1. Axiomatic definition of probability

1.1. **Probability space.** Let $\Omega \neq \emptyset$, and $\mathcal{A} \subseteq 2^{\Omega}$ be a σ-algebra on Ω, and P be a measure on $\mathcal{A}$ with $P(\Omega) = 1$, i.e. P is a probability measure. Then the triplet $(\Omega, \mathcal{A}, P)$ is called a ***probability space***.

Terminology

- $\Omega \ldots$ sure event
- $\omega \in \Omega \ldots$ elementary event
- $A \in \mathcal{A} \ldots$ random event
- $P(A) \ldots$ probability of A
- $A \cap B = \emptyset \ldots$ the events A, B are incompatible (disjoint)

Examples

(1) $\Omega \neq \emptyset$ is at most countable set, and we require that $\{\omega\} \in \mathcal{A}$ whenever $\omega \in \Omega$. In order to $\mathcal{A}$ satisfy the axioms of σ-algebra, we have to put $\mathcal{A} = 2^{\Omega}$. If P is a probability on $(\Omega, \mathcal{A})$, it has to be of the following form

$$P(A) = \sum_{\omega \in A} p_\omega, \quad \text{where} \quad p_\omega \geq 0 \quad \text{and} \quad \sum_{\omega \in \Omega} p_\omega = 1.$$

In this case $(\Omega, \mathcal{A}, P)$ is called a ***discrete probability space***.

(2) $\Omega \in \mathcal{B}(\mathbb{R}^k)$ is uncountable and even of positive k-dimensional Lebesgue measure, where $k \in \mathbb{N}$. Further, we put $\mathcal{A} = \{B \in \mathcal{B}(\mathbb{R}^k), B \subseteq \Omega\}$ and

$$P(A) = \int_A f(\omega)\,\mathrm{d}\omega, \quad \text{where} \quad f \geq 0 \quad \text{and} \quad \int_\Omega f(\omega)\,\mathrm{d}\omega = 1.$$

Then f is called a density and $(\Omega, \mathcal{A}, P)$ a ***continuous probability space***.

1.2. **Random variable.** Let $\Omega \neq \emptyset$, let $(\Omega, \mathcal{A}), (E, \mathcal{E})$ be measurable spaces. We say that $X : \Omega \to E$ is a ***random variable*** if X is measurable, i.e. $X^{-1}B = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{A}$ holds whenever $B \in \mathcal{E}$. Moreover $(E, \mathcal{E})$ is called a ***state space*** and $(\Omega, \mathcal{A})$ an ***underlying space***.

Remark Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ a random variable. Then $(E, \mathcal{E}, P_X)$ is a probability space, where $P_X(B) = P(X \in B)$ if $B \in \mathcal{E}$. We also write $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_X)$. The probability measure P_X is called a ***distribution of random variable*** X.

Proof: $P_X(E) = P(X \in E) = 1$, and if $B_n \in \mathcal{E}$ are pairwise disjoint, then the measure of the countable union $\cup_n B_n$ is $P_X(\cup_n B_n) = P(X \in \cup_n B_n) = P(\cup_n [X \in B_n]) = \sum_n P(X \in B_n) = \sum_n P_X(B_n)$. $\square$

If $(E, \mathcal{E}, P_X)$ is a discrete probability space, then we say that X is a ***discrete random variable***. If $(E, \mathcal{E}, P_X)$ is a continuous probability space, we say that X is a ***continuous random variable***.

Theorem 1 Let $(E, \mathcal{E}, \mu)$ be a probability space. Then there exists a random variable $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{E}, \mu)$, i.e. the random variable X has the distribution $P_X = \mu$.

Proof: Put $\Omega = E, \mathcal{A} = \mathcal{E}, P = \mu$ and $X : e \in E \mapsto e \in E$, i.e. X is indentity. Then we have that $P_X(B) = P(X \in B) = P(B) = \mu(B)$, whenever $B \in \mathcal{E}$, i.e. $P_X = \mu$. $\square$

The random variable X from the proof of theorem 1 is called a ***canonical random variable***, and $(\Omega, \mathcal{A}, P)$ from the proof is called a ***canonical probability space*** in the circumstances of theorem 1.

1.3. **Random variables with values in product spaces.** A random variable $X = (X_1, \ldots, X_k)^{\intercal}$ with values in the product space $(\prod_{n=1}^k E_n, \otimes_{n=1}^k \mathcal{E}_n)$ is called a ***k-dimensional random vector***. Moreover if $(E_n, \mathcal{E}_n) = (E, \mathcal{E})$, then we say that X is a ***k-dimensional random vector with the state space*** $(\boldsymbol{E}, \boldsymbol{\mathcal{E}})$. Finally, if $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, then X is called a ***k-dimensional real-valued random vector***.

A random variable $X = (X_n, n \in \mathbb{N})$ with values in the product space $(\prod_{n=1}^\infty E_n, \otimes_{n=1}^\infty \mathcal{E}_n)$ is called a ***random sequence***. Moreover, if $(E_n, \mathcal{E}_n) = (E, \mathcal{E})$ holds for every $n \in \mathbb{N}$, then we say that X is a ***random sequence with values in*** $(\boldsymbol{E}, \boldsymbol{\mathcal{E}})$. Finally, if $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, we say that X is a ***real-valued random sequence***.

A random variable $X = (X_t, t \in T)$ with values in the product space $(\prod_{t\in T} E_t, \otimes_{t\in T}\mathcal{E}_t)$ is called a ***random process (indexed by T)***. Moreover, if $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$ holds for every $t \in T$, then we say that X is a ***random process with values in $(E, \mathcal{E})$ (indexed by T)***. Finally, if $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, then we say that X is a ***real-valued random process (indexed by T)***.

Remark We remind the definition of the product σ-algebra

$$\bigotimes_{t\in T} \mathcal{E}_t = \sigma(\{\prod_{t\in T} B_t : B_t \in \mathcal{E},\text{ and } B_t \neq E_t \text{ holds only for finitely many } t \in T\}).$$

The set $B = \prod_{t\in T} B_t$ is called a ***measurable cylinder (with finite dimensional base***, where $B_t \in E_t$ holds for every $t \in T$, and where $T_0 := \{t \in T, B_t \neq E_t\}$ is finite set (of important indices corresponding to the measurable cylinder B). We note that the set of all measurable cylinders are closed under finite intersections.

Theorem 2 Let $(\Omega, \mathcal{A}, P)$ be a probability space and $T \neq 0$. Let $X_t : (\Omega, \mathcal{A}) \to (E_t, \mathcal{E}_t)$ be a random variable whenever $t \in T$. Then $X = (X_t, t \in T)$ is a random process indexed by T.

Proof: Denote $\mathcal{E} = \otimes_{t\in T}\mathcal{E}_t$ and $\mathcal{M} = \{B \in \mathcal{E} : [X \in B] \in \mathcal{A}\} \subseteq \mathcal{E}$. Then $\mathcal{M}$ is a σ-algebra, and we will show that it contains measurable cylinders. Let $B = \prod_{t\in T} B_t$ be a cylinder with $T_0 = \{t \in T, B_t \neq E_t\}$. Then

$$[X \in B] = [(X_t, t \in T) \in \prod_{t\in T} B_t] = \bigcap_{t\in T}[X_t \in B_t] = \bigcap_{t\in T_0}[X_t \in B_t] \in \mathcal{A}, \qquad \text{i.e.} \quad B \in \mathcal{M}.$$

Then we get that $\mathcal{M} = \mathcal{E} = \otimes_{t\in T}\mathcal{E}_t$, which means that X is a random process indexed by T. □

Corollary Let $X_n : (\Omega, \mathcal{A}) \to (E_n, \mathcal{E}_n)$ holds for $n \in T$.

(1) If $T = \{1, \dots, k\}$, then $X = (X_1, \dots, X_k)^\intercal$ is a k-dimensional random vector.
(2) If $T = \mathbb{N}$, then $X = (X_n, n \in \mathbb{N})$ is a random sequence.

Lemma A Let $(E_n, d_n), n \in N$ be at most countable system of separable metric spaces, where $N \subseteq \mathbb{N}$, with Borel σ-algebra $\mathcal{B}(E_n) = \mathcal{E}_n$. Then their product (E, d) is again a separable metric space with Borel σ-algebra $\mathcal{B}(E) = \otimes_{n\in N}\mathcal{B}(E_n)$.

Remark The separability assumption in the previous lemma is essential. Generally, we have only

$$\bigotimes_{t\in T} \mathcal{B}(E_t) \subseteq \mathcal{B}(\prod_{t\in T} E_t).$$

(1) If $N = \{1, \dots, k\}$, then d from the previous lemma can be of the form $d(x, y) = \sum_{n=1}^{k} d_n(x_n, y_n)$.
(2) If $N = \mathbb{N}$, then we can consider $d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min\{1, d_n(x_n, y_n)\}$. Note that $\min\{1, d_n(x_n, y_n)\}$ plays the role of metric on E equivalent with d_n, but bounded by 1. Then the weights 2^{-n} ensure the convergence of the corresponding sum.

Theorem 3 Let $(\Omega, \mathcal{A})$ be a measurable space and $(E_n, d_n), n \in N \subseteq \mathbb{N}$ be at most countable system of metric spaces, and let $X_n : (\Omega, \mathcal{A}) \to (E_n, \mathcal{B}(E_n))$ be random variables. Then

$$X = (X_n, n \in N) : (\Omega, \mathcal{A}) \to (E, \mathcal{B}(E))$$

is also a (measurable) random variable, where $E = \prod_{n\in N} E_n$.

Proof: By the previous lemma $\mathcal{B}(E) = \otimes_{n\in N}\mathcal{B}(E_n)$. So, we have to show that X is a random process indexed by N, but it follows from theorem 2. □

Let $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ be a random variable. Then the σ-algebra $\sigma(X) = \{[X \in B] : B \in \mathcal{E}\}$ is called a ***σ-algebra generated by the random variable*** X.

Theorem 4 Let $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ be a random variable and $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a real-valued random variable such that $\sigma(Y) \subseteq \sigma(X)$. Then there exists a measurable function $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $Y = f(X)$.

Proof: If $q \in \mathbb{Q}$, then $[Y < q] \in \sigma(Y) \subseteq \sigma(X) = \{[X \in B] : B \in \mathcal{E}\}$. Hence, there exists $B_q \in \mathcal{E}$ such that $[Y < q] = [X \in B_q]$. Put

$$h(x) = \inf\{q \in \mathbb{Q} : x \in B_q\} \quad \text{and} \quad f(x) = h(x) \cdot 1_{[h(x) \in \mathbb{R}]}.$$

Then

$$h(X) = \inf\{q \in \mathbb{Q} : X \in B_q\} = \inf\{q \in \mathbb{Q} : Y < q\} = Y \in \mathbb{R},$$

and therefore also $f(X) = Y$. Thus, it remains to show measurability of f. Instead, we show that h is measurable. Let $c \in \mathbb{R}$, then

$$[h < c] = \{x \in E : \inf\{q \in \mathbb{Q} : x \in B_q\} < c\} = \bigcup_{c > q \in \mathbb{Q}} \{x \in E : x \in B_q\} = \bigcup_{c > q \in \mathbb{Q}} B_q \in \mathcal{E}$$

and then the measurability of the real-valued function f follows from the measurability of a generalized function h that may attain non-real values $\pm\infty$. $\square$

Remark Let $X = (X_t, t \in T)$ be a random process, then $\sigma(X) = \sigma(\cup_{t \in T} \sigma(X))$.
In particular, $\sigma(X) = \sigma(\sigma(X_1) \cup \ldots \cup \sigma(X_k))$ holds if $T = \{1, \ldots k\}$ and we have a similar equality for random sequence, when $T = \mathbb{N}$.

Proof: Immediately, we have that $\sigma(X_t) \subseteq \sigma(X)$ holds for every $t \in T$, which yields that $\sigma(X) \supseteq \sigma(\cup_{t \in T} \sigma(X_t))$. Hence, we will show the reverse inclusion in the following. Let $B \in \otimes_{t \in T} \mathcal{E}_t$ be a measurable cylinder, then there exist $B_t \in \mathcal{E}_t$ such that the set $T_0 = \{t \in T : B_t \neq E_t\}$ is finite and such that $B = \prod_{t \in T} B_t$. Then

$$[X \in B] = \bigcap_{t \in T_0} [X_t \in B_t] \in \sigma(\cup_{t \in T} \sigma(X_t)).$$

Then set of all measurable cylinders is a subset of the σ-algebra

$$\mathcal{M} = \{B \in \bigotimes_{t \in T} \mathcal{E}_t : [X \in B] \in \sigma(\cup_{t \in T} \sigma(X_t))\},$$

which is generated by the set of measurable cylinders, and so we get that $\mathcal{M} = \otimes_{t \in T} \mathcal{E}_t$. Then we obtain from the definition that $\sigma(X) = \{[X \in B] : B \in \otimes_{t \in T} \mathcal{E}_t\} \subseteq \sigma(\cup_{t \in T} \sigma(X_t))$. $\square$

Remark Every σ-algebra $\mathcal{F}$ is generated by a random variable, namely by a real-valued process indexed by $\mathcal{F}$ in the form $\mathbf{1}_{\mathcal{F}} = (1_F : F \in \mathcal{F})$, where $1_F(\omega) = 1$ if $\omega \in F$ and $1_F(\omega) = 0$ if $\omega \notin F$.

Examples

(1) Note that $\sigma(1_F) = \{\emptyset, F, \Omega \backslash F, \Omega\}$ holds in the previous remark if $F \in \mathcal{F} \subseteq 2^\Omega$.

(2) If random variable $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ is constant x, then $\sigma(X) = \{\emptyset, \Omega\}$, and this set is called a ***trivial σ-algebra*** on Ω. If P is a probability on $(\Omega, \mathcal{A})$, then

$$P_X(B) = 1_{[x \in B]} = 1_B(x) =: \delta_x(B),$$

where δ_x defined above is called a ***Dirac measure*** at point x.

(3) Let $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ attain values in $\{x_n : n \in N\} = \text{range} X$, where $N \subseteq \mathbb{N}$. Then

$$\sigma(X) = \sigma(\{[X = x_n], n \in N\}) = \{[X \in M] : M \subseteq N\}.$$

Let P be a probability measure on $(\Omega, \mathcal{A})$ and denote $p_n = P(X = n)$. Then the distribution P_X of X under P is given by the following formula

$$P_X(B) = \sum_{n:\, x_n \in B} p_n = \sum_{n \in N} p_n 1_{[x_n \in B]} = \sum_{n \in N} p_n 1_B(x_n) = \sum_{n \in N} p_n \delta_{x_n}(B),$$

and it is an (infinite) convex combination of Dirac measures δ_{x_n} at points x_n with weights p_n.

Remark Let P be a probability measure on $(\Omega, \mathcal{A})$ and $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. Then $(\Omega, \sigma(X), P|\sigma(X))$ is a probability space.

2. Distribution functions

(1) **One-dimensional case.** A real-valued function F defined on $\mathbb{R}$ is called a ***distribution function*** if it is non-decreasing left-continuous with $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Let μ be a Borel probability measure on $\mathbb{R}$, then $F(x) = \mu(-\infty, x)$ is a distribution function, and if X is a random variable with the distribution $P_X = \mu$, then $F(x) = P(X < x)$ is a distribution function, and it is called a ***distribution function of variable*** X and it denoted by $F_X(x)$. Then we have that $P(a \le X < b) = F_X(b) - F_X(a)$ holds if $a \le b$.

Examples

(a) Let $X \equiv x$ be a constant variable. Then $P_X = \delta_x$ and $F_X(y) = \delta_x(-\infty, y) = 1_{[x<y]}$ is the distribution function corresponding to Dirac measure δ_x at point x.

(b) Let X be a discrete random variable with values in $\{x_n, n \in N\}$, where $N \subseteq \mathbb{N}$. Then

$$F_X(x) = \sum_{n:\, x_n < x} p_n = \sum_n p_n 1_{[x_n < x]},$$

where $p_n = P(X = x_n)$, and F_X can be regarded as an (infinite) convex combination of distribution functions corresponding to Dirac measures δ_{x_n} at points x_n.

(2) **Multidimensional case.** Let $k \in \mathbb{N}$ and $x, y \in \mathbb{R}^k$, we write $x \le y$ if $x_n \le y_n$ holds for every $n \le k$. Similarly, we write $x < y$ if $x_n < y_n$ holds for every $n \le k$, and finally we write $x = y$ if $x_n = y_n$ holds for every $n \le k$.

Let $X = (X_1, \dots, X_k)^\intercal$ be a k-dimensional real-valued random vector[1]. Then the following function

$$F_X(x) = P(X < x) = P_X(-\infty, x), \quad x \in \mathbb{R}^k$$

is called a ***distribution function of random vector*** X, where $x = (x_1, \dots, x_k)^\intercal$ and $(-\infty, x) = \prod_{n=1}^k (-\infty, x_n)$.

Theorem 5 Let $k \in \mathbb{N}$. Then $F : \mathbb{R}^k \to [0, 1]$ is a distribution function of a random vector if and only if it satisfies the following conditions

(1) $\forall\, x \in \mathbb{R}^k \quad \exists\, \{x^n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}^k$ such that $x^n < x \quad \& \quad F(x^n) \to F(x)$ as $n \to \infty$.

(2) $\exists\, \{x^n\}_{n\in\mathbb{N}} \subseteq \mathbb{R}^k$ such that $x^n \to (\infty, \dots, \infty) \quad \& \quad F(x^n) \to 1$ as $n \to \infty$.

(3) $\forall\, n \in \{1, \dots, k\} \quad \forall\, x \in \mathbb{R}^k \quad F(x_1, \dots, x_{n-1}, y, x_{n+1}, \dots, x_k) \to 0$ as $y \to -\infty$.

(4) $\forall\, x, y \in \mathbb{R}^k$

$$x < y \quad \Rightarrow \quad \sum_{\delta \in \prod_{j=1}^k \{x_j, y_j\}} (-1)^{\sum_{j=1}^k 1_{[\delta_j = x_j]}}\; F(\delta) \ge 0$$

Remark

(1) If $k = 1$, then the condition (4) says that F is non-decreasing, and if $k = 1$ and this condition is satisfied, then (1) corresponds to the left continuity of F, (2) corresponds to $\lim_{x\to\infty} F(x) = 1$ and (3) to $\lim_{x\to-\infty} F(x) = 0$.

(2) If $k = 2$, then (4) is of the form: if $x < y$ then $F(x) + F(y) - F(x_1, y_2) - F(y_{1,2}) \ge 0$. If $F = F_X$, then the left-hand side is just $P(x \le X < Y) = P_X([x_1, y_1) \times [x_2, y_2))$ and it is natural that this value should be non-negative.

Lemma B Let $S \neq \emptyset, \mathcal{S} \subseteq \mathcal{M} \subseteq 2^S$. If $\mathcal{S}$ is closed under finite intersections and $\mathcal{M}$ is a Dynkin system, i.e.

(1) $S \in \mathcal{M}$

(2) $[A, B \in \mathcal{M} \;\&\; B \supseteq A] \;\Rightarrow\; B \backslash A \in \mathcal{M}$

(3) if $B_n \in \mathcal{M}$ are pairwise disjoint, then the countable union $\cup_n B_n \in \mathcal{M}$.

Then $\sigma(\mathcal{S}) \subseteq \mathcal{M}$.

Lemma C Let $(E, \mathcal{E})$ be a measurable space and μ, ν two probability measures on $(E, \mathcal{E})$ that agree on a system $\mathcal{S} \subseteq \mathcal{E}$ closed under finite intersection such that $\sigma(\mathcal{S}) = \mathcal{E}$. Then $\mu = \nu$.

[1] We know that X is a random variable with values in $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.

Proof: Denote $\mathcal{M} = \{B \in \mathcal{E} : \mu(B) = \nu(B)\}$ and realize that it is a Dynkin system in order to obtain from Dynkin lemma that $\mathcal{E} = \sigma(\mathcal{S}) \subseteq \mathcal{M} \subseteq \mathcal{E}$. □

Corollary Let X, Y be two k-dimensional real valued random vectors with the same distribution function $F_X = F_Y$. Then they have the same distribution $P_X = P_Y$.

Proof: Denote $\mathcal{S} = \{(-\infty, x), x \in \mathbb{R}^k\}$. Then $\mathcal{S}$ is a system closed under finite intersection generating Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ and $P_X(-\infty, x) = F_X(x) = F_Y(x) = P_Y(-\infty, x)$ hold whenever $\in \mathbb{R}^k$, i.e. $P_X = P_Y$ holds on $\mathcal{S}$. By the previous lemma, $P_X = P_Y$ holds on $\mathcal{B}(\mathbb{R}^k)$. □

Let $X = (X_t, t \in T)$ be a random process with the index set T and $T_0 \subseteq T$ a subset. Then we denote by $\boldsymbol{X|_{T_0} = (X_t, t \in T_0)}$ its ***restriction to the index set*** $\boldsymbol{T_0}$. Further, we denote by $\boldsymbol{\mathcal{K}(T)}$ the ***system of all finite subsets of*** $\boldsymbol{T}$. [2] By the ***system of finite-demiensional distributions*** of the proces X we mean the system of distributions $\boldsymbol{(P_{X|_{T_0}} : T_0 \in \mathcal{K}(T))}$.

Corollary Let X, Y be two random processes indexed by T with the same finite-dimensional distributions, then they have the same distribution $P_X = P_Y$.

Proof: Let us denote by $(E_t, \mathcal{E}_t)$ the state space of random variables X_t, Y_t. Note that they have to have the same state space as they have by assumption the same distribution. Let $\mathcal{S}$ be the system of all measurable cylinders generating the product σ-algebra $\otimes_{t\in T}\mathcal{E}_t$. By the previous lemma, we are only to show that P_X, P_Y agree on $\mathcal{S}$. Let $B \in \mathcal{S}$, then there are $B_t \in \mathcal{E}_t$ if $t \in T$ such that $B = \prod_{t\in T} B_t$ and that $T_0 = \{t \in T, B_t \neq E_t\}$ is finite. Then

$$P_X(B) = P(\bigcap_{t\in T}[X_t \in B_t]) = P(\bigcap_{t\in T_0}[X_t \in B_t]) = P_{(X_t, t\in T_0)}(\prod_{t\in T_0} B_t) = P_{(Y_t, t\in T_0)}(\prod_{t\in T_0} B_t) = \ldots = P_Y(B).$$

Thus, $P_X = P_Y$ holds on $\mathcal{S}$ and the previous lemma gives $P_X = P_Y$. □

3. Independence

3.1. **Independence of random variables.** Let $T \neq \emptyset$ and $X_t : (\Omega, \mathcal{A}, P) \to (E_t, \mathcal{E}_t, P_{X_t})$ be random variable if $t \in T$. We say that the ***random variables*** $X_t, t \in T$ are ***independent*** if for every finite subset $T_0 \subseteq T$

$$P(\bigcap_{t\in T_0}[X \in B_t]) = \prod_{t\in T_0} P(X_t \in B_t)$$

holds whenever $B_t \in \mathcal{E}_t, t \in T_0$

Theorem 6 Let $T \neq \emptyset$ and $(X_t, t \in T)$ be a random process, then the variables $X_t, t \in T$ are independent if and only if $P_X = \otimes_{t\in T} P_{X_t}$.

Proof: Let $X_t, t \in T$ be independent random variables and let $B = \prod_{t\in T} B_t$ be a measurable cylinder with the finite set T_0 of important indices. Then

$$P_X(B) = P(\bigcap_{t\in T}[X_t \in B_t]) = P\bigcap_{t\in T_0}[X_t \in B_t]) = \prod_{t\in T_0} P(X_t \in B_t) = \prod_{t\in T_0} P_{X_t}(B_t) = [\bigotimes_{t\in T} P_{X_t}](B).$$

It means that P_X and $\otimes_{t\in T} P_{X_t}$ agree on the system of measurable cylinders closed under intersections and generating the product σ-algebra. By lemma from the previous section, $P_X = \otimes_{t\in T} P_{X_t}$ holds.

Now assume that $P_X = \otimes_{t\in T} P_{X_t}$ holds, we are going to show that the random variables $X_t, t \in T$ are independent. Let T_0 be a finite subset of T and let $B_t \in \mathcal{E}_t$ if $t \in T_0$. We put $B_t = E_t$ if $t \in T \backslash T_0$ in order to be able to introduce a measurable cylinder $B = \prod_{t\in T} B_t$. By the definition of the product measure, we get that

$$P(\bigcap_{t\in T_0}[X_t \in B_t]) = P_X(B) = [\bigotimes_{t\in T} P_{X_t}](B) = \prod_{t\in T_0} P_{X_t}(B_t) = \prod_{t\in T_0} P(X_t \in B_t).$$

Thus, the variables $X_t, t \in T$ are independent. □

Corollary Let X_t, Y_t be random variables with the same state space $(E_t, \mathcal{E}_t)$ whenever $t \in T$. Let

(1) $X_t, t \in T$ be independent variables

[2] If T is a topological or metric space, it is usual to denote by $\mathcal{K}(T)$ the set of all kompact sets in T. Here, we can imagine that T is endowed with the discrete topology so that $\mathcal{K}(T)$ is just the set of all finite subsets of T. Further note that compact sets share some properties of finite set and that this was the motivation for introducing the notion of compact sets.

(2) $Y_t, t \in T$ be also independent variables.

If $P_{X_t} = P_{Y_t}$ holds for every $t \in T$, then $P_X = P_Y$ holds.

Proof: By theorem 6, $P_X = \otimes_{t\in T} P_{X_t} = \otimes_{t\in T} P_{Y_t} = P_Y$. $\square$

Remark The assumption of independence cannot be omitted. Let Z be a random variable with uniform distribution on $(0, 1)$, i.e. a continuous random variable with the density $f(z) = 1_{(0,1)}(z)$. Then $X = (Z, Z)^{\intercal}$ and $Y = (Z, 1 - Z)^{\intercal}$ are two random vectors with completely different distribution but with the same marginal distributions $P_{X_1} = P_{Y_1} = P_Z = P_{X_2} = P_{Y_2}$.

Theorem 7 Let $X = (X_1, \ldots, X_k)^{\intercal}$ be a k-dimension real-valued random vector, where $k \in \mathbb{N}$. Then the variables $X_1, \ldots, X_k$ are independent if and only if

$$F_X(x) = \prod_{n=1}^{k} F_{X_n}(x_n) \tag{1}$$

holds for every $x = (x_1, \ldots, x_k)^{\intercal} \in \mathbb{R}^k$.

Proof: Let us assume that (1) holds. Then

$$P_X(-\infty, x) = F_X(x) = \prod_{n=1}^{k} F_{X_n}(x_n) = \prod_{n=1}^{k} P_{X_n}(-\infty, x_n) = [\bigotimes_{n=1}^{k} P_{X_n}](-\infty, x).$$

Thus, P_X and $\otimes_{n=1}^{k} P_{X_n}$ agree on $\{(-\infty, x), x \in \mathbb{R}^k\}$ which is a system closed under finite intersections generating $\mathcal{B}(\mathbb{R}^k)$. By lemma from the previous section, $P_X = \otimes_{n=1}^{k} P_{X_n}$, and theorem 6 gives that $X_1, \ldots, X_k$ are independent variables. On the other hand, if $X_1, \ldots, X_k$ are independent, we put $B = (-\infty, x) = \prod_{n=1}^{k} B_n$, where $B_n = (-\infty, x_n)$, and then

$$F_X(x) = P(X \in B) = \prod_{n=1}^{k} P(X_n \in B_n) = \prod_{n=1}^{k} F_{X_n}(x_n)$$

holds for arbitrary $x = (x_1, \ldots, x_k)^{\intercal} \in \mathbb{R}^k$. $\square$

Theorem 8 Let $X = (X_1, \ldots, X_k)^{\intercal}$ be a k-dimensional

(1) discrete random vector with values in at most countable set $D = \prod_{n=1}^{k} D_n$. Then the random variables X_n are also discrete distribution with values in D_n with

$$P(X_n = x_n) = \sum_{x_j, j\neq n} P(X_1 = x_1, \ldots, X_k = x_k).$$

Further, the variables $X_1, \ldots, X_k$ are independent if and only if

$$P(X_1 = x_1, \ldots, X_k = x_k) = \prod_{n=1}^{k} P(X_n = x_n)$$

holds for every $x = (x_1, \ldots x_k) \in D$.

(2) continuous real-valued random vector with the density $f_X(x)$. Then the variables X_n are again continuous with the density

$$f_{X_n}(x_n) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(x_1, \ldots, x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}\, \mathrm{d}x_{n+1} \cdots \mathrm{d}x_k$$

and they are independent if and only if

$$f_X(x_1, \ldots, x_k) = \prod_{n=1}^{k} f_{X_n}(x_n) \quad \text{holds for a.e.} \quad x \in \mathbb{R}^k.$$

Proof: The first equality in the statement follows from the theorem on full probability. If the discrete variables are independent, we put $B_n = [X_n = x_n]$ and the definition of independence gives the second equality in the statement. The reverse implication is left to the reader.

It follows from Fubini theorem that

$$F_{X_n}(x_n) = \lim_{x_j\to\infty, j\neq n} F_X(x) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(y_1,\dots,y_k) 1_{[y_n<x_n]} \, \mathrm{d}y_1 \cdots \mathrm{d}y_k$$
$$= \int_{-\infty}^{x_n} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(y_1,\dots,y_k) \, \mathrm{d}y_1 \cdots \mathrm{d}y_{n-1} \, \mathrm{d}y_{n+1} \cdots y_k \, \mathrm{d}y_n.$$

Thus, we get that F_{X_n} is an absolutely continuous function and the right hand-side of third equality in the statement can play the role of the corresponding density. It follows from Radon-Nikodym theorem that the density is unique up to a set of zero measure, which is here a k-dimensional Lebesgue measure.

Further, if the last equality in the statement holds, then $F_X(x) = \prod_{n=1}^{k} F_{X_n}(x_n)$ holds by integrating and then theorem 7 gives that the variables $X_1,\dots,X_k$ are independent. On the other hand, if the variables are independent, we have the equality $F_X(x) = \prod_{n=1}^{k} F_{X_n}(x_n)$ again by theorem 7, and Fubini theorem gives that

$$F_X(x) = \prod_{n=1}^{k} F_{X_n}(x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{n-1}} \int_{-\infty}^{x_{n+1}} \cdots \int_{-\infty}^{x_k} \prod_{m=1}^{k} f_{X_m}(y_m) \, \mathrm{d}y_k \cdots \mathrm{d}y_{n+1} \, \mathrm{d}y_{n-1} \cdots \mathrm{d}y_1 \, \mathrm{d}y_n.$$

Now, we see that the right-hand side of the last equality in the statement can play the role of the density of random vector X, and therefore we obtain the last equality in the statement holds almost everywhere by Random-Nikodym theorem. □

3.2. **Elementary conditioning.**

Remark Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ with $P(B) > 0$, then

$$P_{|B} : A \in \mathcal{A} \mapsto P(A|B) = \tfrac{P(A\cap B)}{P(B)}$$

is a probability measure on $(\Omega, \mathcal{A})$ with $P_{|B}(B) = 1$ and it is called a ***conditional probability*** by the event B under the measure P.[3] Moreover, if $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ is a random variable, we call its distribution $P_{|B}X^{-1}$ under the probability $P_{|B}$ as the ***conditional distribution of X given B*** and we write $P_{X|B} = P_{|B}X^{-1}$, i.e.

$$P_{X|B}(C) = P_{|B}[X \in C] = P(X \in C|B) \quad \text{if} \quad C \in \mathcal{E}.$$

Theorem 9 Let $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_X)$ be a continuous random vector with the density f_X and $B \in \sigma(X)$ with $P(B) > 0$. Then $X : (\Omega, \mathcal{A}, P_{|B}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{X|B})$ is again a continuous random variable with the density given by

$$f_{X|B}(x) = f_X(x)\tfrac{1_C(x)}{P(X\in C)}, \quad \text{where} \quad C \in \mathcal{B}(\mathbb{R}^k) \quad \text{is such that} \quad B = [X \in C]. \tag{2}$$

Proof: Let $A \in \mathcal{B}(\mathbb{R}^k)$. Since $B = [X \in C]$, we obtain that

$$\textstyle\int_A f_{X|B}(x) \, \mathrm{d}x = \frac{1}{P(B)} \int_A f_X(x) 1_C(x) \, \mathrm{d}x = \frac{1}{P(B)} \int_{A\cap C} f_X(x) \, \mathrm{d}x = \frac{P([X\in A]\cap B)}{P(B)} = P(X \in A|B),$$

and therefore $f_{X|B}$ given by (2) can play the role of the conditional density $f_{X|B}$ of X given B. □

Remark Generally (and roughly speaking), if X is a continuous random variable, then it is also continuous given B, and similarly, if X is a discrete random variable, then it is also discrete given B.

Theorem 10 Let $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_X)$ be a random variable and $Y : (\Omega, \mathcal{A}, P) \to (H, \mathcal{H}, P_Y)$ be a discrete random variable with $D = \{y \in H : P(Y = y) > 0\}$. Then the random variables X, Y are independent if and only if $P_{X|Y=y} = P_X$ holds for every $y \in D$.

[3] Do not confuse conditional probability $P_{|B}$ by the event B with the restriction of probability P on the set B denoted as $P|_B$, which is a measure that does not have to be probability one. But if we normalize the restriction $P|_B$ by $P(B)$, we obtain a probability measure, which is just $P_{|B}$.

Proof: Let us assume that the conditional distribution does not depends on the condition, i.e. $P_{X|Y=y} = P_X$ holds for every $y \in D$, and let $B \in \mathcal{E}, C \in \mathcal{H}$. Then

$$\begin{aligned} P(X \in B, Y \in C) &= \sum_{y \in C \cap D} P(X \in B | Y = y) P(Y = y) = \sum_{y \in C \cap D} P_{X|Y=y}(B) P(Y = y) \\ &= \sum_{y \in C \cap D} P_X(B) P(Y = y) = P_X(B) P(Y \in C \cap D) = P(X \in B) P(Y \in C), \end{aligned}$$

and we get that the variables X, Y are independent. On the other hand, if the variables X, Y are independent and $B \in \mathcal{E}, y \in D$, then

$$P_{X|Y=y}(B) = P(X \in B | Y = y) = \tfrac{P(X \in B, Y = y)}{P(Y=y)} = P(X \in B) = P_X(B),$$

i.e. $P_{X|Y=y} = P_X$ holds for every $y \in D$. □

Corollary If the variable X in theorem 10 is continuous with the density $f_X : \mathbb{R}^k \to [0, \infty)$, then X, Y are independent if and only if $f_{X|Y=y}(x) \overset{\text{ae}}{=} f_X(x)$ holds whenever $y \in D$, where $\overset{\text{ae}}{=}$ stands for the equality almost everywhere w.r.t. k-dimensional Lebesgue measure here.

Proof: If X, Y are independent, then theorem 10 gives that $P_{X|Y=y} = P_X \ll \lambda^k$ holds for every $y \in D$, where λ^k is a k-dimensional Lebesgue measure, and

$$\tfrac{\mathrm{d}P_{X|Y=y}}{\mathrm{d}\lambda^k} \overset{\text{ae}}{=} \tfrac{\mathrm{d}P_X}{\mathrm{d}\lambda^k} \overset{\text{ae}}{=} f_X.$$

Then the uniqueness of the Radon-Nikodym derivatives gives that $f_{X|Y=y} \overset{\text{ae}}{=} f_X$ holds. On the contrary, if $f_{X|Y=y} \overset{\text{ae}}{=} f_X$ holds for every $y \in D$, then $P_{X|Y=y} = P_X$, and we obtain from theorem 10 that the variables X, Y are independent. □

3.3. **Independence of systems of events.**

Let $T \neq \emptyset$, let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{S}_t \subseteq \mathcal{A}$ whenever $t \in T$. We say that a system of the events $\mathcal{S}_t, t \in T$ are (mutually, stochastically) independent if for every $T_0 \in T$ finite and $A_t \in \mathcal{S}_t, t \in T_0$ the following equality holds

$$P(\bigcap_{t \in T_0} A_t) = \prod_{t \in T_0} P(A_t).$$

Lemma D Random variables $X_t, t \in T$ are independent if and only if the $\sigma(X_t), t \in T$ are independent systems of random events.

Proof: It follows from the both definition of independence and from the definition of σ-algebra generated by a random variable in form $\sigma(X) = \{[X \in B] : B \in \mathcal{E}\}$. □

Theorem 11 Let $T \neq \emptyset$ and let $\mathcal{S}_t$ be a system of event closed under finite intersections whenever $t \in T$. Let us assume that the systems $\mathcal{S}_t, t \in T$ are independent, then also the corresponding σ-hulls $\sigma(\mathcal{S}_t), t \in T$ are independent.

Proof: Let $T_0 = \{t_1, \dots, t_k\} \subseteq T, k \in \mathbb{N}_0$. Denote

$$V(n) = [\sigma(\mathcal{S}_{t_1}), \dots, \sigma(\mathcal{S}_{t_n}), \mathcal{S}_{t_{n+1}}, \dots, \mathcal{S}_{t_k} \text{ are independent}].$$

First, realize that $V(0)$ is just our assumption, and therefore it holds. Further, $V(k)$ is what we want to show, and this is the reason why to use induction by n. Let us assume that $V(n-1)$ holds, where $n \in \mathbb{N}, n < k$, we will show that $V(n)$ holds also. By assumption

$$\mathcal{S}_{t_n} \subseteq \mathcal{M} = \{A_n \in \sigma(\mathcal{S}_{t_n}) : P(\bigcap_{j=1}^n A_j) = \prod_{j=1}^n P(A_j); j < n \Rightarrow A_j \in \sigma(\mathcal{S}_{t_j}), j > n \Rightarrow A_j \in \mathcal{S}_j\}.$$

By Dynkin lemma it is sufficient to realize that $\mathcal{M}$ is a Dynkin system in order to obtain that $\mathcal{M} = \sigma(\mathcal{S}_{t_n})$, which is just $V(n)$. □

Theorem 12 Let $T = \cup_{i \in I} T^i$ be a non-trivial $(I \neq \emptyset)$ disjoint union of non-empty sets $T^i \neq \emptyset$. Let $X_t : (\Omega, \mathcal{A}, P) \to (E_t, \mathcal{E}_t, P_{X_t})$ be a random variable whenever $t \in T$ and let us assume that such variables are independent. Then we get that the following random variables are also independent (under P)

$$X^i = (X_t, t \in T^i) : (\Omega, \mathcal{A}) \to (\prod_{t \in T^i} E_t, \bigotimes_{t \in T^i} \mathcal{E}_t), i \in I.$$

Proof: By lemma, it is sufficient to show that the systems of events $\sigma(X^i), i \in I$ are independent. By theorem 11, it is sufficient to show that the following systems of events closed under intersections and generating $\sigma(X^i) = \sigma(X_t, t \in T^i)$ are independent

$$\mathcal{S}_i = \{\bigcap_{t \in T_0^i} A_t, A_t \in \sigma(X_t) \text{ if } t \in T_0^i, \text{ and } T_0^i \text{ is a finite subset of } T^i\}, i \in I.$$

Let $I_0 \subseteq I$ be a finite set and $A^i \in \mathcal{S}_i$ if $i \in I_0$. Then there exist $T_i^0 \subseteq T^i$ finite and $A_t \in \sigma(X_t)$ if $t \in T_0^i$ such that $A^i = \cap_{t \in T^i} A_t$. Denote $T_0 = \cup_{i \in I} T_0^i$ and remind that $\sigma(X_t), t \in T$ are independent. Then

$$P(\bigcap_{i \in I_0} A^i) = P(\bigcap_{i \in I_0} \bigcap_{t \in T_0^i} A_t) = P(\bigcap_{t \in T_0} A_t) = \prod_{t \in S} P(A_t) = \prod_{i \in I_0} \prod_{t \in T_0^i} P(A_t) = \prod_{t \in I_0} P(\bigcap_{t \in T_0^i} A_t) = \prod_{i \in I_0} P(A^i).$$

□

Lemma E Let $X_t : (\Omega, \mathcal{A}, P) \to (E_t, \mathcal{E}_t, P_X), t \in T$ be independent random variables and let $f_t : (E_t, \mathcal{E}_t) \to (H_t, \mathcal{H}_t)$ be also measurable. Then $f_t(X_t), t \in T$ are also independent.

Proof: $\sigma(f(X_t)) = \{[f_t(X_t) \in B] : B \in \mathcal{H}_t\} = \{[X \in f_t^{-1}B] : B \in \mathcal{H}_t\} \subseteq \sigma(X_t)$ as $f_t^{-1}B \in \mathcal{E}_t$. □

Corollary Let $T = \cup_{i \in I} T_i$ be a non-trivial ($I \neq \emptyset$) disjoint union of non-empty sets $T_i \neq \emptyset$, and let $X_t, t \in T$ be independent random variables. If

$$f_i : (\prod_{t \in T_i} E_t, \bigotimes_{t \in T_i} \mathcal{E}_t) \to (H_i, \mathcal{H}_i) \text{ are measurable,}$$

then $f_i(X_t, t \in T_i)$ are independent random variables.

Proof: It follows from lemma with $X^i = (X_t, t \in T_i)$ and $\mathcal{E} = \otimes_{t \in T_i} \mathcal{E}_t, E_i = \prod_{t \in T_i} E_t$ and theorem 12. □

---

4. Mean (expected) value of real-valued random variable

Denote $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ and also $\mathcal{B}(\bar{\mathbb{R}}) = \sigma([-\infty, x); x \in \mathbb{R})$ the corresponding Borel σ-algebra. Let $(\Omega, \mathcal{A})$ be a measurable space such that $\Omega \neq 0$, denote

$$\mathbb{L}^* = \mathbb{L}^*(\Omega, \mathcal{A}) = \{X : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))\}$$

the set of all random variables on $(\Omega, \mathcal{A})$ with the state space $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$. Similarly,

$$\mathbb{L} = \mathbb{L}(\Omega, \mathcal{A}) = \{X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$$

the set of all random variables on $(\Omega, \mathcal{A})$ with the state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Further, put

$$\mathbb{L}^+(\Omega, \mathcal{A}) = \{f \in \mathbb{L}^*(\Omega, \mathcal{A}) : f \geq 0\}.$$

Let us assume that P is a probability measure on $(\Omega, \mathcal{A})$. We say that $X \in \mathbb{L}^*$ ***has a mean (expected) value*** $EX = \int_\Omega X \, \mathrm{d}P$ if the right-hand side Lebesgue integral exists in $\bar{\mathbb{R}}$. Otherwise, we say that X ***does not have a mean (expected) value***. The set of all real-valued random variable on $(\Omega, \mathcal{A})$ with a mean value is denoted as

$$\bar{\mathbb{L}}^* = \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P) = \{X \in \mathbb{L}^*(\Omega, \mathcal{A}) : E[X^+] < \infty, \ E[X^-] < \infty\}.$$

Further, let $p \in [0, \infty)$, then we denote

$$\mathbb{L}_p^* = \mathbb{L}_p^*(\Omega, \mathcal{A}, P) = \{X \in \mathbb{L}^*(\Omega, \mathcal{A}) : E|X|^p < \infty\}$$
$$\mathbb{L}_p = \mathbb{L}_p(\Omega, \mathcal{A}, P) = \mathbb{L}_p^*(\Omega, \mathcal{A}, P) \cap \mathbb{L}(\Omega, \mathcal{A}).$$

By definition „$0 \cdot \pm\infty = \pm\infty \cdot 0 = 0$".

Let $Z \subseteq \mathbb{R}$ be a locally finite division of $\mathbb{R}$, i.e. for every $a, b \in \mathbb{R}$ such that $a < b$ the set $(a, b) \cap Z$ is finite. Then the following generalized function

$$x \in \bar{\mathbb{R}} \mapsto \lfloor x \rfloor_Z := \sup\{z \in Z : z \leq x\}$$

is measurable from $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ to $(\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$.

Lemma Let $\mathcal{L} \subseteq \mathbb{L}^+(\Omega, \mathcal{A})$ satisfy

(1) $1_A \in \mathcal{L}$ holds whenever $A \in \mathcal{A}$

(2) if $f, g \in \mathcal{L}$ and $a, b \geq 0$ then $af + bg \in \mathcal{L}$
(3) if $f_n \in \mathcal{L}$ and $f_n \uparrow f \in \mathbb{L}^+(\Omega, \mathcal{A})$, then $f \in \mathcal{L}$.

then $\mathcal{L} = \mathbb{L}^+(\Omega, \mathcal{A})$.

Proof: If $f \in \mathbb{L}^+(\Omega, \mathcal{A})$ attains values in $\{0,1\}$. Then $F = [f = 1] \in \mathcal{A}$ and $f = 1_F \in \mathcal{L}$ holds by (1). If f attains values in a finite set $K \subseteq [0, \infty)$, then $1_{[f=k]} \in \mathcal{L}$ if $k \in K$, and we get from (2) that also

$$f = \sum_{k \in K} k 1_F \in \mathcal{L}.$$

If $f \in \mathbb{L}^+(\Omega, \mathcal{A})$, we put $K_n = \{j2^{-n}; j = 0, \ldots, n2^n\}$. Then $K_n \subseteq K_{n+1}$ and $0 \leq f_n := \lfloor f \rfloor_{K_n} \uparrow f$. Since $f_n \in \mathbb{L}^+(\Omega, \mathcal{A})$ attains values in the finite set K_n, we have that $f_n \in \mathcal{L}$ and point (3) gives that $f \in \mathcal{L}$. $\square$

Theorem 13 (*Properties of mean value*) Let $(\Omega, \mathcal{A}, P)$ be a probability space.

(1) If $A \in \mathcal{A}$, then $E[1_A] = P(A)$.
(2) If $a, b, c \in \mathbb{R}$ and $X, Y \in \mathbb{L}_1^*$, then $E(aX + bY + c) = aEX + bEY + c$.
(3) If $X, Y \in \bar{\mathbb{L}}^*$ and $X \leq Y$, then $EX \leq EY$.
(4) *Monotone convergence theorem*: Let $X_{n-1} \leq X_n$ hold almost surely and $X_n \in \mathbb{L}^*$ if $n \in \mathbb{N}$. Let $X_n \to X$ as $n \to \infty$ hold almost surely. If $X_1 \in \bar{\mathbb{L}}^*$ is such that $EX > -\infty$, then $X, X_n \in \bar{\mathbb{L}}^*$ and $EX = \lim_n EX_n$.
(5) *Fatou's lemma*: Let $X_n \in \mathbb{L}^*$ hold for every $n \in \mathbb{N}$. If there exists $Z \in \bar{\mathbb{L}}^*$ with $EZ > -\infty$ such that $Z \leq X_n$ holds almost surely for every $n \in \mathbb{N}$, then
$$\liminf_{n\to\infty} X_n \in \bar{\mathbb{L}}^* \quad \& \quad EZ \leq E \liminf_{n\to\infty} X_n \leq \liminf_{n\to\infty} EX_n.$$
(6) *Dominated Convergence Theorem*: Let $X_n \in \mathbb{L}^*, n \in \mathbb{N}$ and $Z \in \mathbb{L}_1^*$ be such that $|X_n| \leq Z$ holds almost surely whenever $n \in \mathbb{N}$. Then $X_n \in \mathbb{L}_1^*$ and if there exists $X \in \mathbb{L}^*$ such that $X_n \to X$ as $n \to \infty$ almost surely, then $X \in \mathbb{L}_1^*$ and $EX_n \to EX$ as $n \to \infty$.

Theorem 14 Let $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_X)$ be a random variable and $G : (E, \mathcal{E}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$. Then $G(X) \in \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P)$ if and only if $G \in \bar{\mathbb{L}}^*(E, \mathcal{E}, P_X)$. Further, if both sides hold, then

$$EG(X) = \int_\Omega G(X)\,\mathrm{d}P = \int_E G(x)\,\mathrm{d}P_X(x).$$

Proof: Denote $\mathcal{L} = \{F \in \mathbb{L}^+(E, \mathcal{E}) : EF(X) = \int_E F(x)\,\mathrm{d}P_X(x)\}$. Then (i) if $F \in \mathbb{L}^+(E, \mathcal{E})$ attains values in $\{0,1\}$, then

$$EF(X) = P(F(X) = 1) = P_X(\{x \in E : F(x) = 1\}) = \int_E F(x)\,\mathrm{d}P_X(x), \quad \text{i.e.} \quad F \in \mathcal{L}.$$

(ii) If $a, b \geq 0$ and $F, G \in \mathcal{L}$, then we get that $aF + bG \in \mathcal{L}$ as follows

$$E[aF(X) + bG(X)] = aEF(X) + bEG(X) = a\int_E F\,\mathrm{d}P_X + b\int_E G\,\mathrm{d}P_X = \int_E (aF + bG)\,\mathrm{d}P_X.$$

(iii) Let $F_n \in \mathcal{L}$ and let $F_n \uparrow F \in \mathbb{L}^+(E, \mathcal{E})$. Then we obtain from Monotone Convergence Theorem that

$$EF(X) = \lim_{n\to\infty} EF_n(X) = \lim_{n\to\infty} \int_E F_n\,\mathrm{d}P_X = \int_E F\,\mathrm{d}P_X, \quad \text{i.e.} \quad F \in \mathcal{L}.$$

By lemma, $\mathcal{L} = \mathbb{L}^+(E, \mathcal{E})$. If $G \in \mathbb{L}^*(E, \mathcal{E})$, then $G^\pm \in \mathbb{L}^+(E, \mathcal{E})$, and the first part of the proof gives that $E[G(X)^\pm] = \int_E G^\pm\,\mathrm{d}P_X$. This gives the first part of the statement. If $G \in \bar{\mathbb{L}}^*(E, \mathcal{E}, P_X)$, then we get that

$$EG(X) = E[G(X)^+] - E[G(X)^-] = \int_E G^+\,\mathrm{d}P_X - \int_E G^-\,\mathrm{d}P_X = \int_E G\,\mathrm{d}P_X.$$

$\square$

Theorem 15 Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ with $P(B) > 0$. If $X \in \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P)$, then $X \in \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P_{|B})$ and

$$E[X|B] = \int X\,\mathrm{d}P_{|B} = \tfrac{1}{P(B)} \int_B X\,\mathrm{d}P = \tfrac{1}{P(B)}\, E[X\, 1_B].$$

Proof: Denote

$$\mathcal{L} = \{Y \in \mathbb{L}^+(\Omega, \mathcal{A}) : E[X|B] = \tfrac{1}{P(B)} E[X\, 1_B]\}.$$

(i) If $A \in \mathcal{A}$, we show that $1_A \in \mathcal{L}$ as follows

$$E[1_A|B] = P(A|B) = \tfrac{P(A\cap B)}{P(B)} = \tfrac{1}{P(B)} E[1_A 1_B].$$

(ii) If $a, b \geq 0$ and $Y, Z \in \mathcal{L}$, then we get that $aY + bZ \in \mathcal{L}$ as follows

$$E[aY + bZ|B] = aE[Y|B] + bE[Z|B] = \tfrac{a}{P(B)}E[Y\,1_B] + \tfrac{b}{P(B)}E[Z\,1_B] = \tfrac{1}{P(B)}E[(aY + bZ)1_B].$$

(iii) If $\mathcal{L} \ni Y_n \uparrow Y \in \mathbb{L}^+(\Omega, \mathcal{A})$, then Monotone Convergence Theorem gives that

$$E[Y|B] = \lim_{n\to\infty} E[Y_n|B] = \lim_{n\to\infty} \tfrac{1}{P(B)}E[Y_n 1_B] = \tfrac{1}{P(B)}E[Y 1_B].$$

(iv) By lemma, $\mathcal{L} = \mathbb{L}^+(\Omega, \mathcal{A})$. If $X \in \bar{\mathbb{L}}^*$, then $X^\pm \in \mathbb{L}^+(\Omega, \mathcal{A})$ and we get that

$$E[X|B] = E[X^+|B] - E[X^-|B] = \tfrac{E[X^+ 1_B] - E[X^- 1_B]}{P(B)} = \tfrac{E[X 1_B]}{P(B)}.$$

Theorem 16 Let $X_1, \ldots, X_k$ be independent variables.
(1) If $X_j \geq 0$ holds for every $j \leq k$ or
(2) if $X_j \in \mathbb{L}_1^*$ holds for every $j \leq k$, then

$$E\prod_{j=1}^k X_j = \prod_{j=1}^k EX_j.$$

Proof: By lemma X_1 and $\prod_{j=2}^k X_j$ are independent, and therefore we may assume that $k = 2$. Then the correct proof can be obtained by induction. Let $k = 2$. First, we assume that $X_1 = X, X_2 = Y \geq 0$.

(1) If X, Y attain values in $\{0, 1\}$, then $E[XY] = P(X = 1, Y = 1) = P(X = 1)P(Y = 1) = EX \cdot EY$.

(2) If Y attains values in $\{0, 1\}$, then we put

$$\mathcal{L}_Y = \{Z \in \mathbb{L}^+(\Omega, \sigma(X)) : E[YZ] = EY \cdot EZ\}.$$

By (1) $\mathcal{L}_Y$ contains $\{0, 1\}$-valued $\sigma(X)$-measurable r.v.'s. If $a, b \geq 0$ and $Z_1, Z_2 \in \mathcal{L}_Y$, then

$$E[Y(aZ_1 + bZ_2)] = aE[YZ_1] + bE[YZ_2] = aEY \cdot EZ_1 + bEY \cdot EZ_2 = EY \cdot E[aZ_1 + bZ_2].$$

Hence, $aZ_1 + bZ_2 \in \mathcal{L}$. Finally, if $Z \in \mathbb{L}^+(\Omega, \sigma(X))$, then $0 \leq Z_n \uparrow Z$ as $n \to \infty$ holds with $Z_n = \lfloor Z \rfloor_{K_n}$, where K_n is defined as in the proof of the lemma above. Then we get that $Z \in \mathcal{L}$ as follows

$$E[YZ] = \lim_{n\to\infty} E[YZ_n] = \lim_{n\to\infty} EY \cdot EZ_n = EY \cdot EZ.$$

By the same lemma $\mathcal{L} = \mathbb{L}^+(\Omega, \sigma(X))$, and the statement of the theorem is proved for $\{0, 1\}$ valued random variable Y.

(3) Denote $\mathcal{L}^X = \{U \in \mathbb{L}^+(\Omega, \sigma(Y)) : E[XU] = EX \cdot EU\}$. Then $\mathcal{L}^X$ contains $\{0, 1\}$-valued $\sigma(Y)$-measurable random variables. It can be showed that it is closed under non-negative linear combinations and that it is closed under monotone convergence similarly as in step (2). Then we get from the lemma, we have already used, that $\mathcal{L}^X = \mathbb{L}^+(\Omega, \sigma(Y))$.

If $X, Y \in \mathbb{L}_1^*$, then we get from the first part of the proof that

$$\begin{aligned}
E[XY] &= E[X^+Y^+] + E[X^-Y^-] - E[X^+Y^-] - E[X^-Y^+] \\
&= E[X^+]E[Y^+] + E[X^-]E[Y^-] - E[X^+]E[Y^-] - E[X^-]E[Y^+] \\
&= (E[X^+] - E[X^-])(E[Y^+] - E[Y^-]) = EX \cdot EY.
\end{aligned}$$

□

Lemma If $X \geq 0$ is a real-valued random variable, then $EX = \int_0^\infty (1 - F_X(x))\,\mathrm{d}x$.

Proof: By definition $1 - F_X(x) = 1 - P(X < x) = P(X \geq x)$ and we get from Fubini theorem that

$$EX = \int_\Omega X\,\mathrm{d}P = \int_{[0,\infty)} x\,\mathrm{d}P_X(x) = \int_{[0,\infty)} \int_0^t \mathrm{d}s\,\mathrm{d}P_X(x) = \int_0^\infty \int_{[s,\infty)} \mathrm{d}P_X(x)\,\mathrm{d}s = \int_0^\infty P(X \geq s)\,\mathrm{d}s$$

□

Lemma Let X be a real-valued random variable.
(1) Then $X \in \mathbb{L}_1$ holds if and only if $\sum_n P(|X| \geq n) < \infty$
(2) Moreover,

$$\sum_{n\in\mathbb{N}} P(|X| \geq n) \leq EX \leq 1 + \sum_{n\in\mathbb{N}} P(|X| \geq n).$$

Proof: Since $X \in \mathbb{L}_1$ holds if and only if $|X| \in \mathbb{L}_1$, we may prove the statement for $Y = |X| \geq 0$. Denote $Z = \lfloor Y \rfloor_{\mathbb{Z}}$. Then $Z \leq Y \leq Z+1$, and therefore $EZ \leq E|X| \leq 1+EZ$, and $X \in \mathbb{L}_1$ if and only if $EZ < \infty$. Let us compute

$$EZ = \sum_{n=1}^{\infty} nP(Z=n) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} P(Z=n) = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} P(Z=n) = \sum_{k=1}^{\infty} P(Z \geq k) = \sum_{k\in\mathbb{N}} P(|X| \geq k)$$

□

Theorem 17 (*Jensen inequality*) Let $D \subseteq \mathbb{R}^k$ be a convex set, where $k \in \mathbb{N}$. Let X be a k-dimensional real-valued random vector on $(\Omega, \mathcal{A}, P)$ with $X(\Omega) \subseteq D$. Let $EX := (EX_1, \ldots, E_k)^\intercal \in \mathbb{R}^k$. Then $EX \in D$. If $G : D \to \mathbb{R}$ is a convex function, then $G(X) \in \bar{\mathbb{L}}^*$ and $EG(X) \geq G(EX)$.

If G is even strictly convex, and $X \neq EX$ holds with positive probability, then $EG(X) > G(EX)$.

Supporting theorems from convex analysis
(i) Let $D \subseteq \mathbb{R}^k$, where $k \in \mathbb{N}$, and $x \notin \operatorname{int} D$. Then there exists $a \in \mathbb{R}^k, a \neq 0$ such that $a^\intercal x \leq \inf_{d\in D} a^\intercal d$.
(ii) Let $D \subseteq \mathbb{R}^k$, where $k \in \mathbb{N}$, and $x_0 \in \operatorname{int} D$. Let $G : D \to \mathbb{R}$ be a convex function. Then there exists $a \in \mathbb{R}^k$ such that $G(x) \geq G(x_0) + a^\intercal(x - x_0)$ holds for every $x \in D$.

Remark Let assume that the assumptions of (ii) are satisfied and further assume that G is even strictly convex. Then we have the strict inequality $G(x) > G(x_0) + a^\intercal(x - x_0)$ for every $x \in D\backslash\{x_0\}$.

Proof: Let us assume the contrary, i.e. that there exists $x \in D\backslash\{x_0\}$ such that the strict inequality does not hold, i.e. we have the equality $G(x) = G(x_0) + a^\intercal(x - x_0)$, thus G is linear on the line with the end points x and x_0, which contradicts the assumption that G is strictly convex. □

Proof of Theorem 17: Without loss of generality we assume that $P(a_1^\intercal X = b) < 1$ holds for every $0 \neq a_1 \in \mathbb{R}^k$ and $b \neq 0$.[4] Then we show that $EX \in \operatorname{int} D$. Let us assume the contrary, i.e. let $EX \notin \operatorname{int} D$. By (i) the first theorem from convex analysis, there exists $0 \neq a \in \mathbb{R}^k$ such that $a^\intercal EX \leq \inf_{d\in D} a^\intercal d$. Since X attains values in D, we get that $a^\intercal X \geq a^\intercal EX$. By additional assumption, $P(a^\intercal X = a^\intercal EX) < 1$ holds and therefore we get the strict inequality $a^\intercal X > a^\intercal EX$ holds with positive probability. Hence, we have the strict inequality for the mean value $a^\intercal EX = Ea^\intercal X > a^\intercal EX$, and this is a contradiction. Therefore $EX \in \operatorname{int} D$.

By the second theorem from convex analysis, there exists $a \in \mathbb{R}^k$ such that

$$G(x) \geq G(EX) + a^\intercal(x - EX) =: L(x), \quad x \in D.$$

Since L is an affine function and $EX \in \mathbb{R}^k$, we get that $L(X) \in \mathbb{L}_1$, and therefore $G(X) \in \bar{\mathbb{L}}^*$ as $G(X) \geq L(X)$. Further, we get that $EG(X) \geq G(EX)$.

If G is a strict convex function and $X \neq EX$ holds with a positive probability, then $G(X) > G(EX) + a^\intercal(X - EX)$ holds also with a positive probability, and we obtain that $EG(X) > G(EX)$. □

Corollary

(1) Let $0 < p < q < \infty$ and put $G(x) := x^{q/p}$. Then G is a strictly convex function on $D = [0, \infty)$. Let $X \in \mathbb{L}_p$, put $Y = |X|^p$. Then Jensen inequality gives that $E|X|^q = EG(Y) \geq G(EY) = (E|X|^p)^{q/p}$, and therefore $(E|X|^q)^{1/q} \leq (E|X|^p)^{1/p}$. Moreover, if $|X|$ a non-degenerate random variable, we get that we have the strict inequality.

(2) *Schwartz inequality*
Let $X, Y \in \mathbb{L}(\Omega, \mathcal{A})$ be such that $XY \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$. Put $Z = (|X|^2, |Y|^2)^\intercal$ and $G(x, y) = -\sqrt{xy}, D = [0, \infty)^2$. Then Jensen inequality gives that

$$E|XY| \leq \sqrt{E|X|^2 E|Y|^2}.$$

[4] Otherwise we would consider the orthogonal base of $\mathbb{R}^k$ of the form $(a_1, \ldots, a_k) =: A$ and we define vector $Y = AX$. It has values in the convex set $\mathbb{D} := AD$. Further, $\mathbb{G} : y \in \mathbb{D} \mapsto G(A^{-1}y)$ is a convex function. Since $Y_1 = b$ holds almost surely, we can reduce dimension. Put $Z = (Y_2, \ldots, Y_n)^\intercal$ and $C = \{c \in \mathbb{R}^{k-1} : (b, c^\intercal)^\intercal \in \mathbb{D}\}$. Then C is a convex set. Let us consider $F(c) = \mathbb{G}(b, c)$, then F is a convex function on C such that $F(Z) = \mathbb{G}(Y) = G(X)$ holds almost surely. If $EF(Z) \geq F(EZ)$, then $EG(X) \geq G(EX)$ and similarly if $EF(Z) > F(EZ)$, then $EG(X) > G(EX)$. This is a suggestion how to reduce the dimension of the problem. We have showed how difficult would be the complete proof of Jensen inequality. Such a proof would have to contain an induction containing the above suggested reduction of the dimension.

(3) *Minkovsky inequality*
Let $X, Y \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$. Put $G(x,y) = -(\sqrt{x}+\sqrt{y})^2$ on $D = [0,\infty)^2$ and $Z = (|X|^2, |Y|^2)^\intercal$. Then Jensen inequality gives that

$$-E(\sqrt{|X|^2+|Y|^2})^2 = EG(Z) \geq G(EZ) = -(\sqrt{E|X|^2}+\sqrt{E|Y|^2})^2$$

and we get that

$$\sqrt{E|X+Y|^2} \leq \sqrt{E(|X|+|Y|)^2} \leq \sqrt{E|X|^2}+\sqrt{E|Y|^2}.$$

Let $T \neq \emptyset$, we say that real-valued random variables $X_t, t \in T$ defined on $(\Omega, \mathcal{A}, P)$ ***are uniformly integrable*** if

$$\lim_{c\to\infty} \sup_{t\in T} E[|X_t|; |X_t| \geq c] = 0,$$

where $E[X; A] \triangleq E[X \cdot 1_A]$ stands for the expected value of $X \in \mathbb{L}(\Omega, \mathcal{A})$ on $A \in \mathcal{A}$ if the value is well defined.

Lemma Let $T \neq \emptyset$, and assume that there exists $Y \in \mathbb{L}_1$ such that $|X_t| \leq Y$ holds for every $t \in T$. Then $X_t, t \in T$ are uniformly integrable.

Proof: By assumption $E[|X_t|; |X_t| \geq c] \leq E[|Y|; |Y| \geq c] \to 0$ as $c \to \infty$. □

Let $T \neq \emptyset$, we say that real-valued random variables $X_t, t \in T$ defined on $(\Omega, \mathcal{A}, P)$ are ***equally integrable*** if

$$\sup_{t\in T} E|X_t| < \infty.$$

Lemma Let $\delta > 0$ be such that $|X_t|^{1+\delta}, t \in T$ are equally integrable, then $X_t, t \in T$ are uniformly integrable.

Proof: Let us compute

$$E[|X_t|; |X_t| \geq c] = c \cdot E[\tfrac{|X_t|}{c}; |X_t| \geq c] \leq c \cdot E[(\tfrac{|X_t|}{c})^{1+\delta}] \leq c^{-\delta} \sup_{s\in T} E[|X_s|^{1+\delta}] \to 0$$

as $c \to \infty$. □

Let $T \neq \emptyset$, we say that real-valued random variables $X_t, t \in T$ defined on $(\Omega, \mathcal{A}, P)$ have ***equally absolutely continuous integrals*** if

$$\lim_{\delta\to 0^+} \sup\{E[|X_t|; A] : A \in \mathcal{A}, P(A) < \delta, t \in T\} = 0.$$

Theorem 18 Random variables $X_t, t \in T$ are uniformly integrable if and only if they are equally integrable and if they have equally absolutely continuous integrals.

Proof: Let $X_t, t \in T$ be equally integrable with equally absolutely continuous integrals. Then

$$P(|X_t| \geq c) \leq \tfrac{1}{c} E|X_t| \leq \tfrac{1}{c} \sup_{s\in T} E|X_s| < \tfrac{1}{c}\,(1 + \sup_{s\in T} E|X_s|) =: \delta(c) \to 0$$

as $c \to \infty$, and

$$E[|X_t|; |X_t| \geq c] \leq \sup\{E[|X_s|; A] : A \in \mathcal{A}, P(A) < \delta(c), s \in T\} \to 0$$

as $c \to \infty$. On the other hand, let us assume that the random variables $X_t, t \in T$ are uniformly integrable, then there exists c_0 such that $E|X_t| \leq E[|X_t|; |X_t| \geq c_0] \leq 1$ holds whenever $t \in T$.

$$E|X_t| \leq E[|X_t|; |X_t| \geq c_0] + E[|X_t|; |X_t| < c_0] \leq 1 + c_0 < \infty.$$

Further, if $c \in (0, \infty)$, then

$$E[|X_t|; A] \leq E[|X_t|; |X_t| \geq c] + c \cdot P(A).$$

Let $\varepsilon > 0$, then there exists $c \in (0, \infty)$ such that $E[|X_t|; |X_t| \geq c] < \frac{\varepsilon}{2}$ holds whenever $t \in T$. Further, put $\delta_0 = \frac{\varepsilon}{2c}$. If $\delta \in (0, \delta_0)$, then $c \cdot P(A) < \frac{\varepsilon}{2}$ holds if $A \in \mathcal{A}$ is such that $P(A) < \delta$, and therefore

$$\sup\{E[|X_t|; A] : A \in \mathcal{A}, P(A) < \delta, t \in T\} < \varepsilon$$

holds whenever $\delta \in (0, \delta_0)$. □

Theorem 19 (*Vallé-Poussin*) Real-valued random variables $X_t, t \in T$ are uniformly integrable if and only if there exist a non-decreasing function $G : [0, \infty) \to [0, \infty)$ with $G(t) \to \infty$ as $t \to \infty$ such that $|X_t|G(|X_t|), t \in T$ are equally integrable random variables.

Proof: Let $|X_t|G(|X_t|), t \in T$ be equally integrable random variables. Then

$$E[|X_t|; |X_t| \geq c] \leq \tfrac{1}{G(c)} E[|X_t|G(|X_t|)] \leq \tfrac{1}{G(c)} \sup_{s \in T} E[|X_s|G(|X_s|)] \to 0$$

as $c \to \infty$, since $G(c) \to \infty$. On the other hand, assume that the random variables $X_t, t \in T$ are uniformly integrable. Then there exists a sequence $0 \leq c_k \uparrow \infty$ such that

$$E[|X_t|; |X_t| \geq c_k] \leq 4^{-k} \quad \text{and put} \quad G = \textstyle\sum_{n \in \mathbb{N}} 2^n \cdot 1_{[c_n, c_{n+1})} .$$

Then G is really a non-decreasing function with $G(t) \to \infty$ as $t \to \infty$, and

$$E[|X_t|G(|X_t|)] = \sum_{n \in \mathbb{N}} 2^n E[|X_t| \cdot 1_{[c_n, c_{n+1})}(|X_t|)] \leq \sum_{n \in \mathbb{N}} 2^n E[|X_t|; |X_t| \geq c_n] \leq \sum_{n \in \mathbb{N}} 2^{-n} = 1.$$

$\square$

Theorem 20 (*Wald equalities*) Let $X_n, n \in \mathbb{N}$ be a sequence of independent identically distributed real-valued random variables on a probability space $(\Omega, \mathcal{A}, P)$ independent with a random variable N attaining only values in $\mathbb{N}_0$. Let us consider the following random sum

$$S = \sum_{n=1}^{N} X_n.$$

(1) If $X_1, N \in \mathbb{L}_1$, then $S \in \mathbb{L}_1$ and $ES = EN \cdot EX_1$, and $\mathrm{var}(S) = EN \cdot \mathrm{var}(X_1) + \mathrm{var}N \cdot (EX_1)^2$.
(2) If $\alpha \in \mathbb{R} \backslash \{0\}$ is such that $e^{\alpha X_1} \in \mathbb{L}_1$, then

$$E[e^{\alpha S}(Ee^{\alpha X_1})^{-N}] = 1, \quad \text{i.e.} \quad E \prod_{n=1}^{N} \tfrac{\exp\{\alpha X_n\}}{E \exp\{\alpha X_n\}} = 1$$

Proof: (1a) If $X_n \geq 0$, then we get from Fubini theorem that

$$ES = \sum_{n=1}^{\infty} E\big(1_{[N=n]} \sum_{k=1}^{n} X_k\big) = \sum_{n=1}^{\infty} P(N = n) \cdot E\big(\sum_{k=1}^{n} X_k\big) = EX_1 \cdot \sum_{n=1}^{\infty} nP(N = n) = EX_1 \cdot EN$$

as $1_{[N=n]}$ and $\sum_{k=1}^{n} X_k$ are independent real-valued integrable random variables. (1b) If X_n may attain all real values, then $X_n^{\pm}$ satisfy assumptions of (1a). Then we get that $ES_{\pm} = E[X_1^{\pm}] \cdot EN < \infty$, where

$$S_{\pm} = \sum_{n=1}^{N} X_n^{\pm}.$$

Then $S_{\pm} \in \mathbb{L}_1$, $S = S_+ - S_- \in \mathbb{L}_1$, and

$$ES = ES_+ - ES_- = (E[X_1^+] - E[X_1^-]) \cdot EN = E[X_1] \cdot EN.$$

(1c) Since $S^2 \geq 0$ holds, we obtain from Monotone Convergence Theorem that as $n \to \infty$

$$\begin{aligned} ES^2 \leftarrow E[S^2 1_{[N \leq n]}] &= \sum_{k=1}^{n} \sum_{m,j=1}^{k} E(X_m X_j 1_{[N=k]}) = \sum_{k=1}^{n} \sum_{m,j=1}^{k} E(X_m X_j) \cdot P(N = k) \\ &= \sum_{k=1}^{n} P(N = k)\big(k\, EX_1^2 + k(k-1)\,(EX_1)^2\big) \to EX_1^2 \cdot EN + (EX_1)^2 \cdot EN(N-1). \end{aligned}$$

Hence, $ES^2 = EX_1^2 \cdot EN + (EX_1)^2 \cdot EN(N-1) = EN \cdot \mathrm{var}(X_1) + (EX_1)^2 \cdot EN^2$. Then

$$\begin{aligned} \mathrm{var}(S) = ES^2 - (ES)^2 &= EN \cdot \mathrm{var}(X_1) + (EX_1)^2 \cdot EN^2 - (EX_1)^2 \cdot (EN)^2 \\ &= EN \cdot \mathrm{var}(X_1) + \mathrm{var}N \cdot (EX_1)^2. \end{aligned}$$

(2) Let $\alpha \neq 0$ and $Ee^{\alpha X_1} < \infty$. Again by Monotone Convergence Theorem, we get that

$$Ee^{\alpha S}(Ee^{\alpha X_1})^{-N} \leftarrow E[e^{\alpha S}(Ee^{\alpha X_1})^{-N} 1_{[N \le n]}] = \sum_{k=1}^{n} E\big(1_{[N=k]} \cdot \exp\{\alpha \sum_{j=1}^{k} X_j\}\big)$$

$$= \sum_{k=1}^{n} P(N=k) \cdot E \prod_{j=1}^{k} \frac{\exp\{\alpha X_j\}}{E\exp\{\alpha X_1\}} = \sum_{k=1}^{n} P(N=k) \cdot 1 = P(N \le n) \to 1.$$

Hence, $Ee^{\alpha S}(Ee^{\alpha X_1})^{-N} = 1$. □

4.1. **Mean value of a complex random variable.**

Let X be a random variable with the state space $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$. We say that it has a ***mean value c*** $\in \mathbb{C}$ if $\Re X, \Im X \in \mathbb{L}_1$ and $\Re c = E[\Re X], \Im c = E[\Im X]$.

Remark A complex-valued random variable has a mean value if and only if $E|X| < \infty$ and if this holds, we have the following inequality $|EX| \le E|X|$.

Theorem 20* Let the general assumptions of theorem 20 be satisfied.

(1) If $s \in \mathbb{R}$, then $Ee^{sS} = E(E\exp\{s^{X_1}\})^N$.

(2) If $s \in \mathbb{C}$ is such that $Ee^{tS} < \infty$ holds with $t = \Re s$, then $Ee^{sS} = E(E\exp\{sX_1\})^N$.

Proof: (1) By Fubini theorem

$$Ee^{sS} = E\sum_{n=1}^{\infty} 1_{[N=n]} e^{s\sum_{k=1}^{n} X_k} = \sum_{n=1}^{\infty} E1_{[N=n]} e^{s\sum_{k=1}^{n} X_k} = \sum_{n=1}^{\infty} P(N=n) \cdot E\prod_{k=1}^{n} e^{sX_k}$$

$$= \sum_{n=1}^{\infty} P(N=n) \cdot (Ee^{sX_1})^n = E[(Ee^{sX_1})^N].$$

(2) Obviously, $E|e^{sS}| = Ee^{S \cdot \Re s} < \infty$ holds by assumption and therefore e^{sS} has a mean value. Again, we use Fubini theorem and the same calculation gives the desired result. If we write $t = \Re s$ instead of s, we have an integrable function, which dominates the one with s, and this allows us to use Fubini Theorem. □

5. Convergence of Random variables

We say that a sequence $X_n, n \in \mathbb{N}$ of real-valued random variables ***converges almost surely*** to a real valued-random variable X if there exists $A \in \mathcal{A}$ with $P(A) = 1$ such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$ holds whenever $\omega \in A$. As $[\lim_n X_n = X] \in \mathcal{A}$, we get that $X_n \to X$ a.s. iff (if and only if) $P(\lim_n X_n = X) = 1$.

We say that a sequence $X_n, n \in \mathbb{N}$ of real-valued random variables ***converges in probability*** to a real valued-random variable X if

$$\forall\, \varepsilon > 0 \quad \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

We say that a sequence $X_n, n \in \mathbb{N}$ of real-valued random variables ***converges in*** $\mathbb{L}_p$ to a real valued-random variable X, where $p \in [1, \infty)$ if $X_n, X \in \mathbb{L}_p$ and $E|X_n - X|^p \to 0$ as $n \to \infty$.

Remark The limit almost surely, in probability and in $\mathbb{L}_p$ is determined uniquely up to a P-null set.
(i) Let $X_n \to X$ almost surely and $X_n \to Y$ almost surely as $n \to \infty$, then

$$P(X=Y) = P(X=Y, \lim_{n\to\infty} X_n = X, \lim_{n\to\infty} X_n = Y) = P(\lim_{n\to\infty} X_n = X, \lim_{n\to\infty} X_n = Y) = 1.$$

(ii) Let $X_n \to X$ in probability, then

$$P(|X-Y| > \varepsilon) \le P(|X - X_n| + |X_n - Y| > \varepsilon) \le P(|X_n - X| > \tfrac{\varepsilon}{2}) + P(|X_n - Y| > \tfrac{\varepsilon}{2} \to 0$$

as $n \to \infty$. In particular, $P(|X-Y| > \varepsilon) = 0$ holds for every $\varepsilon > 0$, i.e. $X = Y$ holds almost surely.
(iii) Let $X_n \to X$ in $\mathbb{L}_p$ and $X_n \to Y$ in $\mathbb{L}_p$, where $p \in [1, \infty)$. Then

$$(E|X-Y|^p)^{1/p} \le (E|X - X_n|^p)^{1/p} + (E|X_n - Y|^p)^{1/p} \to 0$$

as $n \to \infty$, and therefore $X = Y$ holds almost surely.

- If $X_n \to X$ as $n \to \infty$ almost surely, we briefly write $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$.
- If $X_n \to X$ as $n \to \infty$ in probability P, we briefly write $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$.
- If $X_n \to X$ as $n \to \infty$ in $\mathbb{L}_p$, we briefly write $X_n \xrightarrow{\mathbb{L}_p} X$ as $n \to \infty$.
- If $X = Y$ holds almost surely, we briefly write $X \overset{\text{as}}{=} Y$.

Remark Let $(\Omega, \mathcal{A}, P)$ be a probability space. If $X \in \mathbb{L}_p(\Omega, \mathcal{A})$ holds, where $p \in [1, \infty)$, we denote $||X||_p = (E|X|^p)^{1/p}$. Then $||.||_p$ is a pseudonorm on $\mathbb{L}_p(\Omega, \mathcal{A}, P)$ and $||X||_p = 0$ if and only if $X \overset{\text{as}}{=} 0$.

Then $\varrho_p(X, Y) = ||X - Y||_p$ is a pseudometric on $\mathbb{L}_p(\Omega, \mathcal{A}, P)$ such that $\varrho_p(X, Y) = 0$ holds if and only if $X \overset{\text{as}}{=} Y$. If $X_n, X \in L_p(\Omega, \mathcal{A}, P)$, then we get the by definition of the convergence in $\mathbb{L}_p$ that

$$X_n \to X \quad \text{as} \quad n \to \infty \quad \text{in} \quad \mathbb{L}_p \quad \equiv \quad \varrho_p(X_n, X) \to 0 \quad \text{as} \quad n \to \infty.$$

Remark: Let us denote $\varphi_0(x) = \frac{x}{1+x}$ and $\varphi_1(x) = 1 \wedge x$, where $x \wedge y := \min\{x, y\}$. Further, denote

$$\psi_0(x, y) := \varphi_0(|x - y|) \leq \varphi_1(|x - y|) =: \psi_1(x, y).$$

If $j \in \{0, 1\}$, it can be easily verified that $\psi_j(x, y)$ is a metric on $\mathbb{R}$ equivalent to $|x - y|$ bounded by 1 from above.

Proof: It is left to the reader up to the triangle inequality of ψ_0. We are going to show that $\varphi_0(a + b) \leq \varphi_0(a) + \varphi_0(b)$ holds whenever $a, b \geq 0$. As φ_0 is increasing on $[0, \infty)$, we obtain that

$$\psi_0(x, z) = \varphi_0(|x - z|) \leq \varphi_0(|x - y| + |y - z|) \leq \varphi_0(|x - y|) + \varphi_0(|y - z|) = \psi_0(x, y) + \psi_0(y, z)$$

holds whenever $x, y, z \in \mathbb{R}$. First, $\varphi_0'(x) = \frac{\mathrm{d}}{\mathrm{d}x}[1 - \frac{1}{x}] = (1 + x)^{-2}$ holds if $x > 0$. Since $\varphi'(x)$ is decreasing, we obtain that

$$\varphi_0(a + b) - \varphi_0(a) = \int_a^{a+b} \varphi_0'(x)\,\mathrm{d}x \leq \int_0^b \varphi_0'(x)\,\mathrm{d}x = \varphi_0(b) - \varphi_0(0) = \varphi_0(b).$$

$\square$

Theorem 21 Let $j \in \{0, 1\}$ and $(\Omega, \mathcal{A}, P)$ be a probability space. Then $\rho_j(X, Y) = E\rho_j(X, Y)$ is a pseudometric on $\mathbb{L}(\Omega, \mathcal{A})$ such that $\rho_j(X, Y) = 0$ holds if and only if $X \overset{\text{as}}{=} Y$ and such that

$$X_n \xrightarrow{\text{P}} X \quad \text{as} \quad n \to \infty \quad \equiv \quad \rho_j(X_n, X) \to 0 \quad \text{as} \quad n \to \infty$$

holds whenever $X_n, X \in \mathbb{L}(\Omega, \mathcal{A})$.

Proof: Obviously, ρ_j is a pseudometric on $\mathbb{L}(\Omega, \mathcal{A})$ as ψ_j is a metric on $\mathbb{R}$. If $0 = \rho_j(X, Y) = E\psi_j(X, Y)$, then $\psi_j(X, Y) \overset{\text{as}}{=} 0$ and since ψ_j is a metric, we get that $X \overset{\text{as}}{=} Y$. Let $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$ and $\varepsilon > 0$. Then

$$0 \leq \rho_0(X_n, X) \leq \rho_1(X_n, X) = E[1 \wedge |X_n - X|] \leq \varepsilon + P(|X_n - X| > \varepsilon).$$

As $\varepsilon > 0$ was arbitrary, we obtain convergence $\rho_j(X_n, X) \to 0$ as $n \to \infty$.

Let $\rho_j(X_n, X) \to 0$ as $n \to \infty$, then we get that $\rho_0(X_n, X) \to 0$ as $n \to \infty$, and then

$$P(|X_n - X| > \varepsilon) \leq E\varphi_0(|X_n - X|)/\varphi_0(\varepsilon) = \rho_0(X_n, X)/\varphi_0(\varepsilon) \to 0$$

as $n \to \infty$, since $\varphi_0(\varepsilon) \cdot 1_{[|X_n - X| > \varepsilon]} \leq \varphi_0(|X_n - X|)$. $\square$

Theorem 22 Let $(\Omega, \mathcal{A}, P)$ be a probability space.

(1) Let $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$, then $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$.
(2) Let $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$, then there exists $n_k \to \infty$ such that $X_{n_k} \xrightarrow{\text{as}} X$ as $k \to \infty$.

Proof: (1) Let $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$, then $\psi_1(X_n, X) \xrightarrow{\text{as}} 0$ as $n \to \infty$, and Dominated Convergence Theorem gives that $\rho_1(X, Y) = E\psi_1(X_n, X) \to 0$ as $n \to \infty$. By theorem 21, $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$.

(2) Let $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$. By theorem 21, $\rho_1(X_n, X) \to 0$ as $n \to \infty$, and therefore there exists a sequence $n_k \to \infty$ such that $\rho_1(X_{n_k}, X) < 2^{-k}$. Then Fubini theorem gives that

$$E \sum_{k=1}^{\infty} \psi_1(X_{n_k}, X) = \sum_{k=1}^{\infty} E\psi_1(X_{n_k}, X) = \sum_{k=1}^{\infty} \rho_1(X_{n_k}, X) \leq 1.$$

In particular, $\sum_k \psi(X_{n_k}, X)$ converges almost surely, and we get that $\psi(X_{n_k}, X) \xrightarrow{\text{as}} 0$ as $k \to \infty$. Then we have that $X_{n_k} \xrightarrow{\text{as}} X$ as $k \to \infty$. $\square$

Example Let $\Omega = [0, 1], \mathcal{A} = \mathcal{B}[0, 1]$ and P be a uniform distribution on $[0, 1]$. Let $A_n \in \mathcal{A}$ be a sequence with $P(A_n) \to 0$ as $n \to \infty$ such that $\limsup_n A_n = [0, 1]$. Then $X_n = 1_{A_n} \xrightarrow{\text{P}} 0$ and also in $\mathbb{L}_p$ whenever $p \in [1, \infty)$, but $X_n \overset{\text{as}}{\not\to} 0$ as $n \to \infty$.

Theorem 23 Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X_n, X \in \mathbb{L}(\Omega, \mathcal{A}), n \in \mathbb{N}$. Then

$$X_n \xrightarrow{\text{P}} X \quad \text{as} \quad n \to \infty \quad \text{if and only if} \quad \forall\, n_k \uparrow \infty\ \exists\, k_m \uparrow \infty \quad X_{n_{k_m}} \xrightarrow{\text{as}} X \quad \text{as} \quad n \to \infty. \tag{3}$$

Proof: Let $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$ and let $n_k \to \infty$, then obviously $X_{n_k} \xrightarrow{\mathrm{P}} X$ as $k \to \infty$. By theorem 22, there exists $k_m \uparrow \infty$ such that $X_{n_{k_m}} \xrightarrow{\mathrm{as}} X$ as $n \to \infty$. Now, assume that $X_n \not\xrightarrow{\mathrm{P}} X$ as $n \to \infty$. By theorem 21, there exists $\varepsilon > 0$ and $n_k \uparrow \infty$ such that $\rho_1(X_{n_k}, X) \geq \varepsilon$ holds for every $k \in \mathbb{N}$. If $k_m \uparrow \infty$, then $\rho_1(X_{n_{k_m}}, X) \geq \varepsilon > 0$ holds for every $m \in \mathbb{N}$, and therefore $X_{n_{k_m}} \not\xrightarrow{\mathrm{P}} X$ as $m \to \infty$ again by theorem 21. $\square$

Remark There does not exists a pseudometric, say ρ, on $\mathbb{L}(\Omega, \mathcal{A}, P)$ such that $X_n \xrightarrow{\mathrm{as}} X$ if and only if $\rho(X_n, X) \to 0$ as $n \to \infty$ in general, otherwise theorem 23 gives that $X_n \xrightarrow{\mathrm{as}} X$ as $n \to \infty$ if and only if $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$, i.e. these two convergence agree, which generally does not hold, see the example above.[5]

Lemma Let $\{X_t\}_{t \in T}, \{Y_t\}_{t \in T} \subseteq \mathbb{L}_1(\Omega, \mathcal{A}, P)$.

(1) If T is finite, then $X_t, t \in T$ are uniformly integrable (UI) random variables.
(2) If $T = T_1 \cup T_2$ and $X_t, t \in T_i$ are UI for $i \in \{1, 2\}$, then $X_t, t \in T$ are UI.

Further assume that $X_t, t \in T$ are UI

(a) If $a \in \mathbb{R}$, then $aX_t, t \in T$ are UI.
(b) If $|Y_t| \leq X_t, t \in T$ holds, then $Y_t, t \in T$ are UI.
(c) If $Y_t, t \in T$ are UI, then $X_t + Y_t, t \in T$ are UI.

Proof: (2): Let $\varepsilon > 0$ and $c_j \in (0, \infty)$ be such that $\sup_{t \in T_j} E[|X_t|; |X_t| \geq c_j] < \varepsilon$. Put $c = \max\{c_1, c_2\}$, then $\sup_{t \in T} E[|X_t|; |X_t| \geq c] < \varepsilon$.

(1): If $T = \{t\}$, then (1) obviously holds. Otherwise use (2) and induction.

(a): If $a = 0$, then (a) obviously holds. Otherwise, let $c \in (0, \infty)$, then we obtain from assumption that

$$\sup_{t \in T} E[|aX_t|; |aX_t| \geq c] \leq |a| \sup_{t \in T} E[|X_t|; |X_t| \geq c/|a|] \to 0 \quad \text{as} \quad c \to \infty.$$

(b): By assumption $\sup_{t \in T} E[|Y_t|; |Y_t| \geq c] \leq \sup_{t \in T} E[|X_t|; |X_t| \geq c] \to 0$ as $c \to \infty$.

(c): By theorem 18, X_t, Y_t are for $t \in T$ equally integrable, and therefore we obtain immediately from the definition that $X_t + Y_t, t \in T$ are also equally integrable as follows

$$\sup_{t \in T} E|X_t + Y_t| \leq \sup_{t \in T} E|X_t| + \sup_{t \in T} E|Y_t| < \infty.$$

Further, theorem 18 gives that $X_t, t \in T$ and also $Y_t, t \in T$ have equally absolutely continuous integrals. The same theorem says that it is enough to show that also $X_t + Y_t, t \in T$ have equally absolutely continuous integrals. For $\delta > 0$ denote $\mathcal{A}_{\delta,P} = \{A \in \mathcal{A} : P(A) < \delta\}$. Then

$$\sup_{\substack{t \in T \\ A \in \mathcal{A}_{\delta,P}}} E[|X_t + Y_t|; A] \leq \sup_{\substack{t \in T \\ A \in \mathcal{A}_{\delta,P}}} E[|X_t|; A] + \sup_{\substack{t \in T \\ A \in \mathcal{A}_{\delta,P}}} E[|Y_t|; A] \to 0$$

as $\delta \to 0^+$. $\square$

Remark: If $X_n \xrightarrow{\mathrm{P}} X$ and $c > 0$, then $X_n/c \xrightarrow{\mathrm{P}} X/c$ as $n \to \infty$.

Proof: Let $\varepsilon > 0$, then $P(|\frac{X_n}{c} - \frac{X}{c}| > \varepsilon) = P(|X_n - X| > c\varepsilon) \to 0$ as $n \to \infty$.

Theorem 24 Let $p \in [1, \infty)$ and $X_n, X \in \mathbb{L}(\Omega, \mathcal{A})$. Then the following conditions are equivalent

(1) $X_n \xrightarrow{\mathbb{L}_p} X$ and $X_n, X \in \mathbb{L}_p(\Omega, \mathcal{A}, P)$
(2) $X_n \xrightarrow{\mathrm{P}} X$ and $|X_n|^p, n \in \mathbb{N}$ are uniformly integrable.

Proof: Let (1) hold and let $\varepsilon > 0$. Then $P(|X_n - X| > \varepsilon) \leq \varepsilon^{-p} E|X_n - X| \to 0$ as $n \to \infty$. Hence $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$. Now, we show that $Y_n, n \in \mathbb{N}$ are uniformly integrable, where $Y_n = |X_n - X|^p$. Since $Y_n \in L_1(\Omega, \mathcal{A}, P)$ and $EY_n \to 0$ as $n \to \infty$, we get that Y_n are equally integrable. We will show that they have also equally absolutely continuous integrals. Let $\varepsilon > 0$, then there exists n_0 such that $\sup_{n \geq n_0} E|Y_n| < \varepsilon$. Since $Y_n \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, we get from lemma that $Y_n, n \leq n_0$ are uniformly integrable, and therefore they have equally absolutely continuous integrals. In particular, there exists $\delta_0 > 0$ such that

$$\sup\{E[|Y_n|; A] : A \in \mathcal{A}, P(A) < \delta, n \leq n_0\} < \varepsilon.$$

[5] In the special case $\Omega = \{0\}$, there is only one probability measure $P = \delta_0$ on $(\Omega, 2^\Omega)$, and convergences almost surely and in probability (and in $\mathbb{L}_p$) agree with convergence everywhere. Further, think of the case, where Ω is finite or countable.

Then $\sup\{E[|Y_n|;A] : A \in \mathcal{A}, P(A) < \delta, n \in \mathbb{N}\} < \varepsilon$, and we have that $|X_n - X|^p, n \in \mathbb{N}$ are uniformly integrable. Since $X \in \mathbb{L}_p$, we obtain that $|X_n|^p, n \in \mathbb{N}$ are uniformly integrable from lemma and from the following inequalities

$$|X_n|^p \le (|X_n - X| + |X|)^p \le 2^p \max\{|X_n - X|^p, |X|^p\} \le 2^p(|X_n - X|^p + |X|^p).$$

Let (2) hold. First, we show that $X \in \mathbb{L}_p(\Omega, \mathcal{A}, P)$. By theorem 22, there exists $n_k \uparrow \infty$ such that $X_{n_k} \xrightarrow{\text{as}} X$ as $k \to \infty$. By Fatou's Lemma

$$E|X|^p \le \liminf_{k\to\infty} E|X_{n_k}|^p \le \sup_{n\in\mathbb{N}} E|X_n|^p < \infty.$$

Second, we show that $Y_n = |X_n - X|^p, n \in \mathbb{N}$ are uniformly integrable. If follows from lemma and the following inequality

$$|X_n - X|^p \le 2^p(|X_n|^p + |X|^p).$$

Third, let $\varepsilon > 0$ we will show that there exists $n_0 \in \mathbb{N}$ such that $E|X_n - X|^p < \varepsilon$ holds whenever $n \ge n_1$. Since $Y_n = |X_n - X|^p$ are uniformly integrable, there exists $c \in (0, \infty)$ such that

$$\sup_{n\in\mathbb{N}} E[|Y_n|; |Y_n| \ge c^p] < \varepsilon.$$

Since $x^p \le x$ holds for $x \in [0,1], p \in [1, \infty)$, we obtain from the previous remark that

$$EY_n < \varepsilon + E[c^p \wedge |X_n - X|^p] \le \varepsilon + c^p E(1 \wedge |\tfrac{X_n - X}{c}|) = \varepsilon + c^p \rho_1(\tfrac{X_n}{c}, \tfrac{X}{c}) \to \varepsilon.$$

Hence, we get that for each $\varepsilon > 0$ we have that

$$\limsup_{n\to\infty} E|X_n - X|^p \le \varepsilon.$$

□

Let $X_n, n \in \mathbb{N}$ be a sequence of real valued random variables. We say that it is a ***Cauchy sequence almost surely*** if there exists $A \in \mathcal{A}$ with $P(A) = 1$ such that $X_n(\omega), n \in \mathbb{N}$ is a Cauchy sequence whenever $\omega \in A$, i.e.

$$\sup_{p\in\mathbb{N}} |X_{n+p} - X_n| \xrightarrow{\text{as}} 0 \quad \text{as} \quad n \to \infty.$$

We say that $X_n, n \in \mathbb{N}$ is a ***Cauchy sequence in probability*** if

$$\forall\, \varepsilon > 0 \;\; \exists\, n_0 \in \mathbb{N} \;\; \forall\, n, m \ge n_0 \quad P(|X_n - X_m| > \varepsilon) < \varepsilon.$$

Let $p \in [1, \infty)$, we say that $X_n, n \in \mathbb{N}$ is a ***Cauchy sequence in*** $\mathbb{L}_p$ if $X_n \in \mathbb{L}_p, n \in \mathbb{N}$ and if

$$\forall\, \varepsilon > 0 \;\; \exists\, n_0 \in \mathbb{N} \;\; \forall\, n, m \ge n_0 \quad E|X_n - X_m|^p < \varepsilon.$$

Theorem 25 Let $X_n, n \in \mathbb{N}$ be a sequence of real valued random variables on $(\Omega, \mathcal{A}, P)$.

(1) Then $X_n, n \in \mathbb{N}$ is a Cauchy sequence a.s. iff there exists $X \in \mathbb{L}(\Omega, \mathcal{A})$ s.t. $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$.
(2) $X_n, n \in \mathbb{N}$ is a Cauchy sequence in probability iff there exists $X \in \mathbb{L}(\Omega, \mathcal{A})$ s.t. $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$.
(3) Let $p \in [1, \infty)$, then $X_n, n \in \mathbb{N}$ is a Cauchy in $\mathbb{L}_p$ iff $\exists\, X \in \mathbb{L}_p(\Omega, \mathcal{A}, P)$ s.t. $X_n \xrightarrow{\mathbb{L}_p} X$ as $n \to \infty$.

Proof: a) Let us assume that $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$. Then $A = [\lim_n X_n = X] \in \mathcal{A}$ and $P(A) = 1$, and if $\omega \in A$, then $X_n(\omega) \to X(\omega)$ as $n \to \infty$, and we get that $X_n(\omega)$ is a Cauchy sequence.

b) Let $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$ and $\varepsilon > 0$, then there exists $n_0 \in \mathbb{N}$ such that $P(|X_n - X| > \frac{\varepsilon}{2}) < \frac{\varepsilon}{2}$. If $m, n \ge n_0$, then

$$P(|X_n - X_m| > \varepsilon) \le P(|X_n - X| + |X - X_m| > \varepsilon) \le P(|X_n - X| > \tfrac{\varepsilon}{2}) + P(|X_m - n| > \tfrac{\varepsilon}{2}) < \varepsilon.$$

c) Let $X_n \xrightarrow{\mathbb{L}_p} X$ as $n \to \infty$, then there exists $n_0 \in \mathbb{N}$ such that $(E|X_n - X|^p)^{1/p} < \frac{\varepsilon}{2}$ holds if $n \ge n_0$. If $m, n \ge n_0$, then

$$(E|X_m - X_n|^p)^{1/p} \le (E|X_m - X|^p)^{1/p} + (E|X_n - X|^p)^{1/p} < \varepsilon.$$

d) Let X_n be a Cauchy sequence almost surely, then there exists $A \in \mathcal{A}$ with $P(A) = 1$ such that $X_n(\omega)$ is a Cauchy sequence if $\omega \in A$. Then we put $X(\omega) := \lim_n X_n(\omega)$ if $\omega \in A$ and $X(\omega) := 0$ if $\omega \in \Omega \backslash A$. It follows from the definition that $X \in \mathbb{L}(\Omega, \mathcal{A})$ and also $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$.

e) Let X_n be a Cauchy sequence in probability. We will show that

$$\lim_{n\to\infty} \sup_{j\in\mathbb{N}} \rho_1(X_n, X_{n+j}) = 0$$

Let $\varepsilon > 0$. By assumption there exists n_0 such that $P(|X_n - X_m| > \frac{\varepsilon}{2}) < \frac{\varepsilon}{2}$ holds whenever $n \geq n_0$. Then

$$\rho_1(X_n, X_{n+j}) = E[1 \wedge |X_n - X_{n+j}|] \leq \varepsilon$$

holds if $n \geq n_0$ and $j \in \mathbb{N}$. Let $n_k \uparrow \infty$ be such that $\rho_1(X_{n_k}, X_{n_{k+1}}) < 2^{-k}$. Then

$$E \sum_{k \in \mathbb{N}} \psi_1(X_{n_k}, X_{n_{k+1}}) = \sum_{k \in \mathbb{N}} \rho_1(X_{n_k}, X_{n_{k+1}}) \leq 1,$$

and we get $\sum_k \psi_1(X_{n_k}, X_{n_{k+1}}) < \infty$ holds almost surely, which gives that X_{n_k} is almost surely a Cauchy sequence as

$$\psi_1(X_{n_k}, X_{n_{k+j}}) \leq \sum_{i=k}^{j-1} \psi_1(X_{n_i}, X_{n_{i+1}}) \leq \sum_{i=k}^{\infty} \psi_1(X_{n_i}, X_{n_{i+1}}) \to 0$$

as $k \to \infty$ holds almost surely. By step d) there exists $X \in \mathbb{L}(\Omega, \mathcal{A})$ such that $X_{n_k} \xrightarrow{\text{as}} X$ as $k \to \infty$.

$$\rho_1(X_n, X) \leq \rho_1(X_n, X_{n_k}) + \rho_1(X_{n_k}, X).$$

Then

$$\rho_1(X_n, X) \leq \limsup_{k \to \infty} \rho_1(X_n, X_{n_k}) \leq \sup_{j \in \mathbb{N}} \rho_1(X_n, X_{n+j}) \to 0$$

as $n \to \infty$, i.e. $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$.

f) Let X_n be a Cauchy sequence in $\mathbb{L}_p$, where $p \in [1, \infty)$. Then X_n is a Cauchy sequence in probability: if $\varepsilon > 0$, then there exists $n_0 \in \mathbb{N}$ such that $E|X_n - X_m|^p < \varepsilon^{1+p}$ holds whenever $m, n \geq n_0$. Then

$$P(|X_n - X_m| > \varepsilon) \leq \varepsilon^{-p} E|X_n - X_m|^p < \varepsilon$$

holds if $m, n \geq n_0$. In particular, there exists $X \in \mathbb{L}(\Omega, \mathcal{A})$ such that $X_n \xrightarrow{\text{P}} X$ as $n \to \infty$. By theorem 24, it is enough to show that $|X_n|^p, n \in \mathbb{N}$ are uniformly integrable. First, we show that they are equally integrable. By assumption there exists $n_0 \in \mathbb{N}$ such that $E|X_{n_0} - X_n|^p < 1$ holds whenever $n \geq n_0$. If $n \geq n_0$, we get that

$$E|X_n|^p \leq 2^p(E|X_n - X_{n_0}|^p + E|X_{n_0}|^p) \leq 2^p(1 + E|X_{n_0}|).$$

Since $X_n \in \mathbb{L}_p$, we get that $E|X_n|^p$ is a bounded sequence, and therefore $|X_n|^p, n \in \mathbb{N}$ are equally integrable. Further, we show that $|X_n|^p, n \in \mathbb{N}$ have equally absolutely continuous integrals. Let $\varepsilon > 0$ we consider $n_\varepsilon \in \mathbb{N}$ such that $E|X_n - X_m|^p < \varepsilon$ holds whenever $m, n \geq n_\varepsilon$. If $A \in \mathcal{A}$, then

$$E[|X_n|^p; A] \leq 2^p(E|X_n - X_{n_\varepsilon}|^p + E[|X_{n_\varepsilon}|^p; A]).$$

If $A \in \mathcal{A}$ has probability $P(A)$ small enough such that $E[|X_n|^p; A] < \varepsilon$ holds whenever $n \leq n_\varepsilon$, then we get that $E[|X_n|^p; A] \leq 2^{p+1}\varepsilon$ holds whenever $n \in \mathbb{N}$.

f*) Let X_n be a Cauchy sequence in $\mathbb{L}_p$, where $p \in [1, \infty)$. Then $\sup_j E|X_n - X_{n+j}|^p \to 0$ as $n \to \infty$. Let $n_k \uparrow \infty$ be such that $(E|X_{n_k} - X_{n_{k+1}}|^p)^{1/p} < 2^{-k}$. Then we get from Jensen inequality that

$$E|X_{n_k} - X_{n_{k+1}}| \leq (E|X_{n_k} - X_{n_{k+1}}|^p)^{1/p} < 2^{-k}.$$

Then

$$E \sum_{k \in \mathbb{N}} |X_{n_k} - X_{n_{k+1}}| = \sum_{k \in \mathbb{N}} E|X_{n_k} - X_{n_{k+1}}| \leq 1.$$

In particular, $\sum_k |X_{n_k} - X_{n_{k+1}}| < \infty$ holds almost surely. Similarly as in the step e), we get that X_{n_k} is a Cauchy sequence almost surely, and by d) we obtain that there exists $X \in \mathbb{L}(\Omega, \mathcal{A})$ such that $X_{n_k} \xrightarrow{\text{as}} X$ as $k \to \infty$. Then Fatou's lemma gives that

$$E|X_n - X|^p \leq \liminf_{k \to \infty} E|X_n - X_{n_k}|^p \leq \sup_{j \in \mathbb{N}} E|X_n - X_{n+j}|^p \to 0$$

as $n \to \infty$, and $E|X|^p \leq 2^p[E|X_n - X|^p + E|X_n|^p] < \infty$ holds if n is large enough as $X_n \in \mathbb{L}_p$, and therefore $X \in \mathbb{L}_p$ holds also. $\square$

6. Conditioning

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $B \in \mathcal{A}$ be such that $P(B) > 0$. If $A \in \mathcal{A}$, then we have the conditional probability of A given B in the form

$$P(A|B) = \tfrac{P(A\cap B)}{P(B)} \; (= P_{|B}(A)).$$

Further, if $X \in \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P)$, then $X \in \bar{\mathbb{L}}^*(\Omega, \mathcal{A}, P_{|B})$, and

$$E[X|B] = \int X \,\mathrm{d}P_{|B} = \tfrac{1}{P(B)} \int_B X \,\mathrm{d}P = \tfrac{1}{P(B)} E[X; B]$$

holds by theorem 15.

Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $\mathcal{B} \subseteq \mathcal{A}$ be σ-algebra. By a ***conditional expected value of X given $\mathcal{B}$*** we mean every $Y \in \mathbb{L}_1(\Omega, \mathcal{B}, P|\mathcal{B})$ such that one of the following equivalent conditions holds

(1) $\forall B \in \mathcal{B} \quad \int_B X \,\mathrm{d}P = \int_B Y \,\mathrm{d}P$ (*technical condition used mostly in proofs*)

(2) $\forall B \in \mathcal{B} \quad E[X; B] = E[Y; B]$ (*condition suitable for verifying by computing*)

(3) $\forall B \in \mathcal{B} \quad P(B) > 0 \quad \Rightarrow \quad E[X|B] = E[Y|B]$ (*condition offering interpretation*).

Such a variable Y will be denoted as $E[X|\mathcal{B}]$ and the set of all such values as $\mathbb{E}[X|\mathcal{B}]$.

Remark The condition (1) will be used in proofs, it helps us to avoid to use too many E's in the following proofs, and it is the usually used condition. The condition (2) is obviously equivalent to (1) and it is just the same condition as (1) using expectation E that enables to use our intuition. The relation between the conditions (2) and (3) is the same as the relation between the definition of independence of two random events A, B in the form $P(A \cap B) = P(A)P(B)$ and the condition $P(A|B) = P(B)$ if $P(B) > 0$, which is behind the name of this property "independence of expectation of the random event A on the information that B happened".

Although the third condition is closest to interpretation of the notion of the conditional expectation, it does not really says, what it is. We will be able to say what it really is only in case when $\mathcal{B}$ is finite and to refer the reader to the theory of martingales that

$$E[X|\mathcal{B}_n] \xrightarrow{\text{as}} E[X|\mathcal{B}_\infty]$$

holds whenever $\mathcal{B}_n \subseteq \mathcal{B}_{n+1}$ and $\mathcal{B}_\infty = \sigma(\cup_n \mathcal{B}_n)$. Further comments will be given when we know what is $E[X|\mathcal{B}]$ if B is finite. See theorem 26 and the text bellow.

Radon-Nikodym theorem Let ν, μ be σ-finite measures on a measurable space $(S, \mathcal{S})$ such that $\nu \ll \mu$.[6] then there exists a non-negative $\mathcal{S}$-measurable function f denoted also as $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ such that

$$\nu(A) = \int_A f \,\mathrm{d}\mu, \quad A \in \mathcal{S}. \tag{4}$$

Such a function f is determined uniquely up to a μ-null set, and it is called a ***Radon-Nikodym derivative of ν w.r.t. μ*** or a ***density of ν w.r.t. μ***.

Radon-Nikodym theorem for signed measures Let μ be a σ-finite measure on a measurable space $(S, \mathcal{S})$ and ν be a finite signed measure on $(S, \mathcal{S})$ such that $\nu \ll \mu$[7] Then there exists an $\mathcal{S}$-measurable μ-integrable function f such that (4) holds. Such a function f is determined uniquely up to a μ-null set, and it is called a ***Radon-Nikodym derivative of ν w.r.t. μ*** or a ***density of ν w.r.t. μ***.

Lemma Let ν be a finite signed measure on a measurable space $(S, \mathcal{S})$, then there exist $S_\pm \in \mathcal{S}$ such that $S_+ \cap S_- = \emptyset$ and that $\nu = \nu^+ - \nu_-$, where $\nu^\pm(A) = \nu(A \cap S_\pm)$ are measures on $(S, \mathcal{S})$.

Proof: By assumption $\nu = \nu_+ - \nu_-$, where $\nu_\pm$ are finite measures on $(S, \mathcal{S})$. Obviously, we have that $\nu_\pm \ll \mu := \nu_+ + \nu_-$. By Radon-Nikodym theorem there exist $\mathcal{S}$-measurable functions $f_\pm$ such that $\nu_\pm(A) = \int_A f_\pm \,\mathrm{d}\mu$ holds if $A \in \mathcal{S}$. Put $f = f_+ - f_-$ and $S_+ := [f > 0], S_- := [f < 0]$. Then

$$\nu(A) = \nu_+(A) - \nu_-(A) = \int_A f_+ \,\mathrm{d}\mu - \int_A f_- \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu, \quad A \in \mathcal{S}.$$

[6] i.e. $\forall\ A \in \mathcal{S}\ \mu(A) = 0 \Rightarrow \nu(A) = 0$.

[7] i.e. $\forall\ A \in \mathcal{S}\ \mu(A) = 0 \Rightarrow \nu(A) = 0$.

In particular, $\nu^{\pm}(A) = \nu(A \cap S_{\pm}) = \int_{A\cap S_{\pm}} f \, \mathrm{d}\mu$, and therefore

$$\begin{aligned} \nu^+(A) - \nu^-(A) &= \nu(A \cap S_+) - \nu(A \cap S_-) = \int_{A\cap S_+} f \, \mathrm{d}\mu - \int_{A\cap S_-} f \, \mathrm{d}\mu \\ &= \int_A f 1_{[f>0]} \, \mathrm{d}\mu - \int_A f 1_{[f<0]} \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu = \nu(A) \end{aligned}$$

holds whenever $A \in \mathcal{S}$. □

Proof of Radon-Nikodym theorem for signed measures: By lemma there exist disjoint measurable sets $S_{\pm} \in \mathcal{S}$ such that

$$\nu = \nu^+ - \nu^-, \text{ where } \nu^{\pm} \text{ are finite measures on } (S, \mathcal{S}) \text{ such that } \nu^{\pm}(A) = \nu(A \cap S_{\pm}), \; A \in \mathcal{S}.$$

Let $A \in \mathcal{S}$ be such that $\mu(A) = 0$, then $\mu(A \cap S_{\pm}) = 0$. By assumption $\nu \ll \mu$ we get that

$$\nu^{\pm}(A) = \nu(A \cap S_{\pm}) = 0.$$

Hence, we have that $\nu^{\pm} \ll \mu$, and Radon-Nikodym theorem gives that there are $f_{\pm} : (S, \mathcal{S}) \to ([0, \infty), \mathcal{B}[0, \infty))$ such that $\nu^{\pm}(A) = \int_A f_{\pm} \, \mathrm{d}\mu$ holds if $A \in \mathcal{S}$. Then $f = f_+ - f_-$ is an $\mathcal{S}$-measurable function with

$$\int |f| \, \mathrm{d}\mu \le \int f_+ \, \mathrm{d}\mu + \int f_- \, \mathrm{d}\mu = \nu^+(S) + \nu^-(S) < \infty.$$

Further, if $A \in \mathcal{S}$, then

$$\int_A f \, \mathrm{d}\mu = \int_A f_+ \, \mathrm{d}\mu - \int_A f_- \, \mathrm{d}\mu = \nu^+(A) - \nu^-(A) = \nu(A).$$

Let g, h arbitrary $\mathcal{S}$-measurable μ-integrable function such that $\nu(A) = \int_A g \, \mathrm{d}\mu$ holds if $\in \mathcal{S}$. Put $A_+ = [f > g], A_- = [f < g]$. Then $A_{\pm} \in \mathcal{S}$, and

$$\int_{A_{\pm}} (f - g) \, \mathrm{d}\mu = \int_{A_{\pm}} f \, \mathrm{d}\mu - \int_{A_{\pm}} g \, \mathrm{d}\mu = \nu(A_{\pm}) - \nu(A_{\pm}) = 0.$$

This is possible only if $\mu(A_{\pm}) = 0$, and we get that $f = g$ holds μ-almost everywhere. □

Theorem 26 Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. Then $\emptyset \neq \mathbb{E}[X|\mathcal{F}]$ is a class of equivalent elements of $\mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$ w.r.t. equality almost surely and it is the set of all Radon-Nikodym derivatives of

$$\nu(B) = \int_B X \, \mathrm{d}P = E[X; B]$$

w.r.t. $P|\mathcal{F}$. In particular, $E[X|\mathcal{F}]$ exists and it is determined uniquely almost surely.

Proof: Obviously, $\nu \ll P|\mathcal{F}$, and Radon-Nikodym theorem gives that $\frac{\mathrm{d}\nu}{\mathrm{d}P|\mathcal{F}}$ exists and it is determined uniquely up to a P-null set. Hence, we are now only to show that Y is a Radon-Nikodym derivative of ν w.r.t. $P|\mathcal{F}$ if and only if $Y \in \mathbb{E}[X|\mathcal{F}]$. If $Y \in \mathbb{E}[X|\mathcal{F}]$, then $Y \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$ and $\int_B Y \, \mathrm{d}P = \int_B X \, \mathrm{d}P = \nu(B)$ holds whenever $B \in \mathcal{F}$, and therefore Y is a Radon-Nikodym derivative of ν w.r.t. $P|\mathcal{F}$. On the other hand, let Y be a Radon-Nikodym derivative of ν w.r.t. $P|\mathcal{F}$, then Y is $\mathcal{F}$-measurable and P-integrable function such that $\int_B Y \, \mathrm{d}P = \nu(B) = \int_B X \, \mathrm{d}P$ holds whenever $B \in \mathcal{F}$, and there $Y \in \mathbb{E}[X|\mathcal{F}]$. □

Theorem 27 (*Elementary properties of conditional expectation*) Let $X, Y \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $\mathcal{F} \subseteq \mathcal{A}$ be σ-algebra.

(1) Let $a, b, c \in \mathbb{R}$, then $E[aX + bY + c|\mathcal{F}] \overset{\text{as}}{=} aE[X|\mathcal{F}] + bE[Y|\mathcal{F}] + c$.
(2) Let $X \le Y$ hold almost surely, then $E[X|\mathcal{F}] \le E[Y|\mathcal{F}]$ holds almost surely.
(3) $E[E(X|\mathcal{F})] = EX$.
(4) If $X \in L(\Omega, \mathcal{F})$, then $E[X|\mathcal{F}] \overset{\text{as}}{=} X$.
(5) Let $\mathcal{C} \subseteq \mathcal{F}$ be a σ-algebra, then $E[E(X|\mathcal{F})|\mathcal{C}] \overset{\text{as}}{=} E[E(X|\mathcal{C})|\mathcal{F}] \overset{\text{as}}{=} E[X|\mathcal{C}]$.
(6) If $\sigma(X)$ and $\mathcal{F}$ are independent, then $E[X|\mathcal{F}] \overset{\text{as}}{=} EX$.

Proof:

(1) By definition $Z = aE[X|\mathcal{F}] + bE[Y|\mathcal{F}] + c \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$ and if $B \in \mathcal{F}$, then

$$\int_B Z \, \mathrm{d}P = a \int_B E(X|\mathcal{F}) \, \mathrm{d}P + b \int_B E(X|\mathcal{F}) \, \mathrm{d}P + c = a \int_B X \, \mathrm{d}P + b \int_B X \, \mathrm{d}P + c = \int_B (aX + bY + c) \, \mathrm{d}P.$$

(2) Put $B = [E(X|\mathcal{F}) > E(Y|\mathcal{F})] \in \mathcal{F}$, and therefore we obtain from the following that $P(B) = 0$

$$\int_B (E[X|\mathcal{F}] - E[Y|\mathcal{F}]) \, \mathrm{d}P = \int_B E[X|\mathcal{F}] \, \mathrm{d}P - \int_B E[Y|\mathcal{F}] \, \mathrm{d}P = \int_B X \, \mathrm{d}P - \int_B Y \, \mathrm{d}P = \int_B (X - Y) \, \mathrm{d}P \le 0.$$

(3) $E[E(X|\mathcal{F})] = \int_\Omega E(X|\mathcal{F}) \, \mathrm{d}P = \int_\Omega X \, \mathrm{d}P = EX$.

(4) Denote $B = [X > E(X|\mathcal{F})] \in \mathcal{F}$. Then
$$\int_B (X - E[X|\mathcal{F}])\,\mathrm{d}P = \int_B X\,\mathrm{d}P - \int_B E[X|\mathcal{F}]\,\mathrm{d}P = \int_B X\,\mathrm{d}P - \int_B X\,\mathrm{d}P = 0.$$
This gives that $X \le E[X|\mathcal{F}]$ holds almost surely, and the reverse inequality can be obtained similarly.

(5) By assumption $E[X|\mathcal{C}] \in L_1(\Omega, \mathcal{C}, P|\mathcal{C}) \subseteq L_1(\Omega, \mathcal{F}, P|\mathcal{F})$, and therefore $E[E(X|\mathcal{C})|\mathcal{F}] \stackrel{\text{as}}{=} E[X|\mathcal{C}]$. Further, we will show that $E[E(X|\mathcal{F})|\mathcal{C}] \in \mathbb{E}[X|\mathcal{C}]$. Obviously, $E[E(X|\mathcal{F})|\mathcal{C}] \in \mathbb{L}_1(\Omega, \mathcal{C}, P|\mathcal{C})$. If $B \in \mathcal{C}$, then
$$\int_B E[E(X|\mathcal{F})|\mathcal{C}]\,\mathrm{d}P = \int_B E(X|\mathcal{F})\,\mathrm{d}P = \int_B X\,\mathrm{d}P.$$

(6) Obviously, $EX \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$. Since $\sigma(X)$ and $\mathcal{F}$ are independent, we get that
$$\int_B EX\,\mathrm{d}P = EX \cdot P(B) = E[X; B] = \int_B X\,\mathrm{d}P.$$
□

We say that $A \in \mathcal{A}$ is an ***atom of σ-algebra*** $\mathcal{A}$ if $\forall B \in \mathcal{A}\ B \subseteq A \Rightarrow B = \emptyset$ or $B = A$. We say that $A \in \mathcal{A}$ is an ***atom***[8] ***of probability space*** $(\Omega, \mathcal{A}, P)$ if $\forall B \in \mathcal{A}\ B \subseteq A \Rightarrow P(B) = 0$ or $P(B) = P(A)$.

Example $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ and $e \in \mathbb{E}$ then $[Y = e]$ is an atom of σ-algebra $\sigma(Y)$.

Theorem 28 Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra.

(1) If B is an atom of $\mathcal{F}$ with $P(B) > 0$, then
$$\forall \omega \in B \quad E[X|\mathcal{F}](\omega) = E[X|B], \quad \text{i.e.} \quad E[X|\mathcal{F}]1_B = E[X|B]1_B.$$

(2) If B is an atom of $(\Omega, \mathcal{F}, P|\mathcal{F})$ with $P(B) > 0$, then
$$E[X|\mathcal{F}] = E[X|B] \quad \text{holds almost surely on } B, \quad \text{i.e.} \quad E[X|\mathcal{F}]1_B \stackrel{\text{as}}{=} E[X|B]1_B.$$

Proof: (2) Let B be an atom of $(\Omega, \mathcal{F}, P|\mathcal{F})$ with $P(B) > 0$. Then $C = [E(X|\mathcal{F})1_B > E(X|B)1_B] \in \mathcal{F}$ and $C \subseteq B$. By assumption $P(C) = 0$ or $P(B \backslash C) = 0$. We are going to show that $P(C) = 0$. So, let us assume that $P(B \backslash C) = 0$. Then
$$\begin{aligned}\int_C [E(X|\mathcal{F}) - E(X|B)]\,\mathrm{d}P &= \int_C E(X|\mathcal{F})\,\mathrm{d}P - P(C)E(X|B)] \\ &= \int_B E(X|\mathcal{F})\,\mathrm{d}P - P(B)E(X|B)] = \int_B X\,\mathrm{d}P - E[X1_B] = 0,\end{aligned}$$
and we get that $P(C) = 0$, i.e. $E(X|\mathcal{F})1_B \le E(X|B)1_B$ almost surely. The contrary inequality can be obtained similarly.

(1) Put $N = [E(X|\mathcal{F})1_B \neq E(X|\mathcal{F})1_B] \in \mathcal{F}$. Then $N \subseteq B$. Since B is assumed to be an atom of $\mathcal{F}$, it is also an atom of $(\Omega, \mathcal{F}, P|\mathcal{F})$, and therefore $P(N) = 0$. Since B is assumed to be an atom of $\mathcal{F}$, and $N \subseteq B, N \in \mathcal{F}$, we get that $N = B$ or $N = \emptyset$. Since $P(B) > 0 = P(N)$, we get that $N = \emptyset$. □

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $A \in \mathcal{A}$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. Then we denote $\boldsymbol{P(A|\mathcal{F})}$= $E[1_A|\mathcal{F}]$ and it is called a ***conditional probability of A given $\mathcal{F}$***.

Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. We denote $\boldsymbol{E[X|Y]} = E[X|\sigma(Y)]$ and it is called a ***conditional expectation of real valued r.v. X given r.v. Y***. If $A \in \mathcal{A}$, then we denote $\boldsymbol{P(A|Y)}$= $P(A|\sigma(Y))$ and it is called a ***conditional probability of A given r.v. Y***.

By theorem 4, there exists a measurable function $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $E[X|Y] = f(Y)$ or $P(A|Y) = f(Y)$, respectively. Such a function will be denoted as $\boldsymbol{E[X|Y = y]}$= $f(y)$ or $\boldsymbol{P(A|Y = y)}$= $f(y)$. Note that such functions $E[X|Y = y], P(A|Y = y)$ are determined uniquely up to a P_Y-null set.

Remark If $y \in E$ is such that $P(Y = y) > 0$ then $E[X|Y = y] = E[X|B]$ holds by theorem 28 with $B = [Y = y]$, and therefore the newly introduced notation is not confusing. If $P(Y = y) = 0$, then $E[X|Y = y]$ can be arbitrary real value similarly as $f(y)$ can be arbitrary real value if f is a density of a continuous real-valued random variable.

It may happen that we are interested in $E[X|Y = y]$ similarly as we may be interested in the value $f(y)$ of the density f at point y. It is the case when we are considering a version of such a function, which is continuous at the point y.

---

[8] Usually, A is atom of a σ-algebra only if $A \neq \emptyset$ and of a probability space only if $P(A) > 0$.

Claim Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $Y : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, $k \in \mathbb{N}$. Let $f, g : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be such that $f(Y) \overset{\text{as}}{=} E[X|Y] \overset{\text{as}}{=} g(Y)$. Let $y_0 \in \mathbb{R}^k$ be such that

(1) f, g are continuous at the point y_0
(2) $\forall\, \varepsilon > 0$ $P(||Y - y_0|| < \varepsilon) > 0$.

Then $f(y_0) = g(y_0)$.

Proof: Let us assume that $f(y_0) \neq g(y_0)$. Since, f, g are continuous at y_0, we get that there exists $\varepsilon > 0$ such that $f(y) \neq g(y)$ holds whenever $||y - y_0|| < \varepsilon$. Then we get that $f(Y) \neq g(Y)$ holds with a positive probability, which is a contradiction with assumption that $f(Y) \overset{\text{as}}{=} E[X|Y] \overset{\text{as}}{=} g(Y)$. □

Corollary of theorem 28 Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_Y)$ be a discrete random variable. Then

$$E[X|Y] \overset{\text{as}}{=} \sum_{y \in E} E[X|Y = y] \cdot 1_{[Y=y]}$$

Moreover, if $P(Y = y) > 0$ holds for every $y \in E$, then

$$E[X|Y] = \sum_{y \in E} E[X|Y = y] \cdot 1_{[Y=y]}.$$

Theorem 29 Let $Y \in \mathbb{L}(\Omega, \mathcal{F})$, and $X, XY \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, where $\mathcal{F} \subseteq \mathcal{A}$ are σ-algebras. Then

$$E[XY|\mathcal{F}] \overset{\text{as}}{=} YE[X|\mathcal{F}].$$

Proof: First, we assume that $X, Y \geq 0$. Put

$$\mathcal{L} = \{U \in \mathbb{L}^+(\Omega, \mathcal{F}) : \forall\, B \in \mathcal{F}\ \textstyle\int_B UX \,\mathrm{d}P = \int_B UE[X|\mathcal{F}] \,\mathrm{d}P\}.$$

(1) $1_F \in \mathcal{L}$ holds if $F \in \mathcal{F}$, since

$$\textstyle\int_B 1_F E(X|\mathcal{F}) \,\mathrm{d}P = \int_{B\cap F} E(X|\mathcal{F}) \,\mathrm{d}P = \int_{B\cap F} X \,\mathrm{d}P = \int_B 1_F X \,\mathrm{d}P.$$

(2) Let $a, b \geq 0$ and $U, V \in \mathcal{L}$, then $W = aU + bV \in \mathcal{L}$, since

$$\begin{aligned}\textstyle\int_B WE[X|\mathcal{F}] \,\mathrm{d}P &= \textstyle a\int_B UE[X|\mathcal{F}] \,\mathrm{d}P + b\int_B VE[X|\mathcal{F}] \,\mathrm{d}P \\ &= \textstyle a\int_B UX \,\mathrm{d}P + b\int_B VX \,\mathrm{d}P = \int_B WX \,\mathrm{d}P.\end{aligned}$$

(3) If $U_n \geq 0, U_n \in \mathcal{L}, U_n \uparrow U \in \mathbb{L}^+(\Omega, \mathcal{F})$. Then $U \in \mathcal{L}$ as

$$\textstyle\int_B UE[X|\mathcal{F}] \,\mathrm{d}P = \lim\limits_{n\to\infty} \int_B U_n E[X|\mathcal{F}] \,\mathrm{d}P = \lim\limits_{n\to\infty} \int_B U_n X \,\mathrm{d}P = \int_B UX \,\mathrm{d}P.$$

Hence, we get by lemma that $\mathcal{L} = \mathbb{L}^+(\Omega, \mathcal{F}, P|\mathcal{F})$. In order to show the statement for $X, Y \geq 0$, we need to show that $YE[X|\mathcal{F}] \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$. Obviously, $YE[X|\mathcal{F}] \in \mathbb{L}(\Omega, \mathcal{F})$. Further, we obtain from the previous part of the proof that $E[YE(X|\mathcal{F})] = EXY < \infty$.

Let us consider the general case. By the first part of the proof, the statement holds for $X^\pm$ and $Y^\pm$. In particular, $Y^\pm E[X^\pm|\mathcal{F}] \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$, and

$$\begin{aligned}E[XY|\mathcal{F}] &\overset{\text{as}}{=} E[X^+Y^+ + X^-Y^- - X^-Y^+ - X^+Y^-|\mathcal{F}] \\ &\overset{\text{as}}{=} E[X^+Y^+|\mathcal{F}] + E[X^-Y^-|\mathcal{F}] - E[X^-Y^+|\mathcal{F}] - E[X^+Y^-|\mathcal{F}] \\ &\overset{\text{as}}{=} Y^+E[X^+|\mathcal{F}] + Y^-E[X^-|\mathcal{F}] - Y^+E[X^-|\mathcal{F}] - Y^-E[X^+|\mathcal{F}] \\ &\overset{\text{as}}{=} (Y^+ - Y^-)(E[X^+|\mathcal{F}] - E[X^-|\mathcal{F}]) \overset{\text{as}}{=} YE[X|\mathcal{F}].\end{aligned}$$

□

Theorem 30 Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P$, and $Y, Z : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$. Let $Y = Z$ hold on $A \in \sigma(Y) \cap \sigma(Z)$, then $E[X|Y]1_A \overset{\text{as}}{=} E[X|Z]1_A$.

Proof: By assumption, there exist $B, C \in \mathcal{E}$ such that $A = [Y \in B] = [Z \in C]$. Let us denote $\mathcal{F} = \sigma(Y) \cap \sigma(Z)$. We are going to show that $E[X|Y]1_A, E[X|Z]1_A \in \mathbb{E}[X1_A|\mathcal{F}]$. Obviously, $E[X|Y]1_A, E[X|Z]1_A$ are integrable variables. Let $f(y) = E[X|Y = y]$ and $g(z) = E[X|Z = z]$. Then

$$E[X|Y]1_A = f(Y)1_{[Y\in B]} = f(Y)1_A = f(Z)1_{[Z\in C]},$$

and therefore $E[X|Y]1_A$ is $\sigma(Y), \sigma(Z)$-measurable. If $c \in \mathbb{R}$, we get that

$$[E(X|Y)1_A < c] \in \sigma(Y) \cap \sigma(Z) = \mathcal{F},$$

i.e. $E[X|Y]1_A$ is $\mathcal{F}$-measurable, and therefore $E[X|Y]1_A \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$. Then

$$E[X|Y]1_A \overset{\text{as}}{=} E(E[X|Y]1_A|\mathcal{F}) \overset{\text{as}}{=} E(E[X1_A|Y]|\mathcal{F}) \overset{\text{as}}{=} E(X1_A|\mathcal{F}),$$

and similarly we would obtain that $E[X|Z]1_A \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$ and $E[X|Z]1_A \overset{\text{as}}{=} E(X1_A|\mathcal{F})$. □

Example Let $X \in L_1(\Omega, \mathcal{A}, P)$. Determine $E[X|X^+]$. Then

$$E[X|X^+]1_{[X^+>0]} \overset{\text{as}}{=} E[X|X]1_{[X>0]} \overset{\text{as}}{=} X1_{[X>0]} = X^+.$$

(1) If $P(X^+ = 0) = 0$, then $E[X|X^+] \overset{\text{as}}{=} X^+$.
(2) If $P(X^+ = 0) > 0$, then $E[X|X^+] \overset{\text{as}}{=} X^+ + E[X|X^+ = 0]1_{[X^+=0]} = X^+ + E[X|X \leq 0]1_{[X\leq 0]}$.

Lemma Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, let $\mathcal{C} \subseteq \mathcal{A}$ be a system closed under finite intersections. Denote $\mathcal{F} = \sigma(\mathcal{C})$ and assume that $Y \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$ is such that
(1) $EY = EX$
(2) $\forall\ B \in \mathcal{C} \quad E[X; B] = E[Y; B]$.
Then $Y \in \mathbb{E}[X|\mathcal{F}]$.

Proof: Let us denote $\mathcal{L} = \{B \in \mathcal{F} : E[X; B] = E[Y; B]\}$. Obviously, $\mathcal{L}$ is a Dynkin system containing a system $\mathcal{C}$ closed under finite intersections. By Dynkin lemma, $\mathcal{L} \supseteq \sigma(\mathcal{C}) = \mathcal{F} \supseteq \mathcal{L}$. □

Theorem 31 Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E}), Z : (\Omega, \mathcal{A}) \to (H, \mathcal{H})$. Let Z and (X, Y) be independent variables, then $E[X|(Y, Z)] \overset{\text{as}}{=} E[X|Y]$.

Proof: Put $\mathcal{F} = \sigma(Y, Z)$. Obviously, $E[X|Y] \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$, where $\mathcal{F} = \sigma(\mathcal{C})$, and where

$$\mathcal{C} = \{[Y \in B] \cap [Z \in C] : B \in \mathcal{E}, C \in \mathcal{H}\}$$

is a system closed under finite intersections. If $B \in \mathcal{E}, C \in \mathcal{H}$, and $A = [Y \in B] \cap [Z \in C]$, then

$$\begin{aligned}\int_A E[X|Y]\,\mathrm{d}P &= E[E(X|Y); Y \in B, Z \in C] = E[E(X|Y); Y \in B] \cdot P(Z \in C)\\ &= E[E(X; Y \in B|Y)] \cdot P(Z \in C) = E[X; Y \in B] \cdot P(Z \in C)\\ &= E[X; Y \in B, Z \in C] = \int_A X\,\mathrm{d}P.\end{aligned}$$

By the previous lemma, $E[X|Y] \in \mathbb{E}[X|(Y, Z)]$. □

Let $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{S}, P_X)$ and $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_Y)$. We say that the ***variables X, Y are regularly dependent*** if $P_{X,Y} \ll P_X \otimes P_Y$. Then we denote $\boldsymbol{k_{X,Y}} = \frac{\mathrm{d}P_{X,Y}}{\mathrm{d}P_X \otimes P_Y}$.

Remark If X, Y are independent, then X, Y are regularly dependent with $k_{X,Y} \equiv 1$.

Theorem 32 Let $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{S}, P_X)$ and $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_Y)$ be regularly dependent and $G \in \mathbb{L}_1(S \times E, \mathcal{S} \otimes \mathcal{E}, P_{X,Y})$. Then

$$E[G(X, Y)|Y = y] = E[G(X, y)k_{X,Y}(X, y)]$$

holds for P_Y-almost every $y \in E$.

Proof: By assumption,

$$\infty > E|G(X, Y)| = \int_{S\times E} |G(x, y)|\,\mathrm{d}P_{X,Y}(x, y) = \int_E \int_S |G(x, y)|k_{X,Y}(x, y)\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y).$$

Then $P(A) = 1$, where

$$A = \{y \in E : \int_S |G(x, y)|k_{X,Y}(x, y)\,\mathrm{d}P_X(x) < \infty\} \in \mathcal{E}.$$

Put $g(y) = \int_S G(x, y)k_{X,Y}(x, y)\,\mathrm{d}P_X(x) \cdot 1_A(y) \in \mathbb{L}_1(E, \mathcal{E}, P_Y)$. Then $g(Y) \in \mathbb{L}_1(\Omega, \sigma(Y), P|\sigma(Y))$. Let $B \in \mathcal{E}$. We get from Fubini theorem that

$$\begin{aligned}\int_{[Y\in B]} g(Y)\,\mathrm{d}P &= \int_B g\,\mathrm{d}P_Y = \int_B \int_S G(x, y)k_{X,Y}(x, y)\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y)\\ &= \int_{S\times B} G(x, y)k_{X,Y}(x, y)\,\mathrm{d}P_X \otimes \mathrm{d}P_Y(x, y)\\ &= \int_{S\times B} G(x, y)\,\mathrm{d}P_{X,Y}(x, y) = \int_{[Y\in B]} G(X, Y)\,\mathrm{d}P.\end{aligned}$$

Hence, $g(Y) \in \mathbb{E}[G(X, Y)|Y]$, and therefore we get that $E[G(X, Y)|Y = y] = g(y) = E[G(X, y)k_{X,Y}(X, y)]$ holds P_Y-almost everywhere. □

Theorem 33 Let $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{S}, P_X)$, $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_Y)$ have a joint density $f_{X,Y}$ with respect to $\mu \otimes \nu$, where μ and ν are σ-finite measures on $(S, \mathcal{S})$ and $(E, \mathcal{E})$, respectively. Then X, Y are regularly dependent variables with

$$k_{X,Y}(x,y) = \tfrac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} \cdot 1_{[f_X(x) f_Y(y) \neq 0]} \quad P_X \otimes P_Y\text{-almost everywhere,} \tag{5}$$

where

$$f_X(x) = \textstyle\int f_{X,Y}(x,y)\, \mathrm{d}\nu(y) \cdot 1_{[\int |f_{X,Y}(x,y)|\, \mathrm{d}\nu(y) < \infty]} \quad \text{is a density of r.v. } X \text{ w.r.t. } \mu\text{, i.e.} \quad f_X = \tfrac{\mathrm{d}P_X}{\mathrm{d}\mu}.$$

$$f_Y(y) = \textstyle\int f_{X,Y}(x,y)\, \mathrm{d}\mu(x) \cdot 1_{[\int |f_{X,Y}(x,y)|\, \mathrm{d}\mu(x) < \infty]} \quad \text{is a density of r.v. } Y \text{ w.r.t. } \nu\text{, i.e.} \quad f_Y = \tfrac{\mathrm{d}P_Y}{\mathrm{d}\nu}.$$

Further, $\tilde{f}_{X,Y}(x,y) = f_{X,Y}(x,y) \cdot 1_{[f_X(x) f_Y(y) \neq 0]}$ is a density of $P_{X,Y}$ w.r.t. $\mu \otimes \nu$.

Proof: First, we get from Fubini theorem that

$$\begin{aligned} P(f_X(X) = 0) &= \textstyle\int_{\{(x,y): f_X(x) = 0\}} f_{X,Y}(x,y)\, \mathrm{d}(\mu \otimes \nu)(x,y) \\ &= \textstyle\int_{\{x: f_X(x) = 0\}} \int_E f_{X,Y}(x,y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x) = \int_{\{x: f_X(x) = 0\}} f_X(x)\, \mathrm{d}\mu(x) = 0, \end{aligned}$$

and similarly, we would obtain that $P(f_Y(Y) = 0) = 0$. Second, we show that f_X is a version of $\frac{\mathrm{d}P_X}{\mathrm{d}\mu}$. Let $B \in \mathcal{S}$ and $\tilde{B} = B \cap [f_X \neq 0]$. Then $P_X(B \backslash \tilde{B}) = 0$, and

$$\textstyle\int_B f_X(x)\, \mathrm{d}\mu(x) = \int_{\tilde{B}} f_X(x)\, \mathrm{d}\mu(x) = \int_{\tilde{B}} \int_E f_{X,Y}(x,y)\, \mathrm{d}\nu(y)\, \mathrm{d}\mu(x) = P(X \in \tilde{B}) = P(X \in B).$$

Similarly, we would obtain that f_Y is a version of $\frac{\mathrm{d}P_Y}{\mathrm{d}\nu}$. Third, we show that

$$\tilde{f}_{X,Y}(x,y) = f_{X,Y}(x,y) \cdot 1_{[f_X(x) f_Y(y) \neq 0]}$$

is a density of $P_{X,Y}$ with respect to $\mu \otimes \nu$. If $B \in \mathcal{S}$ and $C \in \mathcal{E}$, we denote $\tilde{B} = \{x \in B : f_X(x) \neq 0\} \in \mathcal{S}$ and $\tilde{C} = \{y \in E : f_Y(y) \neq 0\} \in \mathcal{E}$. By the previous part of the proof $P(X \in B \backslash \tilde{B}) = 0 = P(Y \in C \backslash \tilde{C})$, and therefore

$$P_{X,Y}(B \times C) = P_{X,Y}(\tilde{B} \times \tilde{C}) = \textstyle\int_{\tilde{B} \times \tilde{C}} f_{X,Y}(x,y)\, \mathrm{d}\mu \otimes \nu(x,y) = \int_{B \times C} \tilde{f}_{X,Y}(x,y)\, \mathrm{d}\mu \otimes \nu(x,y)$$

Since $\{B \times C : B \in \mathcal{S}, C \in \mathcal{E}\}$ is a system closed under intersections generating $\mathcal{S} \otimes \mathcal{E}$, we get that it determines a probability measure on $\mathcal{S} \otimes \mathcal{E}$, and we obtain that $\mathrm{d}P_{X,Y} = \tilde{f}_{X,Y}\, \mathrm{d}\mu \otimes \nu$. Finally, we will show that the following Dynkin system

$$\mathcal{M} = \{A \in \mathcal{S} \otimes \mathcal{E} : \textstyle\int_A \tilde{k}_{X,Y}(x,y)\, \mathrm{d}[P_X \otimes P_Y](x,y) = P_{X,Y}(A)\}$$

contains the following system generating $\mathcal{S} \otimes \mathcal{E}$ closed under finite intersections

$$\mathcal{L} = \{B \times F : S \in \mathcal{S}, F \in \mathcal{H}\},$$

where $\tilde{k}_{X,Y}(x,y)$ stands for the right-hand side of (5). Then Dynkin lemma gives (5). Let $B \in \mathcal{S}, C \in \mathcal{E}$.

$$\begin{aligned} P_{X,Y}(B \times C) &= P(X \in B, Y \in C) = \textstyle\int_{B \times C} \tilde{f}_{X,Y}(x,y)\, \mathrm{d}\mu \otimes \nu(x,y) \\ &= \textstyle\int_{B \times C} \frac{\tilde{f}_{X,Y}(x,y)}{f_X(x) f_Y(y)} f_X(x) f_Y(y)\, \mathrm{d}\mu \otimes \nu(x,y) = \int_B \int_C \tilde{k}_{X,Y}(x,y) f_Y(y)\, \mathrm{d}\nu(y) f_X(x)\, \mathrm{d}\mu(x) \\ &= \textstyle\int_B \int_C \tilde{k}_{X,Y}(x,y)\, \mathrm{d}P_Y(y)\, \mathrm{d}P_X(x) = \int_{B \times C} \tilde{k}_{X,Y}(x,y)\, \mathrm{d}(P_X \otimes P_Y)(x,y). \end{aligned}$$

□

Corollary Let $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{S}, P_X)$, $Y : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E}, P_Y)$ have a joint density $f_{X,Y}$ with respect to $\mu \otimes \nu$, where μ and ν are σ-finite measures on $(S, \mathcal{S})$ and $(E, \mathcal{E})$, respectively. Let $\tilde{f}_{X,Y}, f_X, f_Y$ be densities as in theorem 33, if $G \in \mathbb{L}_1(E \times S, \mathcal{S} \otimes \mathcal{E}, P_{X,Y})$, then

$$E[\tilde{k}_{X,Y}(X,y) G(X,y)] = \frac{\int \tilde{f}_{X,Y}(x,y)\, G(x,y)\, \mathrm{d}\mu(x)}{f_Y(y)},$$

and therefore

$$E[G(X,Y)|Y = y] = \frac{\int \tilde{f}_{X,Y}(x,y)\, G(x,y)\, \mathrm{d}\mu(x)}{f_Y(y)}$$

holds for P_Y-almost every $y \in E$.

Proof: Obviously, $E[\tilde{k}_{X,Y}(X,y) G(X,y)] = \int \tilde{k}_{X,Y}(x,y) G(x,y) f_X(x)\, \mathrm{d}\mu(x) = \frac{\int \tilde{f}_{X,Y}(x,y) G(x,y)\, \mathrm{d}\mu(x)}{f_Y(y)}$.

Theorem 34 (*Jensen inequality*) Let $D \subseteq \mathbb{R}^k$ be a non-empty closed convex set and $G : D \to \mathbb{R}$ be a continuous convex function, where $k \in \mathbb{N}$. Let $X = (X_1, \ldots, X_k)^\intercal \in \mathbb{L}_1(\Omega, \mathcal{A}, P)^k$ attain values in D and let $G(X) \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$. Let $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra, then $E[X|\mathcal{F}] := (E[X_1|\mathcal{F}], \ldots, E[X_k|\mathcal{F}])^\intercal \in D$ holds almost surely, and

$$G(E[X|\mathcal{F}]) \leq E[G(X)|\mathcal{F}]$$

holds almost surely.

Theorem from convex analysis

(1) Let $K \subseteq \mathbb{R}^k$ be a convex compact set and $F \subseteq \mathbb{R}^k \backslash K$ be a closed set. Then there exists $a \in \mathbb{R}^k$ such that

$$\sup_{\kappa \in K} a^\intercal \kappa < \inf_{\varphi \in F} a^\intercal \varphi.$$

(2) Let $D \subseteq \mathbb{R}^k$ be a non-empty closed convex set and $G : D \to \mathbb{R}$ be a continuous convex function, where $k \in \mathbb{N}$, then

$$G(x) = \sup\{\mathtt{a}(x) : \mathtt{a} \text{ is affine}^9, \mathtt{a} \leq G \text{ on } D\}.$$

Proof: Let us assume that $E[X|\mathcal{F}] \notin D$ holds with a positive probability. Since $\mathbb{R}^k \backslash D$ is an open subset of $\mathbb{R}^k$, it is a countable union of closed[10] balls $B_n, n \in \mathbb{N}$, and therefore there exists $n \in \mathbb{N}$ such that $E[X|\mathcal{F}] \in B_n$ holds with a positive probability. By the above-mentioned theorem from convex analysis, there exists $a \in \mathbb{R}^k$ such that

$$\sup_{b \in B_n} a^\intercal b < \inf_{d \in D} a^\intercal d = \Delta.$$

Since X attains values in D, we get that $\Delta \leq a^\intercal X$, and therefore $\Delta \leq E[a^\intercal X|\mathcal{F}] = a^\intercal E[X|\mathcal{F}]$ holds a.s. Since $E[X|\mathcal{F}] \in B_n$ holds with a positive probability, we a contradiction that

$$a^\intercal E[X|\mathcal{F}] \leq \sup_{b \in B_n} a^\intercal b < \inf_{b \in B_n} a^\intercal b = \Delta \leq a^\intercal E[X|\mathcal{F}]$$

holds with a positive probability. Thus, we have that $E[X|\mathcal{F}] \in D$ holds almost surely.

Since G a continuous convex function on D it can be rewritten in the form

$$G(x) = \sup\{\mathtt{a}(x) : \mathtt{a} \text{ is affine}, \mathtt{a} \leq G \text{ on } D\}.$$

In order to be able to substitute $E[X|\mathcal{F}]$ into G in order to obtain equality almost surely, we need to find a countable set

$$\mathtt{B} \subseteq \mathtt{A} = \{\mathtt{a} \text{ affine} : \mathtt{a} \leq G \text{ on } D\} \quad \text{s.t.} \quad G(x) = \sup\{\mathtt{a}(x) : \mathtt{a} \in \mathtt{B}\}, \quad x \in D.$$

Now, assume that $\mathtt{a} \in \mathtt{B}$, then there exist $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$ such that $\mathtt{a}(x) = a^\intercal x + b$, and we get that

$$\mathtt{a}(E[X|\mathcal{F}]) \overset{\text{as}}{=} a^\intercal E[X|\mathcal{F}] + b \overset{\text{as}}{=} E[a^\intercal X + b|\mathcal{F}] = \overset{\text{as}}{=} E[\mathtt{a}(X)|\mathcal{F}] \leq E[G(X)|\mathcal{F}]$$

holds almost surely as $\mathtt{a}(X) \leq G(X)$ holds whenever $a \in \mathtt{B} \subseteq A$. Since $\mathtt{B}$ is a countable set, we obtain that

$$P\big(\forall \mathtt{a} \in \mathtt{B} \;\; \mathtt{a}(E[X|\mathcal{F}]) \leq E[G(X)|\mathcal{F}]\big) = 1,$$

and therefore we obtain that

$$G(E[X|\mathcal{F}]) \overset{\text{as}}{=} \sup_{\mathtt{a} \in \mathtt{B}} \mathtt{a}(E[X|\mathcal{F}]) \leq E[G(X)|\mathcal{F}].$$

holds almost surely. Now, we are going to show that the above mentioned countable subset $\mathtt{B} \subseteq \mathbb{A}$ exists. Since $D \neq \emptyset$, we get that $\mathtt{A} \neq \emptyset$. Let $\mathtt{a}_0 \in \mathtt{A}$. Then there exists $a_0 \in \mathbb{R}^k$ and $b_0 \in \mathbb{R}$ such that $\mathtt{a}_0(x) = a_0^\intercal x + b_0$ holds if $x \in D$ and we put

$$\mathtt{B} := \{\mathtt{a} \in \mathtt{A} : \forall x \in \mathbb{R}^k \; \mathtt{a}(x) = a^\intercal x + b, a - a_0 \in \mathbb{Q}^k, b - b_0 \in \mathbb{Q}\}.$$

□

Remark

(1) It is not correct just to select from $\mathtt{A}$ such affine functions $\mathtt{a}(x) = a^\intercal x + b$ such that $a \in \mathbb{Q}^k, b \in \mathbb{Q}$. Is is enough to consider the case $G(x) = a^\intercal x$, where $a \in \mathbb{R}^k \backslash \mathbb{Q}^k$ and $D = \mathbb{R}^k$.

---

[9] i.e. there exists $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$ such that $\mathtt{a}(u) = a^\intercal u + b$ whenever $u \in \mathbb{R}^k$.

[10] Since $\mathbb{R}^n$ is a separable metric space, we get that every open set is a countable union of some open balls, but every open ball is a countable union of closed balls with the same center and smaller radius.

(2) We show why we have to consider a supremum over a countable set B in the proof of Jensen inequality. Let X be a random variable with the uniform distribution on $(0,1)$. Then

$$\sup\{E[1_{\{x\}}(X)|X] : x \in (0,1)\}$$

is not defined correctly, since $E[1_{\{x\}}(X)|X]$ is determined uniquely only up to a P-null set. It can be seen as follows as $1_{\{x\}}(X), 1_{\emptyset}(X) \in \mathbb{E}[1_{\{x\}}(X)|X]$ and

$$\sup_{x\in(0,1)} E[1_{\{x\}}(X)|X] \overset{\text{as?}}{=} \sup_{x\in(0,1)} 1_{\{x\}}(X) = 1 \neq 0 = \sup_{x\in(0,1)} 1_{\emptyset}(X) \overset{\text{as?}}{=} \sup_{x\in(0,1)} E[1_{\{x\}}(X)|X].$$

Corollary (*of Jensen inequality*) Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F} \subseteq \mathcal{A}$ a σ-algebra.

(1) Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$. Then $|E[X|\mathcal{F}]| \leq E[|X||\mathcal{F}]$ holds almost surely.

(2) Let $X, X_n \in \mathbb{L}_1(\Omega, \mathcal{A}, P), n \in \mathbb{N}$ be such that $X_n \xrightarrow{\mathbb{L}_1} X$ as $n \to \infty$, then $E[X_n|\mathcal{F}] \xrightarrow{\mathbb{L}_1} E[X|\mathcal{F}]$ as

$$E\big|E[X_n|\mathcal{F}] - E[X|\mathcal{F}]\big| = E\big|E[X_n - X|\mathcal{F}]\big| \leq E[E(|X - X_n||\mathcal{F})] = E|X_n - X| \to 0 \quad \text{as} \quad n \to \infty.$$

Theorem 35 (*Lévy*) Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra.

(1) Let $X_n \in \mathbb{L}(\Omega, \mathcal{A})$ be such that $0 \leq X_n \leq X_{n+1}$ holds almost surely for every $n \in \mathbb{N}$.

(2) Let $X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ be such that $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$.

Then $E[X_n|\mathcal{F}] \xrightarrow{\text{as}} E[X|\mathcal{F}]$ as $n \to \infty$.

Proof: Since $0 \leq X_n \leq X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ holds almost surely, we get that $X_n \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, and that $0 \leq Y_n \leq E[X|\mathcal{F}]$ holds almost surely, where $Y_n := E[X_n|\mathcal{F}]$. Then we get that

$$Y := \sup_{n\in\mathbb{N}} Y_n \in \mathbb{L}_1^*(\Omega, \mathcal{F}, P|\mathcal{F})$$

as $0 \leq Y \leq E[X|\mathcal{F}]$ holds a.s. Then $Z = Y 1_{[Y<\infty]} \in \mathbb{L}_1(\Omega, \mathcal{F}, P|\mathcal{F})$. We are going to show that $Z \in \mathbb{E}[X|\mathcal{F}]$. Let $B \in \mathcal{F}$, then

$$\int_B Z \,\mathrm{d}P = \int_B Y \,\mathrm{d}P = \lim_{n\to\infty} \int_B Y_n \,\mathrm{d}P = \lim_{n\to\infty} \int_B X_n \,\mathrm{d}P = \int_B X \,\mathrm{d}P$$

holds by Monotone Convergence Theorem as $0 \leq Y_n \uparrow Y$ holds almost surely as $n \to \infty$. □

Theorem 36 (*Lebesgue*) Let $X, X_n \in \mathbb{L}(\Omega, \mathcal{A}), n \in \mathbb{N}$ and $Z \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ be such that

(1) $|X_n| \leq Z$ holds almost surely for every $n \in \mathbb{N}$.

(2) $X_n \xrightarrow{\text{as}} X$ as $n \to \infty$

Then $X_n, X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $E[X_n|\mathcal{F}] \xrightarrow{\text{as}} E[X|\mathcal{F}]$ as $n \to \infty$ whenever $\mathcal{F} \subseteq \mathcal{A}$ is a σ-algebra.

Proof: By theorem 13 (6), $X_n, X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$. Further,

$$X_n \leq \bar{X}_n = \sup_{k\geq n} X_k \xrightarrow{\text{as}} X \quad \text{as} \quad n \to \infty,$$

and

$$0 \leq Y_n := Z - \bar{X}_n \leq Y_{n+1} \xrightarrow{\text{as}} Y := Z - X \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$$

as $n \to \infty$. By theorem 35,

$$E[X_n|\mathcal{F}] \overset{\text{as}}{\leq} E[\bar{X}_n|\mathcal{F}] \overset{\text{as}}{=} E[Z|\mathcal{F}] - E[Y_n|\mathcal{F}] \xrightarrow{\text{as}} E[Z|\mathcal{F}] - E[Y|\mathcal{F}] \overset{\text{as}}{=} E[X|\mathcal{F}].$$

Hence, we get that

$$\limsup_{n\to\infty} E[X_n|\mathcal{F}] \overset{\text{as}}{\leq} E[X|\mathcal{F}].$$

The same inequality for $V_n = -X_n$ and $V = -X$ gives that

$$\liminf_{n\to\infty} E[X_n|\mathcal{F}] \overset{\text{as}}{=} -\limsup_{n\to\infty} E[V_n|\mathcal{F}] \overset{\text{as}}{\geq} -E[V|\mathcal{F}] \overset{\text{as}}{=} E[X|\mathcal{F}],$$

and therefore $E[X_n|\mathcal{F}] \xrightarrow{\text{as}} E[X|\mathcal{F}]$. □

Theorem 37 Let $X \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. Then

(1) $E[X|\mathcal{F}] \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$.

(2) If $Y \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$, then $EY(X - E[X|\mathcal{F}]) = 0$.

(3) If $Z \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$, then $E(X - E[X|\mathcal{F}])^2 \leq E(X - Z)^2$, i.e.

$$E(X - E[X|\mathcal{F}])^2 = \min\{E(X - Z)^2 : Z \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})\}.$$

Proof: (1) We obtain from Jensen inequality, theorem 34, that $(E[X|\mathcal{F}])^2 \overset{\text{as}}{\le} E[X^2|\mathcal{F}]$, and therefore

$$E|E[X|\mathcal{F}]|^2 \le EE[X^2|\mathcal{F}] = EX^2 < \infty.$$

(2) By (1), $Z := X - E[X|\mathcal{F}] \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$. By Schwartz inequality $E|ZY| \le \sqrt{EZ^2EY^2} < \infty$ holds if $Y \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$. Let $Y \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$, then $Y(X - E[X|\mathcal{F}]) \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, and

$$EY(X - E[X|\mathcal{F}]) = EE[Y(X - E[X|\mathcal{F}])|\mathcal{F}] = E\{YE(X - E[X|\mathcal{F}]|\mathcal{F})\} = E\{Y \cdot 0\} = 0.$$

(3) Let $Z \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$, then $Y = E[X|\mathcal{F}] - Z \in \mathbb{L}_2(\Omega, \mathcal{F}, P|\mathcal{F})$, and (2) gives that

$$\begin{aligned} E(X - Z)^2 = E(X - E[X|\mathcal{F}] + Y)^2 &= E(X - E[X|\mathcal{F}])^2 + 2EY(X - E[X|\mathcal{F}]) + EY^2 \\ &= E(X - E[X|\mathcal{F}])^2 + EY^2 \ge E(X - E[X|\mathcal{F}])^2. \end{aligned}$$

□

Let $X_k \in \mathbb{L}_1(\Omega, \mathcal{F}, P), k \le n \in \mathbb{N}$. Then

$$\mathrm{Var}(X) = E(X - EX)(X - EX)^{\intercal}$$

is called a ***variance matrix*** of random vector $X = (X_1, \dots, X_n)^{\intercal}$. If $A, B \in \mathbb{R}^{n\times n}$ we write $\boldsymbol{A \le B}$ if $B - A$ is a positively definite matrix, i.e. if $C = B - A \ge 0$. This means that

$$\forall \lambda \in \mathbb{R}^n \quad \lambda^{\intercal} C\lambda \ge 0, \qquad \text{i.e.} \qquad \lambda^{\intercal} A\lambda \le \lambda^{\intercal} B\lambda.$$

Theorem 38 Let $X_k, Y_k \in \mathbb{L}_2(\Omega, \mathcal{A}, P), k \le n$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. If $\sigma(Y) \subseteq \mathcal{F}$, then

$$\mathrm{Var}(X - E[X|\mathcal{F}]) \le E(X - Y)(X - Y)^{\intercal},$$

where $X = (X_1, \dots, X_n)^{\intercal}, Y = (Y_1, \dots, Y_n)^{\intercal}$.

Proof: Let $\lambda \in \mathbb{R}^n$, then $\lambda^{\intercal} X, \lambda^{\intercal} Y \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$ and $E[\lambda^{\intercal} X|\mathcal{F}] = \lambda^{\intercal} E[X|\mathcal{F}]$. By theorem 37

$$\begin{aligned} \lambda^{\intercal} E(X - Y)(X - Y)^{\intercal}\lambda &= E[(\lambda^{\intercal} X - \lambda^{\intercal} Y)(\lambda^{\intercal} X - \lambda^{\intercal} Y)^{\intercal}] \\ &\ge E[(\lambda^{\intercal} X - E[\lambda^{\intercal} X|\mathcal{F}])(\lambda^{\intercal} X - E[\lambda^{\intercal} X|\mathcal{F}])^{\intercal}] \\ &= \mathrm{Var}(\lambda^{\intercal} X - E[\lambda^{\intercal} X|\mathcal{F}]) = \lambda^{\intercal}\mathrm{Var}(X - E[X|\mathcal{F}])\lambda. \end{aligned}$$

□

Theorem 39 (*Wald*) Let $X_n, n \in \mathbb{N}$ be a sequence of independent identically distributed real-valued random variables on a probability space $(\Omega, \mathcal{A}, P)$ independent with a random variable N attaining only values in $\mathbb{N}_0$. Let us consider the following random sum

$$S = \sum_{n=1}^{N} X_n.$$

(1) If $X_1, N \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$, then

$$E[S|N] \overset{\text{as}}{=} N \cdot EX_1.$$

(2) If $N \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$ and $X_1 \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$, then

$$\mathrm{var}(S|N) := E[(S - E[S|N])^2|N] \overset{\text{as}}{=} N \cdot \mathrm{var}(X_1).$$

(3) If $\alpha \in \mathbb{R}\backslash\{0\}$ is such that $e^{\alpha X_1} \in \mathbb{L}_1$, then

$$E[e^{\alpha S}(Ee^{\alpha X_1})^{-N}|N] \overset{\text{as}}{=} 1, \quad \text{i.e.} \quad E\left[\prod_{n=1}^{N} \frac{\exp\{\alpha X_n\}}{E\exp\{\alpha X_n\}}\middle| N\right] \overset{\text{as}}{=} 1.$$

Proof: (1) By theorem 20, $S \in \mathbb{L}_1(\Omega, \mathcal{A}, P)$. Let $n \in \mathbb{N}_0$, then

$$E[S1_{[N=n]}] = E\left(\sum_{k=1}^{n} X_k 1_{[N=n]}\right) = EX_1\, nP(N = n)$$

If $P(N = n) > 0$, then $E[S|N = n] = \frac{E[S1_{[N=n]}]}{P(N=n)} = nEX_1$, and therefore

$$E[S|N] \overset{\text{as}}{=} \sum_{n=0}^{\infty} E[S|N = n]1_{[N=n]} \overset{\text{as}}{=} \sum_{n=0}^{\infty} nEX_1 1_{[N=n]} = N \cdot EX_1.$$

(2) Denote $\mathbb{S} = S - NEX_1 \overset{\text{as}}{=} S - E[S|N]$. If $n \in \mathbb{N}_0$, then

$$E[\mathbb{S}^2 1_{[N=n]}] = E\big[\big(\sum_{k=1}^{n}(X_k - EX_k)\big)^2 1_{[N=n]}\big] = P(N=n)\, E\big(\sum_{k=1}^{n} X_k - EX_k\big)^2$$
$$= P(N=n)\operatorname{var}\big(\sum_{k=1}^{n} X_k\big) = P(N=n)\, n \operatorname{var}(X_1),$$

and therefore if $P(N=n) > 0$, then $E[\mathbb{S}^2|N=n] = n\operatorname{var}(X_1)$. Hence,

$$\operatorname{var}(S|N) = E[\mathbb{S}^2|N] \overset{\text{as}}{=} \sum_{n=0}^{\infty} E[\mathbb{S}^2|N=n]\cdot 1_{[N=n]} \overset{\text{as}}{=} \sum_{n=0}^{\infty} n\operatorname{var}(X_1)\cdot 1_{[N=n]} = N\cdot\operatorname{var}(X_1).$$

(3) Denote $Y_n = \exp\{\alpha X_n\}/E\exp\{\alpha X_n\}$ and $Z = \prod_{n=1}^{N} Y_n$. Let $n \in \mathbb{N}_0$, then

$$E[Z1_{[N=n]}] = E\left(1_{[N=n]}\cdot\prod_{k=1}^{n} Y_k\right) = P(N=n)\prod_{k=1}^{n} EY_k = P(N=n)$$

as $EY_k = 1$. Then

$$E[e^{\alpha S}(Ee^{\alpha X_1})^{-N}|N] \overset{\text{as}}{=} E[Z|N] \overset{\text{as}}{=} \sum_{n=0}^{\infty} E[Z|N=n]\cdot 1_{[N=n]} \overset{\text{as}}{=} 1.$$

□

7. Random measures

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{F} \subseteq \mathcal{A}$ a σ-algebra, and $(S, \mathcal{S})$ be a measurable space. A function $\mu : \mathcal{S}\times\Omega \to \mathbb{R}$ is called an ***$(\mathcal{S},\mathcal{F})$-random probability measure*** if

(1) $\omega \in \Omega \mapsto \mu(D,\omega)$ is an $\mathcal{F}$-measurable function whenever $D \in \mathcal{S}$.
(2) $\mu(S,\omega) \overset{\text{as}}{=} 1$, $\mu(D,\omega) \overset{\text{as}}{\geq} 0$ whenever $D \in \mathcal{S}$, and if $D_n \in \mathcal{S}$ are pairwise disjoint, then

$$\mu\big(\bigcup_{n\in\mathbb{N}} D_n, \omega\big) \overset{\text{as}}{=} \sum_{n\in\mathbb{N}} \mu(D_n,\omega).$$

Example Let $X : (\Omega,\mathcal{A}) \to (S,\mathcal{S})$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. Put $\mu(D,\omega) = P(X \in D|\mathcal{F})(\omega)$. Then μ is an $(\mathcal{S},\mathcal{F})$-random probability measure. First, if $D \in \mathcal{F}$, then $P(X \in D|\mathcal{F}) \in \mathbb{L}_1(\Omega,\mathcal{F},P|\mathcal{F})$ holds by the definition. Second, $P(X \in S|\mathcal{F}) \overset{\text{as}}{=} 1$ and $P(X \in D|\mathcal{F}) \overset{\text{as}}{\geq} 0$ if $D \in \mathcal{S}$ hold by the elementary properties of conditional expectation. Further, if $D_n \in \mathcal{S}$ are pairwise disjoint, then we obtain from the definition that

$$P(X \in \cup_n D_n|\mathcal{F}) \overset{\text{as}}{=} E[1_{[X\in\cup_n D_n]}|\mathcal{F}] \overset{\text{as}}{=} E[\textstyle\sum_n 1_{[X\in D_n]}|\mathcal{F}] \overset{\text{as}}{=} \textstyle\sum_n E[1_{[X\in D_n]}|\mathcal{F}] \overset{\text{as}}{=} \textstyle\sum_n P(X \in D_n|\mathcal{F}).$$

Let $(\Omega,\mathcal{A},P)$ be a probability space and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra and $(S,\mathcal{S})$ be a measurable space. Let $\mu,\nu : \mathcal{S}\times\Omega \to \mathbb{R}$ be $(\mathcal{S},\mathcal{F})$-random probability measures. We say that ν is a ***regular version*** of μ if

(1) $\nu(D,\omega) \overset{\text{as}}{=} \mu(D,\omega)$ whenever $D \in \mathcal{S}$
(2) $D \in \mathcal{S} \mapsto \nu(D,\omega)$ is a probability measure whenever $\omega \in \Omega$.

Theorem 40 Let (S,d) be a separable and complete metric space (or generally a Polish space[11]). Let $(\Omega,\mathcal{A},P)$ be a probability space and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra. Then every $(\mathcal{B}(S),\mathcal{F})$-random probability measure has a regular version.

Let $X : (\Omega,\mathcal{A},P) \to (S,\mathcal{S},P_X)$ and $Y : (\Omega,\mathcal{A},P) \to (H,\mathcal{H},P_Y)$ be random variables. A ***conditional distribution of X given Y*** is a function $\boldsymbol{P_{X|Y}} : (B|y) \in \mathcal{S}\times H \mapsto P_{X|Y}(B|y) \in [0,1]$ satisfying

(1) $B \in \mathcal{S} \mapsto P_{X|Y}(B|y)$ is a probability measure whenever $y \in H$
(2) $y \in H \mapsto P_{X|Y}(B|y)$ is an $\mathcal{H}$-measurable function whenever $B \in \mathcal{S}$.
(3) If $B \in \mathcal{S}$ and $C \in \mathcal{H}$, then

$$P(X \in B, Y \in C) = \textstyle\int_C P_{X|Y}(B|y)\,\mathrm{d}P_Y(y).$$

[11]A separable topological space is called a Polish space if there exists a complete metric generating the same topology. A metric space is called Polish its generate a Polish topology, i.e. if there exists an equivalent complete and separable metric.

The set of all conditional distributions of X given Y is denoted as $\boldsymbol{\mathcal{L}(X|Y)}$.

Theorem 41 Let (S,d) be a Polish space, $X:(\Omega,\mathcal{A},P)\to(S,\mathcal{B}(S),P_X), Y:(\Omega,\mathcal{A},P)\to(H,\mathcal{H},P_Y)$ be random variables. Then there exists a conditional distribution of X given Y.

Proof: If $D\in\mathcal{B}(S)$, we put $\nu(D,y)=P(X\in D|Y=y)$, and we immediately see that $y\in H\mapsto\nu(D,y)$ is an $\mathcal{H}$-measurable function. Further,

(1) $\mu(S,y)=P(X\in S|Y=y)=1$ holds for P_Y-almost every $y\in H$ as $P(X\in S|Y)\overset{\text{as}}{=}1$.
(2) $\mu(D,y)=P(X\in D|Y=y)\geq 0$ holds for P_Y-almost every $y\in H$ as $P(X\in B|Y)\overset{\text{as}}{\geq}0$.
(3) If $D_n\in\mathcal{B}(S)$ are pairwise disjoint, then

$$P_Y\{y\in H:\mu(\cup_n D_n,y)=\textstyle\sum_n\mu(D_n,y)\}=P[\mu(\cup_n D_n,Y)=\sum_n\mu_n(D_n,Y)]=1$$

as

$$\mu(\cup_n D_n,Y)\overset{\text{as}}{=}P(X\in\cup_n D_n|Y)\overset{\text{as}}{=}\textstyle\sum_n P(X\in D_n|Y)\overset{\text{as}}{=}\sum_n\mu_n(D_n,Y).$$

Hence, μ is an $(\mathcal{B}(S),\mathcal{H})$-random probability measure. By theorem 40, it has a regular version ν, i.e.

(1) $\nu(D,y)=\mu(D,y)=P(X\in D|Y=y)$ holds for P_Y-almost every $y\in H$ whenever $D\in\mathcal{B}(S)$
(2) $D\in\mathcal{B}(S)\mapsto\nu(D,y)$ is a probability measure whenever $y\in H$
(3) $y\in H\mapsto\nu(D,y)$ is an $\mathcal{H}$-measurable function whenever $D\in\mathcal{B}(S)$.

Further, if $B\in\mathcal{B}(S)$ and $C\in\mathcal{H}$, then

$$\begin{aligned}\textstyle\int_C\nu(B,y)\,\mathrm{d}P_Y(y)&=\textstyle\int_C P(X\in B|Y=y)\,\mathrm{d}P_Y(y)=\int_{[Y\in C]}P(X\in B|Y)\,\mathrm{d}P\\&=\textstyle\int_{[Y\in C]}1_{[X\in B]}\,\mathrm{d}P=P(X\in B,Y\in C).\end{aligned}$$

Hence, ν is a conditional distribution of X given Y. □

Theorem 42 Let $X:(\Omega,\mathcal{A},P)\to(S,\mathcal{S},P_X)$ and $Y:(\Omega,\mathcal{A},P)\to(H,\mathcal{H},P_Y)$ be random elements. Let $G\in\mathbb{L}_1(S\times H,\mathcal{S}\otimes\mathcal{H},P_{X,Y})$ and $P_{X|Y}$ be a conditional distribution of X given Y. Then

$$E[G(X,Y)|Y=y]=\textstyle\int_S G(x,y)\,\mathrm{d}P_{X|Y}(x|y)\qquad P_Y\text{-a.e.}\tag{6}$$

Proof: If $F=B\times C$, where $B\in\mathcal{S}$ and $C\in\mathcal{H}$, put

$$m_F(y)=\textstyle\int_S G(x,y)\,\mathrm{d}P_{X|Y}(x|y)=\int_B 1_C(y)\,\mathrm{d}P_{X|Y}(x|y)=P_{X|Y}(B|y)\cdot 1_C(y).$$

By definition, $m_F(y)$ is an $\mathcal{H}$-measurable and with values in $[0,1]$. Hence, $m_F\in\mathbb{L}_1(H,\mathcal{H},P_Y)$, and therefore $m_F(Y)\in\mathbb{L}_1(\Omega,\sigma(Y),P|\sigma(Y))$. Let $D\in\mathcal{H}$, then

$$\begin{aligned}\textstyle\int_{[Y\in D]}m_F(Y)\,\mathrm{d}P&=\textstyle\int_D m_F(y)\,\mathrm{d}P_Y(y)=\int_D P_{X|Y}(B|y)\cdot 1_C(y)\,\mathrm{d}P_Y(y)\\&=\textstyle\int_{C\cap D}P_{X|Y}(B|y)\,\mathrm{d}P_Y(y)=P(X\in B,Y\in C\cap D)=\int_{[Y\in D]}1_F(X,Y)\,\mathrm{d}P.\end{aligned}$$

Hence, $m_F(Y)\in\mathbb{E}[1_F(X,Y)|Y]$, and therefore

$$F\in\mathcal{L}:=\{B\times C:B\in\mathcal{S},C\in\mathcal{H}\}\subseteq\{F\in\mathcal{S}\otimes\mathcal{H},(6)\text{ holds for }G=1_F\}=:\mathcal{M}.$$

Obviously, $\sigma(\mathcal{L})=\mathcal{S}\otimes\mathcal{H}$, the set $\mathcal{L}$ is closed under finite intersections, and $\mathcal{M}$ is a Dynkin system. By Dynkin lemma, $\sigma(\mathcal{L})=\mathcal{S}\otimes\mathcal{H}\subseteq\mathcal{M}$. Hence,

(1) $1_F\in\mathcal{K}=\{G\in\mathbb{L}^+(S\times H,\mathcal{S}\otimes\mathcal{H}):(6)\text{ holds }\}$ if $F\in\mathcal{S}\otimes\mathcal{H}$.
(2) If $a,b\geq 0$ and $G_1,G_2\in\mathcal{K}$, then obviously $aG_1+bG_2\in\mathcal{K}$.
(3) If $G_n\in\mathcal{K}$ and $G_n\uparrow G\in\mathbb{L}^+(S\times H,\mathcal{S}\otimes\mathcal{H})$. Then $G\in\mathcal{K}$ holds by Monotone Convergence Theorem.

By lemma, $\mathcal{K}=\mathbb{L}^+(S\times H,\mathcal{S}\otimes\mathcal{H})$. Then $G^\pm\in\mathcal{K}$ and $G^\pm(X,Y)\in\mathbb{L}_1(\Omega,\mathcal{A},P)$ and therefore

$$\begin{aligned}E[G(X,Y)|Y=y]&=E[G^+(X,Y)|Y=y]-E[G^-(X,Y)|Y=y]\\&=\textstyle\int_S G^+(x,y)\,\mathrm{d}P_{X|Y}(x|y)-\int_S G^-(x,y)\,\mathrm{d}P_{X|Y}(x|y)=\int_S G(x,y)\,\mathrm{d}P_{X|Y}(x|y)\end{aligned}$$

holds for P_Y-almost every $y\in H$. □

Theorem 43 Let $X:(\Omega,\mathcal{A},P)\to(S,\mathcal{S},P_X)$ and $Y:(\Omega,\mathcal{A},P)\to(H,\mathcal{H},P_Y)$ be random elements and $G:(S\times H,\mathcal{S}\otimes\mathcal{H})\to(T,\mathcal{T})$ and $P_{X|Y}$ be a conditional distribution of X given Y. Then $\zeta\in\mathcal{L}(G(X,Y)|Y)$, where

$$\zeta:(B,y)\in\mathcal{T}\times H\mapsto P(G(X,y)\in B)\in[0,1].$$

Proof:

(1) Let $y\in E$, then $B\in\mathcal{T}\mapsto P(G(X,y)\in B)$ is a probability measure.

(2) Let $B \in \mathcal{T}$, then $y \in H \mapsto P(G(X,y) \in B)$ is an $\mathcal{H}$-measurable function.
(3) Let $B \in \mathcal{T}, C \in \mathcal{H}$, then

$$\begin{aligned}\int_C \zeta(B,y)\,\mathrm{d}P_Y(y) &= \int_C P(G(X,y) \in B)\,\mathrm{d}P_Y(y) = \int_C \int_S 1_{[G(x,y)\in B]}\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y) \\ &= \int_{S\times C} 1_{[G(x,y)\in B]}\,\mathrm{d}(P_X \otimes \mathrm{d}P_Y)(x,y) = \int_{S\times C} 1_{[G(x,y)\in B]}\,\mathrm{d}P_{X,Y}(x,y) \\ &= P(G(X,Y) \in B, Y \in C).\end{aligned}$$

□

Theorem 44 Let $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{S}, P_X)$ and $Y : (\Omega, \mathcal{A}, P) \to (H, \mathcal{H}, P_Y)$ have a joint density $f_{X,Y}$ with respect to $\mu \otimes \nu$, where μ, ν are σ-finite measures on $(S, \mathcal{S})$ and $(H, \mathcal{H})$, respectively. Denote

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\, 1_{[f_Y(y)\neq 0]}.$$

Then $\zeta \in \mathcal{L}(X|Y)$, where

$$\zeta(B|y) = \int_B f_{X|Y}(x|y)\,\mathrm{d}\mu(x) \cdot 1_{[f_Y(y)>0]} + P_Y(B) \cdot 1_{[f_Y(y)=0]}, \quad B \in \mathcal{S}, y \in E.$$

The function ***$f_{X|Y}(x|y)$*** is called a ***conditional density of X given Y*** with respect to $\mu \otimes \nu$.

Proof: We know that $f_Y \geq 0$ is a density of Y with respect to ν.
(1) Let $y \in E$, then $B \in \mathcal{S} \mapsto \zeta(B|y)$ is a measure on $(S, \mathcal{S})$ with $\zeta(S|y) = 1$.
(2) Let $B \in \mathcal{S}$, then Fubini theorem gives that $y \in H \mapsto \zeta(B|y)$ is an $\mathcal{H}$-measurable, since $f_{X|Y}$ is $\mathcal{S} \otimes \mathcal{E}$-measurable and $N = \{y \in H : f_Y(y) \neq 0\} \in \mathcal{E}$.
(3) Let $B \in \mathcal{S}, C \in \mathcal{E}$. Then

$$\begin{aligned}\int_C \zeta(B|y)\,\mathrm{d}P_Y(y) &= \int_C \zeta(B|y) f_Y(y)\,\mathrm{d}\nu(y) = \int_C \int_B f_{X|Y}(x|y) f_Y(y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \\ &= \int_{\tilde{C}} \int_B f_{X,Y}(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) = P_{X,Y}(B \times \tilde{C}) \\ &= P(X \in B, Y \in \tilde{C}) = P(X \in B, Y \in C),\end{aligned}$$

where $\tilde{C} = \{y \in C : f_Y(y) > 0\} \in \mathcal{H}$ is such that $P_Y(C \setminus \tilde{C}) = \int_{C\setminus\tilde{C}} f_Y(y)\,\mathrm{d}\nu(y) = 0$. □

8. 0-1 laws

Let $(\Omega, \mathcal{A})$ be a measurable space and $\mathcal{F}_n \subseteq \mathcal{A}, n \in \mathbb{N}$ be a sequence of σ-algebras. A ***residual σ-algebra*** is the following σ-algebra

$$\mathcal{F}^\infty = \bigcap_{n\in\mathbb{N}} \sigma\Big(\bigcup_{k\geq n} \mathcal{F}_k\Big).$$

Its elements are called ***residual events***.

Example Let $F_n \in \mathcal{F}_n$, then

$$\limsup_{n\to\infty} F_n = \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} F_n \in \mathcal{F}^\infty \quad \& \quad \liminf_{n\to\infty} F_n = \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} F_n \in \mathcal{F}^\infty.$$

Theorem 45 (*Kolmogorov 0-1 law*) Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $\mathcal{F}_n \subseteq \mathcal{A}, n \in \mathbb{N}$ be independent σ-algebras, then $P(F) \in \{0,1\}$ holds whenever $F \in \mathcal{F}^\infty$.

Proof: Let $n \in \mathbb{N}, F_k \in \mathcal{F}_k, n \leq K \in \mathbb{N}$, then

$$P\Big(\bigcap_{k=1}^K F_k\Big) = \prod_{k=1}^K P(F_k) = \prod_{k=1}^n P(F_k) \prod_{k=n+1}^K P(F_k) = P\Big(\bigcap_{k=1}^n F_k\Big) \cdot P\Big(\bigcap_{k=n+1}^K F_k\Big).$$

Hence, the system

$$\mathcal{L}_n = \{\bigcap_{k=1}^n F_k : F_k \in \mathcal{F}_k : k \leq n\}$$

is closed under finite intersections and it is independent with

$$\mathcal{L}^n = \{\bigcap_{k=n+1}^K F_k : F_k \in \mathcal{F}_k, k > n, K > n\},$$

which is also closed under finite intersections. Hence, $\sigma(\mathcal{L}_n) = \sigma(\mathcal{F}_1 \cup \ldots \cup \mathcal{F}_n)$ and $\sigma(\mathcal{L}^n) = \sigma(\cup_{k \geq n} \mathcal{F}_k)$ are also independent. Since $\mathcal{F}^\infty \subseteq \sigma(\cup_{k>n} \mathcal{F}_k)$, we get that $\mathcal{F}^\infty$ is a system independent with $\sigma(\mathcal{F}_1 \cup \ldots \cup \mathcal{F}_n)$. Then we get that $\mathcal{F}^\infty$ is a system independent with

$$\mathcal{L} = \bigcup_{n \in \mathbb{N}} \sigma(\mathcal{F}_1 \cup \ldots \cup \mathcal{F}_n).$$

Hence, we get that $\mathcal{F}^\infty$ are independent with $\sigma(\mathcal{L}) = \sigma(\cup_n \mathcal{F}_n) \supseteq \mathcal{F}^\infty$. Hence, if $F \in \mathcal{F}^\infty$, then F, F are independent sets. Then $P(F) = P(F \cap F) = P(F)^2$, and therefore $P(F) \in \{0, 1\}$. $\square$

Lemma (*Cantelli*) Let $F_n \in \mathcal{A}, n \in \mathbb{N}$ be such that $\sum_n P(F_n) < \infty$, then $P(\limsup_n F_n) = 0$.

Theorem 46 (*Borel-Cantelli*) Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $F_n \in \mathcal{A}, n \in \mathbb{N}$ be independent. Then

$$P(\limsup_{n \to \infty} F_n) = 0 \quad \equiv \quad \sum_{n=1}^{\infty} P(F_n) < \infty \qquad \& \qquad P(\limsup_{n \to \infty} F_n) = 1 \quad \equiv \quad \sum_{n=1}^{\infty} P(F_n) = \infty.$$

Theorem 47 Let $X_n, n \in \mathbb{N}$ be independent random variables. Denote $X = (X_n, n \in \mathbb{N})$ and

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \{\mathbb{R}^n \times A : A \in \mathcal{B}(\mathbb{R}^\mathbb{N})\}$$

σ-algebra on $\mathbb{R}^\mathbb{N}$. If $g : (\mathbb{R}^\mathbb{N}, \mathcal{T}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, then $g(X)$ is a degenerate random variable.[12]

Proof: It is enough to show that $c \in \mathbb{R} \mapsto P(g(X) < c)$ attains values in $\{0, 1\}$. Let $c \in \mathbb{R}$, then

$$B = \{x \in \mathbb{R}^\mathbb{N} : g(x) < c\} \in \mathcal{T}.$$

If $n \in \mathbb{N}$, then there exists $A_n \in \mathcal{B}(\mathbb{R}^\mathbb{N})$ such that $B = \mathbb{R}^n \times A_n$. Denote $\mathcal{F}_n = \sigma(X_n)$. Then

$$[g(X) < c] = [(X_{n+k}, k \in \mathbb{N}) \in A_n] \in \sigma(\bigcup_{k \in \mathbb{N}} \mathcal{F}_{n+k}).$$

Hence,

$$[g(X) < c] \in \bigcup_{n \in \mathbb{N}} \sigma(\bigcup_{k \geq n} \mathcal{F}_k) = \mathcal{F}^\infty$$

Since $\mathcal{F}_n = \sigma(X_n)$ are independent, we get by theorem 45 that $P[g(X) < c] \in \{0, 1\}$. $\square$

Theorem 48 Let $X_n, n \in \mathbb{N}$ be sequence of independent identically distribution real-valued random variables on $(\Omega, \mathcal{A}, P)$. Let $0 < b_n \to \infty$ as $n \to \infty$, then there exists $c, d \in \bar{\mathbb{R}}$ such that

$$\limsup_{n \to \infty} \frac{1}{b_n} \sum_{k=1}^{n} X_k \stackrel{\text{as}}{=} c \quad \& \quad \liminf_{n \to \infty} \frac{1}{b_n} \sum_{k=1}^{n} X_k \stackrel{\text{as}}{=} d.$$

Proof: We show only the first part of the statement, the second part can be obtained from the first one immediately. Obviously, if $m \in \mathbb{N}$, then

$$g(x) := \limsup_{n \to \infty} \frac{1}{b_n} \sum_{k=1}^{n} x_k = \limsup_{n \to \infty} \frac{1}{b_n} \sum_{k=m+1}^{n} x_k =: g_m(x)$$

holds. Let $c \in \mathbb{R}$, then there exist $A_m \in \mathcal{B}(\mathbb{R}^\mathbb{N})$ such that

$$\{x \in \mathbb{R}^\mathbb{N} : g(x) < c\} = \{x \in \mathbb{R}^\mathbb{N} : g_m(x) < c\} = \mathbb{R}^m \times A_m$$

whenever $m \in \mathbb{N}$. Hence, $\{x \in \mathbb{R}^\mathbb{N} : g(x) < c\} \in \mathcal{T}$. Now, it is enough to apply theorem 47. $\square$

---

[12] i.e. there exists $c \in \bar{\mathbb{R}}$ such that $g(X) = \bar{c}$ holds almost surely.

9. Summability of sequence of real-valued random elements

Lemma(*Skorochod inequality*) Let $X_1, \ldots, X_n$ be independent random variables, denote $S_n = \sum_{k=1}^n X_k$. Let $\varepsilon > 0$. Then

$$P(|S_n| > \varepsilon) \geq P(\max_{k=1,\ldots,n} |S_k| > 2\varepsilon) \cdot \min_{k=1,\ldots,n} P(|S_n - S_k| \leq \varepsilon)$$

Proof: Denote $T = \inf\{k \in \{1, \ldots, n\} : |S_k| > 2\varepsilon\}$. Then

$$P(|S_n| > \varepsilon) \geq P(|S_n| > \varepsilon, T < \infty) = \sum_{k=1}^n P(|S_n| > \varepsilon, T = k) \geq \sum_{k=1}^n P(|S_k| - |S_n - S_k| > \varepsilon, T = k)$$
$$\geq \sum_{k=1}^n P(|S_n - S_k| \leq \varepsilon, T = k) = \sum_{k=1}^n P(|S_n - S_k| \leq \varepsilon) P(T = k)$$
$$\geq \min_{k=1,\ldots,n} P(|S_n - S_k| \leq \varepsilon) \cdot P(T < \infty).$$

□

Let $X_n \in \mathbb{L}(\Omega, \mathcal{A}, P), n \in \mathbb{N}$. We say that the following sum of variables $\sum_{n=1}^\infty X_n$ is ... almost surely (in probability or in $\mathbb{L}_p, p \geq 1$) if the sequence $S_n = \sum_{k=1}^n X_k$ of partial sums converges to some $S \in \mathbb{L}(\Omega, \mathcal{A}, P)$ almost surely (in probability or in $\mathbb{L}_p$). Then the symbol $\boldsymbol{\sum_{k=1}^\infty X_k}$ stands for such a variable S, which is determined uniquely up to a P-null set.

Remark As it is seen from the definition, we do not emphasize the type of convergent of the sum $\sum_{n=1}^\infty X_n$ we are considering. But as we know, such a sum always converges in probability, and therefore we can always regard $S = \sum_{n=1}^\infty X_n$ as the limit of $S_n = \sum_{k=1}^\infty X_k$ in probability, i.e. $S_n \xrightarrow{\text{P}} S$ as $t \to \infty$.

Theorem 49 Let $X_n \in \mathbb{L}(\Omega, \mathcal{A}, P), n \in \mathbb{N}$ be independent random variables. Then

$$\sum_{n=1}^\infty X_n \text{ is summable a.s.} \quad \equiv \quad \sum_{n=1}^\infty X_n \text{ is summable in probability.}$$

Proof: Denote $S_n = \sum_{k=1}^n X_k$. If $S_n \xrightarrow{\text{as}} S$, then $S_n \xrightarrow{\text{P}} S$ as $n \to \infty$ by theorem 22. Let $S_n \xrightarrow{\text{P}} S$, then S_n is a Cauchy sequence in probability. Let $\varepsilon, \delta \in (0,1)$, then there exists $n_0 \in \mathbb{N}$ such that $P(|S_n - S_{n_0}| > \frac{\varepsilon}{2}) \leq \frac{\delta}{2}$ holds whenever $n \geq n_0$. Then

$$P(|S_m - S_n| > \varepsilon) \leq P(|S_m - S_{n_0}| > \tfrac{\varepsilon}{2}) + P(|S_n - S_{n_0}| > \tfrac{\varepsilon}{2}) \leq \delta$$

and therefore

$$\min_{n_0 \leq n \leq m} P(|S_m - S_n| \leq \varepsilon) \geq 1 - \delta$$

and Skorochod inequality gives that

$$P(\bigcup_{n \geq n_0} [|S_n - S_{n_0}| > 2\varepsilon]) = \lim_{m \to \infty} P(\max_{n_0 \leq n \leq m} |S_n - S_{n_0}| > 2\varepsilon) \leq \limsup_{m \to \infty} \frac{P(|S_m - S_{n_0}| > \varepsilon)}{\min_{n_0 \leq n \leq m} P(|S_m - S_n| \leq \varepsilon)} \leq \frac{\delta}{1-\delta}.$$

and we get that

$$P(\bigcup_{n,m \geq n_0} [|S_n - S_m| > 4\varepsilon]) \leq P(\bigcup_{n,m \geq n_0} [|S_n - S_{n_0}| + |S_m - S_{n_0}| > 4\varepsilon]) \leq P(\bigcup_{n \geq n_0} [|S_n - S_{n_0}| > 2\varepsilon]) \leq \frac{\delta}{1-\delta}.$$

Hence, if $\varepsilon > 0$ and $\delta \in (0,1)$, we have that

$$P(\bigcap_{n_0 \in \mathbb{N}} \bigcup_{n,m \geq n_0} [|S_n - S_m| > 4\varepsilon]) \leq \lim_{n_0 \to \infty} P(\bigcup_{n,m \geq n_0} [|S_n - S_m| > 4\varepsilon]) \leq \frac{\delta}{1-\delta}.$$

Since $\delta \in (0,1)$ was arbitrary, we get that

$$P(\bigcap_{n_0 \in \mathbb{N}} \bigcup_{n,m \geq n_0} [|S_n - S_m| > 4\varepsilon]) = 0,$$

and therefore

$$P(\bigcap_{\varepsilon > 0} \bigcup_{n_0 \in \mathbb{N}} \bigcap_{m,n \geq n_0} [|S_m - S_n| \leq 4\varepsilon]) = 1 - \lim_{\varepsilon \to 0^+} P(\bigcap_{n_0 \in \mathbb{N}} \bigcup_{m,n \geq n_0} [|S_m - S_n| > 4\varepsilon]) = 1.$$

Hence, S_n is a Cauchy sequence almost surely, and theorem 25 gives that S_n converges almost surely. □

Theorem 50 Let $X_n \in \mathbb{L}_2(\Omega, \mathcal{A}, P), n \in \mathbb{N}$ be uncorrelated random variables. Then

$$\sum_{n=1}^{\infty}(X_n - EX_n) \quad \text{is summable in } \mathbb{L}_2 \quad \equiv \quad \sum_{k=1}^{\infty} \mathrm{var}(X_n) < \infty.$$

Proof: By theorem 25, $S_n = \sum_{k=1}^{n}(X_k - EX_k)$ converges in $\mathbb{L}_2$ iff S_n is a Cauchy sequence in $\mathbb{L}_2$, i.e. iff

$$\lim_{n\to\infty} \sup_{m,k\geq n} E|S_m - S_k|^2 = 0 \quad \equiv \quad \lim_{n\to\infty} \lim_{m\to\infty} \sum_{k=n}^{m} \mathrm{var}(X_k) = 0 \quad \equiv \quad \sum_{k=1}^{\infty} \mathrm{var}(X_k) < \infty.$$

□

Theorem 51 Let $X_n \in \mathbb{L}_2(\Omega, \mathcal{A}, P)$ be independent variables with $\sum_n \mathrm{var}(X_n) < \infty$, then $\sum_{n=1}^{\infty}(X_n - EX_n)$ is summable almost surely, in $\mathbb{L}_2$ (and also in $\mathbb{L}_1$ and in probability).

Proof: It follows from theorems 49 and 50. □

A complete characterization of summability of independent variables is given by the following theorem.

Theorem 52 (*Kolmogorov*) Let $X_n \in \mathbb{L}(\Omega, \mathcal{A}, P)$ be independent variables, then the following conditions are equivalent.

(1) $\sum_{n=1}^{\infty} X_n$ is summable.
(2) There exists $c \in (0, \infty)$ such that (7) holds.
(3) (7) holds, whenever $c \in (0, \infty)$, where

$$\text{(7)} \qquad \sum_{n=1}^{\infty} P(|X_n| > c) < \infty \quad \& \quad \sum_{n=1}^{\infty} E[X_n; |X_n| \leq c] \quad \text{is summable} \quad \& \quad \sum_{n=1}^{\infty} \mathrm{var}(X_n 1_{[|X_n|\leq c]}) < \infty.$$

Remark Let $X_n \in \mathbb{L}_1, n \in \mathbb{N}$ be such that $\sum_{n=1}^{\infty} E|X_n| < \infty$, then $\sum_n X_n$ is summable in $\mathbb{L}_1$ and also almost surely.

Proof: First, we show that $S_n = \sum_{k=1}^{n} X_k$ is a Cauchy sequence in $\mathbb{L}_1$. Obviously.

$$E|S_{n+p} - S_n| = E|\sum_{k=n+1}^{n+p} X_k| \leq \sum_{k=n+1}^{\infty} E|X_k| \to 0 \quad \text{as} \quad n \to \infty$$

uniformly in $p \in \mathbb{N}$. Second, $Y = \sum_{n=1}^{\infty} |X_n| \in \mathbb{L}_1$ holds by assumption, and therefore $\sum_{n=1}^{\infty} |X_n| < \infty$ holds almost surely. In particular, $\sum_{n=1}^{\infty} X_n$ is summable almost surely. □

10. Laws of large numbers

Cronecker lemma Let $0 < b_n \uparrow \infty$ and $a_n \in \mathbb{R}$ be such that $S_n = \sum_{k=1}^{n} a_k$ is a convergent sequence in $\mathbb{R}$, i.e. $\sum_n a_n$ is summable. Then

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k a_k \to 0$$

as $n \to \infty$.

Corollary Let $0 < b_n \uparrow \infty$ be such that $\sum_{n=1}^{\infty} b_n^{-1} X_n$ is summable almost surely. Then

$$\text{(8)} \qquad \frac{1}{b_n} \sum_{k=1}^{n} X_k \overset{\text{as}}{\to} 0$$

Remark If $X_n \in \mathbb{L}_1$ are such that $\sum_{n=1}^{\infty} E|\frac{X_n}{b_n}| < \infty$, then (8) holds.

Theorem 53 Let $X_n \in \mathbb{L}_2, n \in \mathbb{N}$ be independent random variable and let $0 < b_n \uparrow \infty$ be such that

$$\sum_{n=1}^{\infty} \mathrm{var}(\tfrac{X_n}{b_n}) < \infty.$$

Then

$$\text{(9)} \qquad \frac{1}{b_n} \sum_{k=1}^{n}(X_k - EX_k) \overset{\text{as}}{\to} 0.$$

Proof: By theorem 51, $\sum_{k=1}^{\infty} \frac{X_k - EX_k}{b_n}$ is summable almost surely, and the above corollary gives (9). $\square$

Theorem 54 Let $X_n, n \in \mathbb{N}$ be independent identically distributed real valued random variables, then

$$P\left(\limsup_{n\to\infty} \frac{1}{n}\left|\sum_{k=1}^{n} X_k\right| < \infty\right) > 0 \qquad \equiv \qquad X_1 \in \mathbb{L}_1 \qquad \equiv \qquad \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{as}} EX_1$$

as $n \to \infty$.

Theorem 55 (Weak law of large numbers,*Čebyšev*) Let $X_n \in \mathbb{L}_2, n \in \mathbb{N}$ be such that $\operatorname{cov}(X_i, X_j) = 0$ if $i \neq j$, and that

$$\frac{1}{b_n^2}\sum_{k=1}^{n} \operatorname{var}(X_k) \to 0, \qquad \text{then} \qquad Y_n := \frac{1}{b_n}\sum_{k=1}^{n}(X_k - EX_k) \xrightarrow{\text{P}} 0.$$

Proof: By assumption $EY_n = 0$ and $E|Y_n|^2 = \operatorname{var}(Y_n) \to 0$ as $n \to \infty$. Hence, $Y_n \xrightarrow{\mathbb{L}_2} 0$, which gives that also $Y_n \xrightarrow{\text{P}} 0$ as $n \to \infty$. $\square$

Remark The direct proof can be obtained from the definition of convergence in probability with the help of so called Čebyšev inequality for $X \in \mathbb{L}_1$ and $\varepsilon > 0$ in the form

$$P(|X - EX| > \varepsilon) \leq \varepsilon^{-2}\operatorname{var}(X).$$

It is a special case of so called Markov inequality in the form

$$P(|X| \geq \varepsilon) \leq \varepsilon^{-r} E|X|^r$$

if $\varepsilon, r > 0$.

Theorem 56 (Law of iterated logarithm, *Hartmann - Wintner*) Let $X_n, n \in \mathbb{N}$ be iid random variables with $\mu = EX_n \in \mathbb{R}$ and $\sigma^2 = \operatorname{var}(X_n) \in (0, \infty)$. Denote

$$Y_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - \mu),$$

then

$$\limsup_{n\to\infty} \frac{Y_n}{\sqrt{\ln\ln n}} \overset{\text{as}}{=} \sqrt{2\sigma^2} \qquad \& \qquad \liminf_{n\to\infty} \frac{Y_n}{\sqrt{\ln\ln n}} \overset{\text{as}}{=} -\sqrt{2\sigma^2}.$$

11. Weak convergence

Let P_n, P be Borel probability measures on a metric space (S, d). We say that ***P_n converge to P as $n \to \infty$ weakly*** and we write ***$P_n \xrightarrow{w} P$*** as $n \to \infty$ if

$$\int f \, \mathrm{d}P_n \to \int f \, \mathrm{d}P$$

holds as $n \to \infty$ whenever f is a bounded continuous function on (S, d). The set of all bounded continuous function on S will be denoted as ***$C_b(S)$***

Theorem 57 Let (S, d) be a metric space and $P, Q, P_n, n \in \mathbb{N}$ be probability measures on $(S, \mathcal{B}(S))$.

(1) If $\int_E f \, \mathrm{d}P = \int_E f \, \mathrm{d}Q$ holds for every $f \in C_b(S)$. Then $P = Q$.
(2) If $P_n \xrightarrow{w} P$ and $P_n \xrightarrow{w} Q$ as $n \to \infty$. Then $P = Q$.

Proof: (1) We will show that the following system $\mathcal{M} = \{B \in \mathcal{B}(S); P(B) = Q(B)\}$ contains closed set in (S, d) and we obtain from Dynkin lemma that $Q = P$. Obviously $\emptyset \in \mathcal{M}$. Let $F \neq \emptyset$ be closed in (S, d) and put $f_n(x) = (1 - nd(x, F))^+$. Then $C_b(S) \ni f_n \downarrow 1_F$. By assumption and Dominated Convergence Theorem,

$$P(F) = \lim_{n\to\infty} \int f_n \, \mathrm{d}P = \lim_{n\to\infty} \int f_n \, \mathrm{d}Q = Q(F).$$

As mentioned above, the closed set determine a Borel probability measure, and therefore $P = Q$.
(2) It follows from the definition that

$$\int f \, \mathrm{d}P = \lim_{n\to\infty} \int f \, \mathrm{d}P_n = \int f \, \mathrm{d}Q, \quad f \in C_b(S)$$

and the first part of the statement gives that $P = Q$.

Theorem 58 (*Portmanteau lemma*) Let (S,d) be a metric space and P_n, P be Borel probability measures on S. Then the following conditions are equivalent

(1) $P_n \xrightarrow{w} P$ as $n \to \infty$.
(2) $\limsup_n P_n(F) \le P(F)$ as $n \to \infty$ holds whenever F is a closed set in (S,d).
(3) $\liminf_n P_n(G) \ge P(G)$ as $n \to \infty$ holds whenever F is an open set in (S,d).
(4) $\lim_n P_n(B) = P(B)$ as $n \to \infty$ holds for every Borel subset B of (S,d) such that $P(\partial B) = 0$, where $\partial B = \mathit{closure}(B) \setminus \mathit{interior}(B)$ is a border of B.

Proof: (1)⇒(2): Let F be a closed set in (S,d). Then $C_b(S) \ni f_k(x) = (1 - kd(x,F))^+ \downarrow 1_F$ are functions with values in $[0,1]$. Then Dominated Convergence Theorem gives that

$$\limsup_{n\to\infty} P_n(F) = \limsup_{n\to\infty} \int_S 1_F \, \mathrm{d}P_n \le \lim_{n\to\infty} \int_S f_k \, \mathrm{d}P_n = \int_S f_k \, \mathrm{d}P \to \int_S 1_F \, \mathrm{d}P = P(F).$$

(2)≡(3): It is obvious as $F = S \setminus G$ is closed if and only if $G = S \setminus F$ is open and $P(F) + P(G) = 1$.
(2,3)⇒(4): Let B be a Borel set with $P(\partial B) = 0$, Then $0 \overset{\text{as}}{=} 1_{\partial B} = 1_F - 1_G$, where $F = \mathrm{cl}(B), G = \mathrm{int}(B)$, and therefore

$$P(B) = P(G) \le \liminf_{n\to\infty} P_n(G) \le \liminf_{n\to\infty} P_n(B) \le \limsup_{n\to\infty} P_n(B) \le \limsup_{n\to\infty} P_n(F) \le P(F) = P(B).$$

(4)⇒(1): Let $f : S \to (a,b)$ be a continuous function, where $-\infty < a < b < \infty$, and $\varepsilon > 0$ be arbitrary. As

$$\mathcal{M} = \{t \in (a,b) : P[f = t] > 0\} = \bigcup_{k\in\mathbb{N}} \{t \in (a,b) : P[f = t] \ge \tfrac{1}{k}\}$$

is a countable union of finite sets (with at most k-elements), it is countable. Hence, there exists a division $D = \{a = t_0 < \ldots < t_m = b\} \subseteq [a,b] \setminus \mathcal{M}$ with $||D|| = \max_{j\le m}\{t_j - t_{j-1}\} < \varepsilon$. Then $|f - g| \le \varepsilon$, where

$$g = \lfloor f \rfloor_D = \sum_{i=1}^{n} t_{i-1} 1_{B_i}, \qquad \text{where} \qquad B_i = f^{-1}[t_{i-1}, t_i).$$

As $\mathrm{int}(B_i) = f^{-1}(t_{i-1}, t_i)$ and $\mathrm{cl}(B_i) = f^{-1}[t_{i-1}, t_i]$, we get that $\partial B_i = f^{-1}\{t_{i-1}, t_i\}$ has $P(\partial B_i) = 0$. By (4),

$$P(B_i) = \lim_{n\to\infty} P_n(B_i), \qquad \text{and therefore} \qquad \int g \, \mathrm{d}P_n = \sum_{i=1}^{m} t_{i-1} P_n(B_i) \to \sum_{i=1}^{m} t_{i-1} P(B_i) = \int g \, \mathrm{d}P$$

as $n \to \infty$. Further, as $|f - g| \le \varepsilon$, we obtain that

$$\left| \int f \, \mathrm{d}P_n - \int f \, \mathrm{d}P \right| \le 2\varepsilon + \left| \int g \, \mathrm{d}P_n - \int g \, \mathrm{d}P \right| \to 2\varepsilon$$

as $n \to \infty$. As $\varepsilon > 0$ was arbitrary, we get that $\int f \, \mathrm{d}P_n \to \int f \, \mathrm{d}P$ as $n \to \infty$ holds for every $f \in C_b(S)$. □

Theorem 59 Let (S,d) be a metric space and P_n, P be Borel probability measures on S. Then $P_n \xrightarrow{w} P$ as $n \to \infty$ holds if and only if each subsequence of P_n has a subsequence, which converges to P weakly.

Proof: If $P_n \xrightarrow{w} P$ as $n \to \infty$ and $\mathbb{N} \ni n_k \uparrow \infty$, then we obtain from the definition that $P_{m_l} \xrightarrow{w} P$ as $l \to \infty$ holds whenever $m_k = n_{k_l}$ and $\mathbb{N} \ni k_l \uparrow \infty$.
On the other hand, let $P_n \not\xrightarrow{w} P$ as $n \to \infty$, then there exists $f \in C_b(S)$ such that $\int f \, \mathrm{d}P_n \not\to \int f \, \mathrm{d}P$ as $n \to \infty$. Then there exist $\varepsilon > 0$ and $\mathbb{N} \ni n_k \uparrow \infty$ such that $|\int f \, \mathrm{d}P_{n_k} - \int f \, \mathrm{d}P| \ge \varepsilon$ holds for every $k \in \mathbb{N}$. Then no subsequence of P_{n_k} converges to P weakly. □

Lemma Let (S,d) be a metric space and $x, x_n \in S, n \in \mathbb{N}$. Then $\delta_{x_n} \xrightarrow{w} \delta_x$ as $n \to \infty$ if and only if $x_n \to x$ in S, where $\delta_y(B) = 1_B(y)$ is a Dirac measure at $y \in S$.

Proof: Let $x_n \to x$ in S and $f \in C_b(S)$. Then $\int f \, \mathrm{d}\delta_{x_n} = f(x_n) \to f(x) = \int f \, \mathrm{d}\delta_x$. On the other hand let $\delta_{x_n} \xrightarrow{w} \delta_x$ and $\varepsilon > 0$. Then $G = \{y \in S : d(x,y) < \varepsilon\}$ is an open set in (S,d). By Portmanteau lemma

$$\liminf_{n\to\infty} 1_G(x_n) = \liminf_{n\to\infty} \delta_{x_n}(G) \ge \delta_x(G) = 1_G(x) = 1.$$

Hence, there exists $n_0 \in \mathbb{N}$ such that $x_n \in G$ holds for every $n \ge n_0$. □

Let $\mathcal{M}$ be a subset of the set of all Borel probability meaures on a metric space (S,d). We say that $\mathcal{M}$ is

(1) ***tight*** if for every $\varepsilon > 0$ there exists a compact set K in (S,d) s.t. $\mu(K) > 1 - \varepsilon$ whenever $\mu \in \mathcal{M}$.
(2) ***relatively weakly compact*** if every sequence μ_n in $\mathcal{M}$ has a weakly convergent subsequence.

Theorem 60 (*Prochorov*) Let (S,d) be a separable metric space and $\mathcal{M}$ be a subset of the set of all Borel probability measures on (S,d). Then $\mathcal{M}$ is relatively weakly compact if and only if $\mathcal{M}$ is tight.

12. Convergence in distribution

Let (S,d) be a separable metric space and $X, X_n, n \in \mathbb{N}$ be Borel measurable random variables with values in S. If $P_{X_n} \xrightarrow{w} P_X$ as $n \to \infty$, we say that ***X_n converge to X in distribution*** and write ***$X_n \xrightarrow{\mathcal{D}} X$*** as $n \to \infty$.

Note that each variable can be defined on own probability space $(\Omega_n, \mathcal{A}_n, P_n)$, then $P_{X_n} = P_n X_n^{-1}$. Further, if $X_n \xrightarrow{\mathcal{D}} X$ and $X_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$, then $P_X = P_Y$ but X, Y does not have to be equal a.s., since they can be even defined on a different probability space.

Remark (*Portmanteau lemma for convergence in distribution*)
Let (S,d) be a metric space and X, X_n be Borel measurable random variables defined on $(\Omega, \mathcal{A}, P)$ or $(\Omega_n, \mathcal{A}_n, P_n)$, respectively, with values in S. Then the following conditions are equivalent

(1) $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.
(2) $Ef(X_n) \to Ef(X)$ as $n \to \infty$ holds for every $f \in C_b(S)$.
(3) $\limsup_n P_n(X_n \in F) \leq P(X \in F)$ as $n \to \infty$ holds whenever F is a closed set in (S,d).
(4) $\liminf_n P_n(X_n \in G) \geq P(X \in G)$ as $n \to \infty$ holds whenever F is an open set in (S,d).
(5) $\lim_n P_n(X_n \in B) = P(X \in B), n \to \infty$ holds whenever B is a Borel subset of (S,d) s.t. $P(\partial B) = 0$,

where $\partial B = \mathit{closure}(B) \setminus \mathit{interior}(B)$ is a border of B.

Theorem 61 Let $X, X_n : (\Omega, \mathcal{A}, P) \to (S, \mathcal{B}(S)), n \in \mathbb{N}$ be random variables defined on the same probability space, where (S,d) is a metric space and $\mathcal{B}(S)$ is its Borel σ-algebra. Then

(1) If $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$, then $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.
(2) If $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ and $X \overset{\mathrm{as}}{=} c \in S$, then $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$.

Proof: (1): Let $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$ and assume that $X_n \not\xrightarrow{\mathcal{D}} X$ as $n \to \infty$. Then there exists $f \in C_b(S), \mathbb{N} \ni n_k \uparrow \infty$ and $\varepsilon > 0$ such that

$$|\int f(X_{n_k})\,\mathrm{d}P - \int f(X)\,\mathrm{d}P| \geq \varepsilon.$$

Let $\mathbb{N} \ni k_l \uparrow \infty$ be such that $X_{m_l} \xrightarrow{\mathrm{as}} X$ as $l \to \infty$, where $m_l = n_{k_l}$. Then $f(X_{m_l}) \xrightarrow{\mathrm{as}} f(X)$ as $l \to \infty$ and Dominated Convergence Theorem gives a contradiction with the above inequality.

(2): Let $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ and $\varepsilon > 0$. Then $F = \{x \in S : d(x,c) \geq \varepsilon\}$ is an closed set in (S,d) and the above remark gives that

$$\liminf_{n\to\infty} P(d(X_n, c) > \varepsilon) \leq \limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F) = 1_F(c) = 0$$

and therefore $X_n \xrightarrow{\mathrm{P}} X$ as $n \to \infty$. □

Example There exists a sequence of $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ such that $X_n \not\xrightarrow{\mathrm{P}} X$ as $n \to \infty$ even if X_n, X are defined on the same probability space. It is sufficient to consider equally distributed random variables X_n such that $X_n \overset{\mathrm{as}}{=} X_1$ if n is odd and $X_n \overset{\mathrm{as}}{=} X_2 \overset{\mathrm{as}}{\neq} X_1$ if n is even.
Then $X_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$ holds whenever $P_Y = P_{X_1}$, but X_n does not converge in probability.

Theorem 62 Let $(S,d), (H,\rho)$ be metric spaces and $g : S \to H$ continuous. If X_n, X are Borel measurable random variables with values in S such that

$$X_n \xrightarrow{\mathcal{D}} X, \quad n \to \infty, \quad \text{then also} \quad Y_n = g(X_n) \xrightarrow{\mathcal{D}} g(X) = Y, \quad n \to \infty.$$

Proof: Let $f \in C_b(H)$, then $h = f \circ g \in C_b(S)$, and we obtain from the above remark that

$$Ef(Y_n) = Ef(g(X_n)) = Eh(X_n) \to Eh(X) = Ef(g(X)) = Ef(Y).$$

Again, we obtain from remark above that $Y_n \xrightarrow{\mathcal{D}} Y$ as $n \to \infty$. □

Corollary Let $X^{(n)} : (\Omega_n, \mathcal{A}_n, P_n) \to (\mathbb{R}^k, \mathcal{B}(R^k)), X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}(R^k))$, let $X^{(n)} \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.

(1) Then $\lambda^{\mathsf{T}} X^{(n)} \xrightarrow{\mathcal{D}} \lambda^{\mathsf{T}} X$ as $n \to \infty$ holds for every $\lambda \in \mathbb{R}^k$.
(2) In particular, $X_j^{(n)} \xrightarrow{\mathcal{D}} X_j$ as $n \to \infty$ holds for every $j \in \{1, \ldots, k\}$.

See thereom 78 later on that the first implication (1) from the above corollary can be reversed.

Theorem 63 Let (S,d) be a separable metric space and X_n, Y_n be Borel measurable random variables with values in S such that X_n, Y_n are defined on $(\Omega_n, \mathcal{A}_n, P_n)$. Let $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ and $d(X_n, Y_n) \xrightarrow{\mathcal{D}} 0$ as $n \to \infty$, where $X : (\Omega, \mathcal{A}, P) \to (S, \mathcal{B}(S))$. Then $Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.

Remark We assume that the metric space (S,d) is separable in order to ensure that $d(X_n, Y_n)$ are ($\mathcal{A}_n$-measurable) random variables. Obviously $d : S \times S \to [0, \infty)$ is a continuous function. Hence $U_\varepsilon = \{(x,y) \in S \times S : d(x,y) < \varepsilon\}$ is a Borel measurable set. If (S,d) is separable, then $\mathcal{B}(S \times S) = \mathcal{B}(S) \otimes \mathcal{B}(S)$, and thefore

$$U_\varepsilon \in \mathcal{B}(S \times S) = \mathcal{B}(S) \otimes \mathcal{B}(S) \quad \Rightarrow \quad [d(X_n, Y_n) < \varepsilon] = (X_n, Y_n)^{-1} U_\varepsilon \in \mathcal{A}_n.$$

Proof of theorem 63: Since $h(x) = 1 \wedge |kx|$ is a bounded continuous function on $\mathbb{R}$, we get that

$$P_n(d(X_n, Y_n) > \tfrac{1}{k}) \le E \min\{1, k\, d(X_n, Y_n)\} \to 0 \quad \text{as} \quad n \to \infty.$$

Let F be a closed set in (S,d) and put $F_k = \{s \in S; d(x,F) \le \frac{1}{k}\}$. Then

$$\limsup_{n\to\infty} P_n(Y_n \in F) \le \limsup_{n\to\infty} P_n(X_n \in F_k) + \limsup_{n\to\infty} P_n(d(X_n, Y_n) > \tfrac{1}{k}) \le P(X \in F_k)$$

holds since $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$ and F_k is a closed set. Since $F = \cap_{k=1}^\infty F_k$, we obtain that

$$\limsup_{n\to\infty} P_n(Y_n \in F) \le P(X \in F_k) \downarrow P(X \in F)$$

as $k \to \infty$. Thus, $Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$. □

Corollary Let $X^{(n)}, Y^{(n)} : (\Omega_n, \mathcal{A}_n, P_n) \to (\mathbb{R}^k, \mathcal{B}(R^k))$ and $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^k, \mathcal{B}(R^k))$. Denote $||x||^2 = x^\intercal x$ if $x \in \mathbb{R}^k$. Let $X^{(n)} \xrightarrow{\mathcal{D}} X$ and $||Y_n|| \xrightarrow{\mathcal{D}} 0$ as $n \to \infty$, then $Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.

If X is a k-dimensional real-valued random vector, we know that its distribution is determined by its distribution function. As we will see later on, also the convergence of such vectors in distribution can be characterized in term of distribution functions.

Theorem 64 Let $X^{(n)}, X$ be k-dimensional real-valued random vectors. Then the following conditions are equivalent

(1) $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$

(2) $F_{X_n}(x) \to F_X(x)$ holds for every $x \in \mathbb{R}^k$ such that F_X is continuous at x.

Proof: (1)⇒(2): Let us assume that F is continuous at $x \in \mathbb{R}^k$ and that $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$. If $x < y \downarrow x$, then $F(y) \downarrow F(x)$, and therefore we obtain that $P_X(-\infty, x) = P_X(-\infty, x]$. Since $P(\partial(-\infty, x)) = P_X(-\infty, x] - P_X(-\infty, x) = 0$, we get that by (Portmanteau) remark on convergence in distribution that

$$\lim_{n\to\infty} F_{X_n}(x) = \lim_{n\to\infty} P_{X_n}(-\infty, x) = P_X(-\infty, x) = F_X(x).$$

(2)⇒(1): On the other hand, let us assume that (2) holds. Obviously, $\mathcal{M}_j = \{x \in \mathbb{R}; P(X_j = x) > 0\}$ are countable sets. If $x \in \mathcal{L} = \prod_{j=1}^k (\mathbb{R} \backslash \mathcal{M}_j)$, we get that

$$|F_X(y) - F_X(x)| \le \sum_{j=1}^k P(|X_j - x_j| \le \delta_{x,y}) \to 0, \quad \text{where} \quad \delta_{x,y} = \max_{j \le k} |x_j - y_j| \to 0$$

as $y \to x$ holds, i.e. F_X is continuous at x. Thus, we get that F_X is continuous at each point of the set $\mathcal{L}$. Since we assume (2), we get that $F_{X_n}(x) \to F_X(x)$ holds if $x \in \mathcal{L}$. If $x, y \in \mathcal{L}$ are such that $x \le y$, then

$$P_{X_n}[x, y) = \eta_{x,y}(F_{X_n}) \to \eta_{x,y}(F_X) = P_X[x, y)$$

holds as $n \to \infty$, where

$$\eta_{x,y}(F) = \sum_{z \in \prod_{j=1}^k \{x_j, y_j\}} (-1)^{\sum_{j=1}^k 1_{[z_j = x_j]}} F(z)$$

is a function that assigns to a distribution function F of a random vector, say Y, the corresponding probability $P_Y[x,y) = \eta_{x,y}(F)$. Let $G \subseteq \mathbb{R}^k$ be an open set, then it is a disjoint countable union of sets of the form $[x, y)$, $x, y \in \mathcal{L}$, say $G = \dot\cup_{i \in \mathbb{N}} [x^{(i)}, y^{(i)})$, where $x^{(i)} \le y^{(i)}$ and $x^{(i)}, y^{(i)} \in \mathcal{L}$ hold for every $i \in \mathbb{N}$. Then Fatou's lemma gives that

$$\liminf_{n\to\infty} P_{X_n}(G) = \liminf_{n\to\infty} \sum_{i \in \mathbb{N}} P_{X_n}[x^{(i)}, y^{(i)}) \ge \sum_{i \in \mathbb{N}} \liminf_{n\to\infty} P_{X_n}[x^{(i)}, y^{(i)}) = \sum_{i \in \mathbb{N}} P_X[x^{(i)}, y^{(i)}) = P_X(G)$$

holds. Then (Portmanteau) remark on convergence in distribution gives that $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$. $\square$

13. Characteristic functions

Let X be a n-dimensional real-valued random vector, by its ***characteristic functions*** we mean

$$\hat{P}_X(t) = Ee^{it^\intercal X}, \quad t \in \mathbb{R}^n.$$

Note that such a function is defined correctly as $\exp\{it^\intercal X\}$ is a bounded complex-valued random variable.

Let $r > 0$ be fixed and denote by $\mathtt{Gon}_r$ the set of all goniometric polynomials on $\mathbb{R}^k$ that are r-periodic in each coordinate, i.e.

$$\mathtt{Gon}_r = \left\{ \Re \sum_{j=1}^n \lambda_j e^{\frac{2\pi i}{r} m_j^\intercal x}; \lambda = (\lambda_1, \ldots, \lambda_n)^\intercal \in \mathbb{C}^n, m = (m_1, \ldots, m_n) \in \mathbb{Z}^{k \times n}, n \in \mathbb{N} \right\}.$$

Lemma Let $f : \mathbb{R}^k \to \mathbb{R}$ be a bounded continuous function, $r > 0, \varepsilon \in (0, 1]$. Then there exists $g \in \mathtt{Gon}_{4r}$ such that

$$\max_{x \in [-r,r]^k} |f(x) - g(x)| < \varepsilon \qquad \& \qquad \max_{x \in \mathbb{R}^k} |g(x)| \le \sup_{x \in \mathbb{R}^k} |f(x)| + 1. \tag{10}$$

Proof: Denote $\mathcal{P}_{4r|r} = \{g|_{[-r,r]^k}; g \in \mathtt{Gon}_{4r}\}$. Note that the functions from $\mathcal{P}_{4r|r}$ separate points from $[-r, r]^k$, i.e. if $x, y \in [-r, r]^k$ and $x \neq y$, then there is $g \in \mathcal{P}_{4r|r}$ such that $g(x) \neq g(y)$. By Stone-Weierstrass theorem, there exists $q \in \mathtt{Gon}_{4r}$ such that

$$\max_{x \in [-r,r]^k} |f(x) - q(x)| < \tfrac{\varepsilon}{2}. \tag{11}$$

Put $K = \sup\{|f(x)| : x \in \mathbb{R}^k\} + \frac{\varepsilon}{2}$ and

$$z(y) = (-K) \vee y \wedge K.$$

Then $|q(x)| \le K$ holds whenever $x \in [-r, r]^k$ by (11) and therefore $z(q(x)) = q(x)$ holds if $x \in [-r, r]^k$. By Weierstrass theorem, there exists a polynomial p such that

$$\max_{y \in [a,b]} |p(y) - z(y)| < \tfrac{\varepsilon}{2}, \tag{12}$$

where $[a, b] = \{q(x) : x \in \mathbb{R}^k\}$. Then $g(x) = p(q(x)) \in \mathtt{Gon}_{4r}$ and (12) gives that

$$\max_{x \in \mathbb{R}^k} |g(x)| = \max_{x \in \mathbb{R}^k} |p(q(x))| \le \max_{y \in [a,b]} |p(y)| \le \max_{y \in [a,b]} |z(y)| + \tfrac{\varepsilon}{2} \le K + \tfrac{\varepsilon}{2} \le \sup_{x \in \mathbb{R}^k} |f(x)| + 1$$

as $\varepsilon \in (0, 1]$. Since $z(q(x)) = q(x)$ holds if $x \in [-r, r]^k$, we obtain from (11) and (12) that

$$\max_{x \in [-r,r]^k} |f(x) - g(x)| \le \tfrac{\varepsilon}{2} + \max_{x \in [-r,r]^k} |q(x) - p(q(x))| \le \tfrac{\varepsilon}{2} + \max_{y \in [\alpha,\beta]} |z(y) - p(y)| < \varepsilon.$$

$\square$

Theorem 65 The characteristic function determines the distribution of a real-valued random vector, i.e. if $k \in \mathbb{N}$ and $\hat{P}_X(t) = \hat{P}_Y(t)$ holds for every $t \in \mathbb{R}^k$, then $P_X = P_Y$.

Proof: It is sufficient to show that $\int f \, dP_X = \int f \, dP_Y$ holds for every bounded continuous functions f on $\mathbb{R}^k$ by theorem 57. If $f \in \mathtt{Gon}_r$ holds for some $r > 0$, then the above equality holds as

$$\begin{aligned} \textstyle\int f \, dP_X = Ef(X) &= E\Re \textstyle\sum_{j=1}^n \lambda_j e^{\frac{2\pi i}{r} m_j^\intercal X} = \Re \sum_{j=1}^n \lambda_j \hat{P}_X(\frac{2\pi i}{r} m_j) = \Re \sum_{j=1}^n \lambda_j \hat{P}_Y(\frac{2\pi i}{r} m_j) \\ &= E\Re \textstyle\sum_{j=1}^n \lambda_j e^{\frac{2\pi i}{r} m_j^\intercal Y} = Ef(Y) = \int f \, dP_Y. \end{aligned}$$

Let us fix $f : \mathbb{R}^k \to \mathbb{R}$ bounded and continuous and $\varepsilon, r > 0$ be arbitrary. By lemma above, there exists $g \in \mathtt{Gon}_{4r}$ such that (10) holds. If μ is a Borel probability measures on $\mathbb{R}^k$, we have that

$$|\textstyle\int f \, d\mu - \int g \, d\mu| \le \varepsilon + \int_{\mathbb{R}^k \setminus \in [-r,r]^k} (|f| + |g|) \, d\mu \le \varepsilon + (2 \max\limits_{x \in \mathbb{R}^k} |f(x)| + 1) \cdot \mu(\mathbb{R}^k \setminus [-r, r]^k).$$

Hence, there exists a sequence $g_n \in \mathtt{Gon}_{4r_n}$, where $r_n \to \infty$ such that

$$\limsup_{n \to \infty} |\textstyle\int f \, d\mu - \int g_n \, d\mu| \le \varepsilon$$

holds whenever $\varepsilon > 0$. If μ is P_X or P_Y, we obtain that

$$Ef(X) = \textstyle\int f \, dP_X = \lim\limits_{n \to \infty} \int g_n \, dP_X = \lim\limits_{n \to \infty} \int g_n \, dP_Y = \int f \, dP_Y = Ef(Y).$$

□

Theorem 66 (*Cramer-Wold I*) The distribution of a k-dimensional real-valued random vector X is determined by the distributions of $\lambda^{\mathsf T}X, \lambda \in \mathbb{R}^k$, i.e. if X, Y are both k-dimensional real-valued random vectors such that $P_{\lambda^{\mathsf T}X} = P_{\lambda^{\mathsf T}Y}$ holds for every $\lambda \in \mathbb{R}^k$. Then $P_X = P_Y$.

Proof: Let $X, Y \in \mathbb{L}(\Omega, \mathcal{A}, P)^k$ be such that $P_{\lambda^{\mathsf T}X} = P_{\lambda^{\mathsf T}Y}$ holds whenever $\lambda \in \mathbb{R}^k$. Then

$$\hat{P}_X(t) = Ee^{it^{\mathsf T}X} = \hat{P}_{t^{\mathsf T}X}(1) = \hat{P}_{t^{\mathsf T}Y}(1) = Ee^{it^{\mathsf T}Y} = \hat{P}_Y(t)$$

holds for every $t \in \mathbb{R}^k$. Then theorem 65 gives that $P_X = P_Y$. □

Theorem 67 Let $X = (X_1, \ldots, X_k)^{\mathsf T}$ be a k-dimensional real-valued random vector. Then

(1) $\hat{P}_X(0) = 1 \geq |\hat{P}_X(t)|$ holds whenever $t \in \mathbb{R}^k$.
(2) $\overline{\hat{P}_X(t)} = \hat{P}_X(-t)$ holds whenever $t \in \mathbb{R}^k$, i.e. $\Re\hat{P}_X(t) = \Re\hat{P}_X(-t)$ and $\Im\hat{P}_X(t) + \Im\hat{P}_X(-t) = 0$.
(3) $\hat{P}_X(t)$ is a uniformly continuous function on $\mathbb{R}^k$.
(4) $\hat{P}_X$ is a real-valued function if and only if X has a symmetric distribution, i.e. $P_X = P_{-X}$.
(5) $\hat{P}_{a+BX}(t) = e^{ia^{\mathsf T}t} \cdot \hat{P}_X(B^{\mathsf T}t)$ holds if $a, t \in \mathbb{R}^m$ and $B \in \mathbb{R}^{m\times k}$.

Proof: (1) Obviously, $1 = E|e^{it^{\mathsf T}X}| \leq |Ee^{it^{\mathsf T}X}| = |\hat{P}_X(t)|$ and $\hat{P}_X(0) = Ee^{i0^{\mathsf T}X} = Ee^0 = 1$. (2) Further,

$$\overline{\hat{P}_X(t)} = \overline{Ee^{it^{\mathsf T}X}} = E\overline{e^{it^{\mathsf T}X}} = Ee^{-it^{\mathsf T}X} = \hat{P}_X(-t).$$

(3) Let $t, h \in \mathbb{R}^k$. Then Dominated Convergence Theorem gives that

$$|\hat{P}_X(t+h) - \hat{P}_X(t)| \leq E|e^{i(t+h)^{\mathsf T}X} - e^{it^{\mathsf T}X}| = E|e^{ih^{\mathsf T}X} - 1| \to 0$$

as $h \to 0 \in \mathbb{R}^k$ uniformly in $t \in \mathbb{R}^k$.

(4) If $P_X = P_{-X}$, then $\hat{P}_X(t) = \hat{P}_{-X}(t) = Ee^{-it^{\mathsf T}X} = \hat{P}_X(-t)$, and therefore $\Im\hat{P}_X(t) = 0$ holds by (2). On the other hand, if $\Im\hat{P}_X(t) = 0$ holds for every $t \in \mathbb{R}^k$, then (2) gives that $\hat{P}_X(t) = \hat{P}_X(-t) = Ee^{-it^{\mathsf T}X} = \hat{P}_{-X}(t)$. Then theorem 65 gives that $P_X = P_{-X}$.

(5) Obviously,

$$\hat{P}_{a+BX}(t) = Ee^{it^{\mathsf T}(a+BX)} = e^{ia^{\mathsf T}t} \cdot Ee^{it^{\mathsf T}BX} = e^{ia^{\mathsf T}t} \cdot \hat{P}_X(B^{\mathsf T}t).$$

□

Theorem 68 The variables $X_1, \ldots, X_k \in \mathbb{L}(\Omega, \mathcal{A}, P)$ are independent if and only if $\hat{P}_X(t) = \prod_{j=1}^k \hat{P}_{X_j}(t_j)$ holds for every $t \in \mathbb{R}^k$.

Proof: Let $X_1, \ldots, X_k$ be independent, then

$$\hat{P}_X(t) = Ee^{it^{\mathsf T}X} = E\prod_{j=1}^k e^{it_jX_j} = \prod_{j=1}^k \hat{P}_{X_j}(t_j).$$

Now, assume that $\hat{P}_X(t) = \prod_{j=1}^k \hat{P}_{X_j}(t_j)$. Let Y_j have the same distribution as $X_j, j = 1, \ldots, k$ be such that $Y_1, \ldots, Y_k$ are independent. Denote $Y = (Y_1, \ldots, Y_k)^{\mathsf T}$. Then the first part of the proof gives that

$$\hat{P}_Y(t) = \prod_{j=1}^k \hat{P}_{Y_j}(t_j) = \prod_{j=1}^k \hat{P}_{X_j}(t_j) = \hat{P}_X(t).$$

Then we get that $P_X = P_Y = \otimes_{j=1}^k P_{Y_j} = \otimes_{j=1}^k P_{X_j}$, i.e. $X_1, \ldots, X_k$ are independent. □

Theorem 69 Let $X = (X_1, \ldots, X_k)^{\mathsf T}$ and $Y = (Y_1, \ldots, Y_k)^{\mathsf T}$ be independent real-valued random vectors. Then $\hat{P}_{X+Y}(t) = \hat{P}_X(t)\hat{P}_Y(t)$ holds for every $t \in \mathbb{R}^k$.

Proof: As X, Y are independent, $\hat{P}_{X+Y}(t) = Ee^{it^{\mathsf T}(X+Y)} = E[e^{it^{\mathsf T}X}e^{it^{\mathsf T}Y}] = Ee^{it^{\mathsf T}X}Ee^{it^{\mathsf T}Y} = \hat{P}_X(t)\hat{P}_Y(t)$. □

Theorem 70 A continuous function $\zeta : \mathbb{R}^k \to \mathbb{C}$ is a characteristic function of a k-dimensional real-valued random vector if and only if ζ is ***positively semidefinite***, i.e. if

$$\forall\, m \in \mathbb{N} \;\; \forall\, \alpha \in \mathbb{C}^m, T = (t_1, \ldots, t_m)^{\mathsf T} \in (\mathbb{R}^k)^m \quad \alpha^{\mathsf T} A_T(\zeta)\bar{\alpha} \geq 0,$$

where $A_T(\zeta) = \{\zeta(t_j - t_l)\}_{j,l=1}^m$. Equivalently, we can say that ζ is a positively semidefinite function if and only if $A_T(\zeta)$ is a positively semidefinite (complex-valued) matrix for every $T \in \mathbb{R}^{m\times k}$ and $m \in \mathbb{N}$.

Theorem 71 Let $X_{n,n\in\mathbb{N}_0}$ be a real valued random sequence independent with $N : \Omega \to \mathbb{N}_0$. Then the real variable $Y = X_N$ has the following characteristic function

$$\hat{P}_Y(t) = \sum_{n=0}^{\infty} P(N = n)\hat{P}_{X_n}(t).$$

Proof: A straightforward computation gives that

$$\hat{P}_Y(t) = Ee^{itX_N} = \sum_{n=0}^{\infty} P(N = n)E[e^{itX_N}|N = n] = \sum_{n=0}^{\infty} P(N = n)E[e^{itX_n}|N = n] = \sum_{n=0}^{\infty} P(N = n)\hat{P}_{X_n}(t),$$

since X_n and N are independent variables. □

Theorem 72 Let $X_{n,n\in\mathbb{N}_0}$ be a real valued random sequence of i.i.d. real-valued variables independent with $N : \Omega \to \mathbb{N}_0$. Then $S_N = \sum_{k=1}^N X_k$ has the following characteristic function

$$\hat{P}_{S_N}(t) = A_N(\hat{P}_{X_1}(t)), \qquad \text{where} \qquad A_N(s) = Es^N.$$

Proof: Denote $Y_k = \sum_{j=1}^k X_k$, then $Y = (Y_k, k \in \mathbb{N})$ is independent with N as $X = (X_k, k \in \mathbb{N})$ is. Then $S_N = Y_N$ by the previous theorem has the following characteristic function

$$\hat{P}_{S_N}(t) = \sum_{n=0}^{\infty} P(N = n)\hat{P}_{Y_n}(t) = \sum_{n=0}^{\infty} P(N = n)(\hat{P}_{X_1}(t))^n = A_N(\hat{P}_{X_1}(t))$$

as $\hat{P}_{Y_n}(t) = \prod_{k=1}^n \hat{P}_{X_k}(t) = \hat{P}_{X_1}(t)^n$. □

Lemma Denote

$$\pi_n(x) = e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k\,!}$$

whenever $n \in \mathbb{N}_0$. Then

$$|\pi_n(x)| \le \min\{2\tfrac{|x|^n}{n\,!}, \tfrac{|x|^{n+1}}{(n+1)!}\}$$

holds for every $n \in \mathbb{N}_0$ and $x \in \mathbb{R}$.

Proof: Obviously, $|\pi_0(x)| = |e^{ix} - 1| \le 2$ and

$$|\pi_0(x)| = |e^{ix} - 1| = |e^{ix} - e^{i0}| = |\textstyle\int_0^x e^{iu}\,\mathrm{d}u| \le \int_0^{|x|} |e^{iu}|\,\mathrm{d}u = |x|.$$

Hence, $|\pi_0(x)| \le \min\{2, |x|\}$ and the statement of lemma holds for $n = 0$. Further, if $n \in \mathbb{N}$, then $\pi_n(0) = 0$ and $|\pi_n'(x)| = |\pi_{n-1}(x)|$, and therefore by induction, we obtain that

$$|\pi_n(x)| \le \textstyle\int_0^{|x|} |\pi_{n-1}(u)|\,\mathrm{d}u \le \int_0^{|x|} \min\{2\frac{|u|^{n-1}}{(n-1)\,!}, \frac{|u|^n}{n!}\}\,\mathrm{d}u \le \min\{2\frac{|x|^n}{n\,!}, \frac{|x|^{n+1}}{(n+1)!}\}.$$

□

Theorem 73 Let $p \in \mathbb{N}$ and $X \in \mathbb{L}_p$.

(1) Then $\hat{P}_X$ has continuous derivatives up to order p. They are bounded on $\mathbb{R}$, and

$$\hat{P}_X^{(k)}(t) := \tfrac{\mathrm{d}^k}{\mathrm{d}t^k} P_X(t) = i^k E[X^k e^{itX}]$$

holds if $t \in \mathbb{R}$ and $k = 1, \dots, p$.

(2) The characteristic function $\hat{P}_X$ has a finite Taylor expansion

$$\hat{P}_X(t + h) = \sum_{k=0}^{p} \tfrac{\hat{P}_X^{(k)}(t)}{k\,!}\, h^k + \rho_p(h, t), \quad s, t \in \mathbb{R}, \tag{13}$$

where

$$\varrho_p(h) := \sup_{t\in\mathbb{R}} |\rho_p(h, t)| = o(h^p) \quad \text{as} \quad h \to 0, \qquad \& \qquad \varrho_p(h) \le \tfrac{E|X|^p}{p\,!}\, 2h^p.$$

Proof: (1): If $p=0$, then the statement holds with $\mathbb{L}_0=\mathbb{L}$. Further, we assume that the statement holds for $p\in\mathbb{N}$ and assume that $X\in\mathbb{L}_{p+1}$. Then Dominated Convergence Theorem gives that

$$\frac{\hat{P}_X^{(p)}(t+h)-\hat{P}_X^{(p)}(t)}{h}=i^pE[X^pe^{itX}\,\frac{e^{ihX}-1}{h}]\to i^{p+1}E[X^{p+1}e^{itX}],\quad h\to 0,$$

since $|X^pe^{itX}\,\frac{e^{ihX}-1}{h}]|\le|X|^{p+1}\in\mathbb{L}_1$. Thus, the statement holds with p replaced by $p+1$.

(2): Since $X\in\mathbb{L}_p$, we get that

$$\hat{P}_X(t+h)=\sum_{k=0}^{p}\tfrac{(ih)^k}{k!}E[X^ke^{itX}]+E[e^{itX}\pi_p(hX)]$$

By the previous lemma, $\varrho_p(h)\le E|\pi_p(hX)|\le\frac{2h^p}{p!}E|X|^p$. Similarly, we obtain that

$$h^{-p}\varrho_p(h)\le h^{-p}E|\pi_p(hX)|\le E\min\{\tfrac{2|X|^p}{p!},\tfrac{h|X|^{n+1}}{(p+1)!}\}\to 0$$

as $h\to 0$ holds by Dominated Convergence Theorem. $\square$

Let X be a real-valued random variable. Then the function $\psi_X:t\in\mathbb{R}\mapsto Ee^{tX}\in\bar{\mathbb{R}}$ is called a ***moment generating function*** of X.

Theorem 74 Let $X\in\mathbb{L}_p$ hold for every $p\in\mathbb{N}$ and assume that the power series $\sum_{k=0}^{\infty}\frac{h^k}{k!}EX^k$ has a positive radius of convergence $R>0$. Then

$$\hat{P}_X(t+h)=\sum_{k=0}^{\infty}\tfrac{h^k}{k!}\,\hat{P}_X^{(k)}(t),\quad t,h\in\mathbb{R},|h|<R. \tag{14}$$

In particular,

$$\hat{P}_X(h)=\sum_{k=0}^{\infty}\tfrac{h^k}{k!}\,i^kEX^k,\quad h\in\mathbb{R},|h|<R, \tag{15}$$

and $\hat{P}_X$ (and also the distribution of X) is uniquely determined by the moments $EX^k,k\in\mathbb{N}$. Further, moment generating function is

$$\psi_X(t)=Ee^{tX}=\sum_{k=0}^{\infty}\tfrac{t^k}{k!}EX^k,\quad t\in\mathbb{R},|t|<R. \tag{16}$$

Proof: We will show that the radius of convergence $\tilde{R}$ of the following power series $\sum_{k=0}^{\infty}\frac{t^k}{k!}E|X|^k$ is not smaller than

$$R=\liminf_{k\to\infty}\big(\tfrac{|EX^k|}{k!}\big)^{-1/k}.$$

Now, we concentrate on the radius of convergence corresponding to odd moments. We get from Jensen inequality that $(E|X|^{2k-1})^{1/(2k-1)}\le(EX^{2k})^{1/(2k)}$. Then

$$\bar{R}=\liminf_{k\to\infty}\big(\tfrac{E|X|^{2k-1}}{(2k-1)!}\big)^{-\frac{1}{2k-1}}\ge\liminf_{k\to\infty}\big(\tfrac{E|X|^{2k}}{(2k)!}\big)^{-\frac{1}{2k}}(2k)^{-\frac{1}{2k}}[(2k-1)!]^{\frac{1}{2k-1}-\frac{1}{2k}}=\liminf_{k\to\infty}\big(\tfrac{E|X|^{2k}}{(2k)!}\big)^{-\frac{1}{2k}}=\hat{R}$$

as obviously $(2k)^{-\frac{1}{2k}}\to 1$ as $k\to\infty$ and similarly

$$0\le(\tfrac{1}{2k-1}-\tfrac{1}{2k})\ln[(2k-1)!]=\tfrac{1}{2k(2k-1)}\sum_{j=1}^{2k-1}\ln j\le\tfrac{\ln(2k)}{2k}\to 0.$$

Since $\bar{R}\ge\hat{R}$, we get that

$$\tilde{R}=\liminf_{k\to\infty}\big(\tfrac{E|X|^k}{k!}\big)^{-\frac{1}{k}}\ge\liminf_{k\to\infty}\big(\tfrac{EX^{2k}}{(2k)!}\big)^{-\frac{1}{2k}}=\hat{R}.$$

Since adding new coefficients does not increase the radius of convergence of a power series, we get that

$$\tilde{R}\ge\hat{R}=\liminf_{k\to\infty}\big(\tfrac{EX^{2k}}{(2k)!}\big)^{-\frac{1}{2k}}\ge\liminf_{k\to\infty}\big(\tfrac{|EX^k|}{k!}\big)^{-\frac{1}{k}}=R.$$

Then the radius of convergence of the power series $\sum_{k=0}^{\infty}\frac{t^k}{k!}\hat{P}_X^{(k)}(s)$ is

$$\ddot{R}=\liminf_{k\to\infty}\big(\tfrac{|\hat{P}_X^{(k)}(s)|}{k!}\big)^{-\frac{1}{k}}=\liminf_{k\to\infty}\big(\tfrac{|EX^ke^{isX}|}{k!}\big)^{-\frac{1}{k}}\ge\liminf_{k\to\infty}\big(\tfrac{E|X|^k}{k!}\big)^{-\frac{1}{k}}=\tilde{R}\ge R.$$

Now, let us consider $s, t \in \mathbb{R}$ such that $|t| < R$. By theorem 73 (2),

$$|\hat{P}_X(s+t) - \sum_{k=0}^{\infty} \tfrac{t^k}{k!} \hat{P}_X^{(k)}(s)| \le |\sum_{k=p+1}^{\infty} \tfrac{t^k}{k!} \hat{P}_X^{(k)}(s)| + \tfrac{2|t|^p}{p!} E|X|^p \le \sum_{k=p}^{\infty} \tfrac{2|t|^k}{k!} E|X|^k \to 0$$

holds whenever $t \in \mathbb{R}$ is such that $|t| < R$, i.e. (14) holds. Let us consider the following function

$$\rho : s \in M \mapsto Ee^{sX} \quad \text{on} \quad M = \{s \in \mathbb{C} : |\Re s| < R\},$$

We will show that it is holomorphic on set M, we show that is has a derivative in the complex variable

$$\tfrac{\mathrm{d}}{\mathrm{d}s}\rho(s) = \tfrac{\mathrm{d}}{\mathrm{d}s} Ee^{sX} = E[Xe^{sX}].$$

To verify that the above calculation is correct, we need to show that $E|Xe^{sX}| < \infty$ if $s \in M$, but

$$E|Xe^{sX}| = E|Xe^{hX}| \le E\left|\sum_{k=0}^{\infty} \frac{X^{k+1}}{k!} h^k\right| \le \sum_{k=0}^{\infty} \frac{E|X|^{k+1}}{k!} |h|^k < \infty$$

holds with $h = \Re s \in (-R, R)$ as the power series

$$\sum_{k=0}^{\infty} \tfrac{E|X|^{k+1}}{k!} h^k = \sum_{n=1}^{\infty} \tfrac{E|X|^n}{(n-1)!} h^{n-1} = \sum_{n=0}^{\infty} \tfrac{E|X|^n}{n!} n h^{n-1}$$

has the same radius of convergence $\tilde{R} \ge R$ as the series $\sum_k \frac{E|X|^k}{k!} h^k$.

Since the characteristic function $\hat{P}_X(t) = \rho(it), t \in \mathbb{R}$ determines the distribution of X, we obtain by the theorem on uniqueness of holomorphic functions that the distribution P_X is uniquely determined by $\rho(it), t \in (-R, R)$, i.e. by the moments $EX^k, k \in \mathbb{N}$ and also by the moment generating function $\psi_X(h) = \rho(h), h \in (-R, R)$. Obviously, (16) holds as the radius of convergence of the power series on the right-hand side of (16) is at least R. □

Counterexample There exist

$$X, Y \in \bigcap_{p \in \mathbb{N}} \mathbb{L}_p$$

such that $EX^k = EY^k, k \in \mathbb{N}$, but $P_X \ne P_Y$. Let $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ have a density

$$f(x) = c\exp\{-\alpha x^\lambda\} \cdot 1_{(0,\infty)}(x), \quad \alpha > 0, \lambda \in (0, \tfrac{1}{2}).$$

Let us consider $\beta = \alpha \tan(\lambda\pi)$ and $\varepsilon \in (-1, 1)$ and

$$f_\varepsilon(x) = c\exp\{-\alpha x^\lambda\}(1 + \varepsilon \sin(\beta x^\lambda)) \cdot 1_{(0,\infty)}(x)$$

Let us consider a rela-valued random variable X_ε with the density f_ε. Then

$$EX_\varepsilon^n = \int_0^\infty x^n c e^{-\alpha x^\lambda}(1 + \varepsilon \sin(\beta x^\lambda))\,\mathrm{d}x,$$

and therefore we get with $u = x^\lambda$ that

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}\varepsilon} EX_\varepsilon^n &= \int_0^\infty x^n c e^{-\alpha x^\lambda} \sin(\beta x^\lambda)\,\mathrm{d}x \\
&= \int_0^\infty x^n c\, \Im\, e^{-(\alpha+i\beta)x^\lambda}\,\mathrm{d}x, \\
&= \tfrac{1}{\lambda} \int_0^\infty x^{\lambda(\frac{n+1}{\lambda}-1)} c\, \Im\, e^{-(\alpha+i\beta)x^\lambda} (\lambda x^{\lambda-1})\,\mathrm{d}x, \\
&= \tfrac{1}{\lambda} \int_0^\infty u^{\frac{n+1}{\lambda}-1} c\, \Im\, [e^{-(\alpha+i\beta)u}]\,\mathrm{d}u, \\
&= \tfrac{1}{\lambda} c\, \Gamma(\tfrac{n+1}{\lambda})\, \Im[(\alpha+i\beta)^{-\frac{n+1}{\lambda}}] = 0
\end{aligned}$$

as

$$(\alpha + i\beta)^{-\frac{n+1}{\lambda}} = \alpha^{-\frac{n+1}{\lambda}}(1 + i\tan(\lambda\pi))^{-\frac{n+1}{\lambda}} = \alpha^{-\frac{n+1}{\lambda}}(1 + \tan^2(\lambda\pi))^{-\frac{n+1}{2\lambda}} e^{-(n+1)\pi i} \in \mathbb{R}$$

and as

$$\int_0^\infty \tfrac{a^p x^{p-1}}{\Gamma(p)} e^{-ax}\,\mathrm{d}x = 1$$

hold whenever $a \in \mathbb{C}$ is such that $\Re a > 0$. The last equality can be verified by taking derivative of the left-hand side with respect to $a \in \mathbb{C}$ such that $\Re a > 0$ and showing that the derivative is equal to zero by per partes.

Hence, we get that EX_ε^n does not depend on $\varepsilon \in (-1, 1)$ and therefore the variables X_ε have the same moments, but their distributions are different.

Theorem 75 Let $p \in \mathbb{N}$ and $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$, be such that $\hat{P}_X^{(2p)}(0) \in \mathbb{R}$, then $X \in \mathbb{L}_{2p}(\Omega, \mathcal{A}, P)$.

Proof: If $p = 0$, then the statement of the theorem holds with the notation $\mathbb{L}_0 = \mathbb{L}$. Let us assume that the statement holds for p and that $\hat{P}_X^{(2k+2)}(0) \in \mathbb{R}$. Theorem 73 gives that $\hat{P}_X \in C^{2p}(\mathbb{R})$ and that $h(t) = (-1)^p \hat{P}_X^{(2p)}(t) = E[X^{2p} e^{itX}]$. As $h''(0) \in \mathbb{R}$ holds by assumption, we use $2\times$l'Hopital to obtain that

$$\lim_{t\to 0} \tfrac{h(t)+h(-t)-2h(0)}{t^2} = \lim_{t\to 0} \tfrac{h'(t)-h'(-t)}{2t} = h''(0).$$

Then Fatou's lemma gives that

$$EX^{2p+2} \le \lim_{t\to 0} E[X^{2p} \cdot \tfrac{1-\cos(tX)}{t^2/2}] = \lim_{t\to 0} \tfrac{2h(0)-h(t)-h(-t)}{t^2} = -h''(0) = (-1)^{p+1} \hat{P}_X^{(2p+2)}(0) < \infty.$$

□

Theorem 76 Let $X(n) \in \mathbb{L}(\Omega_n, \mathcal{A}_n, P_n)^k, n \in \mathbb{N}_0$ be such that $X(n) \to X(0)$ in distribution as $n \to \infty$. Then $\hat{P}_{X(n)}(t) \to \hat{P}_{X(0)}(t)$ as $n \to \infty$ holds for every $t \in \mathbb{R}^k$.

Proof: Let $X(n) \to X(0)$ in distribution as $n \to \infty$. Since $x \in \mathbb{R}^k \mapsto \cos(t^\intercal x), \sin(t^\intercal x), t \in \mathbb{R}^k$ are bounded continuous functions, we get that

$$\hat{P}_{X(n)}(t) = Ee^{it^\intercal X(n)} = E\cos(t^\intercal X(n)) + iE\sin(t^\intercal X(n)) \to E\cos(t^\intercal X(0)) + iE\sin(t^\intercal X(0)) \to \hat{P}_{X(0)}(t).$$

Lemma Let $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ and $u > 0$. Then $P(|uX| > 2) \le \frac{1}{u}\int_{-u}^{u}(1 - \hat{P}_X(t))\,\mathrm{d}t$.

Proof: Let $U \sim R(-u, u)$ be independent with X. Then $\hat{P}_U(x) = \frac{\sin(ux)}{ux}$ and
$\frac{1}{2u}\int_{-u}^{u}(1 - \hat{P}_X(u))\,\mathrm{d}u = 1 - E\hat{P}_X(U) = 1 - Ee^{iXU} = 1 - E\hat{P}_U(X) = E(1 - \frac{\sin(uX)}{uX}) \ge \frac{1}{2}P(|uX| > 2)$,
since $h(x) = 1 - \frac{\sin x}{x} \ge \frac{1}{2} \cdot 1_{(2,\infty)}(|x|)$. □

Theorem 77 Let $X^{(n)} \in \mathbb{L}(\Omega_n, \mathcal{A}_n, P_n)^k, n \in \mathbb{N}$ be such that $\hat{P}_{X^{(n)}}(t) \to \zeta(t)$ holds whenever $t \in \mathbb{R}^k$, where $k \in \mathbb{N}$ is fixed. If ζ is continuous at 0, then there exists $X \in \mathbb{L}(\Omega, \mathcal{A}, P)^k$ such that $X^{(n)} \to X$ in distribution as $n \to \infty$ and $\hat{P}_X = \zeta$.

Proof of theorem 77 based on theorem 70: By theorem 70, $t \in \mathbb{R}^k \mapsto \hat{P}_{X^{(n)}}(t)$ is a positively semidefinite function. Then the limit function ζ of $\hat{P}_{X^{(n)}}$ is again a positively semidefinite. Since it is also continuous at zero $0 \in \mathbb{R}^k$, there exists a k-dimension real-valued random X with $\hat{P}_X(t) = \zeta(t)$. Let $f : \mathbb{R}^k \to \mathbb{R}$ be a continuous function bounded by $c \in (0, \infty)$, i.e. $|f| \le c < \infty$. We are going to show that $Ef(X^{(n)}) \to Ef(X)$ as $n \to \infty$. By the above lemma,

$$P_n(||X^{(n)}||_m > \tfrac{2}{u}) \le \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1 - \hat{P}_{X_j^{(n)}}(t))\,\mathrm{d}t \to \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1 - \hat{P}_{X_j}(t))\,\mathrm{d}t$$

as $n \to \infty$, where $||x||_m = \max\{|x_1|, \ldots |x_k|\}$. Then

$$\limsup_{n\to\infty} P_n(||X^{(n)}||_m > \tfrac{2}{u}) \le \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1 - \hat{P}_{X_j}(t))\,\mathrm{d}t \to 0 \tag{17}$$

as $u \to 0^+$. Let $\varepsilon > 0$, then there exists $n_0 \in \mathbb{N}$ and $u_0 > 0$ large enough so that

$$P_n(||X^{(n)}||_m > \tfrac{2}{u}) + P(||X|| > \tfrac{2}{u}) < \varepsilon$$

holds whenever $n \ge n_0$ and $u \ge u_0$. Put $f_u(x) = f(x)1_{[||x||_m > 2/u]}$. Then

$$|Ef_u(X^{(n)}) - Ef_u(X)| \le \sup_{x\in\mathbb{R}} |f(x)| \cdot [P_n(||X^{(n)}||_m > \tfrac{2}{u}) + P(||X|| > \tfrac{2}{u})] < c\varepsilon \tag{18}$$

holds if $u \ge u_0$ and $n \ge n_0$. By the first lemma in this section, there exists $g \in \mathtt{Gon}_{4r}$ with $r = \frac{2}{u}$ such that (10) holds. Put $f_{[u]} = f - f_u$ and $g_{[u]} = g - g_u$. Then a similar inequality as in (18) gives by (10) that

$$|Ef_{[u]}(X^{(n)}) - Ef_{[u]}(X)| \le 2\varepsilon + |Eg_{[u]}(X^{(n)}) - Eg_{[u]}(X)| \le (3 + c)\varepsilon + |Eg(X^{(n)}) - Eg(X)| \to (3 + c)\varepsilon$$

as $n \to \infty$ since $\hat{P}_{X^{(n)}}(t) \to \hat{P}_X(t)$ and $g \in \mathtt{Gon}_{4r}$. Then we obtain that for every $\varepsilon > 0$

$$\limsup_{n\to\infty} |Ef(X^{(n)}) - Ef(X)| \le (3 + 2c)\varepsilon.$$

$\square$

Proof based on Prochorov theorem: Put $||x||_m = \max\{|x_1|, \dots |x_k|\}$. By the above lemma,

$$P_n(||X^{(n)}||_m > \tfrac{2}{u}) \le \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1-\hat{P}_{X_j^{(n)}}(t))\,\mathrm{d}t \to \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1-\zeta(t\mathrm{e}_j))\,\mathrm{d}t,$$

where $\mathrm{e}_j \in \mathbb{R}^k$ is a unit vector such that $\mathrm{e}_j^{\intercal}x$ is a j-th coordinate of x whenever $x \in \mathbb{R}^k$. Then

$$\limsup_{n\to\infty} P_n(||X^{(n)}||_m > \tfrac{2}{u}) \le \sum_{j=1}^{k} \tfrac{1}{u}\int_{-u}^{u}(1-\hat{P}_{X_j}(t))\,\mathrm{d}t \to 0 \tag{19}$$

as $u \to 0^+$. Let $\varepsilon > 0$, then there exists $n_0 \in \mathbb{N}$ and $u_0 > 0$ large enough so that

$$P_n(||X^{(n)}||_m > \tfrac{2}{u}) < \varepsilon \tag{20}$$

holds whenever $n \ge n_0, u \ge u_0$. Since the set $\{P_{X^{(n)}}; n \le n_0\}$ is finite, it is by Prochorov theorem tight, and therefore there exists $u \ge u_0$ such that (20) holds for every $n \le n_0$ (and therefore for every $n \in \mathbb{N}$). Thus, we have verified that $\{P_{X^{(n)}}, n \in \mathbb{N}\}$ is tight and we get from Prochorov theorem that if $\mathbb{N} \ni n_k \uparrow \infty$, there exist $\mathbb{N} \ni k_l \uparrow \infty$ such that $P_{X^{(m_l)}}$ is weakly convergent, where $m_l = n_{k_l}$. Let P be the corresponding limit Borel probability on $\mathbb{R}^k$, and $X : x \in \mathbb{R}^k \mapsto x \in \mathbb{R}^k$ be the canonical random element. Then $P_X = P$, and therefore $X^{(m_l)} \to X$ in distribution as $l \to \infty$. Then we get that

$$\hat{P}_X(t) = \lim_{l\to\infty} \hat{P}_{X^{(m_l)}}(t) = \zeta(t),$$

i.e. ζ is a characteristic function of X. If $\mathbb{N} \ni \tilde{n}_k \uparrow \infty$ is another sequence, we can again find $\mathbb{N} \ni \tilde{k}_l \uparrow \infty$ such that $X^{\tilde{m}_l} \to X$ in distribution, where $\tilde{m}_k = \tilde{n}_{\tilde{k}_l}$, since the characteristic function determines the distribution. Thus, we obtain from the properties of convergence in distribution that $X^{(n)} \to X$ as $n \to \infty$ in distribution. $\square$

Corollary Let $k \in \mathbb{N}$ and $X(n) \in \mathbb{L}(\Omega_n, \mathcal{A}_n, P_n)^k, n \in \mathbb{N}_0$. Then $X(n) \to X(0)$ as $n \to \infty$ in distribution if and only if $\hat{P}_{X(n)}(t) \to \hat{P}_{X(0)}(t), n \to \infty$ holds for every $t \in \mathbb{R}^k$.

Theorem 78 (*Cramer-Wold II*) Let $k \in \mathbb{N}$ and $X(n) \in \mathbb{L}(\Omega_n, \mathcal{A}_n, P_n)^k, n \in \mathbb{N}_0$. Then $X(n) \to X(0)$ as $n \to \infty$ in distribution if and only if $\lambda^{\intercal} X(n) \to \lambda^{\intercal} X(0), n \to \infty$ in distribution holds for every $\lambda \in \mathbb{R}^k$.

Proof: Let $X(n) \to X := X(0)$ in distribution as $n \to \infty$ and let $\lambda \in \mathbb{R}^k$. Since $x \in \mathbb{R}^k \mapsto \lambda^{\intercal} x$ is a continuous function, we get that $\lambda^{\intercal} X(n) \to \lambda^{\intercal} X$ in distribution as $n \to \infty$. On the other hand, if $\lambda^{\intercal} X(n) \to \lambda^{\intercal} X$ in distribution as $n \to \infty$ hods for every $\lambda \in \mathbb{R}^k$, then we get that

$$\hat{P}_{X(n)}(t) = Ee^{it^{\intercal}X(n)} = E\cos(t^{\intercal}X(n)) + iE\sin(t^{\intercal}X(n)) \to E\cos(t^{\intercal}X) + iE\sin(t^{\intercal}X) \to \hat{P}_X(t)$$

as $n \to \infty$ as $\sin, \cos$ are bounded continuous functions. Now, it is enough to use the above corollary. $\square$

14. Inversion formulas

Theorem 79 Let X be a real-valued random variables such that $P(X \in \mathbb{Z}) = 1$. Then

$$P(X = k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{P}_X(t)\, e^{-ikt}\,\mathrm{d}t$$

Proof: Let $k \in \mathbb{Z}$, then

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt} \sum_{n=-\infty}^{\infty} e^{int} P(X=n)\,\mathrm{d}t &= \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} P(X=n)\int_{-\pi}^{\pi} e^{i(n-k)t}\,\mathrm{d}t \\ &= \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} P(X=n)\cdot 2\pi \cdot 1_{[n=k]} = P(X=k).\end{aligned}$$

Theorem 80 Let $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ satisfy $\int |\hat{P}_X(t)|\,\mathrm{d}t < \infty$. Then X has a bounded continuous density

$$f_X(x) = \tfrac{1}{2\pi}\int_{\mathbb{R}} \hat{P}_X(t)\, e^{-itx}\,\mathrm{d}t. \tag{21}$$

Lemma Let $X_n, X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ be such that

(1) X_n have bounded continuous densities f_{X_n} given by the formulas $f_{X_n}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{P}_{X_n}(t)\, e^{-itx}\, \mathrm{d}t$.
(2) $X_n \to X$ in distribution, i.e. $\hat{P}_{X_n}(t) \to \hat{P}_X(t)$ as $n \to \infty$ whenever $t \in \mathbb{R}$.
(3) There exists an integrable function h such that $|\hat{P}_{X_n}(t)|, |\hat{P}_X(t)| \le h(t)$.

Then X has also a bounded continuous density given by the formula (21) and

$$\sup_{x\in\mathbb{R}} |f_{X_n}(x) - f_X(x)| \to 0, \quad n \to \infty.$$

Proof: Obviously, our assumptions together with Dominated Convergence Theorem (DCT) give that

$$\lim_{n\to\infty} \sup_{p\in\mathbb{N}} \sup_{x\in\mathbb{R}} |f_{X_n}(x) - f_{X_{n+p}}(x)| \le \lim_{n\to\infty} \tfrac{1}{2\pi}\int_{\mathbb{R}} |\hat{P}_{X_n}(t) - \hat{P}_{X_{n+p}}(t)|\, \mathrm{d}t = 0$$

Hence, there exists a continuous bounded function f such that $f_n(x) \to f(x)$ as $n \to \infty$ uniformly in x. Again, Dominated Convergence Theorem gives that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |f_{X_n}(x) - f(x)| \le \lim_{n\to\infty} \tfrac{1}{2\pi}\int_{\mathbb{R}} |\hat{P}_{X_n}(t) - \hat{P}_X(t)|\, \mathrm{d}t = 0$$

and therefore we obtain again from Dominated Convergence Theorem that

$$f(x) = \lim_{n\to\infty} f_{X_n}(x) = \lim_{n\to\infty} \tfrac{1}{2\pi}\int_{\mathbb{R}} \hat{P}_{X_n}(t)\, e^{-itx}\, \mathrm{d}t = \tfrac{1}{2\pi}\int_{\mathbb{R}} \hat{P}_X(t)\, e^{-itx}\, \mathrm{d}t.$$

Obviously, $f(x) \ge 0$ if $x \in \mathbb{R}$, and if F_X is continuous at the point $x < y$, then assumption (2) gives that

$$F_X(y) - F_X(x) = \lim_{n\to\infty} [F_{X_n}(y) - F_{X_n}(x)] = \lim_{n\to\infty} \int_x^y f_n(u)\, \mathrm{d}u = \int_x^y f(u)\, \mathrm{d}u.$$

Then we get that $F_X(y) = \int_{-\infty}^y f(u)\, \mathrm{d}u$ holds whenever F_X is continuous at y, and we immediately obtain that F_X is a continuous function and $f(x)$ is a density of X. □

Proof of theorem 80: Obviously, if (21) holds, then $f_X(x)$ is a continuous function by Dominated Convergence Theorem bounded by the value $\int |\hat{P}_X(t)|\, \mathrm{d}t$.

First, we show that formula (21) holds for $X \sim N(\mu, \sigma^2)$ if $\sigma^2 \in (0, \infty)$.

$$\tfrac{1}{2\pi}\int e^{-itx}\hat{P}_X(t)\, \mathrm{d}t = \tfrac{1}{2\pi}\int e^{-itx} e^{it\mu - \frac{1}{2}\sigma^2 t^2}\, \mathrm{d}t = \tfrac{1}{2\pi}\int e^{-\frac{1}{2}\sigma^2(t + i\frac{x-\mu}{\sigma^2})^2}\sigma\, \mathrm{d}t \cdot \tfrac{1}{\sigma}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

holds, since $\int \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}(s-a)^2}\, \mathrm{d}s = 1$ holds for every $a \in \mathbb{C}$, where $s = \sigma t$. Second, we show that (21) holds for $Z = X + Y$ whenever X is as above and $Y \in \mathbb{L}(\Omega, \mathcal{A}, P)$ is independent with X. Then

$$F_Z(z) = P(Z < z) = \int \int_{-\infty}^{z-y} f_X(x)\, \mathrm{d}x\, \mathrm{d}P_Y(y) = \int \int_{-\infty}^{z} f_X(x-y)\, \mathrm{d}x\, \mathrm{d}P_Y(y) = \int_{-\infty}^{z} \int f_X(x-y)\, \mathrm{d}P_Y(y)\, \mathrm{d}x,$$

and therefore Z has a density

$$f_Z(z) = \int f_X(x-y)\, \mathrm{d}P_Y(y).$$

Further Z has a characteristic function in the form

$$\hat{P}_Z(t) = \hat{P}_X(t)\hat{P}_Y(t) = \hat{P}_X(t)\int e^{ity}\, \mathrm{d}P(y) = \int \hat{P}_{X+y}(t)\, \mathrm{d}P_Y(y).$$

Since $|\hat{P}_{X+y}(t)| = |\hat{P}_X(t)|$ is an integrable function, we obtain that

$$\tfrac{1}{2\pi}\int e^{-itz}\hat{P}_Z(t)\, \mathrm{d}t = \tfrac{1}{2\pi}\int e^{-itz}\int \hat{P}_{X+y}(t)\, \mathrm{d}P_Y(y)\, \mathrm{d}t = \int \tfrac{1}{2\pi}\int e^{-itz}\hat{P}_{X+y}(t)\, \mathrm{d}t\, \mathrm{d}P_Y(y) = \int f_{X+y}(z)\, \mathrm{d}P_Y(y),$$

which verifies (21) for Z as the right-hand side is just $f_Z(z)$.

Third, let X be arbitrary random variable with $\int h(t)\, \mathrm{d}t < \infty$, where $h(t) = |\hat{P}_X(t)|$. Let $Y_n \sim N(0, 1/n)$ be independent with X. Then we know that $X_n := X + Y_n \to X$ as $n \to \infty$ in distribution and (21) holds for X_n. Further since $|\hat{P}_{Y_n}(t)| \le 1$, we get that

$$|\hat{P}_{X_n}(t)| = |\hat{P}_X(t)| \cdot |\hat{P}_{Y_n}(t)| \le |\hat{P}_X(t)| = h(t).$$

By lemma, we obtain that (21) holds also for X. □

Theorem 81 Let $X \in \mathbb{L}_1$ and $\alpha < \beta$ be real values. Then

$$\frac{F_X(\beta) + F_X(\beta_+)}{2} - \frac{F_X(\alpha) + F_X(\alpha_+)}{2} = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \tfrac{e^{-iu\alpha} - e^{-iu\beta}}{iu}\, \hat{P}_X(u)\, \mathrm{d}u.$$

Moreover, if F_X is continuous at the points α, β, then

$$F_X(\beta) - F_X(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Re\big(\tfrac{e^{-iu\alpha} - e^{-iu\beta}}{iu}\hat{P}_X(u)\big)\, \mathrm{d}u$$

Remark If $Y \sim R(\alpha,\beta)$ is independent with X. Then $\hat{P}_Y(u) = \frac{e^{i\beta u} - e^{i\alpha u}}{(\beta-\alpha)iu}$ and $Z = X - Y$ has a characteristic function

$$\hat{P}_Z(u) = \hat{P}_X(u)\hat{P}_Y(-u) = \tfrac{e^{-iu\alpha} - e^{-iu\beta}}{(\beta-\alpha)iu}\hat{P}_X(u).$$

The above theorem together with the previous one also says that if F_X is continuous at the points $\{\alpha,\beta\}$, and if $\hat{P}_Z(t)$ is an integrable characteristic function, then Z has a continuous density, which is given by the inversion formula for the densities

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izu}\hat{P}_Z(u)\,\mathrm{d}u$$

and therefore $f_Z(0) = \frac{F_X(\beta) - F_X(\alpha)}{\beta-\alpha} = \frac{P(\alpha<X<\beta)}{\beta-\alpha}$.

We say that $X \in \mathbb{L}(\Omega,\mathcal{A},P)$ has an ***equidistant distribution*** if there are $a \in \mathbb{R}$ called an ***origin*** and $d > 0$ called a ***step*** such that $Y = (X-a)/d$ attains integer values almost surely, i.e. $P(Y \in \mathbb{Z}) = 1$. If there is no bigger step $\tilde{d} > d$ of the equidistant distribution, then d is called a ***maximal step***.

Theorem 82 Let $X \in \mathbb{L}(\Omega,\mathcal{A},P)$ have an equidistant distribution with a step $d > 0$ and an origin $a \in \mathbb{R}$. Then $|\hat{P}_X(t)|$ is a $\frac{2\pi}{d}$-periodic function and $\hat{P}_X(\frac{2\pi}{d}) = e^{\frac{2\pi i a}{d}}$.

Proof: Denote $Y = (X-a)/d$. By assumption $Y \in \mathbb{Z}$ holds almost surely. Then $e^{2\pi Y i} \overset{\text{as}}{=} 1$, and therefore

$$\hat{P}_X(t+\tfrac{2\pi}{d}) = Ee^{i(t+\frac{2\pi}{d})X} = Ee^{i(t+\frac{2\pi}{d})(a+dY)} = e^{i(t+\frac{2\pi}{d})a}Ee^{2\pi Y i}e^{itdY} = e^{\frac{2\pi i a}{d}}Ee^{it(a+dY)} = e^{\frac{2\pi i a}{d}}\hat{P}_X(t).$$

□

Theorem 83 Let $X \in \mathbb{L}(\Omega,\mathcal{A},P)$ and $t_0 > 0$ be such that $|\hat{P}_X(t_0)| = 1$. Then X has an equidistant distribution with a step $d = \frac{2\pi}{t_0}$. Moreover, if $a \in \mathbb{R}$ is such that $\hat{P}_X(t_0) = e^{it_0 a}$, then a is an origin of the equidistant distribution of X corresponding to the step d.

Proof: Let $a \in \mathbb{R}$ be such that $\hat{P}_X(t_0) = e^{iat_0}$. Then $Z = X - a$ is such that $\hat{P}_Z(t_0) = e^{-iat_0}\hat{P}_X(t_0) = 1$, i.e.

$$1 = E\cos(t_0 Z) + iE\sin(t_0 Z).$$

In particular, $1 = E\cos(t_0 Z)$, and therefore $Y := t_0 Z/(2\pi) \in \mathbb{Z}$ holds almost surely, and

$$X = a + Z = a + \tfrac{2\pi}{t_0}Y.$$

□

Corollary Let $X \in \mathbb{L}(\Omega,\mathcal{A},P)$. It there exists $t_0 > 0$ such that $|\hat{P}_X(t_0)| = 1$, then $|\hat{P}_X(t)|$ is a t_0-periodic function.

In particular, if $h(t) : \mathbb{R} \to \mathbb{C}$ and $t_0 \in \mathbb{R}$ is such that $|h(t_0)| = 1$, but $h(t)$ is not a t_0-periodic function, then $h(t)$ is not a characteristic function of a real-valued random variable.

Examples The following functions are not characteristic functions of a real-valued random variable: $f_1(t) = \cot(t^2)$, $f_2(t) = \cos(\sqrt{|t|})$, $f_3(t) = \cos(\ln(1+|t|))$.

Corollary Let $X \in \mathbb{L}(\Omega,\mathcal{A},P)$. Then X has an equidistant distribution if and only if there exists $t_0 \in (0,\infty)$ such that $|\hat{P}_X(t_0)| = 1$.

Theorem 84 Let $X \in \mathbb{L}(\Omega,\mathcal{A},P)$ have a non-degenerate equidistant distribution, i.e. let $a, b \in \mathbb{R}$ be such that $a < b$ and $P(X=a)P(X=b) > 0$. Then the equidistant distribution P_X has a maximal step $\hat{d} > 0$ in the form

$$\hat{d} = 2\pi/\hat{t}, \qquad \text{where} \qquad \hat{t} = \min\{t > 0 : |\hat{P}_X(t)| = 1\} \in (0,\infty).$$

Proof: Let $a, b \in \mathbb{R}$ be such that $a < b$ and $P(X=a)P(X=b) > 0$, and put $\hat{t} = \inf\{t > 0; |\hat{P}_X(t)| = 1\}$. By theorem 82, $\hat{t} < \infty$ and theorem 83 gives that $\hat{t} \geq \frac{2\pi}{b-a}$. Otherwise, there exists $t \in (0, \frac{2\pi}{b-a})$ such that $|\hat{P}_X(t)| = 1$, which means that X has an equidistant distribution with a step $d = 2\pi/t > b-a$. Thus, $\hat{t} \in (0,\infty)$. Since $|\hat{P}_X(t)|$ is a continuous function, we get that $|\hat{P}_X(\hat{t})| = 1$, and therefore we can write min instead of inf in the definition of $\hat{t}$. By theorem 83, X has an equidistant distribution with a step $\hat{d}$. If $\tilde{d} > \hat{d}$ was a step of P_X, then theorem 82 gives that $1 = |\hat{P}_X(\tilde{t})|$, where $\tilde{t} = 2\pi/\tilde{d} \in (0,\hat{t})$, which is not possible by the definition of $\hat{t}$.

□

Lemma Let $a_{n,k} \in \mathbb{C}, k = 1, \ldots, k_n \in \mathbb{N}$ be such that

$$\sum_{k=1}^{k_n} a_{n,k} \to a \in \mathbb{C}, \quad \limsup_{n\to\infty} \sum_{k=1}^{k_n} |a_{n,k}| < \infty, \quad \lim_{n\to\infty} \sum_{k=1}^{k_n} |a_{n,k}|^2 = 0.$$

Then

$$\lim_{n\to\infty} \prod_{k=1}^{k_n} (1 + a_{n,k}) = e^a.$$

Proof: By triangle inequality

$$\begin{aligned}
\left| \prod_{k=1}^{k_n} (1 + a_{n,k}) - e^{\sum_{k=1}^{k_n} a_{n,k}} \right| &\le \sum_{m=1}^{k_n} \left| \prod_{k=1}^{m-1} (1 + a_{n,k})(1 + a_{n,m} - e^{a_{n,m}})\, e^{\sum_{k=m+1}^{k_n} a_{n,k}} \right| \\
&\le \sum_{m=1}^{k_n} \prod_{k=1}^{m-1} (1 + |a_{n,k}|) \left(\sum_{j=2}^{\infty} \frac{|a_{n,m}|^j}{j!} \right) \exp\left\{ \sum_{k=m+1}^{k_n} |a_{n,k}| \right\} \\
&\le \sum_{m=1}^{k_n} \prod_{k=1}^{m-1} \exp\{|a_{n,k}|\} \frac{|a_{n,m}|^2}{2} e^{|a_{n,m}|} \exp\left\{ \sum_{k=m+1}^{k_n} |a_{n,k}| \right\} \\
&= \exp\left\{ \sum_{k=1}^{k_n} |a_{n,k}| \right\} \sum_{m=1}^{k_n} \frac{|a_{n,m}|^2}{2} \to 0
\end{aligned}$$

as $n \to \infty$. □

Theorem 85=82 (*Poisson*) Let $X_n \sim \texttt{Bi}(p_n, n), n \in \mathbb{N}$. Let $np_n \to \lambda > 0$ as $n \to \infty$, then

$$X_n \to X \sim \texttt{Po}\,(\lambda)$$

as $n \to \infty$ in distribution.

Proof: We will show that $\hat{P}_{X_n}(t) \to \hat{P}_X(t)$. Obviously,

$$\hat{P}_{X_n}(t) = (1 - p_n + p_n e^{it})^n = (1 + p_n(e^{it} - 1))^n.$$

Put $a_{n,k} = p_n(e^{it} - 1), k_n = n$, then

$$\begin{aligned}
\sum_{k=1}^{k_n} a_{n,k} &= np_n(e^{it} - 1) \to \lambda(e^{it} - 1) \\
\sum_{k=1}^{k_n} |a_{n,k}| &= np_n|e^{it} - 1| \le 2np_n \to 2\lambda < \infty \\
\sum_{k=1}^{k_n} |a_{n,k}|^2 &= np_n^2|e^{it} - 1|^2 \le \frac{4n^2p_n^2}{n} \to 0
\end{aligned}$$

as $n \to \infty$. By lemma

$$\hat{P}_{X_n}(t) = \prod_{k=1}^{n} (1 + a_{n,k}) = o(1) + \exp\left\{ \sum_{k=1}^{k_n} a_{n,k} \right\} = \exp\{\lambda(e^{it} - 1)\}$$

as $n \to \infty$. If $X \sim \texttt{Po}(\lambda)$, then $\hat{P}_X(t) = \exp\{\lambda(e^{it} - 1)\}$. Hence, $\hat{P}_{X_n}(t) \to \hat{P}_X(t)$ as $n \to \infty$, and therefore $X_n \to X$ in distribution. □

15.1. **Definition of multi-dimensional normal distribution.** We say that a random vector $X = (X_1, \ldots, X_k)^\intercal$ has a k-dimensional normal distribution $N_k(\mu, \Sigma)$ with the vector $\mu \in \mathbb{R}^k$ of mean values and variance matrix $\Sigma \in \mathbb{R}^{n\times n}$, where Σ is a positively semidefinite matrix, if

$$\forall\, \lambda \in \mathbb{R}^k \quad \lambda^\intercal X \sim N(\lambda^\intercal \mu, \lambda \Sigma \lambda).$$

We recall that $X \sim N(\mu, \sigma^2)$, i.e. X has one-dimensional normal distribution with mean value μ and variance $\sigma^2 \geq 0$ if $\sigma^2 = 0$ and $\mu \overset{\text{as}}{=} X$ or if $\sigma^2 > 0$ and X has the following density

$$f(x) = \tfrac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\tfrac{(x-\mu)^2}{2\sigma^2}\}.$$

Note that if $X \sim N(\mu, \sigma^2)$ and $\sigma^2 > 0$, then $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$ has a density $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$.

So, we can imagine[13]
that $X \overset{\text{as}}{=} \mu + \sigma Y$ holds for some $Y \sim N(0,1)$ whenever $X \sim N(\mu, \sigma^2)$.

Remark It following from Cramer-Wold theorem I that such a k-dimensional distribution is determined uniquely. Thus, we are going to show that such a distribution exists. First assume that $\Sigma = I_k \in \mathbb{R}^{k\times k}$ and that $\mu = 0 \in \mathbb{R}^k$. Let $X_1, \dots, X_k \sim N(0,1)$ be independent and put $X = (X_1, \dots, X_k)^{\intercal}$. Then

$$\hat{P}_X(t) = \prod_{j=1}^{k} \hat{P}_{X_j}(t_j) = \prod_{j=1}^{k} e^{-\frac{1}{2}t_j^2} = e^{-\frac{1}{2}\, t^{\intercal} t},$$

and therefore if $\lambda \in \mathbb{R}^k$, then

$$\hat{P}_{\lambda^{\intercal} X}(s) = E e^{is\lambda^{\intercal} X} = \hat{P}_X(s\lambda) = e^{-\frac{1}{2}\,(s\lambda)^{\intercal}(s\lambda)} = e^{-\frac{1}{2}s^2\,\lambda^{\intercal}\lambda} = \hat{P}_Y(s),$$

where $Y \sim N(0, \lambda^{\intercal}\lambda)$. Hence, $\lambda^{\intercal} X \sim N(0, \lambda^{\intercal}\lambda)$ holds for every $\lambda \in \mathbb{R}^k$ which is nothing else but $X \sim N_k(0, I_k)$ by the definition.

Second, let $X \sim N_k(0, I_k)$ as above and let $\mu \in \mathbb{R}^k$ and $\Sigma \in \mathbb{R}^{k\times k}$ be positively semi-definite matrix. Then there exists a positively definite (and symmetric) matrix $\Sigma^{1/2}$ such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. Put $Z = \mu + \Sigma^{1/2}X$. Then

$$\hat{P}_{\lambda^{\intercal} Z}(s) = E e^{is\lambda^{\intercal} Z} = E e^{is\lambda^{\intercal}(\mu + \Sigma^{1/2}X)} = e^{is\lambda^{\intercal}\mu}\hat{P}_X(s\Sigma^{1/2}\lambda) = e^{is\lambda^{\intercal}\mu}e^{-\frac{1}{2}s^2\lambda^{\intercal}\Sigma\lambda} = e^{is\lambda^{\intercal}\mu - \frac{1}{2}s^2\lambda^{\intercal}\Sigma\lambda} = \hat{P}_V(s),$$

where $V \sim N(\lambda^{\intercal}\mu, \lambda^{\intercal}\Sigma\lambda)$. Hence, $Z \sim N_k(\mu, \Sigma)$ holds by the definition. In particular, $N_k(\mu, \Sigma)$ exists.

Remark A k-dimensional real valued random vector X has a normal distribution $N(\mu, \Sigma)$ if and only if $\hat{P}_X(t) = \exp\{it^{\intercal}\mu - \frac{1}{2}\, t^{\intercal}\Sigma t\}$.

15.2. **Central limit theorems.**

Theorem 86 (Feller-Lindeberg 83) Let $(\Omega_n, \mathcal{A}_n, P_n), n \in \mathbb{N}$ be a sequence of probability spaces. Let $X_{n,1}, \dots, X_{n,k_n} \in \mathbb{L}_2(\Omega_n, \mathcal{A}_n, P_n)$ be independent centered variables, i.e. $EX_{n,1} = \ldots = X_{n,k_n} = 0$, where $k_n \in \mathbb{N}$, whenever $n \in \mathbb{N}$. Denote $Y_n = \sum_{k=1}^{n} X_{n,k}$. Let

$$\text{var}(Y_n) = \sum_{k=1}^{k_n} \text{var}(X_{n,k}) = \sum_{k=1}^{k_n} EX_{n,k}^2 \to 1, \quad n \to \infty \tag{22}$$

$$(\forall\, \varepsilon > 0) \quad \sum_{k=1}^{k_n} E[X_{n,k}^2; |X_{n,k}| \geq \varepsilon] \to 0, \quad n \to \infty. \tag{23}$$

Then $Y_n \to Y$ in distribution, where $Y \sim N(0,1)$.

Proof: Obviously, $\hat{P}_{Y_n}(t) = \prod_{k=1}^{k_n} \hat{P}_{X_{n,k}}(t)$. In order to obtain that $\hat{P}_{Y_n}(t) \to e^{-\frac{1}{2}t^2}$, we use lemma above with $a_{n,k} = \hat{P}_{X_{n,k}}(t) - 1$. First, since $EX_{n,k} = 0$, we obtain from lemma above theorem 73 that

$$|a_{n,k}| = |Ee^{itX_{n,k}} - 1| \leq E|\pi_1(tX_{n,k})| \leq E\min\{2|tX_{n,k}|, \tfrac{(tX_{n,k})^2}{2}\} \leq E\tfrac{(tX_{n,k})^2}{2} = \tfrac{1}{2}t^2 EX_{n,k}^2.$$

Then we immediately obtain that

$$\sum_{k=1}^{k_n} |a_{n,k}| \leq \tfrac{1}{2}\, t^2 \sum_{k=1}^{k_n} EX_{n,k}^2 \to \tfrac{1}{2}\, t^2$$

as $n \to \infty$. Further, we will show that $m_n := \max\{|a_{n,k}|; k = 1, \dots, k_n\} \to 0$ as $n \to \infty$, which gives that $\sum_{k=1}^{k_n} |a_{n,k}|^2 \leq m_n \cdot \sum_{k=1}^{k_n} |a_{n,k}| \to 0$ as $n \to \infty$. Let $\varepsilon > 0$, then

$$|a_{n,k}| \leq \tfrac{1}{2}t^2 EX_{n,k}^2 \leq \tfrac{1}{2}t^2\varepsilon^2 + \tfrac{1}{2}t^2 \textstyle\sum_{k=1}^{k_n} E[X_{n,k}^2; |X_{n,k}| \geq \varepsilon] \to \tfrac{1}{2}\, t^2\varepsilon^2.$$

[13] If $(\Omega, \mathcal{A}, P) = (\{0\}, \{\emptyset, \{0\}\}, \delta_0)$, then the canonical random variable $X \sim N(0,0)$, but there exists no $Y \sim N(0,1)$ such that $X \overset{\text{as}}{=} 0 \cdot Y = 0$. On the other hand, if the underlying probability space admits $Y \sim N(0,1)$, then $X \overset{\text{as}}{=} Y \cdot 0$.

Since $\varepsilon > 0$ was arbitrary, we obtain $m_n \to 0$ as $n \to \infty$. Thus, it remains to show that $\sum_{k=1}^{k_n} a_{n,k} \to -\frac{1}{2}t^2$. By lemma above theorem 73,

$$|a_{n,k} + \tfrac{t^2}{2}\, EX_{n,k}^2| = |Ee^{itX_{n,k}} - 1 - itEX_{n,k} + \tfrac{t^2}{2}\, EX_{n,k}^2| = |E\pi_2(tX_{n,k})| \le E[t^2 X_{n,k}^2 \min\{1, |\tfrac{t}{6}\, X_{n,k}|\}]$$

Let $\varepsilon > 0$ be arbitrary. Then the same steps as above give that

$$|\sum_{k=1}^{k_n} a_{n,k} + \tfrac{t^2}{2}\sum_{k=1}^{k_n} EX_{n,k}^2| \le t^2 \sum_{k=1}^{k_n} E[X_{n,k}^2; |X_{n,k}| \ge \varepsilon] + \tfrac{|t|^3\varepsilon}{6}\sum_{k=1}^{k_n} EX_{n,k}^2 \to \tfrac{|t|^3\varepsilon}{6}.$$

Since $\varepsilon > 0$ was arbitrary, we obtain the desired convergence $\sum_{k=1}^{k_n} a_{n,k} \to -\frac{1}{2}t^2$. Then the above mentioned lemma gives that $\hat{P}_{Y_n}(t) \to \hat{P}_Y(t)$, where $Y \sim N(0,1)$, and therefore $Y_n \to Y$ in distribution. □

Theorem 87 (Lévy-Lindeberg 84) Let $X_n, n \in \mathbb{N}$ be independent identically distributed random variables with $\mu = EX_1$ and $\mathrm{var}(X_1) = \sigma^2 \in (0, \infty)$. Then

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{k=1}^{n} X_k - n\mu\right) \to Y$$

as $n \to \infty$ in distribution, where $Y \sim N(0,1)$.

Proof: Put $X_{n,k} = \frac{X_k - \mu}{\sigma\sqrt{n}}$ and $k_n = n$. Then $X_{n,1}, \ldots, X_{n,k_n} \in \mathbb{L}_2$ are independent centered variables with $\mathrm{var}(X_1) = \frac{1}{n}$. Then $\mathrm{var}(Y_n) = \sum_{k=1}^{n} \mathrm{var}(X_{n,k}) = 1$. We verify (23). Let $\varepsilon > 0$. Then

$$\sum_{k=1}^{n} E[X_{n,k}^2; |X_{n,k}| \ge \varepsilon] = nE[|\tfrac{X_1-\mu}{\sigma\sqrt{n}}|^2; |\tfrac{X_1-\mu}{\sigma\sqrt{n}}| \ge \varepsilon] = \sigma^{-2}E[|X_1 - \mu|^2; |X_1 - \mu| \ge \varepsilon\sigma\sqrt{n}] \to 0$$

as $n \to \infty$ since $X_1 \in \mathbb{L}_2$. By theorem 86, $Y_n \to Y$ in distribution as $n \to \infty$, where $Y \sim N(0,1)$. □

Theorem 88 (Moivre-Laplace) Let $X_n \sim \mathtt{Bi}(n,p), n \in \mathbb{N}$. Then

$$\frac{X_n - np}{\sqrt{np(1-p)}} \to Y$$

in distribution as $n \to \infty$, where $Y \sim N(0,1)$.

Proof: Let us consider a Bernoulli sequence of independent random variables Y_k with alternative distribution with parameter $p \in (0,1)$, i.e. $Y_k \sim \mathtt{Bi}(1,p)$. Then $Z_n = \sum_{k=1}^{n} Y_k \sim B(n,p) \sim X_n$, and theorem 87 says that $(Z_n - EZ_n)/\sqrt{\mathrm{var}(Z_n)} \to Y$, in distribution as $n \to \infty$, where $Y \sim N(0,1)$. □

Theorem 89 (Ljapunov) Let $(\Omega_n, \mathcal{A}_n, P_n), n \in \mathbb{N}$ be a sequence of probability spaces. Let $X_{n,1}, \ldots, X_{n,k_n} \in \mathbb{L}_2(\Omega_n, \mathcal{A}_n, P_n)$ be independent centered variables, i.e. $EX_{n,1} = \ldots = X_{n,k_n} = 0$, where $k_n \in \mathbb{N}$, whenever $n \in \mathbb{N}$. Denote $Y_n = \sum_{k=1}^{n} X_{n,k}$. Let

$$\mathrm{var}(Y_n) = \sum_{k=1}^{k_n} \mathrm{var}(X_{n,k}) = \sum_{k=1}^{k_n} EX_{n,k}^2 \to 1, \quad n \to \infty \tag{24}$$

$$(\exists\, \delta > 0) \quad \sum_{k=1}^{k_n} E|X_{n,k}|^{2+\delta} \to 0, \quad n \to \infty. \tag{25}$$

Then $Y_n \to Y$ in distribution, where $Y \sim N(0,1)$.

Proof: We verify the Feller-Lindeberg condition (23) in the statement of the theorem 86. Let $\varepsilon > 0$, then

$$\sum_{k=1}^{k_n} E[X_{n,k}^2; |X_{n,k}| \ge \varepsilon] \le \varepsilon^{-\delta} \sum_{k=1}^{k_n} E|X_{n,k}|^{2+\delta} \to 0, \quad n \to \infty.$$

Then $Y_n \to Y$ in distribution holds by theorem 86, where $Y \sim N(0,1)$. □

Theorem 90 (Feller-Lindeberg, multi-dimensional CLT) Let $(\Omega_n, \mathcal{A}_n, P_n), n \in \mathbb{N}$ be a sequence of probability spaces and $d \in \mathbb{N}$. Let $X_{n,1}, \ldots, X_{n,k_n} \in \mathbb{L}_2(\Omega_n, \mathcal{A}_n, P_n)^d$ be independent centered random vectors, i.e. $EX_{n,1} = \ldots = X_{n,k_n} = 0 \in \mathbb{R}^d$, where $k_n \in \mathbb{N}$, whenever $n \in \mathbb{N}$. Denote $Y_n = \sum_{k=1}^n X_{n,k}$. Let

$$\text{(26)} \qquad \operatorname{var}(Y_n) = \sum_{k=1}^{k_n} \operatorname{var}(X_{n,k}) = \sum_{k=1}^{k_n} EX_{n,k}X_{n,k}^{\intercal} \to \Sigma \in \mathbb{R}^d, \quad n \to \infty$$

$$\text{(27)} \qquad (\forall\, \varepsilon > 0) \quad \sum_{k=1}^{k_n} E[||X_{n,k}||^2; ||X_{n,k}|| \geq \varepsilon] \to 0, \quad n \to \infty,$$

where $||X_{n,k}||^2 = X_{n,k}^{\intercal} X_{n,k}$. Then $Y_n \to Y$ in distribution, where $Y \sim N_d(0, \Sigma)$.

Proof: We will use Cramer-Wold theorem II saying that convergence of random vectors in distribution can be verified by verifying of the convergence of all linear combinations in distribution. Let $\lambda \in \mathbb{R}^d$, we are going to show that $\lambda^{\intercal} Y_n \to \lambda^{\intercal} Y$ in distribution as $n \to \infty$. Let us consider the first case $\lambda^{\intercal}\Sigma\lambda = 0$. Then

$$E|\lambda^{\intercal} Y_n|^2 = \operatorname{var}(\lambda^{\intercal} Y_n) = \lambda^{\intercal} \operatorname{var}(Y_n)\lambda \to \lambda^{\intercal}\Sigma\lambda = 0$$

as $n \to \infty$, which means[14] that $\lambda^{\intercal} Y_n \xrightarrow{\mathbb{L}_2} 0$ as $n \to \infty$. Then $\lambda^{\intercal} Y_n \xrightarrow{\mathrm{P}} 0$, and therefore $\lambda^{\intercal} Y_n \to 0$ as $n \to \infty$ in distribution. Now assume that $\lambda^{\intercal}\Sigma\lambda > 0$ and put

$$Y_{n,k} = \frac{\lambda^{\intercal} X_{n,k}}{\sqrt{\lambda^{\intercal}\Sigma\lambda}}.$$

Then $Y_{n,k}, k = 1, \ldots, k_n$ are independent centered random variables. Denote

$$Z_n = \sum_{k=1}^{k_n} Y_{n,k}.$$

Then

$$\operatorname{var}(Z_n) = \sum_{k=1}^{k_n} \operatorname{var}(Y_{n,k}) = \sum_{k=1}^{k_n} \frac{\operatorname{var}(\lambda^{\intercal} X_{n,k})}{\lambda^{\intercal}\Sigma\lambda} \to 1$$

as $n \to \infty$. Let $\varepsilon > 0$. Since

$$|Y_{n,k}| = |\tfrac{\lambda^{\intercal} X_{n,k}}{\sqrt{\lambda^{\intercal}\Sigma\lambda}}| \leq \tfrac{||\lambda||}{\sqrt{\lambda^{\intercal}\Sigma\lambda}} \cdot ||X_{n,k}||,$$

we get that

$$\sum_{k=1}^{k_n} E[|Y_{n,k}|^2; |Y_{n,k}| \geq \varepsilon] \leq \frac{||\lambda||^2}{\lambda^{\intercal}\Sigma\lambda} \sum_{k=1}^{k_n} E[||X_{n,k}||^2; ||X_{n,k}|| \geq \frac{\varepsilon\sqrt{\lambda^{\intercal}\Sigma\lambda}}{||\lambda||}] \to 0, \quad n \to \infty.$$

Then theorem 86 gives that $Z_n \to Z$ as $n \to \infty$ in distribution, where $Z \sim N(0,1)$, and therefore $\lambda^{\intercal} Y_n = \sqrt{\lambda^{\intercal}\Sigma\lambda} \cdot Z_n \to \lambda^{\intercal} Y$ in distribution as $n \to \infty$, since $\lambda^{\intercal} Y \sim N(0, \lambda^{\intercal}\Sigma\lambda)$ holds as $Y \sim N_d(0, \Sigma)$. $\square$

Theorem 91 (Lévy-Lindeberg, multi-dimensional CLT) Let $d \in \mathbb{N}$ and $X_n \in \mathbb{L}_2^d, n \in \mathbb{N}$ be independent identically distributed random vectors with $\mu = EX_1 \in \mathbb{R}^d$ and $\operatorname{var}(X_1) = \Sigma \in \mathbb{R}^{d \times d}$. Then

$$\frac{1}{\sqrt{n}} \left(\sum_{k=1}^n X_k - n\mu \right) \to Y$$

as $n \to \infty$ in distribution, where $Y \sim N(0, \Sigma)$.

Proof: Put $k_n = n$ and $X_{n,k} = n^{-1/2}(X_k - \mu) \in \mathbb{L}_2^d$. Then $EX_{n,k} = 0 \in \mathbb{R}^d$ and $X_{n,k}, k = 1, \ldots, n$ are independent variables with

$$\sum_{k=1}^n \operatorname{var}(X_{n,k}) = n \operatorname{var}(X_{n,1}) = \operatorname{var}(X_1) = \Sigma.$$

---

[14] In fact, we should assume (without loss of generality) that Y_n are defined on the same probability space. For example, we can consider $(\Omega, \mathcal{A}, P) = \otimes_{n\in\mathbb{N}}(\Omega_n, \mathcal{A}_n, P_n)$ and $\tilde{Y}_n(\omega_k, k \in \mathbb{N}) = Y_n(\omega_n)$. Then $P_{\tilde{Y}_n} = P_n \circ Y_n^{-1}$, and therefore $E|\lambda^{\intercal} Y_n|^2 = E|\lambda^{\intercal} \tilde{Y}_n|^2 \to 0$ as $n \to \infty$, which means that $\lambda^{\intercal} \tilde{Y}_n \xrightarrow{\mathbb{L}_2} 0$ as $n \to \infty$, and therefore $\lambda^{\intercal} \tilde{Y}_n \xrightarrow{\mathrm{P}} 0$, which implies that $\lambda^{\intercal} \tilde{Y}_n \to 0$ in distribution. This immediately gives that also $\lambda^{\intercal} Y_n$ having the same distribution as $\lambda^{\intercal} \tilde{Y}_n$ convergences to zero in distribution.

Let $\varepsilon > 0$. Then

$$\sum_{k=1}^{k_n} E[||X_{n,k}||^2; ||X_{n,k}|| \geq \varepsilon] = n\, E[||\tfrac{X_1 - EX_1}{\sqrt{n}}||^2; [||\tfrac{X_1 - EX_1}{\sqrt{n}}|| \geq \varepsilon]$$
$$= E[||X_1 - EX_1||^2; ||X_1 - EX_1|| \geq \varepsilon\sqrt{n}] \to 0$$

as $n \to \infty$, since $||X_1 - EX_1|| \in \mathbb{L}_2$. By theorem 90, $\sum_{k=1}^n X_{n,k} \to Y$ in distribution as $n \to \infty$. □

Theorem 92 (CLT for multidimensional distribution) Let $d \in \mathbb{N}$ and $X_n \sim \mathtt{M}(n,p), n \in \mathbb{N}$ be a sequence of random vectors, where $\mathtt{M}(n,p)$ stands for the d-dimensional multinomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]^d$ is such that $\sum_{k=1}^d p_k = 1$. Then

$$n^{-1/2}(X_n - EX_n) \to X, \qquad \text{in distribution, where} \quad X \sim N(0, \operatorname{diag} p - pp^{\mathsf{T}}).$$

Proof: Let $Y_n \sim \mathtt{M}(1,p), n \in \mathbb{N}$ be independent variables. Then $X_n \sim \sum_{k=1}^n Y_k$ and the statement follows from theorem 91 as

$$\operatorname{var}(X_1) = EX_1X_1^{\mathsf{T}} - EX_1EX_1^{\mathsf{T}} = E\operatorname{diag} X_1 - EX_1EX_1^{\mathsf{T}} = \operatorname{diag} p - pp^{\mathsf{T}}.$$

□

Theorem 93 (Ljapunov, multi-dimensional CLT) Let $(\Omega_n, \mathcal{A}_n, P_n), n \in \mathbb{N}$ be a sequence of probability spaces and $d \in \mathbb{N}$. Let $X_{n,1}, \ldots, X_{n,k_n} \in \mathbb{L}_2(\Omega_n, \mathcal{A}_n, P_n)^d$ be independent centered random vectors, i.e. $EX_{n,1} = \ldots = X_{n,k_n} = 0 \in \mathbb{R}^d$, where $k_n \in \mathbb{N}$, whenever $n \in \mathbb{N}$. Denote $Y_n = \sum_{k=1}^n X_{n,k}$. Let

$$\operatorname{var}(Y_n) = \sum_{k=1}^{k_n} \operatorname{var}(X_{n,k}) = \sum_{k=1}^{k_n} EX_{n,k}X_{n,k}^{\mathsf{T}} \to \Sigma \in \mathbb{R}^d, \quad n \to \infty \tag{28}$$

$$(\exists\, \delta > 0) \quad \sum_{k=1}^{k_n} E||X_{n,k}||^{2+\delta} \to 0, \quad n \to \infty. \tag{29}$$

where $||X_{n,k}||^2 = X_{n,k}^{\mathsf{T}}X_{n,k}$. Then $Y_n \to Y$ in distribution, where $Y \sim N_d(0,\Sigma)$.

Proof: We verify the Feller-Lindeberg condition (27) in the statement of the theorem 90. Let $\varepsilon > 0$, then

$$\sum_{k=1}^{k_n} E[||X_{n,k}||^2; ||X_{n,k}|| \geq \varepsilon] \leq \varepsilon^{-\delta} \sum_{k=1}^{k_n} E||X_{n,k}||^{2+\delta} \to 0, \quad n \to \infty.$$

Then $Y_n \to Y$ in distribution holds by theorem 90, where $Y \sim N_d(0,\Sigma)$. □

16. Convergence of distribution functions

Theorem 94 Let $X_n \in \mathbb{L}(\Omega_n, \mathcal{A}_n, P_n), n \in \mathbb{N}$ and $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ be such that $X_n \to X$ as $n \to \infty$ in distribution. If $F_X(x) = P(X < x)$ is a continuous function, then

$$\sup_{x \in \mathbb{R}} |F_{X_n}(x) - F_X(x)| \to 0$$

as $n \to \infty$.

Proof: Obviously $F_X(-\infty_+) = 0, F_X(\infty_-) = 1$. Let $\varepsilon > 0$ be arbitrary. Then there exists $k \in \mathbb{N}$ and real values $x_1 \leq \ldots \leq x_k$ such that

$$|F_X(x_j) - F_X(x_{j-1})| \leq \varepsilon, \quad j = 1, \ldots, k+1,$$

where $x_0 = -\infty, x_{k+1} = +\infty$, and $F_X(-\infty) := 0, F_X(\infty) := 1$. Since $F_{X_n}(x) \to F_X(x)$ holds at each point $x \in \mathbb{R}$ such that F_X is continuous at x and since F_X is a continuous function, we get the pointwise convergence $F_{X_n} \to F_X$. Thus, there exists $n_0 \in \mathbb{N}$ such that

$$\forall\, n \geq n_0 \quad \forall j = 1, \ldots, k \quad |F_{X_n}(x_j) - F_X(x_j)| \leq \varepsilon$$

Let $x \in \mathbb{R}$, then there exists $j \in \{1, \ldots, k+1\}$ such that $x \in \mathbb{R} \cap [x_{j-1}, x_j]$. Then

$$F_{X_n}(x) - F_X(x) \leq F_{X_n}(x_j) - F_X(x_{j-1}) = F_{X_n}(x_j) - F_X(x_j) + F_X(x_j) - F_X(x_{j-1}) \leq 2\varepsilon$$
$$F_{X_n}(x) - F_X(x) \geq F_{X_n}(x_{j-1}) - F_X(x_j) = F_{X_n}(x_{j-1}) - F_X(x_{j-1}) + F_X(x_{j-1}) - F_X(x_j) \geq -2\varepsilon$$

Thus, $|F_{X_n}(x) - F_X(x)| \leq 2\varepsilon$ holds for every $x \in \mathbb{R}$ whenever $n \geq n_0$. □

Corollary Let $X_n \in L(\Omega_n, \mathcal{A}_n, P_n)$ be such that $X_n \to X$ in distribution as $n \to \infty$, where $X \sim N(0,1)$, then

$$\sup_{x\in\mathbb{R}} |F_{X_n}(x) - \Phi(x)| \to 0$$

as $n \to \infty$, where $\Phi(x) = P(X < x)$.

17. Local limit theorems

Theorem 95 Let $X_n \in \mathbb{L}_2(\Omega, \mathcal{A}, P), n \in \mathbb{N}$ be independent identically distributed random variables with $EX_n = \mu$ and $\mathrm{var}(X_n) = \sigma^2 \in (0,\infty)$ and with equidistant distribution with an origin $a \in \mathbb{R}$ and a maximal step $d \in (0,\infty)$. Denote $S_n = \sum_{k=1}^n X_k$. Then

$$\sqrt{n} \cdot \sup_{x\in L_n} |\tfrac{P(S_n=x)}{d} - \tfrac{1}{\sigma\sqrt{n}}\,\varphi(\tfrac{x-n\mu}{\sigma\sqrt{n}})| \to 0$$

as $n \to \infty$, where $L_n = \{an + kd; k \in \mathbb{Z}\}$.

Proof: Without loss of generality, we assume that $d = 1$ and $a = 0$, otherwise we consider i.i.d. variables $Y_k = (X_k - a)/d$. It follows from inversion formula for equidistant random variables that

$$\begin{aligned}\sigma\sqrt{n}\ P(S_n = x) &= \tfrac{\sigma\sqrt{n}}{2\pi}\int_{-\pi}^{\pi} e^{-isx}\hat{P}_{S_n}(s)\,\mathrm{d}s = \tfrac{1}{2\pi}\int_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} \exp\{\tfrac{-itx}{\sigma\sqrt{n}}\}\hat{P}_{S_n}(\tfrac{t}{\sigma\sqrt{n}})\,\mathrm{d}t\\ &= \tfrac{1}{2\pi}\int_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} \exp\{\tfrac{-it(x-n\mu)}{\sigma\sqrt{n}}\}\hat{P}_{S_n-n\mu}(\tfrac{t}{\sigma\sqrt{n}})\,\mathrm{d}t\end{aligned}$$

holds if $x \in L_n$ and from inversion formula for the densities that

$$\varphi(\tfrac{x-n\mu}{\sigma\sqrt{n}}) = \tfrac{1}{2\pi}\int_{\mathbb{R}} e^{-it\frac{x-n\mu}{\sigma\sqrt{n}}}\, e^{-\frac{1}{2}t^2}\,\mathrm{d}t.$$

Then

$$|\sigma\sqrt{n}\ P(S_n = x) - \varphi(\tfrac{x-n\mu}{\sigma\sqrt{n}})| \le I_n(\varepsilon) + J_n(\varepsilon) + K_n(\varepsilon),$$

where

$$I_n(\varepsilon) = \tfrac{1}{2\pi}\int_{\mathbb{R}} e^{-\frac{1}{2}t^2} 1_{[|t|>\varepsilon\sigma\sqrt{n}]}\,\mathrm{d}t \to 0$$

as $n \to \infty$ and

$$\begin{aligned}J_n(\varepsilon) &= \tfrac{1}{2\pi}\int_{\mathbb{R}} |\hat{P}_{S_n-n\mu}(\tfrac{t}{\sigma\sqrt{n}})| \cdot 1_{[\varepsilon<\frac{|t|}{\sigma\sqrt{n}}\le\pi]}\,\mathrm{d}t\\ K_n(\varepsilon) &= \tfrac{1}{2\pi}\int_{\mathbb{R}} |\hat{P}_{S_n-n\mu}(\tfrac{t}{\sigma\sqrt{n}}) - e^{-\frac{1}{2}t^2}| \cdot 1_{[\frac{|t|}{\sigma\sqrt{n}}\le\varepsilon]}\,\mathrm{d}t.\end{aligned}$$

Since $X_1 - \mu$ is a random variable with equidistant distribution with a maximal step 1, we get that

$$c(\varepsilon) = \sup_{\varepsilon\le|s|\le\pi} |\hat{P}_{X_1-\mu}(s)| \in [0,1),$$

and therefore $J_n(\varepsilon) \le \frac{1}{2\pi}\,c(\varepsilon)^n\pi\sigma\sqrt{n} \to 0$ as $n \to \infty$ holds whenever $\varepsilon \in (0,\pi)$. Finally, we will show that there exists $\varepsilon > 0$ small enough so that $K_n(\varepsilon) \to 0$ as $n \to \infty$. Since $Z_1 = \frac{X_1-\mu}{\sigma}$ has $EZ_1 = 0$ and $EZ_1^2 = 1$, we get that

$$\hat{P}_{X_1-\mu}(\tfrac{r}{\sigma}) = \hat{P}_{Z_1}(r) = 1 - \tfrac{1}{2}r^2 + o(r^2)$$

as $r \to 0$. Hence, there exists $\delta > 0$ such that

$$|\hat{P}_{X_1-\mu}(\tfrac{r}{\sigma})| \le 1 - \tfrac{1}{4}r^2 \le e^{-\frac{1}{4}r^2}$$

holds whenever $|r| \le \delta$. Then we obtain that

$$|\hat{P}_{S_n-n\mu}(\tfrac{t}{\sigma\sqrt{n}})| \cdot 1_{[\frac{|t|}{\sigma\sqrt{n}}\le\varepsilon]} \le [e^{\frac{1}{4}(t/\sqrt{n})^2}]^n = e^{\frac{1}{4}t^2}$$

holds if $\varepsilon \in (0, \delta/\sigma]$. Thus, we may use Dominated Convergence Theorem in order to obtain that $K_n(\varepsilon) \to 0$ as $n \to \infty$ if $\varepsilon = \delta/\sigma$, since

$$\hat{P}_{S_n-n\mu}(\tfrac{t}{\sigma\sqrt{n}}) = \hat{P}_{Y_n}(t) \to e^{-\frac{1}{2}t^2}$$

as $n \to \infty$, where $Y_n = (S_n - n\mu)/(\sigma\sqrt{n}) \to Y$ in distribution, where $Y \sim N(0,1)$. □

Lemma Let $X \in \mathbb{L}(\Omega, \mathcal{A}, P)$ have a bounded density $f_X(x) \ge 0$ and a non-negative characteristic function $\hat{P}_X(t)$. Then $\hat{P}_X(t)$ is an integrable function[15].

[15] And therefore there exists a continuous version of the density of X.

Proof: Let $Y_n \sim N(0,\frac{1}{n})$ and $Y \equiv 0$, then $\hat{P}_{Y_n}(t) = e^{-\frac{t^2}{2n}} \to 1 = \hat{P}_Y(t)$, where $Y \equiv 0$. Since $\hat{P}_X(t)$ is assumed to be non-negative, we obtain from Fatou's lemma that

$$\int_{\mathbb{R}} \hat{P}_X(t)\,\mathrm{d}t \le \liminf_{n\to\infty} \int_{\mathbb{R}} \hat{P}_X(t) e^{-\frac{t^2}{2n}}\,\mathrm{d}t$$

Let $n \in \mathbb{N}$ be fixed. Since $e^{-\frac{t^2}{2n}}$ is an integrable function, we obtain from Fubini theorem that

$$\int_{\mathbb{R}} \hat{P}_X(t) e^{-\frac{t^2}{2n}}\,\mathrm{d}t = \int_{\mathbb{R}} \big(\int_{\mathbb{R}} f_X(x) e^{itx}\,\mathrm{d}x\big) e^{-\frac{t^2}{2n}}\,\mathrm{d}t = \int_{\mathbb{R}} \big(\int_{\mathbb{R}} e^{itx} e^{-\frac{t^2}{2n}}\,\mathrm{d}t\big) f_X(x)\,\mathrm{d}x$$

Further, we obtain from inversion formula for the density $f_{N(0,1/n)}(x)$ of $N(0,\frac{1}{n})$ that

$$\tfrac{1}{2\pi} \int_{\mathbb{R}} e^{itx} e^{-\frac{t^2}{2n}}\,\mathrm{d}t = f_{N(0,1/n)}(x).$$

Let $c \in (0,\infty)$ be such that $f_X(x) \le c$ hold for every $x \in \mathbb{R}$, then

$$\int_{\mathbb{R}} \hat{P}_X(t) e^{-\frac{t^2}{2n}}\,\mathrm{d}t = 2\pi \int_{\mathbb{R}} f_{N(0,1/n)}(x) f_X(x)\,\mathrm{d}x \le 2\pi c \cdot \int_{\mathbb{R}} f_{N(0,1/n)}(x)\,\mathrm{d}x \le 2\pi c$$

Thus, we get that $\int_{\mathbb{R}} \hat{P}_X(t)\,\mathrm{d}t \le 2\pi c < \infty$. □

Theorem 96 Let $X_n \in \mathbb{L}_2(\Omega,\mathcal{A},P), n \in \mathbb{N}$ be independent identically distributed random centered variables with a bounded density $f(x)$ and $\mathrm{var}(X_1) = \sigma^2 \in (0,\infty)$. Denote $S_n = \sum_{k=1}^n X_k$ and $Y_n = \frac{1}{\sigma\sqrt{n}} S_n$. Then Y_n has a continuous density $f_{Y_n}(y)$ for $n \ge 2$ and

$$\sup_{x\in\mathbb{R}} |f_{Y_n}(y) - \varphi(y)| \to 0, \quad n \to \infty.$$

Proof: First, we show that Y_n has a continuous density if $n \ge 2$. Put $h(t) = \hat{P}_{X_1}(t)$. Then $|h(t)|^2 = \hat{P}_{X_1}(t)\hat{P}_{-X_2}(t) = \hat{P}_{X_1-X_2}(t)$. By the previous lemma in order to show that $|h(t)|^2$ is integrable, it is enough to show that there exists a bounded density of $Z = X_1 - X_2$, i.e. that the distribution function of Z is Lipschitz. Obviously,

$$F_Z(z) = P(X_1 - X_2 < z) = P(X_1 < X_2 + z) = \int P(X_1 < x + z) f_{X_1}(x)\,\mathrm{d}x,$$

and therefore the corresponding density of Z can be considered in the form

$$f_Z(z) = \int f_{X_1}(x+z) f_{X_1}(x)\,\mathrm{d}x.$$

Now, it is seen that $f_Z(z) \le c$ holds if $0 \le f_{X_1}(x) \le c$ holds for every $x \in \mathbb{R}$. Obviously,

$$|\hat{P}_{Y_n}(t)| = |h(\tfrac{t}{\sigma\sqrt{n}})|^n \le |h(\tfrac{t}{\sigma\sqrt{n}})|^2 \in \mathbb{L}_1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$$

and we obtain from inversion formula for the densities that f_{Y_n} has a bounded and continuous version given by the inversion formula.

Second, we use the above-mentioned inversion formula for densities in order to obtain that

$$|f_{Y_n}(y) - \varphi(y)| = \tfrac{1}{2\pi} \big| \int e^{-ity} \big(h(\tfrac{t}{\sigma\sqrt{n}})^n - e^{-\frac{1}{2}t^2}\big)\,\mathrm{d}t \big| \le \tfrac{1}{2\pi} \int |h(\tfrac{t}{\sigma\sqrt{n}})^n - e^{-\frac{1}{2}t^2}|\,\mathrm{d}t.$$

Let $\varepsilon > 0$ be fixed now and put

$$\begin{aligned} I_n(\varepsilon) &= \int e^{-\frac{1}{2}t^2} 1_{[|t|\ge\varepsilon\sigma\sqrt{n}]}\,\mathrm{d}t \\ J_n(\varepsilon) &= \int |h(\tfrac{t}{\sigma\sqrt{n}})|^n \cdot 1_{[|t|\ge\varepsilon\sigma\sqrt{n}]}\,\mathrm{d}t \\ K_n(\varepsilon) &= \int |h(\tfrac{t}{\sigma\sqrt{n}})^n - e^{-\frac{1}{2}t^2}| \cdot 1_{[|t|<\varepsilon\sigma\sqrt{n}]}\,\mathrm{d}t. \end{aligned}$$

Since $\int e^{-\frac{1}{2}t^2}\,\mathrm{d}t < \infty$, we immediately obtain that $I_n(\varepsilon) \to 0$ as $n \to \infty$ holds for every $\varepsilon > 0$. Since $\int |h(t)|^2\,\mathrm{d}t < \infty$ and $h(t)$ is a uniformly continuous function, we obtain that $|h(t)| \to 0$ as $|t| \to \infty$, otherwise we would find $\varepsilon, \delta > 0$ and a sequence of t_n such that $(t_n - \delta, t_n + \delta), n \in \mathbb{N}$ are disjoint intervals such that $|h(t)| \ge \varepsilon$ holds on their union, which contradicts integrability condition. Since X_1 has a density, it does not have an equidistant distribution, which means that $|h(t)| < 1$ holds for every $t \ne 0$. Since $|h(t)|$ is a continuous function tending to zero as $|t| \to \infty$ and attaining values in $[0,1)$ on $\mathbb{R}\backslash\{0\}$, we get that

$$c(\varepsilon) := \sup\{|h(t)|; |t| \ge \varepsilon\} \in [0,1)$$

holds for every $\varepsilon > 0$. Then

$$J_n(\varepsilon) \le c(\varepsilon)^{n-2} \int_{\mathbb{R}} |h(\tfrac{t}{\sigma\sqrt{n}})|^2\,\mathrm{d}t = c(\varepsilon)^{n-2} \sigma\sqrt{n} \int_{\mathbb{R}} |h(t)|^2\,\mathrm{d}t \to 0$$

as $n \to \infty$. Since $EX_1 = 0, E(X_1/\sigma)^2 = 1$, we get that $h(\frac{r}{\sigma}) = \hat{P}_{X_1/\sigma}(r) = 1 - \frac{1}{2}r^2 + o(r^2)$ as $r \to 0$. Hence, there exists $\delta > 0$ such that

$$|h(\tfrac{r}{\sigma})| \leq 1 - \tfrac{r^2}{4} \leq e^{-\frac{r^2}{4}}$$

holds whenever $|r| \leq \delta$. If $\varepsilon \in (0, \delta/\sigma]$, then

$$|h(\tfrac{t}{\sigma\sqrt{n}})|^n \cdot 1_{[|t|<\varepsilon\sigma\sqrt{n}]} \leq [e^{-\frac{1}{4}(\frac{t}{\sqrt{n}})^2}]^n = e^{-\frac{1}{4}t^2}.$$

Hence, we have a convergent majorant $e^{-\frac{1}{4}t^2} + e^{-\frac{1}{2}t^2}$, and Dominated Convergence Theorem gives that $K_n(\varepsilon) \to 0$ as $n \to \infty$, since

$$h(\tfrac{t}{\sigma\sqrt{n}})^n = \hat{P}_{Y_n}(t) \to e^{-\frac{1}{2}t^2}$$

as $n \to \infty$ by Lévy-Lindeberg central limit theorem. $\square$