Math 481 Review Problems

Be sure to justify your arguments: point out, for example, where you use linearity of E or independence, show clearly why any cancellations occur, etc.

1. Let $X_1, \dots X_n$ be a random sample from an infinite population.

(a) Derive the formula

$$\sum_{i-1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2.$$

(b) Use the formula in (a) to show that S^2 is an unbiased estimator of the population variance σ^2.

2. (a) Suppose that $X_1, \dots, X_n$ are independent random variables having chi-square distribution with 1 degree of freedom. Show that $X_1 + \cdots + X_n$ has chi-square distribution with n degrees of freedom.

(b) Suppose that Y is a random variable having gamma distribution with parameters α and β, and that c is a positive number. Identify the distribution of cY as some known distribution.

3. A certain machine can fill soft drink cans with a (programmable) mean fill of μ fluid ounces; the amount put in each can is an independent random variable with standard deviation 0.1 fluid ounce.

(a) The company needs to fill a large order, which requires that with probability 95% the cans contain at least 12 fluid ounces of soft drink. Product engineer A, who likes Chebyshev's inequality, uses it to calculate an appropriate value μ_A to which the machine should be set to accomplish this; product engineer B assumes that the distribution of the amount of soft drink in a can is a normal random variable, and on that basis calculates a value μ_B. Which number is larger? No calculation should be needed!

(b) Check your answer to (a) by calculating μ_A and μ_B.

(c) On a different occasion the company needs to fill an order for 10,000 cans of soft drink, and sets $\mu = 12.1$ fluid ounces. Engineers A and B each calculate how much bulk soft drink must be available so that the probability that the order can be filled is at least 99%; B proceeds as above, while A uses the central limit theorem. What are their answers?

4. Nine independent observations of a normal population are to be used to estimate the mean and variance of this population, using $\bar{X}$ and S^2.

(a) Suppose that the variance of the population is known to be equal to 15. What is the probability that the estimate of the mean will be correct to within 3 units?

(b) Suppose now that the variance σ^2 of the population is not known. Find a number a such that there is at most a 1% chance that the estimate of the variance will be greater than $a\sigma^2$.

(c) Suppose again that the variance is not known, but that $S^2 = 15$. Find a number b such that there is at most a 5% chance that the estimate of the mean will be wrong by more than b units.

5. A random sample of size 4 is drawn from a population uniformly distributed on the interval $[0, \beta]$. Let Y_3 denote the second largest sample value.

(a) Find the expected value of Y_3.

(b) Find a number c such that cY_3 is an unbiased estimator of β.

(c) Using the notation of (b), find the efficiency of cY_3 relative to the standard unbiased estimator $5Y_4/4$.

6. A statistician plans to draw a random sample of size n from a normal population; she wants to be sure that there will be at most a 1% chance that the magnitude of the difference between the sample mean and the true mean will be more than $3S$, where S^2 is the sample variance. How large must n be?

7. A random sample of size n is drawn from a population uniformly distributed on the interval $[0, \beta]$. Let Y_{n-1} denote the second largest sample value. Show that, as an estimator of β, Y_{n-1} is

(a) Biased but asymptotically unbiased;

(b) Consistent.

8. A scientific programmer has available a "standard normal" random number generator: a call to this program returns a (pseudo)-random number Z having standard normal distribution:

$$P(Z \leq z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-x^2/2}\, dx.$$

How can he use this to write a program which generates random numbers having chi-square distribution with 8 degrees of freedom?

9. Use the central limit theorem to find an approximate value of $\chi_{\alpha,\nu}$ for $\nu = 50$ and for $\alpha = 0.01$, $\alpha = 0.99$.

10. Discuss whether each of the equations below is true in general, under some conditions on the random variables X_i, or never. What has either equation to do with *linearity* of E?

$$E\left(\sum_{i-1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i); \qquad \operatorname{Var}\left(\sum_{i-1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i).$$

11. What is a *sufficient* statistic? Give an example, with a proof.