CHAPTER 5

Estimation

5.1 INTRODUCTION
5.2 ESTIMATING PARAMETERS: THE METHOD OF MAXIMUM LIKELIHOOD AND THE METHOD OF MOMENTS
5.3 INTERVAL ESTIMATION
5.4 PROPERTIES OF ESTIMATORS
5.5 MINIMUM-VARIANCE ESTIMATORS: THE CRAMÉR-RAO LOWER BOUND
5.6 SUFFICIENT ESTIMATORS
5.7 CONSISTENCY
5.8 BAYESIAN ESTIMATION
5.9 TAKING A SECOND LOOK AT STATISTICS (REVISITING THE MARGIN OF ERROR)
APPENDIX 5.A.1 MINITAB APPLICATIONS

Ronald Aylmer Fisher

A towering figure in the development of both applied and mathematical statistics, Fisher took formal training in mathematics and theoretical physics, graduating from Cambridge in 1912. After a brief career as a teacher, he accepted a post in 1919 as statistician at the Rothamsted Experimental Station. There the day-to-day problems he encountered in collecting and interpreting agricultural data led directly to much of his most important work in the theory of estimation and experimental design. Fisher was also a prominent geneticist and devoted considerable time to the development of a quantitative argument that would support Darwin’s theory of natural selection. He returned to academia in 1933, succeeding Karl Pearson as the Galton Professor of Eugenics at the University of London. Fisher was knighted in 1952.

—Ronald Aylmer Fisher (1890–1962)
5.1 INTRODUCTION

The ability of probability functions to describe, or model, experimental data was demonstrated in numerous examples in Chapter 4. In Section 4.2, for example, the Poisson distribution was shown to predict very well the number of alpha emissions from a radioactive source as well as the number of fumbles made by a college football team. In Section 4.3 another probability model, the normal curve, was applied to phenomena as diverse as breath analyzer readings and IQ scores. Other models illustrated in Chapter 4 included the exponential, negative binomial, and gamma distributions.

All of these probability functions, of course, are actually families of models in the sense that each includes one or more parameters. The Poisson model, for instance, is indexed by the occurrence rate, \( \lambda \). Changing \( \lambda \) changes the probabilities associated with \( p_X(k) \) [see Figure 5.1.1, which compares \( p_X(k) = e^{-\lambda} \lambda^k / k! \), \( k = 0, 1, 2, \ldots \) for \( \lambda = 1 \) and \( \lambda = 4 \)]. Similarly, the binomial model is defined in terms of the success probability \( p \); the normal distribution, by the two parameters \( \mu \) and \( \sigma \).

Before any of these models can be applied, values need to be assigned to their parameters. Typically, this is done by taking a random sample (of \( n \) observations) and using those measurements to estimate the unknown parameter(s).

![Figure 5.1.1](image)

**EXAMPLE 5.1.1**

Imagine being handed a coin whose probability, \( p \), of coming up heads is unknown. Your assignment is to toss the coin three times and use the resulting sequence of Hs and Ts to suggest a value for \( p \). Suppose the sequence of three tosses turns out to be HHT. Based on those outcomes, what can be reasonably inferred about \( p \)?

Start by defining the random variable \( X \) to be the number of heads on a given toss. Then

\[
X = \begin{cases} 
1 & \text{if a toss comes up heads} \\
0 & \text{if a toss comes up tails}
\end{cases}
\]
and the theoretical probability model for $X$ is the function

$$p_X(k) = p^k(1 - p)^{1-k} = \begin{cases} p & \text{for } k = 1 \\ 1 - p & \text{for } k = 0 \end{cases}$$

Expressed in terms of $X$, the sequence HHT corresponds to a sample of size $n = 3$, where $X_1 = 1$, $X_2 = 1$, and $X_3 = 0$.

Since the $X_i$s are independent random variables, the probability associated with the sample is $p^2(1 - p)$:

$$P(X_1 = 1 \cap X_2 = 1 \cap X_3 = 0) = P(X_1 = 1) \cdot P(X_2 = 1) \cdot P(X_3 = 0) = p^2(1 - p)$$

Knowing that our objective is to identify a plausible value (i.e., an "estimate") for $p$, it could be argued that a reasonable choice for that parameter would be the value that maximizes the probability of the sample. Figure 5.1.2 shows $P(X_1 = 1, X_2 = 1, X_3 = 0)$ as a function of $p$. By inspection, we see that the value that maximizes the probability of HHT is $p = \frac{2}{3}$.

More generally, suppose we toss the coin $n$ times and record a set of outcomes $X_1 = k_1$, $X_2 = k_2, \ldots$, and $X_n = k_n$. Then

$$P(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) = p^{k_1}(1 - p)^{1-k_1} \cdots p^{k_n}(1 - p)^{1-k_n} = \prod_{i=1}^{n} p^{k_i} (1 - p)^{n-k_i}$$

![Figure 5.1.2](attachment:image.png)
The value of \( p \) that maximizes \( P(X_1 = k_1, \ldots, X_n = k_n) \) is, of course, the value for which the derivative of \( p^{\sum_{i=1}^n k_i} (1 - p)^{n - \sum_{i=1}^n k_i} \) with respect to \( p \) is 0. But

\[
d/dp \left[ \sum_{i=1}^n k_i \left( \frac{p^{\sum_{i=1}^n k_i - n}}{p^{\sum_{i=1}^n k_i - n}} \right)^{n-\sum_{i=1}^n k_i} \right] = \sum_{i=1}^n k_i \left[ \sum_{i=1}^n k_i - 1 \right] (1 - p)^{n-\sum_{i=1}^n k_i - 1} \]

(5.1.1)

If the derivative is set equal to zero, Equation 5.1.1 reduces to

\[
\sum_{i=1}^n k_i (1 - p) + \left( \sum_{i=1}^n k_i - n \right) p = 0
\]

Solving for \( p \) identifies

\[
\left( \frac{1}{n} \right) \sum_{i=1}^n k_i
\]

as the value of the parameter that is most consistent with the \( n \) observations \( k_1, k_2, \ldots, k_n \).

**Comment.** Any function of a random sample whose objective is to approximate a parameter is called a statistic, or an estimator. If \( \theta \) is the parameter being approximated, its estimator will be denoted \( \hat{\theta} \). When an estimator is evaluated (by substituting the actual measurements recorded), the resulting number is called an estimate. In Example 5.1.1, the function \( \left( \frac{1}{n} \right) \sum_{i=1}^n X_i \) is an estimator for \( p \); the value \( \frac{2}{3} \) that is calculated when the \( n = 3 \) observations are \( X_1 = 1, X_2 = 1, \) and \( X_3 = 0 \) is an estimate of \( p \). More specifically, \( \left( \frac{1}{n} \right) \sum_{i=1}^n X_i \) is a maximum likelihood estimator (for \( p \)) and \( \frac{2}{3} \) \( = \left( \frac{1}{n} \right) \sum_{i=1}^n k_i \left( = \frac{1}{3} \right) \) (2) is a maximum likelihood estimate (for \( p \)).

In this chapter, we look at some of the practical, as well as the mathematical, issues involved in the problem of estimating parameters. How is the functional form of an estimator determined? What statistical properties does a given estimator have? What properties would we like an estimator to have? As we answer these questions, our focus will begin to shift away from the study of probability and toward the study of statistics.

### 5.2 ESTIMATING PARAMETERS: THE METHOD OF MAXIMUM LIKELIHOOD AND THE METHOD OF MOMENTS

Suppose \( Y_1, Y_2, \ldots, Y_n \) is a random sample from a continuous pdf \( f_Y(y) \), whose unknown parameter is \( \theta \). [Note: To emphasize that our focus is on the parameter, we will identify continuous pdf's in this chapter as \( f_Y(y; \theta) \); similarly, discrete probability models with an
unknown parameter \( \theta \) will be denoted \( p_X(k; \theta) \). The question is, how should we use the data to approximate \( \theta \)?

In Example 5.1.1, we saw that the parameter \( p \) in the discrete probability model \( f_X(k; p) = p^k(1-p)^{1-k}, k = 0, 1 \) could reasonably be estimated by the function \( \frac{1}{n} \sum_{i=1}^{n} k_i, \) based on the random sample \( X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n \). How would the form of the estimate change if the data came from, say, an exponential distribution? Or a Poisson distribution?

In this section we introduce two techniques for finding estimates—the method of maximum likelihood and the method of moments. Others are available, but these are the two that are the most widely used. Often, but not always, they give the same answer.

**The Method of Maximum Likelihood**

The basic idea behind maximum likelihood estimation is the rationale that was appealed to in Example 5.1.1. That is, it seems plausible to choose as the estimate for \( \theta \) that value of the parameter that maximizes the "likelihood" of the sample. The latter is measured by a **likelihood function**, which is simply the product of the underlying pdf evaluated for each of the data points. In Example 5.1.1, the likelihood function for the sample HHT (i.e., for \( X_1 = 1, X_2 = 1, \) and \( X_3 = 0 \)) is the product \( p^2(1-p) \).

**Definition 5.2.1.** Let \( k_1, k_2, \ldots, k_n \) be a random sample of size \( n \) from the discrete pdf \( p_X(k; \theta) \), where \( \theta \) is an unknown parameter. The **likelihood function**, \( L(\theta) \), is the product of the pdf evaluated at the \( n \) \( k_i \)s. That is,

\[
L(\theta) = \prod_{i=1}^{n} p_X(k_i; \theta)
\]

If \( y_1, y_2, \ldots, y_n \) is a random sample of size \( n \) from a continuous pdf, \( f_Y(y; \theta) \), where \( \theta \) is an unknown parameter, the likelihood function is written

\[
L(\theta) = \prod_{i=1}^{n} f_Y(y_i; \theta)
\]

**Comment.** Joint pdf's and likelihood functions look the same, but the two are interpreted differently. A joint pdf defined for a set of \( n \) random variables is a multivariate function of the values of those \( n \) random variables, either \( k_1, k_2, \ldots, k_n \) or \( y_1, y_2, \ldots, y_n \). By contrast, \( L \) is a function of \( \theta \); it should not be considered a function of either the \( k_i \)s or \( y_i \)s.

**Definition 5.2.2.** Let \( L(\theta) = \prod_{i=1}^{n} p_X(k_i; \theta) \) and \( L(\theta) = \prod_{i=1}^{n} f_Y(y_i; \theta) \) be the likelihood functions corresponding to random samples \( k_1, k_2, \ldots, k_n \) and \( y_1, y_2, \ldots, y_n \) drawn from the discrete pdf \( p_X(k; \theta) \) and continuous pdf \( f_Y(y; \theta) \), respectively, where \( \theta \) is an unknown parameter. In each case, let \( \theta_e \) be a value of the parameter such that \( L(\theta_e) \geq L(\theta) \) for all possible values of \( \theta \). Then \( \theta_e \) is called a **maximum likelihood estimate** for \( \theta \).
Applying the Method of Maximum Likelihood

We will see in Example 5.2.1 and many subsequent examples that finding the \( \theta_e \) that maximizes a likelihood function is often an application of the calculus. Specifically, we solve the equation \( \frac{d}{d\theta} L(\theta) = 0 \) for \( \theta \). In some cases, a more tractable equation results by setting the derivative of \( \ln L(\theta) \) equal to 0. Since \( \ln L(\theta) \) increases with \( L(\theta) \), the same \( \theta_e \) that maximizes \( L(\theta) \) also maximizes \( \ln L(\theta) \).

**EXAMPLE 5.2.1**

Suppose that \( X_1 = 3, X_2 = 2, X_3 = 1, \) and \( X_4 = 3 \) is a set of four independent observations representing the geometric probability model, \( p_X(k) = (1 - p)^{k-1} p, \ k = 1, 2, 3, \ldots \) Find the maximum likelihood estimate for \( p \).

According to Definition 5.2.1,

\[
L(p) = [(1 - p)^{3-1} p][(1 - p)^{2-1} p][(1 - p)^{1-1} p][(1 - p)^{3-1} p]
= (1 - p)^5 p^4
\]

Then \( \ln L(p) = 5 \ln(1 - p) + 4 \ln p \). Differentiating \( \ln L(p) \) with respect to \( p \) gives

\[
\frac{d \ln L(p)}{dp} = \frac{-5}{1 - p} + \frac{4}{p}
\]

To find the \( p \) that maximizes \( L(p) \), we set the derivative equal to zero. Here, \( \frac{-5}{1 - p} + \frac{4}{p} = 0 \) implies that \( -5p + 4(1 - p) = 0 \), and the solution to the latter is \( p = \frac{4}{9} \).

Notice, also, that the second derivative of \( \ln L(p) \left( = \frac{-5}{(1 - p)^2} - \frac{4}{p^2} \right) \) is negative for all \( 0 < p < 1 \), so \( p = \frac{4}{9} \) is, indeed, a true maximum of the likelihood function. (Following the notation introduced in Definition 5.2.2, \( \frac{4}{9} \) is called the maximum likelihood estimate for \( p \), and we would write \( p_e = \frac{4}{9} \).)

**Comment.** There is a better way to answer the question posed in Example 5.2.1. Rather than evaluate—and differentiate—the likelihood function for the particular sample observed (in this case, the four observations 3, 2, 1, and 3), we can get a more informative answer by considering the more general problem of taking a random sample of size \( n \) from \( p_X(k) = (1 - p)^{k-1} p \) and using the outcomes—\( X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n \)—to find a formula for the maximum likelihood estimate.

For the geometric pdf, the likelihood function based on such a sample would be written

\[
L(p) = \prod_{i=1}^{n} (1 - p)^{k_i-1} p
= (1 - p)^{\sum_{i=1}^{n} k_i - n} p^n
\]
As was the case in Example 5.2.1, it will be easier to work with ln \( L(p) \) than \( L(p) \). Here,

\[
\ln L(p) = \left( \sum_{i=1}^{n} k_i - n \right) \cdot \ln(1 - p) + n \ln p
\]

and

\[
d \ln L(p) / dp = \left( n - \sum_{i=1}^{n} k_i \right) / (1 - p) + n / p
\]

Setting the derivative equal to 0 gives

\[
p \left( n - \sum_{i=1}^{n} k_i \right) + (1 - p)n = 0
\]

which implies that

\[
p_c = n \sum_{i=1}^{n} k_i
\]

(Reassuringly, for the particular sample assumed in Example 5.2.1—\( n = 4 \) and \( \sum_{i=1}^{4} k_i = 3 + 2 + 1 + 3 = 9 \)—the formula just derived reduces to the maximum likelihood estimate of \( \frac{4}{9} \) that we found at the outset.)

**Comment.** Implicit in Example 5.2.1 and the Comment that followed is the important distinction between a maximum likelihood estimate and a maximum likelihood estimator. The first is a number (or refers to a number); the second is a random variable (recall the Comment on p. 346).

Both \( \frac{4}{9} \) and the formula \( \frac{n}{\sum_{i=1}^{n} k_i} \) are maximum likelihood estimates (for \( p \)) and would be denoted \( p_e \), because both are numerical constants. In the first case, the actual values of the \( k_i \)'s are provided and \( p_e (= \frac{4}{9}) \) can be calculated. In the second case, the \( k_i \)'s are not identified but they are constants nonetheless.

If, on the other hand, we imagine the measurements before they are recorded—that is, as being the random variables \( X_1, X_2, \ldots, X_n \)—then the formula \( \frac{n}{\sum_{i=1}^{n} k_i} \) is more properly written as the quotient

\[
\frac{n}{\sum_{i=1}^{n} X_i}
\]

The latter, a random variable, is the maximum likelihood estimator (for \( p \)) and would be denoted \( \hat{p} \). Maximum likelihood estimators, such as \( \hat{p} \), have pdfs, expected values, and variances; maximum likelihood estimates, such as \( p_e \), have none of those statistical properties.
EXAMPLE 5.2.2
An experimenter has reason to believe that the pdf describing the variability in a certain type of measurement is the continuous model

\[ f_Y(y; \theta) = \frac{1}{\theta^2} ye^{-y/\theta}, \quad 0 < y < \infty; 0 < \theta < \infty \]

Five data points have been collected—9.2, 5.6, 18.4, 12.1, and 10.7. Find the maximum likelihood estimate for \( \theta \).

Following the advice given in the Comment on p. 348, we begin by deriving a general formula for \( \theta_e \)—that is, by assuming that the data are the \( n \) observations, \( y_1, y_2, \ldots, y_n \). The likelihood function, then, becomes

\[
L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta^2} ye^{-y_i/\theta}
\]

\[
= \theta^{-2n} \prod_{i=1}^{n} y_i e^{-y_i/\theta}
\]

and

\[
\ln L(\theta) = -2n \ln \theta + \ln \prod_{i=1}^{n} y_i - \frac{1}{\theta} \sum_{i=1}^{n} y_i
\]

Setting the derivative of \( \ln L(\theta) \) equal to 0 gives

\[
\frac{d \ln L(\theta)}{d\theta} = -2n/\theta + 1/\theta^2 \sum_{i=1}^{n} y_i = 0
\]

which implies that

\[
\theta_e = \frac{1}{2n} \sum_{i=1}^{n} y_i
\]

The final step is to evaluate numerically the formula for \( \theta_e \). Substituting the actual \( n = 5 \) sample values recorded gives \( \sum_{i=1}^{5} y_i = 9.2 + 5.6 + 18.4 + 12.1 + 10.7 = 56.0 \), so

\[
\theta_e = \frac{1}{2(5)} (56.0) = 5.6
\]

Using Order Statistics as Maximum Likelihood Estimates

Situations exist for which the equations \( \frac{dL(\theta)}{d\theta} = 0 \) or \( \frac{d \ln L(\theta)}{d\theta} = 0 \) are not meaningful and neither will yield a solution for \( \theta_e \). These occur when the range of the pdf from which the data are drawn is a function of the parameter being estimated. (This happens, for instance, when the sample of \( y_i \)s come from the uniform pdf, \( f_Y(y; \theta) = 1/\theta, 0 \leq y \leq \theta \).) The maximum likelihood estimates in these cases will be an order statistic, typically either \( y_{\text{min}} \) or \( y_{\text{max}} \).
EXAMPLE 5.2.3

Suppose \( y_1, y_2, \ldots, y_n \) is a set of measurements representing an exponential pdf with \( \lambda = 1 \) but with an unknown "threshold" parameter, \( \theta \). That is,

\[
f_Y(y; \theta) = e^{-(y-\theta)}, \quad y \geq \theta; \quad \theta > 0
\]

(see Figure 5.2.1). Find the maximum likelihood estimate for \( \theta \).

![Graph showing the pdf](image)

**FIGURE 5.2.1**

Proceeding in the usual fashion, we start by deriving an expression for the likelihood function:

\[
L(\theta) = \prod_{i=1}^{n} e^{-(y_i-\theta)} = e^{-\sum_{i=1}^{n} y_i + n\theta}
\]

Here, finding \( \theta_e \) by solving the equation \( \frac{d \ln L(\theta)}{d\theta} = 0 \) will not work because \( \frac{d \ln L(\theta)}{d\theta} = \frac{d}{d\theta} \left( -\sum_{i=1}^{n} y_i + n\theta \right) = n \). Instead, we need to look at the likelihood function directly.

Notice that \( L(\theta) = e^{-\sum_{i=1}^{n} y_i + n\theta} \) is maximized when the exponent of \( e \) is maximized. But for given \( y_1, y_2, \ldots, y_n \) (and \( n \)), making \( -\sum_{i=1}^{n} y_i + n\theta \) as large as possible requires that \( \theta \) be as large as possible. Figure 5.2.1 shows how large \( \theta \) can be: It can be moved to the right only as far as the smallest order statistic. Any value of \( \theta \) larger than \( y_{\text{min}} \) would violate the condition on \( f_Y(y; \theta) \) that \( y \geq \theta \). Therefore, \( \theta_e = y_{\text{min}} \).
CASE STUDY 5.2.1

"What are you majoring in?" may be the most common question asked of a college student. For some, the answer is simple: Having decided on a field of study, they doggedly stay with it all the way to graduation. For many, though, the path is not so straight. Premeds losing the battle with organic chemistry and engineers unable to appreciate the joy of secants may find their roads to commencement taking a few detours.

Listed in the first two columns of Table 5.2.1 are the results of a "major" poll conducted at the University of West Florida (114). Recorded for each of 356 upperclassmen was the number of times, \( X \), that he or she had switched majors.

Based on the nature of these data, it would not be unreasonable to hypothesize that \( X \) has a Poisson distribution (recall the discussion of the law of small numbers in Section 4.2). Do the actual frequencies support that contention?

<table>
<thead>
<tr>
<th>Number of Major Changes</th>
<th>Observed Frequency</th>
<th>Expected Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>237</td>
<td>230.4</td>
</tr>
<tr>
<td>1</td>
<td>90</td>
<td>100.2</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>21.8</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>356</td>
<td>356.0</td>
</tr>
</tbody>
</table>

To see if \( p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!} \) can provide an adequate fit to these 356 observations requires that we first find an estimate for \( \lambda \). Given that \( X_1 = k_1, X_2 = k_2, \ldots, \) and \( X_n = k_n \),

\[
L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{k_i}}{k_i!} = \frac{e^{-n\lambda}\sum_{i=1}^{n} k_i}{\prod_{i=1}^{n} k_i!} \\
\ln L(\lambda) = -n\lambda + \left( \sum_{i=1}^{n} k_i \right) \ln \lambda - \ln \prod_{i=1}^{n} k_i! \\
\frac{d \ln L(\lambda)}{d \lambda} = -n + \frac{\sum_{i=1}^{n} k_i}{\lambda}
\]

(Continued on next page)
Setting the derivative equal to zero shows that the maximum likelihood estimate for $\lambda$ is the sample mean:

$$\lambda_e = \frac{1}{n} \sum_{i=1}^{n} k_i$$ (5.2.1)

According to the information appearing in Table 5.2.1, 237 of the $k_i$’s were equal to zero, ninety were equal to one, and so on. Substituting into Equation 5.2.1, then, gives

$$\lambda_e = \frac{1}{356} [237 \cdot 0 + 90 \cdot 1 + 22 \cdot 2 + 7 \cdot 3]$$

$$= 0.435$$

so the specific model being proposed is

$$p_X(k) = \frac{e^{-0.435}(0.435)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

The corresponding expected frequencies [ = 356 $\cdot$ $p_X(k)$] for each value of $X$ are listed in column 3 of Table 5.2.1. Agreement with the observed frequencies appears to be quite good. Our conclusion would be that nothing in these data rules out using the Poisson as a “major-change” model. [Formal procedures, known as goodness-of-fit tests, have been developed for assessing the agreement (or lack of agreement) between a set of observed and expected frequencies. These will be taken up in Chapter 10.]

Finding Maximum Likelihood Estimates When More Than One Parameter Is Unknown

If a family of probability models is indexed by two or more unknown parameters—say, $\theta_1, \theta_2, \ldots, \theta_k$—finding maximum likelihood estimates for the $\theta_i$s requires the solution of a set of $k$ simultaneous equations. If $k = 2$, for example, we would typically need to solve the system

$$\frac{\partial \ln L(\theta_1, \theta_2)}{\partial \theta_1} = 0$$

$$\frac{\partial \ln L(\theta_1, \theta_2)}{\partial \theta_2} = 0$$

Example 5.2.4

Suppose a random sample of size $n$ is drawn from the two-parameter normal pdf,

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}, \quad -\infty < y < \infty; -\infty < \mu < \infty; \sigma^2 > 0$$

Use the method of maximum likelihood to find formulas for $\mu_e$ and $\sigma_e^2$. 
We start by finding $L(\mu, \sigma^2)$ and $\ln L(\mu, \sigma^2)$:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left( \frac{y_i - \mu}{\sigma} \right)^2}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^{n} \left( \frac{y_i - \mu}{\sigma} \right)^2$$

Moreover,

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = -\sum_{i=1}^{n} \left( \frac{y_i - \mu}{\sigma} \right) \left( -\frac{1}{\sigma} \right)$$

and

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{2\sigma^2} \cdot 2\pi - \frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^2 \left( -\frac{1}{\sigma^4} \right)$$

Setting the two derivatives equal to zero gives the equations

$$\sum_{i=1}^{n} (y_i - \mu) = 0 \quad (5.2.2)$$

and

$$-n\sigma^2 + \sum_{i=1}^{n} (y_i - \mu)^2 = 0 \quad (5.2.3)$$

Equation 5.2.2 simplifies to

$$\sum_{i=1}^{n} y_i = n\mu$$

which implies that $\mu_e = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$. Substituting $\mu_e$, then, into Equation 5.2.3 gives

$$-n\sigma^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0$$

or

$$\sigma^2_e = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

**Comment.** The method of maximum likelihood has a long history: Daniel Bernoulli was using it as early as 1777 (136). It was Ronald Fisher, though, in the early years of the twentieth century, who first studied the mathematical properties of likelihood estimation in any detail, and the procedure is often credited to him.
QUESTIONS

5.2.1. A random sample of size 8—\( X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 1, X_7 = 1, \) and \( X_8 = 0 \)—is taken from the probability function

\[
p_X(k; \theta) = \theta^k (1 - \theta)^{1-k}, \quad k = 0, 1; \quad 0 < \theta < 1
\]

Find the maximum likelihood estimate for \( \theta \).

5.2.2. The number of red chips and white chips in an urn is unknown, but the proportion, \( p \), of reds is either \( \frac{1}{3} \) or \( \frac{1}{4} \). A sample of size 5, drawn with replacement, yields the sequence red, white, white, red, and white. What is the maximum likelihood estimate for \( p \)?

5.2.3. Use the sample \( Y_1 = 8.2, Y_2 = 9.1, Y_3 = 10.6, \) and \( Y_4 = 4.9 \) to calculate the maximum likelihood estimate for \( \lambda \) in the exponential pdf

\[
f_Y(y; \lambda) = \lambda e^{-\lambda y}, \quad y \geq 0
\]

5.2.4. Suppose a random sample of size \( n \) is drawn from the probability model

\[
p_X(k; \theta) = \frac{\theta^k e^{-\theta}}{k!}, \quad k = 0, 1, 2, \ldots
\]

Find a formula for the maximum likelihood estimator, \( \hat{\theta} \).

5.2.5. Given that \( Y_1 = 2.3, Y_2 = 1.9, \) and \( Y_3 = 4.6 \) is a random sample from

\[
f_Y(y; \theta) = \frac{y^3 e^{-y/\theta}}{664}, \quad y \geq 0
\]

calculate the maximum likelihood estimate for \( \theta \).

5.2.6. Use the method of maximum likelihood to estimate \( \theta \) in the pdf

\[
f_Y(y; \theta) = \frac{\theta}{2\sqrt{y}} e^{-\theta \sqrt{y}}, \quad y > 0
\]

Evaluate \( \theta_e \) for the following random sample of size 4: \( Y_1 = 6.2, Y_2 = 7.0, Y_3 = 2.5, \) and \( Y_4 = 4.2 \).

5.2.7. An engineer is creating a project scheduling program and recognizes that the tasks making up the project are not always completed on time. However, the completion proportion tends to be fairly high. To reflect this condition, he uses the pdf

\[
f_Y(y; \theta) = \theta y^{\theta - 1}, \quad 0 \leq y \leq 1, \quad \text{and} \quad 0 < \theta
\]

where \( y \) is the proportion of the task completed. Suppose in his previous project, the proportion of tasks completed were 0.77, 0.82, 0.92, 0.94, and 0.98. Estimate \( \theta \).
5.2.8. The following data show the number of occupants in passenger cars observed during
one hour at a busy intersection in Los Angeles (68). Suppose it can be assumed that
these data follow a geometric distribution, \( p_X(k; p) = (1 - p)^{k-1} p, k = 1, 2, \ldots \).
Estimate \( p \) and compare the observed and expected frequencies for each value of \( X \).

<table>
<thead>
<tr>
<th>Number of Occupants</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>678</td>
</tr>
<tr>
<td>2</td>
<td>227</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6+</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>1011</td>
</tr>
</tbody>
</table>

5.2.9. (a) Based on the random sample \( Y_1 = 6.3, Y_2 = 1.8, Y_3 = 14.2, \) and \( Y_4 = 7.6, \) use
the method of maximum likelihood to estimate the parameter \( \theta \) in the uniform pdf

\[
f_Y(y; \theta) = \frac{1}{\theta}, \quad 0 \leq y \leq \theta
\]

(b) Suppose the random sample in Part (a) represents the two-parameter uniform pdf

\[
f_Y(y; \theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1}, \quad \theta_1 \leq y \leq \theta_2
\]

Find the maximum likelihood estimates for \( \theta_1 \) and \( \theta_2 \).

5.2.10. Find the maximum likelihood estimate for \( \theta \) in the pdf

\[
f_Y(y; \theta) = \frac{2y}{1 - \theta^2}, \quad 0 \leq y \leq 1
\]

if a random sample of size 6 yielded the measurements 0.70, 0.63, 0.92, 0.86, 0.43, and 0.21.

5.2.11. A random sample of size \( n \) is taken from the pdf

\[
f_Y(y; \theta) = 2y\theta^2, \quad 0 \leq y \leq \frac{1}{\theta}
\]

Find an expression for \( \hat{\theta} \), the maximum likelihood estimator for \( \theta \).

5.2.12. If the random variable \( Y \) denotes an individual's income, Pareto's law claims that

\[
P(Y \geq y) = \left( \frac{k}{y} \right)^\theta, \quad \text{where } k \text{ is the entire population's minimum income. It follows that}
\]

\[
F_Y(y) = 1 - \left( \frac{k}{y} \right)^\theta, \quad \text{and, by differentiation,}
\]

\[
f_Y(y; \theta) = \theta k^\theta \left( \frac{1}{y} \right)^{\theta+1}, \quad y \geq k; \quad \theta \geq 1
\]

Assume \( k \) is known. Find the maximum likelihood estimator for \( \theta \) if income information
has been collected on a random sample of 25 individuals.
5.2.13. The exponential pdf is a measure of lifetimes of devices that do not age (see Question 3.11.11). However, the exponential pdf is a special case of the Weibull distribution, which measures time to failure of devices where the probability of failure increases as time does. A Weibull random variable $Y$ has pdf $f_Y(y; \alpha, \beta) = \alpha \beta y^{\beta - 1} e^{-\alpha y^\beta}$, $0 \leq y$, $0 < \alpha, 0 < \beta$

(a) Find the maximum likelihood estimator for $\alpha$ assuming that $\beta$ is known.

(b) Suppose $\alpha$ and $\beta$ are both unknown. Write down the equations that would be solved simultaneously to find the maximum likelihood estimators of $\alpha$ and $\beta$.

5.2.14. Suppose a random sample of size $n$ is drawn from a normal pdf where the mean $\mu$ is known but the variance $\sigma^2$ is unknown. Use the method of maximum likelihood to find a formula for $\hat{\sigma}^2$. Compare your answer to the maximum likelihood estimator found in Example 5.2.4.

The Method of Moments

A second procedure for estimating parameters is the method of moments. Proposed near the turn of the twentieth century by the great British statistician, Karl Pearson, the method of moments is often more tractable than the method of maximum likelihood in situations where the underlying probability model has multiple parameters.

Suppose that $Y$ is a continuous random variable and its pdf is a function of $s$ unknown parameters, $\theta_1, \theta_2, \ldots, \theta_s$. The first $s$ moments of $Y$, if they exist, are given by the integrals

$$E(Y^j) = \int_{-\infty}^{\infty} y^j \cdot f_Y(y; \theta_1, \theta_2, \ldots, \theta_s) \, dy, \quad j = 1, 2, \ldots, s$$

In general, each $E(Y^j)$ will be a different function of the $s$ parameters. That is,

$$E(Y^1) = g_1(\theta_1, \theta_2, \ldots, \theta_s)$$

$$E(Y^2) = g_2(\theta_1, \theta_2, \ldots, \theta_s)$$

$$\vdots$$

$$E(Y^n) = g_s(\theta_1, \theta_2, \ldots, \theta_s)$$

Corresponding to each theoretical moment, $E(Y^j)$, is a sample moment, $\frac{1}{n} \sum_{i=1}^{n} y_i^j$.

Intuitively, the $j$th sample moment is an approximation to the $j$th theoretical moment. Setting the two equal for each $j$ produces a system of $s$ simultaneous equations, the solutions to which are the desired set of estimates, $\hat{\theta}_{1e}, \hat{\theta}_{2e}, \ldots, \hat{\theta}_{se}$.

**Definition 5.2.3.** Let $y_1, y_2, \ldots, y_n$ be a random sample from the continuous pdf $f_Y(y; \theta_1, \theta_2, \ldots, \theta_s)$. The method of moments estimates, $\hat{\theta}_{1e}, \hat{\theta}_{2e}, \ldots, \hat{\theta}_{se}$, for the
model's unknown parameters are the solutions of the \( s \) simultaneous equations

\[
\int_{-\infty}^{\infty} y f_Y(y; \theta_1, \theta_2, \ldots, \theta_s) \, dy = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i
\]

\[
\int_{-\infty}^{\infty} y^2 f_Y(y; \theta_1, \theta_2, \ldots, \theta_s) \, dy = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i^2
\]

\[
\vdots \quad \vdots
\]

\[
\int_{-\infty}^{\infty} y^s f_Y(y; \theta_1, \theta_2, \ldots, \theta_s) \, dy = \left(\frac{1}{n}\right) \sum_{i=1}^{n} y_i^s
\]

Note: If the underlying random variable is discrete with pdf \( p_X(k; \theta_1, \theta_2, \ldots, \theta_s) \), the method of moments estimates are the solutions of the system of equations,

\[
\sum_{k} k^j p_X(k; \theta_1, \theta_2, \ldots, \theta_s) = \left(\frac{1}{n}\right) \sum_{k} k^j, \quad j = 1, 2, \ldots, s
\]

**EXAMPLE 5.2.5**

Suppose that \( Y_1 = 0.42, Y_2 = 0.10, Y_3 = 0.65, \) and \( Y_4 = 0.23 \) is a random sample of size four from the pdf

\[ f_Y(y; \theta) = \theta y^{\theta-1}, \quad 0 \leq y \leq 1 \]

Find the method of moments estimate for \( \theta \).

Taking the same approach that was followed in finding maximum likelihood estimates, we will derive a general expression for the method of moments estimate before making any use of the four data points. Notice that only one equation needs to be solved because the pdf is indexed by just a single parameter.

The first theoretical moment of \( Y \) is \( \frac{\theta}{\theta + 1} \):

\[
E(Y) = \int_0^1 y \cdot \theta y^{\theta-1} \, dy
\]

\[
= \theta \cdot \left[ \frac{y^{\theta+1}}{\theta + 1} \right]_0^1
\]

\[
= \frac{\theta}{\theta + 1}
\]

Setting \( E(Y) \) equal to \( \frac{1}{n} \sum_{i=1}^{n} y_i (= \bar{y}) \), the first sample moment, gives

\[
\frac{\theta}{\theta + 1} = \bar{y}
\]
which implies that the method of moments estimate for $\theta$ is

$$\theta_e = \frac{\bar{y}}{1 - \bar{y}}$$

Here, $\bar{y} = \frac{1}{4}(0.42 + 0.10 + 0.65 + 0.23) = 0.35$, so

$$\theta_e = \frac{0.35}{1 - 0.35} = 0.54$$

**CASE STUDY 5.2.2**

Although hurricanes generally strike only the eastern and southern coastal regions of the United States, they do occasionally sweep inland before completely dissipating. The U.S. Weather Bureau confirms that in the period from 1900 to 1969 a total of thirty-six hurricanes moved as far as the Appalachians. In Table 5.2.2 are listed the maximum twenty-four-hour precipitation levels recorded for those thirty-six storms during the time they were over the mountains (67).

Figure 5.2.2 shows the data's density-scaled histogram. Its skewed shape suggests that $Y$, the maximum twenty-four-hour precipitation associated with inland hurricanes, can be modeled by the two-parameter gamma pdf,

$$f_Y(y; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}, \quad y > 0$$

Use the method of moments to estimate $r$ and $\lambda$; then superimpose $f_Y(y; r_e, \lambda)$ on a graph of the density-scaled histogram of the 36 $y$'s.

From Theorem 4.6.3,

$$E(Y) = \frac{r}{\lambda}$$

and

$$\text{Var}(Y) = \frac{r}{\lambda^2} = E(Y^2) - [E(Y)]^2$$

so

$$E(Y^2) = \frac{r}{\lambda^2} + \frac{r^2}{\lambda^2} = \frac{r(r + 1)}{\lambda^2}$$

(Continued on next page)
(Case Study 5.2.2 continued)

<table>
<thead>
<tr>
<th>Year</th>
<th>Name</th>
<th>Location</th>
<th>Maximum Precipitation (inches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1969</td>
<td>Camille</td>
<td>Tye River, Va.</td>
<td>31.00</td>
</tr>
<tr>
<td>1968</td>
<td>Candy</td>
<td>Hickley, N.Y.</td>
<td>2.82</td>
</tr>
<tr>
<td>1965</td>
<td>Betsy</td>
<td>Haywood Gap, N.C.</td>
<td>3.98</td>
</tr>
<tr>
<td>1960</td>
<td>Brenda</td>
<td>Cairo, N.Y.</td>
<td>4.02</td>
</tr>
<tr>
<td>1959</td>
<td>Gracie</td>
<td>Big Meadows, Va.</td>
<td>9.50</td>
</tr>
<tr>
<td>1957</td>
<td>Audrey</td>
<td>Russels Point, Ohio</td>
<td>4.50</td>
</tr>
<tr>
<td>1955</td>
<td>Connie</td>
<td>Slide Mt., N.Y.</td>
<td>11.40</td>
</tr>
<tr>
<td>1954</td>
<td>Hazel</td>
<td>Big Meadows, Va.</td>
<td>10.71</td>
</tr>
<tr>
<td>1954</td>
<td>Carol</td>
<td>Eagles Mere, Pa.</td>
<td>6.31</td>
</tr>
<tr>
<td>1952</td>
<td>Able</td>
<td>Bloserville 1-N, Pa.</td>
<td>4.95</td>
</tr>
<tr>
<td>1949</td>
<td></td>
<td>North Ford #1, N.C.</td>
<td>5.64</td>
</tr>
<tr>
<td>1945</td>
<td></td>
<td>Crossnore, N.C.</td>
<td>5.51</td>
</tr>
<tr>
<td>1942</td>
<td></td>
<td>Big Meadows, Va.</td>
<td>13.40</td>
</tr>
<tr>
<td>1940</td>
<td></td>
<td>Rhodhiss Dam, N.C.</td>
<td>9.72</td>
</tr>
<tr>
<td>1939</td>
<td></td>
<td>Caesars Head, S.C.</td>
<td>6.47</td>
</tr>
<tr>
<td>1938</td>
<td></td>
<td>Hubbardston, Mass.</td>
<td>10.16</td>
</tr>
<tr>
<td>1934</td>
<td></td>
<td>Balcony Falls, Va.</td>
<td>4.21</td>
</tr>
<tr>
<td>1933</td>
<td></td>
<td>Peekamoose, N.Y.</td>
<td>11.60</td>
</tr>
<tr>
<td>1932</td>
<td></td>
<td>Caesars Head, S.C.</td>
<td>4.75</td>
</tr>
<tr>
<td>1932</td>
<td></td>
<td>Rockhouse, N.C.</td>
<td>6.85</td>
</tr>
<tr>
<td>1929</td>
<td></td>
<td>Rockhouse, N.C.</td>
<td>6.25</td>
</tr>
<tr>
<td>1928</td>
<td></td>
<td>Roanoke, Va.</td>
<td>3.42</td>
</tr>
<tr>
<td>1928</td>
<td></td>
<td>Caesars Head, S.C.</td>
<td>11.80</td>
</tr>
<tr>
<td>1923</td>
<td></td>
<td>Mohonk Lake, N.Y.</td>
<td>0.80</td>
</tr>
<tr>
<td>1923</td>
<td></td>
<td>Wappigners Falls, N.Y.</td>
<td>3.69</td>
</tr>
<tr>
<td>1920</td>
<td></td>
<td>Landrum, S.C.</td>
<td>3.10</td>
</tr>
<tr>
<td>1916</td>
<td></td>
<td>Altapass, N.C.</td>
<td>22.22</td>
</tr>
<tr>
<td>1916</td>
<td></td>
<td>Highlands, N.C.</td>
<td>7.43</td>
</tr>
<tr>
<td>1915</td>
<td></td>
<td>Lookout Mt., Tenn.</td>
<td>5.00</td>
</tr>
<tr>
<td>1915</td>
<td></td>
<td>Highlands, N.C.</td>
<td>4.58</td>
</tr>
<tr>
<td>1912</td>
<td></td>
<td>Norcross, Ga.</td>
<td>4.46</td>
</tr>
<tr>
<td>1906</td>
<td></td>
<td>Horse Cove, N.C.</td>
<td>8.00</td>
</tr>
<tr>
<td>1902</td>
<td></td>
<td>Sewancee, Tenn.</td>
<td>3.73</td>
</tr>
<tr>
<td>1901</td>
<td></td>
<td>Linville, N.C.</td>
<td>3.50</td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td>Marrobone, Ky.</td>
<td>6.20</td>
</tr>
<tr>
<td>1900</td>
<td></td>
<td>St. Johnsbury, Vt.</td>
<td>0.67</td>
</tr>
</tbody>
</table>

(Continued on next page)
FIGURE 5.2.2

Also, according to the figures in Table 5.2.2,

\[
\frac{1}{36} \sum_{i=1}^{36} y_i = 7.29
\]

and

\[
\frac{1}{36} \sum_{i=1}^{36} y_i^2 = 85.59
\]

To find \( r_c \) and \( \lambda_c \), then, we need to solve the two equations

\[
\frac{r}{\lambda} = 7.29
\]

and

\[
\frac{r(r + 1)}{\lambda^2} = 85.59
\]

Substituting \( r = 7.29\lambda \) into the second equation gives

\[
\frac{(7.29\lambda)(7.29\lambda + 1)}{\lambda^2} = 85.59
\]

or \( \lambda_c = 0.22 \). Then, from the first equation, \( r_c = 1.60 \) \( [= 7.29(0.22)] \).

The estimated model,

\[
f_r(y; 1.60, 0.22) = \frac{(0.22)^{1.60} y^{1.60 - 1} e^{-0.22y}}{\Gamma(1.60)}
\]

is superimposed on the data’s density-scaled histogram in Figure 5.2.3. Considering the relatively small number of observations in the sample, the agreement is quite

(Continued on next page)
5.2.15. Let $y_1, y_2, \ldots, y_n$ be a random sample of size $n$ from the uniform pdf, $f_Y(y; \theta) = \frac{1}{\theta}$, $0 \leq y \leq \theta$. Find a formula for the method of moments estimate for $\theta$. Compare the values of the method of moments estimate and the maximum likelihood estimate if a random sample of size 5 consists of the numbers 17, 92, 46, 39, and 56 (recall Question 5.2.9).

5.2.16. Use the method of moments to estimate $\theta$ in the pdf

$$f_Y(y; \theta) = (\theta^2 + \theta)y^{\theta-1}(1 - y), \quad 0 < y < 1$$

Assume that a random sample of size $n$ has been collected.

5.2.17. A criminologist is searching through FBI files to document the prevalence of a rare double-whorl fingerprint. Among six consecutive sets of 100,000 prints scanned by a computer, the numbers of persons having the abnormality are 3, 0, 3, 4, 2, and 1, respectively. Assume that double whorls are Poisson events. Use the method of moments to estimate their occurrence rate, $\lambda$. How would your answer change if $\lambda$ were estimated using the method of maximum likelihood?

5.2.18. Find the method of moments estimate for $\lambda$ if a random sample of size $n$ is taken from the exponential pdf, $f_Y(y; \lambda) = \lambda e^{-\lambda y}$, $y \geq 0$.

5.2.19. Suppose that $Y_1 = 8.3$, $Y_2 = 4.9$, $Y_3 = 2.6$, and $Y_4 = 6.5$ is a random sample of size 4 from the two-parameter uniform pdf,

$$f_Y(y; \theta_1, \theta_2) = \frac{1}{2\theta_2}, \quad \theta_1 - \theta_2 \leq y \leq \theta_1 + \theta_2$$

Use the method of moments to calculate $\theta_{1e}$ and $\theta_{2e}$.
5.2.20. Find a formula for the method of moments estimate for the parameter \( \theta \) in the Pareto pdf,

\[
f_Y(y; \theta) = \theta k^\theta \left(\frac{1}{y}\right)^{\theta + 1}, \quad y \geq k; \quad \theta \geq 1
\]

Assume that \( k \) is known and the data consist of a random sample of size \( n \). Compare your answer to the maximum likelihood estimator found in Question 5.2.12.

5.2.21. Calculate the method of moments estimate for the parameter \( \theta \) in the probability function

\[
p_X(k; \theta) = \theta^k (1 - \theta)^{1-k}, \quad k = 0, 1
\]

if a sample of size 5 is the set of numbers 0, 0, 1, 0, 1.

5.2.22. Find the method of moments estimates for \( \mu \) and \( \sigma^2 \), based on a random sample of size \( n \) drawn from a normal pdf, where \( \mu = E(Y) \) and \( \sigma^2 = \text{Var}(Y) \). Compare your answers with the maximum likelihood estimates derived in Example 5.2.4.

5.2.23. Use the method of moments to derive formulas for estimating the parameters \( r \) and \( p \) in the negative binomial pdf,

\[
p_X(k; r, p) = \binom{k - 1}{r - 1} p^r (1 - p)^{k-r}, \quad k = r, r + 1, \ldots
\]

5.2.24. Bird songs can be characterized by the number of clusters of “syllables” that are strung together in rapid succession. If the last cluster is defined as a “success,” it may be reasonable to treat the number of clusters in a song as a geometric random variable. Does the model \( p_X(k) = (1 - p)^{k-1} p, k = 1, 2, \ldots \) adequately describe the following distribution of 250 song lengths (102)? Begin by finding the method of moments estimate for \( p \). Then calculate the set of “expected” frequencies.

<table>
<thead>
<tr>
<th>No. of Clusters/Song</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>250</td>
</tr>
</tbody>
</table>

5.3 INTERVAL ESTIMATION

Point estimates, no matter how they are determined, share the same fundamental weakness: They provide no indication of their inherent precision. We know, for instance, that \( \hat{\lambda} = \bar{X} \) is both the maximum likelihood and the method of moments estimator for the Poisson parameter, \( \lambda \). But suppose a sample of size six is taken from the probability model \( p_X(k) = e^{-\lambda} \lambda^k / k! \) and we find that \( \hat{\lambda} = 6.8 \). Does it follow that the true \( \lambda \) is likely