Lecture 22: Asymptotic Evaluations of Hypothesis Tests

Mathematical Statistics II, MATH 60062/70062

Thursday April 21, 2022

Reference: Casella & Berger, 10.3

Large sample hypothesis tests

Based on finite sample criteria, optimal tests (e.g., UMP tests) are available for just a small collection of problems (some of which are not realistic).

We will discuss three large sample approaches to formulate hypothesis tests:

- **Wald tests**
- **Score tests** (also known as Lagrange multiplier tests)
- **Likelihood ratio tests**

These are known as the "large sample likelihood based tests."

Basis for the Wald test

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Under certain regularity conditions, an MLE $\hat{\theta}$ satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, v(\theta)),$$

where

$$v(\theta) = \frac{1}{I_1(\theta)}.$$

If $v(\theta)$ is a continuous function of θ, then

$$v(\hat{\theta}) \xrightarrow{p} v(\theta),$$

for all θ; i.e., $v(\hat{\theta})$ is a consistent estimator of $v(\theta)$. Therefore

$$Z_n^* = \frac{\hat{\theta} - \theta}{\sqrt{\frac{v(\hat{\theta})}{n}}} = \frac{\hat{\theta} - \theta}{\sqrt{\frac{v(\theta)}{n}}} \sqrt{\frac{v(\theta)}{v(\hat{\theta})}} \xrightarrow{d} \mathcal{N}(0, 1),$$

by Slutsky's Theorem.

Wald statistic

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Under H_0,

$$Z_n^W = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{v(\hat{\theta})}{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore,

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^W| \geq z_{\alpha/2}\},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the $\mathcal{N}(0, 1)$ distribution, is an approximate size α rejection region for testing H_0 versus H_1.

Wald statistic

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Under H_0,

$$Z_n^W = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{v(\hat{\theta})}{n}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore,

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^W| \geq z_{\alpha/2}\},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the $\mathcal{N}(0, 1)$ distribution, is an approximate size α rejection region for testing H_0 versus H_1.

One sided tests also use Z_n^W, with a modified form of R.

Bernoulli Wald test

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(p)$, where $0 < p < 1$. Consider testing

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p \neq p_0.$$

The MLE of p is the sample proportion

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Because $\hat{p}$ is an MLE, we know that

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} \mathcal{N}(0, v(p)),$$

where

$$v(p) = \frac{1}{I_1(p)}.$$

The PMF of X is

$$f_X(x \mid p) = p^x(1-p)^{1-x}.$$

Therefore,

$$\log f_X(x \mid p) = x \log p + (1-x)\log(1-p).$$

The derivatives of $\log f_X(x \mid p)$ are

$$\begin{aligned}
\frac{\partial}{\partial p} \log f_X(x \mid p) &= \frac{x}{p} - \frac{1-x}{1-p} \\
\frac{\partial^2}{\partial p^2} \log f_X(x \mid p) &= -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.
\end{aligned}$$

Therefore,

$$I_1(p) = -E_p\left[\frac{\partial^2}{\partial p^2} \log f_X(x \mid p)\right] = E_p\left[\frac{X}{p^2} + \frac{1-X}{(1-p)^2}\right] = \frac{1}{p(1-p)}$$

E_p(X) = p, E_p(1 - X) = 1 - p

and

$$v(p) = \frac{1}{I_1(p)} = p(1-p).$$

Therefore, we have

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} \mathcal{N}(0, p(1-p)).$$

Because the asymptotic variance $v(p) = p(1-p)$ is a continuous function of p, it can be consistently estimated by $v(\hat{p}) = \hat{p}(1-\hat{p})$. The Wald statistic to test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$ is given by

$$Z_n^W = \frac{\hat{p} - p_0}{\sqrt{\frac{v(\hat{p})}{n}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}.$$

An approximate size α rejection region is

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^W| \geq z_{\alpha/2}\}.$$

Basis for the score test

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$.

The score function, when viewed as random, is

$$S(\theta \mid \boldsymbol{X}) = \frac{\partial}{\partial\theta} \log L(\theta \mid \boldsymbol{X}) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_X(X_i \mid \theta),$$

a sum of iid random variables. Recall that

$$E_\theta\left[\frac{\partial}{\partial\theta} \log f_X(X \mid \theta)\right] = 0$$

$$\mathrm{Var}_\theta\left[\frac{\partial}{\partial\theta} \log f_X(X \mid \theta)\right] = E_\theta\left\{\left[\frac{\partial}{\partial\theta} \log f_X(X \mid \theta)\right]^2\right\} = I_1(\theta).$$

Lec 8, page 17

Therefore, applying the CLT to the sum above, we have

$$\sqrt{n}\left(\frac{1}{n}S(\theta \mid \boldsymbol{X}) - 0\right) \xrightarrow{d} \mathcal{N}(0, I_1(\theta)),$$

which means

$$\frac{\frac{1}{n}S(\theta \mid \boldsymbol{X})}{\sqrt{\frac{I_1(\theta)}{n}}} = \frac{S(\theta \mid \boldsymbol{X})}{\sqrt{nI_1(\theta)}} = \frac{S(\theta \mid \boldsymbol{X})}{\sqrt{I_n(\theta)}} \xrightarrow{d} \mathcal{N}(0, 1),$$

where $I_n(\theta) = nI_1(\theta)$ is the Fisher Information based on all n iid observations.

Therefore, the score function divided by the square root of the Fisher information (based on all n observations) behaves asymptotically like a $\mathcal{N}(0, 1)$ random variable.

Score statistic

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Under H_0,

$$Z_n^S = \frac{S(\theta_0 \mid \boldsymbol{X})}{\sqrt{I_n(\theta_0)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

The denominator is I_n, different from 1/I_n in the Wald method

Therefore,

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^S| \geq z_{\alpha/2}\},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the $\mathcal{N}(0, 1)$ distribution, is an approximate size α rejection region for testing H_0 versus H_1.

Score statistic

Suppose $X_1, \ldots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

Under H_0,

$$Z_n^S = \frac{S(\theta_0 \mid \boldsymbol{X})}{\sqrt{I_n(\theta_0)}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Therefore,

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^S| \geq z_{\alpha/2}\},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the $\mathcal{N}(0, 1)$ distribution, is an approximate size α rejection region for testing H_0 versus H_1.

One sided tests also use Z_n^S, with a modified form of R.

Bernoulli score test

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(p)$, where $0 < p < 1$. Consider testing

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p \neq p_0.$$

The likelihood function is given by

$$L(p \mid \boldsymbol{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}.$$

The log-likelihood function is

$$\log L(p \mid \boldsymbol{x}) = \sum_{i=1}^{n} x_i \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p).$$

The score function is

$$S(p \mid \boldsymbol{x}) = \frac{\partial}{\partial p} \log L(p \mid \boldsymbol{x}) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}.$$

We already calculated

$$I_1(p) = \frac{1}{p(1-p)}.$$

Therefore, the score statistic is

$$Z_n^S = \frac{S(p_0 \mid \boldsymbol{X})}{\sqrt{I_n(p_0)}} = \frac{\frac{\sum_{i=1}^n X_i}{p_0} - \frac{n-\sum_{i=1}^n X_i}{1-p_0}}{\sqrt{\frac{n}{p_0(1-p_0)}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

An approximate size α rejection region is

$$R = \{\boldsymbol{x} \in \mathcal{X} : |z_n^S| \geq z_{\alpha/2}\}.$$

Bernoulli Wald and score statistics

The Wald statistic

$$Z_n^W = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

and the score statistic

$$Z_n^S = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

differ only in how the standard error of $\hat{p}$ (as a point estimator of p) is calculated.

- The Wald statistic uses the estimated standard error.
- The score statistic uses the standard error under the assumption that $H_0 : p = p_0$ is true (noting is being estimated).

This is an argument in favor of the score statistic.