Likelihood Ratio Test

April 29, 2020

1 Likelihood Ratio Test

Here is another example of the Likelihood Ratio Test.

EXAMPLE 1.1. *Let $Y_1, Y_2, ..., Y_n$ be a random sample from a normal distribution with unknown mean μ and variance σ^2.*

Show that the Likelihood Ratio Test to test the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$ versus the alternative hypothesis $H_0 : \sigma^2 > \sigma_0^2$ is equivalent to the small sample size test for the variance of a random sample of normal random variables.

Our goal is to show the the RR region we get from this likelihood ratio test is can be computed in terms of the χ^2-distributed random variable $W = \sum_{i=1}^n (Y_i - \overline{Y})$.

The unknown vector is $\Theta = (\mu, \sigma^2)$. The hypothesis test is not simple because it does not specify μ. The null hypothesis corresponds to the region

$$\Omega_0 = \{(\mu, \sigma^2) | \mu \in \mathbb{R} \text{ and } \sigma^2 = \sigma_0^2\}.$$

(You can graph this set on the (μ, σ^2) plane to visualize it, it will be a line)

The alternative hypothesis corresponds to the region

$$\Omega_a = \{(\mu, \sigma^2) | \mu \in \mathbb{R} \text{ and } \sigma^2 > \sigma_0^2\}.$$

(You can graph this set to visualize it, it will be a half of a plane, it's boundary is the half line from Ω_0)

So the region of the parameter space of interest is

$$\Omega = \Omega_0 \cup \Omega_a = \{(\mu, \sigma^2) | \mu \in \mathbb{R} \text{ and } \sigma^2 \geq \sigma_0^2\}.$$

Now our goal is determine the maximum value of $L(\Theta)$ on Ω, given the observation $Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n$, on the sets Ω_a and Ω_0. So our first step is to compute the likelihood function:

$$L(\Theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2}}$$

Note that we only write σ^2, and don't simplify the square root.

Our goal is to now find where the maximum of this function occurs, so as usual we take its ln.

$$\ln(L(\Theta)) = \frac{n}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) - \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma^2} = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \mu)^2$$

We now wish find where the max of this function occurs. From calculus, we know this value occurs when the derivative is zero, or at the boundary of the domain. The second case will be important for us, because we are considering maximums restricted to certain subsets of the entire parameter space.

Because we have a function of two parameters, μ and σ^2, we take the derivative of each parameter.

$$\frac{d}{d\mu}\ln(L(\Theta)) = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)$$

$$\frac{d}{d\sigma^2}\ln(L(\Theta)) = -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

The derivative w.r.t. μ is 0 at

$$-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)=0$$

or

$$\mu=\frac{1}{n}\sum_{i=1}^{n}y_i=\overline{y}$$

Since there are no restrictions on μ in Ω, this point is contained in Ω.

We now find the maximum where the maximum occurs in the σ^2 coordinate.

Solving

$$\frac{d}{d\sigma^2}\ln(L(\Theta)) = -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2=0$$

for σ^2 we find that to be at a maximum of $L(\Theta)$ we get:

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2$$

But this point might not be in Ω, so now we check the sign of the derivative. If σ^2 is large then the first term of $-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2$ will be bigger, so the derivative is negative, on the other hand if σ^2 is small the second term is bigger and the derivative is positive.

So combining this with the maximum in the μ occurs at $\hat{\mu}$ we have that the maximum of $L(\Theta)$ occurs at:

$$\hat{\Theta}=(\hat{\mu},\hat{\sigma}^2)$$

where

$$\hat{\mu}=\overline{y}$$

and

$$\hat{\sigma}^2=\begin{cases}\sigma_0^2, & \text{if } \frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2<\sigma_0^2\\ \frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2 & \text{if } \frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2\geq\sigma_0^2\end{cases}$$

We substituted $\mu=\hat{\mu}=\overline{y}$ into the σ^2 maximum.

We now evaluate the Likelihood function at this point:

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

(I've just written the same thing two different ways.) Note that the term in the exponent is different depending on the value of σ^2

$$-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\overline{y})^2=\begin{cases} -\frac{1}{2\sigma_0^2}\sum_{i=1}^n(y_i-\overline{y})^2, & \text{if } \frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2<\sigma_0^2 \\ -\frac{n}{2} & \text{if } \frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2\geq\sigma_0^2 \end{cases}$$

We want to consider the ratio

$$\frac{\max_{\Theta\in\Omega_0}L(\Theta)}{\max_{\Theta\in\Omega}L(\Theta)}$$

If $\hat{\sigma}^2=\sigma_0^2$ then the maximum over Ω_0 and Ω occur at the same point, and this ratio is 1. So we'll now assume $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2$, and thus the maximum of L on Ω is:

$$L(\hat{\Theta})=\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n e^{-n/2}=\left(\frac{1}{2\pi\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2}\right)^{n/2}e^{-n/2}$$

So the likelihood ratio is:

$$\frac{\max_{\Theta\in\Omega_0}L(\Theta)}{\max_{\Theta\in\Omega}L(\Theta)}=\frac{\left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2}e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n(y_i-\overline{y})^2}}{\left(\frac{1}{2\pi\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2}\right)^{n/2}e^{-n/2}}=\left(\frac{\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2}{2\pi\sigma_0^2}\right)^{n/2}e^{\frac{n}{2}-\frac{1}{2}\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}}$$

The region rejection will then be determined by when this quantity is $<k$ for some small k. How small k needs to be is determined by the choice of α. In order go from a level α to a value for k, we need to know the distribution of the ratio. At this stage, we do not it's distribution, so we will manipulate it until we have a random variable that we know. Just as in the last section, this will create a complicated function of k. Fortunately, we don't have to follow it to closely, the most important thing will be the direction of the inequality.

When H_0 holds $\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}$ is a χ^2 random variable with $(n-1)$ df. So we should try to isolate it. Unfortunately, we can't solve for it very explicitly, but the the likelihood ratio is a function of this quantity. Namely, the function

$$g(x)=\left(\frac{x}{2\pi n}\right)^{n/2}e^{\frac{n}{2}-\frac{1}{2}x}$$

evaluated at $x=\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}$ is this likelihood ratio.

So we want to characterize the region $g(x)<k$, so we'll determine when the function $g(x)$ is decreasing:

$$g'(x)=\left(\frac{1}{2\pi n}\right)^{n/2}\frac{n}{2}x^{n/2-1}e^{\frac{n}{2}-\frac{1}{2}x}-\left(\frac{1}{2\pi n}\right)^{n/2}x^{n/2}\frac{1}{2}e^{\frac{n}{2}-\frac{1}{2}x}$$

so we see this function is decreasing when $x>n$. Recall that we have assumed $\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2\geq\sigma_0^2$ (Otherwise we would automatically fail to reject H_0). So the likelihood ratio is a decreasing function of $\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}$.

This means we reject H_0 when $\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}$ is large. The RR is therefore

$$RR=\sum_{i=1}^n\frac{(y_i-\overline{y})^2}{\sigma_0^2}>a$$

where a is such that

$$\mathbb{P}(W > a) = \alpha$$

where W is a χ^2 random variable with $(n-1)$ df.

So we have derived the small-sample size test for the varaince of normal random variables. This is the end of problem.

The last two examples used very strong properties of the distribution of the random sample $Y_1,...,Y_n$ and did some clever manipulations to find the rejection region. Generally we cannot hope to be able to do this, fortunately in the large sample size regime, we have the following theorem:

THEOREM 1.2. *Let $Y_1,Y_2,...,Y_n$ be a random sample with likelihood function $L(\Theta)$, where the unknown parameter $\Theta = (\theta_1,\theta_2,...,\theta_r)$ has r coordinates. Let $H_0 : \Theta \in \Omega_0$ be a set that specifies d_0 of the r coordinates of Θ. Let $H_a : \Theta \in \Omega_a$ and $\Omega = \Omega_0 \cup \Omega_a$. Let d of the r coordinates of Ω be specified. Let λ we the likelihood ratio:*

$$\lambda = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)},$$

then for large n, under H_0, the random variable $-2\ln(\lambda)$ has an approximately χ^2 distribution with $(d_0 - d)$ df.

For example, in the last examples we considered $\Theta = (\mu,\sigma^2)$, so $r = 2$. Under H_0, either μ or σ^2 is specified, so the other parameter is free and $r_0 = 1$. Under H_a neither parameter is specified so $d = 0$. Typically $d = 0$, because the parameters fixed by Ω could just not be included in Θ.

If the ratio $\lambda < k$ then $-2\ln(\lambda) > -2\ln(k) = k^*$, so if the test statistic $-2\ln(\lambda)$ is large we reject H_0. Let's look at an example:

EXAMPLE 1.3. *A company wants to know if its product is used equally or not in different cities. To see this, they conduct a random sample in each of the three cities and ask if their product is used or not. They collect the following data:*

	City 1	*City 2*	*City 3*
Sample size:	100	100	200
Number who use product:	14	27	47

Use the likelihood ration at level $\alpha = .05$ to determine if the null hypothesis that the same percentage of the population in each city uses the product.

Solution:

Let p_i for $i = 1,2,3$ be the proportion of people in each city that use the product. Then we are testing the null hypothesis $H_0 : p_1 = p_2 = p_3$ against the alternative hypothesis $H_a : p_i \neq p_j$ for $i \neq j$.

So the set Ω_0 is $\{(p_1,p_2,p_3) | 0 \leq p_1 = p_2 = p_3 \leq 1\}$, so there is 1 fixed parameter and 2 free parameters. The set Ω_a is $\{(p_1,p_2,p_3) | 0 \leq p_i \leq 1$ and $p_i \neq p_j$ for $i \neq j.\}$. So then $\Omega = \Omega_0 \cup \Omega_a$ is $\{(p_1,p_2,p_3) | 0 \leq p_i \leq 1\}$ so there are no fixed parameters, all 3 are free parameters.

The estimator for p_i that was used is a Binomial random variable with n equal to the sample size in the city (either 100 or 200) and unknown p_i.Let y_i be the total number of people in city i who use the product. Then the Likelihood function is

$$L(p_1,p_2,p_3) = \binom{100}{y_1} p_1^{y_1}(1-p_1)^{100-y_1} \binom{100}{y_2} p_2^{y_2}(1-p_2)^{100-y_2} \binom{200}{y_3} p_3^{y_3}(1-p_3)^{200-y_3}.$$

This is just the product of the pdfs of 3 binomial random variables, with sample size 100, 100, and 200.

We'll derive the rejection region for this test and then substitute the data we observed to see if we should reject H_0 or not.

When H_0 holds then $p_1 = p_2 = p_3$. Now we find the the maximum of L on this set:

$$L(p_1,p_2,p_3) = L(p) = \binom{100}{y_1} p^{y_1}(1-p)^{100-y_1} \binom{100}{y_2} p^{y_2}(1-p)^{100-y_2} \binom{200}{y_3} p^{y_3}(1-p)^{200-y_3}$$

$$= \binom{100}{y_1}\binom{100}{y_2}\binom{200}{y_3} p^{y_1+y_2+y_3}(1-p)^{400-y_1-y_2-y_3}$$

so

$$\frac{d}{dp}L(p) = \binom{100}{y_1}\binom{100}{y_2}\binom{200}{y_3}\Big((y_1+y_2+y_3)p^{y_1+y_2+y_3-1}(1-p)^{400-y_1-y_2-y_3}$$
$$-p^{y_1+y_2+y_3}(400-y_1-y_2-y_3)(1-p)^{400-y_1-y_2-y_3-1}\Big)$$

which equals zero at $p = \frac{y_1+y_2+y_3}{400}$.

On the other hand on the entire Ω

$$\frac{d}{dp_1}L(p_1,p_2,p_3) = \binom{100}{y_1}\binom{100}{y_2}\binom{200}{y_3}\Big(y_1 p^{y_1-1}(1-p)^{100-y_1} - (100-y_1)p^{y_1}(1-p)^{100-y_1-1}\Big)$$
$$\times p_2^{y_2}(1-p_2)^{100-y_2} p_3^{y_3}(1-p_3)^{200-y_3}$$

which is 0 when $p_1 = \frac{y_1}{100}$.

Computing the derivatives in the other coordinates one sees that with function is maximized at $p_1 = \frac{y_1}{100}, p_2 = \frac{y_2}{100}, p_3 = \frac{y_3}{200}$.

Now we take the ratio of the likelihood functions evaluated at these two points:

$$\lambda = \frac{\max_{\Theta\in\Omega_0} L(\Theta)}{\max_{\Theta\in\Omega} L(\Theta)}$$
$$= \frac{\left(\frac{y_1+y_2+y_3}{400}\right)^{y_1+y_2+y_3}(1-\left(\frac{y_1+y_2+y_3}{400}\right))^{400-y_1-y_2-y_3}}{\left(\frac{y_1}{100}\right)^{y_1}(1-\left(\frac{y_1}{100}\right))^{100-y_1}\left(\frac{y_2}{100}\right)^{y_2}(1-\left(\frac{y_2}{100}\right))^{100-y_2}\left(\frac{y_3}{200}\right)^{y_3}(1-\left(\frac{y_3}{200}\right))^{200-y_3}}$$

The binomial coefficients cancel, so they're not written. Under H_0, we have no idea the distribution of the above random variable (when y_i is replaced by Y_i), and it's also not clear how to manipulate it into something we know the distribution of. Fortunately we have a theorem that tells us $-2\ln(\lambda)$ is a χ^2 random variable with $2-0$ df.

From the table we have

$$\mathbb{P}(W > 5.99) = .05$$

where W is a χ^2 r.v. with 2 df.

So we just need to compute the above fraction at with the data give and then check if $-2\ln(\lambda)$ is sufficiently large.

For our observation we have $\lambda = .056$ so $-2\ln(\lambda) = 5.78$, since this is less than 5.99, we do not reject H_0 at the $\alpha = .05$ level.