Probability and Convergence

Robert A. Miller

Structural Econometrics

November 2024

- Let Ω denote a collection of events and $\phi \subseteq \Omega$ the empty set.
- Let $N \equiv \{1, 2, \ldots\}$ denote the counting numbers.
- Define $A^c \subseteq \Omega$, the *complement* of $A \subseteq \Omega$ such that:
  - $A \bigcap A^c = \phi$ and $A \bigcup A^c = \Omega$
- Let $B \setminus A$denote the *set difference* of subtracting A from B:
  - $B \setminus A \equiv (A \bigcap B)^c \bigcap B$
- A countable collection of events $\{A_n\}_{n \in N}$ *partitions* Ω iff:
  - $A_n \bigcap A_{n'} = \phi$ for all $A_n \in \Omega$ and $A_{n'} \in \Omega$
  - $\bigcup_{n \in N} A_n = \Omega$
- A σ-algebra on Ω, denoted by $\mathcal{F}$, is a collection of subsets containing Ω, and closed under:
  - complements: $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
  - countable unions: $A_n \in \mathcal{F}$ for all $n \in N \implies \bigcup_{n \in N} A_n \in \mathcal{F}$

- A *probability measure* $\mathbb{P}$ is a real valued function, defined on $A \subseteq \Omega$ satisfying the following axioms:
    1. $\mathbb{P}(A) \geq 0$ for every event $A \subseteq \Omega$.
    2. $\mathbb{P}(\Omega) = 1$
    3. If $A_m \bigcap A_n = \phi$ for all $n \neq m$ then $\mathbb{P}\left(\bigcup_{n \in N} A_n\right) = \sum_{n \in N} \mathbb{P}(A_n)$.
- The first two axioms essentially bound how events are weighed.
- In this way we avoid the conundrum of adding infinities to each other or subtracting infinity from infinity.
- Aside from that, the choices of "0" and "1" in the first two axioms is for convenience.
- The third axiom, known as *countable additivity*, helps define sums.
- Note the interval $[0, 1]$ has length one but is the union of an uncountable of points $r \in [0, 1]$, each point having length zero.

- Given $(\Omega, \mathcal{F}, P)$ suppose $B \in \mathcal{F}$ and $P(B) \neq 0$.
- Then for all $A \in \mathcal{F}$ we define the probability of A conditional on B as:

$$P(A|B) \equiv P(A \cap B) / P(B)$$

- More generally, setting $B = \bigcap_{n \in N} A_n$ we obtain:

$$P\left(\bigcap_{n \in N} A_n\right) \equiv P(A_1) \prod_{n=2}^{N} P\left(A_n \middle| \bigcap_{k=1}^{n-1} A_k\right)$$

- We say A and B are independent events iff:

$$P(A \cap B) = P(A) P(B) \text{ or equivalently } P(A|B) = P(A)$$

More generally:

$$P\left(\bigcap_{n \in N} A_n\right) = \prod_{n \in N} P(A_n)$$

- Denote the elements of Ω by $\omega \in \Omega$.
- Note, depending on how $\mathcal{F}$ is defined, perhaps $\omega \notin \mathcal{F}$.
- Real valued, measurable functions of Ω are called random variables.
- That is $x(\omega) : \Omega \to \mathbb{R}$, and for all Borel sets $B \subseteq \mathcal{B}$:

$$\{\omega : x(\omega) \in B\} \in \mathcal{F}$$

- We write the conditional expectation of $x(\omega)$ as:

$$\mathrm{E}(x \,|B) = \frac{1}{P(B)} \int 1\{\omega : x(\omega) \in B\} \, x(\omega) \, \mathrm{d}P$$

- If $x(\omega) : \Omega \to \mathbb{R}^\infty$ denoted by $x(\omega) = (x_1(\omega), x_2(\omega), \ldots)$ is measurable with respect to $\mathcal{B}_\infty$, then:

$$\begin{aligned} P(x(\omega)) &= P\left(\bigcap_{n=1}^{\infty} x_n(\omega)\right) \\ &= P(x_1(\omega)) \prod_{n=2}^{\infty} P(x_n(\omega) \mid x_1(\omega), \ldots, x_{n-1}(\omega)) \end{aligned}$$

Probability Measure

Independence and identically distributed random variables

- We say $x(\omega)$ is independent iff:

$$P(x(\omega)) = \prod_{n=1}^{\infty} P(x_n(\omega))$$

- Supposing $x(\omega)$ is independent, then it is identically distributed as well if:

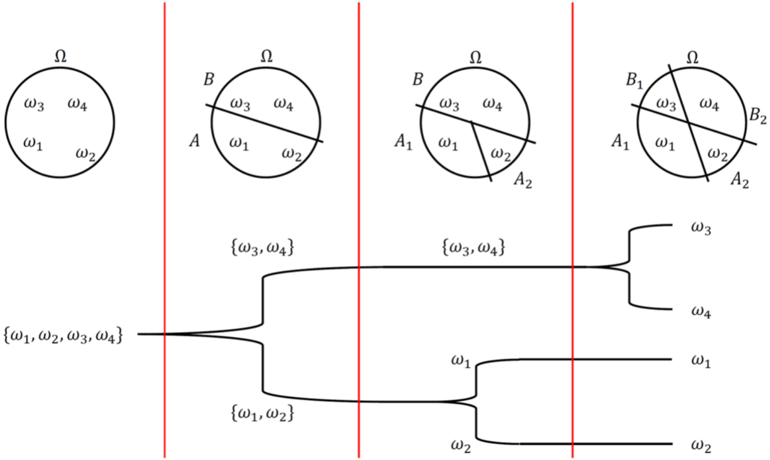
$$P(x_n(\omega)) = P(x_m(\omega))$$

for all n, m, ω.

- For each $N = 1, 2, \ldots$, let $(\Omega, \mathcal{F}_N, P)$ denote a probability space, where we now interpret:
  - $\omega \in \Omega$: a world history, or an ordering of the population, from beginning to end.
  - N : sample size showing how far history (many individuals) has progressed (have been interviewed).
  - $\mathcal{F}_N$: set of all possible (partial) histories showing the distinguishing features of ω that can be observed from sample of size N. By construction $\mathcal{F}_N \subseteq \mathcal{F}_{N+1} \subseteq \ldots$.
  - P : probability measure over possible (world) histories.
  - $\theta \in \Theta$: unknown parameters of interest defining P.
  - $h_N(\omega; \theta) : \Omega \times \Theta \to \mathbb{R}^q$: a function of the sample, measurable with respect to $(\Omega, \mathcal{F}_N, P)$.

Statistical Inference

Illustrating histories to show how information sets evolve



- Consider a statistic:

$$\theta\left(x^{(N)}\right) \equiv \theta_N$$

- On basis of this statistic, suppose we want to estimate θ_0, the parameters defining the data generating process.
- Ideally this involves deriving the probability $\theta_N \leq \theta$ when θ_0 generated the data:

$$\Pr\left\{\theta_N \leq \theta; \theta_0\right\} \qquad \forall\theta$$

- Continuing this wish list, we might hope to form a θ_N such that $\Pr\left\{\theta_N \leq \theta; \theta_0\right\}$ is a step function with a single step at θ_0.
- We often focus on the first two moments:

$$E\left[\theta_N; \theta_0\right] \text{ and } \operatorname{var}\left(\theta_N; \theta_0\right)$$

- This approach is usually intractable for nonlinear models.

- Rather than ask what is the probability distribution of θ_N given θ_0 we ask some less demanding, and less revealing, questions about the limiting properties of θ_N as $N \to \infty$.
- **First question:** For some definition of "$\longrightarrow$", is there a sense in which:

$$\theta_N \text{``} \longrightarrow \text{''}\ \theta_0?$$

- **Second question:** Is there a sense in which how quickly θ_N "$\longrightarrow$" θ_0? We might ask for what values of $\alpha > 0$ does:

$$N^{\alpha} \left(\theta_N - \theta_0 \right) \text{``} \longrightarrow \text{''}\ 0?$$

- **Third question:** For the "biggest" α^* almost satisfying $\mu_N^{\alpha} \equiv N^{\alpha} \left(\theta_N - \theta_0 \right)$ "$\longrightarrow$" 0, let $G_N \left(\mu_N^{\alpha^*}; \theta_0 \right)$ denote the probability distribution function of $\mu_N^{\alpha^*}$ given θ_0 and conditional on $\left\{ x_n \right\}_{n=1}^{N}$. Is there a probablity distribution $G \left(\mu_N^{\alpha^*}; \theta_0 \right)$ such that:

$$G_N \left(\mu_N^{\alpha^*}; \theta_0 \right)\ \text{``} \longrightarrow \text{''}\ G \left(\mu_N^{\alpha^*}; \theta_0 \right)?$$

Convergence

Exception sets

- For $(\Omega, \mathcal{F}, P)$, a probability space induced by θ_0, and $\omega \in \Omega$, a history, define an exception set as:

$$A_{Nr} = \left\{ \omega : \left| \theta_N \left(\omega \right) - \theta_0 \right| > \frac{1}{r} \right\}$$

- Thus A_{Nr} depicts for which histories it is true that our estimator $\theta_N \left(\omega \right)$ differs from θ by more than than $\frac{1}{r}$.
- Note that $A_{Nr} \subseteq A_{N,r+1} \subseteq \cdots$
- Convergence is based on the notion that A_{Nr} shrinks as N increases.

- Convergence in probability means the probability of an exception set occurring converges to zero.
- We denote convergence in probability by:

$$x_N \xrightarrow{p} x \qquad (\text{or } \operatorname{plim} x_N = x)$$

- Formally $x_N \xrightarrow{p} x$ if and only if for all $r \in \{1, 2, \ldots\}$:

$$\begin{aligned} 0 &= \lim_{N\to\infty} \Pr\left(A_{Nr}\right) \\ &= \lim_{N\to\infty} \Pr\left\{\omega : \left|x_N\left(\omega\right) - x\left(\omega\right)\right| > \frac{1}{r}\right\} \qquad \forall r = 1, 2, \ldots \end{aligned}$$

- That is for each $(r, \delta) \in \{1, 2, \ldots\} \times \mathbb{R}$ there exists some N^* such that for all $N > N^*$:

$$\Pr\left\{\omega : \left|x_N\left(\omega\right) - x\left(\omega\right)\right| > \frac{1}{r}\right\} \equiv \Pr\left(A_{Nr}\right) < \delta$$

- Let:

$$\begin{aligned}\Omega &= [0,1], \\ \mathcal{F} &\equiv \mathcal{B}[0,1] \text{ for the Borel sets, and} \\ \mathcal{L} &\equiv \Pr\{\omega \in [0,a]\} = a \qquad \text{on } [0,1]\end{aligned}$$

- Let $\delta_1, \delta_2, \ldots$ be a sequence of real numbers and:

$$x_N(\omega) \equiv \begin{cases} 1 & \text{if } \omega \leq \delta_N \\ 0 & \text{if } \omega > \delta_N. \end{cases}$$

- Then

$$x_N(\omega) \xrightarrow{p} 0 \quad \text{iff } \delta_N \to 0.$$

- Now suppose $x_N(\omega)$ and $x(\omega)$ are K dimensional vector of random variables expressed as:

$$\begin{aligned} x_N(\omega) &\equiv (x_{1N}(\omega), \ldots, x_{KN}(\omega)) \\ x(\omega) &\equiv (x_1(\omega), \ldots, x_K(\omega)) \end{aligned}$$

- We say $x_N(\omega) \xrightarrow{p} x(\omega)$ iff:

$$\begin{aligned} & \lim_{N\to\infty} \Pr\left\{\omega : \|x_N(\omega) - x(\omega)\| > \frac{1}{r}\right\} \\ \equiv \; & \lim_{N\to\infty} \Pr\left\{\omega : \textstyle\sum_{k=1}^{K} [x_{kN}(\omega) - x_k(\omega)]^2 > r^{-2}\right\} \\ = \; & 0 \end{aligned}$$

- Note that $x_N(\omega) \xrightarrow{p} x(\omega)$ iff $x_{kN}(\omega) \xrightarrow{p} x_k(\omega)$ for all k.

- Consider subsets $A_N \subset \Omega$ for $N = 1, 2, \ldots$ and define:

$$\limsup_N A_N \equiv \bigcap_{N=1}^{\infty} \bigcup_{K=N}^{\infty} A_K \text{ and } \liminf_N A_N \equiv \bigcup_{N=1}^{\infty} \bigcap_{K=N}^{\infty} A_K$$

- This leads us to focus on:

$$A_r^* \equiv \limsup_N A_{Nr} = \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \left\{ \omega : |x_k(\omega) - x(\omega)| > \frac{1}{r} \right\}$$

- We say $x_N \stackrel{a.s.}{\rightarrow} x$ (or x_N converges to x with probability 1) if and only if:

$$0 = \Pr\left(\bigcup_{r=1}^{\infty} A_r^* \right) = \Pr\left(\bigcup_{r=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \left\{ \omega : |x_k(\omega) - x(\omega)| > \frac{1}{r} \right\} \right)$$

- We say x_N converges to x in L_p for $p = \{1, 2, \ldots\}$ or $x_N \xrightarrow{L_p} x$ iff:

$$\lim_{N\to\infty} \mathrm{E}\left[|x_N - x|^p\right] = 0$$

- If $p = 2$ we say "convergence in quadratic mean" or "convergence in mean square".

- Chebychev's inequality is a useful tool for understanding how the relationship between these convergence concepts.
- It states: If $\varphi : \mathbb{R} \to \mathbb{R}$ is positive, increasing on $(0, \infty)$, symmetric $(\varphi(u) = \varphi(-u))$, and x is a random variable on $(\Omega, \mathcal{F}, P)$, then for all $u > 0$:

$$\varphi(u) P(|x| \geq u) \leq \mathrm{E}(\varphi(x))$$

- To prove Chebychev's inequality, note that:

$$\begin{aligned}
\mathrm{E}(\varphi(x)) &= \int_\Omega \varphi(x(\omega))\, dP \\
&\geq \int_\Omega 1_{\{|x| \geq u\}}(\omega)\, \varphi(x(\omega))\, dP \\
&\geq \int_\Omega 1_{\{|x| \geq u\}}(\omega)\, \varphi(u)\, dP \\
&= \varphi(u) P(|x| \geq u)
\end{aligned}$$

- To illustrate Chebychev's inequality set:

$$\varphi(x) = x^2 \quad \text{and} \quad u = \sigma$$

to obtain

$$\Pr(|x| \geq \sigma) \leq \frac{\mathrm{E}(x^2)}{\sigma^2}$$

Lemma

If $x_N \xrightarrow{L_p} x$ then $x_N \xrightarrow{p} x$.

How are these Convergence Concepts Related?

Proof.

For $p > 0$ set:

$$\varphi\left(u\right) = \left|u\right|^{p} \text{ and } u = \frac{1}{r}$$

Since $\varphi\left(u\right)$ is symmetric and increasing, we can apply Chebychev's inequality to obtain:

$$\Pr\left(\left|x_N - x\right| \geq \frac{1}{r}\right) \leq \mathrm{E}\left[\left|x_N - x\right|^{p}\right] r^{p}$$

By hypothesis:

$$\lim_{N\rightarrow\infty} \mathrm{E}\left[\left|x_N - x\right|^{p}\right] = 0$$

so for any $p > 0$ and $r \in \{1, 2, \ldots\}$:

$$\lim_{N} \Pr\left(\left|x_N - x\right| \geq \frac{1}{r}\right) = 0$$

How are these Convergence Concepts Related?

Lemma

$x_N \stackrel{a.s.}{\longrightarrow} x$ does not imply $x_N \stackrel{L_p}{\longrightarrow} x$.

How are these Convergence Concepts Related?

Proof.

Let $(\Omega, \mathcal{F}, P) \equiv \{[0,1], \mathcal{B}[0,1], \mathcal{L}\}$ and define:

$$x_N(\omega) = \begin{cases} 2^N \text{ if } \omega \in (0, 1/N) \\ 0 \text{ otherwise} \end{cases}$$

$$\begin{aligned} \text{Then } \Pr\left(\bigcup_{r=1}^{\infty} A_r^*\right) &= \Pr\left(\bigcup_{r=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\omega : x_k(\omega) > 1/r\}\right) \\ &\leq \Pr\left(\bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\omega : x_k(\omega) \neq 0\}\right) \\ &= \Pr\left[\bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} \{\omega \in (0, 1/N)\}\right] \\ &= \Pr\left[\bigcap_{N=1}^{\infty} \{\omega \in (0, 1/N)\}\right] \\ &= 0 \end{aligned}$$

Therefore $x_N \overset{a.s}{\to} 0$. But $E\left[|x_N|^p\right] = \left(2^N/N\right)^p \to \infty$ so $x_N \overset{L_p}{\nrightarrow} 0$.

□

How are these Convergence Concepts Related?

Almost sure convergence implies convergence in probability

Theorem

(Theorem 18.3 J. Davidson, Stochastic limit theory, p. 283): $x_N \xrightarrow{a.s.} x$ *if and only if for all* $\varepsilon > 0$*:*

$$\lim_{m\to\infty} \Pr\left[\sup_{N\geq m} |x_N - x| \leq \varepsilon\right] = 1$$

Lemma

$$x_N \xrightarrow{a.s} x \quad \Longrightarrow \quad x_N \xrightarrow{p} x$$

Proof.

$$
\begin{aligned}
& \sup_{N\geq m} |x_N - x| \geq |x_m - x| \\
\Rightarrow\quad & \Pr\left[\sup_{N\geq m} |x_N - x| \leq \varepsilon\right] \leq \Pr\left[|x_m - x| \leq \varepsilon\right] \\
\Rightarrow\quad & \lim_{m\to\infty} \Pr\left[\sup_{N\geq m} |x_N - x| \leq \varepsilon\right] = 1 \leq \lim_{m\to\infty} \Pr\left[|x_m - x| \leq \varepsilon\right] \\
\Rightarrow\quad & \lim_{m\to\infty} \Pr\left[|x_m - x| \leq \varepsilon\right] = 1
\end{aligned}
$$

But:

$$
\begin{aligned}
& \lim_{m\to\infty} \Pr\left[|x_m - x| \leq \varepsilon\right] = 1 \\
\Rightarrow\quad & \lim_{m\to\infty} \Pr\left[|x_m - x| > \varepsilon\right] = 0.
\end{aligned}
$$

How are these Convergence Concepts Related?

Lemma

$$x_N \overset{L_p}{\rightarrow} x \textit{ does not imply } x_N \overset{a.s.}{\rightarrow} x$$

Proof: Let $(\Omega, \mathcal{F}, P) \equiv \{[0,1], \mathcal{B}[0,1], \mathcal{L}\}$ and define $x_N(\omega)$ as:

$$\begin{aligned} x_1(\omega) &= 1_{[0,1/2]}(\omega), \ x_2(\omega) = 1_{[1/2,1]}(\omega) \\ x_3(\omega) &= 1_{[0,1/3]}(\omega), \ x_4(\omega) = 1_{[1/3,2/3]}(\omega), \ x_5(\omega) = 1_{[2/3,1]}(\omega), \ \ldots \end{aligned}$$

Relative to $x(\omega) = 0$, for all p:

$$\mathrm{E}(|x_N|^p) = \int_\Omega 1_{A_N(\omega)} dP = \frac{1}{k+1} \text{ if } \frac{k(k+1)}{2} \leq N \leq \frac{(k+1)(k+2)}{2}$$

Therefore $x_N \overset{L_p}{\rightarrow} 0$ for all p because $(k+1)^{-1} \rightarrow 0$ as $N \rightarrow \infty$.
But $\forall \omega \in \Omega$ and $N \in \{1, 2, \ldots\}$ there exists k such that $x_k(\omega) = 1$:

$$\bigcup_{r=1}^{\infty} A_r^* = \bigcup_{r=1}^{\infty} \lim_{N\rightarrow\infty} \bigcup_{k=N}^{\infty} \left\{\omega : x_k(\omega) > \frac{1}{r}\right\} = \bigcup_{r=1}^{\infty} \{\omega : \omega \in \Omega\} = \Omega$$

How are these Convergence Concepts Related?

Lemma

Suppose

$$|x_N(\omega)| \leq y(\omega) \in L^p$$

If $x_N(\omega) \xrightarrow{p} 0$, *then* $x_N(\omega) \xrightarrow{L^p} 0$.

Proof.

See Chung (1974, page 67): *A Course in Probability Theory* □

- Let $\{y_N\}_{N=1}^{\infty}$ denote a sequence of real numbers, and $\{g_N\}_{N=1}^{\infty}$ a sequence of positive real numbers.
- We say y_N is at most of order g_N, and write $y_N = O(g_N)$ if there exists a real number M such that for all N:

$$\frac{|y_N|}{g_N} \leq M$$

- We say y_N is of smaller order than g_N and write $y_N = o(g_N)$ if and only if:

$$\frac{y_N}{g_N} \to 0.$$

  That is for any $\varepsilon > 0$ there exists an N_ε such that $\frac{|y_N|}{g_N} < \varepsilon$ for all $N \in \{N_\varepsilon + 1, N_\varepsilon + 2, \ldots\}$.

- Now let x_N be a random variable with respect to $(\Omega, \mathcal{F}, P)$.
- We say x_N is at most of order y_N writing:

$$x_N = O_p\left(g_N\right)$$

if there exists a positive number M_r for all $r \in \{1, 2, \ldots\}$ such that for all N:

$$\Pr\left[\frac{|x_N|}{g_N} \geq M_r\right] \leq \frac{1}{r}$$

- We say that x_N is of smaller order than g_N in probability, writing:

$$x_N = o_p\left(g_N\right)$$

if:

$$\frac{x_N}{g_N} \xrightarrow{p} 0.$$

Order in Magnitude

Second moments imply being bounded in probability

Lemma

Let $x_1, x_2, \ldots$ be a sequence of random variables with respect to $(\Omega, \mathcal{F}, P)$, and $g_1, g_2, \ldots$ a sequence of positive real numbers such that

$$E\left(x_N^2\right) = O\left(g_N^2\right)$$

Then

$$x_N = O_p\left(g_N\right)$$

- For example suppose $x_N \sim \mathcal{N}\left(0, \sigma^2\right)$ that for all N. Then:

$$E\left(x_N^2\right) = \sigma^2 = O\left(1\right)$$

  and by the lemma $x_N = O_p\left(1\right)$.

Proof.

By the premise there exists $M < \infty$, such that:

$$\mathrm{E}\left[x_N^2 \,/ g_N^2\right] \leq M$$

Also by Chebychev's inequality, for all $M_r > 0$:

$$\Pr\left[\frac{|x_N|}{g_N} \geq M_r\right] \leq \frac{\mathrm{E}\left[x_N^2 \,/ g_N^2\right]}{M_r^2}$$

For any r choose M_r to satisfy the inequality:

$$M_r > (Mr)^{1/2} \Rightarrow M_r^2 > Mr \Rightarrow \frac{M}{M_r^2} \leq \frac{1}{r}$$

$$\text{Then } \Pr\left[\frac{|x_N|}{g_N} \geq M_r\right] \leq \frac{\mathrm{E}\left[x_N^2 \,/ g_N^2\right]}{M_r^2} \leq \frac{M}{M_r^2} \leq \frac{1}{r}.$$