§2 方差和标准差

实际背景

例 设有两种牌号的手表，其走时误差情况如下表

日误差(秒)	-3	-2	-1	0	1	2	3
概率(牌号甲)	0.10	0.15	0.15	0.20	0.15	0.15	0.10
概率(牌号乙)	0.05	0.05	0.10	0.60	0.10	0.05	0.05

试问哪种牌号的手表质量较好？

分析 设两种手表的走时误差分别为 X, Y, 则

$$E(X)=\sum_{k=1}^{7} x_k P\{X=x_k\}=0$$

$$E(Y)=\sum_{k=1}^{7} y_k P\{Y=y_k\}=0$$

可见两种手表的平均误差一样.

质 量 是 否 一 样

?

从偏离平均值的大小来考虑

对 r.v X 考虑偏差

$$|X-E(X)|$$

偏差越小，说明质量越稳定

绝对值运算不方便

平方偏差

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

平方偏差仍是r.v

平方偏差的平均值

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

方差反映r.v偏离平均值的平均大小

定义 对r.v X，若

$$\mathrm{Var}(X)\triangleq D(X)\triangleq E(X-E(X))^2$$

存在，则称 $D(X)$为r.v X的**方差**(variance)，$\sqrt{D(X)}$称为**标准差**(standard deviation).

例 设甲、乙两射手击中环数分别为 X, Y, 频率函数为

X	8	9	10
p_k	0.15	0.40	0.45

,

Y	8	9	10
p_k	0.35	0.10	0.55

试评估两人的射击技术.

解 先计算数学期望

$$E(X) = 8\times 0.15 + 9\times 0.40 + 10\times 0.45 = 9.3$$

$$E(Y) = 8\times 0.35 + 9\times 0.10 + 10\times 0.55 = 9.2$$

又

$$D(X) = (8-9.3)^2\cdot 0.15 + (9-9.3)^2\cdot 0.4 + (10-9.3)^2\cdot 0.45$$

$$= 0.51$$

$$D(Y) = (8-9.2)^2\cdot 0.35 + (9-9.2)^2\cdot 0.1 + (10-9.2)^2\cdot 0.55$$

$$= 0.86$$

可见甲的射击水平比乙略好, 且甲的技术比乙要稳定.

§2 方差和标准差

实际意义

数学期望 —— r.v的平均值

方　　差 —— r.v与平均值的平均偏离程度

方差的计算

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

视为 $g(X)=(X-E(X))^2$ 的数学期望，则有

① 设 X 的频率函数为

$$P\{X=x_k\}=p_k, k=1,2,\cdots$$

则

$$D(X)=\sum_{k=1}^{\infty}(x_k-E(X))^2\cdot p_k$$

② 设 X 的概率密度为 $f(x)$,则

$$D(X)=\int_{-\infty}^{\infty}(x-E(X))^2 f(x)dx$$

实际意义

数学期望 —— r.v的平均值

方　　差 —— r.v与平均值的平均偏离程度

方差的计算

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

视为 $g(X)=(X-E(X))^2$ 的数学期望，则有

E(X) 是常数

③ $$\begin{aligned}D(X)&=E(X-E(X))^2\\&=E(X^2)-2E[XE(X)]+E[E(X)]^2\\&=E(X^2)-2E(X)E(X)+[E(X)]^2\\&=E(X^2)-[E(X)]^2\end{aligned}$$

例 设 $X \sim P(\lambda)$, 求 $D(X)$.

通常利用下述公式计算

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

解 由上节计算得 $E(X) = \lambda$,

$$E(X^2) = E[X(X-1) + X]$$

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

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

$$= \lambda^2 + \lambda$$

$$\therefore \quad D(X) = E(X^2) - [E(X)]^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda$$

例 设 $X \sim U(a,b)$, 求 $D(X)$.

解 由上节计算得 $E(X)=(a+b)/2$, X 的密度为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}\end{cases}$$

$$\therefore \quad E(X^2)=\int_a^b \frac{x^2}{b-a}dx$$

$$=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}$$

$$\therefore \quad D(X)=E(X^2)-[E(X)]^2$$

$$=\frac{a^2+ab+b^2}{3}-\left(\frac{a+b}{2}\right)^2$$

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

例 设 X 的密度函数为

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

求 $D(X)$.

解 由 $E(X)=\theta$, 故

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

$$=\int_0^{\infty}x^2\frac{1}{\theta}e^{-x/\theta}dx-\theta^2$$

$$=-x^2e^{-x/\theta}\Big|_0^{\infty}+2\int_0^{\infty}xe^{-x/\theta}dx-\theta^2$$

$$=2\theta^2-\theta^2=\theta^2$$

方差的基本性质

① 若 $X \overset{a.e}{=} c$(常数),则 $D(X)=0$

② 设 c为常数,则 $D(cX)=c^2D(X)$

证 $$\begin{aligned} D(cX) &= E(cX-E(cX))^2 \\ &= E(cX-cE(X))^2 \\ &= c^2E(X-E(X))^2 = c^2D(X) \end{aligned}$$

③ 对于r.v X,Y 有

$$D(X+Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]$$

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

方差的基本性质

① 若 $X \overset{a.e}{=} c$（常数），则 $D(X)=0$

② 设 c为常数，则 $D(cX)=c^2D(X)$

③ 对于 r.v X,Y 有

$$D(X+Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]$$

特别当 X,Y 独立时，有

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

证 $\because$ X,Y 独立 $\therefore$ $X-E(X)$, $Y-E(Y)$独立

$$\therefore\ E[(X-E(X))(Y-E(Y))]=E[X-E(X)]\cdot E[(Y-E(Y)]$$

$$=[E(X)-E(X)]\cdot[E(Y)-E(Y)]$$

$$=0$$

问 当 X,Y独立时，有

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

方差的基本性质

① 若 $X \overset{a.e}{=} c$(常数), 则 $D(X)=0$

② 设 c为常数, 则 $D(cX)=c^2D(X)$

③ 对于 r.v X,Y 有

$$D(X+Y)=D(X)+D(Y)+2E[(X-E(X))(Y-E(Y))]$$

特别当 X,Y 独立时, 有

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

若 $X_1,X_2,\cdots,X_n$独立, 则

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

④ $D(X)=0 \Leftrightarrow X \overset{a.e}{=} c$ (常数)

例 设 $X \sim b(n,p)$, 求 $D(X)$.

解 因为二项分布来自 n 重伯努利试验, 故有

$$X = X_1 + X_2 + \cdots + X_n$$

其中

$$X_i = \begin{cases} 1, \text{第 } i \text{ 次伯努利试验事件 } A \text{ 发生} \\ 0, \text{第 } i \text{ 次伯努利试验事件 } \bar{A} \text{ 发生} \end{cases} \quad (i = 1, 2, \cdots, n)$$

且 $X_1, X_2, \cdots, X_n$ 独立同分布, 其频率函数为

$$P\{X_i = 1\} = p, \; P\{X_i = 0\} = 1 - p$$

$$\therefore \quad E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$= nE(X_1) = np$$

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

$$= n[(1-p)^2 \cdot p + (0-p)^2 \cdot (1-p)]$$

$$= np(1-p)$$

例 设 $X \sim N(\mu,\sigma^2)$, 求 $D(X)$.

解 由上节计算得 $E(X)=\mu$, 故

$$D(X)=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

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

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2 e^{-\frac{t^2}{2}}dt$$

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

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

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

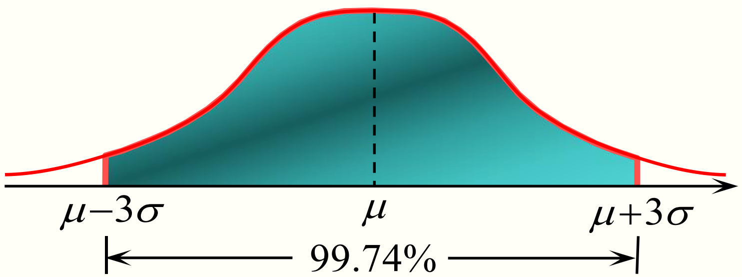
$$=\sigma^2$$

3σ 原则： 正态r.v的值几乎都落在$(\mu-3\sigma,\mu+3\sigma)$内.

即 $X\sim N(\mu,\sigma^2)$, 则

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

$$P\{|X-\mu|\geq+3\sigma\}=0.0026$$



问 对一般的 r.v X，如何估计概率

$$P\{|X-\mu|\geq\varepsilon\}\leq ?$$

其中$\mu=E(X),\varepsilon>0$是任意实数.

定理（切比雪夫Chebyshev不等式）

设$E(X)\triangleq\mu, D(X)\triangleq\sigma^2$都存在, 则$\forall\ \varepsilon>0$, 有

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$$

定理（切比雪夫Chebyshev不等式）

设$E(X)\triangleq\mu, D(X)\triangleq\sigma^2$ 都存在，则$\forall\ \varepsilon>0$，有

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$$

证 只证连续型情形.

设 r.v X的密度函数为$f(x)$,则

$$P\{|X-\mu|\geq\varepsilon\}=\int_{|x-\mu|\geq\varepsilon} f(x)dx$$

$$\leq\int_{|x-\mu|\geq\varepsilon}\frac{|x-\mu|^2}{\varepsilon^2}f(x)dx$$

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

$$=\frac{1}{\varepsilon^2}D(X)$$

$$=\frac{\sigma^2}{\varepsilon^2}$$

定理（切比雪夫Chebyshev不等式）

设 $E(X)\triangleq\mu, D(X)\triangleq\sigma^2$ 都存在，则 $\forall\ \varepsilon>0$，有

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$$

证　还可以利用Markov不等式.

令 $Y=(X-\mu)^2$，　则 $E(Y)=\sigma^2$.

对 Y 利用Markov不等式即得.

定理（马尔可夫Markov不等式）

设 r.v. X 满足 $P\{X\geq 0\}=1$，且 $E(X)$ 存在，则

$$P\{X\geq t\}\leq\frac{E(X)}{t}.$$

定理（切比雪夫Chebyshev不等式）

设$E(X)\triangleq\mu, D(X)\triangleq\sigma^2$都存在，则$\forall\ \varepsilon>0$，有

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$$

改写为

$$P\{|X-\mu|<\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

分别取$\varepsilon=3\sigma,4\sigma$，则有

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{1}{9}=88.90\%$$

$$P\{|X-\mu|<4\sigma\}\geq 1-\frac{1}{16}=93.75\%$$

即使对于一般的r. v,
3σ原则的可信度也近90%

推论

若$\sigma^2=0$，则$P\{X=\mu\}=1$.

课后作业

P118：49、50、55，补充题1，2

1. 设X, Y是互相独立的随机变量，且有$E(X)=3$, $E(Y)=1$, $D(X)=4$, $D(Y)=9$. 令$Z=5X-2Y+15$, 求 $E(Z)$ 和 $D(Z)$.

2. 设随机变量X_1, X_2, X_3, X_4互相独立，且有$E(X_i)=2i$, $D(X_i)=5-i$, 其中$i=1,2,3,4$. 令$Z=2X_1-X_2+3X_3-\frac{1}{2}X_4$, 求 $E(Z)$ 和 $D(Z)$.