§2 方差

- 方差的定义
- 方差的性质
- 切比雪夫不等式
- 几种重要的随机变量的期望与方差

§2　方差

在实际问题中常关心随机变量与均值的偏离程度，可用 $E|X-E(X)|$，但不方便；所以通常用 $E\left[(X-E(X))^2\right]$来度量随机变量 X 与其均值 E(X)的偏离程度。

1、定义　设 X 是随机变量，若 $E\left[(X-E(X))^2\right]$存在，称其为随机变量 X 的*方差*（variance），记作 D(X)，Var(X)，即：$D(X)=\mathrm{Var}(X)=E\left[(X-E(X))^2\right]$。$\sqrt{D(X)}$称为*标准差*。

离散型：$$D(X)=E\left((X-E(X))^2\right)=\sum_{i=1}^{\infty}(x_i-E(X))^2\cdot p_i\ ,$$

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

注：方差描述了随机变量的取值与其均值的偏离程度。

方差也可由下面公式求得：

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

证明：　$D(X)=E\left[\left(X-E(X)\right)^2\right]$

$$=E\left[X^2-2E(X)\cdot X+E(X)^2\right]$$

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

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

例1

甲、乙两人射击，他们的射击水平由下表给出：

X：甲击中的环数；　　Y：乙击中的环数；

X	8	9	10
P	0.3	0.2	0.5

Y	8	9	10
P	0.2	0.4	0.4

试问哪一个人的射击水平较高？

解：　比较两个人的平均环数.

甲的平均环数为

$$E(X)=8\times0.3+9\times0.2+10\times0.5=9.2\ (\text{环})$$

乙的平均环数为

$$E(Y)=8\times0.2+9\times0.4+10\times0.4=9.2\ (\text{环})$$

因此，从平均环数上看，甲乙两人的射击水平是一样的，但两个人射击环数的方差分别为

$$D(X)=(8-9.2)^2\times0.3+(9-9.2)^2\times0.2+(10-9.2)^2\times0.5=0.76$$

$$D(Y)=(8-9.2)^2\times0.2+(9-9.2)^2\times0.4+(10-9.2)^2\times0.4=0.624$$

由于 $D(Y)<D(X)$，这表明乙的射击水平比甲稳定.

2、方差的性质

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

1) D(X)≥0；若 C 是常数，则 D(C)=0

2) $D(CX)=C^2D(X)$

3) $D(aX+bY)=a^2D(X)+b^2D(Y)+2abE\left[(X-E(X))(Y-E(Y))\right]$,

a，b 是常数。若 X，Y 独立，

则 $D(aX+bY)=a^2D(X)+b^2D(Y)$

4) D(X)=0 ⇔ P{X=c}=1，c=E(X)

注：（见书中例1）

令，$Y=(X-E(X))/\sqrt{D(X)}$ 　则 **E(Y)=0, D(Y)=1**。

称**Y**是随机变量**X**的**标准化**了的随机变量。

3、几种重要随机变量的数学期望及方差

- 两点分布
- 二项分布
- 泊松分布
- 均匀分布
- 正态分布

1）两点分布

X	0	1
p_k	$1-p$	p

$$E(X)=p,\quad D(X)=E(X^2)-(E(X))^2=p-p^2=pq$$

2） 二项分布

方法1：

设$X_1,\ldots,X_n$相互独立，都服从参数为 p 的两点分布，

$$P\{X_i=0\}=q,\quad P\{X_i=1\}=p, i=1,2,\cdots,n.$$

令 $X=X_1+\cdots+X_n$ 则 $X\sim B(n,p)$，即

$$P\{X=k\}=C_n^k p^k q^{n-k},\qquad k=0,\cdots,n$$

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

$$D(X)=\sum_{i=1}^{n}D(X_i)=\sum_{i=1}^{n}pq=npq.$$

,

方法1说明了二项分布与两点分布的关系。

方法2：

$$P\{X=k\}=C_n^k p^k q^{n-k}, k=0,1,\cdots,n$$

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

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

$$=np\sum_{k=1}^{n} C_{n-1}^{k-1}p^{k-1}q^{n-1-(k-1)}=np\sum_{i=0}^{n-1} C_{n-1}^{i}p^{i}q^{n-1-i}$$

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

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

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

$$=\sum_{k=0}^{n}k^2\cdot\frac{n!}{k!(n-k)!}p^k q^{n-k}=p\sum_{k=1}^{n}k\frac{n!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}$$

$$=p\sum_{k=1}^{n}(k-1)\frac{n!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}+p\sum_{k=1}^{n}\frac{n!}{(k-1)!(n-k)!}p^{k-1}q^{n-k}$$

$$=n(n-1)p^2\sum_{k=2}^{n}\frac{(n-2)!}{(k-2)!(n-2-(k-2))!}p^{k-2}q^{n-2-(k-2)}+np$$

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

$$D(X)=E(X^2)-(E(X))^2$$
$$=n^2p^2-n\ \ p^2+np-n^2p^2$$
$$=np(1-p)=npq$$

3）泊松分布

设 X 服从参数为 λ 的泊松分布，

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots$$

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

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

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

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

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

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

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

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

4）均匀分布

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

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{a+b}{2}$$

$$D(X)=E(X^2)-(E(X))^2=\int_a^b x^2\frac{1}{b-a}dx-(\frac{a+b}{2})^2$$
$$=\frac{(b-a)^2}{12}$$

5）正态分布

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$$

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

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

$$=\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}te^{-\frac{t^2}{2}}dt+\frac{\mu}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=0+\mu=\mu$$

$$D(X)=E\left[(X-\mu)^2\right]$$

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

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

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

说明： 若 $X\sim N(0,1)$, 则 $D(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt=1$

3、定理（切比雪夫(Chebyshev)不等式）

设随机变量X有数学期望 $E(X)=\mu$, 方差 $D(X)=\sigma^2$**，对任意** ε **>0, 不等式**

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

成立。

证明：（只证 X 是连续型）

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

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

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

这个不等式给出了随机变量 X 的分布未知情况下，事件$\{|X-\mu|<\varepsilon\}$的概率的一种估计方法。

例如：在上面不等式中，取$\varepsilon=3\sigma$，4σ，有：

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

$$P\{|X-\mu|<4\sigma\}\geq 0.9375$$

例 3　假设一批种子的良种率为$\frac{1}{6}$，从中任意选出600粒，试用切比雪夫（Chebyshev）不等式估计：这600粒种子中良种所占比例与$\frac{1}{6}$之差的绝对值不超过0.02的概率。

解：设X表示600粒种子中的良种数,则 $X \sim B(600,\frac{1}{6})$.

$$E(X)=600\times\frac{1}{6},\qquad D(X)=600\times\frac{1}{6}\times\frac{5}{6}.$$

由切比雪夫不等式有

$$P\{\left|\frac{X}{600}-\frac{1}{6}\right|\le 0.02\}= P\{\left|\frac{X-100}{600}\right|\le 0.02\}$$

$$=P\{|X-100|\le 12\}\ge 1-\frac{D(X)}{12^2}=1-\frac{600\times\frac{1}{6}\times\frac{5}{6}}{144}=0.4213$$

例 4 利用Chebyshev不等式证明：

若 $D(X)=0$ ，则 $P\{X=E(X)\}=1.$

证明：

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

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

而 $P\{|X-E(X)|\neq 0\}=P\left(\bigcup_{n=1}^{\infty}\left\{|X-E(X)|\geq \frac{1}{n}\right\}\right)$

$$\leq \sum_{n=1}^{\infty}P\left\{|X-E(X)|\geq \frac{1}{n}\right\}$$ (概率的次可列可加性)

由概率的非负性及Chebyshev不等式，得

$$0 \le P\left\{\left|X - E(X)\right| \ge \frac{1}{n}\right\} \le \frac{D(X)}{\left(\frac{1}{n}\right)^2} = 0$$

所以，$P\left\{\left|X - E(X)\right| \ge \frac{1}{n}\right\} = 0 \quad (n=1,\ 2,\ \cdots)$

所以，$0 \le P\left\{\left|X - E(X)\right| \ne 0\right\} \le \sum_{n=1}^{\infty} 0 = 0$

所以，$P\left\{\left|X - E(X)\right| \ne 0\right\} = 0$ 因此，$P\left\{X = E(X)\right\} = 1$.

由此例及方差的性质，我们有：

$$P\{X = C\} = 1 \ (C\text{为常数})$$

的充分必要条件为 $D(X) = 0$.

设$X \sim N(\mu,\sigma)$

$$P\{|X-\mu|\le\sigma\}=P\{\mu-\sigma\le X\le\mu+\sigma\}$$

$$=\Phi(\frac{\mu+\sigma-\mu}{\sigma})-\Phi(\frac{\mu-\sigma-\mu}{\sigma})$$

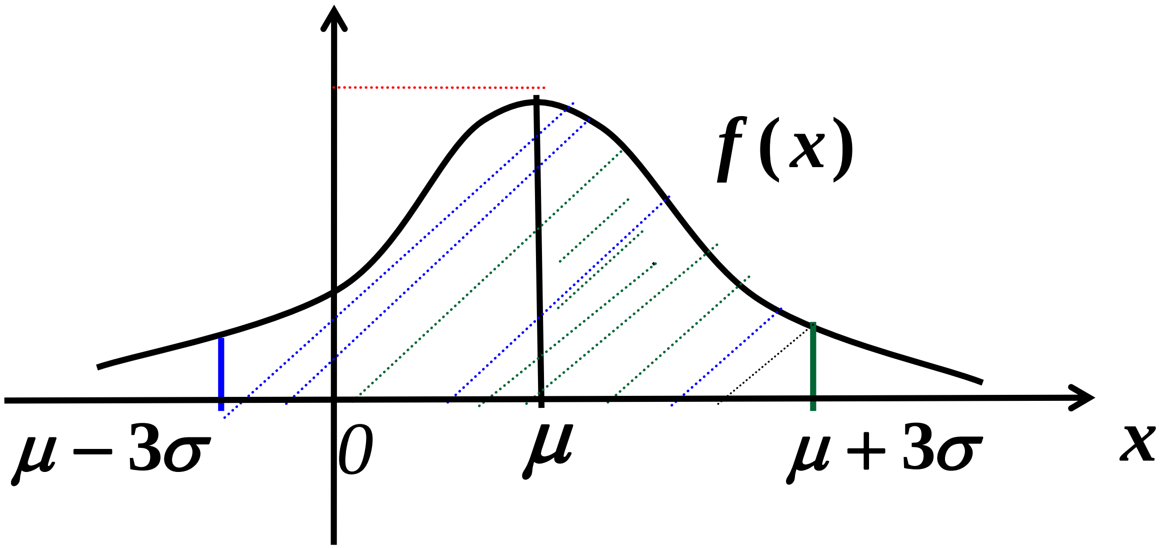
$$=\Phi(1)-\Phi(-1)=2\Phi(1)-1=0.6826$$

$$P\{|X-\mu|\le 2\sigma\}=P\{\mu-2\sigma\le X\le\mu+2\sigma\}$$

$$=2\Phi(2)-1=0.9544$$

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

$$=2\Phi(3)-1=0.9974$$



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

因此，对于正态随机变量X来说，它的值落在区间

$[\mu-3\sigma,\mu+3\sigma]$ 内几乎是肯定的。

注意：若用切比晓夫不等式估计概率有

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

小结：

1）方差的定义；

2）方差的性质；

3）切比晓夫不等式。

4）熟记两点分布、二项分布、泊松分布、均匀分布、正态分布的期望值和方差值。