LECTURE 10: TESTING IN LARGE SAMPLES

Michal Kolesár*

September 30, 2024

REFERENCE CB, Chapter 10.3, or HMC, Chapter 6.3, 6.5.

Suppose we have a random sample $X_1, \dots, X_n$ from a distribution with density $f(x \mid \theta)$, $\theta \in \mathbb{R}$. We are interested in testing the $H_0 \colon \theta = \theta_0$ against $H_1 \colon \theta \neq \theta_0$. We now develop three tests of these hypotheses that are justified by large-sample arguments: the likelihood ratio (LR) test, the Lagrange multiplier (LM) test, and the Wald test. Furthermore, the three methods are asymptotically equivalent.

1. MAXIMIZED LIKELIHOOD RATIO TEST

We have seen that rejecting for large values of the LR $f(x \mid \theta_1)/f(x \mid \theta_0)$ is a good idea when both the null and alternative are simple. In this case, the alternative is not simple, so perhaps we can replace $f(x \mid \theta_1)$ by its maximized value under the alternative, $\sup_{\theta \in \Theta_1} f(x \mid \theta) = f(x \mid \hat{\theta}_{ML})$. Thus, define the maximized LR statistic,

$$\mathrm{MLR}(X) = \frac{\mathcal{L}(\hat{\theta}_{ML} \mid X)}{\mathcal{L}(\theta_0 \mid X)}.$$

Rejecting for large values of $\mathrm{MLR}(X)$ is equivalent to rejecting for large values of

$$\zeta_{LR} = 2 \log \mathrm{MLR}(X) = 2(\ell_n(\hat{\theta}_{ML}) - \ell_n(\theta_0)). \tag{1}$$

Typically, confusingly, people omit “maximized” and refer to this test as the “likelihood ratio” test. However, unlike the LR test (the use of which is justified by the Neyman-Pearson lemma), the maximized LR test has no finite-sample justification. The test is due to Neyman and Pearson (1928).

The next result allows us to compute the appropriate critical value that will ensure that the test satisfies the level constraint in large samples.

Theorem 1 (HMC, Theorem 6.3.1., CB, Theorem 10.3.1). Under the null hypothesis and regularity conditions similar to those leading to normality of maximum likelihood estimator (MLE), $\zeta_{LR} \Rightarrow \chi^2(1)$.

*Email: mkolesar@princeton.edu.

Proof. Since $\ell_n'(\hat{\theta}_{ML}) = 0$, expanding $\ell_n(\theta_0)$ around $\ell_n(\hat{\theta}_{ML})$ yields, for some $\tilde{\theta}$ between $\hat{\theta}_{ML}$ and θ_0,

$$\ell_n(\theta_0) = \ell_n(\hat{\theta}_{ML}) + \frac{1}{2}\ell_n''(\tilde{\theta})(\theta_0 - \hat{\theta}_{ML})^2.$$

Since $\hat{\theta}_{ML} \xrightarrow{p} \theta_0$ by MLE theory, we have $\tilde{\theta} \xrightarrow{p} \theta_0$. Furthermore, $\ell_n''(\theta_0)/n \xrightarrow{p} -\mathcal{I}(\theta_0)$ by law of large numbers (LLN). This suggests, but does not imply that

$$\ell_n''(\tilde{\theta})/n \xrightarrow{p} -\mathcal{I}(\theta_0).$$

However, one can show using more advanced tools that the above display is nonetheless true. Moreover, by the Delta method and MLE normality,

$$n(\hat{\theta}_{ML} - \theta_0)^2\mathcal{I}(\theta_0) \Rightarrow \chi^2(1).$$

Putting together the previous three displays and applying Slutsky's theorem then yields the result:

$$\zeta_{LR} = 2(\ell_n(\hat{\theta}_{ML}) - \ell_n(\theta_0)) = -\ell_n''(\tilde{\theta})(\theta_0 - \hat{\theta}_{ML})^2 \Rightarrow \mathcal{I}(\theta_0) \cdot \chi^2(1)/\mathcal{I}(\theta_0) = \chi^2(1). \qquad \square$$

It follows from the theorem that the test that rejects when $\zeta_{LR} \geq \chi^{-2}_{1-\alpha}(1) = z^2_{1-\alpha/2}$ (where $\chi^{-2}_{\alpha}(1)$ is the α quantile of a $\chi^2(1)$ distribution and z_α the α quantile of a normal distribution) has size α in large samples, in the sense that the null rejection probability converges to α. In finite samples, the size of this test may be greater than α.

Example 1. Let $X_1, \dots, X_n$ be a random sample from $\mathcal{N}(\theta, \sigma^2)$ distribution, with σ^2 known. Suppose we want to test the null hypothesis, $H_0\colon \theta = \theta_0$ against $H_1\colon \theta \neq \theta_0$. Then

$$\begin{aligned}
\zeta_{LR} &= 2(\ell_n(\overline{X}_n) - \ell_n(\theta_0)) \\
&= -\frac{1}{\sigma^2}\sum_i (X_i - \overline{X}_n)^2 + \frac{1}{\sigma^2}\sum_i (X_i - \theta_0)^2 \\
&= \frac{n}{\sigma^2}\left(\overline{X}_n^2 + \theta_0^2 - 2\theta_0\overline{X}_n\right) = \frac{n}{\sigma^2}(\overline{X}_n - \theta_0)^2.
\end{aligned}$$

Therefore, the LR test rejects if $\frac{n}{\sigma^2}(\overline{X}_n - \theta_0)^2 \geq z^2_{1-\alpha/2}$. Note this is actually the uniformly most powerful (UMP) unbiased test we derived in previous lecture note, and that the test has size α in finite samples. $\boxtimes$

Example 2. Let $X_1, \dots, X_n$ be a random sample from a Poisson(θ) distribution, which has density $\theta^x e^{-\theta}/x!$, for $x = 0, 1, \dots$ and $\theta \geq 0$. The log-likelihood is

$$\ell_n(\theta) = n\overline{X}_n \log(\theta) - n\theta - \log(\textstyle\prod_{i=1}^n X_i!),$$

which implies that $\hat{\theta}_{ML} = \overline{X}_n$, and

$$\begin{aligned}
\zeta_{LR} &= 2\left(n\overline{X}_n \log(\hat{\theta}_{ML}/\theta_0) - n\hat{\theta}_{ML} + n\theta_0\right) \\
&\qquad = 2n\left(\overline{X}_n \log(\overline{X}_n/\theta_0) - (\overline{X}_n - \theta_0)\right) = 2\left(S\log(S/\tilde{\theta}_0) - S + \tilde{\theta}_0\right),
\end{aligned}$$

where θ_0 is the null value, $S = \sum_{i=1}^n X_i$, and $\tilde{\theta}_0 = n\theta_0$. So the test rejects if

$$\log(S) > \log(\tilde{\theta}_0) + \frac{z_{1-\alpha/2}^2/2 - \tilde{\theta}_0}{S}.$$

For $\tilde{\theta}_0 \le z_{1-\alpha/2}^2/2$, the left-hand side (LHS) is increasing, and the right-hand side (RHS) is decreasing in S, so the acceptance region has the form $[0, U(\tilde{\theta}_0)]$. Otherwise, the acceptance region has the form $[L(\tilde{\theta}_0), U(\tilde{\theta}_0)]$ with $L(\tilde{\theta}_0) > 0$. The functions U and L can be solved for numerically.

Since $S \sim \text{Poisson}(\tilde{\theta}_0)$ under the null, we can numerically compute the exact finite-sample size of the test by computing $1 - P(S \in [L(\tilde{\theta}_0), U(\tilde{\theta}_0)])$. Note the size depends on n and θ_0 only through their product, $\tilde{\theta}_0 = n\theta_0$. Figure 1 plots the test size. Can you explain the spike at $\tilde{\theta}_0 = 1.921$ and the erratic behavior of the test size?

As a numerical example, suppose $\theta_0 = 6$, $\overline{X}_n = 5$, and $n = 100$. Then

$$\zeta_{LR} = 200(5\log(5/6) + 1) \approx 17.6.$$

The critical value is 1.96^2. So the test rejects. ☒

2. WALD TEST

In the proof of Theorem 1, we have shown that the LR test equals

$$\zeta_{LR} = \left(-\frac{\ell_n''(\tilde{\theta})}{n}\right) n(\theta_0 - \hat{\theta}_{ML})^2,$$

where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}_{ML}$. We can think of $-\ell_n''(\tilde{\theta})$ as an estimate of the information matrix. Alternatively, estimate the information matrix by evaluating ℓ_n'' at $\hat{\theta}_{ML}$, which yields the Wald (1943) statistic

$$\zeta_{Wald} = \left(-\frac{\ell_n''(\hat{\theta}_{ML})}{n}\right) n(\theta_0 - \hat{\theta}_{ML})^2. \tag{2}$$

This has the advantage that it doesn't involve any computation under the null. Since both $\hat{\theta}_{ML}$ and $\tilde{\theta}$ converge in probability to θ_0 under the null, this substitution doesn't matter, and we have

$$\zeta_{Wald} - \zeta_{LR} \xrightarrow{p}_{H_0} 0,$$

so that $\chi_{1-\alpha}^{-2}(1) = z_{1-\alpha/2}^2$ critical value is again appropriate, and the LR and Wald statistics are asymptotically equivalent (they are different in finite samples though unless the information matrix doesn't depend on θ: when is that the case?).

Remark 2. The Wald statistic, unlike the LR statistic, is not invariant to reparametrization.

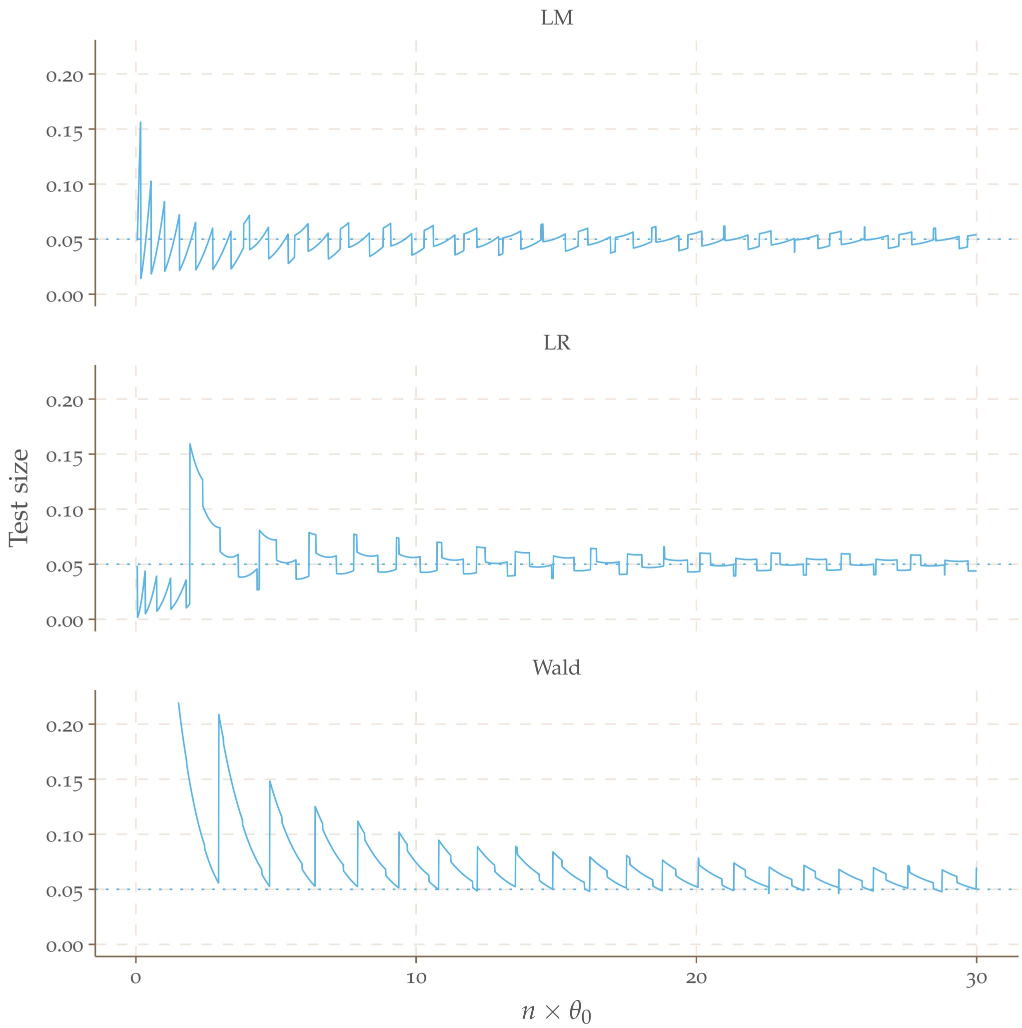

Figure 1: Exact size of the LM, LR, and Wald tests for the Poisson distribution (Example 2).

Another way of motivating this statistic is to note that, under the null

$$\sqrt{n}(\hat{\theta}-\theta_0) \Rightarrow \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1}),$$

and that $(-\frac{\ell_n''(\hat{\theta}_{ML})}{n})^{-1}$ is a consistent estimator of $\mathcal{I}(\theta_0)^{-1}$. More generally, whenever we have an estimator $\hat{\theta}$ such that $\sqrt{n}(\hat{\theta}-\theta) \Rightarrow \mathcal{N}(0, V)$, and a consistent estimator $\hat{V}$ of the asymptotic variance V, then by Slutsky theorem, under the null $\theta = \theta_0$,

$$\sqrt{n}(\hat{\theta}-\theta_0)/\sqrt{\hat{V}} \Rightarrow \mathcal{N}(0, 1),$$

so that an asymptotically level α test rejects the null if

$$\frac{n(\hat{\theta}-\theta_0)^2}{\hat{V}} \geq z_{1-\alpha/2}^2.$$

If the alternative hypothesis is one-sided, $\Theta_1 = \{\theta \colon \theta > \theta_0\}$, then it is better (more powerful) to reject whenever

$$\frac{\sqrt{n}(\hat{\theta}-\theta_0)}{\sqrt{\hat{V}}} \geq z_{1-\alpha}.$$

Example 1 (continued). Here $\zeta_{Wald} = \zeta_{LR}$. ⊠

Example 2 (continued). Differentiating the log-likelihood yields

$$\partial \ell_n(\theta)/\partial\theta = n\overline{X}_n/\theta - n, \qquad \partial^2 \ell_n(\theta)/\partial\theta^2 = -n\overline{X}_n/\theta^2.$$

Thus,

$$\mathcal{I}(\theta) = E\left[-\frac{1}{n}\frac{\partial^2 \ell_n(\theta)}{\partial\theta^2}\right] = \frac{1}{\theta},$$

and the Wald statistic uses the estimator $-\frac{1}{n}\partial^2 \ell_n(\hat{\theta}_{ML})/\partial\theta^2 = 1/\hat{\theta}_{ML}$, so that

$$\zeta_{Wald} = \frac{n(\hat{\theta}_{ML}-\theta_0)^2}{\hat{\theta}_{ML}} = \frac{(S-\tilde{\theta}_0)^2}{S}.$$

The test thus rejects if $(S-\tilde{\theta}_0)^2 > z_{1-\alpha/2}^2 S$, which is a quadratic equation in S, so that the acceptance region has the form $[\tilde{\theta}_0 + z_{1-\alpha/2}^2/2 \pm z_{1-\alpha/2}\sqrt{4\tilde{\theta}_0 + z_{1-\alpha/2}^2}/2]$. Figure 1 plots the test size. The test can be seen to be size-distorted, and the issue is particularly severe for small $\tilde{\theta}_0$. In our numerical example,

$$\zeta_{Wald} = 100 \cdot (5-6)^2 \cdot (1/5) = 20$$

So the test based on the Wald statistic rejects the null hypothesis with an even smaller p-value than the test based on the LR statistic. ⊠

3. SCORE TEST

Recall that the score is defined by

$$S_n(\theta) = \ell_n'(\theta) = \frac{\partial \log \mathcal{L}_n(\theta \mid X)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(X_i \mid \theta)}{\partial \theta}.$$

By the first information equality,

$$E[S_n(\theta_0)] = \sum_{i=1}^{n} E\left[\frac{\partial \log f(X_i|\theta_0)}{\partial \theta}\right] = 0.$$

By definition of Fisher information,

$$E\left[\left(\frac{\partial \log f(X_i|\theta)}{\partial \theta}\right)^2\right] = \mathcal{I}(\theta_0).$$

So, by the Central limit theorem, under the null hypothesis

$$\frac{\frac{1}{\sqrt{n}} S_n(\theta_0)}{\sqrt{\mathcal{I}(\theta_0)}} \Rightarrow N(0,1).$$

Therefore, by the continuous mapping theorem,

$$\zeta_{LM} = \frac{S_n(\theta_0)^2}{n\mathcal{I}(\theta_0)} \Rightarrow \chi^2(1). \tag{3}$$

The test that rejects whenever $\zeta_{LM} > \chi_{1-\alpha}^{-2}(1)$ is known as the Rao (1948)'s score test, or the Lagrange multiplier (LM) test. The latter name is due to Silvey (1959). To see why Silvey called it the LM test, set up the Lagrangian for maximizing the likelihood subject to the null constraint $\theta = \theta_0$,

$$\log \mathcal{L}(\theta \mid x) - \lambda(\theta - \theta_0).$$

(Here the maximization is trivial, but the idea of the LM test extends to cases in which the null is composite, as we discuss below). The first-order condition is

$$S_n(\theta_0) = \lambda.$$

We know that $\hat{\theta}_{ML}$ maximizes $\log \mathcal{L}(\theta|x)$ and, in large samples, $\hat{\theta}_{ML}$ will be close to the true parameter value θ with large probability. If the true value of parameter θ is far from θ_0, then $\hat{\theta}_{ML}$ will be far from θ_0. So, if θ_0 maximizes the Lagrangian, λ will be large. As a result, $S_n(\theta_0)$ and the ζ_{LM} will both be large.

In fact, $\zeta_{LM} - \zeta_{LR} \xrightarrow{p} 0$, so that this statistic is asymptotically equivalent to the Wald and LR statistics. To show this result, note that by the mean value theorem,

$$S_n(\theta_0) = S_n(\theta_0) - S_n(\hat{\theta}_{ML}) = \ell_n''(\bar{\theta})(\theta_0 - \hat{\theta}_{ML}),$$

where $\bar{\theta}$ is between θ_0 and $\hat{\theta}_{ML}$. Thus, $\zeta_{LM} = n(\theta_0 - \hat{\theta}_{ML})^2(\ell_n''(\bar{\theta})/n)^2/\mathcal{I}(\theta_0)$. As argued above, $-(\ell_n''(\bar{\theta})/n) \xrightarrow{p} \mathcal{I}(\theta_0)$. Therefore, by Slutsky's theorem,

$$\zeta_{LM} - \zeta_{LR} = \left(\frac{(\ell_n''(\bar{\theta})/n)^2}{\mathcal{I}(\theta_0)} - \ell_n''(\tilde{\theta})/n \right) n(\theta_0 - \hat{\theta}_{ML})^2 \xrightarrow{p} 0 \cdot \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})^2 = 0.$$

Thus, we have shown that LR, Wald, and LM statistics are all asymptotically equivalent under the null hypothesis. However, they differ in finite samples.

Remark 3. An advantage of the LM statistic is that it only includes calculations based on the restricted estimator θ_0. When is this particularly advantageous?

Example 1 (continued). Here $S_n(\theta_0) = n(\overline{X}_n - \theta_0)/\sigma^2$, and $\mathcal{I}(\theta_0) = \sigma^2$, so that $\zeta_{LM} = n(\overline{X}_n - \theta_0)^2/\sigma^2$, which is exactly equivalent to ζ_{LR} and ζ_{Wald}. ⊠

Example 2 (continued). We have

$$S_n(\theta_0) = \sum_{i=1}^{n} X_i/\theta_0 - n = n(\overline{X}_n/\theta_0 - 1),$$

and $\mathcal{I}(\theta_0) = 1/\theta_0$, so that

$$\zeta_{LM} = \frac{n^2(\overline{X}_n/\theta_0 - 1)^2}{n/\theta_0} = n\theta_0(\overline{X}_n/\theta_0 - 1)^2.$$

So the test rejects if $\tilde{\theta}_0(S/\tilde{\theta}_0 - 1)^2 \geq z^2_{1-\alpha/2}$. The acceptance region for S is therefore again an interval, $[\tilde{\theta}_0 \pm \sqrt{\tilde{\theta}_0} z_{1-\alpha/2}]$, and the exact finite-sample size can be computed analytically—it's plotted in Figure 1. The size is similar to that of the LR test, but without the spike at $\tilde{\theta}_0 = 1.93$—instead, we have a spike close to zero. In the numerical example,

$$\zeta_{LM}(x) = 100 \cdot 6 \cdot (5/6 - 1)^2 = 100/6 \approx 16.67,$$

which gives the biggest p-value among the three tests. ⊠

4. GENERALIZATIONS AND SUMMARY

In general, $\theta \in \Theta \subseteq \mathbb{R}^d$, and we may be interested in testing the null $H_0\colon g(\theta) = 0$, or $H_0\colon \theta \in \Theta_0 = \{\theta \in \Theta\colon g(\theta) = 0\}$. The function $g\colon \mathbb{R}^d \to \mathbb{R}^p$ collects smooth, possibly nonlinear restrictions (these restrictions are independent in the sense that we cannot drop any one g_j without changing Θ_0). The alternative is $H_1\colon \theta \in \Theta_1 = \{\theta \in \Theta\colon g(\theta) \neq 0\}$.

Let $\hat{\theta}_0 = \operatorname{argmax}_{\theta \in \Theta_0} \mathcal{L}(\theta \mid x)$ denote the MLE in the restricted model, and let $\hat{\theta}_{ML} = \operatorname{argmax}_{\theta \in \Theta_1} \mathcal{L}(\theta \mid x) = \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}(\theta \mid x)$ denote the unrestricted estimate.

Then under the null hypothesis, the test statistics given in eqs. (1), (2) and (3) can be generalized to

$$
\begin{aligned}
\zeta_{LR} &= 2(\ell_n(\hat{\theta}_{ML}) - \ell_n(\hat{\theta}_0)) \Rightarrow \chi^2(p) \\
\zeta_{Wald} &= \sqrt{n}(\hat{\theta}_{ML} - \hat{\theta}_0)'[-\nabla^2 \ell_n(\hat{\theta}_{ML})/n]\sqrt{n}(\hat{\theta}_{ML} - \hat{\theta}_0) \Rightarrow \chi^2(p) \\
\zeta_{LM} &= \frac{1}{\sqrt{n}} S_n(\hat{\theta}_0)' \mathcal{I}(\hat{\theta}_0)^{-1} \frac{1}{\sqrt{n}} S_n(\hat{\theta}_0) \Rightarrow \chi^2(p),
\end{aligned}
$$

and all of them are asymptotically equivalent to each other. One can use different estimators of the information matrix in the score or Wald test: any one of $-\nabla^2 \ell_n(\hat{\theta}_{ML})/n$, $\mathcal{I}(\hat{\theta}_0)$, $\mathcal{I}(\hat{\theta}_{ML})$, or $-\nabla^2 \ell_n(\hat{\theta}_0)/n$ will do, depending on what's most convenient. One can also use an estimator based on the outer product of the scores.

The Wald test has yet another form based on the observation that by the Delta method, under the null, $\sqrt{n} g(\hat{\theta}_{ML}) \Rightarrow \mathcal{N}(G\mathcal{I}(\theta)^{-1}G')$, where $G_{ij} = \partial g_i(\theta)/\partial \theta_j$. We can estimate $\hat{G}_{ij} = \partial g_i(\hat{\theta}_{ML})/\partial \theta_j$, which leads to the test statistic

$$
\tilde{\zeta}_{Wald} = \sqrt{n} g(\hat{\theta}_{ML})' \left(\hat{G} \mathcal{I}(\hat{\theta}_{ML})^{-1} \hat{G}' \right)^{-1} \sqrt{n} g(\hat{\theta}_{ML}) \Rightarrow \chi^2(p).
$$

A leading special case is when the restrictions are linear, $g(\theta) = R\theta - r_0$. In this case, the alternative Wald test is particularly easy,

$$
\tilde{\zeta}_{Wald} = \sqrt{n}(R\hat{\theta}_{ML} - r_0)' \left(R \mathcal{I}(\hat{\theta}_{ML})^{-1} R' \right)^{-1} \sqrt{n}(R\hat{\theta}_{ML} - r_0).
$$

A special case of this special case obtains when $\theta = (\beta, \gamma)$, γ is a nuisance parameter of dimension $d - p$, $\beta \in \mathbb{R}^p$, and we're interested in a two-sided test of the hypothesis $\beta = \beta_0$. Then

$$
\tilde{\zeta}_{Wald} = \sqrt{n}(\hat{\beta}_{ML} - \beta_0)' \hat{V}_{\beta\beta}^{-1} \sqrt{n}(\hat{\beta}_{ML} - \beta_0) \Rightarrow \chi^2(p),
$$

where $\hat{V}_{\beta\beta}$ is the upper $p \times p$ block of the $d \times d$ matrix $[-\nabla^2 \ell_n(\hat{\theta}_{ML})/n]^{-1}$. This upper block converges in probability to the upper block of $\mathcal{I}(\theta_0)^{-1}$, which corresponds to the asymptotic variance of $\hat{\beta}_{ML}$, since $\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \Rightarrow_{H_0} \mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$. Note that in general, $\hat{V}_{\beta\beta}^{-1}$ is *not* equal to the upper $p \times p$ block of the matrix $[-\nabla^2 \ell_n(\hat{\theta}_{ML})/n]$.

Remark 4. One can also show that the power against any fixed alternative $\theta_1 \in \Theta_1$ for any of these tests converges to one. When the power function satisfies $\beta(\theta_1) \to 1$, we say that the test is *consistent* against the alternative $H_1 \colon \theta = \theta_1$.

Remark 5. An attractive feature of the Wald test $\tilde{\zeta}_{Wald}$ is that it works whenever we have an asymptotically normal estimator $\hat{\theta}$ of θ with consistently estimable asymptotic variance. The model doesn't even have to be fully parametric.

5. OPTIMALITY

We have derived three tests for restrictions in parametric models that are asymptotically equivalent under the null. But what about optimality? We motivated the Wald statistic by observing that the t-type statistic $\hat{\theta}_{ML} \approx \mathcal{N}(\theta, \mathcal{I}(\hat{\theta}_{ML})^{-1}/n)$, so the problem of testing the null $H_0\colon \theta = \theta_0$ is in large samples analogous to the problem of testing a normal mean. We saw in previous lectures that the UMP unbiased test for this problem is the two-sided z-test. This suggests that the two-sided Wald test should be asymptotically UMP unbiased. This is indeed true, and what's more, the LR and LM variants also share this optimality property.

Just like the large-sample optimality of MLE as a point estimator, properly stating this large-sample optimality of the trinity of tests requires quite a bit of machinery, best left for a second-year course. But loosely speaking, the optimality result means that for values of θ over a shrinking neighborhood of θ_0, (i) the power functions of all three tests are the same, and (ii) no other test that is asymptotically unbiased and controls size in large samples has a power function that's better. This is the real justification for using these tests.

REFERENCES

Neyman, Jerzy, and Egon Sharpe Pearson. 1928. "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I." *Biometrika* 20A, nos. 1/2 (July): 175–240. https://doi.org/10.1093/biomet/20A.1-2.175.

Rao, C. Radhakrishna. 1948. "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation." *Mathematical Proceedings of the Cambridge Philosophical Society* 44, no. 1 (January): 50–57. https://doi.org/10.1017/S0305004100023987.

Silvey, Samuel David. 1959. "The Lagrangian Multiplier Test." *Annals of Mathematical Statistics* 30, no. 2 (June): 389–407. https://doi.org/10.1214/aoms/1177706259.

Wald, Abraham. 1943. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large." *Transactions of the American Mathematical Society* 54, no. 3 (November): 426–482. https://doi.org/10.2307/1990256.