Math3806 Lecture Note 5 Appendix

Heng Peng

March 27, 2020

P5. Result 5.1. Proof: Let $\mathbf{a}^* = \mathbf{a}/\|\mathbf{a}\|$, then

$$\begin{aligned}
\frac{\mathbf{a}^T\Sigma\mathbf{a}}{\mathbf{a}^T\mathbf{a}} &= \mathbf{a}^{*T}\Sigma\mathbf{a}^* = \mathbf{a}^{*T}\left\{\sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i^T\right\}\mathbf{a}^* = \sum_{i=1}^{p}\lambda_i(\mathbf{a}^{*T}\mathbf{e}_i)^2 \\
&\leq \sum_{i=1}^{p}\lambda_1(\mathbf{a}^{*T}\mathbf{e}_i)^2 = \lambda_1\left\{\sum_{i=1}^{p}\mathbf{a}^{*T}\mathbf{e}_i\mathbf{e}_i^T\mathbf{a}^*\right\} \\
&= \lambda_1\mathbf{a}^{*T}PP^T\mathbf{a}^* = \lambda_1\|\mathbf{a}^*\|^2 = \lambda_1
\end{aligned}$$

Hence when $\mathbf{a} = \mathbf{e}_1$

$$\max_{\mathbf{a}\neq 0}\frac{\mathbf{a}^T\Sigma\mathbf{a}}{\mathbf{a}^T\mathbf{a}} = \lambda_1$$

Similar Let $\mathbf{a}^* = \mathbf{a}/\|\mathbf{a}\|$ with $\mathbf{a}^T\mathbf{e}_i = 0, i = 1, \ldots, k$

$$\begin{aligned}
\frac{\mathbf{a}^T\Sigma\mathbf{a}}{\mathbf{a}^T\mathbf{a}} &= \mathbf{a}^{*T}\Sigma\mathbf{a}^* = \mathbf{a}^{*T}\left\{\sum_{i=1}^{p}\lambda_i\mathbf{e}_i\mathbf{e}_i^T\right\}\mathbf{a}^* = \sum_{i=1}^{p}\lambda_i(\mathbf{a}^{*T}\mathbf{e}_i)^2 \\
&\leq \sum_{i=k+1}^{p}\lambda_{k+1}(\mathbf{a}^{*T}\mathbf{e}_i)^2 = \lambda_{k+1}\left\{\sum_{i=1}^{p}\mathbf{a}^{*T}\mathbf{e}_i\mathbf{e}_i^T\mathbf{a}^*\right\} \\
&= \lambda_{k+1}\mathbf{a}^{*T}PP^T\mathbf{a}^* = \lambda_{k+1}\|\mathbf{a}^*\|^2 = \lambda_{k+1}
\end{aligned}$$

So when $\mathbf{a} = \mathbf{e}_{k+1}$

$$\max_{\mathbf{a}\neq 0, \mathbf{a}\perp\mathbf{e}_1,\ldots,\mathbf{e}_k}\frac{\mathbf{a}^T\Sigma\mathbf{a}}{\mathbf{a}^T\mathbf{a}} = \lambda_{k+1}$$

- Result 5.1 continuous.

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathrm{Var}(\mathbf{e}_i^T \boldsymbol{X}) = \mathbf{e}_i^T \mathrm{Var}(\boldsymbol{X})\mathbf{e}_i = \mathbf{e}_i^T \Sigma \mathbf{e}_i \\
&= \mathbf{e}_i^T(\lambda_i \mathbf{e}_i) = \lambda_i \mathbf{e}_i^T \mathbf{e}_i = \lambda_i, i = 1, \ldots, p.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_k) &= \mathrm{Cov}(\mathbf{e}_i^T \boldsymbol{X}, \mathbf{e}_k^T \boldsymbol{X}) = \mathbf{e}_i^T \mathrm{Cov}(\boldsymbol{X})\mathbf{e}_k \\
&= \mathbf{e}_i^T \Sigma \mathbf{e}_k = \mathbf{e}_i^T(\lambda_k \mathbf{e}_k) = \lambda_k \mathbf{e}_i^T \mathbf{e}_k = 0, i \neq k.
\end{aligned}$$

P5. Result 5.2.

$$\begin{aligned}
\sigma_{11} + \cdots + \sigma_{pp} &= \sum_{i=1}^{p} \mathrm{Var}(X_i) = \mathrm{Trace}(\Sigma) = \mathrm{Trace}(P\Lambda P^T) \\
&= \mathrm{Trace}(\Lambda P^T P) = \mathrm{Trace}(\Lambda) = \lambda_1 + \cdots + \lambda_p.
\end{aligned}$$

Proportion of total population variance due to kth principle component is

$$\frac{\lambda_k}{\lambda_1 + \cdots + \lambda_p} = \frac{\lambda_k}{\sigma_1 + \cdots + \sigma_p}$$

P6. Result 5.3. Let $\mathbf{a}_k^T = [0, \ldots, 0, 1, 0, \ldots, 0]$ so that $X_k = \mathbf{a}_k^T \boldsymbol{X}$ and then

$$\text{Cov}(X_k, Y_i) = \text{Cov}(\mathbf{a}_k^T \boldsymbol{X}, \mathbf{e}_i^T \boldsymbol{X}) = \mathbf{a}_k^T(\lambda_i \mathbf{e}_i) = \lambda_i \mathbf{a}_k^T \mathbf{e}_i = \lambda_i e_{ik}$$

Hence

$$\rho_{Y_i, X_k} = \frac{\text{Cov}(Y_i, X_k)}{\sqrt{\text{Var}(Y_i)}\sqrt{\text{Var}(X_k)}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$$

- Example 5.1.

$$\lambda_1 = 5.83, \ \mathbf{e}_1^T = [.383, -.924, 0]$$
$$\lambda_2 = 2.00, \ \mathbf{e}_2^T = [0, 0, 1]$$
$$\lambda_3 = 0.17, \ \mathbf{e}_3^T = [.924, .383, 0]$$

Therefore, the principal components are

$$Y_1 = \mathbf{e}_1^T \boldsymbol{X} = .383X_1 - .924X_2$$

$$Y_2 = \mathbf{e}_2^T \boldsymbol{X} = X_3$$

$$Y_3 = \mathbf{e}_3^T \boldsymbol{X} = .924X_1 + .383X_2$$

$$\begin{aligned} \text{Var}(Y_1) &= (.383)^2\text{Var}(X_1) + (-.924)^2\text{Var}(X_2) + 2(.383)(-.924)\text{Cov}(X_1, X_2) \\ &= .147(1) + .854(5) - .708(-2) = 5.83 = \lambda_1 \end{aligned}$$

- Example 5.1 continuous.

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= \text{Cov}(.383X_1 - .924X_2, X_3) \\ &= .383\text{Cov}(X_1, X_3) - .924\text{Cov}(X_2, X_3) = 0 \end{aligned}$$

$$\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 5 + 2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2.00 + .17$$

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = .73, \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = .98$$

$$\rho_{Y_1, X_1} = \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{.383\sqrt{5.83}}{\sqrt{1}} = .925$$

$$\rho_{Y_1, X_2} = \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-.924\sqrt{5.83}}{\sqrt{5}} = -.998$$

$$\rho_{Y_1, X_3} = \frac{e_{13}\sqrt{\lambda_1}}{\sqrt{\sigma_{33}}} = \frac{0\sqrt{5.83}}{\sqrt{2}} = 0$$

P9. Result 5.4. The result from the above result with $Z_1, \ldots, Z_p$ in place of $X_1, \ldots$ and $\boldsymbol{\rho}$ in place of Σ.

- Proportion of (standardized) population variance due to kth principle component is $\frac{\lambda_k}{p}$, $k = 1, \ldots, p$ where the $\lambda_k's$ are the eigenvalue of $\boldsymbol{\rho}$.

P10. Example 5.2. The eigenvalue-eigenvector pairs from Σ are

$$\lambda_1 = 100.16, \mathbf{e}_1^T = [.040, .999]$$

$$\lambda_2 = .84, \mathbf{e}_2^T = [.999, -.040]$$

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{100.16}{101} = .992$$

The respective principal components based on Σ are

$$Y_1 = .040X_1 + .999X_2, Y_2 = .999X_1 - .040X_2.$$

Similarly, the eigenvalue-eigenvector pairs from $\boldsymbol{\rho}$ are

$$\lambda_1 = 1 + \rho = 1.4, \mathbf{e}_1^T = [.707, .707]$$

$$\lambda_2 = 1 - \rho = .6, \mathbf{e}_2^T = [.707, -.707]$$

Then the respective principal component based on $\boldsymbol{\rho}$ are

$$\begin{aligned} Y_1 &= .707Z_1 + .707Z_2 = .707\frac{X_1 - \mu_1}{1} + 0707\frac{X_2 - \mu_2}{10} \\ &= .707(X_1 - \mu_1) + .0707(X_2 - \mu_2) \\ Y_2 &= .707Z_1 - .707Z_2 = .707\frac{X_1 - \mu_1}{1} - 0707\frac{X_2 - \mu_2}{10} \\ &= .707(X_1 - \mu_1) - .0707(X_2 - \mu_2) \end{aligned}$$

$$\frac{\lambda_1}{p} = \frac{1.4}{2} = .7$$

P10. (1) For the diagonal covariance or correlation matrix, $(\sigma_{ii}, \mathbf{e}_i)$ is the ith eigenvalue-eigenvector pair with $\mathbf{e}_i^T \boldsymbol{X} = X_i$ or $\mathbf{e}_i^T = [0, \ldots, 1, \ldots, 0]$. Hence the set of principle components is just the original set of uncorrelated random variables.

- (2) It is not difficult to show (as excercise)that the p eigenvalues of the correlation matrix can be divided into two groups. When ρ is positive, the largest is

$$\lambda_1 = 1 + (p-1)\rho$$

with associate eigenvector

$$\mathbf{e}_1^T = \left[\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}} \cdots, \frac{1}{\sqrt{p}}\right]$$

The remaining $p-1$ eigenvalues are

$$\lambda_2 = \lambda_3 = \cdots = \lambda_p = 1 - \rho$$

- Continuous: and their eigenvectors are

$$\mathbf{e}_2^T = \left[\frac{1}{\sqrt{1\times 2}}, \frac{-1}{\sqrt{1\times 2}}, 0, \ldots, 0\right]$$

$$\mathbf{e}_3^T = \left[\frac{1}{\sqrt{2\times 3}}, \frac{1}{\sqrt{2\times 3}}, \frac{-2}{\sqrt{2\times 3}} 0, \ldots, 0\right]$$

$$\cdots\cdots \qquad \cdots\cdots$$

$$\mathbf{e}_i^T = \left[\frac{1}{\sqrt{(i-1)\times i}}, \cdots, \frac{1}{\sqrt{(i-1)\times i}}, \frac{-(i-1)}{\sqrt{(i-1)\times i}}, 0, \ldots, 0\right]$$

$$\cdots\cdots \qquad \cdots\cdots$$

$$\mathbf{e}_p^T = \left[\frac{1}{\sqrt{(p-1)\times p}}, \cdots, \frac{1}{\sqrt{(p-1)\times p}}, \frac{-(p-1)}{\sqrt{(p-1)\times p}}\right]$$

The first principal component

$$Y_1 = \mathbf{e}_1^T Z = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} Z_i$$

explaining a proportion

$$\frac{\lambda_1}{p} = \frac{1+(p-1)\rho}{p} = \rho + \frac{1-\rho}{p}$$

of the total population variation.

P16. Example 5.4. The natural logarithms of the dimension of 24 male turtles have sample mean vector $\bar{\mathbf{x}}^T = [4.725, 4.478, 3.703]$ and covariance matri

$$\mathbf{S} = 10^{-3} \begin{bmatrix} 11.072 & 8.019 & 8.160 \\ 8.019 & 6.417 & 6.005 \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

The first principal component

$$\hat{y}_1 = .683 \ln(\text{length}) + .510 \ln(\text{width}) + .523 \ln(\text{height})$$

which explains 96% of the total variance.

P23. Example 5.5. Let $x_1, x_2, \ldots, x_5$ denote the observed weekly rates of return for JP Morgan, Citibank, Well Fargo, Royal Dutch Shell, and ExxonMobil, respectively. Then

$$\bar{\mathbf{x}}^T = [.0011, .0007, .0016, .0040, .0040]$$

and

$$\mathbf{R} = \begin{bmatrix} 1.000 & .632 & .511 & .115 & .155 \\ .632 & 1.000 & .574 & .322 & .213 \\ .511 & .574 & 1.000 & .183 & .146 \\ .115 & .322 & .183 & 1.000 & .683 \\ .155 & .213 & .146 & .683 & 1.000 \end{bmatrix}$$

We note that $\mathbf{R}$ is the covariance matrix of the standardized observations

$$z_1 = \frac{x_1 - \bar{x}_1}{\sqrt{s_{11}}}, z_2 = \frac{x_2 - \bar{x}_2}{\sqrt{s_{22}}}, \ldots, z_1 = \frac{x_5 - \bar{x}_5}{\sqrt{s_{55}}}$$

- Example 5.5. Continuous. The eigenvalues and corresponding normalized eigenvectors of **R**, determined by a computer, are

$$
\begin{aligned}
\hat{\lambda}_1 &= 2.437, \ \hat{\mathbf{e}}_1^T = [.469, .532, .465, .387, .361] \\
\hat{\lambda}_2 &= 1.407, \ \hat{\mathbf{e}}_2^T = [-.368, -.236, -.315, .585, .606] \\
\hat{\lambda}_3 &= .501, \ \hat{\mathbf{e}}_3^T = [-.604, -.136, .772, .093, -.109] \\
\hat{\lambda}_4 &= .400, \ \hat{\mathbf{e}}_4^T = [.363, -.629, .289, -.381, .493] \\
\hat{\lambda}_5 &= .255, \ \hat{\mathbf{e}}_5^T = [.384, -.496, .071, .595, -.498]
\end{aligned}
$$

  Under the standardized variables, we obtain the first two sample principal components:

$$
\begin{aligned}
\hat{y}_1 &= \mathbf{e}_1^T\mathbf{z} = .469z_1 + .532z_2 + .465z_3 + .387z_4 + .361z_5 \\
\hat{y}_2 &= \mathbf{e}_2^T\mathbf{z} = -.368z_1 - .236z_2 - .315z_3 + .585z_4 + .606z_5
\end{aligned}
$$

- These components, which account for

$$
\left(\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p}\right) 100\% = \left(\frac{2.437 + 1.407}{5}\right) 100\% = 77\%
$$

- The first principle component: *a general stock-market component*, The second principle component: *an industry component*.

P24. Example 5.6. The eigenvalues of the covariance matrix are

$$\hat{\lambda}_1 = 3.085, \hat{\lambda}_2 = .383, \hat{\lambda}_3 = .342, \hat{\lambda}_4 = .217$$

We note that the first eigenvalue is nearly equal to

$$1 + (p-1)\bar{r} = 1 + (4-1)(.6854) = 3.056$$

where $\bar{r}$ is the arithmetic average of the off-diagonal elements of $\mathbf{R}$.

- The remaining eigenvalues are small and about equal, though $\hat{\lambda}_4$ is somewhat smaller than $\hat{\lambda}_2$ and $\hat{\lambda}_3$. So there is some evidence that the corresponding population correlation matrix $\boldsymbol{\rho}$ may be of the "equal-correlation" form.
- The first component

$$\hat{y}_1 = \hat{\mathbf{e}}_1 \mathbf{z} = .49z_1 + .52z_2 + .49z_3 + .50z_4$$

accounts for

$$100(\hat{\lambda}_1/p)\% = 100(3.058)/4\% = 76\%$$

of the total variance.

*** **Comment**: Although "large" eigenvalues and the corresponding eigenvectors are important in a principle component analysis, eigenvalues very close to zero should not be routinely ignored. The eigenvectors associated with these latter eigenvalues may point out linear dependencies in the data set that can cause interpretive and computational problems in a subsequent analysis.

P27.

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T &= \mathrm{E}(\mathbf{LF}+\varepsilon)(\mathbf{LF}+\varepsilon)^T \\
&= \mathrm{E}\mathbf{LF}(\mathbf{LF})^T + \mathrm{E}\varepsilon(\mathbf{LF})^T + \mathrm{E}\mathbf{LF}\varepsilon^T + \mathrm{E}\varepsilon\varepsilon^T \\
&= \mathbf{L}\mathrm{E}(\mathbf{FF}^T)\mathbf{L}^T + \mathrm{E}(\varepsilon\mathbf{F}^T)\mathbf{L}^T + \mathbf{L}\mathrm{E}(\mathbf{F}\varepsilon^T) + \mathrm{E}\varepsilon\varepsilon^T \\
&= \mathbf{LL}^T + \Psi
\end{aligned}
$$

- Communality:

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2$$

Specific variance: ψ_i

P27. Example 5.7.

$$\Sigma = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} = \mathbf{L}\mathbf{L}^T + \Psi$$

The Communality of X_1 is

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17$$

The special variance is $\psi_1 = 2$. Hence

$$19 = 4^2 + 1^2 + 2 = 17 + 2$$

P28. Example 5.8. If Σ can be factored by a factor analysis model with $m = 1$, then

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11} F_1 + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21} F_1 + \varepsilon_2 \\ X_3 - \mu_3 &= \ell_{31} F_1 + \varepsilon_3 \end{aligned}$$

or

$$1 = \ell_{11}^2 + \psi_1, \ .90 = \ell_{11}\ell_{21}, \ .70 = \ell_{11}\ell_{31}$$
$$1 = \ell_{21}^2 + \psi_2, \ .40 = \ell_{21}\ell_{31}, \ 1 = \ell_{31}^2 + \psi_3$$

The pair of equations

$$.70 = \ell_{11}\ell_{31}, \ .40 = \ell_{21}\ell_{31}$$

implies that

$$\ell_{21} = \left(\frac{.40}{.70}\right) \ell_{11}$$

Substituting this result for ℓ_{21} in the equation $.90 = \ell_{11}\ell_{21}$, yields

$$\ell_{11}^2 = 1.575$$

or $\ell_{11} = \pm 1.255$.

- Example 5.8. Continuous.
  Since $\mathrm{Var}(F_1) = 1$ by assumption and $\mathrm{Var}(X_1) = 1$,
  $\ell_{11} = \mathrm{Cov}(X_1, F_1) = \mathrm{Corr}(X_1, F_1)$ which cannot be greater than unity. So from this point of view $|\ell_{11}| = 1.225$ is too large. Also, the equation

  $$1 = \ell_{11}^2 + \psi_1$$

  gives

  $$\psi_1 = 1 - 1.575 = -.575$$

  which is unsatisfactory, since it gives a negative value for $\mathrm{Var}(\varepsilon) = \psi_1$. So the solution is not consistent, and is not a proper solution.

P29. if $\hat\lambda_1, \ldots, \hat\lambda_m$ are relative large compared to $\hat\lambda_{m+1}, \ldots, \hat\lambda_p$

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = \hat\lambda_1 \hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + \cdots + \hat\lambda_m \hat{\mathbf{e}}_m \hat{\mathbf{e}}_m^T \approx \mathbf{S}$$

$$\tilde\psi_i = s_{ii} - \sum_{j=1}^{m} \tilde\ell_{ij}^2$$

P30.

$$\tilde\ell_{11}^2 + \tilde\ell_{21}^2 + \cdots + \tilde\ell_{p1}^2 = (\sqrt{\hat\lambda_1}\hat{\mathbf{e}}_1)^T(\sqrt{\hat\lambda_1}\hat{\mathbf{e}}_1) = \hat\lambda_1.$$

P31 Example 5.9.

$$\begin{aligned}\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T + \tilde{\mathbf{\Psi}} &= \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix} \\ &+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1 & .01 & .97 & .44 & .00 \\ & 1 & .11 & .79 & .91 \\ & & 1 & .53 & .11 \\ & & & 1 & .81 \\ & & & & 1 \end{bmatrix}\end{aligned}$$

P33. Example 5.10.

$$\mathbf{R}-\tilde{\mathbf{L}}\tilde{\mathbf{L}}^T-\tilde{\Psi}=\begin{bmatrix} 0 & -.099 & -.185 & -.025 & .056 \\ -.099 & 0 & -.134 & .014 & -.054 \\ -.185 & -.134 & 0 & .003 & .006 \\ -.025 & .014 & .003 & 0 & -.156 \\ .056 & -.054 & .006 & -.156 & 0 \end{bmatrix}$$

P35. Corrected:

$$\hat{\mathbf{L}}_z=\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}},\ \hat{\Psi}_z=\hat{\mathbf{V}}^{-1/2}\hat{\Psi}\hat{\mathbf{V}}^{-1/2}$$

Or given the estimated loadings $\hat{\mathbf{L}}_z$ and specific variance $\hat{\Psi}_z$ obtained from $\mathbf{R}$, the resulting maximum likelihood estimates for a factor analysis of the covariance matrix $[(n-1)/n]\mathbf{S}$ are

$$\hat{\mathbf{L}}=\hat{\mathbf{V}}^{1/2}\hat{\mathbf{L}}_z,\ \hat{\Psi}=\hat{\mathbf{V}}^{1/2}\hat{\Psi}_z\hat{\mathbf{V}}^{1/2}$$

or

$$\hat{\ell}_{ij}=\hat{\ell}_{z,ij}\sqrt{\hat{\sigma}_{ij}} \quad \text{and} \quad \hat{\psi}_i=\hat{\psi}_{z,i}\hat{\sigma}_{ii}$$

P33. Example 5.11.

$$\mathbf{R} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T - \tilde{\Psi} = \begin{bmatrix} 0 & .001 & -.002 & .000 & .052 \\ .001 & 0 & .002 & .000 & -.033 \\ -.002 & .002 & 0 & .000 & .001 \\ .000 & .000 & .000 & 0 & .000 \\ .052 & -.033 & .001 & .000 & 0 \end{bmatrix}$$

Factor 1, *market factor*, Factor 2, *banking factor*

P40. Example 5.13. Factor 1: *General Intelligence*, Factor 2: *bipolar factor* . After Rotation, Factor 1: *mathematical ability* , Factor 2: *verbal ability*

P43. Example 5.14. After rotation, Factor 1: *nutritional factor*, Factor 2: *taste factor*

P45. Example 5.15. After rotation, Factor 1: *Unique economic forces that cause bank stocks to move together*, Factor 2: *economic conditions affecting oil stocks.*

P46. Example 5.12 and 5.16. principle component estimate: Factor 1: *general athletic ability*, Factor 2, *running endurance factor*, The remaining factors cannot be easily interpreted to our minds.

Maximum likelihood estimate: Before rotation: Factor 1: *General athletic ability* , Factor 2, *strength ability* , Factor 3, *running endurance ability*, Factor 4, Jumping ability or leg ability ???

After rotation, Factor 1: *explosive arm strength*, Factor 2: *explosive leg ability.* Factor 3, *running speed*, Factor 4, *running endurance*

P50. The joint distribution of $(\boldsymbol{X} - \boldsymbol{\mu})$ and $\mathbf{F}$ is

$$\Sigma^* = \begin{bmatrix} \Sigma = \mathbf{L}\mathbf{L}^T + \Psi & \mathbf{L} \\ \mathbf{L}^T & \mathbf{I} \end{bmatrix}$$

Then

$$\mathrm{E}(\mathbf{F}|\mathbf{x}) = \mathbf{L}^{\mathbf{T}}\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T + \Psi)^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

and

$$\mathrm{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}^T\Sigma^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}^T\Sigma^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}^T(\mathbf{L}\mathbf{L}^T + \Psi)^{-1}\mathbf{L}$$

- If rotated loadings $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ are used in place of the original loadings, the subsequent factor scores $\hat{\mathbf{f}}_j^*$ are related to $\hat{\mathbf{f}}_j$ by

$$\hat{\mathbf{f}}_j^* = \mathbf{T}\hat{\mathbf{f}}_j, j = 1, 2, \ldots, n.$$

P51. Example 5.16.

$$\hat{\mathbf{L}}_z^* = \begin{bmatrix} .763 & .024 \\ .821 & .227 \\ .669 & .104 \\ .118 & .993 \\ .113 & .675 \end{bmatrix} \quad \text{and} \quad \hat{\Psi}_z = \begin{bmatrix} .42 & 0 & 0 & 0 & 0 \\ 0 & .27 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .00 & 0 \\ 0 & 0 & 0 & 0 & .53 \end{bmatrix}$$

The vector of standardized observations,

$$\mathbf{z}^T = [.50, -1.40, -.20, -.70, 1.40]$$

Yield the following scores on factor 1 and factor 2

$$\hat{\mathbf{f}} = (\hat{\mathbf{L}}_z^{*T} \hat{\Psi}_z^{-1} \hat{\mathbf{L}}_z^*)^{-1} \hat{\mathbf{L}}_z^{*T} \hat{\Psi}_z^{-1} \mathbf{z} = \begin{bmatrix} -.61 \\ -.61 \end{bmatrix}$$