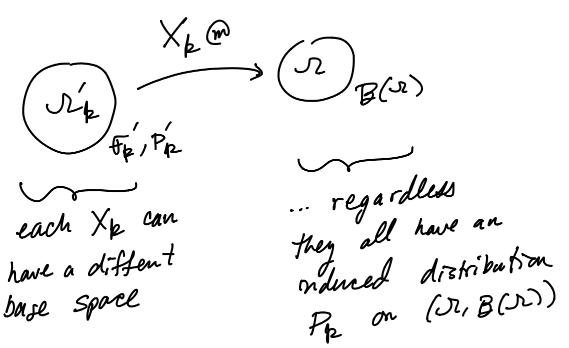
Lecture 14: Convergence in distribution and the Central limit theorem

Convergence in distribution is probably the most important notion of a limit of r.v.s $X_1, X_2, \dots$ or a sequence of probability measures $P_1, P_2, \dots$ on $(\Omega, \mathcal{B}(\Omega))$.

Just as in last lecture we will always assume $(\Omega, \mathcal{B}(\Omega))$ is Polish w.r.t. metric d.

Let $P, P_1, P_2, \dots$ be probability measures on $(\Omega, \mathcal{B}(\Omega))$ and/or $X, X_1, X_2, \dots$ a sequence of ⓜ maps from some prob. space into $(\Omega, \mathcal{B}(\Omega))$



Definition:

$P_n \xrightarrow{\mathcal{D}} P$ iff $\forall f \in C_b(\Omega)$, $\int_\Omega f\, dP_n \longrightarrow \int_\Omega f\, dP$.

$X_n \xrightarrow{\mathcal{D}} X$ iff $\forall f \in C_b(\Omega)$, $E f(X_n) \longrightarrow E f(X)$

Called "convergence in distribution" or "weak convergence".

Remark: This notion of convergence is equiv to weak-$*$ convergence in functional analysis. Its easier to formally see the connection when $P_1, P_2, \dots, P$ have densities $\nu_1, \nu_2, \dots, \nu$ w.r.t. some base measure μ on $(\Omega, \mathcal{B}(\Omega))$, i.e.

$$dP_n = \nu_n d\mu \quad \& \quad dP = \nu d\mu$$

so that $P_n \xrightarrow{\mathcal{D}} P$

$$\Updownarrow$$

$$\int f \nu_n d\mu \longrightarrow \int f \nu d\mu, \ \forall f \in C_b(\Omega)$$

Remark: You can loosely interpret $X_n \xrightarrow{\mathcal{D}} X$ as meaning that for large n, m both X_n and X_m resemble random draws from X but that X_n & X_m are unrelated...

Warning: This is only a loose interpretation since it is possible that $\exists A \in \mathcal{B}(\Omega)$ s.t.

$$P(X_n \in A) \not\to P(X \in A)$$

Most of the examples of $\xrightarrow{\mathcal{D}}$ we will work with come from the central limit theorem ... which effectively says:

If $X_1, X_2, \dots$ are independent r.v.s (all defined on a common $(\Omega', \mathcal{F}', P')$) with

$$E X_n = \mu \quad \& \quad \mathrm{var}(X_n) = \sigma^2 < \infty$$

then $\sqrt{n}\,\bar{X} = \dfrac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N(\mu, \sigma^2)$.

We will derive this near the end of this lecture. However a good (but somewhat degenerate) example which helps interpret future results is as follows.

Example: $\Omega = \mathbb{R}$, $P(X_n = \frac{1}{n}) = 1$, $P(X=0)=1$.

$$\therefore \quad \underbrace{Ef(X_n)}_{f(\frac{1}{n})} \xrightarrow{n\to\infty} \underbrace{Ef(X)}_{f(0)} \quad \forall f \in C_b(\mathbb{R})$$

so $X_n \xrightarrow{\mathcal{D}} X$ but notice

$$\underbrace{P(X_n \le 0)}_{=0} \not\to \underbrace{P(X \le 0)}_{=1}$$

mass can magically appear on the boundaries of closed sets

$$\underbrace{P(X_n > 0)}_{=1} \not\to \underbrace{P(X>0)}_{=0}$$

mass can magically disappear on the boundaries of open sets

Definition: $\forall A \subset \Omega$ define

$\bar{A} :=$ closure of A (w.r.t. the Polish metric d)

$A^\circ :=$ open interior of A (all $x \in A$ s.t. $\exists \varepsilon > 0$ s.t. $B_\varepsilon(x) \subset A$)

$\partial A :=$ boundary of $A := \bar{A} - A^\circ$.

Here are some Portmanteau (French for coat hanger) results which give $\xrightarrow{\mathcal{D}}$ equivalence.

Theorem (Portmanteau I):
Let $P_1, P_2, \dots, P$ be probability measures on a Polish space $(\Omega, \mathcal{B}(\Omega))$.
Then the following are equivalent.

(a) $P_n \xrightarrow{\mathcal{D}} P$

(b) $\int_\Omega f\,dP_n \to \int_\Omega f\,dP$, $\forall f \in Lip(\Omega) \cap C_b(\Omega)$

Possible magically appearing mass

(c) $\limsup_n P_n(F) \le P(F)$, $\forall$ closed $F \subset \Omega$.

Possible magically disappearing mass

(d) $P(G) \le \liminf_n P_n(G)$, $\forall$ open $G \subset \Omega$.

(e) $\lim_n P_n(A) = P(A)$, $\forall A \in \mathcal{B}(\Omega)$ s.t. $P(\partial A) = 0$

Proof:

(a) $\Rightarrow$ (b): Trivial.

(b) $\Rightarrow$ (c): Let $F \subset \Omega$ be closed. As in the proof of the separating class Thm let

$$f_\varepsilon(\omega) = \left(1 - \frac{d(\omega, F)}{\varepsilon}\right)^+$$

so that $f_\varepsilon \in Lip(\Omega) \cap C_b(\Omega)$ and $\forall$ prob Q on $(\Omega, \mathcal{B}(\Omega))$

$$\int_\Omega I_F\,dQ \le \int_\Omega f_\varepsilon\,dQ \le \int_\Omega I_{F^\varepsilon}\,dQ \xrightarrow{\varepsilon\to 0} Q(F). \quad (*)$$

$$\therefore \limsup_n P_n(F) = \limsup_n \int_\Omega I_F\,dP_n$$
$$\le \limsup_n \int_\Omega f_\varepsilon\,dP_n, \text{ by } (*)$$
$$= \int_\Omega f_\varepsilon\,dP, \quad f_\varepsilon \in Lip(\Omega)\cap C_b(\Omega)$$
$$\xrightarrow{\varepsilon\to 0} P(F), \text{ by } (*)$$

(c) $\Leftrightarrow$ (d): Take complements of (c)

(c) & (d) $\Rightarrow$ (e): Suppose $P(\partial A) = 0$

$$\therefore \quad 0 = P(\bar{A} - A^\circ) = P(\bar{A}) - P(A^\circ)$$

by nested set subtraction props of P

and $P(A^\circ) \le \liminf_n P_n(A^\circ)$, by (c)

$\le \limsup_n P_n(\bar{A})$, $\quad$ inf $\le$ sup & $P(A^\circ) \le P(\bar{A})$

$\liminf_n P_n(A)$ & $\limsup_n P_n(A)$ sandwiched in here.

$\le P(\bar{A})$, by (d)

Since $P(A^\circ) \subset P(A) \subset P(\bar{A})$ & $P(\bar{A}) - P(A^\circ) = 0$

$$\lim_n P_n(A) = P(A)$$

(e) $\Rightarrow$ (a):

Let $f \in C_b(\Omega)$ & show $\int_\Omega f\,dP_n \to \int_\Omega f\,dP$.

Adjust f by adding a constant and re-scaling we can assume w.l.g that

$$0 < f < 1.$$

Recall Thm from lecture 11 that says

r.v. $X \geq 0 \Rightarrow E(X) = \int_0^\infty P(X > t)\,dt$

This applies to f so that

$$(*) \quad \int_\Omega f\,dP_n = \int_0^1 P_n(f > t)\,dt$$

$\downarrow ?$ as $n \to \infty$

$$(**) \quad \int_\Omega f\,dP = \int_0^1 P(f>t)\,dt.$$

Moreover continuity implies

$\{f > t\} = f^{-1}((t,\infty)) = \text{open} = \{f>t\}^\circ$

$\{f \geq t\} = (f^{-1}((-\infty,t)))^c = \text{closed} = \overline{\{f>t\}}$

$\therefore \partial\{f>t\} = \{f \geq t\} - \{f>t\} = \underbrace{\{f = t\}}$

has non-zero P mass for at most countably many t

$\therefore$ (e) implies

$$P_n(f>t) \xrightarrow{n\to\infty} P(f>t)$$

for $\mathcal{L}^1$-a.e. t

$\therefore$ DCT implies

$$\int_0^1 P_n(f>t)\,dt \xrightarrow{n\to\infty} \int_0^1 P(f>t)\,dt$$

$\| (*) \qquad\qquad \| (**)$

$$\int_\Omega f\,dP_n \qquad\qquad \int_\Omega f\,dP$$

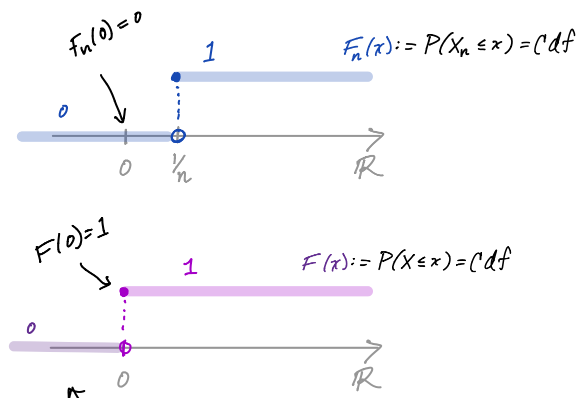
QED

The next result covers the special case of univariate real valued r.v.s.

Recall earlier example

$$X_n = \frac{1}{n} \xrightarrow{\mathcal{D}} X = 0$$

Note $F_n(x) \to F(x)$ $\forall x \neq 0$. Here is the picture



The problem is that $A = (-\infty, 0)$ is s.t. $P(X \in \partial A) \neq 0$

Define $C_F := \{x \in \mathbb{R} : F \text{ is continuous at } x\}$

so that $x \in C_F \Leftrightarrow 0 = F(x) - F(x-)$ ← F is always right cont.

$\Leftrightarrow 0 = P(X = x)$

$\Leftrightarrow PX^{-1}(\partial(-\infty, x]) = 0$

Theorem (Portmanteau II):

Let $X_1, X_2, \ldots, X$ be real-valued r.v.s with cdfs $F_1, F_2, \ldots, F$. Then

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{iff} \quad F_n(x) \xrightarrow{\mathcal{D}} F(x),\ \forall x \in C_F$$

Proof:

Let $P_1, P_2, \ldots, P$ be the induced measures of $X_1, X_2, \ldots, X$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Now

$$X_n \xrightarrow{\mathcal{D}} X \Leftrightarrow P_n \xrightarrow{\mathcal{D}} P$$

$$\Rightarrow P_n((-\infty, x]) \to P((-\infty, x]) \ \forall x \in C_F$$

since $x \in C_F \Rightarrow P(\partial(-\infty, x]) = 0$

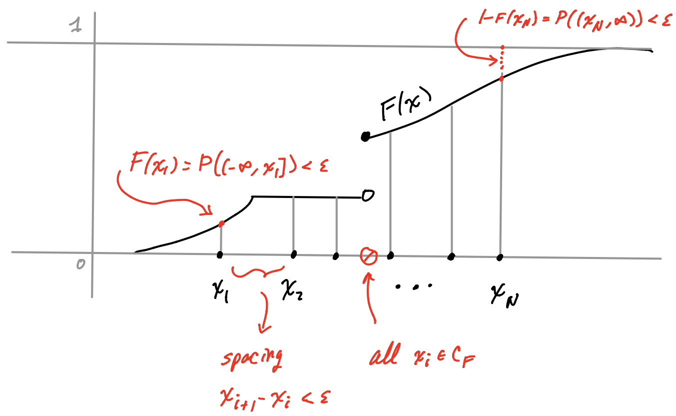
$$\Rightarrow F_n(x) \to F(x) \quad \forall x \in C_F$$

For the other direction suppose $F_n(x) \to F(x)$ $\forall x \in C_F$.

Using Portmanteau I it will be sufficient to show

$$\int_{\mathbb{R}} f\,dP_n \to \int_{\mathbb{R}} f\,dP, \quad \forall f \in Lip(\mathbb{R}) \cap C_b(\mathbb{R}).$$

Now for $\varepsilon > 0$ choose $x_1, \dots, x_N$ points s.t.



These exist since the results in Lecture 11 & 8 imply C_F^c is at most countable, $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Let
$$A_1 = (-\infty, x_1]$$
$$A_2 = (x_1, x_2]$$
$$\vdots$$
$$A_N = (x_{N-1}, x_N]$$
$$A_{N+1} = (x_N, \infty)$$

Note that by design $P(\partial A_i) = 0$ so $P_n(A_i) \to P(A_i)$

Now let $f \in Lip(\mathbb{R}) \cap C_b(\mathbb{R})$.
By rescaling we can suppose w.l.g that
$$0 \le f \le 1.$$

$$\therefore \limsup_n \int_{A_1} f\,dP_n \le \limsup_n \int_{A_1} 1\,dP_n$$

$$= \int_{A_1} 1\,P \quad \text{since } P(\partial A_1) = 0$$

$$\le \varepsilon + \int_{A_1} f\,dP$$
positive

Similarly

$$\limsup_n \int_{A_{N+1}} f\,dP_n \le \varepsilon + \int_{A_{N+1}} f\,dP$$

Also

$$\limsup_n \int_{A_2 \cup \dots \cup A_N} f\,dP_n$$

$$= \limsup_n \sum_{i=2}^{N} \int_{A_i} f\,dP_n$$

$$\le \limsup_n \sum_{i=2}^{N} (f(x_i) + c\varepsilon) \int_{A_i} 1\,dP_n$$

since $|f(x) - f(x_i)| \le c\varepsilon$ for $x \in A_i$ & $2 \le i \le N$

$$= \sum_{i=2}^{N} (f(x_i) + c\varepsilon) \int_{A_i} dP$$

since $P_n(A_i) \to P(A_i)$

$$= \sum_{i=2}^{N} \int_{A_i} f(x_i)\,dP + c\varepsilon \int_{\mathbb{R}} dP$$

$$\le \sum_{i=2}^{N} \int_{A_i} f\,dP + 2c\varepsilon$$

$$= \int_{A_2 \cup \dots \cup A_N} f\,dP + 2c\varepsilon$$

Putting everything together & letting $\varepsilon \to 0$

$$\limsup_n \int_{\mathbb{R}} f\,dP_n \le \int_{\mathbb{R}} f\,dP$$

Replacing f with $1-f$ above gives

$$\int_{\mathbb{R}} f\,dP_n \to \int_{\mathbb{R}} f\,dP$$

QED

Uniqueness of limits:

Continuous mapping:

Prohorov's Thm $\{P_1, P_2, ..\} = \mathcal{P}$

$\mathcal{P}$ is tight iff

$\mathcal{P}$ is sequentially compact

w.r.t. $\xrightarrow{\mathcal{D}}$

Thm: if

(i) $\{P_n\}_n$ is tight

(ii) $\int_\Omega f dP_n \to \int_\Omega f dP$ $\forall f \in$ separating class fcs $\{P_n\}$

then $P_n \xrightarrow{\mathcal{D}} P$.

Portmanteau III

(a) if Ω is locally compact then

$P_n \xrightarrow{\mathcal{D}} P \iff \int f dP_n \to \int f dP$ $\forall f \in C_c(\Omega)$

(b) if $\Omega = \mathbb{R}^d$

$P_n \xrightarrow{\mathcal{D}} P \iff \int f dP_n \to \int f dP$ $\forall f \in C_0^\infty(\mathbb{R})$

RHS $\Rightarrow$ tightness.

$\vdots$ char fns & monomials.

Skorokhod

Delta method

CLT

Examples

Kolmogorov 3 series Thm

Thm: scaled random walk $\xrightarrow{\mathcal{D}}$ Brownian Motion

Thm: $\sqrt{n}\,(F_n(t) - F(t)) \xrightarrow{\mathcal{D}}$ Brownian Bridge

Thm: Cumulative integral of scaled Binomial $\xrightarrow{\mathcal{D}}$ Poisson point process