

Lecture 21: October 16

Lecturer: Siva Balakrishnan

In the last class we showed that the Neyman-Pearson test is optimal for testing simple versus simple hypothesis tests. Today we will develop some generalizations and tests that are useful in other more complex settings.

21.1 The Wald Test

When we are testing a simple null hypothesis against a possibly composite alternative, the NP test is no longer applicable and a general alternative is to use the Wald test.

We are interested in testing the hypotheses in a parametric model:

$$\begin{aligned}H_0 &: \theta = \theta_0 \\H_1 &: \theta \neq \theta_0.\end{aligned}$$

The Wald test most generally is based on an asymptotically normal estimator, i.e. we suppose that we have access to an estimator $\hat{\theta}$ which under the null satisfies the property that:

$$\hat{\theta} \xrightarrow{d} N(\theta_0, \sigma_0^2),$$

where σ_0^2 is the variance of the estimator under the null. The canonical example is when $\hat{\theta}$ is taken to be the MLE.

In this case, we could consider the statistic:

$$T_n = \frac{\hat{\theta} - \theta_0}{\sigma_0},$$

or if σ_0 is not known we can plug-in an estimate to obtain the statistic,

$$T_n = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_0}.$$

Under the null $T_n \xrightarrow{d} N(0, 1)$, so we simply reject the null if: $|T_n| \geq z_{\alpha/2}$. This controls the Type-I error only asymptotically (i.e. only if $n \rightarrow \infty$) but this is relatively standard in applications.

Example: Suppose we considered the problem of testing the parameter of a Bernoulli, i.e. we observe $X_1, \dots, X_n \sim \text{Ber}(p)$, and the null is that $p = p_0$. Defining $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$. A Wald test could be constructed based on the statistic:

$$T_n = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}},$$

which has an asymptotic $N(0, 1)$ distribution. An alternative would be to use a slightly different estimated standard deviation, i.e. to define,

$$T_n = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}.$$

Observe that this alternative test statistic also has an asymptotically standard normal distribution under the null. Its behaviour under the alternate is a bit more pleasant as we will see.

21.1.1 Power of the Wald Test

To get some idea of what happens under the alternate, suppose we are in some situation where the MLE has “standard asymptotics”, i.e. $\hat{\theta} - \theta \xrightarrow{d} N(0, 1/(nI_1(\theta)))$. Suppose that we use the statistic:

$$T_n = \sqrt{nI_1(\hat{\theta})}(\hat{\theta} - \theta_0),$$

and that the true value of the parameter is $\theta_1 \neq \theta_0$. Let us define:

$$\Delta = \sqrt{nI_1(\theta_1)}(\theta_0 - \theta_1),$$

then the probability that the Wald test rejects the null hypothesis is asymptotically:

$$1 - \Phi(\Delta + z_{\alpha/2}) + \Phi(\Delta - z_{\alpha/2}).$$

You will prove this on your HW (it is some simple re-arrangement, similar to what we have done previously when computing the power function in a Gaussian model). There are some aspects to notice:

1. If the difference between θ_0 and θ_1 is very small the power will tend to α , i.e. if $\Delta \approx 0$ then the test will have trivial power.
2. As $n \rightarrow \infty$ the two Φ terms will approach either 0 or 1, and so the power will approach 1.
3. As a rule of thumb the Wald test will have non-trivial power if $|\theta_0 - \theta_1| \gg \frac{1}{\sqrt{nI_1(\theta_1)}}$.

21.2 Likelihood Ratio Test (LRT)

To test composite versus composite hypotheses the general method is to use something called the (generalized) likelihood ratio test.

We want to test:

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \\ H_1 &: \theta \notin \Theta_0. \end{aligned}$$

This test is simple: reject H_0 if $\lambda(X_1, \dots, X_n) \leq c$ where

$$\lambda(X_1, \dots, X_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

where $\hat{\theta}_0$ maximizes $L(\theta)$ subject to $\theta \in \Theta_0$.

We can simplify the LRT by using an asymptotic approximation. This fact that the LRT generally has a simple asymptotic approximation is known as *Wilks' phenomenon*. First, some notation:

Notation: Let $W \sim \chi_p^2$. Define $\chi_{p,\alpha}^2$ by

$$P(W > \chi_{p,\alpha}^2) = \alpha.$$

We let $\ell(\theta)$ denote the log-likelihood in what follows.

Theorem 21.1 Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ where $\theta \in \mathbb{R}$. Under H_0 ,

$$-2 \log \lambda(X_1, \dots, X_n) \rightsquigarrow \chi_1^2.$$

Hence, if we let $T_n = -2 \log \lambda(X^n)$ then

$$P_{\theta_0}(T_n > \chi_{1,\alpha}^2) \rightarrow \alpha$$

as $n \rightarrow \infty$.

Proof: Using a Taylor expansion:

$$\ell(\theta) \approx \ell(\hat{\theta}) + \ell'(\hat{\theta})(\theta - \hat{\theta}) + \ell''(\hat{\theta}) \frac{(\theta - \hat{\theta})^2}{2} = \ell(\hat{\theta}) + \ell''(\hat{\theta}) \frac{(\theta - \hat{\theta})^2}{2}$$

and so

$$\begin{aligned} -2 \log \lambda(x_1, \dots, x_n) &= 2\ell(\hat{\theta}) - 2\ell(\theta_0) \\ &\approx 2\ell(\hat{\theta}) - 2\ell(\hat{\theta}) - \ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2 = -\ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2 \\ &= \frac{-\frac{1}{n}\ell''(\hat{\theta})}{I_1(\theta_0)}(\sqrt{nI_1(\theta_0)}(\hat{\theta} - \theta_0))^2 = A_n \times B_n. \end{aligned}$$

Now $A_n \xrightarrow{p} 1$ by the WLLN and $\sqrt{B_n} \rightsquigarrow N(0, 1)$. The result follows by Slutsky's theorem. ■

Example 21.2 $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$. We want to test $H_0 : \lambda = \lambda_0$ versus $H_1 : \lambda \neq \lambda_0$. Then

$$-2 \log \lambda(x^n) = 2n[(\lambda_0 - \hat{\lambda}) - \hat{\lambda} \log(\lambda_0/\hat{\lambda})].$$

We reject H_0 when $-2 \log \lambda(x^n) > \chi_{1,\alpha}^2$.

Now suppose that $\theta = (\theta_1, \dots, \theta_k)$. Suppose that $H_0 : \theta \in \Theta_0$ fixes some of the parameters. Then, under conditions,

$$T_n = -2 \log \lambda(X_1, \dots, X_n) \rightsquigarrow \chi_\nu^2$$

where

$$\nu = \dim(\Theta) - \dim(\Theta_0).$$

Therefore, an asymptotic level α test is: reject H_0 when $T_n > \chi_{\nu,\alpha}^2$.

Example 21.3 Consider a multinomial with $\theta = (p_1, \dots, p_5)$. So

$$L(\theta) = p_1^{y_1} \cdots p_5^{y_5}.$$

Suppose we want to test

$$H_0 : p_1 = p_2 = p_3 \quad \text{and} \quad p_4 = p_5$$

versus the alternative that H_0 is false. In this case

$$\nu = 4 - 1 = 3.$$

The LRT test statistic is

$$\lambda(x_1, \dots, x_n) = \frac{\prod_{j=1}^5 \hat{p}_{0j}^{Y_j}}{\prod_{j=1}^5 \hat{p}_j^{Y_j}}$$

where $\hat{p}_j = Y_j/n$, $\hat{p}_{01} = \hat{p}_{02} = \hat{p}_{03} = (Y_1 + Y_2 + Y_3)/n$, $\hat{p}_{04} = \hat{p}_{05} = (1 - 3\hat{p}_{01})/2$. Now we reject H_0 if $-2 \log \lambda(X_1, \dots, X_n) > \chi_{3,\alpha}^2$. □

21.3 p-values

When we test at a given level α we will reject or not reject. It is useful to summarize what levels we would reject at and what levels we would not reject at.

The p-value is the smallest α at which we would reject H_0 .

In other words, we reject at all $\alpha \geq p$. So, if the pvalue is 0.03, then we would reject at $\alpha = 0.05$ but not at $\alpha = 0.01$.

Hence, to test at level α , we reject when $p < \alpha$.

Theorem 21.4 *Suppose we have a test of the form: reject when $T(X_1, \dots, X_n) > c$. Then the p-value is*

$$p = \sup_{\theta \in \Theta_0} P_{\theta}(T_n(X_1, \dots, X_n) \geq T_n(x_1, \dots, x_n))$$

where x_1, \dots, x_n are the observed data and $X_1, \dots, X_n \sim p_{\theta_0}$.

Example 21.5 $X_1, \dots, X_n \sim N(\theta, 1)$. Test that $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. We reject when $|T_n|$ is large, where $T_n = \sqrt{n}(\bar{X}_n - \theta_0)$. Let t_n be the observed value of T_n . Let $Z \sim N(0, 1)$. Then,

$$p = P_{\theta_0} (|\sqrt{n}(\bar{X}_n - \theta_0)| > t_n) = P(|Z| > t_n) = 2\Phi(-|t_n|).$$

The p-value is a random variable. Under some assumptions that you will see in your HW the p-value will be uniformly distributed on $[0, 1]$ under the null.

Important. Note that p is NOT equal to $\mathbb{P}(H_0|X_1, \dots, X_n)$. The latter is a Bayesian quantity which we will discuss later.