STAT 3690 Homework 3

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Due at Apr 11 11:59 pm (Central Time)

Answers must be submitted electronically via Crowdmark. Please enclose your R code trunks (if applicable) as well.

1. Figure out the covariance matrix for random vector $\mathbf{Y} = [Y_1, Y_2, Y_3]^\top$, when $\mathbf{Y}$ satisfies the following one-factor model:
   - $Y_1 = .6F + \epsilon_1$, $Y_2 = .8F + \epsilon_2$, and $Y_3 = .5F + \epsilon_3$;
   - Factors F, ϵ_1, ϵ_2, and ϵ_3 are uncorrelated with each other;
   - $\mathrm{var}(F) = 1$, $\mathrm{var}(\epsilon_1) = .64$, $\mathrm{var}(\epsilon_2) = .36$, and $\mathrm{var}(\epsilon_3) = .75$.

Answer to Q1.

```
##      [,1] [,2] [,3]
## [1,] 1.00 0.48  0.3
## [2,] 0.48 1.00  0.4
## [3,] 0.30 0.40  1.0
```

2. We can formally test whether a covariance matrix is diagonal via likelihood ratio test (LRT). Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\mathbf{X}_i = [X_{i1}, \ldots, X_{ip}]^\top$, $i = 1, \ldots, n$.
   a. When $\boldsymbol{\Sigma}$ is diagonal, i.e.,
   $$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{bmatrix},$$
   please point out WITHOUT proof the maximum likelihood estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.
   b. Use the result of part a to derive the LRT statistic λ for hypotheses $H_0 : \boldsymbol{\Sigma}$ is diagonal vs. H_1 : otherwise. Specifically,
   $$\lambda = \frac{\text{maximum likelihood when } (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}_0}{\text{maximum likelihood when } (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}},$$
   where $\boldsymbol{\Theta}_0 = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p} \text{ is diagonal and positive-definite}\}$ and $\boldsymbol{\Theta} = \{(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{\Sigma} \in \mathbb{R}^{p\times p} \text{ is positive-definite}\}$.

Answer to Q2a. When $\boldsymbol{\Sigma}$ is diagonal, X_{ij}'s are independent of each other and $X_{ij} \sim N(\mu_j, \sigma_j^2)$ for all i, where μ_j is the jth entry of $\boldsymbol{\mu}$. So, the maximum likelihood estimator (MLE) for μ_j is $\hat{\mu}_j = \bar{X}_{\cdot j} = n^{-1}\sum_{i=1}^n X_{ij}$, whereas the MLE for σ_j^2 is $\hat{\sigma}_j^2 = n^{-1}\sum_{i=1}^n (X_{ij} - \bar{X}_{\cdot j})^2$. As a result, the MLEs for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = n^{-1}\sum_{i=1}^n \mathbf{X}_i$ and

$$\widehat{\boldsymbol{\Sigma}}_0 = \begin{bmatrix} \hat{\sigma}_1^2 & & \\ & \ddots & \\ & & \hat{\sigma}_p^2 \end{bmatrix} = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix} (X_{i1} - \bar{X}_{\cdot 1})^2 & & \\ & \ddots & \\ & & (X_{ip} - \bar{X}_{\cdot p})^2 \end{bmatrix},$$

respectively.

Answer to Q2b. The log likelihood function is

$$\begin{aligned}\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu}) \\ &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\boldsymbol{\Sigma}) - \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^\top\right\}\end{aligned}$$

If $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}_0$, the MLE for $\boldsymbol{\Sigma}$, say $\widehat{\boldsymbol{\Sigma}}_0$, is given as in the answer to Q2a; if $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \boldsymbol{\Theta}$, the MLE for $\boldsymbol{\Sigma}$ is known to be $\widehat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top$. Note that in both cases the MLE for $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$. Then, after plug-in,

$$\lambda = \exp\{\ell(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}_0) - \ell(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})\} = \left\{\frac{\det(\widehat{\boldsymbol{\Sigma}})}{\prod_{j=1}^{p}\hat{\sigma}_j^2}\right\}^{n/2},$$

since

$$\ell(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}_0) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\widehat{\boldsymbol{\Sigma}}_0) - \frac{np}{2} = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\prod_{j=1}^{p}\hat{\sigma}_j^2\right) - \frac{np}{2}$$

and

$$\ell(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\widehat{\boldsymbol{\Sigma}}) - \frac{np}{2}.$$

Alternatively, noting that $\mathbf{S} = n(n-1)^{-1}\widehat{\boldsymbol{\Sigma}}$ and $s_j^2 = n(n-1)^{-1}\hat{\sigma}_j^2$,

$$\lambda = \left\{\frac{\det(\mathbf{S})}{\prod_{j=1}^{p}s_j^2}\right\}^{n/2}.$$

3. Typically, for the LRT in Q2b, the null distribution of $-2\ln\lambda$ is approximated by $\chi^2(p(p-1)/2)$.
   a. If $p = 3$, what is the rejection region for λ at level $\alpha = .01$?
   b. A simulation study may be used to investigate the Type I error rate of the LRT in Q2b. Generate a dataset of size $n = 100$ with $\boldsymbol{\mu} = [1, 2, 3]^\top$ and $\boldsymbol{\Sigma} = \mathbf{I}_3$ and then calculate λ for this dataset. Repeat $B = 1000$ times (i.e. you will have B realizations of λ). With each realization of λ, test hypotheses $H_0 : \boldsymbol{\Sigma}$ is diagonal vs. H_1 : otherwise. Get a conclusion on the goodness of χ^2-approximation by comparing $\alpha = .01$ with the Type I error rate (i.e., the rejection proportion in the B tests).

Answer to Q3a. The rejcetion region for λ is

```
## [ 0 , 0.003439 ]
```

Answer to Q3b. The simulated Type I error rate is .007 (varying with the random seed and the function that generates multivariate normal samples), close to stated level .01. So, the χ^2-approximation works fine.

4. Suppose

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & .48 & .3 \\ .48 & 1 & .4 \\ .3 & .4 & 1 \end{bmatrix}.$$

   Carry on a factor analysis with one common factor via the principal component method without any rotation. Compare your results with the model posited in Q1.

Answer to Q4. Since $\boldsymbol{\Sigma}$ is identical to the covariance matrix of $\mathbf{Y}$ in Q1, the population value of loading matrix and covariance matrix of error are given in Q1. Comparing estimates and their theoretical counterparts, we may see that the second component of $\mathbf{Y}$ is better approximated than the other two.

5. We are going to use dataset `OTE::Body` on 21 body measurements.

a. Fit a factor model with $q = 3$ common factors for the 21 body measurements of female subjects only, using the maximum likelihood method. What are the estimates for the varimax loading matrix and score matrix (via weighted least squares), respectively? (**If the two matrices are huge, you may choose to report the first several rows of them by using `R` function `head()`.**)
b. Illustrate and explain the association between each factor and Age, Weight, and Height.
c. Refit the model for male observations only and compare the resulting loadings to those obtained in part a. Given these results, do you think the factor analysis is capturing the same unobserved structure for both genders?

```
dataset = OTE::Body[OTE::Body$Gender==0,!names(OTE::Body) %in% c('Gender','Age','Weight','Height')]
names(dataset)
```

```
##  [1] "Biacrom"   "Biiliac"   "Bitro"     "ChestDp"   "ChestD"    "ElbowD"
##  [7] "WristD"    "KneeD"     "AnkleD"    "ShoulderG" "ChestG"    "WaistG"
## [13] "AbdG"      "HipG"      "ThighG"    "BicepG"    "ForearmG"  "KneeG"
## [19] "CalfG"     "AnkleG"    "WristG"
```

Answer to Q5a. See below for (the first six rows of) the varimax loading matrix and score matrix, respectively.

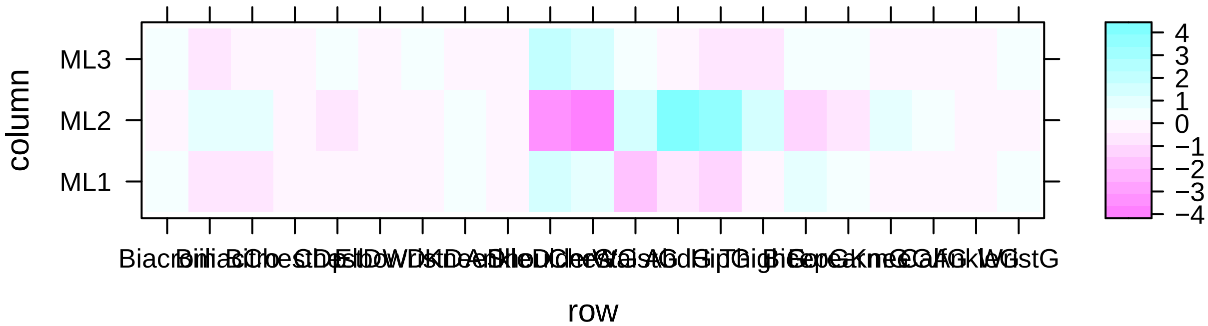
```
##            ML1    ML2    ML3
## Biacrom 0.3584 0.2252 0.9723
## Biiliac 0.5351 1.1458 0.3884
## Bitro   0.4063 1.3096 0.8208
## ChestDp 0.9890 0.6767 0.2566
## ChestD  1.0568 0.2194 0.8909
## ElbowD  0.2359 0.1905 0.5600
```

```
##          ML1     ML2      ML3
## 248 -0.01624 -0.0801 -1.70284
## 249 -2.14983  0.1561  1.45785
## 250 -1.62653  0.4222 -1.25502
## 251  0.05444  1.2493 -0.67861
## 252  1.56403 -1.1538 -1.14004
## 253 -1.30243  0.3438  0.02507
```

Answer to Q5b. Variable `Weight` is correlated with all the three factors, while `Height` is mainly associated with the third factor. It seems that `Age` merely enjoys a weak correlation with the 1st and 2nd factors.

```
##             ML1    ML2      ML3
## Age     0.17958 0.1879 0.003078
## Weight  0.52822 0.5728 0.483722
## Height -0.06531 0.2044 0.418218
```

Answer to Q5c. See below for the plot on the difference between the two loading matrices. The difference is apparent, especially in the loadings of the 2nd factor. We can see that again variable `Weight` is correlated with all the three factors, while `Height` is mainly associated with the third factor. But now the correlation between `Age` and the 1st factor is enhanced.




```
##          ML1      ML2     ML3
## Age    0.4770 -0.08409 -0.1192
## Weight 0.6381  0.45639  0.4692
## Height 0.1412  0.09127  0.4638
```

Appendix

```
options(digits = 4)
set.seed(3690)
## Q1
L = c(.6, .8, .5)
Psi = diag(c(.64, .36, .75))
tcrossprod(L) + Psi
## Q3a
cat("[",0, ",",exp(qchisq(1-.01, 3)/(-2)),"]")
## Q3b
B = 1000
n = 100
Mu = 1:3
Sigma = diag(3)
lambdas = numeric(0)
for (b in 1:B){
  X = MASS::mvrnorm(n, Mu, Sigma)
  S_ml = ((n-1)/n)*cov(X)
  lambdas = c(lambdas, (det(S_ml)/prod(diag(S_ml)))^(n/2))
}
mean(lambdas<=exp(qchisq(1-.01, 3)/(-2)))
```

```r
## Q4
Sigma = matrix(
  c(1, .48, .3,
    .48, 1, .4,
    .3, .4, 1),
ncol = 3)
pc_decomp = eigen(Sigma)
L_pc = pc_decomp$vectors[,1] * pc_decomp$values[1]^.5
Psi_pc = diag(diag(Sigma - tcrossprod(L_pc)))
# estimate - population value
L_pc - L
Psi_pc - Psi
## Q5a
q = 3
fa_decomp_f = psych::fa(r=dataset, covar=T, nfactors=q, rotate="varimax", fm ="ml")
L_ml_f = fa_decomp_f$loadings
head(L_ml_f)
Psi_ml_f = diag(fa_decomp_f$uniquenesses)
Psi_ml_f_inv = diag(fa_decomp_f$uniquenesses^-1)
Weight_mat_f = solve(t(L_ml_f) %*% Psi_ml_f_inv %*% L_ml_f) %*% t(L_ml_f) %*% Psi_ml_f_inv
scores_wls_f = scale(dataset, center = T, scale = F) %*% t(Weight_mat_f)
head(scores_wls_f)
## Q5b
dataset2 = OTE::Body[OTE::Body$Gender==0,names(OTE::Body) %in% c('Age','Weight','Height')]
cor(dataset2, scores_wls_f)
## Q5c
dataset_m = OTE::Body[OTE::Body$Gender==1,!names(OTE::Body) %in% c('Gender','Age','Weight','Height')]
fa_decomp_m = psych::fa(r=dataset_m, covar=T, nfactors=q, rotate="varimax", fm ="ml")
L_ml_m = fa_decomp_m$loadings
lattice::levelplot(unclass(L_ml_f - L_ml_m))
Psi_ml_m = diag(fa_decomp_m$uniquenesses)
Psi_ml_m_inv = diag(fa_decomp_m$uniquenesses^-1)
Weight_mat_m = solve(t(L_ml_m) %*% Psi_ml_m_inv %*% L_ml_m) %*% t(L_ml_m) %*% Psi_ml_m_inv
scores_wls_m = scale(dataset_m, center = T, scale = F) %*% t(Weight_mat_m)
dataset3 = OTE::Body[OTE::Body$Gender==1,names(OTE::Body) %in% c('Age','Weight','Height')]
cor(dataset3, scores_wls_m)
```