1 Introduction

1.1 Notations and Abbreviations

R.v. — random variable

i.i.d. — independent identically distributed

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

F, G — distribution function, f — density

P — probability and probability measure, **E** — expectation, **D** — variance

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

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

Standart families of distributions

$U[a, b]$ $G(p)$ $E(\alpha)$ $B(m, p)$ $N(a, \sigma^2)$ $\Pi(\lambda)$
$I(A)$ — indicator function.

Convergence

$\xi_n \xrightarrow{\text{a.s.}} \xi$ means $\mathbf{P}(\lim \xi_n = \xi) = 1$

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

Definition 1 *Weak convergence: $F_n \Rightarrow F$, 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 each g — bounded continuous,

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

I will write also: $\xi_n \Rightarrow \xi$. It means: $\xi_n \Subset F_n$, $\xi \Subset F$ and $F_n \Rightarrow F$.

$\hat{\xi}$ 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.

Coupling

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

For F_1, F_2 – d.f., their coupling is a construction of two r.v.'s $\xi_1 \Subset F_1$ and $\xi_2 \Subset F_2$ on a *common probability space*. The same — for more than two r.v.'s.

(b) Coupling of two random variables.

Let ξ_1 be defined on $\langle \mathbf{\Omega_1}, \mathcal{F}_\mathbf{1}, \mathbf{P_1} \rangle$ and ξ_2 be defined on $\langle \mathbf{\Omega_2}, \mathcal{F}_\mathbf{2}, \mathbf{P_2} \rangle$. Their coupling: $\langle \mathbf{\Omega}, \mathcal{F}, \mathbf{P} \rangle$ and $\hat{\xi}_1, \hat{\xi}_2$ on it: $\hat{\xi}_1 \stackrel{\mathrm{D}}{=} \xi_1$, $\hat{\xi}_2 \stackrel{\mathrm{D}}{=} \xi_2$.

1.2 Weak and "strong" convergence

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

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

Proof. For a d.f. F, define F^{-1}:

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

Put $\mathbf{\Omega}$ =(0,1), $\mathcal{F}$ — σ-algebra of Borel subsets in (0,1), $\mathbf{P}$ — Lebesgue measure on (0,1).

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

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

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

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

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), then 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!

And $\underline{\xi}_n \leq \xi$ a.s., that is ψ exists: $\underline{\xi}_n \nearrow \psi$ a.s., $\psi \leq \xi$ a.s.

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

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

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

□

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

1.3 Uniform integrability

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

Definition 2 *$\{\xi_n\}$ are uniformly integrable (UI), if $\mathbf{E}|\xi_n| < \infty \quad \forall n$ and, moreover,*

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

We can assume the upper bound $h(x)$ to be monotone and right-continuous.

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) $\to$ (i). For each n,

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

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

(i) $\to$ (ii). Assume that $h(x) > 0$ for all x (otherwise, the statement is trivial).

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

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

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

Note: A_m is an interval, and if z_m is its left boundary point, then $z_m \in A_m$. Therefore,

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

Remark 1 *As a corollary, one can get the following: if* $\mathbf{E}|\xi| < \infty$*, then* $\exists$ *g from Lemma 1 such that* $\mathbf{E}\{|\xi| \cdot g(|\xi|)\} < \infty$*.*

"If $\exists$ *the first moment"* $\models\!\!\Longrightarrow$ *"* $\exists$ *something more".*

Lemma 2 *Assume $\xi_n \Rightarrow \xi$. Then holds*

(1) $\{\xi_n\}$ *are UI* $\Longrightarrow$ $\mathbf{E}|\xi| < \infty$ *and* $\mathbf{E}|\xi_n| \to \mathbf{E}|\xi|$*;*

(2) $\mathbf{P}(\xi_n \geq 0) = 1 \ \forall\, n$*;* $\mathbf{E}|\xi_n| < \infty \ \forall\, n$*;* $\mathbf{E}|\xi_n| \to \mathbf{E}|\xi| < \infty \Longrightarrow \{\xi_n\}$ *are UI.*

Remark 2 *In* (2)*, the condition* $\mathbf{P}(\xi_n \geq 0) = 1$ *may be weakened in a natural way. But it cannot be eliminated.*

Example.

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

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

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 that there exists N: $\mathbf{P}(|\xi_n| \leq N) = 1$ for each n (this is a special case of UI).

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

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

Therefore, $\mathbf{E}|\xi_n| \to \mathbf{E}|\xi|$.

(b) Assume now that $\mathbf{P}(|\xi_n| \leq N) < 1$ for each N and for some n.

Since $\xi_n \stackrel{\text{a.s.}}{\to} \xi$, then $\forall x > 0$

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

Then,

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

and

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

(c) Prove: $\mathbf{E}|\xi| < \infty$. Indeed,

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

(d) $\forall\, \varepsilon > 0$, choose x: $h(x) \le \varepsilon$ and $\mathbf{E}\{|\xi| \cdot I(|\xi| \ge x)\} \le \varepsilon$.
Then,

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

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

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

□

Prove (2).

$$\mathbf{E}\xi < \infty \models\Rightarrow \ \forall\, \varepsilon > 0\ \exists x_0 = x_0(\varepsilon) : \mathbf{E}\{\xi \cdot I(\xi \ge x_0)\} \le \varepsilon/2.$$

Use (b) from the proof of (1): for a given x_0,

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

Therefore, $\exists\, n(\varepsilon)$:

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

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

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

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

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

Therefore,

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

□

1.4 Some useful properties

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

Indeed, $\forall\, x$

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

$\square$

Property 2 *If $\{\eta_n\}$ are such that there exists an i.i.d. sequence $\{\xi_n\}$:*
(a) $|\eta_n| \le |\xi_n|$ *a.s.,*
(b) $\mathbf{E}|\xi_1| < \infty$*,*
then the sequence $\{\psi_n\}$, $\psi_n = \dfrac{\eta_1 + \ldots + \eta_n}{n}$, is UI.

Indeed,

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

(i) $\mathbf{E}\phi_n = \mathbf{E}|\xi_1|\ \ \forall\, n$
(ii) SLLN:

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

$\models\!\!\Rightarrow$ From Lemma 2, (2), $\{\phi_n\}$ are UI.
$\models\!\!\Rightarrow$ From Property 1.1, $\{\psi_n\}$ are UI. $\square$

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 $|\eta_n| \le_{\text{st}} |\xi_n|$ (that means: $\mathbf{P}(|\eta_n| > x) \le \mathbf{P}(|\xi_n| > x)\ \ \forall x$). In particular, if $|\eta_n| \le_{\text{st}} |\xi|\ \ \forall n$ (ξ is the same $\forall n$) and if $\mathbf{E}|\xi| < \infty$, then $\{\eta_n\}$ are UI.*

Remark 3 *Consider, instead of a sequence $\{\xi_n\}_{n\ge1}$, a family of r.v.'s $\{\xi_t\}_{t\in T}$, where T is an arbitrary set. Then one can introduce the following*

Definition 3 (compare with Definition 1).
$\{\xi_t\}_{t\in T}$ *are UI, if* $\mathbf{E}|\xi_t| < \infty \;\; \forall t \in T$ *and, moreover,*

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

Then

(a) The statement and the proof of Lemma 1 will not change, if we replace "n" by "$t \in T$".

(b) For $T = [0, \infty)$, the statement and the proof of Lemma 2 will not change, too.

(c) Properties 1.1 and 1.3 still will be true.

1.5 Coupling inequality. Maximal coupling. Dobrushin's theorem

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

Coupling inequality

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

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

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

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

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

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

Maximal coupling

Let's reformulate the statement. Note that l.h.s. 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 given $\mathbf{P}_1$ and $\mathbf{P}_2$ and for any their coupling (*) takes place. Or, equivalently,

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

(?) May be, in $(**)$ is equality?

(??) If "yes", then does there exists such a coupling that

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

Both answers are positive! And this is the statement of **Dobrushin's theorem**.

Proof. $\mu(B) = \mathbf{P}_1(B) - \mathbf{P}_2(B)$ is a signed measure. Therefore, Banach theorem states that

there exists a subset $C \subset \mathcal{X}$:

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

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

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):

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

$$Q_{2,1} : Q_{2,1}(B) = \frac{\mathbf{P}_2(\overline{C} \cap B) - \mathbf{P}_1(\overline{C} \cap B)}{-\mu(\overline{C})},\ B \in \mathcal{B}_{\mathcal{X}}.$$

Similarly,

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

$$Q_{1,2} : Q_{1,2}(B) = \frac{\mathbf{P}_1(C \cap B) - \mathbf{P}_2(C \cap B)}{\mu(C)},\ B \in \mathcal{B}_{\mathcal{X}}.$$

Then, define 5 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 "construct" ξ_1 and ξ_2 :

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

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

Simple calculations show that $\xi_i \Subset \mathbf{P}_i$, $i = 1, 2$.

Problem No 3. "Indeed, ..."

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,

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

and the proof is completed. □

1.6 Probabilistic Metrics

The Dobrushin's theorem gives a positive solution of one of the important problems that arise in the theory of probabilistic metrics. Let us describe briefly some concepts of this theory. ($\mathcal{X}$,$\mathcal{B}_{\mathcal{X}}$)— complete separable metric space,

$\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}$,

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

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

Problem No 4. Prove: $\mathrm{diag}(\mathcal{X}^2) \in \mathcal{B}^2_{\mathcal{X}}$.

Let $\mathbf{P}$ be any probability distribution on $(\mathcal{X}^2, \mathcal{B}^2_{\mathcal{X}})$. Denote by $\mathbf{P}_1$ its first marginal distribution, and by $\mathbf{P}_2$ — second one:

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

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

Definition 4 *A function $d : \mathcal{P} \longrightarrow [0, \infty)$ is called a probabilistic metric, if it satisfies:*

(1) $\mathbf{P}(\mathrm{diag}(\mathcal{X}^2)) = 1 \ \models\!\!\Rightarrow d(\mathbf{P}) = 0$;

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

(3) $\mathbf{P}^{(1)}$ *has marginals* $\mathbf{P}_1$ *and* $\mathbf{P}_2$
$\mathbf{P}^{(2)}$ *has marginals* $\mathbf{P}_2$ *and* $\mathbf{P}_1$ $\models\!\!\Rightarrow d(\mathbf{P}^{(1)}) = d(\mathbf{P}^{(2)})$;

(4) *"triangle inequiality":*
$\mathbf{P}^{(1)}$ *has marginals* $\mathbf{P}_1$ *and* $\mathbf{P}_2$
$\mathbf{P}^{(2)}$ *has marginals* $\mathbf{P}_1$ *and* $\mathbf{P}_3$ $\models\!\!\Rightarrow d(\mathbf{P}^{(1)}) \le d(\mathbf{P}^{(2)}) + d(\mathbf{P}^{(3)})$;
$\mathbf{P}^{(3)}$ *has marginals* $\mathbf{P}_3$ *and* $\mathbf{P}_2$

Definition 5 *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 simple metric, it is natural to write $d(\mathbf{P}_1, \mathbf{P}_2)$ instead of $d(\mathbf{P})$, so d is some "distance" between $\mathbf{P}_1$ and $\mathbf{P}_2$.

For complex metric, we can 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 joint distribution $\mathbf{P}$:

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

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.

Examples

Simple	Complex
1) $\sup_{B \in \mathcal{B}} \lvert \mathbf{P}_1(B) - \mathbf{P}_2(B) \rvert$ (Total variation norm (T.V.N.))	2) $\mathbf{P}(\xi_1 \neq \xi_2) \equiv \mathbf{P}(\mathcal{X}^2 - \text{diag}(\mathcal{X}^2))$ (Indicator metric (I.M.))
For real-valued r.v.'s: 3) $\sup_x \lvert F_1(x) - F_2(x) \rvert$ (Uniform metric (U.M.))	5) $\inf\{\varepsilon > 0 : \mathbf{P}(\lvert \xi_1 - \xi_2 \rvert > \varepsilon) < \varepsilon\}$ (Ki Fan (???) metric (K.F.M.))
4) $\inf\{\varepsilon > 0 :$ $\mathcal{F}_1(x - \varepsilon) - \varepsilon \leq F_2(x) \leq F_1(x + \varepsilon) + \varepsilon\ \forall x\}$ (Levy metric (L.M.))	

One of the general problem in the theory of probabilistic metrics is:

Assume some simple metric to be given.

Does there exist a complex metric $\tilde{d}$ such that

(a) the following coupling inequality holds:

$$d(\xi_1, \xi_2) \leq \inf_{\text{all couplings}} \tilde{d}(\xi_1, \xi_2) \quad \text{(compare with } (**))$$

(b) "$\leq$" $\mapsto$ "$=$" in (a) ? (And vice versa...)

(c) $\exists$ a coupling: $d(\xi_1, \xi_2) = \tilde{d}(\xi_1, \xi_2)$?

Theorem 1 *The answer on the above question is positive for the metrics:*

$d =$ T.V.N. $\longleftrightarrow \tilde{d} =$ I.M.

$d =$ L.M. $\longleftrightarrow \tilde{d} =$ K.F.M.

1.7 Stopping times

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

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

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

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

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

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

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

where $\mathcal{B}^{n-k+1} = \sigma$-algebra of Borel sets in $\mathbf{R}^{n-k+1}$.

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

Good Property : $\forall\, A \in \mathcal{F}_{[1,\infty)}$, $\exists\, \{A_n\}_{n \geq 1}$, $A_n \in \mathcal{F}_{[1,n]}$, *such that:* $\mathbf{P}(A \setminus A_n) + \mathbf{P}(A_n \setminus A) \to 0$ as $n \to \infty$.

Let now $\mu : \mathbf{\Omega} \to \{1, 2, \ldots, n, \ldots\}$ be an integer-valued r.v.

Definition 6 *μ is called a stopping time (ST) with respect to $\{\xi_n\}$, if $\forall\, n \geq 1$,*

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

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

Another variant of definition is:

Definition 7 *μ is a ST, if $\exists$ a family of functions $h_n : \mathbf{R}^n \to \{0, 1\}$ such that:*

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

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

Examples ...

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

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

Lemma 3 1) $\{\tilde{\xi}_i\}$ *is an i.i.d. sequence;*

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

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

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

Proof of Lemma 3. We have to show that

$(*)$ $\forall\, k \geq 1$, $\forall\, m \geq 1$, $\forall$ Borel sets $B_1, \ldots, B_k$ and $C_1, \ldots, C_m$,

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

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

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

$(**)$

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

In particular, $\forall\, j \geq 1 \,\forall\, C_j$ take $m \geq j$ and $C_i = \mathbf{R}$ for $i \neq j$.

Then
the l.h.s. of $(**) = \mathbf{P}(\tilde{\xi}_j \in C_j)$,
the r.h.s. of $(**) = \mathbf{P}(\xi_1 \in C_j)$ $\models\!\!\Rightarrow$ 2)

Now, take any $C_1, \ldots, C_m$ and replace in $(**)$

$\prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i)$ by $\prod_{i=1}^{m} \mathbf{P}(\tilde{\xi}_1 \in C_i) \models\Rightarrow$ 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 $\prod_{i=1}^{m} \mathbf{P}(\tilde{\xi}_i \in C_i) \models\Rightarrow$ 3)

So, let's prove $(*)$:

$$\mathbf{P}(\{\mu = k;\, \xi_1 \in B_1, \ldots, \xi_k \in B_k\} \cap \{\tilde{\xi}_1 \in C_1, \ldots, \tilde{\xi}_m \in C_m\}) =$$

$$\mathbf{P}(\underbrace{\{h_k(\xi_1, \ldots, \xi_k) = 1;\, \xi_1 \in B_1, \ldots, \xi_k \in B_k\}}_{\in \mathcal{F}_{[1,k]}} \cap \underbrace{\{\xi_{k+1} \in C_1, \ldots, \xi_{k+m} \in C_m\}}_{\in \mathcal{F}_{[k+1,k+m]}}) =$$

$$= \mathbf{P}(\ldots) \cdot \mathbf{P}(\ldots) =$$

$$= \mathbf{P}(\ldots) \cdot \prod_{i=1}^{m} \mathbf{P}(\xi_{k+i} \in C_i) = \mathbf{P}(\ldots) \cdot \prod_{i=1}^{m} \mathbf{P}(\xi_1 \in C_i).$$

□

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 $\mathbf{E}|S_\mu| < \infty$.

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

Note, that $I(\mu \ge n) = 1 - I(\mu \le n-1)$, and $\{\mu \le n-1\} \in \mathcal{F}_{[1,n-1]}$

$\models\Rightarrow$ ξ_n and $I(\mu \ge n)$ are independent $\models\Rightarrow$ $|\xi_n|$ and $I(\mu \ge n)$ are independent

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

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

(b) Therefore,

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

□

"Induction..."

Lemma 5 *Let*

$\{\xi_n\}_{n\geq 1}$ *be an i.i.d. sequence;*

μ *be a ST w.r. to* $\{\xi_n\}_{n\geq 1}$, $\mathbf{P}(\mu < \infty) = 1$;

$\{\tilde{\xi}_i\}_{i\geq 1}$ *be as defined above;*

$\tilde{\mu}$ *be a ST w.r. to* $\{\tilde{\xi}_i\}_{i\geq 1}$, $\mathbf{P}(\tilde{\mu} < \infty) = 1$.

Then

$\mu + \tilde{\mu}$ *is a ST w.r. to* $\{\xi_n\}_{n\geq 1}$.

Proof.

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

$$\Longrightarrow \quad \cap \ldots \in \mathcal{F}_{[1,k]} \ \forall\, k \quad \Longrightarrow \quad \bigcup \ldots \in \mathcal{F}_{[1,k]}.$$

□

Let's 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 a ST w.r. to* $\{\xi_i\}_{i\geq 1}$.

Problem No 5. To prove Lemma 6.

1.8 Generalization onto 2-dimensional case

Let $\{\xi_{n,1}\}_{n\geq 1}$ and $\{\xi_{n,2}\}_{n\geq 1}$ be two sequences; $\mathcal{F}_{[k_1,n_1]\times[k_2,n_2]}$ be 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 8 *A pair of r.v.'s* $\mu_1,\mu_2:\mathbf{\Omega}\to\{1,2,\ldots\}$ *is a 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 a ST, then*

1) *each of the sequences*

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

is i.i.d., and they are mutually independent;

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

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 – omitted.

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

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

Then the 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) $\forall\, n,\ \forall\, B$

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

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

2) **Problem No 6.** To prove for joint distributions — by induction arguments. □

Another variant of generalization on 2-dimensional case.

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* $\{\xi_{n,1}\}_{n\geq 1}$ *and* $\{\xi_{n,2}\}_{n\geq 1}$ *is an i.i.d. sequence;*

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

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

Then

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

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

Proof is very similar to that of Lemma 8. — omitted.

Finally, the last generalization (of Lemma 9).

Lemma 10 *Replace in the statement of Lemma 9 (if* $\exists\, m_1 \geq 1, m_2 \geq 1$*:)*
(i) *by*

(i') $\vec{\xi}_n = (\xi_{(n-1)m_1+1,1}, \ldots, \xi_{nm_1,1}; \xi_{(n-1)m_2+1,2}, \ldots, \xi_{nm_2,2})$ *is an i.i.d. sequence;*

and
(iv) *by*

(iv') (μ_1, μ_2) *is a ST,*
$\mathbf{P}(\mu_1 \in \{m_1, 2m_1, \ldots\}) = \mathbf{P}(\mu_2 \in \{m_2, 2m_2, \ldots\}) = 1$
and $\dfrac{\mu_1}{m_1} \equiv \dfrac{\mu_2}{m_2}$.

Then

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

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

Problem No 7. Prove Lemma 10.

1.9 Stationary Sequences and Processes

Discrete Time

Definition 9 *(a) Let $\{\xi_n\}_{n \ge 0}$ be a sequence of r.v.'s.*
It is stationary, if $\forall\, l = 1, 2, \ldots,\ \forall\, 0 \le i_1 < i_2 < \ldots < i_l,\ \forall\, B_1, \ldots, B_l \subseteq \mathcal{B}$, $\forall\, m = 1, 2, \ldots$

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

(b) Similarly, $\{\xi_n\}_{n=-\infty}^{\infty}$ is stationary, if $\ldots,\ \forall\, m \in \mathbf{Z}$, the above equality holds.

Continuous Time

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

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

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

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

Assume $\{A_n\}_{n=-\infty}^{\infty}$ to be stationary, $\mathbf{P}(A_0) > 0$, $\mathbf{P}(\cup_{n=0}^{\infty} A_n) = 1$.
Introduce r.v.'s:

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

Lemma 11 (a) $\quad \nu \stackrel{\mathrm{D}}{=} \nu^-$;

(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 *It is not obvious, in general. Examples:* $\{\xi_n\}$ *— i.i.d.,* $\mathbf{P}(\xi_n > 0) > 0$.
a) $A_n = \{\xi_n > 0\}$; b) $A_n = \{\xi_n + \xi_{-n} > 0\}$.

Proof of Lemma 11.
(a)

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

(b)

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

(c)

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

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

□

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

Proof. Note:

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

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

$$\models\!\!\Longrightarrow \quad \mathbf{E}\nu^k = \sum_{n=1}^{\infty} n^k \mathbf{P}(\nu = n) = \mathbf{P}(A_0)\cdot \sum_{n=1}^{\infty} n^k \mathbf{P}(\tau \ge n) =$$

$$= \mathbf{P}(A_0)\cdot \sum_{n=1}^{\infty} n^k \sum_{l=n}^{\infty} \mathbf{P}(\tau = l) == \mathbf{P}(A_0)\cdot \sum_{l=1}^{\infty} \mathbf{P}(\tau = l) \sum_{n=1}^{l} n^k \le$$

$$\le \frac{\mathbf{P}(A_0)}{k+1}\cdot 2^{k+1}\cdot \sum_{l=1}^{\infty} \mathbf{P}(\tau = l) l^{k+1} = \frac{\mathbf{P}(A_0)}{k+1}\cdot 2^{k+1}\cdot \mathbf{E}\tau^{k+1}$$

and $\quad \mathbf{E}\nu^k \ge \dfrac{\mathbf{P}(A_0)}{k+1}\cdot \mathbf{E}\tau^{k+1}.$

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

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

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

For $A, B \in \mathcal{F}$, put

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

(A) Remind some properties of σ-algebras.

1) If $\mathcal{F}^{(1)}, \mathcal{F}^{(2)}$ are σ-algebras on $\mathbf{\Omega}$ $\models\!\!\Longrightarrow$ $\mathcal{F}^{(1)} \cap \mathcal{F}^{(2)}$ is σ-algebra, too, but $\mathcal{F}^{(1)} \cup \mathcal{F}^{(2)}$ — not! (in general).

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

Therefore, $\mathcal{F}_{[1,\infty)}$ is a minimal σ-algebra, $\supseteq \mathcal{F}_{[1,n]}$ $\forall\, n$ $\iff$ an intersection of all σ-algebras, that $\supseteq \mathcal{F}_{[1,n]}$ $\forall\, n$.

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

(B) Some properties of d:

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

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

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

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

Similarly,

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

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

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

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

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

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

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

Proof. Let U be the set of events $A \in \mathcal{F} : \quad \exists \, \{A_n\}_{n \ge 1}, \, A_n \in \mathcal{F}_{[1,n]} \; : \; d(A, A_n) \to 0$.

1) $U \supseteq \mathcal{F}_{[1,m]} \; \forall \, m = 1, 2, \ldots$. Indeed, $\forall \, m, \; \forall \, A \in \mathcal{F}_{[1,m]}$, take

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

Therefore, it is sufficient to show that U is σ-algebra. Then $U \supseteq \mathcal{F}_{[1,\infty)}$, and the proof is completed.

2) Prove that U is an algebra, i.e.

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

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

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

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

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

3) Prove that U is a σ-algebra:
(iii') $A^{(1)}, A^{(2)} \ldots \in U \implies A \equiv \cup_{j=1}^{\infty} A^{(j)} \in U$.
Put $B^{(k)} = \cup_{j=1}^{k} A^{(j)}$, $B^{(k)} \nearrow A$ and $\mathbf{P}(B^{(k)}) \nearrow \mathbf{P}(A)$.

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

Choose

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

and, for $k \ge 1$,

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

Finally, put

$$A_n = \begin{cases} \emptyset, & \text{if } n < n(1); \\ B_{n(k)}^{(k)}, & \text{if } n(k) \ge n < n(k+1), \end{cases} \quad A_n \in \mathcal{F}_{[1,n]}.$$

Then $d(A, A_n) \le d(A, B^{(k)}) + 1/2^k$, for $n(k) \ge n < n(k+1)$. 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. Proof — for you!!!

(2). A sequence of independent r.v.'s.

Definition 10 *For a sequence $\{\xi_n\}_{n \ge 1}$, the tail σ-algebra is*

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

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

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 right tail σ-algebra and

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

is 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. Proof — for you!!!

(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 a 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: $\psi \overset{\mathrm{D}}{=} \theta\psi$.

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

$$A \in \mathcal{F}_{[1,\infty)} \quad \Longleftrightarrow \quad I(A) \text{ is } \mathcal{F}_{[1,\infty)}\text{– measurable} \quad \Longleftrightarrow \quad \exists\, h : I(A) = h(\ldots, \xi_n, \xi_{n+1}, \ldots),$$

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

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

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. And θ^0 — identical transformation.

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

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

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 $\forall\ A \in \mathcal{F}_{[1,\infty)}$ ($A \in \mathcal{F}_{[1,\infty)}$),*

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

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

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 $\forall\ A \in \mathcal{F}_{(-\infty,\infty)}$) the 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 egrodic, then $\forall\ A \in \mathcal{F}_{[1,\infty)}$ (or $\forall\ A \in \mathcal{F}_{(-\infty,\infty)}$), $\mathbf{P}(A) > 0$*

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

Proof. (1) follows from definitions.

(2) Set $B = \cup_{n=l}^{\infty}\theta^n A$, then

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

and $B \supseteq \theta B$

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

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

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

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

Proof. (a) The case $\mathcal{F}_{[1,\infty)}$; (b) the case $\mathcal{F}_{(-\infty,\infty)}$.

Problem No 10. Proof at (b) — for you!!

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

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

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

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

Set

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

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

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

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

For instance, introduce the following "mixing" coefficients:

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

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

But, 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

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

Next Part