STAT 3690 Lecture Note

Week Five (Feb 6, 8, & 10, 2023)

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

2023/Feb/03 12:49:04

Multivariate normal (MVN) distribution (con'd)

Checking/testing the normality (con'd, J&W Sec 4.6)

- Checkcing the univariate marginal distributions
  - Normal Q-Q plot
    - qqnorm(); car::qqPlot()
  - Univariate normality test
    - shapiro.test(); nortest::ad.test(); MVN::mvn()
- Checkcing the multivariate normality
  - 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()
  - Multivariate normality test
    - MVN::mvn()

```
options(digits = 4)
library(datasets)
data(iris)
head(iris)
(iris_setosa = iris[iris$Species=='setosa', 1:3])
p = ncol(iris_setosa)
n = nrow(iris_setosa)

# Marginal normal Q-Q plot
car::qqPlot(rnorm(n), id = F)
car::qqPlot(iris_setosa[,1], id = F)
car::qqPlot(iris_setosa[,2], id = F)
car::qqPlot(iris_setosa[,3], id = F)

# Univariate normality test
## Shapiro-Wilk Normality Test
shapiro.test(rnorm(n))
shapiro.test(iris_setosa[,1])
shapiro.test(iris_setosa[,2])
shapiro.test(iris_setosa[,3])
## Anderson-Darling test for normality
```

```r
nortest::ad.test(iris_setosa[,1])
nortest::ad.test(iris_setosa[,2])
nortest::ad.test(iris_setosa[,3])
MVN::mvn(
  iris_setosa,
  univariateTest = "AD" # "SW"/"CVM"/"Lillie"/"SF"/"AD"
)$univariateNormality

# chi^2 Q-Q plot
d_square = diag(
  as.matrix(sweep(iris_setosa, 2, colMeans(iris_setosa))) %*%
    solve(var(iris_setosa)) %*%
    t(as.matrix(sweep(iris_setosa, 2, colMeans(iris_setosa))))
)
car::qqPlot(d_square, dist="chisq", df = p, id = F)
MVN::mvn(
  iris_setosa,
  multivariatePlot = "qq"
)

# Multivariate normality test
MVN::mvn(
  iris_setosa,
  mvnTest = "dh" # "mardia"/"hz"/"royston"/"dh"/"energy"
)$multivariateNormality
```

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: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
- Likelihood function

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

- 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

$$(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}}) = \arg\max_{\boldsymbol{\mu}\in\mathbb{R}^p, \boldsymbol{\Sigma}\in\mathbb{R}^{p\times p}, \boldsymbol{\Sigma}>0} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\bar{\boldsymbol{X}}, \frac{n-1}{n}\mathbf{S})$$

  - Consistency: $(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$ approaches $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (in certain sense) as $n \to \infty$
  - Efficiency: the covariance matrix of $(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$ 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 $g(\hat{\boldsymbol{\mu}}_{\mathrm{ML}}, \widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}})$.

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
  - $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \mathrm{MVN}_p\,(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
  - $\mathbf{S} \perp\!\!\!\perp \bar{\boldsymbol{X}}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}} \perp\!\!\!\perp \hat{\boldsymbol{\mu}}_{\mathrm{ML}}$
  - $\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I})$
  - $(n-1)\mathbf{S} = n\widehat{\boldsymbol{\Sigma}}_{\mathrm{ML}} \sim W_p(\boldsymbol{\Sigma}, n-1)$
  - $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim$ Hotelling's $T^2(p, n-1)$

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

- Hotelling's T^2 distribution
  - A generalization of (Student's) t-distribution

- If $\boldsymbol{X} \sim \mathrm{MVN}_p(\mathbf{0}, \mathbf{I})$ 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 $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1)$ 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 $\boldsymbol{\mu}$ (under the normality assumption)

Likelihood ratio test (LRT)

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

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

  - $\boldsymbol{x}$: 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

$$R_\alpha = \{\boldsymbol{x} : -2 \ln \lambda(\boldsymbol{x}) \geq \chi^2_{\nu, 1-\alpha}\}$$

  - I.e., reject H_0 when $-2 \ln \lambda(\boldsymbol{x}) \geq \chi^2_{\nu, 1-\alpha}$
  - $\chi^2_{\nu, 1-\alpha}$ 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

$$p(\boldsymbol{x}) = 1 - F_{\chi^2(\nu)}\{-2 \ln \lambda(\boldsymbol{x})\}$$

  - $F_{\chi^2(\nu)}(\cdot)$ is the cdf of $\chi^2(\nu)$