Math 481/581 Mathematical Statistics I Final Exam Fall 2007

name ____________________

Take Home Exam. Due: Thursday, Dec. 13, 4:30, pm.

1. Let X have the probability density $f(x) = (1/2)e^{-x/2}$, $x > 0$, and $f(x) = 0$, $x \leq 0$.
(a) Show by direct calculation that the moment generating function of X is

$$M(t) = (1 - 2t)^{-1}, \quad t < 1/2.$$

(b) The moment generating function of a Chi-square random variable U_1 with 1 degree of freedom is known to be

$$M_1(t) = (1 - 2t)^{-1/2}, \quad t < 1/2.$$

Note that $M(t) = M_1(t)^2$. Explain why X itself is a Chi-square random variable and identify its degrees of freedom.

2. (a) Let U_1 be a Chi-square random variable with 1 degree of freedom. Conclude by the moment generating function of problem 1(b) that the variance of U_1 is 2.
(b) Let U_{36} be a Chi-square random variable with 36 degrees of freedom. Find $E(U_{36})$ and $Var(U_{36})$.
(c) Use the Central Limit Theorem to approximate the following probability: $\mathrm{P}(U_{36} > 48)$. Why should the Central Limit Theorem apply?

3. Four balls are placed in an urn. Each ball has a single number written on it. The numbers of the balls are respectively 1, 2, 4, and 5. Two balls are selected at random (without replacement) from the urn. Call the selected values X_1 and X_2.
(a) Find the probability mass function of the sample mean $\overline{X} := (X_1 + X_2)/2$.
(b) Calculate the mean and variance of $\overline{X}$ directly from the probability mass function of part (a).
(c) Let μ and σ^2 denote the population mean variance respectively of the original population of four values. Show that the answers to part (b) satisfy $E(\overline{X}) = \mu$ and $Var(\overline{X}) = \sigma^2/3$. Would these answers be the same if the sampling had been done *with* replacement? Why or why not?

4. A certain housing development contains 12,000 condominium units. Suppose the actual fraction p of all condominium units that are planned to be sold within the coming year is unknown. A simple random sample of 400 of these units reveals that 40 of the 400 are planned to be sold.

(a) Find a 95% confidence interval for p.

(b) Assume that $p \leq .20$. How large must the sample size be to estimate p within .02 with 90% confidence? First ignore the finite population correction in your calculation. Second, include this correction in a re-calculation of the sample size.

5. A population consists of $N = 4$ pairs (x, y) as follows: (1,1), (2,1), (4,2), (5,4).

(a) Show that the population parameters for means, variances and the covariance are as follows: $\mu_x = 3$, $\mu_y = 2$, $\sigma_x^2 = 5/2$, $\sigma_y^2 = 3/2$, and $\sigma_{xy} = 7/4$.

(b) We take a simple random sample of the pairs of size 2. Show all 4 choose 2 (that is six) possible random samples explicitly. Then compute the ratio estimator of μ_y, namely, $\overline{Y}_R := \mu_x \overline{Y}/\overline{X}$, for each of these six possible samples.

(c) Find the actual bias and variance of $\overline{Y}_R$ (these are different from the theoretical approximate bias and approximate variance). Then compare the MSE (mean squared error) of $\overline{Y}_R$ with the MSE of $\overline{Y}$ for estimating μ_y.

6. We are interested in the average value μ of an inventory of 1,000 items. We know that 75% of these items have a "low" value, while the other 25% have a "high" value, and thus we have two strata. We take a simple random sample of size n_1 from the stratum of low values, and independently a sample of size n_2 from the stratum of high values. Denote the subpopulation means of the inventory values by μ_1 and μ_2, respectively. Assume that the corresponding subpopulation standard deviations satisfy $\sigma_1 = \sigma$ and $\sigma_2 = (1.5)\sigma$ for some $\sigma > 0$. Define the stratified estimator of the overall popoulation mean μ by

$$\overline{X}_s := .75\overline{X}_1 + .25\overline{X}_2$$

where $\overline{X}_1$ and $\overline{X}_2$ are the sample means from low and high values respectively.
(a) Show that the expected value of $\overline{X}_s$ is the overall population mean μ, independent of the choices of n_1 and n_2.
(b) Calculate the variance of $\overline{X}_s$ in terms of σ in each of the following cases.
(i) $n_1 = n_2 = 60$.
(ii) $n_1 = 90$ and $n_2 = 30$.
(c) Find the optimal allocation of sample sizes for a total sample size of $n = n_1 + n_2 = 120$ (ignore the f.p.c. to find the allocation), and then compute the variance of $\overline{X}_s$ in this case as well.
(d) Suppose instead it is known that $\sigma_1 = \sigma_2$. Ignoring f.p.c., what is the optimal allocation? What is the name of this type of allocation?

7. (extra credit.) Assume the context of problem 6 verbatim, except do NOT assume a total sample size of $n = 120$. Instead we assume that the cost of auditing the inventory values depends on the stratum, such that the cost of determining the value of a single "low" valued item is $C_1 = 1$, while the cost of determining the value of a single "high" valued item is $C_2 = 4$. The total cost of auditing is then assumed to be $C = C_1 n_1 + C_2 n_2 = n_1 + 4n_2$. We assume further that the total cost allowed is $C = 320$. Determine the optimal allocation of sample sizes subject to this cost constraint to minimize the variance of $\overline{X}_s$. Ignore the f.p.c. in your calculations.