Likelihood Ratio Test

April 27, 2020

1 Likelihood Ratio Test

In the last section, the Neyman-Pearson Lemma gave us the most powerful test when testing two simple hypothesis. We also saw that that when the alternative hypothesis is one-sided, the Neyman-Pearson Lemma might also give the uniformly most powerful test. Unfortunately, beyond these cases the parameter space becomes too large to hope for a uniformly most powerful test. Nevertheless, the **Likelihood Ratio Test** is a natural generalization of the test from the Neyman-Pearson Lemma, when the null hypothesis is not simple.

HYPOTHESIS TEST 1.1 (Likelihood Ratio Test). *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample with likelihood function $L(\Theta)$ where Θ is an unknown vector of parameters. Let Ω_0 be the set of all parameters in H_0 and Ω_a be the set of all parameters in H_a. Let $\Omega = \Omega_0 \cup \Omega_a$.*

The RR of the Likelihood Ratio Test is of the form:

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

for some $1 > k > 0$.

We note that when the maximum of the likelihood function occurs in Ω_0 then the above ratio is 1 and we fail to reject the null hypothesis. When the maximum value of $L(\Theta)$ when evaluated at $\Theta \in \Omega_a$ is much bigger than the maximum value in Ω_0 then the above ratio will be small, and depending on our desired confidence we should be led to reject H_0.

Before looking at the test in more detail let's consider an example. (Warning: these problems are long.)

EXAMPLE 1.2. *Let $Y_1, Y_2, \ldots, Y_n$ 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 $H_0 : \mu = \mu_0$ versus the alternative hypothesis $H_0 : \mu < \mu_0$ is equivalent to the small sample size test for the mean 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 t-distributed random variable $T = \frac{\sqrt{n}(\overline{Y} - \mu_0)}{S}$. This will be a rather involved computation. My compute doesn't seem to like putting the box around such a long answer, so the end of this problem is on page 6 where it says "This is the end of problem."

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

$$\Omega_0=\{(\mu,\sigma^2)|\mu=\mu_0 \text{ and } \sigma^2>0\}.$$

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

The alternative hypothesis corresponds to the region

$$\Omega_a=\{(\mu,\sigma^2)|\mu<\mu_0 \text{ and } \sigma^2>0\}.$$

(You can graph this set to visualize it, it will be a quarter of a plane, it's boundary is the half line from Ω_0 and the line $\sigma^2=0$)

So the region of the parameter space of interest is

$$\Omega=\Omega_0\cup\Omega_a=\{(\mu,\sigma^2)|\mu\leq\mu_0 \text{ and } \sigma^2>0\}.$$

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.

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

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

Without any restrictions $\frac{d}{d\mu}\ln(L(\Theta))=0$ when

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

or

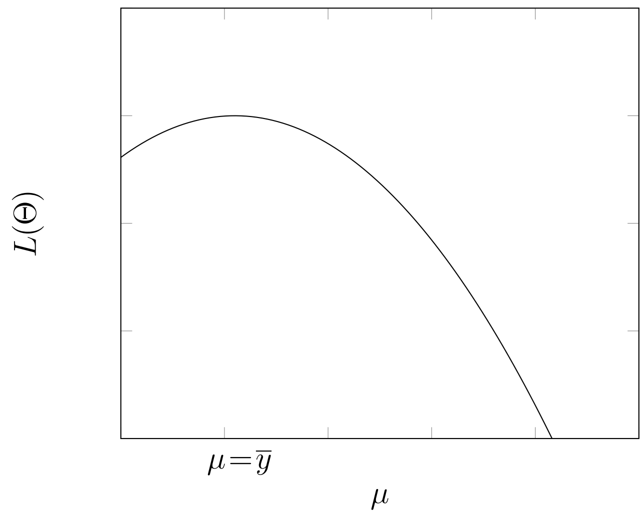
$$\mu=\frac{1}{n}\sum_{i=1}^n y_i=\overline{y}$$

but this point might not be in Ω or Ω_0. We therefore compute the sign of the derivative, to see where it is increasing or decreasing.

$$\frac{d}{d\mu}\ln(L(\Theta)) < 0 \text{ if } \mu > \overline{y}$$

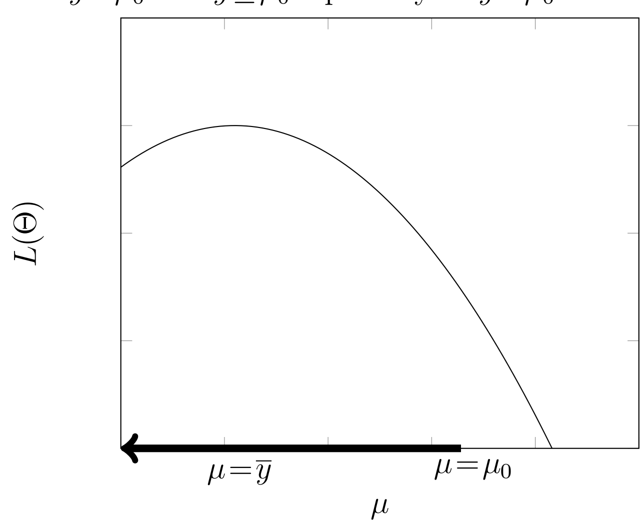
$$\frac{d}{d\mu}\ln(L(\Theta)) > 0 \text{ if } \mu < \overline{y}$$

So we have the picture:



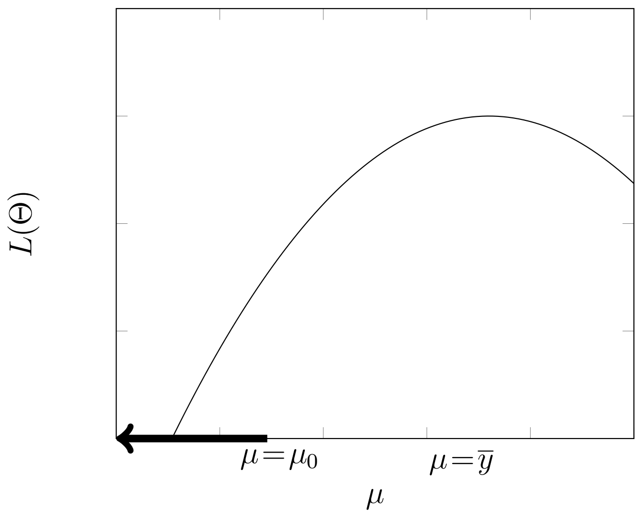
In this case it turns out the location of the maximum doesn't depend on σ^2.

To determine $\hat{\mu}$, the location of the maximum on $\Omega = \{(\mu,\sigma^2)|\mu \leq \mu_0 \text{ and } \sigma^2 > 0\}$, we need to consider the case $\overline{y} < \mu_0$ and $\overline{y} \geq \mu_0$ separately. If $\overline{y} < \mu_0$ then our picture is



and we see the maximum occurs at $\mu = \overline{y}$.

On the other hand, if $\overline{y} \geq \mu_0$ then our picture is



and we see the maximum (when restricted to the set $\mu \leq \mu_0$) occurs at $\mu = \mu_0$.
Putting these two cases together we have $L(\Theta)$ is maximized on Ω when μ equals

$$\hat{\mu} = \begin{cases} \overline{y}, & \text{if } \overline{y} < \mu_0 \\ \mu_0 & \text{if } \overline{y} \geq \mu_0 \end{cases}$$

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 require:

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

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

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

and

$$\hat{\mu} = \begin{cases} \overline{y}, & \text{if } \overline{y} < \mu_0 \\ \mu_0 & \text{if } \overline{y} \geq \mu_0 \end{cases}$$

Note that $\hat{\sigma}^2$ is always in Ω we don't need to check any boundary conditions (if you want to be very careful note that as $\sigma^2 \to 0, \infty$, $L \to 0$).
In contrast to some of the previous sections, we actually want to know the value of the likelihood function at it's maximum, so now we evaluate it:

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

(I've just written the same thing two different ways.) Now we simplify the exponent

Then we substitute $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(y_i-\hat{\mu})^2$ and get that

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

Note that this holds for either value of $\hat{\mu}$.

So

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

We've just substituted $\hat{\sigma}$ and moved the square root into the exponent.

Now we consider the ratio

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

If $\hat{\mu}=\mu_0$, then the maximum over Ω is the same as the maximum over the smaller set Ω_a, and this ratio is 1. In this case we will definitely not reject H_0. (Once you sort through the meaning of everything, this is actually obvious, $\overline{y}\geq\mu_0$ so we don't want to accept $H_a:\mu<\mu_0$.)

The more interesting case is when $\overline{y}<\mu_0$ and $\hat{\mu}=\overline{y}$, in this case the likelihood ratio is:

$$\frac{\max_{\Theta\in\Omega_0}L(\Theta)}{\max_{\Theta\in\Omega}L(\Theta)}=\frac{\left(\frac{1}{2\pi\frac{1}{n}\sum_{i=1}^n(y_i-\mu_0)^2}\right)^{n/2}e^{-n/2}}{\left(\frac{1}{2\pi\frac{1}{n}\sum_{i=1}^n(y_i-\overline{y})^2}\right)^{n/2}e^{-n/2}}=\left(\frac{\sum_{i=1}^n(y_i-\overline{y})^2}{\sum_{i=1}^n(y_i-\mu_0)^2}\right)^{n/2}$$

To compute the max on Ω_0, we used that $\mu=\mu_0$, and all the previous analysis is the exact same.

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.

In order to make the numerator and denominator look more similar we will use the identity:

$$\begin{aligned}\sum_{i=1}^n(y_i-\mu_0)^2&=\sum_{i=1}^n((y_i-\overline{y})-(\overline{y}-\mu_0))^2\\&=\sum_{i=1}^n\left((y_i-\overline{y})^2-2(y_i-\overline{y})(\overline{y}-\mu_0)+(\overline{y}-\mu_0)^2\right)\\&=\sum_{i=1}^n(y_i-\overline{y})^2+\sum_{i=1}^n(\overline{y}-\mu_0)^2\\&=\sum_{i=1}^n(y_i-\overline{y})^2+n(\overline{y}-\mu_0)^2\end{aligned}$$

where we used $\sum_{i=1}^n-2(y_i-\overline{y})(\overline{y}-\mu_0)=-2(\overline{y}-\mu_0)\sum_{i=1}^n(y_i-\overline{y})=0$.

$$\left(\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2}\right)^{n/2} < k$$

$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \mu_0)^2} < k^{2/n} = k'$$

$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2 + n(\overline{y} - \mu_0)^2} < k'$$

$$\frac{1}{1 + \frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2}} < k'$$

where the in the second line, we just take the $n/2^{th}$ root of both sides. The third line applies the identity from above and the final line divides the numerator and denominator of the fraction by $\sum_{i=1}^n (y_i - \overline{y})^2$. Now solving for $\frac{n(\overline{y}-\mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2}$ gives

$$\frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2} > 1/k' - 1 = k''$$

We notice this is very close to a t-distribution. So we multiple the denominator by $\frac{1}{n-1}$ to make it s^2, the empirical variance.

$$\frac{n(\overline{y} - \mu_0)^2}{\sum_{i=1}^n (y_i - \overline{y})^2} > (n-1)k''$$

then we take the square root of both sides:

$$\frac{\sqrt{n}(\overline{y} - \mu_0)}{\left(\frac{1}{n-1}\sum_{i=1}^n (y_i - \overline{y})^2\right)^{1/2}} < \sqrt{(n-1)k''}$$

The direction of the inequality changed because $\overline{y} < \mu_0$. The object on the left-side has a t-distribution. So we have derived the small-sample size test for the mean of normal random variables. This is the end of problem.

The general idea of Likelihood ratio test is to determine all possible parameters in the null and alternative hypothesis, we call this Ω. Then find the maximum of the the likelihood function over Ω. This involves setting the derivative (of $\ln(L)$) with respect to each parameter equal to 0 and then checking its value along the boundary of region.

Once you have found the maximum on all of Ω, you find the maximum on the smaller set Ω_0. This usually involves making some of the parameters constant, so the analysis already done can be reused.

The Likelihood ratio Test then gives RR region to be the ratio of these two maximums. Then you need to manipulate this ratio into a random variable you know.

This last step is rather tricky and often cannot be done. In the next notes we'll do one more example where we can get an explicit distribution. Then we'll see a Theorem that gives the distribution in the the large sample size limit it has a nice distribution. We've seen this phenomenon several times in this class, for a very small class of random sample distributions we can actually compute the the exact distribution of the statistic of interest. For the rest of the cases, we have some kind of CLT like theorem that holds as long as the sample size is large.