

LECTURE 14 NOTES

1. Asymptotic power of tests.

DEFINITION 1.1. A sequence of α -level tests $\{\varphi_n(\mathbf{x})\}$ is consistent if

$$\beta_n(\theta) := \mathbf{E}_\theta[\varphi_n(\mathbf{x})] \rightarrow 1 \text{ as } n \rightarrow \infty,$$

for any $\theta \in \Theta_1$.

Just like consistency of a sequence of estimators, Definition 1.1 is a basic notion of “correctness”. In fact, most tests are consistent. In the rest of the section, we refrain from presenting mathematically rigorous results because the level of the subject is such that it is difficult to state the even the assumptions without introducing additional technical concepts.

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ by a Wald test. That is, consider the test

$$\varphi(\mathbf{x}) = \mathbf{1}_{(\chi_{p,\alpha}^2, \infty)}(\|\mathbf{w}_n\|_2^2),$$

where \mathbf{w}_n is the Wald test statistic:

$$(1.1) \quad \mathbf{w}_n := \sqrt{n\widehat{V}_n}^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)$$

($\hat{\theta}_n$ is an asymptotically normal estimator of θ , and \widehat{V}_n is a consistent estimator of its asymptotic variance). The power is

$$\beta(\theta) = \mathbf{P}_\theta(\|\mathbf{w}_n\|_2^2 \geq \chi_{p,\alpha}^2)$$

At any $\theta_1 \neq \theta_0$, the Wald statistic diverges. Indeed,

$$\sqrt{n\widehat{V}_n}^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0) = \sqrt{n\widehat{V}_n}^{-\frac{1}{2}}(\hat{\theta}_n - \theta_1) + \sqrt{n\widehat{V}_n}^{-\frac{1}{2}}(\theta_1 - \theta_0).$$

We recognize the first term is $O_P(1)$ ($\hat{\theta}_n$ is asymptotically normal), but the second term diverges. Thus the power tends to one. It is possible to show that the LR and score tests are consistent by similar arguments.

Consistency ensures the power of a test grows to one as the sample size grows. However, the rate of convergence is unclear. When we encountered a similar problem when evaluating point estimators, we “blew up” the error by \sqrt{n} and studied the limiting distribution of

$$\sqrt{n}(\hat{\theta}_n - \theta^*).$$

The analogous trick here is to study the limiting distribution of the Wald statistic under a sequence of *local alternatives*:

$$(1.2) \quad \theta_n := \theta_0 + \frac{h}{\sqrt{n}}.$$

Formally, consider a triangular array

$$\begin{array}{ccccccc} \mathbf{x}_{1,1} & & & & & & \\ \mathbf{x}_{2,1} & \mathbf{x}_{2,2} & & & & & \\ \vdots & \vdots & \ddots & & & & \\ \mathbf{x}_{n,1} & \mathbf{x}_{n,2} & \dots & \mathbf{x}_{n,n} & & & \\ \vdots & \vdots & & & \ddots & & \end{array},$$

where $\mathbf{x}_{n,i} \stackrel{i.i.d.}{\sim} F_{\theta_n}$. We remark that observations in different rows of the array are not identically distributed. Let $\hat{\theta}_n$ be an asymptotically normal estimator of θ_n based on observations $\{\mathbf{x}_{n,i}\}_{i \in [n]}$. The Wald statistic is

$$\sqrt{n}\widehat{V}_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0) = \sqrt{n}\widehat{V}_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_n) + \widehat{V}_n^{-\frac{1}{2}}h.$$

Intuitively, the first term converges in distribution to a $\mathcal{N}(0, I_p)$ random variable, and the second term converges to $\text{Avar}(\hat{\theta}_n)^{-\frac{1}{2}}h$. Thus

$$(1.3) \quad \sqrt{n}\widehat{V}_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(\text{Avar}(\hat{\theta}_n)^{-\frac{1}{2}}h, I_p).$$

and the power function converges to

$$\beta(\theta_n) \rightarrow \mathbf{P}\left(\|\mathbf{z} + \text{Avar}(\hat{\theta}_n)^{-\frac{1}{2}}h\|_2^2 \geq \chi_{p,\alpha}^2\right),$$

where $\mathbf{z} \sim \mathcal{N}(0, I_p)$. The preceding limit of the power function is called the *asymptotic power* of the Wald test. Evidently, the larger $\widehat{V}_n^{-\frac{1}{2}}h$ is, the higher is the asymptotic power. Thus Wald tests based on efficient estimators are more powerful. There is a similar story for the LR and score tests.

We remark that $\|\mathbf{z} + \mu\|_2^2$ is distributed as a *non-central* χ^2 random variable: $\|\mu\|_2^2$ is the non-centrality parameter.

EXAMPLE 1.2. Let $\mathbf{x}_i \stackrel{i.i.d.}{\sim} \text{Ber}(p)$. We wish to test $H_0 : p = p_0$ by a Wald test. We know

1. the MLE of p is $\hat{p}_n = \bar{\mathbf{x}}_n$,
2. the asymptotic variance of \hat{p}_n is $p(1-p)$.

If $p \neq p_0$, the power of the Wald test is approximately

$$\begin{aligned}\beta(p) &= \mathbf{P}_p \left(\left(\frac{\sqrt{n}(\hat{p}_n - p_0)}{\sqrt{p(1-p)}} \right)^2 \geq \chi_{1,\alpha}^2 \right) \\ &= \mathbf{P}_p \left(\left(\frac{\sqrt{n}(\hat{p}_n - p)}{\sqrt{p(1-p)}} + \frac{\sqrt{n}(p - p_0)}{\sqrt{p(1-p)}} \right)^2 \geq \chi_{1,\alpha}^2 \right) \\ &\approx \mathbf{P} \left(\left(\mathbf{z} + \frac{\sqrt{n}(p - p_0)}{\sqrt{p(1-p)}} \right)^2 \geq \chi_{1,\alpha}^2 \right),\end{aligned}$$

where \mathbf{z} is a standard normal random variable. The non-centrality parameter is $\frac{n(p-p_0)^2}{p(1-p)}$.

To illustrate use of the preceding approximate power function in a concrete setting, consider the design question: how many samples are required to achieve 0.9 power against the alternative $H_1 : p_1 = p_0 + 0.1$? By the properties of the non-central χ_1^2 distribution, to ensure

$$\mathbf{P}((\mathbf{z} + \mu)^2 \geq \chi_{1,\alpha}^2) \geq 0.90,$$

the non-centrality parameter μ^2 must be at least 10.51. Recall the non-centrality parameter is $\frac{n(p_1 - p_0)^2}{p_1(1-p_1)}$. We solve for n to deduce

$$n > \frac{10.51 p_1 (1 - p_1)}{(p_1 - p_0)^2} = 262.65.$$

It is possible to rigorously justify (1.1) under suitable conditions by appealing to the theory of *local asymptotic normality*, much of which was developed by Lucien Le Cam at Berkeley.

2. Interval estimation. In the first part of the course, we considered the task of point estimation, where the goal is to provide a single point that is a guess for the value of the unknown parameter. The goal of interval estimation is to provide a set that contains the unknown parameter with some prescribed probability.

DEFINITION 2.1. Let $\mathcal{C}(\mathbf{x}) \subset \Theta$ be a set-valued random variable. It is a $1 - \alpha$ -confidence set for a parameter θ if

$$\mathbf{P}_\theta(\theta \in \mathcal{C}(\mathbf{x})) \geq 1 - \alpha.$$

If $\mathcal{C}(\mathbf{x})$ is an interval on \mathbf{R} , we call $\mathcal{C}(\mathbf{x})$ a confidence interval.

We emphasize that the set $\mathcal{C}(\mathbf{x})$ not the parameter θ is the random quantity in Definition 2.1. Observing $[l(\mathbf{x}), u(\mathbf{x})] = [l, u]$ should not be interpreted as “ $\theta \in [l, u]$ with probability at least $1 - \alpha$ ”: θ is a deterministic

quantity, so it is nonsense to consider the probability of $\theta \in [l, u]$. A correct statement is “ $\theta \in [l(\mathbf{x}), u(\mathbf{x})]$ with probability at least $1 - \alpha$ ”.

As we shall see, forming interval estimators essentially boil down to “inverting” hypothesis tests.

LEMMA 2.2. *Let $\mathcal{A}(\theta_0) \subset \mathcal{X}$ be the acceptance region of a α -level test of $H_0 : \theta = \theta_0$. The set $\mathcal{A}^{-1}(\mathbf{x})$, where*

$$\mathcal{A}^{-1}(x) := \{\theta \in \Theta : x \in \mathcal{A}(\theta)\}$$

is a $1 - \alpha$ confidence set for θ_0 . That is, the set of parameters θ_0 at which an α -level test of $H_0 : \theta = \theta_0$ accepts is a $1 - \alpha$ -confidence set for θ_0 .

PROOF. By definition of $\mathcal{A}^{-1}(x)$, the event $\{\mathbf{x} \in \mathcal{A}(\theta_0)\}$ is equivalent to $\{\theta_0 \in \mathcal{A}^{-1}(\mathbf{x})\}$. Under H_0 , an α -level test accepts with probability at least $1 - \alpha$:

$$\mathbf{P}_{\theta_0}(\mathbf{x} \in \mathcal{A}(\theta_0)) \geq \alpha.$$

Thus $\mathcal{A}^{-1}(x)$ is a $1 - \alpha$ confidence set:

$$\mathbf{P}_{\theta_0}(\theta_0 \in \mathcal{A}^{-1}(\mathbf{x})) = \mathbf{P}_{\theta_0}(\mathbf{x} \in \mathcal{A}(\theta_0)) \geq 1 - \alpha.$$

□

Lemma 2.2 is essentially a tautology: $\mathcal{A}^{-1}(\mathbf{x})$ is a $1 - \alpha$ -confidence set because the event $\{\theta \in \mathcal{A}^{-1}(\mathbf{x})\}$ is equivalent to $\{\mathbf{x} \in \mathcal{A}(\theta)\}$. Since $\mathcal{A}(\theta)$ is the acceptance region of an α -level test, $\mathbf{P}_{\theta_0}(\mathbf{x} \in \mathcal{A}(\theta_0)) \geq 1 - \alpha$.

EXAMPLE 2.3. *Let $\mathbf{x} \sim \mathcal{N}(\mu, 1)$. Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. We showed that the α -level LRT of $H_0 : \mu = \mu_0$ rejects if*

$$|\mathbf{x} - \mu_0| > z_{\frac{\alpha}{2}}.$$

The acceptance region $\mathcal{A}(\mu_0)$ is $\{x \in \mathbf{R} : |x - \mu_0| \leq z_{\frac{\alpha}{2}}\}$. Thus

$$\begin{aligned} \mathcal{A}^{-1}(\mathbf{x}) &= \{\mu_0 \in \mathbf{R} : |\mathbf{x} - \mu_0| \leq z_{\frac{\alpha}{2}}\} \\ &= [\mathbf{x} - z_{\frac{\alpha}{2}}, \mathbf{x} + z_{\frac{\alpha}{2}}] \end{aligned}$$

is a $1 - \alpha$ -confidence interval for μ_0 .

EXAMPLE 2.4. *Let $\mathbf{x}_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, where σ^2 is unknown. Consider testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. We showed that the α -level LRT of $H_0 : \mu = \mu_0$ rejects if $|\phi(\mathbf{x})| > t_{\frac{\alpha}{2}}$, where*

$$\phi(\mathbf{x}) = \left| \frac{\sqrt{n}(\bar{\mathbf{x}} - \mu_0)}{\hat{\mathbf{s}}} \right|.$$

is the t -statistic. The acceptance region is $\{x \in \mathbf{R} : |\phi(\mathbf{x})| \leq t_{\frac{\alpha}{2}}\}$. Thus

$$\begin{aligned}\mathcal{A}^{-1}(\mathbf{x}) &= \{\mu_0 \in \mathbf{R} : |\phi(\mathbf{x})| \leq t_{\frac{\alpha}{2}}\} \\ &= \left[\bar{\mathbf{x}} - \frac{\hat{s}}{\sqrt{n}}t_{\frac{\alpha}{2}}, \bar{\mathbf{x}} + \frac{\hat{s}}{\sqrt{n}}t_{\frac{\alpha}{2}}\right]\end{aligned}$$

is a $1 - \alpha$ -confidence interval for μ_0 .

We remark that Lemma 2.2 has a converse: it is possible to obtain an α -level test from a $1 - \alpha$ -confidence set.

LEMMA 2.5. *Let $\mathcal{C}(\mathbf{x})$ is a $1 - \alpha$ -confidence interval for θ . The test*

$$\varphi(\mathbf{x}) = 1 - \mathbf{1}_{\mathcal{C}(\mathbf{x})}(\theta_0),$$

$H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. That is, the test that rejects when $\mathcal{C}(\mathbf{x})$ does not contain θ_0 is an α -level test.

PROOF. Since $\mathcal{C}(\mathbf{x})$ is a $1 - \alpha$ confidence interval for θ_0 , $\mathbf{P}_0(\theta_0 \in \mathcal{C}(\mathbf{x})) \geq 1 - \alpha$. Thus

$$\begin{aligned}\mathbf{E}_0[1 - \varphi(\mathbf{x})] &= 1 - \mathbf{P}_0(\theta_0 \in \mathcal{C}(\mathbf{x})) \\ &\leq 1 - (1 - \alpha).\end{aligned}$$

□

A convenient formalism that highlights the connection between hypothesis testing and interval is that of *pivot* or *pivotal quantity*.

DEFINITION 2.6. *A function $\phi(\mathbf{x}, \theta)$ is a pivot for a parametric model if the its distribution under $\mathbf{x} \sim F_\theta$ does not depend on θ . That is,*

$$\mathbf{P}_\theta(\phi(\mathbf{x}, \theta) \in \mathcal{C})$$

does not depend on θ .

We remark that a pivot is technically not a statistic because the function depends on the unknown parameter θ . The canonical example of a pivot is the z -statistic $\frac{\mathbf{x} - \mu}{\sigma}$. It is a pivot for the normal location-scale model.

Given a pivot, it is possible to

1. test the hypothesis $H_0 : \theta = \theta_0$. Under H_0 , we know the distribution of the pivot $\phi(\mathbf{x}, \theta_0)$. Thus comparing the observation $\phi(\mathbf{x}, \theta_0)$ to the known distribution is the basis of a test.

2. form a confidence interval for θ . As we shall see, inverting the pivot in its second argument leads to a confidence interval for θ .

Let $\mathcal{C} \subset \phi(\mathcal{X}, \theta)$ be a set of $1 - \alpha$ mass under the distribution of the pivot:

$$\mathbf{P}_\theta(\phi(\mathbf{x}, \theta) \in \mathcal{C}).$$

If we pin the second argument of the pivot at θ_0 (letting $\phi_{\theta_0}(x) = \phi(x, \theta_0)$) and invert the pivot in its first argument, we obtain

$$\phi_{\theta_0}^{-1}(\mathcal{C}) := \{x \in \mathcal{X} : \phi(x, \theta_0) \in \mathcal{C}\},$$

which is the acceptance region of an α -level test of $H_0 : \theta = \theta_0$. Indeed,

$$\mathbf{P}_{\theta_0}(\mathbf{x} \in \phi_{\theta_0}^{-1}(\mathcal{C})) = \mathbf{P}_{\theta_0}(\phi(\mathbf{x}, \theta_0) \in \mathcal{C}) = 1 - \alpha,$$

If we pin the first argument of the pivot at x (letting $\phi_x(\theta) = \phi(x, \theta)$) and invert the pivot in its second argument, we obtain

$$\phi_x^{-1}(\mathcal{C}) := \{\theta \in \Theta : \phi(x, \theta) \in \mathcal{C}\},$$

which is a $1 - \alpha$ -confidence set for θ .

EXAMPLE 2.7 (Example 2.3 continued). *We know $\mathbf{x} - \mu$ is a pivot for the Gaussian location model: if $\mathbf{x} \sim \mathcal{N}(\mu, 1)$, $\phi(\mathbf{x}, \mu) := \mathbf{x} - \mu$ is a pivot. If we pin the second argument of the pivot at μ_0 , the pre-image of $[z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]$ under the pivot is*

$$\phi_{\mu_0}^{-1}([z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]) = \{x \in \mathbf{R} : x - \mu \in [z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]\},$$

which is the acceptance region of the α -level LRT of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. The pre-image of $(-\infty, z_\alpha]$ is

$$\phi_{\mu_0}^{-1}((-\infty, z_\alpha]) = \{x \in \mathbf{R} : x - \mu \in (-\infty, z_\alpha]\},$$

which is the acceptance region of the α -level UMPU test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$. Finally, if we pin the first argument of the pivot at μ_0 and invert the pivot in its second argument, we obtain

$$\phi_x^{-1}([z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]) = \{\mu \in \mathbf{R} : x - \mu \in [z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}]\},$$

which is a $1 - \alpha$ -confidence interval for μ .

In practice, it is usually not possible to derive an exact pivot. However, asymptotic pivots are easier to obtain. Formally, an asymptotic pivot for a parametric model is a function whose asymptotic distribution does not depend on the parameter. The canonical example of an asymptotic pivot is:

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

Unsurprisingly, inverting asymptotic pivots leads to asymptotically α -level tests and $1 - \alpha$ -confidence intervals.

2.1. Most accurate interval estimators.

DEFINITION 2.8. A $1 - \alpha$ confidence set $\mathcal{C}(\mathbf{x})$ is most accurate at θ if

$$\mathbf{P}_\theta(\theta' \in \mathcal{C}(\mathbf{x})) \leq \mathbf{P}_\theta(\theta' \in \mathcal{C}'(\mathbf{x}))$$

for any $\theta' \neq \theta$ and any other $1 - \alpha$ confidence set $\mathcal{C}'(\mathbf{x})$. It is uniformly most accurate (UMA) on Θ if it is most accurate at any $\theta \in \Theta$.

Intuitively, a most accurate $1 - \alpha$ confidence set is least likely (among $1 - \alpha$ confidence sets) to include “incorrect” parameters. A straightforward calculation shows that a most accurate $1 - \alpha$ confidence set has the smallest expected (Lebesgue) volume among $1 - \alpha$ confidence sets. Indeed,

$$\begin{aligned} \mathbf{E}_\theta[\text{vol}[\mathcal{C}(\mathbf{x})]] &= \int_{\mathcal{X}} \left(\int_{\Theta} \mathbf{1}_{\mathcal{C}(x)}(\theta') d\theta' \right) f_\theta(x) dx \\ &= \int_{\Theta} \left(\int_{\mathcal{X}} \mathbf{1}_{\mathcal{C}(x)}(\theta') f_\theta(x) dx \right) d\theta' \\ &= \int_{\Theta} \mathbf{P}_\theta(\theta' \in \mathcal{C}(x)) d\theta', \end{aligned}$$

which is minimized when the integrand is minimized on Θ .

THEOREM 2.9. A $1 - \alpha$ -UMA confidence set on Θ is the inverse of α -level UMP tests of $H_0 : \theta = \theta'$ versus $H_1 : \theta \in \Theta \setminus \{\theta'\}$.

PROOF. Let $\mathcal{A}(\theta')$ be the acceptance area of a UMP test of $H_0 : \theta = \theta'$ versus $H_1 : \theta \in \Theta \setminus \{\theta'\}$. Since the test is UMP,

$$\mathbf{P}_\theta(\mathbf{x} \in \mathcal{X} \setminus \mathcal{A}(\theta')) \geq \mathbf{P}_\theta(\mathbf{x} \in \mathcal{X} \setminus \mathcal{A}'(\theta'))$$

for any $\theta \in \Theta \setminus \{\theta'\}$ and any other the acceptance area of any other α -level test $\mathcal{A}'(\theta')$. Equivalently,

$$\mathbf{P}_\theta(\mathbf{x} \in \mathcal{A}(\theta')) \leq \mathbf{P}_\theta(\mathbf{x} \in \mathcal{A}'(\theta')).$$

We invert the acceptance area to obtain

$$\mathbf{P}_\theta(\theta' \in \mathcal{A}^{-1}(\mathbf{x})) \leq \mathbf{P}_\theta(\theta' \in \mathcal{A}'^{-1}(\mathbf{x})).$$

□

We observe that to obtain a most accurate confidence set at a point, it is necessary to invert a family of most powerful tests. Indeed, by Theorem 2.9, the most accurate confidence set at $\bar{\theta}$ is the “inverse” of α -level most powerful tests of $H_0 : \theta = \theta'$ versus $H_1 : \theta = \bar{\theta}$.

In hypothesis testing, there is often no UMP tests for a prescribed pair of null and alternative hypotheses. Similarly, in interval estimation, there is usually no UMA confidence set. To make the notion of most powerful test tractable, we restricted our attention to unbiased tests and studied UMPU tests. There is a similar notion of unbiasedness for confidence sets and inverting unbiased tests lead to unbiased confidence sets.

DEFINITION 2.10. A $1 - \alpha$ confidence set is unbiased if

$$\mathbf{P}_\theta(\theta' \in \mathcal{C}(\mathbf{x})) \leq 1 - \alpha \text{ for any } \theta' \neq \theta.$$

It is a uniformly most accurate unbiased (UMAU) confidence set on Θ if it is uniformly most accurate on Θ among unbiased confidence sets.

LEMMA 2.11. A $1 - \alpha$ -unbiased confidence set on Θ is the inverse of α -level unbiased tests of $H_0 : \theta = \theta'$ versus $H_1 : \theta \in \Theta \setminus \{\theta'\}$.

PROOF. The lemma is almost a tautology. Let $\mathcal{A}(\theta')$ be the acceptance region of an unbiased test of $H_0 : \theta = \theta'$ versus $H_1 : \theta \in \Theta \setminus \{\theta'\}$:

$$\mathbf{P}_\theta(\mathbf{x} \in \mathcal{A}^{-1}(\theta')) \leq 1 - \alpha.$$

We invert the acceptance region to obtain the stated conclusion. □

We combine Lemma 2.11 and Theorem 2.9 to obtain the (unsurprising) result that the inverse of the acceptance region of a UMPU test is a UMAU confidence set. The proof is very similar to the proof of Theorem 2.9, and we skip the details here.

THEOREM 2.12. A $1 - \alpha$ -UMAU confidence set on Θ is the inverse of α -level UMPU tests of $H_0 : \theta = \theta'$ versus $H_1 : \theta \in \Theta \setminus \{\theta'\}$.

YUEKAI SUN
BERKELEY, CALIFORNIA
DECEMBER 2, 2015