balance between the risks of type I and type II errors, but both α and β remain disconcertingly large. *How can we reduce both α and β?* The answer is intuitively clear: Shed more light on the true nature of the population by increasing the sample size. For almost all statistical tests, if α is fixed at some acceptably small value, β decreases as the sample size increases. ■

In this section, we have defined the essential elements of any statistical test. We have seen that two possible types of error can be made when testing hypotheses: type I and type II errors. The probabilities of these errors serve as criteria for evaluating a testing procedure. In the next few sections, we will use the sampling distributions derived in Chapter 7 to develop methods for testing hypotheses about parameters of frequent practical interest.

Exercises

10.1 Define α and β for a statistical test of hypotheses.

10.2 An experimenter has prepared a drug dosage level that she claims will induce sleep for 80% of people suffering from insomnia. After examining the dosage, we feel that her claims regarding the effectiveness of the dosage are inflated. In an attempt to disprove her claim, we administer her prescribed dosage to 20 insomniacs and we observe Y, the number for whom the drug dose induces sleep. We wish to test the hypothesis $H_0: p = .8$ versus the alternative, $H_a: p < .8$. Assume that the rejection region $\{y \leq 12\}$ is used.

a In terms of this problem, what is a type I error?

b Find α.

c In terms of this problem, what is a type II error?

d Find β when $p = .6$.

e Find β when $p = .4$.

10.3 Refer to Exercise 10.2.

a Find the rejection region of the form $\{y \leq c\}$ so that $\alpha \approx .01$.

b For the rejection region in part (a), find β when $p = .6$.

c For the rejection region in part (a), find β when $p = .4$.

10.4 Suppose that we wish to test the null hypothesis H_0 that the proportion p of ledger sheets with errors is equal to .05 versus the alternative H_a, that the proportion is larger than .05, by using the following scheme. Two ledger sheets are selected at random. If both are error free, we reject H_0. If one or more contains an error, we look at a third sheet. If the third sheet is error free, we reject H_0. In all other cases, we accept H_0.

a In terms of this problem, what is a type I error?

b What is the value of α associated with this test?

c In terms of this problem, what is a type II error?

d Calculate $\beta = P(\text{type II error})$ as a function of p.

females and Y_2 the number of mature males in the sample. If the population contains proportions p_1 and p_2 of mature females and males, respectively (with $p_1 + p_2 < 1$), find expressions for

$$E\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right) \quad \text{and} \quad V\left(\frac{Y_1}{n} - \frac{Y_2}{n}\right).$$

5.118 The total sustained load on the concrete footing of a planned building is the sum of the dead load plus the occupancy load. Suppose that the dead load X_1 has a gamma distribution with $\alpha_1 = 50$ and $\beta_1 = 2$, whereas the occupancy load X_2 has a gamma distribution with $\alpha_2 = 20$ and $\beta_2 = 2$. (Units are in kips.) Assume that X_1 and X_2 are independent.

a Find the mean and variance of the total sustained load on the footing.

b Find a value for the sustained load that will be exceeded with probability less than 1/16.

5.9 The Multinomial Probability Distribution

Recall from Chapter 3 that a binomial random variable results from an experiment consisting of n trials with two possible outcomes per trial. Frequently we encounter similar situations in which the number of possible outcomes per trial is more than two. For example, experiments that involve blood typing typically have at least four possible outcomes per trial. Experiments that involve sampling for defectives may categorize the type of defects observed into more than two classes.

A multinomial experiment is a generalization of the binomial experiment.

DEFINITION 5.11 A *multinomial experiment* possesses the following properties:

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcome of a single trial falls into cell i, is p_i, $i = 1, 2, \ldots, k$ and remains the same from trial to trial. Notice that $p_1 + p_2 + p_3 + \cdots + p_k = 1$.
4. The trials are independent.
5. The random variables of interest are $Y_1, Y_2, \ldots, Y_k$, where Y_i equals the number of trials for which the outcome falls into cell i. Notice that $Y_1 + Y_2 + Y_3 + \cdots + Y_k = n$.

The joint probability function for $Y_1, Y_2, \ldots, Y_k$ is given by

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k},$$

where

$$\sum_{i=1}^{k} p_i = 1 \quad \text{and} \quad \sum_{i=1}^{k} y_i = n.$$

Finding the probability that the n trials in a multinomial experiment result in $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k)$ is an excellent application of the probabilistic methods of Chapter 2. We leave this problem as an exercise.

DEFINITION 5.12 Assume that $p_1, p_2, \ldots, p_k$ are such that $\sum_{i=1}^{k} p_i = 1$, and $p_i > 0$ for $i = 1, 2, \ldots, k$. The random variables $Y_1, Y_2, \ldots, Y_k$, are said to have a *multinomial distribution* with parameters n and $p_1, p_2, \ldots, p_k$ if the joint probability function of $Y_1, Y_2, \ldots, Y_k$ is given by

$$p(y_1, y_2, \ldots, y_k) = \frac{n!}{y_1! y_2! \cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k},$$

where, for each i, $y_i = 0, 1, 2, \ldots, n$ and $\sum_{i=1}^{k} y_i = n$.

Many experiments involving classification are multinomial experiments. For example, classifying people into five income brackets results in an enumeration or count corresponding to each of five income classes. Or we might be interested in studying the reaction of mice to a particular stimulus in a psychological experiment. If the mice can react in one of three ways when the stimulus is applied, the experiment yields the number of mice falling into each reaction class. Similarly, a traffic study might require a count and classification of the types of motor vehicles using a section of highway. An industrial process might manufacture items that fall into one of three quality classes: acceptable, seconds, and rejects. A student of the arts might classify paintings into one of k categories according to style and period, or we might wish to classify philosophical ideas of authors in a study of literature. The result of an advertising campaign might yield count data indicating a classification of consumer reactions. Many observations in the physical sciences are not amenable to measurement on a continuous scale and hence result in enumerative data that correspond to the numbers of observations falling into various classes.

Notice that the binomial experiment is a special case of the multinomial experiment (when there are $k = 2$ classes).

EXAMPLE 5.30 According to recent census figures, the proportions of adults (persons over 18 years of age) in the United States associated with five age categories are as given in the following table.

Age	Proportion
18–24	.18
25–34	.23
35–44	.16
45–64	.27
65↑	.16

If these figures are accurate and five adults are randomly sampled, find the probability that the sample contains one person between the ages of 18 and 24, two between the ages of 25 and 34, and two between the ages of 45 and 64.

Solution We will number the five age classes 1, 2, 3, 4, and 5 from top to bottom and will assume that the proportions given are the probabilities associated with each of the classes. Then we wish to find

$$p(y_1, y_2, y_3, y_4, y_5) = \frac{n!}{y_1!\, y_2!\, y_3!\, y_4!\, y_5!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} p_5^{y_5},$$

for $n = 5$ and $y_1 = 1$, $y_2 = 2$, $y_3 = 0$, $y_4 = 2$, and $y_5 = 0$. Substituting these values into the formula for the joint probability function, we obtain

$$\begin{aligned} p(1, 2, 0, 2, 0) &= \frac{5!}{1!\,2!\,0!\,2!\,0!}(.18)^1(.23)^2(.16)^0(.27)^2(.16)^0 \\ &= 30(.18)(.23)^2(.27)^2 = .0208. \end{aligned}$$

■

THEOREM 5.13 If $Y_1, Y_2, \ldots, Y_k$ have a multinomial distribution with parameters n and $p_1, p_2, \ldots, p_k$, then

1. $E(Y_i) = np_i$, $V(Y_i) = np_iq_i$.
2. $\text{Cov}(Y_s, Y_t) = -np_sp_t$, if $s \neq t$.

Proof The marginal distribution of Y_i can be used to derive the mean and variance. Recall that Y_i may be interpreted as the number of trials falling into cell i. Imagine all of the cells, excluding cell i, combined into a single large cell. Then every trial will result in cell i or in a cell other than cell i, with probabilities p_i and $1 - p_i$, respectively. Thus, Y_i possesses a binomial marginal probability distribution. Consequently,

$$E(Y_i) = np_i \quad \text{and} \quad V(Y_i) = np_iq_i, \qquad \text{where } q_i = 1 - p_i.$$

The same results can be obtained by setting up the expectations and evaluating. For example,

$$E(Y_1) = \sum_{y_1}\sum_{y_2} \cdots \sum_{y_k} y_1 \frac{n!}{y_1!y_2!\cdots y_k!} p_1^{y_1} p_2^{y_2} \cdots p_k^{y_k}.$$

Because we have already derived the expected value and variance of Y_i, we leave the summation of this expectation to the interested reader.

The proof of part 2 uses Theorem 5.12. Think of the multinomial experiment as a sequence of n independent trials and define, for $s \neq t$,

$$U_i = \begin{cases} 1, & \text{if trial } i \text{ results in class } s, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$W_i = \begin{cases} 1, & \text{if trial } i \text{ results in class } t, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$Y_s = \sum_{i=1}^{n} U_i \quad \text{and} \quad Y_t = \sum_{j=1}^{n} W_j.$$

(Because $U_i = 1$ or 0 depending upon whether the ith trial resulted in class s, Y_s is simply the sum of a series of 0s and 1s. A 1 occurs in the sum everytime we observe an item from class s, and a 0 occurs everytime we observe any other class. Thus, Y_s is simply the number of times class s is observed. A similar interpretation applies to Y_t.)

Notice that U_i and W_i cannot both equal 1 (the ith item cannot simultaneously be in classes s and t). Thus, the product $U_i W_i$ always equals zero, and $E(U_i W_i) = 0$. The following results allow us to evaluate $\text{Cov}(Y_s, Y_t)$:

$$E(U_i) = p_s$$

$$E(W_j) = p_t$$

$$\text{Cov}(U_i, W_j) = 0, \quad \text{if } i \neq j \text{ because the trials are independent}$$

$$\text{Cov}(U_i, W_i) = E(U_i W_i) - E(U_i)E(W_i) = 0 - p_s p_t$$

From Theorem 5.12, we then have

$$\begin{aligned}\text{Cov}(Y_s, Y_t) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(U_i, W_j) \\ &= \sum_{i=1}^{n} \text{Cov}(U_i, W_i) + \sum\sum_{i \neq j} \text{Cov}(U_i, W_j) \\ &= \sum_{i=1}^{n} (-p_s p_t) + \sum\sum_{i \neq j} 0 = -n p_s p_t.\end{aligned}$$

The covariance here is negative, which is to be expected because a large number of outcomes in cell s would force the number in cell t to be small.

Inferential problems associated with the multinomial experiment will be discussed later.

Exercises

5.119 A learning experiment requires a rat to run a maze (a network of pathways) until it locates one of three possible exits. Exit 1 presents a reward of food, but exits 2 and 3 do not. (If the rat eventually selects exit 1 almost every time, learning may have taken place.) Let Y_i denote the number of times exit i is chosen in successive runnings. For the following, assume that the rat chooses an exit at random on each run.

a Find the probability that $n = 6$ runs result in $Y_1 = 3$, $Y_2 = 1$, and $Y_3 = 2$.

b For general n, find $E(Y_1)$ and $V(Y_1)$.

c Find $\text{Cov}(Y_2, Y_3)$ for general n.

d To check for the rat's preference between exits 2 and 3, we may look at $Y_2 - Y_3$. Find $E(Y_2 - Y_3)$ and $V(Y_2 - Y_3)$ for general n.

5.120 A sample of size n is selected from a large lot of items in which a proportion p_1 contains exactly one defect and a proportion p_2 contains more than one defect (with $p_1 + p_2 < 1$). The cost of repairing the defective items in the sample is $C = Y_1 + 3Y_2$, where Y_1 denotes the number of

***9.93** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3}, & \theta < y < \infty, \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.53, you showed that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

a Find the MLE for θ. [*Hint:* See Example 9.16.]

b Find a function of the MLE in part (a) that is a pivotal quantity.

c Use the pivotal quantity from part (b) to find a $100(1-\alpha)\%$ confidence interval for θ.

***9.94** Suppose that $\hat{\theta}$ is the MLE for a parameter θ. Let $t(\theta)$ be a function of θ that possesses a unique inverse [that is, if $\beta = t(\theta)$, then $\theta = t^{-1}(\beta)$]. Show that $t(\hat{\theta})$ is the MLE of $t(\theta)$.

***9.95** A random sample of n items is selected from the large number of items produced by a certain production line in one day. Find the MLE of the ratio R, the proportion of defective items divided by the proportion of good items.

HW8 **9.96** Consider a random sample of size n from a normal population with mean μ and variance σ^2, both unknown. Derive the MLE of σ.

9.97 The geometric probability mass function is given by

HW8

$$p(y \mid p) = p(1-p)^{y-1}, \qquad y = 1, 2, 3, \ldots.$$

A random sample of size n is taken from a population with a geometric distribution.

a Find the method-of-moments estimator for p.

b Find the MLE for p.

9.8 Some Large-Sample Properties of Maximum-Likelihood Estimators (Optional)

Maximum-likelihood estimators also have interesting large-sample properties. Suppose that $t(\theta)$ is a differentiable function of θ. In Section 9.7, we argued by the invariance property that if $\hat{\theta}$ is the MLE of θ, then the MLE of $t(\theta)$ is given by $t(\hat{\theta})$. Under some conditions of regularity that hold for the distributions that we will consider, $t(\hat{\theta})$ is a *consistent* estimator for $t(\theta)$. In addition, for large sample sizes,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\left[\dfrac{\partial t(\theta)}{\partial \theta}\right]^2 \Bigg/ nE\left[-\dfrac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]}}$$

has approximately a standard normal distribution. In this expression, the quantity $f(Y \mid \theta)$ in the denominator is the density function corresponding to the continuous distribution of interest, evaluated at the random value Y. In the discrete case, the analogous result holds with the probability function evaluated at the random value Y, $p(Y \mid \theta)$ substituted for the density $f(Y \mid \theta)$. If we desire a confidence interval for $t(\theta)$, we can use quantity Z as a pivotal quantity. If we proceed as in Section 8.6, we obtain the following approximate large-sample $100(1-\alpha)\%$ confidence interval for $t(\theta)$:

$$t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]}$$

$$\approx t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left(\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]\right)\Bigg|_{\theta=\hat{\theta}}}.$$

We illustrate this with the following example.

EXAMPLE 9.18 For random variable with a Bernoulli distribution, $p(y \mid p) = p^y(1-p)^{1-y}$, for $y = 0, 1$. If $Y_1, Y_2, \dots, Y_n$ denote a random sample of size n from this distribution, derive a $100(1-\alpha)\%$ confidence interval for $p(1-p)$, the variance associated with this distribution.

Solution As in Example 9.14, the MLE of the parameter p is given by $\hat{p} = W/n$ where $W = \sum_{i=1}^{n} Y_i$. It follows that the MLE for $t(p) = p(1-p)$ is $\widehat{t(p)} = \hat{p}(1-\hat{p})$. In this case,

$$t(p) = p(1-p) = p - p^2 \quad \text{and} \quad \frac{\partial t(p)}{\partial p} = 1 - 2p.$$

Also,

$$p(y \mid p) = p^y(1-p)^{1-y}$$

$$\ln[p(y \mid p)] = y(\ln p) + (1-y)\ln(1-p)$$

$$\frac{\partial \ln[p(y \mid p)]}{\partial p} = \frac{y}{p} - \frac{1-y}{1-p}$$

$$\frac{\partial^2 \ln[p(y \mid p)]}{\partial p^2} = -\frac{y}{p^2} - \frac{1-y}{(1-p)^2}$$

$$E\left\{-\frac{\partial^2 \ln[p(Y \mid p)]}{\partial p^2}\right\} = E\left[\frac{Y}{p^2} + \frac{1-Y}{(1-p)^2}\right]$$

$$= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

Substituting into the earlier formula for the confidence interval for $t(\theta)$, we obtain

$$t(\hat{p}) \pm z_{\alpha/2}\sqrt{\left\{\left[\frac{\partial t(p)}{\partial p}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln p(Y \mid p)}{\partial p^2}\right]\right\}\Bigg|_{p=\hat{p}}}$$

$$= \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\left\{(1-2p)^2 \Bigg/ n\left[\frac{1}{p(1-p)}\right]\right\}\Bigg|_{p=\hat{p}}}$$

$$= \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})(1-2\hat{p})^2}{n}}$$

as the desired confidence interval for $p(1-p)$. ■

Exercises

***9.98** Refer to Exercise 9.97. What is the approximate variance of the MLE?

***9.99** Consider the distribution discussed in Example 9.18. Use the method presented in Section 9.8 to derive a $100(1-\alpha)\%$ confidence interval for $t(p) = p$. Is the resulting interval familiar to you?

***9.100** Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample of size n from an exponential distribution with mean θ. Find a $100(1-\alpha)\%$ confidence interval for $t(\theta) = \theta^2$.

***9.101** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a Poisson distribution with mean λ. Find a $100(1-\alpha)\%$ confidence interval for $t(\lambda) = e^{-\lambda} = P(Y = 0)$.

***9.102** Refer to Exercises 9.97 and 9.98. If a sample of size 30 yields $\overline{y} = 4.4$, find a 95% confidence interval for p.

9.9 Summary

In this chapter, we continued and extended the discussion of estimation begun in Chapter 8. Good estimators are consistent and efficient when compared to other estimators. The most efficient estimators, those with the smallest variances, are functions of the sufficient statistics that best summarize all of the information about the parameter of interest.

Two methods of finding estimators—the method of moments and the method of maximum likelihood—were presented. Moment estimators are consistent but generally not very efficient. MLEs, on the other hand, are consistent and, if adjusted to be unbiased, often lead to minimum-variance unbiased estimators. Because they have many good properties, MLEs are often used in practice.

References and Further Readings

Casella, G., and R. L. Berger. 2002. *Statistical Inference*, 2d ed. Pacific Grove, Calif.: Duxbury.

Cramer, H. 1973. *The Elements of Probability Theory and Some of Its Applications*, 2d ed. Huntington, N.Y.: Krieger.

Hogg, R. V., A. T. Craig, and J. W. McKean. 2005. *Introduction to Mathematical Statistics*, 6th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Lindgren, B. W. 1993. *Statistical Theory*, 4th ed. Boca Raton, Fla.: Chapman and Hall/CRC.

Miller, I., and M. Miller. 2003. *John E. Freund's Mathematical Statistics with Applications*, 7th ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Mood, A. M., F. A. Graybill, and D. Boes. 1974. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill.

Serfling, R. J. 2002. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Wilks, S. S. 1963. *Mathematical Statistics*. New York: Wiley.