**208711 Statistical Theory 1** **Fall 2019**

# Lecture 27: November 19

*Lecturer: Donlapark Pornnopparath*

## 27.1 Likelihood ratio test revisited

In previous lecture, we have defined the likelihood ratio test for a simple vs simple hypothesis testing. We now expand the definition to a more general case.

**Definition 27.1.** Consider the hypothesis test:

$$\begin{aligned} H_0 &: \theta \in \Omega_0 \\ H_1 &: \theta \in \Omega_1, \end{aligned}$$

and $\Omega = \Omega_0 \cup \Omega_1$. The *likelihood ratio statistic* from a random sample $X$ is defined by

$$L(X) = \frac{\sup_{\theta\in\Omega_0} \mathcal{L}(\theta|X)}{\sup_{\theta\in\Omega} \mathcal{L}(\theta|X)}.$$

The *likelihood ratio test* rejects $H_0$ if $L(x) \leq c$.

In general, it is not easy to find the supremum of the likelihood function on a restricted set of parameters. That's why we had to resort to monotonicity of $L(X)$ for the one-sided test. However, if the null hypothesis only has a single point, this is quite easy to compute.

**Example 27.2.** Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$. Consider the two-sided hypothesis test with an unknown $\sigma^2$:

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta \neq \theta_0. \end{aligned}$$

Thus the likelihood ratio is

$$L(X) = \frac{\sup_{\sigma_0^2} \mathcal{L}(\theta_0|X)}{\sup_{\theta\neq\theta_0,\sigma^2} \mathcal{L}(\theta|X)} = \frac{\sup_{\sigma_0^2} (2\pi\sigma_0^2)^{n/2} \exp\left(-\sum_i (x_i-\theta_0)^2/(2\sigma_0^2)\right)}{\sup_{\theta\neq\theta_0,\sigma^2} (2\pi\sigma^2)^{n/2} \exp\left(-\sum_i (x_i-\theta)^2/(2\sigma^2)\right)}$$

Let $\hat{\sigma}_0^2, \hat{\theta}$ and $\hat{\sigma}^2$ be the maximizers of the ratio. It can be showed from the standard calculus that

$$\begin{aligned} \hat{\sigma}_0^2 &= \frac{1}{n}\sum_i (x_i-\theta_0)^2 \\ \hat{\theta} &= \bar{x} \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_i (x_i-\bar{x})^2. \end{aligned}$$

Therefore, the ratio is

$$\begin{aligned} L(x) &= \frac{[(2\pi/n)\sum_i (x_i - \theta_0)^2]^{-n/2} e^{-n/2}}{[(2\pi/n)\sum_i (x_i - \bar{x})^2]^{-n/2} e^{-n/2}} \\ &= \left[\frac{\sum_i (x_i - \theta_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{-n/2} \\ &= \left[\frac{\sum_i (x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{-n/2} \\ &= \left[1 + \frac{n(\bar{x} - \theta_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{-n/2}. \end{aligned}$$

The test rejects $H_0$ if $L(x) \le c$ for a specified $c$. This corresponds to

$$|T_n| \ge k$$

for some $k > 0$, where $T_n = (\bar{x} - \theta_0)/(S/\sqrt{n}) \sim t_{n-1}$. Thus, the likelihood ratio test corresponds to the t-test of a normal random sample. ◇

## 27.2 Wald test

An alternative method to test the two-sided hypothesis is by approximating by the standard normal distribution. Suppose that $\hat{\theta}$ is an MLE estimator of $\theta_0$. Then we have that

$$\sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \to N(0, 1)$$

in distribution as $n \to \infty$. Thus, for a sufficiently large value of $n$, we can use $T_n = \sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta_0)$ as our test statistic with standard normality assumption. Hence, we reject the null hypothesis if $|T_n| > z_{\alpha/2}$. The level of this test is not exactly, but *asymptotically* $\alpha$:

$$\mathbb{P}_{\theta_0}(|T_n| > z_{\alpha/2}) \to \alpha \qquad \text{as } n \to \infty.$$

**Example 27.3.** Take a Bernoulli trial $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$ and record the MLE estimator $\hat{p} = \overline{X}$ in order to test $H_0 : p = p_0$ versus $H_1 : p \ne p_0$. We have already computed that $I(\hat{p}) = 1/[\hat{p}(1 - \hat{p})]$, so the Wald statistic is

$$T_n = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}},$$

and we would reject $H_0$ if $|T_n| \ge z_{\alpha/2}$. ◇

**Example 27.4.** Suppose we want to compare performance of two classification models by their predictions on different test sets: Model 1 on a test set of size $m$ and Model 2 on a test set of size $n$. Let $X$ be the number of samples correctly classified by the first model and $Y$ for the second model. Then $X \sim (m, p_1)$ and $Y \sim (n, p_2)$ where $p_1$ and $p_2$ are true rate of classification error. We want to make a hypothesis test to see if both models have comparable performance. If we let $\delta = p_2 - p_1$ then our hypothesis test is

$$\begin{aligned} H_0 &: \delta = 0 \\ H_1 &: \delta \ne 0 \end{aligned}$$

In this case, the MLE for both models are the sample proportions $\hat{p}_1$ and $\hat{p}_2$. The sample variance can be estimated as

$$\text{Var}(\hat{\delta}) = \frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n},$$

and we would reject the null hypothesis if

$$\frac{|\hat{\delta} - 0|}{\sqrt{\text{Var}(\hat{\delta})}} \geq z_{\alpha/2}.$$

◇

## 27.3 Bayesian hypothesis tests

In Bayesian approach, the posterior density depends on the value of an observed variable $x$, allowing us to have nontrivial values of $\mathbb{P}(H_0 \text{ is true } |x)$ and $\mathbb{P}(H_0 \text{ is false } |x)$. From this, there are a couple of ways to make a decision on whether to accept or reject $H_0$. :

- Accept $H_0$ if $\mathbb{P}(\theta \in \Omega_0|x) \geq \mathbb{P}(\theta \in \Omega_1|x)$ and reject it otherwise.
- Accept $H_0$ if $\mathbb{P}(\theta \in \Omega_0|x)$ exceeds some threshold $c$. For example, to prevent false rejections, one might set $c = 0.99$.

**Example 27.5.** Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ be iid with the prior distribution $N(\theta, \tau^2)$. Suppose that we want to test the hypotheses $H_0 : \mu \leq \mu_0$ versus $\mu > \mu_0$ and we decide to accept $H_0$ if $\mathbb{P}(\mu \leq \mu_0|x) \geq \mathbb{P}(\mu > \mu_0|x)$. In other words,

$$\mathbb{P}(\mu \leq \mu_0) \geq \frac{1}{2}.$$

By the symmetry of the normal distribution, this holds true only if $\mu_0 \geq \mathbb{E}[\mu|x]$. We recall that

$$\mathbb{E}[\mu|x] = \frac{n\tau^2\bar{x} + \sigma^2\theta}{n\tau^2 + \sigma^2},$$

so we accept $H_0$ if

$$\bar{x} \leq \frac{\mu_0(n\tau^2 + \sigma^2) - \sigma^2\theta}{n\tau^2} = \mu_0 + \frac{\sigma^2(\mu_0 - \theta)}{n\tau^2}.$$

In particular, if the parameter we want to test $\mu_0$ is the same as the prior mean $\theta$, then we would accept $H_0$ if $\bar{x} \leq \theta$ and reject it otherwise. ◇