Mathematics 376 – Probability and Statistics 2
Review Sheet – Exam II
April 20, 2012

General Information

As announced in the course syllabus, the second midterm exam this semester will be given on *Thursday, April 26*, at 7:00pm in Swords 302. The exam will cover the material we have discussed since the last exam, starting from the material in Chapter 9 that we covered in class, including the material from Chapter 10 on hypothesis testing, culminating in Lab Project 4, and finishing with the material on linear models and least squares estimators from 11.1,2,3, and 11.10. There is a more detailed breakdown of topics given below.

Format

As on the last exam, you may prepare a sheet (single side of standard 8 1/2 by 11 inch paper) of formulas and any other information you want to include and consult it during the exam at any time. But you should still prepare carefully for the exam and understand the key concepts we have talked about. Know how to apply the different techniques we have studied and how to select the most appropriate method when there are several possible choices. If you need to find examples similar to the test questions to get started on a problem, the exam will take much longer to complete than is necessary.

Topics

The topics to be covered (not in chronological order, but according to logical connections):

1) The method of moments and the method of maximum likelihood for deriving estimators.
2) Consistency of estimators (know the theorem giving a sufficient condition for consistency), sufficient statistics (know the factorization criterion).
3) Hypothesis testing – the general concepts: null hypothesis, alternative hypothesis, test statistic, rejection region, Type I error probability ($= \alpha$, or level of test), Type II error probability ($= \beta$), attained significance level (p-value of a test), interpretation of results.
4) The connection between confidence intervals and rejection/"acceptance" regions for tests.
5) Large sample (Z-) tests and related confidence intervals for means and proportions (Note: some of this overlaps material from Midterm 1!). Questions here might also ask you to design tests with a given α-value to achieve a certain β-value by selecting sample size appropriately.
6) Small sample (t-) tests for means and related confidence intervals.
7) There may be a part of a question dealing with χ^2- and F-tests for variances and ratios of variances and related confidence intervals (as on Lab Project 4 and in Section 10.9).

8) Setting up the normal equations to estimate the coefficients β_i in a linear model using the matrix formulation:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

(Note: Because of the time constraints, and because we did these computations using `R` in class and on Problem Set 8, I will not ask you to actually solve for the β_i.)

Comment: As you should be able to tell, the justifications for the methods we have developed here depend heavily on the probability topics we learned last semester, as well as the sampling distribution theory (Z, χ^2, t, F distributions, etc.) and the Central Limit Theorem. If you are feeling "rusty" on any of that, start by reviewing that material.

Review Session

We will review for the exam in class on Wednesday, April 25, and I will be available for questions during regular office hours.

Suggested Review Problems

From the text:

- Chapter 9/3,15,42,53,69,77,91,93;
- Chapter 10/6,19,23,25,23,37,51,65ac,71a,79ab;
- Chapter 11/67, 69 (just set up the normal equations corresponding to the models in these using the matrix formulation).

Sample Exam

Disclaimer: A reasonable 1-2 hour exam cannot cover every topic we have discussed in this section of the course (in particular every type of estimation and hypothesis testing problem we have seen). But you should be prepared for all the possibilities. This was the exam given in the 2010 offering of MATH 376. They indicate the approximate level of difficulty and cover a possible subset of the topics on the upcoming exam; the actual exam problems may differ substantially from these and might deal with different situations.

Comments: Any general formula we have studied can be used without comment (i.e. you don't need to rederive it in you solution).

I. Let $Y_1, \ldots, Y_n$ be independent samples from the distribution with pdf containing the unknown parameter $\theta > 0$:

$$f(y|\theta) = \begin{cases} \frac{1}{\theta} y^{1/\theta - 1} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

A) Determine the method of moments estimator for θ.

B) Determine the maximum likelihood estimator for θ.

II. A shop manufactures O-rings for the Space Shuttle booster rockets for NASA. Let d be the proportion of defectives in the shop's output. A random sample of size $n = 70$ O-rings produced 6 defectives.

A) Test the hypothesis $H_0 : d = .1$ versus $H_a : d \neq .1$ using this data. Take $\alpha = .02$ (probability of Type I error). Also give the p-value of your test.

B) For the test in part A, what is β (probability of Type II error) if the true value of d is .03?

III. Consider the following measurements of the heat-producing capacity of the natural gas produced by two fields of gas wells (in calories per cubic meter):

Field 1 :	8.26	8.13	8.35	8.07	8.34	
Field 2 :	7.95	7.89	7.90	8.14	7.92	7.84

Let μ_i $(i = 1, 2)$ be the population mean heat-producing capacity of the natural gas from field i.

A) (10) ("Essay Question" – No Computations) What assumptions do you need to make in order to use the basic test of $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$? How would you decide whether those conditions were satisfied?

B) (15) Carry out the basic test from part A with $\alpha = .05$ and state your conclusion clearly and succinctly.

C) (5) Construct a two-sided 95% confidence interval for the difference of the population mean heat producing capacities $\mu_1 - \mu_2$. How is this related to your answer in B?

IV. (15) Let (x_i, y_i), $i = 1, \ldots, n$ be a collection of data points in the plane. Using the matrix formulation, derive the normal equations for the least squares estimators for the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ in the model $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$ fitting the data. (Note: You *do not need to try to solve the equations*!)

Extra Credit (10) In the situation of question I, suppose you set up a test of the hypothesis $H_0 : \theta = 1$ versus the alternative $H_a : \theta > 1$ using Y_1, Y_2 (exactly two of the samples). If you reject H_0 when $Y_1 + Y_2 > 1.8$, what is the Type I error probability, α?