LECTURE 13 NOTES

1. Other large sample tests. We observe that the Wald test is based on the fact that $\hat{\theta}_n$ is an asymptotically normal estimator of θ. In some sources, the term Wald test is an umbrella term for tests based on asymptotically normal estimators. If

1. $\hat{\theta}_n$ is an asymptotically normal estimator of θ,
2. $\widehat{V}_n$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_n$,

by Slutsky's theorem,

$$\sqrt{n}\widehat{V}_n^{-\frac{1}{2}}\big(\hat{\theta}_n - \theta\big) \xrightarrow{d} \mathcal{N}\big(0, I_p\big). \tag{1.1}$$

Thus replacing θ by θ_0 and comparing $\sqrt{n}V_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ to the standard normal distribution is the basis of a test of $H_0 : \theta = \theta_0$.

EXAMPLE 1.1. *Let* $\mathbf{x}_i \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$. *Consider testing* $H_0 : p = p_0$ *versus* $H_1 : p \neq p_0$. *The MLE of* p *is* $\hat{p} := \bar{\mathbf{x}}$. *By the CLT,*

$$\sqrt{n}\big(\hat{p}_n - p\big) \xrightarrow{d} \mathcal{N}\big(0, p(1-p)\big),$$

Since $\hat{p}_n$ *is a consistent estimator of* p, $\big(\hat{p}_n(1-\hat{p}_n)\big)$ *is a consistent estimator of the asymptotic variance. Thus*

$$\frac{\sqrt{n}\big(\hat{p}_n - p\big)}{\sqrt{(\hat{p}(1-\hat{p})}} \xrightarrow{d} \mathcal{N}(0,1).$$

The Wald test statistic replaces p *by* p_0*:*

$$\mathbf{w}_n = \frac{\sqrt{n}\big(\hat{p}_n - p_0\big)}{\sqrt{(\hat{p}(1-\hat{p})}},$$

and the test rejects H_0 *if* $\mathbf{w}_n$ *is an "extreme" realization of a* $\mathcal{N}(0,1)$ *random variable.*

We observe that the Rao test is based on the fact that $\frac{1}{\sqrt{n}}\nabla\ell_n(\theta^*)$ is asymptotically normal. More generally, tests that are based on the score are called *score tests*. Under $H_0 : \theta = \theta_0$, the score at θ_0 should be asymptotically normal:

$$\tfrac{1}{\sqrt{n}}\nabla\ell_n(\theta_0) \xrightarrow{d} \mathcal{N}\big(0, I(\theta_0)\big).$$

A score test compares $\frac{1}{\sqrt{n}}\nabla\ell_n(\theta_0)$ to the $\mathcal{N}\big(0, I(\theta_0)\big)$ distribution.

Example 1.2 (Example 1.1 continued). *The score is*

$$\frac{1}{\sqrt{n}}\nabla\ell_n(p) = \tfrac{1}{\sqrt{n}}\textstyle\sum_{i\in[n]}\mathbf{x}_i\big(\tfrac{1}{p}+\tfrac{1}{1-p}\big) - \tfrac{\sqrt{n}}{1-p}.$$

By the CLT, it is asymptotically normal:

$$\tfrac{1}{\sqrt{n}}\nabla\ell_n(p) \xrightarrow{d} \mathcal{N}\big(0, p(1-p)\big).$$

The score test rejects $H_0 : p = p_0$ *if* $\frac{1}{\sqrt{n}}\nabla\ell_n(p_0)$ *is an "extreme" realization of a* $\mathcal{N}\big(0, p_0(1-p_0)\big)$ *random variable.*

2. Most powerful tests. In the classical approach to hypothesis testing, the investigator only considers tests of level (at most) α. Thus the evaluation of tests focuses on minimizing Type II errors, or equivalently, maximizing power. Thus the optimal test is *uniformly most powerful* (UMP); i.e. its power function is uniformly greater than that of any other α-level test on Θ_1.

We begin by considering testing simple nulls against simple alternatives. That is, testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, where $\Theta_0, \Theta_1 \in \mathbf{R}^p$ are singletons. The Neyman-Pearson theorem show that a test based on the likelihood ratio is most powerful for testing simple hypotheses.

Theorem 2.1 (Neyman-Pearson). *Consider testing* $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta = \theta_1$. *The* α*-level test* $\varphi_{\mathrm{NP}}(\mathbf{x}) := \mathbf{1}_{(t,\infty)}(\lambda(\mathbf{x}))$, *where*

$$\lambda(\mathbf{x}) = \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})},$$

is most powerful among all α*-level tests. Further, any other most powerful* α*-level test is almost everywhere equal to* $\varphi_{\mathrm{NP}}(\mathbf{x})$.

Proof. The most powerful test is the solution to

$$\begin{aligned} &\underset{\varphi:\mathcal{X}\to\{0,1\}}{\text{maximize}} && \mathbf{E}_1\big[\varphi(\mathbf{x})\big] \\ &\text{subject to} && \mathbf{E}_0\big[\varphi(\mathbf{x})\big] = \alpha : t. \end{aligned} \tag{2.1}$$

where t is a Lagrange multiplier. The Lagrangian is

$$\begin{aligned}
L(\varphi, t) &= \mathbf{E}_1\big[\varphi(\mathbf{x})\big] - t\big(\mathbf{E}_0\big[\varphi(\mathbf{x})\big] - \alpha\big) \\
&= \int_{\mathcal{X}} \varphi(x)(f_1(x) - tf_0(x))dx - t\alpha \\
&= \int_{\mathcal{X}} \varphi(x)\,|f_1(x) - tf_0(x)|\,\mathbf{1}\Big\{\frac{f_1(x)}{f_0(x)} > t\Big\}dx \\
&\quad - \int_{\mathcal{X}} \varphi(x)\,|f_1(x) - tf_0(x)|\,\mathbf{1}\Big\{\frac{f_1(x)}{f_0(x)} < t\Big\}dx - t\alpha.
\end{aligned}$$

It is clear that

$$\varphi^*(x) = \mathbf{1}\{f_1(x) > \nu f_0(x)\}$$

maximizes the Lagrangian, which has the form of a LRT. Further, any exact α-level test that maximizes the Lagrangian is also a maximizer of (2.1):

$$\begin{aligned}
\mathbf{E}_1\big[\varphi(\mathbf{x})\big] &= \mathbf{E}_1\big[\varphi(\mathbf{x})\big] - t\big(\mathbf{E}_0\big[\varphi(\mathbf{x})\big] - \alpha\big) \\
&\le \mathbf{E}_1\big[\varphi^*(\mathbf{x})\big] - t\big(\mathbf{E}_0\big[\varphi^*(\mathbf{x})\big] - \alpha\big) \\
&= \mathbf{E}_1\big[\varphi^*(\mathbf{x})\big].
\end{aligned}$$

Thus the α-level LRT is most powerful for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

To complete the proof, we show that any most powerful α-level test is essentially equal to φ_{NP}. Let φ_{NP} be the α-level Neyman-Pearson test and φ' be any other α-level test. It is possible to check that

$$(\varphi_{\mathrm{NP}}(x) - \varphi'(x))(tf_0(x) - f_1(x)) \le 0.$$

Indeed,

1. when $\varphi_{\mathrm{NP}}(x) = 1$, $\varphi_{\mathrm{NP}}(x) - \varphi'(x) \ge 0$ and $tf_0(x) - f_1(x) \le 0$.
2. when $\varphi_{\mathrm{NP}}(x) = 0$, $\varphi_{\mathrm{NP}}(x) - \varphi'(x) \le 0$ and $tf_0(x) - f_1(x) \ge 0$.

We rearrange and integrate over $\mathcal{X}$ to obtain

$$t\int_{\mathcal{X}} (\varphi_{\mathrm{NP}}(x) - \varphi'(x))f_0(x) \le \int_{\mathcal{X}} (\varphi_{\mathrm{NP}}(x) - \varphi'(x))f_1(x)dx,$$

or equivalently,

$$t(\beta_{\mathrm{NP}}(\theta_0) - \beta'(\theta_0)) \le \beta_{\mathrm{NP}}(\theta_1) - \beta'(\theta_1),$$

where β_{NP} and β' are the power functions of φ_{NP} and φ'. Since φ' is a α-level test and most powerful,

$$t(\beta(\theta_0) - \beta'(\theta_0)) - \beta(\theta_1) - \beta'(\theta_1) = 0.$$

The non-positive integrand $(\varphi(x)-\varphi'(x))(tf_0(x)-f_1(x))$ only integrates to zero if it vanishes almost everywhere:

$$(\varphi(x)-\varphi'(x))(tf_0(x)-f_1(x))=0,$$

which is akin to $\varphi'(x)=1$ for any $\frac{f_1(x)}{f_0(x)}>t$ and $\varphi'(x)=0$ for any $\frac{f_1(x)}{f_0(x)}<t$ except on a set of measure zero. □

2.1. *Uniformly most powerful tests.* The Neyman-Pearson Theorem is a complete solution to the problem of testing simple nulls against simple alternatives: the Neyman-Pearson test is most powerful. When either the null or the alternative is composite, the Neyman-Pearson test is sometimes uniformly most powerful (UMP).

EXAMPLE 2.2. *Let $\mathbf{x}\sim\mathcal{N}(\mu,1)$. Consider testing $H_0:\mu=0$ versus $H_1:\mu>0$. We pick $\mu_1>0$ and consider the α-level Neyman-Pearson test for testing H_0 versus $H_1:\mu=\mu_1$. The Neyman-Pearson statistic is*

$$\lambda(\mathbf{x})=\frac{(2\pi)^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu_1)^2\right)}{(2\pi)^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^2\right)}=\exp\left(\mu_1\mathbf{x}-\frac{\mu_1^2}{2}\right).$$

Since $\mu_1>0$, $\lambda(\mathbf{x})>t$ is equivalent to

$$\mathbf{x}>t'=\tfrac{1}{\mu_1}\big(\log t+\tfrac{\mu_1}{2}\big).$$

Thus the Neyman Pearson test rejects H_0 if $\mathbf{x}$ exceeds some threshold. Under H_0, $\mathbf{x}\sim\mathcal{N}(0,1)$, so the critical region is (z_α,∞). We observe that as long as $\mu_1>0$, the critical region does not depend on μ_1.

By the Neyman-Pearson theorem, the preceding test is the most powerful α-level tests for testing H_0 versus $H_1:\mu=\mu_1$. Since the test does not depend on the choice of μ_1, it is UMP (among α-level tests).

Situations like Example 2.2, where the LRT does not depend on the alternative, occur more generally when the parametric model has *monotone likelihood ratios*, i.e. when $\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}$, for any $\theta_1,\theta_2\in\Theta$, is a monotone function of x.

THEOREM 2.3 (Karlin-Rubin). *Consider testing $H_0:\theta\le\theta_0$ versus $H_1:\theta>\theta_0$. If the likelihood ratio $\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$ is non-decreasing in x for any $\theta_1>\theta_0$, the test that rejects H_0 when $\mathbf{x}>t$, for some t such that $\alpha:=\mathbf{P}_{\theta_0}(\mathbf{x}>t)$, is a UMP α-level test.*

PROOF. The proof consists of two parts:

1. show that $\mathbf{1}_{(t,\infty)}(\mathbf{x})$ is UMP for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$.
2. show that its power function is non-decreasing in θ.

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ for some $\theta_1 > \theta_0$. The α-level Neyman Pearson test rejects H_0 when $\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > t$. By assumption, $\frac{f_1(x)}{f_0(x)}$ is non-decreasing in x by assumption. Thus the critical region is necessarily (t, ∞), where t is the $1 - \alpha$ quantile of $\mathbf{x}$ under H_0.

We know the α-level Neyman Pearson test is the most powerful α-level test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Since the critical region of the test does not depend on θ_1, it is UMP for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$, which establishes (1).

To complete the proof, we must show that the power function of the preceding test is non-decreasing in θ, which establishes the preceding test is a α-level test for the composite null $H_0 : \theta \le \theta_0$. For any $\theta_0' \le \theta_0$

$$c := \inf\Big\{x \in \mathbf{R} : \tfrac{f_0(x)}{f_0'(x)} \ge 0\Big\} = \inf\{x \in \mathbf{R} : f_0(x) \ge f_0'(x)\}.$$

If $t \ge c$, by integrating, we obtain $\int_{(t,\infty)} f_0(x)dx \ge \int_{(t,\infty)} f_0'(x)dx$, which is $\beta(\theta_0) \ge \beta(\theta_0')$. If $t < c$, we integrate to obtain

$$1 - \int_{-\infty}^{t} f_0(x)dx \ge 1 - \int_{-\infty}^{t} f_0'(x)dx = \beta(\theta_0'),$$

which again is $\beta(\theta_0) \ge \beta(\theta_0')$. □

However, try as we might, a UMP test simply does not exist for testing some pairs of nulls and alternative hypotheses. The typical example is testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \ne \mu_0$.

EXAMPLE 2.4. *Let $\mathbf{x} \sim \mathcal{N}(\mu, 1)$. Consider testing $H_0 : \mu = 0$ versus $H_1 : \mu \ne 0$. In Example 2.2, we showed that the UMP test for testing H_0 versus $H_1 : \mu = \mu_1 > 0$ is the α-level Neyman-Pearson test:*

$$\varphi_{\mathrm{NP}}(\mathbf{x}) = \mathbf{1}_{(t,\infty)}(\lambda(\mathbf{x})),$$

where $\lambda(x) = \frac{f_{\mu_1}(x)}{f_{\mu_0}(x)}$. By the Neyman-Pearson theorem, any test that has as high a power as φ_{NP} for testing H_0 versus $H_1 : \mu = \mu_1 > 0$ is a.e. equal to φ_{NP}. Thus, if there is a UMP α-level test for testing H_0 versus $H_1 : \mu \ne 0$, it is φ_{NP}.

Consider the test $\phi'(\mathbf{x}) = \mathbf{1}_{(-\infty, z_\alpha)}(\mathbf{x})$*. It is a* α*-level test, and for any* $\mu_1 < \mu_0$*,*

$$\begin{aligned}\beta'(\mu_1) &= \mathbf{P}_1(\mathbf{x} < z_\alpha) = \mathbf{P}(\mathbf{z} < z_\alpha - \mu_1) \\ &> \mathbf{P}(\mathbf{z} < z_\alpha) \\ &= \mathbf{P}(\mathbf{z} > z_\alpha) = \mathbf{P}_1(\mathbf{x} > z_\alpha + \mu_1) \\ &> \mathbf{P}_1(\mathbf{x} > z_\alpha) = \beta_{\mathrm{NP}}(\mu_1).\end{aligned}$$

Thus φ_{NP} *is not UMP. By Theorem 2.1, if there is a UMP, it is a.e. equal to* φ_{NP}*. Therefore there is no UMP.*

2.2. *Uniformly most powerful unbiased tests.* There is no UMP because the class of all α-level tests is so large that no one test is uniformly most powerful on Θ_1. One way to make the notion of a most powerful test tractable is to consider only *unbiased tests*.

Definition 2.5. *A test is unbiased if* $\inf_{\theta\in\Theta_1} \beta(\theta) \geq \sup_{\theta\in\Theta_0} \beta(\theta)$*.*

Intuitively, unbiasedness requires the power function to be larger under any alternative than it is under any null. Thus a α-level unbiased test has power at least α at any $\theta_1 \in \Theta_1$.

Example 2.6 (Example 2.4 continued). *Although there is no UMP test for testing* $H_0 : \mu = 0$ *versus* $H_1 : \mu \neq 0$*, as we shall see, the LRT is a uniformly most powerful unbiased (UMPU) test. That is, its power function is uniformly greater than that of any other unbiased* α*-level test.*

We proceed by maximizing the power function subject to the constraints that (i) the test is α*-level, (ii) the test is unbiased. Pick* $\mu_1 \neq \mu_0$*. We wish to optimize*

$$\begin{aligned} &\underset{\varphi:\mathbf{R}\to\{0,1\}}{\text{maximize}} && \mathbf{E}_1\big[\varphi(\mathbf{x})\big] \\ &\textit{subject to} && \mathbf{E}_0\big[\varphi(\mathbf{x})\big] = \alpha : t_1 \\ & && \partial_\mu \mathbf{E}_0\big[\varphi(\mathbf{x})\big] = \int_{\mathbf{R}} \varphi(x)\partial_\mu f_{\mu_0}(x)dx = 0 : t_2. \end{aligned} \tag{2.2}$$

The second constraint enforces unbiasedness. Since the null is simple, unbiasedness is equivalent to $\mu_0 \in \arg\min_{\mu\in\mathbf{R}} \beta(\mu)$*. Thus a necessary condition for unbiasedness is* $\partial_\mu\beta(\mu_0)$ *vanishes.*

The Lagrangian is

$$\begin{aligned} L(\varphi,t) &= \mathbf{E}_1\big[\varphi(\mathbf{x})\big] - t_1\big(\mathbf{E}_0\big[\varphi(\mathbf{x})\big] - \alpha\big) - t_2\partial_\mu \mathbf{E}_0\big[\varphi(\mathbf{x})\big] \\ &= \int_{\mathcal{X}} \varphi(x)\big(f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2\partial_\mu f_{\mu_0}(x)\big)dx - t\alpha \\ &= \int_{\mathcal{X}} \varphi(x)\,|f_1(x) - tf_0(x)|\,\mathbf{1}\big\{f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2\partial_\mu f_{\mu_0}(x) > 0\big\}dx \\ &\quad - \int_{\mathcal{X}} \varphi(x)\,|f_1(x) - tf_0(x)|\,\mathbf{1}\big\{f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2\partial_\mu f_{\mu_0}(x) \le 0\big\}dx. \end{aligned}$$

Thus the UMPU test has a critical region defined by

$$f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2\partial_\mu f_{\mu_0}(x) > 0,$$

which, by rearranging, is equivalent to

$$\exp\big(x(\mu_1 - \mu_0) - \tfrac{1}{2}(\mu_1^2 - \mu_0^2)\big) > t_1 + t_2(x - \mu_0).$$

The critical region is a set on which an exponential function of x exceeds a linear function of x. Such a set has three possible forms: a union of two intervals $(-\infty, t_1') \cup (t_2', \infty)$, $(-\infty, t_1')$, or (t_2', ∞).

Since the Gaussian density is an even function (in μ), its gradient is odd. Thus, by the unbiasedness constraint

$$\int_{\mathbf{R}} \varphi(x)\partial_\mu f_{\mu_0}(x)dx = 0,$$

the critical region is necessarily symmetric around μ_0, which, together with the aforementioned constraints on the form of the critical region, leads to a critical region of the form $(-\infty, t') \cup (t', \infty)$. We deduce the test that rejects if $|\mathbf{x} - \mu_0| > z_{\alpha/2}$ is the most powerful unbiased α-level test for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu = \mu_1$.

We recognize the preceding critical region as that of the α-level LRT for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Since the critical region does not depend on μ_1, the α-level LRT is UMPU for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Figure 1 plots the power function of the equal-tailed test: it is at least α on $\mathbf{R} \setminus \{\mu_0\}$.

Yuekai Sun
Berkeley, California
December 2, 2015

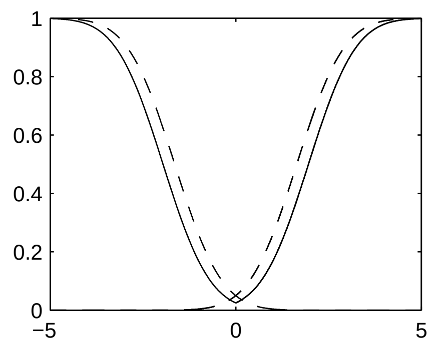

Fig 1: The power functions of the UMPU test for testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ (solid curve) and the UMP tests for testing $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ and $H_0 : \mu \geq 0$ versus $H_1 : \mu < 0$ (dotted curves).