Numerical Characteristics of Multi-dimensional Random Vectors

December 3, 2024

回顾: 多维正态分布的标准化

Focus: 设 n 维随机向量 $X = (X_1, X_2, \ldots, X_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 以及正定矩阵 $\boldsymbol{\Sigma}$ 的特征值分解 $\boldsymbol{\Sigma} = \mathbf{U}^\top \boldsymbol{\Lambda} \mathbf{U}$, 则随机向量

$$Y = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}(X - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{I}_n)\ .$$

我们需要明确两件事情:

- $\mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$ 的分量之间是独立的. 进一步, 当 $\boldsymbol{\Sigma}$ 为对角阵的时候, 多元正态分布的随机变量之间是独立的.
- $\boldsymbol{\Lambda}^{-1/2}\mathbf{U}(X - \boldsymbol{\mu})$ 所带来的线性变换可以将 X 标准化.

线性运算的基本性质

- 回忆: 矩阵的特征值和特征向量. 以二维为例, 矩阵 $\mathbf{A}$ 拥有特征值 λ_1 和 λ_2, 分别对应特征向量 $\boldsymbol{u}_1, \boldsymbol{u}_2$. 则有

$$\mathbf{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i \,, \quad i = 1, 2$$

进而有 $\mathbf{A} = \mathbf{U}^\top \mathbf{\Lambda} \mathbf{U}$ with $\mathbf{U}^\top = \mathbf{U}^{-1}$, 即

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

- 几何性质. For the case of $\mathbf{A}\boldsymbol{v}$ given any $\boldsymbol{v}$,
  - 矩阵 $\mathbf{U}$ 负责对 $\boldsymbol{v}$ 进行旋转
  - 矩阵 $\mathbf{\Lambda}$ 负责对 $\boldsymbol{v}$ 进行放缩

线性运算的基本性质

举个栗子, $\mathbf{A} = [1, 2; 2, 1]$

- 求特征值. $|\lambda\mathbf{I} - \mathbf{A}| = 0$, 得 $\lambda_1 = -1$ 和 $\lambda_2 = 3$.
- 求特征向量. 当 $\lambda_1 = -1$ 时,

$$(\lambda_i\mathbf{I} - \mathbf{A})\boldsymbol{u}_1 = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix}$$

  所以, $\boldsymbol{u}_1 = (1/\sqrt{2}, -1/\sqrt{2})$. 同理, 有 $\boldsymbol{u}_2 = (1/\sqrt{2}, 1/\sqrt{2})$.
- 根据 $\mathbf{UA} = \mathbf{\Lambda U}$, 可理解为
  - 正交坐标轴 $\mathbf{U}$ 上的向量 = 正交坐标轴 $\mathbf{U}$ 的线性组合
  - 协交坐标轴 $\mathbf{A}$ 上的向量 = 协交坐标轴 $\mathbf{A}$ 的线性组合
- For the case of $\mathbf{A}\boldsymbol{v}$, $\mathbf{U}\boldsymbol{v}$ 实现了对 $\boldsymbol{v}$ 的旋转, 而 $\mathbf{\Lambda}$ 实现了对 $|\boldsymbol{v}|$ 的放缩.

为了授课方便, 这里只关心对称的方阵. 同理, 可以推至非方阵的情况.

线性运算的基本性质

如果我们面对一个运算 $\mathbf{A}\boldsymbol{x}+\boldsymbol{b}$,

- $\boldsymbol{b}$ 是平移
- $\mathbf{U}$ 是旋转, if $|\mathbf{U}|=1$
- $\boldsymbol{\Lambda}$ 是放缩, as $|\mathbf{A}\boldsymbol{x}|=|\boldsymbol{\Lambda}\boldsymbol{x}|$

面对 $Y=\boldsymbol{\Lambda}^{-1/2}\mathbf{U}(X-\boldsymbol{\mu})=\boldsymbol{\Sigma}^{-1/2}(X-\boldsymbol{\mu})$

- $-\boldsymbol{\mu}$ 或者 $-\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ 是平移, 使得 $X-\boldsymbol{u}$ 以原点为中心 (X 以 $\boldsymbol{u}$ 为中心)
- $\mathbf{U}$ 是旋转
- $\boldsymbol{\Lambda}^{-1/2}$ 是放缩, 尤其是在 $\boldsymbol{\Lambda}^{-1/2}\bar{X}$ 的情形下

线性变换可以将 X 标准化

我们有

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = \boldsymbol{y}^\top \boldsymbol{y} = \boldsymbol{y}^\top \mathbf{I}_n \boldsymbol{y}$$

- 一方面，

$$\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}^\top \mathbf{U} \boldsymbol{\Lambda}^{-1} \mathbf{U}^\top \bar{\boldsymbol{x}} = \left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top \boldsymbol{\Lambda}^{-1} \left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)$$

- 另一方面，

$$|\boldsymbol{\Sigma}^{-1/2} \bar{X}| = |\boldsymbol{\Lambda}^{-1/2} \bar{X}| = |Y|$$

所以，

$$\begin{aligned}
|\bar{\boldsymbol{x}}^\top \boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{x}}| &= |\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top \boldsymbol{\Lambda}^{-1} \left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)| \\
&= |\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)^\top| \, |\boldsymbol{\Lambda}^{-1}| \, |\left(\mathbf{U}^\top \bar{\boldsymbol{x}}\right)| \\
&= |\boldsymbol{x}^\top| \, |\boldsymbol{\Lambda}^{-1}| \, |\boldsymbol{x}| \\
&= |\boldsymbol{y}^\top| \, |\boldsymbol{y}|
\end{aligned}$$

线性变换可以将 X 标准化

根据

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right)$$

和

$$Y = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}(X-\boldsymbol{\mu})$$

有

$$f_Y(\boldsymbol{y}) = (2\pi)^{-n/2}\exp\left(-\boldsymbol{y}^\top\boldsymbol{y}/2\right)$$

对角的协方差矩阵对应独立性

求证: 当 $\boldsymbol{\Sigma}$ 为对角阵的时候, 多元正态分布的随机变量之间是独立的.

回顾: 多维正态分布的密度函数

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2\right)$$

如果 $\boldsymbol{\Sigma}$ 为对角阵, 假设为 $\boldsymbol{\Sigma}^{-1} = \mathrm{Diag}\{\lambda_1^{-1},\ldots,\lambda_n^{-1}\}$. 则指数项可以拆解为

$$\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]$$

进而有

$$\begin{aligned}
f(\boldsymbol{x}) &= (2\pi)^{-n/2}\left(\prod_{i=1}^n\lambda_i^{-1/2}\right)\exp\left(\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]\right)\\
&= \prod_{i=1}^n\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\right)\exp\left[-\frac{(\bar{x}_1)^2}{2\lambda_1}-\cdots-\frac{(\bar{x}_n)^2}{2\lambda_n}\right]
\end{aligned}$$

对角的协方差矩阵对应着独立性

重新整理

$$
\begin{aligned}
f(\boldsymbol{x}) &= (2\pi)^{-n/2}\left(\prod_{i=1}^{n}\lambda_i^{-1/2}\right)\exp\left(\frac{-1}{2}\left[\lambda_1^{-1}(\bar{x}_1)^2+\cdots+\lambda_n^{-1}(\bar{x}_n)^2\right]\right)\\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\right)\exp\left[-\frac{(\bar{x}_1)^2}{2\lambda_1}-\cdots-\frac{(\bar{x}_n)^2}{2\lambda_n}\right]\\
&= \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}}\exp\left(-\frac{(\bar{x}_i)^2}{2\lambda_i}\right)\right)
\end{aligned}
$$

因此, 有

$$
f(\boldsymbol{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)
$$

基于上述运算, 我们可以推导多维正态分布其他性质 – 证明已留作思考题

设随机向量 $X = (X_1, X_2, \ldots, X_n)^\top$ 和 $Y = (Y_1, Y_2, \ldots, Y_m)^\top$, 以及

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_{xx}} & \boldsymbol{\Sigma_{xy}} \\ \boldsymbol{\Sigma_{yx}} & \boldsymbol{\Sigma_{yy}} \end{pmatrix}\right),$$

则有

- 随机向量 X 和 Y 的边缘分布为 $X \sim \mathcal{N}(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_{xx}})$ 和 $Y \sim \mathcal{N}(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_{yy}})$
- 随机向量 X 和 Y 相互独立的充要条件是 $\boldsymbol{\Sigma_{xy}} = \mathbf{0}_{m\times n}$ (元素全为零的 $m \times n$ 矩阵)
- 在 $X = \boldsymbol{x}$ 的条件下, 随机向量 Y 的分布

$$Y \mid X = \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_y} + \boldsymbol{\Sigma_{yx}}\boldsymbol{\Sigma_{xx}^{-1}}(\boldsymbol{x} - \boldsymbol{\mu_x}), \boldsymbol{\Sigma_{yy}} - \boldsymbol{\Sigma_{yx}}\boldsymbol{\Sigma_{xx}^{-1}}\boldsymbol{\Sigma_{xy}})$$

- 在 $Y = \boldsymbol{y}$ 的条件下, 随机向量 X 的分布

$$X \mid Y = \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu_x} + \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}(\boldsymbol{y} - \boldsymbol{\mu_y}), \boldsymbol{\Sigma_{xx}} - \boldsymbol{\Sigma_{xy}}\boldsymbol{\Sigma_{yy}^{-1}}\boldsymbol{\Sigma_{yx}})$$

数字特征的引言

类似于一维随机变量的数字特征, 多维随机变量也有数字特征

- 一方面是各分量自己的数字特征, 比如: 期望、方差、标准差等
- 另一方面是分量之间的关联程度, 反映随机变量间相依关系的数字特征, 即协方差与相关系数.

多维随机向量函数的期望

定义 0.54 [Informally, 期望 = < 取值, 概率 >]

离散随机变量 (X, Y) 的分布列为 $p_{ij} = P(X = x_i, Y = y_j)$, 则随机变量 $Z = g(X, Y)$ 的期望为

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

连续随机变量 (X, Y) 的概率密度为 $f(x, y)$, 则随机变量 $Z = g(X, Y)$ 的期望为

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) \,\mathrm{d}x \,\mathrm{d}y.$$

期望的性质

- 若随机向量 $X \geq Y$, 则 $\mathbb{E}[X] \geq \mathbb{E}[Y]$;
- 线性性. 对任意随机向量 X 和 Y, 有 $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$;
- 对任意随机向量 X 和 Y, 有 Cauchy-Schwartz 不等式:

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

- 独立可乘. 若随机变量 X 与 Y 相互独立, 有 $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$;
  - 独立方差. 若随机变量 X 与 Y 相互独立, 有 $\mathbb{VAR}(X+Y) = \mathbb{VAR}(X) + \mathbb{VAR}(Y)$.

多维随机向量函数的期望：例 0.120

例 0.120 设随机变量 X 与 Y 相互独立且都服从标准正态分布 $\mathcal{N}(0,1)$，求 $\mathbb{E}[\max(X,Y)]$.

解答：例 0.120

题目: 设随机变量 X 与 Y 相互独立且都服从标准正态分布 $\mathcal{N}(0,1)$, 求 $\mathbb{E}[\max(X,Y)]$.

解答:

- 由题易知 X 与 Y 的联合密度函数为

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

- 将二维平面区域分为两部分 $D_1=\{(x,y):x\geq y\}$, $D_2=\{(x,y):x<y\}$, 于是得到

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

多维随机向量函数的期望：例 0.121

例 0.121 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,求期望意义下水果超市的最优进货策略.

解答：例 0.121

题目: 如上所述.

解答:

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

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

- 周利润 Y 是关于 X 的随机变量, 考虑在期望下的最优策略

$$\mathbb{E}[Y] = \sum_{i=10}^{n-1}(10i - 4(n-i))P(X=i) + \sum_{i=n}^{20} 10nP(X=i)$$
$$= \sum_{i=10}^{n-1}\frac{14i - 4n}{10} + \sum_{i=n}^{20} n = \frac{-7n^2 + 243n + 630}{10}.$$

上式对 n 求导并令导数为零, 求解可得 $n = 17.36$. 则 n 可能取 17 或者 18, 经验证, 最后取 $n = 17$.

多维随机向量函数的条件期望

回顾: 对二维随机向量 (X,Y) 而言, 随机变量 X 的条件分布, 即给定随机变量 Y 取值的条件下 X 的概率分布. 而条件期望是条件分布的数学期望, 具体定义如下:

定义 0.55 设 (X,Y) 为离散型随机向量, 在 $Y=y$ 的条件下 X 的条件分布列为 $P(X=x_i|Y=y)$, 称

$$\mathbb{E}(X|y)=\mathbb{E}(X|Y=y)=\sum_i x_i P(X=x_i|Y=y)$$

为在 $Y=y$ 的条件下 X 的条件期望. 设 (X,Y) 为连续型随机向量, 在 $Y=y$ 的条件下 X 的条件密度函数为 $f_{X|Y}(x|y)$, 称

$$\mathbb{E}(X|y)=\mathbb{E}(X|Y=y)=\int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,\mathrm{d}x$$

为在 $Y=y$ 的条件下 X 的条件期望.

无条件期望 $\mathbb{E}(X)$ 是一个常数, 而条件期望 $\mathbb{E}(X|Y=y)$ 是 y 的函数; 例如用 X 表示中国成年人的身高, Y 表示中国成年人的足长, 我国公安部门研究得足长为 y 的中国成年人平均身高为 $\mathbb{E}(X|Y=y)=6.876y$, 此公式常用于公安痕迹侦查中.

条件期望的性质

- 线性性. 对任意常数 a, b 有 $\mathbb{E}(aX_1+bX_2|Y)=a\mathbb{E}(X_1|Y)+b\mathbb{E}(X_2|Y)$;

- 函数型. 对离散型随机向量 (X,Y) 和函数 $g(X)$, 有

$$\mathbb{E}(g(X)|Y)=\sum_i g(x_i)P(X=x_i|Y=y)$$

  对连续型随机向量 (X,Y) 和函数 $g(X)$, 有

$$\mathbb{E}(g(X)|Y)=\int_{-\infty}^{+\infty} g(x)f(x|Y=y)\,\mathrm{d}x$$

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

$$\mathbb{E}(X|y)=\mu_x-\frac{\rho\sigma_x(y-\mu_y)}{\sigma_y}$$

多维随机向量函数的条件期望：例 0.122

例 0.122 设随机向量 (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 $\mathbb{E}(X|y)$.

解答：例 0.122

题目: 如上所述.

解答:

- 根据条件期望的定义 $\mathbb{E}(X|y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\,\mathrm{d}x$, 计算 Y 的边缘密度函数, 当 $y > 0$ 时

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

由此得到条件分布

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

由此可得

$$\mathbb{E}(X|y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y)\ \mathrm{d}x = \int_0^y \frac{x}{y}\ \mathrm{d}x = \frac{y}{2}.$$

二重期望公式

定理 0.31 [二重期望公式] 设 (X, Y) 是二维随机向量, 且 $\mathbb{E}(X)$ 存在, 则

$$
\begin{aligned}
&\mathbb{E}(X)=\mathbb{E}(\mathbb{E}(X|Y)) \\
&=\begin{cases}\sum_i \mathbb{E}(X|Y=y_i)P(Y=y_i), & \text{若 } Y \text{ 是一个离散型随机变量} \\ \int_{-\infty}^{+\infty}\mathbb{E}(X|Y=y_i)f_Y(y)\ \mathrm{d}y, & \text{若 } Y \text{ 是一个连续型随机变量}\end{cases}
\end{aligned}
$$

二重期望公式在实际中很有用, 譬如, 在计算取值范围很大的 X 的期望 $\mathbb{E}(X)$ 时, 可以通过构建与 X 有关的量 Y, 通过 Y 的不同取值将大范围划分成若干小区域. 先在小区域上求 X 的平均, 再对此类平均求加权平均, 即可得到大范围上 X 的期望 $\mathbb{E}(X)$.

推论: 全期望公式

回顾: 对任意事件 A 而言, 根据全概率公式有 $P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i)$; 即通过对样本空间的切割划分, 将一个复杂事件化成相对简单的事件来求其概率; 借鉴这种化繁为简的思想, 也可以通过全期望公式计算复杂事件的期望, 具体如下:

定理 0.32 设 $A_1,A_2,\ldots,A_n$ 是样本空间 Ω 一个分割, $A_iA_j=\emptyset$ 和 $\Omega=\bigcup_{i=1}^{n}A_i$. 对任意随机变量 X 有

$$\mathbb{E}(X)=\mathbb{E}(X|A_1)P(A_1)+\mathbb{E}(X|A_2)P(A_2)+\cdots+\mathbb{E}(X|A_n)P(A_n),$$

特别的, 随机事件 A 及其对立事件 $\bar{A}$ 构成空间 Ω 一个分割, 对任意随机变量 X 有

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

可证.

证明: 全期望公式

根据二重期望公式定理 0.31, 假设 Y 有 m 个取值

$$
\begin{aligned}
\mathbb{E}(X) = \mathbb{E}_Y\left[\mathbb{E}(X \mid Y)\right] &= \sum_{i=1}^{m} \mathbb{E}(X \mid Y = y_i) P(Y = y_i) \\
&= \sum_{i=1}^{m} \left[\sum_{j} x_j P(x_j \mid Y = y_i)\right] P(Y = y_i) \\
&= \sum_{i=1}^{m} \sum_{j} x_j P(x_j \mid Y = y_i) P(Y = y_i) \\
&= \sum_{k=1}^{n} \sum_{j} x_j P(x_j \mid A_k) P(A_k)
\end{aligned}
$$

最后一步成立于

$$\begin{aligned}\sum_{k=1}^{n} P(x_j \mid A_k) P(A_k) &= \sum_{k=1}^{n} \sum_{k_i : y_{k_i} \in A_k} P(x_j \mid y_{k_i}) P(y_{k_i} \mid A_k) P(A_k) \\ &= \sum_{i=1}^{m} P(x_j \mid Y = y_i) P(Y = y_i)\end{aligned}$$

多维随机向量函数的条件期望：例 0.123

例 0.123 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 小时后回到原处; 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?

解答：例 0.123

题目: 如上所述.

解答:

- 设 X 为该矿工回到安全地点所需的时间, 易知 X 的取值为

$$3, 5+3, 7+3, 5+5+3, 5+7+3, 7+7+3, \ldots$$

直接计算取值范围很大的 X 的期望 $\mathbb{E}(X)$ 较困难, 因此构建变量 Y 表示该矿工第一次选择的门的序号. 因此

$$P(Y=1) = P(Y=2) = P(Y=3) = \frac{1}{3},$$

- 若矿工选择第一个门, 则 $\mathbb{E}(X|Y=1) = 3$; 若矿工选择第二个门, 则 $\mathbb{E}(X|Y=2) = 5+\mathbb{E}(X)$; 若矿工选择第三个门, 则 $\mathbb{E}(X|Y=3) = 7+\mathbb{E}(X)$; 综上所述, 由定理 0.31 有

$$\mathbb{E}(X) = \frac{1}{3}[3+5+\mathbb{E}(X)+7+\mathbb{E}(X)] = 5+\frac{2}{3}\mathbb{E}(X),$$

由此可得 $\mathbb{E}(X) = 15$.

多维随机向量函数的协方差

定义 0.56 设二维随机向量 (X,Y) 的期望 $\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$ 存在, 则称其为X 与 Y 的协方差, 记为

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

从协方差的定义可以看出, 它是 X 的偏差 “$X-\mathbb{E}[X]$” 与 Y 的偏差 “$Y-\mathbb{E}[Y]$” 乘积的数学期望, 由于偏差可正可负, 故协方差也可正可负, 也可为零.

协方差的性质

- 对任意随机变量 X 与 Y, 有

$$\mathrm{Cov}(X,X)=\mathbb{VAR}(X)\quad \text{和} \quad \mathbb{VAR}(X\pm Y)=\mathbb{VAR}(X)+\mathbb{VAR}(Y)\pm 2\mathrm{Cov}(X,Y)$$

- 对任意随机变量 X 与 Y 和常数 c, 有

$$\mathrm{Cov}(X,c)=0 \qquad \text{和} \qquad \mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$$

- 对任意随机变量 X_1、X_2 和 Y, 有

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

- 若随机变量 X 与 Y 相互独立, 则有 $\mathrm{Cov}(X,Y)=0$, 但反之不成立;
- 对任意随机变量 X 与 Y 有

$$(\mathrm{Cov}(X,Y))^2\leq \mathbb{VAR}(X)\cdot\mathbb{VAR}(Y)$$

等号成立的充要条件是 $Y=aX+b$ 几乎处处成立, 即 X 与 Y 之间几乎处处存在线性关系. (可证)

多维随机向量函数的协方差：例 0.124

例 0.124 设随机变量 (X,Y) 的联合概率密度为

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

求协方差 $\mathrm{Cov}(X,Y)$ 和方差 $\sigma(X+Y)$.

解答：例 0.124

题目: 如上所述.

解答:

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

  $$\mathbb{E}[X] = \mathbb{E}[Y] = \int_0^2 \int_0^2 \frac{x(x+y)}{8}\,\mathrm{d}x\,\mathrm{d}y = \frac{7}{6}, \quad \mathbb{E}[XY] = \int_0^2 \int_0^2 \frac{xu(x+y)}{8}\,\mathrm{d}x\,\mathrm{d}y = \frac{4}{3}$$

  由此可得 $\mathrm{Cov}(X, Y) = -1/36$.
- 根据方差的定义 $\sigma(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, 计算

  $$\mathbb{E}[X^2] = \mathbb{E}[Y^2] = \int_0^2 \int_0^2 x^2(x+y)/8\,\mathrm{d}x\,\mathrm{d}y = 5/3$$

  得 $\sigma(X) = \sigma(Y) = 11/36$. 最后得到

  $$\sigma(X+Y) = \sigma(X) + \sigma(Y) + 2\mathrm{Cov}(X, Y) = 5/9.$$

多维随机向量函数的协方差：例 0.125

例 0.125 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.

解答：例 0.125

题目: 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.

解答:

- 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 $X = X_1, X_2, \ldots, X_n$. 随机变量 X_i 得分布列为

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

于是得到期望

$$\mathbb{E}[X] = \mathbb{E}(X_1, X_2, \ldots, X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) \cdots + \mathbb{E}(X_n) = 1.$$

- 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

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

最后根据协方差的性质有

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

插播: 线性运算的基本性质

- 回忆: 矩阵的特征值和特征向量. 以二维为例, 矩阵 $\mathbf{A}$ 拥有特征值 λ_1 和 λ_2, 分别对应特征向量 $\boldsymbol{u}_1, \boldsymbol{u}_2$. 则有

$$\mathbf{A}\boldsymbol{u}_i = \boldsymbol{\Lambda}_i\boldsymbol{u}_i \ , \quad i = 1, 2$$

进而有 $\mathbf{A} = \mathbf{U}^\top \boldsymbol{\Lambda} \mathbf{U}$ with $\mathbf{U}^\top = \mathbf{U}^{-1}$, 即

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1 & 0 \\ 0 & \boldsymbol{\Lambda}_2 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix}$$

- 几何性质. For the case of $\mathbf{A}\boldsymbol{v}$ given any $\boldsymbol{v}$,
  - 矩阵 $\mathbf{U}$ 负责对 $\boldsymbol{v}$ 进行旋转
  - 矩阵 $\boldsymbol{\Lambda}$ 负责对 $\boldsymbol{v}$ 进行放缩

插播: 线性运算的基本性质

如果我们面对一个运算 $\mathbf{A}\boldsymbol{x} + \boldsymbol{b}$,

- $\boldsymbol{b}$ 是平移
- $\mathbf{U}$ 是旋转
- $\boldsymbol{\Lambda}$ 是放缩

二维正态分布的协方差

定理 0.33 若随机向量 $(X,Y)\sim\mathcal{N}(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$, 则

$$\mathrm{Cov}(X,Y)=\rho\,\sigma_x\sigma_y$$

推论 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 相互独立的充要条件是 $\mathrm{Cov}(X,Y)=0$.

可证.

证明：二维正态分布的协方差

二维正态分布的密度函数为

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

这里用到了坐标变换 (归一化做法)

$$f(u,v)=f(x,y)\,|\mathbf{J}|\,,\quad (u,v)=\left(\frac{x-\mu_x}{\sigma_x},\frac{y-\mu_y}{\sigma_y}\right)\,,\quad |\mathbf{J}|=\sigma_x\sigma_y\,.$$

则有

$$\begin{aligned}
\mathrm{Cov}(X,Y)&=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]\\
&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_x)(y-\mu_y)f(x,y)\ \mathrm{d}x\,\mathrm{d}y\\
&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}}\,uv\exp\left(-\frac{u^2+v^2-2\rho uv}{2(1-\rho^2)}\right)\ \mathrm{d}u\,\mathrm{d}v\\
&=\frac{\sigma_x\sigma_y}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}v\exp\left(-\frac{v^2}{2}\right)\left[\int_{-\infty}^{+\infty}u\exp\left(-\frac{(u-\rho v)^2}{2(1-\rho^2)}\right)\mathrm{d}u\right]\mathrm{d}v\\
&=\frac{\sigma_x\sigma_y}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\rho v^2\exp\left(-\frac{v^2}{2}\right)\ \mathrm{d}v=\rho\sigma_x\sigma_y
\end{aligned}$$

协方差与方差

- 方差. 衡量单变量自身的波动性或者偏离性.

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

- 协方差. 衡量变量间的偏离性

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

- 可以定义矩阵 $\mathbf{\Sigma}$ 用以衡量多变量的偏离程度

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbb{VAR}(X) & \mathrm{Cov}(X, Y) \\ \mathrm{Cov}(X, Y) & \mathbb{VAR}(Y) \end{pmatrix}$$

随机向量的数学期望与协方差阵

n 维随机向量的数学期望及方差可以通过矩阵形式给出.

定义 0.57 设 n 维随机向量为 $\boldsymbol{X}=(X_1,X_2,\ldots,X_n)'$, 若每个分量的数学期望都存在, 则称

$$\mathbb{E}[\boldsymbol{X}]=(\mathbb{E}[X_1],\mathbb{E}[X_2],\ldots,\mathbb{E}[X_n])',$$

为 $\boldsymbol{X}$ 的数学期望向量, 简称 $\boldsymbol{X}$ 的数学期望. 而称

$$\begin{aligned}&\mathbb{E}[(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}])(\boldsymbol{X}-\mathbb{E}[\boldsymbol{X}])']\\=&\begin{bmatrix}\sigma_{X_1}^2 & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \sigma_{X_2}^2 & \cdots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & & \vdots\\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \cdots & \sigma_{X_n}^2\end{bmatrix}\end{aligned}$$

为 $\boldsymbol{X}$ 的方差-协方差矩阵, 简称 $\boldsymbol{X}$ 的协方差阵.

随机向量的协方差阵的性质

通过定义 0.57 可以看到 n 维随机向量的各分量的方差构成了协方差阵对角线上的元素, 非对角线的元素为协方差.

定理 0.34 n 维随机向量的协方差阵 $\mathrm{Cov}(\boldsymbol{X}) = (\mathrm{Cov}(X_i, X_j))_{n\times n}$ 是一个对称的非负定矩阵.

Remarks:

- 这说明, 协方差矩阵的特征值是实数的、非负的.
- 多维随机向量函数的协方差：例 0.126

例 0.126 设随机变量 $X_1, X_2, \ldots, X_n$ 相互独立且服从正态分布, 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.

解答：例 0.126

题目：设随机变量 $X_1, X_2, \ldots, X_n$ 相互独立且服从正态分布，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.

解答:

- 根据正态分布的性质易知 $\bar{X}$ 和 $\bar{X} - X_i$ 都服从正态分布 (线性性), 根据定理 0.33 可知正态分布的独立性可通过协方差来研究. 根据协方差的性质有

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

- 根据 $X_1, X_2, \ldots, X_n$ 相互独立有

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

于是得到 $\operatorname{Cov}(\bar{X}, \bar{X} - X_i) = 0$, 根据定理 0.33 的推论可知 $\bar{X}$ 和 $\bar{X} - X_i$ 相互独立.

多维正态分布的联合分布和边缘分布

<u>[命题一：联合推边缘]</u> 若 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$，则有 $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ 和 $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$.

<u>[命题二：边缘推联合]</u> 若 $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ 和 $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$，则 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$?

- X and Y must be linearly correlated.

多维正态分布的联合分布和边缘分布

Counterexamples:

- If X and Y have a nonlinear relationship, they can still be marginally Gaussian, but their joint distribution will not be Gaussian.
  - $X \sim \mathcal{N}(0,1)$ for $X > 0$ and $-|X| \sim \mathcal{N}(0,1)$ for $X \le 0$; and $Y = X$ for $Y > 0$ and $-|Y| \sim \mathcal{N}(0,1)$ for $Y \le 0$.
- If X and Y are drawn from a mixture of Gaussians, then each marginal distribution could be Gaussian, but the joint distribution is not jointly Gaussian.
- If the relationship between X and Y involves some discrete transformation, such as flipping or taking absolute values, the marginals might still be Gaussian, but the joint distribution will not be Gaussian.
  - $X \sim \mathcal{N}(0,1)$ and $Y \sim \text{sign}(X)$.

多维随机向量函数的相关系数

两个随机变量之间的关系可以分为独立与非独立, 其中非独立关系中又可以分为线性关系和非线性关系, 线性相关程度通过线性相关系数来定义.

定义 0.58 设 (X, Y) 为二维随机向量, 如果它们的标准差 σ_x 和 σ_y 存在且都不为零, 则称

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\mathbb{VAR}(X)\,\mathbb{VAR}(Y)}}.$$

为 X 与 Y 的线性相关系数, 简称 相关系数.

相关系数的性质

- 对任意随机变量 X 与 Y, 有 $|\rho_{XY}| \leq 1$. 等号成立的充要条件是 $Y = aX + b$ 几乎处处成立, 即 X 与 Y 之间几乎处处存在线性关系.
  - 若 $\rho_{XY} = 0$, 称 X 与 Y 不相关. 不相关是指 X 与 Y 之间没有线性关系, 但 X 与 Y 之间可能存在其他的函数关系, 比如平方关系、对数关系等;
  - 若 $\rho_{XY} = 1$, 称 X 与 Y 完全正相关; 若 $\rho_{XY} = -1$, 称 X 与 Y 完全负相关;
  - 若 $0 < |\rho_{XY}| < 1$, 称 X 与 Y "有一定程度" 的线性关系; 若 $|\rho_{XY}|$ 越接近于 1, 则线性相关程度越高; 若 $|\rho_{XY}|$ 越接近于 0, 则线性相关程度越低.
- 若随机变量 X 与 Y 相互独立, 则 X 与 Y 不相关, 但反之不成立.

正态分布的相关系数

定理 0.35 若随机向量 $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则有

- X 与 Y 的线性相关系数 $\rho_{XY} = \rho$
- X 与 Y 相互独立充要条件是 X 与 Y 不相关, 即 $\rho = 0$.

Remarks: 独立与不相关的等价性仅限于正态分布随机变量, 对于其他类型不一定成立.

不相关的等价条件

定理 0.36 若随机变量 X 与 Y 的方差存在且都不为零, 以下几个条件相互等价:

- X 与 Y 独立 ;
- X 与 Y 不相关, 即 $\rho_{XY} = 0$;
- 协方差 $\mathrm{Cov}(X, Y) = 0$;
- $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$;
- $\mathbb{VAR}(X \pm Y) = \mathbb{VAR}(X) + \mathbb{VAR}(Y)$.

多维随机向量函数的相关系数：例 0.127

例 0.127 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 $(\alpha, \beta \neq 1)$.

解答：例 0.127

题目: 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 $(\alpha, \beta \neq 1)$.

解答:

- 根据相关系数的定义

$$\rho_{Z_1 Z_2} = \frac{\mathrm{Cov}(Z_1, Z_2)}{\sigma_{Z_1}\sigma_{Z_2}}$$

计算

$$\mathrm{Cov}(Z_1, Z_2) = \mathrm{Cov}(\alpha X + \beta Y, \alpha X - \beta Y) = (\alpha^2 - \beta^2)\sigma^2$$

$$\sigma_{Z_1}^2 = \mathrm{Cov}(\alpha X + \beta Y, \alpha X + \beta Y) = (\alpha^2 + \beta^2)\sigma^2$$

$$\sigma_{Z_2}^2 = \mathrm{Cov}(\alpha X - \beta Y, \alpha X - \beta Y) = (\alpha^2 + \beta^2)\sigma^2$$

由此可得

$$\rho_{Z_1 Z_2} = \frac{\alpha^2 - \beta^2}{\alpha^2 + \beta^2}\,.$$

多维随机向量函数的相关系数：例 0.128

例 0.128 设随机向量 $(X_1, X_2, \ldots, X_n)$ 服从多项分布 $M(m, p_1, p_2, \ldots, p_n)$, 对任意 $i \neq j$, 求 X_i 和 X_j 的相关系数.

解答：例 0.128

题目: 设随机向量 $(X_1, X_2, \ldots, X_n)$ 服从多项分布 $M(m, p_1, p_2, \ldots, p_n)$, 对任意 $i \neq j$, 求 X_i 和 X_j 的相关系数.

解答:

- 根据多项分布的性质, 有边缘分布

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

由此可得 $\sigma(X_i) = mp_i(1 - p_i)$ 和 $\sigma(X_j) = mp_j(1 - p_j)$.

- 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

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

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

$$\mathrm{Cov}(Y_i^k, Y_j^l) = 0 \qquad \text{和} \qquad \mathrm{Cov}(Y_i^k, Y_j^k) = \mathbb{E}[Y_i^k Y_j^k] - \mathbb{E}[Y_i^k]\mathbb{E}[Y_j^k] = -p_i p_j$$

根据协方差的性质有

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

由此可得 X_i 和 X_j 的相关系数

$$\rho = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma(X_i)\sigma(X_j)} = \frac{-m p_i p_j}{\sqrt{m p_i(1-p_i)}\sqrt{m p_j(1-p_j)}} = -\frac{p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}$$

二维正态分布的相关总结

二维随机变量 $(X,Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, 则随机变量 X 和 Y 的边缘分布为

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

- 边缘分布服从正态分布 $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ 和 $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$
- 条件分布服从正态分布

$$(X \mid Y = y) \sim \mathcal{N}\left(\mu_x + \rho(y-\mu_y)\frac{\sigma_x}{\sigma_y}, (1-\rho^2)\sigma_x^2\right)$$

- 正态分布之和是正态分布

$$X + Y \sim \mathcal{N}\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

- 协方差 $\mathrm{Cov}(X,Y) = \rho\,\sigma_x\sigma_y$
- 相关系数

$$\rho = \rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}$$

- X 与 Y 独立, 充要条件, $\mathrm{Cov}(X,Y) = 0$ 或者 $\rho = \rho_{XY} = 0$