L04: Population parameter

1. Population mean vector

   (1) Population mean vector

   The mean vector for population $\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ is $\mu = E(\mathbf{x}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$ where

   $$\mu_i = E(X_i) = \int_R x_i f_i(x_i) dx_i = \iint_{R^p} x_i f(x_1, .., x_p) dx_1, ...dx_p$$

   is the average value of X_i. $\mu = E(\mathbf{x})$ is the average value of $\mathbf{x}$. Since

   $$\mu \text{ exists} \Longleftrightarrow \mu_i \text{ exists for all } i \Longleftrightarrow f_i(x_i) \text{ exists for all } i \Longleftarrow f(x_1, .., x_p) \text{ exists}$$

   there are cases where $\mathbf{x}$ does not have joint pdf but $E(\mathbf{x})$ exists.

   (2) Expectation of random matrix
   For random matrix $\mathbf{Z} = (Z_{ij})_{a\times b}$, the expectation $E(\mathbf{Z}) = (E(Z_{ij}))_{a\times b}$. To have $E(\mathbf{Z})$ we have to assume that the distributions of Z_{ij} are known for all i and j. If $\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ has joint pdf $f(\mathbf{x})$ and $Z_{ij} = g_{ij}(\mathbf{x})$, then $E(Z_{ij}) = \iint_{R^p} g_{ij}(\mathbf{x})\, f(\mathbf{x}) dx_1, .., dx_p$.

   (3) A formula for expectation
   Let X, Y, U, V be random matrices. $E(AXB + CYD + \alpha) = A[E(X)]B + C[E(Y)]D + \alpha$.

   Ex1: $\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ has joint pdf $f(\mathbf{x})$. Then $E(\mathbf{x}\mathbf{x}') = E[(X_iX_j)_{p\times p}] = (E(X_iX_j))_{p\times p}$ where
   $E(X_iX_j) = \iint_{R^2} x_ix_j f_{ij}(x_i, x_j)\, dx_idx_j = \iint_{R^p} x_ix_j f(\mathbf{x}) dx_1, ..., dx_p$.
   $E(\mathbf{x}\mathbf{x}')$ exists if the marginal distributions for X_i and X_j are known for all (i, j).

2. Population variance-covariance matrix

   (1) Population variance-covariance matrix

   The variance-covariance matrix for population $\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ is

   $$\Sigma = \begin{pmatrix} \text{var}(X_1) & \cdots & \text{cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \text{cov}(X_p, X_1) & \cdots & \text{var}(X_p) \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_p^2 \end{pmatrix}.$$

   With $\sigma_{ij} = \text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_iX_j) - \mu_i\mu_j$,

   $$\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)'] = E(\mathbf{x}\mathbf{x}'] - \mu\mu'.$$

   We often write $\mathbf{x} \sim (\mu, \Sigma)$ to indicate $E(\mathbf{x}) = \mu$ and $\text{Cov}(X) = \Sigma$.

   (2) Covariance matrix
   The covariance matrix for random vector $\mathbf{x} \in R^p$ and random vector $\mathbf{y} \in R^q$ is

   $$\Sigma_{xy} = \begin{pmatrix} \text{cov}(x_1, y_1) & \cdots & \text{cov}(x_1, y_q) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_p, y_1) & \cdots & \text{cov}(x_p, y_q) \end{pmatrix} \in R^{p\times q}.$$

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

$$\text{Cov}(\mathbf{x},\, \mathbf{y}) = E[(\mathbf{x} - \mu_x)(\mathbf{y} - \mu_y)'] = E(\mathbf{x}\mathbf{y}') - \mu_x\mu_y'.$$

Clearly $\text{Cov}(\mathbf{x}) = \text{Cov}(\mathbf{x},\, \mathbf{x})$ and $\text{Cov}(\mathbf{y},\, \mathbf{x}) = [\text{Cov}(\mathbf{x},\, \mathbf{y})]'$.

(3) A formula for covariance matrix
Let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{u}$ and $\mathbf{v}$ be random vectors. Then

$$\text{Cov}(A\mathbf{x} + B\mathbf{y} + \alpha,\, C\mathbf{u} + D\mathbf{v} + \beta)$$

$$= A\,\text{cov}(\mathbf{x},\, \mathbf{u})C' + A\,\text{Cov}(\mathbf{x},\, \mathbf{v})D' + B\,\text{Cov}(\mathbf{y},\, \mathbf{u})C' + B\,\text{Cov}(\mathbf{y},\, \mathbf{v})\,D'$$

Ex2: $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim (\mu,\, \Sigma)$ where $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Let $\mathbf{z} = A\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ with $A = \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}$. Then $\mathbf{z} \sim \left(\begin{pmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{pmatrix},\, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{pmatrix}\right)$.

3. Population correlation matrix

(1) Population correlation matrix

Let Σ be the variance-covariance matrix for the population $\mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$.

Then $\mathbf{\Delta}^2 = \text{diag}(\Sigma) = \text{diag}(\sigma_1^2, ..., \sigma_p^2)$ is the population variance matrix and
$\mathbf{\Delta} = \text{diag}(\sigma_1, ..., \sigma_p)$ is the population standard deviation matrix.

$$\rho = \mathbf{\Delta}^{-1}\mathbf{\Sigma}\mathbf{\Delta}^{-1} = \begin{pmatrix} \frac{\sigma_1^2}{\sigma_1\,\sigma_1} & \cdots & \frac{\sigma_{1p}}{\sigma_1\,\sigma_p} \\ \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_p\,\sigma_1} & \cdots & \frac{\sigma p^2}{\sigma_p\,\sigma_p} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \cdots & 1 \end{pmatrix}$$

is the population correlation matrix.

(2) Properties
$-1 \le \rho_{ij} = \frac{\sigma_{ij}}{\sigma_i\,\sigma_j} \le 1$
$\rho_{ij} = 1 \Longleftrightarrow X_i = aX_j + b$ with $a > 0$ and $\rho_{ij} = -1 \Longleftrightarrow X_i = aX_j + b$ with $a < 0$.

4. Parameters for conditional distributions

(1) Conditional mean vector
Conditional pdf of $\mathbf{y} \in R^q$ given $\mathbf{x} \in R^p$ is $f(\mathbf{y}|\mathbf{x})$. Then the conditional mean of $\mathbf{y}$ given $\mathbf{x}$

$$E(\mathbf{y}|\mathbf{x}) = \begin{pmatrix} E(y_1|\mathbf{x}) \\ \vdots \\ E(y_q|\mathbf{x}) \end{pmatrix} \text{ is vector valued function of } \mathbf{x}$$

where $E(y_j|\mathbf{x}) = \int_R y_j f(\mathbf{y}|\mathbf{x})dy_1, .., dy_q$ is a function of $\mathbf{x}$.

(2) Conditional variance-covariance matrix
The conditional variance-covariance matrix of $\mathbf{y}$ given $\mathbf{x}$

$$\text{Cov}(\mathbf{y}|\mathbf{x}) = E\{[\mathbf{y} - E(\mathbf{y}|\mathbf{x})][\mathbf{y} - E(\mathbf{y}|\mathbf{x})]'|\mathbf{x}\} = E(\mathbf{y}\mathbf{y}'|\mathbf{x}) - [E(\mathbf{y}|\mathbf{x})][E(\mathbf{y}|\mathbf{x})]'$$

is a matrix-valued function of $\mathbf{x}$.

L05: Normal distributions

1. Parameters of conditional distributions

   (1) Expectation of conditional expectations
   $\mathbf{x} \in R^p$, $\mathbf{y} \in R^q$ and $\mathbf{Z} = (g_{ij}(\mathbf{y}))_{a\times b}$. Conditional expectation $E(\mathbf{Z}|\mathbf{x}) = (E(g_{ij}(\mathbf{y})|\mathbf{x})_{a\times b}$ is a matrix-valued function of $\mathbf{x}$. Then $E[E(\mathbf{Z}|\mathbf{x})] = E(\mathbf{Z})$.
   Proof We show $E[E(g_{ij}(\mathbf{y})|\mathbf{x})] = E(g_{ij}(\mathbf{y})) = E(\mathbf{Z}_{ij})$.

$$\begin{aligned} E\left[E\left(g_{ij}(\mathbf{y})|\mathbf{x}\right)\right] &= \textstyle\iint_{R^p} E\left(g_{ij}(\mathbf{y}) \mid \mathbf{x}\right)\, f_x(\mathbf{x})\, dx_1,..,dx_p \\ &= \textstyle\iint_{R^p} \left[\iint_{R^q} g_{ij}(\mathbf{y})\, f(\mathbf{y}|\mathbf{x}) dy_1,..,dy_q\right]\, f_X(\mathbf{x})\, dx_1,..,dx_p \\ &= \textstyle\iint_{R^{p+q}} g_{ij}(\mathbf{y})\, f(\mathbf{x},\, \mathbf{y}) dx_1,..,dx_p dy_1,..,dy_q = E(g_{ij}(\mathbf{y}) = E(\mathbf{Z}_{ij}). \end{aligned}$$

   (2) $\mathrm{Cov}(\mathbf{y}) = E[\mathrm{Cov}(\mathbf{y} \mid \mathbf{x})] + \mathrm{Cov}(E(\mathbf{y} \mid \mathbf{x}))$.
   Proof Note that

$$\begin{aligned} E[\mathrm{Cov}(\mathbf{y} \mid \mathbf{x})] &= E\left\{E\left(\mathbf{y}\mathbf{y}' \mid \mathbf{x}\right) - [E(\mathbf{y} \mid \mathbf{x})][E(\mathbf{y} \mid \mathbf{x})]'\right\} \\ &= E(\mathbf{y}\mathbf{y}') - E\{[E(\mathbf{y} \mid \mathbf{x})][E(\mathbf{y} \mid \mathbf{x})]'\}. \\ \mathrm{Cov}[E(\mathbf{y} \mid \mathbf{x})] &= E\{[E(\mathbf{y} \mid \mathbf{x})][E(\mathbf{y} \mid \mathbf{x})]'\} - \{E[E(\mathbf{y} \mid \mathbf{x})]\}\{E[E(\mathbf{y} \mid \mathbf{x})]\}' \\ &= E\{[E(\mathbf{y} \mid \mathbf{x})][E(\mathbf{y} \mid \mathbf{x})]'\} - [E(\mathbf{y})][E(\mathbf{y})]'. \end{aligned}$$

   Thus $\mathrm{Cov}[E(\mathbf{y} \mid \mathbf{x})] + E[\mathrm{Cov}(\mathbf{y} \mid \mathbf{x})] = E(\mathbf{y}\mathbf{y}') - [E(\mathbf{y})][E(\mathbf{y})]' = \mathrm{Cov}(\mathbf{y})$.
   Comment: $g(\mathbf{x}) = E(\mathbf{y} \mid \mathbf{x})$ is a function of $\mathbf{x}$. $E(g(\mathbf{x})) = E(y)$. $\mathrm{Cov}(g(\mathbf{x})) \leq \mathrm{Cov}(y)$.

   (3) Independence and parameters

$$\begin{aligned} \mathbf{x} \in R^p \text{ and } \mathbf{y} \in R^q \text{ are indep.} &\overset{def}{\Longleftrightarrow} f(\mathbf{x},\, \mathbf{y}) = f_x(\mathbf{x})\, f_y(\mathbf{y}) \\ &\overset{*}{\Longrightarrow} E(\mathbf{x}\mathbf{y}') = E(\mathbf{x})\, [E(\mathbf{y})]' \overset{**}{\Longleftrightarrow} \mathrm{Cov}(\mathbf{x},\, \mathbf{y}) = 0 \\ &\overset{def}{\Longleftrightarrow} \mathbf{x} \text{ and } \mathbf{y} \text{ are uncorrelated} \end{aligned}$$

   $\overset{*}{\Rightarrow}$**:** With $E(\mathbf{x}\mathbf{y}') = (E(X_iY_j))_{p\times q}$ and $E(\mathbf{x})[E(\mathbf{y})]' = (E(X_i)\, E(Y_j))_{p\times q}$
   we show $E(X_iY_j) = E(X_i)\, E(Y_j)$.

$$\begin{aligned} E(X_iY_j) &= \textstyle\iint_{R^{p+q}} x_i y_j f(\mathbf{x},\, \mathbf{y})\, dx_1,..,dx_p dy_1,..,dy_q \\ &= \textstyle\iint_{R^p} x_i f_x(\mathbf{x}) dx_1,..,dx_p \iint_{R^q} y_j f_y(\mathbf{y})\, dy_1,..,dy_q = E(X_i)\, E(Y_j) \end{aligned}$$

   $\overset{**}{\Leftrightarrow}$**:** Since $\mathrm{Cov}(\mathbf{x},\, \mathbf{y}) = E(\mathbf{x}\mathbf{y}') - E(\mathbf{x})[E(\mathbf{y})]'$,
   $\mathrm{Cov}(\mathbf{x},\, \mathbf{y}) = 0$ and $E(\mathbf{x}\mathbf{y}') = E(\mathbf{x})\, [E(\mathbf{y})]'$ are equivalent.

2. Multivariate normal distributions

   (1) Definition of normal distributions
   With $\mu \in R^p$ and positive definite $\Sigma \in R^{p\times p}$, let

$$f(x; \mu,\, \Sigma) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right].$$

   Then $f(\mathbf{x}; \mu,\, \Sigma) > 0$ and by the substitution of $z = \Sigma^{-1/2}(x-\mu) \Longleftrightarrow x = \Sigma^{1/2}z + \mu$ 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 $f(\mathbf{x}; \mu, \Sigma)$ is a pdf. The distribution of $\mathbf{x} \in R^p$ with pdf $f(\mathbf{x}; \mu, \Sigma)$ is called a normal distribution denoted by $\mathbf{x} \sim N(\mu, \Sigma)$.

(2) Parameters of $N(\mu, \Sigma)$
Calculation shows $\mathbf{x} \sim N(\mu, \Sigma) \Longrightarrow \mathbf{x} \sim (\mu, \Sigma)$.

(3) Transformation
If $\mathbf{x} \sim N(\mu, \Sigma)$ and $A \in R^{q\times p}$ has full row rank, then $\mathbf{y} = A\mathbf{x} + b \sim N(A\mu + b, A\Sigma A')$

Proof Skipped. **Comment:** Recall $\mathbf{x} \sim (\mu, \Sigma) \Longrightarrow A\mathbf{x} + b \sim (A\mu + b, A\Sigma A')$.

3. Marginal distributions and conditional distributions

(1) Marginal distributions
By the transformation one can easily see that the marginal distributions of a normal distribution are normal distributions.

Ex1: $\mathbf{x} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}\right)$.

Then $X_2 = (0, 1, 0)\mathbf{x} \sim N(\mu_2, \sigma_2^2)$ and
$\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{31} & \sigma_3^2 \end{pmatrix}\right)$.

(2) Independence
Suppose $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix}\right)$. Then

$$\mathbf{x} \text{ and } \mathbf{y} \text{ are indep.} \Longleftrightarrow \Sigma_{xy} = 0 \text{ and } \Sigma_{yx} = 0.$$

Proof Only give a sketch.
⇒: $\mathbf{x}$ and $\mathbf{y}$ are indep$\Longrightarrow$ $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated.
⇐: $f(\mathbf{x}, \mathbf{y}) = f_x(\mathbf{x})\, f_y(\mathbf{y})$.

(3) Conditional distributions
Suppose $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$.
Let $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.
Then $\mathbf{x} \mid \mathbf{y} \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y} - \mu_2), \Sigma_{11.2})$ and
$\mathbf{y} \mid \mathbf{x} \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x} - \mu_1), \Sigma_{22.1})$.

Proof Only show the second one.

$$\begin{aligned} \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} &= \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{x} \end{pmatrix} \\ &\sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{pmatrix}\right). \end{aligned}$$

So $\mathbf{z}_1 = \mathbf{x} \sim N(\mu_1, \Sigma_{11})$ and $\mathbf{z}_2 = \mathbf{y} - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{x} \sim N(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22.1})$ are independent.
Thus $\mathbf{z}_2 \mid \mathbf{z}_1 \sim \mathbf{z}_2 \sim N(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1, \Sigma_{22.1})$.
Therefore $\mathbf{z}_2 + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{z}_1 \mid \mathbf{z}_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{z}_1 - \mu_1), \Sigma_{22.1})$, i.e.,

$$\mathbf{y} \mid \mathbf{x} \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x} - \mu_1), \Sigma_{22.1}).$$