第四章 随机变量的数字特征

第一节 数学期望

第二节 方差

第三节 协方差及相关系数

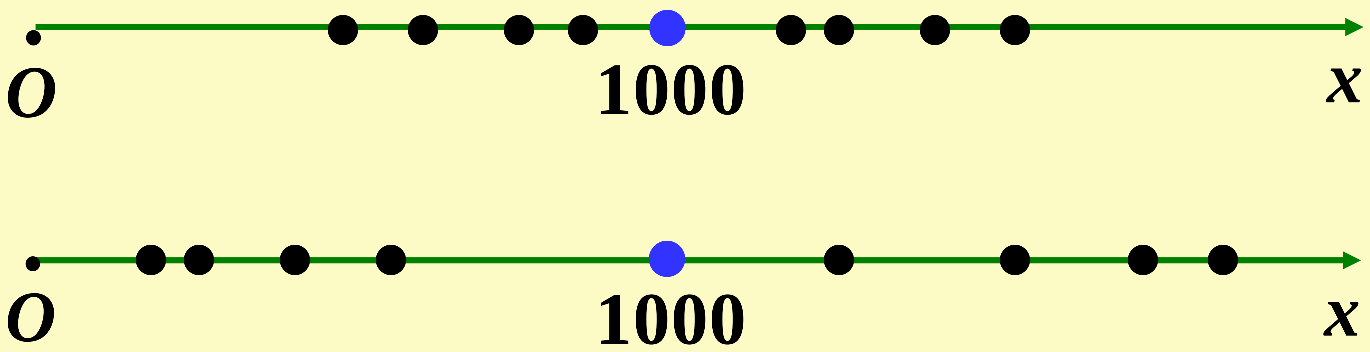
第四节 矩、协方差矩阵

§4.2 方差

概念的引入

方差是一个常用来体现随机变量取值分散程度的量.

实例 有两批灯泡,平均寿命都是 $E(X)=1000$小时.



1. 定义 设X是一随机变量，若$E\{[X-E(X)]^2\}$ 存在，则称为随机变量 X 的**方差**，记为$D(X)$或$\mathrm{Var}(X)$．即

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

并$\sqrt{D(X)}$ 称为X随机变量的**均方差**或**标准差**，记$\sigma(X)$.

1. 离散型: $$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k$$

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

3. 计算公式: $$\boxed{D(X)=E(X^2)-[E(X)]^2}$$

计算公式的推导：

由方差的定义及数学期望的性质，有

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

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

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

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

方差的意义

按定义，随机变量X的方差表达了X的取值与其数学期望的偏离程度. 若$D(X)$较小意味着X的取值比较集中在$E(X)$的附近，反之，若$D(X)$较大则表示X的取值较分散. 因此，$D(X)$是刻画X取值分散程度的一个量，它是衡量X取值分散程度的一个尺度.

例1 设甲、乙两射手在同样条件下进行射击，其命中环数分别用X、Y表示，分布律分别为

X	10	9	8	7
	0.5	0.1	0.2	0.2

Y	10	9	8	7
	0.4	0.3	0.1	0.2

试评定甲、乙的技术水平.

解 甲平均命中环数: $E(X)=10\times0.5+9\times0.1+8\times 0.2+7\times 0.2=8.9$ (环),

乙平均命中环数: $E(Y)=10\times0.4+9\times0.3+8\times 0.1+7\times 0.2=8.9$ (环),

故从平均水平看，甲、乙的技术水平不相上下，

进一步考虑他们射击的稳定性，即D(X), D(Y),

由于$E(X^2)=80.7$, $E(Y^2)=80.5$

于是 $D(X)= E(X^2)-[E(X)]^2 =1.49$

$D(Y)= E(Y^2)-[E(Y)]^2 =1.29$

所以，从稳定性来看，射手乙的技术水平略高于射手甲.

例2 设随机变量X具有概率密度

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

求D(X)

解: $$E(X)=\int_{-1}^{0}x(1+x)dx+\int_{0}^{1}x(1-x)dx=0$$

$$E(X^2)=\int_{-1}^{0}x^2(1+x)dx+\int_{0}^{1}x^2(1-x)dx=\frac{1}{6}$$

于是

$$D(X)=E(X^2)-[E(X)]^2=\frac{1}{6}$$

设随机变量X具有**(0 – 1)分布**，其分布率为

$$P\{X=1\}=p,\ P\{X=0\}=1-p$$

求$D(X)$.

解

$$E(X)=0\times(1-p)+1\times p=p$$

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

由公式

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

因此, 0 – 1分布

$$E(X)=p,\ D(X)=p(1-p)$$

二项分布

设随机变量$X \sim b(n, p)$, 其概率分布为:

$$P\{X=k\}=C_n^k p^k q^{n-k}, \quad k=0,1,2,\cdots,n, \quad q=1-p$$

则 $D(X) = E(X^2) - [E(X)]^2$, 其中

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

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

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

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

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

$$=n(n-1)p^2\sum_{l=0}^{n-2}\frac{(n-2)!}{l!(n-2-l)!}p^{l}q^{n-2-l}+np$$

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

所以

$$\begin{aligned} D(X) &= E(X^2)-[E(X)]^2 \\ &= n(n-1)p^2+np-n^2p^2 \\ &= np(1-p) \end{aligned}$$

泊松分布

设$X\sim\pi(\lambda)$，求$D(X)$。

解　X的分布律为

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

已算得

$$E(X)=\lambda,\text{而}$$

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

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

$$=\lambda^2 e^{-\lambda}e^{\lambda}+\lambda=\lambda^2+\lambda$$

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

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

因此, 泊松分布

$$E(X) = D(X) = \lambda$$

由此可知，泊松分布的 数学期望与方差相等，均为λ

均匀分布

设$X \sim U(a,b)$，求$D(X)$。

解　X的概率密度为

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

已求得$E(X)=\dfrac{a+b}{2}$。方差为

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

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

因此, 均匀分布

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

指数分布

例 设随机变量X服从指数分布,其概率密度为

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

其中$\theta>0$，求$D(X)$.

解

$$E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\theta \qquad E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)\mathrm{d}x=2\theta^2$$

因此$D(X)=\theta^2$.

由此可知, 指数分布

$$E(X)=\theta,\quad D(X)=\theta^2.$$

正态分布　　设$X\sim N(\mu,\sigma^2)$概率密度为：

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

$$E(X)=\mu$$

$$D(X)=E[X-E(X)]^2=\int_{-\infty}^{+\infty}(x-\mu)^2\cdot\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\mathrm{d}x$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}t^2\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t\quad(t=\frac{x-\mu}{\sigma})$$

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

服从正态分布的随机变量的分布完全由其数学期望和方差所确定.

2. 方差的性质 (假设下列方差均存在)

(1) $D(C)=0$, (C为常数)

(2) $D(CX)=C^2 D(X)$, $D(X+C)=D(X)$, (C为常数)

(3) $D(X\pm Y)=D(X)+D(Y)\pm 2E\{[X-E(X)][Y-E(Y)]\}$

特别：若X与Y相互独立，有

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

(4) $D(X)=0 \Leftrightarrow P\{X=C\}=1$,其中$C=E(X)$.

推广：设$X_1,\ldots,X_n$是n个相互独立的随机变量，则

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

1° 设 C 是常数，则 $D(C)=0.$

证 $D(C)=E\{[C-E(C)]^2\}=0.$

2° 设 X 是一个随机变量，C 是常数，则有

$$D(CX)=C^2D(X),\quad D(X+C)=D(X).$$

证 $$D(CX)=E\{[CX-E(CX)]^2\}$$
$$=C^2E\{[X-E(X)]^2\}$$
$$=C^2D(X).$$

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

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

$$= D(X).$$

3° 设 X,Y 是两个随机变量，则有

$$D(X+Y)$$

$$= D(X)+D(Y)+2E\{(X-E(X))(Y-E(Y))\}.$$

若 X,Y 相互独立，则有

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

证 $D(X+Y)$

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

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

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

$$+2E\{[X-E(X)][Y-E(Y)]\}$$

$$=D(X)+D(Y)+2E\{(X-E(X))(Y-E(Y))\}.$$

上式右端第三项:

$$2E\{(X-E(X))(Y-E(Y))\}$$

$$=2E\{XY-XE(Y)-YE(X)+E(X)E(Y)\}$$

$$= 2\{E(XY) - E(X)E(Y) - E(Y)E(X)$$

$$+ E(X)E(Y)\}$$

$$= 2\{E(XY) - E(X)E(Y)\}.$$

若 X,Y 相互独立,由数学期望的性质4°知道上式右端为0, 于是

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

这一性质可以推广到任意有限多个相互独立的随机变量之和的情况.

推广　若 $X_1, X_2, \cdots, X_n$ 相互独立, 则有

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

(4) $D(X) = 0$ 的充要条件是 X 以概率1取常数 C,

即　$\boxed{P\{X = C\} = 1.}$

证明：充分性利用定义证明，必要性利用切比雪夫不等式证明

切比雪夫不等式:

定理 设随机变量X的数学期望$E(X)=\mu$，方差$D(X)=\sigma^2$，则对任意的正数ε，有

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

--------切比雪夫(chebyshev)不等式.

证 （仅就X为连续型时来证）设X的概率密度为$f(x)$，则

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

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

[注] 此不等式给出了在随机变量的分布未知的情况下事件$\{|X-E(X)|<\varepsilon\}$的概率的一种估计方法。

$\varepsilon=3\sigma,4\sigma$ $P\{|X-\mu|<3\sigma\}\geq 0.8889$ $P\{|X-\mu|<4\sigma\}\geq 0.9375$

更一般的结论（马尔科夫不等式）：

设非负随机变量X的数学期望$E(X)$存在，则对任意的正数ε，有 $P\{Y \ge \varepsilon\} \le E(Y)/\varepsilon$

--------马尔科夫(Markov)不等式.

方差性质4°必要性的证明:

设 $D(X)=0$，要证 $P\{X=E(X)\}=1$.

证 用反证法 假设 $P\{X=E(X)\}<1$，则对于某一个数 $\varepsilon>0$，有 $P\{|X-E(X)|\ge\varepsilon\}>0$，但由切比雪夫不等式，对于任意 $\varepsilon>0$，因 $\sigma^2=0$，有

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

矛盾，于是 $P\{X=E(X)\}=1$.

例3 设$X_1, \ldots, X_n$相互独立，且服从同一参数为p的(0-1)分布，证明：$X = X_1 + \ldots + X_n$服从参数为n, p的二项分布，并求$E(X)$，$D(X)$.

证 X的所有可能取值为$0,1,\ldots,n$，$X=k$表示$X_1,\ldots,X_n$中有k个取1，$n-k$个取0，共有C_n^k种方式，故

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0,1,2,\cdots,n.$$

即X服从参数为n, p的二项分布.

$$E(X_i) = p, D(X_i) = p(1-p)$$

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

$$D(X) = D(X_1) + D(X_2) + \cdots + D(X_n) = np(1-p)$$

$$\boxed{E(X) = np, D(X) = np(1-p)}$$

例4一台设备由10个独立工作的元件组成，每一元件在时间T发生故障的概率为0.05．设在时间T发生故障的元件数为X，试用切比雪夫不等式估计随机变量X与其数学期望的偏差（a）小于2；(b)不小于2的概率．

解 （a）由题意知X~b(10, 0.05)，且

$$E(X)=0.5$$

$$D(X)=0.475$$

由切比雪夫不等式，得

$$P\{|X-0.5|<2\}\geq 0.88125$$

(b) $$P\{|X-0.5|\geq 2\}\leq 0.11875$$

例5 设随机变量X有期望$E(X)=\mu$, 方差$D(X)=\sigma^2\neq 0$.

记 $X^*=\dfrac{X-\mu}{\sigma}$ --------称为**X的标准化变量,**

则 $E(X^*)=0, D(X^*)=1$

证明: $E(X^*)=\dfrac{1}{\sigma}E(X-\mu)=\dfrac{1}{\sigma}[E(X)-\mu]=0;$

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

$$=E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]=\frac{1}{\sigma^2}E[(X-\mu)^2]=\frac{\sigma^2}{\sigma^2}=1.$$

例6 设 $X \sim N(\mu, \sigma^2)$, 求 $E(X)$, $D(X)$.

解 先求标准正态变量

$$Z = \frac{X - \mu}{\sigma}$$

的数学期望和方差. Z 的概率密度为

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2},$$

于是 $E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \mathrm{e}^{-t^2/2} \, \mathrm{d}t = \frac{-1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \Big|_{-\infty}^{\infty} = 0,$

$$D(Z) = E(Z^2)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \mathrm{e}^{-t^2/2} \, \mathrm{d}t$$

$$= \frac{-1}{\sqrt{2\pi}} t\mathrm{e}^{-t^2/2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-t^2/2} \, \mathrm{d}t$$

$$= 1,$$

因 $X = \mu + \sigma Z,$

即得 $E(X) = E(\mu + \sigma Z) = \mu,$

$$D(X) = D(\mu + \sigma Z) = D(\sigma Z) = \sigma^2 D(Z) = \sigma^2$$

推广: 若$X_i \sim N(\mu_i,\sigma_i^2), i=1,2,\cdots,n,$ 且它们相互独立,则它们的线性组合:$C_1X_1+C_2X_2+\cdots+C_nX_n$ ($C_1,C_2,\cdots,C_n$是不全为0的常数)仍然服从正态分布,于是由数学期望和方差的性质知道

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

例如，若 $X \sim N(1,3), Y \sim N(2,4)$ 且 X, Y 相互独立，则 $Z = 2X - 3Y$ 也服从正态分布，而

$$E(Z) = 2\times 1 - 3\times 2 = -4,$$

$$D(Z) = D(2X - 3Y) = 4D(X) + 9D(Y) = 48.$$

故有 $Z \sim N(-4,48)$.

例7 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.

解 按题意需求 $P\{X < Y\} = P\{X - Y < 0\}$.

由于 $X - Y \sim N(-0.10, 0.0025)$,

故有 $P\{X < Y\} = P\{X - Y < 0\}$

$$= P\left\{\frac{(X-Y)-(-0.10)}{\sqrt{0.0025}} < \frac{0-(-0.10)}{\sqrt{0.0025}}\right\} = \Phi\left(\frac{0.10}{0.05}\right)$$

$$= \Phi(2) = 0.9772.$$

例8 设 $E(X)=-2, D(X)=1$，$E(Y)=2, D(Y)=4$，且 X与Y独立，根据切比雪夫不等式估 $P\{|X+Y|\geq 5\}$.

答：$\leq 1/5$

例9 设随机变量X服从泊松分布，且

$$3P\{X=1\}+2P\{X=2\}=4P\{X=0\},$$

求X的数学期望与方差. **答：** $[\lambda=1, E(X)=D(X)=1]$

例10 设 $X\sim N(1,2)$，Y服从参数为 3 的泊松分布，且 X与Y独立，求 $D(XY)$.

答：27

几种重要随机变量的数学期望及方差

分　　布	参数	数学期望	方差
两点分布	$0 < p < 1$	p	$p(1-p)$
二项分布	$n \geq 1,$ $0 < p < 1$	np	$np(1-p)$
泊松分布	$\lambda > 0$	λ	λ
均匀分布	$a < b$	$(a+b)/2$	$(b-a)^2/12$
指数分布	$\theta > 0$	θ	θ^2
正态分布	$\mu, \sigma > 0$	μ	σ^2

（详见附表1）

小结

1. $D(X)$是刻画X取值分散程度的一个量,它是衡量X取值分散程度的一个尺度.若$D(X)$较小意味着X的取值比较集中在$E(X)$的附近,若$D(X)$较大则表示X的取值较分散.

2. 方差的计算公式

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

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

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

3. 方差的性质

1° $D(C)=0;$

2° $D(CX)=C^2D(X);$

3° $D(X\pm Y)=D(X)+D(Y).$

4.切比雪夫不等式

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

作业

- **第四章习题17, 18, 22(1), 24**