UNIVERSITY OF TORONTO MISSISSAUGA

December 2007 Examination
STA312H5S
Professor Jerry Brunner
Duration: 3 hours

Aids allowed: Calculator, formula sheet (supplied)

Students may be charged with an academic offence for possessing the following items during the writing of an exam unless otherwise specified: any unauthorized aids, including but not limited to calculators, cell phones, pagers, wristwatch computers, personal digital assistants (PDAs), IPODS, MP3 players, or any other electronic device. If any of these items are in your possession, please turn them off and put them with your belongings at the front of the room before the examination begins and no penalty will be imposed. A penalty MAY BE imposed if any of these items are kept with you during the writing of your exam.

Please note, students are **not** allowed to petition to RE-WRITE a final examination.

Please write your answers in the exam booklet.

20 points

1. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent Multivariate Normal $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors, and let $\boldsymbol{\Sigma}$ be fixed and *known*. Derive the maximum likelihood estimate of $\boldsymbol{\mu}$. "Derive" means show all the work. Where do you use the fact that $\boldsymbol{\Sigma}^{-1}$ is positive definite? Indicate this clearly.

20 points

2. The Poisson distribution has probability mass function

$$p(y) = \frac{e^{-\lambda}\lambda^y}{y!},$$

where $\lambda > 0$ and y is a non-negative integer. You may use the fact that the MLE of λ is $\overline{y}$; you don't have to prove this.

Given two independent random samples $X_1, \ldots, X_{n_1} \stackrel{i.i.d.}{\sim}$ Poisson(λ_1) and $Y_1, \ldots, Y_{n_2} \stackrel{i.i.d.}{\sim}$ Poisson(λ_2), we are interested in testing $H_0 : \lambda_1 = \lambda_2$.

(a) What is the parameter space Θ?

(b) What is the restricted parameter space Θ_0?

(c) Construct and *simplify* the large-sample likelihood ratio test statistic G. Show your work.

(d) Given $n_1 = 50$, $n_2 = 40$, $\overline{x} = 1.88$ and $\overline{x} = 2.625$, calculate the value of G.

(e) What are the degrees of freedom for this test? The answer is a number.

(f) The critical value of this test at $\alpha = 0.05$ is 3.841459. Do you reject the null hypothesis? Answer Yes or No.

continued on page 3

20 points

3. The following is a multivariate regression model with measurement error in the independent variables. Let

$$\begin{aligned} \mathbf{X}_1 &= \boldsymbol{\xi} + \boldsymbol{\delta}_1 \\ \mathbf{X}_2 &= \boldsymbol{\xi} + \boldsymbol{\delta}_2 \\ \mathbf{Y} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \end{aligned}$$

with $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ all independent, $V(\boldsymbol{\xi}) = \boldsymbol{\Phi}$, $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$, $V(\boldsymbol{\delta}_1) = \boldsymbol{\Theta}_1$, and $V(\boldsymbol{\delta}_2) = \boldsymbol{\Theta}_2$.

   (a) Give the variance-covariance matrix of the observed variables $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Y}$.

   (b) Is the model identified? Answer Yes or No and prove your answer.

15 points

4. Consider the general factor analysis model

$$\mathbf{X} = \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e},$$

where $\boldsymbol{\Lambda}$ is a p by m matrix of factor loadings, the vector of factors $\mathbf{F}$ is multivariate normal with expected value zero and covariance matrix $\mathbf{I}_m$ (the identity), and $\mathbf{e}$ is multivariate normal independent of $\mathbf{F}$, with expected value zero and covariance matrix $\boldsymbol{\Psi}$, a p by p diagonal matrix of error variances, all strictly greater than zero. *To exclude trivial cases, assume that the number of factors is at least two, the number of variables per factor is at least three, all the factor loading are non-zero, and at least one loading per factor is known to be positive.*

   (a) Calculate the matrix of covariances between the observable variables $\mathbf{X}$ and the underlying factors $\mathbf{F}$.

   (b) Give the covariance matrix of $\mathbf{X}$.

   (c) Is this model identified? Answer Yes or No and prove your answer.

continued on page 4

10 points

5. Here is a factor analysis model with three underlying factors; all the observed variables X_j are standardized; that is, they have mean zero and variance one.

$$
\begin{aligned}
X_1 &= \lambda_{11}F_1 + \lambda_{12}F_2 + \lambda_{13}F_3 + e_1 \\
X_2 &= \lambda_{21}F_1 + \lambda_{22}F_2 + \lambda_{23}F_3 + e_2 \\
X_3 &= \lambda_{31}F_1 + \lambda_{32}F_2 + \lambda_{33}F_3 + e_3 \\
X_4 &= \lambda_{41}F_1 + \lambda_{42}F_2 + \lambda_{43}F_3 + e_4 \\
X_5 &= \lambda_{51}F_1 + \lambda_{52}F_2 + \lambda_{53}F_3 + e_5 \\
X_6 &= \lambda_{61}F_1 + \lambda_{62}F_2 + \lambda_{63}F_3 + e_6 \\
X_7 &= \lambda_{71}F_1 + \lambda_{72}F_2 + \lambda_{73}F_3 + e_7 \\
X_8 &= \lambda_{81}F_1 + \lambda_{82}F_2 + \lambda_{83}F_3 + e_8
\end{aligned}
$$

where F_1, F_2, F_3 and $e_1, \ldots, e_8$, are all independent of one another with expected value zero, $V(F_1) = V(F_2) = V(F_3) = 1$, and the λ_{ij} quantities are fixed constants.

(a) What is the correlation between X_2 and F_3?

(b) What is $V(e_4)$? Don't forget that $V(X_4) = 1$.

(c) Give the communality of X_6.

(d) What is $Cov(X_1, X_2)$?

continued on page 5

15 points

6. This factor analysis model has a single underlying factor; all the observed variables are standardized.

$$
\begin{aligned}
X_1 &= \lambda_1 F + e_1 \\
X_2 &= \lambda_2 F + e_2 \\
X_3 &= \lambda_3 F + e_3,
\end{aligned}
$$

where $F \sim N(0,1)$, e_1, e_2 and e_3 are normal and independent of F and each other with expected value zero, $V(X_1) = V(X_2) = V(X_3) = 1$, and λ_1, λ_2 and λ_3 are nonzero constants; they may be either positive or negative, so this is a little different from the homework.

   (a) Give the variance-covariance (correlation) matrix of the observed variables.

   (b) Is the model identified? Answer Yes or No and prove your answer.

Total Marks = 100 points

STA 312F07 Formulas

Here are some useful formulas. You may use any them without proof, unless you are explicitly asked to derive it. You will not need them all.

- If $\mathbf{A}$ is $n \times r$ and $\mathbf{B}$ is $r \times m$, then $\mathbf{AB} = [\sum_{=1}^{r} a_{ik}b_{kj}]$.
- If $\mathbf{A}$ is $n \times r$ and $\mathbf{B}$ is $r \times n$, then $tr(\mathbf{AB}) = tr(\mathbf{BA})$.
- The $r \times r$ matrix $\mathbf{A}$ is said to be positive definite if $\mathbf{b}'\mathbf{Ab} > 0$ for every nonzero $r \times 1$ vector $\mathbf{b}$.
- The inverse of a square symmetric matrix exists if and only if it is positive definite.
- $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$
- $V(\mathbf{X}) = E\left((\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{X} - \boldsymbol{\mu}_x)'\right)$
- $C(\mathbf{X}, \mathbf{Y}) = E\Big((\mathbf{X} - \boldsymbol{\mu}_x)(\mathbf{Y} - \boldsymbol{\mu}_y)'\Big)$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, $E(\mathbf{XY}) = E(\mathbf{X})E(\mathbf{Y})$.
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, $V(\mathbf{X} + \mathbf{Y}) = V(\mathbf{X}) + V(\mathbf{Y})$.
- The multivariate normal density is $f(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(2\pi)^{\frac{p}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right]$.
- If $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}$ is a matrix of constants, $\mathbf{AX} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.
- For the multivariate normal, $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$ and $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$
- For the multivariate normal, $-2\log L(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = n\log|\widehat{\boldsymbol{\Sigma}}| + np[1 + \log(2\pi)]$.