NANJING UNIVERSITY

第四章 随机变量的数字特征

分布律、概率密度、分布函数各自都能完整地描述随机变量的数量规律，但实际应用中并不需要知道它们的具体表达，而只需知道随机变量的某些数字特征.

例如：

判断棉花质量时，既要看**纤维的平均长度**，又要看**纤维长度与平均长度的偏离程度**，平均长度越长，偏离程度越小，质量就越好.

再如：考察一射手的水平, 既要看他的**平均环数**是否高, 还要看他着弹点的范围是否小, 即**数据的波动**是否小.

由上面例子看到，**与随机变量有关的某些数值，**虽不能完整地描述随机变量但能清晰地描述随机变量在某些方面的重要特征, 这些数字特征在理论和实践上都具有重要意义.

随机变量某一方面的数量特征都可用**数值**来描述

随机变量的平均取值 —— 数学期望

随机变量取值平均偏离均值的情况

—— 方差

描述两随机变量间的某种关系的数值

—— 协方差与相关系数

主要内容

- 4.1 数学期望
- 4.2 方 差
- 4.3 协方差及相关系数
- 4.4 矩、协方差矩阵

4.1 数学期望

数学期望的定义

设 X 为离散随机变量，其分布律为：

$$P(X = x_k) = p_k, \quad k = 1, 2, \cdots$$

若无穷级数 $\sum_{k=1}^{+\infty} x_k p_k$ 绝对收敛, 则称

其为 X 的数学期望，记作 $E(X)$，即

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

数学期望的定义

设**连续型随机变量** X 的 概率密度为 $f(x)$

若广义积分

$$\int_{-\infty}^{+\infty} xf(x)dx$$

绝对收敛, 则称此积分为 X 的**数学期望**,

记作 $E(X)$, 即

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

数学期望的本质 —— 加权平均,
它是一个数, 不再是随机变量.

期望反映了随机变量取值的平均，故又称为**均值**！

二项分布的期望

例：$X \sim b(n, p)$，求 $E(X)$.

解：$$E(X)=\sum_{k=0}^{n} kC_n^k p^k (1-p)^{n-k}$$

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

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

特例：若 $Y \sim b(1, p)$，则 $E(Y)=p$.

泊松分布的期望

例： 设 $X \sim \pi(\lambda)$， 求$E(X)$.

解：

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

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

均匀分布的期望

例： 设 $X \sim U(a, b)$， 求$E(X)$.

解：

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

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

$$=\frac{a+b}{2}$$

正态分布的期望

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

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

$$\overset{令\frac{x-\mu}{\sigma}=u}{=}\int_{-\infty}^{+\infty}(u\sigma+\mu)\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=\mu$$

常见随机变量的数学期望

分布类型	概率分布	期望
参数为p 的0-1分布	$P(X=1)=p$ $P(X=0)=1-p$	p
$b(n,p)$	$P(X=k)=C_n^k p^k (1-p)^{n-k}$ $k=0,1,2,\cdots,n$	np
$\pi(\lambda)$	$P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}$ $k=0,1,2,\cdots$	λ

常见随机变量的数学期望

分布类型	概率密度	期望
$U(a,b)$	$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}\end{cases}$	$\dfrac{a+b}{2}$
$E(\theta)$	$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它}\end{cases}$	θ
$N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ

随机变量函数的数学期望

设离散型随机变量 X 的分布律为：

$$P(X = x_i) = p_i, \quad i = 1, 2, \cdots$$

随机变量$Y = g(X)$，$g(x)$是连续函数，

若无穷级数 $\sum_{i=1}^{\infty} g(x_i)p_i$ 绝对收敛，则

$$E(Y) = \sum_{i=1}^{\infty} g(x_i)p_i$$

随机变量函数的数学期望

设连续型随机变量X的概率密度为$f(x)$，

随机变量$Y = g(X)$，$g(x)$是连续函数，

若广义积分 $\int_{-\infty}^{\infty} g(x)f(x)dx$ 绝对收敛，则

$$E(Y)=\int_{-\infty}^{\infty} g(x)f(x)dx$$

随机变量函数的数学期望

设离散型随机变量 (X,Y) 的分布律为：

$$P(X = x_i, Y = y_j) = p_{ij}, \quad i, j = 1, 2, \cdots$$

随机变量$Z = g(X, Y)$, $g(x)$是连续函数，

若无穷级数 $\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} g(x_i, y_j)p_{ij}$ 绝对收敛，则

$$E(Z) = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} g(x_i, y_j)p_{ij}$$

随机变量函数的数学期望

设连续型随机变量(X, Y)的概率密度为$f(x,y)$,

随机变量$Z = g(X, Y)$, $g(x)$是连续函数,

若广义积分 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)dxdy$

绝对收敛, 则

$$E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)dxdy$$

例： 设风速V在(0, a)上服从均匀分布，即具有概率密度

$$f(v)=\begin{cases}\dfrac{1}{a}, & 0<v<a,\\ 0, & \text{其他}.\end{cases}$$

又设飞机机翼受到的正压力W是V的函数：$W=kV^2$，k为大于0的常数，求W的数学期望。

解：

$$E(W)=\int_{-\infty}^{\infty}kv^2 f(v)dv=\int_0^a kv^2\frac{1}{a}dv=\frac{1}{3}ka^2$$

例： 设随机变量（X，Y）的概率密度为

$$f(x,y)=\begin{cases}\dfrac{3}{2x^3y^2}, & \dfrac{1}{x}<y<x, x>1, \\ 0, & \text{其他}.\end{cases}$$

求数学期望$E(Y), E\left(\dfrac{1}{XY}\right)$.

解： $E(Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$

$$=\int_1^{\infty}\int_{\frac{1}{x}}^{x}\frac{3}{2x^3y}dydx$$

$$=\int_1^{\infty}\frac{3\ln x}{x^3}dx$$

$$=\int_1^{\infty}3\ln x d(\frac{x^{-2}}{-2})$$

$$=(-\frac{3\ln x}{2x^2})\Big|_1^{\infty}+\frac{3}{2}\int_1^{\infty}\frac{1}{x^3}dx \qquad \text{(分部积分法)}$$

$$=0+\frac{3}{4}=\frac{3}{4}$$

$$E(\frac{1}{XY})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{xy}f(x,y)dxdy$$

$$=\int_{1}^{\infty}\int_{\frac{1}{x}}^{x}\frac{3}{2x^4y^3}dydx$$

$$=\int_{1}^{\infty}-\frac{3}{4}(x^{-6}-x^{-2})dx$$

$$=-\frac{3}{4}(\frac{x^{-5}}{-5}-\frac{x^{-1}}{-1})\Big|_1^{\infty}$$

$$=-\frac{3}{4}(0-\frac{4}{5})=\frac{3}{5}$$

数学期望的性质

$$E(C)=C \quad (\text{C为常数})$$

$$E(CX)=CE(X)$$

$$E(X+Y)=E(X)+E(Y)$$

$$E\left(\sum_{i=1}^{n}a_iX_i+C\right)=\sum_{i=1}^{n}a_iE(X_i)+C$$

当X,Y独立时，$E(XY)=E(X)E(Y)$.

数学期望的性质

例：设一电路中电流$I(A)$与电阻$R(\Omega)$是两个相互独立的随机变量，其概率密度分别为

$$g(i)=\begin{cases}2i, & 0\le i\le 1,\\ 0, & \text{其他},\end{cases}\qquad h(r)=\begin{cases}\dfrac{r^2}{9}, & 0\le r\le 3,\\ 0, & \text{其他}.\end{cases}$$

试求电压$V=IR$的均值.

解：$E(V)=E(IR)=E(I)E(R)$

数学期望的性质

$$= [\int_{-\infty}^{\infty} ig(i)di][\int_{-\infty}^{\infty} rh(r)dr]$$

$$= (\int_0^1 2i^2 di)(\int_0^3 \frac{r^3}{9} dr)$$

$$= \frac{2}{3} \bullet \frac{9}{4}$$

$$= \frac{3}{2}$$

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

$\overline{\text{甲}} = 8.3$, $\overline{\text{乙}} = 8.3$

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

$+(7-8.3)^2+(6-8.3)^2]$

$=13.34/6=2.22 \triangleq \sum_{k=1}^{5}\left(x_k-E(X)\right)^2 p_k$

乙 $\frac{1}{6}[(10-8.3)^2+(9-8.3)^2$

$+3\times(8-8.3)^2+(7-8.3)^2]$

$=5.34/6=0.89 \triangleq \sum_{k=1}^{4}\left(x_k-E(X)\right)^2 p_k$

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

方差的定义

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**.

两者量纲相同

$D(X)$ —— 描述随机变量X的取值偏离平均值的平均偏离程度 —— **数值**

方差的计算

若 X 为离散型随机变量，分布律为

$$P(X=x_k)=p_k,\quad k=1,2,\cdots$$

则：$$DX=\sum_{k=1}^{+\infty}\left(x_k-E(X)\right)^2 p_k$$

若 X 为连续型随机变量，概率密度为 $f(x)$

则：$$DX=\int_{-\infty}^{+\infty}\left(x-E(X)\right)^2 f(x)dx$$

计算方差的常用公式：

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

例：设随机变量X服从两点分布，求$D(X)$.

解：

$X = x_k$	1	0
P_k	p	$1-p$

$$0 < p < 1$$

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

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

$$D(X) = E(X^2) - (EX)^2 = p - p^2 = p(1-p).$$

例：设随机变量$X\sim E(\theta)$，求$D(X)$.

解：

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

$\theta>0$ 为常数

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

$$=-\int_{0}^{+\infty}x\cdot e^{-x/\theta}d\left(-\frac{x}{\theta}\right)=-\int_{0}^{+\infty}x\cdot d\left(e^{-x/\theta}\right)$$

$$=-\left(xe^{-x/\theta}\Big|_{0}^{\infty}-\int_{0}^{+\infty}e^{-x/\theta}dx\right)=\int_{0}^{+\infty}e^{-x/\theta}dx$$

$$=-\theta\int_0^{+\infty} e^{-x/\theta} d\left(-\frac{x}{\theta}\right) = -\theta\left(e^{-x/\theta}\Big|_0^{\infty}\right) = -\theta(0-1) = \theta$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^{+\infty} x^2 \cdot \frac{1}{\theta} e^{-x/\theta} dx$$

$$= -\int_0^{+\infty} x^2 \cdot e^{-x/\theta} d\left(-\frac{x}{\theta}\right) = -\int_0^{+\infty} x^2 \cdot d\left(e^{-x/\theta}\right)$$

$$= -\left(x^2 e^{-x/\theta}\Big|_0^{\infty} - \int_0^{\infty} e^{-x/\theta} dx^2\right) = \int_0^{\infty} 2x e^{-x/\theta} dx$$

$$= 2\theta^2$$

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

常见随机变量的方差

分布类型	概率分布	方差
参数为p 的0-1分布	$P(X=1)=p$ $P(X=0)=1-p$	$p(1-p)$
$b(n,p)$	$P(X=k)=C_n^k p^k(1-p)^{n-k}$ $k=0,1,2,\cdots,n$	$np(1-p)$
$\pi(\lambda)$	$P(X=k)=\dfrac{\lambda^k e^{-\lambda}}{k!}$ $k=0,1,2,\cdots$	λ

分布类型	概率密度	方差
区间(a,b)上的均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a}, & a<x<b,\\ 0, & \text{其它}\end{cases}$	$\dfrac{(b-a)^2}{12}$
$E(\theta)$	$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它}\end{cases}$	θ^2
$N(\mu,\sigma^2)$	$f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	σ^2

方差的性质

$$\left.\begin{array}{l} D(C)=0 \ (\text{C为常数}) \\ \\ D(CX)=C^2D(X) \end{array}\right\} \Rightarrow D(CX+b)=C^2D(X)$$

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

特别地，若X,Y相互独立，则

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

仅知随机变量的期望与方差并不能确定其分布，也说明了期望与方差并不能全面描述随机变量。

例：

X	-1	0	1
P	0.1	0.8	0.1

$$E(X)=0, D(X)=0.2$$

与

Y	-2	0	2
P	0.025	0.95	0.025

$$E(Y)=0, D(Y)=0.2$$

有相同的期望方差，但是分布却不相同.

在已知某些分布类型时，若知道其期望和方差，便常能确定分布。

例： 已知 X 服从正态分布，$E(X) = 1.7$，$D(X) = 3$，求X的密度函数.

解： $\mu = 1.7$

$\sigma^2 = 3$

$$f(x) = \frac{1}{\sqrt{6\pi}} e^{-\frac{(x-1.7)^2}{6}}, \quad -\infty < x < +\infty$$

4.3 协方差及相关系数

协方差

对于二维随机变量(X,Y)，当它们不相互独立时：

已知联合分布 $\underset{\not\longleftarrow}{\longrightarrow}$ 边缘分布

此时表明X和Y之间存在某种联系。

问题：用一个怎样的指标去反映这种联系？

协方差

定义 称 $E\{[X-E(X)][Y-E(Y)]\}$

为 X,Y 的**协方差**. 记为:

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

相关系数

若$D(X)>0, D(Y)>0$，称

$$E\left(\frac{(X-E(X))(Y-E(Y))}{\sqrt{D(X)}\sqrt{D(Y)}}\right)=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为X,Y的 **相关系数**，记为

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \longrightarrow \text{无量纲的量}$$

若 $\rho_{XY}=0$，称 X,Y **不相关**.

协方差和相关系数的计算公式

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

$$\rho_{XY}=\frac{E(XY)-(EX)(EY)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

例： 已知 X,Y 的联合分布律为

$Y \backslash X$ (p_{ij})	1	0
1	p	0
0	0	q

$0 < p < 1$
$p + q = 1$

求 $\mathrm{Cov}(X,Y)$, ρ_{XY}

解：

X	1	0
P	p	q

Y	1	0
P	p	q

XY	1	0
P	p	q

$$\left.\begin{array}{l} E(X)=p,\ \ E(Y)=p, \\ D(X)=pq,\ \ D(Y)=pq, \\ E(XY)=p, \end{array}\right\} \Longrightarrow$$

$$Cov(X,Y)=pq,\ \ \rho_{XY}=1$$

例： 设二维随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}8xy, & 0\le x\le y, 0\le y\le 1,\\ 0, & \text{其他}\end{cases}$$

求X与Y的相关系数ρ_{XY}.

解： $$\rho_{XY}=\frac{E(XY)-(EX)(EY)}{\sqrt{DX}\sqrt{DY}}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\cdot f(x,y)\,dxdy$$

$$=\int_0^1\int_0^y 8x^2y^2\,dxdy$$

$$=\int_0^1 \frac{8y^5}{3}dy$$

$$=\frac{4}{9}$$

$$EX=\int_0^1\int_0^y x\cdot 8xy\,dxdy$$

$$=\int_0^1 \frac{8y^4}{3}dy$$

$$=\frac{8}{15}$$

$$EY = \int_0^1 \int_0^y y \cdot 8xy\,dx\,dy$$

$$= \int_0^1 4y^4\,dy$$

$$= \frac{4}{5}$$

$$DX = E(X^2) - (EX)^2$$

$$= \int_0^1 \int_0^y x^2 \cdot 8xy\,dx\,dy - \left(\frac{8}{15}\right)^2$$

$$= \frac{1}{3} - \left(\frac{8}{15}\right)^2$$

$$= \frac{11}{225}$$

$$DY = E(Y^2) - (EY)^2$$

$$= \int_0^1 \int_0^y y^2 \cdot 8xy\,dx\,dy - (\frac{4}{5})^2$$

$$= \frac{2}{3} - (\frac{4}{5})^2$$

$$= \frac{2}{75}$$

$$\rho_{XY} = \frac{E(XY) - (EX)(EY)}{\sqrt{DX}\sqrt{DY}}$$

$$= \frac{\frac{4}{9} - \frac{8}{15} \cdot \frac{4}{5}}{\sqrt{\frac{11}{225}}\sqrt{\frac{2}{75}}}$$

$$= \frac{2\sqrt{66}}{33}$$

协方差的性质

$$1.\ Cov(X,Y)=Cov(Y,X)$$

$$2.\ Cov(X,X)=DX$$

$$3.\ Cov(aX,bY)=abCov(X,Y)$$

$$4.\ Cov(X_1+X_2,Y)$$

$$=Cov(X_1,Y)+Cov(X_2,Y)$$

相关系数的性质

$1. |\rho_{XY}| \leq 1$

$2. |\rho_{XY}| = 1$ $\Longleftrightarrow$ Y 与 X 有线性关系的概率等于1

$3. \rho_{XY} = 0$ $\Longleftrightarrow$ X, Y 不相关

$\Longleftrightarrow$ $Cov(X, Y) = 0$

$\Longleftrightarrow$ $E(XY) = E(X)E(Y)$

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

X,Y 相互独立 $\underset{\not\Leftarrow}{\Rightarrow}$ X,Y 不相关

反例：设$X \sim N(0,1), Y = X^2$,求证相关系数$\rho = 0$.

证：易知 $EX = 0, DX = 1.$

因此$EY = EX^2 = DX + (EX)^2 = 1.$

$$Cov(X,Y) = E(XY) - EX \cdot EY$$

$$= EX^3 = 0$$

因此$\rho = 0.$

但X与Y显然不相互独立.

4. 4 矩、协方差矩阵

k阶原点矩

- 若$E(X^k)$, $k = 1, 2, 3, \ldots$ 存在，称它为X 的k阶原点矩。
- 当$k = 1$时，$E(X^k) = E(X)$, 因此，期望是1阶原点矩。

k阶中心矩

- 若$E\{[X\text{-}E(X)]^k\}$, $k = 2, 3, \ldots$ 存在，称它为X的k阶中心矩。
- 当$k = 2$时，$E\{[X\text{-}E(X)]^k\}=E\{[X\text{-}E(X)]^2\}=D(X)$, 因此，方差是2阶中心矩。

k+l 阶混合中心矩

- 若$E\{[X-E(X)]^k[Y-E(Y)]^l\}$, $k, l = 1, 2, 3, \ldots$ 存在，称它为X 和Y 的k+l 阶混合中心矩。
- 当$k = l = 1$时，$E\{[X-E(X)]^k[Y-E(Y)]^l\} =$
  $E\{[X-E(X)][Y-E(Y)]\}=Cov(X, Y)$, 因此，
  协方差是2阶混合中心矩。

二维随机变量的协方差矩阵

$$\begin{array}{cc} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{pmatrix} Cov(X,X) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{pmatrix} \end{array}$$

即

$$\begin{array}{cc} & \begin{array}{cc} X & Y \end{array} \\ \begin{array}{c} X \\ Y \end{array} & \begin{pmatrix} D(X) & Cov(X,Y) \\ Cov(Y,X) & D(Y) \end{pmatrix} \end{array}$$

THE END