ECON 481
LECTURE 10: INTRO TO UNIFORMITY

Ivan A. Canay
Northwestern University

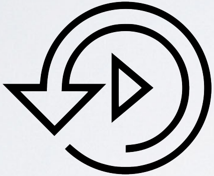
Past & Future

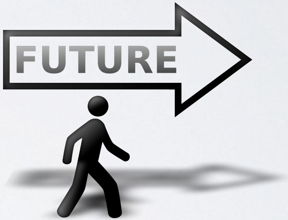
PARTS I & II

- **Part I**: Treatment Effects
- **Part II**: Asymptotic Approximations

PART III

- Why uniform inference?
- What's the problem with pointwise approximations?
- When if uniformity a technicality and when it's a real issue.
- Inference in moment inequalities models.





Introduction

- Let $\{X_i : i = 1, \ldots, n\}$ be an i.i.d. sample from some distribution $P \in \mathbf{P}$.
- We wish to test the null hypothesis $H_0 : P \in \mathbf{P}_0 \subseteq \mathbf{P}$.
- To this end, one may consider a **test function** $\phi_n = \phi_n(X_1, \ldots, X_n)$ (that maps data into a binary decision) such that it controls the probability of a Type I error in some sense.
- We must distinguish the exact size of a test from its approximate or asymptotic size.
- **Ideal**: we would like the test to satisfy

$$E_P[\phi_n] \leqslant \alpha \text{ for all } P \in \mathbf{P}_0 \text{ and } n \geqslant 1 , \tag{1}$$

  but many times this is too demanding of a requirement.
- We therefore rely on approximations that take n to infinity. But, **HOW?**

Pointwise approximation

- **Alternative**: we may settle instead for tests such that
$$\limsup_{n\to\infty} E_P[\phi_n] \leqslant \alpha \text{ for all } P \in \mathbf{P}_0 \ . \tag{2}$$

- Test satisfying (1) are said to be of **level** α for $P \in \mathbf{P}_0$, whereas tests satisfying (2) are said to be **pointwise asymptotically of level** α for $P \in \mathbf{P}_0$.

- The hope is that if (2) holds, then (1) holds approximately, at least for large enough n.

- However, asymptotic constructions that merely assert a result under any **fixed** distribution P may lead to deceiving results.

- All that (2) ensures is that for each $P \in \mathbf{P}_0$ and $\epsilon > 0$ there is an $N(P)$ such that for all $n > N(P)$
$$E_P[\phi_n] \leqslant \alpha + \epsilon \ .$$
Note: the sample size required for the approximation to work, $N(P)$, may depend on P. It could be the case that for **every sample size** n (even, e.g., for $n = 10^{10}$) there could be $P = P_n \in \mathbf{P}_0$ such that
$$E_P[\phi_n] \gg \alpha \ .$$

Example

Example

Suppose $\mathbf{P} = \{P \text{ on } \mathbf{R} : 0 < \sigma^2(P) < \infty\}$ and $\mathbf{P}_0 = \{P \in \mathbf{P} : \mu(P) = 0\}$. Let ϕ_n be the t-test; that is, $\phi_n = I\{\sqrt{n}\bar{X}_n > \hat{\sigma}_n z_{1-\alpha}\}$, where $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution. We know

$$E_P[\phi_n] \to \alpha \text{ for all } P \in \mathbf{P}_0 \ .$$

However, we can show that for every $0 < c < 1$ and every sample size n there exists a $P_{n,c} \in \mathbf{P}_0$ such that

$$E_{P_{n,c}}[\phi_n] \geqslant c \ .$$

Uniform approximation

- **To rule this out**: we need to ensure that the convergence in (2) is uniform for $P \in \mathbf{P}_0$.

Definition

The sequence $\{\phi_n\}$ is **uniformly asymptotically of level** α if

$$\limsup_{n\to\infty} \sup_{P\in\mathbf{P}_0} E_P[\phi_n] \leqslant \alpha \,. \tag{3}$$

The LHS of (3) is called the **asymptotic size** of $\{\phi_n\}$.

- The requirement in (3) implies that for each $\epsilon > 0$ there is an N (independent of P) st for all $n > N$

$$E_P[\phi_n] \leqslant \alpha + \epsilon \,.$$

- For the t-test, the example shows this is not true for $\mathbf{P} = \{P \text{ on } \mathbf{R} : 0 < \sigma^2(P) < \infty\}$ and $\mathbf{P}_0 = \{P \in \mathbf{P} : \mu(P) = 0\}$.

Definition

The sequence $\{\phi_n\}$ is pointwise consistent in power if, for an $P \in \mathbf{P}_1$, $E_P[\phi_n] \to 1$ as $n \to \infty$.

QUESTIONS?

A result of Bahadur and Savage (1956)

- Given the example, we know that for the t-test, when $\mathbf{P} = \{P \text{ on } \mathbf{R} : 0 < \sigma^2(P) < \infty\}$ and $\mathbf{P}_0 = \{P \in \mathbf{P} : \mu(P) = 0\}$,

$$\sup_{P \in \mathbf{P}_0} E_P[\phi_n] = 1 \,.$$

- Perhaps a bit shocking, but it is unique to the t-test? Unfortunately, the answer is no.

Theorem (Bahadur-Savage)

Let $\{X_i : i = 1, \ldots, n\}$ be i.i.d. with distribution $P \in \mathbf{P}$ where $\mathbf{P}$ is a class of distributions on $\mathbf{R}$ such that

(I) *For every $P \in \mathbf{P}$, $\mu(P)$ exists and is finite;*

(II) *For every $m \in \mathbf{R}$, there is $P \in \mathbf{P}$ such that $\mu(P) = m$;*

(III) *$\mathbf{P}$ is convex in the sense that if $P_1, P_2 \in \mathbf{P}$, then $\gamma P_1 + (1-\gamma) P_2 \in \mathbf{P}$.*

Then, for $H_0 : \mu(P) = 0$, the following statements hold .

(A) *Any test of H_0 which has size α for $\mathbf{P}$ has power $\leqslant \alpha$ for any alternative $P \in \mathbf{P}$.*

(B) *Any test of H_0 which has power β against some alternative $P \in \mathbf{P}$ has size $\geqslant \beta$.*

A useful Lemma

Lemma

Let $\{X_i : i = 1, \ldots, n\}$ *be i.i.d. with distribution* $P \in \mathbf{P}$*, where* $\mathbf{P}$ *is the class of distributions on* $\mathbf{R}$ *satisfying (i)-(iii) in the previous theorem. Let* $\phi_n(X_1, \ldots, X_n)$ *be any test function and define*

$$\mathbf{P}_m = \{P \in \mathbf{P} : \mu(P) = m\} .$$

Then,

$$\inf_{P \in \mathbf{P}_m} E_P[\phi_n] \quad \textit{and} \quad \sup_{P \in \mathbf{P}_m} E_P[\phi_n]$$

are independent of m*.*

Proof of the Lemma I

Proof of the Lemma II

Proof of the Theorem

Remarks

- The Bahadur-Savage result holds in the **multivariate case** as well. The theorem reads exactly, except that $\mathbf{P}$ refers to a family of distributions on $\mathbf{R}^k$ satisfying (i)-(iii) above with m a vector.

- The class of distributions with finite second moment satisfies the requirements of the theorem, as does the class of distributions with **infinitely many moments**. Thus, the failure of the t-test is not special to the t-test; in this setting, there simply exist no **"reasonable"** tests.

- **Problem**: the mean $\mu(P)$ is quite sensitive for the tails of P, and one sample yields little information about the tails.

- **Not all hope is lost**: the t-test does satisfy (3) for certain large classes of distributions that are somewhat smaller than $\mathbf{P}$ in the theorem. We will discuss this next class.

QUESTIONS?

Extension of the Result by Bahadur-Savage

- We now generalize the result of Bahadur and Savage following Romano (2004), by providing a constructive **sufficient condition** that applies to other testing problems as well.

- The idea of the Theorem was key in proving results about the testability of completeness conditions in non-parametric models with endogeneity, see Canay, Santos, and Shaikh (2013).

- Suppose data X are observed on a sample space S with probability law P. A model is assumed only in the sense that P is known to belong to $\mathbf{P}$, some family of distributions on S.

- **Testing Problem**: $H_0 : P \in \mathbf{P}_0$ versus the alternative hypothesis $H_1 : P \in \mathbf{P}_1 = \mathbf{P} \setminus \mathbf{P}_0$.

- A convenient way to discuss the non-existence of tests with good power properties is in terms of the **total variation metric**, defined by

$$\tau(P,Q) \equiv \sup_{\{g:|g|\leqslant 1\}} \left| \int g dQ - \int g dP \right| . \quad (4)$$

Conditions

Condition (A)

For every $Q \in \mathbf{P}_1$ there exists a sequence $P_k \in \mathbf{P}_0$ such that $\tau(Q, P_k) \to 0$ as $k \to \infty$.

Condition A asserts that $\mathbf{P}_0$ is dense in $\mathbf{P}$ with respect to the metric τ. In some settings assuming the following (stronger) conditions simplifies the arguments.

Condition (B)

For every $Q \in \mathbf{P}_1$ and any $\epsilon > 0$, there exists a subset $A = A_\epsilon$ of S satisfying $Q(A_\epsilon) \geqslant 1 - \epsilon$ and such that, if X has distribution Q, the conditional distribution of X given $X \in A_\epsilon$ is a distribution in $\mathbf{P}_0$.

We now prove, under conditions (A) or (B), that no test has power against Q greater than the size of the test.

THEOREM

Let $\phi_n(X)$ be any test of $\mathbf{P}_0$ versus $\mathbf{P}_1$.

(I) *If Condition (A) holds, then*

$$\sup_{Q \in \mathbf{P}_1} E_Q[\phi_n(X)] \leqslant \sup_{P \in \mathbf{P}_0} E_P[\phi_n(X)]. \tag{5}$$

Hence, if ϕ_n has size α, then

$$\sup_{Q \in \mathbf{P}_1} E_Q[\phi_n(X)] \leqslant \alpha; \tag{6}$$

that is, the power function if bounded by α.

(II) *Assume Condition (B) holds. Then Condition (A) holds and therefore* (5) *and* (6) *hold as well.*

Proof of the Theorem: part (i)

Proof of the Theorem: part (ii)

Let $\epsilon_k \to 0$ and A_{ϵ_k} be as in (B). Let P_k be the dist. of X given $X \in A_{\epsilon_k}$ when $X \sim Q$.

Remarks

- **Note**: the hypothesis testing framework does not have to be cast in terms of testing a particular parameter as $\mathbf{P}_0$ and $\mathbf{P}_1$ are quite general.
- **Important**: Condition (B), while stronger than condition (A), is easily verified is some novel examples.
- When $X_1, \ldots, X_n$ is a vector of i.i.d. random variables, then it suffices to verify condition (B) for $n = 1$.

  To produce the set A_ϵ, for X, simply take n-fold product set A_δ obtained from the case $n = 1$, where δ is taken small enough to guarantee with probability $1 - \epsilon$ that all n observations fall in A_δ.

  The chance that all observations fall in A_δ is at least $(1 - \delta)^n$. Thus, choose δ no bigger than $1 - (1 - \epsilon)^{1/n}$ - same trick we used in the Example!

Finite versus not finite mean

Example (Finite versus not finite mean)

- Let $X = (X_1, \ldots, X_n)$ be n i.i.d. observations on the real line.
- Bahadur and Savage (1956) conjectured that tests of the existence of $\mu = E[X]$ suffer from the previous problem.
- Let $\mathbf{P}_0$ be the family of distributions on the real line with a finite mean, and let $\mathbf{P}_1$ be the distributions without a finite mean.
- **Result**: Condition (B) readily holds.
- **Proof**: Let Q be a distribution without a mean. Given ϵ, let $A_\epsilon \in \mathbf{R}$ be any bounded subset st

$$Q(A_\epsilon) \geqslant 1 - \epsilon \; .$$

  Note: if $X \sim Q$ then $X | X \in A_\epsilon$ has a distribution in $\mathbf{P}_0$ (as such a rv has support on a bounded set). Hence, (B) holds, the conclusion of theorem holds, and it is impossible to construct a test with power greater than the size of the test.

Introducing Condition (C)

- For a real-valued parameter θ the impossibility of testing a hypothesis like $H_0 : \theta \neq \theta_0$ versus $H_1 : \theta = \theta_0$ is well known.
- **Generally**: impossible to test $H_0 : P \in \mathbf{P}_0$ versus $H_1 : P \in \mathbf{P}_1$ when $\mathbf{P}_0$ is dense in $\mathbf{P} = \mathbf{P}_1 \cup \mathbf{P}_0$.
- **Goodness-of-fit**: impossible to conclude that the underlying distribution is normal, or any other family that falls in a lower dimensional subspace of the a priori model space.
- To make this precise, consider the following condition.

Condition (C)

For any $Q \in \mathbf{P}_1$ and any $\epsilon > 0$, there exists some distribution R such that $(1-\epsilon)Q + \epsilon R \in \mathbf{P}_0$.

Condition (C) implies Condition (A)

- Pick $Q \in \mathbf{P}_1$ and let $\epsilon_k > 0$. Flip a coin with probability $1 - \epsilon_k$ of heads, and let
$$A_{\epsilon_k} = \{\text{ the toss is a head }\}.$$

- Let $Y(\omega)$ be a random variable (on some probability space) that has distribution Q conditional on $\omega \in A_{\epsilon_k}$, and has distribution R condition on $\omega \in A^c_{\epsilon_k}$, for some distribution R.

- **Condition (C)**: for any $Q \in \mathbf{P}_1$ and $\epsilon_k > 0$, $\exists\, Y$ with distribution $P_k = (1-\epsilon_k)Q + \epsilon_k R \in \mathbf{P}_0$ and a subset $A = A_{\epsilon_k}$, with $P_k(A) \geqslant 1 - \epsilon_k$ such that the conditional distribution of $Y|A \sim Q$.

  Then, for any g such that $|g| \leqslant 1$,
$$\begin{aligned} E_{P_k}[g(Y)] &= E_{P_k}[g(Y)|A_{\epsilon_k}]P_k(A_{\epsilon_k}) + E_{P_k}[g(Y)|A^c_{\epsilon_k}]P_k(A^c_{\epsilon_k}), \\ &\leqslant E_Q[g(Y)] + P_k(A^c_{\epsilon_k}), \\ &\leqslant E_Q[g(Y)] + \epsilon_k\,. \end{aligned}$$

  Similarly,
$$E_{P_k}[g(Y)] \geqslant E_Q[g(Y)](1-\epsilon_k) - \epsilon_k \geqslant E_Q[g(Y)] - 2\epsilon_k\,.$$

  Hence, $\tau(Q, P_k) \leqslant 2\epsilon_k$ and Condition (A) follows by letting $\epsilon_k \to 0$.

EXAMPLE (GOODNESS-OF-FIT TESTING)

- The usual approach to testing goodness-of-fit runs as follows.
- Assume $X_1, \ldots, X_n$ are i.i.d. S-valued random variables with distribution P.
- The null hypothesis asserts P belongs to some class $\{P_\theta : \theta \in \Theta\}$ and the alternative hypothesis asserts $P \in \bar{\mathbf{P}}$, the family of all other distributions on S.
- Reversing the roles of the null and alternative is **not possible**.
- **Example**: consider the problem of testing uniformity on $S = (0, 1)$.
- **Condition (C)**: $\mathbf{P}_1$ consists of U, the uniform distribution on $(0, 1)$. Then, for any other distribution R and any $\epsilon > 0$, $(1 - \epsilon)U + \epsilon R$ is not U, and so condition (C) holds.
- **Other examples**: Specification testing (condition (C)), Vishal's paper on RDD (condition (A))

THE END