Sec. 0103 Fall 2003

Sample Problems for Stat 400 Final Exam

Instructions. The exam itself will contain 8 problems, or 9 problems from which you are to choose 8. The following problems are typical, in coverage and level, but not exhaustive of all types of problems that might be asked. You will also be given any tables which are needed to do the problems on the exam.

#1: *Combinatorial probability basics.* You are dealt a 2-card hand from a well-shuffled deck of 52.

(a) What is the probability of the union of the event that both cards are picture cards (J, Q, or K) with the event that both cards are black (clubs or spades)?

(b) What is the conditional probability that the second card is an Ace given that at least one of the two cards is an Ace?

#2: *Moments and distributional facts about sums of r.v.'s.* Suppose that X, Y are independent random variables with $\mu_X = 1, \sigma^2_X = 2, \mu_Y = 4, \sigma^2_Y = 3$. (a) Find the mean and variance of $5X - 4Y$. (b) If X, Y were each normally distributed in part (a), then find $P(5X - 4Y \geq 1)$.

#3: *Transformation & simulation of continuous random variables.* Suppose that the r.v. X is defined as a function $X = \sqrt{U}$ of a *Uniform*$(1, 2)$ random variable.

(a). Find the density of the random variable X.

(b). Find the expectation of X^4.

(c). Suppose that a sequence of 1000 independent random variables $X_i = \sqrt{U_i}$, $i = 1, \ldots, 1000$, were simulated in this way from *iid Uniform*$(1, 2)$ r.v.'s U_i. Approximately what fraction of these 1000 r.v.'s X_i would have values ≤ 2 ?

#4: *Bayes' rule, urn problems.* Consider the following two-stage experiment. Suppose that I have two decks of cards, the first a well-shuffled ordinary deck of 52 card and the second a well-shuffled deck of 24 containing only the cards of each suit for the values 9, 10, J, Q, K, A. Suppose that I first flip a fair coin, and pick up the 52-card deck if the coin falls Heads, and the 24-card deck otherwise.

(a) Find the probability that of three cards dealt from the top of the chosen deck, all are Pictures (J,Q, or K).

(b) If the three cards dealt from the top of the chosen deck are pictures, then what is the probability that the coin used to choose ther deck had come up Heads ?

#5: *Confidence intervals & hypothesis tests.* Based upon values $\overline{X} = 25.4$, $s^2 = 6.25$ calculated from a random sample of 150 data-values,

(a) give a 90% two-sided confidence interval for μ, and

(b) test the hypothesis at significance level $\alpha = 0.01$ that $\mu = E(\overline{X}) = 26$ versus the alternative $H_A : \mu < 26$.

#6: *Approximation of probability via CLT* Suppose we collect independent *Binomial*$(3, p)$ observations $X_1, \ldots, X_n$ (each X_i can be 0, 1, 2, or 3). Approximately how large must n be before we can say with probability at least 0.9 that $|\overline{X} - 3p| \le 0.1$?

#7: *Mechanics with densities, expectations.* If the continuous random variable Y has density $f_Y(t) = \frac{2}{27}, t(t+1)$ for $0 \le t \le 3$ (and $= 0$ otherwise), then find the distribution function and median value for Y, and find the expected value for $1/Y$.

#8: *Method of moments.* Suppose that $W_1, \ldots W_{100}$ are a random sample of discrete random variables with $P(W = 0) = (1-a)/4$, $P(W = 1) = (1+2a)/2$, and $P(W = 2) = (1-3a)/4$, where $0 < a < \frac{1}{3}$ is an unknown parameter. Find the method of moments estimator of a in terms of $W_1, \ldots, W_{100}$. What is its (approximate) probability (*i.e., sampling*) distribution ?

#9: *Distribution of Event Counts, and Approximate Distributions.* 1000 values of independent *Exponential(2)* random variables W_j, $j = 1, \ldots, 1000$ (with density $2e^{-2x}$, $x > 0$) are observed.

(a) Let N be the number among the variables W_j which are larger than $\ln(20)$. What is the exact probability distribution of N ?

(b) What is the approximate probability that N is 3 or less ?

(c) Find the approximate probability distribution of the number M of the r.v.'s W_j, $1 \le j \le 1000$, which are larger than $\ln(2)$, and use it to approximate $P(M > 230)$.

#10. *Confidence interval & sample-size, in sampling setting.* A warehouse-lot of 20000 machined parts is to be inspected for defects. A random sample of the parts is to be taken without replacement and inspected, and the proportion $\hat{p}$ of defectives in the sample will be used to decide whether the parts are being manufactured to specifications of at most 5% defective.

(a) If it is believed that in reality only 1000 parts out of the 20000 are defective, then how many of the parts must be sampled for inspection if you want the width of the resulting 95% confidence interval to be ≤ 0.02 ?

(b) Suppose that you sample $n = 500$ of the parts, and observe 30 to be defective. Give a 95% two-sided confidence interval for the true proportion of defective parts (among the whole 20000).

#11. *CI, hypothesis test, in normal-data context.* Suppose that you observe the scores of 10 students on a standardized-test on a scale from 200 to 800, and that you know from past experience that the actual distribution of individuals' test-scores are $\mathcal{N}(\mu, \sigma^2)$ but that different groups of students may have different μ, σ^2 parameter values. Suppose that the true average μ for this group of 10 students is to be compared with the national average parameter $\mu_0 = 500$, and that these students' data yield $\overline{X} = 560$, $s = 70$

(a) Give a 90% 2-sided confidence interval for the true μ for these students.

(b) Test at significance level $\alpha = 0.05$ the hypothesis that the average for these students was the same as the national average, against the alternative that the true mean is actually higher for these students than the national average. Report your rejection region and also your conclusion (whether to accept or reject the null.)

(c) Give as accurately as you can the borderline value of α at which the hypothesis test as in (b) would be able to reject, for the given data.

#12. *Scaled relative frequency histogram.* Suppose that you have tabulated a sample of 2000 observations from a $\mathcal{N}(100,\, 10^2)$ distribution into 8 class intervals, as follows.

Interval	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140
Count	2	38	277	696	696	252	33	6

(a) Calculate the bar-heights for a scaled relative frequency histogram for these data.

(b) Explain what the area in the bar over the interval $110 - 120$ represents, and give the theoretical value (from $\mathcal{N}(100,\, 10^2)$ distribution) which it estimates.