- <u>Definition</u> (Converges weakly, Convergence in distribution)

Define F_n converges weakly to F, written $F_n \Rightarrow F$, if

$\lim_n F_n(x) = F(x)$ for every x at which F is continuous.

$F_n \Rightarrow F$ iff $\lim_n \mu_n(A) = \mu(A)$ for $A = (-\infty, x]$ for which $\mu\{x\} = 0$, written $\mu_n \Rightarrow \mu$.

Define X_n converges in distribution to X, written $X_n \Rightarrow X$, if $F_n \Rightarrow F$.

Remark:

- The r.v.'s X_n may be defined on entirely different $(\Omega_n, \mathcal{F}_n, P_n)$.
- To study the approximate distribution for a r.v. Y_n it is often necessary to study instead the normalized r.v. $\frac{Y_n - b_n}{a_n}$ for appropriate a_n and b_n.

  If Y_n has distribution function F_n and if $a_n > 0$, then

  $P\left[\frac{Y_n - b_n}{a_n} < x\right] = P[Y_n < a_n x + b_n] = F_n(a_n x + b_n)$, i.e.

  $\frac{Y_n - b_n}{a_n}$ has distribution function $F_n(a_n x + b_n)$.

  So weak convergence often appears in the form

  $F_n(a_n x + b_n) \Rightarrow F(x)$.

- <u>Proposition</u> (Discontinuity Points of F, Unique Weak Limit)

(a) F have only countably many points of discontinuity; and hence the set of continuity points is dense.

(b) A sequence F_n can have at most one weak limit.

pf: (a) If the jumps $\mu\{x\} = F(x) - F(x-)$ exceeds ϵ at $x_1 < \cdots < x_n$,

then $F(x_i) - F(x_{i-1}) > \epsilon$ (take $x_0 < x_1$), and so

$n\epsilon \le F(x_n) - F(x_0) \le 1$, $n \le 1/\epsilon$; hence

$\{x : \mu\{x\} > \epsilon\}$ is finite and $\{x : \mu\{x\} > 0\}$ is countable.

(b) Suppose that $F_n \Rightarrow F$ and $F_n \Rightarrow G$. Then

$F(x) = \lim_n F_n(x) = G(x)$ if F and G are both continuous at x.

Since F and G each have only countably many points of discontinuity, the set of common continuity points is dense, and it follows by right continuity that F and G are equal. #

- <u>Definition</u> (Type, Degenerate Distribution Function)

(a) Distribution functions F and G are of the same "type" if there exists constants a and b, $a>0$, s.t.

$F(ax+b) = G(x)$ for all x. $\Rightarrow$ location-scale family: $\frac{X-b}{a} \overset{d}{=} Y$; $F(x) = G(\frac{x-b}{a})$

(b) A distribution function is "degenerate" if it has the form

$$\Delta(x-x_0) = \begin{cases} 0 & \text{if } x-x_0<0;\ x<x_0 \\ 1 & \text{if } x-x_0\ge 0;\ x\ge x_0. \end{cases}$$ for some x_0; $\Rightarrow$ has unit mass at x_0.

Otherwise it is nondegenerate.

- <u>Theorem</u> (Convergence of Types Thm)

Suppose that $F_n(u_nx+v_n) \Rightarrow F(x)$, $F_n(a_nx+b_n) \Rightarrow G(x)$, where $u_n>0$, $a_n>0$, and F and G are nondegenerate.

Then there exist a and b, $a>0$, s.t.

$a_n/u_n \to a$, $(b_n-v_n)/u_n \to b$, and

$F(ax+b) = G(x)$.

Remark:

- Thus there can only one possible limit type and essentially one possible sequence of norming constants.
- $F_n(u_nx+v_n) \Rightarrow F(x)$,

$F_n(u_n(\frac{a_n}{u_n}x + \frac{b_n-v_n}{u_n}) + v_n) \Rightarrow F(ax+b)$ by $\frac{a_n}{u_n} \to a$, $\frac{b_n-v_n}{u_n} \to b$

$= F_n(a_nx+b_n) \Rightarrow G(x)$

$F(ax+b) = G(x)$ by $F_n(a_nx+b_n)$ can only have one weak limit.

• **Definition** (Distribution Function of the Maxima $M_n = \max\{X_1, \ldots, X_n\}$)

To find the distribution function of the maxima

$M_n = \max\{X_1, \cdots X_n\}$,

(i) If the X_n are independent r.v.'s:

$$F_n(x) = P[M_n \le x] = P[X_i \le x, i=1, \ldots, n] = \prod_{i=1}^{n} P[X_i \le x]$$

(ii) If the X_n are iid r.v.'s with distribution function G:

$$F_n(x) = \prod_{i=1}^{n} P[X_i \le x] = G^n(x).$$

• **Definition** (Extremal Distributions)

A distribution function F is called a "extremal distribution" if

(i) it is nondegenerate; and

(ii) for some distribution function G and constants $a_n > 0$ and b_n,

$$G^n(a_n x + b_n) \Rightarrow F(x)$$

Remark:

- These are the possible limiting distributions of normalized maxima.
- **Theorem.** (Extremal Distribution Functions)

  The class of extremal distribution functions consists exactly of the distribution functions of the 3 types below:

  (i) $F_1(x) = e^{-e^{-x}}, \quad x > 0$

  (ii) $F_{2,\alpha}(x) = \begin{cases} 0 & \text{if } x < 0 \\ e^{-x^{-\alpha}} & \text{if } x \ge 0 \end{cases}$

  (iii) $F_{3,\alpha}(x) = \begin{cases} e^{-(-x)^{\alpha}} & \text{if } x \le 0 \\ 1 & \text{if } x \ge 0. \end{cases}$

- <u>Example</u> (Extremal Distribution of Type 1)
  - Suppose the X_n are iid exponentially distributed r.v.'s with df. G s.t.

    $G(x) = 1 - e^{-\alpha x}$, $x \geq 0$.
  - Let $M_n = \max\{X_1, \dots, X_n\}$ with df F_n s.t

    $F_n(x) = G^n(x) = (1 - e^{-\alpha x})^n$, $x \geq 0$.

    For each x, $\lim_n F_n(x) = 0$,

    which means that M_n tends to be large for n large.
  - But $P[M_n - \alpha^{-1}\log n \leq x] = F_n(x + \alpha^{-1}\log n)$,

    this is the distribution function of $M_n - \alpha^{-1}\log n$, and

    $F_n(x + \alpha^{-1}\log n) = (1 - e^{-(\alpha x + \log n)})^n$

    $= (1 - \frac{1}{n} e^{-\alpha x})^n \to e^{-e^{-\alpha x}}$ as $n \to \infty$. if $\log n \geq -\alpha x$.

    So the limit holds for all x.
  - This gives for large n the approximate distribution of the normalized r.v. $M_n - \alpha^{-1}\log n$:

    $\lim_n F_n(x + \alpha^{-1}\log n) = e^{-e^{-\alpha x}}$ for $-\infty < x < \infty$.

    where the norming constants $a_n = 1$, $b_n = \alpha^{-1}\log n$.

#

- <u>Example</u> (Extremal Distribution of Type 2)

- Suppose the X_n are iid with distribution function G,

$$G(x) = \begin{cases} 0 & \text{if } x \le 1 \\ 1 - x^{-\alpha} & \text{if } x \ge 1 \end{cases} \quad \text{where } \alpha > 0.$$

- Let $M_n = \max\{X_1, \cdots, X_n\}$, with distribution function F_n,

$$F_n(x) = P[M_n \le x] = G^n(x) = (1 - x^{-\alpha})^n, \quad x \ge 1$$

- $$P[n^{-1/\alpha} M_n \le x] = F_n(n^{1/\alpha} x)$$
$$= \left(1 - \tfrac{1}{n} x^{-\alpha}\right)^n \quad \text{for } n^{1/\alpha} x \ge 1,\ x \ge n^{-1/\alpha}$$
$$\to e^{-x^{-\alpha}} \quad \text{as } n \to \infty \text{ if } x \ge 0$$

- This gives for large n the approximate distribution of the normalized r.v. $n^{-1/\alpha} M_n$:

$$\lim_n F_n(n^{1/\alpha} x) = \begin{cases} 0 & \text{if } x < 0 \\ e^{-x^{-\alpha}} & \text{if } x \ge 0 \end{cases}$$

where the norming constants $a_n = n^{1/\alpha}$, $b_n = 0$. #

- Example (Extremal Distribution of Type 3)
  - Suppose the X_n are iid with distribution function G,

$$G(x)=\begin{cases}0 & \text{if } x\le 0\\ 1-(1-x)^{\alpha} & \text{if } 0\le x\le 1\\ 1 & \text{if } x\ge 1\end{cases} \quad \text{where } \alpha>0.$$

  - Let $M_n=\max\{X_1,\dots,X_n\}$ with distribution function F_n

$$F_n(x)=P[M_n\le x]=G^n(x)$$

  - $$F_n(n^{-1/\alpha}x+1)=(1-(1-n^{-1/\alpha}x-1)^{\alpha})^n$$
$$=(1-\tfrac{1}{n}(-x)^{\alpha})^n \quad \text{if } 0\le n^{-1/\alpha}x+1\le 1,\ -n^{1/\alpha}\le x\le 0.$$
$$\to e^{-(-x)^{\alpha}} \quad \text{as } n\to\infty \text{ if } x\le 0.$$

  - $$F_n(n^{-1/\alpha}x+1)=P[M_n\le n^{-1/\alpha}x+1]$$
$$=P[n^{1/\alpha}M_n-n^{1/\alpha}\le x]$$

  This gives for large n the approximate distribution of the normalized r.v. $n^{1/\alpha}M_n-n^{1/\alpha}$:

$$\lim_n F_n(n^{-1/\alpha}x+1)=\begin{cases}e^{-(-x)^{\alpha}} & \text{if } x\le 0\\ 1 & \text{if } x\ge 0.\end{cases}$$

  where the norming constants $a_n=n^{-1/\alpha}$, $b_n=1$.

#

- **Definition** (Total Variation Convergence)

Suppose F_n, F are distribution functions.

Define $F_n \to F$ in total variation if

$$\sup_{A \in \mathcal{B}(\mathbb{R})} |F_n(A) - F(A)| \to 0.$$

- **Theorem** (Total Variation Convergence $\overset{\Rightarrow}{\not\Leftarrow}$ Weak Convergence)

$F_n \to F$ in total variation $\overset{\Rightarrow}{\not\Leftarrow}$ $F_n \Rightarrow F$

Pf: • If $F_n \to F$ in total variation,

$$\sup_{A \in \mathcal{B}(\mathbb{R})} |F_n(A) - F(A)| \to 0.$$

For $A = (-\infty, x]$ for each continuity points x of F,

$$F_n((-\infty, x]) - F((-\infty, x]) = F_n(x) - F(x) \to 0, \quad F_n(x) \to F(x).$$

• We show the converse does not hold.

Suppose F_n puts mass $\frac{1}{n}$ at points $\{\frac{1}{n}, \frac{2}{n}, \cdots, \frac{n}{n}\}$.

If $F(x) = x$, $0 \le x \le 1$ is the uniform distribution on $[0,1]$,

then for $x \in (0,1)$, $F_n(x) = \frac{[nx]}{n} \to x = F(x)$,

thus we have weak convergence $F_n \Rightarrow F$.

However, if Q is the set of rationals in $[0,1]$,

$$F_n(Q) = 1 \not\to F(Q) = 0, \text{ so } \sup_{A \in \mathcal{B}(\mathbb{R})} |F_n(A) - F(A)| \not\to 0.$$

#

• Theorem (Scheffé's Lemma)

Suppose $\{F, F_n, n\geq 1\}$ are probability distribution functions with densities $\{f, f_n, n\geq 1\}$. Then

$$\sup_{B\in\mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| = \frac{1}{2}\int |f_n(x) - f(x)|\,dx \quad (*).$$

If $f_n \to f$ a.e. (that is, for all x except a set of Lebesgue measure 0), then $\int |f_n(x) - f(x)|\,dx \to 0$

and thus $F_n \to F$ in total variation and hence $F_n \Rightarrow F$.

Remark:

$$f_n \to f \text{ a.e} \Leftarrow\!\!\!\!/\; F_n \to F \text{ in total variation} \overset{\Rightarrow}{\not\Leftarrow} F_n \Rightarrow F.$$

Pf: • Let $B\in\mathcal{B}(\mathbb{R})$. Then $\int (f_n(x) - f(x))\,dx = 1-1=0$.

$\because\quad 0 = \int_B (f_n(x) - f(x))\,dx + \int_{B^c}(f_n(x) - f(x))\,dx,$

$$\left|\int_B (f_n(x) - f(x))\,dx\right| = \left|\int_{B^c}(f_n(x) - f(x))\,dx\right|$$

• Then, $2|F_n(B) - F(B)| = 2\left|\int_B (f_n(x) - f(x))\,dx\right|$

$$= \left|\int_B (f_n(x) - f(x))\,dx\right| + \left|\int_{B^c}(f_n(x) - f(x))\,dx\right|$$

$$\leq \int_B |f_n(x) - f(x)|\,dx + \int_{B^c}|f_n(x) - f(x)|\,dx$$

$$= \int |f_n(x) - f(x)|\,dx$$

$$\therefore \sup_{B\in\mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| \leq \frac{1}{2}\int |f_n(x) - f(x)|\,dx$$

• If we find some set B for which equality actually holds, then we will have shown $(*)$. Set $B = [f_n \geq f]$, then

$$2|F_n(B) - F(B)| = \left|\int_B (f_n(x) - f(x))\,dx\right| + \left|\int_{B^c}(f_n(x) - f(x))\,dx\right|$$

$$= \int_B |f_n(x) - f(x)|\,dx + \int_{B^c}|f_n(x) - f(x)|\,dx$$

$\because\ f_n(x) - f(x) \geq 0$ for B, $f_n(x) - f(x) < 0$ for B^c.

$$= \int |f_n(x) - f(x)|\,dx.$$

So equality holds in $(*)$.

(next pg. cont.)

- Now suppose $f_n \to f$ a.e. So $f - f_n \to 0$ a.e., $(f-f_n)^+ \to 0$ a.e.
- Since if $f - f_n \geq 0$, $(f-f_n)^+ = f - f_n \leq f$; if $f - f_n < 0$, $(f-f_n)^+ = 0 < f$,

  so $(f-f_n)^+ \leq f \in L_1$ $\because$ f is a density.

  Since $0 = \int (f(x) - f_n(x))\,dx = \int (f(x) - f_n(x))^+ dx - \int (f(x) - f_n(x))^- dx$,

$$\int |f_n(x) - f(x)|\,dx = \int (f(x) - f_n(x))^+ dx + \int (f(x) - f_n(x))^- dx$$

$$= 2\int (f(x) - f_n(x))^+ dx \to 0 \text{ by DCT.}$$

#

- Remark:
  - $F_n \to F$ in total variation:

    $\sup_{B \in \mathcal{B}(\mathbb{R})} |F_n(B) - F(x)| \to 0$.
  - Scheffe's lemma

    (i) $\sup_{B \in \mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| = \frac{1}{2}\int |f_n(x) - f(x)|\,dx$.

    (ii) Suppose $f_n \to f$ a.e., then $f - f_n \to 0$, so

    ① $(f-f_n)^+ \to 0$

    ② $(f-f_n)^+ \leq f \in L_1$,

    by the dominated convergence thm,

    $\Rightarrow \int |f_n(x) - f(x)|\,dx = 2\int (f(x) - f_n(x))^+ dx \to 0$.

    (iii) $\Rightarrow \sup_{B \in \mathcal{B}(\mathbb{R})} |F_n(B) - F(B)| \to 0$, so

    $F_n \to F$ in total variation.

    (iv) Hence $F_n \Rightarrow F$.

• **Theorem** (Convergence a.s. $\Rightarrow$ Convergence i.p. $\Rightarrow$ Weak Convergence)

Suppose that X_n and X are r.v.'s on the same probability space.
If $X_n \to_{a.s.} X$, then $X_n \to_p X$. If $X_n \to_p X$, then $X_n \Rightarrow X$.

pf: It suffices to show that if $X_n \to_p X$, then $X_n \Rightarrow X$

• Suppose $X_n \to_p X$, i.e. $\lim_n P[|X_n - X| > \epsilon] = 0$ for each ϵ

• $P[X_n \le x]$

$$= P[X_n \le x, |X_n - X| \le \epsilon] + P[X_n \le x, |X_n - X| > \epsilon]$$

$$\le P[X \le x+\epsilon, |X_n - X| \le \epsilon] + P[|X_n - X| > \epsilon]$$

$$\le P[X \le x+\epsilon] + P[|X_n - X| > \epsilon]$$

• So replace X_n by X, x by $x-\epsilon$,

$$P[X \le x-\epsilon] \le P[X_n \le x] + P[|X_n - X| > \epsilon].$$

Hence

• $$P[X \le x-\epsilon] - P[|X_n - X| > \epsilon] \le P[X_n \le x] \le P[X \le x+\epsilon] + P[|X_n - X| > \epsilon]$$

Letting n tend to ∞ and letting ϵ tend to 0 shows that

$$P[X < x] \le \liminf_n P[X_n \le x] \le \limsup_n P[X_n \le x] \le P[X \le x].$$

• Thus $P[X_n \le x] \to P[X \le x]$ if $P[X = x] = 0$, and so $X_n \Rightarrow X$.

#

★ • **Example** (Weak Convergence but NOT Convergence i.p.)

1. If X and Y are independent and assume the values 0 and 1 with probability $\frac{1}{2}$ each, and if $X_n = Y$, then $X_n \Rightarrow X$, but $X_n \to_p X$ cannot hold because $P[|X-Y| = 1] = (\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{2}$.
2. $\lim_n P[|X_n - X| > \epsilon] = 0$ is impossible if X and the X_n are defined on different probability spaces, as may happen in the case of convergence in distribution.

#

- <u>Theorem</u> (Convergence i.p. to a iff Convergence in distribution to a)

$X_n \to_p a$ iff $X_n \Rightarrow a$, that is,

$\lim_n P[|X_n - a| > \epsilon] = 0$ iff $\lim_n P[X_n \le x] = \begin{cases} 0 & \text{if } x < a \\ 1 & \text{if } x \ge a. \end{cases}$

Remark:

This condition, $\lim_n P[|X_n - a| > \epsilon] = 0$, makes sense even if the space of X_n does vary with n. Now a can be regarded as a r.v. (on any probability space at all).

pf: • Suppose $\lim_n P[|X_n - a| > \epsilon] = 0$. Put $\epsilon = |x-a|$.

- If $x > a$,

$$P[X_n \le x] = P[X_n - a \le x - a] = P[X_n - a \le |x-a|] \quad \because |x-a| = x-a.$$
$$\ge P[|X_n - a| \le \epsilon] \to 1 \quad \because X_n - a \le |X_n - a|.$$

- If $x < a$,

$$P[X_n \le x] = P[X_n - a \le x - a] = P[X_n - a \le -|x-a|]$$
$$\le P[-|X_n - a| \le -|x-a|] \quad \because X_n - a \ge -|X_n - a|$$
$$= P[|X_n - a| \ge \epsilon] \to 0$$

• Conversely, suppose $X_n \Rightarrow a$.

$$P[|X_n - a| > \epsilon] = P[X_n - a > \epsilon] + P[X_n - a < -\epsilon]$$
$$= P[X_n > a + \epsilon] + P[X_n < a - \epsilon]$$
$$= \{1 - P[X_n \le a + \epsilon]\} + P[X_n < a - \epsilon]$$
$$\to (1-1) + 0 = 0.$$

#

- <u>Theorem</u> (Skorohod's Thm)

Suppose $\{X, X_n, n\geq 1\}$ are r.v.'s with distribution functions $\{F, F_n, n\geq 1\}$, and $F_n \Rightarrow F$ (and hence $X_n \Rightarrow X$). Then there exists r.v.'s $\{Y, Y_n, n\geq 1\}$ on a common probability space $(\Omega, \mathcal{F}, P)$

s.t. for $n\geq 1$, $Y_n \stackrel{d}{=} X_n, n\geq 1$, $Y \stackrel{d}{=} X$, and $Y_n \to_{a.s.} Y$.

Pf: Set $(\Omega, \mathcal{F}, P) = ((0,1), \mathcal{B}((0,1)), P = \text{Lebesgue measure})$.

For $0<\omega<1$, put $Y_n(\omega) = \inf[x : \omega \leq F_n(x)]$ and $Y(\omega) = \inf[x : \omega \leq F(x)]$.

That is, Y_n and Y are "quantile functions" of F_n and F.

Since $\omega \leq F_n(x)$ iff $Y_n(\omega) \leq x$,

$P[\omega : Y_n(\omega) \leq x] = P[\omega : \omega \leq F_n(x)] = F_n(x)$.

Thus Y_n has distribution function F_n; similarly, Y has distribution function F.

It remains to show that $Y_n(\omega) \to Y(\omega)$ a.s.

Suppose that $0<\omega<1$. Given $\epsilon>0$, choose x s.t.

$Y(\omega)-\epsilon < x < Y(\omega)$ and $P[X=x]=0$, i.e. continuity point of F.

Then $F(x)<\omega$; $F_n(x) \to F(x)$ now implies that, for n large enough,

$F_n(x)<\omega$, then $Y_n(\omega) > x$, and hence $Y(\omega)-\epsilon < x < Y_n(\omega)$.

Thus $Y(\omega) \leq \liminf_n Y_n(\omega)$.

If $\omega<\omega'$ and $\epsilon>0$, choose y s.t.

$Y(\omega') < y < Y(\omega')+\epsilon$ and $P[X=y]=0$, i.e. continuity point of F.

Now $\omega<\omega' \leq F(Y(\omega')) \leq F(y)$, and so, for n large enough,

$F_n(y) \geq \omega$, then $Y_n(\omega) \leq y$, and hence $Y_n(\omega) \leq y < Y(\omega')+\epsilon$.

Thus $\limsup_n Y_n(\omega) \leq Y(\omega')$ if $\omega<\omega'$.

Let $\omega' \downarrow \omega$, we therefore have $Y_n(\omega) \to Y(\omega)$ if Y is continuous at ω.

Since Y is nondecreasing on $(0,1)$, it has at most countably many discontinuities.

Hence $Y_n(\omega) \to Y(\omega)$ for all ω but a null set.

#

- **Theorem** (The Mapping Thm)

  Suppose that $h: R^1 \to R^1$ is measurable and that the set D_h of its discontinuities is measurable.
  If $\mu_n \Rightarrow \mu$ (that is, $F_n \Rightarrow F$) and $\mu(D_h)=0$, then
  $\mu_n h^{-1} \Rightarrow \mu h^{-1}$

  Pf: · By Skorohod's thm, there exist
  $Y_n \overset{d}{=} X_n$, $Y \overset{d}{=} X$ s.t. $Y_n \to_{a.s.} Y$.

  · if $Y_n(\omega) \notin D_h$, then $h(Y_n(\omega)) \to h(Y(\omega))$ a.s.

  · Hence $h(Y_n) \Rightarrow h(Y)$.

  · Since $P[h(Y) \in A] = P[Y \in h^{-1}A] = \mu(h^{-1}A)$,
  $h(Y)$ has distribution μh^{-1};
  similarly, $h(Y_n)$ has distribution $\mu_n h^{-1}$.

  · Thus $h(Y_n) \Rightarrow h(Y)$ is the same thing as $\mu_n h^{-1} \Rightarrow \mu h^{-1}$.

  #

- **Corollary** (Cor's of the Mapping Thm)

  (i) If $X_n \Rightarrow X$ and $P[X \in D_h]=0$, then $h(X_n) \Rightarrow h(X)$

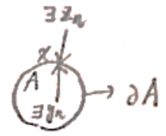
  (ii) If $X_n \Rightarrow a$ and h is continuous at a, then $h(X_n) \Rightarrow h(a)$.

• **Definition** (The Boundary of Set, μ-Continuity Set).

∂A = the "boundary" of A

$= A^- \setminus A^\circ$ = the closure of A minus the interior of A

$= \{x : \exists y_n \in A \text{ s.t. } y_n \to x \text{ and } \exists z_n \in A^c \text{ s.t. } z_n \to x\}$



= the points that are limits of sequences in A and are also limits of sequences in A^c.

Define A is a "μ-continuity set" if $A \in \mathcal{R}'$ and $\mu(\partial A) = 0$.

• **Theorem** (Portmanteau Thm: Charaterizing Weak Convergence)

The following three conditions are equivalent:

(i) $\mu_n \Rightarrow \mu$ (that is, $F_n \Rightarrow F$, with corresponding df's F_n, F);

(ii) $\int f d\mu_n \to \int f d\mu$ for every bounded, continuous real ft f; $E[f(X_n)] \to E[f(X)]$.

(iii) $\mu_n(A) \to \mu(A)$ for every μ-continuity set A.

Remark:

(ii)' $\int f d\mu_n \to \int f d\mu$ for every bounded, uniformly continuous real ft f;

Pf: • (i) $\Rightarrow$ (ii). By Skorohod's Thm, consider the r.v.'s Y_n and Y with Y_n has distribution μ_n, Y has distribution μ, s.t. $Y_n \to_{a.s.} Y$. Suppose that f is a bounded function s.t. $\mu(D_f) = 0$, where D_f is the set of discontinuity points of f. From $P[Y \in D_f] = \mu(D_f) = 0$ it follows that $f(Y_n) \to_{a.s.} f(Y)$. By the bounded convergence thm,

$$\int f d\mu_n = E[f(Y_n)] \to E[f(Y)] = \int f d\mu.$$ Thus (ii) holds.

• (i) $\Rightarrow$ (iii). Let $f = I_A$. A is a μ-continuity set, then $\partial A = D_f$. From $\mu(\partial A) = \partial(D_f) = 0$ and $\mu_n \Rightarrow \mu$ follows

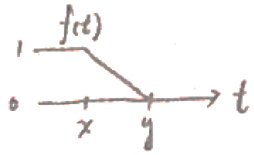
$$\mu_n(A) = \int f d\mu_n \to \int f d\mu = \mu(A).$$ Thus (iii) holds.

• (iii) $\Rightarrow$ (i). Let x be s.t. $\mu\{x\} = 0$. Since $\partial(-\infty, x] = \{x\}$, it is μ-continuity set, then $\mu_n(-\infty, x] = F_n(x) \to F(x) = \mu(-\infty, x]$. Thus (i) holds.

(next pg. cont.)

• (ii) ⇒ (i). Consider the corresponding distribution functions F_n, F.
Suppose that $x<y$, and let

$$f(t) = \begin{cases} 1 & \text{for } t \le x \\ \frac{y-t}{y-x} & \text{for } x \le t \le y \\ 0 & \text{for } t \ge y \end{cases}$$



Then f is a bounded and (uniformly) continuous function.

• Since $F_n(x) = \int I_{(-\infty,x]}\, d\mu_n \le \int f\, d\mu_n$ and $\int f\, d\mu \le \int I_{(-\infty,y]}\, d\mu = F(y)$,
it follows from (ii) that $\limsup_n F_n(x) \le F(y)$;

• letting $y \downarrow x$ shows that $\limsup_n F_n(x) \le F(x)$.
Similarly, $F(x-) \le \liminf_n F_n(x)$.
This implies convergence at continuity points of F. Thus (i). #

- <u>Theorem</u> (Slutsky's Thm)

If $X_n \Rightarrow X$, and $Y_n \xrightarrow{P} c \in \mathbb{R}$ (iff $Y_n \Rightarrow c$), then.

(a) $X_n + Y_n \Rightarrow X + c$

(b) $X_n Y_n \Rightarrow Xc$

Pf: (a). It suffices to show, by Portmanteau's thm, that

$$E[f(X_n+Y_n)] \to E[f(X+c)]$$

for every bounded, uniformly continuous function f.

- By uniform continuity,

given $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ s.t. $|f(x) - f(y)| < \varepsilon$ if $|x-y| < \delta$.

- $|E[f(X_n+Y_n)] - E[f(X+c)]|$

$$\le |E[f(X_n+Y_n)] - E[f(X_n+c)]| + |E[f(X_n+c)] - E[f(X+c)]|$$

$$\le E|f(X_n+Y_n) - f(X_n+c)| 1_{[|Y_n-c|>\delta]} + E|f(X_n+Y_n) - f(X_n+c)| 1_{[|Y_n-c|\le\delta]}$$

$$+ |E[f(X_n+c)] - E[f(X+c)]|$$

$$\le 2\|f\| \cdot P[|Y_n - c| > \delta] + \varepsilon + |E[f(X_n+c)] - E[f(X+c)]|, \quad \|f\| = \sup_x |f(x)| < \infty.$$

$\to 0 + \varepsilon + 0 = \varepsilon$ since

① $Y_n \xrightarrow{P} c \Rightarrow P[|Y_n - c| > \delta] \to 0$ as $n \to \infty$.

② It remains to show $X_n + c \Rightarrow X + c$,

For f bounded, uniformly continuous, so is $h(x) = f(x+c)$.

Thus $E[f(X_n+c)] = E[h(X_n)] \to E[h(X)] = E[f(X+c)]$ as $n \to \infty$.

Hence $|E[f(X_n+c)] - E[f(X+c)]| \to 0$ as $n \to \infty$.

Since ε is arbitrary, the result follows by Portmanteau's thm, again.

(next pg. cont.)

(b). It suffices to show, by Portmanteau's thm, that

$$E[f(X_nY_n)] \to E[f(cX)]$$

for every bounded, uniformly continuous function f.

- First, choose K s.t. $\pm K$ are continuity point of F_X and $P[|X|>k]<\varepsilon$, which means that

$$P[|X_n|>2k] \le P[|X_n-X|>k]+P[|X|>k] < 2\varepsilon \text{ for large } n.$$

- By uniform continuity, given $\varepsilon>0$, there is $\delta=\delta(\varepsilon)>0$ s.t. $|f(x)-f(y)|<\varepsilon$ if $|x-y|<\delta$.

- $$|E[f(X_nY_n)]-E[f(cX)]|$$

$$\le |E[f(X_nY_n)]-E[f(X_nc)]|+|E[f(X_nc)]-E[f(Xc)]|$$

$$\le E|f(X_nY_n)-f(X_nc)|1_{[|Y_n-c|>\delta/2k]}+E|f(X_nY_n)-f(X_nc)|1_{[|Y_n-c|\le\delta/2k]}$$
$$+|E[f(X_nc)]-E[f(Xc)]|$$

$$\le 2\|f\|\cdot P[|Y_n-c|>\delta/k] \qquad \because \|f\|=\sup_x|f(x)|<\infty.$$
$$+E|f(X_nY_n)-f(X_nc)|1_{[|Y_n-c|\le\delta/2k,|X_n|>2k]}$$
$$+E|f(X_nY_n)-f(X_nc)|1_{[|Y_n-c|\le\delta/2k,|X_n|\le 2k]}$$
$$+|E[f(X_nc)]-E[f(Xc)]|$$

$$\le 2\|f\|P[|Y_n-c|>\delta]+2\|f\|P[|X_n|>2k]+\varepsilon \qquad \because |Y_n-c|\le\delta/2k \text{ and } |X_n|\le 2k$$
$$+|E[f(X_nc)]-E[f(Xc)]| \qquad \Rightarrow |X_nY_n-X_nc|\le\delta \Rightarrow |f(X_nY_n)-f(X_nc)|<\varepsilon$$

$$\le 0+4\|f\|\varepsilon+\varepsilon+|E[f(X_nc)]-E[f(Xc)]|$$

- It remains to show that $cX_n \Rightarrow cX$, but $h(x)=f(cx)$ is bounded and uniformly continuous, s.t.

$$E[f(cX_n)]=E[h(X_n)]\to E[h(X)]=E[f(cX)]$$

Hence $|E[f(cX_n)]-E[f(cX)]|\to 0$ as $n\to\infty$

Since ε is arbitrary, the result follows by Portmanteau's thm.

#

- **Theorem** (Slutsky's Thm: Another Proof)

(a) If $X_n \Rightarrow X$ and $\delta_n \to_p 0$ (iff $\delta_n \Rightarrow 0$), then $\delta_n X_n \Rightarrow 0$

(b) If $X_n \Rightarrow X$ and $X_n - Y_n \to_p 0$ (iff $X_n - Y_n \Rightarrow 0$), then $Y_n \Rightarrow X$.

Pf: (a) • Given ϵ and η, choose x s.t. $P[|X| \geq x] < \eta$ and $P[X = \pm x] = 0$,

• and then choose n_0 s.t. $n \geq n_0$ implies that

$$P[|\delta_n| > \epsilon/x] < \eta \text{ and } |P[X_n \leq y] - P[X \leq y]| < \eta \text{ for } y = \pm x.$$

Then

• $P[|\delta_n X_n| \geq \epsilon] = P[|\delta_n X_n| \geq \epsilon, |\delta_n| \leq \epsilon/x] + P[|\delta_n X_n| \geq \epsilon, |\delta_n| > \epsilon/x]$

$\leq P[|X_n| \geq x] + P[|\delta_n| > \epsilon/x]$ $\because |\delta_n X_n| \geq \epsilon \ \& \ |\delta_n|^{-1} \geq \frac{x}{\epsilon} \Rightarrow |X_n| \geq x.$

$\leq |P[|X_n| \geq x] - P[|X| \geq x]| + P[|X| \geq x] + P[|\delta_n| > \epsilon/x]$

$\because |P[|X_n| \geq x] - P[|X| \geq x]|$

$= |(P[X_n > x] - P[X > x]) + (P[X_n \leq -x] - P[X \leq -x])|$

$\leq |(1 - P[X_n \leq x]) - (1 - P[X \leq x])| + |P[X_n \leq -x] - P[X \leq -x]| < 2\eta$

$\leq 2\eta + \eta + \eta = 4\eta.$

(b) • Suppose that $y' < x < y''$ and $P[X = y'] = P[X = y''] = 0$.

• If $y' < x - \epsilon < x < x + \epsilon < y''$, then

$P[Y_n \leq x] = P[Y_n \leq x, |X_n - Y_n| \leq \epsilon] + P[Y_n \leq x, |X_n - Y_n| > \epsilon]$

$\because Y_n \leq x \ \& \ -\epsilon \leq X_n - Y_n \leq \epsilon \Rightarrow x - \epsilon \leq X_n \leq x + \epsilon$

$\leq P[X_n \leq x + \epsilon] + P[|X_n - Y_n| > \epsilon] < P[X_n \leq y''] + P[|X_n - Y_n| > \epsilon]$

• Interchang X_n and Y_n, and replace x by $x - \epsilon$,

$P[X_n \leq x - \epsilon] \leq P[Y_n \leq x] + P[|X_n - Y_n| > \epsilon]$,

$P[Y_n \leq x] \geq P[X_n \leq x - \epsilon] - P[|X_n - Y_n| > \epsilon] \geq P[X_n \leq y'] - P[|X_n - Y_n| > \epsilon]$.

• Since $X_n \Rightarrow X$, letting $n \to \infty$ gives

$P[X < y'] \leq \liminf_n P[Y_n \leq x] \leq \limsup_n P[Y_n \leq x] \leq P[X < y'']$

Since $P[X = y] = 0$ for all but countably many y, if $P[X = x] = 0$, then y' and y'' can further be chosen s.t. $P[X \leq y']$ and $P[X \leq y'']$ are arbitrary near $P[X \leq x]$; hence $P[Y_n \leq x] \to P[X \leq x]$.

#

- <u>Example</u> (Application of the Mapping Thm, Slutsky's Thm).

If $X_n \Rightarrow X$, $a_n \to a$, and $b_n \to b$, then $a_n X_n + b_n \Rightarrow aX+b$.

Pf: • From $X_n \Rightarrow X$, it follows by the mapping thm that
$aX_n + b \Rightarrow aX+b$ by $f(x) = ax+b$ is a continuous ft.

- Suppose also $a_n \to a$, $b_n \to b$.
- Then $(a_n - a)X_n \Rightarrow 0$ by Slutsky's thm, hence
$(a_n X_n + b_n) - (aX_n + b)$
$= (a_n - a)X_n + (b_n - b) \Rightarrow 0$ by Slutsky's thm.
- Slutsky thm finally implies that
$(a_n X_n + b_n) \Rightarrow aX+b$.

#

- <u>Theorem</u> (The Diagonal Method)

Suppose that each row of the array

$$(1) \quad \begin{matrix} x_{1,1} & x_{1,2} & x_{1,3} \cdots \\ x_{2,1} & x_{2,2} & x_{2,3} \cdots \\ \vdots & \vdots & \vdots \end{matrix}$$

is a bounded sequence of real numbers.

Then there exists an increasing sequence $\{n_k\}$ of integers s.t the limit $\lim_k x_{r,n_k}$ exists for $r=1,2,\dots$, i.e. for all r

pf: From the first row, from boundedness we could select a convergent subsequence

(2) $x_{1,n_{1,1}}, x_{1,n_{1,2}}, x_{1,n_{1,3}}, \cdots$

here $\{n_{1,k}\}$ is an increasing sequence of integers and $\lim_k x_{1,n_{1,k}}$ exists.

Look next at the second row of (1) along the sequence $n_{1,1}, n_{1,2}, \cdots$:

(3) $x_{2,n_{1,1}}, x_{2,n_{1,2}}, x_{2,n_{1,3}}, \cdots$

As the subsequence of the second row of (1), (3) is bounded.

Select from it a convergent subsequence

$$x_{2,n_{2,1}}, x_{2,n_{2,2}}, x_{2,n_{2,3}}, \cdots$$

here $\{n_{2,k}\}$ is an increasing sequence of integers, a subsequence of $\{n_{1,k}\}$, and $\lim_k x_{2,n_{2,k}}$ exists.

Continue inductively in the same way. This gives an array

$$(4) \quad \begin{matrix} n_{1,1} & n_{1,2} & n_{1,3} \cdots \\ n_{2,1} & n_{2,2} & n_{2,3} \cdots \\ n_{3,1} & n_{3,2} & n_{3,3} \cdots \\ \vdots & \vdots & \vdots \end{matrix}$$

with the properties

(i) Each row of (4) is an increasing sequence of integers.

(ii) The rth row is a subsequence of the $(r-1)$st.

(iii) For each r, $\lim_k x_{r,n_{r,k}}$ exists. Thus

(5) $x_{r,n_{r,1}}, x_{r,n_{r,2}}, x_{r,n_{r,3}}, \cdots$

is a convergent subsequence of the rth row of (1).

(next pg. cont.)

Put $n_k = n_{k,k}$, then $\{n_k\}$ is the diagonal of the array (4).
Since each row is increasing and is contained in the preceding row, $\{n_k\}$ is an increasing sequence of integers.
Futhermore, $n_r, n_{r+1}, n_{r+2}, \cdots$ is a subsequence of the rth row of (4).
Thus $x_{r,n_r}, x_{r,n_{r+1}}, x_{r,n_{r+2}}, \cdots$ is a subsequence of (5) and is therefore convergent. Thus $\lim_k x_{r,n_k,k}$ does exist for each rational r.

#

- <u>Theorem</u> (Helly's Selection Thm)

For every sequence $\{F_n\}$ of distribution functions, there exists a subsequence $\{F_{n_k}\}$ and a "nondecreasing", "right-continuous" function F, where F "need not be a distribution function", s.t.

$\lim_k F_{n_k}(x) = F(x)$ at continuity points x of F. i.e. so $F_{n_k} \not\Rightarrow F$.

Pf: • An application of the diagonal method gives a sequence $\{n_k\}$ $\because \{F_n\}$ bounded. of integers along which the limit $G(r) = \lim_k F_{n_k}(r)$ exists for every "rational r."

• Define $F(x) = \inf[G(r) : x < r]$, clearly F is nondecreasing.

• To each x and ϵ there is an r for which $x < r$ and $G(r) < F(x) + \epsilon$.

If $x \le y < r$, then $F(y) \le G(r) < F(x) + \epsilon$.

Hence F is continuous from the right.

• If F is continuous at x, choose $y < x$ s.t. $F(x) - \epsilon < F(y)$; $\because F(y) < F(x)$

• now choose rational r and s s.t. $y < r < x < s$ and $G(s) < F(x) + \epsilon$.

From $F(x) - \epsilon < F(y) \le G(r) \le G(s) < F(x) + \epsilon$ and $F_n(r) \le F_n(x) \le F_n(s)$

it follows that $F_{n_k}(r) \le F_{n_k}(x) \le F_{n_k}(s)$ and then

$$G(r) = \lim_k F_{n_k}(r) \le \liminf_k F_{n_k}(x) \le \limsup_k F_{n_k}(x) \le \lim_k F_{n_k}(s) = G(s),$$

hence $F(x) - \epsilon < \liminf_k F_{n_k}(x) \le \limsup_k F_{n_k}(x) < F(x) - \epsilon$.

Since ϵ is arbitrary, we get $\lim_k F_{n_k}(x) = F(x)$ at continuity points x of F.

#

- <u>Definition</u> (Tightness)

A sequence of probability measures μ_n on $(R^1, \mathcal{R}^1)$ is "tight" if for each ϵ there exists a "finite interval" $(a, b]$ s.t. $\mu_n(a, b] > 1 - \epsilon$ "for all n"

In terms of the corresponding distribution functions,

$\mu_n(a, b] = F_n(b) - F_n(a)$, so for each ϵ there exist a and b s.t.

$$\Rightarrow \begin{cases} 1 - F_n(a) \ge F_n(b) - F_n(a) > 1 - \epsilon \Rightarrow F_n(a) = P[X_n < a] < \epsilon \\ F_n(b) \ge F_n(b) - F_n(a) > 1 - \epsilon \Rightarrow 1 - F_n(b) = P[X_n > b] < \epsilon \end{cases} \text{ for all } n$$

Remark:

- If μ_n is a unit mass at n, $\{\mu_n\}$ is not tight, i.e. the mass of μ_n escapes to infinity.
- Tightness is the condition which ensures that for some subsequence $\{F_{n_k}\}$, $F_{n_k} \Rightarrow F$, where F is a distribution function.

- **Theorem** (Prohorov's Thm)

Tightness is a necessary and sufficient condition that for every subsequence $\{\mu_{n_k}\}$ there exists a further subsequence $\{\mu_{n_{k(j)}}\}$ and a probability measure μ s.t. $\mu_{n_{k(j)}} \Rightarrow \mu$ as $j \to \infty$.

Pf: "Sufficiency". Apply Helly's selection thm to the subsequence $\{F_{n_k}\}$ of the corresponding distribution functions. There exists a further subsequence $\{F_{n_{k(j)}}\}$ s.t. $\lim_j F_{n_{k(j)}}(x) = F(x)$ at continuity points x of F, where F is nondecreasing and right-continuous.
There exists a measure μ on $(R^1, \mathcal{R}^1)$ s.t. $\mu(a,b] = F(b) - F(a)$.

- Given ϵ, choose a and b s.t. $\mu_n(a,b] > 1-\epsilon$ for all n, by tightness. By decreasing a and increasing b, one can ensure that they are continuity points of F. Then $\mu_{n_{k(j)}}(a,b] = F_{n_{k(j)}}(b) - F_{n_{k(j)}}(a)$. Thus.

$$1-\epsilon \le \lim_j \mu_{n_{k(j)}}(a,b] = \lim_j F_{n_{k(j)}}(b) - \lim_j F_{n_{k(j)}}(a) = F(b) - F(a) = \mu(a,b].$$

Therefore, μ is a probability measure, i.e. $\mu(R)=1$, and hence $\mu_{n_{k(j)}} \Rightarrow \mu$.

- "Necessity". If μ_n is not tight, there exists a positive ϵ s.t. for each finite interval $(a,b]$, $\mu_n(a,b] \le 1-\epsilon$ for some n.
- Choose n_k s.t. $\mu_{n_k}(-k,k] \le 1-\epsilon$. Suppose that some subsequence $\{\mu_{n_{k(j)}}\}$ of $\{\mu_{n_k}\}$ were to converge weakly to some probability measure μ, i.e. $\mu_{n_{k(j)}} \Rightarrow \mu$. Choose $(a,b]$ s.t. $\mu\{a\} = \mu\{b\} = 0$ and $\mu(a,b] > 1-\epsilon$.
- For large enough j, $(a,b] \subset (-k(j), k(j)]$, and so

$$1-\epsilon \ge \mu_{n_{k(j)}}(-k(j), k(j)] \ge \mu_{n_{k(j)}}(a,b] \to \mu(a,b].$$

Thus $\mu(a,b] \le 1-\epsilon$, a contradiction.

#

- Corollary (Cor. of Prohorov's Thm)

① If $\{\mu_n\}$ is tight sequence of probability measures, and
② if each subsequence $\{\mu_{n_k}\}$ that converges weakly at all converges weakly to the same probability measure μ, then $\mu_n \Rightarrow \mu$.

Remark:
"Tightness of sequence of probability measures" is analogous to to "boundedness of sequences of real numbers":
If $\{x_n\}$ is bounded, and if each subsequence that converges at all converges to x, then $\lim_n x_n = x$.

Pf: By Prohorov's thm and the hypothesis, "every" subsequence $\{\mu_{n_k}\}$ contains a further subsequence $\{\mu_{n_{k(j)}}\}$ s.t. $\mu_{n_{k(j)}} \Rightarrow \mu$.
- Suppose that $\mu_n \Rightarrow \mu$ is false. Then there exists x s.t. $\mu\{x\}=0$ but $\mu_n(-\infty, x] \not\to \mu(-\infty, x]$. But then there exists a positive ϵ s.t. $|\mu_n(-\infty, x] - \mu(-\infty, x]| \geq \epsilon$ for an infinite sequence $\{n_k\}$, and no subsequence $\{\mu_{n_{k(j)}}\}$ of $\{\mu_{n_k}\}$ can converge weakly to μ, a contradiction. Thus $\mu_n \Rightarrow \mu$.

#

- Summary

(1) Helly's selection thm (always holds!):
for every subsequence $\{F_{n_k}\}$ of $\{F_n\}$, $\exists$ a further subsequence $\{F_{n_{k(j)}}\}$ & a F s.t.
$\lim_j F_{n_{k(j)}}(x) = F(x)$ at continuity points x of F
where F need not to be a distribution function.

(2) Helly's selection thm + Tightness $\Rightarrow$ Prohorov's thm:
$F_{n_{k(j)}} \Rightarrow F$, now F is a distribution function.

(3) Prohorov's thm + each subsequence $\{\mu_{n_k}\}$ s.t. $\mu_{n_k} \Rightarrow \mu$ (same μ):
$F_n \Rightarrow F$, so $\mu_n \Rightarrow \mu$.

- <u>Example</u> (Tightness Condition for Normal r.v.'s)

Let μ_n be the normal distribution with mean m_n and variance σ_n^2.

- If m_n and σ_n^2 are bounded, then $\{\mu_n\}$ is tight:

1. If $m_n > b$, then $\mu_n(b,\infty) \geq \frac{1}{2}$; if $m_n < a$, then $\mu_n(-\infty,a] \geq \frac{1}{2}$.

   Hence $\{\mu_n\}$ cannot be tight if m_n is unbounded.

2. If m_n is bounded, say by k, i.e. $|m_n| \leq k$, then

   $\mu_n(-\infty, a] \geq \nu(-\infty, (a-k)\sigma_n^{-1}]$, where ν is the standard normal distribution.

   $\cdot$ If σ_n is unbounded, then $\nu(-\infty, (a-k)\sigma_n^{-1}] \to \Phi(0) = \frac{1}{2}$ along some subsequence.

   and $\{\mu_n\}$ cannot be tight.

#