Biostatistics 602 - Statistical Inference
Lecture 22
p-Values

Hyun Min Kang

April 9th, 2013

Last Lecture

- Is the exact distribution of LRT statistic typically easy to obtain?
- How about its asymptotic distribution? For testing which null/alternative hypotheses is the asymptotic distribution valid?
- What is a Wald Test?
- Describe a typical way to construct a Wald Test.

Asymptotics of LRT

Theorem 10.3.1

Consider testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Suppose $X_1, \cdots, X_n$ are iid samples from $f(x|\theta)$, and $\hat{\theta}$ is the MLE of θ, and $f(x|\theta)$ satisfies certain "regularity conditions" (e.g. see misc 10.6.2), then under H_0:

$$-2 \log \lambda(\mathbf{x}) \xrightarrow{\text{d}} \chi^2_1$$

as $n \to \infty$.

Wald Test

Wald test relates point estimator of θ to hypothesis testing about θ.

Definition

Suppose W_n is an estimator of θ and $W_n \sim \mathcal{AN}(\theta, \sigma^2_W)$. Then Wald test statistic is defined as

$$Z_n = \frac{W_n - \theta_0}{S_n}$$

where θ_0 is the value of θ under H_0 and S_n is a consistent estimator of σ_W

p-Values

Conclusions from Hypothesis Testing

- Reject H_0 or accept H_0.
- If size of the test is (α) small, the decision to reject H_0 is convincing.
- If α is large, the decision may not be very convincing.

Definition: p-Value

A *p-value* $p(\mathbf{X})$ is a test statistic satisfying $0 \leq p(\mathbf{x}) \leq 1$ for every sample point $\mathbf{x}$. Small values of $p(\mathbf{X})$ given evidence that H_1 is true. A *p-value* is *valid* if, for every $\theta \in \Omega_0$ and every $0 \leq \alpha \leq 1$,

$$\Pr(p(\mathbf{X}) \leq \alpha|\theta) \leq \alpha$$

Advantage to reporting a test result via a p-value

- The size α does not need to be predefined
  - Each reader can choose the α he or she considers appropriate
  - And then can compare the reported $p(\mathbf{x})$ to α
  - So that each reader can individually determine whether these data lead to acceptance or rejection to H_0.
- The p-value quantifies the evidence against H_0.
  - The smaller the p-value, the stronger, the evidence for rejecting H_0.
  - A p-value reports the results of a test on a more continuous scale
  - Rather than just the dichotomous decision "Accept H_0" or "Reject H_0".

Constructing a valid p-value

Theorem 8.3.27.

Let $W(\mathbf{X})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\mathbf{x}$, define

$$p(\mathbf{x}) = \sup_{\theta \in \Omega_0} \Pr(W(\mathbf{X}) \geq W(\mathbf{x})|\theta)$$

Then $p(\mathbf{X})$ is a valid p-value.

Example : Two-sided normal p-value

Problem

Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a random sample from a $\mathcal{N}(\theta, \sigma^2)$ population. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

1. Construct a size α LRT test.
2. Find a valid p-value, as a function of $\mathbf{x}$.

Solution - Constructing LRT

$$\begin{aligned}\Omega &= \{(\theta,\sigma^2) : \theta \in \mathbb{R}, \sigma^2 > 0\} \\ \Omega_0 &= \{(\theta,\sigma^2) : \theta = \theta_0, \sigma^2 > 0\} \\ \lambda(\mathbf{x}) &= \frac{\sup_{\{(\theta,\sigma^2):\theta=\theta_0,\sigma^2>0\}} L(\theta,\sigma^2|\mathbf{x})}{\sup_{\{(\theta,\sigma^2):\theta\in\mathbb{R},\sigma^2>0\}} L(\theta,\sigma^2|\mathbf{x})}\end{aligned}$$

For the denominator, the MLE of θ and σ^2 are

$$\begin{cases} \hat{\theta} = \overline{X} \\ \hat{\sigma}^2 = \frac{\sum(X_i-\overline{X})^2}{n} = \frac{n-1}{n}s^2_{\mathbf{X}} \end{cases}$$

For the numerator, the MLE of θ and σ^2 are

$$\begin{cases} \hat{\theta}_0 = \theta_0 \\ \hat{\sigma}^2_0 = \frac{\sum(X_i-\theta_0)^2}{n} \end{cases}$$

Solution - Constructing and Simplifying the Test

Combining the results together

$$\lambda(\mathbf{x}) = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2_0}\right)^{n/2}$$

LRT test rejects H_0 if and only if

$$\begin{aligned}\left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2_0}\right)^{n/2} &\leq c \\ \left(\frac{\sum(x_i-\overline{x})^2/n}{\sum(x_i-\theta_0)^2/n}\right)^{n/2} &\leq c \\ \frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\theta_0)^2} &\leq c^*\end{aligned}$$

Solution - Simplifying the LRT

$$\begin{aligned}\frac{\sum(x_i-\overline{x})^2}{\sum(x_i-\overline{s})^2 + n(\overline{x}-\theta_0)^2} &\leq c^* \\ \frac{1}{1+\frac{n(\overline{x}-\theta_0)^2}{\sum(x_i-\overline{x})^2}} &\leq c^* \\ \frac{n(\overline{x}-\theta_0)^2}{\sum(x_i-\overline{x})^2} &\geq c^{**} \\ \left|\frac{\overline{x}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}}\right| &\geq c^{***}\end{aligned}$$

LRT test rejects H_0 if $\left|\frac{\overline{x}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}}\right| \geq c^{***}$. The next step is specify c to get size α test.

Solution - Obtaining size α test

Under H_0

$$\begin{aligned}\frac{\overline{X}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}} &\sim T_{n-1} \\ \Pr\left(\left|\frac{\overline{X}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}}\right| \geq c^{***}\right) &= \alpha \\ \Pr\left(|T_{n-1}| \geq c^{***}\right) &= \alpha \\ c^{***} &= t_{n-1,\alpha/2}\end{aligned}$$

Therefore, size α LRT test rejects H_0 if and only if $\left|\frac{\overline{x}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}}\right| \geq t_{n-1,\alpha/2}$

Solution - p-value from two-sided test

For a test statistic $W(\mathbf{X}) = \left|\frac{\overline{X}-\theta_0}{s_{\mathbf{X}}/\sqrt{n}}\right|$, under H_0, regardless of the value of σ^2, $W(\mathbf{X}) \sim T_{n-1}$. Then, a valid p-value can be defined by

$$\begin{aligned} p(\mathbf{x}) &= \sup_{\theta \in \Omega_0} \Pr(W(\mathbf{X}) \geq W(\mathbf{x}) | \theta, \sigma^2) \\ &= \Pr(W(\mathbf{X}) \geq W(\mathbf{x}) | \theta_0, \sigma^2) \\ &= 2 \Pr(T_{n-1} \geq W(\mathbf{x})) \\ &= 2 \left[1 - F_{T_{n-1}}^{-1}\{W(\mathbf{x})\}\right] \end{aligned}$$

where $F_{T_{n-1}}^{-1}(\cdot)$ is the inverse CDF of t-distribution with $n-1$ degrees of freedom.

Example : One-sided normal p-value

Problem

Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a random sample from a $\mathcal{N}(\theta, \sigma^2)$ population. Consider testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

1. Construct a size α LRT test.
2. Find a valid p-value, as a function of $\mathbf{x}$.

Constructing LRT test

As shown in previous lectures, the LRT size α test rejects H_0 if

$$W(\mathbf{x}) = \frac{\overline{x} - \theta_0}{s_{\mathbf{X}}/\sqrt{n}} \geq t_{n-1,\alpha}$$

Because the null hypothesis contains multiple possible $\theta \leq \theta_0$, we first want to show that the supreme in the definition of p-value

$$p(\mathbf{x}) = \sup_{\theta \in \Omega_0} \Pr(W(\mathbf{X}) \geq W(\mathbf{x}) | \theta, \sigma^2)$$

always occurs at when $\theta = \theta_0$, and the value of σ does not matter.

Obtaining one-sided p-value

Consider any $\theta \leq \theta_0$ and any σ.

$$\begin{aligned} \Pr(W(\mathbf{X}) \geq W(\mathbf{x}) | \theta, \sigma^2) &= \Pr\left(\frac{\overline{X} - \theta_0}{s_{\mathbf{X}}/\sqrt{n}} \geq W(\mathbf{x}) | \theta, \sigma^2\right) \\ &= \Pr\left(\frac{\overline{X} - \theta}{s_{\mathbf{X}}/\sqrt{n}} \geq W(\mathbf{x}) + \frac{\theta_0 - \theta}{s_{\mathbf{X}}/\sqrt{n}} | \theta, \sigma^2\right) \\ &= \Pr\left(T_{n-1} \geq W(\mathbf{x}) + \frac{\theta_0 - \theta}{s_{\mathbf{X}}/\sqrt{n}} | \theta, \sigma^2\right) \\ &\leq \Pr\left(T_{n-1} \geq W(\mathbf{x})\right) \\ &= \Pr\left(W(\mathbf{X}) \geq W(\mathbf{x}) | \theta_0, \sigma^2\right) \end{aligned}$$

Obtaining one-sided p-value (cont'd)

Thus, the p-value for this one-side test is

$$\begin{aligned} p(\mathbf{x}) &= \sup_{\theta \in \Omega_0} \Pr(W(\mathbf{X}) \geq W(\mathbf{x})|\theta, \sigma^2) \\ &= \Pr(W(\mathbf{X}) \geq W(\mathbf{x})|\theta_0, \sigma^2) \\ &= \Pr(T_{n-1} \geq W(\mathbf{x})) = 1 - F^{-1}_{T_{n-1}}[W(\mathbf{x})] \end{aligned}$$

p-Values by conditioning on on sufficient statistic

Suppose $S(\mathbf{X})$ is a sufficient statistic for the model $\{f(\mathbf{x}|\theta) : \theta \in \Omega_0\}$. (not necessarily including alternative hypothesis). If the null hypothesis is true, the conditional distribution of $\mathbf{X}$ given $S = s$ does not depend on θ. Again, let $W(\mathbf{X})$ denote a test statistic where large value give evidence that H_1 is true. Define

$$p(\mathbf{x}) = \Pr(W(\mathbf{X}) \geq W(\mathbf{x})|S = S(\mathbf{x}))$$

If we consider only the conditional distribution, by Theorem 8.3.27, this is a valid p-value, meaning that

$$\Pr(p(\mathbf{X}) \leq \alpha | S = s) \leq \alpha$$

p-Values by conditioning on sufficient statistic (cont'd)

Then for any $\theta \in \Omega_0$, unconditionally we have

$$\begin{aligned} \Pr(p(\mathbf{X}) \leq \alpha|\theta) &= \sum_s \Pr(p(\mathbf{X}) \leq \alpha | S = s) \Pr(S = s|\theta) \\ &\leq \sum_s \alpha \Pr(S = s|\theta) = \alpha \end{aligned}$$

Thus, $p(\mathbf{X})$ is a valid p-value.

Example - Fisher's Exact Test

Problem

Let X_1 and X_2 be independent observations with $X_1 \sim \text{Binomial}(n_1, p_1)$, and $X_2 \sim \text{Binomial}(n_2, p_2)$. Consider testing $H_0 : p_1 = p_2$ versus $H_1 : p_1 > p_2$. Find a valid p-value function.

Solution

Under H_0, if we let p denote the common value of $p_1 = p_2$. Then the join pmf of (X_1, X_2) is

$$\begin{aligned} f(x_1, x_2|p) &= \binom{n_1}{x_1} p^{x_1}(1-p)^{n_1 - x_1} \binom{n_2}{x_2} p^{x_2}(1-p)^{n_2-x_2} \\ &= \binom{n_1}{x_1}\binom{n_2}{x_2} p^{x_1+x_2}(1-p)^{n_1+n_2-x_1-x_2} \end{aligned}$$

Therefore $S = X_1 + X_2$ is a sufficient statistic under H_0.

Solution - Fisher's Exact Test (cont'd)

Given the value of $S = s$, it is reasonable to use X_1 as a test statistic and reject H_0 in favor of H_1 for large values of X_1,because large values of X_1 correspond to small values of $X_2 = s - X_1$. The conditional distribution of X_1 given $S = s$ is a hypergeometric distribution.

$$f(X_1 = x_1|s) = \frac{\binom{n_1}{x_1}\binom{n_2}{s-x_1}}{\binom{n_1+n_2}{s}}$$

Thus, the p-value conditional on the sufficient statistic $s = x_1 + x_2$ is

$$p(x_1, x_2) = \sum_{j=x_1}^{\min(n_1,s)} f(j|s)$$

Exercise 8.1

Problem

In 1,000 tosses of a coin, 560 heads and 440 tails appear. Is it reasonable to assume that the coin is fair? Justify your answer.

Hypothesis

Let $\theta \in (0, 1)$ be the probability of head.

1. $H_0 : \theta = 1/2$
2. $H_1 : \theta \neq 1/2$

Two possible strategies

Performing size α Hypothesis Testing

1. Define a level α test for a reasonably small α.
2. Test whether the observation rejects H_0 or not.
3. Conclude that H_0 is true or false at level α

Obtaining p-value

1. Obtain a p-value function $p(\mathbf{X})$.
2. Compute p-value as a quantitative support for the null hypothesis.

Asymptotic size α test

1,000 tosses are large enough to approximate using CLT.

$$\overline{X} \sim \mathcal{AN}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

A two-sided Wald test statistic can be defined by

$$Z(\mathbf{X}) = \frac{\overline{X} - \theta_0}{\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}}$$

At level α, the H_0 is rejected if and only if

$$|Z(\mathbf{x})| > z_{\alpha/2}$$

$$\frac{0.56 - 0.5}{\sqrt{\frac{0.56\times 0.44}{1000}}} = 3.822 > z_{\alpha/2}$$

Hypothesis Testing

Since $z_{\alpha/2}$ is 1.96, 2.57, and 4.42 for $\alpha = 0.05, 0.01$, and 10^{-5}, respectively, we can conclude that the coin is biased at level 0.05 and 0.01. However, at the level of 10^{-5}, the coin can be assumed to be fair.

Using p-value function

If the normal approximation is used, the p-value can be obtained as

$$\begin{aligned} \Pr(|Z(\mathbf{X})| \geq |Z(\mathbf{x})|) &= \Pr(|Z(\mathbf{X})| \geq 3.795) \\ &= 1.32 \times 10^{-4} \end{aligned}$$

So, under the null hypothesis, the size of test is less than 1.32×10^{-4}, suggesting a strong evidence for rejecting H_0.

Exercise 8.2

Problem

In a given city, it is assume that the number of automobile accidents in a given year follows a Poisson distribution. In past years, the average number of accidents per year was 15, and this year it was 10. Is it justified to claim that the accident rate has dropped?

Solution - Hypothesis

$X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$.

1. $H_0 : \lambda_1 = \lambda_2$.
2. $H_1 : \lambda_1 \neq \lambda_2$.

Constructing a test based on sufficient statistic

Under H_0, let $\lambda_1 = \lambda_2 = \lambda$.

$$\begin{aligned} f_{\mathbf{X}}(x_1, x_2|\lambda) &= \Pr(X = x_1|\lambda)\Pr(X = x_2|\lambda) \\ &= \frac{e^{-2\lambda}\lambda^{x_1+x_2}}{x_1! x_2!} \end{aligned}$$

Let $S = X_1 + X_2$. S is sufficient statistic for λ under H_0. $S \sim \text{Poisson}(2\lambda)$.

$$\begin{aligned} f_S(s|\lambda) &= \Pr(S = s|2\lambda) \\ &= \frac{e^{-2\lambda}\lambda^s}{s!} \end{aligned}$$

Constructing a test based on sufficient statistic (cont'd)

The conditional distribution of $\mathbf{x}$ given s is

$$\begin{aligned} f(x_1, x_2 | s) &= \frac{f_{\mathbf{X}}(x_1, x_2|\lambda)}{f_S(s|\lambda)} \\ &= \frac{\frac{e^{-2\lambda}\lambda^{x_1+x_2}}{x_1!x_2!}}{\frac{e^{-2\lambda}(2\lambda)^s}{s!}} \\ &= \frac{s!}{2^s x_1! x_2!} = \frac{\binom{s}{x_1}}{2^s} \end{aligned}$$

Constructing a test based on sufficient statistic (cont'd)

Let $W(\mathbf{X}) = X_1$, then the p-value conditioned on sufficient statistic is

$$\begin{aligned} p(\mathbf{x}) &= \Pr(W(\mathbf{X}) \geq W(\mathbf{x}) | S = S(\mathbf{x})) \\ &= \Pr(X_1 \geq x_1 | S = s) \\ &= \sum_{j=x_1}^{s} \frac{\binom{s}{x_1}}{2^s} = \sum_{j=x_1}^{x_1+x_2} \frac{\binom{x_1+x_2}{x_1}}{2^{x_1+x_2}} \approx 0.21 \end{aligned}$$

where $x_1 = 15$, $x_2 = 10$. Therefore, H_0 is not rejected when $\alpha < .05$, and it is not reasonable to claim that the accident rate has dropped.

Summary

Today

- p-Value
- Fisher's Exact Test
- Examples of Hypothesis Testing

Next Lectures

- Interval Estimation
- Confidence Interval