

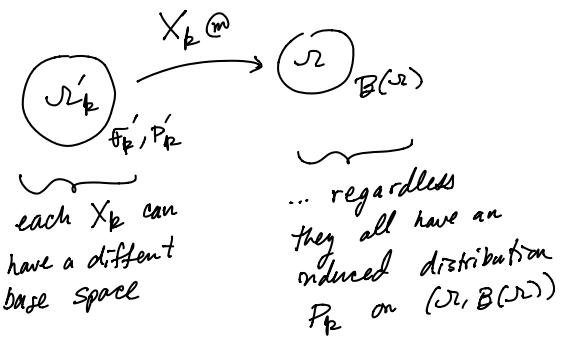
Lecture 14: Convergence in distribution and the Central limit theorem

(1)

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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

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

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



Definition:

$P_n \xrightarrow{\mathcal{D}} P$ iff $\forall f \in C_b(\mathbb{R})$, $\int f dP_n \rightarrow \int f dP$.

$X_n \xrightarrow{\mathcal{D}} X$ iff $\forall f \in C_b(\mathbb{R})$, $E f(X_n) \rightarrow E f(X)$

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 v_1, v_2, \dots, v w.r.t. some base measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e.

$$dP_n = v_n d\mu \quad \& \quad dP = v d\mu$$

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

$$\int f v_n d\mu \rightarrow \int f v d\mu, \quad \forall f \in C_b(\mathbb{R})$$

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 $\exists A \in \mathcal{B}(\mathbb{R})$ s.t.

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

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 (Ω, \mathcal{F}, P)) with

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

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

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: $\Omega = \mathbb{R}$, $P(X_n = \frac{1}{n}) = 1$, $P(X = 0) = 1$. (3)

$$\therefore \underbrace{\mathbb{E} f(X_n)}_{f(\frac{1}{n})} \xrightarrow{n \rightarrow \infty} \underbrace{\mathbb{E} f(X)}_{f(0)} \text{ if } f \in C_b(\mathbb{R})$$

so $X_n \xrightarrow{d} X$ but notice

$$\left. \begin{array}{l} P(X_n \leq 0) \xrightarrow{=} P(X \leq 0) \\ = 1 \end{array} \right\} \begin{array}{l} \text{mass can} \\ \text{magically} \\ \text{appear on the} \\ \text{boundaries of} \\ \text{closed sets} \end{array}$$

$$\left. \begin{array}{l} P(X_n > 0) \xrightarrow{=} P(X > 0) \\ = 0 \end{array} \right\} \begin{array}{l} \text{mass can} \\ \text{magically} \\ \text{disappear on} \\ \text{the boundaries} \\ \text{of open sets} \end{array}$$

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{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) $P_n \xrightarrow{d} P$

(b) $\int_{\Omega} f dP_n \rightarrow \int_{\Omega} f dP$, $\forall f \in \text{Lip}(\Omega) \cap C_b(\Omega)$

(c) $\limsup_n P_n(F) \leq P(F)$, $\forall \text{closed } F \subset \Omega$.

(d) $P(G) \leq \liminf_n P_n(G)$, $\forall \text{open } G \subset \Omega$.

(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 then let

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

so that $f_\varepsilon \in \text{Lip}(\Omega) \cap C_b(\Omega)$ and $\mathbb{E} f_\varepsilon dQ$ on $(\Omega, \mathcal{B}(\Omega))$

$$\int_{\Omega} f_F dQ \leq \int_{\Omega} f_\varepsilon dQ \leq \left(\int_{\Omega} f_\varepsilon dQ \xrightarrow{\varepsilon \downarrow 0} Q(F) \right). \text{ (*)}$$

$$\therefore \limsup_n P_n(F) = \limsup_n \int_{\Omega} f_F dP_n$$

$$\leq \limsup_n \int_{\Omega} f_\varepsilon dP_n, \text{ by (*)}$$

$$= \int_{\Omega} f_\varepsilon dP, \quad f_\varepsilon \in \text{Lip}(\Omega) \cap C_b(\Omega)$$

$$\leq \int_{\Omega} f_F dP$$

$$\xrightarrow{\varepsilon \downarrow 0} P(F), \text{ by (*)}$$

(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

$$\text{and } P(A^\circ) \leq \liminf_n P_n(A^\circ), \text{ by (c)}$$

$$\begin{aligned} &\leq \limsup_n P_n(\bar{A}), \quad \text{int } \sup \bar{P}_n(\bar{A}) \\ &\leq P(\bar{A}), \quad \text{by (d)} \end{aligned}$$

$\text{int } \sup \bar{P}_n(\bar{A}) \leq P(\bar{A}) \leq P(\bar{A})$

sandwiched in here.

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

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

(4)

(e) \Rightarrow (a):

$$\text{Let } f \in C_b(\mathbb{R}) \text{ & show } \int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP.$$

Adjust f by adding a constant and re-scaling we can assume w.l.g. that $0 < f < 1$.

Recall Thm from lecture II that says

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

This applies to f so that

$$(*) \quad \int_{\mathbb{R}} f dP_n = \int_0^1 P_n(f > t) dt \quad \downarrow ? \text{ as } n \rightarrow \infty$$

$$(**) \quad \int_{\mathbb{R}} f dP = \int_0^1 P(f > t) dt.$$

Moreover continuity implies

$$\begin{aligned} \{f > t\} &= f^{-1}((t, \infty)) = \text{open} = \{f > t\}^o \\ \{f \geq t\} &= (f^{-1}((-\infty, t]))^c = \text{closed} = \overline{\{f > t\}} \end{aligned}$$

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

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

\therefore (e) implies

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

for \mathbb{P} -a.e. t

\therefore DCT implies

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

"(x)" "(**)"

$$\int_{\mathbb{R}} f dP_n \quad \int_{\mathbb{R}} f dP$$

$\square \quad \square$

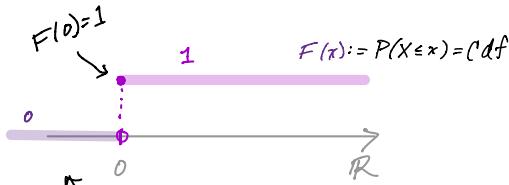
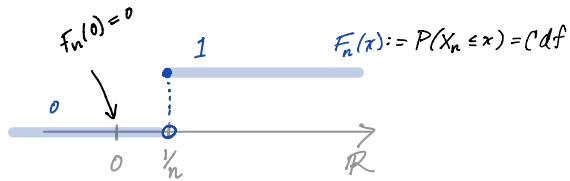
(5)

The next result covers the special case of univariate real valued r.v.s. (6)

Recall earlier example

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

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



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

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

so that $x \in C_F \iff 0 = F(x) - F(x^-) \iff$ right cont.
 $\iff 0 = P(X=x)$
 $\iff P X^{-1}(\partial(-\infty, x]) = 0$

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

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

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

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

→ the left-continuous quasi-inverse defined on $u \in C_F$ in lecture 8

Proof:

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

$(i) \Rightarrow (ii)$

(7)

$$X_n \xrightarrow{d} X \Leftrightarrow P_n \xrightarrow{d} P$$

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

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

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

$(ii) \Rightarrow (i)$

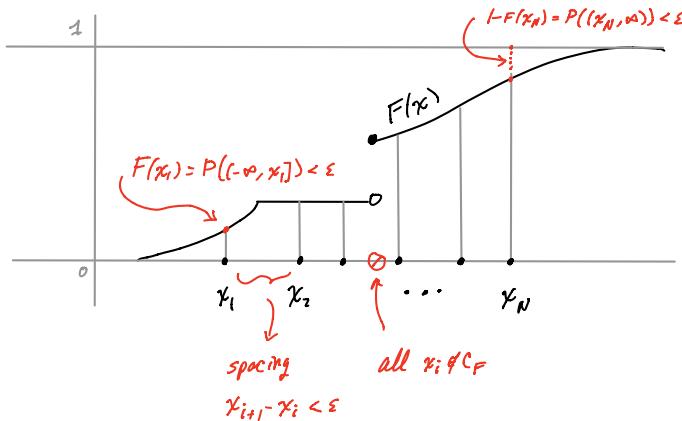
For the other direction suppose

$$F_n(x) \rightarrow F(x) \quad \forall x \in C_F.$$

Using Portmanteau I it will be sufficient to show

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

For $\epsilon > 0$ 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, $\lim_{x \rightarrow -\infty} F(x) = 0$

and $\lim_{x \rightarrow \infty} F(x) = 1$.

$$\text{Let } A_1 = (-\infty, x_1]$$

$$A_2 = (x_1, x_2]$$

:

$$A_N = (x_{N-1}, x_N]$$

$$A_{N+1} = (x_N, \infty)$$

Note that by

design $P(\partial A_i) = 0$

so $P_n(A_i) \rightarrow P(A_i)$

$$\begin{aligned} \therefore \limsup_n \int_{A_1} f dP_n &\leq \limsup_n \int_{A_1} 1 dP_n \\ &= \limsup_n F_n(x_1) \\ &= F(x_1) \quad \text{since } x_1 \in C_F \end{aligned}$$

$$\leq \epsilon$$

$$\leq \epsilon + \int_{A_1} f dP \underset{\text{positive}}{\nwarrow}$$

Similarly

$$\limsup_n \int_{A_{N+1}} f dP_n \leq \epsilon + \int_{A_{N+1}} f dP$$

Also

$$\begin{aligned} \limsup_n \int_{A_2 \cup \dots \cup A_N} f dP_n &= \limsup_n \sum_{i=2}^N \int_{A_i} f dP_n \\ &\leq \limsup_n \sum_{i=2}^N (f(x_i) + c\epsilon) \int_{A_i} 1 dP_n \\ &\quad \text{since } |f(x) - f(x_i)| \leq c\epsilon \\ &\quad \text{for } x \in A_i \text{ & } 2 \leq i \leq N \end{aligned}$$

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

$$\text{Since } P(A_i) = F_n(x_i) - F_n(x_{i-1}) \rightarrow P(A_i)$$

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

$$\leq \sum_{i=2}^N \int_{A_i} f dP + 2c\epsilon$$

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

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

$$\limsup_n \int_{\Omega} f dP_n \leq \int_{\Omega} f dP$$

Replacing f with $1-f$ above gives

$$\int_{\Omega} f dP_n \rightarrow \int_{\Omega} f dP$$

$(ii) \Rightarrow (iii)$

We show that $\forall \varepsilon > 0$,

$$F^{-1}(u) \leq \liminf_n F_n^{-1}(u) \leq \limsup_n F_n^{-1}(u) \leq F^{-1}(u) \quad (**)$$

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

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

$\therefore F_n(x) < u \quad \forall \text{large } n$

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

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

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

by switching again $(F(x-) \leq u \Leftrightarrow x \leq F^{-1}(u))$
 $\therefore F(x-) > u \Leftrightarrow x > F^{-1}(u)$

$\therefore u < F_n(x) \quad \forall \text{large } n$

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

\therefore contradiction

$((iii) \Rightarrow (ii))$ Similcn.

QED.

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

(10)

Let P, Q, P_1, P_2, \dots 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 $\forall f \in C_b(\Omega)$

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

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

QED

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

Let P, P_1, P_2, \dots be a collection of
probability measures on a Polish $(\Omega, \mathcal{B}(\Omega))$.

If \forall sub-seg n_k , \exists a sub-sub-seg n_{k_j}

s.t. $P_{n_{k_j}} \xrightarrow{\mathcal{D}} P$ as $j \rightarrow \infty$ then

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

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

$$\int_{\Omega} f dP_n \not\rightarrow \int_{\Omega} f dP.$$

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

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

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

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

\therefore contradiction

QED

Skorokhod Representation

(11)

Skorokhod's representation theorem is really handy and effectively says

$$X_n \xrightarrow{d} X \Rightarrow \exists X^*, X_n^* \text{ defined on a common probability space s.t. } \begin{array}{ccc} X_n^* & \xrightarrow{\text{a.e.}} & X^* \\ \parallel \mathcal{D} & & \parallel \mathcal{D} \\ X_n & & X \end{array}$$

Here is the full version of the theorem.

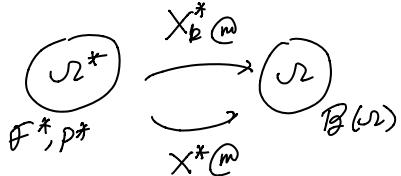
Theorem: (Skorokhod Representation Theorem)

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

$$P_n \xrightarrow{d} P.$$

Then there exist a probability space $(\mathcal{S}^*, \mathcal{F}^*, P^*)$ & maps X^*, X_n^*, \dots

s.t.



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

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

for all $w \in \mathcal{S}^*$.

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 $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$.

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

(12)

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

Define c.d.f.s

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

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

By Portmanteau II

$$F_n^{-1}(u) \rightarrow F^{-1}(u) \quad \forall u \in C_F \quad (*)$$

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

and set

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

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

$$\forall u \in \mathcal{S}^*$$

$\therefore X_n^*(u) \rightarrow X^*(u)$, $\forall u \in C_F$ by (*).

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

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

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

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

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

this set is finite.

QED

(13)

The following result demonstrates Skorohod's usefulness.

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

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

If $g: \mathcal{S} \rightarrow \mathbb{R}$ satisfies

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

then

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

Proof:

By Skorohod's Rep Thm $\exists X_n^*, X^*$ defined on $(\mathcal{S}^*, \mathcal{F}^*, P^*)$ s.t.

$$\lim_{n \rightarrow \infty} X_n^*(w) = X^*(w) \quad \forall w \in \mathcal{S}^*$$

$$\text{where } X_n = X_n^*, \quad X = X^*.$$

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

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

$$\text{Since } X = X^*, \quad P(A) = 1.$$

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

$$\therefore g(X_n^*) \xrightarrow{\mathcal{D}} g(X^*) \text{ by corollary above}$$

$$\begin{array}{ccc} \parallel \mathcal{D} & & \parallel \mathcal{D} \\ g(X_n) & & g(X) \end{array}$$

GED

(14)

Skorohod's Representation Theorem also gives us extensions of Faton, 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 $E|X_n| \rightarrow E|X| < \infty$ and

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

Proof:

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

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

since $X_n = X_n^*$ the X_n^* 's are also UI.

\therefore by old UI Theorem we have

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

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

but again since $X_n^* = X_n$ & $X^* = X$

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

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

QED

Skorohod's Rep Thm also gives us the Delta method

(15)

Theorem: (Delta method)

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

$$c_n(X_n - x_0) \xrightarrow{D} Z \quad (\star)$$

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

If $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is differentiable ∂x_0 then

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

Remark: $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$ differentiable ∂x_0 means that

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

as $v \rightarrow 0$ 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 (\star) suggest

$$X_n \xrightarrow{D} x_0 + \underbrace{\frac{Z}{c_n}}_{\rightarrow 0 \text{ as } n \rightarrow \infty}$$

$$\text{so } g(X_n) \approx g(x_0 + \frac{Z}{c_n}) \xrightarrow{D} g(x_0) + Dg(x_0) \frac{Z}{c_n}$$

Proof:

(16)

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

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

Now

define this to be $g'(x_0)$
when $X_n^* - x_0 = 0$

$$c_n(g(X_n^*) - g(x_0))$$

$$= c_n(X_n^* - x_0) \cdot \underbrace{\left(\frac{g(X_n^*) - g(x_0)}{X_n^* - x_0} \right)}_{\xrightarrow{a.e.} g'(x_0) \text{ since by assumption}}$$

$$\lim_n (X_n^* - x_0) \xrightarrow{a.e.} \lim_n \frac{Z^*}{c_n} = 0$$

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

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

QED

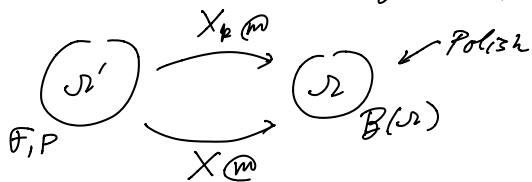
Slutsky

(17)

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 ω maps



Then

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

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

Remark:

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

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

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

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

Since

$$\begin{aligned} \limsup_n P(d(X_n, X) \geq \varepsilon) &\stackrel{\text{Fatou}}{\leq} P(d(X_n, X) \geq \varepsilon \text{ i.o.}) \\ &\leq \underbrace{P(d(X_n, X) \neq 0)}_{=0} \end{aligned}$$

Theorem: (Slutsky)

(18)

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 $B(\Omega)$) all defined on the same probability space. Then

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

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 $\forall \varepsilon > 0$

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

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

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

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

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

since $\bar{F}^\varepsilon \downarrow F$ as $\varepsilon \downarrow 0$

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

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 $\xrightarrow{a.e.}$, \xrightarrow{P} and a.e. convergence.

(19)

Corollary:

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

↑
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 $\exists X_n \neq X$ 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

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

and $X_n := (-1)^n X$ so that $X_n \xrightarrow{\mathcal{D}} X$
but $P(|X_n - X| \geq \varepsilon) = 1$ when n is odd.

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

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

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

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

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

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

a.e.d.

Here is another useful corollary of Slutsky's Thm

(20)

Corollary: Suppose X, X_n, Y_n are r.v.s

taking values in a polish space $(\mathcal{R}, \mathcal{B}(\mathcal{R}))$ all defined on the same probability space.

If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c \in \mathcal{R}$
then $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$.

Proof:

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

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

since $f(x, y) \in C_b(\mathcal{R} \times \mathcal{R})$ implies
 $f(x, c) \in C_b(\mathcal{R})$ so that
 $E f(X_n, c) \rightarrow E f(X, c)$.

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

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

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:

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

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

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

Tightness and Porhorov's Thm

(21)

An extremely 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. $x_{n_k} \xrightarrow{k \rightarrow \infty} x$

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 existence 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 $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then

\mathbb{P} is tight iff \exists a compact $K \subset \mathcal{X}$ 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

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

Theorem: (Porhorov)

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

\mathbb{P} is tight



$\left\{ \begin{array}{l} \forall P_1, P_2, \dots \in \mathbb{P} \text{ there exists a sub-seq } n_k \\ (\star) \text{ and a prob measure } P \text{ on } (\mathcal{X}, \mathcal{B}(\mathcal{X})) \\ \text{s.t. } P_{n_k} \xrightarrow{\mathcal{D}} P \end{array} \right.$

not necessarily in \mathbb{P} .

Note: If \mathbb{P} satisfies (\star) 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 probabilistic intuition so we will skip it and simply cite Billingsley's book on convergence of Probability measures.

(23)

Here is an example of how one uses Prohorov.

Theorem: (Portmanteau III)

Let X_1, X_2, \dots, X_n be r.v.s taking values in $C[0,1]$ w.r.t $\mathcal{B}(C[0,1])$. Then

The following are equivalent:

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

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

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

Proof:

$$(A) \xrightarrow{\quad} (B):$$

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.

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

by the continuous mapping theorem.

$$(B) \xrightarrow{\quad} (A): \text{ Assume}$$

$$(\ast) \pi(X_n) \rightarrow \pi(X), \forall \pi \in \{\pi_{t_1, \dots, t_m}: t_i \in [0,1]\}$$

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

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

By "sub-sub-seg check" it is sufficient to

$$\text{show } \forall n_k \exists n_{k_j} \text{ s.t. } X_{n_{k_j}} \xrightarrow[j \rightarrow \infty]{\mathcal{D}} X.$$

(24)

For a given n_p use $(\ast\ast)$ with Prohorov to find n_{k_j} and Y s.t.

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

By the continuous mapping thm

$$\pi(X_{n_{k_j}}) \xrightarrow{\mathcal{D}} \pi(Y), \forall \pi = \pi_{t_1, \dots, t_m}$$

$$\mathcal{D} \downarrow \text{by } (\ast)$$

$$\pi(X)$$

$$\therefore \pi(X) = \pi(Y), \forall \pi = \pi_{t_1, \dots, t_m} \text{ by uniqueness of limits.}$$

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

$$\therefore X_{n_{k_j}} \xrightarrow[j \rightarrow \infty]{\mathcal{D}} X \text{ by } (\ast\ast)$$

Since n_p was arb the "sub-sub check" Thm

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

QED

(25)

Here is another consequence of Prohorov.

Theorem: (Portmanteau IV)

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

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

(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

$$Eg(X_n) \rightarrow Eg(X) \quad \forall g \in \Gamma. \quad (*)$$

Proof:

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

(ii) \Rightarrow (i):

Let n_p be an arb sub-seq. By "sub-sub-seq check" it will be sufficient to show

$$\exists n_{p_j} \text{ s.t. } X_{n_{p_j}} \xrightarrow{\mathcal{D}} X \text{ as } j \rightarrow \infty.$$

Now by tightness & Prohorov's Thm

$$\exists Y \text{ & } \exists n_{p_j} \text{ s.t.}$$

$$X_{n_{p_j}} \xrightarrow{j \rightarrow \infty} Y \quad (**)$$

(26)

$$\therefore E f(X_{n_{p_j}}) \xrightarrow{j \rightarrow \infty} E f(Y) \quad \forall f \in G_b(\Omega).$$

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

$$Eg(X_{n_{p_j}}) \xrightarrow{j \rightarrow \infty} Eg(Y) \quad \forall g \in \Gamma$$

by (ii) ↓

$$Eg(X)$$

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

$$\therefore X \xrightarrow{\mathcal{D}} Y \text{ since } \Gamma \text{ separates}$$

$$\therefore X_{n_{p_j}} \xrightarrow{j \rightarrow \infty} X \text{ by } (**)$$

as was to be shown.

QED.

(27)

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

Theorem: (Portmanteau IV)

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

$$X_n \xrightarrow{\text{d}} X \text{ iff } Ef(X_n) \rightarrow Ef(X) \quad \forall f \in \Gamma$$

whenever

(I) $\Gamma = C_c(\mathcal{S})$ and \mathcal{S} is locally compact

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

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

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

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) and (II) choose compact K s.t.

$$P(X \in K) > 1 - \varepsilon_2$$

existence since \mathcal{S} is Polish so that

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

(28)

By similar techniques as in Lecture 13 we can find an approximation

$$f \in \Gamma \text{ to } I_K \text{ s.t.}$$

$$I_K \geq f \quad \& \quad \left| Ef(x) - \underbrace{E I_K(x)}_{= P(X \in K)} \right| \leq \frac{\varepsilon}{2}$$

$$\therefore P(X_n \in K) \geq Ef(X_n) \xrightarrow{n \rightarrow \infty} Ef(x) \quad \text{within } \varepsilon_2 \text{ of } P(X \in K) > 1 - \varepsilon_2$$

$$\therefore P(X_n \in K) \geq 1 - \varepsilon \quad \forall \text{ suff large } n.$$

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

For (III): $\{X_n\}_{n \geq 1}$ is tight since \mathcal{S} is compact.

For (IV): Trivial since $|X_n| = O_p(1)$ means

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

QED

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

Corollary: (characteristic functions for $\xrightarrow{\text{d}}$)

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

$$\text{Then } X_n \xrightarrow{\text{d}} X$$

if

$$\begin{aligned} & \left(\phi_n(k) \rightarrow \phi(k) \quad \forall k \in \mathbb{Z}^d \right) \\ & \text{and } |X_n| = O_p(1) \end{aligned} \quad (*)$$

Proof:

(29)

This follows immediately from
Portmanteau IV since $\phi_n(k) \rightarrow \phi(k) \forall k \in \mathbb{R}^d$
means $E f(X_n) \rightarrow E f(X)$ for all
 $f \in \{x \mapsto e^{ixk} : k \in \mathbb{R}^d\}$.

BED

This means that whenever $\phi_n(k) \rightarrow \phi(k)$

$\forall k \in \mathbb{R}^d$ it is sufficient to simply show

$X_n = O_p(1)$ for establishing
 $X_n \xrightarrow{d} X$.

Many of the conditions for $X_n \xrightarrow{d} X$

we learn as an undergrad are

simply conditions for establishing $(*)$.

Here are a few examples:

$$\left(\phi_n(k) \rightarrow \phi(k) \quad \forall k \in \mathbb{R}^d \right) \Rightarrow (*)$$

and $\sup_n E|X_n| < \infty$

Proof: Use Markov's Thm

$$\begin{aligned} P(|X_n| \geq N) &\leq \frac{E|X_n|}{N} \\ &\leq \frac{\sup_n E|X_n|}{N} < \infty \end{aligned}$$

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

$$P(X_n \notin \overline{B_N}(0)) < \varepsilon$$

ball of radius N

$$\therefore X_n = O_p(1).$$

(30)

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

Proof:

See Achim Klenke's book

"Probability theory" p. 309

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

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

Hence the previous result applies.

The next condition is both sufficient
and necessary ... its called The

Cramér-Wald device:-

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

Proof:

(\Leftarrow)

use the continuous mapping Thm.

(31)

 \Leftrightarrow write $X_n = (X_{n,1}, \dots, X_{n,d})$ andsuppose $\langle b, X_n \rangle \xrightarrow{\phi} \langle b, x \rangle$.

$$\therefore |\langle b, X_n \rangle| = \mathcal{O}_p(1) \text{ if}$$

by the "char fun check corollary"

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

 \therefore Given $\varepsilon > 0$ we can find $N_1, \dots, N_d > 0$

$$\text{s.t. } P(X_{n,i} \notin [-N_i, N_i]) < \frac{\varepsilon}{d}$$

for $i=1, \dots, d$ and hence

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

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

To finish note $\langle b, X_n \rangle \xrightarrow{\phi} \langle b, x \rangle$

implies

$$\underbrace{E e^{i \langle b, X_n \rangle}}_{= \phi_n(b)} \rightarrow \underbrace{E e^{i \langle b, x \rangle}}_{\phi(b)}$$

by the continuous mapping thm.

QED