is the Student \((m + n - 2)\) distribution; see Problem 7.

(a) Use Problem 8.2.6 to justify the preceding distribution for \(t\).

(b) Describe the procedure for testing the hypothesis: \(\mu_1 = \mu_2\).

(c) Determine a \(1 - \alpha\) confidence interval for \(\mu_2 - \mu_1\).

(10) (continuation) Use the data and model from Problem 8 to determine a 95 percent confidence interval for the difference \(\mu_B - \mu_A\) in mean lifetimes.

11 Eleven pieces of material were sampled from a lot; five chosen at random were subjected to a first treatment, and the remaining six to a second treatment. The resulting response measurement yields

\[
\bar{y}_1 = 0.275, \quad \bar{y}_2 = 0.293,
\]

\[
s_1^2 = 0.00045, \quad s_2^2 = 0.00039.
\]

On the basis of the model in Problem 9,

(a) Test the hypothesis \(\mu_1 = \mu_2\).

(b) Form a 95 percent confidence interval for \(\delta = \mu_2 - \mu_1\).

12 For the model as given in Problem 8.2.9, we have that

\[
\frac{\sigma_2^2}{\sigma_1^2} = \frac{s_2^2/\sigma_1^2}{s_2^2/\sigma_2^2}
\]

is the \(F\) distribution \((m - 1, n - 1)\).

(a) Describe the procedure for testing the hypothesis \(\sigma_1^2 = \sigma_2^2\).

(b) Determine a \(1 - \alpha\) confidence interval for \(\gamma = \sigma_1^2/\sigma_2^2\).

13 (continuation) For the data in Problem 11, test the hypothesis \(\sigma_1^2 = \sigma_2^2\).

\[8.4\]

AN IMPORTANT STATISTIC

For the normal(\(\theta, \sigma_0\)) model we have seen (Example 8.1.1) that each sample point \((y_1, \ldots, y_n)\) that has a particular value for the function \(\bar{y}\) gives the same likelihood function; accordingly, as we have noted, the model says the same thing about each of these sample points. As a result we restricted our attention for statistical inference to \(\bar{y}\) and its marginal model.

For this example we can see that a particular value for \(\bar{y}\) determines the particular likelihood function, and, indeed, a particular likelihood function in return determines the particular value for \(\bar{y}\). Thus \(\bar{y}\) indexes the possible likelihood functions. Such an essentially unique statistic that indexes the likelihood functions is called the likelihood statistic. The likelihood statistic is a fundamental part of the analysis of most statistical models; indeed, it guides the analysis in almost all areas of application.

Of course, for this example the function \(\bar{y}\) is the maximum likelihood estimate (MLE). For some more general problems, however, we may need to know more than the MLE estimate in order to determine the particular likelihood function. Thus we should not expect generally that a maximum likelihood estimate will serve as the likelihood statistic; recall Exercise 8.1.4.

A THE LIKELIHOOD STATISTIC

First, consider the normal model with known variance \(\sigma_0^2\). The likelihood function \(L(y | \cdot)\) from a possible sample \(y = (y_1, \ldots, y_n)\) is given by
(1) \[ L(y | \theta) = c \exp \left( -\frac{n}{2\sigma^2_0} (\bar{y} - \theta)^2 \right); \]

the arbitrary constant \( c \) generates the equivalence class of similarly shaped functions of \( \theta \). We can eliminate the constant by choosing, say, the representative that is standardized with respect to maximum likelihood:

(2) \[ L_2(y | \theta) = \exp \left( -\frac{n}{2\sigma^2_0} (\bar{y} - \theta)^2 \right). \]

Note that the likelihood function \( L_2(y | \cdot) \) has normal shape, is scaled by \( \sigma_0/\sqrt{n} \), and has maximum value at the point \( \theta = \bar{y} \).

Each sample point \((y_1, \ldots, y_n)\) that has \( \bar{y} \) equal to \( k \) produces a particular likelihood function located at \( k \), and each point that has \( \bar{y} \) equal to a different \( k' \) produces a particular likelihood function located at a different point \( k' \); see Figure 8.10. Thus the values for \( \bar{y} \) can be put in one-to-one correspondence with the possible likelihood functions; that is, \( \bar{y} \) indexes the possible likelihood functions. Any other function that indexes the likelihood functions is of course one-to-one equivalent to \( \bar{y} \); for example \( \Sigma y_i \) or \( \bar{y}^2 \), or \( \exp \{ \bar{y} \} \); recall Section 3.1D. Accordingly, we call \( \bar{y} \) the likelihood statistic for the normal model with known variance; and indeed we can call \( \bar{y}^3 \) the likelihood statistic.

Now consider a statistical model on \( R^n \) with density function \( f(\cdot | \theta) \) with \( \theta \) in \( \Omega \); we assume some regularity, as discussed in Section 8.1A. The likelihood function from a sample point \( y \) is \( L(y | \cdot) \), where

(3) \[ L(y | \theta) = cf(y | \theta); \]

the arbitrary constant \( c \) generates the equivalence class of similarly shaped functions of \( \theta \).

We now consider a particular likelihood function and we identify the preimage set of points that produce that likelihood function; in a similar way, we consider some other likelihood function and we identify the preimage set of points that produce that likelihood function; see Figure 8.11. Suppose that we can find a function \( s \) on \( R^n \) that indexes the likelihood functions; then the set of points that produce a particular likelihood function is the set of points for which \( s(y) \) takes some value \( k \).

**FIGURE 8.10**
Likelihood function from points having \( \bar{y} = k \); likelihood function from points having \( \bar{y} = k' \).
**FIGURE 8.11**
Likelihood function from points having $s(y) = k$; likelihood function from points having $s(y) = k'$.

**DEFINITION 1**

The **likelihood statistic** is a function on the sample space that indexes the possible likelihood functions; the likelihood statistic is essentially unique up to a one-to-one equivalence.

In conclusion we remark that this concept can be examined from a very general point of view. Let $\mathcal{L}$ be the set of possible likelihood functions. We can view $L(y | \cdot)$ as a function that carries a point $y$ in $\mathbb{R}^n$ to a particular likelihood function in $\mathcal{L}$; call this function the **likelihood map**. The likelihood map itself is an example of the likelihood statistic; in the preceding paragraph, however, we had in mind a simple function such as $s(y) = \bar{y}$.

**B EXAMPLES OF LIKELIHOOD STATISTICS**

We now determine the likelihood statistic for several examples.

**EXAMPLE 1**

The **location-scale normal**: Let $(y_1, \ldots, y_n)$ be a sample from the normal($\mu$, $\sigma$) distribution with $(\mu, \sigma)$ in $\mathbb{R} \times \mathbb{R}^+$. From Example 8.1.3 we know that $(\bar{y}, s_n^2)$ indexes the likelihood functions. Thus $s(y) = (\bar{y}, s_n^2)$ is the likelihood statistic. Note some equivalent forms: $(\bar{y}, \sum (y_i - \bar{y})^2), (\sum y_i, \sum y_i^2)$.

**EXAMPLE 2**

The **Bernoulli**: Let $(x_1, \ldots, x_n)$ be a sample from the Bernoulli($p$) with $p$ in $[0, 1]$. From Example 8.1.2 we know that $\sum x_i$ indexes the likelihood functions. Thus $s(x) = \sum x_i$ is the likelihood statistic. An equivalent statistic is the sample proportion $\hat{p} = \sum x_i/n$.

**EXAMPLE 3**

The **uniform**: Let $(y_1, \ldots, y_n)$ be a sample from a uniform(0, $\theta$) distribution with $\theta$ in (0, $\infty$). For notation, let $\phi$ be the indicator function for the interval (0, 1).
FIGURE 8.12
Likelihood function $f(y|\theta)$ from a sample point $y$.

Then for a single response we have

$$f(y|\theta) = \theta^{-1} \phi(y/\theta),$$

and for a sample of $n$ we have

$$f(y|\theta) = \theta^{-n} \prod_{i=1}^{n} \phi(y_i/\theta).$$

Thus the likelihood function from a sample point $y$ is

$$L(y|\theta) = c \theta^{-n} \prod_{i=1}^{n} \phi(y_i/\theta).$$

This function has the form $c \theta^{-n}$ whenever the product of the $\phi$'s is equal to 1, that is, whenever $\theta$ is greater than all the $y$'s or whenever $\theta$ is greater than max $y_i = y_{(n)}$, and it is zero for $\theta \leq y_{(n)}$. See Figure 8.12. Note that $y_{(n)}$ determines the likelihood function and the likelihood function determines $y_{(n)}$. Thus the function $s(y) = L_{n}$ indexes the likelihood functions; it is the likelihood statistic. It follows that $y_{(n)}$ is all that is needed from an observed response.

EXAMPLE 4

Let $\{y_1, \ldots, y_n\}$ be a sample from a distribution given by $f(y|\theta)$ with $\theta$ in $\{\theta_1, \theta_2\}$. The likelihood function can be expressed conveniently by using the representative

$$L_2(y) = \frac{\prod_{i=1}^{n} f(y_i|\theta_2)}{\prod_{i=1}^{n} f(y_i|\theta_1)},$$

mentioned in Section 8.1D. For given $y$ note that $L_2(y)$ is just a real number or $+\infty$ (nonzero numerator, zero denominator); it is called the likelihood ratio and it gives the ratio of the probability at $y$ from $\theta_2$ to that from $\theta_1$. Note that $L_2(y)$ as a function on the sample space $\mathbb{R}^{n}$ maps into $\mathbb{R}$ or into $\mathbb{R} \cup \{+\infty\}$. Thus $s(y) = L_2(y)$ indexes the likelihood functions; it is a real-valued likelihood statistic.
THE LIKELIHOOD STATISTIC
WITH INDEPENDENCE

Now consider the likelihood statistic when independent models are combined. We see first that likelihood functions from independent systems combine in a very simple way.

Consider a statistical model for $y_1$ given by $f'(y_1 | \theta)$ with $\theta$ in $\Omega$ and a statistical model for an independent $y_2$ given by $f''(y_2 | \theta)$ with the same $\theta$ in $\Omega$. The statistical model for the combined $(y_1, y_2)$ is then given by $f'(y_1 | \theta)f''(y_2 | \theta)$ with $\theta$ in $\Omega$.

The likelihood functions from the components $y_1$ and $y_2$ are

$$L'(y_1 | \theta) = cf'(y_1 | \theta), \quad L''(y_2 | \theta) = cf''(y_2 | \theta).$$

And the likelihood function from the combined $(y_1, y_2)$ is

$$(6) \quad L(y_1, y_2 | \theta) = cf'(y_1 | \theta)f''(y_2 | \theta) = L'(y_1 | \theta)L''(y_2 | \theta).$$

Thus when independent models are combined, the likelihood functions are multiplied. Correspondingly, the log-likelihoods or the scores are added:

$$(7) \quad l(y_1, y_2 | \theta) = l'(y_1 | \theta) + l''(y_2 | \theta).$$

$$(8) \quad S(y_1, y_2 | \theta) = S'(y_1 | \theta) + S''(y_2 | \theta).$$

Thus when two statistical models are combined, we have the simple formulas (6), (7), and (8) for combining the likelihood, log-likelihood, or score functions. Accordingly, we will almost always determine the likelihood statistic for a combined model directly from properties of the simply obtained likelihood function.

THE LIKELIHOOD STATISTIC
IS SUFFICIENT

We have seen that the likelihood function presents all that the model has to say about an observed response. We have also seen that a range of sample points can have the same likelihood function—that the model is saying the same thing about each of the points in the range. Accordingly, we used $\bar{y}$ with the normal$(\mu, \sigma_0)$ and we used $(\bar{y}, s_n^2)$ with the normal$(\mu, \sigma)$—on the grounds that the model does not distinguish among the points having a particular value for the likelihood statistic. And, of course, any implications concerning the true parameter value would be the same from each of the points in the range.

We now consider a distribution property of the likelihood statistic, a property that is somewhat weaker than the preceding result. Suppose that we have an investigator who has collected data from a system being investigated and has reported to us the value of the likelihood statistic. With this value in hand we are, of course, free to contemplate the possibilities for the antecedent data, that is, to contemplate the conditional distribution given the value of the likelihood statistic. This conditional distribution cannot involve the parameter $\theta$, for if it did, the model would have to say different things about various possibilities for the antecedent data. We examine this more specifically.

Consider a statistical model on $\mathbb{R}^n$ with density function $f(\cdot | \theta)$ with $\theta$ in $\Omega$; we assume some regularity, as discussed in Section 8.1A. In the preceding paragraph we
mentioned a conditional distribution being independent of a parameter; for this we introduce the following definition.

**DEFINITION 2**

A function \( t \) on the sample space \( R^n \) is a **sufficient statistic** if the conditional distribution given \( t \) does not depend on the parameter \( \theta \).

As support for the term "sufficient," suppose that we have been given the value \( t(y) = t' \) for the sufficient statistic in an application. Then it is reasonable to inquire about the antecedent data. The definition says that the distribution describing possible antecedent data does not involve \( \theta \). This suggests that the antecedent data would be of no additional use—that the statistic \( t \) is sufficient. Of course, \( t \) can be vector-valued.

Consider briefly an example.

**EXAMPLE 5**

Let \( (y_1, y_2) \) be a sample of \( n = 2 \) from the normal(\( \theta \), \( \sigma_0 \)) with \( \theta \) in \( R \). From Section 8.2C we know that

\[
  u_1 = \frac{y_1 + y_2}{\sqrt{2}}, \quad u_2 = \frac{y_1 - y_2}{\sqrt{2}}
\]

are statistically independent, that \( u_1 \) is normal(\( \sqrt{2}\theta \), \( \sigma_0 \)), and that \( u_2 \) is normal(0, \( \sigma_0 \)). We examine the conditional distribution given \( u_1 \): the conditional distribution is described by \( u_2 \) and is normal(0, \( \sigma_0 \)); it does not involve \( \theta \). Thus \( u_1 \) is sufficient. Of course, we note that \( u_1 = \sqrt{2}\bar{y} \) is the likelihood statistic.

Our introductory remarks can now be formalized as the following theorem.

**THEOREM 3**

The likelihood statistic is sufficient.

*Proof* Our consideration of conditional distributions in Section 4.2 was limited to cases in which we could find a complementing function \( u \) such that there is a one-to-one correspondence, \( y \leftrightarrow (s(y), u(y)) \); and in the absolutely continuous case, we require a continuous Jacobian each way. We record the proof for this absolutely continuous case; the discrete case is a direct analog without the Jacobians.

The likelihood function \( L(y \mid \cdot) \) is determined by the likelihood statistic \( s(y) \). Let \( L^*(s \mid \cdot) \) be a particular representative in the class of similarly shaped likelihood functions determined by \( s(y) = s \). Note that \( f(y \mid \theta) \) and \( L^*(s(y) \mid \theta) \) are similarly shaped functions of \( \theta \); let \( a(y) \) give the proportionality constant,

\[
  (9) \quad f(y \mid \theta) = a(y)L^*(s(y) \mid \theta).
\]

Thus the probability differential for \( y \) can be written

\[
  (10) \quad f(y \mid \theta) \, dy = a(y)L^*(s(y) \mid \theta) \, dy.
\]
The joint distribution for \((s, u)\) can be obtained from Section 2.4E. Let 
\(J(s, u) = |\partial y/\partial(s, u)|_+\) be the Jacobian determinant. Then substitution in (10) gives the probability differential for \((s, u)\),

\[
(11) \quad a(y)L^*(s|\theta)J(s, u) \, ds \, du.
\]

The marginal distribution for \(s\) can be obtained by integrating out the variable \(u\). The variable \(u\) occurs only in the factors \(a(y)\) and \(J(s, u)\); hence we obtain the marginal differential

\[
(12) \quad L^*(s|\theta)k(s) \, ds,
\]

where \(k\) is the obvious integral involving \(a(y)\) and \(J(s, u)\).

The conditional distribution for \(u\) given \(s\) can be obtained by dividing (11) by (12); we obtain

\[
(13) \quad \frac{a(y)J(s, u)}{k(s)} \, du.
\]

Note that this conditional distribution does not depend on \(\theta\); thus \(s\) is a sufficient statistic.

The preceding proof has an immediate corollary:

**COROLLARY 4**

The likelihood function obtained from an observed \(y\) is the same as the likelihood function obtained from the likelihood statistic \(s(y)\) and the marginal model for \(s\).

**Proof** The likelihood function from \(s(y)\) and its marginal model is available from (12):

\[
(14) \quad cL^*(s(y)|\theta)k(s(y)) = cL^*(s(y)|\theta).
\]

But this is just the likelihood function from \(y\) as produced by (9).

**E THE LIKELIHOOD STATISTIC IS MINIMAL SUFFICIENT**

A sufficient statistic has an attractive property: it provides a reduction or simplification on the original response, and yet it records all that seems necessary concerning the original response. Clearly we would be interested in a sufficient statistic that makes the largest reduction—records as little as possible concerning the original response and yet records all that is necessary. For this we present the following definition:

**DEFINITION 5**

A sufficient statistic \(s\) is **minimal** if for any other sufficient statistic \(t\) we can find a function \(h\) so that \(s(y) = h(t(y))\).

The definition says that a minimal sufficient statistic is a reduction on any other sufficient statistic. We now have the following property for the likelihood statistic.
Theorem 6

The likelihood statistic is the minimal sufficient statistic.

Proof Consider a sufficient statistic \( t \). We confine our proof as with Theorem 3 to the case where we have a complementing function \( u \) such that there is a one-to-one correspondence \( y \leftrightarrow (t, u(y)) \); and in the absolutely continuous case we require a continuous Jacobian each way. We then obtain

\[
(15) \quad f(y|\theta) \, dy = f(y|\theta)J(t, u) \, dt \, du = f_1(u|\theta) f_2(t|\theta) \, dt \, du,
\]

where the middle expression comes by a change of variable and the last expression is an ordinary conditional-times-marginal expression in which the conditional density \( f_1(u|t) \), by sufficiency, does not depend on \( \theta \). The likelihood function \( L(y|\cdot) \) can now be calculated and we have

\[
(16) \quad L(y|\theta) = cnf(y|\theta) = cf_2(t|y) \theta.
\]

Theorem states in effect that the likelihood statistic is the best sufficient statistic. The concept of a sufficient statistic has been widely used in statistical theory. The concept of the likelihood statistic, however, has recently emerged as a more appropriate concept: The likelihood statistic has the strong justification based on the likelihood function in Section 8.1, a justification that is substantially stronger than the distribution property of a sufficient statistic; the likelihood function itself admits direct interpretation and immediate usefulness in statistical inference; and, besides, the likelihood statistic is the fastest route to the minimal sufficient statistic.

F Exercises

1. **Poisson distribution:** Let \( y_1, \ldots, y_n \) be a sample from the Poisson(\( \lambda \)) distribution with \( \lambda \) in \((0, \infty)\). Show that the likelihood statistic can be given by \( \sum y_i \) or by \( \bar{y} \).

2. **Binomial distribution:** Let \( x_1, \ldots, x_n \) be a sample from the Bernoulli(\( p \)) distribution with \( p \) in \([0, 1]\). Show that the likelihood statistic can be given by \( \sum x_i \) or by \( \bar{x} \).

3. **Scale normal:** Let \( y_1, \ldots, y_n \) be a sample from the normal(\( \mu_\circ, \sigma^2 \)) distribution with \( \sigma^2 \) in \((0, \infty)\). Show that the likelihood statistic is \( \sum (y_i - \mu_\circ)^2 \).

4. **Uniform \( (\theta, \theta + 1) \):** Let \( y_1, \ldots, y_n \) be a sample from the uniform(\( \theta, \theta + 1 \)) distribution with \( \theta \) in \( R \). Show that the likelihood statistic is \( (y_{(1)}, y_{(n)}) \).

5. **Scale-exponential:** Let \( y_1, \ldots, y_n \) be a sample from the exponential(\( \theta \)) with \( \theta \) in \((0, \infty)\). Show that the likelihood statistic is \( \sum y_i \).

6. **Scale gamma or chi-square:** Let \( y_1, \ldots, y_n \) be a sample from the scale gamma having density \( f(y|\theta) = \Gamma^{-1}(p)\theta^{-y}e^{-y/\theta} \) on \((0, \infty)\) and 0 otherwise, with \( \theta \) in \((0, \infty)\). Show that the likelihood statistic is \( \sum y_i \).

7. **Pareto:** Let \( y_1, \ldots, y_n \) be a sample from the Pareto distribution \( f(y|\theta) = \theta(1 + y)^{-\theta - 1} \) on \((0, \infty)\) and 0 otherwise, with \( \theta \) in \((0, \infty)\). Show that the likelihood statistic is \( \Pi (1 + y_i) \).

8. **Order statistic:** Let \( y_1, \ldots, y_n \) be a sample of \( n \) from an arbitrary continuous density on \( R \). Determine the conditional distribution of \( y_{(1)}, \ldots, y_{(n)} \) given \( y_{(1)}, \ldots, y_{(n)} \); and deduce that the order statistic \( (y_{(1)}, \ldots, y_{(n)}) \) is sufficient; see Section 3.6E.
9 **Negative binomial distribution:** Let \((x_1, x_2, \ldots)\) be a sample sequence from the Bernoulli(\(p\)) distribution with \(p\) in \((0, 1)\), and suppose the sequence is terminated when \(\sum x_i = \text{a given } k\). Show that the likelihood statistic is given by \(n\).

10 **Hypergeometric:** Let \((x_1, \ldots, x_n)\) be a random sample from a finite population of \(D\) 0’s and \(N - D\) 1’s with \(D\) in \((0, 1, \ldots, N)\). Show that the likelihood statistic is \(\sum x_i\) (the hypergeometric variable).

11 **Multinomial:** Let \((y_1, \ldots, y_n)\) be a sample from the multinomial \((n, p_1, \ldots, p_k)\) with \(\sum p_i = 1, p_i \geq 0\). Show that the likelihood statistic is \(\sum y_i\).

12 **Multivariate hypergeometric:** Let \((x_1, \ldots, x_n)\) be a random sample from a finite population containing \(Np_1\) elements \(b_1, \ldots, Np_k\) elements \(b_k\), where the possible \(p_1, \ldots, p_k\) satisfy \(\sum p_i = 1, p_i \geq 0\). Show that the likelihood statistic is \((t_1, \ldots, t_k)\), where \(t_j\) is the number of \(b_j\) in \((x_1, \ldots, x_n)\).

13 **Variable carrier:** Let \((y_1, \ldots, y_n)\) be a sample from \(f(y|\theta) = k(\theta)c(y - \theta)h(y)\), where \(c\) is the positive indicator function defined in Problem 7.4.8. Show that the likelihood statistic is \(y_{11}\).

14 **Uniform(\(\theta_1, \theta_2\):** Let \((y_1, \ldots, y_n)\) be a sample from the uniform distribution \((\theta_1, \theta_2)\) with \(\theta_1 < \theta_2\) and \(\theta_i\) in \(R\). Show that the likelihood statistic is \((y_{11}, y_{1n})\).

15 **Location-scale exponential:** Let \((y_1, \ldots, y_n)\) be a sample from \(f(y|\theta, \tau) = \tau^{-1} \exp\left(-\frac{y - \theta}{\tau}\right)\) on \((\theta, \infty)\) and 0 otherwise, with \(\theta, \tau\) in \(R \times (0, \infty)\). Show that the likelihood statistic is \((y_{11}, \sum y_i)\).

16 A drug is administered to \(n\) animals at each of \(r\) doses \(x_1, \ldots, x_r\). The “logistic” model is sometimes used for the probability \(p_i\) of reaction at dose \(x_i\),

\[
p_i = \frac{\alpha e^{\beta x_i}}{1 + \alpha e^{\beta x_i}}
\]

with parameters \(\alpha\) and \(\beta\). Let \(y_1, \ldots, y_r\) be the number of reactions, respectively, at doses \(x_1, \ldots, x_r\). Show that the likelihood statistic is \((\sum y_i, \sum x_i y_i)\).

17 Let \(p_i\) be the probability of exactly \(i\) of \(n\) female children in a family having \(3\) children, assuming independence of successive births, and let \(\theta\) be the probability of a female child. For a sample of \(n\) families having \(3\) children, let \(y_i\) be the number of families having \(i\) of \(3\) female children. Show that \(y_1 + 2y_2 + 3y_3\) is the likelihood statistic.

18 (a) Let \((x_1, \ldots, x_n)\) be a sample from the Bernoulli(\(p\)) distribution and write \(y = \sum x_i\). Determine the likelihood function.

(b) Let \(x_1, x_2, \ldots\) be statistically independent with the Bernoulli(\(p\)) distribution. Let \(n\) be the coordinate number at which \(\sum x_i = y\), where \(y\) is specified. Determine the likelihood function.

(c) If the values of \((n, y)\) in (a) and (b) are the same, show that the likelihood functions are the same.

19 Consider a population containing \(D\) 1’s and \(N - D\) 0’s with parameter \(D\) in \(\Omega = \{0, 1, \ldots, N\}\).

(a) Let \(y\) be the number of 1’s in a random sample of \(n\). Determine the likelihood function from an observed \(y\).

(b) Suppose that elements are sampled randomly until a specified number \(y\) of 1’s has been obtained. Let \(n\) be the number of elements sampled. Determine the likelihood function from an observed \(n\).

(c) If the values of \((n, y)\) in (a) and (b) are the same, show that the likelihood functions are the same.

20 Suppose that the parameter space \(\Omega\) is a connected open set in \(R^t\). Then show that the score function \(S_j(y|\theta) = \left. \frac{\partial}{\partial \theta} \ln \left| f(y|\theta) \right| \right| \) on \(\Omega\) can be calculated from the log-likelihood, and conversely; the score components are

\[
S_j(y|\theta) = \frac{\partial}{\partial \theta_j} \ln f(y|\theta).
\]
21 From formula (16) we know that the likelihood function from an observed value of a sufficient statistic as calculated from the marginal model is the same as the likelihood function from the antecedent response as calculated from the full model. Let \( y_1, \ldots, y_n \) be a sample from a distribution with density \( f(y|\theta) \) with \( \theta \) in \( (\theta_1, \theta_2) \) and let \( L_3(y) \) be the likelihood ratio given in formula (5). Then show that the likelihood function (third form) from the observed \( L_3(y) \) is obtained from \( L_3 \) by the identity function.

22 Let \( (T_1, \ldots, T_n) \) be a sample from the canonical Student(\( \lambda \)) distribution with parameter \( \lambda \) in \( (0, \infty) \). Determine the likelihood statistic.

23 Let \( (G_1, \ldots, G_n) \) be a sample from the canonical \( F(\lambda_1, \lambda_2) \) distribution with parameter \( (\lambda_1, \lambda_2) \) in \( (0, \infty) \times (0, \infty) \). Determine the likelihood statistic.

8.5 MODELS THAT HAVE EXPONENTIAL FORM

We have seen that, for a sample \( (y_1, \ldots, y_{100}) \) of 100 from the normal(\( \mu, \sigma \)) model, it suffices to record the value of \( (\bar{y}, S)^2 \) for purposes of inference. This is a huge reduction—from a point in \( R^{100} \) to a point in \( R^2 \). In this section we investigate what kind of functional form for the model makes such reductions possible—in short, what makes the normal work.

Consider this normal example in detail. The density for a single \( y \) from the normal(\( \mu, \sigma \)) can be expressed in the form

\[
(1) \quad f(y|\mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\} \cdot \exp \left\{ y^2 - \frac{\mu^2}{\sigma^2} - \frac{1}{2\sigma^2} \right\}.
\]

Note the simple form in the right-hand exponent—a linear combination of statistics and parameters. We obtain the density for a sample of \( n \) by adding the exponents:

\[
(2) \quad f(y|\mu, \sigma) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \cdot \exp \left\{ \sum y_i^2 - \sum y_i^2 - \frac{1}{2\sigma^2} \right\}.
\]

Recall that \( (\Sigma y_i, \Sigma y_i^2) \) is the likelihood statistic.

Consider the Bernoulli(\( \rho \)) model. The density for a single \( x \) can be expressed in the form

\[
(3) \quad f(x|\rho) = \rho^x q^{1-x} = q \cdot \exp \left\{ x (\ln \rho - \ln q) \right\}.
\]

Note the simple form in the exponent—a linear combination of a statistic and a parameter. We obtain the density for a sample of \( n \) by adding the exponents:

\[
(4) \quad f(x|\rho) = \rho^x q^{n-x} = q^n \cdot \exp \left\{ \sum x_i (\ln \rho - \ln q) \right\}.
\]

Recall that \( \sum x_i \) is the likelihood statistic.

In this section we investigate statistical models that have this linear form in the exponent, and we see how this form relates to having a simple likelihood statistic regardless of the sample size.

A EXPONENTIAL FORM

The linear properties in the exponent lead to the following definition of exponential form.