

STAT 3690 Lecture Note

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

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

2023/Jan/30 22:12:42

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

Definition

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

-
- Exercise: Density of $\text{MVN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ evaluated at $(4, 7)$, where

$$\boldsymbol{\mu} = [3, 6]^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 10 & 2 \\ 2 & 5 \end{bmatrix}.$$

Properties of MVN

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

-
- Exercise: Generate six iid samples following bivariate normal $\text{MVN}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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 $\mathbf{X} = [X_1, X_2]^\top$ 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 $Y \perp\!\!\!\perp X_1$.
-

Marginal and conditional MVN

- If $\mathbf{X} \sim \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

with

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

- Exercise: The argument $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2 \Leftrightarrow \boldsymbol{\Sigma}_{12} = \mathbf{0}$ is based on $[\mathbf{X}_1^\top, \mathbf{X}_2^\top]^\top \sim \text{MVN}$. That is, if \mathbf{X}_1 and \mathbf{X}_2 are both MVN BUT they are not jointly normal, the zero $\boldsymbol{\Sigma}_{12}$ doesn't suffice for the independence between \mathbf{X}_1 and \mathbf{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
 - * $D_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})^\top \mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}) \approx \chi^2(p)$ if $\mathbf{X}_i \stackrel{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
 - * `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 μ and Σ (J&W Sec 4.3)

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

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

- Log likelihood

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

- ML estimator

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

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

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

- Recall the univariate case
 - $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$
 - $S^2 \perp\!\!\!\perp \bar{X}$
 - * 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
 - $\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{iid}}{\sim} \text{MVN}_p(\mu, \Sigma)$, $n > p$
 - $\mathbf{S} \perp\!\!\!\perp \bar{\mathbf{X}}$, i.e., $\hat{\Sigma}_{\text{ML}} \perp\!\!\!\perp \hat{\mu}_{\text{ML}}$
 - $\sqrt{n}\Sigma^{-1/2}(\bar{\mathbf{X}} - \mu) \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$

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

- Wishart distribution

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

$$\frac{\mathbf{a}^\top \mathbf{W} \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a}} \sim \chi^2(n)$$

- * $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n), \mathbf{a} \in \mathbb{R}^p$ and $n \geq p \Rightarrow$

$$\frac{\mathbf{a}^\top \boldsymbol{\Sigma}^{-1} \mathbf{a}}{\mathbf{a}^\top \mathbf{W}^{-1} \mathbf{a}} \sim \chi^2(n - p + 1)$$

- * $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow$

$$\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W}) \sim \chi^2(np)$$

- Hotelling's T^2 distribution

- A generalization of (Student's) t -distribution
- If $\mathbf{X} \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W} \sim W_p(\mathbf{I}, n)$, then

$$\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X} \sim T^2(p, n)$$

- $Y \sim T^2(p, n) \Leftrightarrow \frac{n-p+1}{np} Y \sim F(p, n-p+1)$
-

- 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 $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1)$ and $\mathbf{W}_2 \sim W_p(\boldsymbol{\Sigma}, n_2)$ 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

- Is it a squirrel?



Figure 1: Squirrel (Photograph by the Lacoste Garden Centre)



Figure 2: Flying Squirrel (Photograph by Joel Sartore)



Figure 3: Flying Squirrel (Photograph by Alex Badyaev)

-
- Null and alternative hypotheses, say H_0 and H_1 , resp.
 - Name of approach
 - Test statistic (not unique) and corresponding level α rejection region R_α
 - $\Pr(\text{test statistic} \in R_\alpha \mid H_0) \leq \alpha$
 - Reject H_0 if the value of test statistic $\in R_\alpha$
 - * Type I error: H_0 is incorrectly rejected; i.e., H_0 is correct but rejected

- * Type II error: H_0 is incorrectly accepted i.e., H_0 is wrong but NOT rejected
- p -value: a special test statistic with a default level α rejection region $[0, \alpha]$
- Necessary components in reporting a testing result
 1. Hypotheses
 2. Name of approach
 3. Level α
 4. (Value of test statistic AND rejection region) OR p -value
 5. Conclusion: e.g., at the α level, we reject/do not reject H_0 , i.e., we believe that...

Likelihood ratio test (LRT)

- Minimize the type II error rate subject to a capped type I error rate (under certain classical circumstances)
- Test statistic

$$\lambda(\mathbf{x}) = \frac{L(\hat{\boldsymbol{\theta}}_0; \mathbf{x})}{L(\hat{\boldsymbol{\theta}}; \mathbf{x})}$$

- \mathbf{x} : all the observations
- L : the likelihood function
- $\boldsymbol{\theta}$: the unknown parameter(s)
- $\hat{\boldsymbol{\theta}}_0$: ML estimator for $\boldsymbol{\theta}$ under H_0
- $\hat{\boldsymbol{\theta}}$: ML estimator for $\boldsymbol{\theta}$
- (Asymptotic) rejection region

$$R_\alpha = \{\mathbf{x} : -2 \ln \lambda(\mathbf{x}) \geq \chi_{\nu, 1-\alpha}^2\}$$
 - I.e., reject H_0 when $-2 \ln \lambda(\mathbf{x}) \geq \chi_{\nu, 1-\alpha}^2$
 - $\chi_{\nu, 1-\alpha}^2$ is the $(1 - \alpha)$ -quantile of $\chi^2(\nu)$
 - ν : the difference in numbers of free parameters between H_0 and H_1
- (Asymptotic) p -value

$$p(\mathbf{x}) = 1 - F_{\chi^2(\nu)}\{-2 \ln \lambda(\mathbf{x})\}$$
 - $F_{\chi^2(\nu)}(\cdot)$ is the cdf of $\chi^2(\nu)$