and, solving for α and β, we get the following formulas for estimating the two parameters of the gamma distribution:

$$\hat{\alpha} = \frac{(m_1')^2}{m_2' - (m_1')^2} \quad \text{and} \quad \hat{\beta} = \frac{m_2' - (m_1')^2}{m_1'}$$

Since $m_1' = \dfrac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$ and $m_2' = \dfrac{\sum_{i=1}^{n} x_i^2}{n}$, we can write

$$\hat{\alpha} = \frac{n\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \quad \text{and} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n\overline{x}}$$

in terms of the original observations.

In these examples we were concerned with the parameters of a specific population. It is important to note, however, that when the parameters to be estimated are the moments of the population, then the method of moments can be used without any knowledge about the nature, or functional form, of the population.

8 The Method of Maximum Likelihood

In two papers published early in the last century, R. A. Fisher proposed a general method of estimation called the **method of maximum likelihood**. He also demonstrated the advantages of this method by showing that it yields sufficient estimators whenever they exist and that maximum likelihood estimators are asymptotically minimum variance unbiased estimators.

To help to understand the principle on which the method of maximum likelihood is based, suppose that four letters arrive in somebody's morning mail, but unfortunately one of them is misplaced before the recipient has a chance to open it. If, among the remaining three letters, two contain credit-card billings and the other one does not, what might be a good estimate of k, the total number of credit-card billings among the four letters received? Clearly, k must be two or three, and if we assume that each letter had the same chance of being misplaced, we find that the probability of the observed data (two of the three remaining letters contain credit-card billings) is

$$\frac{\binom{2}{2}\binom{2}{1}}{\binom{4}{3}} = \frac{1}{2}$$

for $k = 2$ and

$$\frac{\binom{3}{2}\binom{1}{1}}{\binom{4}{3}} = \frac{3}{4}$$

for $k = 3$. Therefore, if we choose as our estimate of k the value that maximizes the probability of getting the observed data, we obtain $k = 3$. We call this estimate a **maximum likelihood estimate**, and the method by which it was obtained is called the method of maximum likelihood.

Thus, the essential feature of the method of maximum likelihood is that we look at the sample values and then choose as our estimates of the unknown parameters the values for which the probability or probability density of getting the sample values is a maximum. In what follows, we shall limit ourselves to the one-parameter case; but, as we shall see in Example 18, the general idea applies also when there are several unknown parameters. In the discrete case, if the observed sample values are $x_1, x_2, \ldots, x_n$, the probability of getting them is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots, x_n; \theta)$$

which is just the value of the joint probability distribution of the random variables $X_1, X_2, \ldots, X_n$ at $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. Since the sample values have been observed and are therefore fixed numbers, we regard $f(x_1, x_2, \ldots, x_n; \theta)$ as a value of a function of θ, and we refer to this function as the **likelihood function**. An analogous definition applies when the random sample comes from a continuous population, but in that case $f(x_1, x_2, \ldots, x_n; \theta)$ is the value of the joint probability density of the random variables $X_1, X_2, \ldots, X_n$ at $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

DEFINITION 8. MAXIMUM LIKELIHOOD ESTIMATOR. *If* $x_1, x_2, \ldots, x_n$ *are the values of a random sample from a population with the parameter* θ*, the* ***likelihood function*** *of the sample is given by*

$$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta)$$

for values of θ *within a given domain. Here,* $f(x_1, x_2, \ldots, x_n; \theta)$ *is the value of the joint probability distribution or the joint probability density of the random variables* $X_1, X_2, \ldots, X_n$ *at* $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$*. We refer to the value of* θ *that maximizes* $L(\theta)$ *as the* ***maximum likelihood estimator*** *of* θ*.*

EXAMPLE 15

Given x "successes" in n trials, find the maximum likelihood estimate of the parameter θ of the corresponding binomial distribution.

Solution

To find the value of θ that maximizes

$$L(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

it will be convenient to make use of the fact that the value of θ that maximizes $L(\theta)$ will also maximize

$$\ln\ L(\theta) = \ln \binom{n}{x} + x \cdot \ln \theta + (n-x) \cdot \ln(1-\theta)$$

Thus, we get

$$\frac{d[\ln L(\theta)]}{d\theta} = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$

and, equating this derivative to 0 and solving for θ, we find that the likelihood function has a maximum at $\theta = \frac{x}{n}$. This is the maximum likelihood estimate of the binomial parameter θ, and we refer to $\hat{\Theta} = \frac{X}{n}$ as the corresponding **maximum likelihood estimator**.

EXAMPLE 16

If $x_1, x_2, \ldots, x_n$ are the values of a random sample from an exponential population, find the maximum likelihood estimator of its parameter θ.

Solution

Since the likelihood function is given by

$$\begin{aligned} L(\theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\ &= \prod_{i=1}^{n} f(x_i; \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot e^{-\frac{1}{\theta}\left(\sum_{i=1}^{n} x_i\right)} \end{aligned}$$

differentiation of $\ln L(\theta)$ with respect to θ yields

$$\frac{d[\ln L(\theta)]}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \cdot \sum_{i=1}^{n} x_i$$

Equating this derivative to zero and solving for θ, we get the maximum likelihood estimate

$$\hat{\theta} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \overline{x}$$

Hence, the maximum likelihood estimator is $\hat{\Theta} = \overline{X}$.

Now let us consider an example in which straightforward differentiation cannot be used to find the maximum value of the likelihood function.

EXAMPLE 17

If $x_1, x_2, \ldots, x_n$ are the values of a random sample of size n from a uniform population with $\alpha = 0$ (as in Example 4), find the maximum likelihood estimator of β.

Solution

The likelihood function is given by

$$L(\beta) = \prod_{i=1}^{n} f(x_i; \beta) = \left(\frac{1}{\beta}\right)^n$$

for β greater than or equal to the largest of the x's and 0 otherwise. Since the value of this likelihood function increases as β decreases, we must make β as small as possible, and it follows that the maximum likelihood estimator of β is Y_n, the nth order statistic.

Comparing the result of this example with that of Example 4, we find that maximum likelihood estimators need not be unbiased. However, the ones of Examples 15 and 16 were unbiased.

The method of maximum likelihood can also be used for the simultaneous estimation of several parameters of a given population. In that case we must find the values of the parameters that jointly maximize the likelihood function.

EXAMPLE 18

If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a normal population with the mean μ and the variance σ^2, find joint maximum likelihood estimates of these two parameters.

Solution

Since the likelihood function is given by

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

partial differentiation of $\ln L(\mu, \sigma^2)$ with respect to μ and σ^2 yields

$$\frac{\partial[\ln L(\mu, \sigma^2)]}{\partial\mu} = \frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(x_i - \mu)$$

and

$$\frac{\partial[\ln L(\mu, \sigma^2)]}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\cdot\sum_{i=1}^{n}(x_i - \mu)^2$$

Equating the first of these two partial derivatives to zero and solving for μ, we get

$$\hat{\mu} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i = \overline{x}$$

and equating the second of these partial derivatives to zero and solving for σ^2 after substituting $\mu = \overline{x}$, we get

$$\hat{\sigma}^2 = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \overline{x})^2$$

It should be observed that we did not show that $\hat{\sigma}$ is a maximum likelihood estimate of σ, only that $\hat{\sigma}^2$ is a maximum likelihood estimate of σ^2. However, it can be shown (see reference at the end of this chapter) that maximum likelihood estimators have the **invariance property** that if $\hat{\Theta}$ is a maximum likelihood estimator of θ and the function given by $g(\theta)$ is continuous, then $g(\hat{\Theta})$ is also a maximum likelihood estimator of $g(\theta)$. It follows that

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

which differs from s in that we divide by n instead of $n-1$, is a maximum likelihood estimate of σ.

In Examples 15, 16, and 18 we maximized the logarithm of the likelihood function instead of the likelihood function itself, but this is by no means necessary. It just so happened that it was convenient in each case.

Exercises

50. If $X_1, X_2, \ldots, X_n$ constitute a random sample from a population with the mean μ and the variance σ^2, use the method of moments to find estimators for μ and σ^2.

51. Given a random sample of size n from an exponential population, use the method of moments to find an estimator of the parameter θ.

52. Given a random sample of size n from a uniform population with $\alpha = 0$, find an estimator for β by the method of moments.

53. Given a random sample of size n from a Poisson population, use the method of moments to obtain an estimator for the parameter λ.

54. Given a random sample of size n from a beta population with $\beta = 1$, use the method of moments to find a formula for estimating the parameter α.

55. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a population given by

$$f(x;\theta) = \begin{cases} \dfrac{2(\theta - x)}{\theta^2} & \text{for } 0 < x < \theta \\ 0 & \text{elsewhere} \end{cases}$$

find an estimator for θ by the method of moments.

56. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a population given by

$$g(x;\theta,\delta) = \begin{cases} \dfrac{1}{\theta} \cdot e^{-\frac{x-\delta}{\theta}} & \text{for } x > \delta \\ 0 & \text{elsewhere} \end{cases}$$

find estimators for δ and θ by the method of moments. This distribution is sometimes referred to as the **two-parameter exponential distribution**, and for $\theta = 1$ it is the distribution of Example 3.

57. Given a random sample of size n from a continuous uniform population, use the method of moments to find formulas for estimating the parameters α and β.

58. Consider N independent random variables having identical binomial distributions with the parameters θ and $n = 3$. If n_0 of them take on the value 0, n_1 take on the value 1, n_2 take on the value 2, and n_3 take on the value 3, use the method of moments to find a formula for estimating θ.

59. Use the method of maximum likelihood to rework Exercise 53.

60. Use the method of maximum likelihood to rework Exercise 54.

61. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a gamma population with $\alpha = 2$, use the method of maximum likelihood to find a formula for estimating β.

62. Given a random sample of size n from a normal population with the known mean μ, find the maximum likelihood estimator for σ.

63. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a geometric population, find formulas for estimating its parameter θ by using
(a) the method of moments;
(b) the method of maximum likelihood.

64. Given a random sample of size n from a Rayleigh population, find an estimator for its parameter α by the method of maximum likelihood.

65. Given a random sample of size n from a Pareto population, use the method of maximum likelihood to find a formula for estimating its parameter α.

66. Use the method of maximum likelihood to rework Exercise 56.

67. Use the method of maximum likelihood to rework Exercise 57.

68. Use the method of maximum likelihood to rework Exercise 58.

69. Given a random sample of size n from a gamma population with the known parameter α, find the maximum likelihood estimator for
(a) β; **(b)** $\tau = (2\beta - 1)^2$.

70. If $V_1, V_2, \ldots, V_n$ and $W_1, W_2, \ldots, W_n$ are independent random samples of size n from normal populations with the means $\mu_1 = \alpha + \beta$ and $\mu_2 = \alpha - \beta$ and the common variance $\sigma^2 = 1$, find maximum likelihood estimators for α and β.

71. If $V_1, V_2, \ldots, V_{n_1}$ and $W_1, W_2, \ldots, W_{n_2}$ are independent random samples of sizes n_1 and n_2 from normal populations with the means μ_1 and μ_2 and the common variance σ^2, find maximum likelihood estimators for μ_1, μ_2, and σ^2.

72. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the uniform population given by

$$f(x;\theta) = \begin{cases} 1 & \text{for } \theta - \dfrac{1}{2} < x < \theta + \dfrac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

Show that if Y_1 and Y_n are the first and nth order statistic, any estimator $\hat{\Theta}$ such that

$$Y_n - \frac{1}{2} \leqq \hat{\Theta} \leqq Y_1 + \frac{1}{2}$$

can serve as a maximum likelihood estimator of θ. This shows that maximum likelihood estimators need not be unique.

73. With reference to Exercise 72, check whether the following estimators are maximum likelihood estimators of θ:
(a) $\frac{1}{2}(Y_1 + Y_n)$; **(b)** $\frac{1}{3}(Y_1 + 2Y_2)$.

9 Bayesian Estimation†

So far we have assumed in this chapter that the parameters that we want to estimate are unknown constants; in Bayesian estimation the parameters are looked upon as random variables having **prior distributions**, usually reflecting the strength of one's belief about the possible values that they can assume.

The main problem of Bayesian estimation is that of combining prior feelings about a parameter with direct sample evidence, and this is accomplished by determining $\varphi(\theta|x)$, the conditional density of Θ given $X = x$. In contrast to the prior distribution of Θ, this conditional distribution (which also reflects the direct sample evidence) is called the **posterior distribution** of Θ. In general, if $h(\theta)$ is the value of the prior distribution of Θ at θ and we want to combine the information that it conveys with direct sample evidence about Θ, for instance, the value of a statistic $W = u(X_1, X_2, \ldots, X_n)$, we determine the posterior distribution of Θ by means of the formula

$$\varphi(\theta|w) = \frac{f(\theta, w)}{g(w)} = \frac{h(\theta) \cdot f(w|\theta)}{g(w)}$$

Here $f(w|\theta)$ is the value of the sampling distribution of W given $\Theta = \theta$ at w, $f(\theta, w)$ is the value of the joint distribution of Θ and W at θ and w, and $g(w)$ is the value of the marginal distribution of W at w. Note that the preceding formula for $\varphi(\theta|w)$ is, in fact, an extension of Bayes' theorem to the continuous case. Hence, the term "Bayesian estimation."

† This section may be omitted with no loss of continuity.

Once the posterior distribution of a parameter has been obtained, it can be used to make estimates, or it can be used to make probability statements about the parameter, as will be illustrated in Example 20. Although the method we have described has extensive applications, we shall limit our discussion here to inferences about the parameter Θ of a binomial population and the mean of a normal population; inferences about the parameter of a Poisson population are treated in Exercise 77.

THEOREM 5. If X is a binomial random variable and the prior distribution of Θ is a beta distribution with the parameters α and β, then the posterior distribution of Θ given $X = x$ is a beta distribution with the parameters $x+\alpha$ and $n-x+\beta$.

Proof For $\Theta = \theta$ we have

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

$$h(\theta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{for } 0<\theta<1 \\ 0 & \text{elsewhere} \end{cases}$$

and hence

$$\begin{aligned} f(\theta, x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1}\times\binom{n}{x}\theta^x(1-\theta)^{n-x} \\ &= \binom{n}{x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \end{aligned}$$

for $0<\theta<1$ and $x = 0, 1, 2, \ldots, n$, and $f(\theta, x) = 0$ elsewhere. To obtain the marginal density of X, let us make use of the fact that the integral of the beta density from 0 to 1 equals 1; that is,

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Thus, we get

$$g(x) = \binom{n}{x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+x)\cdot\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

for $x = 0, 1, \ldots, n$, and hence

$$\varphi(\theta|x) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\cdot\Gamma(n-x+\beta)}\cdot\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

for $0<\theta<1$, and $\varphi(\theta|x) = 0$ elsewhere. As can be seen by inspection, this is a beta density with the parameters $x+\alpha$ and $n-x+\beta$.

To make use of this theorem, let us refer to the result that (under very general conditions) the mean of the posterior distribution minimizes the Bayes risk when the loss function is quadratic, that is, when the loss function is given by

$$L[d(x), \theta] = c[d(x) - \theta]^2$$

where c is a positive constant. Since the posterior distribution of Θ is a beta distribution with parameters $x + \alpha$ and $n - x + \beta$, it follows from the theorem "The mean and the variance of the beta distribution are given by $\mu = \frac{\alpha}{\alpha+\beta}$ and $\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$" that

$$E(\Theta|x) = \frac{x + \alpha}{\alpha + \beta + n}$$

is a value of an estimator of θ that minimizes the Bayes risk when the loss function is quadratic and the prior distribution of Θ is of the given form.

EXAMPLE 19

Find the mean of the posterior distribution as an estimate of the "true" probability of a success if 42 successes are obtained in 120 binomial trials and the prior distribution of Θ is a beta distribution with $\alpha = \beta = 40$.

Solution

Substituting $x = 42, n = 120, \alpha = 40$, and $\beta = 40$ into the formula for $E(\Theta|x)$, we get

$$E(\Theta|42) = \frac{42 + 40}{40 + 40 + 120} = 0.41$$

Note that without knowledge of the prior distribution of Θ, the minimum variance unbiased estimate of θ (see Exercise 14) would be the sample proportion

$$\hat{\theta} = \frac{x}{n} = \frac{42}{120} = 0.35$$

THEOREM 6. If $\overline{X}$ is the mean of a random sample of size n from a normal population with the known variance σ^2 and the prior distribution of M (capital Greek *mu*) is a normal distribution with the mean μ_0 and the variance σ_0^2, then the posterior distribution of M given $\overline{X} = \overline{x}$ is a normal distribution with the mean μ_1 and the variance σ_1^2, where

$$\mu_1 = \frac{n\overline{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2} \quad \text{and} \quad \frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$$

Proof For M $= \mu$ we have

$$f(\overline{x}|\mu) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}\right)^2} \qquad \text{for } -\infty < \overline{x} < \infty$$

and

$$h(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} \qquad \text{for } -\infty < \mu < \infty$$

so that

$$\begin{aligned}\varphi(\mu|\bar{x}) &= \frac{h(\mu)\cdot f(\bar{x}|\mu)}{g(\bar{x})} \\ &= \frac{\sqrt{n}}{2\pi\sigma\sigma_0 g(\bar{x})} \cdot e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2 - \frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} \qquad \text{for } -\infty < \mu < \infty\end{aligned}$$

Now, if we collect powers of μ in the exponent of e, we get

$$-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2}\right)\mu^2 + \left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\sigma_0^2}\right)\mu - \frac{1}{2}\left(\frac{n\bar{x}^2}{\sigma^2}+\frac{\mu_0^2}{\sigma_0^2}\right)$$

and if we let

$$\frac{1}{\sigma_1^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \quad \text{and} \quad \mu_1 = \frac{n\bar{x}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$$

factor out $-\frac{1}{2\sigma_1^2}$, and complete the square, the exponent of e in the expression for $\varphi(\mu|\bar{x})$ becomes

$$-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2 + R$$

where R involves $n, \bar{x}, \mu_0, \sigma$, and σ_0, but not μ. Thus, the posterior distribution of M becomes

$$\varphi(\mu|\bar{x}) = \frac{\sqrt{n}\cdot e^R}{2\pi\sigma\sigma_0 g(\bar{x})} \cdot e^{-\frac{1}{2\sigma_1^2}(\mu-\mu_1)^2} \qquad \text{for } -\infty < \mu < \infty$$

which is easily identified as a normal distribution with the mean μ_1 and the variance σ_1^2. Hence, it can be written as

$$\varphi(\mu|\bar{x}) = \frac{1}{\sigma_1\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\mu-\mu_1}{\sigma_1}\right)^2} \qquad \text{for } -\infty < \mu < \infty$$

where μ_1 and σ_1 are defined above. Note that we did not have to determine $g(\bar{x})$ as it was absorbed in the constant in the final result.

EXAMPLE 20

A distributor of soft-drink vending machines feels that in a supermarket one of his machines will sell on the average $\mu_0 = 738$ drinks per week. Of course, the mean will vary somewhat from market to market, and the distributor feels that this variation is measured by the standard deviation $\sigma_0 = 13.4$. As far as a machine placed in a particular market is concerned, the number of drinks sold will vary from week to

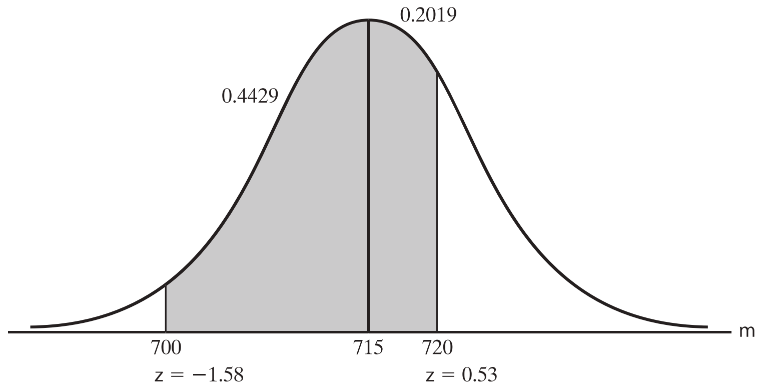

Figure 1. Diagram for Example 20.

week, and this variation is measured by the standard deviation $\sigma = 42.5$. If one of the distributor's machines put into a new supermarket averaged $\overline{x} = 692$ during the first 10 weeks, what is the probability (the distributor's personal probability) that for this market the value of M is actually between 700 and 720?

Solution

Assuming that the population sampled is approximately normal and that it is reasonable to treat the prior distribution of M as a normal distribution with the mean μ_0 and the standard deviation $\sigma_0 = 13.4$, we find that substitution into the two formulas of Theorem 6 yields

$$\mu_1 = \frac{10.692(13.4)^2 + 738(42.5)^2}{10(13.4)^2 + (42.5)^2} = 715$$

and

$$\frac{1}{\sigma_1^2} = \frac{10}{(42.5)^2} + \frac{1}{(13.4)^2} = 0.0111$$

so that $\sigma_1^2 = 90.0$ and $\sigma_1 = 9.5$. Now, the answer to our question is given by the area of the shaded region of Figure 1, that is, the area under the standard normal curve between

$$z = \frac{700 - 715}{9.5} = -1.58 \quad \text{and} \quad z = \frac{720 - 715}{9.5} = 0.53$$

Thus, the probability that the value of M is between 700 and 720 is $0.4429 + 0.2019 = 0.6448$, or approximately 0.645.

Exercises

74. This question has been intentionally omitted for this edition.

75. This question has been intentionally omitted for this edition.

76. Show that the mean of the posterior distribution of M given in Theorem 6 can be written as

$$\mu_1 = w \cdot \overline{x} + (1 - w) \cdot \mu_0$$

that is, as a weighted mean of $\overline{x}$ and μ_0, where

$$w = \frac{n}{n + \frac{\sigma^2}{\sigma_0^2}}$$

77. If X has a Poisson distribution and the prior distribution of its parameter Λ (capital Greek *lambda*) is a gamma distribution with the parameters α and β, show that

(a) the posterior distribution of Λ given $X = x$ is a gamma distribution with the parameters $\alpha + x$ and $\frac{\beta}{\beta+1}$;

(b) the mean of the posterior distribution of Λ is

$$\mu_1 = \frac{\beta(\alpha + x)}{\beta + 1}$$

10 The Theory in Practice

The sample mean, $\overline{x}$, is most frequently used to estimate the mean of a distribution from a random sample taken from that distribution. It has been shown to be the minimum variance unbiased estimator as well as a sufficient estimator for the mean of a normal distribution. It is at least asymptotically unbiased as an estimator for the mean of most frequently encountered distributions.

In spite of these desirable properties of the sample mean as an estimator for a population mean, we know that the sample mean will never equal the population mean. Let us examine the error we make when using $\overline{x}$ to estimate $\mu, E = |\overline{x} - \mu|$. If the sample size, n, is large, the quantity

$$\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

is a value of a random variable having approximately the standard normal distribution. Thus, we can state with probability $1 - \alpha$ that

$$\frac{|\overline{x} - \mu|}{\sigma/\sqrt{n}} \le z_{\alpha/2}$$

or

$$E \le z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

EXAMPLE 21

A pollster wishes to estimate the percentage of voters who favor a certain candidate. She wishes to be sure with a probability of 0.95 that the error in the resulting estimate will not exceed 3 percent. How many registered voters should she interview?

Solution

We shall use the normal approximation to the binomial distribution, assuming that n will turn out to be large. As per the theorem "If X has a binomial distribution with the parameters n and θ and $Y = \frac{X}{n}$, then $E(Y) = \theta$ and $\sigma_Y^2 = \frac{\theta(1-\theta)}{n}$" we know that $\sigma_{X/n}^2 = \frac{\theta(1-\theta)}{n}$, where θ is the parameter of the binomial distribution. Since this quantity is maximized when $\theta = \frac{1}{2}$, the maximum value of σ is $\frac{1}{2\sqrt{n}}$. Since the maximum error is to be 0.03, the inequality for E can be written as

$$E \le z_{\alpha/2} \cdot \frac{1}{2\sqrt{n}}$$

Noting that $z_{\alpha/2} = z_{.025} = 1.96$, and solving this inequality for n, we obtain for the sample size that assures, with probability 0.95, that the error in the resulting estimate will not exceed 3 percent

$$n \leq \frac{z^2\alpha/2}{4E^2} = \frac{(1.96)^2}{4(.03)^2} = 1{,}068$$

(Note that, in performing this calculation, we always round *up* to the nearest integer.)

It should not be surprising in view of this result that most such polls use sample sizes of about 1,000.

Another consideration related to the accuracy of a sample estimate deals with the concept of **sampling bias**. Sampling bias occurs when a sample is chosen that does not accurately represent the population from which it is taken. For example, a national poll based on automobile registrations in each of the 50 states probably is biased, because it omits people who do not own cars. Such people may well have different opinions than those who do. A sample of product stored on shelves in a warehouse is likely to be biased if all units of product in the sample were selected from the bottom shelf. Ambient conditions, such as temperature and humidity, may well have affected the top-shelf units differently than those on the bottom shelf.

The mean square error defined above can be viewed as the expected squared error loss encountered when we estimate the parameter θ with the point estimator $\hat{\Theta}$. We can write

$$\begin{aligned}
\text{MSE}(\hat{\Theta}) &= E(\hat{\Theta} - \theta)^2 \\
&= E[\hat{\Theta} - E(\hat{\Theta}) + E(\hat{\Theta}) - \theta]^2 \\
&= E[\hat{\Theta} - E(\hat{\Theta})]^2 + [E(\hat{\Theta}) - \theta]^2 + 2\{E[\hat{\Theta} - E(\hat{\Theta})][E(\hat{\Theta}) - \theta]\}
\end{aligned}$$

The first term of the cross product, $E[\hat{\Theta} - E(\hat{\Theta})] = E(\hat{\Theta}) - E(\hat{\Theta}) = 0$, and we are left with

$$\text{MSE}(\hat{\Theta}) = E[\hat{\Theta} - E(\hat{\Theta})]^2 + [E(\hat{\Theta}) - \theta]^2$$

The first term is readily seen to be the variance of $\hat{\Theta}$ and the second term is the square of the bias, the difference between the expected value of the estimate of the parameter θ and its true value. Thus, we can write

$$\text{MSE}(\hat{\Theta}) = \sigma^2_{\hat{\Theta}} + [\text{Bias}]^2$$

While it is possible to estimate the variance of $\hat{\Theta}$ in most applications, the sampling bias usually is unknown. Great care should be taken to avoid, or at least minimize sampling bias, for it can be much greater than the sampling variance $\sigma^2_{\hat{\Theta}}$. This can be done by carefully calibrating all instruments to be used in measuring the sample units, by eliminating human subjectivity as much as possible, and by assuring that the method of sampling is appropriately randomized over the entire population for which sampling estimates are to be made. These and other related issues are more thoroughly discussed in the book by Hogg and Tanis, referenced at the end of this chapter.

Applied Exercises

SECS. 1–3

78. Independent random samples of sizes n_1 and n_2 are taken from a normal population with the mean μ and the variance σ^2. If $n_1 = 25, n_2 = 50$, $\bar{x}_1 = 27.6$, and $\bar{x}_2 = 38.1$, estimate μ using the estimator of Exercise 23.

79. Random samples of size n are taken from normal populations with the mean μ and the variances $\sigma_1^2 = 4$ and $\sigma_2^2 = 9$. If $\bar{x}_1 = 26.0$ and $\bar{x}_2 = 32.5$, estimate μ using the estimator of part (b) of Exercise 21.

80. A country's military intelligence knows that an enemy built certain new tanks numbered serially from 1 to k. If three of these tanks are captured and their serial numbers are 210, 38, and 155, use the estimator of part (b) of Exercise 12 to estimate k.

SECS. 4–8

81. On 12 days selected at random, a city's consumption of electricity was 6.4, 4.5, 10.8, 7.2, 6.8, 4.9, 3.5, 16.3, 4.8, 7.0, 8.8, and 5.4 million kilowatt-hours. Assuming that these data may be looked upon as a random sample from a gamma population, use the estimators obtained in Example 14 to estimate the parameters α and β.

82. Certain radial tires had useful lives of 35,200, 41,000, 44,700, 38,600, and 41,500 miles. Assuming that these data can be looked upon as a random sample from an exponential population, use the estimator obtained in Exercise 51 to estimate the parameter θ.

83. The size of an animal population is sometimes estimated by the **capture–recapture method**. In this method, n_1 of the animals are captured in the area under consideration, tagged, and released. Later, n_2 of the animals are captured, X of them are found to be tagged, and this information is used to estimate N, the total number of animals of the given kind in the area under consideration. If $n_1 = 3$ rare owls are captured in a section of a forest, tagged, and released, and later $n_2 = 4$ such owls are captured and only one of them is found to be tagged, estimate N by the method of maximum likelihood. (*Hint*: Try $N = 9, 10, 11, 12, 13$, and 14.)

84. Among six measurements of the boiling point of a silicon compound, the size of the error was 0.07, 0.03, 0.14, 0.04, 0.08, and 0.03°C. Assuming that these data can be looked upon as a random sample from the population of Exercise 55, use the estimator obtained there by the method of moments to estimate the parameter θ.

85. Not counting the ones that failed immediately, certain light bulbs had useful lives of 415, 433, 489, 531, 466, 410, 479, 403, 562, 422, 475, and 439 hours. Assuming that these data can be looked upon as a random sample from a two-parameter exponential population, use the estimators obtained in Exercise 56 to estimate the parameters δ and θ.

86. Rework Exercise 85 using the estimators obtained in Exercise 66 by the method of maximum likelihood.

87. Data collected over a number of years show that when a broker called a random sample of eight of her clients, she got a busy signal 6.5, 10.6, 8.1, 4.1, 9.3, 11.5, 7.3, and 5.7 percent of the time. Assuming that these figures can be looked upon as a random sample from a continuous uniform population, use the estimators obtained in Exercise 57 to estimate the parameters α and β.

88. Rework Exercise 87 using the estimators obtained in Exercise 67.

89. In a random sample of the teachers in a large school district, their annual salaries were \$23,900, \$21,500, \$26,400, \$24,800, \$33,600, \$24,500, \$29,200, \$36,200, \$22,400, \$21,500, \$28,300, \$26,800, \$31,400, \$22,700, and \$23,100. Assuming that these data can be looked upon as a random sample from a Pareto population, use the estimator obtained in Exercise 65 to estimate the parameter α.

90. Every time Mr. Jones goes to the race track he bets on three races. In a random sample of 20 of his visits to the race track, he lost all his bets 11 times, won once 7 times, and won twice on 2 occasions. If θ is the probability that he will win any one of his bets, estimate it by using the maximum likelihood estimator obtained in Exercise 68.

91. On 20 very cold days, a farmer got her tractor started on the first, third, fifth, first, second, first, third, seventh, second, fourth, fourth, eighth, first, third, sixth, fifth, second, first, sixth, and second try. Assuming that these data can be looked upon as a random sample from a geometric population, estimate its parameter θ by either of the methods of Exercise 63.

92. The I.Q.'s of 10 teenagers belonging to one ethnic group are 98, 114, 105, 101, 123, 117, 106, 92, 110, and 108, whereas those of 6 teenagers belonging to another ethnic group are 122, 105, 99, 126, 114, and 108. Assuming that these data can be looked upon as independent random samples from normal populations with the means μ_1 and μ_2 and the common variance σ^2, estimate these parameters by means of the maximum likelihood estimators obtained in Exercise 71.

SEC. 9

93. The output of a certain integrated-circuit production line is checked daily by inspecting a sample of 100 units. Over a long period of time, the process has maintained a yield of 80 percent, that is, a proportion defective of 20 percent, and the variation of the proportion defective from day to day is measured by a standard deviation of 0.04. If on a certain day the sample contains

38 defectives, find the mean of the posterior distribution of Θ as an estimate of that day's proportion defective. Assume that the prior distribution of Θ is a beta distribution.

94. Records of a university (collected over many years) show that on the average 74 percent of all incoming freshmen have I.Q.'s of at least 115. Of course, the percentage varies somewhat from year to year, and this variation is measured by a standard deviation of 3 percent. If a sample check of 30 freshmen entering the university in 2003 showed that only 18 of them have I.Q.'s of at least 115, estimate the true proportion of students with I.Q.'s of at least 115 in that freshman class using

(a) only the prior information;

(b) only the direct information;

(c) the result of Exercise 74 to combine the prior information with the direct information.

95. With reference to Example 20, find $P(712 < M < 725|\bar{x} = 692)$.

96. A history professor is making up a final examination that is to be given to a very large group of students. His feelings about the average grade that they should get is expressed subjectively by a normal distribution with the mean $\mu_0 = 65.2$ and the standard deviation $\sigma_0 = 1.5$.

(a) What prior probability does the professor assign to the actual average grade being somewhere on the interval from 63.0 to 68.0?

(b) What posterior probability would he assign to this event if the examination is tried on a random sample of 40 students whose grades have a mean of 72.9 and a standard deviation of 7.4? Use $s = 7.4$ as an estimate of σ.

97. An office manager feels that for a certain kind of business the daily number of incoming telephone calls is a random variable having a Poisson distribution, whose parameter has a prior gamma distribution with $\alpha = 50$ and $\beta = 2$. Being told that one such business had 112 incoming calls on a given day, what would be her estimate of that particular business's average daily number of incoming calls if she considers

(a) only the prior information;

(b) only the direct information;

(c) both kinds of information and the theory of Exercise 77?

SEC. 10

98. How large a random sample is required from a population whose standard deviation is 4.2 so that the sample estimate of the mean will have an error of at most 0.5 with a probability of 0.99?

99. A random sample of 36 resistors is taken from a production line manufacturing resistors to a specification of 40 ohms. Assuming a standard deviation of 1 ohm, is this sample adequate to ensure, with 95 percent probability, that the sample mean will be within 1.5 ohms of the mean of the population of resistors being produced?

100. Sections of sheet metal of various lengths are lined up on a conveyor belt that moves at a constant speed. A sample of these sections is taken for inspection by taking whatever section is passing in front of the inspection station at each five-minute interval. If the purpose is to estimate the number of defects per section in the population of all such manufactured sections, explain how this sampling procedure could be biased.

101. Comment on the sampling bias (if any) of a poll taken by asking how people will vote in an election if the sample is confined to the person claiming to be the head of household.

References

Various properties of sufficient estimators are discussed in

LEHMANN, E. L., *Theory of Point Estimation*. New York: John Wiley & Sons, Inc., 1983,

WILKS, S. S., *Mathematical Statistics*. New York: John Wiley & Sons, Inc., 1962,

and a proof of Theorem 4 may be found in

HOGG, R. V., and TANIS, E. A., *Probability and Statistical Inference*, 6th ed. Upper Saddle River, N.J.: Prentice Hall, 1995.

Important properties of maximum likelihood estimators are discussed in

KEEPING, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962,

and a derivation of the Cramér–Rao inequality, as well as the most general conditions under which it applies, may be found in

RAO, C. R., *Advanced Statistical Methods in Biometric Research*. New York: John Wiley & Sons, Inc., 1952.

Answers to Odd-Numbered Exercises

1 $\sum_{i=1}^{n} a_i = 1$.

9 $(n+1)Y_1$.

25 $\frac{8}{9}$.

29 **(a)** $\frac{3}{4}$; **(b)** $\frac{3}{5}$.

37 Yes.

45 Yes.

51 $\hat{\theta} = m_1'$.

53 $\hat{\lambda} = m_1'$.

55 $\hat{\theta} = 3m_1'$.

57 $\hat{\beta} = m_1' + \sqrt{3[m_2' - (m_1')^2]}$.

59 $\hat{\lambda} = \bar{x}$.

61 $\hat{\beta} = \frac{\bar{x}}{2}$.

63 **(a)** $\hat{\theta} = \dfrac{1}{\bar{x}}$; **(b)** $\hat{\theta} = \dfrac{1}{\bar{x}}$.

65 $\hat{\alpha} = \dfrac{n}{\sum_{i=1}^{n} \ln x_i}$.

67 $\hat{\alpha} = y_1, \hat{\beta} = y_n$.

69 **(a)** $\hat{\beta} = \dfrac{\bar{x}}{\alpha}$; $\hat{\tau} = \left(\dfrac{2\bar{x}}{\alpha} - 1\right)^2$.

71 $\mu_1' = \bar{v}$; $\mu_2' = \bar{v}, \hat{\sigma}^2 = \dfrac{\sum(v-\bar{v})^2 + \sum(w-\bar{w})^2}{n_1+n_2}$.

73 **(a)** Yes; **(b)** No.

75 $\mu = \frac{1}{2}$; $\sigma^2 = \frac{1}{18}$; symmetrical about $x = \frac{1}{2}$.

79 $\hat{\mu} = 28$.

81 $\hat{\alpha} = 4.627$ and $\hat{\beta} = 1.556$.

83 $N = 11$ or 12.

85 $\hat{\theta} = 47.69$ and $\hat{\delta} = 412.64$.

87 $\hat{\alpha} = 3.83$ and $\hat{\beta} = 11.95$.

91 $\hat{\theta} = 0.30$.

93 $E(\Theta|38) = 0.29$.

95 0.4786.

97 **(a)** $\hat{\mu} = 100$; **(b)** $\hat{\mu} = 112$; **(c)** $\hat{\mu} = 108$.

99 Yes.

This page intentionally left blank