STAT 3690 Lecture Note

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

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

2023/Feb/10 14:13:50

Multivariate normal (MVN) distribution (con'd)

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

- Checkcing the univariate normality
  - 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])
## via MVN::mvn()
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
- Checking 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.

```r
library(datasets)
data(iris)
head(iris)
iris_setosa = iris[iris$Species=='setosa', 1:3]

iris_setosa = iris_setosa - min(iris_setosa) + 1 # make sure all the entries are positive

(lambda = EnvStats::boxcox(iris_setosa[,2], optimize=T)$lambda)
if (lambda != 0){
```

```
  df_new = (iris_setosa[,2]^lambda-1)/lambda
}else df_new = log(iris_setosa[,2])

car::qqPlot(df_new, id = F)
shapiro.test(df_new)
nortest::ad.test(df_new)
```

- Multivariate Box-Cox transformation

```
(lambdas = MVN::mvn(
  iris_setosa,
  bc = T,
  bcType = 'optimal'
)$BoxCoxPowerTransformation)
for (i in 1:length(lambdas)){
  if (lambdas[i] != 0){
    iris_setosa_new[,i] = (iris_setosa[,i]^lambdas[i]-1)/lambdas[i]
  }else iris_setosa_new[,i] = log(iris_setosa[,i])
}
MVN::mvn(
  iris_setosa_new,
  mvnTest = "energy" # "mardia"/"hz"/"royston"/"dh"/"energy"
)$multivariateNormality
```

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

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

- 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

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

  - Consistency: $(\hat{\boldsymbol{\mu}}_{\text{ML}}, \widehat{\boldsymbol{\Sigma}}_{\text{ML}})$ 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 $\mathbf{S}$ (J&W Sec 4.4)

- Recall the univariate case: if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, then
  - $s^2 \perp\!\!\!\perp \bar{X}$

    - 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: if $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$, then
  - $\mathbf{S} \perp\!\!\!\perp \bar{\boldsymbol{X}}$, i.e., $\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \perp\!\!\!\perp \hat{\boldsymbol{\mu}}_{\text{ML}}$
  - $\sqrt{n}\boldsymbol{\Sigma}^{-1/2}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$
  - $(n-1)\mathbf{S} = n\widehat{\boldsymbol{\Sigma}}_{\text{ML}} \sim W_p(\boldsymbol{\Sigma}, n-1)$
  - $n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \mathbf{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \sim$ 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 $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\mathbf{0}, \boldsymbol{\Sigma})$
    - A generalization of χ^2-distribution: $W_p(\boldsymbol{\Sigma}, n) = \chi^2(n)$ if $p = \boldsymbol{\Sigma} = 1$
  - 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 $n \geq p \Rightarrow$
$$\frac{\boldsymbol{a}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{W}^{-1} \boldsymbol{a}} \sim \chi^2(n-p+1)$$
    - $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n) \Rightarrow$
$$\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{W}) \sim \chi^2(np)$$

- Hotelling's T^2 distribution
  - A generalization of (Student's) t-distribution
  - If $\boldsymbol{X} \sim \text{MVN}_p(\mathbf{0}, \mathbf{I})$ 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 $\mathbf{W}_1 \sim W_p(\boldsymbol{\Sigma}, n_1)$ 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
$$\lambda(\mathcal{X}) = \frac{L(\hat{\boldsymbol{\theta}}_0; \mathcal{X})}{L(\hat{\boldsymbol{\theta}}; \mathcal{X})}$$
  - $\mathcal{X}$: all the observations/the entire dataset
  - 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) level α rejection region (with respect to $\lambda(\mathcal{X})$)
$$R_\alpha = \{\lambda(\mathcal{X}) : -2\ln\lambda(\mathcal{X}) \geq \chi^2_{1-\alpha,\nu}\}$$
  - I.e., reject H_0 when $-2\ln\lambda(\mathcal{X}) \geq \chi^2_{1-\alpha,\nu}$
  - $\chi^2_{1-\alpha,\nu}$ is the $(1-\alpha)$-quantile of $\chi^2(\nu)$
  - ν: the difference in numbers of free parameters without/with H_0
- (Asymptotic) p-value
$$p(\mathcal{X}) = 1 - F_{\chi^2(\nu)}\{-2\ln\lambda(\mathcal{X})\}$$
  - $F_{\chi^2(\nu)}(\cdot)$ is the cdf of $\chi^2(\nu)$

Testing $\boldsymbol{\mu}$ (J&W Sec. 5.2 & 5.3)

- Sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} \text{MVN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $n > p$
  - $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, the set of all the data
- $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ v.s. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$
- Recall the univariate case ($p = 1$)
  - The model reduces to $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$
  - Hypotheses reduces to $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$
  - $\bar{X}$ and s^2 are sample mean and sample variance, respectively
  - Known σ^2
    * Name of approach: Z-test (equiv. LRT)
    * Test statistic: $T(\mathcal{X}) = \sqrt{n}(\bar{X} - \mu_0)/\sigma$ ($\sim \mathcal{N}(0,1)$ under H_0)
    * Level α Rejection region (with respect to $T(\mathcal{X})$): $R_\alpha = \{T(\mathcal{X}) : |T(\mathcal{X})| \geq \Phi^{-1}_{1-\alpha/2}\}$, i.e., reject H_0 if $|T(\mathcal{X})| \geq \Phi^{-1}_{1-\alpha/2}$
      · Critical point: $\Phi^{-1}_{1-\alpha/2}$, the $(1-\alpha/2)$-quantile of $\mathcal{N}(0,1)$
  - Unknown σ^2
    * Name of approach: t-test (equiv. LRT)
    * Test statistic: $T = \sqrt{n}(\bar{X} - \mu_0)/s$ ($\sim t(n-1)$ under H_0)
    * Level α rejection region (with respect to $T(\mathcal{X})$): $R_\alpha = \{T(\mathcal{X}) : |T(\mathcal{X})| \geq t_{1-\alpha/2,n-1}\}$, i.e., reject H_0 if $|T(\mathcal{X})| \geq t_{1-\alpha/2,n-1}$
      · Critical point: $t_{1-\alpha/2,n-1}$, the $(1-\alpha/2)$-quantile of $t(n-1)$

- Multivariate case (with known $\boldsymbol{\Sigma}$)
  - Name of approach: LRT
  - Test statistic: $T(\mathcal{X}) = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)$ ($\sim \chi^2(p)$ under H_0)
  - Level α rejection region (with respect to $T(\mathcal{X})$): $R_\alpha = \{T(\mathcal{X}) : T(\mathcal{X}) \geq \chi^2_{1-\alpha,p}\}$, i.e., reject H_0 if $T(\mathcal{X}) \geq \chi^2_{1-\alpha,p}$
    * Critical point: $\chi^2_{1-\alpha,p}$, the $(1-\alpha)$-quantile of $\chi^2(p)$
  - p-value: $p(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = 1 - F_{\chi^2(p)}(T)$
    * $F_{\chi^2(p)}(\cdot)$: the cdf of $\chi^2(p)$

```
options(digits = 4)
install.packages(c("dslabs"))
library(dslabs)
data("gapminder")
head(gapminder)
dataset = as.matrix(gapminder[
  !is.na(gapminder$infant_mortality),
  c("infant_mortality", "life_expectancy", "fertility")])

# Assume we know Sigma
Sigma <- matrix(c(555, -170, 30,
                  -170, 65, -10,
                  30, -10, 2), ncol = 3)

(mu_hat <- colMeans(dataset))

# Test mu = mu_0
mu_0 <- c(25, 50, 3)
n = nrow(dataset)
p = ncol(dataset)
(test.stat <- drop(
   n * t(mu_hat - mu_0) %*% solve(Sigma) %*% (mu_hat - mu_0)
))
test.stat >= qchisq(0.95, df=p)
(p.val = 1-pchisq(test.stat, df=p))
```

- Report: Testing hypotheses $H_0 : \boldsymbol{\mu} = [25, 50, 3]^\top$ v.s. $H_1 : \boldsymbol{\mu} \neq [25, 50, 3]^\top$, we carried on the LRT and obtained 450477 as the value of test statistic and $[7.815, \infty)$ as the corresponding level .05 rejection region. In addition, the p-value was around 0. So, at the .05 level, there was a strong statistical evidence implying the rejection of H_0, i.e., we believed that the population mean vector was not $[25, 50, 3]^\top$.