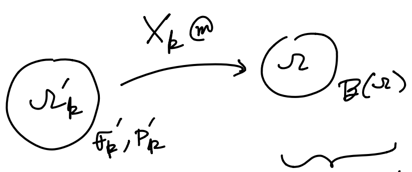
Lecture 14: Convergence in distribution and the Central limit theorem

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 (m) maps from some prob. space into $(\Omega, \mathcal{B}(\Omega))$



each X_k can have a different base space

... regardless they all have an induced distribution P_k on $(\Omega, \mathcal{B}(\Omega))$

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

$$P_n \xrightarrow{\mathcal{D}} P$$
$$\Updownarrow$$
$$\int f \nu_n d\mu \to \int f \nu d\mu, \ \forall f \in C_b(\Omega)$$

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 <u>central limit theorem</u> ... which effectively says:

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

$$E X_n = 0 \quad \& \quad \text{var}(X_n) = \sigma^2 < \infty$$

then $\sqrt{n}\,\bar{X}_n = \dfrac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} Z \sim N(0, \sigma^2)$.

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

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

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

$$\begin{aligned}
\therefore \limsup_n P_n(F) &= \limsup_n \int_\Omega I_F\,dP_n \\
&\le \limsup_n \int_\Omega f_\varepsilon\,dP_n, \quad \text{by } (*) \\
&= \int_\Omega f_\varepsilon\,dP, \quad f_\varepsilon \in \mathrm{Lip}(\Omega) \cap C_b(\Omega) \\
&\le \int_\Omega I_{F^\varepsilon}\,dP \\
&\xrightarrow{\varepsilon\downarrow 0} P(F), \quad \text{by } (*)
\end{aligned}$$

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

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

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

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): 5

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

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

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

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

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

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

$\| (*)$ $\quad\quad$ $\| (**)$

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

QED

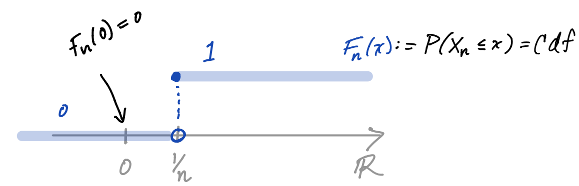
6

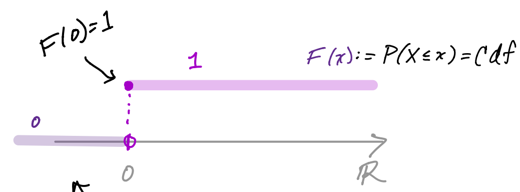
The next result covers the special case of univariate real valued r.v.s.

Recall earlier example

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

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





The problem is that $A = (-\infty, 0]$ has $PX^{-1}(\partial A) = P(X=0) \neq 0$

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 $X_1, X_2, \dots, X$ be real-valued r.v.s with cdfs $F_1, F_2, \dots, F$. Then the following are equivalent

(i) $X_n \xrightarrow{\mathcal{D}} X$

(ii) $F_n(x) \to F(x)$, $\forall x \in C_F$

(iii) $F_n^{-1}(u) \to F^{-1}(u)$, $\forall u \in C_{F^{-1}}$

the left-continuous quasi-inverse defined on $u \in (0,1)$ in lecture 8

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

((i) ⇒ (ii))

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

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

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

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

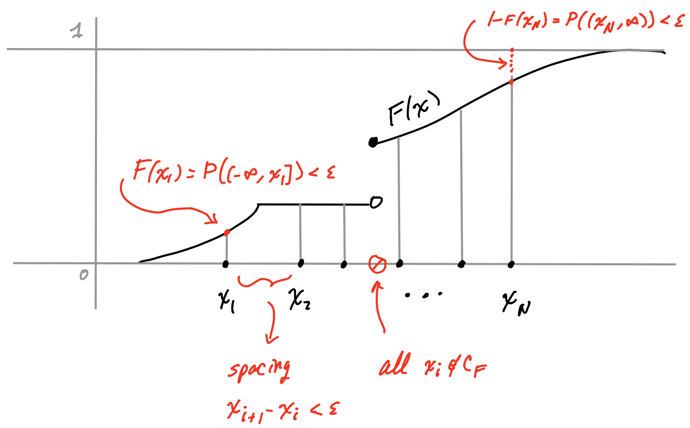
((ii) ⇒ (i))

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

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

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

$$0 \le f \le 1.$$

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

$$= \limsup_n F_n(x_1)$$

$$= F(x_1) \quad \text{since } x_1 \in C_F$$

$$\le \varepsilon$$

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

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 $P_n(A_i) = F_n(x_i) - F_n(x_{i-1}) \to P(A_i)$

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

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

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

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

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

$\big((ii) \Rightarrow (iii)\big)$

we show that $\forall u \in (0,1)$

$$F^{-1}(u) \underset{(*)}{\le} \liminf_n F_n^{-1}(u) \le \limsup_n F_n^{-1}(u) \underset{(**)}{\le} F^{-1}(u+)$$

To show $(*)$ suppose not.

Now we can choose $x \in C_F$ s.t.

$$\liminf_n F_n^{-1}(u) < x < F^{-1}(u).$$

By the switching formula $F(x) < u$ so that

$$F_n(x) \to F(x) < u$$

$\therefore$ $F_n(x) < u$ $\forall$ large n

$\therefore$ $x < F_n^{-1}(u)$ $\forall$ large n by "switching" again

which contradicts $\liminf_n F_n^{-1}(u) < x$.

To show $(**)$ use the same trick & suppose not.

$\therefore$ $\exists x \in C_F$ s.t.

$$\underbrace{F^{-1}(u+) < x}_{} < \limsup F_n^{-1}(u)$$

$\therefore u < F(x-) = F(x)$

by switching again $\left(\begin{array}{l} F(x-) \le u \iff x \le F^{-1}(u+) \\ \therefore F(x-) > u \iff x > F^{-1}(u+) \end{array}\right)$

$\therefore$ $u < F_n(x)$ $\forall$ large n

$\therefore$ $F_n^{-1}(u) < x$ $\forall$ large n

$\therefore$ contradiction

$\big((iii) \Rightarrow (ii)\big)$ Similar.

QED.

Theorem: (uniqueness of $\xrightarrow{\mathcal{D}}$ limits)

Let $P, Q, P_1, P_2, \ldots$ be probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$. If $P_n \xrightarrow{\mathcal{D}} P$ and $P_n \xrightarrow{\mathcal{D}} Q$ then $P = Q$.

Proof:

If $P_n \xrightarrow{\mathcal{D}} P$ & $P_n \xrightarrow{\mathcal{D}} Q$ then $\forall f \in C_b(\Omega)$

$$\lim_n \int_\Omega f\,dP_n = \underbrace{\int_\Omega f\,dQ = \int_\Omega f\,dP}$$

since $Lip(\Omega) \cap C_b(\Omega) \subset C_b(\Omega)$ is a separating class this implies $P = Q$.

QED

Theorem: (sub-sub-seq check for $\xrightarrow{\mathcal{D}}$)

Let $P, P_1, P_2, \ldots$ be a collection of Probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$. If $\forall$ sub-seq n_k, $\exists$ a sub-sub-seq n_{k_j} s.t. $P_{n_{k_j}} \xrightarrow{\mathcal{D}} P$ as $j \to \infty$ then

$$P_n \xrightarrow{\mathcal{D}} P \text{ as } n \to \infty.$$

Proof: Suppose not. Then $\exists f \in C_b(\Omega)$ s.t.

$$\int_\Omega f\,dP_n \not\to \int_\Omega f\,dP.$$

$\therefore$ $\exists n_k$ & $\varepsilon > 0$ s.t.

$$\left| \int_\Omega f\,dP_{n_k} - \int_\Omega f\,dP \right| \ge \varepsilon > 0, \ \forall k.$$

Now by assumption we can choose n_{k_j} s.t.

$$\int_\Omega f\,dP_{n_{k_j}} \longrightarrow \int_\Omega f\,dP.$$

$\therefore$ contradiction

QED

Skorokhod Representation

Skorokhods Representation theorem is really handy and effectively says

$$X_n \xrightarrow{\mathcal{D}} X \implies \exists X^*, X_n^* \text{ defined on a common probability space}$$

$$\text{s.t.} \quad \begin{array}{ccc} X_n^* & \xrightarrow{a.e.} & X^* \\ \| \mathcal{D} & & \| \mathcal{D} \\ X_n & & X \end{array}$$

Here is the full version of the Theorem.

Theorem: (Skorokhod Representation Theorem)

Let $P, P_1, P_2, \ldots$ be probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$ s.t.

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

Then there exist a probability space $(\Omega^*, \mathcal{F}^*, P^*)$ & maps $X^*, X_1^*, \ldots$

s.t.

$$\begin{tikzcd} \underset{\mathcal{F}^*, P^*}{\Omega^*} \arrow[r, bend left, "X_n^*(\omega)"] \arrow[r, bend right, "X^*(\omega)"'] & \underset{\mathcal{B}(\Omega)}{\Omega} \end{tikzcd}$$

s.t. $\mathcal{L}(X^*) = P$, $\mathcal{L}(X_n^*) = P_n$ and

$$\lim_{n\to\infty} X_n^*(w) = X^*(w)$$

for all $w \in \Omega^*$.

Notation: In the above theorem I'm using $\mathcal{L}(X^*)$, $\mathcal{L}(X_n^*)$ as a shorthand for $P^*(X^*)^{-1}$ & $P^*(X_n^*)^{-1}$, i.e. the induced distributions on $(\Omega, \mathcal{B}(\Omega))$.

The proof of the general Skorokhod's result is nasty and doesn't really provide any insight. We will just show the case $\Omega = \mathbb{R}$ and cite the general result in Billingsley's book "Convergence of Prob. Meas.".

Proof of Skorokhod's Rep when $\Omega = \mathbb{R}$:

Define c.d.f.s

$$F_n(x) = P_n((-\infty, x])$$

$$F(x) = P((-\infty, x]).$$

By Portmanteau II

$$F_n^{-1}(u) \longrightarrow F^{-1}(u) \quad \forall u \in C_{F^{-1}} \qquad (*)$$

Let $\Omega^* = (0,1)$, $\mathcal{F}^* = \mathcal{B}((0,1))$, $\underbrace{P^* = \mathcal{L}^1}_{\text{uniform measure on } (0,1)}$.

and set

$$X_n^*(u) := F_n^{-1}(u)$$

$$X^*(u) := F^{-1}(u)$$

$\forall u \in \Omega^*$.

$$\therefore \quad X_n^*(u) \to X^*(u), \ \forall u \in C_{F^{-1}} \text{ by } (*).$$

Finally notice that $C_{F^{-1}}^c$ must be countable since

$$C_{F^{-1}}^c \subset \bigcup_{\substack{\varepsilon > 0 \\ \varepsilon \in \mathbb{Q}}} \ \bigcup_{\substack{0<a<b<1 \\ a,b \in \mathbb{Q}}} \underbrace{\{u \in [a,b] : F^{-1}(u+) - F^{-1}(u-) > \varepsilon\}}$$

Since F^{-1} is monotonic on $[a,b]$ the sum of the jumps of size exceeding ε must be no greater than $F^{-1}(b) - F^{-1}(a)$ ... hence this set is finite.

$$\therefore P^*(C_{F^{-1}}) = 1$$

$$\therefore X_n^* \xrightarrow{a.e.} X^*$$

QED

The following result demonstraits Skorokhods' usefullness.

Theorem: (continuous mapping Thm for $\xrightarrow{\mathcal{D}}$)

Let $X, X_1, X_2, \ldots$ be generalized r.v.s taking values in a Polish space Ω (with σ-field $\mathcal{B}(\Omega)$) s.t. $X_n \xrightarrow{\mathcal{D}} X$.

If $g: \Omega \to \mathbb{R}$ satisfies

$$P(g \text{ is continuous @ } X) = 1$$

then

$$g(X_n) \xrightarrow{\mathcal{D}} g(X).$$

Proof:

By Skorokhods Rep Thm $\exists X_n^*, X^*$ defined on $(\Omega^*, \mathcal{F}^*, P^*)$ s.t.

$$\lim_{n\to\infty} X_n^*(w) = X^*(w) \quad \forall w \in \Omega^*$$

where $X_n \overset{\mathcal{D}}{=} X_n^*$, $X \overset{\mathcal{D}}{=} X^*$.

Let $A := \{w \in \Omega^* : g \text{ is continuous at } X^*(w)\}$

$$\therefore\ w \in A \Rightarrow \lim_{n\to\infty} g(X_n^*(w)) = g(X^*(w))$$

Since $X \overset{\mathcal{D}}{=} X^*$, $P^*(A) = 1$.

$$\therefore\ g(X_n^*) \xrightarrow{a.e.} g(X^*)$$

$$\therefore\ \underset{\displaystyle g(X_n)}{\overset{\displaystyle g(X_n^*)}{\shortparallel\mathcal{D}}} \xrightarrow{\mathcal{D}} \underset{\displaystyle g(X)}{\overset{\displaystyle g(X^*)}{\shortparallel\mathcal{D}}} \quad \text{by corollary above}$$

QED

Skorokhods Representation Theorem also gives us extensions of Fatou, DCT, ... which apply to the case $X_n \xrightarrow{\mathcal{D}} X$.

Here is an example.

Theorem: (UI extension for $\xrightarrow{\mathcal{D}}$)

If $X_n \xrightarrow{\mathcal{D}} X$ and the X_n's are UI then $E|X_n| \to E|X| < \infty$ and

$$E(X_n) \to E(X) < \infty.$$

Proof:

By Skorokhod $\exists X_n^*, X^*$ s.t.

$$X_n \overset{\mathcal{D}}{=} X_n^* \xrightarrow{a.e.} X^* \overset{\mathcal{D}}{=} X$$

Since $X_n \overset{\mathcal{D}}{=} X_n^*$ the X_n^*'s are also UI.

$\therefore$ by old UI Theorem we have

$$E|X_n^*| \to E|X^*| < \infty \text{ and}$$

$$E(X_n^*) \to E(X^*) < \infty$$

but again since $X_n^* \overset{\mathcal{D}}{=} X_n$ & $X^* \overset{\mathcal{D}}{=} X$

$$E|X_n| \to E|X| < \infty \text{ and}$$

$$E(X_n) \to E(X) < \infty.$$

QED

Skorokhod's Rep Thm also gives us the Delta method

Theorem: (Delta method)

Let $Z, X_1, X_2, \dots$ be random d-dimensional real vectors s.t.

$$c_n(X_n - x_0) \xrightarrow{\mathcal{D}} Z \qquad (*)$$

where $0 < c_n \to \infty$ as $n \to \infty$ & $x_0 \in \mathbb{R}$.

If $g: \mathbb{R}^d \to \mathbb{R}^d$ is differentiable @ x_0 then

$$c_n\big(g(X_n) - g(x_0)\big) \xrightarrow{\mathcal{D}} Dg(x_0) Z$$

Remark: $g: \mathbb{R}^d \to \mathbb{R}^d$ differentiable @ x_0 means that

$$\lim_{v \to 0} \frac{g(x_0 + v) - g(x_0) - Dg(x_0)v}{|v|} = 0$$

as $\varepsilon \to 0$ where $Dg(x_0) = \left(\frac{\partial}{\partial x_1} g, \dots, \frac{\partial}{\partial x_d} g\right)\Big|_{x = x_0}$

Remark: The intuitive way to understand this theorem is that $(*)$ suggest

$$X_n \overset{\mathcal{D}}{\approx} x_0 + \underbrace{\frac{Z}{c_n}}_{\to 0 \text{ as } n \to \infty}$$

so $g(X_n) \overset{\mathcal{D}}{\approx} g\left(x_0 + \frac{Z}{c_n}\right) \overset{\mathcal{D}}{\approx} g(x_0) + Dg(x_0)\frac{Z}{c_n}$

Proof:

Lets prove this for $d = 1$ ... the case $d > 0$ is similar. we can use $(*)$ along with Skorokhod's Rep to get X_n^* and Z^* s.t.

$$c_n(X_n^* - x_0) \xrightarrow{a.e.} Z^*.$$

Now

$$c_n\big(g(X_n^*) - g(x_0)\big) = \underbrace{c_n(X_n^* - x_0)}_{\substack{\xrightarrow{a.e.} Z^* \\ \text{by assumption}}} \cdot \underbrace{\overbrace{\left(\frac{g(X_n^*) - g(x_0)}{X_n^* - x_0}\right)}^{\text{define this to be } g'(x_0) \text{ when } X_n^* - x_0 = 0}}_{\substack{\xrightarrow{a.e.} g'(x_0) \text{ since} \\ \lim_n (X_n^* - x_0) \overset{a.e.}{=} \lim_n \frac{Z^*}{c_n} = 0}}$$

$$\therefore \; c_n\big(g(X_n^*) - g(x_0)\big) \xrightarrow{a.e.} g'(x_0) Z^*$$

$$\therefore \; c_n\big(g(X_n) - g(x_0)\big) \xrightarrow{\mathcal{D}} g'(x_0) Z$$

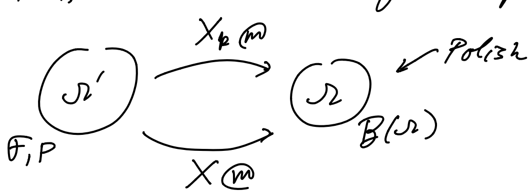
QED

Slutzky

The next theorem deals with the notion of convergence in probability. We will cover this in a later lecture but give some brief notation for it now, along with almost everywhere convergence.

Definition: ($\xrightarrow{P}$ & $\xrightarrow{a.e.}$)

Let $X, X_1, \dots$ be a collection of ⓜ maps



Then

$X_n \xrightarrow{P} X$ iff $\forall \varepsilon > 0$, $P(d(X_n, X) \geq \varepsilon) \xrightarrow{n\to\infty} 0$

$X_n \xrightarrow{a.e.} X$ iff $P(\lim_n d(X_n, X) = 0) = 1$.

Remark:

$$X_n \xrightarrow{P} X \iff d(X_n, X) \xrightarrow{P} 0$$

$$X_n \xrightarrow{a.e.} X \iff d(X_n, X) \xrightarrow{a.e.} 0$$

Remark: In lecture 6 we remarked that SLLN $\Rightarrow$ WLLN. Indeed

$$X_n \xrightarrow{a.e.} X \Rightarrow X_n \xrightarrow{P} X$$

since

$$\limsup_n P(d(X_n, X) \geq \varepsilon) \overset{\text{Fatou}}{\leq} P(d(X_n, X) \geq \varepsilon \text{ i.o.}_n)$$

$$\leq \underbrace{P(d(X_n, X) \not\to 0)}_{=0}$$

Theorem: (Slutzky)

Let $X, X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be collections of generalized r.v.s taking values in a Polish space Ω (with σ-field $\mathcal{B}(\Omega)$) all defined on the same probability space.

Then

$$X_n \xrightarrow{\mathcal{D}} X \ \&\ d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \xrightarrow{\mathcal{D}} X.$$

Proof:

We use Portmanteau I and show

$$\limsup_n P(Y_n \in F) \leq P(X \in F)$$

for all closed $F \subset \Omega$.

First note $\forall \varepsilon > 0$

$$\{Y_n \in F\} \subset \{d(X_n, Y_n) \geq \varepsilon\} \cup \{X_n \in F^\varepsilon\}$$

$$\therefore \limsup_n P(Y_n \in F)$$

$$\leq \underbrace{\limsup_n P(d(X_n, Y_n) \geq \varepsilon)}_{=0 \text{ by assumption}} + \limsup_n P(X_n \in F^\varepsilon)$$

$$\leq \limsup_n P(X_n \in \overline{F^\varepsilon})$$

$$\leq P(X \in \overline{F^\varepsilon}) \text{ by Portmanteau I}$$

Since $\overline{F^\varepsilon} \downarrow F$ as $\varepsilon \downarrow 0$

$$P(X \in \overline{F^\varepsilon}) \downarrow P(X \in F)$$

as $\varepsilon \downarrow 0$ by CFA.

$$\therefore \limsup_n P(Y_n \in F) \leq P(X \in F)$$

as was to be shown

QED.

Slutsky's theorem gives us a nice corollary that relates $\xrightarrow{\mathcal{D}}$, $\xrightarrow{P}$ and a.e. convergence.

Corollary:

$$X_n \xrightarrow{a.e.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{D}} X.$$

$\uparrow$ use $X \xrightarrow{\mathcal{D}} X$, $d(X_n, X) \xrightarrow{P} 0$

Warning: The reverse implications do not hold in general.

The "moving spike" showed $\exists X_n$ & X s.t.

$$X_n \xrightarrow{P} X \text{ but } X_n \not\xrightarrow{a.e.} X$$

To see why $\xrightarrow{\mathcal{D}} \not\Rightarrow \xrightarrow{P}$ consider

$$X = \begin{cases} -1 & \text{w.p. } 1/2 \\ 1 & \text{w.p. } 1/2 \end{cases}$$

and $X_n := (-1)^n X$ so that $X_n \overset{\mathcal{D}}{=} X$

but $P(|X_n - X| \geq \varepsilon) = 1$ when n is odd.

$$\therefore \quad X_n \xrightarrow{\mathcal{D}} X \text{ but } X_n \not\xrightarrow{P} X.$$

However, the following corollary is one special case where the reverse implication does hold.

Corollary: If $X_1, X_2, \ldots$ are Ω-valued r.v.s where $(\Omega, \mathcal{B}(\Omega))$ is Polish all defined on the same prob. space then

$$X_n \xrightarrow{\mathcal{D}} c \iff X_n \xrightarrow{P} c$$

Proof: If $X_n \xrightarrow{\mathcal{D}} c$ then

$$\limsup_n P(X_n \in \underbrace{B_\varepsilon(c)^c}_{\text{closed}}) \leq P(c \in B_\varepsilon(c)^c) = 0$$

QED.

Here is another useful corollary of Slutsky's thm

Corollary: Suppose X, X_n, Y_n are r.v.s taking values in a polish space $(\Omega, \mathcal{B}(\Omega))$ all defined on the same probability space.

If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c \in \Omega$

then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X_n, c)$.

Proof:

Suppose $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c \in \Omega$.

$$\therefore \quad (X_n, c) \xrightarrow{\mathcal{D}} (X, c) \qquad (*)$$

since $f(x,y) \in C_b(\Omega \times \Omega)$ implies $f(x,c) \in C_b(\Omega)$ so that $Ef(X_n, c) \to Ef(X, c)$.

Let $\tilde{d}$ be the product dist on $\Omega \times \Omega$ so that

$$\tilde{d}((X_n, c), (X_n, Y_n)) = d(c, Y_n) \xrightarrow{P} 0. \quad (**)$$

Now $(*)$ and $(**)$ implies

$$(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$$

by Slutsky's theorem.

QED

Now under the same assumptions as in the previous corollary we can use the continuous mapping theorem to obtain:

$$X_n + Y_n \xrightarrow{\mathcal{D}} X + c$$

$$X_n Y_n \xrightarrow{\mathcal{D}} Xc$$

$$\frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X}{c} \quad \text{provided } c \neq 0$$

Tightness and Porhorov's Thm

An extremly useful fact, when working with a sequence of real numbers $x_1, x_2, \dots$ is that if $\{x_n\}_{n\geq 1}$ is bounded then there exists a sub-sequence n_k and a real number x s.t. $x_{n_k} \xrightarrow{k\to\infty} x$

We would like to have something similar for a sequence $P_1, P_2, \dots$ of probability measures on a Polish space. The problem is to find the right generalization of "boundedness" to guarantee the existance of a probability measure P & a sub-sequence n_k s.t.

$$P_{n_k} \xrightarrow{\mathcal{D}} P.$$

It turns out the right definition is our old friend "tightness" from the homeworks last lecture.

Definition:

Let $\mathbb{P}$ be a collection of Probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$. Then

$\mathbb{P}$ is tight iff $\exists$ a compact $K \subset \Omega$ s.t.

$$\sup_{P\in\mathbb{P}} P(K^c) < \varepsilon$$

Notation: When $\{X_n\}_{n\geq 1}$ are random vectors taking values in $\mathbb{R}^d$ a here is a fancy notational shorthand:

$|X_n| = \mathcal{O}_p(1)$ means $\{X_n\}_{n\geq 1}$ is tight

Prohorov's Thm shows tightness is the right definition of "boundedness."

Theorem: (Prohorov)

Let $\mathbb{P}$ be a collection of Probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$. Then

$$\mathbb{P} \text{ is tight}$$
$$\Updownarrow$$

(*) $\forall P_1, P_2, \dots \in \mathbb{P}$ there exists a sub-seq n_k and a prob measure P on $(\Omega, \mathcal{B}(\Omega))$ s.t. $P_{n_k} \xrightarrow{\mathcal{D}} P$

not necessarily in $\mathbb{P}$.

Note: If $\mathbb{P}$ satisfies (*) then $\mathbb{P}$ is said to be relatively compact w.r.t. $\xrightarrow{\mathcal{D}}$.

Proof: This proof is a bit nasty and doesn't really provide much probabalistic intuition so we will skip it and simply cite Billingsley's book on convergence of Probability measures.

Here is an example of how one uses Prohorov.

Theorem: (Portmanteau III)

Let $X_1, X_2, \ldots, X$ be r.v.s taking values in $C[0,1]$ ⓜ w.r.t $\mathcal{B}(C[0,1])$. Then the following are equivalent:

(A) $X_n \xrightarrow{\mathcal{D}} X$.

(B) $\{X_n\}_{n\geq 1}$ is tight and $\forall t_1,\ldots,t_m \in [0,1]$

$$\pi_{t_1,\ldots,t_m}(X_n) \xrightarrow{\mathcal{D}} \pi_{t_1,\ldots,t_m}(X).$$

Proof:

(A) $\Rightarrow$ (B):

If $X_n \xrightarrow{\mathcal{D}} X$ the sub-sub check for $\xrightarrow{\mathcal{D}}$ implies $\{X_n\}_{n\geq 1}$ is relatively compact.

$\therefore$ $\{X_n\}_{n\geq 1}$ is tight by Prohorov.

Also $\pi_{t_1,\ldots,t_m}(X_n) \xrightarrow[n\to\infty]{\mathcal{D}} \pi_{t_1,\ldots,t_m}(X)$

by the continuous mapping theorem.

(B) $\Rightarrow$ (A): Assume

(*) $\pi(X_n) \to \pi(X)$, $\forall \pi \in \{\pi_{t_1,\ldots,t_m} : t_i \in [0,1]\}$

(**) $\{X_n\}_{n\geq 1}$ is tight.

show $X_n \xrightarrow{\mathcal{D}} X$.

By "sub-sub-seq check" it is sufficient to show $\forall n_k\ \exists n_{k_j}$ s.t. $X_{n_{k_j}} \xrightarrow[j\to\infty]{\mathcal{D}} X$.

For a given n_k use (**) with Prohorov to find n_{k_j} and Y s.t.

(***) $$X_{n_{k_j}} \xrightarrow[j\to\infty]{\mathcal{D}} Y$$

By the continuous mapping thm

$$\begin{array}{ccc} \pi(X_{n_{k_j}}) & \xrightarrow{\mathcal{D}} & \pi(Y), \quad \forall \pi = \pi_{t_1,\ldots,t_m} \\ \mathcal{D}\big\downarrow \text{by } (*) & & \\ \pi(X) & & \end{array}$$

$\therefore$ $\pi(X) \overset{\mathcal{D}}{=} \pi(Y)$, $\forall \pi = \pi_{t_1,\ldots,t_m}$ by uniqueness of limits.

$\therefore$ $X \overset{\mathcal{D}}{=} Y$ since the f.d.d characterize probs on $(C[0,1], \mathcal{B}(C[0,1]))$.

$\therefore$ $X_{n_{k_j}} \xrightarrow[j\to\infty]{\mathcal{D}} X$ by (***)

Since n_k was arb the "sub-sub check" Thm gives $X_n \xrightarrow{\mathcal{D}} X$.

QED

Here is another consequence of Prohorov.

Theorem: (Portmanteau IV)

Let $X_1, X_2, \ldots, X$ be r.v.s taking values in a Polish $(\Omega, \mathcal{B}(\Omega))$. Then the following are equivalent:

(i) $X_n \xrightarrow{\mathcal{D}} X$.

(ii) $\{X_n\}_{n\geq 1}$ is tight and there exists a separating family Γ for $(\Omega, \mathcal{B}(\Omega))$ s.t. $\Gamma \subset C_b(\Omega)$ and

$$E g(X_n) \to E g(X) \quad \forall g \in \Gamma. \quad (*)$$

Proof:

(i) $\Rightarrow$ (ii): follows immediately since (i) implies $\{X_n\}_{n\geq 1}$ is tight and trivially $(*)$ holds by def of $\xrightarrow{\mathcal{D}}$.

(ii) $\Rightarrow$ (i):

Let n_k be an arb sub-seq. By "sub-sub-seq check" it will be sufficient to show $\exists\, n_{k_j}$ s.t. $X_{n_{k_j}} \xrightarrow{\mathcal{D}} X$ as $j \to \infty$.

Now by tightness & Porhorov's Thm $\exists\, Y$ & $\exists\, n_{k_j}$ s.t.

$$X_{n_{k_j}} \xrightarrow[j\to\infty]{\mathcal{D}} Y \quad (**)$$

$$\therefore \quad E f(X_{n_{k_j}}) \xrightarrow[j\to\infty]{} E f(Y) \quad \forall f \in C_b(\Omega).$$

Since $\Gamma \subset C_b(\Omega)$ we therefore have

$$E g(X_{n_{k_j}}) \xrightarrow[j\to\infty]{} E g(Y) \quad \forall g \in \Gamma$$

by (ii) $\downarrow$

$$E g(X)$$

$\therefore\ E g(X) = E g(Y) \ \forall g \in \Gamma$

$\therefore\ X \overset{\mathcal{D}}{=} Y$ since Γ separates

$\therefore\ X_{n_{k_j}} \xrightarrow[j\to\infty]{\mathcal{D}} X$ by $(**)$

as was to be shown.

QED.

Now we can use our results on separating classes in Lecture 13 to get the following.

Theorem: (Portmanteau V)

Let $X_1, X_2, \ldots, X$ be r.v.s taking values in a Polish $(\Omega, \mathcal{B}(\Omega))$. Then

$$X_n \xrightarrow{\mathcal{D}} X \text{ iff } Ef(X_n) \to Ef(X) \;\forall f \in \Gamma$$

whenever

(I) $\Gamma = C_c(\Omega)$ and Ω is locally compact

(II) $\Gamma = C_c^\infty(\Omega)$ and $\Omega = \mathbb{R}^d$

(III) $\Gamma = \{\text{monomials}\}$ and Ω is a compact subset of $\mathbb{R}^d$

(IV) $\Gamma = \{e^{ix\cdot k} : k \in \mathbb{R}^d\}$ and $\Omega = \mathbb{R}^d$ and $|X_n| = \mathcal{O}_p(1)$.

Proof:

Notice that all the above function classes Γ form separating classes by the results in Lecture 13.

$\therefore$ Portmanteau IV implies we just need to show tightness of $\{X_n\}_{n\geq 1}$ each case.

For (I) an (II) choose compact K s.t.

$$P(X \in K) > 1 - \varepsilon/2$$

existance since Ω is Polish so that

$$P(X \in A) = \sup\{P(X \in K) : \text{compact } K \subset \Omega\}$$

By similar techiques as in Lecture 13 we can find an approximation $f \in \Gamma$ to I_K s.t.

$$I_K \leq f \;\&\; |Ef(X) - \underbrace{EI_K(X)}_{= P(X \in K) > 1 - \varepsilon/2}| \leq \frac{\varepsilon}{2}$$

$$\therefore P(X_n \in K) \leq Ef(X_n) \xrightarrow{n \to \infty} \underbrace{Ef(X)}_{\text{within } \varepsilon/2 \text{ of } P(X \in K)}$$

$\therefore P(X_n \in K) > 1 - \varepsilon \quad \forall$ suff large n.

$\therefore \{X_n\}_{n\geq 1}$ is tight.

For (III), $\{X_n\}_{n\geq 1}$ is tight since Ω is compact.

For (IV) the requirement that $\sup_n E(|X_n|) < \infty$ allows us to use Markov's Thm

$$P(|X_n| \geq a) \leq \frac{E|X_n|}{a}$$ can be made to be $< \varepsilon$ by choosing a suff large

Now choose $K := \{x \in \mathbb{R}^d : |x| \leq a\}$ for a suff large so that

$$P(X_n \notin K) < \varepsilon$$

$\therefore \{X_n\}_{n\geq 1}$ is tight

QED

Notice that part (IV) of Portmanteau V can be rephrased in terms of characteristic functions

Corollary: (characteristic functions for $\xrightarrow{\mathcal{D}}$)

Suppose $X_1, X_2, \ldots, X$ are random vectors with characteristic functions $\phi_1, \phi_2, \ldots, \phi$.

Then $X_n \xrightarrow{\mathcal{D}} X$

$$\Updownarrow$$

$$\left(\begin{array}{l} \phi_n(k) \to \phi(k) \quad \forall k \in \mathbb{R}^d \\ \text{and} \quad |X_n| = \mathcal{O}_p(1) \end{array}\right) \quad (*)$$

Proof:

This follows immediately from Portmanteau IV since $\phi_n(k) \to \phi(k)$ $\forall k \in \mathbb{R}^d$ means

$$Ef(X_n) \to Ef(X)$$

$\forall f \in \{x \mapsto e^{ix \cdot k} : k \in \mathbb{R}^d\}$.

QED

This means that whenever $\phi_n(k) \to \phi(k)$ $\forall k \in \mathbb{R}^d$ it is sufficient to simply show $X_n = \mathcal{O}_p(1)$ for establishing

$$X_n \xrightarrow{\mathcal{D}} X.$$

Many of the conditions for $X_n \xrightarrow{\mathcal{D}} X$ as an undergrad are simply conditions for establishing $(*)$.

Here are a few examples:

$$\left(\begin{array}{l} \phi_n(k) \to \phi(k) \quad \forall k \in \mathbb{R}^d \\ \text{and} \quad \sup_n E|X_n| < \infty \end{array}\right) \implies (*)$$

Proof: Use Markov's Thm

$$P(|X_n| \geq N) \leq \frac{E|X_n|}{N}$$

$$\leq \frac{\sup_n E|X_n|}{N} < \infty$$

$\therefore$ $\forall \varepsilon > 0$ one can choose N large enough s.t.

$$P(X_n \notin \underbrace{\overline{B_N(0)}}_{\text{ball of radius } N}) < \varepsilon$$

$\therefore$ $X_n = \mathcal{O}_p(1)$.

$$\left(\begin{array}{l} \phi_n(k) \to \phi(k) \quad \forall k \in \mathbb{R}^d \\ \phi(k) \text{ is continuous at } k = 0 \end{array}\right) \implies (*)$$

Proof:

See Achim Klenke's book "Probability theory" p. 309

$$\left(\begin{array}{l} M_n(k) \to M(k) \quad \forall k \in \mathbb{R}^d \\ \& \ M(k) < \infty \text{ in a} \\ \text{neighborhood of } k = 0 \end{array}\right) \implies (*)$$

Proof: By analyzing the Taylor series of M_n & ϕ_n one can show $\phi_n(k) \to \phi(k)$ $\forall k$ and $\phi(k)$ is continuous @ $k = 0$.

Hence the previous result applies.

The last condition is both sufficient and necessary ... its called the

Cramér-Wald device:

$$\left(\langle k, X_n \rangle \xrightarrow{\mathcal{D}} \langle k, X \rangle \quad \forall k \in \mathbb{R}^d \right) \iff (*)$$

Proof:

($\Leftarrow$)

use the continuous mapping thm.

($\Rightarrow$)

write $X_n = (X_{n,1}, \dots, X_{n,d})$ and suppose $\langle k, X_n \rangle \xrightarrow{\mathcal{D}} \langle k, X \rangle$.

$\therefore$ $|\langle k, X_n \rangle| = \mathcal{O}_p(1)$ $\forall k$

by the "char fun check corollary"

$\therefore$ $|X_{n,i}| = \mathcal{O}_p(1)$ $\forall i = 1, 2, \dots, d$

$\therefore$ Given $\varepsilon > 0$ we can find $N_1, \dots, N_d > 0$ s.t. $P(X_{n,i} \notin [-N_i, N_i]) < \frac{\varepsilon}{d}$ for $i = 1, \dots, d$ and hence

$$P\left(X_n \notin \prod_{i=1}^{d} [-N_i, N_i]\right) \leq \sum_{i=1}^{d} P(X_{n,i} \notin [-N_i, N_i]) \leq \varepsilon$$

$\therefore$ $|X_n| = \mathcal{O}_p(1)$

To finish notice $\langle k, X_n \rangle \xrightarrow{\mathcal{D}} \langle k, X \rangle$ implies

$$\underbrace{E e^{i\langle k, X_n \rangle}}_{=\phi_n(k)} \longrightarrow \underbrace{E e^{i\langle k, X \rangle}}_{\phi(k)}$$

QED