Mathematics 376 – Mathematical Statistics
Review Sheet – Exam I
February 24, 2012

General Information

As announced in the course syllabus, the first midterm exam this semester will be given *Friday, March 16* (the week after Spring Break), or possibly in the evening of Thursday, March 15. The exact scheduling will be determined in class on Monday, February 27. The exam will cover the material we have discussed since the start of the semester, up to and including the material on confidence intervals for variances (this is section 7.2, section 6.7 on the order statistics, and all of Chapter 8), plus the material on Lab Project/Problem Set 4. There is a more detailed breakdown of topics given below.

Format

This will be an *closed book exam*, but you may place any information you think might be useful on one side of a standard 8.5×11 sheet of paper, bring that to the exam and consult it at any time. I will also provide copies of any tables from the text you will need to use. Even though you will have all that information to use, you should still prepare very carefully for the exam and understand the key concepts we have talked about. Know how to apply the different techniques we have studied. If you are consulting your sheet for *every* formula you need or examples, etc. you *may not be able to finish the exam in the allotted time.*

Topics

The topics to be covered are:

1) Sampling distributions related to the normal distribution (χ^2, t, F distributions) – know how to tell when a random variable has one of these distributions, how to use the tables for each in the text, etc.
2) Point estimators for distribution parameters, bias, mean square error, standard error. Be sure you understand Table 8.1 on page 397 of the text, where all the entries come from, and how they are used. Also be able to analyze estimators to determine whether they are biased or not, construct unbiased estimators, etc. (see Problem Set 2).
3) The pdf's for order statistics (especially the sample maximum and minimum), and how they can be used for estimation problems – this was covered in class; also see section 6.7 in the text.
4) The concept of a confidence interval (interval estimator), derivation of formulas for interval estimators via pivotal quantities.
5) Large-sample confidence intervals for means, differences of means, proportions, differences of proportions (constructed using pivotal quantities that are normal)

6) Small-sample confidence intervals for means, differences of means (constructed using pivotal quantities that have t-distributions)
7) Confidence intervals for variances, ratios of variances.
8) From Lab Project/Problem Set 4: Normal quantile-quantile plots as a way to test for normality.

Comment: As you can tell, much of this depends heavily on the probability topics we learned last semester. If you are feeling "rusty" on general properties of expected values or variances like the important relations

$$E(aX + bY) = aE(X) + bE(Y)$$

and

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\text{Cov}(X, Y),$$

or on computing expected values or variances of random variables with given distribution (i.e. given pdf), start by reviewing that material.

Review Session

We will review in class on Wednesday, March 14.

Suggested Review Problems

From the text:
Chapter 7/15,20,33,37,38,39,95;
Chapter 8/9,15,19,23,43,61,67,83,91,95,103

Sample Exam

Disclaimer: The following problems represent the approximate level of difficulty of the upcoming exam; the actual exam problems may differ substantially from these and these questions are somewhat longer than the actual exam will be to give an idea of the range of different sorts of things that might be asked.

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 $Z_1, Z_2, \ldots, Z_7$ be independent random samples from a standard normal distribution. Let

$$W = Z_1^2 + Z_2^2 + \cdots + Z_7^2$$

A) What is the distribution of W? Explain.
B) Find $P(1.68987 \le W \le 14.0671)$.
C) Let $U = Z_1^2 + Z_2^2 + \cdots + Z_6^2$ (not including Z_7). What is the distribution of $V = \frac{6Z_7^2}{U}$. Explain.

D) For which value of y is $P(V \geq y) = .05$?
E) How would you find the y such that $P(V \geq y) = .95$?
F) Use the Z_i to construct a random variable T that has a t-distribution with $\nu = 5$ degrees of freedom.

II. Let $Y_1, \ldots, Y_n$ (n fixed but arbitrary) be random samples from an exponential distribution with pdf

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

A) What is $\lim_{n\to\infty} P(Y_{(n)} \geq 2\theta)$? ($Y_{(n)}$ is the sample maximum, as usual.) Is your answer intuitively reasonable, given what we know about exponential distributions?
B) Let $\widehat{\theta} = Y_{(1)}$ (the sample minimum). Find a constant c so that $c\widehat{\theta}$ is an unbiased estimator for θ.

III. A farm grows grapes for jelly. The following data are measurements of grape sugar levels from 8 random samples:

$$16.0, 15.2, 12.0, 16.9, 14.4, 16.3, 15.6, 12.9.$$

A) Find point estimates μ and σ from the given information.
B) Construct a 90% confidence interval for the population mean μ from the given information. Explain how you know which method to use, and state all assumptions you are making for this method to be applicable. What method(s) could you use to test whether these assumptions are reasonable?
C) From your answer to part B, can we rule out the possibility that the value of μ is 16?
D) Construct a 95% confidence interval for the population variance σ^2.

IV. Two machine shops manufacture toggle levers (whatever those are!). A random sample of size 642 from the output of shop A produced 24 defective toggle levers. A random sample of size 500 from the output of shop B produced 22 defective toggle levers.
A) Give a point estimate of p_A, the proportion of defectives in the output of shop A.
B) Derive a 95% confidence interval for p_A.
C) Derive a 95% confidence interval the difference $p_A - p_B$ where p_B is the proportion of defectives in the output of shop B. Is there good reason to suppose the output of shop A has higher quality than the output of shop B? Explain.
D) How large a sample would be needed from the output of shop A in order that

$$P(|\widehat{p_A} - p_A| < .01) \geq .95$$

(i.e. the probability that the estimator is within .01 of the true value is at least .95).