4.2　随机变量的方差

1. 概念的引入

引例：甲、乙两射手，各射击十次，X, Y分别表示他们射中的环数，如表：

$X_甲$	8	9	10
P	0.2	0.6	0.2

$Y_乙$	8	9	10
P	0.4	0.2	0.4

问哪一个选手技术较好？

解：E(X)=9.0；E(Y)=9.0.

但直观上，他们射击的水平有差异，甲较稳定，相对于E(X)的偏离较小，所以甲的技术较好。

2. 方差的定义

设 X 是一个随机变量, 若 $E\{[X-E(X)]^2\}$ 存在,

则称 $E\{[X-E(X)]^2\}$ 为 X 的方差,

记为 $D(X)$ 或 $\mathrm{Var}(X)$, 即

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

称 $\sqrt{D(X)}$ 为 $r.v.X$ 的标准差或均方差, 记为 $\sigma(X)$.

两者量纲相同

$D(X)$ —— 描述 r.v. X 的取值偏离平均值的平均偏离程度 —— 数

3. 方差的意义

方差是一个常用来体现随机变量 X 取值分散程度的量. 如果 $D(X)$ 值大, 表示 X 取值分散程度大, $E(X)$ 的代表性差; 而如果 $D(X)$ 值小, 则表示X的取值比较集中, 以 $E(X)$ 作为随机变量的代表性好.

称为均方误差.

用$X-E(X)$, $E(X-E(X))$可以吗?

用$E|X-E(X)|$, $E(X-E(X))^r$可以吗?

4. 随机变量方差的计算

(1) 利用定义计算

离散型随机变量的方差

$$D(X)=\sum_{k=1}^{+\infty}[x_k-E(X)]^2 p_k,$$

其中 $P\{X=x_k\}=p_k, k=1,2,\cdots$ 是 X 的分布律.

连续型随机变量的方差

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

其中 $f(x)$ 为X的概率密度.

(2) 利用公式计算

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

证明

$$\begin{aligned}D(X)&=E\{[X-E(X)]^2\}\\&=E\{X^2-2XE(X)+[E(X)]^2\}\\&=E(X^2)-2E(X)E(X)+[E(X)]^2\\&=E(X^2)-[E(X)]^2\\&=E(X^2)-E^2(X).\end{aligned}$$

5.方差的性质

$$\left.\begin{aligned}&1.\ D(C)=0\\&2.\ D(CX)=C^2D(X)\end{aligned}\right\}\ D(aX+b)=a^2D(X)$$

3. $$\boxed{\begin{aligned}D(X\pm Y)&=D(X)+D(Y)\\&\pm 2E\big((X-E(X))(Y-E(Y))\big)\end{aligned}}$$

特别地，若X,Y相互独立，则

$$\boxed{D(X\pm Y)=D(X)+D(Y)}$$

若 $X_1,\cdots,X_n$ 相互独立，$a_1,a_2,\cdots,a_n,b$ 为常数

则 $D\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 D(X_i)$

若X,Y相互独立 $\underset{\not\Leftarrow}{\Rightarrow}$ $D(X \pm Y) = D(X) + D(Y)$

$\Longleftrightarrow$ $E(XY) = E(X)E(Y)$

4. 对任意常数C, $D(X) \le E(X-C)^2$,

当且仅当$C = E(X)$时等号成立

5. $D(X) = 0$ $\Longleftrightarrow$ $P\{X = E(X)\} = 1$

称为X依概率 1 等于常数 $E(X)$

性质 1 的证明：

$$D(C) = E(C - E(C))^2 = 0$$

性质 2 的证明：

$$D(aX + b) = E((aX + b) - E(aX + b))^2$$

$$= E(a(X - E(X)) + (b - E(b)))^2$$

$$= E(a^2(X - E(X))^2)$$

$$= a^2 D(X)$$

性质 3 的证明：

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

$$= E(X - E(X))^2 + E(Y - E(Y))^2$$

$$\pm 2E\left((X - E(X))(Y - E(Y))\right)$$

$$= D(X) + D(Y)$$

$$\pm 2E\left((X - E(X))(Y - E(Y))\right)$$

注意到，$E\left((X - E(X))(Y - E(Y))\right)$

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

当 X, Y 相互独立时，

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

性质 4 的证明：

$$E(X-C)^2 = E((X-E(X))+(E(X-C))^2$$

$$= E(X-E(X))^2+(C-E(X))^2+2E(X-E(X))(C-E(X))$$

$$= D(X)+(C-E(X))^2$$

当$C=E(X)$时，显然等号成立；

当$C\neq E(X)$时，$(C-E(X))^2>0$

$$E(X-C)^2>D(X)$$

6. 几种重要随机变量的数学期望和方差

（一） X具有(0-1)分布，其分布律为$P\{X=0\}=1-p$，$P\{X=1\}=p$，则$D(X)=p(1-p)$。

证： $E(X)=0\cdot(1-p)+1\cdot p=p$ ，$E(X^2)=0^2\cdot(1-p)+1^2\cdot p=p$，

$D(X)=E(X^2)-[E(X)]^2=p-p^2=p(1-p)$。

（二）二项分布 设$X \sim B(n, p)$，求$D(X)$.

解一 仿照上例求$D(X)$.

解二 引入随机变量 $X_1, X_2, \cdots, X_n$

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验事件 } A \text{ 发生} \\ 0, & \text{第 } i \text{ 次试验事件 } \overline{A} \text{ 发生} \end{cases}$$

$$D(X_i) = p(1-p) \qquad i = 1, 2, \cdots, n$$

$X_1, X_2, \cdots, X_n$ 相互独立，$X = \sum_{i=1}^{n} X_i$

故 $D(X) = \sum_{i=1}^{n} D(X_i) = np(1-p)$

（三）泊松分布 设$X \sim P(\lambda)$, 求$D(X)$.

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

$$\left.\begin{aligned} E(X^2)&=E(X(X-1))+E(X)\\ E(X(X-1))&=\sum_{k=1}^{+\infty} k(k-1)\cdot\frac{\lambda^k e^{-\lambda}}{k!}\\ &=\lambda^2 e^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!}=\lambda^2 \end{aligned}\right\} \Rightarrow E(X^2)=\lambda^2+\lambda$$

$$D(X)=E(X^2)-E^2(X)=\lambda$$

(四) 均匀分布

设X在区间(a，b)上服从均匀分布，则E(X)=(a+b)/2，

$$D(X)=\frac{(b-a)^2}{12}.$$

证：$D(X)=E(X^2)-[E(X)]^2$

$$=\int_a^b x^2\frac{1}{b-a}dx-\left(\frac{a+b}{2}\right)^2=\frac{(b-a)^2}{12}$$

(五)指数分布

参数为θ的指数分布X的概率密度为

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

$$E(X)=\int_0^{+\infty}x\times\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\theta$$

$$E\left(X^2\right)=\int_0^{+\infty}x^2\times\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=2\theta\int_0^{+\infty}x\times\frac{1}{\theta}e^{-\frac{x}{\theta}}dx$$

$$=2\theta^2$$

$$D(X)=E\left(X^2\right)-\left(E(X)\right)^2=\theta^2$$

（六）正态分布 设 $X \sim N(\mu, \sigma^2)$, 求 $D(X)$

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

$$\overset{令\frac{x-\mu}{\sigma}=t}{=} \int_{-\infty}^{+\infty} \sigma^2 t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} -t de^{-\frac{t^2}{2}}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} [(-te^{-\frac{t^2}{2}}) \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt] = \sigma^2$$

例4 已知X,Y相互独立, 且都服从 $N(0,0.5)$, 求 $E(|X-Y|)$.

解 $X \sim N(0,0.5), Y \sim N(0,0.5)$

$$E(X-Y)=0, D(X-Y)=1$$

故 $X-Y \sim N(0,1)$

$$E(|X-Y|)=\int_{-\infty}^{+\infty}|z|\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz$$

$$=\frac{2}{\sqrt{2\pi}}\int_{0}^{+\infty}ze^{-\frac{z^2}{2}}dz=\sqrt{\frac{2}{\pi}}$$

例5 将 编号分别为 $1\sim n$ 的 n 个球随机地放入编号分别为 $1\sim n$ 的n 只盒子中，每盒一 球. 若球的号码与盒子的号码一致，则称为一个配对. 求配对个数 X 的期望与方差.

解 $X_i=\begin{cases}1, & i\text{ 号球放入 } i\text{ 号盒}\\ 0, & \text{其它}\end{cases}$ $\quad i=1,2,\cdots,n$

则 $X=\sum_{i=1}^{n}X_i$

但 $X_1,X_2,\cdots,X_n$ 不相互独立，

X_i	1	0
P	$\frac{1}{n}$	$1-\frac{1}{n}$

$i = 1,2,\cdots,n$

$$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot \frac{1}{n} = 1$$

$$E(X^2) = E\left(\sum_{i=1}^{n} X_i\right)^2 = E\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{1\le i<j\le n} X_i X_j\right)$$

$$= \sum_{i=1}^{n} E(X_i^2) + 2\sum_{1\le i<j\le n} E(X_i X_j)$$

X_i^2	1	0
P	$\frac{1}{n}$	$1-\frac{1}{n}$

$$E(X_i^2)=\frac{1}{n} \qquad i=1,2,\cdots,n$$

X_iX_j	1	0
P	$\frac{1}{n(n-1)}$	$1-\frac{1}{n(n-1)}$

$$E(X_iX_j)=\frac{1}{n(n-1)} \qquad i,j=1,2,\cdots,n$$

$$E(X^2)=\sum_{i=1}^{n}E(X_i^2)+2\sum_{1\le i<j\le n}^{n}E(X_iX_j)$$

$$=\sum_{i=1}^{n}\frac{1}{n}+2\sum_{1\le i<j\le n}^{n}\frac{1}{n(n-1)}$$

$$=n\cdot\frac{1}{n}+2\cdot C_n^2\cdot\frac{1}{n(n-1)}$$

$$=2$$

$$D(X)=E(X^2)-E^2(X)=1$$

标准化随机变量

设随机变量 X 的期望$E(X)$、方差$D(X)$都存在，且$D(X) \neq 0$，则称

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}$$

为 X 的标准化随机变量. 显然，

$$E(X^*) = 0, \quad D(X^*) = 1$$

仅知 r.v. 的期望与方差并不能确定其分布

例如

X	-1	0	1
P	0.1	0.8	0.1

$$E(X)=0,D(X)=0.2$$

与

Y	-2	0	2
P	0.025	0.95	0.025

$$E(Y)=0,D(Y)=0.2$$

有相同的期望方差但是分布却不相同

6、切比雪夫不等式

设随机变量*X*有期望*E*(*X*)和方差σ^2，则对于任给 >0,

$$P\{|X-E(X)|\geq \varepsilon\}\leq \frac{\sigma^2}{\varepsilon^2}$$

或 $$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

由切比雪夫不等式可以看出，若 σ^2越小，则事件{|*X*-*E*(*X*)|< }的概率越大，即随机变量*X*集中在期望附近的可能性越大.

❖证：只对连续型情况给出证明。设X的概率密度为$f(x)$，记μ=E(X)，则有

$$P\{|X-\mu|\geq\varepsilon\}=\int\limits_{|x-\mu|\geq\varepsilon} f(x)dx \leq \int\limits_{|x-\mu|\geq\varepsilon} f(x)\frac{(x-\mu)^2}{\varepsilon^2}dx$$

$$\leq\int_{-\infty}^{+\infty} f(x)\frac{(x-\mu)^2}{\varepsilon^2}dx == \frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty} f(x)(x-\mu)^2dx$$

$$=\frac{D(X)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2}.$$

意义：

(1)这个不等式给出了在随机变量X的分布未知的情况下事件 $|X-\mu|<\varepsilon$ 的概率的一种估计方法。例如：

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{1}{9}=0.8889;$$

$$P\{|X-\mu|<4\sigma\}\geq 1-\frac{1}{16}=0.9375.$$

(2) 切比雪夫不等式也从另一角度体现了方差D(X)的意义。从切比雪夫不等式可以看出，随机变量X的方差越小，则X的取值越集中在其中心E(X)的附近。方差越小，X取值越集中在区间(E(X)−ε，E(X)+ε)之内。

(3)已知$X \sim N(\mu, \sigma^2)$,

$P\{|X-\mu|<\sigma\}=2\Phi(1)-1=0.6826$

$P\{|X-\mu|<2\sigma\}=2\Phi(2)-1=0.9544$

$P\{|X-\mu|<3\sigma\}=2\Phi(3)-1=0.9944$

说明：$X \sim N(\mu, \sigma^2)$落在$(\mu-3\sigma, \mu+3\sigma)$内的概率为0.9944, 这一事实称为“$3\sigma$规则”

例6 已知正常男性成人血液中，每一毫升白细胞数平均是7300，均方差是700 . 利用切比雪夫不等式估计每毫升白细胞数在5200~9400之间的概率 .

解：设每毫升白细胞数为X

依题意，$E(X)=7300, D(X)=700^2$

所求为 $P(5200 \quad X \quad 9400)$

$$P(5200 \quad X \quad 9400)$$

$$=P(5200-7300 \quad X-7300 \quad 9400-7300)$$

$$=P(-2100 \quad X-E(X) \quad 2100)$$

$$=P\{|X-E(X)| \quad 2100\}$$

由切比雪夫不等式

$$P\{|X-E(X)| \quad 2100\} \geq 1-\frac{D(X)}{(2100)^2}$$

$$=1-\left(\frac{700}{2100}\right)^2=1-\frac{1}{9}=\frac{8}{9}$$

即估计每毫升白细胞数在5200~9400之间的概率不小于8/9 .

例7 在每次试验中，事件A发生的概率为0.75，利用切比雪夫不等式求：n需要多么大时，才能使得在n次独立重复试验中，事件A出现的频率在0.74~0.76之间的概率至少为0.90?

解：设X为n次试验中，事件A出现的次数，

则 $X \sim B(n, 0.75)$

$E(X)=0.75n$, $D(X)=0.75*0.25n=0.1875n$

所求为满足

$$P\left(0.74 < \frac{X}{n} < 0.76\right) \geq 0.90$$

的最小的n.

$P(0.74 < \frac{X}{n} < 0.76)$ 可改写为

$$P(0.74n < X < 0.76n)$$

$$= P(-0.01n < X - 0.75n < 0.01n)$$

$$= P\{|X - E(X)| < 0.01n\}$$

在切比雪夫不等式中取 $\varepsilon = 0.01n$，则

$$P(0.74 < \frac{X}{n} < 0.76) = P\{|X - E(X)| < 0.01n\}$$

$$\geq 1 - \frac{D(X)}{(0.01n)^2} = 1 - \frac{0.1875n}{0.0001n^2} = 1 - \frac{1875}{n}$$

依题意，取 $1-\dfrac{1875}{n}\geq 0.9$

解得

$$n\geq \frac{1875}{1-0.9}=18750$$

即n取18750时，可以使得在n次独立重复试验中,事件A出现的频率在0.74~0.76之间的概率至少为0.90 .

方差性质5: $P\{X=E(X)\}=1 \Leftrightarrow D(X)=0$

证明:充分性:设$P\{X=E(X)\}=1$,

$$P\{X-E(X)=0\}=1,$$

则$D(X)=E(X-E(X))^2=0$

必要性：用反证法。

假设$P\{X=E(X)\}<1$,

则存在常数$\varepsilon>0$，有$P\{|X-E(X)|>\varepsilon\}>0$,

而由切比雪夫不等式知，对于任意常数$\varepsilon>0$,

有$P\{|X-E(X)|>\varepsilon\}\le \frac{D(X)}{\varepsilon^2}=0$,

即$P\{|X-E(X)|>\varepsilon\}=0$，矛盾,

所以$P\{X=E(X)\}=1$

另证：

$$\mathbf{P}\{\mathbf{X} \neq \mathbf{E}(\mathbf{X})\} = \mathbf{P}\{\bigcup_{n=1}^{\infty}\{|X - E(X)| \geq \frac{1}{n}\}\}$$

$$\leq \sum_{n=1}^{\infty} P\left\{|X - E(X)| \geq \frac{1}{n}\right\} \leq \sum_{n=1}^{\infty} \frac{D(X)}{\frac{1}{n^2}} = 0$$

则$\mathbf{P}\{\mathbf{X} = \mathbf{E}(\mathbf{X})\} = \mathbf{1}$

7. 矩

1.定义

设 X 和 Y 是随机变量, 若 $E(X^k)$, $k=1,2,\cdots$ 存在, 称它为 X 的 k 阶原点矩, 简称 k 阶矩.

若 $E\{[X-E(X)]^k\}$, $k=2,3,\cdots$ 存在, 称它为 X 的 k 阶中心矩.

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

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

2. 说明

(1) 以上数字特征都是随机变量函数的数学期望;

(2) 随机变量 X 的数学期望 $E(X)$ 是 X 的一阶原点矩, 方差为二阶中心矩, 协方差 $\mathrm{Cov}(X,Y)$ 是 X 与 Y 的二阶混合中心矩;

(3) 在实际应用中, 高于4阶的矩很少使用.

三阶中心矩 $E\{[X-E(X)]^3\}$ 主要用来衡量随机变量的分布是否有偏.

四阶中心矩 $E\{[X-E(X)]^4\}$ 主要用来衡量随机变量的分布在均值附近的陡峭程度如何.