Introduction to Probability and Statistics

Spring 2010

Review of topics to be covered on the final exam

The final exam will cover the following sections of the textbook:

- **Chapter 3**: All
- **Chapter 4**: All except 4.3.2, and 4.8
- **Chapter 5**: All except 5.1.1, 5.2.1, 5.6.1, 5.7, 5.8.3 and 5.9
- **Chapter 6**: All
- **Chapter 7**: All except 7.2.1, 7.6, 7.7 and 7.8
- **Chapter 8**: All except 8.4.4, 8.5.1, 8.6, 8.6.1, 8.7 and 8.7.1

It will be cumulative, but perhaps with a slight leaning towards the post-second-midterm material.

Here is a more detailed lists of the things that you should be able to do:

Basic Probability

You should be able to

- Set up the sample space of possible outcomes for an experiment, and assign probabilities to events
- Compute the probabilities of unions, intersections and complements of events
- Verify relations between events involving unions, intersections and complements (e.g., by using Venn diagrams)
- Identify when two events are mutually exclusive
- Compute probabilities of events in uniform probability spaces (all outcomes equally likely) using the basic principles of counting, the factorial function and binomial coefficients

- Compute conditional probabilities
- Explain what is meant by two events being independent, and identify when pairs of events are independent
- Apply Bayes' formula to compute probabilities

Random Variables

You should be able to

- Explain what a random variable is, what the cumulative distribution function, mass function and density functions are, what a discrete random variable is, and what a continuous random variable is
- Calculate the mass function of a simple discrete random variable associated with an explicitly described experiment (such as rolling a pair of dice and taking the product of the numbers that come up)
- Compute probabilities involving a random variable using its cumulative distribution function, mass function and/or density function
- Compute the distribution function, mass function and/or density function of a function of a random variable, using the distribution function, mass function and/or density function of the random variable
- Explain what the joint cumulative distribution function, the joint mass function and the joint density functions of a pair of random variables are, and what it means to say that two random variables are independent
- Identify when two random variables are independent, using their joint distribution, joint mass and/or joint density functions
- Compute probabilities involving a pair of random variables using the cumulative distribution functions, mass functions and/or density functions of the individual random variables (e.g., compute the probability that the sum of two random variables takes on a value in a certain range)
- Compute the expectation and variance of discrete and continuous random variables
- Compute the expectation and variance of linear combinations of discrete and continuous random variables
- Compute and interpret the covariance of a pair of random variables
- Use Markov's inequality to obtain information about the probability that a non-negative random variable takes a value greater than or equal to a specified value

- Use Tchebychev's inequality to obtain information about the probability that a random variable takes a value differing from the mean by more than a specified value

Special Random Variables

For each of a number of special random variables, you should

- Identify the parameters that the random variable depends on
- Identify situations where it is useful and appropriate to use that random variable
- Use the mass function/density function of the random variable to compute probabilities
- Compute the expectation and variance of the random variable

The discrete random variables:

- Bernoulli, Binomial, Poisson and Hypergeometric (for hypergeometric, expectation but not variance)

The continuous random variables:

- Uniform, Exponential, Normal and Standard Normal

You should also be able to do the following

- Compute the distribution function of the sum of independent Poisson random variables
- Identify when it is appropriate to approximate the binomial distribution by the Poisson distribution, and compute approximate binomial probabilities using this method
- Use the memoryless property of the exponential random variable
- Compute the distribution function of the minimum of independent exponential random variables
- Convert a general normal random variable into a standard normal
- Compute the distribution function of a linear combination of independent normal random variables
- Use a standard normal table to compute probabilities involving normal random variables
- Identify when it is appropriate to use a χ-squared random variable, and use a χ-squared table to compute probabilities involving a χ-squared random variable

Sampling

You should be able to

- Compute the sample mean, sample variance and sample standard deviation of a data set
- Compute the mean and variance of the sample mean, and the mean of the sample variance, in terms of the mean and variance of the population distribution
- Use the Central Limit Theorem to compute the approximate distribution of the sample mean when the sample size is large, and to compute probabilities associated with the sample mean
- Compute the exact distribution of the sample mean and sample variance when the population distribution is normal
- Give an unbiased point estimate for the population mean and population variance using sample data
- Compute the maximum likelihood estimator of a parameter of a random variable using sample data

Confidence Intervals

You should be able to construct confidence intervals, at various percentage levels, both two-sided and one-sided, in each of the following situations. You should also be able to determine how large to make the sample size to make sure that the confidence interval has a specified width.

- Mean of a normally distributed population when variance is known
- Mean of a normally distributed population when variance is unknown
- Variance of a normally distributed population
- Difference of means of two normally distributed populations when both variances are known
- Difference of means of two normally distributed populations when both variances are unknown but assumed equal (this includes computing the pooled estimator for the common variance)
- Proportion of a Bernoulli population

Hypothesis tests

You should be able to set up a hypothesis test, deciding what is the appropriate null hypothesis and what is the appropriate alternate hypothesis, and deciding whether the test should be one- or two-sided in each of the following situations.

- Mean of a normally distributed population when variance is known
- Mean of a normally distributed population when variance is unknown
- Variance of a normally distributed population
- Difference of means of two normally distributed populations when both variances are known
- Difference of means of two normally distributed populations when both variances are unknown but assumed equal
- Difference of means of two normally distributed populations when both variances are unknown (not assumed equal), with large samples (approximate)

You should also be able to carry out the test, at various significance levels, correctly interpret the results, and compute the p-value of the data, using whichever of these tables is appropriate:

- Standard Normal Table
- t-table
- χ-squared table

For general hypothesis tests, you should be able to explain what each of these terms mean:

- Significance
- Power
- Type I error
- Type II error
- p-value

For the test of the mean of a normal population, variance known, you should be able to calculate how large a sample size to take to make the power a specified value when the difference between the actual mean and hypothesisized mean is specified.