I. Events, probability, random variables

Let Ω be a set of points ω.

Def 1) A nonempty system of subsets of Ω is called an algebra $\mathcal{A}$ if $\Omega \in \mathcal{A}$, $A \cup B$, $A \cap B$, $A^c = \Omega \setminus A$ are elements of $\mathcal{A}$ whenever $A, B \in \mathcal{A}$.

A function $\mu : \mathcal{A} \to [0, \infty]$ is called finitely additive measure if for any disjoint $A, B \in \mathcal{A}$

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(Note that then $\forall A, B \in \mathcal{A}$:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \,)$$

2) An algebra $\mathcal{A}$ is called σ-algebra $\mathcal{F}$ if countable union $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ whenever $A_1, A_2 \ldots \in \mathcal{F}$.

(Note that then also $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$: consider $\Omega \setminus A_k = \hat{A}_k$).

A function $\mu : \mathcal{F} \to [0, \infty]$ is called σ-additive if for any disjoint $A_1, A_2, \ldots \in \mathcal{F}$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Such μ is called a <u>measure</u> on $\mathcal{F}$. A measure μ is called a <u>probability measure</u> if $\mu(\Omega) = 1$.

(Note that $\mu(\emptyset) = 0$ since
$\mu(\emptyset) = \mu(\emptyset \cup \emptyset) = 2\mu(\emptyset)$)
A measure is called σ-finite if there exists a representation $\Omega = \bigcup_1^\infty \Omega_k$, Ω_k - pairwise disjoint, $\mu(\Omega_k) < \infty$, $k = 1, 2 \ldots$

<u>Def</u> A <u>probability space</u> is a triple $(\Omega, \mathcal{F}, P)$, where Ω is a set called sample space, $\mathcal{F}$ is a σ-algebra of subsets of Ω, P - probability measure on $\mathcal{F}$. Any element of $\mathcal{F}$ is called event.

We also say:
$P(A \cup B)$ - probability that either A or B occurs;
$P(A \cap B)$ - probability that both A and B occur.
$P(\Omega \setminus A)$ - probability that A does not occur.

<u>Lemma 1</u> (continuity of measure)
a) If $A_n \in \mathcal{F}$, $A_1 \subset A_2 \subset \cdots$ then
$P(\bigcup_{n=1}^\infty A_n) = \lim_{n \to \infty} P(A_n)$ — continuity from below
b) If $B_n \in \mathcal{F}$, $B_1 \supset B_2 \supset \cdots$ then
$P(\bigcap_{n=1}^\infty B_n) = \lim_{n \to \infty} P(B_n)$ - continuity from above.

Since $\bigcup_n A_n = (\bigcap_n A_n^c)^c$, a) and b) are equivalent.

So sufficient to show b). To show b), it suffices to show it for the case $\bigcap_{n=1}^{\infty} B_n = \emptyset$, as otherwise we replace B_n by $B_n \setminus \bigcap_{n=1}^{\infty} B_n$.

<u>Remark</u> b) with $\bigcap_{n=1}^{\infty} B_n = \emptyset$ is called continuity at zero.

To prove continuity at zero, i.e. the fact that $\lim_{n\to\infty} P(B_n) = 0$, consider $\Omega \setminus B_1$, $B_1 \setminus B_2$ ....

They are disjoint.

$$\emptyset = \bigcap_1^{\infty} B_n = \Omega \setminus \bigcup_1^{\infty} (\Omega \setminus B_n) =$$
$$= \Omega \setminus ((\Omega \setminus B_1) \cup (B_1 \setminus B_2) \cup \cdots).$$

Therefore

$$1 = P(\Omega \setminus B_1) + P(B_1 \setminus B_2) + \cdots =$$
$$= \sum_{j=0}^{\infty} P(B_j \setminus B_{j+1}), \quad B_0 = \Omega.$$

So $\forall \varepsilon > 0 \ \exists n_0$ s.t. $\forall n > n_0$

$$1 - \sum_{j=0}^{n-1} P(B_j \setminus B_{j+1}) = P(B_n) < \varepsilon$$

Any finitely additive measure μ on $\mathcal{F}$ satisfies

$$\sum_{n=1}^{\infty} \mu(A_n) \leqslant \mu\left(\bigcup_{n=1}^{\infty} A_n\right) \text{ if } A_j\text{'s are disjoint, } A_1, A_2, \ldots \in \mathcal{F}.$$

When is it a measure ("=")?

<u>Lemma 2</u> A finitely additive probability measure on $\mathcal{F}$ is a probability measure iff it is continuous at zero.

$\triangleleft$ ($\Rightarrow$) If P is a probability measure then continuity at zero already shown.

($\Leftarrow$) Let P be a finitely additive probability measure on $\mathcal{F}$ and for any $B_1 \supset B_2 \supset \cdots$, $\bigcap_{1}^{\infty} B_n = \emptyset$, $B_1, B_2, \ldots \in \mathcal{F}$ we have $\lim\limits_{n\to\infty} P(B_n) = 0$.

Hence a) of Lemma 1 holds since a), b) are equivalent and equivalent to continuity at zero.

For any disjoint $C_1, C_2, \ldots \in \mathcal{F}$ define $A_n = \bigcup_{k=1}^{n} C_k$. Then $A_1 \subset A_2 \subset \cdots$ and by a)

$$P\left(\bigcup_{k=1}^{\infty} C_k\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) =$$

$$= \lim_{n\to\infty} \sum_{k=1}^{n} P(C_k) = \sum_{k=1}^{\infty} P(C_k) \quad \triangleright$$

Examples of σ-algebras on Ω:

$\mathcal{F}_* = \{\emptyset, \Omega\}$; $\mathcal{F}^* = \{A : A \subset \Omega\} = 2^{\Omega}$;

σ-algebra generated by partitions:

$\sigma(D) = \{\bigcup_{j \in I} D_j : I \subset \mathbb{N}\}$ where

$D = \{D_1, D_2, \dots\}$ is a partition of Ω into countable union of disjoint D_j, $\Omega = \bigcup_1^{\infty} D_j$.

<u>Lemma</u> For any collection $\mathcal{E}$ of subsets of Ω, there exists minimal algebra $a(\mathcal{E})$ and minimal σ-algebra $\sigma(\mathcal{E})$ that contains all elements of $\mathcal{E}$ (intersection of all algebras (resp., σ-algebras) containing $\mathcal{E}$).

$\triangle$ intersection, countable or uncountable, of algebras (resp., σ-algebras) containing $\mathcal{E}$ is an algebra (resp. σ-algebra) containing $\mathcal{E}$ $\blacktriangleright$.

We say that $\sigma(\mathcal{E})$ is generated by $\mathcal{E}$.

<u>Def</u> A measurable space is a pair $(E, \mathcal{E})$ where E is a set, and $\mathcal{E}$ is a σ-algebra on E.

Examples:

1) $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where

$\mathcal{B}(\mathbb{R}^n) = \sigma(\{A \subset \mathbb{R}^n : A\text{-open}\})$ - Borel σ-algebra.

For $n=1$ recall that

$$\mathcal{B}(\mathbb{R}) = \sigma(\{\text{open subsets of } \mathbb{R}\}) =$$
$$= \sigma(\{\text{open intervals}\}) = \sigma(\{\text{closed intervals})$$
$$= \sigma(\{\text{closed half-lines } (-\infty, x],\ x \in \mathbb{R}\})$$

2) Product σ-algebra.

Consider $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ where $(\Omega_j, \mathcal{F}_j)$ - ~~measure~~ - measurable spaces, $j = 1, 2$.

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\})$$

<u>Lemma</u>: $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

$\triangleleft$ 1) $\mathcal{B}(\mathbb{R}^2) \subset \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, since any open $A \subset \mathbb{R}^2$ can be written as follows:

$$A = \bigcup_{x \in A \cap \mathbb{Q}^2} R(x, r(x)) \in \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}),$$

$R(x, r)$ - open square centered at x of ~~side~~ side length $r(x)$.

2) $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\mathbb{R}^2)$?

Sufficient to check that $B_1 \times B_2 \in \mathcal{B}(\mathbb{R}^2)$ for any 2 Borel sets B_1, B_2.

Note that $B_1 \times \mathbb{R} \in \mathcal{B}(\mathbb{R}^2)$ since

$B_1 \times \mathbb{R} \in \sigma(\{\text{open subsets of } \mathbb{R}\}) \times \mathbb{R} =$
$= \sigma(\{\text{open subsets of } \mathbb{R}\} \times \mathbb{R})$,

Similarly, $\mathbb{R} \times B_2 \in \mathcal{B}(\mathbb{R}^2)$, and so
$B_1 \times B_2 = (B_1 \times \mathbb{R}) \cap (\mathbb{R} \times B_2) \in \mathcal{B}(\mathbb{R}^2)$ $\boxed{\mathcal{B}}$

3) Cylindrical σ-algebra

Let $\mathbb{R}^\infty = \{ x = (x_1, x_2, \dots), \ x_k \in \mathbb{R} \}$

<u>Def</u> A set $C \subset \mathbb{R}^\infty$ is called cylindrical if it is of the form $C = \{ x \in \mathbb{R}^\infty : (x_1, x_2 \dots x_n) \in \tilde{C}_n \}$ for some $n \geq 1$ and $\tilde{C}_n \in \mathcal{B}(\mathbb{R}^n)$.

Cylindrical sets form an algebra (check!) which generates a σ-algebra called cylindrical σ-algebra, denoted $\mathcal{B}(\mathbb{R}^\infty)$.

One can verify that

$$\mathcal{B}(\mathbb{R}^\infty) = \sigma(\{ A_1 \times A_2 \times \dots \subset \mathbb{R}^\infty, \ A_k \in \mathcal{B}(\mathbb{R}) \})$$

Example :

$\forall c \in \mathbb{R}$, let

$A = \{ x \in \mathbb{R}^\infty : \limsup_n x_n = \inf_n \sup_{k>n} x_k > c \}$

We have $A \in \mathcal{B}(\mathbb{R}^\infty)$: indeed

$$A = \bigcap_{n=1}^{\infty} \bigcup_{k=n+1}^{\infty} \{ x \in \mathbb{R}^\infty : x_k > c \} = (x_k > c \text{ i.o.}) .$$

$\forall c$, let

$B = \{ x \in \mathbb{R}^\infty : \liminf_n x_n = \sup_n \inf_{k>n} x_k > c \}$

We have $B \in \mathcal{B}(\mathbb{R}^\infty)$: indeed

$$B = \bigcup_{n=1}^{\infty} \bigcap_{k=n+1}^{\infty} \{ x \in \mathbb{R}^\infty : x_k > c \} = (x_k > c \text{ ev.}) .$$

$\forall c$, $D = \{ x \in \mathbb{R}^\infty : \lim_{n\to\infty} x_n = c \} \in \mathcal{B}(\mathbb{R}^\infty)$

exercise .

Recall : nondecreasing function $g(x)$ on $\mathbb{R}$ is continuous up to possibly countably many discontinuities of first kind : $f(x+0)$, $f(x-0)$ Both exist but $f(x+0) - f(x-0) = h_x > 0$.
Moreover, the derivative $g'(x)$ exist Lebesgue a.e.

Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ - probability space,
$F(x) = P(-\infty, x]$, $x \in \mathbb{R}$. Then (exercise)

(1) $F(x)$ is nondecreasing
(2) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$
(3) $F(x)$ is continuous on the right $\forall x \in \mathbb{R}$.

Def Any function $F: \mathbb{R} \to [0,1]$ satisfying (1), (2), (3) is called a distribution function (on $\mathbb{R}$).

We have seen that P gives rise to F. In fact, the opposite is true and there exists 1-1 correspondence between distribution functions and probability measures:

Thm Let $F(x)$ be a distribution function on $\mathbb{R}$. Then there exists a unique probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t. $P(-\infty, x] = F(x)$.

Caratheodory thm: Let μ_0 be a σ-additive measure on $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is an algebra of subsets of Ω. Then there exists a unique measure μ on $(\Omega, \sigma(\mathcal{A}))$ s.t.
$\mu(A) = \mu_0(A)$ $\forall A \in \mathcal{A}$.

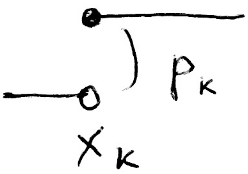
Remark: A measure μ on a σ-algebra Σ on Ω is called complete if any subset of a set of measure zero is measurable, i.e. belongs to Σ.
If a measure μ on Σ is not complete, it can be completed by extending Σ to
$\overline{\Sigma} = \sigma(\Sigma \cup \{B \in \Omega : B \subset A \in \Sigma, \mu(A) = 0\})$.
Clearly, we must set $\mu(B) = 0$ $\forall B \subset A$, $\mu(A) = 0$.
If P is complete, we say that $(\Omega, \mathcal{F}, P)$ is a complete probability space.
Note that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ considered above is not complete ($\exists$ subsets of a Borel set which are not Borel). Its completion extends $\mathcal{B}(\mathbb{R})$ to $\overline{\mathcal{B}(\mathbb{R})} = \mathcal{M}(\mathbb{R})$ - σ-algebra of Lebesgue measurable sets, and P becomes Lebesgue-Stiltjes measure. In particular, the distribution function $F(x) = x$ corresponds to the Lebesgue measure on $\mathbb{R}$.

3 types of distribution functions

(probability measures — Lebesgue-Stieltjes probability measures)

1) Discrete:

$$F(x) = F_{disc}(x) = \sum_{x_k \le x} p_k \quad , \ldots, p_1, p_2 \ldots \ge 0$$



Note that $F'(x) = 0$ a.e.

Let $\Omega = \mathbb{N}$ (instead of $\mathbb{R}$)

$$P(A) \equiv P_{disc}(A) = \sum_{k \in A} p_k$$

Examples:

discrete uniform distribution: $p_k = \frac{1}{N}$, $k = 1, 2, \ldots, N$, N - fixed

Bernoulli: $p_1 = p$, $p_2 = 1 - p$, $0 \le p \le 1$.

Binomial: $p_k = C_n^k p^k q^{n-k}$, $q = 1 - p$, $0 \le p \le 1$, $k = 0, 1, \ldots, n$.

Poisson: $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ $\lambda > 0$, $k = 0, 1, \ldots$

2) Absolutely continuous

there exists an integrable $f(x) \ge 0$ called the density of the distribution F s.t.

$$F(x) = F_{ac}(x) = \int_{-\infty}^{x} f(t)\,dt \quad \text{— w.r.t. Lebesgue measure.}$$

Then $P_{ac}(A) = \int_A f(t)\,dt$, $A \in \mathcal{F}$

◁ First, $P_{ac}(a,b] = \int_a^b f(t)\,dt$,
then use Caratheodory thm to extend
~~th~~ to σ-algebra ▷

In particular, if $\mu(A) = 0$ - Lebesgue measure
then $P_{ac}(A) = 0$.

Note that $F'_{ac}(x) = f(x)$ a.e.; $F_{ac}(x)$ is absolutely continuous function, in particular continuous at any x.

Examples:

Uniform distribution on $[a,b]$: $f(x) = \dfrac{1}{b-a}$, $a \le x \le b$,
$f(x) = 0$ otherwise.

Normal (or Gaussian): $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$,
$x \in \mathbb{R}$, $\sigma > 0$.

Gamma: $f(x) = \dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$, $x \ge 0$, $\alpha, \beta > 0$

Exponential $(\alpha = 1, \beta = 1/\lambda)$: $f(x) = \lambda e^{-\lambda x}$, $x \ge 0$, $\lambda > 0$.

Chi-squared χ^2: $(\alpha = n/2, \beta = 2)$ $f(x) = 2^{-n/2} x^{n/2 - 1} e^{-x/2} \dfrac{1}{\Gamma(n/2)}$ $x \ge 0$, $n \in \mathbb{N}$

3) Singular continuous :

Recall : a measure ν is said to be concentrated on a measurable set A if $\nu(E)=0$ for any $E \subset \mathbb{R} \setminus A$.

$F(x) = F_{sc}(x)$ is continuous at any x and P_{sc} is concentrated on a set of Lebesgue measure zero. It is the set where $F_{sc}'(x) \neq 0$ or does not exist. Thus $F_{sc}'(x) = 0$ a.e.

Note that, by continuity, $P_{sc}\{x\} = 0$ $\forall$ point $x \in \mathbb{R}$.

Example : Cantor staircase.

Thm (Hahn decomposition) Any probability distribution function has a representation :

$F(x) = a_1 F_{disc}(x) + a_2 F_{ac}(x) + a_3 F_{sc}(x)$,

$a_1 + a_2 + a_3 = 1$.

- Distribution functions on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ are defined similarly, e.g.

  $F(x,y) = P((-\infty, x] \times (-\infty, y])$, $n=2$.

- Product measure on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ is defined as follows :

First, set $P_0(A_1 \times A_2) = P_1(A_1)\, P_2(A_2)$ for $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$.

(probability spaces $(\Omega_1, \mathcal{F}_1, P_1)$, $(\Omega_2, \mathcal{F}_2, P_2)$)

Then extend P_0 to the algebra generated by $A_1 \times A_2$, show that P_0 is a σ-additive measure on this algebra, and apply Caratheodory thm to obtain the extension. This extension is called the product measure, denoted $P_1 \otimes P_2$.

- Probability measures on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$.

The sequence P_n of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is said to be consistent if

$$P_{n+1}(B_n \times \mathbb{R}) = P_n(B_n) \quad \forall \text{ Borel } B_n \in \mathbb{R}^n \ \forall n.$$

<u>Thm</u> (Kolmogorov) For any consistent sequence P_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, there exists a unique probability measure P on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ s.t.

$$P(\{x \in \mathbb{R}^\infty : (x_1, x_2, \dots x_n) \in B_n\}) = P_n(B_n)$$

$\forall n \geq 1$, Borel $B_n \subset \mathbb{R}^n$.

Why measurable sets? $\rightarrow$

<u>Thm</u> (Banach-Tarski paradox) A unit ball in $\mathbb{R}^3$ can be partitioned into 5 parts that can be reassembled (shifts + rotations) to form 2 nonintersecting unit balls.

Random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

<u>Def</u> A function $\xi : \Omega \to \mathbb{R}$ is called a random variable (r.v.) if it is P-measurable, i.e. for any Borel B, $\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \mathcal{F}$

The distribution P_ξ of ξ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by $P_\xi(B) \equiv P(\xi^{-1}(B))$ $\forall B \in \mathcal{B}$. The distribution function

$F_\xi(x) \equiv P_\xi(-\infty, x] = P(\omega \in \Omega : \xi(\omega) \leq x)$ $\forall x \in \mathbb{R}$.

<u>Remark</u> P on $\Omega \to P_\xi$ on $\mathbb{R}$.

Introduce also $\mathcal{F}_\xi \equiv \{\xi^{-1}(B) : B \in \mathcal{B}(\mathcal{B})\}$ - it is a σ-algebra (check!), $\mathcal{F}_\xi \subset \mathcal{F}$.

<u>Lemma</u> Let $\mathcal{D}$ be a collection of subsets on $\mathbb{R}$ s.t. $\sigma(\mathcal{D}) = \mathcal{B}(\mathbb{R})$ Then ξ is a r.v. if $\xi^{-1}(D) \in \mathcal{F}$ $\forall D \in \mathcal{D}$. (proof - exercise)

Example: $\{\omega : \xi(\omega) \leq x\} \in \mathcal{F}$ $\forall x \in \mathbb{R}$
$\Rightarrow \xi$ is r.v.

<u>Lemma</u>: If f, g are r.v. Then $f+g$, $f-g$, $c\cdot f$, $c \in \mathbb{R}$, $|f|$, $f \cdot g$, f/g if $g \neq 0$, $\max(f,g)(x)$, $\min(f,g)$ are r.v.

If f_n, $n = 1, 2, \ldots$ are r.v. and if there exist $\forall \omega$ $\sup_n f_n(\omega) = s(\omega)$ (or $\inf_n$ or $\lim_n$) then $s(\omega)$ is r.v.

<u>Lemma</u>: If ξ is r.v. and $f: \mathbb{R} \to \mathbb{R}$ is continuous then $f(\xi)$ is r.v.

◁ $f^{-1}(A)$ is open for any open $A \subset \mathbb{R}$. Therefore $f^{-1}(B)$ is Borel for any Borel B ▷

<u>Def</u> A r.v. ξ is called simple if

$$\xi(\omega) = \sum_{j=1}^{n} x_j \chi_{D_j}(\omega) \text{ for some } n \geq 1$$

and a partition $D_1, \ldots, D_n$ of Ω,

$$\chi_D(\omega) = \begin{cases} 1, & \omega \in D \\ 0, & \text{otherwise} \end{cases}.$$

<u>Lemma</u> For any r.v. $\xi(\omega) \geq 0$ there exists a point-wise nondecreasing sequence of simple r.v. $\xi_1(\omega) \leq \xi_2(\omega) \leq \cdots \leq \xi(\omega)$ s.t.

$$\lim_{n \to \infty} \xi_n(\omega) = \xi(\omega) \quad \forall \omega \in \Omega \quad (\text{in short, } \xi_n \nearrow \xi)$$

◁ Set $\xi_n(\omega) = \sum_{j=0}^{n2^n-1} \frac{j}{2^n} \chi_{\{\omega : \frac{j}{2^n} \le \xi(\omega) < \frac{j+1}{2^n}\}}$

$$+ n \chi_{\{\omega : \xi(\omega) \ge n\}}$$ ▷

Remark If ξ is not positive,
$\xi = \xi^+ - \xi^-$, $\xi^+ = \max(\xi, 0)$, $\xi^- = \max(-\xi, 0)$.

Extensions :

- $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$: $\xi : \Omega \to \mathbb{R}^n$, $\xi = (\xi_1, \dots \xi_n)$ is called a random vector if for any $B \subset \mathcal{B}(\mathbb{R}^n)$, $\xi^{-1}(B) \in \mathcal{F}$. P_ξ is also defined as before and we say that $P_\xi = P_{(\xi_1, \dots \xi_n)}$ - is a joint distribution of $\xi_1, \xi_2 \dots, \xi_n$.
Note that ξ is a random vector iff $\xi_1, \xi_2 \dots \xi_n$ are random variables (check!)

Def r.v. ξ and η are called independent if $P_{(\xi,\eta)}(\xi \in A, \eta \in B) = P_\xi(\xi \in A) P_\eta(\eta \in B)$ for any Borel A, B.

- $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ - we can define random sequences $\xi = (\xi_1, \xi_2 \dots)$

Def Expectation $E(\xi)$ of a r.v. ξ is the Lebesgue integral w.r.t. P

$$E(\xi) = \int_{\Omega} \xi \, dP \quad \text{if it exists}$$

We say that ξ is integrable if $E(|\xi|)$ exists and finite ($E|\xi|$ exists and finite $\Leftrightarrow E(\xi)$ exists and finite)

Recall constuction of Lebesgue integral:
First define the integral of a simple r.v. $\xi = \sum_1^n x_j \chi_{D_j}$ by $E\xi = \sum_1^n x_j P(D_j)$.
For general r.v. $\xi \geq 0$ take a sequence of simple r.v. $\xi_n \nearrow \xi$ and set $E(\xi) = \lim_{n\to\infty} E(\xi_n)$
- finite or infinite exists since $E(\xi_n) \leq E(\xi_{n+1})$
but one has to show it is independent of the choice of $\xi_n \nearrow \xi$. In general use the fact $\xi = \xi^+ - \xi^-$.

<u>Properties</u>: Let ξ, η be integrable r.v. Then
$E(\text{const}) = \text{const}$, $E(\text{const}\cdot\xi) = \text{const}\, E(\xi)$,
$\xi+\eta$ is integrable and $E(\xi+\eta) = E(\xi) + E(\eta)$;
$\xi \leq \eta \Rightarrow E(\xi) \leq E(\eta)$; if $\xi = \eta$ a.e. w.r.t. P
(a.e $\equiv$ almost surely (a.s.) i.e up to sets of P-measure zero)

then $E(\xi) = E(\eta)$;

if $\xi \geq 0$ and $E\xi = 0$ then $\xi = 0$ a.s.

Thm (dominated convergence or Lebesgue thm)

Let ξ_n, $n = 1, 2 \ldots$ be r.v. s.t. $\xi_n \to \xi$, $n \to \infty$ a.s.

If there exists an integrable r.v. η s.t. $|\xi_n| \leq \eta$ $\forall n$ then ξ is integrable and $\lim\limits_{n\to\infty} E\xi_n = E\xi$.

Thm (monotone convergence, Levi thm)

Let $0 \leq \xi_1 \leq \xi_2 \leq \cdots$ be r.v.

Then there exists (finite or infinite)

$$\lim_{n\to\infty} E(\xi_n) = E \lim_{n\to\infty} \xi_n$$

Remarks: 1) $0 \leq \xi_1 \leq \cdots$ can be replaced $\eta \leq \xi_1 \leq \cdots$ with $E\eta > -\infty$ (just consider $\xi_n - \eta$ instead of ξ_n).

2) $0 \leq \xi_1 \leq \cdots$ can be replaced by $\cdots \leq \xi_2 \leq \xi_1 \leq \eta$, $E\eta < +\infty$.

Corollary If $\eta_k \geq 0$, $k = 1, 2, \ldots$

then $E \sum\limits_1^\infty \eta_k = \sum\limits_1^\infty E\eta_k$.

Thm (Fatou lemma) Let $\xi_n \geq 0$, $n=1,2\dots$

Then $E \liminf_n \xi_n \leqslant \liminf_n E\xi_n$.

Remarks 1) $\xi_n \geq 0$ can be replaced by $\xi_n \geq \eta$, $E\eta > -\infty$

2) If $\xi_n \leq \eta$, $E\eta < +\infty$, the statement holds for limsup instead.

Hint: apply monotone convergence thm to $\lambda_n = \inf_{k>n} \xi_k$

In all the above thm's integral over Ω can be replaced by integral over any measurable $\hat{A} \subset \Omega$.

Thm (change of variables)

For a r.v. ξ, measurable $g(x)$, measurable A,

$$\int_A g(x)\, dP_\xi = \int_{\xi^{-1}(A)} g(\xi(\omega))\, dP$$ where both

integrals exist or not simultaneously.

In particular $E\, g(\xi) = \int_\Omega g(\xi(\omega))\, dP = \int_{-\infty}^{\infty} g(x)\, dP_\xi \equiv$

$\equiv \int_{-\infty}^{\infty} g(x)\, dF_\xi$.

The result clearly holds for $g = \chi_B(x)$, $B \in \mathcal{B}(\mathbb{R})$

Therefore, also holds for simple $g(x)$ by linearity of the integral.

For any measurable $g(x) \geq 0$ consider a sequence of simple $g_n \nearrow g$. The result for g then follows from monotone convergence thm. For arbitrary measurable $g(x)$ we use $g(x) = g_+ - g_-$ (3)

Example : $g(x) = x$.

Remarks 1) If ξ is discrete taking values $x_1, x_2 \dots$ with probabilities $p_1, p_2 \dots$ then thm gives $\mathbb{E}\xi = \sum_j x_j p_j$. ($\xi$ discrete means F_ξ is discrete)

2) If ξ is a.c. (i.e. F_ξ is a.c.) with density $f(x)$, then $\mathbb{E}\, g(\xi) = \int_{-\infty}^{\infty} g(x) f(x) dx$.

Thm (Fubini) Let $(E_1, \mathcal{E}_1, \mu_1)$, $(E_2, \mathcal{E}_2, \mu_2)$ - measure spaces, μ_1, μ_2 - σ-finite measures. Then ~~there exist~~ for any $\mathcal{E}_1 \otimes \mathcal{E}_2$ - measurable $g(x,y): E_1 \times E_2 \to \mathbb{R}$; $g(x, y_0)$ is $\mathcal{E}_1$ - measurable $\forall y_0 \in E_2$; $g(x_0, y)$ is $\mathcal{E}_2$ - measurable $\forall x_0 \in E_1$; $\int_{E_1} g \, d\mu_1$ and $\int_{E_2} g \, d\mu_2$ are $\mathcal{E}_2$ and $\mathcal{E}_1$ (resp.) measurable.

If $g(x,y)$ is integrable, then repeated integrals exist and

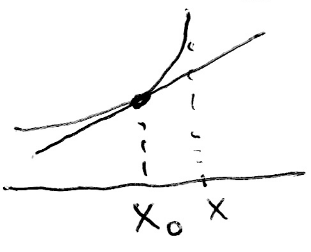
$$\int_{E_1\times E_2} g \, d\mu_1 \otimes d\mu_2 = \int_{E_1} \Big(\int_{E_2} g \, d\mu_2 \Big) d\mu_1 = \qquad (F)$$

$$= \int_{E_2} \Big(\int_{E_1} g \, d\mu_1 \Big) d\mu_2 .$$

If $\int_{E_1} \Big(\int_{E_2} |g| \, d\mu_2 \Big) d\mu_1 < \infty$ then g is integrable and (F) holds.

<u>Lemma</u> (Jensen ineq.) Let ξ be an integrable r.v. and $g(x)$ be measurable, convex downwards. Then $g(E(\xi)) \le E\, g(\xi)$

◁ Convex downwards means that $\forall x_0 \in \mathbb{R} \ \exists \lambda(x_0)$ s.t. $g(x) \ge g(x_0) + (x - x_0)\lambda(x_0)$.



Set $x = \xi$, $x_0 = E\xi$:

$$g(\xi) \ge g(E\xi) - (\xi - E\xi)\lambda(E\xi).$$

taking E of both sides proves the result ▷

<u>Corollary</u> (Lyapunov ineq.)

$$(E|\xi|^s)^{1/s} \le (E|\xi|^t)^{1/t} \text{ if } 0 < s < t$$

◁ Use Jensen ineq. with $g(x) = |x|^{t/s}$ and replace ξ by $|\xi|^s$ ▷

Thus $E|\xi| \leq (E|\xi|^2)^{1/2} \leq \dots \leq (E|\xi|^n)^{1/n}$ 23

Thus $E|\xi| \leq (E|\xi|^2)^{1/2} \leq \dots \leq (E|\xi|^n)^{1/n}$.

<u>Lemma</u> (Markov ineq.)

Let $\xi \geq 0$ - integrable r.v., $c > 0$

Then $P(\xi \geq c) \leq \dfrac{E\xi}{c}$

$\triangle$ $E\xi \geq E(\xi \cdot \chi_{\xi \geq c}) \geq c\, E\chi_{\xi \geq c} = c\,P(\xi \geq c)$ $\blacktriangle$

<u>Def</u> The variance (dispersion) of a random variable ξ is defined as

$V\xi = E[(\xi - E\xi)^2]$;

$\delta = \sqrt{V\xi}$ is called standard deviation.

Note: $V\xi = E\xi^2 - (E\xi)^2$.

<u>Def</u> The covariance of r.v. ξ and η is defined as $\operatorname{cov}(\xi, \eta) = E((\xi - E\xi)(\eta - E\eta))$.

Note: $V(\xi + \eta) = V(\xi) + V(\eta) + 2\operatorname{cov}(\xi, \eta)$.

(so if $\operatorname{cov} = 0$, $V(\xi + \eta) = V(\xi) + V(\eta)$).

<u>Lemma</u> (Chebyshev ineq.) Let ξ be integrable r.v., $\varepsilon > 0$.

Then $P(|\xi - E\xi| \geq \varepsilon) \leq \dfrac{V(\xi)}{\varepsilon^2}$.

◁ By Markov ineq. for $\xi \geq 0$,

$$P(\xi \geq \varepsilon) = P(\xi^2 \geq \varepsilon^2) \leqslant \frac{E\xi^2}{\varepsilon^2},$$ ▷

now replace ξ by $|\xi - E\xi|$

<u>Lemma</u> (exponential Chebyshev ineq.)

Let $\xi \geq 0$ be r.v., $\varepsilon > 0$, $t > 0$, $\xi, e^{t\xi}$ - integrable. Then

$$P(\xi \geq \varepsilon) \leqslant e^{-t\varepsilon} E(e^{t\xi})$$

◁ $P(\xi \geq \varepsilon) = P(e^{t\xi} \geq e^{t\varepsilon}) \leqslant \dfrac{E(e^{t\xi})}{e^{t\varepsilon}}$ ▷

<u>Lemma</u> (E and tail probabilities)

Let $\xi \geq 0$ be an integrable r.v.

Then $\quad E\xi = \int_{[0,\infty)} P(\xi \geq x)\, dx$

◁ $E\xi = \int_{[0,\infty)} x\, dP_\xi = \int_{[0,\infty)} \left(\int_0^x dt\right) dP_\xi(x) =$

$= \left[\text{apply Fubini thm to } g(t,x) = \begin{cases} 1, & 0 \leq t \leq x \\ 0, & \text{otherwise} \end{cases}\right]$

$= \int_{[0,\infty)} P(\xi \geq t)\, dt$ ▷

<u>Exercise</u>

Let ξ be the normal random variable with density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad \sigma > 0.$$

Show that $m = E\xi$, $\sigma^2 = V\xi$. Thus, the distribution is fully determined by m, σ^2 (!)

We denote this r.v. by $\mathcal{N}(m, \sigma^2)$.

<u>Def</u> $E\xi^k$, $k = 1, 2, 3, \ldots$ are called moments of ξ

$E\xi$ - 1'st moment, $V\xi$ - combination of 1'st and 2'nd.

Let $F_\xi(x)$ be distribution function of a r.v. ξ and $\varphi(x)$ be a continuous function. What is $F_\eta(x)$, where $\eta = \varphi(\xi)$?

$$F_\eta(y) = P(\eta \le y) = P(\xi \in \varphi^{-1}(-\infty, y]) =$$

$$= \int_{\varphi^{-1}(-\infty, y]} dF_\xi$$

<u>Examples</u>: 1) $\eta = a\xi + b$, $a > 0$.

$$F_\eta(y) = P(\eta \le y) = P\left(\xi \le \frac{y-b}{a}\right) = F_\xi\left(\frac{y-b}{a}\right)$$

2) $\eta = \xi^2$. Then $F_\eta(y) = 0$ if $y < 0$

If $y \geq 0$, $F_\eta(y) = P(\xi^2 \leq y) =$

$= P(-\sqrt{y} \leq \xi \leq \sqrt{y}) = P_\xi(-\infty, \sqrt{y}] - P_\xi(-\infty, -\sqrt{y})$

$= F_\xi(\sqrt{y}) - F_\xi(-\sqrt{y}) + P(\xi = -\sqrt{y})$.

Let now ξ be a a.c. r.v. What is the density $f_\eta(y)$, $\eta = \varphi(\xi)$?

Let $\xi \in I$ - finite or infinite open interval,

$\varphi(x)$ - continuously differentiable, strictly increasing on I, so $\varphi'(x) \neq 0$.

Denote $h(y) = \varphi^{-1}(y)$ (well-defined, differentiable)

Then for $y \in \varphi(I)$,

$F_\eta(y) = P(\eta \leq y) = P(\xi \leq \varphi^{-1}(y)) = \int_{-\infty}^{h(y)} f_\xi(x)\,dx =$

$= \int_{-\infty}^{y} f_\xi(h(z))\, h'(z)\, dz$.

$\Rightarrow \underline{f_\eta(y)} = f_\xi(h(y))\, h'(y) = \underline{f_\xi(h(y))\, |h'(y)|}$

Similarly, $f_\eta(y) = f_\xi(h(y))\, |h'(y)|$ if φ is strictly decreasing, cont. differentiable on I.

Example 1) $\eta = \xi^2$. By above formula $F_{\xi^2}(y) = P(-\sqrt{y} \le \xi \le \sqrt{y})$, and since $h_1(y) = -\sqrt{y}$, $h_2(y) = \sqrt{y}$ are strictly monotone, $y \ge 0$,

$$f_\eta(y) = \begin{cases} 0, & y \le 0 \\ \frac{1}{2\sqrt{y}}\left(f_\xi(\sqrt{y}) + f_\xi(-\sqrt{y})\right), & y > 0 \end{cases}$$

(check!)

In particular, if ξ is $\mathcal{N}(0,1)$,

$$f_{\xi^2}(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0.$$

Random vectors: if (ξ, η) - random vector, joint distribution $F_{(\xi,\eta)}(x,y)$, $\varphi(x,y)$ - continuous

then $F_{\varphi(\xi,\eta)}(z) = \int_{\varphi^{-1}(-\infty, z]} dF_{(\xi,\eta)}$.

Some properties of <u>independent</u> r.v.

1) Let ξ, η - be independent, so $F_{(\xi,\eta)}(x,y) = F_\xi(x) F_\eta(y)$. Consider $\xi + \eta$.

$$F_{\xi+\eta}(z) = \int_{\varphi^{-1}(-\infty, z]} dF_\xi \, dF_\eta = \int_{\mathbb{R}^2} \chi_{\{x+y\le z\}} \, dF_\xi(x) \, dF_\eta(y)$$

$$\varphi(x,y) = x + y.$$

$$= \int_{-\infty}^{\infty} dF_\xi(x) \Big(\int_{-\infty}^{\infty} \chi_{\{x+y \le z\}} \, dF_\eta(y) \Big) =$$

$$= \int_{-\infty}^{\infty} F_\eta(z-x)\, dF_\xi(x) = \int_{-\infty}^{\infty} F_\xi(z-y)\, dF_\eta(y).$$

Thus, we obtained

__Lemma__ The distribution function $F_{\xi+\eta}$ of the sum of 2 independent r.v. is the convolution of their distribution functions.

Similarly, we easily obtain

__Lemma__ If ξ, η - independent a.c. r.v. then the density $f_{\xi+\eta}$ is the convolution of the densities.

__Examples__ (ξ, η - independent normal r.v.)

a) Let ξ be $\mathcal{N}(m_1, \sigma_1^2)$, η be $\mathcal{N}(m_2, \sigma_2^2)$,

i.e. $f_\xi(x) = \frac{1}{\sigma_1} \varphi\left(\frac{x-m_1}{\sigma_1}\right)$, $f_\eta = \frac{1}{\sigma_2} \varphi\left(\frac{x-m_2}{\sigma_2}\right)$,

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

check!

Then $f_{\xi+\eta}(z) = \int_{-\infty}^{\infty} f_\eta(z-x) f_\xi(x)\, dx = \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}} \varphi\left(\frac{z-(m_1+m_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)$

Thus the sum of 2 independent normal r.v. is the normal r.v. $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$.

6) Let $\xi_1, \dots, \xi_n$ be independent $N(0,1)$

Then $$f_{\xi_1^2+\dots+\xi_n^2}(x) = \frac{1}{2^{n/2}\,\Gamma(n/2)}\, x^{n/2-1} e^{-x/2}, \quad x>0 \qquad (*)$$

(follows from the formula for f_{ξ^2} using induction – – check!)

The r.v. $\xi_1^2+\dots+\xi_n^2$ is called χ_n^2 and $(*)$ is called χ^2-distribution with n degrees of freedom.

2) <u>Thm</u> Let ξ, η be independent integrable r.v. Then $\xi\cdot\eta$ is integrable and $E\xi\cdot\eta = E\xi\cdot E\eta$.

◁ First assume $\xi, \eta \geq 0$. Consider

$$\xi_n = \sum_{k=0}^{\infty} \frac{k}{n}\, \chi_{\{k/n \leq \xi(\omega) < \frac{k+1}{n}\}}$$

$$\eta_n = \sum_{k=0}^{\infty} \frac{k}{n}\, \chi_{\{k/n \leq \eta(\omega) < \frac{k+1}{n}\}}.$$

Then $\xi_n \leq \xi$, $\eta_n \leq \eta$, $|\xi-\xi_n| \leq \frac{1}{n}$, $|\eta-\eta_n| \leq \frac{1}{n}$, $\forall n$

Since ξ, η are integrable, by dominated convergence,

$$\lim_{n\to\infty} E\xi_n = E\xi, \quad \lim_{n\to\infty} E\eta_n = E\eta.$$

Now write

$$E\zeta_n\eta_n \underset{\substack{\text{monotone}\\ \text{convergence}}}{=} \sum_{j,k\ge 0} \frac{jk}{n^2} E\left(\chi_{\{\frac{j}{n}\le\zeta<\frac{j+1}{n}\}}\chi_{\{\frac{k}{n}\le\eta<\frac{k+1}{n}\}}\right)$$

$$\underset{\text{independence}}{=} \sum_{j,k\ge 0} \frac{j\cdot k}{n^2} E\left(\chi_{\{\frac{j}{n}\le\zeta<\frac{j+1}{n}\}}\right) E\left(\chi_{\{\frac{k}{n}\le\eta<\frac{k+1}{n}\}}\right)$$

$$= E\zeta_n\cdot E\eta_n$$

Since

$$|E\zeta\cdot\eta - E\zeta_n\eta_n| \le E|\zeta\eta-\zeta_n\eta_n| =$$

$$= E|\zeta(\eta-\eta_n)+\eta_n(\zeta-\zeta_n)| \le \frac{1}{n}E|\zeta| + \frac{1}{n}E\left(|\eta|+\frac{1}{n}\right)$$

$$\longrightarrow 0, \quad n\to\infty,$$

we have that

$$E\zeta\eta = \lim_{n\to\infty} E\zeta_n\eta_n = \lim_{n\to\infty} E\zeta_n \cdot \lim_{n\to\infty} E\eta_n = E\zeta\cdot E\eta,$$

and $E\zeta\eta<\infty$.

The result in general case follows by using representations $\zeta=\zeta^+-\zeta^-$, $\eta=\eta^+-\eta^-$

<u>Def</u> r.v. ζ and η are called <u>uncorrelated</u> if $\operatorname{cov}(\zeta,\eta)=0$.

<u>Corollary of thm</u>: Independent r.v. are uncorrelated.

Indeed, $\operatorname{cov}(\zeta,\eta) = E\zeta\eta - E\zeta\cdot E\eta$.

The converse is not true:

r.v. α takes values $0, \pi/2, \pi$ with probability $1/3$. Then $\xi = \sin\alpha$, $\eta = \cos\alpha$ are uncorrelated (note: $E\xi = 1/3$, $E\eta = 0$)

but $P(\xi=1, \eta=1) = 0 \neq 1/9 = P(\xi=1)\cdot P(\eta=1)$.

II Weak Law of Large Numbers LLN and de Moivre-Laplace central limit thm CLT.

Consider $(\Omega_n, \mathcal{A}, P_n)$ with

$\Omega_n = \{\omega : \omega = (a_1, \dots, a_n),\ a_j = 0,1\}$

$\mathcal{A} = \{A : A \subset \Omega_n\}$, $p(\omega) = p^{\sum_1^n a_j} q^{n - \sum_1^n a_j}$,

$0 < p < 1$, $q = 1-p$ (check that $P_n(\Omega_n) = 1$).

– n independent experiments with 2 outcomes or Bernoulli scheme (e.g. throwing of biased coin).

Let $\xi_1, \dots, \xi_n$ be r.v. defined as $\xi_j(\omega) = a_j$, $j = 1, \dots, n$. Note that ξ_j are independent, identically distributed (i.i.d.):

$P_{\xi_j}(\xi_j = 1) = P_n(\omega \in \Omega_n : \xi_j = 1) = p$;

$P_{\xi_j}(\xi_j = 0) = P_n(\omega \in \Omega_n : \xi_j = 0) = q$ $\quad \forall j$,

$P_n(\omega \in \Omega_n : \xi_j \in B_j,\ j = 1, \dots, n) = \prod_{j=1}^n P_n(\omega \in \Omega_n : \xi_j \in B_j)$

B_j - Borel sets in $\mathbb{R}$.

Let $\underline{S_n = \xi_1 + \dots + \xi_n}$

Then $ES_n = \sum_1^n E\xi_j =$

$$= \sum_1^n \left(1 \cdot P_{\xi_j}(\xi_j = 1) + 0 \cdot P_{\xi_j}(\xi_j = 0)\right) = np$$

Thus the mean value of $\frac{1}{n} S_n$ is equal to p.

Question: what is $\left|\frac{1}{n} S_n(\omega) - p\right|$ for large n?

Is it small?

Cannot be that $\left|\frac{1}{n} S_n(\omega) - p\right| \to 0$ uniformly in ω because $\frac{1}{n} S_n(\omega = \{\text{all } a_j = 1\}) = 1$.

But $P_n\left(\frac{S_n}{n} = 1\right) = p^n \to 0$, $n \to \infty$, so there is some hope!

<u>Remark</u> Since sequence P_n, $n = 1, 2 \ldots$ is consistent, according to Kolmogorov extension thm, there exists unique $P \equiv P_{(\xi_1, \xi_2 \ldots)}$ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$

s.t. $P(\xi \in \mathbb{R}^\infty : (\xi_1, \xi_2 \ldots, \xi_n) \in B_n) =$

$$= P_{n\,(\xi_1 \ldots \xi_n)}((\xi_1 \ldots \xi_n) \in B_n)$$

$$(\equiv P_n(\omega \in \Omega_n : (\xi_1 \ldots \xi_n) \in B_n))$$

Using Chebyshev inequality and the fact that independent r.v. are uncorrelated, we obtain:

$$P\left(\left|\frac{S_n}{n}-p\right|\geq \varepsilon\right)\leq$$

$$\leq \frac{V(S_n/n)}{\varepsilon^2} = \frac{1}{\varepsilon^2}\sum_{j=1}^{n} V\left(\frac{\xi_j}{n}\right) = \frac{1}{n^2\varepsilon^2}\sum_{j=1}^{n} V(\xi_j)$$

$$= \frac{1}{n^2\varepsilon^2}\, npq = \frac{pq}{n\varepsilon^2} \to 0, \quad n\to\infty .$$

i.e. $\underline{\dfrac{S_n}{n} \xrightarrow{\text{prob}} p}$ (means conv. in measure, called convergence in probability).

This is Bernoulli LLN.

Remark 1 "time" average $\dfrac{\xi_1+\dots+\xi_n}{n}$ converges in probability to "space" average $E\xi_i = p$.

Remark 2 $P\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right) =$

$$0 \swarrow \quad = \sum_{\{k:\ |\frac{k}{n}-p|\geq\varepsilon\}} P(S_n=k) \quad -$$

~~measure~~ concentrated at $k \sim np$!

More generally:

Let $\xi_1 \dots \xi_n, \dots$ - integrable r.v.

Let $S_n^{(c)} = \sum_{j=1}^{n}(\xi_j - E\xi_j)$ Note $ES_n^{(c)} = 0$.

Thm (weak LLN)

Let $\xi_1, \xi_2 \dots$ be uncorrelated integrable r.v. and s.t. $V(\xi_j) \leq C$ for some $C>0$ and all $n \geq 1$.

Then $\frac{S_n^{(c)}}{n} \xrightarrow{prob} 0$, i.e.

$$\lim_{n\to\infty} P\left(\left| \frac{S_n^{(c)}}{n} \right| \geq \varepsilon \right) = 0 \quad \forall \varepsilon > 0.$$

◁ By Chebyshev ineq. and since ξ_j are uncorrelated,

$$P\left(\left| \frac{S_n^{(c)}}{n} \right| \geq \varepsilon \right) \leq \frac{V\left(\frac{S_n^{(c)}}{n}\right)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{j=1}^{n} V(\xi_j)$$

$$\leq \frac{C}{n\varepsilon^2} \to 0, \quad n \to \infty$$ ▷

Corollary If $\xi_1, \xi_2 \dots$ are integrable i.i.d., $V(\xi_j) < \infty$, then $\frac{1}{n} \sum_{1}^{n} \xi_j \xrightarrow{prob} E\xi_1$

("time" average converges to "space" average in probability).

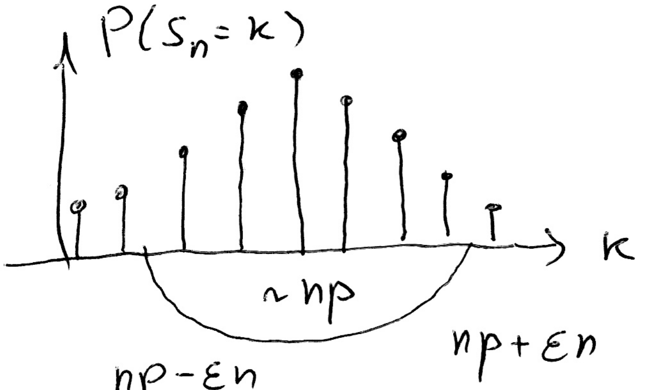
In particular, we reobtain Bernoulli LLN.

Return to Bernoulli: $S_n = \xi_1 + \dots + \xi_n$,

ξ_i - Bernoulli r.v. as above.

Recall Remark 2: S_n tends to be close to $n \cdot p$ for large n.

Question: what is the distribution of S_n for large n?



$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

LLN: contribution of outside of $(np - \varepsilon n, np + \varepsilon n)$ is small.

<u>Exercise</u>: $P(S_n = k)$ is monotone in k below and above its point of maximum.

We will need the O-notation for sequences g_n, f_n:

$g_n = O(|f_n|)$ as $n \to \infty$ if $\left|\frac{g_n}{f_n}\right|$ is bounded for sufficiently large n, i.e. $\exists\, C, N$ s.t. $\forall n > N$: $|g_n| \le C|f|$

$g_n = o(|f_n|)$ as $n \to \infty$ if $\lim\limits_{n \to \infty} \left|\frac{g_n}{f_n}\right| = 0$, i.e. $\forall \varepsilon > 0\ \exists\, N_0(\varepsilon)$ s.t. $\forall n > N_0$ $|g_n| \le \varepsilon |f_n|$.

If $g_n = g_n(\alpha)$, $f_n = f_n(\alpha)$ and $g_n = O(|f_n|)$ with C, N independent of α or $g_n = o(|f_n|)$ with N_0

independent of α, we say that the terms o, O are uniform in α.

The following result compares the probability mass function $P(S_n = k)$, $k = 0, 1, \dots, n$ of r.v. S_n whose expectation is np and variance is $np(1-p)$, with the density of $N(np, np(1-p))$:

<u>Thm</u> (local limit thm) For any $0 < p < 1$,

$$\max_{0 \le k \le n} \left| P(S_n = k) - \frac{1}{\sqrt{2\pi p(1-p)}\,\sqrt{n}} e^{-\frac{x^2}{2p(1-p)}} \right| = o\left(\frac{1}{\sqrt{n}}\right),$$

$n \to \infty$

where $x \equiv x_{k,n} = \dfrac{k - np}{\sqrt{n}}$.

◁ Let $A_n > 0$, $A_n = o(n)$, $n \to \infty$, e.g. $A_n = n^{\varepsilon}$, $0 < \varepsilon < 1$, and consider first only such k that $|x_{k,n}| \le \dfrac{A_n}{\sqrt{n}}$, i.e.

$$np - A_n \le k \le np + A_n,$$

$$n(1-p) - A_n \le n - k \le n(1-p) + A_n.$$

Then k, $n-k \to \infty$, $n \to \infty$ and we can use Stirling formula:

$$P(S_n = k) = \frac{n!}{k!\,(n-k)!}\, p^k (1-p)^{n-k} =$$

$$= \frac{\sqrt{\cancel{2\pi} n}}{\sqrt{\cancel{2\pi} k}\sqrt{2\pi(n-k)}} e^{n\log n - \cancel{n} - k\log k + \cancel{k} - (n-k)\log(n-k) + \cancel{n-k}} \cdot p^k(1-p)^{n-k}\left(1+O\left(\tfrac{1}{n}\right)\right)$$ 38

$$= \left[\begin{array}{ll} k = np\left(1+O\left(\frac{A_n}{n}\right)\right), & n-k = n(1-p)\left(1+O\left(\frac{A_n}{n}\right)\right) \\ k = np + x\sqrt{n} & n-k = n(1-p) - x\sqrt{n} \end{array}\right]$$

$$= \sqrt{\frac{\cancel{n}}{\cancel{n}p\cdot 2\pi n(1-p)}}\left(1+O\left(\frac{A_n}{n}\right)\right)\cdot$$
$$\cdot e^{\cancel{n\log n} - (np+x\sqrt{n})\left(\log(\cancel{np}) + \frac{x}{p\sqrt{n}} - \frac{x^2}{2p^2 n} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3\right)\right)}\cdot$$
$$\cdot e^{-(n(1-p)-x\sqrt{n})\left(\log(\cancel{n(1-p)}) - \frac{x}{(1-p)\sqrt{n}} - \frac{x^2}{2(1-p)^2 n} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3\right)\right)}\cdot$$
$$\cdot e^{\cancel{k\log p + (n-k)\log(1-p)}} =$$

$$= \sqrt{\frac{1}{2\pi p(1-p)n}}\left(1+O\left(\frac{A_n}{n}\right)\right)\cdot$$
$$\cdot e^{-(np+x\sqrt{n})\left(\frac{x}{p\sqrt{n}} - \frac{x^2}{2p^2 n} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3\right)\right)}\cdot$$
$$\cdot e^{+(n(1-p)-x\sqrt{n})\left(\frac{x}{(1-p)\sqrt{n}} + \frac{x^2}{2(1-p)^2 n} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3\right)\right)}$$

$$= \sqrt{\frac{1}{2\pi p(1-p)n}}\left(1+O\left(\frac{A_n}{n}\right)\right)\cdot e^{\cancel{-\sqrt{n}x} + \frac{x^2}{2p} - \frac{x^2}{p} + \frac{x^3}{2p^2\sqrt{n}} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3 n\right)}\cdot$$
$$\cdot e^{\cancel{+\sqrt{n}x} + \frac{x^2}{2(1-p)} - \frac{x^2}{1-p} - \frac{x^3}{2(1-p)^2\sqrt{n}} + O\left(\left(\frac{x}{\sqrt{n}}\right)^3 n\right)}$$

$$= \frac{1+O\left(\frac{A_n}{n}\right)}{\sqrt{2\pi p(1-p)n}} e^{-\frac{x^2}{2}\left(\frac{1}{p}+\frac{1}{1-p}\right)} + O\left(\frac{A_n^3}{\sqrt{n}\,{}^3\sqrt{n}}\right)$$

$$= \frac{1}{\sqrt{2\pi p(1-p)n}} e^{-\frac{x^2}{2p(1-p)}} \left(1 + O\left(\frac{A_n^3}{n^2}\right) + O\left(\frac{A_n}{n}\right)\right),$$

if $A_n = o(n^{2/3})$.

Let $A_n = n^{7/12}$; $A_n^3/n^2 = n^{-1/4}$, $A_n/n = n^{-5/12}$.

Thus the ~~that~~ formula in thm holds for

$$\max_{np-A_n \le k \le np+A_n}$$

But $$\max_{\substack{(k > np+A_n) \cup \\ (k < np-A_n)}} \left| P(S_n=k) - \frac{e^{-x^2/2p(1-p)}}{\sqrt{2\pi p(1-p)n}} \right| \le$$

(by monotonicity in k) $\le \max\left(P(S_n = \lfloor np+A_n \rfloor),\ P(S_n = \lceil np-A_n \rceil)\right)$

$$+ \frac{1}{\sqrt{2\pi p(1-p)n}} e^{-\frac{A_n^2}{2p(1-p)n}}, \quad A_n = n^{7/12} =$$

$$= o\left(\frac{1}{\sqrt{n}}\right), \quad n \to \infty$$

<u>Thm</u> (de Moivre-Laplace CLT)

For any $0<p<1$, $x \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) = \Phi(x),$$

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$ — distribution function of $N(0,1)$.

◁ $$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) = P\left(S_n \le np + x\sqrt{np(1-p)}\right)$$

$$= \sum_{k=0}^{\lfloor np + x\sqrt{np(1-p)}\rfloor} P(S_n = k) =$$ (using local thm, estimates for $P(S_n=k)$), $(A_n = n^{7/12})$,

$$= o(1) + \sum_{k=\lfloor np - A_n\rfloor}^{\lfloor np + x\sqrt{np(1-p)}\rfloor} \frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2}$$

$$\xrightarrow[n\to\infty]{} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$$ ▷

<u>Exercise</u>: Obtain Bernoulli's LLN from this CLT.

<u>Poisson distribution</u>: fix k and let $p = p(n) \to 0$ as $n \to \infty$ s.t. $p(n)\cdot n \to \lambda > 0$.

Then $$P(S_n = k) = \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k \cdot \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} \longrightarrow \frac{1}{k!}\lambda^k \cdot e^{-\lambda}.$$

III Convergence. Characteristic functions.

Consider r.v. on prob. space $(\Omega, \mathcal{F}, P)$

Def A sequence of r.v. $\xi_1, \xi_2 \ldots$ converges to r.v. ξ

1) a.s. (or P-a.s., with prob. 1, a.e. w.r.t. P)

if $\xi_n(\omega) \to \xi(\omega)$ a.e. on Ω (w.r.t. P)

[denoted $\xi_n \xrightarrow{a.s.} \xi$]

2) in probability (or in measure P) if

$\forall \varepsilon > 0 \quad P(|\xi_n - \xi| \geq \varepsilon) \to 0$ [denoted $\xi_n \xrightarrow{P} \xi$]

3) in L^p, $1 \leq p < \infty$ (or in mean of order p) if

$E|\xi_n - \xi|^p \to 0$ [denoted $\xi_n \xrightarrow{L^p} \xi$]

4) in distribution (or weakly) if

$E_n f(\xi_n) \longrightarrow E f(\xi)$ for any bounded, continuous $f(x)$.

[denoted $\xi_n \xrightarrow{d} \xi$ or $\xi_n \xrightarrow{w} \xi$]

Here E_n means that ξ_n can be defined on different prob. spaces.

Exercises:

1) $\xi_n \xrightarrow{a.s.} \xi \underset{\not\Leftarrow}{\Rightarrow} \xi_n \xrightarrow{P} \xi$, however

if $\xi_n \xrightarrow{P} \xi$, $\exists$ subsequence $\xi_{n_k} \xrightarrow{a.s.} \xi$.

2) $\xi_n \xrightarrow{L^P} \xi \underset{\not\Leftarrow}{\Rightarrow} \xi_n \xrightarrow{P} \xi$ (use Markov ineq.)

3) $\xi_n \xrightarrow{a.s.} \xi \underset{\not\Leftarrow}{\not\Rightarrow} \xi_n \xrightarrow{L^P} \xi$

For this we have counterexample:

$\xi_1, \xi_2 \ldots$ independent r.v. $P(\xi_n = 1) = \frac{1}{n}$

$P(\xi_n = 0) = 1 - \frac{1}{n}$

$E|\xi_n - 0|^P = \frac{1}{n} \to 0$, so $\xi_n \xrightarrow{L^P} 0$,

but $\{\omega : \xi_n \to 0\} = (\xi_n = 0 \text{ ev.}) = \bigcup_{n=1}^{\infty} \underbrace{\bigcap_{k \geq n} \{\xi_k = 0\}}_{\text{increasing sequence of sets}}$

so $P(\xi_n \to 0) = \lim_{n\to\infty} P\left(\bigcap_{k\geq n} \{\xi_k = 0\}\right)$

$= \lim_{n\to\infty} \prod_{k\geq n} P\{\xi_k = 0\} = \lim_{n\to\infty} \prod_{k\geq n} (1 - \frac{1}{k}) = 0$

Indeed $\prod_{k\geq n} (1-\frac{1}{k}) = \lim_{N\to\infty} \prod_{k=n}^{N} \frac{k-1}{k} = \lim_{N\to\infty} \frac{n-1}{\cancel{n}} \frac{\cancel{n}}{\cancel{n+1}} \cdots \frac{\cancel{N-1}}{N}$

$= 0$. Thus $\xi_n \not\xrightarrow{a.s.} 0$.

<u>Thm</u> Let $\xi_n \geq 0$, integrable, $n = 1, 2 \ldots$,

$\xi_n \xrightarrow{a.s.} \xi$ integrable,

and $E\xi_n \to E\xi$.

Then $\xi_n \xrightarrow{L^1} \xi$.

◁ $E|\xi_n - \xi| = E(\xi - \xi_n)\chi_{\xi \geq \xi_n} + E(\xi_n - \xi)\chi_{\xi_n > \xi}$

$= 2E(\xi - \xi_n)\chi_{\xi \geq \xi_n} + E(\xi_n - \xi)$,

but $0 \leq (\xi - \xi_n)\chi_{\xi \geq \xi_n} \leq \xi$, so by dominated convergence, $E(\xi - \xi_n)\chi_{\xi \geq \xi_n} \to 0$ ▷

<u>Thm</u>

$$\begin{array}{ccc} \xi_n \xrightarrow{a.s.} \xi & \Rightarrow & \\ & & \xi_n \xrightarrow{P} \xi \Rightarrow \xi_n \xrightarrow{d} \xi \\ \xi_n \xrightarrow{L^p} \xi & \nearrow & \end{array}$$

◁ Remains to prove $\xi_n \xrightarrow{P} \xi \Rightarrow \xi_n \xrightarrow{d} \xi$.

Let $\varepsilon > 0$; $f(x)$ - continuous, $|f(x)| \leq C \ \forall x$.

Choose N s.t. $P(|\xi| > N) \leq \varepsilon/4C$

Choose δ s.t. $|f(x)-f(y)| \leq \varepsilon/2$ (44)

for $|x| \leq N$, $|x-y| \leq \delta$

(continuous f on compact set)

$$E|f(\xi_n)-f(\xi)| = E(1\cdot 1;\ |\xi_n-\xi| \leq \delta,\ |\xi| \leq N) +$$

$$+ E(1\cdot 1;\ |\xi_n-\xi| \leq \delta,\ |\xi| > N) +$$

$$+ E(1\cdot 1;\ |\xi_n-\xi| > \delta) \leq$$

$$\leq \varepsilon/2 + 2c\frac{\varepsilon}{4c} + 2c\,P(|\xi_n-\xi| > \delta)$$

$$\leq \varepsilon/2 + \varepsilon/2 + 2c\,\underbrace{P(|\xi_n-\xi| > \delta)}_{\to 0,\ n\to\infty}$$

since $\xi_n \xrightarrow{P} \xi$.

$\Rightarrow$ $E|f(\xi_n)-f(\xi)| \leq 2\varepsilon$, $\forall n > n_0$ for some $n_0(\varepsilon,\delta)$. $\Rightarrow$ $\lim\limits_{n\to\infty} E|f(\xi_n)-f(\xi)| = 0$

Now note: $Ef_n = E(f-f_n) + Ef$, $f_n = f(\xi_n)$

<u>Thm</u> Let $\xi_1, \xi_2 \dots ; \xi$ - be integrable r.v.
The following statements are equivalent

1) $\xi_n \xrightarrow{d} \xi$

2) $\limsup P_n(\xi_n \in B) \leq P(\xi \in B)$ $\forall$ closed B

3) $\liminf P_n(\xi_n \in A) \geq P(\xi \in A)$ $\forall$ open A

4) $\lim P_n(\xi_n \in C) = P(\xi \in C)$
for any C s.t. $P(\xi \in \partial C) = 0$.

5) distribution functions $F_{\xi_n}(x) \to F_\xi(x)$
at any point of continuity of $F_\xi(x)$.

$\triangleleft$ $1 \Rightarrow 2$ Let B - closed, $f(x) = \chi_B(x)$

$$f_\varepsilon(x) = g\left(\tfrac{1}{\varepsilon}\rho(x, B)\right), \quad \varepsilon > 0$$

$$\rho(x,B) = \inf\{|x-y| : y \in B\}, \quad g(t) = \begin{cases} 1, & t \leq 0 \\ 1-t, & 0 \leq t \leq 1 \\ 0, & t \geq 1 \end{cases}.$$

Let $B_\varepsilon = \{x : \rho(x,B) < \varepsilon\}$

Note $B_\varepsilon \downarrow B$ as $\varepsilon \downarrow 0$ (i.e. $B_{\varepsilon_1} \supset B_{\varepsilon_2}$, $\varepsilon_1 > \varepsilon_2$,
$\bigcap_{\varepsilon > 0} B_\varepsilon = B$ (1))

We have

$$P_n(\xi_n \in B) = \int f \, dP_n \leq \int f_\varepsilon \, dP_n$$

$\Rightarrow \limsup P_n(\xi_n \in B) \leq \limsup \int f_\varepsilon \, dP_n =$

$\underset{\substack{\uparrow \\ \text{since } \xi_n \xrightarrow{d} \xi}}{=} \int f_\varepsilon \, dP \leq P(B_\varepsilon) \underset{\substack{\uparrow \\ \text{by (1)}}}{\longrightarrow} P(B), \ \varepsilon \downarrow 0$

$2 \Rightarrow 3$ Consider $A = \mathbb{R} \setminus B$ — B-closed.

$3 \Rightarrow 4$ Recall $\bar{C} = C \cup \partial C$; $C^\circ = C \setminus \partial C$

Since $P(\xi \in \partial C) = 0$,

$\limsup P_n(C) \leq \limsup P_n(\bar{C}) \leq P(\bar{C}) = P(C)$
$\liminf P_n(C) \geq \liminf P_n(C^\circ) \geq P(C^\circ) = P(C)$

$\Rightarrow \lim P_n(C) = P(C)$

$4 \Rightarrow 5$ obvious

$5 \Rightarrow 1$ Based on the following single prob. space theorem:

Thm (SPS) Suppose that 5 holds. Then there exists prob. space $(\Omega', \mathcal{F}', P')$ and r.v. $X, X_2 \ldots$ on it s.t. $X_n \overset{d}{=} \xi_n$, $n = 1, 2, \ldots$, $X \overset{d}{=} \xi$ and $X_n \to X$ P'-a.s.

Assume this thm. We have to show that $\lim E_n(f(\xi_n)) = E(f(\xi))$ for any continuous, bounded $f(x)$.

For such f, $f(X_n) \to f(X)$ P'-a.s. and (47)
by dominated convergence,

$$E_n f(\xi_n) = E f(X_n) \to E f(X) = E f(\xi)$$

The proof of SPS thm is based on the following

Lemma Let μ be r.v. with distribution function $F(x)$. Let U be r.v. uniformly distributed on $[0,1]$, and let

$$F^{-1}(u) = \sup\{y : F(y) < u\}.$$

Then $\mu \stackrel{d}{=} F^{-1}(U)$.

If $z = F(x)$ is invertible,

$$P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$$

since U is uniform on $[0,1]$.

In general, the same is true since

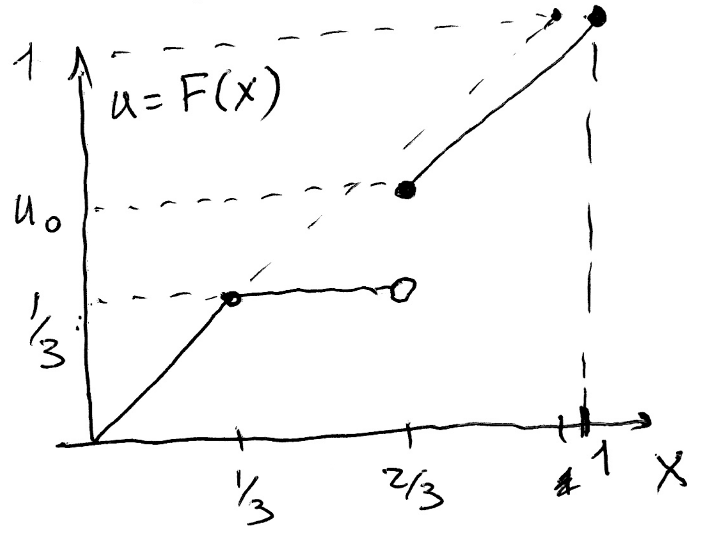
$$u \le F(x) \Leftrightarrow F^{-1}(u) \le x$$

Remains to prove this equivalence.

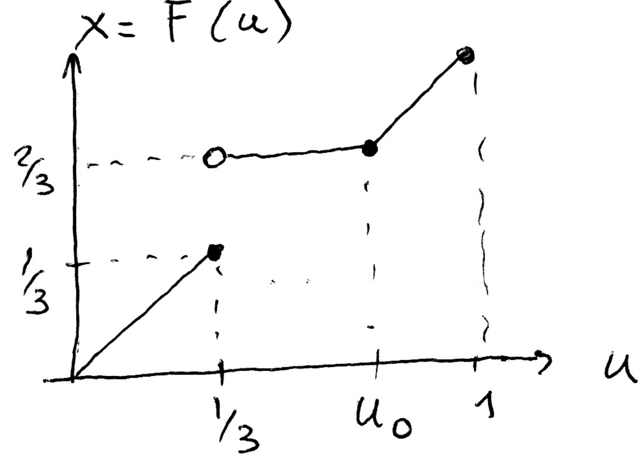
$\Rightarrow$ Let $u \le F(x)$. Assume for contradiction that $F^{-1}(u) > x$, i.e. $\sup\{y : F(y) < u\} > x$,
$\sup\{y : F(y) < u \le F(x)\} > x$ – contradiction as F - nondecreasing. Thus $F^{-1}(u) \le x$.

($\Leftarrow$) Let $x \geq F^{-1}(u) = \sup\{y : F(y) < u\}$. Then $F(x) \geq u$, since $F(x)$ is nondecreasing, continuous from the right $\triangleright$

Example:



– F for ξ uniform on $[0, 1/3] \cup [2/3, 1]$ with an atom at $2/3$



Proof of SPS thm:

Let $(\Omega', \mathcal{F}', P') = ([0,1], \mathcal{B}[0,1], \cancel{\text{Lebes}}\ \lambda)$, λ - Lebesgue measure. We will prove that $F_n^{-1}(u) \longrightarrow F^{-1}(u)$ for any u s.t. preimage of u under F is finite. λ-measure of all other points is zero since they form a countable set. So $X_n = F_n^{-1}(U) \to X = F^{-1}(U)$ a.s. ~~B~~ By Lemma, $X_n \stackrel{d}{=} \xi_n$, $X \stackrel{d}{=} \xi$.

Let u be ~~the~~ point s.t. the preimage of u under F is finite.

1) For any $x < F^{-1}(u)$, by definition of F^{-1}, $F(x) < u$. If $x \in C_F$ (points of continuity of F) then $F_n(x) \to F(x)$ (assumption 5), so $F_n(x) < u$ for all sufficiently large n.

~~Th~~ $\Rightarrow x \leq F_n^{-1}(u) \Rightarrow x \leq \liminf F_n^{-1}(u)$,

Now choosing a sequence $x_k \in C_F$ s.t. $x_k \uparrow F^{-1}(u)$ we obtain: $\underline{F^{-1}(u) \leq \liminf F_n^{-1}(u)}$.

2) For any $x > F^{-1}(u)$, we have $F(x) \geq u$. If $F(x) = u$ then $F(y) = u$ on $[F^{-1}(u), x]$, i.e. the preimage is infinite, but u is not such a point by assumption. So $F(x) > u$. Repeating arguments in 1) we obtain

$\underline{F^{-1}(u) \geq \limsup F_n^{-1}(u)}$

Therefore $F_n^{-1}(u) \to F^{-1}(u)$ ∎

<u>Def</u> A family of prob. measures $\mathcal{P} = \{P_\alpha, \alpha \in A\}$ (and the corresponding set of distribution functions F_α) is called relatively compact if every sequence of measures from $\mathcal{P}$ contains a subsequence that weakly converges to a prob. measure (not necessarily from $\mathcal{P}$ - that is why "relatively")

<u>Remark</u> Weak convergence of measures is metrizable, as one can show. The corresponding metric is called Levy-Prokhorov metric. Relative compactness of $\mathcal{P}$ means that the closure of $\mathcal{P}$ is compact in the topology of Levy-Prokhorov metric.

A given $\mathcal{P}$ is not necessarily relatively compact.

<u>Examples</u>: Let ξ_n, $n = 1, 2 \ldots$ - r.v. and $F_{\xi_n}(x) \to F(x)$ for all $x \in C_F$. Then $F(x)$ is not necessarily a distribution function!

1) Let ξ_n be $U[n, n+1]$. Then $F(x) \equiv 0$ - runaway to ∞.

2) Let ξ_n be $U[-n, n]$. Then $F(x) \equiv 1/2$ - spread to ∞.

We need a criterion that a given $\mathcal{P}$ is relatively compact.

We start with the following auxilliary result:

Denote $G = \{ F : \mathbb{R} \to [0,1]$ s.t. F is nondecreasing, continuous from the right $\}$ – the set of generalized distribution functions.

[Distribution functions form a subset of G for which $F(-\infty) = 0$, $F(\infty) = 1$]

<u>Thm</u> (Helly) The set G is sequentially compact, i.e. for any sequence $F_n \in G$, $n=1,2\ldots$ there exist a function $F \in G$ and a subsequence $F_{n_k} \subset \{F_n\}$ s.t. $F_{n_k}(x) \to F(x)$ $\forall x \in C_F$.

$\triangle$ Let $Q = \{q_1, q_2, \ldots\}$ – <u>the set</u> of rational numbers.

Since $F_n(q_1)$, $n=1,2\ldots$ is bounded, it has a converging subsequence $F_{n_k^{(1)}}(q_1) \to f(q_1)$, $\{n_k^{(1)}\} \subset \{n\}$.

Since $F_{n_k^{(1)}}(q_2)$, $k=1,2\ldots$ is bounded, there exists $F_{n_k^{(2)}}(q_2) \to f(q_2)$, $\{n_k^{(2)}\} \subset \{n_k^{(1)}\}$. And so on.

Denote $n_k \equiv n_k^{(k)}$ – diagonal subsequence.

Then $F_{n_k}(q) \to f(q)$ $\forall q \in \mathbb{Q}$. Note that $f(q)$ is nondecreasing.

Define $F(x) = \inf\{f(q) : q \in \mathbb{Q}, q > x\} =$

$$= \lim_{q \to x+} f(q).$$

Then $F(x) \in G$ (check!). Moreover,

$F(q) \geq f(q)$, $\forall q \in \mathbb{Q}$,

$F(x) \leq f(q)$ if $x < q$, $q \in \mathbb{Q}$.

Remains to show that $F_{n_k}(x) \to F(x)$ $\forall x \in C_F$.

Let $x \in C_F$. Fix $\varepsilon > 0$.

Choose $y < r < x < q$, $r, q \in \mathbb{Q}$ s.t.

$$F(x) - \varepsilon < F(y) \leq F(r) \leq F(x) \leq F(q) < F(x) + \varepsilon$$

$$\Rightarrow F(x) - \varepsilon < f(r) \leq F(x) \leq f(q) < F(x) + \varepsilon$$

So for sufficiently large k:

$$F(x) - \varepsilon < F_{n_k}(r); \quad F_{n_k}(q) < F(x) + \varepsilon$$

$$\Rightarrow F(x) - \varepsilon < F_{n_k}(x) < F(x) + \varepsilon$$

for all such k

$$\Rightarrow F_{n_k}(x) \to F(x)$$

<u>Def</u> A family of prob. measures $\mathcal{P} = \{P_\alpha, \alpha \in A\}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called <u>tight</u> if

$\forall \varepsilon > 0 \ \exists$ compact set $K \subset \mathbb{R}$

s.t. $\sup\limits_{\alpha \in A} P_\alpha(\mathbb{R} \setminus K) \leq \varepsilon$

<u>Thm</u> (Prokhorov)

$\mathcal{P}$ is tight $\Leftrightarrow$ $\mathcal{P}$ is relatively compact.

$\triangle$ ($\Leftarrow$) Let $\mathcal{P}$ be relatively compact.

Suppose it is not tight, i.e.

$\exists \varepsilon > 0$ s.t. for any compact $K \subset \mathbb{R}$

$\sup\limits_{\alpha} P_\alpha(\mathbb{R} \setminus K) > \varepsilon$, in particular

for $K = [-n, n]$ ~~$\neq$~~ with any fixed n.

Therefore, for any n $\exists P_{\alpha_n}$ s.t.

$$P_{\alpha_n}(\mathbb{R} \setminus (-n, n)) > \varepsilon. \quad (*)$$

By relative compactness $\exists P_{\alpha_{n_k}} \xrightarrow{w} Q$ – prob. measure.

Thus $\limsup P_{\alpha_{n_k}}(\mathbb{R} \setminus (-n, n)) \leq Q(\mathbb{R} \setminus (-n, n))$

But $Q(\mathbb{R} \setminus (-n, n)) \to 0$, $n \to \infty$, which contradicts $(*)$.

Therefore $\mathcal{P}$ is tight.

($\Rightarrow$) Let $\mathcal{P}$ be tight; $P_n \in \mathcal{P}$ a sequence, corresponding distribution functions F_n, $n=1,2\ldots$

By Helly's thm, there exists a subsequence $F_{n_k}(x) \to G(x)$ at $x \in C_G$, where $G(x)$ is a generalized distribution function. We now show that $G(-\infty)=0$, $G(\infty)=1$, so G is a distribution function. (which means that $\mathcal{P}$ is relatively compact).

Fix $\varepsilon > 0$, let $I = (a,b]$ be an interval for which $\sup\limits_n P_n(\mathbb{R} \setminus (a,b]) < \varepsilon$ (exists by tightness).

$\sup\limits_n P_n(\mathbb{R} \setminus (a,b]) = 1 - \inf\limits_n P_n(a,b]$, so

$1-\varepsilon < P_n(a,b]$, $n=1,2\ldots$

Let $a' < a < b < b'$, where $a', b' \in C_G$

So $1-\varepsilon < P_{n_k}(a,b] < P_{n_k}(a',b'] = F_{n_k}(b') - F_{n_k}(a')$
$\longrightarrow G(b') - G(a')$

Therefore $G(+\infty) - G(-\infty) \geq 1$.

Since also $G: \mathbb{R} \to [0,1]$,

we have $G(+\infty) = 1$, $G(-\infty) = 0$ $\triangleright$

Remarks

1) By Prokhorov's thm, if $F_{\xi_n}(x) \to F(x)$ $\forall x \in C_F$ for a family of r.v. ξ_n, $n = 1, 2, \dots$ and F is a distribution function then the family $\{\xi_n\}$ is tight.
Without using tightness of $\{\xi_n\}$, it is often easy to establish that this $F(x)$ is a distribution function: indeed by Helly's thm sufficient to check that $F(-\infty) = 0$, $F(+\infty) = 1$.

2) One can show that Prokhorov's thm remains true for measures on $\mathbb{R}^n$, $\mathbb{R}^\infty$, and moreover on any complete separable metric space with a Borel σ-algebra of sets.

Characteristic functions

Def The characteristic function of a r.v. ξ is $\varphi_\xi(t) \equiv \varphi(t) \equiv E e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF_x$,

F_x - distribution function of ξ.

The characteristic function of a random vector $\xi = (\xi_1, \dots, \xi_n)$ is

$$\varphi_\xi(t_1, \dots, t_n) = E e^{i \sum_1^n t_k \xi_k} .$$

Some properties:

1) If ξ - r.v., a, b - constants; $\eta = a\xi + b$ then

$$\varphi_\eta(t) = e^{itb} E e^{iat\xi}$$

2) $|\varphi(t)| \leq \varphi(0) = 1$

3) Let ξ be a r.v. Then $\varphi_\xi(t)$ is uniformly continuous on $\mathbb{R}$

◁ $|\varphi(t+h) - \varphi(t)| = |E e^{it\xi}(e^{ih\xi} - 1)|$

$\leq E|e^{ih\xi} - 1|$,

By dominated convergence, $E|e^{ih\xi} - 1| \to 0$ as $h \to 0$ ▷

4) If $\xi_1, \xi_2 \dots \xi_n$ - independent r.v., $S = \xi_1 + \dots + \xi_n$, then

$$\varphi_S(t) = \prod_{j=1}^{n} \varphi_{\xi_j}(t).$$

Examples

1) For Bernoulli r.v: $\xi = 1$, prob. p;
$\xi = 0$, prob $1-p = q$.

$$\varphi_\xi(t) = pe^{it} + q.$$

2) Let $\xi \sim N(m, \sigma^2)$. Then $\varphi_\xi(t) = e^{itm - \frac{t^2\sigma^2}{2}}$

△ Let $\eta = \frac{\xi - m}{\sigma}$. Then $\eta \sim N(0,1)$, i.e. with density $f = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

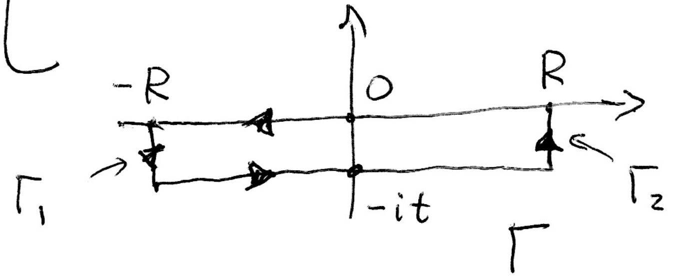
Sufficient to show that $\varphi_\eta(t) = e^{-t^2/2}$.

$$\varphi_\eta(t) = E e^{it\eta} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - x^2/2} dx =$$

$$= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} dx =$$

$$= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty - it}^{\infty - it} e^{-z^2/2} dz$$

[Consider $\int_{\Gamma} = 0$



Show that the $\int_{\Gamma_1} \cdots , \int_{\Gamma_2} \cdots \to 0$ as $R \to \infty$]

$$= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, dx = e^{-t^2/2}$$

3) For Poisson ξ with $P(\xi = k) = \frac{e^{-\lambda} \lambda^k}{k!}$,

$k = 0, 1, \ldots$

$$\varphi_\xi(t) = e^{-\lambda} \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda + \lambda e^{it}} .$$

Thm (moments) Let ξ be r.v. with characteristic function φ and distrib. function F. Then

1) If $E|\xi|^n < \infty$ for some $n \geq 1$ then $\varphi^{(r)}(t)$ exists for any $0 \leq r \leq n$ and

$$\varphi^{(r)}(t) = \int_{-\infty}^{\infty} (ix)^r e^{ixt} dF_x, \quad E\xi^r = i^{-r}\varphi^{(r)}(0),$$

$$(a) \quad \varphi(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!} E\xi^r + \frac{(it)^n}{n!}\varepsilon_n(t), \quad \text{where}$$

$|\varepsilon_n(t)| \leq 3E|\xi|^n$, $\varepsilon_n(t) \to 0$ as $t \to 0$.

2) If $E|\xi|^n < \infty$ for all $n \geq 1$ and

$$(b) \quad \limsup \frac{(E|\xi|^n)^{1/n}}{n} = \frac{1}{e \cdot T} < \infty, \quad T > 0.$$

then $\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} E\xi^n$ converges $\forall |t| < T$.

△ 1) Since $E|\xi^n| < \infty$, we have $E|\xi|^k < \infty$ for all $k \leq n$ (Indeed $E|\xi| \leq (E|\xi|^2)^{1/2} \leq \dots$)

Let $E|\xi| < \infty$. Since $\left|\frac{e^{ihx} - 1}{h}\right| \leq X$, we have by dominated convergence

$$\varphi'(t) = \lim_{h\to 0} \frac{\varphi(t+h)-\varphi(t)}{h} =$$

$$= \lim_{h\to 0} E e^{it\xi} \frac{e^{ih\xi}-1}{h} = i E \xi e^{it\xi} =$$

$$= i \int_{-\infty}^{\infty} x e^{itx} dF_x$$

For general $r \le n$, we apply induction.

We now show (a):

$$e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!}\left(\cos t\xi\theta_1(\omega) + i\sin t\xi\theta_2(\omega)\right)$$

$$\text{So } E e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} E\xi^k + \frac{(it)^n}{n!}\left(E\xi^n + \varepsilon_n(t)\right),$$

$$\varepsilon_n(t) = E\left(\xi^n(\cos t\xi\theta_1 + i\sin t\xi\theta_2 - 1)\right).$$

$\Rightarrow |\varepsilon_n(t)| \le 3E|\xi|^n$ and by dominated convergence $\varepsilon_n(t) \to 0,\ t \to 0$.

2) Let $0 < t_0 < T$. Then

$$\limsup \frac{(E|\xi|^n)^{1/n}}{n} < \frac{1}{e t_0} \Rightarrow \underbrace{\limsup \left(\frac{E|\xi|^n}{n^n} t_0^n e^n\right)^{1/n}}_{L} < 1$$

moreover, $\limsup \left(\frac{E|\xi|^n t_0^n}{n!}\right)^{1/n} < L \le 1$

Therefore the series $\sum_{n=0}^{\infty} \frac{E|\xi|^n}{n!} t_0^n$ converges!

Remark Second part of this thm gives a sufficient condition for the moments $E\xi^n$ to determine $\varphi(t)$ uniquely. Indeed, under condition (6), they already determine $\varphi(t)$ for $-T < t < T$. Take s s.t. $|s| < T/2$. Follow the proof to obtain that

$$\varphi(t) = \sum_{k=0}^{\infty} i^k \frac{(t-s)^k}{k!} \varphi^{(k)}(s), \text{ where}$$

$\varphi^{(k)}(s) = E\xi^k e^{is\xi}$, $-T/2 < s < T/2$, is uniquely determined by $E\xi^n$, $n \geq 1$.

Therefore the moments uniquely determine $\varphi(t)$ for $|t| < \frac{3}{2}T$. Carry on to determine $\varphi(t)$ $\forall t$.

- A stronger sufficient condition for determinacy of the moment problem (i.e. unique determination of φ) is that $\sum_{n=0}^{\infty} \frac{1}{(E\xi^{2n})^{1/2n}} = \infty$ (Carleman's test), no proof.
- If $E\xi^n$ grow too fast, there may be multiple $\varphi(t)$ with these moments (indeterminant moment problem).

<u>Thm</u> (Inversion). Let ξ be r.v. with F, φ.

1) If $a, b \in C_F$, $a<b$ then
$$F(b)-F(a) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,dt$$

2) If $\int_{-\infty}^{\infty} |\varphi(t)|\,dt < \infty$ then F is a.c. with density $f(x)$ and
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt.$$

$\triangle$ 1) If F is a.c., $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$, if $\varphi(t)$ is integrable then by inversion formula for the Fourier transform, if f is sufficiently smooth,
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt.$$

Integrating this and using Fubini,
$$F(b)-F(a) = \int_a^b f(x)\,dx =$$
$$= \frac{1}{2\pi}\int_a^b dx\left(\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)\frac{e^{-ita}-e^{-itb}}{it}\,dt.$$

Now set

$$\Phi_T = \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi\, dt =$$

$$= \int_{-\infty}^{\infty} dF \underbrace{\left(\frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it} e^{itx} dt\right)}_{= \Psi_T(x)}$$

(Fubini works here since

$$\left|\frac{e^{-ita}-e^{-itb}}{it} e^{itx}\right| = \left|\int_a^b e^{-itx} dx\right| \le b-a$$

and so $\int_{-T}^{T}\int_{-\infty}^{\infty}(b-a)\,dF \le 2T(b-a) < \infty$)

Rewrite Ψ_T as follows:

$$\Psi_T(x) = \frac{1}{2\pi}\int_{-T}^{T} \frac{\sin t(x-a) - \sin t(x-b)}{t} dt$$

$$= \frac{1}{2\pi}\left(\int_{-T(x-a)}^{T(x-a)} - \int_{-T(x-b)}^{T(x-b)}\right)\frac{\sin u}{u} du$$

The function $g(s,t) = \int_s^t \frac{\sin u}{u} du$ is continuous in s,t and $g(s,t) \to \pi$ as $\begin{smallmatrix} s \to -\infty \\ t \to +\infty \end{smallmatrix}$

(verify!). Therefore

$\exists C > 0$ s.t. $|\Psi_T(x)| \le C \quad \forall x, T$.

Morever, $\psi_T(x) \xrightarrow[T\to\infty]{} \psi(x) = \begin{cases} 0, & x \notin [a,b] \\ 1/2, & x=a \text{ or } x=b \\ 1, & x \in (a,b) \end{cases}$.

By dominated convergence,

$$\Phi_T = \int_{-\infty}^{\infty} \psi_T \, dF \xrightarrow[T\to\infty]{} \int_{-\infty}^{\infty} \psi(x)\, dF =$$

$$= F(b-) - F(a) + \tfrac{1}{2}\left(F(a) - F(a-) + F(b) - F(b-)\right)$$

$$= F(b) - F(a)$$ since a, b by assumption are in C_F. Thus 1) is proved.

2) Since by assumption, $\int_{-\infty}^{\infty} |\varphi| dt < \infty$, it follows that the following function exists and, by dominated convergence, is continuous (and differentiable); and therefore integrable on $[a,b]$:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt.$$

By Fubini, $\int_a^b f\,dx = \int_{-\infty}^{\infty} \frac{1}{2\pi} \varphi(t) \frac{e^{-ita} - e^{-itb}}{it} dt =$

$= F(b) - F(a)$, where $F(x) = \int_{-\infty}^{x} f(y)\,dy \ \forall x$

So F is a.c. Recalling ~~ar~~guments above, we conclude the proof $\blacksquare$

<u>Corollary</u> Probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and charakteristic functions are in 1-to-1 correspondence

◁ By Thm, part 1, since points of continuity of F are dense ~~ed~~ and therefore we can use continuity of measure to define $F(b) - F(a)$ for any a, b, uniquely ▷

Some useful theorems (without giving proofs)

<u>Thm</u> (Bochner) Let $\varphi(t)$ be continuous on $\mathbb{R}$ and $\varphi(0) = 1$. Then $\varphi(t)$ is a characteristic function <u>iff</u> $\varphi(t)$ is positive semi-definite, i.e. $\sum_{j,k=1}^{n} \varphi(t_j - t_k) \lambda_j \overline{\lambda_k} \geq 0$, $n = 1, 2 \ldots$ $\forall\, t_j \in \mathbb{R}$, $\lambda_j \in \mathbb{C}$.

<u>Thm</u> (Polya). Let $\varphi(t) \geq 0$, continuous, even, $\varphi(0) = 1$, $\varphi(\infty) = 0$, and convex on $0 \leq t < \infty$. Then $\varphi(t)$ is a characteristic function. (e.g. $e^{-|t|}$)

Thm (Marcinkiewicz)

If a characteristic function is of the form $e^{p(t)}$, where $p(t)$ is a polynomial, then $p(t)$ is of degree at most 2.

(e.g. e^{-t^4} is not a characteristic function)

Def If there exists expansion

$$\log \varphi_\xi(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} s_k + o(|t|^n), \quad t \to 0$$

then the coefficients s_k are called cumulants of ξ

Exercise: show that ~~w~~ $E\xi = s_1$, $V\xi = s_2$.

Remark: 1) If $\xi \sim N(m, \sigma^2)$ then $s_1 = m$, $s_2 = \sigma^2$, $s_k = 0$, $k \geq 3$

2) In general, by Marcinkiewicz thm, if for a r.v. ξ $\exists\, n$ s.t. $s_k = 0$, $\forall k \geq n$ then also $s_k = 0$ $\forall k \geq 3$ and $\xi \sim N(s_1, s_2)$

Thm Let $\varphi(t)$ be a characteristic function of ξ and $|\varphi(t_0)| = 1$ for some $t_0 \neq 0$

Then ξ is P.P.

Our assumption means that

$\varphi(t_0) = e^{iat_0}$ for some $a \in \mathbb{R}$.

So $e^{iat_0} = \int e^{it_0 x} dF$

$\Rightarrow \quad 1 = \int \cos t_0(x-a)\, dF$

$\Rightarrow \quad \int (1-\cos t_0(x-a))\, dF = 0$

But $1-\cos \geq 0$ so $\cos t_0(x-a) = 1$ P-a.s. (corollary to Markov ineq), i.e. the measure is concentrated at points $a + \frac{2\pi}{t_0} n$, $n \in \mathbb{Z}$

CLT via characteristic functions

Method Based on

Thm (continuity thm) Let φ_n be a sequence of characteristic functions of distribution functions F_n, $n = 1, \ldots$ $\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n$

1) If $F_n \xrightarrow[\text{(weakly)}]{w} F$, F - distr. function,

then $\varphi_n(t) \to \varphi(t)$ $\forall t \in \mathbb{R}$ where φ is the charact. function of F.

2) If $\lim_n \varphi_n(t)$ exists $\forall t \in \mathbb{R}$ and $\varphi(t) = \lim_n \varphi_n(t)$ is continuous at $t=0$, then $\varphi(t)$ is a characteristic function of some distr. function F and $F_n \xrightarrow{w} F$.

3) If φ_n corresponds to F_n and φ is a characteristic function corresponding to distr. function F then

$$\varphi_n(t) \to \varphi(t) \ \forall t \in \mathbb{R} \iff F_n \xrightarrow{w} F.$$

__Proof__ 1) is obvious by definition of weak convergence applied to $\mathrm{Re}\, e^{it\xi}$, $\mathrm{Im}\, e^{it\xi}$.

2), 3) proof optional – see, e.g. Shiryaev "Probability" ◭

__Thm__ (CLT for i.i.d.) Let $\xi_1, \xi_2 \dots$ – i.i.d. with $E\xi_1^2 < \infty$ and nondegenerate, i.e. $V\xi_1 \neq 0$.

Let $S_n = \xi_1 + \dots + \xi_n$.

Then
$$P\left\{ \frac{S_n - ES_n}{\sqrt{VS_n}} \le x \right\} \xrightarrow[n\to\infty]{} \Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$$
$\forall x \in \mathbb{R}$.

<u>Remark</u> The result can be written also as $\frac{S_n - ES_n}{\sqrt{VS_n}} \xrightarrow{d} \mathcal{N}(0,1)$.

Set $m = E\xi_1$, $\sigma^2 = V\xi_1$, $\varphi(t) = E e^{it(\xi_1 - m)}$,

$$\varphi_n(t) = E e^{\left(it \frac{S_n - ES_n}{\sqrt{VS_n}}\right)}$$

Then, by independence, $\varphi_n(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$.

Since $E\xi_1^2 < \infty$, we have by properties of characteristic functions (Theorem above)

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad t \to 0.$$

So $\varphi_n(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \to e^{-t^2/2} \quad \forall t \in \mathbb{R}$.

This is characteristic function of $N(0,1)$ and so the result follows by continuity thm

Thm (CLT for independent r.v.)

Let $\xi_1, \xi_2 \ldots$ be independent r.v. ~~with~~ with $E\xi_j^2 < \infty$ and distribution fuction F_j $\forall j$.

Let $m_j = E\xi_j$, $\sigma_j^2 = V\xi_j > 0$;

$S_n = \xi_1 + \cdots + \xi_n$, $D_n^2 = \sum_{j=1}^{n} \sigma_j^2$.

Suppose the Lindeberg condition holds, namely,

$$\forall \varepsilon > 0, \quad \frac{1}{D_n^2} \sum_{k=1}^{n} \int_{\{x : |x - m_k| \geq \varepsilon D_n\}} (x - m_k)^2 \, dF_k \to 0, \quad n \to \infty.$$

Then $$\frac{S_n - ES_n}{\sqrt{VS_n}} \xrightarrow{d} N(0,1).$$

We omit the proof.

Some special cases of Lindeberg condition:

1) Lyapunov condition:

$$\exists \delta > 0 \text{ s.t. } \frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} E\left(|\xi_k - m_k|^{2+\delta}\right) \to 0, \quad n \to \infty$$

We now show that it implies Lindeberg condition:

Indeed, let $\varepsilon > 0$. Then

$$E|\xi_k - m_k|^{2+\delta} = \int |x-m_k|^{2+\delta} dF_k \geq$$

$$\geq (\varepsilon D_n)^\delta \int_{\{x : |x-m_k| \geq \varepsilon D_n\}} (x-m_k)^2 dF_k$$

and therefore

$$\frac{1}{D_n^2} \sum_{k=1}^n \int_{\{x:|x-m_k| \geq \varepsilon D_n\}} (x-m_k)^2 dF_k \leq \frac{1}{\varepsilon^\delta} \frac{1}{D_n^{2+\delta}} \sum_1^n E|\xi_k - m_k|^{2+\delta}$$

$$\longrightarrow 0, \quad n \to \infty.$$

2) $\exists K > 0$ s.t. $|\xi_k| \leq K \ \forall k$ and

$D_n \to \infty, \ n \to \infty$

This also implies Lindeberg condition (exercise).

Question: how quickly does convergence to CLT occurs?

Answer:

Thm (Berry-Esseen inequality).

Let $\xi_1, \xi_2 \dots$ be i.i.d. with $E|\xi_1|^3 < \infty$.

Then $\sup_x \left| P\left(\frac{S_n - ES_n}{\sqrt{VS_n}} \leq x\right) - \Phi(x) \right| \leq$

$$\leq \frac{c\, E|\xi_1 - E\xi_1|^3}{\sigma^3 \sqrt{n}},$$

where C is an absolute constant,

$$\frac{1}{\sqrt{2\pi}} \leq C \leq 1/2 .$$

We omit the proof.

<u>Remark</u> $O(1/\sqrt{n})$ is optimal. Indeed, let $\xi_1, \xi_2, \dots$ be i.i.d. Bernoulli, $P(\xi_k = 1) = P(\xi_k = -1) = 1/2$.

Then by symmetry,

$$2P\left(\sum_{k=1}^{2n} \xi_k < 0\right) + P\left(\sum_{k=1}^{2n} \xi_k = 0\right) = 1 \Rightarrow$$

$$\left| P\left(\sum_{k=1}^{2n} \xi_k < 0\right) - 1/2 \right| = \frac{1}{2} P\left(\sum_{1}^{2n} \xi_k = 0\right) =$$

$$= \frac{1}{2} \left(C_{2n}^{n}\right) \frac{1}{2^{2n}} \sim \frac{1}{2\sqrt{\pi n}} = \frac{1}{\sqrt{2\pi}\sqrt{2n}} .$$

$E|\xi_1|^3 = 1 = \delta_1$. So the thm cannot be improved in terms of $O\left(\frac{1}{\sqrt{n}}\right)$ and $C \geq \frac{1}{\sqrt{2\pi}}$.

<u>Question</u> What can happen if $E\xi^2 = \infty$?

Let $\xi_1, \ldots$ i.i.d with Cauchy distribution (density $f = \frac{\theta}{\pi(x^2+\theta^2)}$, $\theta > 0$). Then

$$\varphi_{\xi_1}(t) = \frac{\theta}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{x^2+\theta^2}\,dx = e^{-t\theta}, \quad t > 0$$

similarly, $t < 0$

So $\varphi_{\xi_1}(t) = e^{-\theta|t|}$, $t \in \mathbb{R}$.

$$\Rightarrow \varphi_{S_n/n}(t) = \left(e^{-\theta\frac{|t|}{n}}\right)^n = e^{-\theta|t|}$$

and thus S_n/n also has the Cauchy distribution!

IV Borel-Cantelli, Strong LLN

Let A_n be events. Recall

$$\{A_n \text{ i.o.}\} = \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k ;$$

$$\{A_n \text{ ev.}\} = \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k .$$

Then $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ ev.}\}$

<u>Thm 1</u> (Borel-Cantelli lemma)

1) If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then $P(A_n \text{ i.o.}) = 0$

2) If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and A_n are mutually independent then $P(A_n \text{ i.o.}) = 1$.

$\triangle$ 1) $P(A_n \text{ i.o.}) = \lim_{n\to\infty} P(\bigcup_{k \geq n} A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0$.

by continuity of measure

2) Consider $\{A_n \text{ i.o.}\}^c = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k^c$

$$1 - P(A_n \text{ i.o.}) = \lim_{n\to\infty} P(\bigcap_{k \geq n} A_k^c) ;$$

$P(\bigcap_{k \geq n} A_k^c) = \prod_{k \geq n} P(A_k^c)$ (since true for any finite number of intersections)

Note that $\log(1-x) \leq -x$, $x \in [0,1)$ (since $1-x \leq e^{-x}$), and thus

$$\log P\Big(\bigcap_{k\geq n} A_k^c\Big) = \log \prod_{k \geq n} (1 - P(A_k)) \leq$$

$$\leq -\sum_{k\geq n} P(A_k) = -\infty,$$

i.e. $P\big(\bigcap_{k\geq n} A_k^c\big) = 0 \quad \forall n$ ⊳

Applications:

<u>Lemma 2</u> If $\xi_n \xrightarrow{P} \xi$ then $\exists$ subsequence $\xi_{n_k} \xrightarrow{a.s.} \xi$.

◁ Since $\lim\limits_{n\to\infty} P(|\xi_n - \xi| > 1/k) = 0 \quad \forall k \geq 1$,

we can choose a subsequence s.t.

$$P(|\xi_{n_k} - \xi| > 1/k) \leq 2^{-k} \quad \forall k \geq 1.$$

Since $\sum\limits_{k=1}^{\infty} 2^{-k}$ converge, Borel-Cantelli (1) implies $P(|\xi_{n_k} - \xi| \leq 1/k \text{ ev.}) = 1$

i.e. $\xi_{n_k} \xrightarrow{a.s.} \xi$ ⊳

<u>Lemma 3</u> If $\xi_1 \geq \xi_2 \geq \dots \geq 0$, ξ_n - r.v. and $\xi_n \xrightarrow{P} 0$ then $\xi_n \xrightarrow{a.s.} 0$.

Proof - exercise: consider $\{\limsup |\xi_n| > \varepsilon\}$

Let $\xi_1, \xi_2 \ldots$ – r.v. Denote $S_n = \sum_{j=1}^n \xi_j$

Def The sequence ξ_n is said to satisfy the weak (respectively, strong) LLN, law of large numbers, if ξ_n are integrable $\forall n$ and

$$\frac{S_n - ES_n}{n} \xrightarrow{P} 0 \quad \left(\text{respectively, } \xrightarrow{a.s.} 0\right)$$

Above, we proved weak LLN, in particular, for i.i.d. ξ_n with $\operatorname{Var} \xi_1 < \infty$.

Start with strong LLN under excessive moment assumption:

Lemma 4 (Cantelli's strong LLN)

Let $\xi_1, \xi_2 \ldots$ be i.i.d. s.t. $E\xi_1^4 < \infty$.

Then ~~ξ_1, ξ_2~~ $\{\xi_n\}$ satisfies strong LLN.

◁ Without loss of generality assume $E\xi_1 = 0$.

Let $\varepsilon > 0$, $A_n = \{ |S_n/n| > \varepsilon \}$.

We will show that $P(\limsup |\frac{S_n}{n}| > \varepsilon) =$

$= P(A_n \text{ i.o.}) = 0$. By Borel-Cantelli,

sufficient to show that $\sum_{n=1}^{\infty} P(A_n) < \infty$.

But, by Chebyshev inequality,

$$P(A_n) < \frac{1}{\varepsilon^4} E\left(\left|\frac{S_n}{n}\right|^4\right),$$ so sufficient

to show that $\sum_1^{\infty} E\left(\left|\frac{S_n}{n}\right|^4\right) < \infty$.

$$S_n^4 = (\xi_1 + \cdots + \xi_n)^4 = \sum_1^n \xi_j^4 + \sum_{i<j} \frac{4!}{2!\,2!} \xi_i^2 \xi_j^2$$

$+$ terms of the form $\underset{i \neq j \neq k}{\xi_i^2 \xi_j \xi_k}$, $\underset{i \neq j}{\xi_i^3 \xi_j}$,

$\underset{i \neq j \neq k \neq m}{\xi_i \xi_j \xi_k \xi_m}$.

Since $E\xi_j = 0$,

$$E S_n^4 = n E\xi_1^4 + \frac{n(n-1)}{2} \frac{4!}{2!^2} \left(E\xi_1^2\right)^2$$

$$\leq n E\xi_1^4 + \frac{n(n-1)}{2} 6 E\xi_1^4 \leq n^2 \cdot \text{Const}$$

since $(E\xi^2)^{1/2} \leq (E\xi^4)^{1/4}$

$$\Rightarrow \sum_1^{\infty} E\frac{S_n^4}{n^4} \leq \text{Const} \sum_1^{\infty} \frac{1}{n^2} < \infty$$

Thm 5 (Kolmogorov's inequality)

Let $\xi_1, \xi_2 \dots$ independent r.v. with finite variances. Then $\forall n \geq 1, x > 0$:

$$P\left(\max_{1 \leq k \leq n} | S_k - ES_k | \geq x \right) \leq \frac{\mathrm{Var}\, S_n}{x^2}.$$

◁ Assume without loss $E\xi_j = 0$.

Let $A = \{ \max_{1 \leq k \leq n} | S_k | \geq x \}$,

$A_k = \{ |S_j| < x,\ j = 1, \dots, k-1,\ |S_k| \geq x \}$,

$k = 1, \dots, n$.

Then A_k's are mutually disjoint, $A = \bigcup_1^n A_k$,

$$ES_n^2 \geq ES_n^2 \chi_A = \sum_k E S_n^2 \chi_{A_k}$$

But $ES_n^2 \chi_{A_k} = E(S_k + \xi_{k+1} + \dots + \xi_n)^2 \chi_{A_k}$

$$= ES_k^2 \chi_{A_k} + 2E\, S_k (\xi_{k+1} + \dots + \xi_n) \chi_{A_k} + \qquad (*)$$

$$+ E(\xi_{k+1} + \dots + \xi_n)^2 \chi_{A_k}.$$

By independence,

$$E\, S_k (\xi_{k+1} + \dots + \xi_n) \chi_{A_k} =$$

$$= E S_k \chi_{A_k} \cdot E(\xi_{k+1} + \cdots + \xi_n) = 0,$$

as $E\xi_j = 0$.

So (*) gives:

$$E S_n^2 \chi_{A_k} \geq E S_k^2 \chi_{A_k}$$

Therefore $E S_n^2 \geq \sum_k E S_n^2 \chi_{A_k} \geq \sum_k E S_k^2 \chi_{A_k}$

$$\geq x^2 \sum_k P(A_k) = x^2 P(A) \quad \triangleright$$

<u>Thm</u> 6 (2 series thm) Let $\xi_1, \xi_2 \ldots$ independent r.v.

If $\sum_{n=1}^{\infty} E\xi_n$ and $\sum_{n=1}^{\infty} V\xi_n$ converge

then $\sum_{n=1}^{\infty} \xi_n$ converge a.s.

$\triangleleft$ First assume $E\xi_j = 0$, $j = 1, 2 \ldots$

We now show that $\{S_k\}$ is a fundamental (Cauchy) sequence with prob. 1, i.e.

$\lim_{k\to\infty} \sup_{m,n\geq k} |S_m - S_n| = 0$ a.s. (by Cauchy property this implies our thm). By Lemma 3, it is sufficient to show that

$$\sup_{m,n\geq k} |S_n - S_m| \xrightarrow{P} 0, \quad k\to\infty,$$

as the sequence in question is decreasing.

Note that

$$0 \leq \sup_{m,n\geq k} |S_n - S_m| \leq \sup_{m,n\geq k} \left(|S_n - S_k| + |S_k - S_m|\right)$$

$$= 2\sup_{n\geq k} |S_n - S_k| \equiv 2\delta_k,$$

so we need to show that $\delta_k \xrightarrow{P} 0$.

By Kolmogorov inequality,

$$P(\delta_k \geq x) = \lim_{m\to\infty} P\left(\max_{m\geq n\geq k} |S_n - S_k| \geq x\right)$$

$$\leq \frac{1}{x^2}\lim_{m\to\infty}\sum_{n=k+1}^{m} V\xi_n = \frac{1}{x^2}\sum_{n=k+1}^{\infty} V\xi_n \longrightarrow 0$$

as $k\to\infty$ since $\sum_1^\infty V\xi_k$ converge!

Since holds $\forall x>0$, we have $\delta_k \xrightarrow{P} 0$

If expectations $E\xi_j \neq 0$, consider

$$\sum_{n=1}^{\infty}\xi_n = \sum_{n=1}^{\infty}(\xi_n - E\xi_n) + \underbrace{\sum_{n=1}^{\infty}E\xi_n}_{\text{converge}}$$

For the next thm, we need the following 2 lemmas:

<u>Lemma 7</u> (Toeplitz) Let $a_n \geq 0$, $n=1,2\dots$, $b_n = \sum_{k=1}^{n} a_k$ and $b_n \to \infty$ as $n \to \infty$

and let $x_n \to x$.

Then $\frac{1}{b_n} \sum_{j=1}^{n} a_j \cdot x_j \to x$, $n \to \infty$.

(In particular for $a_j \equiv 1$, $\frac{x_1 + \dots + x_n}{n} \to x$)

◁ Fix $\varepsilon > 0$. Let $n_0(\varepsilon)$ be s.t. $|x_n - x| \leq \varepsilon/2$ $\forall n \geq n_0$.

Choose $n_1 > n_0$ s.t. $\frac{1}{b_{n_1}} \sum_{j=1}^{n_0} |x_j - x| < \varepsilon/2$

Then for $n > n_1$:

$$\left| \frac{1}{b_n} \sum_1^n a_j x_j - x \right| = \left| \frac{1}{b_n} \sum_1^n a_j (x_j - x) \right| \leq$$

$$\leq \frac{1}{b_n} \sum_1^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{n_0+1}^{n} a_j |x_j - x|$$

$$\leq \frac{1}{b_{n_1}} \sum_1^{n_0} a_j |x_j - x| + \frac{1}{b_n} \sum_{n_0+1}^{n} a_j \cdot \varepsilon/2$$

$$\leq \varepsilon/2 + \frac{1}{b_n} (b_n - b_{n_0}) \varepsilon/2 \leq \varepsilon$$ ▷

<u>Lemma 8</u> (Kronecker) Let a_n, b_n, $n=1,2\ldots$ be as above and x_n, $n=1,2\ldots$ be s.t. $\sum_1^\infty x_n$ converge.

Then $\frac{1}{b_n}\sum_1^n b_j x_j \to 0$, $n\to\infty$.

◁ Let $b_0=S_0=0$, $S_n=\sum_1^n x_j$. Then

$$\sum_1^n b_j x_j = \sum_1^n b_j(S_j - S_{j-1}) =$$
$$= b_n S_n - b_0 S_0 - \sum_1^n S_{j-1}(b_j - b_{j-1}).$$

So $\frac{1}{b_n}\sum_1^n b_j x_j = S_n - \frac{1}{b_n}\sum_1^n S_{j-1} a_j \to 0$

because by Toeplitz lemma $\frac{1}{b_n}\sum_1^n S_{j-1} a_j \to$
$$\to \lim_{n\to\infty} S_n$$ ▷

<u>Thm 9</u> Let $\xi_1, \xi_2 \ldots$ independent r.v. with finite variances and let $b_n \geq 0$, $n=1,2\ldots$ s.t. $b_n \to \infty$.

If $\sum_1^\infty \frac{V\xi_n}{b_n^2} < \infty$ then $\frac{S_n - ES_n}{b_n} \xrightarrow{a.s.} 0$.

(in case $b_n = n$, this is a strong LLN)

$\triangleleft$ We have

$$\frac{S_n - ES_n}{b_n} = \frac{1}{b_n} \sum_1^n b_k \frac{\xi_k - E\xi_k}{b_k} \qquad (**)$$

Now $V \sum_1^n \frac{\xi_k - E\xi_k}{b_k} = \sum_1^n \frac{V\xi_k}{b_k^2}$ converge by assumption. By 2-series thm,

$\sum_1^n \frac{\xi_k - E\xi_k}{b_k}$ converge a.s. as $n \to \infty$,

and thm follows from $(**)$ by Kronecker's lemma. $\triangleright$

<u>Example</u> Let $\xi_1, \xi_2 \ldots$ Bernoulli,

$P(\xi_n = 1) = P(\xi_n = -1) = 1/2$.

Since $\cancel{\sum_1^\infty \frac{1}{n^2 \log}}$ $\sum_1^\infty \frac{1}{n \log^2 n}$ converge,

We have $\frac{S_n}{\sqrt{n} \log n} \to 0$ a.s.

Thm 10 (Kolmogorov's strong LLN)

Let $\xi_1, \xi_2, \dots$ i.i.d. with finite expectations $m = E\xi_1$ (i.e. $E|\xi_1| < \infty$)

Then $\frac{S_n}{n} \xrightarrow{a.s.} m$.

Remark $S_n/n \xrightarrow{a.s.} m$ can be written

as $\sum_{k=1}^{n} (\xi_k - E\xi_k) = o(n)$, $n \to \infty$ a.s.

Lemma 11 Let $\xi \geq 0$ - r.v. Then

$$\sum_1^\infty P(\xi \geq n) \leq E\xi \leq 1 + \sum_1^\infty P(\xi \geq n)$$

$$\sum_1^\infty P(\xi \geq n) = \sum_{n=1}^\infty \sum_{k \geq n} P(k \leq \xi < k+1) =$$

$$= \sum_{k=1}^\infty k\, P(k \leq \xi < k+1) = \sum_0^\infty E(k \chi_{[k,k+1)})$$

$$\leq \sum_0^\infty E(\xi \chi_{[k,k+1)}) = E(\xi) \leq$$

$$\leq \sum_0^\infty E((k+1)\chi_{[k,k+1)}) = 1 + \sum_1^\infty P(\xi \geq n)$$

Proof of Thm 10

Without loss, assume $E\xi_1 = 0$.

We need to show that $S_n/n \to 0$ a.s.

By Lemma 11, $E|\xi_1| < \infty \Rightarrow$

$$\Rightarrow \sum_1^\infty P(|\xi_1| \geq n) = \sum_1^\infty P(|\xi_n| \geq n) < \infty$$

By Borel-Cantelli, it follows that

$P(|\xi_n| \geq n \text{ i.o.}) = 0$ i.e. $\underline{P(|\xi_n| < n \text{ ev.}) = 1}$

Therefore, defining

$$\tilde{\xi}_n = \begin{cases} \xi_n, & |\xi_n| < n \\ 0, & |\xi_n| \geq n \end{cases}$$ we have that

$$\frac{S_n}{n} \to 0 \text{ a.s.} \iff \frac{\tilde{\xi}_1 + \dots + \tilde{\xi}_n}{n} \to 0 \text{ a.s.}$$

Note that $E\tilde{\xi}_n = E\xi_n \chi_{|\xi_n| < n} =$

$$= E\xi_1 \chi_{|\xi_1| < n} \xrightarrow[n \to \infty]{} E\xi_1 = 0.$$

By Toeplitz lemma with $x_n = E\tilde{\xi}_n$,

$$\frac{1}{n} \sum_{k=1}^n E\tilde{\xi}_k \to 0, \quad n \to \infty.$$

Therefore,

$$\frac{S_n}{n} \to 0 \text{ a.s.} \iff \frac{1}{n}\sum_{k=1}^{n} \tilde{\xi}_k - E\tilde{\xi}_k \to 0 \text{ a.s.}$$

By Kronecker's lemma (with $b_n = n$, $x_n = \frac{\tilde{\xi}_n - E\tilde{\xi}_n}{n}$)

this is so if $\sum_1^\infty \frac{\tilde{\xi}_n - E\tilde{\xi}_n}{n}$ converge. But this

is, in turn, so if $\sum_1^\infty \frac{V(\tilde{\xi}_n - E\tilde{\xi}_n)}{n^2}$ converge

(by 2-series thm). We have

$$\sum_1^\infty \frac{V(\tilde{\xi}_n - E\tilde{\xi}_n)}{n^2} \le \sum_1^\infty \frac{E\tilde{\xi}_n^2}{n^2} = \sum_1^\infty \frac{1}{n^2} E(\xi_1^2 \chi_{|\xi_1|<n})$$

$$= \sum_{n=1}^\infty \frac{1}{n^2} \sum_{k=1}^{n} E(\xi_1^2 \chi_{k-1 \le |\xi_1| < k}) =$$

$$= \sum_{k=1}^\infty E(\xi_1^2 \chi_{k-1\le|\xi_1|<k}) \sum_{n=k}^\infty \frac{1}{n^2} \quad \le 2/k$$

$$\le 2\sum_{k=1}^\infty E(|\xi_1| \chi_{k-1\le|\xi_1|<k}) = 2E|\xi_1| < \infty.$$

Thm 10 is proved.

<u>Example</u>: Probability space

$([0,1)=\Omega, \mathcal{B}(\Omega), P$ - Lebesgue measure$)$

Numbers $\omega \in \Omega$ have binary representation

$$\omega = \omega_1/2 + \omega_2/2^2 + \cdots = 0,\omega_1\omega_2\cdots,$$

$$\omega_j = \{0,1\}\ \cancel{\in}$$

Let r.v. $\xi_j(\omega) = \omega_j$. Let $x_j = \{0,1\}$

and consider

$$A_{x_1,\dots x_n} = \{\omega : \xi_1 = x_1, \dots \xi_n = x_n\} =$$

$$= \left\{\omega : \frac{x_1}{2} + \dots + \frac{x_n}{2^n} \le \omega < \frac{x_1}{2} + \dots + \frac{x_n}{2^n} + \frac{1}{2^n}\right\}$$

We have $P(A_{x_1,\dots x_n}) = 1/2^n$.

Therefore $\xi_1, \xi_2 \dots$ i.i.d. $P(\xi_1=0) = P(\xi_1=1) = 1/2$.

Strong LLN gives

$$1/n \sum_{k=1}^{n} \xi_k = 1/n \sum_{k=1}^{n} \chi_{\omega_k = 1} \to E\xi_1 = 1/2 \text{ a.s.}$$

i.e. for a.e. number in $[0,1)$ the proportion of 0's and 1's in its binary expansion tends to $1/2$ (such numbers are called normal).

Law of iterated logarithm

We have seen for Bernoulli with $E\xi_1 = 0$ that not only $\frac{S_n}{n} \to 0$ but in fact

$\frac{S_n}{\sqrt{n}\log n} \to 0$ a.s. However, by CLT,

$$\frac{S_n}{\sqrt{n}} \overset{d}{\not\to} 0.$$

Def 1) A function $\varphi^*(n)$ is called upper for S_n if $S_n \le \varphi^*(n)$ $\forall n \ge n_0$ with probability 1

2) A function $\varphi_*(n)$ is called lower for S_n if $S_n > \varphi_*(n)$ for infinitely many n with probability 1.

[e.g., for Bernoulli, $\varepsilon\sqrt{n}\log n$ is upper $\forall \varepsilon > 0$]

For some $\varphi(n)$, consider

$$\left\{ \limsup \frac{S_n}{\varphi(n)} \le 1 \right\} = \left\{ \lim_{n\to\infty} \sup_{m \ge n} \frac{S_m}{\varphi(m)} \le 1 \right\}$$

$$= \left\{ \forall \varepsilon > 0 \ \exists n_1 \text{ s.t. } \sup_{m \ge n} \frac{S_m}{\varphi(m)} \le 1 + \varepsilon \ \forall n \ge n_1 \right\}$$

$$= \left\{ \forall \varepsilon > 0 \ \exists n_1 \text{ s.t. } S_m \le (1+\varepsilon)\varphi(m) \ \forall m \ge n_1 \right\}$$

So if $P(\limsup \frac{S_n}{\varphi(n)} \leq 1) = 1$

then $(1+\varepsilon)\varphi(n)$ is upper for S_n $\forall \varepsilon > 0$.

$$\{\limsup \frac{S_n}{\varphi(n)} \geq 1\} = \{\forall \varepsilon > 0 \ \exists n_1 \text{ s.t. } \sup_{m \geq n} \frac{S_m}{\varphi(m)} \geq 1-\varepsilon \ \forall n > n_1\}$$

$$= \{\forall \varepsilon > 0 : S_m \geq (1-\varepsilon)\varphi(m) \text{ for infinitely many } m\}$$

So if $P(\limsup \frac{S_n}{\varphi(n)} \geq 1) = 1$

then $(1-\varepsilon)\varphi(n)$ is lower for S_n $\forall \varepsilon > 0$.

<u>Thm</u> (law of iterated logarithm).

Let $\xi_1, \xi_2 \dots$ i.i.d. with $E\xi_1 = 0$, $E\xi_1^2 = \sigma^2 > 0$.

Then $P(\limsup \frac{S_n}{\psi(n)} = 1) = 1$,

$$\psi(n) = \sqrt{2\sigma^2 n \log\log n},$$

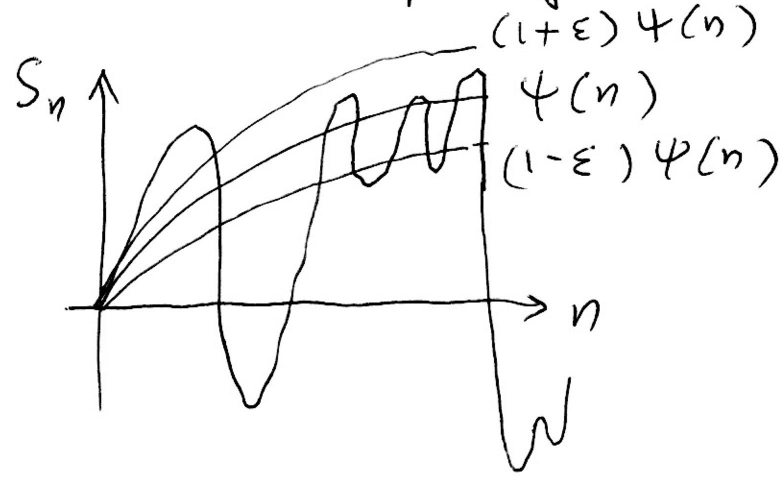
i.e. $\forall \varepsilon > 0$, $(1+\varepsilon)\psi$ is upper, $(1-\varepsilon)\psi$ is lower for S_n.

<u>Remark</u> SLLN: $\sum_1^n \xi_k - E\xi_k = o(n)$ a.s.

Law of iterated log, in particular,

$$\sum_1^n \xi_k - E\xi_k = O(\psi(n)) \text{ a.s.}$$

(we omit the proof)



– typical path.

Zero-one law

Let $\xi_1, \xi_2 \dots$ r.v. on $(\Omega, \mathcal{F}, P)$

Denote $\mathcal{F}_n^{n+k} = \sigma(\xi_n, \dots \xi_{n+k}) =$

$$= \sigma\left(\{\xi_n \le x_n, \dots, \xi_{n+k} \le x_{n+k}\ ;\ x_n, \dots x_{n+k} \in \mathbb{R}\}\right)$$

$$\mathcal{F}_n^{\infty} = \sigma(\xi_n, \dots) = \sigma\left(\bigcup_{k=1}^{\infty} \mathcal{F}_n^{n+k}\right).$$

<u>Def</u> σ-algebra $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}_n^{\infty}$ is called <u>the tail σ-algebra</u>. Events of $\mathcal{T}$ are called tail events.

<u>Examples</u> $\{\xi_n \in B \text{ i.o.}\} =$

$$= \bigcap_{n=1}^{\infty} \bigcup_{k \ge n} \{\xi_k \in B\} \in \mathcal{T},$$

where B ∉ is some $\mathcal{B}(\mathbb{R})$.

$\{ \sum_1^\infty \xi_n \text{ converge} \} \in \mathcal{T}$.

$\{ \xi_n \notin B \ \forall n \}$, $\{ \xi_{10} \in B \}$ may not be in $\mathcal{T}$.

<u>Lemma</u> If $\mathcal{A}_0 \subset \mathcal{F}$, $\mathcal{B}_0 \subset \mathcal{F}$ are independent algebras (i.e. $P(AB) = P(A)P(B)$, $A \in \mathcal{A}_0$, $B \in \mathcal{B}_0$) then $\sigma(\mathcal{A}_0)$, $\sigma(\mathcal{B}_0)$ are also independent.

◁ Let $A \in \mathcal{A}_0$ and consider measures

$$P_A^{(1)}(B) = P(AB), \quad P_A^{(2)}(B) = P(A)P(B).$$

They coincide on $\mathcal{B}_0$, so by Caratheodory thm, they coincide on $\sigma(\mathcal{B}_0)$, i.e.

$$P(AB) = P(A)P(B) \quad \forall A \in \mathcal{A}_0, \ B \in \sigma(\mathcal{B}_0).$$

Now let $B \in \sigma(\mathcal{B}_0)$ and consider

$$Q_B^{(1)}(A) = P(AB), \quad Q_B^{(2)}(A) = P(A)P(B).$$

We similarly conclude that

$$P(AB) = P(A)P(B) \quad \forall A \in \sigma(\mathcal{A}_0), \ B \in \sigma(\mathcal{B}_0).$$

▷

<u>Thm</u> (Kolmogorov 0-1 law)

Let $\xi_1, \xi_2 \cdots$ be <u>independent</u> r.v.

Then for any $A \in \mathcal{T}$ either $P(A) = 0$

or $P(A) = 1$.

⊠ Note that $\mathcal{F}_1^n$ is independent with $\mathcal{F}_{n+1}^{n+k}$ $\forall k$, so $\mathcal{F}_1^n$ is independent with $\bigcup_{k=1}^{\infty} \mathcal{F}_{n+1}^{n+k}$. By lemma, $\mathcal{F}_1^n$ is independent with $\mathcal{F}_{n+1}^{\infty} \supset \mathcal{T}$. So $\bigcup_{n=2}^{\infty} \mathcal{F}_1^n$ is independent with $\mathcal{T}$, and therefore, by lemma, $\mathcal{F}_1^{\infty} \supset \mathcal{T}$ is independent with $\mathcal{T}$. Thus $\mathcal{T}$ is independent with itself! If $A \in \mathcal{T}$ then

$$P(A) = P(A \cap A) \underset{\substack{\uparrow \\ \text{by independence}}}{=} P(A)^2$$

$$\Rightarrow \quad P(A) = 0 \quad \text{or} \quad P(A) = 1.$$

⊠

Examples :

1) By 0-1 law, if $\xi_1, \xi_2 \dots$ are independent r.v. then $\sum_{n=1}^{\infty} \xi_n$ either converge a.s or diverge a.s.

2) Let ξ_n be independent, $A_n = \{ \xi_n \in B_n \}$. Then by 0-1 law $P(A_n \text{ i.o.}) = 0$ or 1.
Note that this also follows from Borel-Cantelli.

V Conditional expectation

Recall:

Radon-Nikodym thm : Let $(\Omega, \mathcal{F})$ - measure space ; μ - a finite measure on $\mathcal{F}$. Let λ be a measure on $\mathcal{F}$ a.c. with respect to μ [i.e. $\lambda(A) = 0$ whenever $\mu(A) = 0$]

Then there exists an $\mathcal{F}$-measurable function f s.t. $\lambda(A) = \int_A f \, d\mu \quad \forall A \in \mathcal{F}$.

This function is determined uniquely, up to sets of measure zero. It is called the derivative of λ w.r.t. μ : $f = \frac{d\lambda}{d\mu}$.

Let $\xi \geq 0$ r.v. on $(\Omega, \mathcal{F}, P)$ and $\mathcal{L} \subset \mathcal{F}$ a σ-algebra. Consider $Q(A) = \int_A \xi \, dP$, $A \in \mathcal{L}$.

Then Q is a.c. measure w.r.t. P (we consider P, Q as measures on $\mathcal{L}$). By Radon-Nikodym thm, there is a unique $\mathcal{L}$-measurable function $\frac{dQ}{dP} \equiv E(\xi \mid \mathcal{L})$. It is called the conditional expectation of ξ w.r.t. $\mathcal{L}$.

In other words,

<u>Def</u> The conditional expectation of a r.v. $\xi \geq 0$ w.r.t. a σ-algebra $\mathcal{L} \subset \mathcal{F}$ is a nonnegative r.v., denoted $E(\xi|\mathcal{L})$ such that

1) $E(\xi|\mathcal{L})$ is $\mathcal{L}$-measurable

2) $\int_A \xi \, dP = \int_A E(\xi|\mathcal{L}) \, dP \quad \forall A \in \mathcal{L}$.

If ξ is a r.v. s.t. $\min\left(E(\xi^+|\mathcal{L}), E(\xi^-|\mathcal{L})\right) < \infty$ a.s.

then $E(\xi|\mathcal{L}) \equiv E(\xi^+|\mathcal{L}) - E(\xi^-|\mathcal{L})$.

<u>Remark</u> $E(\xi|\mathcal{L})$ is defined up to sets of P-measure zero.

<u>Def</u> Let $(\Omega, \mathcal{F}, P)$ - probability space. The conditional probability of an event $B \in \mathcal{F}$ w.r.t. σ-algebra $\mathcal{L} \subset \mathcal{F}$ is

$P(B|\mathcal{L}) \equiv E(\chi_B|\mathcal{L})$.

<u>Remark</u> For fixed B, $P(B|\mathcal{L})$ is an $\mathcal{L}$-measurable r.v.

$$\int_A P(B|\mathcal{L}) \, dP = \int_A \chi_B \, dP = P(A \cap B) \quad \forall A \in \mathcal{L}.$$

Let $\mathcal{L} = \sigma(\{\mathcal{D}_1, \mathcal{D}_2 \dots\})$ for a partition $\Omega = \bigcup_{j=1}^{\infty} \mathcal{D}_j$, $\mathcal{D}_j$ - disjoint, $P(\mathcal{D}_j) > 0$, $j = 1, 2 \dots$

Then all $\mathcal{L}$-measurable functions have the form $f(\omega) = \sum_{j=1}^{\infty} c_j \chi_{\mathcal{D}_j}(\omega)$, in particular, $E(\xi|\mathcal{L})$ has such a form, so it is constant on $\mathcal{D}_j$ (a.e.) $\forall j$. – we call this constant $E(\xi|\mathcal{D}_j)$.

$$\Rightarrow E(\xi \chi_{\mathcal{D}_j}) = \int_{\mathcal{D}_j} \xi \, dP = \int_{\mathcal{D}_j} E(\xi|\mathcal{L}) \, dP = E(\xi|\mathcal{D}_j) P(\mathcal{D}_j),$$

i.e. $E(\xi|\mathcal{D}_j) = \dfrac{E(\xi \chi_{\mathcal{D}_j})}{P(\mathcal{D}_j)}$.

Note: $P(B|\mathcal{D}_j) \equiv E(\chi_B|\mathcal{D}_j) = \dfrac{E(\chi_B \chi_{\mathcal{D}_j})}{P(\mathcal{D}_j)} = \dfrac{P(B \cap \mathcal{D}_j)}{P(\mathcal{D}_j)}$ –

– classical definition of conditional probability.

$$P(B|\mathcal{L}) = E(\chi_B|\mathcal{L}) = \sum_1^{\infty} P(B|\mathcal{D}_j) \chi_{\mathcal{D}_j}(\omega).$$

Note: $P(B|\{\emptyset, \Omega\}) = P(B)$.

<u>Thm</u> Let ξ - r.v. on $(\Omega, \mathcal{F}, P)$,

$\mathcal{L} \subset \mathcal{F}$ is a σ-algebra.

Then, a.s.,

1) $|E(\xi|\mathcal{L})| \leqslant E(|\xi| \,|\, \mathcal{L})$

If $\xi \leqslant \mu$ then $E(\xi|\mathcal{L}) \leqslant E(\mu|\mathcal{L})$

2) $E(a\xi + b\eta|\mathcal{L}) = aE(\xi|\mathcal{L}) + bE(\eta|\mathcal{L})$,

$a, b \in \mathbb{R}$, η - r.v.

3) $E(\xi|\{\phi, \Omega\}) = E\xi$

4) $E(\xi|\mathcal{F}) = \xi$

5) If $\mathcal{L}_1 \subset \mathcal{L}_2$ then $E(E(\xi|\mathcal{L}_2)|\mathcal{L}_1) = E(\xi|\mathcal{L}_1)$

If $\mathcal{L}_1 \supset \mathcal{L}_2$ then $E(E(\xi|\mathcal{L}_2)|\mathcal{L}_1) = E(\xi|\mathcal{L}_2)$

6) $E(E(\xi|\mathcal{L})) = E\xi$

7) If ξ is independent of $\mathcal{L}$ (i.e. independent of χ_B, $B \in \mathcal{L}$),

then $E(\xi|\mathcal{L}) = E\xi$.

1-4 exercise using definition, uniqueness.

5) Let $\mathcal{X}_1 \subset \mathcal{X}_2$, $A \in \mathcal{X}_1 \subset \mathcal{X}_2$

$$\int_A E(\xi|\mathcal{X}_1)\,dP = \int_A \xi\,dP;$$

$$\int_A E(E(\xi|\mathcal{X}_2)|\mathcal{X}_1)\,dP = \int_A E(\xi|\mathcal{X}_2)\,dP = \int_A \xi\,dP$$

The statement follows since $A \in \mathcal{X}_1$ is arbitrary.

Let $\mathcal{X}_1 \supset \mathcal{X}_2$, $A \in \mathcal{X}_1$. $E(\xi|\mathcal{X}_2)$ is $\mathcal{X}_2$-measurable, and since $\mathcal{X}_2 \subset \mathcal{X}_1$, also $\mathcal{X}_1$-measurable.

Hence $\int_A E(\xi|\mathcal{X}_2)\,dP = \int_A E(E(\xi|\mathcal{X}_2)|\mathcal{X}_1)\,dP \quad \forall A \in \mathcal{X}_1$

and the result follows.

6) Take $\mathcal{X}_1 = \{\emptyset, \Omega\}$, $\mathcal{X}_2 = \mathcal{X}$

and apply 5).

7) We have

$$\int_B \xi\,dP = E(\xi\chi_B) = E(\xi)E(\chi_B) = E(\xi)\int_B dP$$

Thm (continued)

8) c-dominated convergence.

Let $\xi_1, \xi_2 \ldots$ be r.v., $\xi_n \to \xi$ a.s.

and $|\xi_n| \leq \eta$, $E\eta < \infty$.

Then $E(\xi_n | \mathcal{L}) \to E(\xi | \mathcal{L})$ a.s.

9) Let η be an $\mathcal{L}$-measurable r.v,

$E|\xi| < \infty$, $E|\xi\eta| < \infty$.

Then $E(\xi \cdot \eta | \mathcal{L}) = \eta\, E(\xi | \mathcal{L})$.

△ 8) Let $\mu_n = \sup_{m \geq n} |\xi_m - \xi|$. Then $\mu_n \to 0$ a.s.

$|E(\xi_n | \mathcal{L}) - E(\xi | \mathcal{L})| = |E(\xi_n - \xi | \mathcal{L})| \leq$

$\leq E(|\xi_n - \xi| \,|\, \mathcal{L}) \leq E(\mu_n | \mathcal{L})$

Since $0 \leq E(\mu_{n+1} | \mathcal{L}) \leq E(\mu_n | \mathcal{L})$, $n = 1, 2, \ldots$

we have $h = \lim_{n \to \infty} E(\mu_n | \mathcal{L})$ exists a.s.

$$0 \leq \int_\Omega h\, dP \leq \int_\Omega E(\mu_n | \mathcal{L})\, dP = \int_\Omega \mu_n\, dP \longrightarrow 0$$

by dominated convergence

since $|\mu_n| \leq 2\eta$

i.e. $\int_\Omega h\,dP = 0$, $h \geq 0$

$\Rightarrow h = 0$ a.e. 8 is proven.

9) Both $E(* \mid \mathcal{X})$, η are $\mathcal{X}$-measurable, so it remains to show

$$E(\chi_B E(\xi\eta \mid \mathcal{X})) = E(\chi_B \eta E(\xi \mid \mathcal{X})), \quad B \in \mathcal{X}$$

First, we show this for $\eta = \chi_A$, $A \in \mathcal{X}$.

For the l.h.s:

$$E(\chi_B E(\xi\eta \mid \mathcal{X})) = E(\chi_B \xi\eta) = E(\chi_{A\cap B}\xi)$$

For the r.h.s:

$$E(\chi_B \eta E(\xi \mid \mathcal{X})) = E(\chi_{A\cap B} E(\xi \mid \mathcal{X})) =$$

$$\underset{\substack{\uparrow \\ A\cap B \in \mathcal{X}}}{=} E(\chi_{A\cap B}\xi).$$

Then by linearity, the statement holds for any simple r.v. η. But any η is approximated by simple η_n s.t. $|\eta_n| \leq |\eta|$.

Then $\eta_n \xi \to \eta\xi$ a.s., $|\eta_n \xi| \leq |\eta\xi|$.

Since $E|\eta\xi| < \infty$, by c-dominated convergence,

$$\eta_n E(\xi|\mathcal{X}) = E(\eta_n \xi|\mathcal{X}) \to E(\xi\eta|\mathcal{X})$$

$$\downarrow$$

$$\eta E(\xi|\mathcal{X})$$

$\Rightarrow$

Def The conditional expectation of a r.v. ξ w.r.t. a r.v. η is defined as follows

$E(\xi|\eta) \equiv E(\xi|\sigma(\eta))$, where $\sigma(\eta)$ is the σ-algebra generated by η.

Thm Let μ, η be r.v. and μ be $\sigma(\eta)$-measurable. The there exists a Borel-measurable function $f: \mathbb{R} \to \mathbb{R}$ s.t. $\mu = f(\eta)$.
In particular, there exists Borel $g(x)$ s.t. $E(\xi|\eta) = g(\eta)$.

$\triangle$ Assume first that μ be simple:

$\mu = \sum_1^n c_j \chi_{A_j}$, $\Omega = \bigcup_1^n A_j$, A_j's - disjoint.

Since μ is $\sigma(\eta)$ measurable, $A_j \in \sigma(\eta)$

i.e. $\exists B_j \in \mathcal{B}(\mathbb{R})$ s.t. $A_j = \eta^{-1}(B_j)$.

Since A_j's are disjoint, B_j's are also disjoint;

$\bigcup_1^n B_j = \eta(\Omega)$.

Set $f(x) = \sum_1^n c_j \chi_{B_j}(x)$, $x \in \bigcup_1^n B_j$, zero otherwise.

We have $f(\eta(\omega)) = \mu(\omega)$.

For general μ, take simple μ_n, $\mu_n(\omega) \to \mu(\omega)$ $\forall \omega \in \Omega$. Then $\mu_n = f_n(\eta)$, f_n - Borel-measurable. Set $f(x) = \lim_n f_n(x)$ is exists (it exists on $\eta(\Omega)$) and $f(x) = 0$ otherwise. Then $f(x)$ is Borel-measurable and

$\mu(\omega) = \lim_n \mu_n(\omega) = \lim_n f_n(\eta(\omega)) = f(\eta(\omega))$ ⊳

Remark (without proof). If $E\xi^2 < \infty$ then

$\min_f E(\xi - f(\eta))^2 = E(\xi - E(\xi|\eta))^2$, where

min is over all $\sigma(\eta)$-measurable f s.t.

$Ef^2(\eta) < \infty$.