

AI 中的数学

第十二讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

① 条件期望

② n 维随机变量

① 条件期望

② n 维随机变量

条件期望的定义：设 X 和 Y 是两个随机变量。

(1) 若在 $Y = y$ 的条件下 X 的可能值是 x_1, x_2, \dots (有限个或无穷可列个)，条件概率分布是 $P(X = x_i | Y = y_i) (i = 1, 2, \dots)$ 则称

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

为在 $Y = y$ 条件下 X 的条件期望，记为 $E(X | Y = y)$ 。

(2) 若在 $Y = y$ 的条件下 X 有条件分布密度 $p_{X|Y}(x|y)$ ，则称积分

$$\int_{-\infty}^{\infty} x p_{X|Y}(x|y) dx$$

为在 $Y = y$ 的条件下 X 的条件期望，记为 $E(X | Y = y)$ 。

关于 y 的条件期望是一个依赖于 y 的随机变量

设二维随机向量 (X, Y) 有联合密度 $p(x, y)$, 有

$$E(X|Y = y) = \frac{1}{p_Y(y)} \int_{-\infty}^{\infty} xp(x, y)dx.$$

定理：设二维随机向量 (X, Y) 有联合密度 $p(x, y)$, 则

$$E(X) = \int_{\{y:p_Y(y)>0\}} E(X|Y = y)p_Y(y)dy.$$

定理：设二维随机向量 (X, Y) 有联合密度 $p(x, y)$ ，则

$$E(X) = \int_{\{y: p_Y(y) > 0\}} E(X|Y=y) p_Y(y) dy.$$

证明：首先，若 $p_Y(y) = 0$ ，则对任何 $A > 0$ 有

$$\left| \int_{-A}^A x p(x, y) dx \right| \leq A \int_{-A}^A p(x, y) dx \leq A \int_{-\infty}^{+\infty} p(x, y) dx = A p_Y(y) = 0,$$

于是

$$\int_{-\infty}^{+\infty} x p(x, y) dx = \lim_{A \rightarrow \infty} \int_{-A}^A x p(x, y) dx = 0$$

可见

$$\begin{aligned} \int_{\{y: p_Y(y) > 0\}} E(X|Y=y) p_Y(y) dy &= \int_{\{y: p_Y(y) > 0\}} \left(\int_{-\infty}^{+\infty} x p(x, y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} x p(x, y) dy \right) dx = \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} p(x, y) dy \right) dx \end{aligned}$$

$$E(X) = \sum_i E(X|Y = y_i)P(Y = y_i).$$
$$\begin{aligned} E(X) &= \sum_k x_k P(X = x_k) = \sum_k x_k \sum_i P(X = x_k, Y = y_i) \\ &= \sum_k \sum_i x_k P(X = x_k | Y = y_i) P(Y = y_i) \\ &= \sum_i E(X | Y = y_i) P(Y = y_i). \end{aligned}$$

例：设 (X, Y) 的联合密度函数为

$$p(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y}, & x, y \in (0, +\infty), \\ 0, & \text{其他.} \end{cases}$$

求 $E(X|Y = y)$ 。

例：设 (X, Y) 的联合密度函数为

$$p(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y}, & x, y \in (0, +\infty), \\ 0, & \text{其他.} \end{cases}$$

求 $E(X|Y = y)$ 。

解：对给定的 $y \in (0, +\infty)$ ，在 $Y = y$ 条件下 X 的条件密度函数为

$$p(x|y) = \begin{cases} \frac{p(x, y)}{p_Y(y)} = \frac{\frac{1}{y} e^{-x/y} e^{-y}}{\int_0^{+\infty} \frac{1}{y} e^{-x/y} e^{-y} dx} = \frac{e^{-x/y}}{y} & x \in (0, +\infty), \\ 0, & x \notin (0, +\infty), \end{cases}$$

因此， X 在给定 $Y = y$ 条件下的条件分布恰好是参数为 $\frac{1}{y}$ 的指数分布。从而

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

例：（对应郑书例 7.7）一矿工在有三个门的矿井中迷了路，第 1 个门通到一个通道，走 2 个小时可到达地面；第 2 个门通到另一个通道，走 3 个小时又回到原处；第 3 个门通到第 3 个通道，沿它走 5 个小时也回到原处，假定该矿工总是等可能从 3 个门选择任意一个进入通道，试问，该矿工到达地面平均需要多长时间。

例：（对应郑书例 7.7）一矿工在有三个门的矿井中迷了路，第 1 个门通到一个通道，走 2 个小时可到达地面；第 2 个门通到另一个通道，走 3 个小时又回到原处；第 3 个门通到第 3 个通道，沿它走 5 个小时也回到原处，假定该矿工总是等可能从 3 个门选择任意一个进入通道，试问，该矿工到达地面平均需要多长时间。

解：设矿工到达地面所需时间为 X ，选择门的编号为 Y ，则 $P(Y = 1) = P(Y = 2) = P(Y = 3) = \frac{1}{3}$ ，于是

$$E(X) = \sum_{i=1}^3 P(Y = i)E(X|Y = i) = \frac{1}{3} \sum_{i=1}^3 E(X|Y = i).$$

易知， $E(X|Y = 1) = 2$ ， $E(X|Y = 2) = E(X) + 3$ ， $E(X|Y = 3) = E(X) + 5$ ，于是

$$E(X) = \frac{1}{3}(2 + E(X) + 3 + E(X) + 5)$$

推知 $E(X) = 10$ ，即矿工到达地面平均要 10 小时。

① 条件期望

② n 维随机变量

- 为 ξ 的联合分布函数, 也记为 F_{ξ} 或 F_{X_1, \dots, X_n} .

- 定义 6.3. 若存在 $p(x_1, \dots, x_n)$ 使得对任意 n 维矩形 D 都有

$$P(\xi \in D) = \int \cdots \int p(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

则称 ξ 为连续型随机向量, 称 $p(x_1, \dots, x_n)$ 为 ξ 的联合密度, 也记为 P_{X_1, \dots, X_n} . (注: 上式对一般 D 都成立).

- 定义 6.4. 对任意 $1 \leq k < n, 1 \leq i_1 < \cdots < i_k \leq n$, 则称 $(X_{i_1}, \dots, X_{i_k})$ 为 ξ 的 (一个 k 维) 边缘, 其分布被称为 ξ 的边缘分布.

例 6.1 (多项分布). 设 U_1, \dots, U_n 相互独立, 都服从如下分布:

$$P(U_i = k) = p_k, \quad k = 1, \dots, t,$$

其中 $t \geq 2, p_k > 0, \forall k$ 且 $p_1 + \dots + p_t = 1$.

- 背景模型: n 次独立重复试验 (投掷一枚 t 面骰子). 记

$$X_k = |\{1 \leq i \leq n : U_i = k\}| = \sum_{i=1}^n 1_{\{U_i=k\}}.$$

- $\xi = (X_1, \dots, X_t)$ 的联合分布列:

$$P(\xi = (i_1, \dots, i_t)) = \frac{n!}{i_1! \dots i_t!} p_1^{i_1} \dots p_t^{i_t}.$$

- 因为 $X_t = n - \sum_{s=1}^{t-1} X_s$, 所以 ξ 与 (X_1, \dots, X_{t-1}) 等价.

例：口袋中有 5 个白球，8 个黑球，从中不放回的依次取出 3 个，若第 i 次取出白球，则 $X_i = 1$ ，否则令 $X_i = 0$ ， $i = 1, 2, 3$ ，求 (X_1, X_2, X_3) 的联合分布列。

解：

$$P(X_1 = 0, X_2 = 0, X_3 = 0) = \frac{8 \times 7 \times 6}{13 \times 12 \times 11} = 0.1958$$

$$\begin{aligned} P(X_1 = 1, X_2 = 0, X_3 = 0) &= P(X_1 = 0, X_2 = 1, X_3 = 0) \\ &= P(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{8 \times 7 \times 5}{13 \times 12 \times 11} = 0.1632 \end{aligned}$$

$$\begin{aligned} P(X_1 = 1, X_2 = 1, X_3 = 0) &= P(X_1 = 0, X_2 = 1, X_3 = 1) \\ &= P(X_1 = 1, X_2 = 0, X_3 = 1) = \frac{5 \times 4 \times 8}{13 \times 12 \times 11} = 0.0932 \end{aligned}$$

$$P(X_1 = 1, X_2 = 1, X_3 = 1) = \frac{5 \times 4 \times 3}{13 \times 12 \times 11} = 0.0035$$

- 定义 6.5. 若对任意 $a_i < b_i, i = 1, \dots, n$ 都有

$$\begin{aligned} & P(a_1 < X_1 < b_1, \dots, a_n < X_n < b_n) \\ &= P(a_1 < X_1 < b_1) \cdots P(a_n < X_n < b_n) \end{aligned}$$

则称 X_1, \dots, X_n 相互独立.

- 若 X_1, \dots, X_n 相互独立, 且 $F_{X_i} = F_{X_1}, i = 2, \dots, n$, 则称 X_1, \dots, X_n 独立同分布.
- 若相互独立, 则上式中的 $a_i < X_i < b_i$ 可以改为 $X_i \in B_i$.

- 相互独立的充要条件与充分条件:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) = f_1(x_1) \cdots f_n(x_n).$$

- 离散型:

$$\begin{aligned} & P(X_1 = x_{i_1}^{(1)}, \dots, X_n = x_{i_n}^{(n)}) \\ &= P(X_1 = x_{i_1}^{(1)}) \cdots P(X_n = x_{i_n}^{(n)}) = p_{i_1}^{(1)} \cdots p_{i_n}^{(n)} \end{aligned}$$

- 连续型 (定理 6.1):

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) = p_1(x_1) \cdots p_n(x_n).$$

- 若 X_i 与 X_j 相互独立, $\forall i \neq j$, 则称 X_1, \dots, X_n 两两独立.
- 例. 甲、乙玩石头剪刀布. 甲出 X , 乙出 Y , 结局为 Z . 则 X, Y, Z 两两独立, 但不相互独立.

例：设随机向量 (X, Y, Z) 在矩形区域
 $a < x < b, c < y < d, e < z < f$ 内服从均匀分布，求 X, Y, Z 的
分布密度函数，以及 X, Y, Z 是否相互独立。

例：设随机向量 (X, Y, Z) 在矩形区域
 $a < x < b, c < y < d, e < z < f$ 内服从均匀分布，求 X, Y, Z 的
分布密度函数，以及 X, Y, Z 是否相互独立。

解：由均匀分布定义

$$f(x, y, z) = \frac{1}{(b-a)(d-c)(f-e)} \quad a < x < b, c < y < d, e < z < f.$$

当 x, y, z 所在边界矩形是独立的，且在矩形内时有：

$$f_X(x) = \int_e^f \int_c^d \frac{1}{(b-a)(d-c)(f-e)} dydz = \frac{1}{b-a}$$

$$f_Y(y) = \int_e^f \int_a^b \frac{1}{(b-a)(d-c)(f-e)} dx dz = \frac{1}{d-c}$$

$$f_Z(z) = \int_c^d \int_a^b \frac{1}{(b-a)(d-c)(f-e)} dx dy = \frac{1}{f-e}.$$

由于 $f(x, y, z) = f_X(x)f_Y(y)f_Z(z)$ ，因此 X, Y, Z 之间相互独立。

定义：设 $\mathbf{X} = (X_1, \dots, X_m)$ 和 $\mathbf{Y} = (Y_1, \dots, Y_n)$ 分别是 m 维和 n 维随机向量，给定 $\mathbf{y} = (y_1, \dots, y_n)$ ，若 $P(\mathbf{Y} = \mathbf{y}) > 0$ ，则 x_1, \dots, x_m 的函数

$$P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathbf{Y} = \mathbf{y})$$

称为在 $\mathbf{Y} = \mathbf{y}$ 条件下 \mathbf{X} 的条件分布函数，记为 $F_{\mathbf{X}|\mathbf{Y}}(x_1, \dots, x_m | \mathbf{y})$.

若 $\mathbf{X} = (X_1, \dots, X_m)$ 和 $\mathbf{Y} = (Y_1, \dots, Y_n)$ 有联合密度 $p(x_1, \dots, x_m, y_1, \dots, y_n)$ ，则

$$F_{\mathbf{X}|\mathbf{Y}}(x_1, \dots, x_m | y_1, \dots, y_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} \frac{p(u_1; \cdots; u_m, y_1; \cdots; y_n)}{p_{\mathbf{Y}}(y_1, \dots, y_n)} du_1 \cdots du_m,$$

这里 $p_{\mathbf{Y}}(y_1, \dots, y_n)$ 是 $\mathbf{Y} = (Y_1, \dots, Y_n)$ 的联合密度，称这里的被积函数为在 $\mathbf{Y} = (y_1, \dots, y_n)$ 条件下 \mathbf{X} 的条件分布密度。

例：设 X_1, X_2, X_3 为独立同分布的连续型随机变量，求 $P(X_3 < X_1 | X_1 = \min \{X_1, X_2\})$.

例：设 X_1, X_2, X_3 为独立同分布的连续型随机变量，求 $P(X_3 < X_1 | X_1 = \min \{X_1, X_2\})$.

解：

$$\begin{aligned} & P(X_3 < X_1 | X_1 = \min \{X_1, X_2\}) \\ &= \frac{P(X_3 < X_1, X_1 = \min \{X_1, X_2\})}{P(X_1 = \min \{X_1, X_2\})} \\ &= \frac{\int_{-\infty}^{+\infty} \int_{x_3}^{+\infty} \int_{x_1}^{+\infty} p(x_2) dx_2 p(x_1) dx_1 p(x_3) dx_3}{\int_{-\infty}^{+\infty} \int_{x_1}^{+\infty} p(x_2) dx_2 p(x_1) dx_1} \\ &= \frac{\int_{-\infty}^{+\infty} \int_{x_3}^{+\infty} (1 - F(x_1)) dF(x_1) dF(x_3)}{\int_{-\infty}^{+\infty} (1 - F(x_1)) dF(x_1)} \\ &= \frac{\int_{-\infty}^{+\infty} \frac{1}{2} - F(x_3) + \frac{1}{2} F^2(x_3) dF(x_3)}{1/2} = \frac{1/6}{1/2} = \frac{1}{3}. \end{aligned}$$

2. n 维随机向量 $\xi = (X_1, \dots, X_n)$ 的数字特征

- 定义 6.6. 称 (EX_1, \dots, EX_n) 为 ξ 的期望, 记为 $E\xi$.
- 定义 6.7. 记 $\sigma_{ij} = \text{cov}(X_i, X_j)$, $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$. 称 $\Sigma = (\sigma_{ij})_{n \times n}$, $\mathbf{R} = (\rho_{ij})_{n \times n}$ 为 ξ 的协方差阵, 相关系数阵.
- 定义 6.8. n 维正态分布. 假设 ξ 有如下的联合密度, 则称 ξ 服从 n 维正态分布, 记为 $\xi \sim N(\vec{\mu}, \Sigma)$.

$$p(\vec{x}) = \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2}(\vec{x} - \vec{\mu})\Sigma^{-1}(\vec{x} - \vec{\mu})^\top \right\}.$$

- $n = 1$ 与 $n = 2$ 的特例已介绍.
- $N(\vec{\mu}, \Sigma)$ 的数字特征: $\mu_i = EX_i, \sigma_{ij} = \text{cov}(X_i, X_j)$.
- X_1, \dots, X_n 相互独立当且仅当 $\sigma_{ij} = 0, \forall i \neq j$.
- 边缘分布, 条件分布都是正态.

例：设随机变量 X_1, X_2, X_3 满足

$$aX_1 + bX_2 + cX_3 = 0,$$

$$E(X_1) = E(X_2) = E(X_3) = d,$$

$$\text{var}(X_1) = \text{var}(X_2) = \text{var}(X_3) = \sigma^2.$$

求相关系数 $\rho_{12}, \rho_{23}, \rho_{31}$.

例：设随机变量 X_1, X_2, X_3 满足

$$aX_1 + bX_2 + cX_3 = 0,$$

$$E(X_1) = E(X_2) = E(X_3) = d,$$

$$\text{var}(X_1) = \text{var}(X_2) = \text{var}(X_3) = \sigma^2.$$

求相关系数 $\rho_{12}, \rho_{23}, \rho_{31}$.

解：对等式 $aX_1 + bX_2 = -cX_3$ 两侧求方差得
 $a^2\sigma^2 + b^2\sigma^2 + 2ab\sigma^2\rho_{12} = c^2\sigma^2$ ，由此解得

$$\rho_{12} = \frac{c^2 - a^2 - b^2}{2ab},$$

同理，对等式 $aX_1 + cX_3 = -bX_2$ 两侧求方差得

$$\rho_{13} = \frac{b^2 - a^2 - c^2}{2ac},$$

同理，对等式 $bX_2 + cX_3 = -aX_1$ 两侧求方差得

$$\rho_{23} = \frac{a^2 - b^2 - c^2}{2bc}.$$

特别的，当 $d \neq 0$ 时，有 $(a + b + c)d = 0$ ，因此 $a + b + c = 0$ ，由此可得

$$c^2 = a^2 + b^2 + 2ab, \quad b^2 = a^2 + c^2 + 2ac, \quad a^2 = b^2 + c^2 + 2bc,$$

代入 $\rho_{12}, \rho_{23}, \rho_{31}$ 表达式得 $\rho_{12} = \rho_{23} = \rho_{31} = 1$ 。

3. n 个随机变量的函数 $Y = f(X_1, \dots, X_n)$

- 定理 6.2 (分布函数法):

$$F_Y(y) = P(f(\xi) \leq y) = \int_{f(x_1, \dots, x_n) \leq y} \cdots \int_1 p(x_1, \dots, x_n) dx_1 \cdots dx_n$$

- 定理 6.3.

$$EY = \int \cdots \int f(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

例 6.3, 6.4, 定义 6.9. 若 X 与 Y 独立, $X \sim \Gamma(r, \lambda)$, $Y \sim \Gamma(s, \lambda)$.
则

$$X + Y \sim \Gamma(r + s, \lambda).$$

- 密度: $p_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}$, $x > 0$.
- $Z = X + Y$: $p_Z(z) = \int p_X(x) p_Y(z - x) dx, \forall z > 0$,

$$\begin{aligned} p_Z(z) &= C \int_0^z x^{r-1} e^{-\lambda x} \cdot (z - x)^{s-1} e^{-\lambda(z-x)} dx \\ &= C e^{-\lambda z} \int_0^1 (tz)^{r-1} ((1-t)z)^{s-1} d(tz) = \hat{C} z^{r+s-1} e^{-\lambda z}. \end{aligned}$$

例: (χ^2 分布) 假设 X_1, \dots, X_n 独立同分布, 都服从 $N(0, 1)$. 于是, $Y_n := X_1^2 + \dots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, 密度为

$$p_n(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad x > 0.$$

其中

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

例: (χ^2 分布) 假设 X_1, \dots, X_n 独立同分布, 都服从 $N(0, 1)$. 于是, $Y_n := X_1^2 + \dots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$, 密度为

$$p_n(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad x > 0.$$

其中

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

利用数学归纳法, 已经证明 (郑书例 5.2) $Y_1 = X_1^2$ 的分布密度是

$$p_1(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}, \quad x > 0.$$

$$X_i^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right).$$

$$\text{于是, } y_n := X_1^2 + \dots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

例：假设 X_1, \dots, X_n 独立同分布，都服从参数为 λ 的指数分布，
则 $Y_n := \sum_{i=1}^n X_i$ ($n \geq 1$) 的分布密度是

$$p_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad (x > 0).$$

例：假设 X_1, \dots, X_n 独立同分布，都服从参数为 λ 的指数分布，
则 $Y_n := \sum_{i=1}^n X_i$ ($n \geq 1$) 的分布密度是

$$p_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad (x > 0).$$

利用数学归纳法假设 X_1, \dots, X_n 独立同分布，都服从 $\text{Exp}(\lambda) = \Gamma(1, \lambda)$.

于是， $T_n := X_1 + \dots + X_n \sim \Gamma(n, \lambda)$.

例 6.6. N 件产品中有 D 件次品. 随机抽 n 件, 包含 X 件次品. 求 EX 与 $\text{var}(X)$. (其中, $N \geq n \geq 2$).

- 随机数目的分解: $X = X_1 + \cdots + X_n$, 其中

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 件是次品;} \\ 0, & \text{若第 } i \text{ 件是合格品.} \end{cases}$$

- 由期望的线性、伯努利分布的期望,

$$EX = \sum_{i=1}^n EX_i = \sum_{i=1}^n P(\text{第 } i \text{ 件是次品}) = n \frac{D}{N}.$$

- $\text{var}(X) = EX^2 - (EX)^2$. 根据对称性,

$$EX^2 = \sum_{i=1}^n EX_i^2 + \sum_{i \neq j} EX_i X_j = nEX_1^2 + n(n-1)EX_1 X_2$$

- 由乘法公式,

$$EX_1 X_2 = P(\text{前两件都是次品}) = \frac{D}{N} \cdot \frac{D-1}{N-1}.$$

- 因此,

$$\begin{aligned}\text{var}(X) &= n \frac{D}{N} + n(n-1) \frac{D}{N} \cdot \frac{D-1}{N-1} - \left(n \frac{D}{N}\right)^2 \\ &= \frac{n(N-n)D(N-D)}{N^2(N-1)}\end{aligned}$$

例：随机向量 $\mathbf{X} = (X_i)_{n \times 1}$, $E(\mathbf{X}) = \mu$, $\text{var}(\mathbf{X}) = \Sigma$, 矩阵 $\mathbf{A}_{n \times n}$, 证明： $E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{A} \Sigma) + \mu^\top \mathbf{A} \mu$.

例：随机向量 $\mathbf{X} = (X_i)_{n \times 1}$, $E(\mathbf{X}) = \mu$, $\text{var}(\mathbf{X}) = \Sigma$, 矩阵 $\mathbf{A}_{n \times n}$, 证明： $E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{A} \Sigma) + \mu^\top \mathbf{A} \mu$.

证明：

$$\begin{aligned} E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) &= E(\text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})) = E(\text{tr}(\mathbf{A} \mathbf{X} \mathbf{X}^\top)) \\ &= \text{tr}(E(\mathbf{A} \mathbf{X} \mathbf{X}^\top)) = \text{tr}(\mathbf{A} E(\mathbf{X} \mathbf{X}^\top)) \\ &= \text{tr}(\mathbf{A} (\text{var}(\mathbf{X}) + \mu \mu^\top)) = \text{tr}(\mathbf{A} \Sigma) + \text{tr}(\mathbf{A} \mu \mu^\top) \\ &= \text{tr}(\mathbf{A} \Sigma) + \mu^\top \mathbf{A} \mu. \end{aligned}$$

定理：设 $\xi = (X_1, \dots, X_n)$ 的协方差阵为 Σ , 且

$$Y_i = \sum_{j=1}^n a_{ij} X_j, j = 1, \dots, m.$$

记 $\mathbf{A} = (a_{ij})_{m \times n}$, $\eta = (Y_1, \dots, Y_m)$, 则

$$(E(\eta))^{\top} = \mathbf{A}(E(\xi))^{\top},$$

$$\text{cov}(\eta, \eta) = \mathbf{A}\Sigma\mathbf{A}^{\top}.$$

定理：设 $\xi = (X_1, \dots, X_n)$ 的协方差阵为 Σ , 且

$$Y_i = \sum_{j=1}^n a_{ij} X_j, j = 1, \dots, m.$$

记 $\mathbf{A} = (a_{ij})_{m \times n}$, $\eta = (Y_1, \dots, Y_m)$, 则

$$(E(\eta))^{\top} = \mathbf{A}(E(\xi))^{\top},$$

$$\text{cov}(\eta, \eta) = \mathbf{A}\Sigma\mathbf{A}^{\top}.$$

证明：由于 $E(Y_i) = \sum_{j=1}^n a_{ij} E(X_j)$, 故 $(E(\eta))^{\top} = \mathbf{A}(E(\xi))^{\top}$ 成立, 又由于 $Y_i - E(Y_i) = \sum_{j=1}^n a_{ij}(X_j - E(X_j))$, 知

$$(Y_i - E(Y_i))(Y_k - E(Y_k)) = \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} (X_j - E(X_j))(X_l - E(X_l)).$$

于是

$$\begin{aligned}\operatorname{cov}(Y_i, Y_k) &= \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} E(X_j - E(X_j))(X_l - E(X_l)) \\ &= \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} \sigma_{jl},\end{aligned}$$

这里 $\sigma_{jl} = \operatorname{cov}(X_j, X_l)$.

由于 $\Sigma = (\sigma_{jl})_{n \times n}$, 知 $\operatorname{cov}(\eta, \eta) = \mathbf{A}\Sigma\mathbf{A}^\top$ 成立。

n 个随机变量的多个函数

- 定理 6.4. 设 $\xi = (X_1, \dots, X_n)$ 为连续型,

$$f: A \rightarrow G, \quad (x_1, \dots, x_n) \mapsto (y_1, \dots, y_n)$$

一对一, C^1 且 $J = \frac{\partial (y_1, \dots, y_n)}{\partial (x_1, \dots, x_n)} \neq 0$. 则 $\eta = (Y_1, \dots, Y_n)$ 是连续型, 且.

$$p_\eta(y_1, \dots, y_n) = p_\xi(x_1, \dots, x_n) |J^{-1}|, (y_1, \dots, y_n) \in G.$$

次序统计量：设 n 个随机变量 X_1, \dots, X_n ，将它们从小到大排列：

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

称 $X_{(k)}$ 为第 k 个次序统计量.

次序统计量：设 n 个随机变量 X_1, \dots, X_n ，将它们从小到大排列：

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

称 $X_{(k)}$ 为第 k 个次序统计量.

例：设 X_1, \dots, X_n 独立同分布，都服从 $U(0, 1)$. 已知对于 $\forall 0 < x < 1$,

$$P(X_{(k)} \leq x) = \sum_{i=k}^n \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i}.$$

求 $E(X_{(k)})$ 与 $\text{var}(X_{(k)})$.

- $\forall 0 < x < 1,$

$$P(X_{(k)} \leq x) = \sum_{i=k}^n \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i}.$$

- $k \leq i \leq n-1$, 上式单项的导数是

$$\begin{aligned} &= \frac{n!}{i!(n-i)!} (ix^{i-1}(1-x)^{n-i} - x^i(n-i)(1-x)^{n-i-1}) \\ &= \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} - \frac{n!}{i!(n-i-1)!} x^i(1-x)^{n-i-1} \\ &= a_{i-1} - a_i, \end{aligned}$$

- $i = n$ 时, $(x^n)' = a_{n-1}$, 于是, $\forall 0 < x < 1,$

$$p_{X_{(k)}}(x) = \sum_{i=k}^{n-1} (a_{i-1} - a_i) + a_{n-1} = a_{k-1}.$$

- 已有 $q_k(x) := p_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k}$.
- $\forall \ell, m \geq 1$,

$$\begin{aligned} \int_0^1 x^\ell (1-x)^m dx &= \frac{1}{\ell+1} \int_0^1 (1-x)^m dx^{\ell+1} \\ &= -\frac{1}{\ell+1} \int_0^1 x^{\ell+1} d(1-x)^m = \frac{m}{\ell+1} \int_0^1 x^{\ell+1} (1-x)^{m-1} dx \\ &= \cdots = \frac{m!}{(\ell+1) \cdots (\ell+m)} \int_0^1 x^{\ell+m} dx = \frac{\ell! m!}{(\ell+m+1)!} \end{aligned}$$

- 期望: 取 $\ell = k, m = n - k$, 知

$$\begin{aligned} EX_{(k)} &= \int_0^1 x q_k(x) dx = \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^k (1-x)^{n-k} dx \\ &= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}. \end{aligned}$$

- 已有 $q_k(x) := p_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1} (1-x)^{n-k}$.

$$\int_0^1 x^\ell (1-x)^m dx = \frac{\ell! m!}{(\ell + m + 1)!}.$$

- 二阶矩：取 $\ell = k + 1, m = n - k$,

$$\begin{aligned} EX_{(k)}^2 &= \int_0^1 x^2 q_k(x) dx = \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^{k+1} (1-x)^{n-k} dx \\ &= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)} \end{aligned}$$

- 方差：

$$\begin{aligned} \text{var}(X_{(k)}) &= EX_{(k)}^2 - (EX_{(k)})^2 = \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} \\ &= \frac{k^2(n+1) + k(n+1) - k^2(n+2)}{(n+1)^2(n+2)} = \frac{k(n+1-k)}{(n+1)^2(n+2)} \end{aligned}$$