

Biostatistics 602 - Statistical Inference

Lecture 21

Asymptotics of LRT

Wald Test

Hyun Min Kang

April 2nd, 2013

Uniformly Most Powerful Test (UMP)

Definition

Let \mathcal{C} be a class of tests between $H_0 : \theta \in \Omega$ vs $H_1 : \theta \in \Omega_0^c$. A test in \mathcal{C} , with power function $\beta(\theta)$ is *uniformly most powerful (UMP) test* in class \mathcal{C} if $\beta(\theta) \geq \beta'(\theta)$ for every $\theta \in \Omega_0^c$ and every $\beta'(\theta)$, which is a power function of another test in \mathcal{C} .

UMP level α test

Consider \mathcal{C} be the set of all the level α test. The UMP test in this class is called a UMP level α test.

UMP level α test has the smallest type II error probability for any $\theta \in \Omega_0^c$ in this class.

Last Lecture

- What is a Uniformly Most Powerful (UMP) Test?
- Does UMP level α test always exist for simple hypothesis testing?
- For composite hypothesis, which property makes it possible to construct a UMP level α test?
- What is a sufficient condition for an exponential family to have MLR property?
- For one-sided composite hypothesis testing, if a sufficient statistic satisfies MLR property, how can a UMP level α test be constructed?

Neyman-Pearson Lemma

Theorem 8.3.12 - Neyman-Pearson Lemma

Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ where the pdf or pmf corresponding the θ_i is $f(\mathbf{x}|\theta_i)$, $i = 0, 1$, using a test with rejection region R that satisfies

$$\mathbf{x} \in R \quad \text{if } f(\mathbf{x}|\theta_1) > kf(\mathbf{x}|\theta_0) \quad (8.3.1) \text{ and}$$

$$\mathbf{x} \in R^c \quad \text{if } f(\mathbf{x}|\theta_1) < kf(\mathbf{x}|\theta_0) \quad (8.3.2)$$

For some $k \geq 0$ and $\alpha = \Pr(\mathbf{X} \in R|\theta_0)$, Then,

- (Sufficiency) Any test that satisfies 8.3.1 and 8.3.2 is a UMP level α test
- (Necessity) if there exist a test satisfying 8.3.1 and 8.3.2 with $k > 0$, then every UMP level α test is a size α test (satisfies 8.3.2), and every UMP level α test satisfies 8.3.1 except perhaps on a set A satisfying $\Pr(\mathbf{X} \in A|\theta_0) = \Pr(\mathbf{X} \in A|\theta_1) = 0$.

Monotone Likelihood Ratio

Definition

A family of pdfs or pmfs $\{g(t|\theta) : \theta \in \Omega\}$ for a univariate random variable T with real-valued parameter θ have a monotone likelihood ratio if $\frac{g(t|\theta_2)}{g(t|\theta_1)}$ is an increasing (or non-decreasing) function of t for every $\theta_2 > \theta_1$ on $\{t : g(t|\theta_1) > 0 \text{ or } g(t|\theta_2) > 0\}$.

Note: we may define MLR using decreasing function of t . But all following theorems are stated according to the definition.

Normal Example with Known Mean

$X_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_0, \sigma^2)$ where σ^2 is unknown and μ_0 is known. Find the UMP level α test for testing $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$. Let $T = \sum_{i=1}^n (X_i - \mu_0)^2$ is sufficient for σ^2 . To check whether T has MLR property, we need to find $g(t|\sigma^2)$.

$$\begin{aligned} \frac{X_i - \mu_0}{\sigma} &\sim \mathcal{N}(0, 1) \\ \left(\frac{X_i - \mu_0}{\sigma}\right)^2 &\sim \chi_1^2 \\ Y &= T/\sigma^2 = \sum_{i=1}^n \left(\frac{X_i - \mu_0}{\sigma}\right)^2 \sim \chi_n^2 \\ f_Y(y) &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \end{aligned}$$

Karlin-Rabin Theorem

Theorem 8.1.17

Suppose $T(\mathbf{X})$ is a sufficient statistic for θ and the family $\{g(t|\theta) : \theta \in \Omega\}$ is an MLR family. Then

- ① For testing $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, the UMP level α test is given by rejecting H_0 if and only if $T > t_0$ where $\alpha = \Pr(T > t_0|\theta_0)$.
- ② For testing $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$, the UMP level α test is given by rejecting H_0 if and only if $T < t_0$ where $\alpha = \Pr(T < t_0|\theta_0)$.

Normal Example with Known Mean (cont'd)

$$\begin{aligned} f_T(t) &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} \left(\frac{t}{\sigma^2}\right)^{\frac{n}{2}-1} e^{-\frac{t}{2\sigma^2}} \left|\frac{dy}{dt}\right| dt \\ &= \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} \left(\frac{t}{\sigma^2}\right)^{\frac{n}{2}-1} e^{-\frac{t}{2\sigma^2}} \frac{1}{\sigma^2} dt \\ &= \frac{t^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right) 2^{n/2}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{t}{2\sigma^2}} dt \\ &= h(t)c(\sigma^2) \exp[w(\sigma^2)t] \end{aligned}$$

where $w(\sigma^2) = -\frac{1}{2\sigma^2}$ is an increasing function in σ^2 . Therefore, $T = \sum_{i=1}^n (X_i - \mu)^2$ has the MLR property.

Normal Example with Known Mean (cont'd)

By Karlin-Rabin Theorem, UMP level α rejects H_0 if and only if $T > t_0$ where t_0 is chosen such that $\alpha = \Pr(T > t_0 | \sigma_0^2)$.

Note that $\frac{T^2}{\sigma^2} \sim \chi_n^2$

$$\begin{aligned} \Pr(T > t_0 | \sigma_0^2) &= \Pr\left(\frac{T}{\sigma_0^2} > \frac{t_0}{\sigma_0^2} | \sigma_0^2\right) \\ \frac{T}{\sigma_0^2} &\sim \chi_n^2 \\ \Pr\left(\chi_n^2 > \frac{t_0}{\sigma_0^2}\right) &= \alpha \\ \frac{t_0}{\sigma_0^2} &= \chi_{n,\alpha}^2 \\ t_0 &= \sigma_0^2 \chi_{n,\alpha}^2 \end{aligned}$$

where $\chi_{n,\alpha}^2$ satisfies $\int_{\chi_{n,\alpha}^2}^{\infty} f_{\chi_n^2}(x) dx = \alpha$.

Distribution of LRT

$$\lambda(\mathbf{x}) = \frac{\sup_{\Omega_0} L(\theta | \mathbf{x})}{\sup_{\Omega} L(\theta | \mathbf{x})}$$

LRT level α test procedure rejects H_0 if and only if $\lambda(\mathbf{x}) \leq c$. c is chosen such that

$$\sup_{\theta \in \Omega_0} \Pr(\lambda(\mathbf{x}) \leq c) \leq \alpha$$

Usually, it is difficult to derive the distribution of $\lambda(\mathbf{x})$ and to solve the equation of c .

Remarks

- For many problems, UMP level α test does not exist (Example 8.3.19).
- In such cases, we can restrict our search among a subset of tests, for example, all unbiased tests.

Asymptotics of LRT

Theorem 10.3.1

Consider testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Suppose X_1, \dots, 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 :

$$-2 \log \lambda(\mathbf{x}) \xrightarrow{d} \chi_1^2$$

as $n \rightarrow \infty$.

Proof

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{\sup_{\theta \in \Omega_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Omega} L(\theta|\mathbf{x})} = \frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})} \\ -2\log \lambda(\mathbf{x}) &= -2 \log \left(\frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})} \right) \\ &= -2 \log L(\theta_0|\mathbf{x}) + 2 \log L(\hat{\theta}|\mathbf{x}) \\ &= -2l(\theta_0|\mathbf{x}) + 2l(\hat{\theta}|\mathbf{x}) \end{aligned}$$

Proof (cont'd)

Expanding $l(\theta|\mathbf{x})$ around $\hat{\theta}$,

$$\begin{aligned} l(\theta|\mathbf{x}) &= l(\hat{\theta}|\mathbf{x}) + l'(\hat{\theta}|\mathbf{x})(\theta - \hat{\theta}) + l''(\hat{\theta}|\mathbf{x}) \frac{(\theta - \hat{\theta})^2}{2} + \dots \\ l'(\hat{\theta}|\mathbf{x}) &= 0 \quad (\text{assuming regularity condition}) \\ l(\theta_0|\mathbf{x}) &\approx l(\hat{\theta}|\mathbf{x}) + l''(\hat{\theta}|\mathbf{x}) \frac{(\theta_0 - \hat{\theta})^2}{2} \\ -2 \log \lambda(\mathbf{x}) &= -2l(\hat{\theta}_0|\mathbf{x}) + 2l(\hat{\theta}|\mathbf{x}) \\ &\approx -(\theta_0 - \hat{\theta})^2 l''(\hat{\theta}|\mathbf{x}) \end{aligned}$$

Proof (cont'd)

Because $\hat{\theta}$ is MLE, under H_0 ,

$$\begin{aligned} \hat{\theta} &\sim \mathcal{N}\left(\theta_0, \frac{1}{I_n(\theta_0)}\right) \\ (\hat{\theta} - \theta_0) \sqrt{I_n(\theta_0)} &\xrightarrow{d} \mathcal{N}(0, 1) \\ (\hat{\theta} - \theta_0)^2 I_n(\theta_0) &\xrightarrow{d} \chi_1^2 \end{aligned}$$

Therefore,

$$\begin{aligned} -2 \log \lambda(\mathbf{x}) &\approx -(\theta_0 - \hat{\theta})^2 l''(\hat{\theta}|\mathbf{x}) \\ &= (\theta_0 - \hat{\theta})^2 I_n(\theta_0) \frac{-\frac{1}{n} l''(\hat{\theta}|\mathbf{x})}{\frac{1}{n} I_n(\theta_0)} \end{aligned}$$

Proof (cont'd)

$$\begin{aligned} -\frac{1}{n} l''(\hat{\theta}|\mathbf{x}) &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} f(x_i|\theta) \Big|_{\theta=\hat{\theta}} \\ &\xrightarrow{P} -E \left(\frac{\partial^2}{\partial \theta^2} f(x|\theta) \right) \Big|_{\theta=\hat{\theta}} = I(\theta_0) \end{aligned}$$

Therefore,

$$\frac{-\frac{1}{n} l''(\hat{\theta}|\mathbf{x})}{\frac{1}{n} I_n(\theta_0)} = \frac{-\frac{1}{n} l''(\hat{\theta}|\mathbf{x})}{I(\theta_0)} \xrightarrow{P} 1$$

By Slutsky's Theorem, under H_0

$$\begin{aligned} -(\hat{\theta} - \theta_0)^2 l''(\hat{\theta}|\mathbf{x}) &\xrightarrow{d} \chi_1^2 \\ -2 \log \lambda(\mathbf{x}) &\xrightarrow{d} \chi_1^2 \end{aligned}$$

Example

$X_1 \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$. Consider testing $H_0 : \lambda = \lambda_0$ vs $H-1 : \lambda \neq \lambda_0$. Using LRT,

$$\lambda(\mathbf{x}) = \frac{L(\lambda_0|\mathbf{x})}{L(\lambda|\mathbf{x})}$$

MLE of λ is $\hat{\lambda} = \bar{X} = \frac{1}{n} \sum X_i$.

$$\begin{aligned} \lambda(\mathbf{x}) &= \frac{\prod_{i=1}^n \frac{e^{-\lambda_0} \lambda_0^{x_i}}{x_i!}}{\prod_{i=1}^n \frac{e^{-\bar{x}} \bar{x}^{x_i}}{x_i!}} \\ &= \frac{e^{-n\lambda_0} \lambda_0^{\sum x_i}}{e^{-n\bar{x}} \bar{x}^{\sum x_i}} \\ &= e^{-n(\lambda_0 - \bar{x})} \left(\frac{\lambda_0}{\bar{x}} \right)^{\sum x_i} \end{aligned}$$

Example (cont'd)

LRT is to reject H_0 when $\lambda(\mathbf{x}) \leq c$

$$\begin{aligned} \alpha &= \Pr(\lambda(\mathbf{x}) \leq c | \lambda_0) \\ -2 \log \lambda(\mathbf{x}) &= -2 \left[-n(\lambda_0 - \bar{X}) + \sum X_i (\log \lambda_0 - \log \bar{X}) \right] \\ &= 2n \left(\lambda_0 - \bar{X} - \bar{X} \log \left(\frac{\lambda_0}{\bar{X}} \right) \right) \xrightarrow{d} \chi_1^2 \end{aligned}$$

under H_0 , (by Theorem 10.3.1).

Example (cont'd)

Therefore, asymptotic size α test is

$$\begin{aligned} \Pr(\lambda(\mathbf{x}) \leq c | \lambda_0) &= \alpha \\ \Pr(-2 \log \lambda(\mathbf{x}) \leq c^* | \lambda_0) &= \alpha \\ \Pr(\chi_1^2 \geq c^*) &\approx \alpha \\ c^* &= \chi_{1,\alpha}^2 \end{aligned}$$

rejects H_0 if and only if $-2 \log \lambda(\mathbf{x}) \geq \chi_{1,\alpha}^2$

Wald Test

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

Definition

Suppose W_n is an estimator of θ and $W_n \sim \mathcal{AN}(\theta, \sigma_W^2)$. 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

Examples of Wald Test

Two-sided Wald Test

$H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, then Wald asymptotic level α test is to reject H_0 if and only if

$$|Z_n| > z_{\alpha/2}$$

One-sided Wald Test

$H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, then Wald asymptotic level α test is to reject H_0 if and only if

$$|Z_n| > z_\alpha$$

Example of Wald Test

Suppose $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$, and consider testing $H_0 : p = p_0$ vs $H_1 : p \neq p_0$. MLE of p is \bar{X} , which follows

$$\bar{X} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

by the Central Limit Theorem. The Wald test statistic is

$$Z_n = \frac{\bar{X} - p_0}{S_n}$$

where S_n is a consistent estimator of $\sqrt{\frac{p(1-p)}{n}}$, whose MLE is

$$S_n = \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

by the invariance property of MLE.

Remarks

- Different estimators of θ leads to different testing procedures.
- One choice of W_n is MLE and we may choose $S_n = \frac{1}{I_n(W_n)}$ or $\frac{1}{\hat{I}_n(W_n)}$ (observed information number) when $\sigma_W^2 = \frac{1}{I_n(\theta)}$.

Example of Wald Test (cont'd)

Therefore, S_n is consistent for $\sqrt{\frac{p(1-p)}{n}}$. The Wald statistic is

$$Z_n = \frac{\bar{X} - p_0}{\sqrt{\bar{X}(1-\bar{X})/n}}$$

An asymptotic level α Wald test rejects H_0 if and only if

$$\left| \frac{\bar{X} - p_0}{\sqrt{\bar{X}(1-\bar{X})/n}} \right| > z_{\alpha/2}$$

Interval Estimation

$\hat{\theta}(\mathbf{X})$ is usually represented as a point estimator

Interval Estimator

Let $[L(\mathbf{X}), U(\mathbf{X})]$, where $L(\mathbf{X})$ and $U(\mathbf{X})$ are functions of sample \mathbf{X} and $L(\mathbf{X}) \leq U(\mathbf{X})$. Based on the observed sample \mathbf{x} , we can make an inference that

$$\theta \in [L(\mathbf{X}), U(\mathbf{X})]$$

Then we call $[L(\mathbf{X}), U(\mathbf{X})]$ an interval estimator of θ .

Three types of intervals

- Two-sided interval $[L(\mathbf{X}), U(\mathbf{X})]$
- One-sided (with lower-bound) interval $[L(\mathbf{X}), \infty)$
- One-sided (with upper-bound) interval $(-\infty, U(\mathbf{X})]$

Definitions

Definition : Coverage Probability

Given an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of θ , its *coverage probability* is defined as

$$\Pr(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$$

In other words, the probability of a random variable in interval $[L(\mathbf{X}), U(\mathbf{X})]$ covers the parameter θ .

Definition: Confidence Coefficient

Confidence coefficient is defined as

$$\inf_{\theta \in \Omega} \Pr(\theta \in [L(\mathbf{X}), U(\mathbf{X})])$$

Example

Let $X_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$. Define 1. A point estimator of $\mu : \bar{X}$

$$\Pr(\bar{X} = \mu) = 0$$

2. An interval estimator of $\mu : [\bar{X} - 1, \bar{X} + 1]$

$$\begin{aligned} \Pr(\mu \in [\bar{X} - 1, \bar{X} + 1]) &= \Pr(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) \\ &= \Pr(\mu - 1 \leq \bar{X} \leq \mu + 1) \\ &= \Pr(-\sqrt{n} \leq \sqrt{n}(\bar{X} - \mu) \leq \sqrt{n}) \\ &= \Pr(-\sqrt{n} \leq Z \leq \sqrt{n}) \xrightarrow{P} 1 \end{aligned}$$

as $n \rightarrow \infty$, where $Z \sim \mathcal{N}(0, 1)$.

Definitions

Definition : Confidence Interval

Given an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of θ , if its confidence coefficient is $1 - \alpha$, we call it a $(1 - \alpha)$ *confidence interval*

Definition: Expected Length

Given an interval estimator $[L(\mathbf{X}), U(\mathbf{X})]$ of θ , its *expected length* is defined as

$$E[U(\mathbf{X}) - L(\mathbf{X})]$$

where \mathbf{X} are random samples from $f_{\mathbf{X}}(\mathbf{x}|\theta)$. In other words, it is the average length of the interval estimator.

How to construct confidence interval?

A confidence interval can be obtained by inverting the acceptance region of a test.

There is a one-to-one correspondence between tests and confidence intervals (or confidence sets).

Example (cont'd)

As this is size α test, the probability of accepting H_0 is $1 - \alpha$.

$$\begin{aligned} 1 - \alpha &= \Pr\left(\theta_0 - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \leq \bar{X} \leq \theta_0 + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right) \\ &= \Pr\left(\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \leq \theta_0 \leq \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right) \end{aligned}$$

Since θ_0 is arbitrary,

$$1 - \alpha = \Pr\left(\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \leq \theta \leq \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

Therefore, $[\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}]$ is $(1 - \alpha)$ confidence interval (CI).

Example

$X_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\theta, \sigma^2)$ where σ^2 is known. Consider $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. As previously shown, level α LRT test reject H_0 if and only if

$$\left| \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right| > z_{\alpha/2}$$

Equivalently, we accept H_0 if $\left| \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \right| \leq z_{\alpha/2}$.

Accepting $H_0 : \theta = \theta_0$ because we believe our data "agrees with" the hypothesis $\theta = \theta_0$.

$$\begin{aligned} -z_{\alpha/2} &\leq \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \\ \theta_0 - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} &\leq \bar{X} \leq \theta_0 + \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \end{aligned}$$

Acceptance region is $\left\{ \mathbf{x} : \theta_0 - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \leq \bar{x} \leq \theta_0 + \frac{\sigma}{\sqrt{n}}z_{\alpha/2} \right\}$

Summary

Today

- Asymptotics of LRT
- Wald Test
- Interval Estimation

Next Week

- More on Interval Estimation