

第6章 多维随机向量的数字特征

6.1 多维随机向量函数的期望

前面介绍了一维随机变量及其函数的期望, 下面研究二维随机向量函数的期望, 同理可推广到维度更多的随机变量.

定理 6.1 设二维离散型随机向量 (X, Y) 的分布列为 $p_{ij} = P(X = x_i, Y = y_j)$, 则随机向量函数 $g(X, Y)$ 的期望为

$$E[g(X, Y)] = \sum_{i,j} g(x_i, y_j) p_{ij} ;$$

设二维连续型随机向量 (X, Y) 的密度函数为 $f(x, y)$, 则随机向量函数 $g(X, Y)$ 的期望为

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

定理的证明超出了本书的范畴而被略去. 当 $g(X, Y) = X$ 时, 对二维连续型随机变量有

$$E[X] = \int_{-\infty}^{+\infty} x dx \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{+\infty} x f_X(x) dx ,$$

随机向量 (X, Y) 中 X 的期望就是其边缘分布的期望, 同理可得 Y 的期望, 对离散情况结论也成立.

当 $g(X, Y) = (X - E(X))^2$ 时有

$$E[g(X, Y)] = \int_{-\infty}^{+\infty} (x - E(X))^2 dx \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx = \text{Var}(X) ,$$

X 的方差就是边缘分布的方差, 同理可得 Y 的方差. 根据期望的定义有

性质 6.1 若随机变量 X, Y 满足 $X \geq Y$, 则有 $E[X] \geq E[Y]$.

性质 6.2 对任意随机变量 X, Y 有 $E[X + Y] = E[X] + E[Y]$.

证明 这里仅仅给出连续情况的证明, 同理可以考虑类型随机变量.

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f(x, y) dx dy = \int_{-\infty}^{+\infty} x dx \int_{-\infty}^{+\infty} f(x, y) dy \\ &\quad + \int_{-\infty}^{+\infty} y dy \int_{-\infty}^{+\infty} f(x, y) dx = \int_{-\infty}^{+\infty} x f_X(x) dx + \int_{-\infty}^{+\infty} y f_Y(y) dy = E(X) + E(Y) , \end{aligned}$$

由此完成证明.

性质 6.2 可推广到 n 个随机变量, 即 $E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$.

性质 6.3 对相互独立的随机变量 X 和 Y , 有

$$E[XY] = E[X]E[Y];$$

对任意随机变量 X 和 Y , 有 Cauchy-Schwartz 不等式

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}.$$

证明 这里仅仅给出连续情况的证明, 同理可以考虑类型随机变量. 根据随机变量 X 与 Y 的独立性有 $f(x, y) = f_X(x)f_Y(y)$, 由此可得

$$\begin{aligned} E[XY] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf_X(x)yf_Y(y)dxdy \\ &= \int_{-\infty}^{+\infty} xf_X(x)dx \int_{-\infty}^{+\infty} yf_Y(y)dy = E(X)E(Y). \end{aligned}$$

对任意随机变量 X 和 Y , 以及对任意 $t \in \mathbb{R}$ 有 $E[(X + tY)^2] \geq 0$, 即任意 $t \in \mathbb{R}$,

$$t^2E[Y^2] + E[X^2] + 2tE[XY] \geq 0.$$

因此有 $\Delta = 4[E(XY)]^2 - 4E[X^2]E[Y^2] \leq 0$, 即 $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$.

根据性质 6.3, 若随机变量 X 和 Y 相互独立, 则有

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)];$$

若随机变量 $X_1, X_2, \cdots, X_n$ 相互独立, 则有

$$E[X_1X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n].$$

性质 6.4 设随机变量 X 与 Y 相互独立, 则有

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y).$$

证明 根据方差的定义有

$$\begin{aligned} \text{Var}(X \pm Y) &= E[(X - EX \pm (Y - EY))^2] \\ &= E(X - EX)^2 + E(Y - EY)^2 \pm 2E[(X - EX)(Y - EY)] \\ &= \text{Var}(X) + \text{Var}(Y) \pm 2E[(X - EX)(Y - EY)]. \end{aligned}$$

根据性质 6.3, 我们有 $E[(X - EX)(Y - EY)]E(X - EX)E(Y - EY) = 0$, 由此完成证明.

例 6.1 设随机变量 $X \sim \mathcal{N}(0, 1)$ 和 $Y \sim \mathcal{N}(0, 1)$ 相互独立, 求 $E[\max(X, Y)]$.

解 根据独立性定义可得随机变量 X 和 Y 的联合概率密度为

$$f(x, y) = f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.$$

考虑区域 $D_1 = \{(x, y): x \geq y\}$ 和 $D_2 = \{(x, y): x < y\}$, 如图 6.1 所示. 于是得到

$$\begin{aligned} E[\max(X, Y)] &= \int \int_{D_1} xf(x, y)dx dy + \int \int_{D_2} yf(x, y)dx dy \\ &= \int_{-\infty}^{+\infty} dy \int_y^{+\infty} xf(x, y)dx + \int_{-\infty}^{+\infty} dx \int_x^{+\infty} yf(x, y)dy \\ &= 2 \int_{-\infty}^{+\infty} dy \int_y^{+\infty} xf(x, y)dx = \frac{1}{\pi} \int_{-\infty}^{+\infty} dy \int_y^{+\infty} xe^{-\frac{x^2+y^2}{2}} dx \\ &= \frac{1}{\pi} \int_{-\infty}^{+\infty} e^{-y^2} dy = \frac{1}{\sqrt{\pi}} \end{aligned}$$

最后一个等式成立是因为 $\int_{-\infty}^{+\infty} e^{-y^2} dy = \sqrt{\pi}$.

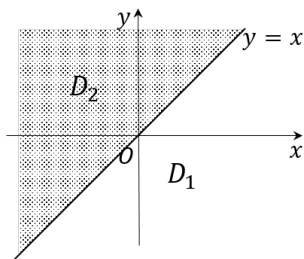


图 6.1 例 6.1 的积分区域 D_1 和 D_2 .

例 6.2 某水果超市在每星期一进货一定数量的新鲜水果, 假设一周内出售水果的件数 $X \sim U(10, 20)$. 若这一周内出售一件水果获利 10 元, 若不能出售则因为水果过期而每件亏损 4 元, 求水果超市的最优进货策略.

解 不妨假设水果超市每周进货 n 件 ($10 \leq n \leq 20$), 则它的周利润为

$$Y = \begin{cases} 10n & X \geq n \\ 10X - 4(n - X) & X < n \end{cases}.$$

由于周利润 Y 是关于 X 的随机变量, 只能考虑在期望下的最优策略

$$E[Y] = \sum_{i=10}^{n-1} (10i - 4(n - i))P(X = i) + \sum_{i=n}^{20} 10nP(X = i)$$

$$= \sum_{i=10}^{n-1} \frac{14i - 4n}{10} + \sum_{i=n}^{20} n = (-7n^2 + 243n + 630)/10 .$$

上式对 n 求一阶导数并令其等于零, 求解可得 $n = 17.36$, 最后取 $n = 17$.

6.2 协方差

随机变量的期望或方差仅涉及变量自身的统计信息, 没有刻画变量之间的统计信息. 本节引入一个新的统计特征: 协方差, 用于描述随机变量 X 和 Y 相互关系的数字特征.

定义 6.1 设二维随机向量 (X, Y) 的期望 $E[(X - E(X))(Y - E(Y))]$ 存在, 则称其为 X 和 Y 的协方差, 记为

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

协方差是两个随机变量与它们各自期望的偏差之积的期望, 由于偏差可正可负, 因此协方差可正可负. 根据协方差的定义容易发现

$$\text{Cov}(X, X) = \text{Var}(X) \quad \text{和} \quad \text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y) .$$

下面进一步研究协方差的性质:

性质 6.5 对任意随机变量 X, Y 和常数 c 有

$$\text{Cov}(X, c) = 0 \quad \text{和} \quad \text{Cov}(X, Y) = \text{Cov}(Y, X) .$$

性质 6.6 对任意随机变量 X, Y 和常数 a, b 有

$$\text{Cov}(aX, bY) = ab\text{Cov}(X, Y) \quad \text{和} \quad \text{Cov}(X + a, Y + b) = \text{Cov}(X, Y) .$$

证明 根据协方差的定义有

$$\text{Cov}(aX, bY) = E[(aX - E(aX))(bY - E(bY))] = abE[(X - E(X))(Y - E(Y))] ,$$

以及

$$\begin{aligned} \text{Cov}(X + a, Y + b) &= E[(X + a - E(X + a))(Y + b - E(Y + b))] \\ &= E[(X - E(X))(Y - E(Y))] . \end{aligned}$$

性质 6.7 对任意随机变量 X_1, X_2, Y 有

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

证明 根据协方差的定义有

$$\begin{aligned} \text{Cov}(X_1 + X_2, Y) &= E[(X_1 + X_2 - E(X_1) - E(X_2))(Y - E(Y))] \\ &= E[(X_1 - E(X_1))(Y - E(Y))] + E[(X_2 - E(X_2))(Y - E(Y))] \\ &= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y) . \end{aligned}$$

可将性质 6.7 推广到多个随机变量: 对随机变量 $X_1, X_2, \dots, X_n$ 和 $Y_1, Y_2, \dots, Y_m$ 有

$$\text{Cov}\left(\sum_i^n X_i, \sum_j^m Y_j\right) = \sum_i^n \sum_j^m \text{Cov}(X_i, Y_j) ,$$

以及进一步有

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) .$$

性质 6.8 若随机变量 X 与 Y 相互独立, 则有 $\text{Cov}(X, Y) = 0$, 但反之不成立.

证明 若 X 与 Y 相互独立, 则有 $E(XY) = E(X)E(Y)$, 于是得到

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 .$$

但反方向并不成立, 即若随机变量 X 与 Y 满足 $\text{Cov}(X, Y) = 0$, 并不能得到 X 与 Y 相互独立. 下面来看一个反例, 设随机变量 X 的分布列为

$$P(X = -1) = P(X = 0) = P(X = 1) = 1/3 ,$$

很容易得到 $E(X) = 0$. 引入一个新的随机变量 Y

$$Y = \begin{cases} 0 & X \neq 0 \\ 1 & X = 0 \end{cases} ,$$

可以发现 $E(XY) = 0$, 最后得到

$$\text{Cov}(X, Y) = E[XY] - E(X)E(Y) = 0 .$$

但随机变量 X 与 Y 显然不是相互独立的, 因为

$$P(X = 0, Y = 0) = 0 \quad \text{而} \quad P(X = 0)P(Y = 0) = 2/9 .$$

由此完成证明.

性质 6.9 对任意随机变量 X 与 Y 有

$$(\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y) ,$$

等式成立的充要条件是 $Y = aX + b$ 几乎处处成立, 即 X 与 Y 几乎处处存在线性关系.

证明 根据 Cauchy-Schwartz 不等式有

$$\begin{aligned} |\text{Cov}(X, Y)| &= |E[(X - E(X))(Y - E(Y))]| \\ &\leq \sqrt{E[(X - E(X))^2]E[(Y - E(Y))^2]} = \sqrt{\text{Var}(X)\text{Var}(Y)} . \end{aligned}$$

下面证明等号成立的充要条件. 若 $Y = aX + b$ 几乎处处成立, 则有

$$\text{Var}(Y) = a^2 \text{Var}(X) \quad \text{和} \quad \text{Cov}(X, Y) = \text{Cov}(X, aX + b) = a \text{Var}(X) ,$$

由此直接验证 $(\text{Cov}(X, Y))^2 = \text{Var}(X)\text{Var}(Y)$. 另一方面, 可以考虑函数

$$\begin{aligned} f(t) &= E[t(X - EX) - (Y - EY)]^2 \\ &= t^2 E[X - E(X)]^2 - 2tE[(X - E(X))(Y - E(Y))] + E[Y - E(Y)]^2 , \end{aligned}$$

根据方程的性质和条件 $(\text{Cov}(X, Y))^2 = \text{Var}(X)\text{Var}(Y)$ 有

$$\Delta = 4(E[(X - EX)(Y - EY)])^2 - 4E(X - EX)^2 E(Y - EY)^2 = 0 .$$

由此存在一个根 t_0 使得 $f(t_0) = 0$ 成立, 即

$$f(t_0) = E[(t_0(X - EX) - (Y - EY))^2] = 0 ,$$

由此可得 $Y = t_0(X - E(X)) + E(Y) = aX + b$ 几乎处处成立.

定理 6.2 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则有 $\text{Cov}(X, Y) = \rho\sigma_x\sigma_y$.

证明 根据协方差的定义有

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_x)(y - \mu_y)f(x, y)dxdy ,$$

其中

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right).$$

考虑变量变换

$$\begin{cases} u = (x - \mu_x)/\sigma_x \\ v = (y - \mu_y)/\sigma_y \end{cases} \quad \text{于是有} \quad \begin{cases} x = u\sigma_x + \mu_x \\ y = v\sigma_y + \mu_y \end{cases}.$$

容易得到其雅可比行列式为 $J = \sigma_x\sigma_y$, 于是有

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} uv \exp\left(-\frac{u^2 + v^2 - 2\rho uv}{2(1-\rho^2)}\right) dudv \\ &= \frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} v \exp(-v^2/2) dv \int_{-\infty}^{+\infty} u \exp\left(-\frac{(u-\rho v)^2}{2(1-\rho^2)}\right) du \\ &= \frac{\sigma_x\sigma_y}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \rho v^2 \exp(-v^2/2) dv = \rho\sigma_x\sigma_y. \end{aligned}$$

其中第三个等式成立是因为利用正太分布 $\mathcal{N}(\rho v, 1-\rho^2)$ 的期望, 而最后一个不等式成立是因为利用标准正太分布的方差.

结合定理 5.4 和定理 6.2 得到

推论 6.1 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则 X 和 Y 相互独立的充要条件是协方差 $\text{Cov}(X, Y) = 0$.

例 6.3 设随机变量 $X_1, X_2, \dots, X_n$ 相互独立且服从正太分布, 方差为 σ^2 . 记 $\bar{X} = \sum_{i=1}^n X_i/n$, 讨论 $\bar{X}$ 和 $\bar{X} - X_i$ 的独立性.

解 根据正太分布的性质可以知道 $\bar{X}$ 和 $\bar{X} - X_i$ 都服从正太分布, 而正太分布的独立性可通过协方差来研究. 根据协方差的性质有

$$\text{Cov}(\bar{X}, \bar{X} - X_i) = \text{Cov}(\bar{X}, \bar{X}) - \text{Cov}(\bar{X}, X_i) = \text{Var}(\bar{X}) - \text{Cov}\left(\sum_{j=1}^n \frac{X_j}{n}, X_i\right).$$

根据 $X_1, X_2, \dots, X_n$ 的相互独立性有

$$\text{Var}(\bar{X}) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n} \quad \text{和} \quad \text{Cov}\left(\sum_{j=1}^n \frac{X_j}{n}, X_i\right) = \frac{1}{n} \text{Cov}(X_i, X_i) = \frac{\sigma^2}{n},$$

于是得到 $\text{Cov}(\bar{X}, \bar{X} - X_i) = 0$. 根据推论 6.1 得到 $\bar{X}$ 和 $\bar{X} - X_i$ 是相互独立的.

例 6.4 随机变量 (X, Y) 联合概率密度为

$$f(x, y) = \begin{cases} (x+y)/8 & 0 \leq x \leq 2, 0 \leq y \leq 2 \\ 0 & \text{其它} \end{cases},$$

求 $\text{Cov}(X, Y)$ 和 $\text{Var}(X+Y)$.

解 根据协方差的定义有 $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$, 需计算

$$E[X] = E[Y] = \int_0^2 \int_0^2 x(x+y)/8 dx dy = 7/6 \quad \text{和} \quad E[XY] = \int_0^2 \int_0^2 xy(x+y)/8 dx dy = 4/3,$$

由此可得 $\text{Cov}(X, Y) = -1/36$. 进一步计算

$$E[X^2] = E[Y^2] = \int_0^2 \int_0^2 x^2(x+y)/8 dx dy = 5/3,$$

由此可得 $\text{Var}(X) = \text{Var}(Y) = 5/3 - (7/6)^2 = 11/36$. 最后得到

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = 11/18 - 1/18 = 5/9.$$

例 6.5 (匹配问题) 有 n 对夫妻参加一次聚会, 将所有参会人员任意分成 n 组, 每组一男一女, 用 X 表示夫妻两人被分到一组的对数, 求 X 的期望和方差.

解 用 X_i 表示第 i 对夫妻是否被分到一组, 即

$$X_i = \begin{cases} 1 & \text{第 } i \text{ 对夫妻被分到一组} \\ 0 & \text{否则} \end{cases},$$

则有 $X = X_1 + X_2 + \cdots + X_n$. 随机变量 X_i 的分布列为

$$P(X_i = 1) = (n-1)!/n! = 1/n \quad \text{和} \quad P(X_i = 0) = 1 - 1/n.$$

于是得到期望

$$E(X) = E(X_1 + X_2 + \cdots, X_n) = E(X_1) + E(X_2) + \cdots + E(X_1) = 1.$$

对任意 $i \neq j$ 有

$$P(X_i = 1, X_j = 1) = (n-2)!/n! = 1/n(n-1),$$

由此得到

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E(X_i)E(X_j) = 1/n^2(n-1),$$

最后根据协方差的性质有

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i \neq j} \text{Cov}(X_i, X_j) = 1 .$$

6.3 相关系数

两个随机变量之间的关系可分为独立与非独立, 在非独立中又可以分为线性关系和非线性关系. 非线性关系较为复杂, 目前尚无好的办法来处理. 但线性相关程度可以通过线性相关系数来刻画, 下面给出具体的定义:

定义 6.2 设 (X, Y) 为二维随机向量, 如果它们的方差 $\text{Var}(X)$ 和 $\text{Var}(Y)$ 存在且都不为零, 则称 $\text{Cov}(X, Y)/\sqrt{\text{Var}(X)\text{Var}(Y)}$ 为 X 与 Y 的 **线性相关系数**, 简称 **相关系数** (correlation coefficient), 记为

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} .$$

若 $\rho_{XY} > 0$, 称 X 与 Y **正相关**; 若 $\rho_{XY} < 0$, 称 X 与 Y **负相关**; 若 $\rho_{XY} = 0$, 称 X 与 Y **不相关**.

相关系数是根据协方差和方差所定义, 与协方差同号, 可以看作是对协方差的一种规范化. 相关系数的很多性质可以通过协方差获得:

- 根据协方差性质 6.9 可知

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

$|\rho_{XY}| = 1$ 的充要条件为 X 与 Y 几乎处处有线性关系 $Y = aX + b$.

- 根据协方差性质 6.8 可知, 若 X 与 Y 相互独立, 则 X 与 Y 不相关 ($\rho_{XY} = 0$), 但反之不成立;
- 随机变量 X 与 Y 不相关, 仅仅表示 X 与 Y 之间不存在线性关系, 可能存在其他关系. 例如, 设随机变量 $X \sim U(-1/2, 1/2)$ 和 $Y = \cos(X)$, 则有

$$\text{Cov}(X, Y) = E(X \cos(X)) - E(X)E(\cos(X)) = E[X \cos(X)] = \int_{-1/2}^{1/2} x \cos(x) dx = 0 .$$

根据定理 6.2 和推论 6.1 有

定理 6.3 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则 X 与 Y 的相关系数 $\rho_{XY} = \rho$, 以及 X 与 Y 独立的充要条件是 X 与 Y 不相关.

独立与不相关的等价性仅限于正太分布随机变量, 对其它类型并不一定成立, 详情见性质 6.8 证明中的反例.

根据相关系数、协方差和方差的定义很容易得到如下几个不相关的等价条件.

定理 6.4 设随机变量 X 和 Y 的方差存在且都不为零, 以下几个条件相互等价:

- X 和 Y 不相关, 即 $\rho_{XY} = 0$;
- 协方差 $\text{Cov}(X, Y) = 0$;
- $E(XY) = E(X)E(Y)$;
- $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$.

例 6.6 设随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$ 和 $Y \sim \mathcal{N}(\mu, \sigma^2)$ 相互独立. 求 $Z_1 = \alpha X + \beta Y$ 和 $Z_2 = \alpha X - \beta Y$ 的相关系数 ($\alpha, \beta \neq 0$).

解 根据正态分布的定义有

$$\begin{aligned}\text{Cov}(Z_1, Z_2) &= \text{Cov}(\alpha X + \beta Y, \alpha X - \beta Y) = (\alpha^2 - \beta^2)\sigma^2 \\ \text{Var}(Z_1) &= \text{Cov}(\alpha X + \beta Y, \alpha X + \beta Y) = (\alpha^2 + \beta^2)\sigma^2 \\ \text{Var}(Z_2) &= \text{Cov}(\alpha X - \beta Y, \alpha X - \beta Y) = (\alpha^2 + \beta^2)\sigma^2,\end{aligned}$$

由此可知 $\rho_{XY} = (\alpha^2 - \beta^2)/(\alpha^2 + \beta^2)$.

例 6.7 设离散型随机向量 $(X_1, X_2, \dots, X_n)$ 服从多项分布 $M(m, p_1, p_2, \dots, p_n)$, 对任意 $i \neq j$, 求 X_i 和 X_j 的相关系数.

根据多项分布的定义有 $X_1 + X_2 + \dots + X_n = m$, 当 X_i 越大时 X_j 应该越小, 因此直观而言这两随机变量之间应该是负相关的.

解 根据多项分布的性质有 X_i 和 X_j 的边缘分布分别

$$X_i \sim B(m, p_i) \quad \text{和} \quad X_j \sim B(m, p_j).$$

由此可得 $\text{Var}(X_i) = mp_i(1 - p_i)$ 和 $\text{Var}(X_j) = mp_j(1 - p_j)$. 直接求解 $\text{Cov}(X_i, X_j)$ 存在一定的难度. 对每个 $k \in [m]$, 引入随机变量

$$Y_i^k = \begin{cases} 1 & \text{若第 } k \text{ 次实验的结果为 } i \\ 0 & \text{其它} \end{cases} \quad \text{和} \quad Y_j^k = \begin{cases} 1 & \text{若第 } k \text{ 次实验的结果为 } j \\ 0 & \text{其它} \end{cases}.$$

由此可得

$$X_i = Y_i^1 + Y_i^2 + \dots + Y_i^m \quad \text{和} \quad X_j = Y_j^1 + Y_j^2 + \dots + Y_j^m.$$

根据第 k 次试验和第 l 次试验相互独立 ($k \neq l$), 以及 $Y_i^k Y_j^k = 0$ 有

$$\text{Cov}(Y_i^k, Y_j^l) = 0, \quad \text{以及} \quad \text{Cov}(Y_i^k, Y_j^k) = E[Y_i^k Y_j^k] - E(Y_i^k)E(Y_j^k) = -p_i p_j,$$

根据协方差的性质有

$$\text{Cov}(X_i, X_j) = \sum_{k=1}^m \text{Cov}(Y_i^k, Y_j^k) + \sum_{k \neq l} \text{Cov}(Y_i^k, Y_j^l) = -mp_i p_j .$$

最后得到 X_i 和 X_j 的相关系数为

$$\rho = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}} = \frac{-mp_i p_j}{\sqrt{mp_i(1-p_i)}\sqrt{mp_j(1-p_j)}} = -\frac{\sqrt{p_i p_j}}{\sqrt{(1-p_i)(1-p_j)}} .$$

6.4 条件期望

前一章介绍了条件分布, 基于条件分布可以考虑条件期望, 分离散和连续性随机变量两种情况.

定义 6.3 设 (X, Y) 为连续型随机变量, 在 $Y = y$ 条件下 X 的条件密度函数为 $f_{X|Y}(x|y)$, 称

$$E(X|y) = E(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx$$

为在 $Y = y$ 条件下 X 的 **条件期望**. 设 (X, Y) 为离散型随机变量, 在 $Y = y$ 条件下 X 的条件分布列为 $P(X = x_i | Y = j)$, 称

$$E(X|y) = E(X|Y = y) = \sum_i x_i P(X = x_i | Y = j)$$

为在 $Y = y$ 条件下 X 的 **条件期望**.

条件期望 $E[X|y]$ 一般都与 y 相关, 是 y 的函数, 而 (无条件) 期望 $E(X)$ 是一个具体的常数. 在上面的定义中, 我们都默认期望存在. 条件期望是条件分布的期望, 具有期望的一切性质:

- 对任意常数 a, b 有 $E(aX_1 + bX_2|Y) = aE(X_1|Y) + bE(X_2|Y)$;
- 对离散型随机变量 (X, Y) 和函数 $g(X)$ 有

$$E(g(X)|Y) = \sum_i g(x_i) P(X = x_i | Y = y) ;$$

对连续型随机变量 (X, Y) 和函数 $g(X)$ 有

$$E(g(X)|Y) = \int_{-\infty}^{+\infty} g(x) f(x|Y = y) dx ;$$

- 设随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则在 $Y = y$ 的条件下随机变量 X 服从正太分布 $\mathcal{N}(\mu_x - \rho\sigma_x(y - \mu_y)/\sigma_y, (1 - \rho^2)\sigma_x^2)$, 由此可得 $E(X|y) = \mu_x - \rho\sigma_x(y - \mu_y)/\sigma_y$.

下面给出了计算期望的另一种方法.

定理 6.5 对二维随机变量 (X, Y) 有

$$E(X) = E_Y(E(X|Y)) = \begin{cases} \sum_{y_j} E(X|y_j)P(Y = y_j) & \text{离散型随机变量,} \\ \int_{-\infty}^{\infty} E(X|y)f_Y(y)dy & \text{连续型随机变量.} \end{cases}$$

证明 对连续型随机变量 (X, Y) , 不妨假设其联合密度函数为 $f(x, y)$, 根据条件概率有

$$E[X] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xf(x, y)dydx = \int_{-\infty}^{+\infty} f_Y(y) \int_{-\infty}^{+\infty} xf_{X|Y}(x|y)dx dy = \int_{-\infty}^{+\infty} E(X|y)f_Y(y)dy .$$

对离散型随机变量 (X, Y) , 根据条件概率和全概率公式有

$$\begin{aligned} E[X] &= \sum_i x_i P_X(X = x_i) = \sum_i \sum_j x_i P(X = x_i, Y = y_j) \\ &= \sum_i \sum_j x_i P(X = x_i|Y = y_j)P(Y = y_j) \\ &= \sum_j P(Y = y_j) \sum_i x_i P(X = x_i|Y = y_j) \\ &= \sum_j P(Y = y_j)E[X|Y = y_j] = E_Y[E[X|Y]]. \end{aligned}$$

下面介绍与全概率公式相对于的一个公式: **全期望公式** (law of total expectation), 在期望的计算起到重要作用.

定理 6.6 设 $A_1, A_2, \dots, A_n$ 是样本空间 Ω 的一个分割, 即 $A_i A_j = \emptyset$ 和 $\Omega = \cup_{i=1}^n A_i$. 对任意随机变量 X 有

$$E[X] = E[X|A_1]P(A_1) + E[X|A_2]P(A_2) + \dots + E[X|A_n]P(A_n) ,$$

特别地, 随机事件 A 与其对立事件 $\bar{A}$ 构成样本空间 Ω 的一个划分, 对任意随机变量 X 有

$$E[X] = E[X|A]P(A) + E[X|\bar{A}]P(\bar{A}) .$$

证明 对于随机变量 X 和 $A_1, A_2, \dots, A_n$, 首先引入新的离散随机变量 $Y = 1, 2, \dots, n$ 满足

随机事件 $Y = i$ 发生的充要条件是 $X \in A_i$.

根据定理 6.5 可知

$$E(X) = E_Y(E(X|Y)) = \sum_{i=1}^n E(X|Y = i)P(Y = i) = \sum_{i=1}^n E(X|A_i)P(A_i) .$$

例 6.8 设 (X, Y) 的联合概率密度为

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

求条件期望 $E(X|y)$.

解 首先计算 Y 的边缘密度函数, 当 $y > 0$ 时

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx = \int_0^y \exp(-y)dx = y \exp(-y),$$

由此得到在 $Y = y$ 的条件下 X 的条件分布

$$f_{X|Y}(x|y) = f(x, y)/f_Y(y) = 1/y \quad (0 < x < y < +\infty).$$

最后得到条件期望

$$E(X|y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)dx = \int_0^y x/y dx = y/2.$$

例 6.9 一矿工被困在有三个门的矿井里, 第一个门通一坑道, 沿此坑道走 3 小时可使他到达安全地点; 第二个门可使他走 5 小时后义回到原地; 第三个门可使他走 7 小时后也回到原地. 如设此矿工在任何时刻都等可能地选定其中一个门, 试问他到达安全地点平均要用多长时间?

解 用 X 为该矿工到达安全地点所需时间, 用 Y 为他所选的门, 根据全期望公式有

$$E(X) = E(X|Y=1)P(Y=1) + E(X|Y=2)P(Y=2) + E(X|Y=3)P(Y=3),$$

其中 $P(Y=1) = P(Y=2) = P(Y=3) = 1/3$, $E(X|Y=1) = 3$. 用 $E(X|Y=2)$ 表示矿工从第二个门出去要到达安全地点所需平均时间. 而他沿此坑道走 5 小时又转回原地, 而一旦返回原地, 问题就与当初他还没有进第二个门之前一样, 因此他要到达安全地点平均还需再用 $E(X)$ 小时. 同理可以考虑 $E(X|Y=2)$, 故有

$$E(X|Y=2) = 5 + E(X) \quad \text{和} \quad E(X|Y=3) = 7 + E(X).$$

于是得到

$$E(X) = (3 + 5 + E(X) + 7 + E(X))/3.$$

求解出 $E(X) = 15$ (小时), 该矿工到达安全地点平均需要 15 小时.

6.5 随机向量的数学期望与协方差阵

定义 6.4 设随机向量 $X = (X_1, X_2, \dots, X_n)^\top$, 称

$$E(X) = (E(X_1), E(X_2), \dots, E(X_n))^\top$$

为随机向量 X 的期望, 以及称

$$\text{Cov}(X) = \Sigma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

为随机变量 X 的协方差矩阵.

下面介绍协方差矩阵的一些性质:

定理 6.7 随机向量 $X = (X_1, X_2, \dots, X_n)$ 的协方差矩阵是对称半正定的矩阵.

证明 对任意 $i \neq j$, 根据协方差的性质

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i),$$

可知协方差矩阵是对称的. 对于半正定性的证明, 首先引入新的函数

$$f(t_1, t_2, \dots, t_n) = (t_1, t_2, \dots, t_n) \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix} (t_1, t_2, \dots, t_n)^\top,$$

由此得到

$$f(t_1, t_2, \dots, t_n) = E \left(\left(\sum_{i=1}^n t_i (X_i - E(X_i)) \right)^2 \right) \geq 0,$$

由此完成证明.

定理 6.8 设多维正态分布 $X = (X_1, X_2, \dots, X_n)^\top \sim N(\mu, \Sigma)$, 则有

$$\mu = (E[X_1], E[X_2], \dots, E[X_n])^\top \quad \text{和} \quad \Sigma = [\text{Cov}(X_i, X_j)]_{n \times n}.$$

6.6 应用案例

有时我们能观察到随机变量 X 的值, 需要对随机变量 Y 的值进行预测, 即选择一个函数 $g(x)$, 使得 $g(X)$ 接近预测值 Y . 选择函数 $g(x)$ 的一个准则是最小化 $E[(Y - g(x))^2]$. 关于最优的函数 $g(X)$, 有如下结论:

定理 6.9 对任意函数 $g(x)$ 和随机变量 X 和 Y , 有

$$E([Y - g(X)]^2) \geq E([Y - E(Y|X)]^2) .$$

证明 根据定理 6.5 只需证明对任意给定 X 有

$$E([Y - g(X)]^2|X) \geq E([Y - E(Y|X)]^2|X) , \quad (6.1)$$

对上式两边分别对 X 求期望可完成证明. 下面考虑如何上面的条件期望不等式, 首先有

$$\begin{aligned} E([Y - g(X)]^2|X) &= E([Y - E(Y|X) + E(Y|X) - g(X)]^2|X) \\ &= E([Y - E(Y|X)]^2|X) + E([E(Y|X) - g(X)]^2|X) + 2E([Y - E(Y|X)][E(Y|X) - g(X)]|X) , \end{aligned}$$

当给定 X 后, $[E(Y|X) - g(X)]^2$ 和 $E(Y|X)$ 都是常数, 因此有

$$E([Y - E(Y|X)][E(Y|X) - g(X)]|X) = [E(Y|X) - g(X)]E([Y - E(Y|X)]|X) = 0 ,$$

结合上面两式完成 (6.1) 的证明.

很多情况下很难知道随机变量 X 和 Y 的联合分布, 有些情况下即使知道联合分布计算 $E(Y|X)$ 也非常复杂. 若已知随机变量 X 和 Y 的一些统计量, 依然可以很好地估计出 X 和 Y 的最优线性预测, 例如,

例 6.10 设随机变量 X 和 Y 的期望、方差、相关系数分别为 $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$, 其中 $\sigma_x > 0, \sigma_y > 0, \rho \in [-1, +1]$, 求解最优的线性预测 $Y = aX + b$ 使得 $E((Y - aX - b)^2)$ 最小化.

解 首先设函数

$$\begin{aligned} F(a, b) &= E((Y - aX - b)^2) \\ &= E(Y^2) - 2aE(Y) - 2bE(XY) + a^2 + 2abE(X) + b^2E(X^2) . \end{aligned}$$

求函数 $F(a, b)$ 的最小值, 可以考虑令 a 和 b 的偏导等于零, 即

$$\begin{cases} \partial F(a, b)/\partial a = 2a + 2bE(X) - 2E(Y) = 0 \\ \partial F(a, b)/\partial b = 2bE(X^2) + 2aE(X) - 2E(XY) = 0 . \end{cases}$$

求解上面的方程组可得

$$\begin{cases} a = E(Y) + bE(X) = \mu_y - \rho\sigma_y\mu_x/\sigma_x \\ b = \frac{E(XY) - E(X)E(Y)}{E(X^2) - (E(X))^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\rho\sigma_x\sigma_y}{\sigma_x^2} = \rho\sigma_y/\sigma_x . \end{cases}$$

由此给出 Y 的最优线性预测为

$$Y = \rho\sigma_y(X - \mu_x)/\sigma_x + \mu_y ,$$

在最优线性预测下预测的均分误差

$$\begin{aligned} & E((Y - \rho\sigma_y(X - \mu_x)/\sigma_x - \mu_y)^2) \\ &= E((Y - \mu_y)^2) + \rho^2\sigma_y^2 E((X - \mu_x)^2) / \sigma_x^2 - 2\rho\sigma_y E((X - \mu_x)(Y - \mu_y)) / \sigma_x \\ &= \sigma_y^2 + \rho^2\sigma_y^2 - 2\rho^2\sigma_y^2 = \sigma_y^2(1 - \rho^2) . \end{aligned}$$

由此可以看出, 当 $\rho^2 \rightarrow 1$ 时最优线性预测的均方误差接近零.

习题

- 6.1 随机变量 X 与 Y 独立, 且 $\text{Var}(X) = 6$ 和 $\text{Var}(Y) = 3$, 求 $\text{Var}(2X \pm Y)$.
- 6.2 设随机变量 $X \sim P(2)$ 和 $Y \sim \mathcal{N}(-2, 4)$ 相互独立, 求 $E[(X - Y)^2]$.
- 6.3 在长度为 l 的线段上任取两点 X, Y , 求期望 $E[\min(X, Y)]$ 和 $E[|X - Y|]$.
- 6.4 随机变量 $X \sim \mathcal{N}(-1, 2)$ 和 $Y \sim \mathcal{N}(1, 8)$, 且相关系数 $\rho_{XY} = -1/2$. 求 $\text{Var}(X + Y)$.
- 6.5 对任意实数 $x, y \in (0, 1)$ 且满足 $x + y \leq 1$, 证明 $xy/(1-x)(1-y) \in (0, 1)$.
- 6.6 证明: 定义 5.9 和定义 5.15 中关于二维正太分布的密度函数完全相等.