4.2 方差

方差 $DX := E(X-EX)^2 = EX^2-(EX)^2$.

标准差 $\sqrt{DX}$. （与 X 量纲相同）.

方差刻画了 r.v. 的集中或分散程度. r.v. 的方差越大, 取值越分散.

离散型 r.v. $P(X=x_i)=p_i$, 则

$$DX=\sum_i (x_i-EX)^2 p_i=\sum_i x_i^2 p_i-(EX)^2$$

连续型 r.v. $f(x)$. 则

$$DX=\int_{-\infty}^{+\infty}(x-EX)^2 f(x)\,dx=\int_{-\infty}^{+\infty}x^2 f(x)\,dx-(EX)^2$$

例 设 $X\sim B(n,p)$. 则 $DX=npq$, $q=1-p$.

证: $EX=np$.

$$EX^2=\sum_{k=0}^{n}k^2 C_k^n p^k q^{n-k}$$

$$\because\ k^2C_k^n=\frac{k^2\,n!}{k!\,(n-k)!}=\frac{k\,n!}{(k-1)!\,(n-k)!}$$

$$=\frac{n!}{(k-2)!\,(n-k)!}+\frac{n!}{(k-1)!\,(n-k)!}$$

$$=n(n-1)C_{n-2}^{k-2}+nC_{n-1}^{k-1}$$

$$\therefore\ EX^2=n(n-1)\sum_{k=2}^{n}C_{n-2}^{k-2}p^kq^{n-k}+n\sum_{k=1}^{n}C_{n-1}^{k-1}p^kq^{n-k}$$

$$=n(n-1)p^2+np$$

$$\therefore\ DX=EX^2-(EX)^2=np-np^2=npq$$ ▯

例 设 $X \sim P(\lambda)$. 则 $DX = \lambda$.

证 $EX = \lambda$.

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

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

$\therefore$ $DX = EX^2 - (EX)^2 = \lambda$. $\square$

例 几何分布. $P(X=k) = p\,q^{k-1}$, $k \geq 1$. 则 $DX = \frac{q}{p^2}$.
$q := 1-p$.

证: $EX = \frac{1}{p}$.

$$EX^2 = \sum_{k=1}^{+\infty} k^2 p\, q^{k-1} = p\sum_{k=1}^{+\infty} \frac{d}{dq}(kq^k)$$

$$= p\frac{d}{dq}\left[\sum_{k=1}^{+\infty} k\, q^{k-1}(1-q) \cdot \frac{q}{1-q}\right]$$

$$= p \cdot \frac{d}{dq} \frac{q}{(1-q)^2}$$

$$= \frac{1+q}{p^2}$$

$\therefore$ $DX = EX^2 - (EX)^2 = \frac{q}{p^2}$.

例. 设 X 服从超几何分布. 求 DX.

解. $P(X=k) = \dfrac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, $0 \leq k \leq n$.

已知 $EX = \frac{M}{N} \cdot n$. 下面计算 EX^2.

$$EX^2 = \sum_{k=0}^{n} \frac{k^2 C_M^k C_{N-M}^{n-k}}{C_N^n}$$

$\because$ $k C_M^k = M C_{M-1}^{k-1}$

$$\because \quad EX^2 = \frac{M}{C_N^n}\sum_{k=1}^{n} k\, C_{M-1}^{k-1} C_{N-M}^{n-k}$$

$$= \frac{M}{C_N^n}\sum_{k=2}^{n} (k-1) C_{M-1}^{k-1} C_{N-M}^{n-k} + \frac{M}{C_N^n}\sum_{k=1}^{n} C_{M-1}^{k-1} C_{N-M}^{n-k}$$

$$= \frac{M(M-1)}{C_N^n}\sum_{k=2}^{n} C_{M-2}^{k-2} C_{N-M}^{n-k} + \frac{M}{C_N^n}\sum_{k=1}^{n} C_{M-1}^{k-1} C_{N-M}^{n-k}$$

又 $\because \quad \sum_{k=0}^{n} C_{N-M}^{n-k} C_M^k = C_N^n$

$$\therefore \quad EX^2 = \frac{M(M-1)\, C_{N-2}^{n-2}}{C_N^n} + \frac{M C_{N-1}^{n-1}}{C_N^n}$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N}$$

$$\therefore \quad DX = EX^2 - (EX)^2$$

$$= \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{Mn}{N} - \left(\frac{Mn}{N}\right)^2$$

$$= \frac{Mn}{N}\,\frac{N-n}{N-1}\left(1-\frac{M}{N}\right). \qquad \square$$

例　设 $X \sim U(a,b)$，则 $DX = \frac{(b-a)^2}{12}$.

证：已知　$EX = \frac{b+a}{2}$.

由定义，

$$EX^2 = \int_a^b \frac{x^2}{b-a}\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+a^2+ab}{3}$$

$$\therefore \quad DX = EX^2 - (EX)^2 = \frac{4(b^2+a^2+ab) - 3(a+b)^2}{12}$$

$$= \frac{(b-a)^2}{12}. \qquad \square$$

例. 设 $X \sim E(\lambda)$. 则 $DX = \frac{1}{\lambda^2}$.

证: $EX = \frac{1}{\lambda}$

$$EX^2 = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = -\int_0^{+\infty} x^2 d e^{-\lambda x}$$

$$= \int_0^{+\infty} 2x e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$\therefore DX = EX^2 - (EX)^2 = \frac{1}{\lambda^2}$$

$\square$

例. 设 $X \sim N(\mu, \sigma^2)$. 则 $DX = \sigma^2$.

证: $EX = \mu$.

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

$(y = \frac{x-\mu}{\sigma})$

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

$$= \sigma^2 \int_{-\infty}^{+\infty} \frac{-y}{\sqrt{2\pi}} d e^{-\frac{y^2}{2}}$$

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

$$= \sigma^2$$

$\square$

方差的性质

1) $\forall$ r.v. X, $D(X) \geq 0$. 且 $D(X)=0 \Leftrightarrow X = C = EX$.

2) $D(a+bX) = b^2 D(X)$.

3) 若 $X_1, \cdots, X_n$ 独立, 则

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

4) $D(X) = E(X-EX)^2 \leq E(X-C)^2, \quad \forall C$.

证明见板书.

<u>切比雪夫不等式.</u> $\forall \varepsilon>0$,

$$P(|X-EX|\geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

<u>标准化</u> 设 $EX=\mu$, $DX=\sigma^2$, 称 $Y=\frac{X-\mu}{\sigma}$ 为 X 的标准化.

显然, $EY=0$, $DY=1$.

<u>例</u>. 设 $X_1, \cdots, X_n$ 相互独立, $EX_i=\mu$, $DX_i=\sigma^2$, $\forall i$.

令 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$. 则 $E\bar{X}=\mu$, $D\bar{X}=\frac{1}{n}\sigma^2$.

实际应用: 多次测量取平均, 减少误差.

<u>例</u>. 设 $X, Y \sim N(0,\frac{1}{2})$ 且独立. 求 $E|X-Y|$, $D|X-Y|$.

<u>解</u>. $X-Y\sim N(0,1)$.

$$E|X-Y| = \int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

$$= 2\int_0^{+\infty}\frac{x}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\sqrt{\frac{2}{\pi}}$$

$$E|X-Y|^2 = 1$$

$$\therefore D|X-Y| = 1-\frac{2}{\pi}.$$ $\square$

<u>例</u>. 设 $X\in[a,b]$, 证

$$a\leq EX\leq b, \quad DX\leq\frac{(b-a)^2}{4}.$$

<u>证</u>. $a\leq EX\leq b$ 显然.

$$DX=E(X-EX)^2 = E\left(X-\frac{a+b}{2}\right)^2-\left(EX-\frac{a+b}{2}\right)^2$$

$$\leq E\left(X-\frac{a+b}{2}\right)^2\leq\frac{(b-a)^2}{4}.$$ $\square$

4.4 协方差和相关系数

协方差： $\text{Cov}(X,Y) := E[(X-EX)(Y-EY)]$

$$= EXY - EXEY.$$

X与Y不相关： $\text{Cov}(X,Y)=0.$

基本性质：

1) X与Y独立 $\Rightarrow$ X与Y不相关.

反之不对. 反例： $X\sim N(0,1)$, $Y=X^2$.

2) $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

3) $\text{Cov}(aX+bY, Z) = a\,\text{Cov}(X,Z) + b\,\text{Cov}(Y,Z)$.

4) $D(k_0 + k_1X_1 + \cdots + k_nX_n)$

$$= \sum_{i=1}^{n} k_i^2 DX_i + 2\sum_{i<j} k_ik_j \text{Cov}(X_i, X_j)$$

例 设 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 求 $\text{Cov}(X,Y)$.

解. $\text{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_1)(y-\mu_2) \times \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]\right\} dx\,dy$$

令 $u = \frac{x-\mu_1}{\sigma_1}$, $v = \frac{y-\mu_2}{\sigma_2}$ 则

$u^2+v^2-2\rho vu = (u-\rho v)^2 + (1-\rho^2)v^2$

$$\text{Cov}(X,Y) = \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} uv\, e^{-\frac{(u-\rho v)^2}{2(1-\rho^2)}} e^{-\frac{v^2}{2}}\, du\, dv$$

$$= \frac{\sigma_1\sigma_2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \rho v^2 e^{-\frac{v^2}{2}} dv = \rho\sigma_1\sigma_2$$

$\square$

例. N件产品中含M件次品，无放回取n件. 令X表示取得的次品数. 求DX.

解：令

$$X_i = \begin{cases} 1, & \text{第}\,i\,\text{次抽到次品} \\ 0, & \text{其它.} \end{cases}$$

则 $X = X_1 + X_2 + \cdots + X_n$.

$$\therefore\ DX = \sum_{i=1}^{n} DX_i + 2\sum_{i<j} Cov(X_i, X_j).$$

$$\because\ P(X_i = 1) = \frac{M}{N}$$

$$\therefore\ DX_1 = EX_i^2 - (EX_i)^2 = \frac{M}{N} - \left(\frac{M}{N}\right)^2 = \frac{M}{N}\left(1 - \frac{M}{N}\right)$$

$i<j$ 时，

$$EX_iX_j = P(X_i = X_j = 1) = P(X_j = 1 \mid X_i = 1)\,P(X_i = 1)$$

$$= \frac{M-1}{N-1} \cdot \frac{M}{N}$$

$$\therefore\ Cov(X_i, X_j) = EX_iX_j - EX_iEX_j$$

$$= \frac{M-1}{N-1} \cdot \frac{M}{N} - \left(\frac{M}{N}\right)^2 = \frac{M}{N} \cdot \frac{M-N}{N\cdot(N-1)}$$

$$\therefore\ DX = \frac{nM}{N}\left(1 - \frac{M}{N}\right) + 2\,C_n^2\,\frac{M}{N}\,\frac{M-N}{N(N-1)}$$

$$= \frac{N-n}{N-1}\, n\, \frac{M}{N}\left(1 - \frac{M}{N}\right).$$

协方差受X与Y本身大小的影响. 如kX与kY的联系与X与Y的联系一样，但 $Cov(kX, kY) = k^2 Cov(X, Y)$. 为此，可以先将$X$与$Y$标准化，再求它们的协方差.

相关系数. $\rho_{XY} := \mathrm{Cov}\left(\frac{X-EX}{\sqrt{DX}}, \frac{Y-EY}{\sqrt{DY}}\right) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}$

(与量纲无关)

定理. $|\rho_{XY}| \le 1$. 且

$|\rho_{XY}| = 1 \Leftrightarrow \exists\, a, b$ s.t. $P(Y = aX + b) = 1$.

证. 由柯西施瓦茨不等式, $|\rho_{XY}| \le 1$.

若 $P(Y = aX+b)$, 则 $\rho_{XY} = \frac{a}{|a|} \in \{\pm 1\}$

若 $|\rho_{XY}| = 1$, 令 $\bar{X} = X - EX$, $\bar{Y} = Y - EY$

则 $g(t) = E(t\bar{X} - \bar{Y})^2$ 有二重根 t_0

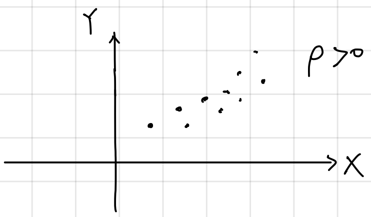
$\therefore E(t_0\bar{X} - \bar{Y})^2 = 0$

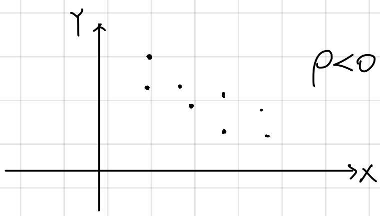
$\Rightarrow P(t_0\bar{X} - \bar{Y} = 0) = 1$

$\Rightarrow P(t_0 X - t_0 EX + EY = Y) = 1$ $\square$

X与Y **不相关**: $\rho = 0$. **完全相关**: $|\rho| = 1$.

正相关: $\rho > 0$. **负相关**: $\rho < 0$.





例. 设 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. 则 $\rho_{XY} = \rho$.

注. 1) 对于多元正态, 独立 $\Leftrightarrow$ 不相关.

2) 独立 $\Rightarrow$ 不相关, 但不相关 $\not\Rightarrow$ 独立

反例. $X \sim N(0,1)$, $Y = X^2$.