Lecture 23: Asymptotic Evaluations of Likelihood Ratio Tests

Mathematical Statistics II, MATH 60062/70062

Tuesday April 26, 2022

Reference: Casella & Berger, 10.3

Recap: 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.

Recap: 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.

Large sample likelihood ratio test statistic

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

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

The likelihood ratio test (LRT) statistic is defined as

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x})}{\sup_{\theta \in \Theta} L(\theta \mid \boldsymbol{x})} = \frac{L(\theta_0 \mid \boldsymbol{x})}{L(\hat{\theta} \mid \boldsymbol{x})} = \frac{L(\theta_0)}{L(\hat{\theta})}.$$

Suppose the regularity conditions needed for MLEs to be consistent and asymptotically normal hold. Under H_0,

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

Small values of $\lambda(\boldsymbol{x})$ are evidence against H_0, so are large values of $-2 \log \lambda(\boldsymbol{x})$. Therefore,

$$R = \{\boldsymbol{x} \in \mathcal{X} : -2 \log \lambda(\boldsymbol{x}) \geq \chi^2_{1,\alpha}\},$$

where $\chi^2_{1,\alpha}$ is the upper α quantile of the χ^2_1 distribution, is an approximate size α rejection region for testing H_0 versus H_1.

Proof of the asymptotic distribution of LRT statistic

Suppose $H_0 : \theta = \theta_0$ is true. First, expand $\log L(\hat{\theta})$ in a Taylor series around θ_0, giving

$$\begin{aligned}\log L(\hat{\theta}) &= \log L(\theta_0) + (\hat{\theta} - \theta_0)\frac{\partial}{\partial \theta}\log L(\theta_0) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial \theta^2}\log L(\hat{\theta}_*) \\ &= \log L(\theta_0) + \sqrt{n}(\hat{\theta} - \theta_0)\frac{1}{\sqrt{n}}\frac{\partial}{\partial \theta}\log L(\theta_0) + \frac{n}{2}(\hat{\theta} - \theta_0)^2 \frac{1}{n}\frac{\partial^2}{\partial \theta^2}\log L(\hat{\theta}_*\end{aligned}$$

where $\hat{\theta}_*$ is between $\hat{\theta}$ and θ_0. Now expand $\frac{\partial}{\partial \theta}\log L(\theta_0)$ in a Taylor series around $\hat{\theta}$,

$$\frac{\partial}{\partial \theta}\log L(\theta_0) = \underbrace{\frac{\partial}{\partial \theta}\log L(\hat{\theta})}_{=0} + (\theta_0 - \hat{\theta})\frac{\partial^2}{\partial \theta^2}\log L(\hat{\theta}_{**}),$$

where $\hat{\theta}_{**}$ is between θ_0 and $\hat{\theta}$. Note that $\frac{\partial}{\partial \theta}\log L(\hat{\theta}) = 0$ because $\hat{\theta}$ solves the score equation. Thus,

$$\frac{1}{\sqrt{n}}\frac{\partial}{\partial \theta}\log L(\theta_0) = \sqrt{n}(\hat{\theta} - \theta_0)\left\{-\frac{1}{n}\frac{\partial^2}{\partial \theta^2}\log L(\hat{\theta}_{**})\right\}.$$

Combining the equations, we have

$$\log L(\hat{\theta}) = \log L(\theta_0) + \sqrt{n}(\hat{\theta} - \theta_0)\sqrt{n}(\hat{\theta} - \theta_0)\left\{-\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\hat{\theta}_{**})\right\} + \frac{n}{2}(\hat{\theta} - \theta_0)^2\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\hat{\theta}_*),$$

so that

$$\log L(\hat{\theta}) - \log L(\theta_0) = -n(\hat{\theta}-\theta_0)^2\left\{\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\hat{\theta}_{**})\right\} + \frac{n}{2}(\hat{\theta}-\theta_0)^2\left\{\frac{1}{n}\frac{\partial^2}{\partial\theta^2}\log L(\hat{\theta}_*)\right\}$$

Because $\hat{\theta}$ is consistent, under $H_0 : \theta = \theta_0$, $\hat{\theta} \xrightarrow{p} \theta_0$, as $n \to \infty$. Also, because both $\hat{\theta}_{**}$ and $\hat{\theta}_*$ are between $\hat{\theta}$ and θ_0, by WLLN both terms in the brackets converge in probability to $E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X \mid \theta)\right] = -I_1(\theta_0)$. Therefore, the RHS of the above equation will behave in the limit the same as

$$\frac{n}{2}(\hat{\theta}-\theta_0)^2 I_1(\theta_0) = \frac{1}{2}\sqrt{n}(\hat{\theta}-\theta_0)\sqrt{n}(\hat{\theta}-\theta_0)I_1(\theta_0) = \frac{1}{2}\underbrace{\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sqrt{\frac{1}{I_1(\theta_0)}}}}_{\xrightarrow{d}\mathcal{N}(0,1)}\underbrace{\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sqrt{\frac{1}{I_1(\theta_0)}}}}_{\xrightarrow{d}\mathcal{N}(0,1)} \xrightarrow{d} \frac{1}{2}\chi_1^2$$

by continuity. Therefore, under $H_0 : \theta = \theta_0$,

$$-2\log\lambda(\boldsymbol{X}) = -2[\log L(\theta_0) - \log L(\hat{\theta})] \xrightarrow{d} \chi_1^2.$$

Large sample Bernoulli LRT

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 LRT statistic is

$$\begin{aligned}
\lambda(\boldsymbol{x}) = \frac{L(p_0 \mid \boldsymbol{x})}{L(\hat{p} \mid \boldsymbol{x})} &= \frac{p_0^{\sum_{i=1}^n x_i}(1-p_0)^{n-\sum_{i=1}^n x_i}}{\hat{p}^{\sum_{i=1}^n x_i}(1-\hat{p})^{n-\sum_{i=1}^n x_i}} \\
&= \left(\frac{p_0}{\hat{p}}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_0}{1-\hat{p}}\right)^{n-\sum_{i=1}^n x_i}.
\end{aligned}$$

To find the rejection region, write

$$\begin{aligned}
-2\log\lambda(\boldsymbol{X}) &= -2\left[\sum_{i=1}^n X_i \log\left(\frac{p_0}{\hat{p}}\right) + \left(n - \sum_{i=1}^n X_i\right)\log\left(\frac{1-p_0}{1-\hat{p}}\right)\right] \\
&= -2\left[n\hat{p}\log\left(\frac{p_0}{\hat{p}}\right) + n(1-\hat{p})\log\left(\frac{1-p_0}{1-\hat{p}}\right)\right].
\end{aligned}$$

An approximate size α rejection region is

$$R = \{\boldsymbol{x} \in \mathcal{X} : -2\log\lambda(\boldsymbol{x}) \geq \chi^2_{1,\alpha}\}.$$ R: lambda(x)<=c

Summary

Suppose $X_1, \dots, X_n$ are iid from $f_X(x \mid \theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$. We have discussed three large sample procedures to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

- **Wald test:**

$$Z_n^W = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{v(\hat{\theta})}{n}}} = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{1}{nI_1(\hat{\theta})}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

- **Score test:**

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

- **LRT:**

$$-2 \log \lambda(\boldsymbol{X}) = -2[\log L(\theta_0 \mid \boldsymbol{X}) - \log L(\hat{\theta} \mid \boldsymbol{X})] \xrightarrow{d} \chi_1^2.$$

Under H_0, $(Z_n^W)^2$, $(Z_n^S)^2$, and $-2 \log \lambda(\boldsymbol{X})$ each converge in distribution to a χ_1^2 as $n \to \infty$. However, their true sizes (and powers) can be quite different in finite samples.

Monte Carlo simulations

Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Bern}(p)$, where $0 < p < 1$. We have derived the Wald, score, and large sample LRT for testing

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

Simulations to assess the finite sample properties of the three tests:

- Sample size $n = 20$, $n = 50$, $n = 100$
- True parameter $p_0 = 0.1$, $p_0 = 0.3$
- 10,000 samples for each pair of n and p_0

Percentage of times that H_0 is (incorrectly) rejected with $\alpha = 0.5$:

		Wald	Score	LRT
	$n = 20$	0.122	0.0433	0.0114
$p_0 = 0.1$	$n = 50$	0.116	0.0289	**0.0536**
	$n = 100$	0.0677	0.0659	**0.0468**
	$n = 20$	**0.0529**	0.0263	**0.0515**
$p_0 = 0.3$	$n = 50$	0.0647	0.0438	0.0438
	$n = 100$	**0.0533**	0.0671	**0.0533**

The percentage is an estimate of the true size of the test.

- With a 99% confidence level, the margin of error associated with each estimate is

$$0.258 \times \sqrt{\frac{0.05(1 - 0.05)}{10000}} \approx 0.0056.$$

- Size estimates between 0.0444 and 0.0556 indicate that the test is operating at the expected level.