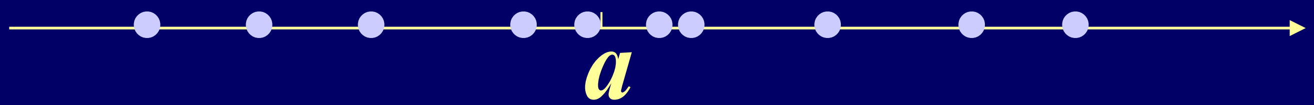
第二节　　方差

- 方差的定义
- 方差的计算
- 方差的性质
- 切比雪夫不等式
- 课堂练习
- 小结　布置作业

上一节我们介绍了随机变量的数学期望，它体现了随机变量取值的平均水平，是随机变量的一个重要的数字特征.

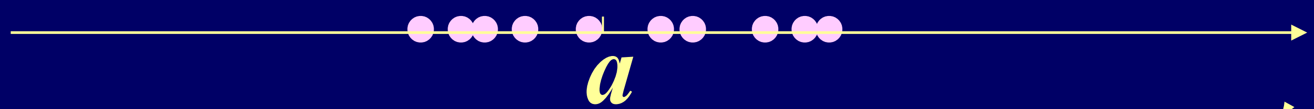
但是在一些场合，仅仅知道平均值是不够的.

例如，某零件的真实长度为a，现用甲、乙两台仪器各测量10次，将测量结果X用坐标上的点表示如图：



测量结果的均值都是 a

甲仪器测量结果



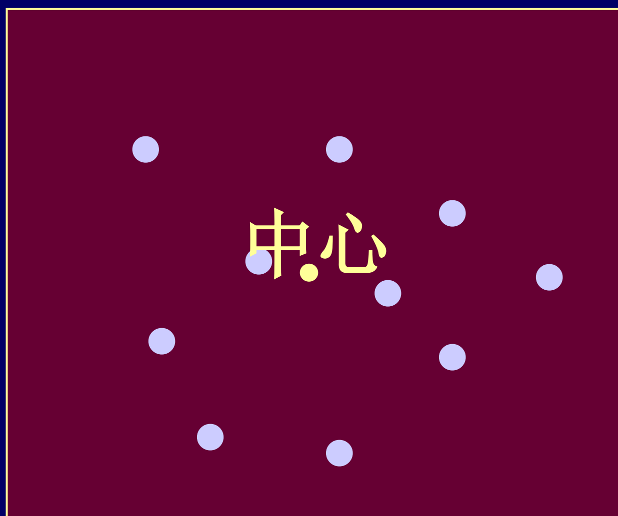
较好

乙仪器测量结果

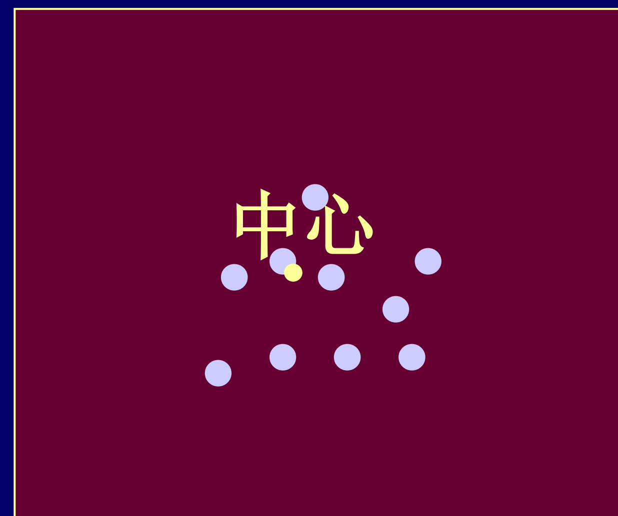
若让你就上述结果评价一下两台仪器的优劣，你认为哪台仪器好一些呢？

因为乙仪器的测量结果集中在均值附近

又如,甲、乙两门炮同时向一目标射击10发炮弹，其落点距目标的位置如图:



甲炮射击结果



乙炮

乙炮射击结果

你认为哪门炮射击效果好一些呢?

因为乙炮的弹着点较集中在中心附近.

由此可见,研究随机变量与其均值的偏离程度是十分必要的.那么,用怎样的量去度量这个偏离程度呢?容易看到

$$E\{|X - E(X)|\}$$

能度量随机变量与其均值E(X)的偏离程度. 但由于上式带有绝对值,运算不方便,通常用量

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

来度量随机变量X与其均值E(X)的偏离程度.

这个数字特征就是我们这一讲要介绍的

方差

一、方差的定义

设X是一个随机变量，若$E[(X-E(X)]^2$存在，称$E[(X-E(X)]^2$ 为 X 的方差. 记为D(X)或Var(X)，即

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

方差的算术平方根$\sqrt{D(X)}$称为X的标准差或均方差记为$\sigma(X)$，它与X具有相同的量纲。

方差刻划了随机变量的取值对于其数学期望的离散程度 .

若*X*的取值比较集中，则方差D(X)较小；

若*X*的取值比较分散，则方差D(X)较大.

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。

二、方差的计算

由定义知，方差是随机变量 X 的函数

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

的数学期望．

$$D(X)=\begin{cases}\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k, \\ \int_{-\infty}^{\infty}[x-E(X)]^2f(x)dx,\end{cases}$$

X为离散型，分布率 $P\{X=x_k\}=p_k$

X为连续型，X概率密度$f(x)$

计算方差的一个简化公式

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

展开

证：$D(X)=E[X-E(X)]^2$

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

利用期望性质

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

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

例1　设随机变量X具有(0—1）分布，其分布率为

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

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

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

例2 设$X \sim \pi(\lambda)$，求$D(X)$。

解 X的分布率为

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

上节已算得 $E(X)=\lambda$,而

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

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

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

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

因此,泊松分布

$$\boxed{E(X)=\lambda, D(X)=\lambda}$$

由此可知，泊松分布的数学期望与方差相等，等于λ。泊松分布的分布率中只含一个参数λ,只要知道λ,泊松分布就被确定了.

例3 设$X \sim U(a,b)$，求$D(X)$。

解 X的概率密度为 $$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & 其它\end{cases}$$

上节已求得$E(X)=\dfrac{a+b}{2}$。方差为

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

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

因此,均匀分布

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

例4 设随机变量X服从指数分布,其概率密度为

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

其中$\theta>0$，求$E(X)$，$D(X)$

解 $$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\theta$$

$$E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_{0}^{+\infty}x^2\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=2\theta^2$$

因此$D(X)=\theta^2$

由此可知,指数分布

$E（X）=\theta$，$D（X）=\theta^2$

三、方差的性质

1. 设C 是常数, 则 $D(C)=0$;

2. 若 C 是常数, 则 $D(CX)=C^2 D(X)$;

3. 设 X 与 Y 是两个随机变量，则

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

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

下面我们证明性质3

证明

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

若 X,Y 相互独立, 由数学期望的性质4得

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

此性质可以推广到有限多个相互独立的随机变量之和的情况.

下面我们举例说明方差性质的应用．

例6　设$X\sim B(n,p)$，求E(X)和D(X).

解　$X\sim B(n,p)$,　　则X表示n重努里试验中的"成功"次数．

若设 $$X_i=\begin{cases}1 & 如第i次试验成功\\ 0 & 如第i次试验失败\end{cases}$$ $i=1,2,\ \ldots,\ n$

则 $X=\sum_{i=1}^{n}X_i$ 是n次试验中"成功"的次数

可知X_i是$0-1$分布，所以

$E(X_i)=p,\ D(X_i)=p(1-p)$ ，$i=1,2$，…，n

由于$X_1, X_2, \ldots, X_n$ 相互独立

于是 $$E(X)=\sum_{i=1}^{n}E(X_i)=np$$

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

若$X \sim B(n,p)$,则

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

例7　设$X \sim N(0,1)$,求$E(X)$和$D(X)$.

解　X的概率密度为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad -\infty<x<+\infty$$

于是　$$E(X)=\int_{-\infty}^{+\infty}x\varphi(x)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}xe^{-\frac{x^2}{2}}dx=0$$

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

若$X \sim N(0,1)$,则　$E(X)=0, D(X)=1$

若$X \sim N(\mu,\sigma^2)$，则$Z=\dfrac{X-\mu}{\sigma} \sim N$（0,1）

$$E(Z)=0, D(Z)=1$$

而$X=\sigma Z+\mu$,由数学期望和方差的性质得

$$E(X)=E(\sigma Z+\mu)=\sigma E(Z)+E(\mu)=\mu$$

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

若$X \sim N(\mu,\sigma^2)$，则 $E(X)=\mu, D(X)=\sigma^2$

这就是说，正态分布的概率密度中的两个参数μ和σ^2分别是该分布的数学期望和方差，因而正态分布完全可由它的数学期望和方差所确定。

例如， 若$X \sim N(1,3), Y \sim N(2,4)$,且$X$和$Y$相互独立,

则$Z = 2X - 3Y$也服从正态分布，而$E(Z) = -4$,

$D(X) = 48$,故有$Z \sim N$（$-4,48$）

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.

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

例8 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和Y相互独立.任取一支活塞,任取一只气缸,求活塞能装入气缸的概率.

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

由于 $X - Y \sim N(-0.10, 0.0025)$

故有
$$P\{X < Y\} = P\{X - Y < 0\}$$
$$= P\{\frac{(X-Y)-(-0.10)}{\sqrt{0.0025}} < \frac{0-(-0.10)}{\sqrt{0.0025}}\}$$
$$= \Phi(\frac{0.10}{0.05}) = \Phi(2) = 0.9772$$

四、切比雪夫不等式

定理　设随机变量X具有数学期望$E(X)=\mu$,方差$D(X)=\sigma^2$,则对于任意正数ε，有不等式

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

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$\{|X-E(X)|<\varepsilon\}$的概率越大，即随机变量X集中在期望附近的可能性越大.

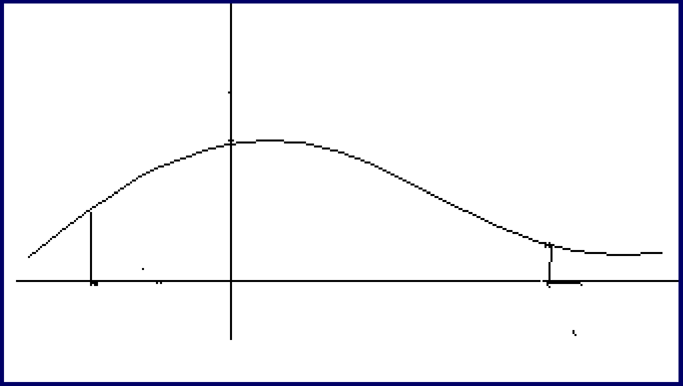
我们只就连续型随机变量的情况来证明.

证 设X的概率密度为$f(x)$，则有

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

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

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



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

当方差已知时，切比雪夫不等式给出了$r.v\ X$与它的期望的偏差不小于 ε 的概率的估计式 .

如取 $\varepsilon = 3\sigma$

$$P\{|X-E(X)|\geq 3\sigma\}\leq \frac{\sigma^2}{9\sigma^2}\approx 0.111$$

可见，对任给的分布，只要期望和方差 σ^2存在，则 $r.v\ X$取值偏离$E(X)$超过 3σ 的概率小于0.111 .

例9 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 $P(5200 \le X \le 9400)$

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

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

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

由切比雪夫不等式

$$P\{\,|X\text{-}E(X)|\le\ 2100\}\quad\ge 1-\frac{D(X)}{(2100)^2}$$

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

即估计每毫升白细胞数在5200~9400之间的概率不小于8/9 .

例10 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 $X \sim B(n, 0.75)$

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

所求为满足 $$P(0.74 < \frac{X}{n} < 0.76) \geq 0.90$$

的最小的n .

$P(0.74 < \frac{X}{n} < 0.76)$ 可改写为

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

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

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

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 .

五、课堂练习

1、 设随机变量X服从几何分布，概率分布为

$$P\{X=k\}=p(1-p)^{k-1},\ k=1,2,\ldots$$

其中$0<p<1$,求E(X),$D(X)$

2、 设随机变量X服从瑞利分布，其概率密度为：

$$f(x)=\begin{cases}\dfrac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} & x>0\\ 0 & x\le 0\end{cases}$$

其中$\sigma>0$是常数，求$E(X),D(X)$.

1、解：　记 q=1-p

$$E(X)=\sum_{k=1}^{\infty}kpq^{k-1}=p\sum_{k=1}^{\infty}(q^k)'$$

$$=p(\sum_{k=1}^{\infty}q^k)'=p(\frac{q}{1-q})'$$

$$=\frac{1}{p}$$

求和与求导

交换次序

无穷递缩等比

级数求和公式

$$E(X^2)=\sum_{k=1}^{\infty}k^2pq^{k-1}$$

$$=p[\sum_{k=1}^{\infty}k(k-1)q^{k-1}+\sum_{k=1}^{\infty}kq^{k-1}]$$

$$=qp(\sum_{k=1}^{\infty}q^k)''+E(X)=qp(\frac{q}{1-q})''+\frac{1}{p}$$

$$=qp\frac{2}{(1-q)^3}+\frac{1}{p}=\frac{2q}{p^2}+\frac{1}{p}=\frac{2-p}{p^2}$$

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

$$=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}$$

2、解 $E(X)=\int_{-\infty}^{+\infty} xf(x)dx$

$$=\int_{0}^{+\infty} x\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}dx=\sqrt{\frac{\pi}{2}}\sigma$$

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

$$=\int_{0}^{+\infty} x^2 f(x)dx-\frac{\pi}{2}\sigma^2$$

$$=\frac{4-\pi}{2}\sigma^2$$

六、小结

这一讲，我们介绍了随机变量的方差.

它是刻划随机变量取值在其中心附近离散程度的一个数字特征 .

下一讲，我们将介绍刻划两$r.v$间线性相关程度的一个重要的数字特征：

协方差、相关系数

七、 布置作业

《概率论与数理统计》作业（四）

一、填空题第2、3小题

二、选择题第3小题

三、解答题第1、5小题