I3–TD1
(Point Estimation)

1. Data on pull-off force (pounds) for connectors used in an automobile engine application are as follows

79.3	75.1	78.2	74.1	73.9	75.0	77.6	77.3	73.8	74.6	75.5	74.0	74.7
75.9	72.9	73.8	74.2	78.1	75.4	76.3	75.3	76.2	74.9	78.0	75.1	76.8

   (a) Calculate a point estimate of the mean pull-off force of all connectors in the population. State which estimator you used and why.

   (b) Calculate a point estimate of the pull-off force value that separates the weakest 50% of the connectors in the population from the strongest 50%.

   (c) Calculate point estimates of the population variance and the population standard deviation.

   (d) Calculate the standard error of the point estimate found in part (a). Interpret the standard error.

   (e) Calculate a point estimate of the proportion of all connectors in the population whose pull-off force is less than 73 pounds

2. (a) A random sample of 10 houses in a particular area, each of which is heated with natural gas, is selected and the amount of gas (therms) used during the month of January is determined for each house. The resulting observations are

   103 156 118 89 125 147 122 109 138 99

   Let μ denote the average gas usage during January by all houses in this area. Compute a point estimate of μ

   (b) Suppose there are 10,000 houses in this area that use natural gas for heating. Let t denote the total amount of gas used by all of these houses during January. Estimate t using the data of part (a). What estimator did you use in computing your estimate?

   (c) Use the data in part (a) to estimate p, the proportion of all houses that used at least 100 therms.

   (d) Give a point estimate of the population median usage (the middle value in the population of all houses) based on the sample of part (a). What estimator did you use?

3. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution having finite variance σ^2. Show that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

   is an unbiased estimator of σ^2. Hint: Write

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

   and compute $E(S^2)$.

4. Suppose that X is the number of observed "successes" in a sample of n observations where p is the probability of success on each observation.

   (a) Show that $\hat{p} = \frac{X}{n}$ is an unbiased estimator of p.

   (b) Show that the standard error of $\hat{p}$ is $\sqrt{p(1-p)/n}$. How would you estimate the standard error?

5. Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from a distribution with mean μ and variance σ^2 and let $a_1, \ldots, a_n$ be real numbers such that $\sum_{i=1}^{n} a_i = 1$.

   Define $\hat{X} = \sum_{i=1}^{n} a_i X_i$.

   (a) Show that $\bar{X}$ is an unbiased estimator of μ.

   (b) Show that $V(\bar{X}) \leq V(\hat{X})$ (hence among all estimators of μ of the form $\sum_{i=1}^{n} a_i X_i$, $\bar{X}$ is the MVUE).

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with unknown mean $-\infty < \mu < +\infty$, and unknown variance $\sigma^2 > 0$. Show that the statistic $\bar{X}$ and $Y = \dfrac{X_1 + 2X_2 + \ldots + nX_n}{\frac{n(n+1)}{2}}$ are both unbiased estimators of μ. Further, show that $V\left(\bar{X}\right) < V\left(Y\right)$.

7. Using a long rod that has length μ, you are going to lay out a square plot in which the length of each side is μ. Thus the area of the plot will be μ^2. However, you do not know the value of μ, so you decide to make n independent measurements $X_1, X_2, ..., X_n$ of the length. Assume that each X_i has mean μ(unbiased measurements) and variance σ^2.

   (a) Show that $\bar{X}^2$ is not an unbiased estimator for μ^2. [Hint: For any rv Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Apply this with $Y = \bar{X}$.]

   (b) For what value of k is the estimator $\bar{X}^2 - kS^2$ unbiased for μ^2? [Hint: Compute $E(\bar{X}^2 - kS^2)$.]

8. Let $X_1, X_2, \ldots, X_n$ be uniformly distributed on the interval $[0, \theta]$. Recall that the maximum likelihood estimator of θ is $\hat{\theta} = max(X_i)$.

   (a) Argue intuitively why $\hat{\theta}$ cannot be an unbiased estimator for θ.

   (b) Suppose that $E(\hat{\theta}) = n\theta/(n+1)$. Is it reasonable that $\hat{\theta}$ consistently underestimates θ? Show that the bias in the estimator approaches zero as n gets large.

   (c) Propose an unbiased estimator for θ.

   (d) Let $Y = max(X_i)$. Use the fact that $Y \leq y$ if and only if each $X_i \leq y$ to derive the cumulative distribution function of Y. Then show that the probability density function of Y is

   $$f(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 \leq y \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

   Use this result to show that the maximum likelihood estimator for θ is biased.

(e) We have two unbiased estimators for θ: the moment estimator $\hat{\theta}_1 = 2\bar{X}$ and $\hat{\theta}_2 = [(n+1)/n]max(X_i)$,, where $max(X_i)$ is the largest observation in a random sample of size n. It can be shown that $V(\hat{\theta}_1) = \theta^2/(3n)$ and that $V(\hat{\theta}_2 = \theta^2/[n(n+2)])$. Show that if $n > 1$, $\hat{\theta}_2$ is a better estimator than $\hat{\theta}_1$. In what sense is it a better estimator of θ?

9. A random sample $X_1, X_2, ..., X_n$ of size n is taken from a Poisson distribution with a mean of $\lambda, 0 < \lambda < \infty$.

   (a) Show that the maximum likelihood estimator for λ is $\hat{\lambda} = \bar{X}$.

   (b) Let X equal the number of flaws per 100 feet of a used computer tape. Assume that X has a Poisson distribution with a mean of λ. If 40 observations of X yielded 5 zeros, 7 ones, 12 twos, 9 threes, 5 fours, 1 five, and 1 six, find the maximum likelihood estimate of λ.

10. Let $f(x) = (1/\theta)x^{(1-\theta)/\theta}, 0 < x < 1, 0 < \theta < \infty$.

    (a) Show that the maximum likelihood estimator of θ is $\hat{\theta} = -(1/n)\sum_{i=1}^{n} \ln X_i$.

    (b) Show that $E(\hat{\theta}) = \theta$ and thus that $\hat{\theta}$ is an unbiased estimator of θ.

11. Let $X_1, X_2, ..., X_n$ be a random sample of size n from the exponential distribution whose pdf is $f(x;\theta) = (1/\theta)e^{-x/\theta}, 0 < x < \infty, 0 < \theta < \infty$.

    (a) Show that $\bar{X}$ is an unbiased estimator of θ.

    (b) Show that the variance of $\bar{X}$ is θ^2/n.

    (c) What is a good estimate of θ if a random sample of size 5 yielded the sample values 3.5, 8.1, 0.9, 4.4, and 0.5?

12. A diagnostic test for a certain disease is applied to n individuals known to not have the disease. Let X =the number among the n test results that are positive (indicating presence of the disease, so X is the number of false positives) and p =the probability that a disease-free individual's test result is positive (i.e., p is the true proportion of test results from disease-free individuals that are positive). Assume that only X is available rather than the actual sequence of test results.

    (a) Derive the maximum likelihood estimator of p. If $n = 20$ and $x = 3$, what is the estimate?

    (b) Is the estimator of part (a) unbiased?

    (c) If $n = 20$ and $x = 3$, what is the MLE of the probability $(1 - p)^5$ that none of the next five tests done on disease free individuals are positive?

13. The shear strength of each of ten test spot welds is determined, yielding the following data (psi):

    392 376 401 367 389 362 409 415 358 375

    (a) Assuming that shear strength is normally distributed, estimate the true average shear strength and standard deviation of shear strength using the method of maximum likelihood.

(b) Again assuming a normal distribution, estimate the strength value below which 95% of all welds will have their strengths. [Hint: What is the 95th percentile in terms of μ and σ? Now use the invariance principle.]

14. At time $t = 0$, 20 identical components are tested. The lifetime distribution of each is exponential with parameter λ. The experimenter then leaves the test facility unmonitored. On his return 24 hours later, the experimenter immediately terminates the test after noticing that $y = 15$ of the 20 components are still in operation (so 5 have failed). Derive the MLE of λ. [Hint: Let Y = the number that survive 24 hours. Then $Y \sim Bin(n, p)$. What is the mle of p? Now notice that $p = P(X_i \geq 24)$, where X_i is exponentially distributed. This relates λ to p, so the former can be estimated once the latter has been.]

15. Let $X_1, X_2, ..., X_n$ be a random sample from $Bin(1, p)$ (i.e., n Bernoulli trials). Thus,

$$Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$$

(a) Show that $\bar{X} = Y/n$ is an unbiased estimator of p.

(b) Show that $Var(\bar{X}) = p(1-p)/n$.

(c) Show that $E[\bar{X}(1-\bar{X})/n] = (n-1)[p(1-p)/n^2]$.

(d) Find the value of c so that $c\bar{X}(1-\bar{X})$ is an unbiased estimator of $Var(\bar{X}) = p(1-p)/n$.

16. Assume that the number of defects in a car has a Poisson distribution with parameter λ. To estimate λ we obtain the random sample $X_1, X_2, ..., X_n$.

(a) Find the Fisher information in a single observation using two methods.

(b) Find the Cramer-Rao lower bound for the variance of an unbiased estimator of λ.

(c) Find the MLE of λ and show that the MLE is an efficient estimator.

17. Suppose the waiting time for a bus is uniformly distributed on $[0, \theta]$ and the results $x_1, ..., x_n$ of a random sample from this distribution have been observed.

(a) Find the MLE $\hat{\theta}$ of θ.

(b) Letting $\tilde{\theta} = \frac{n+1}{n}\hat{\theta}$, show that $\tilde{\theta}$ is unbiased and find its variance.

(c) Find the Cramer-Rao lower bound for the variance of an unbiased estimator of θ.

18. An estimator $\hat{\theta}$ is said to be **consistent** if for any $\epsilon > 0$, $P(|\hat{\theta} - \theta| \geq \epsilon) \to 0$ as $n \to \infty$. That is, $\hat{\theta}$ is consistent if, as the sample size gets larger, it is less and less likely that $\hat{\theta}$ will be further than ϵ from the true value of θ. Show that $\bar{X}$ is a consistent estimator of μ when $\sigma^2 < \infty$ by using Chebyshev's inequality.
Hint: (Chebyshev's inequality) Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k > 0$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$