

STA 2101/442 Assignment Four¹

Except for Questions 11 and 13, these problems are practice for the quiz, and are not to be handed in. Please do the homework using the formula sheet, which is posted on the course home page. **Please bring the R printouts for Questions 11 and 13 to the quiz.**

1. The joint moment-generating function of a p -dimensional random vector \mathbf{X} is defined as $M_{\mathbf{X}}(\mathbf{t}) = E\left(e^{\mathbf{t}^\top \mathbf{X}}\right)$. Find the moment-generating function of
 - (a) Let $\mathbf{Y} = \mathbf{A}\mathbf{X}$, where \mathbf{A} is a matrix of constants. Find the moment-generating function of \mathbf{Y} .
 - (b) Let $\mathbf{Y} = \mathbf{X} + \mathbf{c}$, where \mathbf{c} is a $p \times 1$ vector of constants. Find the moment-generating function of \mathbf{Y} .
2. Let $Z_1, \dots, Z_p \stackrel{i.i.d.}{\sim} N(0, 1)$, and

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_p \end{pmatrix}.$$

- (a) What is the joint moment-generating function of \mathbf{Z} ? Show some work.
 - (b) Let $\mathbf{Y} = \Sigma^{1/2}\mathbf{Z} + \boldsymbol{\mu}$, where Σ is a $p \times p$ symmetric *non-negative definite* matrix and $\boldsymbol{\mu} \in \mathbb{R}^p$.
 - i. What is $E(\mathbf{Y})$?
 - ii. What is the variance-covariance matrix of \mathbf{Y} ? Show some work.
 - iii. What is the moment-generating function of \mathbf{Y} ? Show your work.
3. We say the p -dimensional random vector \mathbf{Y} is multivariate normal with expected value $\boldsymbol{\mu}$ and variance-covariance matrix Σ , and write $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, when \mathbf{Y} has moment-generating function $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \Sigma \mathbf{t}}$.
 - (a) Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{W} = \mathbf{A}\mathbf{Y}$, where \mathbf{A} is an $r \times p$ matrix of constants. What is the distribution of \mathbf{W} ? Show your work.
 - (b) Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{W} = \mathbf{Y} + \mathbf{c}$, where \mathbf{c} is a $p \times 1$ vector of constants. What is the distribution of \mathbf{W} ? Show your work.
 4. Let $\mathbf{Y} \sim N_2(\boldsymbol{\mu}, \Sigma)$, with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

Using moment-generating functions, show Y_1 and Y_2 are independent.

¹Copyright information is at the end of the last page.

5. Let $\mathbf{X} = (X_1, X_2, X_3)'$ be multivariate normal with

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \\ 6 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2 + X_3$. Find the joint distribution of Y_1 and Y_2 .

6. Let X_1 be $\text{Normal}(\mu_1, \sigma_1^2)$, and X_2 be $\text{Normal}(\mu_2, \sigma_2^2)$, independent of X_1 . What is the joint distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$? What is required for Y_1 and Y_2 to be independent? Hint: Use matrices.

7. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where \mathbf{X} is an $n \times p$ matrix of known constants, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constants, and $\boldsymbol{\epsilon}$ is multivariate normal with mean zero and covariance matrix $\sigma^2 \mathbf{I}_n$, where $\sigma^2 > 0$ is a constant. In the following, it may be helpful to recall that $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$.

(a) What is the distribution of \mathbf{Y} ?

(b) The maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. What is the distribution of $\hat{\boldsymbol{\beta}}$? Show the calculations.

(c) Let $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. What is the distribution of $\hat{\mathbf{Y}}$? Show the calculations.

(d) Let the vector of residuals $\mathbf{e} = (\mathbf{Y} - \hat{\mathbf{Y}})$. What is the distribution of \mathbf{e} ? Show the calculations. Simplify both the expected value (which is zero) and the covariance matrix.

8. Show that if $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ positive definite, then $Y = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ has a chi-square distribution with p degrees of freedom.

9. Let X_1, \dots, X_n be a random sample from a $N(\mu, \sigma^2)$ distribution.

(a) Show $\text{Cov}(\bar{X}, (X_j - \bar{X})) = 0$ for $j = 1, \dots, n$.

(b) Show that \bar{X} and S^2 are independent.

10. The statistic

$$W_n = n(\mathbf{L}\mathbf{T}_n - \mathbf{h})' (\mathbf{L}\widehat{\Sigma}_n\mathbf{L}')^{-1} (\mathbf{L}\mathbf{T}_n - \mathbf{h})$$

provides a large-sample test of $H_0 : \mathbf{L}\boldsymbol{\theta} = \mathbf{h}$. Let X_1, \dots, X_n be a random sample from a $B(1, \theta)$ distribution.

- (a) Write down and simplify the W_n statistic for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.
 - (b) Your answer is related to a Z -test of this same null hypothesis. Write down the formula for the Z statistic.
11. In Assignment One, you tested difference between means for the Verbal SAT and the Math SAT score. A link to the SAT data is available from the course home page. Using R, please calculate the W_n statistic to test this hypothesis. Feel free to use my code directly. Note that the statistic \mathbf{T}_n is of dimension *two*. Guided by the usual $\alpha = 0.05$ significance level, what do you conclude? Be able to state your conclusion in plain, non-statistical language. Bring your R printout to the quiz.
12. A team of botanists grew fungus in a nutrient solution in test tubes. Each day for seven days, one of their graduate students carefully measured the length of the fungus in each of n tubes. The scientists were interested in lots of things, including whether average growth was linear or not. Denote the expected amount of fungus at day j by μ_j .
- (a) What is the null hypothesis, in symbols?
 - (b) Assuming that the scientists wish to make as few assumptions as possible and n is large, the W_n statistic is natural for this problem. What is \mathbf{T}_n ?
 - (c) What is \mathbf{L} ?
 - (d) What is \mathbf{h} ?
 - (e) What is a convenient choice for $\widehat{\Sigma}_n$? How many rows and columns?

13. In a study of the psychology of attention, subjects attempted to solve word problems while listening to distracting background noise. The distracting material was either music, or spoken words related to the problem they were trying to solve. The distracting material was presented at three different levels of loudness. Each subject attempted 10 problems at each combination of loudness and type of distraction, for a total of 60 problems. Order of presentation was randomized. Data for each subject are number correct in each of the six treatment combinations. The data are available in the file `distract.data`, available at

http://www.utstat.toronto.edu/~brunner/appliedf14/code_n_data/hw/distract.data.

There is a link from the course website under `Data sets` in case the one in this document does not work. See `help(read.table)` if necessary.

- (a) Produce a table showing the sample mean for each of the six treatment conditions.
- (b) Give a large-sample 95% confidence interval for each treatment mean.
- (c) Now test whether the six treatment means (expected values) are equal; as usual, $\alpha = 0.05$. You may use my `Wtest` function. There is a link from the course website under `Data sets` in case the one in this document does not work. Just to make sure we are doing things the same way, my test statistic value is $W_n = 757.293$. In plain, non-statistical language, what do you conclude?
- (d) Now we will compare *averages* of expected values. Those who have had a course in experimental design will recognize that we are testing differences between marginal means. Test the difference between the average expected test performance for Voice distraction and the average expected test performance for Music distraction. Be able to state a *directional* conclusion in plain, non-statistical language, if a conclusion is justified by the test.
- (e) Now just for Voice distraction, is there any effect of volume? Do the test and state a conclusion in plain language. Don't bother with follow-up tests yet; we'll do that next week.
- (f) Now just for Music distraction, is there any effect of volume? Do the test and state a conclusion in plain language. Don't bother with follow-up tests yet; we'll do that next week.

Please bring your R printout for this question to the quiz.

This assignment was prepared by [Jerry Brunner](#), Department of Statistics, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The \LaTeX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf14>