Chap4.随机变量的数字特征

36.随机变量的数字特征

离散型随机变量的数学期望

定义：设离散型随机变量 X 的分布律为

$$P\{X=x_k\}=p_k, \quad k=1,2,\ldots \tag{1}$$

若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛，则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 X 的数学期望，记为 $E(X)$，即

$$E(X)=\sum_{k=1}^{\infty} x_k p_k \tag{2}$$

连续型随机变量的数学期望

定义：设连续型随机变量 X 的概率密度为 $f(x)$，若积分 $\int_{-\infty}^{\infty} x f(x)\,dx$绝对收敛，则称积分的值 $\int_{-\infty}^{\infty} x f(x)\,dx$为随机变量 X 的数学期望，记为 $E(X)$。即

$$E(X)=\int_{-\infty}^{\infty} x f(x)\,dx \tag{3}$$

37.随机变量函数的数学期望

一维随机变量的函数的数学期望

定义 设 Y 是随机变量 X 的函数：$Y=g(X)$ （g 是连续函数）

- 若 X 是离散型随机变量，其分布律为 $P\{X=x_k\}=p_k, \quad k=1,2,\ldots$，若级数 $\sum_{k=1}^{\infty} g(x_k)p_k$ 绝对收敛，则有

$$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k. \tag{4}$$

- 若 X 是连续型随机变量，其概率密度为 $f(x)$，若 $\int_{-\infty}^{\infty} g(x)f(x)dx$ 绝对收敛，则有

$$E(Y)=E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)dx. \tag{5}$$

1. 设 $X \sim \pi(\lambda)$，求 $E[\frac{1}{X+1}]$。

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,2,\ldots \tag{6}$$

$$E\left[\frac{1}{X+1}\right]=\sum_{k=0}^{\infty}\frac{1}{k+1}\cdot\frac{\lambda^k e^{-\lambda}}{k!} \tag{7}$$

$$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{(k+1)!} \tag{8}$$

$$=e^{-\lambda}\cdot\frac{1}{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!} \tag{9}$$

$$=e^{-\lambda}\cdot\frac{1}{\lambda}\cdot\left[\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}-1\right] \tag{10}$$

$$= e^{-\lambda} \cdot \frac{1}{\lambda} \cdot \left[e^{\lambda} - 1\right] \tag{11}$$

$$= \frac{1 - e^{-\lambda}}{\lambda}. \tag{12}$$

2. 设风速 V 在 $(0, a)$ 上服从均匀分布，即具有概率密度

$$f(v) = \begin{cases} \frac{1}{a}, & 0 < v < a \\ 0, & \text{其他} \end{cases} \tag{13}$$

又设风速 V 受到了正压力 W 的作用，其中 W 是 V 的函数：$W = kV^2, \quad (k > 0$ 常数)，求 W 的数学期望。

解：$E(W) = E[kV^2] = \int_0^a kv^2 \cdot f(v)\, dv$

$$= \int_0^a kv^2 \cdot \frac{1}{a}\, dv = \frac{k}{a} \int_0^a v^2\, dv \tag{14}$$

$$= \frac{k}{a} \left[\frac{v^3}{3}\right]_0^a = \frac{k}{a} \cdot \frac{a^3}{3} = \frac{1}{3} ka^2 \tag{15}$$

二维随机变量的函数的数学期望

定义 设 (X, Y) 是二维随机变量，$Z = g(X, Y)$（g 是连续函数），则 $Z = g(X, Y)$ 是一个随机变量，并且

- 若 (X, Y) 是离散型随机变量，其分布律为

$$P\{X = x_i, Y = y_j\} = p_{ij}, \quad i, j = 1, 2, \ldots \tag{16}$$

则有

$$E(Z) = E[g(X, Y)] = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} g(x_i, y_j) p_{ij} \tag{17}$$

假设式右端的级数绝对收敛。

- 若 (X, Y) 是连续型随机变量，其概率密度为 $f(x, y)$，则有

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy \tag{18}$$

假设式右端的积分绝对收敛。

设随机变量 (X, Y) 的概率密度为

$$f(x, y) = \begin{cases} \frac{3}{2x^3} y^2, & \frac{1}{x} < y < x, \quad x > 1 \\ 0, & \text{其他} \end{cases} \tag{19}$$

求数学期望 $E(Y)$ 和 $E\left(\frac{1}{XY}\right)$。

1. 设 $Y = g(X, Y)$，则 $g(X, Y) = y$。

$$E(Y) = E[g(X, Y)] = \int_1^{\infty} \int_{\frac{1}{x}}^{x} y f(x, y)\, dy\, dx \tag{20}$$

$$= \int_1^{\infty} \frac{3}{2x^3} \int_{\frac{1}{x}}^{x} y^2 \, dy \, dx \tag{21}$$

$$= \frac{3}{2} \int_1^{\infty} \frac{1}{x^3} \left[\int_{\frac{1}{x}}^{x} y^2 \, dy \right] dx \tag{22}$$

继续计算得到：

$$= \frac{3}{2} \int_1^{\infty} \frac{1}{x^3} [\ln x] dx = \frac{3}{4} \tag{23}$$

2. 对于 $E\left(\frac{1}{XY}\right)$，我们有

$$E\left(\frac{1}{XY}\right) = E[g(X,Y)] = \int_1^{\infty} \int_{\frac{1}{x}}^{x} \frac{1}{xy} \cdot \frac{3}{2x^4} y^3 \, dy \, dx = \frac{3}{5} \tag{24}$$

38.数学期望的性质

(1) 设 C 是常数，则 $E(C) = C$。

> 常数是没有随机性的，因此无论进行多少次试验，其期望值始终是该常数本身。

(2) 设 X 是一个随机变量，C 是常数，则 $E(CX) = CE(X)$。

> 常数因子可以从期望值运算中提取出来，这意味着随机变量 X 的期望被常数 C 缩放。

(3) 设 X, Y 是两个随机变量，则

$$E(X+Y) = E(X) + E(Y) \tag{25}$$

推广：

$$E(X_1 + X_2 + X_3 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) \tag{26}$$

> 两个随机变量的和的期望是各自期望的和。这说明期望值运算是**线性**的。
>
> 多个随机变量之和的期望是它们各自期望的和，这同样适用于多个随机变量的情况。

(4) 设 X, Y 是相互独立的随机变量，$E(XY) = E(X)E(Y)$。

推广：

$$X_1, X_2, \ldots, X_n \text{ 相互独立时}, E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2)\cdots E(X_n) \tag{27}$$

> 对于两个独立的随机变量，它们的积的期望值等于它们期望值的乘积。
>
> 独立性意味着一个随机变量的结果不影响另一个变量的结果，因此它们的期望值可以直接相乘。

39.方差

方差的定义

定义：设 X 是一个随机变量，若 $E\left\{[X - E(X)]^2\right\}$ 存在，则称 $E\left\{[X - E(X)]^2\right\}$ 为 X 的方差，记为 $D(X)$，或 $Var(X)$，即

$$D(X) = Var(X) = E\left\{[X - E(X)]^2\right\}. \tag{28}$$

引入：$\sigma(X) = \sqrt{D(X)}$，称为标准差或均方差。

方差 $Var(X)$ 描述了 X 的取值与其数学期望的偏离程度，$Var(X)$ 越小， X 的取值集中在 $E[X]$ 附近，$Var(X)$ 越大，X 的取值越分散。

方差的计算

对离散型随机变量 X：$P\{X=x_k\}=p_k, k=1,2,\ldots$，则

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2p_k=E(X^2)-[E(X)]^2 \tag{29}$$

计算步骤：

1. 先计算 $E(X)$。
2. 计算 X^2 的期望，即计算 $E(X^2)$。
3. 代入公式：$D(X)=E(X^2)-[E(X)]^2$。

对连续型随机变量 X：概率密度为 $f(x)$，则

$$D(X)=\int_{-\infty}^{\infty}[x-E(X)]^2f(x)\,dx=E(X^2)-[E(X)]^2 \tag{30}$$

$$\int_{-\infty}^{\infty}x^2f(x)\,dx-\left[\int_{-\infty}^{\infty}xf(x)\,dx\right]^2 \tag{31}$$

计算步骤：

1. $E(X)=\int_{-\infty}^{\infty}xf(x)\,dx$。
2. $E(X^2)=\int_{-\infty}^{\infty}x^2f(x)\,dx$。
3. $D(X)=E(X^2)-[E(X)]^2$。

设随机变量 X 具有数学期望 $E(X)=\mu$，$D(X)=\sigma^2$，记 $X^*=\frac{X-\mu}{\sigma}$，求 $E(X^*)$，$D(X^*)$

$$X^*=\frac{X-\mu}{\sigma}\quad \text{称为 } X \text{ 的标准化变量.} \tag{32}$$

$$E(X^*)=E\left(\frac{X-\mu}{\sigma}\right)=\frac{1}{\sigma}E(X-\mu)=\frac{1}{\sigma}[E(X)-\mu]=0 \tag{33}$$

$$D(X^*)=E(X^{*2})-[E(X^*)]^2=E\left(\frac{(X-\mu)^2}{\sigma^2}\right)=\frac{1}{\sigma^2}E[(X-E(X))^2] \tag{34}$$

$$=\frac{1}{\sigma^2}E\left[(X-\mu)^2\right]=\frac{1}{\sigma^2}\cdot\sigma^2=1 \tag{35}$$

40.方差的性质

(1) 设 C 是常数，则 $D(C)=0$。

$$D(C)=E\left\{[C-E(C)]^2\right\}=E(0)=0. \tag{36}$$

> 常数是没有随机性的，因此其偏离均值的程度为零，方差为零，即常数的方差为零：

(2) 设 X 是一个随机变量，C 是常数，则

$$D(CX)=C^2D(X), \quad D(C+X)=D(X). \tag{37}$$

> 当随机变量与 C 相乘时，方差会被常数 C^2 放大，即 $D(CX)$ 等于 C^2 乘以 X 的方差
>
> 当随机变量 X 加上常数 C 时，常数不影响方差，所以 $D(C+X)=D(X)$

(3) 设 X,Y 是两个随机变量，则

$$D(X+Y)=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\} \tag{38}$$

$$=D(X)+D(Y)+2\left(E(XY)-E(X)E(Y)\right) \tag{39}$$

特别地，若 X,Y 相互独立，则

$$D(X+Y)=D(X)+D(Y) \tag{40}$$

> 当我们求两个随机变量和的方差时，不仅考虑它们各自的方差 $D(X)$ 和 $D(Y)$，还需要考虑它们之间的协方差，这个部分描述了 X 和 Y 之间的线性关系。

推导：

$$D(X+Y)=E\left\{[X+Y-E(X+Y)]^2\right\}=E\left\{[X-E(X)+Y-E(Y)]^2\right\} \tag{41}$$

$$=E\left\{[X-E(X)]^2+[Y-E(Y)]^2+2[X-E(X)][Y-E(Y)]\right\} \tag{42}$$

$$D(X+Y)=D(X)+D(Y)+2E\{[X-E(X)][Y-E(Y)]\} \tag{43}$$

$$E\{[X-E(X)][Y-E(Y)]\}=E[XY-XE(Y)-YE(X)+E(X)E(Y)]=E(XY)-E(X)E(Y) \tag{44}$$

因此，

$$D(X+Y)=D(X)+D(Y)+2\left[E(XY)-E(X)E(Y)\right] \tag{45}$$

若 X,Y 相互独立，则 $E(XY)=E(X)E(Y)$，所以

$$D(X+Y)=D(X)+D(Y) \tag{46}$$

推论：

- 若 $X_1,X_2,\ldots,X_n$ 相互独立，则

$$D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n) \tag{47}$$

- 若 X,Y,Z 相互独立，则

$$D(aX+bY+cZ)=a^2D(X)+b^2D(Y)+c^2D(Z) \tag{48}$$

(4) $D(X)=0$ 的充要条件是 X 以概率 1 取得常数 $E(X)$，即 $P\{X=E(X)\}=1$。

$$D(X)=E\left\{[X-E(X)]^2\right\}=0. \tag{49}$$

近似：

$$X=E(X) \tag{50}$$

> 当随机变量 X 的方差为零时，说明 X 的值总是等于其期望值，即没有波动。这意味着 X 以概率 1 始终等于常数 $E(X)$。

41.切比雪夫不等式

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：

$$P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}. \tag{51}$$

证明：设 X 为连续型，$f(x)$ 为概率密度。

$$P\{|X-\mu|\geq\varepsilon\}=\int_{|x-\mu|\geq\varepsilon}f(x)\,dx \tag{52}$$

$$=\frac{1}{\varepsilon^2}\int_{|x-\mu|\geq\varepsilon}(x-\mu)^2f(x)\,dx \tag{53}$$

$$\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2f(x)\,dx=\frac{D(X)}{\varepsilon^2}=\frac{\sigma^2}{\varepsilon^2}. \tag{54}$$

切比雪夫不等式给出了一个概率界限，他的意思是：

无论数据的分布如何（即不依赖于数据是否服从正态分布或其他特定分布），如果知道了均值和方差，它就能够估计出数据点偏离均值的概率上界。

42.协方差及相关系数

称量 $E\{[X-E(X)][Y-E(Y)]\}$ 为随机变量 X 与 Y 的协方差，记为 $\mathrm{Cov}(X,Y)$，即

$$\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\} \tag{55}$$

$$=E(XY)-E(X)E(Y) \tag{56}$$

而

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{57}$$

称为随机变量 X 与 Y 的相关系数。

1. $\mathrm{Cov}(X,Y)$ 有量纲（X: kg, Y: m），则 $\mathrm{Cov}(X,Y)$ 也有量纲。
   - 要求一个数值时，需通过描述 X 与 Y 之间某种相关关系，才能将其换算成量纲。
2. ρ_{XY}：标准化后的协方差，是无量纲的量，可反映 X 与 Y 之间的关系。

$\mathrm{Cov}(X,Y)$，ρ_{XY} 到底描述了 X 与 Y 之间什么关系?

1. $\mathrm{Cov}(X,X)=D(X)$

$$\mathrm{Cov}(X,X)=E[(X-E(X))(X-E(X))] \tag{58}$$

$$=E[(X^2)-(E(X))^2]=D(X). \tag{59}$$

> 随机变量 X 与自身的协方差等于 X 的方差

2. $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$; $\mathrm{Cov}(X,c)=0$

$$\operatorname{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))] \quad\Rightarrow\quad \operatorname{Cov}(X,Y)=\operatorname{Cov}(Y,X). \tag{60}$$

$$\operatorname{Cov}(X,c)=E[(X-E(X))(c-E(c))]=c\cdot E(X)-c\cdot E(X)=0. \tag{61}$$

> 协方差是对称的，即随机变量 X 与 Y 之间的协方差与 Y 与 X 之间的协方差是相同的。
>
> 随机变量 X 与常数 c 之间的协方差为 0

3. $\operatorname{Cov}(aX,bY)=ab\operatorname{Cov}(X,Y)$

> 如果对随机变量 X 和 Y 进行线性变换，协方差也会按比例变化。

4. $\operatorname{Cov}(X_1+X_2,Y)=\operatorname{Cov}(X_1,Y)+\operatorname{Cov}(X_2,Y)$

> 多个随机变量之和与另一个随机变量之间的协方差，可以拆解成各个随机变量与这个变量之间协方差的和。

5. $D(X+Y)=D(X)+D(Y)+2\operatorname{Cov}(X,Y)$

> 两个随机变量 X 和 Y，那么它们的和 $X+Y$ 的方差等于各自方差的和，再加上它们之间协方差的两倍。

相关系数 ρ_{XY} 的性质

(1) $|\rho_{XY}|\le 1$；　相关系数的绝对值最大为 1。

(2) $|\rho_{XY}|=1$ 的充要条件是，存在常数 a,b 使得

$$P\{Y=a+bX\}=1,\quad b>0. \tag{62}$$

当相关系数为 1 或 -1 时，两个随机变量之间有完全的线性关系。

不相关与相互独立的关系

不相关指的是，没有线性关系，所以

相互独立一定没有线性关系，相互独立能够推出不相关

不相关只是没有线性关系，并不代表不相关

唯一的特殊情况是二维正态分布，如果服从二维正态分布，不相关等价于相互独立

X,Y 重要关系的判断

X,Y 相互独立 $\iff$

$F(x,y)=F_X(x)\cdot F_Y(y),\quad \forall x,y.$

$\iff\ P_{ij}=P_i\cdot P_j,\quad \forall i,j.$

$\iff\ f(x,y)=f_X(x)\cdot f_Y(y),\quad \forall x,y.$

X,Y 不相关 $\iff$

$\rho_{XY} = 0$
$\iff \mathrm{Cov}(X,Y) = 0$
$\iff E(XY) = E(X)E(Y)$
$\iff D(X+Y) = D(X) + D(Y)$

43.矩、协方差矩阵

k 阶矩

设 X 是随机变量，若 $E(X^k),\quad k=1,2,\ldots$ 存在，称它为 X 的 k 阶原点矩，简称 k 阶矩。
注：X 的数学期望 $E(X)$ 是 X 的一阶原点矩。

k 阶中心矩

设 X 是随机变量，若 $E\{[X-E(X)]^k\},\quad k=2,3,\ldots$ 存在，称它为 X 的 k 阶中心矩。
注：X 的方差 $D(X)$ 是 X 的二阶中心矩。
$k=2$ 时，特定公式为 $E\{[X-E(X)]^2\} = D(X).$

一般而言，
$X \sim f(x):\quad f(x) = \int_{-\infty}^{x}[x-E(X)]^k dx.$

$k+l$ 阶混合矩

设 (X,Y) 是二维随机变量，若 $E(X^kY^l),\quad k,l=1,2,\ldots$ 存在，称它为 X 和 Y 的 $k+l$ 阶混合矩。

$k+l$ 阶混合中心矩

设 (X,Y) 是二维随机变量，若 $E\{[X-E(X)]^k[Y-E(Y)]^l\},\quad k,l=1,2,\ldots$存在，称它为 X 和 Y 的 $k+l$ 阶混合中心矩。

注：协方差 $\mathrm{Cov}(X,Y)$ 是 X 和 Y 的二阶混合中心矩。

$k=1, l=1$ 时，X 和 Y 的二阶混合中心矩为

$$E\{[X-E(X)][Y-E(Y)]\} = \mathrm{Cov}(X,Y). \tag{63}$$

二维随机变量的协方差矩阵

设二维随机变量 (X_1,X_2) 有两个二阶中心矩，分别记为
$c_{11} = E\{(X_1-E(X_1))^2\} = D(X_1) = \mathrm{Cov}(X_1,X_1),$
$c_{12} = E\{(X_1-E(X_1))(X_2-E(X_2))\} = \mathrm{Cov}(X_1,X_2),$
$c_{21} = E\{(X_2-E(X_2))(X_1-E(X_1))\} = \mathrm{Cov}(X_2,X_1),$
$c_{22} = E\{(X_2-E(X_2))^2\} = D(X_2) = \mathrm{Cov}(X_2,X_2).$

将它们排成矩阵的形式

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) \end{pmatrix}. \tag{64}$$

这个矩阵就称为随机变量 (X_1, X_2) 的协方差矩阵。

n 维随机变量 $(X_1, X_2, \ldots, X_n)$ 的协方差矩阵

设 n 维随机变量 $(X_1, X_2, \ldots, X_n)$ 的二阶混合中心矩
$c_{ij} = \mathrm{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}, \quad i, j = 1, 2, \ldots, n$
都存在，则称矩阵为协方差矩阵。
为设 n 维随机变量 $(X_1, X_2, \ldots, X_n)$ 的协方差矩阵：

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{pmatrix}. \tag{65}$$

注：由于 $c_{ij} = c_{ji}$，故上述矩阵是一个对称矩阵。
一般来说，n 维随机变量的分布是不知道的，或者很复杂，以致于在数学上不易处理，因此在实际应用中协方差矩阵就显得尤为重要。

n 维正态随机变量的四条重要性质

n 维正态随机变量 $(X_1, X_2, \ldots, X_n)$ 的每一个分量 $X_i, i = 1, 2, \ldots, n$ 都是正态随机变量；

反之，若 $X_1, X_2, \ldots, X_n$ 都是正态随机变量，且相互独立，则 $(X_1, X_2, \ldots, X_n)$ 是 n 维正态随机变量。

例子：
$X_1 \sim N(1, 2^2), X_2 \sim N(-1, 3^2), X_3 \sim N(0, 5^2)$
X_1, X_2, X_3 独立 $\Rightarrow (X_1, X_2, X_3) \sim 3$维正态分布

n 维随机变量 $(X_1, X_2, \ldots, X_n)$ 服务从 n 维正态分布的充要条件是 $X_1, X_2, \ldots, X_n$ 的任意线性组合
$l_1X_1 + l_2X_2 + \cdots + l_nX_n$ 服从一维正态分布（其中 $l_1, l_2, \ldots, l_n$ 不全为零）。

若 $(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布，设 $Y_1, Y_2, \ldots, Y_k$ 是 $X_j\,(j = 1, 2, \ldots, n)$ 的线性函数，则 $(Y_1, Y_2, \ldots, Y_k)$ 也服从多维正态分布。

例子：
若 (X_1, X_2, X_3) 服从三维正态分布：
$Y_1 = 2X_1 - X_2 + X_3$
$Y_2 = -X_1 + X_2 + 3X_3$
则 (Y_1, Y_2) 服从二维正态分布。

设 $(X_1, X_2, \ldots, X_n)$ 服从 n 维正态分布，则 "$X_1, X_2, \ldots, X_n$ 相互独立" 与 "$X_1, X_2, \ldots, X_n$ 两两不相关" 是等价的。