}\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(y-\mu-2\sigma^2)^2}{2\sigma^2}\}dy$$

$$=\mathrm{e}^{2(\mu+\sigma^2)}$$

$$D(X)=E(X^2)-[E(X)]^2=\mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2}-1).$$

➢ 方差的性质:

1. 设 C 是常数，则 $D(C)=0$;

2. 设 X 是 r.v.，C 是常数，则有 $D(CX) = C^2D(X)$;

3. 设 X, Y 是两个相互独立的随机变量，则有

$D(X \pm Y) = D(X)+D(Y)$;

4. $D(X) = 0$ 的充要条件是 X 以概率 1 取常数 C,

即 $P\{X=C\}=1$. 这里 $C=E(X)$

➢ 性质3的证明：

$$D(X \pm Y) = E((X \pm Y) - E(X \pm Y))^2$$

$$= E[(X - EX)^2 \pm 2(X - EX)(Y - EY) + (Y - EY)^2]$$

$$= D(X) + D(Y) \pm 2E(X - E(X))(Y - E(Y))$$

当X,Y独立时，$E(X - E(X))(Y - E(Y))$

$$= E(XY - X \cdot E(Y) - E(X) \cdot Y + E(X)E(Y))$$

$$= E(XY) - E(X)E(Y) = 0$$

命题得证。

推论：$X_1, X_2, \cdots, X_n$相互独立，则

$$D(X_1 + X_2 + \cdots + X_n) = D(X_1) + D(X_2) + \cdots + D(X_n)$$

$$D(c_1X_1 + c_2X_2 + \cdots + c_nX_n) = c_1^2D(X_1) + \cdots + c_n^2D(X_n)$$

➢ 切比雪夫(Chebyshev)不等式:

设r.v.X具有数学期望$\mathrm{E}(X)=\mu$, 方差$\mathrm{D}(X)=\sigma^2$, 则对$\forall\varepsilon>0$,

有不等式 $P\{|X-\mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$ 或$P\{|X-\mu| < \varepsilon\} \geq 1-\frac{\sigma^2}{\varepsilon^2}$

这一不等式称为Chebyshev不等式.

证：设X的概率密度为$f(x)$,则 $P\{|X-\mu| \geq \varepsilon\} = \int_{|x-\mu|\geq\varepsilon} f(x)\mathrm{d}x$

$$\leq \int_{|x-\mu|\geq\varepsilon} \left(\frac{|x-\mu|^2}{\varepsilon^2}\right) f(x)\mathrm{d}x \leq \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(x-\mu)^2 f(x)\mathrm{d}x = \frac{\sigma^2}{\varepsilon^2}$$

如取 $\varepsilon=2\sigma$ 或 3σ 可得 $P\{|X-\mu|<2\sigma\}\geq 0.7500,$

$$P\{|X-\mu|<3\sigma\}\geq 0.8889$$

这个估计的精度不高，但具有普遍适用性（随机变量的分布）。

$X\sim N(\mu,\sigma^2)$时 $P\{|X-\mu|<2\sigma\}=0.9544,$

$$P\{|X-\mu|<3\sigma\}=0.9974$$

性质4的证明:

4. $\mathrm{D}(X)=0$ 的充要条件是 X 以概率 1 取常数C, 即 $P\{X=C\}=1.$

这里 $C=\mathrm{E}(X)$

充分性证明: $P\{X=C\}=1$，则$P\{X^2=C^2\}=1$，即$\mathrm{E}(X^2)=C^2$；

$\mathrm{D}(X)=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2=0$；

必要性证明：假设 $\mathrm{D}(X)=0$， $P\{X=C\}<1$；

则对 $\forall \varepsilon>0, P\left\{\left|X-C\right|\geq\varepsilon\right\}>0$

由切比雪夫不等式, $P\left\{\left|X-\mathrm{E}(X)\right|\geq\varepsilon\right\}\leq\dfrac{\mathrm{D}(X)}{\varepsilon^2}=0$

矛盾，假设不成立，因此 $P\{X=C\}=1$；

例1. 设 $X \sim b(n, p)$, 分解 X, 求其方差 $\mathrm{D}(X)$.

解：$X = \sum_{i=1}^{n} X_i$,其中

$$X_i = \begin{cases} 0, & \text{第}i\text{次试验}A\text{不发生} \\ 1, & \text{第}i\text{次试验}A\text{发生} \end{cases} \quad (i = 1, 2, \cdots, n)$$

易知 $X_1, X_2, \ldots, X_n$ 独立同服从(0--1)分布，因此

$$\mathrm{D}(X) = \mathrm{D}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{D}(X_i)$$

$$= pq + pq + \cdots + pq = npq$$

例2．证明：若$X_i \sim N(\mu_i,\sigma_i^2)(i=1,2,\cdots,n)$，且相互独立，$k_1,k_2,\cdots,k_n$不全为零，则

$$\sum_{i=1}^{n} k_i X_i \sim N\left(\sum_{i=1}^{n} k_i \mu_i, \sum_{i=1}^{n} k_i^2 \sigma_i^2\right)$$

.

证明：有限个独立的正态随机变量之线性组合仍服从正态分布，

可知$\sum_{i=1}^{n} k_i X_i$服从正态分布。再由期望与方差的性质可得

$$\mathrm{E}\left(\sum_{i=1}^{n} k_i X_i\right) = \sum_{i=1}^{n} \mathrm{E}(k_i X_i) = \sum_{i=1}^{n} k_i \mathrm{E}(X_i) = \sum_{i=1}^{n} k_i \mu_i$$

$$\mathrm{D}\left(\sum_{i=1}^{n} k_i X_i\right) = \sum_{i=1}^{n} \mathrm{D}(k_i X_i) = \sum_{i=1}^{n} k_i^2 \mathrm{D}(X_i) = \sum_{i=1}^{n} k_i^2 \sigma_i^2$$

所以结论成立.

例3. 对任一随机变量X，若其期望 $\mathrm{E}(X)$,方差 $\mathrm{D}(X)$ 均存在，且$\mathrm{D}(X)>0$，则称 $X^*=\dfrac{X-\mathrm{E}(X)}{\sqrt{\mathrm{D}(X)}}$为$X$的标准化随机变量。

试证：$\mathrm{E}(X^*)=0,\ \mathrm{D}(X^*)=1.$

证： $\mathrm{E}(X^*)=\mathrm{E}[\dfrac{X-\mathrm{E}(X)}{\sqrt{\mathrm{D}(X)}}]=\dfrac{1}{\sqrt{\mathrm{D}(X)}}\mathrm{E}[X-\mathrm{E}(X)]=0$

$$\mathrm{D}(X^*)=\mathrm{E}(X^{*2})-[\mathrm{E}(X^*)]^2=\mathrm{E}[\frac{X-\mathrm{E}(X)}{\sqrt{\mathrm{D}(X)}}]^2$$

$$=\frac{1}{\mathrm{D}(X)}\mathrm{E}(X-\mathrm{E}(X))^2$$

$$=\frac{1}{\mathrm{D}(X)}\cdot\mathrm{D}(X)=1$$

例6a 设随机变量 X 服从参数为 α 和 λ 的 Γ 分布，试计算

(a) $E[X]$; (b) $Var(X)$

$$f(x)=\begin{cases}\dfrac{\lambda^p}{\Gamma(p)}x^{p-1}e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

解 (a) 注意到 $\Gamma(\alpha)=(\alpha-1)\int_0^{\infty}e^{-y}y^{\alpha-2}dy$

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

(b) 首先计算$E[X^2]$, 再由方差计算公式可得

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

4.4 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：

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

- 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，$\mathrm{Var}(X|Y)$ 与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

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

$$= \mathrm{E}(X^2) - \mathrm{E}[(\mathrm{E}(X|Y))^2] \quad (4.9)$$

同时

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

$$= \mathrm{E}[(\mathrm{E}(X|Y))^2] - (\mathrm{E}(X))^2 \quad (4.10)$$

将（4.9）与式（4.10）相加，得到如下命题.

命题 [条件方差公式] $\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathrm{E}(X|Y))$

例1 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 $t\geq 0$，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

$$\mathrm{E}(N(Y)|Y=t)=\mathrm{E}(N(t)|Y=t)=\mathrm{E}(N(t))$$ 由Y与$N(t)$的独立性

$$=\lambda t$$ $N(t)$是均值为 λt 的泊松随机变量

因此，$\mathrm{E}(N(Y)|Y)=\lambda Y$

两边取期望可得：$\mathrm{E}(N(Y))=\lambda \mathrm{E}(Y)=\dfrac{\lambda T}{2}$

➢ 为了计算 $\mathrm{Var}(N(Y))$，我们利用条件方差公式

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

➢ 因此，有：

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

➢ 再由条件方差公式，得：

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

➢ 上式利用了 $\mathrm{Var}(Y)=T^2/12$ 的事实

例2 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 $X_i,\ i \geq 1$，计算 $\mathrm{Var}(\sum_{i=1}^{N} X_i)$，

- 先固定 N 的值作为条件

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

- 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

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

命题 [条件方差公式] $\mathrm{Var}(X) = \mathsf{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathsf{E}(X|Y))$

4.5 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的值进行预测。
- 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 $g(X)$接近 Y.选择 g 的一个准则是极小化 $E((Y-g(X))^2)$.下面我们指出在这个准则之下，Y 的最好的预测值为 $g(X)=E(Y|X)$ ；

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

证明：$\mathrm{E}[(Y-g(X))^2|X] = \mathrm{E}[(Y-\mathrm{E}(Y|X)+\mathrm{E}(Y|X)-g(X))^2|X]$

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

$$+ 2\mathrm{E}[(Y-\mathrm{E}(Y|X))(\underline{\mathrm{E}(Y|X)-g(X)})|X]$$

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

于是： $\mathrm{E}[(Y-\mathrm{E}(Y|X))(\mathrm{E}(Y|X)-g(X))|X]$

$$= (\mathrm{E}[Y|X]-g(X))\mathrm{E}[Y-\mathrm{E}[Y|X]|X]$$
$$= (\mathrm{E}(Y|X)-g(X))(\mathrm{E}(Y|X)-\mathrm{E}(Y|X)) = 0$$

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

上式两边再求期望即可得到命题的结论。

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

- **注释：**
  - ✓ 此处可以给出命题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.
  - ✓ 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，于是命题得证。

例1 设父亲的身高为 X 英寸，儿子的身高服从均值为 $X+1$ 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？（1英尺=12英寸）

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）

对于 6 英尺的父亲，其儿子身高的最优预测为 $\mathrm{E}(Y\,|\,\mathrm{X}=72)$，

$$\mathrm{E}(Y|X=72)=\mathrm{E}(X+1+e|X=72)$$

$$=73+\mathrm{E}(e|X=72) \qquad =73+\mathrm{E}(e)=73$$

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

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

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

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

- 其中 C 与 s无关，由上式可知，在给定 $R=r$ 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, 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:

$$E(S|R=r)=\frac{1}{1+\sigma^2}\mu+\frac{\sigma^2}{1+\sigma^2}r$$

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（接收到信号的期望值）的加权平均。而两个权值之比为$1:\sigma^2$，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。

4.6 协方差和相关系数

定义1: 设(X, Y)为二维r.v.，若 $E([X-E(X)][Y-E(Y)])$ 存在, 则把它称作 X和 Y 的协方差, 记作$Cov(X, Y)$, 即 :

$$Cov(X, Y) = E([X-E(X)][Y-E(Y)]).$$

展开可得计算公式: $Cov(X, Y) = E(XY) - E(X)E(Y)$.

由方差性质证明知, 对于任意的两个 r.v. X 和Y, 有:

$$D(X \pm Y)=D(X)+D(Y) \pm 2Cov(X,Y).$$

➢ 协方差的性质:

1、$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$;

2、$\mathrm{Cov}(a_1X+b_1, a_2Y+b_2)=a_1a_2\mathrm{Cov}(X, Y)$, 其中$a_1, a_2, b_1, b_2$是常数;

3、$\mathrm{Cov}(X_1+X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$;

4、 $\mathrm{Cov}(X, a)=0$, $\mathrm{Cov}(X, X) = D(X)$, a为常数;

5、 若 X, Y 相互独立, 则 $\mathrm{Cov}(X, Y)=0$.

6、$|\mathrm{Cov}(X,Y)|^2 \le D(X)\cdot D(Y)$; “=”成立当且仅当 X 与 Y 之间有严格的线性关系。$(Y= aX + b)$

性质6证明： 对任$t \in R$都有 $\mathrm{E}[t(X-\mathrm{E}(X))+(Y-\mathrm{E}(Y))]^2 \geq 0$

展开为 $\mathrm{E}[t^2(X-\mathrm{E}(X))^2+2t(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))+(Y-\mathrm{E}(Y))^2] \geq 0$

即$t^2 \cdot \mathrm{D}(X)+2t\mathrm{Cov}(X,Y)+\mathrm{D}(Y) \geq 0$

此不等式对应的方程无实根或有二重根, 故有

$$\Delta = 4(\mathrm{Cov}(X,Y))^2 - 4\mathrm{D}(X)\mathrm{D}(Y) \leq 0,$$即命题成立。

“$=$” 成立时方程有重根t_0, 即 t_0 满足$t_0(X-\mathrm{E}(X))+(Y-\mathrm{E}(Y))=0.$
X与 Y有线性关系。（要使得该式子成立，须有$Y=-t_0X+b$）

定义2: 若$\mathrm{D}(X)\neq 0$, $\mathrm{D}(Y)\neq 0$, 则称$\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{D}(X)}\sqrt{D(Y)}}$为$X,Y$的相关系数，记为$\rho_{XY}$.

记 $X^*=\dfrac{X-\mathrm{E}(X)}{\sqrt{\mathrm{D}(X)}}$, $Y^*=\dfrac{Y-\mathrm{E}(Y)}{\sqrt{\mathrm{D}(Y)}}$为 X,Y 的标准化随机变量

$$\mathrm{Cov}(X^*,Y^*)=\mathrm{E}(X^*Y^*)-\mathrm{E}(X^*)\mathrm{E}(Y^*)$$

$$=\mathrm{E}\left((\frac{X-\mathrm{E}(X)}{\sqrt{\mathrm{D}(X)}})\cdot(\frac{Y-\mathrm{E}(Y)}{\sqrt{\mathrm{D}(Y)}})\right)$$

$$=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{D}(X)\mathrm{D}(Y)}}=\rho_{XY}$$

显然，相关系数是标准化了的协方差。

➢相关系数的性质:

1 $|\rho_{XY}| \le 1$;.

证：因为$\mathrm{D}(X^* \pm Y^*) = \mathrm{D}(X^*) + \mathrm{D}(Y^*) \pm 2\mathrm{Cov}(X^*, Y^*)$

$$=1 + 1 \pm 2\rho_{XY} \ge 0$$

所以有 $|\rho_{XY}| \le 1$.

2 $|\rho_{XY}| = 1 \Leftrightarrow X$ 和Y以概率1线性相关, 即$P\{Y = aX + b\} = 1$,

其中a, b为常数, 且$a \ne 0$ (由协方差的性质6)

3 若X, Y相互独立，则$\rho_{XY} = 0$.

（若 X, Y 相互独立, 则 Cov (X, Y)=0）

定义3： 若 $\rho_{XY}=0$, 则称 X 和 Y 是不相关的。

> ➢ **注意**：相关系数 ρ_{XY} 刻划了 X, Y 之间的线性相关关系, 当 $\rho_{XY}=0$ 时，称 X, Y 不相关是指它们之间没有线性相关关系. $\rho_{XY}=1$ 或 -1时，X 与 Y 有严格线性关系。

不相关与相互独立的逻辑关系:

a. 若X,Y相互独立，则X, Y不相关（ρ=0);

b. 上面的逆命题一般不真;

$$\left(\begin{array}{l}\text{反例, 二维r.v.}(X,\ Y)\text{的密度函数是} \\ \qquad f(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\le 1, \\ 0, & \text{其它}, \end{cases} \\ \text{其}\rho_{XY}=0,\text{但}f(x,y)\ne f_X(x)f_Y(y).\end{array}\right).$$

c. 当 (X,Y) 服从二维正态分布时，X,Y独立 $\Leftrightarrow$ X,Y不相关

例1. 设(X, Y)服从二维正态分布, 求 X 和 Y 的相关系数.

解：前面在第三章的例子中已经知道 (X,Y) 的边缘概率密度为

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}},\ f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}},\ -\infty<x,y<+\infty,$$

故知 $E(X)=\mu_1,\ E(Y)=\mu_2,\ D(X)=\sigma_1^2,\ D(Y)=\sigma_2^2.$

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}$$

$$-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$

$$-\infty<x<+\infty,-\infty<y<+\infty$$

$$\text{而}\operatorname{Cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_1)(y-\mu_2)f(x,y)\,\mathrm{d}x\mathrm{d}y$$

$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\sigma_1\sigma_2\sqrt{1-\rho^2}\,tu+\rho\sigma_1\sigma_2u^2\right)\mathrm{e}^{-\frac{u^2}{2}-\frac{t^2}{2}}\mathrm{d}u\mathrm{d}t$$

$$\left(\text{其中}\ t=\frac{1}{\sqrt{1-\rho^2}}\left(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}\right),u=\frac{x-\mu_1}{\sigma_1}\right)\ \text{故}$$

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$

$$-\infty<x<+\infty,-\infty<y<+\infty$$

$$\mathrm{Cov}\,(X,Y)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\sigma_1\sigma_2\sqrt{1-\rho^2}\,tu+\rho\sigma_1\sigma_2u^2\right)\mathrm{e}^{-\frac{u^2}{2}-\frac{t^2}{2}}\mathrm{d}u\mathrm{d}t$$

$$\mathrm{Cov}(X,Y)=\frac{\rho\sigma_1\sigma_2}{2\pi}\underline{\left(\int_{-\infty}^{+\infty}u^2\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\right)}\left(\int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)+ \qquad \sqrt{2}\Gamma\left(\frac{1}{2}\right)$$

$$\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}\underline{\left(\int_{-\infty}^{+\infty}u\mathrm{e}^{-\frac{u^2}{2}}\mathrm{d}u\right)\left(\int_{-\infty}^{+\infty}t\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\right)}$$

奇函数！

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\cdot\sqrt{2\pi}\cdot\sqrt{2\pi}=\rho\sigma_1\sigma_2.$$

$$\therefore \rho_{XY}=\rho.$$

- 由第三章我们曾证明过的一个命题, 设 (X, Y) 服从二维正态分布, 则 X, Y 相互独立的充要条件是 $\rho=0$. 知 X 与 Y 不相关与 X 和 Y 相互独立是等价的.

公式：$\mathrm{Cov}(aX+bY, cX+dY) = ac\mathrm{D}(X) + (ad+bc)\mathrm{Cov}(X, Y) + bd\mathrm{D}(Y)$

例2. $X \sim N(2003, 1)$, $Y \sim N(2004, 1)$, 且 X 与 Y 独立，求 $3X-Y$ 与 $X+Y$ 的相关系数。

解：由于 X, Y 独立，则

$$\mathrm{Cov}(3X-Y, X+Y) = 3\mathrm{D}(X)+2\mathrm{Cov}(X,Y)-\mathrm{D}(Y)=3-1=2$$

$$\mathrm{D}(3X-Y)=9\mathrm{D}(X)+\mathrm{D}(Y)=10$$

$$\mathrm{D}(X+Y)=\mathrm{D}(X)+\mathrm{D}(Y)=2$$

相关系数为

$$\rho = \frac{\mathrm{Cov}(3X-Y, X+Y)}{\sqrt{\mathrm{D}(3X-Y)\mathrm{D}(X+Y)}} = \frac{2}{\sqrt{2\times 10}} = \frac{\sqrt{5}}{5}$$

- **例3** 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，$i = 0, \pm 1, \pm 2, \cdots$，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，X 与 Y 之间有如下的关系:

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

  - 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

$$y_i,\ i=0,\ \pm 1,\ \pm 2,\ \cdots$$

使得 $\mathrm{E}((X-Y)^2)$达到极小。

(a)找到最优值 $y_i, i=0,\pm1,\pm2,\cdots$

对于最优Y，证明：

(b) $\mathrm{E}[Y]=\mathrm{E}[X]$ 即均方误差最小意义下的离散化保持均值不变。

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

解：(a) 对于任意的离散化随机变量Y，在给定Y值的条件下

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

如果我们令， $a_i < X \le a_{i+1}$为 $I=i$

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

利用4.5节命题的结论，当

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

$\mathrm{E}((X - y_i)^2 \mid a_i < X \le a_{i+1})$ 达到极小值。

$Y = \mathrm{E}(X|I)$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：

$$(\mathrm{b})\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(X \mid I)) = \mathrm{E}(X)$$

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

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

- 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，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

$$\mathrm{E}[(Y-(a+bX))^2]$$

$$=\mathrm{E}[Y^2-2aY-2bXY+a^2+2abX+b^2X^2]$$

$$=\mathrm{E}(Y^2)-2a\mathrm{E}(Y)-2b\mathrm{E}(XY)+a^2+2ab\mathrm{E}(X)+b^2\mathrm{E}(X^2)]$$

求上式对a和b的偏导数，得到

$$\begin{aligned}
&\frac{\partial}{\partial a}\mathrm{E}[(Y-a-bX)^2]=-2\mathrm{E}(Y)+2a+2b\mathrm{E}(X) \\
&\frac{\partial}{\partial b}\mathrm{E}[(Y-a-bX)^2]=-2\mathrm{E}(XY)+2a\mathrm{E}(X)+2b\mathrm{E}(X^2)
\end{aligned} \qquad (3)$$

令偏导数为0，求解关于(a,b)的方程组(3)，得到

$$\begin{aligned}
&b=\frac{\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)}{\mathrm{E}(X^2)-(\mathrm{E}(X))^2}=\frac{\mathrm{Cov}(X,Y)}{\sigma_x^2}=\rho\frac{\sigma_y}{\sigma_x} \\
&a=\mathrm{E}(Y)-b\mathrm{E}(X)=\mathrm{E}(Y)-\frac{\rho\sigma_y\mathrm{E}(X)}{\sigma_x}
\end{aligned} \qquad (4)$$

- 其中ρ为 X,Y 的相关系数，$\sigma_x^2=\mathrm{Var}(X)$　$\sigma_y^2=\mathrm{Var}(Y)$
- 容易验证由 (4) 给出的 a, b 值使得 $\mathrm{E}[(Y-(a+bX))^2]$达到极小；
- 因此，在均方误差意义下，Y的最优线性预测为，

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

其中，μ_y=E(Y), μ_x=E(X).

➢ 这个线性预测的均方误差为

$$\mathrm{E}[(Y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(X-\mu_x))^2] \tag{5}$$

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

$$=\sigma_y^2+\rho^2\sigma_y^2-2\rho^2\sigma_y^2=\sigma_y^2(1-\rho^2)$$

➢ 由(5)看出，当 ρ 接近于+1或-1时，其最优线性预测的均方误差接近于0.

例4 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 例3 已经给出，在正态情况下，

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

4.7 矩、协方差矩阵

一. 定义: 设 X 和 Y 是随机变量,

(1) 若$E(X^k)$, $k=1, 2, \ldots$存在, 则称它为 X 的 k 阶原点矩.

(2) 若$E[[X-E(X)]^k]$, $k=1, 2, \ldots$存在, 则称它为 X 的 k 阶中心矩.

(3) 若$E(X^k \bullet Y^l)$, $k, l=1, 2, \ldots$存在, 则称它为 X 和 Y 的 $k+l$ 阶混合矩.

(4) 若$E[[X-E(X)]^k \bullet [Y-E(Y)]^l]$, $k, l=1, 2, \ldots$存在, 则称它为 X 和 Y 的 $k+l$ 阶混合中心矩.

- 显然, $E(X)$, $E(Y)$为一阶原点矩, $D(X)$, $D(Y)$为二阶中心矩, $Cov(X,Y)$为二阶混合中心矩.

二、定义: 二维随机变量(X_1,X_2) 有四个二阶中心矩，分别记为

$$c_{11}=\mathrm{E}[[X_1-\mathrm{E}(X_1)]^2]$$
$$c_{12}=\mathrm{E}[[X_1-\mathrm{E}(X_1)][X_2-\mathrm{E}(X_2)]]$$
$$c_{21}=\mathrm{E}[[X_2-\mathrm{E}(X_2)][X_1-\mathrm{E}(X_1)]]$$
$$c_{22}=\mathrm{E}[[X_2-\mathrm{E}(X_2)]^2]$$

将它们排成矩阵形式 $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$ 称这个矩阵为 (X_1, X_2) 的 协方差矩阵。

设n维随机变量$(X_1, X_2, \cdots, X_n)$的二阶中心矩及二阶混合中心矩 $c_{ij} = \mathrm{Cov}(X_i, X_j), \quad i, j = 1, 2, \cdots, n,$ 都存在, 则称矩阵

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为n维随机变量$(X_1, X_2, \cdots, X_n)$的协方差阵.

- 由于$c_{ij}=c_{ji}$ $(i, j=1,2, \ldots, n)$，因此协方差矩阵是一个对称矩阵；

(由协方差的性质 $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$, $c_{ij}= c_{ji}$可得)

三. 协方差阵的性质:

1 C是对称的;

2 $c_{ii}=\mathrm{D}(X_i)$, $i=1, 2, 3, \ldots, n$.

3 $c_{ij}^2 \le c_{ii}c_{jj}$, i、$j=1, 2, \ldots, n$. (由协方差的性质$|\mathrm{Cov}(X, Y)|^2 \le \mathrm{D}(X)\mathrm{D}(Y)$)

4 C是非负定的, 即对任意的向量 $a=(a_1, a_2, \ldots, a_n)^T$, 都有 $a^TCa \ge 0$.

四. n 维正态变量:

1. 定义: 设有n维r.v.$(X_1, X_2, \cdots, X_n)$, 称以 $f(x_1, x_2, \cdots, x_n)$为密度函数的 $(X_1, X_2, \cdots, X_n)$ 作n维正态变量, 记作 $(X_1, X_2, \cdots, X_n) \sim N(\boldsymbol{\mu}, \mathbf{C})$.

其中$\boldsymbol{\mu}$为n维常向量, $\mathbf{C}$是n维对称正定矩阵,

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix},$$

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu)^T \mathbf{C}^{-1} (X-\mu)}$$

2. 性质:

1、n 维正态变量 $(X_1, X_2, \ldots, X_n)$ 的每一个分量 X_i 都是正态变量；反之，*若* $X_1, X_2, \ldots, X_n$ 都是正态变量，且相互独立，则$(X_1, X_2, \ldots, X_n)$ 是 n 维正态变量。

2、n 维 *r.v.* $(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布的充要条件是 $X_1, X_2, \ldots, X_n$ 的任一线性组合 $l_1X_1+l_2X_2+\ldots+l_n X_n$ 服从一维正态分布.

3、若 $(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布, 设 $Y_1, Y_2, \ldots, Y_n$ 是 X_j (j=1, 2, …, n) 的线性函数, 则 $(Y_1, Y_2, \ldots Y_n)$ 也服从多维正态分布.

4、若$(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布, 则 "$X_1, X_2, \ldots, X_n$"相互独立与 "$X_1, X_2, \ldots, X_n$"两两不相关是等价的.

练习

1. 有5个相互独立工作的电子装置, 它们的寿命 $X_k(k=1, \cdots, 5)$服从同一指数分布, 其概率密度为:

$$f(x)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, x>0, \\ 0, \quad x\le 0,\end{cases} \quad \theta>0,$$

(i)若将这5个电子装置串联工作组成整机, 求整机寿命N 的数学期望.

(ii)若将这5个电子装置并联工作组成整机, 求整机寿命M 的数学期望.

2 设随机变量(X,Y)的联合密度为$f(x,y)$，求EX,EY

$$f(x,y)=\begin{cases}1,|y|<x,0<x<1\\0,\text{其它}\end{cases}.$$

3. 将 n 个编号为1-n 的 n 个球随机放入 m 个盒子中去(盒子容量不限)，X 表示有球的盒子数，求 EX

4. 已知$X\sim N(\mu,\sigma^2)$,求$E|X-\mu|$.

5. 当 X 服从离散分布时，证明切比雪夫不等式。

6.已知$f(x,y)=\begin{cases}\frac{1}{8}(x+y),0\le x\le 2,0\le y\le 1\\0,\text{其它}\end{cases}$，求$\rho_{XY}$