Hypothesis Testing in a Likelihood Framework

Eric Zivot

Winter 2013

Hypothesis Testing Framework (Simplified Setting)

Let $X_1, \ldots, X_n$ be iid with density $f(x_i; \theta)$, where θ is a scalar. We estimate θ by ML giving $\hat{\theta}_{mle}$. From the asymptotic properties of MLE we have

$$\hat{\theta}_{mle} \overset{A}{\sim} N\left(\theta, \hat{I}(\hat{\theta}_{mle}|\mathbf{x})^{-1}\right)$$

$$\hat{I}(\hat{\theta}_{mle}|\mathbf{x}) = n\hat{I}(\hat{\theta}_{mle}|x_i) \text{ and } \hat{I}(\hat{\theta}_{mle}|x_i) \xrightarrow{p} I(\theta_0|x_i)$$

where

$$I(\theta_0|x_i) = -E[H(\theta_0|x_i)] = \text{var}(S(\theta_0|x_i))$$

Consider testing

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

Trinity of Test Statistics

There are three asymptotically equivalent test statistics (the so-called trinity of tests): Wald, Lagrange Multiplier (LM)/Score and Likelihood Ratio(LR):

$$
\begin{aligned}
Wald &= \frac{\left(\hat{\theta}_{mle} - \theta_0\right)^2}{\hat{I}(\hat{\theta}_{mle}|\mathbf{x})^{-1}} = \left(\hat{\theta}_{mle} - \theta_0\right) \hat{I}(\hat{\theta}_{mle}|\mathbf{x}) \left(\hat{\theta}_{mle} - \theta_0\right) \\
LM &= \frac{S(\theta_0|\mathbf{x})^2}{I(\theta_0|\mathbf{x})} = S(\theta_0|\mathbf{x}) I(\theta_0|\mathbf{x})^{-1} S(\theta_0|\mathbf{x}) \\
LR &= 2[\ln L(\hat{\theta}_{mle}|\mathbf{x}) - \ln L(\theta_0|\mathbf{x})]
\end{aligned}
$$

such that under $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$

$$Wald, LM, LR \xrightarrow{d} \chi^2(1)$$

Wald Statistic

The Wald statistic is based directly on the asymptotic normal distribution of $\hat{\theta}_{mle}$:

$$\hat{\theta}_{mle} \overset{A}{\sim} N\left(\theta, \hat{I}(\hat{\theta}_{mle}|\mathbf{x})^{-1}\right)$$

An implication of the asymptotic normality result is that the usual t-ratio for testing $H_0 : \theta = \theta_0$

$$\begin{aligned} t &= \frac{\hat{\theta}_{mle} - \theta_0}{\widehat{SE}(\hat{\theta}_{mle})} = \frac{\hat{\theta}_{mle} - \theta_0}{\sqrt{\hat{I}(\hat{\theta}_{mle}|\mathbf{x})^{-1}}} \\ &= \left(\hat{\theta}_{mle} - \theta_0\right)\sqrt{\hat{I}(\hat{\theta}_{mle}|\mathbf{x})} \overset{A}{\sim} N(0,1) \end{aligned}$$

Using the continuous mapping theorem, it follows that

$$\begin{aligned} Wald &= t^2 = \frac{\left(\hat{\theta}_{mle} - \theta_0\right)^2}{\hat{I}(\hat{\theta}_{mle}|\mathbf{x})^{-1}} \\ &= \left(\hat{\theta}_{mle} - \theta_0\right)^2 \hat{I}(\hat{\theta}_{mle}|\mathbf{x}) \overset{A}{\sim} \chi^2(1) \end{aligned}$$

Note: if the curvature of ln $L(\theta|\mathbf{x})$ near $\theta = \hat{\theta}_{mle}$ is big (high information) then the squared distance $\left(\hat{\theta}_{mle} - \theta_0\right)^2$ gets blown up when constructing the Wald statistic. If the curvature of ln $L(\theta|\mathbf{x})$ near $\theta = \hat{\theta}_{mle}$ is low, then $I(\hat{\theta}_{mle}|\mathbf{x})$ is small and the squared distance $\left(\hat{\theta}_{mle} - \theta_0\right)^2$ gets attenuated when constucting the Wald statistic.

Lagrange Multiplier/Score Statistic

With ML estimation, $\hat{\theta}_{mle}$ solves the first order conditions

$$0 = \frac{d \ln L(\hat{\theta}_{mle}|\mathbf{x})}{d\theta} = S(\hat{\theta}_{mle}|\mathbf{x})$$

If $H_0 : \theta = \theta_0$ is true, then we should expect that

$$0 \approx \frac{d \ln L(\theta_0|\mathbf{x})}{d\theta} = S(\theta_0|\mathbf{x})$$

If $H_0 : \theta = \theta_0$ is not true, then we should expect that

$$0 \neq \frac{d \ln L(\theta_0|\mathbf{x})}{d\theta} = S(\theta_0|\mathbf{x})$$

The Lagrange multiplier (score) statistic is based on how far $S(\theta_0|\mathbf{x})$ is from zero.

Recall the following properties of the score $S(\theta|x_i)$. If $H_0 : \theta = \theta_0$ is true then

$$\begin{aligned} E[S(\theta_0|x_i)] &= 0 \\ \text{var}(S(\theta_0|x_i)) &= I(\theta_0|x_i) \end{aligned}$$

Further, it can be shown that

$$\frac{1}{\sqrt{n}} S(\theta_0|\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S(\theta_0|x_i) \xrightarrow{d} N(0, I(\theta_0|x_i))$$

so that

$$S(\theta_0|\mathbf{x}) \overset{A}{\sim} N\left(0, nI(\theta_0|x_i)\right) = N\left(0, I(\theta_0|\mathbf{x})\right)$$

These results motivate the statistic

$$LM = \frac{S(\theta_0|\mathbf{x})^2}{I(\theta_0|\mathbf{x})} = S(\theta_0|\mathbf{x})^2 I(\theta_0|\mathbf{x})^{-1}$$

Under general regularity conditions, if $H_0 : \theta = \theta_0$ is true, then

$$\begin{aligned} LM &= S(\theta_0|\mathbf{x})^2 I(\theta_0|\mathbf{x})^{-1} \\ &= \left(\frac{1}{\sqrt{n}} S(\theta_0|\mathbf{x})\right)^2 \left(\frac{1}{n} I(\theta_0|\mathbf{x})\right)^{-1} \\ &\xrightarrow{d} \left(N(0, I(\theta_0|x_i))\right)^2 I(\theta_0|x_i)^{-1} \sim \chi^2(1) \end{aligned}$$

Remark: Typically, $I(\theta_0|\mathbf{x})$ is not known and is replaced by

$$\tilde{I}(\theta_0|\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} S(\theta_0|x_i)^2 \rightarrow E\left[S(\theta_0|x_i)^2\right] = I(\theta_0|x_i)$$

Lagrange Multiplier Interpretation

Consider the constrained estimation problem

$$\max \ln L(\theta|\mathbf{x}) \text{ s.t. } \theta = \theta_0$$

The Lagrangian is

$$L = \ln L(\theta|\mathbf{x}) + \lambda(\theta - \theta_0)$$

The FOCs are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \theta} = S(\tilde{\theta}|\mathbf{x}) + \tilde{\lambda} \Rightarrow \tilde{\lambda} = -S(\tilde{\theta}|\mathbf{x}) \\ 0 &= \frac{\partial L}{\partial \lambda} = \tilde{\theta} - \theta_0 \Rightarrow \tilde{\theta} = \theta_0 \end{aligned}$$

so that

$$\tilde{\lambda} = -S(\tilde{\theta}|\mathbf{x}) = -S(\theta_0|\mathbf{x})$$

Then the score statistic can be re-written as a function of the Lagrange multiplier $\tilde{\lambda}$

$$\begin{aligned} LM &= S(\theta_0|\mathbf{x})^2 I(\theta_0|\mathbf{x})^{-1} \\ &= \tilde{\lambda}^2 I(\theta_0|\mathbf{x})^{-1} \end{aligned}$$

which explains why the statistic is also called the Lagrange Multiplier statistic.

Likelihood Ratio Statistic

Consider the likelihood ratio

$$\lambda = \frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}_{mle}|\mathbf{x})} = \frac{L(\theta_0|\mathbf{x})}{\max_\theta L(\theta|\mathbf{x})}$$

By construction $0 < \lambda \leq 1$. If $H_0 : \theta = \theta_0$ is true, then we should see $\lambda \approx 1$; if $H_0 : \theta = \theta_0$ is not true then we should see $\lambda < 1$.

The likelihood ratio (LR) statistic is a simple transformation of λ such that the value of LR is small if $H_0 : \theta = \theta_0$ is true, and the value of LR is large when $H_0 : \theta = \theta_0$ is not true.

Formally, the LR statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is

$$\begin{aligned} LR &= -2\ln\lambda = -2\ln\frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}_{mle}|\mathbf{x})} \\ &= -2[\ln L(\theta_0|\mathbf{x}) - \ln L(\hat{\theta}_{mle}|\mathbf{x})] \\ &= 2[\ln L(\hat{\theta}_{mle}|\mathbf{x}) - \ln L(\theta_0|\mathbf{x})] \end{aligned}$$

Under general regularity conditions, if $H_0 : \theta = \theta_0$ is true then

$$LR \xrightarrow{d} \chi^2(1)$$

In general, the degrees of freedom of the chi-square limiting distribution depends on the number of restrictions imposed under the null hypothesis. The decision rule for the LR statistic is to reject $H_0 : \theta = \theta_0$ at the $\alpha \times 100\%$ level if $LR > \chi^2_{1-\alpha}(1)$, where $\chi^2_{1-\alpha}(1)$ is the $(1-\alpha) \times 100\%$ quantile of the chi-square distribution with 1 degree of freedom.

Sketch of proof of chi-square limiting distribution for LR statistic in scalar case

$$LR = 2[\ln L(\hat{\theta}_{mle}|\mathbf{x}) - \ln L(\theta_0|\mathbf{x})]$$

Step 1: Do 2nd order TSE of $\ln L(\hat{\theta}_{mle}|\mathbf{x})$ about θ_0

$$\begin{aligned}\ln L(\hat{\theta}_{mle}|\mathbf{x}) &= \ln L(\theta_0|\mathbf{x}) + S(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0) \\ &\quad + \frac{1}{2}H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + o_p(1)\end{aligned}$$

Step 2: Substitute TSE into expression for $\ln L(\hat{\theta}_{mle}|\mathbf{x})$ in LR statistic

$$\begin{aligned}LR &= 2[\ln L(\hat{\theta}_{mle}|\mathbf{x}) - \ln L(\theta_0|\mathbf{x})] \\ &= 2\ln L(\theta_0|\mathbf{x}) + 2S(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0) \\ &\quad + H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 - 2\ln L(\theta_0|\mathbf{x}) + o_p(1) \\ &= 2S(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0) + H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + o_p(1)\end{aligned}$$

Step 3: Use F.O.C's and TSE for MLE to derive expression for $S(\theta_0|x)$

$$\begin{aligned} 0 &= S(\hat{\theta}_{mle}|\mathbf{x}) = S(\theta_0|\mathbf{x}) + H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0) + o_p(1) \\ &\Rightarrow S(\theta_0|\mathbf{x}) = -H(\theta_0|x)(\hat{\theta}_{mle} - \theta_0) + o_p(1) \end{aligned}$$

Step 4: Substitute expression for $S(\theta_0|\mathbf{x})$ into expression for LR

$$\begin{aligned} LR &= 2S(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0) + H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + o_p(1) \\ &= -2H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + o_p(1) \\ &= -H(\theta_0|\mathbf{x})(\hat{\theta}_{mle} - \theta_0)^2 + o_p(1) \\ &= -\frac{1}{n}H(\theta_0|\mathbf{x})(\sqrt{n}(\hat{\theta}_{mle} - \theta_0))^2 + o_p(1) \\ &\xrightarrow{d} I(\theta_0|x_i) \cdot \Big(N(0, I(\theta_0|x_i)^{-1})\Big)^2 \sim \chi^2(1) \end{aligned}$$