Recitation Note: Which Wald Test?

Tyler Maunu
maunut@mit.edu

September 3, 2019

This note will discuss two ways of carrying out the asymptotic Wald test.

Introduction

Recall that the Wald test uses the central limit theorem or asymptotic normality of the MLE to give an approximate test. Suppose that we wish to estimate a parameter θ with an associated distribution function F_θ. Suppose that we use an estimator $\hat{\theta}$ that is asymptotically normal:

$$\frac{\hat{\theta} - \theta_0}{\sqrt{\operatorname{var}(\hat{\theta})}} \xrightarrow{d} N(0,1), \tag{1}$$

and suppose we want to test

$$H_0 : \theta = \theta_0, \text{ vs. } H_1 : \theta \neq \theta_0, \tag{2}$$

for some specified value of θ_0. Then, the Wald test uses the quantiles of the standard normal distribution to do the test

$$\psi = \mathbb{1}(|T| > q_{\alpha/2}). \tag{3}$$

Here, $T = \frac{\hat{\theta}-\theta_0}{\sqrt{\operatorname{var}(\hat{\theta})}}$ and $q_{\alpha/2}$ is the $1 - \alpha/2$ quantile of $N(0, 1)$.

Usually, $\operatorname{var}(\hat{\theta})$ is itself a function of θ (say it equals $v(\theta)$). So, we can write the test as

$$\psi = \mathbb{1}\left(\left|\frac{\hat{\theta} - \theta_0}{\sqrt{h(\theta)}}\right| > q_{\alpha/2}\right). \tag{4}$$

Now since you don't know θ, how could you possibly do the test? The answer is to either use Slutsky's and plug in $\hat{\theta}$ (and so the denominator becomes $\sqrt{h(\hat{\theta})}$) or plug in the null hypothesis value θ_0 (where the denominator then becomes $\sqrt{h(\theta_0)}$). Both yield asymptotic $N(0, 1)$ (under the null hypothesis) and are thus valid asymptotic tests. They may give different answers in practice though.

Question

Suppose that we observe $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f_\theta$, where

$$f_\theta(x) = \begin{cases} (1-\theta)^2, & x = -1 \\ 2\theta(1-\theta), & x = 0 \\ \theta^2, & x = 1. \end{cases} \tag{5}$$

This is a discrete distribution used to model three distinct outcomes. It is the expected distribution in the Hardy-Weinburg equilibrium model, and is used to model genotype frequencies.

1. Calculate the mean, variance and maximum likelihood estimator for θ under this model.
2. Derive the Wald test.
3. Compare the two different versions of the Wald test on the given data.
4. Compare the performance of these tests.

Solution

1. In our set of observed random variables, let n_{-1} be the number of $X_i = -1$, and similarly for n_0 and n_1. Our likelihood is

$$\begin{aligned} L(X_1, \ldots, X_n; \theta) &= [(1-\theta)^2]^{n_{-1}}[1\theta(1-\theta)]^{n_0}[\theta^2]^{n_1} \\ &= (1-\theta)^{2n_{-1}+n_0}2^{n_0}\theta^{2n_1+n_0}. \end{aligned} \tag{6}$$

The log-likelihood is given by

$$\ell(X_1, \ldots, X_n; \theta) = (2n_{-1} + n_0)\log(1-\theta) + n_0 \log 2 + (2n_1 + n_0)\log\theta. \tag{7}$$

The derivative is

$$\frac{d}{d\theta}\ell(X_1, \ldots, X_n; \theta) = -\frac{2n_{-1} + n_0}{1-\theta} + \frac{2n_1 + n_0}{\theta}. \tag{8}$$

Setting this to zero and solving for θ, we find the MLE is

$$\hat{\theta}_{MLE} = \frac{2n_1 + n_0}{2n}. \tag{9}$$

The Fisher Information is given by

$$\begin{aligned} I(\theta) = -\mathbb{E}\frac{d^2}{d\theta^2}\ell(X;\theta) &= -\mathbb{E}\left[-\frac{2\mathbb{1}(X=-1) + \mathbb{1}(X=0)}{(1-\theta)^2} - \frac{2\mathbb{1}(X=1) + \mathbb{1}(X=0)}{\theta^2}\right] \\ &= \frac{2(1-\theta)^2 + 2\theta(1-\theta))}{(1-\theta)^2} + \frac{2\theta^2 + 2\theta(1-\theta)}{\theta^2} = \frac{2}{\theta(1-\theta)}. \end{aligned} \tag{10}$$

Therefore, by the asymptotic theorem for the MLE, we have

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I^{-1}(\theta)) = N\left(0, \frac{\theta(1-\theta)}{2}\right). \tag{11}$$

2. In its most general form, suppose we want to test

$$H_0 : g(\theta) = 0 \text{ vs } H_1 : g(\theta) \neq 0,$$

and we have n observations which we use to find our MLE $\hat{\theta}_n$. We assume that $g'(\theta)$ is continuous and $g'(\theta) \neq 0$ when $g(\theta) = 0$. We can do a Taylor expansion of $g(\theta)$ around θ to find

$$g(\hat{\theta}_n) = g(\theta) + g'(\tilde{\theta})(\hat{\theta}_n - \theta). \tag{12}$$

Here, $\tilde{\theta}$ lies in between $\hat{\theta}_n$ and θ. Since $\hat{\theta}_n \xrightarrow{p} \theta$, we also have that $\tilde{\theta} \xrightarrow{p} \theta$. Rearranging and applying Slutsky's Theorem, we find

$$\sqrt{n}(g(\hat{\theta}_n - g(\theta)) = \sqrt{n}g'(\tilde{\theta})(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, (g'(\theta)^2)I^{-1}(\theta)), \tag{13}$$

as long as $g'(\theta) \neq 0$. By Slutsky's Theorem, we therefore have

$$\frac{\sqrt{n}(g(\hat{\theta}_n) - g(\theta)}{|g'(\theta)|\sqrt{I^{-1}(\theta)}} \xrightarrow{d} N(0, 1). \tag{14}$$

Equivalently,

$$\frac{\sqrt{n}(g(\hat{\theta}_n) - g(\theta)}{g'(\theta)^2 I^{-1}(\theta)} \xrightarrow{d} \chi_1^2. \tag{15}$$

Under the null hypothesis, we have (since $g(\theta) = 0$)

$$\frac{n(g(\hat{\theta}_n)^2}{g'(\theta)^2 I^{-1}(\theta)} \xrightarrow{d} \chi_1^2 \tag{16}$$

Here, we now have two options. We can solve $g(\theta) = 0$ for θ and then plug this into the denominator. Denote this solution by θ_0. On the other hand, we could use Slutsky's to justify plugging in $\hat{\theta}_n$ into the denominator. These yield the following statements:

$$\begin{aligned} \frac{n(g(\hat{\theta}_n)^2}{g'(\theta_0)^2 I^{-1}(\theta_0)} &\xrightarrow{d} \chi_1^2 \\ \frac{n(g(\hat{\theta}_n)^2}{g'(\hat{\theta}_n)^2 I^{-1}(\hat{\theta}_n)} &\xrightarrow{d} \chi_1^2 \end{aligned} \tag{17}$$

While the left hand sides define two different test statistics, they are both asymptotically valid for this test. They may even give different answers in some cases, which we will see in the next part. This does not mean that one is more right than the other, though – both of these are based on approximations, and it is hard to distinguish if one is better than the other in general. The plug in method (with the MLE) may be nice to use since we do not have to solve the equation $g(\theta) = 0$ explicitly.

3. Suppose we take 100 measurements. Consider the following two cases: We wish to test

$$H_0 : \theta - .2 = 0 \text{ vs } H_1 : \theta - .2 \neq 0.$$

	n_{-1}	n_0	n_1	$\hat{\theta}_n$
Experiment 1	17	18	65	.26
Experiment 2	5	20	75	.15

The test statistics for our two versions of the Wald test are

$$\psi_1 = \mathbb{1}\left(\frac{n(\hat{\theta}_n - .2)^2}{I^{-1}(\theta_0)} > 3.84\right) \tag{18}$$

$$\psi_2 = \mathbb{1}\left(\frac{n(\hat{\theta}_n - .2)^2}{I^{-1}(\hat{\theta}_n)} > 3.84\right).$$

Here, $n = 100$, $\theta_0 = .2$, and 3.84 is the 0.95-quantile of the χ_1^2 distribution.

Denoting the statistics as T_1 and T_2, we find the following values.

	T_1	T_2
Experiment 1	4.5	3.74
Experiment 2	3.125	3.92

As we can see, the tests give different answers in both cases.

4. So how can we compare the performance of these tests in general?

One tool that may help is *simulation studies.* While this is beyond the scope of this course, we can outline the procedure and understand what is happening. It may be a helpful exercise to run some simulation studies in your own favorite programming language (our recommended choices would be to use R or Python).

Suppose that I want to examine the performance of my statistics on the following test:

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta = \theta_1. \tag{19}$$

Type I and Type II error rates are defined as

$$P(\text{type I error}) = P(\text{rejecting } H_0 \text{ when it is true}). \tag{20}$$

$$P(\text{type II error}) = P(\text{accepting } H_0 \text{ when } H_1 \text{ is true}). \tag{21}$$

These are the false positive and false negative rates, respectively.

Suppose I generate 1000 datasets according to H_0. Then, how can I estimate $P(\text{type I error})$? Let ψ_1^j be the output of the test ψ_1 on the jth dataset. By the law of large numbers, a consisten estimate of this quantity is given by

$$\frac{1}{1000}\sum_{j=1}^{1000}\psi_1^j. \tag{22}$$

Similarly, if I generate 1000 datasets from H_1, a consisten estimate of the type II error rate is given by

$$\frac{1}{1000}\sum_{j=1}^{n}(1 - \psi_1^j). \tag{23}$$

Thus, while I cannot explicitly calculate the type I and type II error rates for these two tests, ψ_1 and ψ_2, I can estimate these rates in a consistent way. I can therefore use a simulation study to see which test I prefer for a given dataset size n and set of hypotheses. Some code will be given to run these simulation studies.

The following table presents results from a simulation study.

Hypotheses	Test	Type I Error Rate	Simulated Type II Error Rate
$H_0 : \theta = 0.2$ vs $H_1 : \theta = 0.1$	1	0.455	0.0272
	2	0.617	0.0093
$H_0 : \theta = 0.2$ vs $H_1 : \theta = 0.3$	1	0.406	0.0948
	2	0.589	0.1236

Table 1: Simulation study for the tests ψ_1 and ψ_2. We fix the number of observations as $n = 100$. In each, we generate 10000 datasets from H_0 and 10000 datasets from H_1. We then calculate the output of the tests ψ_1 and ψ_2, and use (22) and (23) to estimate the error rates.