L01: Multivariate distributions

1. Multivariate distributions

   (1) Probability density function

   $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ is a continuous random vector if it assumes values $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in D \subset R^p$ with probability density $f(x_1, ..., x_p)$. This $f(\cdot)$ is called the joint probability density function (pdf) for X. D, the domain of $f(\cdot)$, is called the support of X. For $A \subset D$,

   $$P(X \in A) = \iint_A f(x_1, .., x_p) dx_1, .., dx_p.$$

   One can extend the domain of $f(\cdot)$ from D to R^p by defining $f(x_1, .., x_p) \equiv 0$ on D^c.

   (2) Conditions for $f(x_1, ..., x_p)$ to be a pdf

   Function $f(x_1, ..., x_p)$ could be used as a pdf to define a random vector if

   $$f(x_1, ..., x_p) \geq 0 \text{ and } \iint_{R^p} f(x_1, .., x_p) dx_1, .., dx_p = 1.$$

   (3) pdf of $Y = y(X)$

   $X \in R^p$ is a random vector with pdf $f(x_1, .., x_p)$. Suppose $y = y(x) \in R^p \iff x = x(y) \in R^p$ with Jacobian $J = \frac{\partial\, (x_1, .., x_p)}{\partial\, (y_1, .., y_p)} \in R^{p \times p}$. Then the pdf for $Y = y(X)$ is

   $$g(y) = f(x(y)) \cdot \text{abs}|J|$$

   Proof. First $g(y) = f(x(y))\, \text{abs}|J| \geq 0$. Suppose $x \in A \iff y \in B$. Then By the substitution of $x = x(y)$ in integral,

   $$\begin{aligned} P(Y \in B) &= P(X \in A) = \textstyle\iint_{x \in A} f(x) dx_1, .., dx_p = \iint_{y \in B} f(x(y)) \text{abs}|J| dy_1, .., dy_p \\ &= \textstyle\iint_{y \in B} g(y) dy_1, ..dy_p. \end{aligned}$$

   Hence $g(y)$ is the pdf for Y.

2. Multivariate normal distributions

   (1) Definition of $X \sim N(\mu, \Sigma)$.

   Consider $f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right]$, a function of $x \in R^p$ with parameter vector $\mu \in R^p$ and a positive definite matrix $\Sigma \in R^{p \times p}$. Clearly $f(x) > 0$.

   By the substitution $z = \Sigma^{-1/2}(x - \mu) \iff x = \Sigma^{1/2} z + \mu$ with $J = \frac{\partial (x_1, .., x_p)}{\partial\, (z_1, .., z_p)} = \Sigma^{1/2}$.

   $$\iint_{R^p} f(x) dx_1, .., dx_p = \iint_{R^p} \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2} z' z\right) dz_1, ..dz_p = \prod_{i=1}^{p} \int_R \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} dz_i = 1$$

   So $f(x)$ is a pdf. X with this pdf is called a normal vector denoted by $X \sim N(\mu, \Sigma)$.

   (2) Transformation

   $X \sim N(\mu, \Sigma)$. If $A \in R^{p \times p}$ is non-singular, then $Y = AX + b \sim N(A\mu + b, A\Sigma A')$

   Prof. Transformation $y = Ax + b \iff x = A^{-1}(y - b)$ has $J = \frac{\partial\, (x_1, .., x_p)}{\partial\, (y_1, .., y_p)} = A^{-1}$.

   So the pdf for Y is

   $$\begin{aligned} g(y) &= f(A^{-1}(y - b)) |A^{-1}| \\ &= \tfrac{1}{(2\pi)^{p/2} |A\Sigma A'|^{1/2}} \exp\left[-\tfrac{1}{2}(y - A\mu - b)' (A\Sigma A')^{-1} (y - A\mu - b)\right]. \end{aligned}$$

   Thus $Y \sim N(A\mu + b, A\Sigma A')$.

Ex1: $X \sim N(\mu, \Sigma)$.

For full row rank A there exists B such that $\begin{pmatrix} A \\ B \end{pmatrix} \in R^{p\times p}$ is non-singular and $A\Sigma B' = 0$.

Define $\begin{pmatrix} Y \\ Y_a \end{pmatrix} = \begin{pmatrix} AX + b \\ BX + c \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} X + \begin{pmatrix} b \\ c \end{pmatrix}$. By the transformation in (2),

$$\begin{pmatrix} Y \\ Y_a \end{pmatrix} \sim N\left(\begin{pmatrix} A\mu + b \\ B\mu + c \end{pmatrix}, \begin{pmatrix} A\Sigma A' & 0 \\ 0 & B\Sigma B' \end{pmatrix} \right)$$

By the pdf in (1), $f_{Y,\,Y_a}(y,\,y_a) = f(y) \cdot g(y_a)$ where $f(y)$ is pdf for $N(A\mu + b,\, A\Sigma A')$ and $g(y_a)$ is the pdf for $N(B\mu + c,\, B\Sigma B')$.

3. Marginal distributions and conditional distributions

(1) Definitions

X is a random vector. X_I contains some of the components of X, and X_{II} contains the rest components.

The distribution of X_I, disregarding X_{II} at all, is a marginal distribution of X. Distribution of X_I, knowing that X_{II} is fixed at x_{II}, is the conditional distribution of X_I given $X_{II} = x_{II}$.

For simplicity assume $X = \begin{pmatrix} X_I \\ X_{II} \end{pmatrix}$ where $X_I \in R^k$.

(2) Marginal pdf

$g(x_I) = \iint_{R^{p-k}} f(x_1, .., x_p) dx_{k+1}, ..., dx_p$ is the marginal pdf of X_I.

Pf: First $g(x_1, ..., x_k) \geq 0$. Secondly, for $X_I \in A$,

$$\begin{aligned} P(X_I \in A) &= P(X_I \in A,\, X_{II} \in R^{p-k}) = \iint_{x_I \in A\, x_{II} \in R^{p-k}} f(x_1, .., x_p)\, dx_1, .., dx_p \\ &= \iint_{x_I \in A} \left[\iint_{x_{II} \in R^{p-k}} f(x_1, .., x_p)\, dx_{k+1}, .., dx_p \right] dx_1, .., dx_k \\ &= \iint_{x_I \in A} g(x_1, ..., x_k)\, dx_1, ..., dx_k. \end{aligned}$$

So $g(x_I)$ is the pdf for X_I.

(3) Conditional pdf

The conditional pdf of X_I given $X_{II} = x_{II}$ is $f_{X_I|X_{II}=x_2}(x_I) = \frac{f(x_I, x_{II})}{f_{X_{II}}(x_{II})}$.

Explanation: At fixed $X_I = x_I$, to get the value for the pdf of X_I, we collected pdf values at all points of X_{II}. But if X_{II} is fixed at $X_{II} = x_{II}$, the only value available is $f(x_I, x_{II})$. Therefore initially $f(x_I, x_I I)$ treated as a function of x_I can be regarded as the pdf of X_I while X_{II} is fixed at x_{II}. But it is not a pdf since $\iint_{R^k} f(x_I,\, x_{II}) dx_1, ..dx_k \neq 1$. $\frac{f(x_I, x_{II})}{f_{X_{II}}(x_{II})}$ is proportional to $f(x_I, x_{II})$, and it is a pdf.

Ex2: Improvement to transformation in (2)of 2.

$X \sim N(\mu, \Sigma)$. If A has full row rank, then $Y = AX + b \sim N(A\mu + b,\, A\Sigma A'$.

Proof. In Ex1, we obtained the pdf for $\begin{pmatrix} Y \\ Y_a \end{pmatrix}$. The pdf for Y is

$$\iint_{y_a \in R^{p-k}} f_{Y,Y_a}(y,\, y_a) dy_a = \iint_{y_a \in R^{p-k}} f_1(y) f_2(y_a) dy_a = f_1(y)$$

But $f_1(y)$ is the pdf for $N(A\mu + b,\, A\Sigma A')$. So $Y = AX + B \sim N(A\mu + b,\, A\Sigma A')$.

Ex3: Marginal distributions for $N(\mu, \Sigma)$ can be obtained by the transformations in Ex2. For example with $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix} \right)$, $X_2 = (0, 1, 0)X \sim N(\mu_2,\, \sigma_2^2)$ and $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} X \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{31} & \sigma_3^2 \end{pmatrix} \right)$.

L02: Parameters of multivariate distributions

1. Conditional distribution for normal vectors

   (1) Recall

   (i) For $\begin{pmatrix} X \\ Y \end{pmatrix}$ with pdf $f(x, y)$, the conditional pdf of X given $Y = y$ is $f_{X|y}(x) = \frac{f(x,y)}{f_Y(y)}$ where $f_Y(y)$ is the marginal pdf for y.

   (ii) For $X \sim N(\mu, \Sigma)$, $AX + b \sim N(A\mu + b, A\Sigma A')$ for full row rank A

   (2) For $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$, let $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.
   Then $X|(Y = y) \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2), \Sigma_{11.2}\right)$

   Proof. $\begin{pmatrix} Z \\ Y \end{pmatrix} = \begin{pmatrix} X - \Sigma_{12}\Sigma_{22}^{-1}Y \\ Y \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 - \Sigma_{11}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)$
   and $Y = (0, I)\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu_2, \Sigma_{22})$.
   By examining the pdfs we see $f_{Z,Y}(z, y) = f_1(z)f_2(y)$ where
   $f_1(z)$ is the pdf for $N\left(\mu_1 - \Sigma_{11}\Sigma_{22}^{-1}\mu_2, \Sigma_{11.2}\right)$ and $f_2(y)$ is the pdf for $N(\mu_2, \Sigma_{22})$.
   So $f_{Z|Y=y}(z) = \frac{f_{Z,Y}(z)}{f_Y(y)} = \frac{f_1(z)f_2(y)}{f_2(y)} = f_1(z)$. Thus $Z|y \sim N\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11.2}\right)$.
   Hence $Z + \Sigma_{11}\Sigma_{22}^{-1}y|Y = y \sim N\left(\mu_1 + \Sigma_{11}\Sigma_{22}^{-1}(y - \mu_2), \Sigma_{11.2}\right)$ Therefore

   $$X|y \sim N\left(\mu_1 + \Sigma_{11}\Sigma_{22}^{-1}(y - \mu_2), \Sigma_{11.2}\right)$$

   Ex1: $Y|(X = x) \sim N\left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1), \Sigma_{22.1}\right)$ where $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

2. Expectation matrices

   (1) Definition
   $X \in R^p$ is a random vector with joint pdf $f(x_1, .., x_p)$, and $G(X) = (g_{ij}(X))_{m\times n}$ is a matrix valued function of X. Define the expectation of $G(X)$ by $E[G(X)] = (E[g_{ij}(X)])_{m\times n}$ where $E[g_{ij}(X)] = \iint_{R^p} g(x_1, ..., x_p)\, f(x_1, ..., x_p)\, dx_1, ..., dx_p$.

   (2) Property of the operator $E(\cdot)$
   $E(\cdot)$ is a linear operator, i.e., $E[AG(X)B + CH(X)D + K] = AE[G(X)]B + CE[H(X)]D + K$

   (3) Computation for $E[g_{ij}(X_{k+1}, ..., X_p)]$

   $$\begin{aligned} E[g_{ij}(X_{k+1}, .., X_p)] &= \iint_{R^p} g(x_{k+1}, ..., x_p)\, f(x_1, ..., x_p) dx_1, ..dx_p \\ &= \iint_{R^{p-k}} g(x_{k+1}, ..., x_p) \left[\iint_{R^k} f(x_1, ..., x_p)\, dx_1, ..., dx_k\right] dx_{k+1}, ..., dx_p \\ &= \iint_{R^{p-k}} g(x_{k+1}, .., x_p)\, f_{(X_{k+1}, .., X_p)}(x_{k+1}, ..., x_p)\, dx_{k+1}, ..., dx_p \end{aligned}$$

   Ex2: $XX' = (X_iX_j)_{p\times p}$ is a random matrix. $E(XX') = (E(X_iX_j))_{p\times p}$ where
   $E(X_iX_j) = \iint_{R^p} x_ix_jf(x_1, .., x_p)\, dx_1, ..., dx_p = \iint_{R^2} x_ix_jf_{(X_i, X_j)}(x_i, x_j)dx_idx_j$.

3. Mean, variance, covariance and correlation coefficient

   (1) Mean for $X \in R^p$
   The mean of random vector $X \in R^p$ is $\mu = E(X) \in R^p$.
   The mean of X is regarded as the center of the values of X since $E(X - \mu) = 0$

   (2) Covariance matrix
   The covariance matrix for random vectors $X \in R^p$ and $Y \in R^q$ with joint distributions is

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

Let $\mu_x = E(X) \in R^p$ and $\mu_y = E(Y) \in R^q$. Then $\text{Cov}(X, Y)$ is a $p \times q$ matrix. The element on the ith row and jth column is

$$E[(X_i - \mu_{x_i})(Y_j - \mu_{y_j})] = E(X_iY_j) - E(X_i)E(Y_j) = \text{cov}(X_i, Y_j)$$

Thus $\text{Cov}(X, Y) = (\text{cov}(X_i, Y_j))_{p\times q}$.

(3) Variance-covariance for $X \in R^p$
For random vector $X \in R^p$ with $E(X) = \mu$,

$$\text{Cov}(X, X) = E[(X-\mu)(X-\mu)'] = E(XX') - \mu\mu' = (\text{cov}(X_i, X_j))_{p\times p}\,.$$

With $i = j$, $\text{cov}(X_i, X_i) = \text{var}(X_i)$. Thus the matrix is called the variance-covariance matrix or covariance matrix for X, and is often denoted by $\text{Cov}(X) = \Sigma$.
$E(X) = \mu$ and $\text{Cov}(X) = \Sigma$ are two important parameter vector and matrix for X and is often given using the form $X \sim (\mu, \Sigma)$.
The variance matrix for X is $V_x = \text{diag}(\text{var}(X_1), ...\text{var}(X_p))$. With $\Sigma = \text{Cov}(X)$, V_x is often denoted as $\text{diag}(\Sigma)$.

(4) Correlation matrix
For random vectors $X \in R^p$ and $Y \in R^q$ with joint distributions, the correlation matrix

$$\rho(X, Y) = V_x^{-1/2}\text{Cov}(X, Y)V_y^{-1/2} = \left(\frac{\text{cov}(X_i, Y_j)}{\sqrt{\text{var}(X_i), \text{var}(Y_j)}}\right)_{p\times q} = (\rho(X_i, Y_j))_{p\times q}$$

where $\rho(X_i, Y_j)$ is the correlation of X_i and Y_j.
$\rho(X) = \rho(X, X) = (\rho(X_i, X_j))_{p\times p}$.
$\rho(X_i, X_i) = 1$.

(5) Operator $\text{Cov}(\cdot, \cdot)$.

$$\begin{aligned} & \text{Cov}(AX_I + BX_{II} + \alpha, CY_I + DY_{II} + \beta) \\ = \; & A\text{Cov}(X_I, Y_I)C' + A\text{Cov}(X_I, Y_{II})D' + B\text{Cov}(X_{II}, Y_I)C' + B\text{Cov}X_{II}, Y_{II})D' \end{aligned}\,.$$

(6) Parameters for normal distributions
$X \sim N(\mu, \Sigma) \Longrightarrow X \sim (\mu, \Sigma)$.

Ex3: p107 2.30 (j)

$$X^{(1)} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\; X^{(2)} = \begin{pmatrix} X_3 \\ X_4 \end{pmatrix},\; X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} \sim (\mu, \Sigma),\; \mu = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix},\; \Sigma_x = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 9 & -2 \\ 2 & 0 & -2 & 4 \end{pmatrix}.$$

With $A = (1, 2)$ and $B = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}$, find $\text{Cov}(AX^{(1)}, BX^{(2)})$.

Method I: $AX^{(1)} = X_1 + 2X_2$ and $BX^{(2)} = \begin{pmatrix} X_3 - 2X_4 \\ 2X_3 - X_4 \end{pmatrix} X$. So

$$\begin{aligned} & \text{Cov}(AX^{(1)}, BX^{(2)}) = \text{Cov}\left(X_1 + 2X_2, \begin{pmatrix} X_3 - 2X_4 \\ 2X_3 - X_4 \end{pmatrix}\right) \\ = \; & \begin{pmatrix} \text{cov}(X_1 + 2X_2, X_3 - 2X_4) & \text{cov}(X_1 + 2X_2, 2X_3 - X_4) \end{pmatrix} \\ = \; & \begin{pmatrix} 1\times 2 - 2\times 2 + 2\times 1 - 4\times 0 & 2\times 2 - 1\times 2 + 4\times 1 - 2\times 0 \end{pmatrix} \\ = \; & \begin{pmatrix} 0 & 6 \end{pmatrix} \end{aligned}$$

Method II: $AX^{(1)} = (A, 0)X$ and $BX^{(2)} = (0, B)X$. Then

$$\text{Cov}(AX^{(1)}, BX^{(2)}) = \text{Cov}((A, 0)X, (0, B)X) = (A, 0)\Sigma \begin{pmatrix} 0 \\ B' \end{pmatrix} = (0, 6)$$