Multivariate normal distributions (多元常態分布)

- Definition. The distribution of $(X_1, \ldots, X_m)$ is a multivariate normal distribution means that there exist IID $N(0,1)$ random variables Z_1, ..., Z_ℓ such that each X_i is a linear combination of Z_1, ..., Z_ℓ plus a constant. That is,

$$X_i = \mu_i + a_{i,1}Z_1 + \cdots + a_{i,\ell}Z_\ell, \qquad i = 1, \ldots, m,$$

  where μ_i, $a_{i,1}, \ldots, a_{i,\ell}$ are constants for $i = 1, \ldots, m$. That is,

$$(X_1, \ldots, X_m)^T = (\mu_1, \ldots, \mu_m)^T + A(Z_1, \ldots, Z_\ell)^T,$$

  where A is the $m \times \ell$ matrix whose (i,j)-th element is $a_{i,j}$ for $1 \le i \le m$ and $1 \le j \le \ell$.

  - When $m = 1$, the multivariate normal distribution is a (univariate) normal distribution.
  - When $m = 2$, the multivariate normal distribution is called a bivariate normal distribution (二元常態分布).

- Fact 1 Suppose that the distribution of $(X_1, \ldots, X_m)$ is a multivariate normal distribution. Let $\boldsymbol{\mu} = (E(X_1), \ldots, E(X_m))^T$ and Σ be the covariance matrix of $(X_1, \ldots, X_m)$, and let $M_{X_1,\ldots,X_m}$ be the joint MGF of $(X_1, \ldots, X_m)$, then

$$M_{X_1,\ldots,X_m}(\boldsymbol{t}) = e^{\boldsymbol{\mu}^T\boldsymbol{t}+0.5\boldsymbol{t}^T\Sigma\boldsymbol{t}} \tag{1}$$

  for $\boldsymbol{t} = (t_1, \ldots, t_m)^T \in R^m$. If Σ^{-1} exists, then $(X_1, \ldots, X_m)$ has a joint PDF f, where

$$f(\boldsymbol{x}) = (2\pi)^{-m/2}(\det(\Sigma))^{-1/2}e^{-(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2} \tag{2}$$

  for $\boldsymbol{x} = (x_1, \ldots, x_m)^T \in R^m$. Here $\det(\Sigma)$ denotes the determinant of Σ.

- Note that (1) follows from direct calculation and (2) is left as a homework problem.

- From (1), we know that when the distribution of $\boldsymbol{X} = (X_1, \ldots, X_m)$ is a multivariate normal distribution, then the MGF of the multivariate normal distribution is determined by $\boldsymbol{\mu}$: the mean vector of $\boldsymbol{X}$ and Σ: the covariance matrix of $\boldsymbol{X}$ and we denote this multivariate normal distribution by $N(\boldsymbol{\mu}, \Sigma)$.

- Fact 2 Suppose that the distribution of $(X_1, \ldots, X_m)$ is a multivariate normal, A is an $n \times m$ matrix of constants and $\boldsymbol{b}$ is a column vector of n constants. Then the distribution of $A(X_1, \ldots, X_m)^T + \boldsymbol{b}$ is also a multivariate normal distribution.

- Fact 3 Suppose that the distribution of $(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ is multivariate normal. If $Cov(X_i, Y_j) = 0$ for $1 \le i \le m$, $1 \le j \le n$, then the two random vectors $(X_1, \ldots, X_m)$ and $(Y_1, \ldots, Y_n)$ are independent.

- Results in Facts 2 and 3 can be verified using the MGF of a multivariate normal distribution given in Fact 1.

- Example 1. Suppose that $X_1, \ldots, X_n$ are IID random variables and $X_1 \sim N(\mu, \sigma^2)$, where $\mu \in (-\infty, \infty)$ and $\sigma > 0$. Show that $\bar{X} = \sum_{i=1}^n X_i/n \sim N(\mu, \sigma^2/n)$.

  A sketch of solution. Let $Z_i = (X_i - \mu)/\sigma$ for $i = 1, \ldots, n$, then $Z_1, \ldots, Z_n$ are IID $N(0,1)$ and the distribution of $(X_1, \ldots, X_n)$ is a multivariate distribution. By Fact 2, the distribution of $\bar{X}$ is a univariate normal distribution. Let $N(\mu_1, \sigma_1^2)$ be the distribution of $\bar{X}$, then $\mu_1 = E(\bar{X})$ and $\sigma_1^2 = Var(\bar{X})$. Thus we can obtain the result in this example by calculating $E(\bar{X})$ and $Var(\bar{X})$.

- Example 2. Suppose that

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 3 & 25 \end{pmatrix} \right).$$

  Find a constant b such that $Y - bX$ is independent of X.

  Sol. Since the distribution of (X, Y) is a bivariate normal distribution, by Fact 2, the distribution of $(Y - bX, X)$ is also a bivariate normal distribution. By Fact 3,

$$Y - bX \text{ is independent of } X \Leftrightarrow Cov(Y - bX, X) = 0.$$

  Thus we take

$$b = \frac{Cov(Y, X)}{Var(X)} = \frac{3}{1} = 3$$

  so that $Y - bX$ and X are independent.

- Best linear predictor for the multivariate case. Suppose that $Y, X_1, \ldots, X_k$ are random variables with finite first and second moments. The best linear predictor of Y based on $X_1, \ldots, X_k$ is $a_0 + b_{1,0}X_1 + \cdots + b_{k,0}X_k$, where

$$(a_0, b_{1,0}, \ldots, b_{k,0}) = \arg\min_{a, b_1, \ldots, b_k} E(Y - (a + b_1 X_1 + \cdots + b_k X_k))^2.$$

  Let $\boldsymbol{b}_0 = (b_{1,0}, \ldots, b_{k,0})^T$, $\boldsymbol{X} = (X_1, \ldots, X_k)^T$ and $\Sigma_{\boldsymbol{X}}$ be the covariance matrix of $\boldsymbol{X}$, then it can be shown that

$$\Sigma_{\boldsymbol{X}} \boldsymbol{b}_0 = (Cov(X_1, Y), \ldots, Cov(X_k, Y))^T \qquad (3)$$

and

$$a_0 = E(Y) - E(\boldsymbol{b}_0^T \boldsymbol{X}) = E(Y) - E(b_{1,0}X_1 + \cdots + b_{k,0}X_k). \qquad (4)$$

Note that (3) is equivalent to $Cov(Y - (a_0 + b_{1,0}X_1 + \cdots + b_{k,0}X_k), X_i) = 0$ for $i = 1, \ldots, k$.

- Best linear predictor of a random vector. Suppose that $(Y_1, \ldots, Y_m)^T$ and $\boldsymbol{X}$ are random vectors. For $j \in \{1, \ldots, m\}$, let $\hat{Y}_j$ be the best linear predictor of Y_j based on $\boldsymbol{X}$. Then the best linear predictor of $(Y_1, \ldots, Y_m)^T$ is $(\hat{Y}_1, \ldots, \hat{Y}_m)^T$.

- Fact 4 Suppose that $\boldsymbol{X} = (X_1, \ldots, X_m)^T$, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, and the distribution of $(\boldsymbol{X}^T, \boldsymbol{Y}^T)$ is a multivariate normal distribution. Let Y_i^* be the best linear predictor of Y_i based on $\boldsymbol{X}$ for $i = 1, \ldots, n$, and let $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_n^*)^T$, then (i) and (ii) hold.

  (i) $\boldsymbol{Y} - \boldsymbol{Y}^*$ and $\boldsymbol{X}$ are independent.

  (ii) Let $B\boldsymbol{X} + \boldsymbol{a} = \boldsymbol{Y}^*$. If the covariance matrix of $(\boldsymbol{X}^T, \boldsymbol{Y}^T)$ is invertible, then a conditional PDF of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ is the PDF of $N(\boldsymbol{\mu}, \Sigma)$ given in (2) with $\boldsymbol{\mu} = B\boldsymbol{x} + \boldsymbol{a}$ and $\Sigma = E(\boldsymbol{Y} - \boldsymbol{Y}^*)(\boldsymbol{Y} - \boldsymbol{Y}^*)^T$.

  Note that (i) follows from the fact that the covariance between a component of $\boldsymbol{Y} - \boldsymbol{Y}^*$ and a component of $\boldsymbol{X}$ is zero, and (ii) follows from (i).

- Example 3. Suppose that

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \right).$$

  (a) Find the best linear predictor of X based on Y and Z.

  (b) Find a version of the conditional PDF of X given (Y, Z).

  (c) Find a version of the conditional PDF of (X, Y) given Z.

  A sketch of solution.

  (a) Let $a_0 + b_{1,0}Y + b_{2,0}Z$ be the best linear predictor of X based on Y and Z. Solving $Cov(X - (a_0 + b_{1,0}Y + b_{2,0}Z), Y) = 0$ and $Cov(X - (a_0 + b_{1,0}Y + b_{2,0}Z), Z) = 0$ gives $b_{1,0} = -2/3$ and $b_{2,0} = -1/3$, so $a_0 = E(X) - E(b_{1,0}Y + b_{2,0}Z) = 4/3$ and the best linear predictor of X based on Y and Z is $(-2/3)Y + (-1/3)Z + (4/3)$.

(b) A version of the conditional PDF of X given (Y, Z) is $\{f_{\mu(y,z),\Sigma} : (y, z) \in R^2\}$, where $f_{\mu(y,z),\Sigma}$ is the PDF of $N(\boldsymbol{\mu}, \Sigma)$ given in (2) with $\boldsymbol{\mu} = \mu(y, z) = (-2/3)y+(-1/3)z+(4/3)$ and $\Sigma = E(X-((-2/3)Y+(-1/3)Z + (4/3)))^2 = Var(X + (2/3)Y + (1/3)Z) = 4/3$.

(c) The best linear predictor of Y based on Z is $(-1/2)Z+2$ and the best linear predictor of X based on Z is 0, so a version of the conditional PDF of $(X, Y)^T$ given Z is $\{f_{\mu(z),\Sigma} : z \in R\}$, where $f_{\mu(z),\Sigma}$ is the PDF of $N(\boldsymbol{\mu}, \Sigma)$ given in (2) $\boldsymbol{\mu} = \mu(z) = (0, (-1/2)z + 2)^T)$ and

$$\Sigma = E(X, Y-((-1/2)Z+2))^T(X, Y-((-1/2)Z+2)) = \begin{pmatrix} 2 & -1 \\ -1 & 3/2 \end{pmatrix}.$$