CIMAT

Inferencia Estadística

Maestría en Análisis Estadístico y Computación

Tarea 3

INEGI

Fecha de entrega: Jueves 27 de febrero, 2020

Problemas a entregar: 3, 4, 8, 10, 12, 13, 16, 18.

1. **Wasserman, Cap 3.** Show that if $\mathbb{E}(X|Y=y)=c$ for some constant c, then X and Y are uncorrelated.

2. **Wasserman, Cap 3.** (Computer Experiment: Simulating the Stock Market.) Let $Y_1, Y_2, \ldots$ be independent random variables such that $P(Y_i = 1) = P(Y_i = -1) = 1/2$. Let $X_n = \sum_{i=1}^{n} Y_i$. Think of $Y_i = 1$ as "the stock price increased by one dollar" , $Y_i = -1$ as "the stock price decreased by one dollar", and X_n as the value of the stock on day n.

   a. Find $\mathbb{E}(X_n)$ and $\mathbb{V}(X_n)$

   b. Simulate X_n and plot X_n versus n for $n = 1, 2, \ldots, 10,000$. Repeat the whole simulation several times. Notice two things. First, it's easy to "see" patterns in the sequence even though it is random. Second, you will find that the four runs look very different even though they were generated the same way. How do the calculations in (a) explain the second observation?

3. **Wasserman, Cap 3.** Let $X_1, \ldots, X_n \sim \text{Exp}(\beta)$. Find the moment generating function of X_i Prove that $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \beta)$

4. **Wasserman, Cap 3.** *Let $X_1, \ldots, X_n \sim$ Uniform $(0,1)$ and let $Y_n = \max\{X_1, \ldots, X_n\}$. Find $\mathbb{E}(Y_n)$

5. **Wasserman, Cap 3.** *Suppose we generate a random variable X in the following way. First we flip a fair coin. If the coin is heads, take X to have a Unif$(0,1)$ distribution. If the coin is tails, take X to have a Unif$(3,4)$ distribution. (a) Find the mean of X. (b) Find the standard deviation of X.

6. **Dekking, Cap 9** X and Y be two independent Bernoulli (1/2) random variables. Define random variables U and V by:
$$U = X + Y \quad \text{and} \quad V = |X - Y|$$

   a. Determine the joint and marginal probability distributions of U and V

   b. Find out whether U and V are dependent or independent.

7. **Dekking, Cap 9** Het X and Y be independent random variables with probability distributions given by
$$\mathrm{P}(X=0) = \mathrm{P}(X=1) = \frac{1}{2} \quad \text{and} \quad \mathrm{P}(Y=0) = \mathrm{P}(Y=2) = \frac{1}{2}$$
   Compute the distribution of $Z = X + Y$.

8. **Dekking, Cap 9** The joint probability density function f of the pair (X, Y) is given by
$$f(x,y) = K\left(3x^2 + 8xy\right) \quad \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 2$$
   and $f(x,y) = 0$ for all other values of x and y. Here K is some positive constant.

   a. Find K

b. Determine the probability $P(2X \leq Y)$

9. **Dekking, Cap 10** Let X and Y be two random variables and let r, s, t, and u be arbitrary real numbers.

   a. Derive from the definition that $\operatorname{Cov}(X+s, Y+u) = \operatorname{Cov}(X, Y)$

   b. Derive from the definition that $\operatorname{Cov}(rX, tY) = rt \operatorname{Cov}(X, Y)$

   c. Combine parts a and b to show $\operatorname{Cov}(rX+s, tY+u) = rt \operatorname{Cov}(X, Y)$

10. **Dekking, Cap 14** Let $X_1, X_2, \ldots, X_{144}$ be independent identically distributed random variables, each with expected value $\mu = \mathrm{E}[X_i] = 2$, and variance $\sigma^2 = \operatorname{Var}(X_i) = 4$. Approximate

$$\mathrm{P}\left(X_1 + X_2 + \cdots + X_{144} > 144\right),$$

   using the central limit theorem for these two different cases

   a. $X_1, \ldots, X_{144}$ are continous random variables

   b. $X_1, \ldots, X_{144}$ are discrete random variables. Use the continuity correction.

11. **Dekking, Cap 14** Let $X_1, X_2, \ldots$ be a sequence of independent $N(0,1)$ distributed random variables. For $n = 1, 2, \ldots$, let Y_n be the random variable, defined by

$$Y_n = X_1^2 + \cdots + X_n^2$$

   a. Show that $\mathrm{E}\left[X_i^2\right] = 1$

   b. One can show-using integration by parts-that $\mathrm{E}\left[X_i^4\right] = 3$. Deduce from this that $\operatorname{Var}\left(X_i^2\right) = 2$

   c. Use the central limit theorem to approximate $\mathrm{P}\left(Y_{100} > 110\right)$

12. **Dekking, Cap 19** Suppose our dataset is a realization of a random sample $X_1, X_2, \ldots, X_n$ from a uniform distribution on the interval $[-\theta, \theta]$, where θ is unknown.

   a. Show that

$$T = \frac{3}{n}\left(X_1^2 + X_2^2 + \cdots + X_n^2\right)$$

   is an unbiased estimator for θ^2

   b. Is $\sqrt{T}$ also an unbiased estimator for θ? If not, argue whether it has positive or negative bias.

13. **Dekking, Cap 19** Suppose a dataset is modeled as a realization of a random sample $X_1, X_2, \ldots, X_n$ from an $\operatorname{Exp}(\lambda)$ distribution, where $\lambda > 0$ is unknown. Let μ denote the corresponding expectation and let M_n denote the minimum of $X_1, X_2, \ldots, X_n$.

   a. Show that M_n has an $\operatorname{Exp}(n\lambda)$ distribution.

   b. Find out for which constant c the estimator

$$T = cM_n$$

   is an unbiased estimator for μ

14. **Dekking, Cap 20** Given is a random sample $X_1, X_2, \ldots, X_n$ from a distribution with finite variance σ^2. We estimate the expectation of the distribution with the sample mean $\bar{X}_n$. Argue that the larger our sample, the more efficient our estimator. What is the relative efficiency $\operatorname{Var}\left(\bar{X}_n\right) / \operatorname{Var}\left(\bar{X}_{2n}\right)$ of $\bar{X}_{2n}$ with respect to $\bar{X}_n$?

15. **Dekking, Cap 20** Given are two estimators S and T for a parameter θ. Furthermore it is known that $\operatorname{Var}(S)=40$ and $\operatorname{Var}(T)=4$ a. Suppose that we know that $\mathrm{E}[S]=\theta$ and $\mathrm{E}[T]=\theta+3$. Which estimator would you prefer, and why? b. Suppose that we know that $\mathrm{E}[S]=\theta$ and $\mathrm{E}[T]=\theta+a$ for some positive number a. For each a, which estimator would you prefer, and why?

16. **Dekking, Cap 20** Let $\bar{X}_n$ and $\bar{Y}_m$ be the sample means of two independent random samples of size n (resp. m) from the same distribution with mean μ. We combine these two estimators to a new estimator T by putting
$$T=r\bar{X}_n+(1-r)\bar{Y}_m$$
where r is some number between 0 and 1

    a. Show that T is an unbiased estimator for the mean μ

    b. Show that T is most efficient when $r=n/(n+m)$

17. **Dekking, Cap 20** Given is a random sample $X_1, X_2, \ldots, X_n$ from a $\operatorname{Ber}(p)$ distribution. One considers the estimators
$$T_1=\frac{1}{n}\left(X_1+\cdots+X_n\right) \quad \text{and} \quad T_2=\text{mín}\left\{X_1, \ldots, X_n\right\}$$

    a. Are T_1 and T_2 unbiased estimators for p?

    b. Show that
$$\operatorname{MSE}\left(T_1\right)=\frac{1}{n} p(1-p), \quad \operatorname{MSE}\left(T_2\right)=p^n-2 p^{n+1}+p^2$$

    c. Which estimator is more efficient when $n=2$?

18. **Dekking, Cap 21** Consider the following situation. Suppose we have two fair dice, D_1 with 5 red sides and 1 white side and D_2 with 1 red side and 5 white sides. We pick one of the dice randomly, and throw it repeatedly until red comes up for the first time. With the same die this experiment is repeated two more times. Suppose the following happens:

    - First experiment: first red appears in 3rd throw
    - Second experiment: first red appears in 5th throw
    - Third experiment: first red appears in 4th throw.

    Show that for die D_1 this happens with probability $5.7424 \cdot 10^{-8}$, and for die D_2 the probability with which this happens is $8.9725 \cdot 10^{-4}$. Given these probabilities, which die do you think we picked?