The Generalized Likelihood Ratio

Maths Note | Larry Cui

May 3, 2022

1 Traditional Hypothesis Testing

We've already known that in hypothesis testing, we have a null hypothesis, where $H_0 : \theta = \theta_0$ versus alternative hypothesis $H_A : \theta \neq \theta_0$. We also have a presumed pdf function for the variables. Based on these information, we can construct a critical point/region when whatever level of significance α is given. Then if a sample of size n comes in, we can use the sample mean to decide if we accept the null hypothesis or reject it based on whether it falls within the critical region or not.

Example 1 Let X be variable from a population of a normal distribution with unknown μ but known variance $\sigma^2 = 1$. If we decide to use $\mu_0 = 0$ as null hypothesis, we can construct a critical region of 0.05 significance (two-sided) as follows:

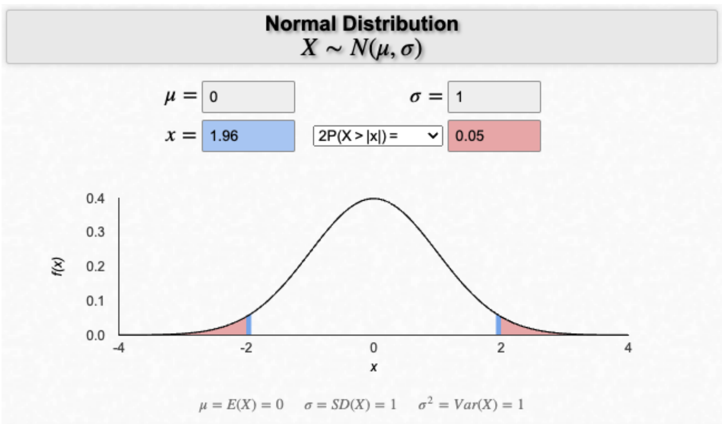

Figure 1: Level of significance at 0.05 (two-sided)

We use Z-transformation to convert the sample mean to z and decide:

$$Z = \frac{\overline{x} - 0}{1/\sqrt{n}}$$

Example 2 For binomial distribution, the parameter in question is p, the success probability. For a sample of size n, the sample mean or sum is also in normal distribution because of CLT. Let k be the total successes in n trials, the Z-transformation for binomial is:

$$Z = \frac{k - np}{\sqrt{np(1-p)}}$$

Of course, we can also construct it based on sample success mean,

$$Z = \frac{k/n - p}{\sqrt{p(1-p)/n}}$$

Comment For discrete distribution, we simply list out options and their corresponding probability, and construct the critical region accordingly.

2 Definition of GLR and GLRT

Notion We assume the pdf of variable y is $f(y;\theta)$ where θ represents ONE or MORE unknown parameters, then

1. Ω denotes the total possible parameter space of θ, that is all possible values of θ.
2. ω denotes possible parameter values admissible ONLY under H_0.
3. ω^C must denote all other values of Ω under H_A.

Let $y_1, y_2, \dots y_n$ be a sample of size n from distribution $f(y;\theta)$, and we pick $\theta = \theta_0$ for H_0. Recall likelihood function,

$$L(\theta_0) = f(y_1;\theta_0)f(y_2;\theta_0)\cdots f(y_n;\theta_0)$$

We know likelihood function is a pdf of parameter θ. But $L(\theta_0)$ may or may not at the peak of the pdf curve. On the other hand, if we put the maximum likelihood estimate θ_e into $L(\theta)$, we can get the maximum of likelihood function,

$$\max_{\theta_e \in \Omega} L(\theta_e) = f(y_1;\theta_e)f(y_2;\theta_e)\cdots f(y_n;\theta_e)$$

$L(\theta_e)$ must be the maximum of likelihood function because we find θ_e by differentiating $L(\theta)$ to get it. Now we can introduce the definition below.

Generalized Likelihood Ratio

$$\lambda = \frac{L(\theta_0)}{L(\theta_e)}\ \text{, where } 0 < \lambda \leqslant 1$$

Comment Given a specific sample, we have λ as a point value. But if we think of sample as n variables from a population of distribution, λ is a function of sample values. Furthermore, we know by intuition that the larger the λ is the better θ_0 matches sample data. Otherwise, θ_0 may not be a good estimate and should be rejected.

Now we can set a test rule about λ to accept or reject H_0.

Definition A generalized likelihood ratio test is one that rejects H_0 whenever

$$0 < \lambda \leqslant \lambda^*$$

where λ^* is chosen so that

$$P(0 < \Lambda \leqslant \lambda^* | H_0) = \alpha$$

By convention, people use Λ to represent λ as a variable (λ is now regarded as a function of sample values).

Apparently, from one end 0, H_0 doesn't match the sample at all, to the other end 1, matches perfectly, we are interested in finding a position λ^* so that the cumulative probability of Λ pdf equals α.

$$\alpha = \int_0^{\lambda^*} f_\Lambda(\lambda)\, d\lambda$$

As λ is a ratio, it might be difficult to find the pdf $f_\Lambda(\lambda)$ directly. But since λ is a function of variable y, we can construct the integral from y pdf.

3 Examples Revisiting

Example 1: revisiting

For a sample of size n from a normal distribution, the maximum likelihood estimate for $\mu_e = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$, so

$$L(\mu_0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{y_i^2}{2}\right]$$

and

$$L(\mu_e) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_i - \overline{y})^2}{2}\right]$$

then

$$\lambda = \frac{L(\mu_0)}{L(\mu_e)} = \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{n} y_i^2\right]}{\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - \overline{y})^2\right]}$$

A little trick on the numerator:

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n}(y_i - \overline{y} + \overline{y})^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 + \underbrace{\sum_{i=1}^{n} 2(y_i - \overline{y})\overline{y}}_{0} + \sum_{i=1}^{n} \overline{y}^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 + n\overline{y}^2$$

put it back to the above equation,

$$\lambda = \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - \overline{y})^2\right] \exp\left[-\frac{1}{2}n\overline{y}^2\right]}{\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(y_i - \overline{y})^2\right]} = \exp\left[-\frac{1}{2}n\overline{y}^2\right]$$

If we pick some number λ^* so that $\lambda \leqslant \lambda^*$, then

$$\begin{aligned}
\exp\left[-\frac{1}{2}n\overline{y}^2\right] &\leqslant \lambda^* \\
\overline{y}^2 &\geqslant \frac{-2\ln\lambda^*}{n} \\
\frac{|\overline{y}-0|}{1/\sqrt{n}} &\geqslant \frac{\sqrt{-(2/n)\ln\lambda^*}}{1/\sqrt{n}} && \triangleright \text{ divide both sides by } 1/\sqrt{n}
\end{aligned}$$

We can tell that the left side is Z-transformation of the sample mean of size n from a normal distribution. If $\alpha = 0.05$, we just need to calculate λ^* by

$$\frac{\sqrt{-(2/n)\ln\lambda^*}}{1/\sqrt{n}} = 1.96$$

Example 3 Suppose a random sample $X_1, X_2, \ldots, X_n$ is taken from a normal distribution population with unknown μ and σ^2. Find the size α likelihood ratio test for testing the null hypothesis $H_0\colon \mu = \mu_0$ against two-sided $H_A\colon \mu \neq \mu_0$.

We find $\max L(\theta_e)$ first. Use the maximum likelihood estimates, respectively,

$$\hat{\mu} = \overline{x} \qquad \text{and} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

then

$$\begin{aligned}
\max L(\theta_e) &= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\hat{\sigma}}\exp\left[-\frac{(x_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right] \\
&= \left[\frac{1}{2\pi(\frac{1}{n})\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}\exp\left[-\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{2(\frac{1}{n})\sum_{i=1}^{n}(x_i-\overline{x})^2}\right] \\
&= \left[\frac{1}{2\pi(\frac{1}{n})\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}\cdot e^{-\frac{n}{2}} \\
&= \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}}
\end{aligned}$$

Under null hypothesis, we have

$$\mu = \mu_0 \qquad \text{and} \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2$$

Comment How do we find σ^2 under null hypothesis? Well, we use maximum likelihood method again. Recall that in a sample of size n,

$$\ln L(\mu,\sigma^2) = -\frac{n}{2}\ln\Big(2\pi\sigma^2\Big) - \frac{1}{2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

let

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\left(\frac{1}{\sigma^2}\right)^2 \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2 \qquad \triangleright\ \mu=\mu_0$$

Okay, now we have null parameters to construct the numerator,

$$\begin{aligned} L(\theta_0) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_i-\mu_0)^2}{2\sigma^2}\right] \\ &= \left[\frac{1}{2\pi(\frac{1}{n})\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2(\frac{1}{n})\sum_{i=1}^{n}(x_i-\mu_0)^2}\right] \\ &= \left[\frac{1}{2\pi(\frac{1}{n})\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}} \cdot e^{-\frac{n}{2}} \\ &= \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}} \end{aligned}$$

Taking the ratio of the two likelihoods,

$$\lambda = \frac{L(\theta_0)}{L(\theta_e)} = \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}} \Bigg/ \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{\frac{n}{2}} = \left[\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}}$$

A small algebraic trick kicks in here:

$$\begin{aligned} \sum_{i=1}^{n}(x_i-\mu_0)^2 &= \sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\mu_0)^2 \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^2 + 2\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})(\overline{x}+\mu_0)}_{0} + \sum_{i=1}^{n}(\overline{x}-\mu_0)^2 \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^2 + \sum_{i=1}^{n}(\overline{x}-\mu_0)^2 \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu_0)^2 \end{aligned}$$

The ratio therefore can be further simplified as:

$$\begin{aligned} \lambda &= \left[\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu_0)^2}\right]^{\frac{n}{2}} \\ &= \left[\frac{1}{1+\dfrac{n(\overline{x}-\mu_0)^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\right]^{\frac{n}{2}} \end{aligned}$$

We then let $\lambda \leqslant \lambda^*$, and integrate pdf of λ to equate α, and find the value of λ^* from the equation,

$$\alpha = \int_0^{\lambda^*} f_\Lambda(\lambda)\, d\lambda$$

It looks intimidating to find the pdf for λ here. However, we can move about the fraction further,

$$\begin{aligned}
\left[\frac{1}{1+\frac{n(\overline{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2}}\right] &\leqslant (\lambda^*)^{2/n} \\
\frac{n(\overline{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2} &\geqslant (\lambda^*)^{-2/n} - 1 \\
\frac{n(\overline{x}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2} &\geqslant (n-1)((\lambda^*)^{-2/n} - 1) \\
\frac{(\overline{x}-\mu_0)^2}{S^2/n} &\geqslant (n-1)[(\lambda^*)^{-2/n} - 1]
\end{aligned}$$

Notice above $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$ is the sample variance, so we use S^2 to denote it. Now to make it tidier, we use c^2 to denote right side of the inequality, i.e., let $c^2 = (n-1)[(\lambda^*)^{-2/n} - 1]$.

$$\begin{aligned}
\frac{(\overline{x}-\mu_0)^2}{S^2/n} &\geqslant c^2 \\
\frac{|\overline{x}-\mu_0|}{S/\sqrt{n}} &\geqslant c
\end{aligned}$$

It turns out the left side of the inequality follows a T distribution with $n-1$ degrees of freedom. So we can use T distribution to choose $c = t_{\alpha/2,n-1}$. Tracing back forth, we get the threshold of λ^*.

4 GLRT is Hypothesis Testing

Example 4 Suppose $y_1, y_2, \ldots, y_n$ is a random sample from a uniform pdf over the interval $[0, \theta]$, where θ is unknown. Test $H_0\colon \theta = \theta_0$ versus $H_A\colon \theta < \theta_0$.

We know pdf of Y is: $1/\theta$, and cdf is:

$$\int_0^y \frac{1}{\theta}\, dy = \frac{y}{\theta}$$

As we discussed before, the best estimator for θ is $y_{\max}$, we now will use it to test null hypothesis.

4.1 traditional way

The pdf of $y_{\max}$ is the first derivative of cdf $y_{\max}$:

$$F(y_{\max}) = \left(\frac{y}{\theta_0}\right)^n \qquad \text{and} \qquad f(y_{\max}) = \frac{d}{dy}F(y_{\max}) = \frac{ny^{n-1}}{\theta_0^n}$$

$f(y_{\max})$ is a pdf on the scope of $[0, \theta_0]$. Yes, to test H_0 we first assume it is true.

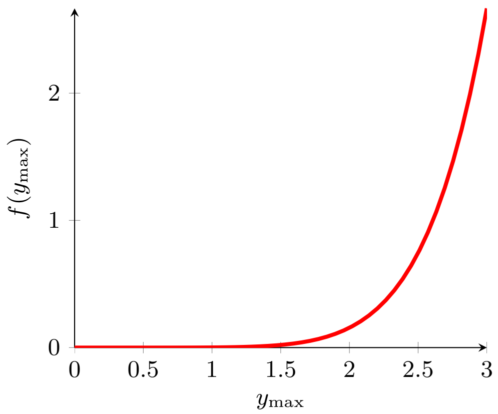

Figure 2: pdf plot of $f(y_{\max})$ when $H_0\colon \theta_0 = 3, n = 8$

By intuition, we see the larger $f(y_{\max})$ is, the more it supports the null hypothesis at θ_0. Apparently, it is a one-sided test. So on a scope from 0 to θ_0, we need to find a point p on x axis that the integral on $[0, p]$ is α, which in turn means the critical region to reject H_0.

$$\alpha = \int_0^p f(y_{\max})\, dy = \int_0^p \frac{ny^{n-1}}{\theta_0^n}\, dy = \left(\frac{p}{\theta_0}\right)^n$$

As a result, when $p \leqslant \theta_0 \sqrt[n]{\alpha}$, we reject H_0.

4.2 GLRT way

For a sample of size n, null likelihood is

$$L(\theta_0) = \left(\frac{1}{\theta_0}\right)^n$$

and maximum estimate likelihood is

$$L(\theta_e) = \left(\frac{1}{y_{\max}}\right)^n$$

then

$$\lambda = \frac{(1/\theta_0)^n}{(1/y_{\max})^n} = \left(\frac{y_{\max}}{\theta_0}\right)^n$$

We don't have directly way to find the pdf of λ, but by differentiating y-denoted λ, we get a pdf of λ wrt y.

$$f_\Lambda(\lambda) = \lambda' = \left[\left(\frac{y_{\max}}{\theta_0}\right)^n\right]' = \frac{ny^{n-1}}{\theta_0^n} = f(y_{\max})$$

A plot graph of $f_\Lambda(\lambda)$ below,

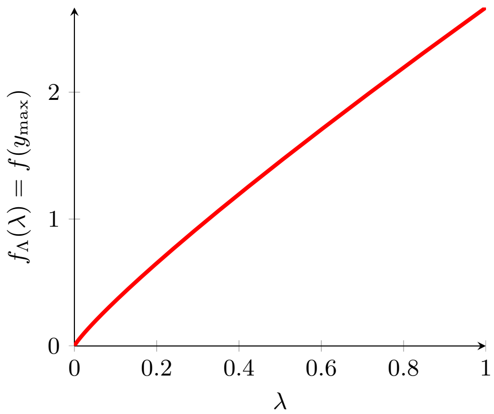

Figure 3: pdf plot of $f_\Lambda(\lambda)$ wrt λ when $H_0: \theta_0 = 3, n = 8$

We need to find a point λ^* on x axis so that

$$\alpha = \int_0^{\lambda^*} f_\Lambda(\lambda)\, d\lambda$$

Let $\lambda^* = \left(\frac{p}{\theta_0}\right)^n$, and expand the scope of integral from $[0, 1]$ to $[0, \theta_0]$, our task becomes finding p so that

$$\alpha = \int_0^p \frac{ny^{n-1}}{\theta_0^n}\, dy$$

As shown above, the critical point is $p = \theta_0 \sqrt[n]{\alpha}$. Substituting back to get our result:

$$\lambda^* = \alpha$$