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 $\theta$. 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 $\theta$,
2. $\hat{V}_n$ is a consistent estimator of the asymptotic variance of $\hat{\theta}_n$,

by Slutsky’s theorem,

\begin{equation}
\frac{1}{\sqrt{n}} \hat{V}_n^{-1/2} (\hat{\theta}_n - \theta) \overset{d}{\to} N(0, I_p).
\end{equation}

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

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

$$\sqrt{n}(\hat{p}_n - p) \overset{d}{\to} N(0, p(1 - p)),$$

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

$$\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{\hat{p}(1 - \hat{p})}} \overset{d}{\to} N(0, 1).$$

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

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

and the test rejects $H_0$ if $w_n$ is an “extreme” realization of a $N(0, 1)$ 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 $\theta_0$ should be asymptotically normal:

$$\frac{1}{\sqrt{n}} \nabla \ell_n(\theta_0) \overset{d}{\to} N(0, I(\theta_0)).$$

A score test compares $\frac{1}{\sqrt{n}} \nabla \ell_n(\theta_0)$ to the $N(0, I(\theta_0))$ distribution.
Example 1.2 (Example 1.1 continued). The score is
\[
\frac{1}{\sqrt{n}} \nabla \ell_n(p) = \frac{1}{\sqrt{n}} \sum_{i \in [n]} x_i \left( \frac{1}{p} + \frac{1}{1-p} \right) - \frac{\sqrt{n}}{1-p}.
\]
By the CLT, it is asymptotically normal:
\[
\frac{1}{\sqrt{n}} \nabla \ell_n(p) \xrightarrow{d} \mathcal{N}(0, p(1-p)).
\]
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 $\mathcal{N}(0, p_0(1-p_0))$ random variable.

2. Most powerful tests. In the classical approach to hypothesis testing, the investigator only considers tests of level (at most) $\alpha$. 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 $\alpha$-level test on $\Theta_1$.

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 $\Theta_0, \Theta_1 \in \mathbb{R}^p$ 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 $\alpha$-level test $\varphi_{NP}(x) := 1_{(t, \infty)}(\lambda(x))$, where
\[
\lambda(x) = \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)},
\]
is most powerful among all $\alpha$-level tests. Further, any other most powerful $\alpha$-level test is almost everywhere equal to $\varphi_{NP}(x)$.

Proof. The most powerful test is the solution to
\[
\begin{align*}
\text{maximize} & \quad \mathbb{E}_1[\varphi(x)] \\
\text{subject to} & \quad \mathbb{E}_0[\varphi(x)] = \alpha : t.
\end{align*}
\]
where \( t \) is a Lagrange multiplier. The Lagrangian is

\[
L(\varphi, t) = E_1[\varphi(x)] - t(E_0[\varphi(x)] - \alpha)
\]

\[
= \int_\mathcal{X} \varphi(x) (f_1(x) - t f_0(x)) dx - t \alpha
\]

\[
= \int_\mathcal{X} \varphi(x) |f_1(x) - t f_0(x)| \left\{ \frac{f_1(x)}{f_0(x)} > t \right\} dx
\]

\[
- \int_\mathcal{X} \varphi(x) |f_1(x) - t f_0(x)| \left\{ \frac{f_1(x)}{f_0(x)} < t \right\} dx - t \alpha.
\]

It is clear that

\[
\varphi^*(x) = 1 \{ f_1(x) > \nu f_0(x) \}
\]

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

\[
E_1[\varphi(x)] = E_1[\varphi(x)] - t(E_0[\varphi(x)] - \alpha)
\]

\[
\leq E_1[\varphi^*(x)] - t(E_0[\varphi^*(x)] - \alpha)
\]

\[
= E_1[\varphi^*(x)].
\]

Thus the \( \alpha \)-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 \( \alpha \)-level test is essentially equal to \( \varphi_{NP} \). Let \( \varphi_{NP} \) be the \( \alpha \)-level Neyman-Pearson test and \( \varphi' \) be any other \( \alpha \)-level test. It is possible to check that

\[
(\varphi_{NP}(x) - \varphi'(x))(t f_0(x) - f_1(x)) \leq 0.
\]

Indeed,

1. when \( \varphi_{NP}(x) = 1 \), \( \varphi_{NP}(x) - \varphi'(x) \geq 0 \) and \( t f_0(x) - f_1(x) \leq 0 \).
2. when \( \varphi_{NP}(x) = 0 \), \( \varphi_{NP}(x) - \varphi'(x) \leq 0 \) and \( t f_0(x) - f_1(x) \geq 0 \).

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

\[
t \int_{\mathcal{X}} (\varphi_{NP}(x) - \varphi'(x)) f_0(x) dx \leq \int_{\mathcal{X}} (\varphi_{NP}(x) - \varphi'(x)) f_1(x) dx,
\]

or equivalently,

\[
t(\beta_{NP}(\theta_0) - \beta'(\theta_0)) \leq \beta_{NP}(\theta_1) - \beta'(\theta_1),
\]

where \( \beta_{NP} \) and \( \beta' \) are the power functions of \( \varphi_{NP} \) and \( \varphi' \). Since \( \varphi' \) is a \( \alpha \)-level test and most powerful,

\[
t(\beta(\theta_0) - \beta'(\theta_0)) - \beta(\theta_1) - \beta'(\theta_1) = 0.
\]
The non-positive integrand \((\varphi(x) - \varphi'(x))(t f_0(x) - f_1(x))\) only integrates to zero if it vanishes almost everywhere:

\[
(\varphi(x) - \varphi'(x))(t f_0(x) - f_1(x)) = 0,
\]

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. 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 \(x \sim N(\mu, 1)\). Consider testing \(H_0 : \mu = 0\) versus \(H_1 : \mu > 0\). We pick \(\mu_1 > 0\) and consider the \(\alpha\)-level Neyman-Pearson test for testing \(H_0\) versus \(H_1 : \mu = \mu_1\). The Neyman-Pearson statistic is

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

Since \(\mu_1 > 0\), \(\lambda(x) > t\) is equivalent to

\[
x > t' = \frac{1}{\mu_1} (\log t + \frac{\mu_1^2}{2}).
\]

Thus the Neyman-Pearson test rejects \(H_0\) if \(x\) exceeds some threshold. Under \(H_0\), \(x \sim N(0, 1)\), so the critical region is \((z_\alpha, \infty)\). We observe that as long as \(\mu_1 > 0\), the critical region does not depend on \(\mu_1\).

By the Neyman-Pearson theorem, the preceding test is the most powerful \(\alpha\)-level tests for testing \(H_0\) versus \(H_1 : \mu = \mu_1\). Since the test does not depend on the choice of \(\mu_1\), it is UMP (among \(\alpha\)-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 \(H_0 : \theta \leq \theta_0\) versus \(H_1 : \theta > \theta_0\). If the likelihood ratio \(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\) is non-decreasing in \(x\) for any \(\theta_1 > \theta_0\), the test that rejects \(H_0\) when \(x > t\), for some \(t\) such that \(\alpha := P_{\theta_0}(x > t)\), is a UMP \(\alpha\)-level test.
PROOF. The proof consists of two parts:

1. show that \( 1_{(t, \infty)}(x) \) 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 \( \theta \).

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

We know the \( \alpha \)-level Neyman Pearson test is the most powerful \( \alpha \)-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 \( \theta_1 \), 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 \( \theta \), which establishes the preceding test is a \( \alpha \)-level test for the composite null \( H_0 : \theta \leq \theta_0 \). For any \( \theta' \leq \theta_0 \)

\[
c := \inf \left\{ x \in \mathbb{R} : \frac{f_0(x)}{f_0'(x)} \geq 0 \right\} = \inf \{ x \in \mathbb{R} : f_0(x) \geq f_0'(x) \}.
\]

If \( t \geq c \), by integrating, we obtain \( \int_{(t, \infty)} f_0(x)dx \geq \int_{(t, \infty)} f_0'(x)dx \), which is \( \beta(\theta_0) \geq \beta(\theta'_0) \). If \( t < c \), we integrate to obtain

\[
1 - \int_{-\infty}^{t} f_0(x)dx \geq 1 - \int_{-\infty}^{t} f_0'(x)dx = \beta(\theta'_0),
\]

which again is \( \beta(\theta_0) \geq \beta(\theta'_0) \).

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 \( H_1 : \mu \neq \mu_0 \).

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

\[
\varphi_{\text{NP}}(x) = 1_{(t, \infty)}(\lambda(x)),
\]

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 \( \varphi_{\text{NP}} \) for testing \( H_0 \) versus \( H_1 : \mu = \mu_1 > 0 \) is a.e. equal to \( \varphi_{\text{NP}} \). Thus, if there is a UMP \( \alpha \)-level test for testing \( H_0 \) versus \( H_1 : \mu \neq 0 \), it is \( \varphi_{\text{NP}} \).
Consider the test \( \phi'(x) = 1_{(-\infty, z_\alpha)}(x) \). It is a \( \alpha \)-level test, and for any \( \mu_1 < \mu_0 \),

\[
\beta'(\mu_1) = P_1(x < z_\alpha) = P(z < z_\alpha - \mu_1) > P(z < z_\alpha)
\]

\[
= P(z > z_\alpha) = P_1(x > z_\alpha + \mu_1) > P_1(x > z_\alpha) = \beta_{NP}(\mu_1).
\]

Thus \( \phi_{NP} \) is not UMP. By Theorem 2.1, if there is a UMP, it is a.e. equal to \( \phi_{NP} \). Therefore there is no UMP.

2.2. Uniformly most powerful unbiased tests. There is no UMP because the class of all \( \alpha \)-level tests is so large that no one test is uniformly most powerful on \( \Theta_1 \). 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

\[
\inf_{\theta \in \Theta} \beta(\theta) \geq \sup_{\theta \in \Theta_0} \beta(\theta).
\]

Intuitively, unbiasedness requires the power function to be larger under any alternative than it is under any null. Thus a \( \alpha \)-level unbiased test has power at least \( \alpha \) 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 \( \alpha \)-level test.

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

\[
\text{maximize}_{\varphi : \mathbb{R} \to \{0,1\}} \ E_1 [\varphi(x)]
\]

subject to

\[
\text{subject to } E_0 [\varphi(x)] = \alpha : t_1
\]

\[
\partial_\mu E_0 [\varphi(x)] = \int_{\mathbb{R}} \varphi(x) \partial_\mu f_{\mu_0}(x) dx = 0 : t_2.
\]

The second constraint enforces unbiasedness. Since the null is simple, unbiasedness is equivalent to \( \mu_0 \in \arg \min_{\mu \in \mathbb{R}} \beta(\mu) \). Thus a necessary condition for unbiasedness is \( \partial_\mu \beta(\mu_0) \) vanishes.
The Lagrangian is
\[
L(\varphi, t) = \mathcal{E}_1[\varphi(x)] - t_1(\mathcal{E}_0[\varphi(x)] - \alpha) - t_2 \partial_{\mu} \mathcal{E}_0[\varphi(x)]
\]
\[
= \int_X \varphi(x) (f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2 \partial_{\mu} f_{\mu_0}(x)) dx - t\alpha
\]
\[
= \int_X \varphi(x) |f_1(x) - tf_0(x)| \mathbf{1}\{f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2 \partial_{\mu} f_{\mu_0}(x) > 0\} dx
\]
\[
- \int_X \varphi(x) |f_1(x) - tf_0(x)| \mathbf{1}\{f_{\mu_1}(x) - t_1 f_{\mu_0}(x) - t_2 \partial_{\mu} f_{\mu_0}(x) \leq 0\} dx.
\]
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
\[
\exp(x(\mu_1 - \mu_0) - \frac{1}{2}(\mu_1^2 - \mu_0^2)) > t_1 + t_2(x - \mu_0).
\]
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 \((t'_{2}, \infty)\).

Since the Gaussian density is an even function (in \(\mu\)), its gradient is odd. Thus, by the unbiasedness constraint
\[
\int_{\mathbb{R}} \varphi(x) \partial_{\mu} f_{\mu_0}(x) dx = 0,
\]
the critical region is necessarily symmetric around \(\mu_0\), which, together with the aforementioned constraints on the form of the critical region, leads to a critical region of the form \((-\infty, t'_{1}) \cup (t'_{2}, \infty)\). We deduce the test that rejects if \(|x - \mu_0| > z_{\alpha/2}\) is the most powerful unbiased \(\alpha\)-level test for testing \(H_0: \mu = \mu_0\) versus \(H_1: \mu \neq \mu_0\).

We recognize the preceding critical region as that of the \(\alpha\)-level LRT for testing \(H_0: \mu = \mu_0\) versus \(H_1: \mu \neq \mu_0\). Since the critical region does not depend on \(\mu_1\), the \(\alpha\)-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 \(\alpha\) on \(\mathbb{R} \setminus \{\mu_0\}\).
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).