Lectures on Stochastic Stability

Sergey FOSS

Heriot-Watt University

This mini-course presents an overview of stochastic stability methods, mostly motivated by (but not limited to) stochastic network applications. We work with stochastic recursive sequences, and, in particular, Markov chains, in a general Polish state space. We discuss and compare methods based on (i) Lyapunov functions, and fluid limits, (ii) explicit coupling (renovating events and Harris chains), (iii) monotonicity, and some others. We also discuss instability methods and perfect simulation methods.

Lectures are based on handouts of my lecture notes (Colorado State Uni, 1996; Novosibirsk State Uni, 1997–2000; Kazakh National University, 2007), on the joint overview paper with Takis Konstantopoulos (2004), on notes written by us for a Short LMS/EPSRC Course for PhD students (September 2006), and on some (more-or-less) recent publications.

Table of Topics

1 Lecture 1. Basic Tools.

1.1 Notation, Acronyms, and Basic Concepts

R.v. — random variable

i.i.d. — independent identically distributed

$X, Y, Z, \xi, \eta, \psi, \ldots$ — for r.v.s

F, G — distribution functions, f — density function

$\mathbf{P}$ — probability and probability measure, $\mathbf{E}$ — expectation, $\mathbf{D}$ — variance

$\xi \Subset F$ means $\mathbf{P}(\xi \le x) = F(x)$ for all x

$\xi \Subset \mathbf{P}$ means $\mathbf{P}(\xi \in B) = \mathbf{P}(B)$, $B \in \mathcal{B}$.

$\mathbf{I}(A)$, or $\mathbf{1}(A)$ the *indicator* function of event A, $\mathbf{I}(A) = 1$ if A occurs, and .$\mathbf{I}(A) = 0$, otherwise.

Here are standart families of distributions:

$U[a, b]$	$G(p)$
$E(\alpha)$	$B(m, p)$
$N(a, \sigma^2)$	$\Pi(\lambda)$

Convergence:

$\xi_n \xrightarrow{\text{a.s.}} \xi$ means $\mathbf{P}(\lim \xi_n = \xi) = 1$, or $\forall \varepsilon > 0,\ \mathbf{P}(\sup_{m \ge n} |\xi_m - \xi| > \varepsilon) \to 0$ as $n \to \infty$.

$\xi_n \xrightarrow{\text{p}} \xi$ means $\forall \varepsilon > 0,\quad \mathbf{P}(|\xi_n - \xi| > \varepsilon) \to 0$ as $n \to \infty$.

The same – for random vectors.

Key Properties of Convergence. Let $\to$ mean either $\stackrel{\text{a.s.}}{\to}$ or $\stackrel{\text{p}}{\to}$.

(1) If $\xi_n \to \xi$ and $\eta_n \to \eta$, then $(\xi_n, \eta_n) \to (\xi, \eta)$

(2) If $\xi_n \to \xi$ and if g is a continuous function, then $g(\xi_n) \to (\xi)$.

(3) More generally, assume that g is not continuous everywhere and denote by D_g a set of its discontinuity points. If $\xi_n \to \xi$ and if $\mathbf{P}(\xi \in D_g) = 0$, then $g(\xi_n) \to (\xi)$.

Weak convergence of distribution functions: $F_n \Rightarrow F$, if, for each x such that $F(x)$ is continuous in x,

$$F_n(x) \to F(x).$$

Equivalent form: $F_n \Rightarrow F$ if, for any bounded and continuous function g,

$$\int g(x)dF_n(x) \to \int g(x)dF(x).$$

Comment on terminology: Weak convergence is the most common term. Other terms are *convergence of/in distribution(s)* and *convergence in law.*

Weak convergence of random variables: $\xi_n \Rightarrow \xi$. It means: $\xi_n \Subset F_n$, $\xi \Subset F$ and $F_n \Rightarrow F$.

Note that $\xi_n \Rightarrow \xi$ is just a convenient notation ! There is no any "real" convergence of random variables on sample paths.

Relations between convergence types:

$$\xi_n \stackrel{\text{a.s.}}{\to} \xi \quad \text{implies} \quad \xi_n \stackrel{\text{p}}{\to} \xi \quad \text{and} \quad \xi_n \stackrel{\text{p}}{\to} \xi \quad \text{implies} \quad \xi_n \Rightarrow \xi.$$

Both converse statements are incorrect. Here are two examples:

Example 1. Weak convergence does not imply convergence in probability. Let $\mathbf{P}(\xi_1 = 1) = \mathbf{P}(\xi_1 = -1) = 1/2$ and $\xi_{n+1} = -\xi_n$, $n = 1, 2, \ldots$.

Example 2. Convergence in probability does not imply a.s. convergence. Let $\langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle$ $= ((0,1], \mathcal{B}_{(0,1]}, \lambda)$ where λ is the Lebesgue measure. Let $\xi_0 \equiv 1$. Let, for $m = 1, 2, \ldots$, for n such that $1 + 2 + \ldots + 2^{m-1} < n \leq 1 + 2 + \ldots + 2^{m-1} + 2^m$, and for $i = n - (1 + 2 + \ldots + 2^{m-1})$,

$$\xi_n(\omega) = 1 \quad \text{if} \quad \omega \in ((i-1)/2^m, i/2^m) \quad \text{and} \quad \xi_n = 0, \quad \text{otherwise.}$$

Laws of Large Numbers.

If $\xi, \xi_1, \xi_2, \ldots$ are i.i.d. random variables with a finite mean, say $a = \mathbf{E}\xi$, then the Weak Law of Large Numbers (WLLN) says:

$$S_n/n \stackrel{\mathrm{p}}{\to} a \quad \text{as} \quad n \to \infty$$

and the Strong Law of Large Numbers (SLLN) says

$$S_n/n \stackrel{\mathrm{a.s.}}{\to} a \quad \text{as} \quad n \to \infty.$$

Lebesgue and Beppo Levy Theorems.

Theorem (Beppo Levy). If $\{\xi_n\}$ is a.s. non-negative and non-decreasing sequence of random variables, then

$$\mathbf{E} \lim_{n\to\infty} \xi_n = \lim_{n\to\infty} \mathbf{E}\xi_n$$

where both sides are either finite or infinite simultaneously.

Coupling.

$\hat{\xi}$ is a copy of $\xi \iff$ they have the same distribution $\iff \hat{\xi} \stackrel{\mathrm{D}}{=} \xi$. In general, $\hat{\xi}$ and ξ may be defined on different probability spaces.

(a) Coupling of distribution functions (d.f.) or of probability measures.

For two d.f.'s F_1 and F_2, their *coupling* is a construction of a two-variate distribution function $F(x_1, x_2)$ such that $F(x_1, \infty) = F_1(x_1)$ and $F(\infty, x_2) = F_2(x_2)$.

Similarly, for two probability measures, $\mathbf{P}_1$ and $\mathbf{P}_2$ on the real line, their coupling is a probability measure on the plane $\mathbf{P}(\cdot)$, such that its projections are $\mathbf{P}_1$ and $\mathbf{P}_2$.

The same definitions of coupling may be introduced for any number of distributions (distribution functions, probability measures).

Such a coupling may also be viewed as follows: we define a probability space $\langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle$ and two random variables ξ_1 and ξ_2 on this space such that $\xi_1 \Subset F_1$ and $\xi_2 \Subset F_2$ (or, in other notation, $\xi_1 \Subset \mathbf{P}_1$ and $\xi_2 \Subset \mathbf{P}_2$). Then their joint distribution, say F, has marginals F_1 and F_2 (or, equivalently, a probablity measure $\mathbf{P}(B) = \mathbf{P}((\xi_1, \xi_2) \in B)$ has martinals $\mathbf{P}_1$ and $\mathbf{P}_2$).

(b) Coupling of two random variables.

Let ξ_1 be defined on $\langle \boldsymbol{\Omega_1}, \boldsymbol{\mathcal{F}_1}, \mathbf{P_1} \rangle$ and ξ_2 be defined on $\langle \boldsymbol{\Omega_2}, \boldsymbol{\mathcal{F}_2}, \mathbf{P_2} \rangle$.
A coupling of these two r.v.'s is defined by, first, an introduction of a new probability space, say $\langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle$ and, then, by defining a pair of two r.v.'s $\widehat{\xi}_1, \widehat{\xi}_2$ on this space such that $\widehat{\xi}_1 \stackrel{\mathrm{D}}{=} \xi_1$, $\widehat{\xi}_2 \stackrel{\mathrm{D}}{=} \xi_2$.

Examples:

(1) $F_1 = U(0,1)$, $F_2 = U(0,1)$;
(2) $F_1 = U(0,1)$, $F_2 = E(1)$;
(3) $F_1 = U(0,1)$, $F_2 = \Pi(1)$;
(4) $F_1 = B(n,p)$, $F_2 = \Pi(np)$;
(5) F_1 has a density $2x\mathbf{I}(x \in (0,1)$ and F_2 a density $2(1-x)\mathbf{I}(x \in (0,1)$.
In each example, there are many couplings !

1.2 Weak and "strong" convergence

Lemma 0. *If $F_n \Rightarrow F$ (all F_n and F are d.f.'s), then $\exists$ a coupling of $\{F_n\}$ and F:*

$$\xi_n \stackrel{\text{a.s.}}{\to} \xi.$$

Proof. For a d.f. F, define its inverse F^{-1} by

$$F^{-1}(z) = \inf\{x \ : F(x) \geq z\}, \quad z \in (0,1).$$

Let $\mathbf{\Omega}$ =(0,1), $\boldsymbol{\mathcal{F}}$ be the σ-algebra of Borel subsets in (0,1), and $\mathbf{P}$ the Lebesgue measure on (0,1).

Set $\eta(\omega) = \omega$, $\omega \in \mathbf{\Omega}$. Then $\eta \Subset U(0,1)$.

Let $\xi_n = F_n^{-1}(\eta), \xi = F^{-1}(\eta)$ and show $\xi_n \stackrel{\text{a.s.}}{\to} \xi$. Note that $\xi_n \Subset F_n, \xi \Subset F$

In order to avoid some technicalities, assume, for simplicity, that all d.f. are continuous. Let

$$\underline{\xi}_n = \inf_{m\geq n} \xi_m, \overline{\xi}_n = \sup_{m\geq n} \xi_m, \underline{F}_n = \sup_{m\geq n} F_m, \overline{F}_n = \inf_{m\geq n} F_m$$

Then $\underline{\xi}_n \Subset \underline{F}_n$, $\overline{\xi}_n \Subset \overline{F}_n$.

Indeed,

$$\begin{aligned}
\mathbf{P}(\underline{\xi}_n \leq x) &= \mathbf{P}(\underline{\xi}_n < x) = \mathbf{P}(\ \exists m \geq n : \xi_m < x) = \\
&= \mathbf{P}(\ \exists m \geq n : F_m^{-1}(\eta) < x) = \mathbf{P}(\ \exists m \geq n : \eta < F_m(x)) = \\
&= \mathbf{P}(\eta < \sup_{m\geq n} F_m(x)) = \underline{F}_n(x)
\end{aligned}$$

Similarly, $\mathbf{P}(\overline{\xi}_n > x) = \ldots = 1 - \overline{F}_n(x)$.

Since $\underline{F}_n \Rightarrow F$ and $\overline{F}_n \Rightarrow F$ (by definition), it is sufficient to show that, for instance, $\underline{\xi}_n \stackrel{\text{a.s.}}{\to} \xi$.

But both $\{\underline{F}_n\}$ and $\{\underline{\xi}_n\}$ are monotone as a function of n!

Then $\underline{\xi}_n \leq \xi$ a.s. and, therefore, there exists ψ such that $\underline{\xi}_n \nearrow \psi$ a.s. Then $\psi \leq \xi$ a.s.

If $\mathbf{P}(\psi \neq \xi) > 0$, then there exists x:

$$\mathbf{P}(\psi \leq x) > \mathbf{P}(\xi \leq x).$$

But $\mathbf{P}(\xi \leq x) = F(x) = \lim \underline{F}_n(x) \geq \mathbf{P}(\psi \leq x)$!

Thus, we got a contradiction, and $\underline{\xi}_n \nearrow \xi$ a.s. By similar arguments, $\overline{\xi}_n \downarrow \xi$ a.s. Therefore, $\xi_n \to \xi$ a.s.

□

Problem No 1. Prove this lemma without the additional assumption that all d.f.'s are continuous

Exercises: What is F^{-1} for the following distribution functions: $U(0,1)$, $E(\alpha)$, $N(0,1)$, $B(1,p)$, $B(n,p$, $\Pi(\lambda)$...

1.3 Uniform integrability

Let $\{\xi_n\}_{n\geq 1}$ be a sequence of real-valued r.v.'s.

Definition 1. *$\{\xi_n\}$ are uniformly integrable (UI), if $\mathbf{E}|\xi_n| < \infty \ \ \forall n$ and, moreover,*

$$\sup_n \mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| \geq x)\} \leq h(x) \to 0 \text{ as } x \to \infty.$$

Comments:

– Actually, we can put "=" instead of "≤"' in the definition above. But I prefer to keep "≤" since I want the upper bound $h(x)$ to be monotone non-increasing and right-continuous.

– Clearly, if $\{\xi_n\}$ are UI, then $\sup_n \mathbf{E}|\xi_n|$ is finite.

Examples:

(1) $\xi_n \Subset E(\alpha_n)$, $n = 1, 2, \ldots$ are UI if and only if $\min_n \alpha_n > 0$.

(2)

ξ_n	$2n$	$-2n$	0
	$\frac{1}{2n}$	$\frac{1}{2n}$	$1-\frac{1}{n}$

$\Longrightarrow$ $\mathbf{E}|\xi_n| = 2, \mathbf{E}\xi_n = 0; \xi \equiv 0,$ but $\{\xi_n\}$ are not UI!

Lemma 1. *The following are equivalent:*
(i) *$\{\xi_n\}$ are UI;*
(ii) *$\exists$ a function $g : [0, \infty) \to [0, \infty)$:*
(a) $g(0) > 0$; $g \nearrow$; $\lim_{x\to\infty} g(x) = \infty$;
(b) $\sup_n \mathbf{E}\{|\xi_n| \cdot g(|\xi_n|)\} < \infty$

Note: $g(0) > 0$ is not essential!

Proof.

(ii) → (i). For each n,

$$\mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| \geq x)\} \equiv \mathbf{E}\{|\xi_n| \cdot \frac{g(|\xi_n|)}{g(|\xi_n|)} \cdot \mathbf{I}(|\xi_n| \geq x)\} \leq$$

$$\leq \frac{1}{g(x)} \cdot \sup_n \mathbf{E}\{|\xi_n| \cdot g(|\xi_n|)\} \to 0 \text{ as } x \to \infty.$$

(i) → (ii). Assume that $h(x) > 0 \ \ \forall x$ (otherwise the statement is trivial).

For $m \in \mathbf{Z}$, let

$$A_m = \{x : \frac{1}{2^{2(m+1)}} < h(x) \leq \frac{1}{2^{2m}}\}$$

and, for $x \in A_m$, let $g(x) = 2^m$. From $h(0) < \infty$, we get $g(0) > 0$.

Note that A_m is an interval which is closed from the left and open from the right. Denote by z_m its left boundary point, $z_m \in A_m$. Then

$$\mathbf{E}\{|\xi_n| \cdot g(|\xi_n|)\} = \sum_m \mathbf{E}\{|\xi_n| \cdot g(|\xi_n|) \cdot \mathbf{I}(|\xi_n| \in A_m)\} =$$

$$= \sum_m \mathbf{E}\{|\xi_n| \cdot 2^m \cdot \mathbf{I}(|\xi_n| \in A_m)\} \le \sum_m 2^m \mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| \ge z_m)\} \le$$
$$\le \sum_m 2^m \cdot h(z_m) \le \sum_m 2^m \cdot \frac{1}{2^{2m}} < \infty$$

Note that, in the proof of Lemma 1, we have proposed a particular choice of a function g which is the step function. But we can make g very "smooth" if we like.

Remark 1. *As a corollary, one can get the following: if $\mathbf{E}|\xi| < \infty$, then $\exists\, g$ from Lemma 1 such that $\mathbf{E}\{|\xi| \cdot g(|\xi|)\} < \infty$. This may be reworded as "$\exists$ the first moment" $\Longrightarrow$ "$\exists$ something more".*

Lemma 2. *Assume $\xi_n \Rightarrow \xi$. Then*
(1) $\{\xi_n\}$ are UI $\Longrightarrow$ $\mathbf{E}|\xi| < \infty$ and $\mathbf{E}|\xi_n| \to \mathbf{E}|\xi|$ (and, therefore, $\mathbf{E}\xi_n \to \mathbf{E}\xi$);
(2) $\mathbf{P}(\xi_n \ge 0) = 1 \ \ \forall\, n$; $\mathbf{E}\xi_n < \infty \ \ \forall\, n$; $\mathbf{E}\xi_n \to \mathbf{E}\xi < \infty$ $\Longrightarrow$ $\{\xi_n\}$ are UI.

Remark 2. *In (2), the condition $\mathbf{P}(\xi_n \ge 0) = 1$ may be weakened in a natural way.* Problem No 2. How? *But it cannot be eliminated.*

Proof of Lemma 2. First, note that both statements (1) and (2) are "marginal", i.e. only marginal distributions are involved. So, we can construct a coupling: $\xi_n \stackrel{\text{a.s.}}{\to} \xi$.

Prove (1).

(a) Assume first that (distributions of) r.v.'s are "uniformly bounded", i.e. $\exists N$: $\mathbf{P}(|\xi_n| \le N) = 1 \ \ \forall n$ (this is a special case of UI).

Then $\mathbf{P}(|\xi| \le N) = 1$ and, $\forall \varepsilon > 0$,

$$\begin{aligned} 0 \le\ & |\mathbf{E}|\xi_n| - \mathbf{E}|\xi|| \le \mathbf{E}||\xi_n| - |\xi|| = \mathbf{E}\{||\xi_n| - |\xi|| \cdot \mathbf{I}(||\xi_n| - |\xi|| \le \varepsilon)\} + \\ +\ & \mathbf{E}\{||\xi_n| - |\xi|| \cdot \mathbf{I}(||\xi_n| - |\xi|| > \varepsilon)\} \le \varepsilon + 2N \cdot \mathbf{P}(||\xi_n| - |\xi|| > \varepsilon) \to \varepsilon \text{ as } n \to \infty \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, $\mathbf{E}|\xi_n| \to \mathbf{E}|\xi|$.

(b) Assume now that at least one of distributions of r.v.'s has an unbounded support, that is, $\mathbf{P}(|\xi_n| \le N) < 1 \ \ \forall N$ and for some n.

(b1) Take any $x > 0$ such that $\mathbf{P}(|\xi| = x) = 0$. Since $\xi_n \stackrel{\text{a.s.}}{\to} \xi$,

$$\eta_n \equiv \xi_n \cdot \mathbf{I}(|\xi_n| < x) \stackrel{\text{a.s.}}{\to} \xi \cdot \mathbf{I}(|\xi| < x) \equiv \eta.$$

Then

$$\forall n,\ \mathbf{P}(|\eta_n| \le x) = \mathbf{P}(|\eta| \le x) = 1 \Longrightarrow \mathbf{E}\eta_n \to \mathbf{E}\eta \text{ (see (a))};$$

and

$$|\eta_n| \le |\xi_n| \text{ a.s. } \Longrightarrow \mathbf{E}|\eta_n| \le \mathbf{E}|\xi_n| \le \sup_n \mathbf{E}|\xi_n| \equiv K \ \forall\, n \ \Longrightarrow \mathbf{E}|\eta| \le K.$$

(b2) Show first that $\mathbf{E}|\xi| < \infty$. Indeed,

$$\mathbf{E}|\xi| = \lim_{x\to\infty} \mathbf{E}\{|\xi| \cdot \mathbf{I}(|\xi| \le N)\} \le K < \infty$$

(b3) Now $\forall\,\varepsilon > 0$, choose x such that $\mathbf{P}(|\xi| = x) = 0$, $h(x) \le \varepsilon$, and $\mathbf{E}\{|\xi| \cdot \mathbf{I}(|\xi| \ge x)\} \le \varepsilon$.

Let

$$\delta_n = \mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| \ge x)\} \quad \text{and} \quad \delta = \mathbf{E}\{|\xi| \cdot \mathbf{I}(|\xi| \ge x)\}.$$

Then

$$\begin{array}{rcl} \mathbf{E}|\xi_n| & = & \mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| < x)\} + \delta_n, \\ & & \downarrow \\ \mathbf{E}\xi & = & \mathbf{E}\{|\xi| \cdot \mathbf{I}(|\xi| < x)\} + \delta. \end{array}$$

Since $\delta_n \le \varepsilon \;\; \forall\, n$ and $|\delta| \le \varepsilon$, then

$$\limsup(\mathbf{E}|\xi_n| - \mathbf{E}|\xi|) \le 2\varepsilon \text{ and}$$
$$\liminf(\mathbf{E}|\xi_n| - \mathbf{E}|\xi|) \ge -2\varepsilon \text{ for any } \varepsilon.$$

Letting ε to 0, we obtain the first statement of the lemma. □

Prove now the second statement. First, from $\mathbf{E}\xi < \infty$, we may take an arbitrary $\varepsilon > 0$ and then choose $x_0 = x_0(\varepsilon)$ such that $\mathbf{P}(\xi = x_0) = 0$ and

$$\mathbf{E}\{\xi \cdot \mathbf{I}(\xi \ge x_0)\} \le \varepsilon/2.$$

Then we may use part (b1) from the proof of (1): for a given x_0,

$$\begin{array}{rl} \mathbf{E}\eta_n \to \mathbf{E}\eta \quad \Longrightarrow & \mathbf{E}\{\xi_n \cdot \mathbf{I}(\xi_n \ge x_0)\} = \mathbf{E}(\xi_n - \eta_n) = \\ & = \mathbf{E}\xi_n - \mathbf{E}\eta_n \to \mathbf{E}\xi - \mathbf{E}\eta = \mathbf{E}\{\xi \cdot \mathbf{I}(\xi \ge x_0)\} \le \varepsilon/2. \end{array}$$

Therefore, $\exists\, n(\varepsilon)$ such that

$$\mathbf{E}\{\xi_n \cdot \mathbf{I}(\xi_n \ge x_0)\} \le \varepsilon \quad \forall n > n(\varepsilon).$$

Now, $\forall\; n = 1, 2, \ldots, n(\varepsilon)$,

$$\mathbf{E}\xi_n < \infty \Longrightarrow \exists\, x_n : \mathbf{E}\{\xi_n \cdot \mathbf{I}(\xi_n \ge x_n)\} \le \varepsilon.$$

Let $x = \max(x_1, \ldots, x_{n(\varepsilon)}, x_0)$. Then

$$\mathbf{E}\{\xi_n \cdot \mathbf{I}(\xi_n \ge x)\} \le \varepsilon \,\forall\, n.$$

Thus,

$$\sup_n \mathbf{E}\{\xi_n \cdot \mathbf{I}(\xi_n \ge x)\} \to 0 \text{ as } x \to \infty.$$

□

1.4 Some useful properties of UI

Property 1. [*If $\{\xi_n\}$ are UI and if $\{\eta_n\}$ are such that $|\eta_n| \le |\xi_n|$ a.s., then $\{\eta_n\}$ are UI.*

Indeed, let $h(x)$ be from Definition 1. Then, $\forall\, x > 0$,

$$\mathbf{E}\{|\eta_n| \cdot \mathbf{I}(|\eta_n| > x)\} \le \mathbf{E}\{|\eta_n| \cdot \mathbf{I}(|\xi_n| > x)\} \le \mathbf{E}\{|\xi_n| \cdot \mathbf{I}(|\xi_n| > x)\} \le h(x).$$

□

Property 2. [*If $\{\xi_n\}$ is an i.i.d. sequence with finite mean, $\mathbf{E}|\xi_1| < \infty$ and if $|\eta_n| \le |\xi_n|$ a.s., then a sequence $\psi_n = \dfrac{\eta_1 + \ldots + \eta_n}{n}$, $n = 1, 2, \ldots$ is UI.*

Indeed,

$$|\psi_n| \le \frac{|\xi_1| + \ldots + |\xi_n|}{n} \equiv \varphi_n,$$

where

(i) $\mathbf{E}\varphi_n = \mathbf{E}|\xi_1| \;\; \forall\, n$ and,

(ii) by the SLLN,

$$\varphi_n \stackrel{\text{a.s.}}{\to} \mathbf{E}|\xi_1|.$$

$\models\Rightarrow$ From Lemma 2, (2), $\{\varphi_n\}$ are UI.

$\models\Rightarrow$ From Property 1.1, $\{\psi_n\}$ are UI. □

Property 3. [*Since the UI property is the property of "marginal" distributions only, one can replace the a.s.-inequality in Property 1.1, $|\eta_n| \le |\xi_n|$, by the weaker one, $|\eta_n| \le_{\rm st} |\xi_n|$ (this means: $\mathbf{P}(|\eta_n| > x) \le \mathbf{P}(|\xi_n| > x) \;\; \forall x$). In particular, if r.v.s $\{\eta_n\}$ admit a stochastic integrable majorant ξ,*

$$|\eta_n| \le_{\rm st} |\xi|, \quad \forall n$$

and if $\mathbf{E}|\xi| < \infty$, then $\{\eta_n\}$ are UI.

Remark 3. Consider, instead of a sequence $\{\xi_n\}_{n \ge 1}$, a family of r.v.'s $\{\xi_t\}_{t \in T}$ indexed by an arbitrary set T. Then one can introduce the following

Definition 2. [(compare with Definition 1).
$\{\xi_t\}_{t \in T}$ are UI, if $\mathbf{E}|\xi_t| < \infty \;\; \forall t \in T$ and, moreover,

$$\sup_{t \in T} \mathbf{E}\{|\xi_t| \cdot \mathbf{I}(|\xi_t| \ge x)\} \le h(x) \to 0, \text{ as } x \to \infty.$$

Then

(a) The statement and the proof of Lemma 1 stay the same if we replace "$n = 1, 2, \ldots$" by "$t \in T$".

(b) Similarly, the statement and the proof of Lemma 2 stay unchanged if we replace "$n = 1, 2, \ldots$" by "$t \in T = [0, \infty)$".

(c) Properties 1 and 3 still hold is we replace "$n = 1, 2, \ldots$" by "$t \in T$".

1.5 Coupling inequality. Maximal coupling. Dobrushin's theorem.

In this section, we assume that random variables are not necessarily real-valued and may take values in a general measurable space $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{B}_{\mathcal{X}}})$ which is assumed to be complete *separable* metric space.

The Coupling Inequality

Let $\xi_1, \xi_2 : \langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle \longrightarrow (\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{B}_{\mathcal{X}}})$ be two $\boldsymbol{\mathcal{X}}$-valued r.v.'s. Let

$$\mathbf{P}_1(B) = \mathbf{P}(\xi_1 \in B),\ \mathbf{P}_2(B) = \mathbf{P}(\xi_2 \in B),\ B \in \boldsymbol{\mathcal{B}_{\mathcal{X}}}.$$

Then, for $B \in \boldsymbol{\mathcal{B}_{\mathcal{X}}}$,

$$\begin{aligned} \mathbf{P}_1(B) - \mathbf{P}_2(B) &= \mathbf{P}(\xi_1 \in B, \xi_1 = \xi_2) + \mathbf{P}(\xi_1 \in B, \xi_1 \neq \xi_2) - \\ &- \mathbf{P}(\xi_2 \in B, \xi_1 = \xi_2) - \mathbf{P}(\xi_2 \in B, \xi_1 \neq \xi_2) = \\ &= \mathbf{P}(\xi_1 \in B, \xi_1 \neq \xi_2) - \mathbf{P}(\xi_2 \in B, \xi_1 \neq \xi_2) \begin{array}{l} \leq \mathbf{P}(\xi_1 \neq \xi_2) \\ \geq -\mathbf{P}(\xi_1 \neq \xi_2) \end{array} \end{aligned}$$

Therefore, for any $B \in \boldsymbol{\mathcal{B}_{\mathcal{X}}}$, $|\mathbf{P}_1(B) - \mathbf{P}_2(B)| \leq \mathbf{P}(\xi_1 \neq \xi_2)$, that is

$$\sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)| \leq \mathbf{P}(\xi_1 \neq \xi_2) \qquad (*)$$

The Maximal Coupling

Now we reformulate the result obtained. Note that the LHS of inequality (*) depends on "marginal" distributions $\mathbf{P}_1$ and $\mathbf{P}_2$ only and does not depend on the joint distribution of ξ_1 and ξ_2. Therefore, we get the following:

for any coupling of marginal distributions $\mathbf{P}_1$ and $\mathbf{P}_2$, inequality (*) holds. Equivalently,

$$\sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)| \leq \inf_{\textit{in all coupling}} \mathbf{P}(\xi_1 \neq \xi_2) \qquad (**)$$

The following questions seem to be natural:

(?) Is there equality in $(**)$?

(??) If the answer is "yes", then does there exist a coupling such that

$$\sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)| = \mathbf{P}(\xi_1 \neq \xi_2)?$$

The answers to both questions are positive! And this is the content of **Dobrushin's theorem**.

Theorem 1. *Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two probability measures on a complete separable metric space $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}})$. There exists a coupling of these probalility measures such that, for $\xi_i \Subset \mathbf{P}_i$, $i = 1, 2$,*

$$\sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)| = \mathbf{P}(\xi_1 \neq \xi_2).$$

Proof. $\mu(B) = \mathbf{P}_1(B) - \mathbf{P}_2(B)$ is a signed measure. Then Banach theorem states that there exists a subset $C \subset \boldsymbol{\mathcal{X}}$ such that

(a) $\mu(B) \geq 0 \ \ \forall\, B \subset C$;

(b) $\mu(B) \leq 0 \ \ \forall\, B \subset \boldsymbol{\mathcal{X}} \setminus C \equiv \overline{C}$.

Note:

1) if $\mu(C) = 0$, then $\mathbf{P}_1 = \mathbf{P}_2$ and the coupling is obvious;

2) $\mu(C) = -\mu(\overline{C})$.

Assume $\mu(C) > 0$. Introduce 4 distributions (probability measures):

$$Q_{1,1} \quad \text{is defined by} \quad \begin{cases} Q_{1,1} = U(\overline{C}), & \text{if } \mathbf{P}_1(\overline{C}) = 0, \\ Q_{1,1}(B) = \dfrac{\mathbf{P}_1(\overline{C} \cap B)}{\mathbf{P}_1(\overline{C})}, \ B \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}, & \text{otherwise.} \end{cases}$$

and

$$Q_{2,1} \quad \text{is defined by} \quad Q_{2,1}(B) = \frac{\mathbf{P}_2(\overline{C} \cap B) - \mathbf{P}_1(\overline{C} \cap B)}{-\mu(\overline{C})}, \ B \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}.$$

Similarly,

$$Q_{2,2} \quad \text{is defined by} \quad \begin{cases} Q_{2,2} = U(C), & \text{if } \mathbf{P}_2(C) = 0, \\ Q_{2,2}(B) = \dfrac{\mathbf{P}_2(C \cap B)}{\mathbf{P}_2(C)}, \ B \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}, & \text{otherwise.} \end{cases}$$

and

$$Q_{1,2} \quad \text{is defined by} \quad Q_{1,2}(B) = \frac{\mathbf{P}_1(C \cap B) - \mathbf{P}_2(C \cap B)}{\mu(C)}, \ B \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}.$$

Then introduce 5 mutually independent r.v.s:

$$\eta_{1,1} \Subset Q_{1,1}, \ \eta_{1,2} \Subset Q_{1,2}, \ \eta_{2,1} \Subset Q_{2,1}, \ \eta_{2,2} \Subset Q_{2,2},$$

and

α	1	2	0
	$\mathbf{P}_1(\overline{C})$	$\mathbf{P}_2(C)$	$\mu(C)$

Now we can define ξ_1 and ξ_2 as follows:

$$\xi_1 = \eta_{1,1} \cdot \mathbf{I}(\alpha = 1) + \eta_{2,2} \cdot \mathbf{I}(\alpha = 2) + \eta_{2,1} \cdot \mathbf{I}(\alpha = 0),$$

$$\xi_2 = \eta_{1,1} \cdot \mathbf{I}(\alpha = 1) + \eta_{2,2} \cdot \mathbf{I}(\alpha = 2) + \eta_{1,2} \cdot \mathbf{I}(\alpha = 0).$$

Simple calculations show that $\xi_i \Subset \mathbf{P}_i$, $i = 1, 2$. This is Problem No 3 for you.

Then,

$$\mathbf{P}(\xi_1 \neq \xi_2) = \mathbf{P}(\alpha = 0) = \mu(C) \leq \sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)|.$$

So,

$$\mathbf{P}(\xi_1 \neq \xi_2) = \sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)|.$$

□

Comment. Banach theorem and Radon-Nykodim theorem are two equivalent statements formulated in slightly different ways.
There is (formally!) another proof (see, e.g. T. Lindvall's book on the coupling method) based on Radon-Nykodim theorem:
Consider a new probability measure $\mathbf{P}(\cdot) = (\mathbf{P}_1(\cdot) + \mathbf{P}_2(\cdot))/2$. Let $f_i = \frac{d\mathbf{P}_i}{d\mathbf{P}}$ be corresponding densities. Then

$$\sup_{B \in \mathcal{B}_{\mathcal{X}}} |\mathbf{P}_1(B) - \mathbf{P}_2(B)| = 1 - \int \min(f_1(x), f_2(x)) \mathbf{P}(dx),$$

and we may repeat the previous construction using densities.

What is the maximal coupling in the following examples:

(1) Two discrete two-point distributions.
(2) Two absolutely continuous distributions on $(0, 1)$ with densities f_1 and f_2.
(3) Bernoulli and Poisson distributions.
(4) Normal and exponential distributions.
(this is another exercise to you)

1.6 Probabilistic Metrics

Dobrushin's theorem provides a positive solution to one of important problems in the theory of Probabilistic Metrics. We will discuss briefly basic concepts of this theory.

Again, consider a complete separable metric space $(\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}})$and introduce the following notation:

1) $\boldsymbol{\mathcal{X}}^2 = \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{X}}$,

2) $\boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}} \otimes \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}$ is a σ-algebra in $\boldsymbol{\mathcal{X}}^2$ generated by all sets $B_1 \times B_2$, $B_1, B_2 \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}$,

3) $\mathrm{diag}(\boldsymbol{\mathcal{X}}^2) = \{(x,x), x \in \boldsymbol{\mathcal{X}}\}$.

Problem No 4. Prove that $\mathrm{diag}(\boldsymbol{\mathcal{X}}^2) \in \boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}}$. (Actually, there is no need to assume that the state space is complete separable metric, and the $minimal$ requirement for $\mathrm{diag}(\boldsymbol{\mathcal{X}}^2) \in \boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}}$ to hold is that the sigma-algebra $\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}$ is countably generated).

Let $\mathbf{P}$ be any probability distribution on $(\boldsymbol{\mathcal{X}}^2, \boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}})$. Denote by $\mathbf{P}_i$, $i = 1, 2$ its marginal distributions:

$$\begin{aligned} \mathbf{P}_1(B) &= \mathbf{P}(B \times \boldsymbol{\mathcal{X}}), \\ \mathbf{P}_2(B) &= \mathbf{P}(\boldsymbol{\mathcal{X}} \times B),\ B \in \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{X}}}. \end{aligned}$$

Let $\boldsymbol{\mathcal{P}}$ be a set of all probability distributions (measures) on $(\boldsymbol{\mathcal{X}}^2, \boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}})$.

Definition 3. *A function $d : \boldsymbol{\mathcal{P}} \longrightarrow [0,\infty)$ is called a probabilistic metric if it satisfies the following conditions:*

(1) $\mathbf{P}(\mathrm{diag}(\boldsymbol{\mathcal{X}}^2)) = 1 \ \Longrightarrow d(\mathbf{P}) = 0$;

(2) $d(\mathbf{P}) = 0 \quad \Longrightarrow \mathbf{P}_1 = \mathbf{P}_2$;

(3) *the "triangle inequiality":*
$\mathbf{P}^{(1)}$ *has marginals* $\mathbf{P}_1$ *and* $\mathbf{P}_2$
$\mathbf{P}^{(2)}$ *has marginals* $\mathbf{P}_1$ *and* $\mathbf{P}_3$ $\quad \Longrightarrow d(\mathbf{P}^{(1)}) \le d(\mathbf{P}^{(2)}) + d(\mathbf{P}^{(3)})$;
$\mathbf{P}^{(3)}$ *has marginals* $\mathbf{P}_3$ *and* $\mathbf{P}_2$

Definition 4. *A probabilistic metric d is simple if it depends on marginal distributions only (i.e. if $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ have the same marginals, then $d(\mathbf{P}^{(1)}) = d(\mathbf{P}^{(2)})$), and complex — otherwise.*

For a simple metric, it is reasonable to write $d(\mathbf{P}_1, \mathbf{P}_2)$ instead of $d(\mathbf{P})$, so d has the meaning of a "distance" between $\mathbf{P}_1$ and $\mathbf{P}_2$.

For a complex metric, we may also write $d(\xi_1, \xi_2)$ instead of $d(\mathbf{P})$ where ξ_1, ξ_2 is a coupling of two r.v.'s with a joint distribution $\mathbf{P}$,

$$\mathbf{P}(B) = \mathbf{P}((\xi_1, \xi_2) \in B),\ B \in \boldsymbol{\mathcal{B}}^2_{\boldsymbol{\mathcal{X}}}.$$

So, $d(\xi_1, \xi_2)$ may be considered as a "distance" between r.v.'s.

We can also write $d(\xi_1, \xi_2)$ for simple metrics. In this case,

$$d(\xi_1, \xi_2) = d(F_1, F_2) = d(\mathbf{P}_1, \mathbf{P}_2).$$

Examples.

Simple	Complex
1) $\sup_{B\in\mathcal{B}} \lvert\mathbf{P}_1(B) - \mathbf{P}_2(B)\rvert$	2) $\mathbf{P}(\xi_1 \neq \xi_2) \equiv \mathbf{P}(\boldsymbol{\mathcal{X}}^2 - \mathrm{diag}(\boldsymbol{\mathcal{X}}^2))$
(Total variation norm (T.V.N.))	(Indicator metric (I.M.))

For real-valued r.v.'s:

3) $\sup_x \lvert F_1(x) - F_2(x)\rvert$	5) $\inf\{\varepsilon > 0 : \mathbf{P}(\lvert\xi_1 - \xi_2\rvert > \varepsilon) < \varepsilon\}$
(Uniform metric (U.M.))	(Ki Fan metric (K.F.M.))

4) $\inf\{\varepsilon > 0 : \boldsymbol{\mathcal{F}}_1(x-\varepsilon) - \varepsilon \leq F_2(x) \leq F_1(x+\varepsilon) + \varepsilon \ \forall x\}$

(Levy metric (L.M.))

One of key problems in the theory of probabilistic metrics is to find answers to the following questions:

Assume a simple metric $d(\mathbf{P}_1, \mathbf{P}_2)$ is given. Does there exist a complex metric $\tilde{d}$ such that

(a) *the following coupling inequality holds:*

$$d(\xi_1, \xi_2) \leq \inf_{\textit{all couplings}} \widetilde{d}(\xi_1, \xi_2) \qquad ?\textit{(compare with } (**)\textit{)}$$

(b) *If yes, then is it possible to replace "$\leq$" by "$=$" in (a) ?*

(c) *Does there exist a coupling such that $d(\xi_1, \xi_2) = \widetilde{d}(\xi_1, \xi_2)$?*

The following result holds:

Theorem 2. *The answers to the above questions are positive for the metrics:*
(1) $d =$ *T.V.N.* $\longleftrightarrow \widetilde{d} =$ *I.M.*
(2) $d =$ *L.M.* $\longleftrightarrow \widetilde{d} =$ *K.F.M.*

Comment. Statement (1) is Dobrushin's theorem. Statement (2) is Strassen's theorem (its proof is omitted).

1.7 Stopping times

Let $\langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle$ be a probability space and $\{\xi_n\}_{n\geq 1}$ a sequence of r.v.'s, $\xi_n : \boldsymbol{\Omega} \to \mathbf{R}$. Denote by $\boldsymbol{\mathcal{F}}_n$ a σ-algebra, generated by ξ_n:

$$\boldsymbol{\mathcal{F}}_n \subseteq \boldsymbol{\mathcal{F}};\ \boldsymbol{\mathcal{F}}_n = \{\xi_n^{-1}(B),\, B \in \boldsymbol{\mathcal{B}}\},$$

where $\boldsymbol{\mathcal{B}}$ is a σ-algebra of Borel sets in $\mathbf{R}$.

Then, for $1 \leq k \leq n$, $\boldsymbol{\mathcal{F}}_{[k,n]}$ is a σ-algebra generated by $\xi_k, \ldots, \xi_n$; i.e.

$\boldsymbol{\mathcal{F}} \supseteq \boldsymbol{\mathcal{F}}_{[k,n]}$ is a minimal σ-algebra such that

$$\boldsymbol{\mathcal{F}}_{[k,n]} \supseteq \boldsymbol{\mathcal{F}}_l \text{ for all } l = k, \ldots, n.$$

Another way to describe $\boldsymbol{\mathcal{F}}_{[k,n]}$ is:
let $\vec{\xi}_{k,n} := (\xi_k, \ldots, \xi_n)$ be a random vector; $\vec{\xi}_{k,n} : \boldsymbol{\Omega} \to \mathbf{R}^{n-k+1}$. Then

$$\boldsymbol{\mathcal{F}}_{[k,n]} = \{\vec{\xi}_{k,n}^{-1}(B),\, B \in \boldsymbol{\mathcal{B}}^{n-k+1}\},$$

where $\boldsymbol{\mathcal{B}}^{n-k+1}$ is a σ-algebra of Borel sets in $\mathbf{R}^{n-k+1}$.

Finally, $\boldsymbol{\mathcal{F}}_{[1,\infty)}$ is a σ-algebra generated by the whole sequence $\{\xi_n\}_{n\geq 1}$.

Good Property : $\left[\begin{array}{l} \forall\ A \in \boldsymbol{\mathcal{F}}_{[1,\infty)},\ \exists\ \textit{a sequence of events}\ \{A_n\}_{n\geq 1},\ A_n \in \boldsymbol{\mathcal{F}}_{[1,n]}\ \textit{such that} \\ \mathbf{P}(A \setminus A_n) + \mathbf{P}(A_n \setminus A) \to 0\ \textit{as}\ n \to \infty. \end{array} \right.$

Let now $\mu : \boldsymbol{\Omega} \to \{1, 2, \ldots, n, \ldots\}$ be an integer-valued r.v. (we say it is a "counting" r.v.)

Definition 5. *μ is a stopping time (ST) with respect to $\{\xi_n\}$, if $\forall\ n \geq 1$,*

$$\{\mu = n\} \in \boldsymbol{\mathcal{F}}_{[1,n]}$$

(or, equivalently — $\{\mu \leq n\} \in \boldsymbol{\mathcal{F}}_{[1,n]}$).

Another variant of a definition of a stopping time is:

Definition 6. *μ is an ST if $\exists$ a family of functions $h_n : \mathbf{R}^n \to \{0,1\}$ such that:*

$$\forall\ n \geq 1,\ \mathbf{I}(\mu = n) = h_n(\xi_k, \ldots, \xi_n) \textit{ a.s.}$$

(or, equivalently — $\mathbf{I}(\mu \leq n) = h_n(\xi_k, \ldots, \xi_n)$ a.s.).

Examples of ST's:
(1) $\mu = \min\{n \geq 1 : \xi_n \geq x\}$;
(2) $\mu = \min\{n \geq 1 : \sum_1^n \xi_i \geq x\}$;
(3) More examples....

Assume now that $\{\xi_n\}$ is an i.i.d. sequence, μ is an ST with $\mathbf{P}(\mu < \infty) = 1$.
Let

$$\tilde{\xi}_1 = \xi_{\mu+1}, \tilde{\xi}_2 = \xi_{\mu+2}, \ldots, \tilde{\xi}_i = \xi_{\mu+i}, \ldots$$

Lemma 3. *The following statements hold:*

1) $\{\tilde{\xi}_i\}$ *is an i.i.d. sequence;*

2) $\tilde{\xi}_i \overset{\mathrm{D}}{=} \xi_1$*;*

3) $\{\tilde{\xi}_i\}_{i\ge 1}$ *and a random vector* $(\mu, \xi_1, \ldots, \xi_\mu)$ *are mutually independent.*

Corollary 1. $\{\tilde{\xi}_i\}_{i\ge 1}$ *and* $S_\mu \equiv \xi_1 + \ldots + \xi_\mu$ *are mutually independent.*

Proof of Lemma 3. It is sufficient to show that
$\forall\, k \ge 1,\ \forall\, m \ge 1,\ \forall$ Borel sets $B_1, \ldots, B_k$ and $C_1, \ldots, C_m$,

$$\begin{aligned}&\mathbf{P}(\{\mu = k;\ \xi_1 \in B_1, \ldots, \xi_k \in B_k\} \cap \{\tilde{\xi}_1 \in C_1, \ldots, \tilde{\xi}_m \in C_m\}) = \\ &= \mathbf{P}(\mu = k;\ \xi_1 \in B_1, \ldots, \xi_k \in B_k)\mathbf{P}(\xi_1 \in C_1, \ldots, \xi_m \in C_m).\end{aligned} \qquad (*)$$

Indeed, $(*)\ \models\!\!\!\Longrightarrow$ 1), 2), and 3).

First, take $B_1 = \ldots = B_k = B_{k+1} = \ldots = \mathbf{R}$. Then, $\forall\, m$,

$$\begin{aligned}&\mathbf{P}(\tilde{\xi}_1 \in C_1, \ldots, \tilde{\xi}_m \in C_m) \overset{t.p.f.}{=} \sum_{k=1}^{\infty} \mathbf{P}(\mu = k; \tilde{\xi}_1 \in C_1, \ldots, \tilde{\xi}_m \in C_m) \\ &\overset{(*)}{=} \sum_{k=1}^{\infty} \mathbf{P}(\mu = k) \prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i) = \prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i).\end{aligned} \qquad (**)$$

In particular, $\forall\, j \ge 1\, \forall\, C_j$, we can take $m \ge j$ and $C_i = \mathbf{R}$ for $i \ne j$.
Then the LHS of $(**)$=$\mathbf{P}(\tilde{\xi}_j \in C_j)$,
the RHS of $(**)$=$\mathbf{P}(\xi_1 \in C_j)$.
$\Bigg\}\ \models\!\!\!\Longrightarrow$ 2)

Now, take any $C_1, \ldots, C_m$ and replace in $(**)$
$\prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i)$ by $\prod_{i=1}^{m} \mathbf{P}(\tilde{\xi}_1 \in C_i)$.
$\Bigg\}\ \models\!\!\!\Longrightarrow$ 1)

Finally, take any $B_1, \ldots, B_k$ and $C_1, \ldots, C_m$ and replace in $(*)$
$\prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i)$ by $\prod_{i=1}^{m} \mathbf{P}(\tilde{\xi}_i \in C_i)$.
$\Bigg\}\ \models\!\!\!\Longrightarrow$ 3)

So, we will prove $(*)$ now:

$$\begin{aligned}&\mathbf{P}(\{\mu = k;\ \xi_1 \in B_1, \ldots, \xi_k \in B_k\} \cap \{\tilde{\xi}_1 \in C_1, \ldots, \tilde{\xi}_m \in C_m\}) = \\ &\mathbf{P}(\underbrace{\{h_k(\xi_1, \ldots, \xi_k) = 1;\ \xi_1 \in B_1, \ldots, \xi_k \in B_k\}}_{\in \boldsymbol{\mathcal{F}}_{[1,k]}} \cap \underbrace{\{\xi_{k+1} \in C_1, \ldots, \xi_{k+m} \in C_m\}}_{\in \boldsymbol{\mathcal{F}}_{[k+1,k+m]}}) = \\ &= \mathbf{P}(\ldots) \cdot \mathbf{P}(\ldots) = \\ &= \mathbf{P}(\ldots) \cdot \prod_{i=1}^{m} \mathbf{P}(\xi_{k+i} \in C_i) = \mathbf{P}(\ldots) \cdot \prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i).\end{aligned}$$

□

Lemma 4. (Wald identity) *Assume that $\mathbf{E}|\xi_1| < \infty$ and $\mathbf{E}\mu < \infty$. Then*

$$\mathbf{E}S_\mu = \mathbf{E}\xi_1 \cdot \mathbf{E}\mu.$$

Proof. (a) Show that $\mathbf{E}|S_\mu| < \infty$.

$$|S_\mu| \le \sum_{n=1}^{\mu} |\xi_n| \equiv \sum_{n=1}^{\infty} |\xi_n| \cdot \mathbf{I}(\mu \ge n).$$

Note, that $\mathbf{I}(\mu \ge n) = 1 - \mathbf{I}(\mu \le n-1)$, and $\{\mu \le n-1\} \in \boldsymbol{\mathcal{F}}_{[1,n-1]}$
$\models\!\!\!\Longrightarrow$ ξ_n and $\mathbf{I}(\mu \ge n)$ are independent $\models\!\!\!\Longrightarrow$ $|\xi_n|$ and $\mathbf{I}(\mu \ge n)$ are independent

$$\models\!\!\!\Longrightarrow \quad \mathbf{E}|S_\mu| \le \mathbf{E}\{\sum_{n=1}^{\infty} |\xi_n| \cdot \mathbf{I}(\mu \ge n)\} = \sum_{n=1}^{\infty} \mathbf{E}\{\ldots\} =$$

$$= \sum_{n=1}^{\infty} \mathbf{E}|\xi_n| \cdot \mathbf{P}(\mu \ge n) = \mathbf{E}|\xi_1| \cdot \sum_{n=1}^{\infty} \mathbf{P}(\mu \ge n) = \mathbf{E}|\xi_1| \cdot \mathbf{E}\mu < \infty.$$

(b) Therefore,

$$\mathbf{E}S_\mu = \mathbf{E}\{\sum_{n=1}^{\infty} \xi_n \cdot \mathbf{I}(\mu \ge n)\} = \ldots = \mathbf{E}\xi_1 \cdot \mathbf{E}\mu.$$

□

Lemma 5. *Let $\{\xi_n\}_{n\ge1}$ be an i.i.d. sequence;*
μ be an ST w.r. to $\{\xi_n\}_{n\ge1}$, $\mathbf{P}(\mu < \infty) = 1$;
$\{\tilde{\xi}_i\}_{i\ge1}$ be as defined above;
$\tilde{\mu}$ be an ST w.r. to $\{\tilde{\xi}_i\}_{i\ge1}$, $\mathbf{P}(\tilde{\mu} < \infty) = 1$.

Then $\mu + \tilde{\mu}$ is a ST w.r. to $\{\xi_n\}_{n\ge1}$.

Proof.

$$\begin{aligned}\{\mu + \tilde{\mu} = k\} &= \bigcup_{l=1}^{k-1} \{\mu = l\} \cap \{\tilde{\mu} = k - l\} \\ &= \bigcup_{l=1}^{k-1} \{h_l(\xi_1, \ldots, \xi_l) = 1\} \cap \{\tilde{h}_{k-l}(\tilde{\xi}_1, \ldots, \tilde{\xi}_{k-l}) = 1\} \\ &= \bigcup_{l=1}^{k-1} \underbrace{\{h_l(\xi_1, \ldots, \xi_l) = 1\}}_{\in \boldsymbol{\mathcal{F}}_{[1,l]}} \cap \underbrace{\{\tilde{h}_{k-l}(\xi_{l+1}, \ldots, \xi_k) = 1\}}_{\in \boldsymbol{\mathcal{F}}_{[l+1,k]}}\end{aligned}$$

$$\models\!\!\Longrightarrow \quad \cap \ldots \in \boldsymbol{\mathcal{F}}_{[1,k]} \ \forall\, k$$

$$\Downarrow$$

$$\models\!\!\Longrightarrow \quad \bigcup \ldots \ \in \boldsymbol{\mathcal{F}}_{[1,k]}.$$

□

Now let us write $\xi_i^{(1)}$ instead of ξ_i
$\mu^{(1)}$ instead of μ
$\xi_i^{(2)}$ ... $\tilde{\xi}_i$
$\mu^{(2)}$ ... $\tilde{\mu}$
$\vdots$ ... $\vdots$

Lemma 6. *If $\mu^{(i)}$ is a ST w.r. to $\{\xi_i^{(j)}\}_{i\geq 1} \ \ \forall\, j = 1, \ldots, J$ and if $\{\xi_i^{(j+1)}\} = \{\tilde{\xi}_i^{(j)}\}$, then $\mu^{(1)} + \ldots + \mu^{(J)}$ is an ST w.r. to $\{\xi_i\}_{i\geq 1}$.*

Problem No 5. Prove Lemma 6.

1.8 Two-dimensional stopping times

Let $\{\xi_{n,1}\}_{n\geq 1}$ and $\{\xi_{n,2}\}_{n\geq 1}$ be two sequences of r.v.'s and $\mathcal{F}_{[k_1,n_1]\times[k_2,n_2]}$ a σ-algebra generated by

$$\xi_{k_1,1}, \xi_{k_1+1,1}, \ldots, \xi_{n_1,1}; \xi_{k_2,2}, \xi_{k_2+1,2}, \ldots, \xi_{n_2,2}.$$

Definition 7. *A pair of r.v.'s $\mu_1, \mu_2 : \mathbf{\Omega} \to \{1, 2, \ldots\}$ is an ST w.r. to $\{\xi_{n,1}\}$ and $\{\xi_{n,2}\}$, if*

$$\forall\, n_1 \geq 1,\ \forall\, n_2 \geq 1 \quad \{\mu_1 = n_1, \mu_2 = n_2\} \in \mathcal{F}_{[1,n_1]\times[1,n_2]}.$$

Lemma 7. *If $\{\xi_{n,1}\}_{n\geq 1}$ and $\{\xi_{n,2}\}_{n\geq 1}$ are two mutually independent sequences and if (μ_1, μ_2) is an ST, then*

1) *each of the sequences*

$$\{\tilde{\xi}_{i,1}\} \equiv \{\xi_{\mu_1+i,1}\} \text{ and } \{\tilde{\xi}_{i,2}\} \equiv \{\xi_{\mu_2+i,2}\}$$

is i.i.d., and these sequences are mutually independent;

2) $\tilde{\xi}_{i,1} \stackrel{\mathrm{D}}{=} \xi_{1,1}$; $\tilde{\xi}_{i,2} \stackrel{\mathrm{D}}{=} \xi_{1,2}$;

3) $\{\{\tilde{\xi}_{i,1}\}_{i\geq 1}; \{\tilde{\xi}_{i,2}\}_{i\geq 1}\}$ *and a random vector*

$$(\mu_1, \mu_2; \xi_{1,1}, \ldots, \xi_{\mu_1,1}; \xi_{1,2}, \ldots, \xi_{\mu_2,2})$$

are mutually independent.

Proof is omitted.

Lemma 8. *In conditions of Lemma 7, assume, in addition, that*

$$\xi_{1,1} \stackrel{\mathrm{D}}{=} \xi_{1,2}.$$

Then a sequence $\{\xi_n\}_{n\geq 1}$,

$$\xi_n = \begin{cases} \xi_{n,1}, & \text{if } n \leq \mu_1 \\ \xi_{n-\mu_1+\mu_2,2}, & \text{if } n > \mu_1 \end{cases}$$

is i.i.d.; $\quad \xi_n \stackrel{\mathrm{D}}{=} \xi_{1,1}$.

Proof. We have to show that $\forall\, n = 1, 2, \ldots$, $\forall\, B_1, \ldots, B_l$

$$\mathbf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \prod_{i=1}^{n} \mathbf{P}(\xi_{1,1} \in B_i).$$

1) $\forall\, n$, $\forall\, B$

$$\mathbf{P}(\xi_n \in B) = \mathbf{P}(\xi_{n,1} \in B; n \leq \mu_1) + \mathbf{P}(\xi_{n-\mu_1+\mu_2,2} \in B; n > \mu_1).$$

$$
\begin{aligned}
\mathbf{P}(\xi_{n,1} \in B; n \le \mu_1) &= \mathbf{P}(\xi_{1,1} \in B) - \mathbf{P}(\xi_{1,1} \in B) \cdot \mathbf{P}(n > \mu_1) = \\
&= \mathbf{P}(\xi_{n,1} \in B) \cdot \mathbf{P}(n \le \mu_1) \\
\mathbf{P}(\xi_{n-\mu_1+\mu_2,2} \in B; n > \mu_1) &= \sum_{l=1}^{n-1} \mathbf{P}(\xi_{\mu_2+n-l,2} \in B; \mu_1 = l) \\
&= \sum_{l=1}^{n-1} \mathbf{P}(\tilde{\xi}_{n-l,2} \in B; \mu_1 = l) \\
&= \ldots = \mathbf{P}(\xi_{1,2} \in B) \cdot \mathbf{P}(\mu_1 < n)
\end{aligned}
$$

2) Problem No 6. Prove the statement for joint distributions. Use the induction arguments. □

Here is another variant of a two-dimensional analogue of Lemma 3.

Lemma 9. *Assume that*

(i) $\vec{\xi}_n = (\xi_{n,1}, \xi_{n,2})$ *is a sequence* ($n = 1, 2, \ldots$) *of independent random vectors;*

(ii) *each of* $\{\xi_{n,1}\}_{n\ge1}$ *and* $\{\xi_{n,2}\}_{n\ge1}$ *is an i.i.d. sequence;*

(iii) $\xi_{1,1} \stackrel{\mathrm{D}}{=} \xi_{1,2}$*;*

(iv) (μ_1, μ_2) *is an ST and* $\mu_1 \equiv \mu_2 = \mu$*.*

Then

$$\xi_n = \begin{cases} \xi_{n,1}, & \text{if } n \le \mu \\ \xi_{n,2}, & \text{if } n > \mu \end{cases}$$

is an i.i.d. sequence; $\xi_n \stackrel{\mathrm{D}}{=} \xi_{1,1}$.

Proof is very similar to that of Lemma 8 (omitted).

Finally, here is a further generalization of Lemma 9.

Lemma 10. *In the statement of Lemma 9, replace*

(i) *by* (i') $= \begin{cases} \exists\, m_1 \ge 1, m_2 \ge 1\text{:} \\ \vec{\xi}_n = (\xi_{(n-1)m_1+1,1}, \ldots, \xi_{nm_1,1}; \xi_{(n-1)m_2+1,2}, \ldots, \xi_{nm_2,2}) \text{ is} \\ \text{an i.i.d. sequence;} \end{cases}$

and

(iv) *by* (iv') $= \begin{cases} (\mu_1, \mu_2) \text{ is an ST,} \\ \mathbf{P}(\mu_1 \in \{m_1, 2m_1, \ldots\}) = \mathbf{P}(\mu_2 \in \{m_2, 2m_2, \ldots\}) = 1 \\ \text{and } \dfrac{\mu_1}{m_1} \equiv \dfrac{\mu_2}{m_2}. \end{cases}$

Then

$$\xi_n = \begin{cases} \xi_{n,1}, & \text{if } n \le \mu_1 \\ \xi_{n-\mu_1+\mu_2,2}, & \text{if } n > \mu_1 \end{cases}$$

is an i.i.d. sequence; $\xi_n \stackrel{\mathrm{D}}{=} \xi_{1,1}$.

Problem No 7. Prove Lemma 10.

1.9 Stationary Sequences and Processes

Discrete Time

Definition 8. *(a) Let $\{\xi_n\}_{n\geq 0}$ be a sequence of r.v.'s. It is stationary if $\forall\ l = 1,2,\ldots,\ \ \forall\ 0 \leq i_1 < i_2 < \ldots < i_l$, $\forall\ B_1,\ldots,B_l \subseteq \boldsymbol{\mathcal{B}}$, $\forall\ m = 1,2,\ldots$*

$$\mathbf{P}(\xi_{i_1} \in B_1,\ldots,\xi_{i_l} \in B_l) = \mathbf{P}(\xi_{i_1+m} \in B_1,\ldots,\xi_{i_l+m} \in B_l). \quad (1)$$

(b) Similarly, a double-infinite sequence $\{\xi_n\}_{n=-\infty}^{\infty}$ is stationary, if (1) holds $\forall\ m \in \mathbf{Z}$ and $\forall\ B_1,\ldots,B_l \subseteq \boldsymbol{\mathcal{B}}$.

Continuous Time

Definition $\ddot{8}$. *(a) Let $\{\xi_t\}_{t\geq 0}$ be a family of r.v.'s. It is stationary, if $\forall\ l = 1,2,\ldots,\ \ \forall\ 0 \leq t_1 < t_2 < \ldots < t_l$, $\forall\ B_1,\ldots,B_l \subseteq \boldsymbol{\mathcal{B}}$, $\forall\ u \geq 0$*

$$\mathbf{P}(\xi_{t_1} \in B_1,\ldots,\xi_{t_l} \in B_l) = \mathbf{P}(\xi_{t_1+u} \in B_1,\ldots,\xi_{t_l+u} \in B_l).$$

(b) Similarly, $\{\xi_t\}_{t=-\infty}^{\infty}$ is stationary, if the above equality holds $\forall\ u \in \mathbf{R}$ and $\forall\ B_1,\ldots,B_l \subseteq \boldsymbol{\mathcal{B}}$.

Definition 9. *A sequence of events $\{A_n\}_{n=-\infty}^{\infty}$ is stationary, if a sequence of random variables $\{\mathbf{I}(A_n)\}_{n=-\infty}^{\infty}$ is stationary.*

Assume that $\{A_n\}_{n=-\infty}^{\infty}$ is a stationary sequence and that $\mathbf{P}(A_0) > 0$ and $\mathbf{P}(\cup_{n=0}^{\infty} A_n) = 1$.

Introduce the following r.v.'s:

$$\begin{aligned}
&\nu \equiv \nu^+ = \min\{n \geq 1 : \mathbf{I}(A_n) = 1\} \equiv \min\{n \geq 1 : \omega \in A_n\}\\
&\nu^- = \min\{n \geq 1 : \mathbf{I}(A_{-n}) = 1\}\\
&\tau \equiv \tau^+ : \mathbf{P}(\tau > n) = \mathbf{P}(\overline{A}_1 \ldots \overline{A}_n | A_0)\\
&\tau^- : \mathbf{P}(\tau^- > n) = \mathbf{P}(\overline{A}_{-1} \ldots \overline{A}_{-n} | A_0)
\end{aligned}$$

Lemma 11.

(a) $\nu \stackrel{\mathrm{D}}{=} \nu^-$;

(b) $\tau \stackrel{\mathrm{D}}{=} \tau^-$;

(c) $\mathbf{P}(\nu = n) = \mathbf{P}(A_0) \cdot \mathbf{P}(\tau \geq n)\ \ \forall\ n = 1,2,\ldots$

Remark 4. [*The statement of the lemma is not obvious, in general.*

Examples: Let $\{\xi_n\}$ be an i.i.d. sequence, $\mathbf{P}(\xi_n > 0) > 0$.
The we can take a) $A_n = \{\xi_n > 0\}$; b) $A_n = \{\xi_n + \xi_{n-1} > 0\}$.

Proof of Lemma 11.

(a)

$$\begin{aligned} \mathbf{P}(\nu > n) &= \mathbf{P}(\overline{A}_1 \dots \overline{A}_n) \overset{\forall m}{\downarrow}{=} \mathbf{P}(\overline{A}_{1+m} \dots \overline{A}_{n+m}) \overset{m=-n-1}{\downarrow}{=} \\ &= \mathbf{P}(\overline{A}_{-n} \dots \overline{A}_{-1}) = \mathbf{P}(\nu^- > n). \end{aligned}$$

(b)

$$\begin{aligned} \mathbf{P}(\tau = n) &= \frac{\mathbf{P}(A_0\overline{A}_1 \dots \overline{A}_{n-1}A_n)}{\mathbf{P}(A_0)} = \frac{\mathbf{P}(A_{-n}\overline{A}_{-n+1} \dots \overline{A}_{-1}A_0)}{\mathbf{P}(A_0)} = \\ &= \mathbf{P}(\tau^- = n). \end{aligned}$$

(c)

$$\begin{aligned} \mathbf{P}(\nu \geq n) &= \mathbf{P}(\overline{A}_1 \dots \overline{A}_{n-1}) = \mathbf{P}(A_0\overline{A}_1 \dots \overline{A}_{n-1}) + \mathbf{P}(\overline{A}_0\overline{A}_1 \dots \overline{A}_{n-1}) \\ &= \mathbf{P}(A_0) \cdot \mathbf{P}(\overline{A}_1 \dots \overline{A}_{n-1} | A_0) + \mathbf{P}(\overline{A}_1 \dots \overline{A}_n) = \\ &= \mathbf{P}(A_0) \cdot \mathbf{P}(\tau \geq n) + \mathbf{P}(\nu \geq n+1). \end{aligned}$$

$$\Longrightarrow \quad \mathbf{P}(\nu = n) = \mathbf{P}(\nu \geq n) - \mathbf{P}(\nu \geq n+1) = \mathbf{P}(A_0) \cdot \mathbf{P}(\tau \geq n).$$

□

Corollary 2. $\forall\, k > 0,\ \mathbf{E}\nu^k < \infty \iff \mathbf{E}\tau^{k+1} < \infty.$

Proof. Note that

$$\sum_{n=1}^{l} n^k \geq \int_0^l x^k dx = \frac{l^{k+1}}{k+1} \quad \text{and}$$

$$\sum_{n=1}^{l} n^k \leq \int_1^{l+1} x^{k+1} dx \leq \frac{(l+1)^{k+1}}{k+1} \leq 2^{k+1}\frac{l^{k+1}}{k+1}.$$

$$\begin{aligned} \Longrightarrow \quad \mathbf{E}\nu^k &= \sum_{n=1}^{\infty} n^k \mathbf{P}(\nu = n) = \mathbf{P}(A_0) \cdot \sum_{n=1}^{\infty} n^k \mathbf{P}(\tau \geq n) = \\ &= \mathbf{P}(A_0) \cdot \sum_{n=1}^{\infty} n^k \sum_{l=n}^{\infty} \mathbf{P}(\tau = l) = \mathbf{P}(A_0) \cdot \sum_{l=1}^{\infty} \mathbf{P}(\tau = l) \sum_{n=1}^{l} n^k \leq \\ &\leq \frac{\mathbf{P}(A_0)}{k+1} \cdot 2^{k+1} \cdot \sum_{l=1}^{\infty} \mathbf{P}(\tau = l) l^{k+1} = \frac{\mathbf{P}(A_0)}{k+1} \cdot 2^{k+1} \cdot \mathbf{E}\tau^{k+1} \end{aligned}$$

and, using similar arguments with the lower bound,

$$\mathbf{E}\nu^k \geq \frac{\mathbf{P}(A_0)}{k+1} \cdot \mathbf{E}\tau^{k+1}.$$

$\Longrightarrow$ $\mathbf{E}\nu^k$ and $\mathbf{E}\tau^{k+1}$ are either finite or infinite simultaneously. $\square$

1.10 On σ-algebras generated by a sequence of r.v.'s.

(1). Let $\langle \boldsymbol{\Omega}, \boldsymbol{\mathcal{F}}, \mathbf{P} \rangle$ be a probabililty space and $\xi_n : \boldsymbol{\Omega} \to \mathbf{R}$, $n = 1, 2, \ldots$ a sequence of r.v.'s. Let $\boldsymbol{\mathcal{F}}_{[k,n]} = \sigma(\xi_k, \ldots, \xi_n)$; $\quad \boldsymbol{\mathcal{F}}_{[k,\infty)} = \sigma(\xi_k, \xi_{k+1} \ldots)$.

For $A, B \in \boldsymbol{\mathcal{F}}$, introduce a distance

$$d(A, B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A).$$

(A) Recall basic properties of σ-algebras.

1) If $\boldsymbol{\mathcal{F}}^{(1)}, \boldsymbol{\mathcal{F}}^{(2)}$ are σ-algebras on $\boldsymbol{\Omega}$ $\quad \models\!\!\Rightarrow \quad$ $\boldsymbol{\mathcal{F}}^{(1)} \cap \boldsymbol{\mathcal{F}}^{(2)}$ is σ-algebra, too, but $\boldsymbol{\mathcal{F}}^{(1)} \cup \boldsymbol{\mathcal{F}}^{(2)}$ may be not, in general.

2) More generally, let T be any parameter set and $\boldsymbol{\mathcal{F}}^{(t)}, t \in T$ σ-algebras on $\boldsymbol{\Omega}$ $\models\!\!\Rightarrow \quad \cap_{t \in T} \boldsymbol{\mathcal{F}}^{(t)}$ is σ-algebra, too.

By definition, $\boldsymbol{\mathcal{F}}_{[1,\infty)}$ is a minimal σ-algebra which contains all σ-algebras $\boldsymbol{\mathcal{F}}_{[1,n]}$, $n = 1, 2, \ldots$ $\iff$ it is an intersection of all σ-algebras $\boldsymbol{\mathcal{F}}_{[1,n]}$ $n = 1, 2, \ldots$.

Since $\boldsymbol{\mathcal{F}} \supseteq \boldsymbol{\mathcal{F}}_{[1,n]} \;\forall\, n \quad \models\!\!\Rightarrow \quad \boldsymbol{\mathcal{F}}_{[1,\infty]} \subseteq \boldsymbol{\mathcal{F}}$.

(B) Now we study properties of the distance d:

(1) Clearly, $d(A, B) = d(B, A) \geq 0$;

(2) $d(A, C) \leq d(A, B) + d(B, C)$ (the triangle inequality);

Indeed, $A \setminus C = (A \setminus B) \cap (A \cap (B \setminus C)) \subseteq (A \setminus B) \cup (B \setminus C)$

$$\models\!\!\Rightarrow \quad \mathbf{P}(A \setminus C) \leq \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus C).$$

Similarly,

$$\mathbf{P}(C \setminus A) \leq \mathbf{P}(B \setminus A) + \mathbf{P}(C \setminus B).$$

(3) $d(A, B) = d(\overline{A}, \overline{B})$ (since $\mathbf{P}(A \setminus B) = \mathbf{P}(\overline{B} \setminus \overline{A})$);

(4) $|\mathbf{P}(A) - \mathbf{P}(B)| \equiv |\mathbf{P}(A \cap B) + \mathbf{P}(A \setminus B) - \mathbf{P}(A \cap B) - \mathbf{P}(B \setminus A)| \leq d(A, B)$;

(5) $d(A_1 \cup A_2, B_1 \cup B_2) \leq d(A_1, B_1) + d(A_2, B_2)$;

Indeed, $(A_1 \cup A_2) \setminus (B_1 \cup B_2) = (A_1 \setminus (B_1 \cup B_2)) \cup (A_2 \setminus (B_1 \cup B_2)) \subseteq (A_1 \setminus B_1) \cup (A_2 \setminus B_2)$

$$\models\!\!\Rightarrow \quad \mathbf{P}((A_1 \cup A_2) \setminus (B_1 \cup B_2)) \leq \mathbf{P}(A_1 \setminus B_1) + \mathbf{P}(A_2 \setminus B_2).$$

Lemma 12. $\Big[\;\; \forall A \in \boldsymbol{\mathcal{F}}_{[1,\infty)}, \quad \exists \, \{A_n\}_{n \geq 1},\, A_n \in \boldsymbol{\mathcal{F}}_{[1,n]} \;:\; d(A, A_n) \to 0.$

Proof. Let U be a set of events $A \in \boldsymbol{\mathcal{F}}$ such that $\exists \{A_n\}_{n\geq 1},\ A_n \in \boldsymbol{\mathcal{F}}_{[1,n]}\ :\ d(A, A_n) \to 0$.

1) One can easily se that $U \supseteq \boldsymbol{\mathcal{F}}_{[1,m]}\ \forall\, m = 1, 2, \ldots$.
Indeed, $\forall\, m,\ \forall\, A \in \boldsymbol{\mathcal{F}}_{[1,m]}$, let

$$A_n = \begin{cases} \emptyset, & \text{if } n < m; \\ A, & \text{if } n \geq m. \end{cases}$$

Therefore, $A \in U$.

2) Thus, it is sufficient to show that U is σ-algebra. Then, with necessity, $U \supseteq \boldsymbol{\mathcal{F}}_{[1,\infty)}$, that completes the proof.

2.1) First we prove that U is an algebra, i.e.

(i) $\boldsymbol{\Omega} \in U$;

(ii) $A \in U \ \models\!\!\!\Longrightarrow\ \overline{A} \in U$;

(iii) $\forall\, k,\ A^{(1)}, \ldots, A^{(k)} \in U \ \models\!\!\!\Longrightarrow\ A^{(1)} \cup \ldots \cup A^{(k)} \in U$.

(i) is obvious, (ii) follows from property (3), and (iii) follows from (5):

$$d(A^{(1)} \cup \ldots \cup A^{(k)}, A_n^{(1)} \cup \ldots \cup A_n^{(k)}) \leq \sum_{j=1}^{k} d(A^{(j)}, A_n^{(j)}) \to 0.$$

2.2) Now we prove that U is a σ-algebra:
(iii') $A^{(1)}, A^{(2)} \ldots \in U \ \models\!\!\!\Longrightarrow\ A \equiv \cup_{j=1}^{\infty} A^{(j)} \in U$.
Let $B^{(k)} = \cup_{j=1}^{k} A^{(j)}$ Then $B^{(k)} \nearrow A$ and $\mathbf{P}(B^{(k)}) \nearrow \mathbf{P}(A)$.

$$\models\!\!\!\Longrightarrow \quad \exists\, \{B_n^{(k)}\}\ :\ B_n^{(k)} \in \boldsymbol{\mathcal{F}}_{[1,n]},\ d(B^{(k)}, B_n^{(k)}) \to 0 \text{ as } n \to \infty.$$

Choose

$$n(1) = \min\{n \geq 1\ :\ d(B^{(1)}, B_l^{(1)}) \leq 1/2\ \forall l \geq n\}$$

and, for $k \geq 1$,

$$n(k+1) = \min\{n \geq n(k)\ :\ d(B^{(k)}, B_l^{(k)}) \leq 1/2^k\ \forall l \geq n\}.$$

Then let

$$A_n = \begin{cases} \emptyset, & \text{if } n < n(1); \\ B_{n(k)}^{(k)}, & \text{if } n(k) \leq n < n(k+1). \end{cases}$$

Clearly, $A_n \in \boldsymbol{\mathcal{F}}_{[1,n]}$. Then $d(A, A_n) \leq d(A, B^{(k)}) + 1/2^k$, for $n(k) \leq n < n(k+1)$. Since $k \to \infty$ as $n \to \infty$, $d(A, A_n) \to 0$. □

Lemma 13. *Let* $\{\xi_n\}_{n=-\infty}^{\infty}$ *be a double-infinite sequence of r.v.'s,*

$$\mathcal{F}_{(-\infty,\infty)} = \sigma\{\ldots,\xi_{-2},\xi_{-1},\xi_0,\xi_1,\xi_2,\ldots\}.$$

Then $\forall\ A \in \mathcal{F}_{(-\infty,\infty)},\ \exists\ \{A_n\}, A_n \in \mathcal{F}_{[-n,n]}\ :\ d(A, A_n) \to 0.$

Problem No 8. Prove Lemma 13.

(2). Sigma-algebras generated by sequences of independent r.v.'s.

Definition 10. *For a sequence* $\{\xi_n\}_{n\geq 1}$ *of r.v.'s, its* tail σ-algebra *is*

$$\mathcal{F}_\infty = \cap_{k=1}^{\infty}\mathcal{F}_{[k,\infty)}.$$

Note: Since $\mathcal{F}_{[k+1,\infty)} \subseteq \mathcal{F}_{[k,\infty)},\ \Longrightarrow\ \mathcal{F}_\infty = \cap_{k=l}^{\infty}\mathcal{F}_{[k,\infty)}\ \ \forall\ l.$

Definition 11. *For a sequence* $\{\xi_n\}_{n=-\infty}^{\infty}$,

$$\mathcal{F}_\infty = \cap_{k=1}^{\infty}\mathcal{F}_{[k,\infty)} \equiv \cap_{k=l}^{\infty}\mathcal{F}_{[k,\infty)},\ \forall\ -\infty < l < \infty$$

is its right *tail* σ-*algebra and*

$$\mathcal{F}_{-\infty} = \cap_{k=-0}^{\infty}\mathcal{F}_{(-\infty,k]} \equiv \cap_{k=l}^{\infty}\mathcal{F}_{(-\infty,k]},\ \forall\ -\infty < l < \infty$$

its left *tail* σ-*algebra.*

Examples...

Lemma 14. *If* $\{\xi_n\}_{n\geq 1}$ *is a sequence of* independent *r.v.'s, then* $\mathcal{F}_\infty$ *is* trivial*, i.e.*

$$\forall\ A \in \mathcal{F}_\infty,\quad \mathbf{P}(A) = 0 \vee 1.$$

Proof.

1) $A \perp \mathcal{F}_{[1,n]}\ \forall\ n;$

2) Since $\mathcal{F}_\infty \subseteq \mathcal{F}_{[1,\infty)},\ \exists\ \{A_n\} \in \mathcal{F}_{[1,n]} : d(A_n, A) \to 0.$

Therefore,

$$\mathbf{P}(A) = \mathbf{P}(A \cap A_n) + \mathbf{P}(A \setminus A_n) = \mathbf{P}(A)\cdot\mathbf{P}(A_n) + \mathbf{P}(A \setminus A_n);$$

$$0 \leq \mathbf{P}(A)[1 - \mathbf{P}(A_n)] = \mathbf{P}(A \setminus A_n) \leq d(A_n, A) \to 0.$$

□

Lemma 15. *If $\{\xi_n\}_{n=-\infty}^{\infty}$ is a sequence of independent r.v.'s, then both $\mathcal{F}_{-\infty}$ and $\mathcal{F}_{\infty}$ are trivial.*

Problem No 9. Prove Lemma 15.

(3). A stationary sequence of r.v.'s.

Definition 12. *A sequence $\{\xi_n\}_{n\geq 1}$ (or $\{\xi_n\}_{n=-\infty}^{\infty}$) is stationary, if*
$\forall\, l \geq 1,\ \forall\, 1 \leq n_1 < n_2 < \ldots < n_l$ *(or without "$1 \leq$"),*
$\forall\, k \geq 1$ *(or $\forall\ -\infty < k < \infty$),*
$\forall\, B_1, \ldots, B_l$

$$\mathbf{P}(\xi_{n_1} \in B_1, \ldots, \xi_{n_l} \in B_l) = \mathbf{P}(\xi_{n_1+k} \in B_1, \ldots, \xi_{n_l+k} \in B_l).$$

In particular, all ξ_n are identically distributed and all finite-dimensional vectors $\vec{\xi_n} = (\xi_n, \xi_{n+1}, \ldots, \xi_{n+l})$ are i.d. (for a fixed l).

Examples

1) $\{\xi_n\}$ — i.i.d.
2) $\xi_n \equiv \xi_1$
3) $\xi_{n+1} = -\xi_n,\ \xi_1 = \begin{cases} 1, & 1/2 \\ -1, & 1/2 \end{cases}$

Introduce the shift transformation θ on the set of $\mathcal{F}_{[1,\infty)}$-measurable (or $\mathcal{F}_{(-\infty,\infty)}$-measurable) r.v.'s:

1) $\theta\xi_n = \xi_{n+1}\ \ \forall\, n$

2) if $\psi = h(\xi_n, \xi_{n+1}, \ldots, \xi_{n+l})$, then $\theta\psi = h(\xi_{n+1}, \xi_{n+2}, \ldots, \xi_{n+l+1})$

3) if $\psi = h(\ldots, \xi_n, \xi_{n+1}, \ldots)$, then $\theta\psi = h(\ldots, \xi_{n+1}, \xi_{n+2}, \ldots)$.

Note: θ is measure-preserving, i.e. $\psi \stackrel{\mathrm{D}}{=} \theta\psi$.

Introduce a shift transformation θ on events from $\mathcal{F}_{[1,\infty)}$ (or from $\mathcal{F}_{(-\infty,\infty)}$):

$A \in \mathcal{F}_{[1,\infty)} \iff \mathbf{I}(A)$ is $\mathcal{F}_{[1,\infty)}-$ measurable $\iff \exists\, h : \mathbf{I}(A) = h(\ldots, \xi_n, \xi_{n+1}, \ldots)$,

h is $\{0,1\}$-valued. Then

$$\theta A = \{h(\ldots, \xi_{n+1}, \xi_{n+2}, \ldots) = 1\} \iff \theta\mathbf{I}(A) = h(\ldots, \xi_{n+1}, \xi_{n+2}, \ldots).$$

For any m, introduce $\theta^m \equiv \underbrace{\theta \cdot \ldots \cdot \theta}_{m}$.

In the case of $\mathcal{F}_{(-\infty,\infty)}$, we can introduce θ^{-m}, too. Finally, θ^0 is the identical transformation.

Definition 13. An $\mathcal{F}_{[1,\infty)}$-measurable (or $\mathcal{F}_{(-\infty,\infty)}$-measurable) r.v. ψ is invariant (w.r.to θ), if

$$\theta\psi = \psi \text{ a.s.} \quad (\text{i.e. } \mathbf{P}(\theta\psi = \psi) = 1).$$

An event $A \in \mathcal{F}_{[1,\infty)}$ (or $A \in \mathcal{F}_{(-\infty,\infty)}$) is invariant (w.r.to θ), if

$$\mathbf{P}(A \cap \theta A) = \mathbf{P}(A).$$

Note that $\theta\psi = \psi$ a.s. $\iff$ $\forall\, x$,

$$\mathbf{P}(\{\psi \le x\} \cap \{\theta\psi \le x\}) = \mathbf{P}(\psi \le x).$$

Comments, examples...

Definition 14. A stationary sequence $\{\xi_n\}$ is ergodic (w.r.to θ), if $\forall\, A \in \mathcal{F}_{[1,\infty)}$ ($A \in \mathcal{F}_{[1,\infty)}$),

$$A \text{ is invariant} \quad \models\!\!\Longrightarrow \quad \mathbf{P}(A) = 0 \vee 1$$

$$(\text{or } \psi \text{ is invariant} \quad \models\!\!\Longrightarrow \quad \psi = \text{const a.s. }).$$

Remark 5. All invariant events (sets) form a σ-algebra $\mathcal{F}^{(inv)}$ (invariant σ-algebra).

Lemma 16.

(1) $\forall\, A \in \mathcal{F}_{[1,\infty)}$ (or $\forall\, A \in \mathcal{F}_{(-\infty,\infty)}$) a sequence of events $\{\theta^n A, n \ge 0\}$ (or $\{\theta^n A, -\infty \le n \le \infty\}$) is stationary;

(2) If $\{\xi_n\}$ is stationary ergodic, then $\forall\ A \in \mathcal{F}_{[1,\infty)}$ (or $\forall\ A \in \mathcal{F}_{(-\infty,\infty)}$), $\mathbf{P}(A) > 0$

$$\models\!\!\Longrightarrow \quad \mathbf{P}(\cup_{n=l}^{\infty}\theta^n A) = 1\ \forall\, l \quad (\text{ and } \mathbf{P}(\cup_{n=l}^{-\infty}\theta^n A) = 1\ \forall\, l).$$

Proof. (1) follows from definitions.

(2) Let $B = \cup_{n=l}^{\infty}\theta^n A$. Then

$$\theta B = \cup_{n=l}^{\infty}\theta(\theta^n A) = \cup_{n=l+1}^{\infty}\theta^n A$$

and $B \supseteq \theta B$

$\models\!\!\Longrightarrow$ $\mathbf{P}(B \cap \theta B) = \mathbf{P}(\theta B) = \mathbf{P}(B)$ $\models\!\!\Longrightarrow$ B is invariant

$\models\!\!\Longrightarrow$ $\mathbf{P}(B) = 0 \vee 1$.

But $\mathbf{P}(B) \ge \mathbf{P}(\theta^l A) = \mathbf{P}(A) > 0$ $\models\!\!\Longrightarrow$ $\mathbf{P}(B) = 1$. □

Lemma 17. If A is invariant, then $\exists\, B \in \mathcal{F}_\infty$ such that $d(A, B) = 0$.

Proof. There are two cases: (a) $\mathcal{F}_{[1,\infty)}$; (b) $\mathcal{F}_{(-\infty,\infty)}$. Here we give a proof in the first case.

Problem No 10. Prove the lemma in the case (b).

1) Let $B_{0,m} = A \cap \theta A \cap \theta^2 A \cap \ldots \cap \theta^m A$, $B_0 = \cap_{n=0}^{\infty} \theta^n A$. Then

$$A = B_{0,0} \supseteq B_{0,1} \supseteq \ldots \supseteq B_{0,m} \supseteq B_{0,m+1} \supseteq \ldots \supseteq B_0$$

and $\mathbf{P}(B_{0,m}) \searrow \mathbf{P}(B_0)$. But

$$\mathbf{P}(B_{0,m}) = \mathbf{P}(A) \ \forall\ m \quad \Longmapsto \quad \mathbf{P}(B_0) = \mathbf{P}(A) \text{ and } d(B_0, A) = 0.$$

2) For $k \geq 1$, put $B_k = \theta^k B_0 \equiv \cap_{n=k}^{\infty} \theta^n A$.
Note that $B_{k+1} \supseteq B_k$ and $B_k \in \mathcal{F}_{[k,\infty)}$,

$$\mathbf{P}(B_k) = \mathbf{P}(B_0) = \mathbf{P}(A) \quad \text{and} \quad d(B_k, A) = 0.$$

Let

$$B = \lim_{k\to\infty} B_k \quad \Longmapsto \quad \mathbf{P}(B) = \mathbf{P}(A) \quad \text{and} \quad d(B, A) = 0.$$

Since $B \in \mathcal{F}_{[k,\infty)}\ \forall\ k \quad \Longmapsto \quad B \in \mathcal{F}_{\infty}$. □

Remark 6. *In the case $\mathcal{F}_{(-\infty,\infty)}$, the "symmetric" statement is true, too: if A is invariant, then $\exists\ B \in \mathcal{F}_{-\infty}$ such that $d(A, B) = 0$.*

Corollary 3. *Any i.i.d. sequence is stationary ergodic.*

Indeed, $\mathcal{F}_{\infty}$ is trivial $\Longmapsto$ if A is invariant, $B \in \mathcal{F}_{\infty}$, $\mathbf{P}(B) = 0 \vee 1$ and $d(A, B) = 0$ $\Longmapsto$ $\mathbf{P}(A) = 0 \vee 1$.

Remark 7. *There exists a number of weaker conditions that imply the "triviality" of the tail σ-algebra $\mathcal{F}_{\infty}$ and, as a corollary, the ergodicity of a stationary sequence.*

For instance, we can introduce the following "mixing" coefficients:

$$d_k = \sup_{B\in\mathcal{F}_{[k,\infty)}, A\in\mathcal{F}_{(-\infty,0]}} |\mathbf{P}(A \cap B) - \mathbf{P}(A) \cdot \mathbf{P}(B)|,$$

and then show that if $d_k \to 0$ as $k \to \infty$, then $\mathcal{F}_{\infty}$ is trivial.

In general, there are examples when $\mathcal{F}_{\infty}$ is not trivial, but $\mathcal{F}^{inv}$ is (i.e. the sequence is ergodic).

Example $\xi_{n+1} = -\xi_n\ \forall\ n$; $\xi_1 = \begin{cases} 1, & \text{w.pr. } 1/2 \\ -1, & \text{w.pr. } 1/2 \end{cases}$ Then

$$\mathcal{F}_{\infty} = \sigma(\xi_1), \quad \mathcal{F}^{inv} = \{\mathbf{\Omega}, \emptyset\}.$$