Mathematics 375 – Probability and Statistics I
Review Sheet, Final Exam
December 7, 2009

General Information

The final examination for this course will be given at 2:30 pm on Monday, December 14 in our regular class room, Stein Hall 316. The exam will be roughly twice the length of one of the two midterms, but you will have the full three hour period from 2:30 pm to 5:30 pm to work on it if you need that much time. As was true on the midterms, I will let you bring a sheet of information for your use on the exam – for this exam, you can place anything you want on *one side of an 8.5 × 11 inch piece of paper.* I will provide copies of whatever tables from the text might be needed to work some of the problems.

Topics to be Covered

1) Descriptive statistics such as the mean, standard deviation, frequency histograms, etc. The "empirical rule".
2) Discrete sample spaces and counting techniques (especially in connection with the "sample point method" for the probability of an event): the $m \times n$ rule, permutations, binomial and multinomial coefficients.
3) The "event composition method" for probabilities
4) Conditional probabilities, independence of events, Bayes' Rule, the Law of Total Probability
5) Discrete random variables: probability distribution functions, cumulative distribution functions, expected values and variances of functions of a discrete random variable, moment generating functions. Know the situations leading to binomial, geometric, Poisson, and hypergeometric random variables and how to apply them.
6) Continuous random variables: probability distribution functions, cumulative distribution functions, expected values and variances of functions of a continuous random variable, moment generating functions. Know the situations leading to uniform, exponential, gamma, beta, and normal random variables and how to apply them.
7) Tchebysheff's Theorem.
8) Multivariate probability distributions: joint densities, marginal and conditional densities, expected values in this setting, conditions for independence, the covariance and the general formula for the variance of a linear combination of random variables.
9) Using moment generating functions and the uniqueness theorem to determine the distribution of a random variable.
10) The method of distribution functions to determine the distribution of a random variable.
11) Know the statement of the Central Limit Theorem.

Suggestions on How to Study

Start by reading the above list of topics carefully. If there are terms there that are unfamiliar or for which you cannot give the precise definition, start by reviewing those topics. Review the class notes. *Everything on the final will be closely related to something we have discussed at some point this semester.* Also look back over your graded problem sets and exams. If there are problems that you did not get the first time around, try them again now, consulting the solutions on reserve in the Science Library as necessary. Then go through the suggested problems from the review sheets. If you have worked these out previously, it is not necessary to do them all again. But try a representative sample "from scratch" – don't just look over your old solutions. Practice thinking through the logic of how the solution is derived again.

Suggested Practice/Review Problems

Look at the problems from the two previous review sheets for the topics from Chapters 1 - 5. From Chapter 5/105 (do all the computations needed to find $V(Y_1 - Y_2)$ using the formula of Theorem 5.12), 147,149,151,161. Chapter 6/93,95,107,111.

Review Session

I will be happy to run a review session for the final exam during study week. We can discuss a time in class on Monday, December 7.

Sample Exam

I. A large number of observations of a certain continuous random variable Y were taken, and a typical subset of $n = 50$ of them were ordered to produce the data set below:

0.10	0.23	0.25	0.26	0.26	0.28	0.31	0.39	0.42	0.44
0.45	0.48	0.53	0.54	0.55	0.66	0.69	0.70	0.72	0.73
0.73	0.74	0.74	0.75	0.77	0.78	0.81	0.95	0.97	1.02
1.03	1.05	1.07	1.13	1.13	1.14	1.20	1.22	1.27	1.28
1.30	1.34	1.34	1.37	1.39	1.40	1.43	1.43	1.46	1.48

A) Construct relative frequency histograms for this data using 5 bins on the interval $[0, 1.5]$, and then 10 bins on the interval $[0, 1.5]$.

B) Two models are proposed for the distribution of Y:
   1. a *uniform* distribution on $[0, 1.5]$, or
   2. a distribution described by the probability density function (pdf)

$$f(y) = \begin{cases} \frac{8y}{9} & \text{if } 0 \leq y \leq 3/2 \\ 0 & \text{otherwise} \end{cases}$$

   Find the expected values and variances of the two proposed model distributions. Does either one of these look like a clearly better match for the data?

C) The actual sample mean is $\overline{y} = .8542$ and the sample standard deviation is $s \doteq .4033$. How well does the "empirical rule" match the distribution of Y?

II. A bin contains three components from supplier A, four from supplier B, and five from supplier C. If four of the components are selected randomly, without replacement, for testing, what is the probability that each supplier will have at least one component tested?

III. Carrot seeds from grower A have an 85% germination rate, while seeds from grower B have a 75% germination and those from grower C have a 90% germination rate. rate. A seed packaging company buys 30% of its seeds from grower A, 40% from grower B, and 30% from grower C, then mixes them thoroughly in making up packets for sale.

A) What is the probability that a randomly selected seed will germinate?
B) Given that a seed did not germinate, what is the probability that it came from grower B?

IV. A permutation of $A = \{1, 2, 3, 4, 5\}$ is a one-to-one, onto mapping from this set to itself. For instance, the f defined by

$$f(1) = 2, f(2) = 4, f(3) = 3, f(4) = 1, f(5) = 5$$

is one such permutation. Consider a permutation selected at random, and let Y be the *number of elements of the set A that are mapped to themselves.* For instance, if the f above was selected, then the value of Y is 2 (since $f(3) = 3$ and $f(5) = 5$). For this problem, you are given the information that the moment-generating function of Y is

$$m(t) = \frac{44}{120} + \frac{45}{120}e^t + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}$$

A) Find the mean and variance of Y.
B) What is the probability that a randomly selected permutation maps *at least one element of A to itself*?

V. A candy maker produces thin chocolate mints (yum!) that have a label weight of 20.4 grams. The manufacturing process is subject to some randomness, though, and the actual weights of the mints are normally distributed with mean $\mu = 20.8$ and $\sigma = 0.3$.

A) What is the probability that a single mint has weight < 20.4 grams?
B) If mints are selected independently and at random from the production line, what is the probability that at least 10 trials will be necessary to find one that has weight < 20.4 grams?
C) If 40 mints are selected independently and at random from the production line, what is the probability that 3 or fewer will have weight < 20.4 grams?

VI. 900 students are given an exam. The average score is $\overline{y} = 83$ and the sample variance is $v = 36$. At least how many students must have scored between 71 and 95?

VII. Telephone calls arrive at a switchboard.

A) The number of calls within a 10 minute interval has a Poisson distribution with mean $\lambda = 5$. What is the probability that more than 10 calls arrive in a 10 minute interval?

Starting from any instant, the elapsed time Y until the 5th call arrives is a random variable with a pdf of the form

$$f(y) = \begin{cases} ky^4 e^{-2y} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

B) Determine the value of k.
C) Find $E(Y)$ and $V(Y)$.
D) The elapsed time Y until the 5th call and the elapsed time Z between the 5th and 10th calls are independent, and both have the same distribution. Find the distribution of the random variable $Y + Z$ by using the appropriate moment-generating functions.

VIII. Let Y_1 and Y_2 be random variables with joint pdf

$$f(y_1, y_2) = \begin{cases} 3y_1^2 y_2 & \text{if } -1 \leq y_1 \leq 1, 0 \leq y_2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

A) What is $P(Y_2 > Y_1)$? (*Hint:* Draw a picture!)
B) Are Y_1 and Y_2 independent? Why or why not?
C) Determine $V(3Y_1 + 2Y_2)$.
D) Determine the p.d.f. for $U = Y_1 + Y_2$ using the method of distribution functions.

IX. Let Y be a continuous random variable whose cumulative distribution function $F(y)$ satisfies $F(y) = 0$ for all $y \leq 0$. Assume that Y has the "memoryless" property:

$$P(Y > a + b | Y > a) = P(Y > b)$$

for all $a, b > 0$. Show that $g(y) = 1 - F(y)$ satisfies

$$g(y + y') = g(y)g(y')$$

for all $y, y' > 0$. What type of distribution does Y have? Prove your assertion.