Lecture 9

Uniformly Most Powerful Tests

1 Some more on bootstrap in testing

Example. Assume you have a random sample from some unknown distribution $X = (X_1, ..., X_n)$ with four finite moments. Denote $\mu = EX_i$, $\sigma^2 = Var(X_i)$. Let $h(\mu, \sigma^2)$ be a twice-continuously-differentiable function of both arguments with a full rank derivative. Assume we wish to test the hypothesis $H_0 : h(\mu, \sigma^2) = 0$ against $H_1 : h(\mu, \sigma^2) \neq 0$.

One way to approach this problem is to define a new parameter $\gamma = h(\mu, \sigma^2)$ for which we have a consistent estimate $\hat{\gamma} = h(\bar{X}, s^2)$. We can even prove asymptotic gaussianity via the delta-method:

$$\sqrt{n}(\hat{\gamma} - \gamma) \Rightarrow N(0, (\nabla h)' V \nabla h),$$

where $\sqrt{n}(\bar{X} - \mu, s^2 - \sigma^2)' \Rightarrow N(0, V)$ and $\nabla h = (\frac{\partial h(\mu,\sigma^2)}{\partial \mu}, \frac{\partial h(\mu,\sigma^2)}{\partial \sigma^2})'$. So, in principle, one may get a natural estimate of the asymptotic variance of $\hat{\gamma}$ and construct a regular t-statistic to conduct a test. Alternatively one can employ a non-parametric bootstrap in the way we discussed in the previous lecture by approximating an unknown, finite-sample distribution of a statistic $z = \sqrt{n}(\hat{\gamma} - 0)$ with $z^* = \sqrt{n}(h(\bar{X}^*, s^{*2}) - \hat{\gamma})$. Let us denote all parameters and statistics related to the bootstraped distribution with stars, that is $\mu^* = \int x d\hat{F} = E^* X_i^* = \bar{X}$, etc. Several comments are needed:

- Notice a "re-centering" of the null hypothesis. We approximate the unknown distribution of X_i with empirical distribution function $\hat{F}$. The size is defined under the null, so, we assume that the unknown distribution of X_i is such that $H_0 : h(\mu, \sigma^2) = 0$. However we know for sure that this null does not hold for the empirical distribution $\hat{F}$ as the mean for it, $\mu^* = \bar{X}$, and the variance is $\sigma^{*2} = \frac{1}{n} \sum_i (X_i - \bar{X})^2$ (almost s^2; the difference is of order $O_p(1/n)$ and will be ignored here as too 'smallish' to care). From this perspective the true value of parameter γ for the empirical distribution is $\hat{\gamma}$ (almost). So, for the bootstrapped samples we should be testing the true null hypothesis $H_0 : h(\mu^*, \sigma^{*2}) = \hat{\gamma}$.

- The justification of bootstrap validity here is based on the validity of asymptotic approximation, both for the statistic itself and for its bootstrapped version:

$$\sqrt{n}(\hat{\gamma}^* - \hat{\gamma}) \Rightarrow N(0, (\nabla h^*)' V^* \nabla h^*),$$

where $\sqrt{n}(\bar{X}^* - \mu^*, s^{*2} - \sigma^{*2})' \Rightarrow N(0, V^*)$ and $\nabla h^* = (\frac{\partial h(\mu^*,\sigma^{*2})}{\partial \mu}, \frac{\partial h(\mu^*,\sigma^{*2})}{\partial \sigma^2})'$. And noticing that $V^* = V + o_p(1)$ and $\nabla h^* = \nabla h + o_p(1)$.

- In the vast majority of cases the justification for the bootstrap comes from the validity of some asymptotic approximation to the test statistic, and so the asymptotic method can be used. There are only a couple of cases I know of when we do not know if a statistic has an asymptotic limit and, if so, what it is, but the bootstrap magically works. But there is a large number of cases (weak identification, moment inequality, boundary problem, post-selection inferences) when there are difficulties in carrying out the asymptotic approach, and when the bootstrap does not work either.

- Another thing the bootstrap is implicitly doing here is bias correction. Remember our discussion about bias of the order $O(1/n)$ present in estimate $\hat{\gamma}$?

- In the example above by using the bootstrap we avoided calculating the standard error for $\hat{\gamma}$. The question arises, if we are willing to calculate standard errors, would still using the bootstrap have any benefits? The answer is yes. Imagine we construct a natural consistent estimator $\hat{V}$ for V and considered t-statistic:
$$t = \sqrt{n}\frac{\hat{\gamma} - 0}{\sqrt{(\nabla\hat{h})'\hat{V}\nabla\hat{h}}},$$
where $\nabla\hat{h}$ is the corresponding derivative evaluated at $\bar{X}, s^2$. We know that under the null $t \Rightarrow N(0,1)$. The bootstrapped statistic t^* is constructed in an analogous way (using re-centering) and $t^* \Rightarrow N(0,1)$. Thus the distribution of t is close to the distribution of t^* in large samples. However, there are results suggesting that the distance between the distributions of t and t^* are asymptotically smaller than the distance between the distribution of t and the standard normal cdf. That is, the bootstrap provides better approximation, or so called second-order refinement. Statements like this one are based on Edgeworth's expansion, which in general is quite complicated to arrive at and we will not derive it. But it is a statement of the following sort:
$$P\{t \le x\} = \Phi(x) + \frac{1}{\sqrt{n}}\phi(x, F) + O(\frac{1}{n}),$$
here $\phi(x, F)$ will depend on F only through several of the first moments of this distribution (I think in our case on first 6). Similarly,
$$P^*\{t^* \le x\} = \Phi(x) + \frac{1}{\sqrt{n}}\phi(x, \hat{F}) + O_p(\frac{1}{n}).$$
One may show that $\phi(x, \hat{F}) - \phi(x, F) = O_p(\frac{1}{\sqrt{n}})$ (CLT and delta-method). Then
$$P\{t \le x\} - P^*\{t^* \le x\} = O_p(\frac{1}{n}),$$
while
$$P\{t \le x\} - \Phi(x) = O_p(\frac{1}{\sqrt{n}}).$$

2 Uniformly Most Powerful Test

Let $\Theta = \Theta_0 \cup \Theta_1$ be a parameter space. Consider a parametric family $\{f(x|\theta),\ \theta \in \Theta\}$. Suppose we want to test the null hypothesis, H_0, that $\theta \in \Theta_0$ against the alternative, H_a, that $\theta \in \Theta_1$. Let C be some critical set. Then the probability that the null hypothesis is rejected is given by $\beta(\theta) = P_\theta\{X \notin C\}$. Recall that the test based on C is of level α if $\alpha \geq \sup_{\theta \in \Theta_0} \beta(\theta)$. The restriction of $\beta(\cdot)$ on Θ_1 is known as the power of the test. Let C' be another critical set. Denote the power of the test based on C' by $\beta'(\theta)$. Suppose that both tests are of level α. Then the test based on C is more powerful than the test based on C' if $\beta(\theta) \geq \beta'(\theta)$ for all $\theta \in \Theta_1$. Any test which is more powerful than any other test in some class $\mathcal{G}$ will be called uniformly the most powerful test in the class $\mathcal{G}$ (UMP test).

As follows from the theorem below, the UMP test exists if both the null and the alternative are simple.

Theorem 1 (Neyman-Pearson Lemma)**.** *Let $f(x|\theta)$ with $\Theta = \{\theta_0, \theta_1\}$ be some parametric family. Suppose we want to test the null hypothesis, H_0, that $\theta = \theta_0$ against the alternative hypothesis, H_a, that $\theta = \theta_1$. Assume that some critical set C satisfies (1) $x \in C$ if $kf(x|\theta_0) > f(x|\theta_1)$ and (2) $x \notin C$ if $kf(x|\theta_0) < f(x|\theta_1)$ where $k \geq 0$ is chosen so that $\alpha = P_{\theta_0}(X \notin C)$. Then the test based on C is the UMP among all tests of level α. In addition, any UMP test of level α satisfies (1) and (2).*

Proof. Denote $\phi(x) = I(x \notin C)$, i.e. $\phi(x) = 1$ if $x \notin C$ and 0 otherwise. Thus, $\phi(x)$ denotes the probability that the test based on C rejects the null hypothesis upon observing data of value x. Consider any other test of level α. Let $\tilde{\phi}(x)$ denote the probability that this test rejects the null hypothesis upon observing data value x. Since this test is of level α,

$$\tilde{\beta}(\theta_0) = \int \tilde{\phi}(x) f(x|\theta_0) dx \leq \alpha,$$

where $\tilde{\beta}(\theta)$ denotes the probability that this test rejects the null hypothesis when the true parameter value is θ.

Note that

$$(\phi(x) - \tilde{\phi}(x))(f(x|\theta_1) - kf(x|\theta_0)) \geq 0$$

for any x. Indeed, if $f(x|\theta_1) - kf(x|\theta_0) \geq 0$, then $\phi(x) = 1$ and $\phi(x) - \tilde{\phi}(x) \geq 0$. If $f(x|\theta_1) - kf(x|\theta_0) < 0$, then $\phi(x) = 0$ and $\phi(x) - \tilde{\phi}(x) \leq 0$. So,

$$0 \leq \int (\phi(x) - \tilde{\phi}(x))(f(x|\theta_1) - kf(x|\theta_0)) dx = \beta(\theta_1) - \tilde{\beta}(\theta_1) - k(\beta(\theta_0) - \tilde{\beta}(\theta_0)),$$

where $\beta(\theta)$ denotes the probability that the test based on C rejects the null hypothesis when the true parameter value is θ. Therefore,

$$\beta(\theta_1) - \tilde{\beta}(\theta_1) \geq k(\beta(\theta_0) - \tilde{\beta}(\theta_0)) \geq k(\alpha - \alpha) = 0,$$

since $\tilde{\beta}(\theta_0) \leq \alpha$ and $\beta(\theta_0) = \alpha$. So the test based on C is more powerful than any other test of level α. So it is the UMP which proves the first statement of the theorem.

If $\tilde{\phi}(\cdot)$ is also a UMP among all tests of level α, then $\tilde{\beta}(\theta_1) = \beta(\theta_1)$. So, $k(\alpha - \tilde{\beta}(\theta_0)) \leq 0$. Therefore $\tilde{\beta}(\theta_0) \geq \alpha$. On the other hand, $\tilde{\beta}(\theta_0) \leq \alpha$, since this test is of level α. We conclude that $\tilde{\beta}(\theta_0) = \alpha$. It follows that

$$\int (\phi(x) - \tilde{\phi}(x))(f(x|\theta_1) - kf(x|\theta_0))dx = 0.$$

Since the integrand is nonnegative for all x,

$$(\phi(x) - \tilde{\phi}(x))(f(x|\theta_1) - kf(x|\theta_0)) = 0.$$

Thus, $\tilde{\phi}(x) = \phi(x)$ whenever $f(x|\theta_1) - kf(x|\theta_0) \neq 0$. So, $\tilde{\phi}(\cdot)$ also satisfies conditions (1) and (2). $\square$

Recall from the proof of the factorization theorem in lecture 4 that if $T(X)$ is a sufficient statistic, then $f(x|\theta) = g(T(x)|\theta)h(x)$ where $g(\cdot)$ denotes pdf of $T(X)$. So, in terms of the pdf of sufficient statistics, the critical set C of the UMP test satisfies (1) $x \in C$ if $kg(T(x)|\theta_0) > g(T(x)|\theta_1)$ and (2) $x \notin C$ if $kg(T(x)|\theta_0) < g(T(x)|\theta_1)$.

Example Let $X_1, ..., X_n$ be a random sample from the $N(\mu, \sigma^2)$ distribution with known σ^2. Suppose we want to test the null hypothesis, H_0, that $\mu = \theta_0$ against the alternative hypothesis, H_a, that $\mu = \theta_1$. Without loss of generality we can assume that $\theta_0 > \theta_1$. We have already seen that the sufficient statistic in this example is given by $\overline{X}_n = \sum_{i=1}^n X_i/n$. We know that $\overline{X}_n \sim N(\mu, \sigma^2/n)$. So,

$$g(t|\theta) = C\exp\{-(n/(2\sigma^2))(t - \theta)^2\}.$$

From the Neyman-Pearson lemma, the UMP test among all tests of level α accepts the null hypothesis if and only if

$$kg(\overline{X}_n|\theta_0) > g(\overline{X}_n|\theta_1)$$

or, equivalently,

$$k\exp\{(n/(\sigma^2))\overline{X}_n(\theta_0 - \theta_1) - (n/(2\sigma^2))(\theta_0^2 - \theta_1^2)\} > 1.$$

Since $\theta_0 > \theta_1$, the test accepts the null hypothesis if and only if $\overline{X}_n > \tilde{k}$ where $\tilde{k}$ is such that $P_{\theta_0}(\overline{X} \leq \tilde{k}) = \alpha$. So, $\tilde{k} = \theta_0 + \sigma z_\alpha/\sqrt{n}$ where z_α denotes an α-quantile of the standard normal distribution.

3 UMP tests with composite hypotheses

The idea of the Neyman-Pearson lemma is to consider an optimization problem $\int \phi(x)f(x|\theta_1)dx \to \max$ subject to $\int \phi(x)f(x|\theta_0)dx \leq \alpha$ and $0 \leq \phi \leq 1$. The solution to this problem gives the UMP test of level α. This problem looks like maximization of utility given some budget constraint. We want to choose the most valuable items with the lowest price. In some special cases we can extend this idea to the case of complex hypotheses. Suppose we want to test the null hypothesis, H_0, that $\theta \leq \theta_0$ against the alternative hypothesis, H_a, that $\theta > \theta_0$. Then the UMP test exists if $f(x|\theta)$ satisfies the monotone likelihood ratio property.

Definition 2. A family $f(x|\theta)$ with $\theta \in \mathbb{R}$ satisfies the monotone likelihood ratio if there exists some function

$T(x)$ such that for any $\theta < \theta'$, $P_{\theta'}(x)/P_\theta(x)$ depends on x only through $T(x)$ and, moreover, $P_{\theta'}(x)/P_\theta(x)$ is a nondecreasing function of $T(x)$.

Theorem 3. *Let $f(x|\theta)$ with $\theta \in \mathbb{R}$ be some parametric family that satisfies the monotone likelihood ratio with function $T(x)$. Suppose we want to test the null hypothesis, H_0, that $\theta \geq \theta_0$ against the alternative hypothesis, H_a, that $\theta < \theta_0$. Then a UMP test of level α exists. It is given by $\phi(X) = 1$ if $T(X) > C$, $\phi(X) = \gamma$ if $T(X) = C$, and 0 otherwise for some constants c and γ such that $E_{\theta_0}[\phi(X)] = \alpha$. In addition, the power of this test $\beta(\theta) = E_\theta[\phi(X)]$ for $\theta > \theta_0$ is strictly decreasing in θ.*

Proof. Choose a simple alternative $\theta_1 < \theta_0$. By the Neyman-Pearson lemma, the UMP test of θ_0 against θ_1 accepts the null hypothesis if $f(X|\theta_0)/f(X|\theta_1) > k$ and rejects it if $f(X|\theta_0)/f(X|\theta_1) < k$. By the monotone likelihood ratio, this test accepts the null hypothesis if $T(X) > C$ and rejects it if $T(X) < C$. When $T(X) = C$, the test rejects with probability, say, γ. The constants C and γ should be chosen such that $E_{\theta_0}[\phi(X)] = \alpha$. Now, note that the same test will be UMP of level α for any other alternative $\theta_2 < \theta_0$ as well. So, this test is UMP of level α for the null hypothesis $\theta = \theta_0$ against the alternative $\theta < \theta_0$. Thus, to show that the same test is UMP of level α for the null $\theta \geq \theta_0$ against the alternative $\theta < \theta_0$, it will be enough to show that this test is of level α, i.e. $\sup_{\theta \geq \theta_0} \beta(\theta) \leq \alpha$.

Since there always exists a test of level α with power α (this test rejects the null hypothesis with probability α independently of the data), $\beta(\theta_0) \leq \beta(\theta_1)$. Since the test based on $T(x)$, C, and γ is also UMP (of some level) for the null hypothesis $\theta = \theta_1$ against the alternative $\theta = \theta_2$ for any $\theta_2 < \theta_1$, the same argument yields $\beta(\theta_1) \leq \beta(\theta_2)$ for any $\theta_1 \geq \theta_2$ which is the second statement of the theorem. The first statement of the theorem follows from $\sup_{\theta \geq \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha$ since $\beta(\theta)$ is decreasing in θ. □

In many cases a UMP test does not exist. Below is an example of such a situation.

Example Let $X_1, ..., X_n$ be a random sample from an $N(\theta, \sigma^2)$ distribution with known σ^2. Suppose we want to test the null hypothesis, H_0, that $\theta = \theta_0$ against the alternative hypothesis, H_a, that $\theta \neq \theta_0$. Consider some $\theta_1 < \theta_0$. The only UMP test of level α of $\theta = \theta_0$ against $\theta = \theta_1$ rejects the null hypothesis if and only if $\overline{X}_n < \theta_0 + \sigma z_\alpha/\sqrt{n}$ as we have already seen. But this test has little power in our situation for any $\theta > \theta_0$. Indeed, $\beta(\theta) < \alpha$ for all $\theta > \theta_0$. So, this test cannot be UMP in this situation, since there always exists a test of level α with power α.

3.1 Unbiased tests

Since there are no UMP tests among all tests of level α in many situations, the question arises whether we can find UMP tests in some smaller, but still reasonably large, classes of tests. The definition below gives a property that reasonable tests should have.

Definition 4. Any test of the null hypothesis $\theta \in \Theta_0$ against the alternative $\theta \in \Theta_1$ is called unbiased if for some $\alpha \in [0, 1]$, $\beta(\theta) \leq \alpha$ for all $\theta \in \Theta_0$ and $\beta(\theta) \geq \alpha$ for all $\theta \in \Theta_1$.

Let $\Theta = \mathbb{R}$ be a parameter space. Suppose we want to test the null hypothesis, H_0, that $\theta = \theta_0$ against the alternative hypothesis, H_a, that $\theta \neq \theta_0$. Consider a test which rejects the null hypothesis with probability

$\phi(x)$ upon observed data value x. As before, denote $\beta(\theta) = \int \phi(x) f(x|\theta) dx$. If $\beta(\theta)$ is differentiable in θ, then for any unbiased test, we necessarily have $\beta'(\theta_0) = 0$. Indeed, if this condition does not hold, then there exists a point θ in the neighborhood of θ_0 such that $\beta(\theta) \leq \beta(\theta_0)$ by definition of derivative. In some situations, there exists a UMP test among all unbiased tests of level α even though there are no UMP tests among all tests of level α.

4 Likelihood Ratio Test

Suppose we want to test the null hypothesis, H_0, that $\theta \in \Theta_0$ against the alternative hypothesis, H_a, that $\theta \in \Theta_1$. Denote $\Theta = \Theta_0 \cup \Theta_1$. Let $\mathcal{L}(\theta|x)$ denote likelihood function. Then we have

Definition 5. A Likelihood ratio test (LRT) statistic is

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta|x)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta|x)}.$$

By definition, $0 \leq \lambda(x) \leq 1$. Small values of the LRT statistic imply that there is a value θ in the alternative hypothesis Θ_1 which gives much greater likelihood than all values in the null hypothesis. So, likelihood ratio tests reject the null hypothesis if and only if $\lambda(x) \leq c$ for some c. As usual, the constant c is chosen according to the desired level of the test.

Let $\hat{\theta}_r = \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta|x)$ be the ML estimator of the restricted model. Let $\hat{\theta}_{ur} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta|x)$ be the ML estimator of the unrestricted model. Then an equivalent way to define LRT statistic is to set

$$\lambda(x) = \frac{\mathcal{L}(\hat{\theta}_r|x)}{\mathcal{L}(\hat{\theta}_{ur}|x)}.$$

Example Let $X_1, ..., X_n$ be a random sample from an $N(\theta, 1)$ distribution. Suppose we want to test the null hypothesis, H_0, that $\theta = \theta_0$ against the alternative hypothesis, H_a, that $\theta \neq \theta_0$. Then $\hat{\theta}_r = \theta_0$ and $\hat{\theta}_{ur} = \hat{\theta}_{MLE} = \overline{X}_n$. So, the LRT statistic is

$$\begin{aligned}
\lambda(x) &= \frac{\mathcal{L}(\theta_0|x)}{\mathcal{L}(\hat{\theta}_{MLE}|x)} \\
&= \frac{(2\pi)^{-n/2} \exp\{-(1/2)\sum_{i=1}^n (X_i - \theta_0)^2\}}{(2\pi)^{-n/2} \exp\{-(1/2)\sum_{i=1}^n (X_i - \overline{X}_n)^2\}} \\
&= \exp\{-(1/2)\sum_{i=1}^n [(X_i - \overline{X}_n + \overline{X}_n - \theta_0)^2 - (X_i - \overline{X}_n)^2]\} \\
&= \exp\{-(n/2)(\overline{X}_n - \theta_0)^2\}
\end{aligned}$$

So, the LRT rejects the null hypothesis if and only if $|\overline{X}_n - \theta_0| > c$. Specifically, the LRT of level α rejects the null hypothesis if and only if $\overline{X}_n - \theta_0 > z_{1-\alpha/2}/\sqrt{n}$ or $\overline{X}_n - \theta_0 < z_{\alpha/2}/\sqrt{n}$ since under the null hypothesis, $\overline{X}_n \sim N(\theta_0, 1/\sqrt{n})$.