Biostat 602 Winter 2017

Lecture Set 18

Hypothesis Testing

Large-Sample Tests

Reading: CB 10.3

Large-sample Results for LRT

Question: Why do we need this?

We have a seen a few examples where the LRT rejection region is equivalent to a rejection region based on a statistic whose distribution is known, at least under H_0, so that the critical (a.k.a. rejection) region could be formed. However, these scenarios are quite limited to some standard distribution examples. In cases where such distributions are not available, one needs to take recourse some approximate method to construct a critical region. The large-sample result of LRT addresses this issue through a general asymptotic (valid for large n) result that applies to a large class of distributions.

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:

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

as $n \to \infty$.

Proof of Theorem 10.3.1: Note that

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

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

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

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

$$
\begin{aligned}
\hat{\theta} &\sim \mathcal{AN}\left(\theta_0, \frac{1}{I_n(\theta_0)}\right) \\
(\hat{\theta} - \theta_0)\sqrt{I_n(\theta_0)} &\xrightarrow{d} \mathcal{N}(0,1) \\
(\hat{\theta} - \theta_0)^2 I_n(\theta_0) &\xrightarrow{d} \chi^2_1
\end{aligned}
$$

Therefore,

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

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

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

By Slutsky's Theorem, under H_0

$$
-(\hat{\theta} - \theta_0)^2 l''(\hat{\theta}|\mathbf{X}) \xrightarrow{d} \chi_1^2
$$

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

The following result is the version of large-sample LRT result that generalizes the above to one with nuisance parameters.

Theorem 10.3.3: Let $X_1, \ldots, X_n$ be a random sample from a pdf or pmf $f(x|\theta)$. (Under the regulatory condition in 10.6.2), if $\theta \in \Omega_0$:

$$
-2 \log \lambda(\mathbf{X}) \xrightarrow{d} \chi_{q-p}^2
$$

if the number of **free** parameters specified by $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega$ are p and q, respectively.

Example 1: Let $X_i \sim Poisson(\lambda)$. Consider testing $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$.

Using LRT,

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

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

$$\lambda(\mathbf{x}) \;=\; \frac{\prod_{i=1}^n \frac{e^{-\lambda_0}\lambda_0^{x_i}}{x_i!}}{\prod_{i=1}^n \frac{e^{-\overline{x}}\;\overline{x}^{x_i}}{x_i!}} \;=\; \frac{e^{-n\lambda_0}\lambda_0^{\sum x_i}}{e^{-n\overline{x}}\;\overline{x}^{\sum x_i}} \;=\; e^{-n(\lambda_0-\overline{x})}\left(\frac{\lambda_0}{\overline{x}}\right)^{\sum x_i}$$

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

$$\alpha \;=\; \Pr(\lambda(\mathbf{X}) \leq c|\lambda_0)$$

$$\begin{aligned}
-2\log\lambda(\mathbf{X}) &= -2\left[-n(\lambda_0 - \overline{X}) + \sum X_i(\log\lambda_0 - \log\overline{X})\right] \\
&= 2n\left(\lambda_0 - \overline{X} - \overline{X}\log\left(\frac{\lambda_0}{\overline{X}}\right)\right) \xrightarrow{d} \chi_1^2
\end{aligned}$$

under H_0, (by Theorem 10.3.1).

Therefore, asymptotic size α test is given by

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

which 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 $W_n \sim \mathcal{AN}(\theta, \sigma^2_{W_n})$. 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_n}

Two-sided Wald Test:

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

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

One-sided Wald Test:

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

$$Z_n > z_\alpha$$

Remarks:

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

Example 2: Suppose $X_i \sim Bernoulli(p)$, and consider testing $H_0 : p = p_0$ vs $H_1 : p \neq p_0$.

MLE of p is $\overline{X}$, which follows

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

by the Central Limit Theorem. So the Wald test statistic is

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

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

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

which is the MLE of $\sqrt{\frac{p(1-p)}{n}}$ by the invariance property of MLE.

The Wald statistic is

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

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

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

Score Test

Definition: Let $S(\theta) = \frac{\partial}{\partial\theta} \log L(\theta|\mathbf{x})$ be a score function. Then the variance of the score function is

$$\mathrm{Var}\left[S(\theta)\right] = \mathrm{E}\left[S^2(\theta)\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2} \log L(\theta|\mathbf{x})\right] = I_n(\theta)$$

if the interchangeability condition holds. The test statistic for score test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is

$$Z_S = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}$$

If H_0 is true

- Z_S has mean 0 and variance 1.
- $Z_S \xrightarrow{d} \mathcal{N}(0, 1)$.

Example 3: Let $X_i \sim Bernoulli(p)$. Consider testing $H_0 : p = p_0$ vs $H_1 : p \neq p_0$.

The likelihood and score function is

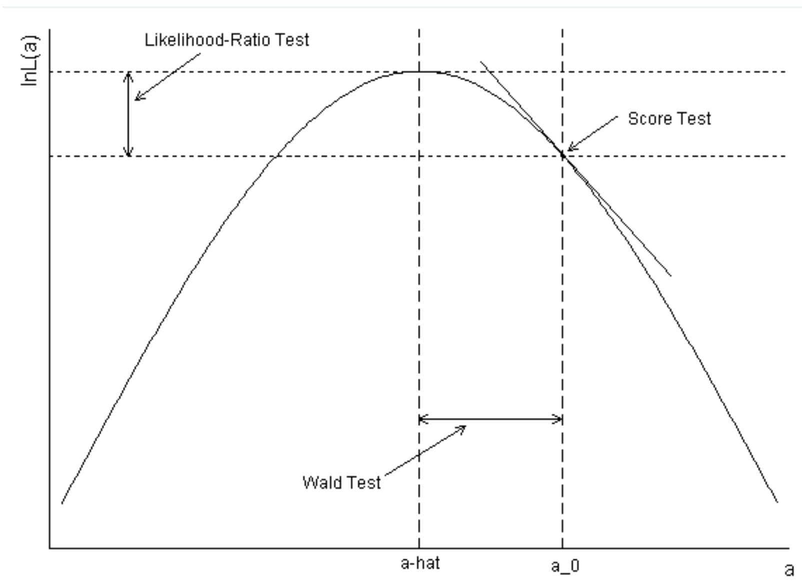
$$\begin{aligned}
\log L(p|\mathbf{x}) &= \sum x_i \log p + (n - \sum x_i) \log(1-p) \\
S(p) &= \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = \frac{\overline{x} - p}{p(1-p)/n} \\
I(p) &= \frac{1}{p(1-p)}
\end{aligned}$$

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

$$|Z_S| = \left|\frac{S(p_0)}{\sqrt{I_n(p_0)}}\right| = \left|\frac{\overline{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right| > z_{\alpha/2}$$

Comparison of the Three Tests

- The three tests are approximately equivalent in terms of asymptotic power.
- For likelihood functions that are not well-behaved, LRT has the best small-sample properties.



Example 4: Let $X_1, \ldots, X_n$ be *i.i.d.* random variables from Exponential(θ) distribution with pdf

$$f_X(x) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) I(x > 0), \qquad \theta > 0$$

(a) Construct a large-sample (asymptotic) size α Wald test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ for an arbitrary $\theta_0 > 0$.

(b) Consider a test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ given by the following rejection region:

$$R = \left\{ \mathbf{X} : \frac{\sqrt{n}(\overline{X} - \theta_0)}{\theta_0} > z_\alpha \right\}$$

where z_α is upper α-quantile of $N(0,1)$. Is the test defined above always more powerful than the Wald test defined in part (a)? Justify your answer.