STAT 3690 Lecture Note

Week Four (Jan 30, Feb 1, & 3, 2023)

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

2023/Jan/27 11:52:30

Multivariate normal (MVN) distribution (con'd, J&W Sec 4.2)

Definition

- Standard MVN
  - $\boldsymbol{Z} = [Z_1, \ldots, Z_p]^\top \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I}) \Leftrightarrow Z_1, \ldots, Z_p \overset{\text{iid}}{\sim} N(0, 1)$
  - 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 $X_1 \sim N(0,1)$. In the following two cases, verify that $X_2 \sim N(0,1)$ 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 $Y \sim \text{Ber}(p)$ 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 $\boldsymbol{X} \sim \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, 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
  - random p_i-vector $\boldsymbol{X}_i$, $i = 1, 2$,
  - p_i-vector $\boldsymbol{\mu}_i$, $i = 1, 2$,
  - $p_i \times p_i$ matrix $\boldsymbol{\Sigma}_{ii} > 0$, $i = 1, 2$,
- then
  - (Marginals of MVN are still MVN) $\boldsymbol{X}_i \sim \text{MVN}_{p_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$
  - (Conditionals of MVN are MVN) $\boldsymbol{X}_i \mid \boldsymbol{X}_j = \boldsymbol{x}_j \sim \text{MVN}_{p_i}(\boldsymbol{\mu}_{i|j}, \boldsymbol{\Sigma}_{i|j})$
    - $\boldsymbol{\mu}_{i|j} = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}_j)$
    - $\boldsymbol{\Sigma}_{i|j} = \boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{ji}$
  - $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2 \Leftrightarrow \boldsymbol{\Sigma}_{12} = \boldsymbol{0}$
    - Warning: the prerequisite for this equivalence is the joint normal of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$.

- Exercise: The argument $\boldsymbol{X}_1 \perp\!\!\!\perp \boldsymbol{X}_2 \Leftrightarrow \boldsymbol{\Sigma}_{12} = 0$ is based on $[\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top]^\top \sim \text{MVN}$. That is, if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are both MVN BUT they are not jointly normal, the zero $\boldsymbol{\Sigma}_{12}$ 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()
  - Univariate normality test
    - shapiro.test(); nortest::ad.test()
- Testing the multivariate normality
  - MVN::mvn()
- 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 $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
    - 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)

- (Original) Box-Cox (power) transformation: transform positive x into

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

  with λ selected with certain criterion
  - If $x \leq 0$, change it to be positive first.
  - See J. Tukey (1977). *Exploratory Data Analysis*. Boston: Addison-Wesley.
- Multivariate Box-Cox transformation

Maximum likelihood (ML) estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (J&W Sec 4.3)

- Sample: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
- Parameter space: $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}, \boldsymbol{\Sigma} > 0\}$
- Likelihood function

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) =& \prod_{i=1}^{n} \left[\frac{1}{\sqrt{(2\pi)^p \det(\boldsymbol{\Sigma})}} \exp\Big\{ -\frac{1}{2} (\boldsymbol{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) \Big\} \right] \\ =& \frac{1}{\sqrt{(2\pi)^{np} \{\det(\boldsymbol{\Sigma})\}^n}} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu}) \Big\} \end{aligned}$$

- Log likelihood

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \ln\{L(\boldsymbol{\mu}, \boldsymbol{\Sigma})\} = -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln\{\det(\boldsymbol{\Sigma})\} - \frac{1}{2} \sum_{i=1}^{n} (\boldsymbol{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{\mu})$$

- ML estimator

$$(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}}) = \arg \max_{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\bar{\boldsymbol{X}}, \frac{n-1}{n} \mathbf{S})$$

  - Consistency: $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$ approaches $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (in certain sense) as $n \to \infty$
  - Efficiency: the covariance matrix of $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$ is approximately optimal (in certain sense) as $n \to \infty$
  - Invariance: For any function g, the ML estimator of $g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $g(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$.

Sampling distributions of $\bar{\boldsymbol{X}}$ and S (J&W Sec 4.4)

- Recall the univariate case
  - $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$
  - $S^2 \perp\!\!\!\perp \bar{X}$
    - Sample variance $S^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
  - $\sqrt{n}(\bar{X} - \mu)/\sigma \sim \mathcal{N}(0, 1)$
  - $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
  - $\sqrt{n}(\bar{X} - \mu)/S \sim t(n-1)$
- The multivariate case
  - $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} rmMVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
  - $\mathbf{S} \perp\!\!\!\perp \bar{\boldsymbol{X}}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \perp\!\!\!\perp \hat{\boldsymbol{\mu}}_{\text{ML}}$
  - $\sqrt{n} \boldsymbol{\Sigma}^{-1/2} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$

- $(n-1)\mathbf{S} = n\widehat{\mathbf{\Sigma}}_{\text{ML}} \sim W_p(\mathbf{\Sigma}, n-1)$
- $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim$ Hotelling's $T^2(p, n-1)$

- Wishart distribution
  - Def: $W_p(\mathbf{\Sigma}, n)$ is the distribution of $\sum_{i=1}^n \boldsymbol{Y}_i \boldsymbol{Y}_i^\top$ with $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\mathbf{0}, \mathbf{\Sigma})$
    - A generalization of χ^2-distribution: $W_p(\mathbf{\Sigma}, n) = \chi^2(n)$ if $p = \mathbf{\Sigma} = 1$
  - Propoties
    - $\mathbf{A}\mathbf{A}^\top > 0$ and $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n) \Rightarrow \mathbf{A}\mathbf{W}\mathbf{A}^\top \sim W_p(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top, n)$
    - $\mathbf{W}_i \overset{\text{iid}}{\sim} W_p(\mathbf{\Sigma}, n_i) \Rightarrow \mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, n_1 + n_2)$
    - $\mathbf{W}_1 \perp\!\!\!\perp \mathbf{W}_2$, $\mathbf{W}_1 + \mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, n)$ and $\mathbf{W}_1 \sim W_p(\mathbf{\Sigma}, n_1) \Rightarrow \mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, n - n_1)$
    - $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n)$ and $\boldsymbol{a} \in \mathbb{R}^p \Rightarrow$
$$\frac{\boldsymbol{a}^\top \mathbf{W} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{\Sigma} \boldsymbol{a}} \sim \chi^2(n)$$
    - $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n)$, $\boldsymbol{a} \in \mathbb{R}^p$ and $n \geq p \Rightarrow$
$$\frac{\boldsymbol{a}^\top \mathbf{\Sigma}^{-1} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{W}^{-1} \boldsymbol{a}} \sim \chi^2(n - p + 1)$$
    - $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n) \Rightarrow$
$$\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{W}) \sim \chi^2(np)$$

- Hotelling's T^2 distribution
  - A generalization of (Student's) t-distribution
  - If $\boldsymbol{X} \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W} \sim W_p(\mathbf{I}, n)$, then
$$\boldsymbol{X}^\top \mathbf{W}^{-1} \boldsymbol{X} \sim T^2(p, n)$$
  - $Y \sim T^2(p, n) \Leftrightarrow \frac{n-p+1}{np} Y \sim F(p, n-p+1)$

- Wilk's lambda distribution
  - Wilks's lambda is to Hotelling's T^2 as F distribution is to Student's t in univariate statistics.
  - Given independent $\mathbf{W}_1 \sim W_p(\mathbf{\Sigma}, n_1)$ and $\mathbf{W}_2 \sim W_p(\mathbf{\Sigma}, n_2)$ with $n_1 \geq p$,
$$\Lambda = \frac{\det(\mathbf{W}_1)}{\det(\mathbf{W}_1 + \mathbf{W}_2)} = \frac{1}{\det(\mathbf{I} + \mathbf{W}_1^{-1}\mathbf{W}_2)} \sim \Lambda(p, n_1, n_2)$$
    - Resort to an approximation in computation: $\{(p - n_2 + 1)/2 - n_1\} \ln \Lambda(p, n_1, n_2) \approx \chi^2(n_2 p)$

Inference on $\boldsymbol{\mu}$

Hypothesis testing

- Model: $\boldsymbol{X} \sim f_{\boldsymbol{\theta}^*} \in \{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbf{\Theta}\}$
  - $\boldsymbol{\theta}^*$: parameters of interest, fixed and unknown
  - $\mathbf{\Theta}$: the parameter space
- Hypotheses $H_0 : \boldsymbol{\theta}^* \in \mathbf{\Theta}_0$ v.s. $H_1 : \boldsymbol{\theta}^* \in \mathbf{\Theta}_1$
  - $\mathbf{\Theta}_0 \cap \mathbf{\Theta}_1 = \emptyset$
  - $\mathbf{\Theta}_0 \cup \mathbf{\Theta}_1 = \mathbf{\Theta}$
- Rejection/critical region R
  - Reject H_0 if $\boldsymbol{X} \in R$
- Level α: $\sup_{\boldsymbol{\theta} \in \mathbf{\Theta}_0} \beta(\boldsymbol{\theta}) \leq \alpha$

- Power function: $\beta(\boldsymbol{\theta}) = \Pr_{\boldsymbol{\theta}}(\boldsymbol{X} \in R)$
- When $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_0$, Pr(type I error) $= \beta(\boldsymbol{\theta}^*) \le \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \beta(\boldsymbol{\theta}) \le \alpha$
  - Type I error: H_0 is incorrectly rejected
- When $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}_1$, Pr(type II error) $= 1 - \beta(\boldsymbol{\theta}^*)$
  - Type II error: H_0 is incorrectly accepted

- p-value: alternative to rejection region
  - Impossible to be well-defined in some cases
  - $p = p(\boldsymbol{x})$ is defined such that $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}\{p(\boldsymbol{x}) \in [0, \alpha)\} \le \alpha$ for all $\alpha \in [0, 1]$
    - $R = \{\boldsymbol{x} : p(\boldsymbol{x}) \in [0, \alpha)\}$
- Necessary components in reporting a testing result
  1. Hypotheses
  2. Name of approach
  3. Value of test statistic
  4. Rejection region/p-value
  5. Conclusion: e.g., at the α level, we reject/do not reject H_0, i.e., we believe...

Likelihood ratio test (LRT)

- Minimize the type II error rate subject to a capped type I error rate (under certain classical circumstances)
- Test statistic
$$\lambda(\boldsymbol{X}) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}; \boldsymbol{X})}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \boldsymbol{X})} = \frac{L(\hat{\boldsymbol{\theta}}_0; \boldsymbol{X})}{L(\hat{\boldsymbol{\theta}}; \boldsymbol{X})}$$
  - $\hat{\boldsymbol{\theta}}_0$: ML estimator for $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$
  - $\hat{\boldsymbol{\theta}}$: ML estimator for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- Rejection region $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c\}$
  - $\boldsymbol{x}$ is the realization of $\boldsymbol{X}$
  - $c \in \mathbb{R}$ is chosen such that
$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}(\lambda(\boldsymbol{X}) \le c) = \alpha.$$
    - Have to know the null distribution of $\lambda(\boldsymbol{X})$, i.e., the distribution of $\lambda(\boldsymbol{X})$ with $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$

- p-value
$$p(\boldsymbol{x}) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} \Pr_{\boldsymbol{\theta}}\{\lambda(\boldsymbol{X}) \le \lambda(\boldsymbol{x})\}$$
- Null distribution of $\lambda(\boldsymbol{X})$
  - Use the accurate distribution of $\lambda(\boldsymbol{X})$ if it is known; otherwise see below for an approximation.
  - As $n \to \infty$,
$$-2 \ln \lambda(\boldsymbol{X}) \sim \chi^2(\nu)$$
    - ν: the difference in numbers of free parameters between H_0 and H_1
    - Leading to an (asymptotic) rejection region $\{\boldsymbol{x} : -2 \ln \lambda(\boldsymbol{x}) \ge \chi^2_{\nu, 1-\alpha}\}$
      - $\chi^2_{\nu, 1-\alpha}$ is the $(1 - \alpha)$- quantile of $\chi^2(\nu)$.