

9.5: Likelihood Ratio Tests

Basic Theory

As usual, our starting point is a random experiment with an underlying sample space, and a probability measure \mathbb{P} . In the basic statistical model, we have an observable random variable \mathbf{X} taking values in a set S . In general, \mathbf{X} can have quite a complicated structure. For example, if the experiment is to sample n objects from a population and record various measurements of interest, then

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (9.5.1)$$

where X_i is the vector of measurements for the i th object. The most important special case occurs when (X_1, X_2, \dots, X_n) are independent and identically distributed. In this case, we have a random sample of size n from the common distribution.

In the previous sections, we developed tests for parameters based on natural test statistics. However, in other cases, the tests may not be parametric, or there may not be an obvious statistic to start with. Thus, we need a more general method for constructing test statistics. Moreover, we do not yet know if the tests constructed so far are the best, in the sense of maximizing the power for the set of alternatives. In this and the next section, we investigate both of these ideas. Likelihood functions, similar to those used in maximum likelihood estimation, will play a key role.

Tests of Simple Hypotheses

Suppose that \mathbf{X} has one of two possible distributions. Our simple hypotheses are

- $H_0 : \mathbf{X}$ has probability density function f_0 .
- $H_1 : \mathbf{X}$ has probability density function f_1 .

We will use subscripts on the probability measure \mathbb{P} to indicate the two hypotheses, and we assume that f_0 and f_1 are positive on S . The test that we will construct is based on the following simple idea: if we observe $\mathbf{X} = \mathbf{x}$, then the condition $f_1(\mathbf{x}) > f_0(\mathbf{x})$ is evidence in favor of the alternative; the opposite inequality is evidence against the alternative.

The *likelihood ratio function* $L : S \rightarrow (0, \infty)$ is defined by

$$L(\mathbf{x}) = \frac{f_0(\mathbf{x})}{f_1(\mathbf{x})}, \quad \mathbf{x} \in S \quad (9.5.2)$$

The statistic $L(\mathbf{X})$ is the *likelihood ratio statistic*.

Restating our earlier observation, note that small values of L are evidence in favor of H_1 . Thus it seems reasonable that the likelihood ratio statistic may be a good test statistic, and that we should consider tests in which we reject H_0 if and only if $L \leq l$, where l is a constant to be determined:

The significance level of the test is $\alpha = \mathbb{P}_0(L \leq l)$.

As usual, we can try to construct a test by choosing l so that α is a prescribed value. If \mathbf{X} has a discrete distribution, this will only be possible when α is a value of the distribution function of $L(\mathbf{X})$.

An important special case of this model occurs when the distribution of \mathbf{X} depends on a parameter θ that has two possible values. Thus, the parameter space is $\{\theta_0, \theta_1\}$, and f_0 denotes the probability density function of \mathbf{X} when $\theta = \theta_0$ and f_1 denotes the probability density function of \mathbf{X} when $\theta = \theta_1$. In this case, the hypotheses are equivalent to $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

As noted earlier, another important special case is when $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from a distribution an underlying random variable X taking values in a set R . In this case, $S = R^n$ and the probability density function f of \mathbf{X} has the form

$$f(x_1, x_2, \dots, x_n) = g(x_1)g(x_2) \cdots g(x_n), \quad (x_1, x_2, \dots, x_n) \in S \quad (9.5.3)$$

where g is the probability density function of X . So the hypotheses simplify to

- $H_0 : X$ has probability density function g_0 .

- $H_1 : X$ has probability density function g_1 .

and the likelihood ratio statistic is

$$L(X_1, X_2, \dots, X_n) = \prod_{i=1}^n \frac{g_0(X_i)}{g_1(X_i)} \quad (9.5.4)$$

In this special case, it turns out that under H_1 , the likelihood ratio statistic, as a function of the sample size n , is a martingale.

The Neyman-Pearson Lemma

The following theorem is the *Neyman-Pearson Lemma*, named for Jerzy Neyman and Egon Pearson. It shows that the test given above is most powerful. Let

$$R = \{\mathbf{x} \in S : L(\mathbf{x}) \leq l\} \quad (9.5.5)$$

and recall that the *size* of a rejection region is the significance of the test with that rejection region.

Consider the tests with rejection regions R given above and arbitrary $A \subseteq S$. If the size of R is at least as large as the size of A then the test with rejection region R is more powerful than the test with rejection region A . That is, if $\mathbb{P}_0(\mathbf{X} \in R) \geq \mathbb{P}_0(\mathbf{X} \in A)$ then $\mathbb{P}_1(\mathbf{X} \in R) \geq \mathbb{P}_1(\mathbf{X} \in A)$.

Proof

First note that from the definitions of L and R that the following inequalities hold:

$$\mathbb{P}_0(\mathbf{X} \in A) \leq l \mathbb{P}_1(\mathbf{X} \in A) \text{ for } A \subseteq R \quad (9.5.6)$$

$$\mathbb{P}_0(\mathbf{X} \in A) \geq l \mathbb{P}_1(\mathbf{X} \in A) \text{ for } A \subseteq R^c \quad (9.5.7)$$

Now for arbitrary $A \subseteq S$, write $R = (R \cap A) \cup (R \setminus A)$ and $A = (A \cap R) \cup (A \setminus R)$. From the additivity of probability and the inequalities above, it follows that

$$\mathbb{P}_1(\mathbf{X} \in R) - \mathbb{P}_1(\mathbf{X} \in A) \geq \frac{1}{l} [\mathbb{P}_0(\mathbf{X} \in R) - \mathbb{P}_0(\mathbf{X} \in A)] \quad (9.5.8)$$

Hence if $\mathbb{P}_0(\mathbf{X} \in R) \geq \mathbb{P}_0(\mathbf{X} \in A)$ then $\mathbb{P}_1(\mathbf{X} \in R) \geq \mathbb{P}_1(\mathbf{X} \in A)$.

The Neyman-Pearson lemma is more useful than might be first apparent. In many important cases, the *same* most powerful test works for a range of alternatives, and thus is a *uniformly* most powerful test for this range. Several special cases are discussed below.

Generalized Likelihood Ratio

The likelihood ratio statistic can be generalized to composite hypotheses. Suppose again that the probability density function f_θ of the data variable \mathbf{X} depends on a parameter θ , taking values in a parameter space Θ . Consider the hypotheses $\theta \in \Theta_0$ versus $\theta \notin \Theta_0$, where $\Theta_0 \subseteq \Theta$.

Define

$$L(\mathbf{x}) = \frac{\sup \{f_\theta(\mathbf{x}) : \theta \in \Theta_0\}}{\sup \{f_\theta(\mathbf{x}) : \theta \in \Theta\}} \quad (9.5.9)$$

The function L is the *likelihood ratio function* and $L(\mathbf{X})$ is the *likelihood ratio statistic*.

By the same reasoning as before, small values of $L(\mathbf{x})$ are evidence in favor of the alternative hypothesis.

Examples and Special Cases

Tests for the Exponential Model

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size $n \in \mathbb{N}_+$ from the exponential distribution with scale parameter $b \in (0, \infty)$. The sample variables might represent the lifetimes from a sample of devices of a certain type. We are interested in testing the simple hypotheses $H_0 : b = b_0$ versus $H_1 : b = b_1$, where $b_0, b_1 \in (0, \infty)$ are distinct specified values.

Recall that the sum of the variables is a sufficient statistic for b :

$$Y = \sum_{i=1}^n X_i \quad (9.5.10)$$

Recall also that Y has the gamma distribution with shape parameter n and scale parameter b . For $\alpha > 0$, we will denote the quantile of order α for this distribution by $\gamma_{n,b}(\alpha)$.

The likelihood ratio statistic is

$$L = \left(\frac{b_1}{b_0}\right)^n \exp\left[\left(\frac{1}{b_1} - \frac{1}{b_0}\right) Y\right] \quad (9.5.11)$$

Proof

Recall that the PDF g of the exponential distribution with scale parameter $b \in (0, \infty)$ is given by $g(x) = (1/b)e^{-x/b}$ for $x \in (0, \infty)$. If g_j denotes the PDF when $b = b_j$ for $j \in \{0, 1\}$ then

$$\frac{g_0(x)}{g_1(x)} = \frac{(1/b_0)e^{-x/b_0}}{(1/b_1)e^{-x/b_1}} = \frac{b_1}{b_0} e^{(1/b_1 - 1/b_0)x}, \quad x \in (0, \infty) \quad (9.5.12)$$

Hence the likelihood ratio function is

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{g_0(x_i)}{g_1(x_i)} = \left(\frac{b_1}{b_0}\right)^n e^{(1/b_1 - 1/b_0)y}, \quad (x_1, x_2, \dots, x_n) \in (0, \infty)^n \quad (9.5.13)$$

where $y = \sum_{i=1}^n x_i$.

The following tests are most powerful test at the α level

1. Suppose that $b_1 > b_0$. Reject $H_0 : b = b_0$ versus $H_1 : b = b_1$ if and only if $Y \geq \gamma_{n,b_0}(1 - \alpha)$.
2. Suppose that $b_1 < b_0$. Reject $H_0 : b = b_0$ versus $H_1 : b = b_1$ if and only if $Y \leq \gamma_{n,b_0}(\alpha)$.

Proof

Under H_0 , Y has the gamma distribution with parameters n and b_0 .

1. If $b_1 > b_0$ then $1/b_1 < 1/b_0$. From simple algebra, a rejection region of the form $L(\mathbf{X}) \leq l$ becomes a rejection region of the form $Y \geq y$. The precise value of y in terms of l is not important. For the test to have significance level α we must choose $y = \gamma_{n,b_0}(1 - \alpha)$.
2. If $b_1 < b_0$ then $1/b_1 > 1/b_0$. From simple algebra, a rejection region of the form $L(\mathbf{X}) \leq l$ becomes a rejection region of the form $Y \leq y$. Again, the precise value of y in terms of l is not important. For the test to have significance level α we must choose $y = \gamma_{n,b_0}(\alpha)$.

Note that these tests do not depend on the value of b_1 . This fact, together with the monotonicity of the power function can be used to show that the tests are uniformly most powerful for the usual one-sided tests.

Suppose that $b_0 \in (0, \infty)$.

1. The decision rule in part (a) above is uniformly most powerful for the test $H_0 : b \leq b_0$ versus $H_1 : b > b_0$.
2. The decision rule in part (b) above is uniformly most powerful for the test $H_0 : b \geq b_0$ versus $H_1 : b < b_0$.

Tests for the Bernoulli Model

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size $n \in \mathbb{N}_+$ from the Bernoulli distribution with success parameter p . The sample could represent the results of tossing a coin n times, where p is the probability of heads. We wish to test the simple hypotheses $H_0 : p = p_0$ versus $H_1 : p = p_1$, where $p_0, p_1 \in (0, 1)$ are distinct specified values. In the coin tossing model, we know that the probability of heads is either p_0 or p_1 , but we don't know which.

Recall that the number of successes is a sufficient statistic for p :

$$Y = \sum_{i=1}^n X_i \quad (9.5.14)$$

Recall also that Y has the binomial distribution with parameters n and p . For $\alpha \in (0, 1)$, we will denote the quantile of order α for this distribution by $b_{n,p}(\alpha)$; although since the distribution is discrete, only certain values of α are possible.

The likelihood ratio statistic is

$$L = \left(\frac{1-p_0}{1-p_1} \right)^n \left[\frac{p_0(1-p_1)}{p_1(1-p_0)} \right]^Y \quad (9.5.15)$$

Proof

Recall that the PDF g of the Bernoulli distribution with parameter $p \in (0, 1)$ is given by $g(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$. If g_j denotes the PDF when $p = p_j$ for $j \in \{0, 1\}$ then

$$\frac{g_0(x)}{g_1(x)} = \frac{p_0^x(1-p_0)^{1-x}}{p_1^x(1-p_1)^{1-x}} = \left(\frac{p_0}{p_1} \right)^x \left(\frac{1-p_0}{1-p_1} \right)^{1-x} = \left(\frac{1-p_0}{1-p_1} \right) \left[\frac{p_0(1-p_1)}{p_1(1-p_0)} \right]^x, \quad x \in \{0, 1\} \quad (9.5.16)$$

Hence the likelihood ratio function is

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{g_0(x_i)}{g_1(x_i)} = \left(\frac{1-p_0}{1-p_1} \right)^n \left[\frac{p_0(1-p_1)}{p_1(1-p_0)} \right]^y, \quad (x_1, x_2, \dots, x_n) \in \{0, 1\}^n \quad (9.5.17)$$

where $y = \sum_{i=1}^n x_i$.

The following tests are most powerful test at the α level

1. Suppose that $p_1 > p_0$. Reject $H_0 : p = p_0$ versus $H_1 : p = p_1$ if and only if $Y \geq b_{n,p_0}(1-\alpha)$.
2. Suppose that $p_1 < p_0$. Reject $p = p_0$ versus $p = p_1$ if and only if $Y \leq b_{n,p_0}(\alpha)$.

Proof

Under H_0 , Y has the binomial distribution with parameters n and p_0 .

1. If $p_1 > p_0$ then $p_0(1-p_1)/p_1(1-p_0) < 1$. From simple algebra, a rejection region of the form $L(\mathbf{X}) \leq l$ becomes a rejection region of the form $Y \geq y$. The precise value of y in terms of l is not important. For the test to have significance level α we must choose $y = b_{n,p_0}(1-\alpha)$.
2. If $p_1 < p_0$ then $p_0(1-p_1)/p_1(1-p_0) > 1$. From simple algebra, a rejection region of the form $L(\mathbf{X}) \leq l$ becomes a rejection region of the form $Y \leq y$. Again, the precise value of y in terms of l is not important. For the test to have significance level α we must choose $y = b_{n,p_0}(\alpha)$.

Note that these tests do not depend on the value of p_1 . This fact, together with the monotonicity of the power function can be used to show that the tests are uniformly most powerful for the usual one-sided tests.

Suppose that $p_0 \in (0, 1)$.

1. The decision rule in part (a) above is uniformly most powerful for the test $H_0 : p \leq p_0$ versus $H_1 : p > p_0$.
2. The decision rule in part (b) above is uniformly most powerful for the test $H_0 : p \geq p_0$ versus $H_1 : p < p_0$.

Tests in the Normal Model

The one-sided tests that we derived in the normal model, for μ with σ known, for μ with σ unknown, and for σ with μ unknown are all uniformly most powerful. On the other hand, none of the two-sided tests are uniformly most powerful.

A Nonparametric Example

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size $n \in \mathbb{N}_+$, either from the Poisson distribution with parameter 1 or from the geometric distribution on \mathbb{N} with parameter $p = \frac{1}{2}$. Note that both distributions have mean 1 (although the Poisson distribution has variance 1 while the geometric distribution has variance 2). So, we wish to test the hypotheses

- $H_0 : X$ has probability density function $g_0(x) = e^{-1} \frac{1}{x!}$ for $x \in \mathbb{N}$.

- $H_1 : X$ has probability density function $g_1(x) = \left(\frac{1}{2}\right)^{x+1}$ for $x \in \mathbb{N}$.

The likelihood ratio statistic is

$$L = 2^n e^{-n} \frac{2^Y}{U} \text{ where } Y = \sum_{i=1}^n X_i \text{ and } U = \prod_{i=1}^n X_i! \quad (9.5.18)$$

Proof

Note that

$$\frac{g_0(x)}{g_1(x)} = \frac{e^{-1}/x!}{(1/2)^{x+1}} = 2e^{-1} \frac{2^x}{x!}, \quad x \in \mathbb{N} \quad (9.5.19)$$

Hence the likelihood ratio function is

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n \frac{g_0(x_i)}{g_1(x_i)} = 2^n e^{-n} \frac{2^y}{u}, \quad (x_1, x_2, \dots, x_n) \in \mathbb{N}^n \quad (9.5.20)$$

where $y = \sum_{i=1}^n x_i$ and $u = \prod_{i=1}^n x_i!$.

The most powerful tests have the following form, where d is a constant: reject H_0 if and only if $\ln(2)Y - \ln(U) \leq d$.

Proof

A rejection region of the form $L(\mathbf{X}) \leq l$ is equivalent to

$$\frac{2^Y}{U} \leq \frac{le^n}{2^n} \quad (9.5.21)$$

Taking the natural logarithm, this is equivalent to $\ln(2)Y - \ln(U) \leq d$ where $d = n + \ln(l) - n \ln(2)$

This page titled [9.5: Likelihood Ratio Tests](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.