Introduction to Statistics

(Lectures 9-10: Testing)

1 Testing

OVERVIEW: Lectures 9-10 will focus on *testing*. The parameter space in the testing game is partitioned into two subsets: the "null" hypothesis Θ_0 and the "alternative" hypothesis Θ_1. There are two actions $\{0, 1\}$, where an action a is interpreted as "reject the null in favor of the alternative" with probability a. The loss function used for this problem is the "0-1 loss": $\mathcal{L}(0, \theta) = 0 = L(1, \theta')$ for any $\theta \in \Theta_0$, $\theta' \in \Theta_1$ and $\mathcal{L}(a, \theta) = 1$ otherwise. This loss gives rise to the popular *Type I/Type II* error performance criterion.

We first study an idealized environment in which the parameter space has only two components: the null (θ_0) and the alternative (θ_1). We show that in this set-up any reasonable test must maximize "power" subject to a "size" constraint. Moreover, we show that power maximizer tests reject whenever the "likelihood ratio" between the alternative and the null is large enough.

We then analyze some commonly used testing strategies for parametric models with "composite" null or alternative hypothesis: the Likelihood Ratio test, the Wald test, and the Score test. We specialize these trinity of tests to the linear regression model with unknown variance.

1.1 The Testing Problem

Let X be a random variable and let $\{f(x|\theta)\}_{\theta \in \Theta}$ be a statistical model.[1] Partition the parameter space into two subsets Θ_0 and Θ_1.[2]The testing problem starts as follows. In the first stage nature selects a parameter $\theta \in \Theta$. Since the null and the alternative hypothesis partition the parameter space, then either $\theta \in \Theta_0$ or $\theta \in \Theta_1$, but not both.

As usual, the econometrician cannot observe the parameter selected by nature but observes data. Based on the data, the statistician would like to decide whether the parameter selected by nature belongs to Θ_0 (the null) or to Θ_1 (the alternative).

In the simplest set-up, the statistician can take only two actions: $a \in \{0, 1\}$. Action a is interpreted as "reject the null in favor of the alternative" with probability a. Hence, if the econometrician picks $a = 1$, he/she will be rejecting the null in favor of the alternative (and thus, saying that $\theta \in \Theta_1$). If, however, $a = 0$ we the econometrician does not reject the null (which we interpret as saying that $\theta \in \Theta_0$).

A *decision rule/algorithm/strategy* for the statistician in the hypothesis testing problem is a mapping

$$\phi : X \to \{0, 1\}$$

[1]Throughout this section, we will work with statistical models in which $f(x|\theta)$ is a p.d.f.

[2]By partition, we mean $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 = \Theta$.

These decision rules are called "tests". The collection of all data sets for which the econometrician rejects the null hypothesis

$$\{x \mid \phi(x) = 1\}$$

is called the critical region of the test ϕ.

1.2 0-1 Loss and Type I/Type II error

The payoff/loss for the econometrician depends on the action taken (1 or 0) and the true state of the world (θ). The following $\{0, 1\}$-loss structure is typical in testing problems:

a/s	$\theta \in \Theta_0$	$\theta \in \Theta_1$
1	1	0
0	0	1

This is intended to model a payoff structure in which the econometrician is only punished when taking the "wrong" decision. The punishments are symmetric, but it is easy to extend this framework.

The expected loss (risk) is computed as follows. When $\theta \in \Theta_0$ the econometrician only experiences a loss if he/she (incorrectly) rejects the null hypothesis. Thus, the expected loss equals the probability of (incorrectly!) rejecting the null hypothesis ($\phi(x) = 1$) when the null hypothesis is true. For any $\theta \in \Theta_0$:

$$\mathbb{E}_{f(x|\theta)}[\mathcal{L}(\phi(x), \theta)] = \mathbb{E}_{f(x|\theta)}[\mathbf{1}\{\phi(x) = 1\}] = P_\theta(\phi(x) = 1). \tag{1.1}$$

This is referred to as the *rate of Type I error* of the test $\phi(x)$ at $\theta \in \Theta_0$. The *largest* rate of Type I error of a test; i.e.,

$$\sup_{\theta \in \Theta_0} P_\theta(\phi(x) = 1)$$

is called the *size* of the test.

When $\theta \in \Theta_1$ the econometrician only experiences a loss if he/she (incorrectly) fails to reject the null hypothesis. Thus, the expected loss equals the probability of (incorrectly!) failing to reject the null hypothesis ($\phi(x) = 0$) when the null hypothesis is not true. For any $\theta \in \Theta_1$:

$$\mathbb{E}_{f(x|\theta)}[\mathcal{L}(\phi(x), \theta)] = \mathbb{E}_{f(x|\theta)}[\mathbf{1}\{\phi(x) = 0\}] = P_\theta(\phi(x) = 0). \tag{1.2}$$

This is the *rate of Type II error* of the test $\phi(x)$ at θ. It is also common to make reference to the *power* of a test at $\theta \in \Theta_1$. The power is defined as the probability of rejecting the null when it is not true; thus, it equals one minus the rate of Type II error; i.e.,

$$1 - P_\theta(\phi(x) = 0) = P_\theta(\phi(x) = 1).$$

1.3 Testing a "simple" null against a "simple" alternative

Suppose that both Θ_0 and Θ_1 are singletons, so that the testing problem becomes

$$\mathbf{H}_0 : \theta = \theta_0 \text{ vs. } \mathbf{H}_1 : \theta = \theta_1.$$

In this problem it is relatively straightforward to characterize the "optimal" test. Define a randomized test as a map

$$\phi : X \to [0, 1]$$

where $\phi(x)$ is interpreted as the probability of "rejecting the null hypothesis" after observing x. See Ferguson (1967), p. 198, 199 (also, think about what we did in the last problem of problem set 3!). The rates of Type I/Type II error of a randomized test ϕ are given by:

$$R(\phi, \theta_0) = \mathbb{E}_{f(x|\theta_0)}[\phi(x)], \quad R(\phi, \theta_1) = 1 - \mathbb{E}_{f(x|\theta_1)}[\phi(x)].$$

The following proposition shows that under a mild assumption, any admissible test for this problem maximizes power subject to a "size control" constraint.

Proposition 1. *Suppose that for any set $A \subseteq \mathbf{X}$*

$$\int_A f(x, \theta_0)dx > 0 \implies \int_A f(x, \theta_1)dx > 0.$$

A randomized test ϕ is admissible if and only if there exists $\alpha \in [0, 1]$ such that ϕ maximizes power subject to having size at most α; that is

$$\phi \in \arg\max_\phi (1 - R(\phi, \theta_1)) \tag{1.3}$$

s.t.

$$R(\phi, \theta_0) \leq \alpha \tag{1.4}$$

Proof. "⇒" First we show that if ϕ is admissible, then ϕ solves (1.3) subject to (1.4) for some α. We show that contrapositive. Let α be the rate of Type I error of ϕ. Suppose ϕ does not solve the con-

strained optimization problem for any such value of α. Then, there is another test ϕ' (namely, the solution to the constrained optimization problem for $\alpha = R(\phi, \theta_0)$) such that $R(\phi, \theta_1) > R(\phi', \theta_1)$ and $R(\phi, \theta_0) \geq R(\phi', \theta_0)$. Hence, ϕ is not admissible.

"$\Leftarrow$" We do the proof by contradiction. Suppose that ϕ solves (1.3) subject to (1.4) for some α, but is not admissible. Then, there exists ϕ' such that $R(\phi, \theta_1) \geq R(\phi', \theta_1)$ and $R(\phi, \theta_0) \geq R(\phi', \theta_0)$ with strict inequality somewhere. We have three cases to consider:

1. If $R(\phi, \theta_0) = R(\phi', \theta_0)$ and $R(\phi, \theta_1) > R(\phi', \theta_1)$. This contradicts the fact that ϕ solved (1.3) subject to (1.4).

2. Suppose $1 > \alpha = R(\phi, \theta_0) > R(\phi', \theta_0)$ and $R(\phi, \theta_1) \geq R(\phi', \theta_1)$. Consider the test ϕ'' that rejects for every value of x. Such test has Type I and II error given by $(1, 0)$. Consider the test

$$\phi'''(x) = \lambda\phi''(x) + (1 - \lambda)\phi'(x), \quad \lambda \in [0, 1]$$

   This randomized test has Type I error $\lambda + (1 - \lambda)R(\phi', \theta_0)$. Since $\alpha \in (R(\phi', \theta_0), 1)$, there exists $\lambda(\alpha) > 0$ such that:

$$(R(\phi''', \theta_0), R(\phi''', \theta_1)) = (\alpha, (1 - \lambda(\alpha))R(\phi'(x), \theta_1)).$$

   This contradicts the fact that ϕ solved (1.3) subject to (1.4).

3. Finally, suppose that $1 = R(\phi, \theta_0) > R(\phi', \theta_0)$ and $R(\phi, \theta_1) \geq R(\phi', \theta_1)$. We have already seen that the test ϕ'' which rejects for every value of x has $R(\phi'', \theta_0) = 1$ and $R(\phi'', \theta_1) = 0$. Since ϕ solves the minimisation problem for $\alpha = 1$, it must be the case that $R(\phi, \theta_1) \leq R(\phi'', \theta_1) = 0$. Then, by assumption $0 \leq R(\phi', \theta_1) \leq R(\phi, \theta_1) \leq 0$. Therefore, $R(\phi', \theta_1) = R(\phi, \theta_1) = 0$. Since $R(\phi', \theta_0) < 1$, this implies

$$\int_{\{x \in \mathbf{X} | \phi'(x) = 0\}} f(x, \theta_0) dx > 0$$

   which implies

$$\int_{\{x \in \mathbf{X} | \phi'(x) = 0\}} f(x, \theta_1) dx > 0$$

which implies $R(\phi', \theta_1) > 0$. Contradiction.

□

The implication of this proposition is important: if you want to use an admissible test for the simple null against simple alternative then you have no choice but to maximize power subject to a size control constraint. The following proposition tells us how to solve this optimization problem: reject the null hypothesis if the likelihood of the data under the alternative is sufficiently higher than the likelihood under the null.

Proposition 2. *(The Neyman-Pearson Lemma) Consider the hypothesis testing problem of a simple null θ_0 against a simple alternative θ_1. Define the* likelihood ratio statistic *as*

$$L(x) \equiv \frac{f(x, \theta_1)}{f(x, \theta_0)}$$

and suppose that for each $\alpha \in (0, 1)$ there exists $c_{LR}(\alpha)$ such that

$$P_{\theta_0}(L(X) > c_{LR}(\alpha)) = \alpha$$

Then the test:

$$\phi(x) = 1_{L(x) > c_{LR}(\alpha)} \tag{1.5}$$

solves:

$$\phi \in \arg\max_{\phi}(1 - R(\phi, \theta_1))$$

subject to

$$R(\phi, \theta_0) \leq \alpha$$

Proof. Let $\phi(x)$ be defined as in (1.5). We would like to show the following: if $t(x)$ is another test of level α, then $P_{\theta_1}(\phi(x) = 1) \geq P_{\theta_1}(t(x) = 1) \geq 0$. The prove proceed goes follows. By definition of ϕ

$$\phi(x) = 1 \qquad \text{if} \quad f(x, \theta_1) > c_{LR}(\alpha) f(x, \theta_0)$$

Since $t(x) \in [0, 1]$ it follows that

$$\phi(x) - t(x) \geq 0 \qquad \text{if} \quad f(x, \theta_1) > c_{LR}(\alpha) f(x, \theta_0)$$

and

$$\phi(x) - t(x) \leq 0 \qquad \text{if} \quad f(x, \theta_1) \leq c_{LR}(\alpha) f(x, \theta_0)$$

Therefore, the function:

$$[\phi(x) - t(x)]\,[f(x,\theta_1) - c_{LR}(\alpha) f(x,\theta_0)] \geq 0$$

Note then that, for all x

$$[\phi(x) - t(x)]\, f(x,\theta_1) \quad \geq \quad [\phi(x) - t(x)]\, c_{LR}(\alpha) f(x,\theta_0)$$

Therefore, integrating with respect to x:

$$\int_{\mathbf{X}} [\phi(x) - t(x)]\, f(x,\theta_1) dx \quad \geq \quad \int_{\mathbf{X}} [\phi(x) - t(x)]\, \lambda f(x,\theta_0) dx$$

Note, that the right hand side of the previous equation is exactly the same as the difference in levels (scaled by λ) of the tests ϕ and t. Since the Neyman-Pearson test has level α and any other competing test has level at most α, then the right-hand side is larger than or equal to zero. Re-arranging the expression we get:

$$\int_{\mathbf{X}} \phi(x) f(x,\theta_1) \geq \int_{\mathbf{X}} t(x) f(x,\theta_1)$$

And this completes the proof (make sure you understand why these integrals equal the power of the tests) □

1.4 Testing hypotheses in Parametric Models

1.4.1 Generalized Likelihood Ratio Test

We have just shown that the test that rejects the null hypothesis whenever the likelihood under the alternative is (sufficiently) larger than the likelihood under the null cannot be dominated. To establish this result we assumed that both the null and the alternative were "simple". Unfortunately, most problems we will encounter in econometric practice involve "composite" hypotheses.

EXAMPLE: Consider the linear regression model with non-stochastic regressors:

$$Y \sim \mathcal{N}_n(X\beta, \sigma^2 \mathbb{I}_n). \tag{1.6}$$

Assume (as we did in the last problem set) that both $\beta \in \mathbb{R}^k$ and $\sigma^2 \in \mathbb{R}_+$ are unknown. Suppose that we are interested in the problem

$$\mathbf{H}_0 : \beta = \beta_0 \text{ vs. } \mathbf{H}_1 : \beta \neq \beta_0. \tag{1.7}$$

Both the null and the alternative are composite. To see this, write

$$\Theta_0 \equiv \{(\beta, \sigma^2) \mid \beta = \beta_0\}, \quad \Theta_1 \equiv \{(\beta, \sigma^2) \mid \beta \neq \beta_0\}.$$

The value of σ^2 is not specified under the null hypothesis, hence any tuple (β_0, σ^2) is plausible under the null. Parameters that are not specified under the null hypothesis are usually called *nuisance* parameters.

The *generalized likelihood ratio* statistic provides a general approach for testing hypothesis in parametric models with composite null and alternative hypotheses. The test rejects Θ_0 in favor of the alternative whenever the *likelihood ratio statistic*

$$2\Big[\max_{\theta \in \Theta_1} \ln f(x|\theta) - \max_{\theta \in \Theta_0} \ln f(x|\theta)\Big] \tag{1.8}$$

is large enough. The first term denotes the largest value that the "log-likelihood" can achieve under the alternative, and the second term denotes the largest value that the log-likelihood can achieve under the null. Let $\widehat{\theta}_1$ and $\widehat{\theta}_0$ denote the values of the parameter that maximize the likelihood under the alternative and the null, respectively. The generalized likelihood ratio test thus rejects whenever

$$2 \ln \Big(f(x|\widehat{\theta}_1) / f(x, \widehat{\theta}_0) \Big)$$

is large enough.

Likelihood Ratio Test for the Linear Regression Model: We now compute the generalized likelihood ratio statistic for the testing problem (1.7) in the linear regression model. First, we maximize the likelihood under the alternative hypothesis $\beta \neq \beta_0$. This is the same as just maximizing the likelihood over the whole parameter space

$$\max_{\beta, \sigma^2} f(Y|\beta, \sigma^2).$$

We solved in the previous problem set and we showed that:

$$\widehat{\beta}_{\text{ML}} = (X'X)^{-1}X'Y, \quad \widehat{\sigma}^2_{\text{ML}} = (Y - X\widehat{\beta}_{\text{ML}})'(Y - X\widehat{\beta}_{\text{ML}})/n.$$

Thus, we can verify that the largest value of the log-likelihood under the alternative is

$$\log f(Y|\widehat{\beta}_{\mathrm{ML}}, \widehat{\sigma}^2_{\mathrm{ML}}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2_{\mathrm{ML}}) - \frac{n}{2}.$$

We now turn to the problem of maximizing the log-likelihood under the null hypothesis:

$$\max_{\sigma^2} f(Y|\beta_0, \sigma^2).$$

Algebra shows that the maximizer is

$$\widehat{\sigma}^2_0 = (Y - X\beta_0)'(Y - X\beta_0)/n,$$

and the maximized value of the log-likelihood under the alternative corresponds to

$$\log f(Y|\widehat{\beta}_0, \widehat{\sigma}^2_0) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widehat{\sigma}^2_0) - \frac{n}{2}.$$

This means that the likelihood ratio test statistic equals

$$n\left(\ln(\widehat{\sigma}^2_0) - \ln(\widehat{\sigma}^2_{\mathrm{ML}})\right). \tag{1.9}$$

The *critical value*—the threshold against which the likelihood ratio statistic is compared against—is usually taken as the $1-\alpha$ quantile of χ^2 distribution with degrees of freedom given by the dimension of β_0. We now explain this approximation for the linear regression model and then present an heuristic derivation of the more general approximation result.

Algebra shows that we can write the likelihood ratio test statistic above as

$$n\left(\ln\left((\widehat{\sigma}^2_0 - \widehat{\sigma}^2_{\mathrm{ML}})/\widehat{\sigma}^2_{\mathrm{ML}} + 1\right)\right).$$

Under the null hypothesis we expect $\widehat{\sigma}^2_0$ and $\widehat{\sigma}^2_{\mathrm{ML}}$ not to be very different. Since a standard first-order Taylor appproximation suggests that $\ln(1+x) \approx x$ when x is small, the likelihood ratio test statistic is expected to be approximately equal to

$$n(\widehat{\sigma}^2_0 - \widehat{\sigma}^2_{\mathrm{ML}})/\widehat{\sigma}^2_{\mathrm{ML}}.$$

Interestingly, in the linear regression model

$$n(\widehat{\sigma}^2_0 - \widehat{\sigma}^2_{\mathrm{ML}}) = (\widehat{\beta}_{\mathrm{ML}} - \beta_0)(X'X)(\widehat{\beta}_{\mathrm{ML}} - \beta_0) = \sigma^2\frac{1}{\sigma^2}(\widehat{\beta}_{\mathrm{ML}} - \beta_0)(X'X)(\widehat{\beta}_{\mathrm{ML}} - \beta_0).$$

Moreover, if the null hypothesis is true

$$(\widehat{\beta}_{\mathrm{ML}} - \beta_0) = (X'X)^{-1}X'(Y - X\beta_0) \sim \mathcal{N}_k(0, \sigma^2(X'X)^{-1}),$$

implying

$$\frac{1}{\sigma^2}(\widehat{\beta}_{\mathrm{ML}} - \beta_0)'(X'X)(\widehat{\beta}_{\mathrm{ML}} - \beta_0)$$

has the distribution of a χ^2 random variable with k degrees of freedom. If we further assume that when the sample size is large $\sigma^2/\widehat{\sigma}^2_{\mathrm{ML}} = 1$, with probability close to one then:

$$n(\widehat{\sigma}^2_0 - \widehat{\sigma}^2_{\mathrm{ML}})/\widehat{\sigma}^2_{\mathrm{ML}} \approx \chi^2_k.$$

1.4.2 The Wald Test

The approximation result above is more general. Consider the problem

$$\mathbf{H}_0 : \theta = \theta_0 \text{ vs. } \mathbf{H}_1 : \theta \neq \theta_0.$$

The maximized log-likelihood ratio under the alternative is

$$\ln f(x|\widehat{\theta}_{\mathrm{ML}}).$$

Under the null, $\widehat{\theta}_{\mathrm{ML}}$ should be close to θ_0. Thus, an heuristic first-order Taylor approximation of $\ln f(x|\theta_0)$ around $\widehat{\theta}_{\mathrm{ML}}$ suggests that

$$\begin{aligned} \ln f(x|\theta_0) &\approx \ln f(x|\widehat{\theta}_{\mathrm{ML}}) + \left(\frac{\partial}{\partial\theta}\ln f(x|\widehat{\theta}_{\mathrm{ML}})\right)'(\theta_0 - \widehat{\theta}_{\mathrm{ML}}) \\ &+ \frac{1}{2}(\widehat{\theta}_{\mathrm{ML}} - \theta_0)'\left(\frac{\partial^2}{\partial\theta\partial\theta}\ln f(x|\widehat{\theta}_{\mathrm{ML}})\right)(\widehat{\theta}_{\mathrm{ML}} - \theta_0). \end{aligned}$$

The term

$$\left(\frac{\partial}{\partial\theta}\ln f(x|\widehat{\theta}_{\mathrm{ML}})\right)$$

is the score (the derivative of the log-likelihood) evaluated at $\widehat{\theta}_{\mathrm{ML}}$. The F.O.C. defining $\widehat{\theta}_{\mathrm{ML}}$ imply

this term equals zero. Therefore, the likelihood ratio statistic is approximately equal to

$$2(\ln f(x|\widehat{\theta}_{\mathrm{ML}}) - \ln f(x|\theta_0)) \approx (\widehat{\theta}_{\mathrm{ML}} - \theta_0)' \underbrace{\left(-\frac{\partial^2}{\partial\theta\partial\theta} \ln f(x|\widehat{\theta}_{\mathrm{ML}})\right)}_{\widehat{\mathcal{I}}\text{: Observed Information Matrix}} (\widehat{\theta}_{\mathrm{ML}} - \theta_0).$$

The last term in the equation above is called the Wald test statistic for the null hypothesis $\theta = \theta_0$. Asymptotic theory (which you will cover in the next block of the course) will show that under very general conditions

$$\widehat{\mathcal{I}}^{1/2}(\widehat{\theta}_{\mathrm{ML}} - \theta_0) \sim \mathcal{N}_{\dim(\theta)}(0, \mathbb{I}_{\dim(\theta)}).$$

Therefore, the Wald test statistic has approximately the same distribution as a χ^2 random variable with $\dim(\theta)$ degrees of freedom. This means that if we let $\chi^2_{\dim(\theta),1-\alpha}$ denote the $1-\alpha$ quantile of a $\chi^2_{\dim(\theta)}$ random variable then the test that rejects whenever

$$(\widehat{\theta}_{\mathrm{ML}} - \theta_0)'\widehat{\mathcal{I}}(\widehat{\theta}_{\mathrm{ML}} - \theta_0) > \chi^2_{\dim(\theta),1-\alpha}$$

will have size of approximately $1-\alpha$. Moreover, the outcome of this Wald test will be approximately the same as that of the Likelihood Ratio Test.

Wald Test for the Linear Regression Model: We derive the Wald test for the problem

$$\mathbf{H}_0 : \beta = \beta_0 \text{ vs. } \mathbf{H}_1 : \beta \neq \beta_0,$$

using the Linear Regression Model with unknown variance. Note that σ^2 is a nuisance parameter, as it is not specified under the null hypothesis. To construct the Wald test statistic we will need to replace σ^2 by its ML estimator.

We showed in the previous problem set that the score is given by

$$\begin{pmatrix} \frac{1}{\sigma^2}X'(Y - X\beta) \\ -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{\sigma^4}(Y - X\beta)'(Y - X\beta) \end{pmatrix}.$$

The upper $k \times k$ block of the sample information matrix is then given by the matrix

$$\frac{1}{\sigma^2}X'X,$$

which does not depend on β, but depends σ^2. The Wald statistic is thus

$$\frac{1}{\widehat{\sigma}^2_{\text{ML}}}(\widehat{\beta}-\beta_0)'(X'X)(\widehat{\beta}-\beta_0).$$

1.4.3 The Score Test

Finally, we introduce the score test. Once again, consider the problem

$$\mathbf{H}_0 : \theta = \theta_0 \text{ vs. } \mathbf{H}_1 : \theta \neq \theta_0.$$

In any parametric model, the ML estimator satisfies the first order condition

$$\frac{\partial}{\partial\theta}\ln f(x|\widehat{\theta}_{\text{ML}}) = 0.$$

An heuristic first-order Taylor approximation thus suggests

$$\underbrace{\frac{\partial}{\partial\theta}\ln f(x|\theta_0)}_{S(x;\theta_0)\text{: Score at } \theta_0} \approx \underbrace{\frac{\partial}{\partial\theta}\ln f(x|\widehat{\theta}_{\text{ML}})}_{S(x;\theta_0)\text{: F.O.C.}} + \underbrace{-\frac{\partial^2}{\partial\theta\partial\theta}\ln f(x|\theta_0)}_{\widehat{I}\text{: observed information}}\left(\widehat{\theta}_{\text{ML}}-\theta_0\right),$$

implying

$$\left(-\frac{\partial^2}{\partial\theta\partial\theta}\ln f(x|\theta_0)\right)^{1/2}\left(\widehat{\theta}_{\text{ML}}-\theta_0\right) \approx \left(-\frac{\partial^2}{\partial\theta\partial\theta}\ln f(x|\theta_0)\right)^{-1/2} S(x;\theta_0).$$

If it is true that $\widehat{\mathcal{I}}^{1/2}(\widehat{\theta}_{\text{ML}}-\theta_0) \sim \mathcal{N}_{\dim(\theta)}(0,\mathbb{I}_{\dim(\theta)})$ (an approximation we have used to motivate the Wald Test and to derive the critical value for the Likelihood Ratio Test) then we must have

$$S(x;\theta_0)'\left(-\frac{\partial^2}{\partial\theta\partial\theta}\ln f(x|\theta_0)\right)^{-1} S(x;\theta_0) \approx \chi^2_{\dim(\theta)}$$

The score test (with nominal size α) rejects if the *score test statistic above* is larger than the $1-\alpha$ quantile of a $\chi^2_{\dim(\theta)}$. Note that the score does not require computing the ML estimator, only the F.O.C.

References

Berger, J. (1985): *Statistical decision theory and Bayesian analysis*, Springer.

Ferguson, T. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, vol. 7, Academic Press New York.