Math 541: Statistical Theory II

Likelihood Ratio Tests

Instructor: Songfeng Zheng

A very popular form of hypothesis test is the likelihood ratio test, which is a generalization of the optimal test for simple null and alternative hypotheses that was developed by Neyman and Pearson (We skipped Neyman-Pearson lemma because we are short of time). The likelihood ratio test is based on the likelihood function $f_n(X-1, \cdots, X_n|\theta)$, and the intuition that the likelihood function tends to be highest near the true value of θ. Indeed, this is also the foundation for maximum likelihood estimation. We will start from a very simple example.

1 The Simplest Case: Simple Hypotheses

Let us first consider the simple hypotheses in which both the null hypothesis and alternative hypothesis consist one value of the parameter. Suppose $X_1, \cdots, X_n$ is a random sample of size n from an exponential distribution

$$f(x|\theta) = \frac{1}{\theta}e^{-x/\theta}; \quad x > 0$$

Conduct the following simple hypothesis testing problem:

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_a : \theta = \theta_1,$$

where $\theta_1 < \theta_0$. Suppose the significant level is α.

If we assume H_0 were correct, then the likelihood function is

$$f_n(X_1, \cdots, X_n|\theta_0) = \prod_{i=1}^{n} \frac{1}{\theta_0}e^{-X_i/\theta_0} = \theta_0^{-n} \exp\{-\sum X_i/\theta_0\}.$$

Similarly, if H_1 were correct, the likelihood function is

$$f_n(X_1, \cdots, X_n|\theta_1) = \theta_1^{-n} \exp\{-\sum X_i/\theta_1\}.$$

We define the likelihood ratio as follows:

$$LR = \frac{f_n(X_1, \cdots, X_n|\theta_0)}{f_n(X_1, \cdots, X_n|\theta_1)} = \frac{\theta_0^{-n} \exp\{-\sum X_i/\theta_0\}}{\theta_1^{-n} \exp\{-\sum X_i/\theta_1\}} = \left(\frac{\theta_0}{\theta_1}\right)^{-n} \exp\left\{\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum X_i\right\}$$

Intuitively, if the evidence (data) supports H_1, then the likelihood function $f_n(X_1, \cdots, X_n|\theta_1)$ should be large, therefore the likelihood ratio is small. Thus, we reject the null hypothesis if the likelihood ratio is small, i.e. $LR \leq k$, where k is a constant such that $P(LR \leq k) = \alpha$ under the null hypothesis ($\theta = \theta_0$).

To find what kind of test results from this criterion, we expand the condition

$$\begin{aligned}
\alpha &= P(LR \leq k) = P\left(\left(\frac{\theta_0}{\theta_1}\right)^{-n} \exp\left\{\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum X_i\right\} \leq k\right) \\
&= P\left(\exp\left\{\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum X_i\right\} \leq \left(\frac{\theta_0}{\theta_1}\right)^{n} k\right) \\
&= P\left(\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum X_i \leq \log\left[\left(\frac{\theta_0}{\theta_1}\right)^{n} k\right]\right) \\
&= P\left(\sum X_i \leq \frac{\log k + n\log\theta_0 - n\log\theta_1}{\frac{1}{\theta_1} - \frac{1}{\theta_0}}\right) \\
&= P\left(\frac{2}{\theta_0}\sum X_i \leq \frac{2}{\theta_0}\frac{\log k + n\log\theta_0 - n\log\theta_1}{\frac{1}{\theta_1} - \frac{1}{\theta_0}}\right) \\
&= P\left(V \leq \frac{2}{\theta_0}\frac{\log k + n\log\theta_0 - n\log\theta_1}{\frac{1}{\theta_1} - \frac{1}{\theta_0}}\right)
\end{aligned}$$

where $V = \frac{2}{\theta_0}\sum X_i$. From the property of exponential distribution, we know under the null hypothesis, $\frac{2}{\theta_0}X_i$ follows χ^2_2 distribution, consequently, V follows a Chi square distribution with $2n$ degrees of freedom. Thus, by looking at the chi-square table, we can find the value of the chi-square statistic with $2n$ degrees of freedom such that the probability that V is less than that number is α, that is, solve for c, such that $P(V \leq c) = \alpha$. Once you find the value of c, you can solve for k and define the test in terms of likelihood ratio.

For example, suppose that $H_0 : \theta = 2$ and $H_a : \theta = 1$, and we want to do the test at a significance level $\alpha = 0.05$ with a random sample of size $n = 5$ from an exponential distribution. We can look at the chi-square table under 10 degrees of freedom to find that 3.94 is the value under which there is 0.05 area. Using this, we can obtain $P(\frac{2}{2}\sum X_i \leq 3.94) = 0.05$. This implies that we should reject the null hypothesis if $\sum X_i \leq 3.94$ in this example.

To find a rejection criterion directly in terms of the likelihood function, we can solve for k by

$$\frac{2}{\theta_0}\frac{\log k + n\log\theta_0 - n\log\theta_1}{\frac{1}{\theta_1} - \frac{1}{\theta_0}} = 3.94,$$

and the solution is $k = 0.8034$. So going back to the original likelihood ratio, we reject the null hypothesis if

$$\left(\frac{\theta_0}{\theta_1}\right)^{-n} \exp\left\{\left(\frac{1}{\theta_1} - \frac{1}{\theta_0}\right)\sum X_i\right\} = \left(\frac{2}{1}\right)^{-5} \exp\left\{\left(\frac{1}{1} - \frac{1}{2}\right)\sum X_i\right\} \leq 0.8034$$

2 General Likelihood Ratio Test

Likelihood ratio tests are useful to test a composite null hypothesis against a composite alternative hypothesis.

Suppose that the null hypothesis specifies that θ (may be a vector) lies in a particular set of possible values, say Θ_0, i.e. $H_0 : \theta \in \Theta_0$; the alternative hypothesis specifies that Θ lies in another set of possible values Θ_a, which does not overlap Θ_0, i.e. $H_a : \theta \in \Theta_a$. Let $\Theta = \Theta_0 \cup \Theta_a$. Either or both of the hypotheses H_0 and H_a can be compositional.

Let $L(\hat{\Theta}_0)$ be the maximum (actually the supremum) of the likelihood function for all $\theta \in \Theta_0$. That is, $L(\hat{\Theta}_0) = \max_{\theta\in\Theta_0} L(\theta)$. $L(\hat{\Theta}_0)$ represents the best explanation for the observed data for all $\theta \in \Theta_0$. Similarly, $L(\hat{\Theta}) = \max_{\theta\in\Theta} L(\theta)$ represents the best explanation for the observed data for all $\theta \in \Theta = \Theta_0 \cup \Theta_a$. If $L(\hat{\Theta}_0) = L(\hat{\Theta})$, then a best explanation for the observed data can be found inside Θ_0 and we should not reject the null hypothesis $H_0 : \theta \in \Theta_0$. However, if $L(\hat{\Theta}_0) < L(\hat{\Theta})$, then the best explanation for the observed data could be found inside Θ_a, and we should consider rejecting H_0 in favor of H_a. A likelihood ratio test is based on the ratio $L(\hat{\Theta}_0)/L(\hat{\Theta})$.

Define the likelihood ratio statistic by

$$\Lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\theta)},$$

A likelihood ratio test of $H_0 : \theta \in \Theta_0$ vs.$H_a : \theta \in \Theta_a$ employs Λ as a test statistic, and the rejection region is determined by $\Lambda \le k$.

Clearly, $0 \le \Lambda \le 1$. A value of Λ close to zero indicates that the likelihood of the sample is much smaller under H_0 than it is under H_a, therefore the data suggest favoring H_a over H_0. The actually value of k is chosen so that α achieves the desired value.

A lot of previously introduced testing procedure can be reformulated as likelihood ratio test, such at the example below:

Example 1: Testing Hypotheses about the mean of a normal distribution with unknown variance. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ is a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. We wish to test the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_a : \mu \neq \mu_0$$

Solution: In this example, the parameter is $\theta = (\mu, \sigma^2)$. Notice that Θ_0 is the set $\{(\mu_0, \sigma^2) : \sigma^2 > 0\}$, and $\Theta_a = \{(\mu, \sigma^2) : \mu \neq \mu_0, \sigma^2 > 0\}$, and hence that $\Theta = \Theta_0 \cup \Theta_a = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$. The value of the constant σ^2 is completely unspecified. We must now find $L(\hat{\Theta}_0)$ and $L(\hat{\Theta})$.

For the normal distribution, we have

$$L(\theta) = L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}\right].$$

Restricting μ to Θ_0 implies that $\mu = \mu_0$, and we can find $L(\hat{\Theta}_0)$ if we can determine the value of σ^2 that maximizes $L(\mu, \sigma^2)$ subject to the constraint that $\mu = \mu_0$. It is easy to see that when $\mu = \mu_0$, the value of σ^2 that maximizes $L(\mu_0, \sigma^2)$ is

$$\hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2.$$

Thus, $L(\hat{\Theta}_0)$ can be obtained by replacing μ with μ_0 and σ^2 with $\hat{\sigma}_0^2$ in $L(\mu, \sigma^2)$, which yields

$$L(\hat{\Theta}_0) = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n \exp\left[-\sum_{i=1}^n \frac{(X_i-\mu_0)^2}{2\hat{\sigma}_0^2}\right] = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{-n/2}.$$

We now turn to finding $L(\hat{\Theta})$. Let $(\hat{\mu}, \hat{\sigma}^2)$ be the point in the set Θ which maximizes the likelihood function $L(\mu, \sigma^2)$, by the method of maximum likelihood estimation, we have

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})^2.$$

Then $L(\hat{\Theta})$ is obtained by replacing μ with $\hat{\mu}$ and σ^2 with $\hat{\sigma}^2$, which gives

$$L(\hat{\Theta}) = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n \exp\left[-\sum_{i=1}^n \frac{(X_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right] = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{-n/2}.$$

Therefore, the likelihood ratio is calculated as

$$\Lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}_0}\right)^n e^{-n/2}}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{-n/2}} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} = \left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \mu_0)^2}\right]^{n/2}$$

Notice that $0 < \Lambda \leq 1$ because $\Theta_0 \subset \Theta$, thus when $\Lambda < k$ we would reject H_0, where $k < 1$ is a constant. Because

$$\sum_{i=1}^n (X_i - \mu_0)^2 = \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu_0)]^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2,$$

the rejection region, $\Lambda < k$, is equivalent to

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \mu_0)^2} < k^{2/n} = k'$$

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2} < k'$$

$$\frac{1}{1 + \frac{n(\bar{X}-\mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} < k'.$$

This inequality is equivalent to

$$\frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} > \frac{1}{k'} - 1 = k''$$

$$\frac{n(\bar{X} - \mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2} > (n-1)k''$$

By defining

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2,$$

the above rejection region is equivalent to

$$\left|\frac{\sqrt{n}(\bar{X} - \mu_0)}{S}\right| > \sqrt{(n-1)k''}.$$

We can recognize that $\sqrt{n}(\bar{X} - \mu_0)/S$ is the t statistic employed in previous sections, and the decision rule is exactly the same as previous. Consequently, in this situation, the likelihood ratio test is equivalent to the t test. For two-sided tests, we can also verify that likelihood ratio test is equivalent to the t test.

Example 2: Suppose $X_1, \cdots, X_n$ from a normal distribution $N(\mu, \sigma^2)$ where both μ and σ are unknown. We wish to test the hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_a : \sigma^2 \neq \sigma_0^2$$

at the level α. Show that the likelihood ratio test is equivalent to the χ^2 test.

Solution: The parameter is $\theta = (\mu, \sigma^2)$. Notice that Θ_0 is the set $\{(\mu, \sigma_0^2) : -\infty < \mu < \infty\}$, and $\Theta_a = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 \neq \sigma_0^2\}$, and hence that $\Theta = \Theta_0 \cup \Theta_a = \{(\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}$. We must now find $L(\hat{\Theta}_0)$ and $L(\hat{\Theta})$.

For the normal distribution, we have

$$L(\theta) = L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^2}\right].$$

In the subset Θ_0, we have $\sigma^2 = \sigma_0^2$, and we can find $L(\hat{\Theta}_0)$ if we can determine the value of μ that maximizes $L(\mu, \sigma^2)$ subject to the constraint that $\sigma^2 = \sigma_0^2$. It is easy to see that the value of μ that maximizes $L(\mu, \sigma_0^2)$ is $\hat{\mu}_0 = \bar{X}$. Thus, $L(\hat{\Theta}_0)$ can be obtained by replacing μ with $\hat{\mu}_0$ and σ^2 with σ_0^2 in $L(\mu, \sigma^2)$, which yields

$$L(\hat{\Theta}_0) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(X_i - \hat{\mu}_0)^2}{2\sigma_0^2}\right].$$

Next, We find $L(\hat{\Theta})$. Let $(\hat{\mu}, \hat{\sigma}^2)$ be the point in the set Θ which maximizes the likelihood function $L(\mu, \sigma^2)$, by the method of maximum likelihood estimation, we have

$$\hat{\mu} = \bar{X} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2.$$

Then $L(\hat{\Theta})$ is obtained by replacing μ with $\hat{\mu}$ and σ^2 with $\hat{\sigma}^2$, which gives

$$L(\hat{\Theta}) = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(X_i - \hat{\mu})^2}{2\hat{\sigma}^2}\right] = \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{-n/2}.$$

Therefore, the likelihood ratio is calculated as

$$\Lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{\left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(X_i-\hat{\mu}_0)^2}{2\sigma_0^2}\right]}{\left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^n e^{-n/2}} = e^{n/2}\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{n/2} \exp\left[-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right]$$

Notice that $0 < \Lambda \leq 1$ because $\Theta_0 \subset \Theta$, thus when $\Lambda < k$ we would reject H_0, where $k < 1$ is a constant. The rejection region, $\Lambda < k$, is equivalent to

$$\left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{n/2} \exp\left[-\frac{n}{2}\frac{\hat{\sigma}^2}{\sigma_0^2}\right] < ke^{-n/2} = k'$$

Viewing the left hand side as a function of $\hat{\sigma}^2/\sigma_0^2$, the above inequality holds if $\hat{\sigma}^2/\sigma_0^2$ is too big or too small, i.e.

$$\frac{\hat{\sigma}^2}{\sigma_0^2} < a \quad \text{or} \quad \frac{\hat{\sigma}^2}{\sigma_0^2} > b$$

This inequality is equivalent to

$$\frac{n\hat{\sigma}^2}{\sigma_0^2} < na \quad \text{or} \quad \frac{n\hat{\sigma}^2}{\sigma_0^2} > nb.$$

We can recognize that $n\hat{\sigma}^2/\sigma_0^2$ is the χ^2 statistic employed in previous sections, and the decision rule is exactly the same as previous. Consequently, in this situation, the likelihood ratio test is equivalent to the χ^2 test.

Likelihood ratio statistic Λ is a function of the sample $X_1, \cdots, X_n$, and we can prove that it only depends on the sample through a sufficient statistic. Formally, suppose $X_1, \cdots, X_n$ is a random sample from the distribution $f(x|\theta)$, where $\theta \in \Theta$ is the unknown parameter (vector). Furthermore, assume that $T(\mathbf{X})$ is a sufficient statistic, then by factorization theorem the joint distribution of $X_1, \cdots, X_n$ can be decomposed as

$$f(\mathbf{x}|\theta) = u(\mathbf{x})v[T(\mathbf{x}), \theta],$$

which is also the likelihood function.

Let us assume we want to test the hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_a : \theta \in \Theta_a$$

where Θ_0 and Θ_a are disjoint subsets of the parameter space Θ, and $\Theta_0 \cup \Theta_a = \Theta$. Using likelihood ratio test, we first need to find the maximal points in Θ_0 and Θ. In Θ_0, let

$$\hat{\theta}_0 = \arg\max_{\theta\in\Theta_0} f(\mathbf{X}|\theta) = \arg\max_{\theta\in\Theta_0} u(\mathbf{X})v[T(\mathbf{X}), \theta] = \arg\max_{\theta\in\Theta_0} v[T(\mathbf{X}), \theta],$$

then clearly $\hat{\theta}_0$ depends on the data $X_1, \cdots, X_n$ only through the sufficient statistic $T(\mathbf{X})$, and let us denote this relation as $\hat{\theta}_0 = g(T(\mathbf{X}))$. Similarly, in the set Θ, we have

$$\hat{\theta} = \arg\max_{\theta\in\Theta} f(\mathbf{X}|\theta) = \arg\max_{\theta\in\Theta} u(\mathbf{X})v[T(\mathbf{X}), \theta] = \arg\max_{\theta\in\Theta} v[T(\mathbf{X}), \theta] = h(T(\mathbf{X})).$$

Therefore, the likelihood ratio statistic Λ can be calculated as

$$\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = \frac{U(\mathbf{X})v[T(\mathbf{X}), \hat{\theta}_0]}{U(\mathbf{X})v[T(\mathbf{X}), \hat{\theta}]} = \frac{v[T(\mathbf{X}), g(T(\mathbf{X}))]}{v[T(\mathbf{X}), h(T(\mathbf{X}))]},$$

which depends on the sufficient statistic only. For example, in example 1, the final likelihood ratio test depends on $\bar{X}$ and S, and we know that $\bar{X}$ is a sufficient statistic for μ, and S is a sufficient statistic for σ.

Here, we see the importance of sufficient statistic another time. Previously, we saw that MLE and Bayesian estimators are functions of sufficient statistics, and in exponential family, the efficient estimator is a linear function of sufficient statistics.

It can be verified that the t test and F test used for two sample hypothesis testing problems can also be reformulated as likelihood ratio test. Unfortunately, the likelihood ratio method does not always produce a test statistic with a known probability distribution. If the sample size is large, however, we can obtain an approximation to the distribution of Λ if some reasonable "regularity conditions" are satisfied by the underlying population distribution(s). These are general conditions that hold for most (but not all) of the distributions that we have considered. The regularity conditions mainly involve the existence of derivatives, with respect to the parameters, of the likelihood function. Another key condition is that the region over which the likelihood function is positive cannot depend on unknown parameter values. In summary, we have the following theorem:

Theorem. Let $X_1, \cdots, X_n$ have joint likelihood function $L(\Theta)$. Let r_0 be the number of free parameters under the null hypothesis $H_0 : \theta \in \Theta_0$, and let r be the number of free parameters under the alternative hypothesis $H_a : \theta \in \Theta_a$. Then, for large sample size n, the null distribution of $-2 \log \Lambda$ has approximately a χ^2 distribution with $r - r_0$ degrees of freedom.

In example 1, the null hypothesis specifies $\mu = \mu_0$ but does not specify σ^2, so there is one free parameter, $r_0 = 1$; under the alternative hypothesis, there are two free parameters, so $r = 2$. For this example, the null distribution of $-2 \log \Lambda$ is exactly χ_1^2.

Example 2: Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a manufacturing plant. 100 independent observations on the number of complaints gave meas $\bar{x} = 20$ for shift 1 and $\bar{y} = 22$ for shift 2. Assume that the number of complaints per week on the i-th shift has a Poisson distribution with mean θ_i, for $i = 1, 2$. Use the likelihood ratio method to test $H_0 : \theta_1 = \theta_2$ versus $H_a : \theta_1 \neq \theta_2$ with significance level $\alpha = 0.01$.

Solution. The likelihood function of the sample is now the joint probability function of all x_i's and y_i's, and is given by

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} \frac{\theta_1^{x_i} e^{-\theta_1}}{x_i!} \prod_{i=1}^{n} \frac{\theta_2^{y_i} e^{-\theta_2}}{y_i!} = \left(\frac{1}{k}\right) \theta_1^{\sum x_i} e^{-n\theta_1} \theta_2^{\sum y_i} e^{-n\theta_2}$$

where $k = x_1! \cdots x_n! y_1! \cdots y_n!$, and $n = 100$. In this example $\Theta_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$, where θ is unknown. Hence, under H_0 the likelihood function is a function of a single parameter θ, and

$$L(\theta) = \left(\frac{1}{k}\right) \theta^{\sum x_i + \sum y_i} e^{-2n\theta}.$$

Maximizing function $L(\theta)$, and we can find

$$\hat{\theta} = \frac{1}{2n} \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \right) = \frac{1}{2}(\bar{x} + \bar{y})$$

In this example, $\Theta_a = \{(\theta_1, \theta_2) : \theta_1 \neq \theta_2\}$, and $\Theta = \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}$. Using the general likelihood function $L(\theta_1, \theta_2)$, we see that $L(\theta_1, \theta_2)$ is maximized when $\hat{\theta}_1 = \bar{x}$ and $\hat{\theta}_2 = \bar{y}$, respectively. That is, $L(\theta_1, \theta_2)$ is maximized when both θ_1 and θ_2 are replaced by their maximum likelihood estimates. Thus,

$$\Lambda = \frac{L(\hat{\Theta}_0)}{L(\hat{\Theta})} = \frac{k^{-1} (\hat{\theta})^{n\bar{x} + n\bar{y}} e^{-2n\hat{\theta}}}{k^{-1} (\hat{\theta}_1)^{n\bar{x}} (\hat{\theta}_2)^{n\bar{y}} e^{-n\theta_1 - n\theta_2}} = \frac{(\hat{\theta})^{n\bar{x} + n\bar{y}}}{(\bar{x})^{n\bar{x}} (\bar{y})^{n\bar{y}}}.$$

The observed value of $\hat{\theta}$ is

$$\hat{\theta} = \frac{1}{2}(\bar{x} + \bar{y}) = (20 + 22)/2 = 21,$$

therefore the observed value of Λ is

$$\lambda = \frac{21^{100(20+22)}}{20^{100(20)}22^{100(22)}} \quad \text{and} \quad -2\log\lambda = 9.53.$$

In this example, the number of free parameters in $\Theta = \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}$ is $r = 2$; the number of free parameters in $\Theta_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$ is $r_0 = 1$. Therefore, the approximate distribution of $-2\log\Lambda$ is a χ^2 distribution with $r - r_0 = 1$ degree of freedom. Small values of λ correspond to large values of $-2\log\lambda$, so the rejection region for a test at level α contains the value of $-2\log\lambda$ that exceed $\chi_1^2(1-\alpha)$. In this example, $\alpha = 0.01$, we have the critical value is 6.635, the value that cuts off an area of 0.99 in the left-hand tail of χ_1^1. Because the observed value of $-2\log\lambda$ is larger than 6.635, we reject the null hypothesis $H_0 : \theta_1 = \theta_2$. We conclude, at the significance level $\alpha = 0.01$ that the mean numbers of complaints filed by the union stewards do differ.

3 Exercises

Exercise 1. Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. We wish to test the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_a : \mu \neq \mu_0$$

at the level α. Show that the likelihood ratio test is equivalent to the z test.

Exercise 2. Suppose $X_1, \cdots, X_n$ from a normal distribution $N(\mu, \sigma^2)$ where both μ and σ are unknown. We wish to test the hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_a : \sigma^2 > \sigma_0^2$$

at the level α. Show that the likelihood ratio test is equivalent to the χ^2 test.