5. Likelihood Ratio Tests

Preliminaries

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 $\boldsymbol{X}$ taking values in a set S. In general, $\boldsymbol{X}$ 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

$$\boldsymbol{X} = (X_1, X_2, ..., X_n)$$

where X_i is the vector of measurements for the i^{th} object. The most important special case occurs when $(X_1, X_2, ..., X_n)$ 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 $\boldsymbol{X}$ has one of two possible distributions. Our simple hypotheses are

$$H_0\text{: } \boldsymbol{X} \text{ has probability density function } f_0 \text{ versus } H_1\text{: } \boldsymbol{X} \text{ has probability density function } f_1$$

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

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

The function L is the **likelihood ratio function** for the hypotheses and $L(\boldsymbol{X})$ 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 of the following form, where k is a constant:

$$\text{Reject } H_0 \text{ if and only if } L \le k$$

1. Show that the significance level of the test is $\alpha = \mathbb{P}_0(L \le k)$

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

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

$$H_0\colon\ \theta = \theta_0 \ \text{ versus } \ H_1\colon\ \theta = \theta_1$$

The Neyman-Pearson Lemma

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

$$R = \{\boldsymbol{x} \in S : L(\boldsymbol{x}) \le k\}$$

2. Use the definitions of L and R to show that

a. $\mathbb{P}_0(\boldsymbol{X} \in A) \le k\,\mathbb{P}_1(\boldsymbol{X} \in A)$ for $A \subseteq R$
b. $\mathbb{P}_0(\boldsymbol{X} \in A) \ge k\,\mathbb{P}_1(\boldsymbol{X} \in A)$ for $A \subseteq R^c$

3. Show that if $A \subseteq S$ then

$$\mathbb{P}_1(\boldsymbol{X} \in R) - \mathbb{P}_1(\boldsymbol{X} \in A) \ge \frac{1}{k}(\mathbb{P}_0(\boldsymbol{X} \in R) - \mathbb{P}_0(\boldsymbol{X} \in A))$$

Hint: Write $R = (R \cap A) \cup (R \setminus A)$ and $A = (A \cap R) \cup (A \setminus R)$. Use the additivity of probability and the results in Exercise 2.

4. Consider the tests with rejection regions R and A. Recall that the *size* of a rejection region is the significance of the test with that rejection region. Use Exercise 3 to show that 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:

$$\mathbb{P}_0(\boldsymbol{X} \in R) \ge \mathbb{P}_0(\boldsymbol{X} \in A) \Rightarrow \mathbb{P}_1(\boldsymbol{X} \in R) \ge \mathbb{P}_1(\boldsymbol{X} \in A)$$

The Neyman-Pearson lemma is a beautiful result, and 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. In the following subsections, we will consider some of these special cases.

Tests for the Exponential Model

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample from the exponential distribution with scale parameter b. The sample variables might represent the lifetimes from a sample of devices of a certain type. We are interested in testing the simple hypotheses $H_0\colon\ b = b_0$ versus $H_1\colon\ b = b_1$, where $b_0 > 0$ and $b_1 > 0$ are distinct specified values.

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

$$Y = \sum_{i=1}^n X_i$$

Recall also that Y has the gamma distribution with shape parameter n and scale parameter b. For $\alpha \in (0,\ 1)$, we will denote the quantile of order α for the this distribution by $\gamma_{n,b}(\alpha)$.

5. Show that the likelihood ratio statistic is

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

6. Show that the following tests are most powerful test at the α level

a. Suppose that $b_1 > b_0$. Reject $H_0\colon\ b = b_0$ versus $H_1\colon\ b = b_1$ if and only if $Y \ge \gamma_{n,b_0}(1 - \alpha)$.
b. Suppose that $b_1 < b_0$. Reject $H_0\colon\ b = b_0$ versus $H_1\colon\ b = b_1$ if and only if $Y \le \gamma_{n,b_0}(\alpha)$.

Note that the tests in Exercise 6 do not depend on the value of b_1. This fact, together with the monotonicity of the power function can be used to shows that the tests are uniformly most powerful for the usual one-sided tests.

7. Show that

a. The test in Exercise 6 (a) is uniformly most powerful for the hypotheses $H_0\colon\ b \le b_0$ versus $H_1\colon\ b > b_0$
b. The test in Exercise 6 (b) is uniformly most powerful for the hypotheses $H_0\colon\ b \ge b_0$ versus $H_1\colon\ b < b_0$

Tests for the Bernoulli Model

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size 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 $H_0\colon\ p = p_0$ versus $H_1\colon\ p = p_1$, where $p_0 \in (0,\ 1)$ and $p_1 \in (0,\ 1)$ 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:

$$Y = \sum_{i=1}^{n} X_i$$

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 the this distribution by $b_{n,p}(\alpha)$; although since the distribution is discrete, only certain values of α are possible.

8. Show that the likelihood ratio statistic is

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

9. Show that the following tests are most powerful test at the α level

a. Suppose that $p_1 > p_0$. Reject $H_0\colon\ p = p_0$ versus $H_1\colon\ p = p_1$ if and only if $Y \ge b_{n,p_0}(1-\alpha)$.
b. Suppose that $p_1 < p_0$. Reject $H_0\colon\ p = p_0$ versus $H_1\colon\ p = p_1$ if and only if $Y \le b_{n,p_0}(\alpha)$.

Note that the tests in Exercise 9 do not depend on the value of p_1. This fact, together with the monotonicity of the power function can be used to shows that the tests are uniformly most powerful for the usual one-sided tests.

10. Show that

a. The test in Exercise 9 (a) is uniformly most powerful for the hypotheses $H_0\colon\ p \le p_0$ versus $H_1\colon\ p > p_0$
b. The test in Exercise 9 (b) is uniformly most powerful for the hypotheses $H_0\colon\ p \ge p_0$ versus $H_1\colon\ p < p_0$

Uniformly Most Powerful Tests

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 $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample, either from the Poisson distribution with parameter 1 or from the geometric distribution on $\mathbb{N}$ with parameter $\frac{1}{2}$. Thus, we wish to test the hypotheses

$$H_0: \; X \text{ has probability density function } \; f_0(x) = e^{-1}\frac{1}{x!}, \; x \in \mathbb{N}$$

$$H_1: \; X \text{ has probability density function } \; f_1(x) = \left(\frac{1}{2}\right)^{x+1}, \; x \in \mathbb{N}$$

11. Show that the likelihood ratio statistic is

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

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

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 $\boldsymbol{X}$ depends on a parameter θ, taking values in a parameter space Θ. Consider the hypotheses $H_0: \; \theta \in \Theta_0$ versus $H_1: \; \theta \notin \Theta_0$, where $\Theta_0 \subseteq \Theta$. We define

$$L(\boldsymbol{x}) = \frac{\max\{f_\theta(\boldsymbol{x}) : \theta \in \Theta_0\}}{\max\{f_\theta(\boldsymbol{x}) : \theta \in \Theta\}}$$

The function L is the **likelihood ratio function** and $L(\boldsymbol{X})$ is the **likelihood ratio statistic**. By the same reasoning as before, small values of $L(\boldsymbol{x})$ are evidence in favor of the alternative hypothesis.
