§4.2 方差

引例 甲、乙两射手各打了**6**发子弹,每发子弹击中的环数分别为:

甲	**10, 7, 9, 8, 10, 6,**
乙	**8, 7, 10, 9, 8, 8,**

有五个不同数

有四个不同数

问哪一个射手的技术较好?

解 首先比较平均环数

$\overline{\text{甲}} = 8.3,\quad \overline{\text{乙}} = 8.3$ 再比较稳定程度

甲:
$$2\times(10-8.3)^2+(9-8.3)^2+(8-8.3)^2+(7-8.3)^2+(6-8.3)^2=13.34$$

乙:
$$(10-8.3)^2+(9-8.3)^2+3\times(8-8.3)^2+(7-8.3)^2=5.34$$

乙比甲技术稳定，故乙技术较好.

进一步比较平均偏离平均值的程度

甲

$$\frac{1}{6}[2\times(10-8.3)^2+(9-8.3)^2+(8-8.3)^2+(7-8.3)^2+(6-8.3)^2]$$

$$=13.34/6=2.22 \triangleq \sum_{k=1}^{6}\left(x_k-E(X)\right)^2 p_k$$

乙

$$\frac{1}{6}[(10-8.3)^2+(9-8.3)^2+3\times(8-8.3)^2+(7-8.3)^2]$$

$$=5.34/6=0.89 \triangleq \sum_{k=1}^{6}\left(x_k-E(X)\right)^2 p_k$$

$$\boxed{E\{(X-E(X))^2\}}$$

方差概念

定义　若$E\{[X-E(X)]^2\}$存在，则称其为随机变量 X 的方差，记为$D(X)$ 或 $Var(X)$

即　$$D(X)=E\{[X-E(X)]^2\}$$

称 $\sqrt{D(X)}$ 为 X 的均方差或标准差.

$D(X)$ — 描述随机变量X的取值偏离平均值的平均偏离程度

若 X 为离散型随机变量，分布律为　$P\{X=x_k\}=p_k,\ k=1,2,\cdots$

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

若 $\boldsymbol{X}$ 为连续型随机变量，概率密度为 $\boldsymbol{f(x)}$

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

计算方差的常用公式:

$$\boxed{D(X)=E(X^2)-E^2(X)}$$

Proof.

$$D(X)=E\{[X-E(X)]^2\}$$
$$=E\{X^2-2XE(X)+[E(X)]^2\}$$
$$=E(X^2)-2E(X)E(X)+[E(X)]^2$$
$$=E(X^2)-[E(X)]^2$$

一维随机变量的方差

◆ 离散型

设离散型随机变量X的概率分布为

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

$$D(X)=\sum_k p_k(x_k-\mu)^2=\sum_k x_k^2 p_k-(\sum_k x_k p_k)^2$$

◆ 连续型

设连续型随机变量X的分布密度为 $f(x)$

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

方差的计算

例　设有两种球形产品，其直径的取值规律如下：

X_1	4	5	6
P	1/4	1/2	1/4

X_2	2	3	5	7	8
P	1/8	1/8	1/2	1/8	1/8

求**$D(X_1)$,$D(X_2)$**

解　　$E(X_1)=5$　　　　$E(X_2)=5$

$$E(X_1^2)=4^2\times\frac{1}{4}+5^2\times\frac{1}{2}+6^2\times\frac{1}{4}=25.5$$

$$E(X_2^2)=2^2\times\frac{1}{8}+3^2\times\frac{1}{8}+5^2\times\frac{1}{2}+7^2\times\frac{1}{8}+8^2\times\frac{1}{8}=28.25$$

$$D(X_1)=E(X_1^2)-E^2(X_1)=0.5$$

$$D(X_2)=E(X_2^2)-E^2(X_2)=3.25$$

0-1分布的方差

◆ 分布律

X	0	1
P	1-p	p

$$E(X) = p$$

◆ 方差

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

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

其中 $q = 1 - p$

泊松分布的方差

◆ 分布律

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}$$

◆ 方差

$$E(X)=\lambda$$

$$\left.\begin{aligned}&E(X^2)=E(X(X-1))+E(X)\\&E(X(X-1))=\sum_{k=0}^{+\infty}k(k-1)\cdot\frac{\lambda^k e^{-\lambda}}{k!}=\lambda^2e^{-\lambda}\sum_{k=2}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!}=\lambda^2\end{aligned}\right\}\Rightarrow E(X^2)=\lambda^2+\lambda$$

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

If $X\sim P(\lambda)$, then $D(X)=\lambda$

均匀分布的方差

◆ 分布密度

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

◆ 方差 $E(X)=\dfrac{1}{2}(a+b)$

$$E(X^2)=\int_a^b \frac{x^2}{b-a}dx=\frac{x^3}{3(b-a)}\Big|_a^b=\frac{a^2+ab+b^2}{3}$$

$$D(X)=E(X^2)-[E(X)]^2=\frac{1}{12}(b-a)^2$$

正态分布的方差

◆ 分布密度 $X \sim N(\mu, \sigma^2)$ $E(X) = \mu$

◆ 方差

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

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

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

指数分布的方差

◆ 分布密度

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

$$E(X)=\theta$$

◆ 方差

$$E(X^2)=\int_0^{+\infty}x^2\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=-x^2e^{-\frac{x}{\theta}}\Big|_0^{+\infty}+\int_0^{+\infty}2xe^{-\frac{x}{\theta}}dx$$

$$=-2\theta\left[xe^{-\frac{x}{\theta}}\Big|_0^{+\infty}-\int_0^{+\infty}e^{-\frac{x}{\theta}}dx\right]=2\theta^2\left[e^{-\frac{x}{\theta}}\Big|_0^{+\infty}\right]=2\theta^2$$

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

方差的计算步骤

Step 1: 计算期望 **E(X)**

$$E(X) = p_1 x_1 + p_2 x_2 + \cdots p_k x_k + \cdots = \sum_k p_k x_k$$ 离散型

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$ 连续型

Step 2: 计算 **E(X²)**

$$E(X^2) = p_1 x_1^2 + p_2 x_2^2 + \cdots p_k x_k^2 + \cdots = \sum_k p_k x_k^2$$ 离散型

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

Step 3: 计算 **D(X)**

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

方差的性质

- $D(C)=0$
- $D(aX+b)=a^2D(X)$
- $\boxed{D(X\pm Y)=D(X)+D(Y)\pm 2E\big((X-E(X))(Y-E(Y))\big)}$

特别地，若$\boldsymbol{X},\boldsymbol{Y}$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

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

- $\boldsymbol{D(X)=0} \longleftrightarrow \boldsymbol{P\{X=E(X)\}=1}$

称为$\boldsymbol{X}$依概率 **1** 等于常数 $\boldsymbol{E(X)}$

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

$\underset{\not\Leftarrow}{\Rightarrow}$ $E(XY)=E(X)E(Y)$

- **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**

性质 **2** 的证明:

$$
\begin{aligned}
D(aX+b) &= E\big((aX+b)-E(aX+b)\big)^2 \\
&= E\big(a(X-E(X))+(b-E(b))\big)^2 \\
&= E\big(a^2(X-E(X))^2\big) = a^2D(X)
\end{aligned}
$$

二项分布的方差

分布律　　X服从二项分布，其概率分布为

$$P\{X=k\}=C_n^k p^k (1-p)^{n-k}$$

数学期望

根据二项分布的定义，随机变量**X**是**n**重伯努利试验中事件**A**发生的次数，且在每次实验中事件**A**发生的概率为**p**。引进随机变量

$$X_i=\begin{cases}0, & A\text{在第}i\text{次试验中不发生}\\ 1, & A\text{在第}i\text{次试验中发生}\end{cases}$$

则$X=\sum_{i=1}^{n}X_i$，　且X_i只依赖于第k次试验，而各次试验相互独立，于是X_1，$X_2\cdots X_n$相互独立，而X_i服从0-1分布，　**即二项分布可表示为** n **个0-1分布的和**

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

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

故

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

If X~B(n, p), then D(X)= np(1-p)

常见分布及其期望和方差列表

分布名称	数学期望E（X）	方差D（X）
0-1分布	p	pq
二项分布	np	npq
泊松分布	λ	λ
均匀分布	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
正态分布	μ	σ^2
指数分布	θ	θ^2

例 已知一批玉米种子的发芽率是75%，播种时每穴种三粒，求每穴发芽种子粒数的数学期望、方差及均方差.

解 设发芽种子数为 X，则 X 服从二项分布，且

$$n = 3, \quad p = 0.75$$

所以

$$E(X) = np = 3 \times 0.75 = 2.25$$

$$D(X) = np(1 - p) = 3 \times 0.75 \times 0.25 = 0.5625$$

$$\sqrt{D(X)} = \sqrt{0.5625} = 0.75$$

例 某动物的寿命X（年）服从指数分布，其中参数θ =10，求这种动物的平均寿命及标准差.

解 因为X服从指数分布，且 $\theta=10$

$$E(X)=\theta=10, \qquad D(X)=\theta^2=100$$

$$\sqrt{D(X)}=\sqrt{100}=10$$

所以这种动物的平均寿命为10年，标准差为10年.

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,

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

结论

若$X_i \sim N(\mu_i, \sigma_i^2), i=1,2,\cdots,n,$且它们相互独立,

且它们的线性组合

$$C_1X_1 + C_2X_2 + \cdots + C_nX_n$$

仍然服从正态分布, 且

$$C_1X_1 + C_2X_2 + \cdots + C_nX_n \sim N(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2)$$

设活塞的直径（以cm计）$X \sim N(22.40, 0.03^2)$，气缸的直径 $Y \sim N(22.50, 0.04^2)$，X,Y相互独立。任取一只活塞，任取一只气缸，求活塞能装入气缸的概率。

练一练

设随机变量X服从参数为1的指数分布，求 $E\{X+e^{-2X}\}$

解 **X**的密度函数为

$$f(x)=\begin{cases} e^{-x}, & (x\geq 0) \\ 0, & (x<0) \end{cases} \qquad EX=1$$

所以 $$E(X+e^{-2X})=EX+E(e^{-2X})$$

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

$$=\int_{0}^{+\infty} e^{-3x}dx=\frac{1}{3}$$

所以 $$E(X+e^{-2X})=\frac{4}{3}$$

二维随机变量的方差

$$\boxed{D(X,Y)=(D(X),D(Y))}$$

◆ **(X,Y)**为二维连续型随机变量

$$D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f_X(x)dx$$

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

$$D(Y)=\int_{-\infty}^{+\infty}[y-E(Y)]^2 f_Y(y)dx$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[y-E(Y)]^2 f(x,y)dxdy$$

练一练

X_1, X_2 是两个相互独立的随机变量，其概率密度分别为

$$f_1(x)=\begin{cases}2x, & 0\le x\le 1,\\ 0, & \text{其它}.\end{cases}\qquad f_2(x)=\begin{cases}e^{-(x-5)}, & x\ge 5,\\ 0, & \text{其它}.\end{cases}$$

求 $D(X_1+X_2)$

解 因为 X_1, X_2 相互独立，所以

$$D(X_1+X_2)=D(X_1)+D(X_2)$$

而 $$E(X_1)=\int_{-\infty}^{+\infty}xf_1(x)dx=\int_0^1 x\cdot 2xdx=\frac{2}{3}$$

$$E(X_2)=\int_{-\infty}^{+\infty}yf_2(y)dy=\int_5^{+\infty}y\cdot e^{-(y-5)}dy=6$$

$$E(X_1^2)=\int_{-\infty}^{+\infty}x^2 f_1(x)dx \ =\int_0^1 x^2\cdot 2xdx=\frac{1}{2}$$

$$E(X_2^2)=\int_{-\infty}^{+\infty}y^2 f_2(y)dy \ =\int_5^{+\infty} y^2\cdot e^{-(y-5)}dy$$

$$=-y^2\cdot e^{-(y-5)}\Big|_5^{+\infty}+\int_5^{+\infty}2y\cdot e^{-(y-5)}dy$$

$$=25-2y\cdot e^{-(y-5)}\Big|_5^{+\infty}+\int_5^{+\infty}2e^{-(y-5)}dy$$

$$=35-2e^{-(y-5)}\Big|_5^{+\infty}=37$$

所以 $D(X_1)=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$ $\quad D(X_2)=37-6^2=1$

$$D(X_1+X_2)=\frac{1}{18}+1=\frac{19}{18}$$