STAT 3690 Lecture Note

Week Three (Jan 23, 25, & 27, 2023)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2023/Jan/24 12:58:34

Statistical modelling (con'd)

Transformation of random vectors

- Derive the pdf of continuous $\boldsymbol{Y} = \boldsymbol{g}(\boldsymbol{X})$ from the pdf of continuous $\boldsymbol{X}$
- Prerequisite
  - $\boldsymbol{X} = [X_1, \ldots, X_p]^\top$ and $\boldsymbol{Y} = [Y_1, \ldots, Y_p]^\top$
  - $\boldsymbol{g} = (g_1, \ldots, g_p)$: $\mathbb{R}^p \to \mathbb{R}^p$ is a continuous one-to-one map with inverse $\boldsymbol{g}^{-1} = (h_1, \ldots, h_p)$, i.e., $Y_i = g_i(\boldsymbol{X})$ and $X_i = h_i(\boldsymbol{Y})$
- Elaborate $\operatorname{supp}(\boldsymbol{Y}) = \{[y_1, \ldots, y_p]^\top : [h_1(y_1, \ldots, y_p), \ldots, h_p(y_1, \ldots, y_p)]^\top \in \operatorname{supp}(\boldsymbol{X})\}$
- Jacobian matrix of $\boldsymbol{g}^{-1}$ is $\mathbf{J}_{\boldsymbol{g}^{-1}} = [\partial x_i/\partial y_j]_{p\times p} = [\partial h_i(y_1, \ldots, y_p)/\partial y_j]_{p\times p}$
  - Also, $|\det(\mathbf{J}_{\boldsymbol{g}^{-1}})| = |\det([\partial y_i/\partial x_j]_{p\times p})|^{-1} = |\det([\partial g_i(x_1, \ldots, x_p)/\partial x_j]_{p\times p})|^{-1}$
- Then

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_p) = f_{\boldsymbol{X}}(h_1(y_1, \ldots, y_p), \ldots, h_p(y_1, \ldots, y_p))|\det(\mathbf{J}_{\boldsymbol{g}^{-1}})|\mathbf{1}_{\operatorname{supp}(\boldsymbol{Y})}(y_1, \ldots, y_p)$$

- Exercise: Let $\boldsymbol{X} = [X_1, X_2]^\top$ follow the standard bivariate normal, i.e., its pdf is

$$f_{\boldsymbol{X}}(x_1, x_2) = (2\pi)^{-1}\exp\{-(x_1^2 + x_2^2)/2\}\mathbf{1}_{\mathbb{R}^2}(x_1, x_2).$$

  Find out the joint pdf of $\boldsymbol{Y} = [Y_1, Y_2]^\top$, where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $0 \le Y_2 < 2\pi$ is the angle from the positive x-axis to the ray from the origin to the point (X_1, X_2), that is, Y is X in the polar coordinate.

- Exercise: Given positive α, β and θ, $\boldsymbol{X} = [X_1, X_2]^\top$ follow

$$f_{\boldsymbol{X}}(x_1, x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} x_1^{\alpha-1} x_2^{\beta-1} \exp\left(-\frac{x_1 + x_2}{\theta}\right)\mathbf{1}_{\mathbb{R}^+\times\mathbb{R}^+}(x_1, x_2),$$

  where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\mathrm{d}t$, e.g., $\Gamma(n) = (n-1)!$ for positive integer n. Find out the joint pdf of $\boldsymbol{Y} = [Y_1, Y_2]^\top$, where $Y_1 = X_1/(X_1 + X_2)$ and $Y_2 = X_1 + X_2$.

Mean matrix

- $\mathrm{E}(\boldsymbol{X}) = [\mathrm{E}(X_{ij})]_{n\times p}$, where
  - Random $n \times p$ matrix $\boldsymbol{X} = [X_{ij}]_{n\times p}$
- (Linearity) $\mathrm{E}(\mathbf{A}\boldsymbol{X} + \mathbf{B}\boldsymbol{Y}) = \mathbf{A}\mathrm{E}(\boldsymbol{X}) + \mathbf{B}\mathrm{E}(\boldsymbol{Y})$, where
  - Fixed $\mathbf{A} \in \mathbb{R}^{\ell\times n}$ and $\mathbf{B} \in \mathbb{R}^{\ell\times m}$
  - Random matrices $\boldsymbol{X} = [X_{ij}]_{n\times p}$ and $\boldsymbol{Y} = [Y_{ij}]_{m\times p}$

Covariance matrix

- Random p-vector $\boldsymbol{X} = [X_1, \dots, X_p]^\top$ and random q-vector $\boldsymbol{Y} = [Y_1, \dots, Y_q]^\top$
- Covariance matrix (defined via expectation) $\boldsymbol{\Sigma}_{\boldsymbol{XY}} = \mathrm{cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}[\{\boldsymbol{X} - \mathrm{E}(\boldsymbol{X})\}\{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\}^\top]$
  - Also, $\boldsymbol{\Sigma}_{\boldsymbol{XY}} = \mathrm{E}(\boldsymbol{X}\boldsymbol{Y}^\top) - \mathrm{E}(\boldsymbol{X})\mathrm{E}(\boldsymbol{Y}^\top)$
  - The (i, j)-entry of $\boldsymbol{\Sigma}_{\boldsymbol{XY}}$ is $\mathrm{cov}(X_i, Y_j)$
- $\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}+\boldsymbol{a}, \mathbf{B}\boldsymbol{Y}+\boldsymbol{b}} = \mathbf{A}\Sigma_{\boldsymbol{XY}}\mathbf{B}^\top$ for fixed $\mathbf{A} \in \mathbb{R}^{m\times p}$, $\boldsymbol{a} \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{\ell\times q}$ and $\boldsymbol{b} \in \mathbb{R}^\ell$
- $\boldsymbol{\Sigma}_{\boldsymbol{X}} \geq 0$, where $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \mathrm{cov}(\boldsymbol{X})$ is short for $\boldsymbol{\Sigma}_{\boldsymbol{XX}} = \mathrm{cov}(\boldsymbol{X}, \boldsymbol{X})$

- Exercise: Verify the following properties of covariance matrix
  1. $\boldsymbol{\Sigma}_{\mathbf{A}\boldsymbol{X}+\boldsymbol{a}, \mathbf{B}\boldsymbol{Y}+\boldsymbol{b}} = \mathbf{A}\Sigma_{\boldsymbol{XY}}\mathbf{B}^\top$
  2. $\boldsymbol{\Sigma}_{\boldsymbol{X}} \geq 0$

Sample covariance matrix

- Samples $\boldsymbol{X}_k = [X_{k1}, \dots, X_{kp}]^\top$ and $\boldsymbol{Y}_k = [Y_{k1}, \dots, Y_{kq}]^\top$, $k = 1, \dots, n$
- $(\boldsymbol{X}_k, \boldsymbol{Y}_k) \overset{\text{iid}}{\sim} (\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{X} = [X_1, \dots, X_p]^\top$ and $\boldsymbol{Y} = [Y_1, \dots, Y_q]^\top$
- Sample mean vectors
  - $\bar{\boldsymbol{X}} = n^{-1}\sum_{k=1}^n \boldsymbol{X}_k = [\bar{X}_{\cdot 1}, \cdots, \bar{X}_{\cdot p}]^\top$
  - $\bar{\boldsymbol{Y}} = n^{-1}\sum_{k=1}^n \boldsymbol{Y}_k = [\bar{Y}_{\cdot 1}, \cdots, \bar{Y}_{\cdot q}]^\top$
- Sample covariance matrix:

$$\mathbf{S}_{\boldsymbol{XY}} = \frac{1}{n-1}\sum_{k=1}^n \{(\boldsymbol{X}_k - \bar{\boldsymbol{X}})(\boldsymbol{Y}_k - \bar{\boldsymbol{Y}})^\top\}$$

  - The (i, j)-entry of $\mathbf{S}_{\boldsymbol{XY}}$ is $(n-1)^{-1}\sum_{k=1}^n (X_{ki} - \bar{X}_{\cdot i})(Y_{kj} - \bar{Y}_{\cdot j})$, i.e., the sample covariance between X_i and Y_j
  - Unbiasedness: $\mathrm{E}(\mathbf{S}_{\boldsymbol{XY}}) = \boldsymbol{\Sigma}_{\boldsymbol{XY}}$
  - $\mathbf{S}_{\mathbf{A}\boldsymbol{X}+\boldsymbol{a}, \mathbf{B}\boldsymbol{Y}+\boldsymbol{b}} = \mathbf{A}\mathbf{S}_{\boldsymbol{XY}}\mathbf{B}^\top$ for $\mathbf{A} \in \mathbb{R}^{m\times p}$, $\boldsymbol{a} \in \mathbb{R}^m$, $\mathbf{B} \in \mathbb{R}^{\ell\times q}$ and $\boldsymbol{b} \in \mathbb{R}^\ell$
  - $\mathbf{S}_{\boldsymbol{X}} \geq 0$
  - Implementation in *R*: cov() (or var() if $\boldsymbol{X} = \boldsymbol{Y}$)

- Exercise: Verify the following properties of sample covariance matrix
  1. $\mathrm{E}(\mathbf{S}_{\boldsymbol{XY}}) = \boldsymbol{\Sigma}_{\boldsymbol{XY}}$
  2. $\mathbf{S}_{\mathbf{A}\boldsymbol{X}+\boldsymbol{a}, \mathbf{B}\boldsymbol{Y}+\boldsymbol{b}} = \mathbf{A}\mathbf{S}_{\boldsymbol{XY}}\mathbf{B}^\top$
  3. $\mathbf{S}_{\boldsymbol{X}} \geq 0$

Computing sample mean vectors and sample covariance matrices via *R*

Multivariate normal (MVN) distribution (J&W Sec 4.2)

Definition

- Standard MVN
  - $\boldsymbol{Z} = [Z_1, \ldots, Z_p]^\top \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I}) \Leftrightarrow Z_1, \ldots, Z_p \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$
  - pdf
$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = (2\pi)^{-p/2} \exp(-\boldsymbol{z}^\top \boldsymbol{z}/2) \cdot \mathbf{1}_{\mathbb{R}^p}(\boldsymbol{z})$$
- General MVN
  - $\boldsymbol{X} = [X_1, \ldots, X_p]^\top \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Leftrightarrow$ there exists $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\boldsymbol{Z} \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I})$ such that $\boldsymbol{X} = \mathbf{A}\boldsymbol{Z} + \boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^\top$
    - Limited to non-degenerate cases, i.e., invertible $\mathbf{A}$ ($\Leftrightarrow \boldsymbol{\Sigma} > 0$)
  - pdf
$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-p/2} (\det\boldsymbol{\Sigma})^{-1/2} \exp\{-(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})/2\} \cdot \mathbf{1}_{\mathbb{R}^p}(\boldsymbol{x})$$

- Exercise: Density of $\mathrm{MVN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ evaluated at $(4, 7)$, where
$$\boldsymbol{\mu} = [3, 6]^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 10 & 2 \\ 2 & 5 \end{bmatrix}.$$

Properties of MVN

- $\boldsymbol{X}$ is of MVN $\Leftrightarrow a^\top \boldsymbol{X}$ is normally distributed for ALL non-zero $a \in \mathbb{R}^p$.
  - Warning: marginal normals do not imply the joint normal.
- If $\boldsymbol{X} \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{A}\boldsymbol{X} + \boldsymbol{b} \sim \mathrm{MVN}_q(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ for $\mathbf{A} \in \mathbb{R}^{q \times p}$ of full-row-rank. Hence,
  - (Stochastic representation of MVN) if $\boldsymbol{X} \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then there is $\boldsymbol{Z} \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I})$ such that $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$. Actually, $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$.
- $(\boldsymbol{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(p)$ if $\boldsymbol{X} \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

- Exercise: Generate six iid samples following bivariate normal $\mathrm{MVN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with
$$\boldsymbol{\mu} = [3, 6]^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 10 & 2 \\ 2 & 5 \end{bmatrix}.$$

- Exercise: Suppose $X_1 \sim \mathcal{N}(0, 1)$. In the following two cases, verify that $X_2 \sim \mathcal{N}(0, 1)$ as well. Does $\boldsymbol{X} = [X_1, X_2]^\top$ follow an MVN in both cases?
  a. $X_2 = -X_1$;
  b. $X_2 = (2Y - 1)X_1$, where $Y \sim \mathrm{Ber}(p)$ and $Y \perp\!\!\!\perp \boldsymbol{X}$. (Hint: $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X} \Leftrightarrow f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{\boldsymbol{X}}(\boldsymbol{x}) f_{\boldsymbol{Y}}(\boldsymbol{y})$, where $\boldsymbol{Z} = [\boldsymbol{X}^\top, \boldsymbol{Y}^\top]^\top$.)

Marginal and conditional MVN

- If $\boldsymbol{X} \sim \mathrm{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where
$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

with $\mathbf{\Sigma}_{11} > 0$ and $\mathbf{\Sigma}_{22} > 0$, then
  - (Marginals of MVN are still MVN) $\boldsymbol{X}_i \sim \text{MVN}_{p_i}(\boldsymbol{\mu}_i, \mathbf{\Sigma}_{ii})$
  - (Conditionals of MVN are MVN) $\boldsymbol{X}_i \mid \boldsymbol{X}_j = \boldsymbol{x}_j \sim \text{MVN}_{p_i}(\boldsymbol{\mu}_{i|j}, \mathbf{\Sigma}_{i|j})$
    * $\boldsymbol{\mu}_{i|j} = \boldsymbol{\mu}_i + \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}_j)$
    * $\mathbf{\Sigma}_{i|j} = \mathbf{\Sigma}_{ii} - \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{ji}$
  - $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2 \Leftrightarrow \mathbf{\Sigma}_{12} = \mathbf{0}$
    * Warning: the prerequisite for this equivalence is the joint normal of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

- Exercise: The argument $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2 \Leftrightarrow \mathbf{\Sigma}_{12} = 0$ is based on $\boldsymbol{X} = [\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top]^\top \sim \text{MVN}$. That is, if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are both MVN BUT they are not jointly normal, the zero $\mathbf{\Sigma}_{12}$ doesn't suffice for the independence between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Recall the instance in the previous exercise: $X_1 \sim \mathcal{N}(0,1)$ and $X_2 = (2Y-1)X_1$. Verify that X_1 and X_2 are not independent of each other.

Checking normality (J&W Sec 4.6)

- Checking the univariate marginal distributions
  - Normal Q-Q plot
    * qqnorm(); car::qqPlot()
  - Normality test
    * shapiro.test()
- Checking the quadratic form
  - χ^2 Q-Q plot
    * $D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top \mathbf{S}^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}) \approx \chi^2(p)$ if $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \mathbf{\Sigma})$
    * qqplot(); car::qqPlot()

Detecting outliers (J&W Sec 4.7)

- Scatter plot of standardized values
- Check the points farthest from the origin in χ^2 Q-Q plot

Improving normality (J&W Sec 4.8)

- Box-Cox transformation: for $x > 0$,

$$x^*(\lambda) = \begin{cases} (x^\lambda - 1)/\lambda & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases}$$

  - If $x \leq 0$, change it to be positive first.
- Exploratory data analysis (EDA)
  - J. Tukey (1977). Exploratory Data Analysis. Addison-Wesley. ISBN 978-0-201-07616-5.

R package "MVN"