

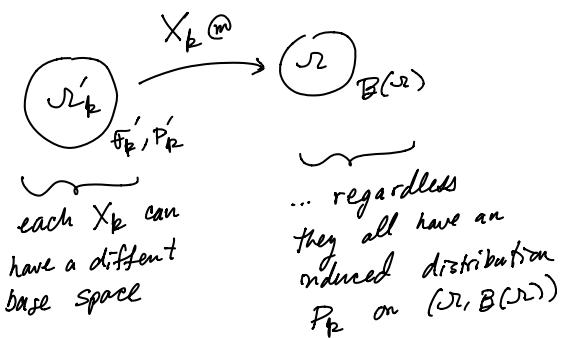
# Lecture 14: Convergence in distribution and the Central limit theorem

(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  $\mu$  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  $(\mathbb{R}', \mathcal{F}', P')$ ) 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)

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

so  $X_n \xrightarrow{\mathcal{D}} X$  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  $\xrightarrow{\mathcal{D}}$  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)  $\int_{\Omega} f dP_n \rightarrow \int_{\Omega} f dP$ ,  $\forall f \in \text{Lip}(\Omega) \cap C_b(\Omega)$

(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  $\int_{\Omega} f_\varepsilon dQ$  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  $P$

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

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

$\text{int } \sup \bar{P}_n(\bar{A})$   
 $\& \limsup_n P_n(\bar{A})$   
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)"      "(\*\*)"

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

$\square \quad \square$

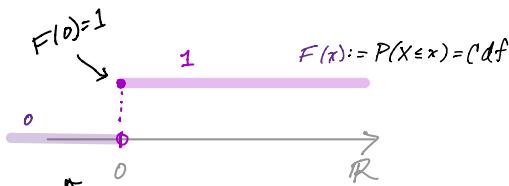
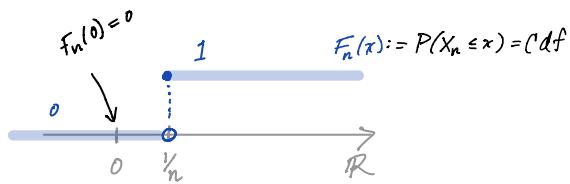
(5)

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  $u \in C_{F^{-1}}$  in lecture 8

Proof:

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

$(i) \Rightarrow (ii)$

(7)

$$\begin{aligned} X_n \xrightarrow{\mathcal{D}} X &\Leftrightarrow P_n \xrightarrow{\mathcal{D}} P \\ &\Rightarrow P_n((-\infty, x]) \rightarrow P((-\infty, x]) \quad \forall x \in C_F \\ &\text{since } x \in C_F \Rightarrow P(\partial(-\infty, x]) = 0 \\ &\Rightarrow F_n(x) \rightarrow F(x) \quad \forall x \in C_F \end{aligned}$$

$(ii) \Rightarrow (i)$

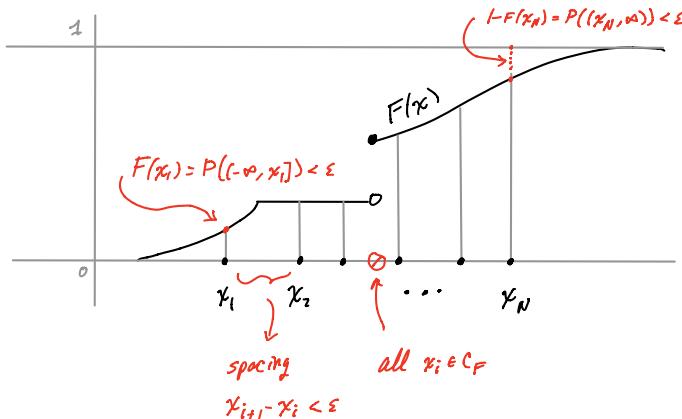
For the other direction suppose

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

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

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

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

$$A_2 = (x_1, x_2]$$

:

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

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

Note that by

design  $P(\partial A_i) = 0$

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

Now let  $f \in \text{Lip}(\mathbb{R}) \cap C_b(\mathbb{R})$ .

By rescaling we can suppose w.l.o.g. that

$$0 \leq f \leq 1.$$

(8)

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

Similarly

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

Also

$$\begin{aligned} \limsup_n \int_{A_2 \cup \dots \cup A_N} f dP_n \\ &= \limsup_n \sum_{i=2}^N \int_{A_i} f dP_n \end{aligned}$$

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

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

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

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

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

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

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

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  $n_{pk}$ :

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

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

Proof:

Suppose not, i.e.  $\exists f \in C_b(\Omega)$  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 (11)

$$p_i \text{ so that } \int_{\Omega} f dP_{n,p_i} \rightarrow \int_{\Omega} f dP.$$

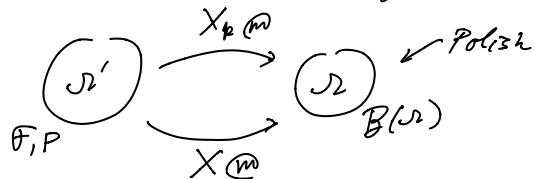
$\therefore$  contradiction

QED

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

Let  $X, X_1, \dots$  be a collection of  $\mathbb{C}^m$  maps



Then

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

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

Remark:

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

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

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

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

Since

$$\begin{aligned} \limsup_n P(d(X_n, X) \geq \varepsilon) &\stackrel{\text{Fact}}{\leq} P(d(X_n, X) \geq \varepsilon \text{ i.o.}) \\ &\leq \underbrace{P(d(X_n, X) \not\rightarrow 0)}_{=0} \end{aligned}$$

Theorem: (Slutsky) (12)

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  $\Omega$  (with  $\sigma$ -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  $\forall \varepsilon > 0$

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

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

$$\leq \limsup_n P(d(X_n, Y_n) \geq \varepsilon) + \limsup_n P(X_n \in F^\varepsilon)$$

$\underbrace{\quad}_{=0 \text{ by assumption}}$

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

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

Since  $F^\varepsilon \downarrow F$  as  $\varepsilon \downarrow 0$

$$P(X \in F^\varepsilon) \downarrow P(X \in F)$$

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.

(13) Slutsky's Theorem gives us a nice corollary that relates  $\xrightarrow{d}$ ,  $\xrightarrow{P}$  and a.e. convergence.

Corollary:

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

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

Warning: The reverse implications do not hold in general.

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

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

To see why  $\xrightarrow{d} \not\Rightarrow \xrightarrow{P}$  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  $X_n \xrightarrow{d} X$

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

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

## Skorokhod Representation

(14)

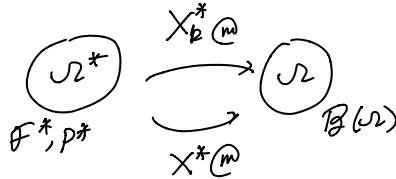
Theorem: (Skorokhod Representation Theorem)

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

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

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

s.t.



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

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

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

Notation: In the above theorem

I'm using  $\mathcal{L}(X^*)$ ,  $\mathcal{L}(X_n^*)$  as a shorthand for  $D^*(X^*)^{-1}$  &  $P^*(X_n^*)^{-1}$

i.e. the induced distributions on  $(\Omega, \mathcal{B}(\Omega))$

I'll use this notation, when convenient, throughout the rest of the course.

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.  $X_n^* \xrightarrow{\text{a.e.}} X^*$   
 $\parallel d \quad \parallel d$   
 $X_n \quad X$

The proof of the general Skorokhod's result is nasty and doesn't really provide any insight. We will just show the case  $\mathcal{D} = \mathbb{R}$  and cite the general result in Billingsley's book "Convergence of Prob. Meas." (15)

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

Define c.d.f.s

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

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

By Portmanteau II

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

Let  $\mathcal{D}^* = (0, 1)$ ,  $\mathcal{F}^* = \mathcal{B}((0, 1))$ ,  $P^* = \mathcal{L}'$

and set

$$X_n^*(u) := F_n^{-1}(u) \quad \text{uniform measure on } (0, 1).$$

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

$\forall u \in \mathcal{D}^*$ .

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

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

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

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

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

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

QED

The following result demonstrates Skorokhod's usefulness. (16)

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

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

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

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

then

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

Proof:

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

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

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

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

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

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

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

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

$$|| \mathcal{D} \quad || \mathcal{D}$$

$$g(X_n) \quad g(X)$$

QED

## Skorohod's Representation Theorem

(17)

also gives us extensions of Fatou, DCT, ... which apply to the case  $X_n \xrightarrow{d} X$ .

Here is an example.

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

If  $X_n \xrightarrow{d} X$  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

Finally lets use Skorohod's

Rep Thm to prove the "Delta-method"

Theorem: (Delta method)

(18)

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

$$c_n(X_n - x_0) \xrightarrow{d} Z \quad (*)$$

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

If  $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is differentiable at  $x_0$  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 at  $x_0$  means that

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

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

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

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

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

Proof:

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

$$c_n(X_n^* - x_0) \xrightarrow{a.e.} Z^* \quad \text{define this to be } g'(x_0)$$

Now

$$c_n(g(X_n^*) - g(x_0)) = \left( \frac{g(X_n^*) - g(x_0)}{X_n^* - x_0} \right) c_n(X_n^* - x_0)$$

Let  $n \rightarrow \infty$  so that  $X_n^* - x_0 \xrightarrow{a.e.} 0$  which implies

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

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

$$\therefore c_n(g(X_n) - g(x_0)) \xrightarrow{d} g'(x_0) Z \quad \text{QED.}$$

## Tightness and Prohorov's Thm

(19)

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.

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

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

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

The following result shows this is the "right" definition of boundedness.

### Theorem: (Prohorov)

(20)

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

$\mathcal{P}$  is tight



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

not necessarily in  $\mathcal{P}$ .

Note: If  $\mathcal{P}$  satisfies \* then  $\mathcal{P}$  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.

Our main use of Prohorov's Thm will be through the following consequence

### Theorem: (tightness + separating class)

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

If  $\{P_n\}_{n \geq 1}$  is tight and there exists a separating family  $\Gamma \subset C_b(\mathbb{R})$  s.t.

$$\int f dP_n \rightarrow \int f dP \text{ iff } f \in \Gamma$$

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

Proof:

(21)

Assume  $P_n \not\rightarrow P$ .

$\therefore \exists f \in C_b(\mathbb{R})$  & a sub-seq  $n_k$  s.t.

$$\left| \int_{\mathbb{R}} f dP_{n_k} - \int_{\mathbb{R}} f dP \right| > \varepsilon \quad \forall k \in \mathbb{N}. \quad (a)$$

If  $\{P_n\}$  is tight Prokhorov's Theorem implies there exists a sub-sub-seq  $n_{k_j}$  and a measure  $G$  on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  s.t.

$$P_{n_{k_j}} \xrightarrow{d} G \quad \text{as } j \rightarrow \infty$$

$$\therefore \int_{\mathbb{R}} f dP_{n_{k_j}} \xrightarrow{j \rightarrow \infty} \int_{\mathbb{R}} f dG \quad \forall f \in C_b(\mathbb{R}) \quad (b)$$

Now by assumption we have

$$\int_{\mathbb{R}} f dP_{n_{k_j}} \xrightarrow{j \rightarrow \infty} \int_{\mathbb{R}} f dP \quad \forall f \in \Gamma \quad (c)$$

Combining (b) and (c) gives

$$P = G$$

since  $\Gamma$  is a separating class &  $\Gamma \subset C_b(\mathbb{R})$ .

Now (a) and (b) contradict each other.

QED

We get moderate utility out of this result for another Portmanteau Theorem

Theorem: (Portmanteau III)

(22)

Let  $P, P_1, P_2, \dots$  be probability measures on a Polish  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Then

$$P_n \xrightarrow{d} P \text{ iff } \int_{\mathbb{R}} f dP_n \rightarrow \int_{\mathbb{R}} f dP \quad \forall f \in \Gamma$$

whenever

(I)  $\Gamma = C_c(\mathbb{R})$  &  $\mathbb{R}$  is locally compact

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

(III)  $\Gamma = \{e^{ix \cdot k} : k \in \mathbb{R}^d\}$  &  $\mathbb{R} = \mathbb{R}^d$  &  
 $\sup_n \int_{\mathbb{R}^d} |x|^2 dP_n(x) < \infty$

(IV)  $\Gamma = \{\text{monomials}\}$  &  $\mathbb{R}$  is a compact  
subset of  $\mathbb{R}^d$

Theorem: (Portmanteau IV)

O

O

Let  $P, P_1, P_2, \dots$  be probability measures on  $(C[0,1], \mathcal{B}(C[0,1]))$ .

Then the following are equivalent:

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

$$(B) P_{\pi_{t_1, \dots, t_m}^{-1}} \xrightarrow{\mathcal{D}} P_{\pi_{t_1, \dots, t_m}^{-1}} \quad \forall t_1, \dots, t_m \in [0,1]$$

The central limit theorem

0

0

Kolmogorov's 3 series theorem

O

O

O

O

O

O