Taking the derivative with respect to λ and setting it equal to 0 yields

$$\frac{d}{d\lambda}\ln f = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

Solving for λ yields the MLE:

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

Desirable Properties of Maximum Likelihood Estimators

Maximum likelihood is the most commonly used method of estimation. The main reason for this is that in most cases that arise in practice, MLEs have two very desirable properties.

1. In most cases, as the sample size n increases, the bias of the MLE converges to 0.
2. In most cases, as the sample size n increases, the variance of the MLE converges to a theoretical minimum.

Together, these two properties imply that when the sample size is sufficiently large, the bias of the MLE will be negligible, and the variance will be nearly as small as is theoretically possible.

Exercises for Section 4.9

1. Choose the best answer to fill in the blank. If an estimator is unbiased, then ____________

  i. the estimator is equal to the true value.
  ii. the estimator is usually close to the true value.
  iii. the mean of the estimator is equal to the true value.
  iv. the mean of the estimator is usually close to the true value.

2. Choose the best answer to fill in the blank. The variance of an estimator measures ____________

  i. how close the estimator is to the true value.
  ii. how close repeated values of the estimator are to each other.
  iii. how close the mean of the estimator is to the true value.
  iv. how close repeated values of the mean of the estimator are to each other.

3. Let X_1 and X_2 be independent, each with unknown mean μ and known variance $\sigma^2 = 1$.

  a. Let $\widehat{\mu}_1 = \dfrac{X_1 + X_2}{2}$. Find the bias, variance, and mean squared error of $\widehat{\mu}_1$.
  b. Let $\widehat{\mu}_2 = \dfrac{X_1 + 2X_2}{3}$. Find the bias, variance, and mean squared error of $\widehat{\mu}_2$.
  c. Let $\widehat{\mu}_3 = \dfrac{X_1 + X_2}{4}$. Find the bias, variance, and mean squared error of $\widehat{\mu}_3$.
  d. For what values of μ does $\widehat{\mu}_3$ have smaller mean squared error than $\widehat{\mu}_1$?
  e. For what values of μ does $\widehat{\mu}_3$ have smaller mean squared error than $\widehat{\mu}_2$?

4. Let $X_1, \ldots, X_n$ be a simple random sample from a $N(\mu, \sigma^2)$ population. For any constant $k > 0$, define $\widehat{\sigma}_k^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{k}$. Consider $\widehat{\sigma}_k^2$ as an estimator of σ^2.

  a. Compute the bias of $\widehat{\sigma}_k^2$ in terms of k. [*Hint:* The sample variance s^2 is unbiased, and $\widehat{\sigma}_k^2 = (n-1)s^2/k$.]
  b. Compute the variance of $\widehat{\sigma}_k^2$ in terms of k. [*Hint:* $\sigma_{s^2}^2 = 2\sigma^4/(n-1)$, and $\widehat{\sigma}_k^2 = (n-1)s^2/k$.]
  c. Compute the mean squared error of $\widehat{\sigma}_k^2$ in terms of k.

d. For what value of k is the mean squared error of $\widehat{\sigma}_k^2$ minimized?

5. Let $X \sim \text{Geom}(p)$. Find the MLE of p.

6. Let $X_1, \dots, X_n$ be a random sample from a population with the Poisson(λ) distribution. Find the MLE of λ.

7. Maximum likelihood estimates possess the property of *functional invariance*, which means that if $\widehat{\theta}$ is the MLE of θ, and $h(\theta)$ is any function of θ, then $h(\widehat{\theta})$ is the MLE of $h(\theta)$.

a. Let $X \sim \text{Bin}(n, p)$ where n is known and p is unknown. Find the MLE of the odds ratio $p/(1-p)$.

b. Use the result of Exercise 5 to find the MLE of the odds ratio $p/(1-p)$ if $X \sim \text{Geom}(p)$.

c. If $X \sim \text{Poisson}(\lambda)$, then $P(X = 0) = e^{-\lambda}$. Use the result of Exercise 6 to find the MLE of $P(X = 0)$ if $X_1, \dots, X_n$ is a random sample from a population with the Poisson(λ) distribution.

8. Let $X_1, \dots, X_n$ be a random sample from a $N(\mu, 1)$ population. Find the MLE of μ.

9. Let $X_1, \dots, X_n$ be a random sample from a $N(0, \sigma^2)$ population. Find the MLE of σ.

10. Let $X_1, \dots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population. Find the MLEs of μ and of σ. (*Hint:* The likelihood function is a function of two parameters, μ and σ. Compute partial derivatives with respect to μ and σ and set them equal to 0 to find the values $\widehat{\mu}$ and $\widehat{\sigma}$ that maximize the likelihood function.)

4.10 Probability Plots

Scientists and engineers frequently work with data that can be thought of as a random sample from some population. In many such cases, it is important to determine a probability distribution that approximately describes that population. In some cases, knowledge of the process that generated the data can guide the decision. More often, though, the only way to determine an appropriate distribution is to examine the sample to find a probability distribution that fits.

Probability plots provide a good way to do this. Given a random sample $X_1, \dots, X_n$, a probability plot can determine whether the sample might plausibly have come from some specified population. We will present the idea behind probability plots with a simple example. A random sample of size 5 is drawn, and we want to determine whether the population from which it came might have been normal. The sample, arranged in increasing order, is

$$3.01,\ 3.35,\ 4.79,\ 5.96,\ 7.89$$

Denote the values, in increasing order, by $X_1, \dots, X_n$ ($n = 5$ in this case). The first thing to do is to assign increasing, evenly spaced values between 0 and 1 to the X_i. There are several acceptable ways to do this; perhaps the simplest is to assign the value $(i - 0.5)/n$ to X_i. The following table shows the assignment for the given sample.

i	X_i	$(i - 0.5)/5$
1	3.01	0.1
2	3.35	0.3
3	4.79	0.5
4	5.96	0.7
5	7.89	0.9