

Probability and Convergence

Robert A. Miller

Structural Econometrics

November 2021

Probability Measure

Collections of sets and sigma algebras

- Let Ω denote a collection of events and $\phi \subseteq \Omega$ the empty set.
- Let $N \equiv \{1, 2, \dots\}$ denote the counting numbers.
- Define $A^c \subseteq \Omega$, the *complement* of $A \subseteq \Omega$ such that:
 - $A \cap A^c = \phi$ and $A \cup A^c = \Omega$
- Let $B \setminus A$ denote the *set difference* of subtracting A from B :
 - $B \setminus A \equiv (A \cap B)^c \cap B$
- A countable collection of events $\{A_n\}_{n \in N}$ *partitions* Ω iff:
 - $A_n \cap A_{n'} = \phi$ 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 $n \in N \implies \bigcup_{n \in N} A_n \in \mathcal{F}$

Probability Measure

Kolmogorov Axioms

- 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 $A_m \cap A_n = \emptyset$ for all $n \neq m$ then $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$.
- 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.

Probability Measure

Conditional probability and independence

- Given (Ω, \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:

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

- We say A and B are independent events iff:

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

or more generally:

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

Probability Measure

Random variables

- 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 $x(\omega) : \Omega \rightarrow \mathbb{R}$, 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:

$$E(x|B) = \frac{1}{P(B)} \int 1_{\{\omega : x(\omega) \in B\}} x(\omega) dP$$

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

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

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 .

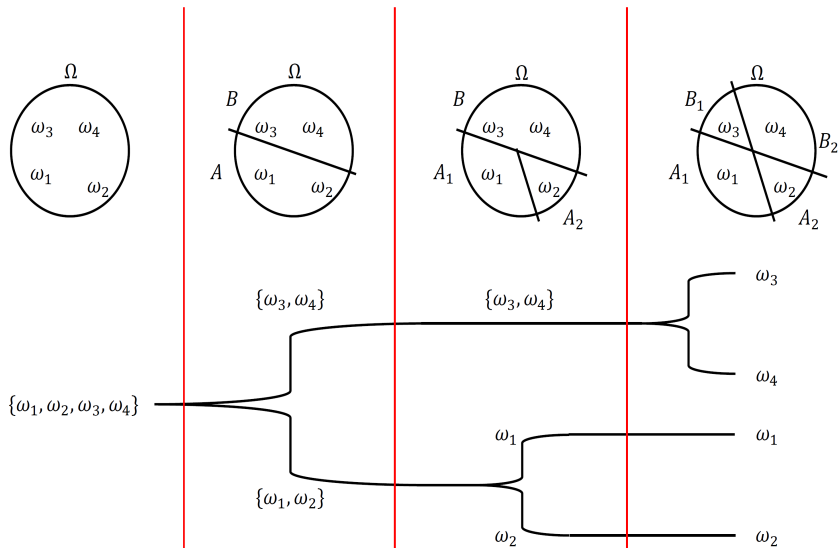
Statistical Inference

Model

- For each $N = 1, 2, \dots$, 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 $\mathcal{F}_N \subseteq \mathcal{F}_{N+1} \subseteq \dots$
 - P : probability measure over possible (world) histories.
 - $\theta \in \Theta$: unknown parameters of interest defining P .
 - $h_N(\omega; \theta) : \Omega \times \Theta \rightarrow \mathbb{R}^q$: a function of the sample, measurable with respect to $(\Omega, \mathcal{F}_N, P)$.

Statistical Inference

Illustrating histories to show how information sets evolve



Statistical Inference

Motivation

- 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:

$$\Pr \{ \theta_N \leq \theta; \theta_0 \} \quad \forall \theta$$

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

$$E [\theta_N; \theta_0] \quad \text{and} \quad \text{var} (\theta_N; \theta_0)$$

- This approach is usually intractable for nonlinear models.

Statistical Inference

Three questions

- 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 $N \rightarrow \infty$.
- **First question:** For some definition of " \rightarrow ", is there a sense in which:

$$\theta_N \text{ " } \rightarrow \text{ " } \theta_0?$$

- **Second question:** Is there a sense in which how quickly $\theta_N \text{ " } \rightarrow \text{ " } \theta_0$? We might ask for what values of $\alpha > 0$ does:

$$N^\alpha (\theta_N - \theta_0) \text{ " } \rightarrow \text{ " } 0?$$

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

$$G_N(\mu_N^{\alpha^*}; \theta_0) \text{ " } \rightarrow \text{ " } G(\mu_N^{\alpha^*}; \theta_0)?$$

Convergence

Exception sets

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

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

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

Convergence

Convergence in probability

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

$$x_N \xrightarrow{P} x \quad (\text{or } \text{plim } x_N = x)$$

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

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

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

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

Convergence

An example of convergence in probability

- Let:

$$\Omega = [0, 1],$$

$$\mathcal{F} \equiv \mathcal{B}[0, 1] \text{ for the Borel sets, and}$$

$$\mathcal{L} \equiv \Pr\{\omega \in [0, a]\} = a \quad \text{on } [0, 1]$$

- Let $\delta_1, \delta_2, \dots$ 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

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

Convergence

Extending convergence in probability to vectors of random variables

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

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

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

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

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

Convergence

Almost sure convergence defined

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

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

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

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

Convergence

L_p convergence

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

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

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

How are these Convergence Concepts Related?

Chebychev's inequality

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

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

- To prove Chebychev's inequality, note that:

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

How are these Convergence Concepts Related?

Two applications of Chebychev's inequality

- To illustrate Chebychev's inequality set:

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

to obtain

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

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:

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

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

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

By hypothesis:

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

so:

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



How are these Convergence Concepts Related?

Lemma

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

How are these Convergence Concepts Related?

Proof.

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

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

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

Therefore $x_N \xrightarrow{a.s.} 0$. But $E[|x_N|^p] = (2^N/N)^p \rightarrow \infty$ so $x_N \not\xrightarrow{L_p} 0$.



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 $\varepsilon > 0$:

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

Lemma

$$x_N \xrightarrow{a.s.} x \implies x_N \xrightarrow{p} x$$

How are these Convergence Concepts Related?

Proof of lemma

Proof.

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

But:

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

How are these Convergence Concepts Related?

Lemma

$$x_N \xrightarrow{L_p} x \text{ does not imply } x_N \xrightarrow{\text{a.s.}} x$$

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

$$x_1(\omega) = 1_{[0, 1/2]}(\omega), \quad x_2(\omega) = 1_{[1/2, 1]}(\omega)$$

$$x_3(\omega) = 1_{[0, 1/3]}(\omega), \quad x_4(\omega) = 1_{[1/3, 2/3]}(\omega), \quad x_5(\omega) = 1_{[2/3, 1]}(\omega), \quad \dots$$

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

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

Therefore $x_N \xrightarrow{L_p} 0$ for all p because $(k+1)^{-1} \rightarrow 0$ as $N \rightarrow \infty$.

But $\forall \omega \in \Omega$ and $N \in \{1, 2, \dots\}$ 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*

Order of Magnitude

Relative magnitude of real sequences

- 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:

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

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

Order of Magnitude

Order in probability

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

$$x_N = O_p(g_N)$$

if there exists a positive number M_r for all $r \in \{1, 2, \dots\}$ 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:

$$x_N = o_p(g_N)$$

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, \dots be a sequence of random variables with respect to (Ω, \mathcal{F}, P) , and g_1, g_2, \dots a sequence of positive real numbers such that

$$E(x_N^2) = O(g_N^2)$$

Then

$$x_N = O_p(g_N)$$

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

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

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

Order in Magnitude

Proving the connection

Proof.

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

$$E [x_N^2 / g_N^2] \leq M$$

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

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

For any r choose M_r to satisfy the inequality:

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

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