Econ 2110, fall 2016, Part IVa
Foundations of Asymptotic Statistics

Maximilian Kasy

Department of Economics, Harvard University

Motivating question

- The world is complicated
  – things don't follow simple parametric models
- Can we still say something general
  about the behavior of statistical procedures?
- Idea of asymptotics:
  in large samples we can
- We approximate behavior of procedures by some limit

Takeaways for this part of class

- How we get our formulas for standard deviations in many settings.
- When and why we can expect asymptotic normality for many estimators (and what that means).
- When we might expect problems to arise for asymptotic approximations.

Textbook

van der Vaart, A. (2000). *Asymptotic statistics*.
Cambridge University Press.

- Part IVa:
  sections 2.1 and 2.2.
- Part IVb:
  sections 3.1, 5.1, 5.2, 5.3, and 5.5.

Roadmap

- IVa
  - Types of convergence
  - Laws of large numbers (LLN) and central limit theorems (CLT)
- IVb
  - The delta method
  - M- and Z-Estimators
  - Special M-Estimators
    - Ordinary least squares (OLS)
    - Maximum likelihood estimation (MLE)
  - Confidence sets

Part IVa

Types of convergence

Laws of large numbers (LLN) and central limit theorems (CLT)

Types of convergence

- Recall convergence of non-stochastic sequences:
- $x_n \to x$
- if and only if:

$$\forall \varepsilon > 0 \, \exists N \, \forall n > N \, \|x_n - x\| < \varepsilon.$$

- For sequences of random variables, there is more than one notion of convergence.

4 types of convergence

1. almost sure convergence
2. convergence in probability
3. convergence in mean,
   convergence in mean squared
4. convergence in distribution

Almost sure convergence

- $X_n \rightarrow^{a.s.} X$
- if and only if

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$$

- We will not use almost sure convergence much

Convergence in probability

- $X_n \to^p X$
- if and only if

$$P(\|X_n - X\| < \varepsilon) \to 1 \quad \forall \varepsilon > 0.$$

- convergence almost surely implies convergence in probability
- the reverse is not true

Convergence in mean and in mean squared

- $X_n \rightarrow^m X$
- if and only if
$$E[\|X_n - X\|] \rightarrow 0.$$
- $X_n \rightarrow^{m.s.} X$
- if and only if
$$E[\|X_n - X\|^2] \rightarrow 0.$$
- can define similar notion for rth mean.

Practice problem

Show that convergence in mean squared implies convergence in mean.

Hint: use Jensen's inequality!

Practice problem

Show that convergence in mean implies convergence in probability.

Hint: use Markov's inequality!

Sketch of solutions:

1. Jensen's inequality for $h(x) = x^2$:

$$E[\|X_n - X\|]^2 \leq E[\|X_n - X\|^2]$$

2. Markov's inequality:

$$P(\|X_n - X\| > \varepsilon) \leq \frac{E[\|X_n - X\|]}{\varepsilon}$$

- Easiest way to show convergence in probability: show convergence in mean squared
- Let X_n, X be random variables,

$$v_n = \mathrm{Var}(X_n - X)$$
$$b_n = E[X_n] - E[X]$$

- Then
$$E[|X_n - X|^2] = v_n + b_n^2.$$
- Convergence in probability therefore follows if

$$v_n \to 0 \text{ and}$$
$$b_n \to 0.$$

Practice problem

Show that the “variance / bias decomposition” holds.

Convergence in distribution

- $X_n \to^d X$
- if and only if for all continuity points of F_X

$$F_{X_n}(x) \to F_X(x).$$

- convergence in probability implies convergence in distribution
- the reverse is not true
- except when X is non-random

Practice problem

1. Let $F_{X_n}(x) = \mathbf{1}(x \geq 1/n)$,
   $F_X(x) = \mathbf{1}(x \geq 0)$
   Does $X_n \to^d X$?
2. Let $Y \sim N(0,1)$,
   $Y_n = (-1)^n \cdot Y$
   Does $Y_n \to^d Y$?
   How about $Y_n \to^p Y$?

Relationship between convergence concepts

$$\begin{array}{ccc}
\text{almost surely} & & \text{in mean squared} \\
\Downarrow & & \Downarrow \\
\text{in probability} & \Longleftarrow & \text{in mean} \\
\Downarrow & & \\
\text{in distribution} & &
\end{array}$$

Theorem (Slutsky's theorem)

- Let c be a constant,
- suppose $X_n \rightarrow^d X$ and $Y_n \rightarrow^p c$
- then
  1. $X_n + Y_n \rightarrow^d X + c$
  2. $X_n Y_n \rightarrow^d Xc$
  3. $X_n / Y_n \rightarrow^d X / c$, provided $c \neq 0$.
- In particular, if $X_n \rightarrow^d X$ and $Y_n \rightarrow^p 0$, then $X_n Y_n \rightarrow^p 0$.

Theorem (Continuous Mapping Theorem (CMT))

- Let g be a continuous function
- If $X_n \to^d X$, then $g(X_n) \to^d g(X)$.
- If $X_n \to^p X$, then $g(X_n) \ \to^p g(X)$.

Example:
Suppose $X_n \to^d \mathcal{N}(0,1)$. Then $X_n^2 \to^d \chi_1^2$.

O_p and o_p Notation

- Let a_n and b_n be two sequences of real numbers. Recall:
- $a_n = o(b_n)$ means that:
  $a_n/b_n \to 0$, and
- $a_n = O(b_n)$ means that:
  there exists a number M
  such that $|a_n/b_n| < M$ for all n.

- similar notation if $a_n = X_n$ is a sequence of random variables:
- $X_n = o_p(b_n)$ means that:
  $X_n / b_n \to^p 0$.
- $X_n = O_p(b_n)$ means that:
  for all $\varepsilon > 0$ there exists a number M such that
  $P(|X_n / b_n| < M) > 1 - \varepsilon$ for all n.

Example

- Let $X_n \sim^{iid} \mathcal{N}(0, n)$.
- Then $X_n = O_p(n^{1/2})$
  - since $X_n / n^{1/2} \sim^{iid} \mathcal{N}(0, 1)$,
  - and for any $\varepsilon > 0$ one can choose M
  - such that $P(|\mathcal{N}(0, 1)| < M) > 1 - \varepsilon$.
- Also, $X_n = o_p(n)$,
  - since $X_n / n \sim \mathcal{N}(0, n^{-1})$,
  - and for any $\varepsilon > 0$

$$P(|\mathcal{N}(0, n^{-1})| > \varepsilon) = P(|\mathcal{N}(0, 1)| > n^{1/2}\varepsilon) \to 0,$$

  - so that $X_n / n \to^p 0$.

Lemma

If $X_n \to^d X$, then $X_n = O_p(1)$

Proof:

- Since X is a random variable, there exists $M > 0$ such that
  1. F_X is continuous at $-M$ and M, and
  2. $P(|X| > M) = F_X(-M) + (1 - F(M)) < \varepsilon/2$.
- Since $X_n \to^d X$, for all large enough n,
  $|F_{X_n}(-M) - F_X(-M)| < \varepsilon/4$ and
  $|F_{X_n}(M) - F_X(M)| < \varepsilon/4$.
- Hence for all n large enough, $P(|X_n| > M) < \varepsilon$.

Laws of large numbers (LLNs) and central limit theorems (CLTs)

- Two basic building blocks from which we build our asymptotic theory
- **LLNs**:
  - sample averages converge in probability to expectations
  - "weak LLN"
  - actually, they even converge almost surely (strong LLN)
- **CLTs**:
  - sample averages,
  - with normalized expectation and variance,
  - converge in distribution to standard normals

Theorem (A Weak Law of Large Numbers)

- Let $X_1, X_2, \ldots$ be a sequence of random variables with
  1. $E[X_i] = \mu_i$,
  2. $Var[X_i] = \sigma_i^2$, and
  3. $Cov(X_i, X_j) = 0$ for $i \neq j$.
- Let
  1. $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$,
  2. $\overline{\mu}_n = n^{-1} \sum_{i=1}^{n} \mu_i$, and
  3. $\overline{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} \sigma_i^2$, where $\overline{\sigma}_n^2 / n \to 0$.
- Then $\overline{X}_n - \overline{\mu}_n \to^p 0$.

- Special case:
  - If X_i are iid., $\text{Var}(X_i) < \infty$
  - Then $\overline{X}_n \to^p E[X]$

Practice problem

Verify that the special case satisfies the assumptions of the theorem.

Practice problem

Prove the theorem.

Hint: Use Markov's inequality

Proof:

1. $E[\overline{X}_n - \overline{\mu}_n] = 0$
2.

$$\begin{aligned}\mathrm{Var}(\overline{X}_n - \overline{\mu}_n) = \mathrm{Var}(\overline{X}_n) &= \frac{1}{n^2}\sum_{ij}\mathrm{Cov}(X_1, X_j) \\ &= \frac{1}{n^2}\sum_i \sigma_i^2 = \overline{\sigma}_n^2/n \to 0.\end{aligned}$$

3. This implies $\overline{X}_n - \overline{\mu}_n \to^{ms} 0$.
4. And thus $\overline{X}_n - \overline{\mu}_n \to^{p} 0$.

Theorem (A Central Limit Theorem)

- Let $X_1, X_2, \dots$ be a sequence of iid random variables with
  1. $E[X_i] = \mu$,
  2. $Var(X_i) = \sigma^2$,
  3. and $0 < \sigma^2 < \infty$.
- Let $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$.
- Then

$$\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu) \to^d \mathcal{N}(0,1).$$

Some remarks

- Taking limits is just a thought experiment
- It's a way to get an approximation for what's happening in a given, finite sample
- Advantage: things get simpler
  eg.: no matter what distribution X_i has, $\overline{X}$ is approximately normal
- Disadvantage: Sometimes the approximation is bad
- There is always more than one way to take a limit