Lecture notes in Probability Theory

By Carlos Buitrago

Moscow Institute of Physics and Technology

These notes contain the most important definitions and theorems which are discussed in my course in Probability Theory that I teach at Moscow Institute of Physics and Technology (MIPT). This Course is based, almost completely, on the lectures in Probability theory and Measure theory taught in the Russian program in Computer Science at the Department of Applied Mathematics and Informatics in MIPT. The aim of these notes is to help my students to better understand the structure of this course. This text and all the material of my lectures is available in my personal page *jcbuitrago.com.* Please contact me if you spot any mistake.

Table of Contents

Part 1 - Discrete Probability

Part 2 - Continuous Distributions

Part I - Discrete Probability

Basics of Combinatorics:

	With repetiton	Without repetiton
Combinations	$\binom{n+k-1}{k}$	$\binom{n}{k}$
Permutations	n^k	$\frac{n!}{(n-k)!}$

Binomial Theorem: $\forall n \in \mathbb{Z}_+ \quad (x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

Main idetities:

(1) $\binom{n}{k} = \binom{n}{n-k}$

(2) $\overline{\binom{n}{k}} = \binom{n+k-1}{k}$

(3) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$

(4) $\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{n} = 2^n$

(5) $\binom{n}{0}^2 + \binom{n}{1}^2 + \dots + \binom{n}{n}^2 = \binom{2n}{n}$

(6) $\binom{n+m}{n} = \binom{n+m-1}{n-1} + \binom{n+m-2}{n-1} + \dots + \binom{n-1}{n-1}$

From (6) We get

(6.1) $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

(6.2) $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

Multinomial Theorem: $(x_1 + \dots + x_k)^n = \sum_{\substack{(n_1,\dots,n_k): \\ \forall i \; n_i \geq 0 \\ n_1+\dots+n_k = n}} \frac{n!}{n_1! \dots n_k!} x^{n_1} \dots x^{n_k}$

Inclusion-Exclusion Principle: $a_1, \dots, a_N$ - objects; $\alpha_1, \dots, \alpha_n$ properties

$N(\alpha_i)$ - # of objects for which property α_i holds.

$N(\alpha_i')$ - # of objects for which property α_i doesn't hold.

$$N(\alpha_1', \dots, \alpha_n') = N - N(\alpha_1) - \dots - N(\alpha_n) + N(\alpha_1, \alpha_2) + \dots + N(\alpha_{n-1}, \alpha_n) - N(\alpha_1, \alpha_2, \alpha_3) - \dots + (-1)^n N(\alpha_1, \dots, \alpha_n)$$

Probability Space $(\Omega, \mathcal{F}, P)$

A probability space consits of three elements:

(1) A sample space Ω, which is the set of all possible outcomes.

(2) A set of events $\mathcal{F}$, which is a σ-algebra on the set Ω.

(3) A probability measure P, which is a function that assigns a number between 0 and 1 to each event.

Def (σ-algebra): A σ-algebra $\mathcal{F}$ on Ω is a collection of subsets of Ω (i.e. $\mathcal{F} \subseteq 2^{\Omega}$) s.t. it satisfies the following conditions:

(1) $\Omega \in \mathcal{F}$
(2) $A \in \mathcal{F} \rightarrow \bar{A} \in \mathcal{F}$
(3) $\{A_n\}_{n=1}^{\infty} \in \mathcal{F} \longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

Def (Probability Measure): A probability measure P on a measurable space $(\Omega, \mathcal{F})$ is a function $P: \mathcal{F} \rightarrow [0,1]$ for which the following conditions hold:

(1) $P(\Omega) = 1$

(2) $\{A_n\}_{n=1}^{\infty} \in \mathcal{F} \wedge \forall i \neq j \; A_i \cap A_j = \emptyset \longrightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

Properties of the Probability Measure P

(1) $P(\bar{A}) = 1 - P(A)$

(2) $A \subseteq B \longrightarrow P(B \setminus A) = P(B) - P(A)$

(3) $A \subseteq B \longrightarrow P(A) \leq P(B)$

(4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(5) Continuity: $A_1 \subset A_2 \subset \ldots \longrightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} P(A_n)$

$A_1 \supset A_2 \supset \ldots \longrightarrow P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} P(A_n)$

Classical Model

(1) Ω - finite, $\Omega = \{\omega_1, \ldots, \omega_n\}$, $|\Omega| = n$.

(2) $\mathcal{F} = 2^{\Omega}$

(3) $\forall i \; P(\{\omega_i\}) = \frac{1}{n}$

Then $\forall A \in \mathcal{F}$ $\boxed{P(A) = \frac{|A|}{|\Omega|}}$

Proposition: If $\Omega < \infty$ or countable and $\mathcal{F} = 2^{\Omega}$, then it is enough to define $\forall i \; P(\{\omega_i\})$ in order to define $P(A)$ $\forall A \in \mathcal{F}$.

Bernoulli Scheme: $k\in\mathbb{N}$, $\Omega=\{(x_1,\dots,x_k), x_i\in\{0,1\}\}$, $|\Omega|=2^k$, $p\in(0,1)$

$$P(\{\omega=(x_1,\dots,x_k)\}) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

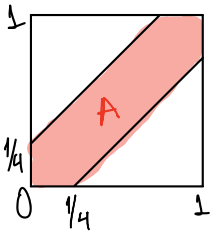
Usually we are interested in the number of successes in n trials: $P\left(\sum_{i=1}^{n} x_i = m\right) = \binom{n}{m} p^m (1-p)^{n-m}$

Notice that $\sum_{\omega\in\Omega} P(\omega) = \sum_{m=0}^{n}\binom{n}{m} p^m (1-p)^m = (p+(1-p))^n = 1.$

Geometric Probability: $\Omega\subset\mathbb{R}^d$: $\mu(\Omega)<\infty$, $\mathcal{F}=\{A\subseteq\Omega:\mu(A)<\infty\}$, $P(A)=\frac{\mu(A)}{\mu(\Omega)}$

Example: Both the bus and you get to the bus stop at random between 12pm and 1pm. When the bus arrives, it waits for 15 mins before leaving. When you arrive, you wait for 15 mins before leaving if the bus doesn't come. What is the probability that you catch the bus?

Solution: $P(A) = \frac{1-5/16}{1} = \frac{7}{16}$

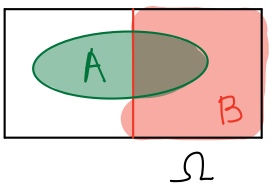


Theorem (Probabilistic Principle of Inclusion-Exclusion): $A_1, A_2, \dots, A_n\in\mathcal{F}$ then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}A_{i_2}) + \dots + (-1)^{n-1} P(A_1\cdot\ldots\cdot A_n)$$

Lemma: $A_1, A_2, \dots, A_n\in\mathcal{F}$ then $P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$.

Conditional Probability: $A, B\in\mathcal{F}$, $P(B)>0$, then the probability of A under the condition B is given by $P(A|B) = \frac{P(A\cap B)}{P(B)}$



Theorem (Multiplication Theorem): $A_1, \dots, A_n\in\mathcal{F}$ s.t. $P(A_1\cdot\ldots\cdot A_k)>0$ $\forall k\in\{1,\dots,n\}$, then

$$P(A_1\cdot\ldots\cdot A_n) = P(A_1)P(A_2|A_1)\cdot\ldots\cdot P(A_n|A_1\cdot\ldots\cdot A_{n-1})$$

It's natural to think that if A and B were independent events, then $P(A|B)=P(A)$

Def (Independence of events): $A, B\in\mathcal{F}$. A and B are called independent ($A \perp\!\!\!\perp B$) if $P(AB)=P(A)\cdot P(B)$.

Proposition: $A \perp\!\!\!\perp B \longrightarrow \bar{A} \perp\!\!\!\perp \bar{B}$

Def (Pairwise Independence): $A_1, \dots, A_n$ are called pairwise independent if $\forall\, i \neq j$ $A_i \perp\!\!\!\perp A_j$.

Def (Mutual Independence): $A_1, \dots, A_n$ are called mutually independent if $\forall\, I \subseteq \{1, \dots, n\}$

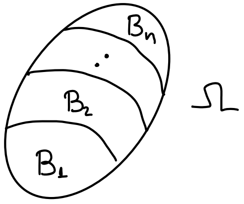
$$P\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} P(A_i).$$

Notice that here we need to verify $\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$ equalities.

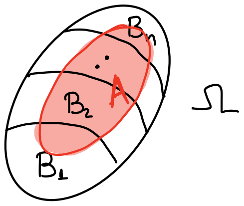
Def (Partition): A collection of events $\{B_n\}_{n=1}^{\infty(N)}$ is called a partition of Ω if the following holds:

(1) $\forall\, i \neq j$ $B_i \cap B_j = \emptyset$

(2) $\bigsqcup_{i=1}^{\infty(N)} B_i = \Omega$



Theorem (Total Probability Formula): Let $\{B_n\}_{n=1}^{\infty(N)}$ be a partition of Ω s.t. $\forall i$ $P(B_i) > 0$.



Then $\forall A \in \mathcal{F}$ $$P(A) = \sum_{n=1}^{\infty(N)} P(A \mid B_n) P(B_n).$$

Theorem (Bayes Formula): Let $\{B_n\}_{n=1}^{\infty(N)}$ be a partition of Ω s.t. $\forall i$ $P(B_i) > 0$. Let $A \in \mathcal{F}$ be s.t. $P(A) > 0$. Then

$$P(B_n \mid A) = \frac{P(A \mid B_n) P(B_n)}{\sum_{n=1}^{\infty(N)} P(A \mid B_n) P(B_n)}$$

Random Variables

Def (Borel σ-algebra): The Borel σ-algebra $(\mathcal{B}(\mathbb{R}))$ is the smallest σ-algebra containing all open sets of $\mathbb{R}$.

Def (Measurable function): Consider the measurable spaces $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. A function $\xi : \Omega \to \mathbb{R}$ is said to be measurable if $\forall\, B \in \mathcal{B}(\mathbb{R})$ it holds that

$$\xi^{-1}(B) := \{\omega \in \Omega \mid \xi(\omega) \in B\} \in \mathcal{F}.$$

Def (Random Variable): Any measurable function on $(\Omega, \mathcal{F}, P)$ is known as a random variable.

Let X_ξ denote the set of all possible values of ξ.

If X_ξ finite or countable, then we call ξ a discrete random variable.

Def (Probability Distribution): We call a probability distribution, any probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

Notice that $\mathbb{P}:\mathcal{B}(\mathbb{R})\to[0,1]$.

If $\exists X$ countable or finite s.t. $\mathbb{P}(X)=1$, then $\mathbb{P}$ is called a Discrete Probability Distribution.

Def (Prob. Distribution of ξ): The probability distribution of a random variable ξ is the prob. measure $\mathbb{P}_\xi$ in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ generated by ξ. That is,

$$\forall B\in\mathcal{B}(\mathbb{R}) \quad \mathbb{P}_\xi(B)=P\big(\{\omega:\xi(\omega)\in B\}\big)=P(\xi\in B)$$

When ξ is discrete, we may use the PMF (Prob. Mass Function) of ξ to completely describe its probability distribution.

The PMF of ξ (p_ξ) is a function that gives the probability that ξ is equal to some value.

Let $X_\xi=\{x_k\}_{k=1}^{\infty(N)}$, then $p_\xi:\mathbb{R}\to[0,1]$ s.t. $\begin{cases}\forall x_k\in X_\xi & p_\xi(x_k)=\mathbb{P}_\xi(\{x_k\})\\ \forall x_k\notin X_\xi & p_\xi(x_k)=0\end{cases}$

We will write $\boxed{p_\xi(x_k)=P(\xi=x_k).}$ Notice that $\sum_{k=1}^{\infty(N)}P(\xi=x_k)=1$.

Def (Independence of Discrete Random Variables): ξ and η are called independent r.v. ($\xi\perp\!\!\!\perp\eta$) if

$$\forall x_k\in X_\xi \text{ and } \forall y_j\in X_\eta \text{ it holds that } P(\xi=x_k,\eta=y_j)=P(\xi=x_k)\,P(\eta=y_j)$$

Similarly as we did for events, we can define mutual and pairwise independence for a set of ≥ 3 random variables.

Examples of Discrete Distributions:

(1) Uniform Distribution $U(X)$, X-finite. $\forall x\in X$ $\mathbb{P}(x)=\frac{1}{|X|}$

(2) Bernoulli Distribution $Ber(p)$, $p\in(0,1)$, $X=\{0,1\}$, $\mathbb{P}(0)=1-p$, $\mathbb{P}(1)=p$

(3) Binomial Distribution $Bin(n,p)$, $p\in(0,1)$, $n\in\mathbb{N}$, $X=\{0,1,\dots,n\}$, $\mathbb{P}(k)=\binom{n}{k}p^k(1-p)^{n-k}$

(4) Geometric Distribution $Geom(p)$, $p\in(0,1)$, $X=\mathbb{N}$, $\mathbb{P}(k)=(1-p)^{k-1}p$

(5) Poisson Distribution $Pois(\lambda)$, $\lambda>0$, $X=\mathbb{N}\cup\{0\}$, $\mathbb{P}(k)=\frac{\lambda^k e^{-\lambda}}{k!}$

Theorem (Poisson Limit Theorem): $\xi\sim Bin(n,p)$, $pn\to\lambda>0$. Then $\forall k\in\mathbb{N}\cup\{0\}$ $P(\xi=k)\longrightarrow\frac{\lambda^k e^{-\lambda}}{k!}$.

Convolution of Discrete Distributions: Let ξ, η be discrete independent random variables. Then

$$P(\xi+\eta=k)=\sum_{i=-\infty}^{+\infty} P(\xi=i)\,P(\eta=k-i).$$

Examples: (1) $\xi \sim Bin(n,p)$, $\eta \sim Bin(m,p)$. Then $\xi+\eta \sim Bin(n+m,p)$

(2) $\xi \sim Pois(\lambda_1)$, $\eta \sim Pois(\lambda_2)$. Then $\xi+\eta \sim Pois(\lambda_1+\lambda_2)$

Numerical Characteristics of Random Variables

Def (Expected Value of a Discrete r.v.): The Expected value of a discrete random variable ξ ($E\xi$) is given by

$$E\xi := \sum_{\omega\in\Omega} \xi(\omega)\,P(\omega),$$ if the series absolutely converges, otherwise it's not defined.

Properties of Expectation (Discrete Case)

(1) $E(a\xi+b)=aE\xi+b$, $a,b\in\mathbb{R}$.

(2) Let $X_\xi=\{x_k\}_{k=1}^{\infty(N)}$, then $E\xi=\sum_{i=1}^{\infty(N)} x_k P(\xi=x_k)$

(3) Let $\varphi(\xi):\mathbb{R}\to\mathbb{R}$. Then $E\varphi(\xi)=\sum_{i=1}^{\infty(N)} \varphi(x_k)P(\xi=x_k)$

(4) $\xi\le\eta \to E\xi\le E\eta$

(5) $|E\xi|\le E|\xi|$

(6) $\xi \perp\!\!\!\perp \eta \longrightarrow E\xi\eta=E\xi E\eta$

Def (Variance of a random Variable): $Var\,\xi := E(\xi-E\xi)^2$

Properties of Variance:

(1) $Var\,\xi = E\xi^2-(E\xi)^2$

(2) $Var\,\xi\ge 0$. Moreover $Var\,\xi=0 \leftrightarrow P(\xi=const)=1$.

(3) $Var(a\xi+b)=a^2\,Var\,\xi$, $a,b\in\mathbb{R}$.

Examples

(1) $\xi\sim U(\{1,\dots,n\})$, $E\xi=\frac{n+1}{2}$, $Var\,\xi=\frac{(n+1)(n-1)}{12}$

(2) $\xi\sim Ber(p)$, $E\xi=p$, $Var\,\xi=p(1-p)$

(3) $\xi\sim Bin(n,p)$, $E\xi=np$, $Var\,\xi=np(1-p)$

(4) $\xi\sim Geom(p)$, $E\xi=\frac{1}{p}$, $Var\,\xi=\frac{1-p}{p^2}$

(5) $\xi\sim Pois(\lambda)$, $E\xi=\lambda$, $Var\,\xi=\lambda$

Def (Covariance): The covariance between two random variables ξ and η is a measure the "linear independence" of these two random variables. It is given by

$$\operatorname{cov}(\xi,\eta) := E[(\xi - E\xi)(\eta - E\eta)]$$

Properties of Covariance:

(1) $\operatorname{cov}(\xi,\eta) = E\xi\eta - E\xi E\eta$

(2) $\operatorname{cov}(\xi,\eta) = \operatorname{cov}(\eta,\xi)$

(3) $\xi \perp\!\!\!\perp \eta \rightarrow \operatorname{cov}(\xi,\eta) = 0$

(4) $\operatorname{cov}(\xi,\xi) = \operatorname{Var}\xi$

(5) $\operatorname{cov}(a\xi_1 + b\xi_2, \eta) = a\operatorname{cov}(\xi_1,\eta) + b\operatorname{cov}(\xi_2,\eta)$

Proposition: $\operatorname{Var}(\xi_1 + \xi_2 + \dots + \xi_n) = \sum_{i,j=1}^{n} \operatorname{cov}(\xi_i,\xi_j)$

Def (Correlation Coefficient): $\operatorname{Corr}(\xi,\eta) = \dfrac{\operatorname{cov}(\xi,\eta)}{\sqrt{\operatorname{Var}\xi}\sqrt{\operatorname{Var}\eta}}$, where $P(\xi,\eta \neq \text{const}) = 1$.

Theorem: $|\operatorname{corr}(\xi,\eta)| \leq 1$.

Proposition: If $\operatorname{corr}(\xi,\eta) = \pm 1$, then ξ and η are linearly dependent each other.

Binomial Random Graph $G(n,p)$

Consider a set $\{1,2,\dots,n\}$ of n vertices. With probability p, we draw an edge in every possible pair of vertices.

Let $\mathcal{G}$ be a fixed graph in n vertices with $|E(\mathcal{G})|$ edges.

Then $P(G(n,p) = \mathcal{G}) = p^{|E(\mathcal{G})|}(1-p)^{\binom{n}{2} - |E(\mathcal{G})|}$.

On the other hand, $P(G(n,p) \text{ has } m \text{ edges}) = \binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2} - m}$

Let ξ be the number of triangles in $G(n,p)$. Then $E\xi = \binom{n}{3} p^3$

Important Inequalities

Markov's Inequality: Let $\xi \geq 0$, $E\xi > 0$, $a > 0$. Then $P(\xi \geq a) \leq \dfrac{E\xi}{a}$.

Chebyshev's Inequality: Let ξ be s.t. $\exists \operatorname{Var}\xi$. $a > 0$. Then $P(|\xi - E\xi| \geq a) \leq \dfrac{\operatorname{Var}\xi}{a^2}$

Couchy-Bunyakovsky Inequality: $E|\xi\eta| \le \sqrt{E\xi^2 E\eta^2}$

Weak Law of Large Numbers (Chebyshev's Form): Let $\xi_1, \xi_2, \dots$ be i.i.d.r.v. s.t. $E\xi_1 = a$, $Var\,\xi_1 = \sigma^2$.

Let $S_n = \xi_1 + \dots + \xi_n$. Then $\forall \varepsilon > 0$ $P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) \longrightarrow 0$ as $n \to +\infty$.

General Form of WLLN: Let $\xi_1, \xi_2, \dots$ be s.t. $\forall i \neq j$ $cov(\xi_i, \xi_j) = 0$ and $\exists C \in \mathbb{R}$ s.t. $\forall i$ $Var\,\xi_i \le C$.

Then $\forall\, \omega(n) \to +\infty$ $P\left(\left|S_n - ES_n\right| > \sqrt{n}\,\omega(n)\right) \longrightarrow 0$ as $n \to +\infty$.

Random Walk

$\xi_1, \xi_2, \dots$ i.i.d.r.v., $S_0 = 0$, $S_n = \xi_1 + \dots + \xi_n$.

$S = \{S_n, n \in \mathbb{Z}_+\}$ – random walk in $\mathbb{R}$.

If $\xi_i = \begin{cases} 1, & 1/2 \\ -1, & 1/2 \end{cases}$, then we have a simple symmetric random walk.

Trajectory $:= \{S_n(\omega), n \in \mathbb{Z}_+\}$, $\omega \in \Omega$.

Main Results

(1) $P(S_n = k) = \dfrac{\binom{n}{\frac{n+k}{2}}}{2^k}$, if $n+k$ – even, otherwise $P(S_n = k) = 0$.

(2) $P(S_1 > 0, \dots, S_{n-1} > 0, S_n = k) = \dfrac{\binom{n-1}{\frac{n+k}{2}-1} - \binom{n-1}{\frac{n+k}{2}}}{2^n}$

(3) $M_n = \max\limits_{1 \le k \le n} S_k$, $P(M_n \ge x) \le 2P(S_n \ge x)$

(4) $X_n = \left|\{k \in \{1, \dots, n\} : S_k = 0\}\right|$, $EX_n = \sqrt{\dfrac{2n}{\pi}} + \underline{\underline{O}}(1)$

(5) $\lim\limits_{n \to \infty} P(S_2 \neq 0, \dots, S_{2n} \neq 0) = \dfrac{\binom{2n}{n}}{2^{2n}}$

(6) $M_n = \max\limits_{1 \le k \le n} S_k$, $P(M_n = N) = \dfrac{1}{2^n}\left[\binom{n}{\frac{n+N+1}{2}} + \binom{n}{\frac{n+N}{2}}\right]$

De Moivre-Laplace Limit Theorem:

Local Theorem: Let $X_n \sim Bin(n,p)$, $\varphi(n) = \bar{\bar{o}}(n^{2/3})$. Then

$$\sup_{\substack{k \in \mathbb{Z}_+ \\ |k-np| \leq \varphi(n)}} \left| \frac{P(X_n = k)}{\frac{1}{\sqrt{2\pi n p(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}} - 1 \right| \longrightarrow 0 \quad \text{as } n \to +\infty$$

Integral Theorem: Let $X_n \sim Bin(n,p)$. Then

$$\sup_{-\infty \leq a < b \leq +\infty} \left| P\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} < b\right) - \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \longrightarrow 0 \quad \text{as } n \to +\infty.$$

Part 2 - Continuous Distributions

We need to remember from part 1 the following concepts:

1) Probability Space $(\Omega, \mathcal{F}, P)$

Def (σ-algebra): A σ-algebra $\mathcal{F}$ on Ω is a collection of subsets of Ω (i.e. $\mathcal{F} \subseteq 2^{\Omega}$) s.t. it satisfies the following conditions:

(1) $\Omega \in \mathcal{F}$

(2) $A \in \mathcal{F} \rightarrow \bar{A} \in \mathcal{F}$

(3) $\{A_n\}_{n=1}^{\infty} \in \mathcal{F} \rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

Def (Probability Measure): A probability measure P on a measurable space $(\Omega, \mathcal{F})$ is a function $P: \mathcal{F} \rightarrow [0,1]$ for which the following conditions hold:

(1) $P(\Omega) = 1$

(2) $\{A_n\}_{n=1}^{\infty} \in \mathcal{F} \wedge \forall i \neq j \ A_i \cap A_j = \emptyset \longrightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$

Properties of the Probability Measure P

(1) $P(\bar{A}) = 1 - P(A)$

(2) $A \subseteq B \rightarrow P(B \setminus A) = P(B) - P(A)$

(3) $A \subseteq B \longrightarrow P(A) \leq P(B)$

(4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(5) Continuity: $A_1 \subset A_2 \subset \dots \rightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$

$A_1 \supset A_2 \supset \dots \rightarrow P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$

Def (Borel σ-algebra): The Borel σ-algebra $(\mathcal{B}(\mathbb{R}))$ is the smallest σ-algebra containing all open sets of $\mathbb{R}$.

Def (Probability Distribution): We call a probability distribution, any probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Notice that $P: \mathcal{B}(\mathbb{R}) \rightarrow [0,1]$.

If $\exists X$ countable or finite s.t. $\mathbb{P}(X) = 1$, then $\mathbb{P}$ is called a Discrete Probability Distribution.

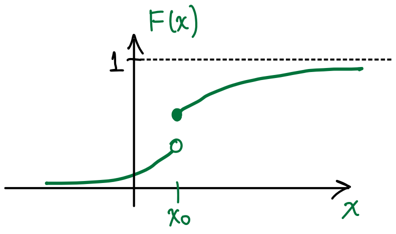
Def (Cumulative Distribution Function): The CDF of a probability distribution is a function $F: \mathbb{R} \rightarrow [0,1]$ defined as $\forall x \in \mathbb{R} \ F(x) = \mathbb{P}((-\infty, x])$.

Proposition: If $\mathbb{P}_1$ and $\mathbb{P}_2$ have the same CDF, then $\mathbb{P}_1 = \mathbb{P}_2$.

Properties of the CDF: (1) $F(x)$ is non-decreasing.

(2) $\lim_{x\to-\infty} F(x)=0$, $\lim_{x\to+\infty} F(x)=1$

(3) $F(x)$ is right-continuous, that is $\lim_{x\to x_0+0} F(x)=F(x_0)$



□ (1) If $x<y$, then $F(y)-F(x)= \underbrace{P((-\infty,y])-P((-\infty,x])}_{\supseteq} = P((-\infty,y]\setminus(-\infty,x]) = P((x,y])\geq 0.$

(2) $x_n\uparrow+\infty$, then $\lim_{n\to+\infty} F(x_n) = \lim_{n\to+\infty} P((-\infty,x_n]) = P\left(\bigcup_n (-\infty,x_n]\right) = P(\mathbb{R}) = 1.$

w.l.o.g. x_n increases and tends to $+\infty$.

$x_n\downarrow-\infty$, then $\lim_{n\to+\infty} F(x_n) = \lim_{n\to+\infty} P((-\infty,x_n]) = P\left(\bigcap_n (-\infty,x_n]\right) = P(\emptyset) = 0.$

(3) $x_n\downarrow x_0$, then $\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P((-\infty,x_n]) = P\left(\bigcap_n (-\infty,x_n]\right) = P((-\infty,x_0]) = F(x_0).$ ▨

Proposition: Every function $F(x)$ satisfying conditions 1), 2), 3) correspond to the CDF of some probability distribution.

There are 3 types of probability distributions:
(1) Discrete
(2) Singular – distribution on an uncountable set of Lebesgue measure zero, the probability of each point in that set is zero.
Example: Cantor distribution, CDF is the devil's staircase.
(3) Absolutely continuous

and any distribution can be decomposed into a mixture of these 3 types.

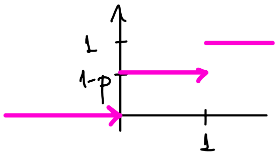
Discrete Distributions:

If $\exists X$ countable or finite s.t. $P(X)=1$, then P is called a Discrete Probability Distribution.

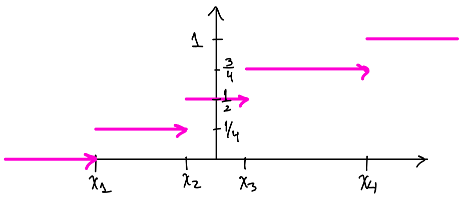
It's CDF is defined as $F(y)=\sum_{\substack{x\in X:\\ x\leq y}} p(x)$, where $p(x)$ – PMF of the distribution.

Examples of CDF of Discrete Distributions:

(1) $Ber(p)$, $X=\{0,1\}$, $P(0)=1-p$
$P(1)=p$



(2) $U(X)$, X-finite, $\forall x\in X$ $P(\{x\})=\frac{1}{|X|}$



Absolutely Continuous Distributions:

A probability distribution is called absolutely continuous if $\exists\ p:\mathbb{R}\to\mathbb{R}_+$ s.t. $\forall x\in\mathbb{R}$ $F(x)=\int_{-\infty}^{x} p(t)\,dt$.

We call $p(t)$ the Probability Density Function (PDF) of the distribution.

In this case $F(x)$ is an absolutely continuous function.

Remark: $F(x)=\mathbb{P}((-\infty,x])=\int_{-\infty}^{x} d\mathbb{P}$, so in case we can replace this Lebesgue integral over measure $\mathbb{P}$ by the Riemann integral by adding a function $p(t)$, we say that $\mathbb{P}$ is absolutely continuous relative to the classical Lebesgue Measure λ.

Properties of the PDF:

(1) If $F(x)$ is differentiable, then $\frac{d}{dx}F(x)=p(x)$

(2) $\int_{-\infty}^{+\infty} p(t)\,dt=1$

(3) $\forall\ B\in\mathfrak{B}(\mathbb{R})$ $\quad \mathbb{P}(B)=\int_B p(t)\,dt$

☐ (1) Let $p(x):=F'(x)$. Then $\int_{-\infty}^{x} p(t)\,dt = F(x)-F(-\infty)=F(x)$.

(2) $\int_{-\infty}^{+\infty} p(t)\,dt = \lim_{x\to+\infty}\int_{-\infty}^{x} p(t)\,dt = \lim_{x\to+\infty} F(x)=1$. ▨

Proposition: If $p:\mathbb{R}\to\mathbb{R}_+$ is s.t. $\int_{\mathbb{R}} p(t)\,dt=1$, then $\exists!$ distribution $\mathbb{P}$ s.t. p is its PDF.

☐ Since we know that every probability distribution is defined by a single CDF, then it's enough to show that for every PDF correspond a single CDF. Notice that $F(x)=\int_{-\infty}^{x} p(t)\,dt$ is uniquely defined.

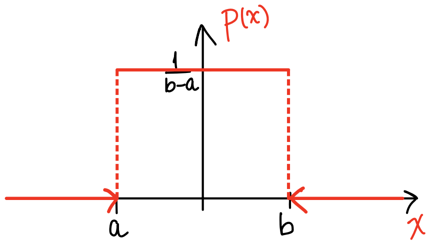
We are left to show tha $F(x)=\int_{-\infty}^{x} p(t)\,dt$ is indeed a CDF (Exercise) ▨

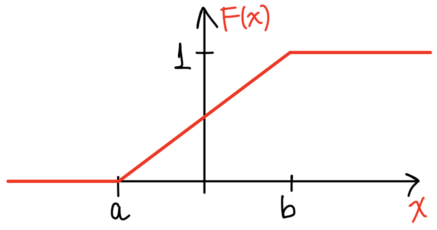
Examples of Absolutely Continuous Distributions:

(1) Uniform Distribution in $[a,b]$, $U([a,b])$

$$p(x)=\frac{1}{b-a}\,I(x\in[a,b])$$

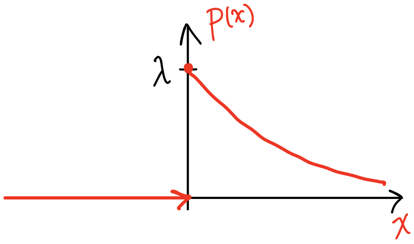
$$F(x)=\int_{-\infty}^{x} p(t)\,dt=\begin{cases}0 & , x<a\\ \frac{x-a}{b-a} & , a\le x\le b\\ 1 & , x>b\end{cases}$$

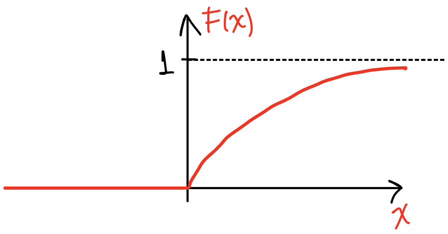




(2) Exponential Distribution with parameter $\lambda>0$, $Exp(\lambda)$

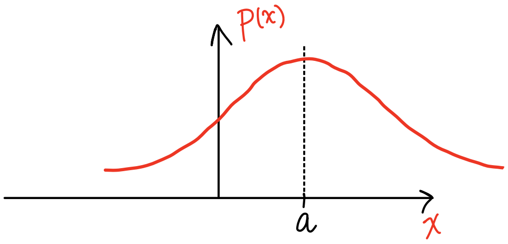
$$p(x) = \lambda e^{-\lambda x} I(x \geq 0) \qquad F(x) = 1 - e^{-\lambda x} I(x \geq 0)$$

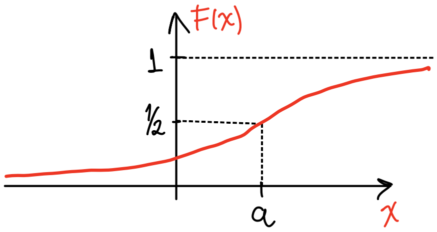




(3) Normal Distribution with parameters $a \in \mathbb{R}$ and $\sigma^2>0$, $N(a,\sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}} \qquad F(x) = \int_{-\infty}^{x} p(t)\,dt$$





The standard Normal distribution $N(0,1)$ is an important special case.

Probability Distributions in $\mathbb{R}^n$

$\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, $\mathbb{P}$ - probability distribution in $\mathbb{R}^n$

$\mathcal{B}(\mathbb{R}^n) = \sigma\left(\{B_1 \times \dots \times B_n ,\ B_i \in \mathcal{B}(\mathbb{R})\}\right) = \sigma\left(\{(-\infty, x_1] \times \dots \times (-\infty, x_n] ,\ x_1, \dots, x_n \in \mathbb{R}\}\right)$

Def (CDF in $\mathbb{R}^n$): A function $F: \mathbb{R}^n \to [0,1]$ is called the CDF of $\mathbb{P}$, if $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$

$$F(x_1, \dots, x_n) = \mathbb{P}\left((-\infty, x_1] \times \dots \times (-\infty, x_n]\right)$$

Proposition: If $F_1 = F_2$, then $\mathbb{P}_1 = \mathbb{P}_2$.

Properties of the CDF in $\mathbb{R}^n$: (1) Let $\Delta^i_{a,b} F(x_1, \dots, x_n) = F(x_1, \dots, x_{i-1}, b, x_{i+1}, \dots, x_n) - F(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n)$

Then $\forall\ a_1 \le b_1, \dots, a_n \le b_n$ $\Delta^1_{a_1,b_1} \cdots \Delta^n_{a_n,b_n} F(x_1, \dots, x_n) \geq 0$.

(2) $\lim\limits_{\substack{x_1 \to +\infty \\ \vdots \\ x_n \to +\infty}} F(x_1, \dots, x_n) = 1$, $\forall i \lim\limits_{x_i \to -\infty} F(x_1, \dots, x_n) = 0$.

(3) $\vec{x}^k \downarrow \vec{x}$ as $k \to +\infty$ (i.e. $\forall i \in \{1, \dots, n\}$ $x_i^k \downarrow x_i$), then $\lim\limits_{k \to +\infty} F(\vec{x}^k) = F(\vec{x})$

☐ (a) It's only necessary to show that $\Delta^1_{a_1,b_1} \cdots \Delta^n_{a_n,b_n} F(x_1, \dots, x_n) = \mathbb{P}\left((a_1,b_1] \times \dots \times (a_n,b_n]\right)$,

since we know that $\mathbb{P}\left((a_1,b_1] \times \dots \times (a_n,b_n]\right) \geq 0$.

Let's do it for n=2. The general case will follow by induction.

$$\Delta^1_{a_1,b_1} \Delta^2_{a_2,b_2} F(x_1, x_2) = \Delta^1_{a_1,b_1}\left(F(x_1, b_2) - F(x_1, a_2)\right) = \Delta^1_{a_1,b_1}\left(\mathbb{P}\big(\underbrace{(-\infty, x_1] \times (-\infty, b_2]}_{\supseteq}\big) - \mathbb{P}\big(\underbrace{(-\infty, x_1] \times (-\infty, a_2]}\big)\right)$$

$$= \Delta^1_{a_1,b_1}\left(\mathbb{P}\left((-\infty, x_1] \times (a_2, b_2]\right)\right) = \mathbb{P}\left((-\infty, b_1] \times (a_2, b_2]\right) - \mathbb{P}\left((-\infty, a_1] \times (a_2, b_2]\right)$$

$$= \mathbb{P}\left((a_1, b_1] \times (a_2, b_2]\right) \geq 0.$$

(b.1) $x_1^k \uparrow +\infty$, $x_2^k \uparrow +\infty$, ..., $x_n^k \uparrow +\infty$; then

$$\lim_{k \to +\infty} F(x_1^k, \dots, x_n^k) = \lim_{k \to +\infty} \mathbb{P}\left((-\infty, x_1^k] \times \dots \times (-\infty, x_n^k]\right) = \mathbb{P}\left(\bigcup_k (-\infty, x_1^k] \times \dots \times (-\infty, x_n^k]\right) = \mathbb{P}(\mathbb{R}^n) = 1$$

(b.2) Let $x_1^k \downarrow -\infty$, then $\lim\limits_{k \to +\infty} F(x_1^k, x_2, \dots, x_n) = \mathbb{P}\left(\bigcap_k (-\infty, x_1^k] \times \dots \times (-\infty, x_n]\right) = \mathbb{P}(\emptyset) = 0$.

(c) Similar to the case on $\mathbb{R}$.

▨

Proposition: Any function with properties (1), (2) and (3) is the CDF of some probability distribution $\mathbb{P}$ in $\mathbb{R}^n$.

Def (Marginal Distribution): Let $\mathbb{P}$ be a distribution in $\mathbb{R}^n$, $i \in \{1, \dots, n\}$, $B \in \mathcal{B}(\mathbb{R})$. We call the following function a marginal distribution:

$$\mathbb{P}_i(B) = \mathbb{P}(\underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{i-1} \times B \times \underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{n-i})$$

Def (Independence of Distributions): Let $\mathbb{P}_1, \dots, \mathbb{P}_n$ be the marginal distributions of $\mathbb{P}$. Distributions $\mathbb{P}_1, \dots, \mathbb{P}_n$ are called independent, if $\forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$

$$\mathbb{P}(B_1 \times \dots \times B_n) = \mathbb{P}\Big(\underbrace{\underbrace{\{B_1 \times \mathbb{R}^{n-1}\}} \cap \underbrace{\{\mathbb{R} \times B_2 \times \mathbb{R}^{n-2}\}} \cap \dots \cap \underbrace{\{\mathbb{R}^{n-1} \times B_n\}}}_{\text{These events are independent}}\Big) = \mathbb{P}(B_1 \times \mathbb{R}^{n-1}) \cdot \mathbb{P}(\mathbb{R} \times B_2 \times \mathbb{R}^{n-2}) \cdot \ldots = \mathbb{P}_1(B_1)\, \mathbb{P}_2(B_2) \cdot \ldots \cdot \mathbb{P}_n(B_n)$$

Similarly, marginal CDFs $F_1, \dots, F_n$ are called independent if $\forall x_1, \dots, x_n \in \mathbb{R}$

$$F(x_1, \dots, x_n) = \mathbb{P}\big((-\infty, x_1] \times \dots \times (-\infty, x_n]\big) = \mathbb{P}_1\big((-\infty, x_1]\big) \cdot \ldots \cdot \mathbb{P}_n\big((-\infty, x_n]\big) = F_1(x_1) \cdot \ldots \cdot F_n(x_n).$$

- If $\mathbb{P}_1, \dots, \mathbb{P}_n$ are discrete distributions in $\mathbb{R}$, then $\mathbb{P} = \mathbb{P}_1 \times \dots \times \mathbb{P}_n$ is a discrete distribution in $\mathbb{R}^n$. In this way, if X_i – finite (or countable) set of possible values of $\mathbb{P}_i$, then $X = X_1 \times \dots \times X_n$ – finite (or countable) is the set of possible values of $\mathbb{P}$.

Absolutely Continuous Distributions in $\mathbb{R}^n$

A probability distribution $\mathbb{P}$ in $\mathbb{R}^n$ is called absolutely continuous if $\exists\ p: \mathbb{R}^n \to \mathbb{R}_+$ s.t.

$\forall (x_1, \dots, x_n) \in \mathbb{R}^n$

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p(t_1, \dots, t_n)\, dt_1 \dots dt_n$$

Proposition: Let $\mathbb{P}_1, \dots, \mathbb{P}_n$ be absolutely continuous distributions in $\mathbb{R}$ with PDFs $p_1, \dots, p_n$. Then $\mathbb{P} = \mathbb{P}_1 \times \dots \times \mathbb{P}_n$ is an absolutely continuous distribution in $\mathbb{R}^n$ with PDF $p(t_1, \dots, t_n) = p_1(t_1) \cdot \ldots \cdot p_n(t_n)$

☐ Let $p(t_1, \dots, t_n) := p_1(t_1) \cdot \ldots \cdot p_n(t_n)$. Let's show that $p(t_1, \dots, t_n)$ is indeed the PDF of $\mathbb{P}$.

$$\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p(t_1, \dots, t_n)\, dt_1 \dots dt_n \underset{\substack{\uparrow \\ \text{Fubini's th. We get iterated integral}}}{=} \int_{-\infty}^{x_1} p(t_1)\, dt_1 \cdot \ldots \cdot \int_{-\infty}^{x_n} p(t_n)\, dt_n = F_1(x_1) \cdot \ldots \cdot F_n(x_n) = F(x_1, \dots, x_n)$$

▨

Properties of the PDF: (1) $\forall B \in \mathcal{B}(\mathbb{R}^n)$ $\quad \mathbb{P}(B) = \int_B p(t_1, \dots, t_n)\, dt_1 \dots dt_n$

(2) If $\forall k \leq n\ \exists F^{(k)}$, then $p(x_1, \dots, x_n) = \dfrac{\partial^n}{\partial x_1 \dots \partial x_n} F$

(3) Let $p:\mathbb{R}^n \to \mathbb{R}_+$ be s.t. $\int\limits_{\mathbb{R}^n} p(t_1,\dots,t_n)\,dt_1\dots dt_n = 1$, then p is the PDF of some $\mathbb{P}$.

Random Elements

It's more convenient to deal with random variables (and its values) than with distributions.

Def: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathfrak{X}, \Sigma)$ a measurable space. Then ξ is called a $(\mathcal{F}|\Sigma)$-measurable function if $\forall A \in \Sigma$ $\xi^{-1}(A) \in \mathcal{F}$. Indeed, if

(1) $(\mathfrak{X}, \Sigma) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then ξ is a random variable.
(2) $(\mathfrak{X}, \Sigma) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, then ξ is a random vector.

Def (σ-algebra generated by ξ): $\mathcal{F}_\xi = \{\xi^{-1}(B), B \in \mathcal{B}(\mathbb{R})\} \subset \mathcal{F}$

Proposition: $\mathcal{F}_\xi$ is a σ-algebra.

□ 1) $\Omega = \xi^{-1}(\mathbb{R}^n) \in \mathcal{F}_\xi$

2) Let $A \in \mathcal{F}_\xi$. $\exists B \in \mathcal{B}(\mathbb{R}^n): \xi^{-1}(B) = A \Rightarrow \overline{A} = \xi^{-1}(\overline{B}) \in \mathcal{F}_\xi$

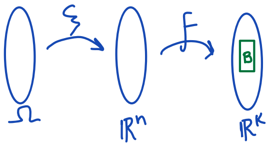
3) Let $A_1, A_2, \dots \in \mathcal{F}_\xi$, $\exists B_1, B_2, \dots \in \mathcal{B}(\mathbb{R}^n): A_i = \xi^{-1}(B_i) \Rightarrow \bigcup A_i = \bigcup \xi^{-1}(B_i) = \xi^{-1}(\bigcup B_i) \in \mathcal{F}_\xi$

▨

Def (Borel function): $f:\mathbb{R}^n \to \mathbb{R}^k$ is called a Borel function, if f is $(\mathcal{B}(\mathbb{R}^n)|\mathcal{B}(\mathbb{R}^k))$-measurable.

Proposition: If f is a continuous function, then f is a Borel function.

Proposition: Let ξ be a n-dimensional random vector. Let $f:\mathbb{R}^n \to \mathbb{R}^k$ be a Borel function. Then $f(\xi)$ is a k-dimensional random vector.



□ We need to show that $\forall B \in \mathcal{B}(\mathbb{R}^k)$ $(f(\xi))^{-1}(B) \in \mathcal{F}$.

$$(f(\xi))^{-1}(B) = \xi^{-1}(\underbrace{f^{-1}(B)}_{\in \mathcal{B}(\mathbb{R}^n)}) \in \mathcal{F} \Rightarrow f(\xi)\text{-measurable}$$

▨

Proposition: Let ξ be a n-dimensional random vector. Let $\eta:\Omega \to \mathbb{R}^k$. Then η is $(\mathcal{F}_\xi|\mathcal{B}(\mathbb{R}^k))$-measurable iff $\exists f:\mathbb{R}^n \to \mathbb{R}^k$ s.t. f is a Borel function and $\eta = f(\xi)$.

□ ($\Leftarrow$) Let $\eta = f(\xi)$. Then $\forall B \in \mathcal{B}(\mathbb{R}^k)$ $\eta^{-1}(B) = \xi^{-1}(\underbrace{f^{-1}(B)}_{\in \mathcal{B}(\mathbb{R}^n)}) \in \mathcal{F}_\xi$.

($\Rightarrow$) This part of the proof is beyond the scope of this course. ▨

Theorem (Measurability Criterion): Let (X,Σ) a measurable space. Let $\xi:\Omega\to X$. Let $\mathcal{E}\subseteq 2^X$ be s.t. $\sigma(\mathcal{E})=\Sigma$.
Then ξ is $(\mathcal{F}|\Sigma)$-measurable iff $\forall A\in\mathcal{E}$ $\xi^{-1}(A)\in\mathcal{F}$.

□ $(\Rightarrow)$ Trivial.

$(\Leftarrow)$ Let $T=\{A\subseteq X : \xi^{-1}(A)\in\mathcal{F}\}$. It's given that $\mathcal{E}\subseteq T$. If T is a σ-algebra, then $\sigma(\mathcal{E})=\Sigma\subseteq T$.

Let's show that T is indeed a σ-algebra: 1) $X\in T$, since $\xi^{-1}(X)=\Omega\in\mathcal{F}$.

2) $A\in T\Rightarrow \xi^{-1}(A)\in\mathcal{F}$ and since $\xi^{-1}(\overline{A})=\overline{\xi^{-1}(A)}\in\mathcal{F}\Rightarrow\overline{A}\in\mathcal{F}$.

3) $A_1,A_2,\ldots\in T\Rightarrow\xi^{-1}(A_i)\in\mathcal{F}$ and since $\xi^{-1}(\cup A_i)=\cup\xi^{-1}(A_i)\in\mathcal{F}\Rightarrow\cup A_i\in T$.

▨

Proposition: $\xi=(\xi_1,\ldots,\xi_n)$ is a random vector iff $\xi_1,\ldots,\xi_n$ are random variables.

□ $(\Rightarrow)$ $i\in\{1,\ldots,n\}$, $B\in\mathcal{B}(\mathbb{R})$ $\quad\xi_i^{-1}(B)=\xi^{-1}(\underbrace{\mathbb{R}^{i-1}\times B\times\mathbb{R}^{n-i}}_{\in\mathcal{B}(\mathbb{R}^n)})\in\mathcal{F}$.

$(\Leftarrow)$ Consider the set of parallelepipeds $\mathcal{E}=\{B_1\times\ldots\times B_n,\ B_i\in\mathcal{B}(\mathbb{R})\}$. $\sigma(\mathcal{E})=\mathcal{B}(\mathbb{R}^n)$

Due to the Measurability Criterion, it's enough to verify that $\forall A\in\mathcal{E}$ $\xi^{-1}(A)\in\mathcal{F}$.

$A=B_1\times\ldots\times B_n$ $\quad\xi^{-1}(B_1\times\ldots\times B_n)=\underbrace{\xi_1^{-1}(B_1)}_{\in\mathcal{F}}\cap\ldots\cap\underbrace{\xi_n^{-1}(B_n)}_{\in\mathcal{F}}\in\mathcal{F}$

▨

Corollary: If ξ and η are random variables, then $\xi+\eta$, $\xi-\eta$, $\xi\eta$, ξ/η are random variables.

□ ξ,η - random variables $\Rightarrow(\xi,\eta)$ - random vector $\Rightarrow f((\xi,\eta))=\xi+\eta$ - random variable (since f - Borel function)

▨

Corollary: Let $\xi_1,\xi_2,\ldots$ be a sequence of random variables defined in the same probability space.
Then $\inf_n\xi_n$, $\sup_n\xi_n$, $\underline{\lim}_{n\to+\infty}\xi_n$, $\overline{\lim}_{n\to+\infty}\xi_n$ are random variables.

Proposition: Let ξ be a n-dimensional random vector. Let $\mathbb{P}_\xi:\mathcal{B}(\mathbb{R}^n)\to[0,1]$ be a function s.t. $\mathbb{P}_\xi(B)=P(\xi\in B)$.
Then $\mathbb{P}_\xi$ is a probability distribution in $\mathbb{R}^n$.

□

1) $\mathbb{P}_\xi(\mathbb{R}^n)=P(\xi\in\mathbb{R}^n)=1$

2) $\{B_i\}_{i=1}^{+\infty}\in\mathcal{B}(\mathbb{R}^n)$ s.t. $\forall i\neq j$ $B_i\cap B_j=\emptyset\Rightarrow\mathbb{P}_\xi(\cup B_i)=\mathbb{P}(\xi^{-1}(\cup B_i))=\mathbb{P}(\underbrace{\cup\xi^{-1}(B_i)}_{\text{non-intersecting}})=\sum\mathbb{P}(\xi^{-1}(B_i))=\sum\mathbb{P}_\xi(B_i)$

▨

We call $\mathbb{P}_\xi$ the probability distribution of ξ, and $F_\xi(x)=\mathbb{P}_\xi((-\infty,x])$ the CDF of ξ. Notice that $F_\xi(x)=P(\xi\le x)$.

If $\mathbb{P}_\xi$ is discrete (absolutely continuous), then we call ξ discrete (absolutely continuous).

If p_ξ is the PDF of $\mathbb{P}_\xi$, the p_ξ is called the PDF of ξ.

Independence of Random Vectors

Def: Let ξ_1,ξ_2 be random vectors of dimensions n_1,n_2 respectively. We say that ξ_1 and ξ_2 are independent, if $\forall\ B_1\in\mathcal{B}(\mathbb{R}^{n_1}),\ B_2\in\mathcal{B}(\mathbb{R}^{n_2})$ $P(\xi_1\in B_1,\ \xi_2\in B_2)=P(\xi_1\in B_1)\,P(\xi_2\in B_2)$. Or, in other words, $\xi_1 \perp\!\!\!\perp \xi_2$ if $\mathcal{F}_{\xi_1}\perp\!\!\!\perp \mathcal{F}_{\xi_2}$.

Def (Mutual Independence): Let Σ be a set of random vectors (not necessarily of the same dimension). These random vectors from Σ are called independent (mutually), if $\forall n\in\mathbb{N}$

$$\left\{\begin{array}{l}\forall \text{ distinct } \xi_1,\dots,\xi_n\in\Sigma\\ \forall\ B_1,\dots,B_n \text{ (of respective dimensions)}\end{array}\right. \longrightarrow P(\xi_1\in B_1,\ \dots,\ \xi_n\in B_n)=P(\xi_1\in B_1)\cdot\ldots\cdot P(\xi_n\in B_n)$$

Theorem (Criterion for Independence of r.v.): Random variables (vectors) $\xi_1,\dots,\xi_n$ are independent iff $\forall\ (x_1,\dots,x_n)\in\mathbb{R}^n$ $(x_1\in\mathbb{R}^{k_1},\ \dots,\ x_n\in\mathbb{R}^{k_n})$ it holds that

$$F_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n)=F_{\xi_1}(x_1)\cdot\ldots\cdot F_{\xi_n}(x_n)$$

Proposition: Let $\xi_1,\dots,\xi_k$ be independent random vectors. Let $f_1,\dots,f_k$ be Borel functions s.t. $f_i:\mathbb{R}^{n_i}\to\mathbb{R}^{m_i}$, where n_i is the dimension of vector ξ_i. Then $f_1(\xi_1),\dots,f_k(\xi_k)$ are independent random vectors.

☐ $B_1\in\mathcal{B}(\mathbb{R}^{m_1}),\ \dots,\ B_k\in\mathcal{B}(\mathbb{R}^{m_k})$

$$P(f_1(\xi_1)\in B_1,\dots,f_k(\xi_k)\in B_k)=P(\xi_1\in \underbrace{f_1^{-1}(B_1)}_{\in\mathcal{B}(\mathbb{R}^{n_1})},\dots,\xi_k\in f_k^{-1}(B_k))=P(\xi_1\in f_1^{-1}(B_1))\cdot\ldots\cdot P(\xi_k\in f_k^{-1}(B_k))$$

$$=P(f_1(\xi_1)\in B_1)\cdot\ldots\cdot P(f_k(\xi_k)\in B_k)$$

▨

Lemma: Let $\xi_1,\dots,\xi_n$ be independent random variables with PDFs $p_1,\dots,p_n$. Then $(\xi_1,\dots,\xi_n)$ is an absolutely continuous random vector with PDF $p(x_1,\dots,x_n)=p_1(x_1)\cdot\ldots\cdot p_n(x_n)$.

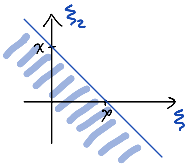
☐ If the statement of the Lemma is true, then we should get

$$\int_{-\infty}^{x_1}\!\!\cdots\int_{-\infty}^{x_n} p(t_1,\dots,t_n)\,dt_1\dots dt_n=F_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n).$$

Indeed,
$$\int_{-\infty}^{x_1}\!\!\cdots\int_{-\infty}^{x_n} p(t_1,\dots,t_n)\,dt_1\dots dt_n=\int_{-\infty}^{x_1}p(t_1)\,dt_1\dots\int_{-\infty}^{x_n}p(t_n)\,dt_n=F_1(x_1)\cdot\ldots\cdot F_n(x_n)=F_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n)$$

▨

Theorem (Convolution's Formula): Let $\xi_1 \perp\!\!\!\perp \xi_2$. Then $F_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} F_{\xi_2}(x-u)\, dP_{\xi_1}$. As a special case, if ξ_1, ξ_2 are absolutely continuous, then $p_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} p_{\xi_2}(x-u)\, p_{\xi_1}(u)\, du$.



$\square$ $$F_{\xi_1+\xi_2}(x) = P(\xi_1+\xi_2 \le x) = \int_{u+v\le x} p_{(\xi_1,\xi_2)}(u,v)\, du\, dv = \int_{u+v\le x} p_{\xi_1}(u)\, p_{\xi_2}(v)\, du\, dv = \int_{-\infty}^{+\infty}\int_{-\infty}^{x-u} p_{\xi_1}(u)\, p_{\xi_2}(v)\, du\, dv = \int_{-\infty}^{+\infty} p_{\xi_1}(u)\, F_{\xi_2}(x-u)\, du$$

If ξ_1, ξ_2 are absolutely continuous, then $$\frac{d}{dx} F_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} p_{\xi_1}(u)\, \frac{d}{dx} F_{\xi_2}(x-u)\, du \Rightarrow p_{\xi_1+\xi_2}(x) = \int_{-\infty}^{+\infty} p_{\xi_1}(u)\, p_{\xi_2}(x-u)\, du$$

▨

Def: Consider the set X. Let $\mathcal{E} \subseteq 2^X$ be a set of subsets of X.
We call $\mathcal{E}$ a π-system, if $\forall A, B \in \mathcal{E}$ $A \cap B \in \mathcal{E}$.
We call $\mathcal{E}$ a λ-system, if
1) $X \in \mathcal{E}$
2) $A, B \in \mathcal{E}$ s.t. $A \subseteq B$, then $B \setminus A \in \mathcal{E}$
3) If $A_n \uparrow A$, $A_i \in \mathcal{E}$, then $A \in \mathcal{E}$.

Lemma: Let $\mathcal{E}$ be a π-system. Then $\sigma(\mathcal{E}) = \lambda(\mathcal{E})$.

Theorem (Criterion for Independence of r.v.): Random variables (vectors) $\xi_1, \dots, \xi_n$ are independent iff $\forall\ (x_1,\dots,x_n)\in\mathbb{R}^n$ $(x_1\in\mathbb{R}^{k_1},\dots,x_n\in\mathbb{R}^{k_n})$ it holds that

$$F_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n)=F_{\xi_1}(x_1)\cdot\ldots\cdot F_{\xi_n}(x_n)$$

□ (⇒) Trivial.

(⇐) Let's prove it using induction. Let's prove that

$$\forall B\in\mathcal{B}(\mathbb{R})\ \forall x_2,\dots,x_n\in\mathbb{R}\quad P(\xi_1\in B,\xi_2\le x_2,\dots,\xi_n\le x_n)=P(\xi_1\in B)\,P(\xi_2\le x_2)\cdot\ldots\cdot P(\xi_n\le x_n)\qquad (1)$$

Let $\Sigma=\{B\in\mathcal{B}(\mathbb{R})$ s.t. statement (1) holds$\}$. It's given that $\Sigma\supset\{(-\infty,x],x\in\mathbb{R}\}$ – π-system.

If Σ - λ-system $\Rightarrow\Sigma\supset\lambda(\{(-\infty,x]\})=\sigma(\{(-\infty,x]\})=\mathcal{B}(\mathbb{R})$ and the proof will be completed.

Let's show that Σ - λ-system:

1) $\mathbb{R}\in\Sigma$ since $P(\xi_1\in\mathbb{R},\xi_2\le x_2,\dots,\xi_n\le x_n)\overset{\text{cont. property}}{=}\lim\limits_{x\to+\infty}P(\xi_1\le x,\xi_2\le x_2,\dots,\xi_n\le x_n)\overset{\text{given conditions}}{=}$

$$=\underbrace{\lim_{x\to+\infty}P(\xi_1\le x)}_{P(\xi_1\in\mathbb{R})=1}P(\xi_2\le x_2)\cdot\ldots\cdot P(\xi_n\le x_n)$$

$$=1\cdot P(\xi_2\le x_2)\cdot\ldots\cdot P(\xi_n\le x_n).$$

2) $A,B\in\Sigma\ \wedge\ A\subseteq B.$ $P(\xi_1\in B\setminus A,\xi_2\le x_2,\dots,\xi_n\le x_n)=P(\xi_1\in B,\dots,\xi_n\le x_n)-P(\xi_1\in A,\dots,\xi_n\le x_n)$

$$=P(\xi_1\in B)\cdot\ldots\cdot P(\xi_n\le x_n)-P(\xi_1\in A)\cdot\ldots\cdot P(\xi_n\le x_n)$$

$$=P(\xi_1\in B\setminus A)\cdot P(\xi_2\le x_2)\cdot\ldots\cdot P(\xi_n\le x_n)\Rightarrow B\setminus A\in\Sigma$$

3) $A_k\uparrow A,\ A_k\in\Sigma.$ $P(\xi_1\in A,\xi_2\le x_2,\dots,\xi_n\le x_n)=\lim\limits_{k\to+\infty}P(\xi_1\in A_k,\xi_2\le x_2,\dots,\xi_n\le x_n)$

$$=\lim_{k\to+\infty}P(\xi_1\in A_k)\cdot\ldots\cdot P(\xi_n\le x_n)$$

$$=P(\xi_1\in A)\cdot\ldots\cdot P(\xi_n\le x_n)\Rightarrow A\in\Sigma$$

Induction Step: Let $\forall\ B_1,\dots,B_m\in\mathcal{B}(\mathbb{R})$ and $\forall\ x_{m+1},\dots,x_n\in\mathbb{R}$ hold the following property

$$P(\xi_1\in B_1,\dots,\xi_m\in B_m,\xi_{m+1}\le x_{m+1},\dots,\xi_n\le x_n)=P(\xi_1\in B_1)\cdot\ldots\cdot P(\xi_m\in B_m)\cdot P(\xi_{m+1}\le x_{m+1})\cdot\ldots\cdot P(\xi_n\le x_n)\qquad (2)$$

Let's prove that property (2) also holds for $m\to m+1$.

$\Sigma=\{B\in\mathcal{B}(\mathbb{R}):\forall B_1,\dots,B_{m+1}\ \forall x_{m+2},\dots,x_n$ property (2) holds$\}$. It remains to prove that Σ - λ-system.

We proceed analogously to the proof of the base of induction. ▨

Def (Expected Value): The expected value of a random variable ξ, defined in $(\Omega, \mathcal{F}, P)$, is given by the Lebesgue integral of $\xi(\omega)$ over the probability measure P.

$$\boxed{E\xi = \int_{\Omega} \xi \, dP}$$

(1) If ξ is discrete, then $E\xi = \sum_{x \in \mathcal{X}} x P(\xi = x)$

(2) If ξ is absolutely continuous, then $E\xi = \int_{\mathbb{R}} x \, p_\xi(x) \, dx$

$\xi = \xi_+ - \xi_-$, where $\xi^+ = \max\{\xi, 0\}$, $\xi^- = \max\{-\xi, 0\}$. Then $E\xi = E\xi^+ - E\xi^-$. If both $E\xi^+, E\xi^- < \infty$, then $E\xi < \infty$. If $E\xi^+ < \infty$ and $E\xi^- = +\infty$, then $E\xi = -\infty$. If $E\xi^+ = \infty$ and $E\xi^- < \infty$, then $E\xi = +\infty$. If $E\xi^+ = E\xi^- = +\infty$, then $E\xi$ doesn't exist.

Example ($E\xi$ doesn't exist): 1) ξ-discrete, $P(\xi = k) = \frac{c}{k^2}$, $k \in \mathbb{Z} \setminus \{0\}$. Since $\sum_{k=1}^{+\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, then $c = \frac{3}{\pi^2}$.

Then $\forall k \in \mathbb{Z} \setminus \{0\}$ $\boxed{P(\xi = k) = \frac{3}{\pi^2 k^2}}$

$$E\xi^+ = \frac{3}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{k} = +\infty = E\xi^- \Rightarrow E\xi \text{ doesn't exist.}$$

2) ξ-absolutely continuous. $p_\xi(x) = \frac{I(|x| \geq 1)}{2x^2}$, then $E\xi^+ = \int_1^{+\infty} \frac{x}{2x^2} dx = +\infty = E\xi^- \Rightarrow E\xi$ doesn't exist.

Properties of Expectation:

(1) $\xi = c \in \mathbb{R} \rightarrow E\xi = c$

(2) $\xi \geq 0 \rightarrow E\xi \geq 0$.
Moreover, $E\xi = 0 \rightarrow P(\xi = 0) = 1$.

(3) $\xi \geq \eta \rightarrow E\xi \geq E\eta$

(4) $E\xi$-finite $\rightarrow |E\xi| \leq E|\xi|$

(5) If $\exists E\xi \rightarrow \forall A \in \mathcal{F}$ $\exists E(\xi I_A)$.
If $|E\xi| < \infty \rightarrow \forall A \in \mathcal{F}$ $|E(\xi I_A)| < \infty$.

(6) $\forall a \in \mathbb{R}$ $E(a\xi + b) = aE\xi + b$

(7) Let $|E\xi|, |E\eta| < \infty$. If $\forall A \in \mathcal{F}$ $E(\xi I_A) \leq E(\eta I_A)$, then $P(\xi \leq \eta) = 1$.

(8) $\xi \perp\!\!\!\perp \eta \rightarrow E\xi\eta = E\xi E\eta$

□ (4) $-|\xi| \leq \xi \leq |\xi| \Rightarrow E(-|\xi|) \leq E\xi \leq E|\xi| \Leftrightarrow -E|\xi| \leq E\xi \leq E|\xi| \Leftrightarrow |E\xi| \leq E|\xi|$.

(5) If $\exists E\xi$, then either $E\xi^+ < \infty$ or $E\xi^- < \infty$. Let $E\xi^+ < \infty$. Then $0 \leq \xi^+ I_A \leq \xi^+ \Rightarrow E(\xi^+ I_A) \leq E\xi^+$
$\Rightarrow E(\xi I_A)^+ = E\xi^+ I_A < \infty$. Similar for $E\xi^- < \infty$.

(7) Let $B = \{\omega : \xi(\omega) > \eta(\omega)\} \in \mathcal{F}$. By the property's conditions, it holds that $E\xi I_B \leq E\eta I_B$. Then

$0 \geq E\xi I_B - E\eta I_B = E(\xi - \eta) I_B \geq 0 \Rightarrow E(\xi - \eta) I_B = 0$, and since $(\xi - \eta) I_B \geq 0 \Rightarrow P((\xi - \eta) I_B = 0) = 1 \Rightarrow P(\xi \leq \eta) = 1$.

(8) $E\xi\eta = \int_{\mathbb{R}^2} xy \, dP_{(\xi,\eta)} = \int_{\mathbb{R}^2} xy \, dP_\xi \, dP_\eta = \int_{\mathbb{R}} x \, dP_\xi \int_{\mathbb{R}} y \, dP_\eta = E\xi E\eta$.

▨

Theorem (Dominated Convergence Theorem): Let $\eta \geqslant 0$, $E\eta < \infty$. Let $\forall n$ $|\xi_n| \leq \eta$. Therefore, if

$$P(\xi_n \to \xi) = 1 \text{, then } E\xi_n \to E\xi.$$

Theorem (On the change of variable): Let ξ be a random variable (vector) s.t. $\exists E\xi$. Let $\varphi : \mathbb{R}^{(n)} \to \mathbb{R}$ be a Borel function. Then

$$E\varphi(\xi) = \int_{\mathbb{R}^{(n)}} \varphi(x)\, dP_\xi$$

□ 1) Let's start with simple functions: $\varphi(x) = I(x \in B)$, $B \in \mathfrak{B}(\mathbb{R})$.

$$E\varphi(\xi) = E\, I(\xi \in B) = P(\xi \in B) = \int_B dP_\xi = \int_{\mathbb{R}} I(x\in B)\, dP_\xi = \int_{\mathbb{R}} \varphi(x)\, dP_\xi.$$

2) $\varphi(x) = \sum_{j=1}^{n} C_j I(x \in B_j)$, $B_j \in \mathfrak{B}(\mathbb{R})$, $\varphi_j(x) = I(x \in B_j)$

$$E\varphi(\xi) = E\sum_{j=1}^{n} C_j \varphi_j(x) = \sum_{j=1}^{n} C_j E\varphi_j(x) = \sum_{j=1}^{n} C_j \int_{\mathbb{R}} \varphi_j(x)\, dP_\xi = \int_{\mathbb{R}} \sum_{j=1}^{n} C_j \varphi_j(x)\, dP_\xi = \int_{\mathbb{R}} \varphi(x)\, dP_\xi$$

3) $\varphi \geqslant 0$ $\exists$ simple $\varphi_n \geqslant 0$ s.t. $\varphi_n \uparrow \varphi$

$$E\varphi(\xi) \underset{\uparrow}{=} \lim_{n\to\infty} E\varphi_n(\xi) = \lim_{n\to\infty} \int_{\mathbb{R}} \varphi_n(x)\, dP_\xi = \int_{\mathbb{R}} \varphi(x)\, dP_\xi$$

Lebesgue integral properties.

4) Consider an arbitrary φ. Let $\varphi = \varphi^+ - \varphi^-$, where $\varphi^+ = \max\{\varphi, 0\}$, $\varphi^- = \max\{-\varphi, 0\}$

$$E\varphi(\xi) = E(\varphi(\xi))^+ - E(\varphi(\xi))^- = E\varphi^+(\xi) - E\varphi^-(\xi) = \int_{\mathbb{R}} \varphi^+(x)\, dP_\xi - \int_{\mathbb{R}} \varphi^-(x)\, dP_\xi = \int_{\mathbb{R}} (\varphi^+ - \varphi^-)\, dP_\xi = \int_{\mathbb{R}} \varphi(x)\, dP_\xi$$

▨

Expectation and Variance of some absolutely continuous distributions:

(1) $\xi \sim U([a,b])$, $E\xi = \frac{b+a}{2}$, $Var\,\xi = \frac{(b-a)^2}{12}$

(2) $\xi \sim Exp(\lambda)$, $E\xi = \frac{1}{\lambda}$, $Var\,\xi = \frac{1}{\lambda^2}$

(3) $\xi \sim \mathcal{N}(a,\sigma^2)$, $E\xi = a$, $Var\,\xi = \sigma^2$

(4) $\xi \sim Gamma(\alpha,\lambda)$, $\alpha,\lambda>0.$ $E\xi = \frac{\alpha}{\lambda}$, $Var\,\xi = \frac{\alpha}{\lambda^2}$, where $p_\xi(x) = \frac{x^{\alpha-1}\lambda^{\alpha}e^{-\lambda x}}{\Gamma(\alpha)} I(x>0)$, $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-t}dt$

(5) $\xi \sim Beta(\alpha,\beta)$, $\alpha,\beta>0.$ $E\xi = \frac{\alpha}{\alpha+\beta}$, $Var\,\xi = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, where $p_\xi(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$, $\mathbb{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$

(6) $\xi \sim Laplace(\mu,\theta)$, $\theta>0.$ $E\xi = \mu$, $Var\,\xi = 2\theta^2$, where $p_\xi(x) = \frac{1}{2\theta}e^{-\frac{|x-\mu|}{\theta}}$

(7) $\xi \sim Cauchy(\theta)$, $\theta>0.$ $E\xi$ doesn't exist, where $p_\xi(x) = \frac{\theta}{\pi(x^2+\theta^2)}$.

Proposition: Let $\alpha>\beta>0$. If $E|\xi|^\alpha<\infty$, then $E|\xi|^\beta<\infty$.

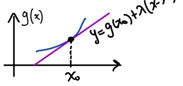
☐ $$E|\xi|^\beta = \int_{\mathbb{R}} |x|^\beta dP_\xi = \int_{|x|\le 1} |x|^\beta dP_\xi + \int_{|x|>1} |x|^\beta dP_\xi \le \int_{|x|\le 1} 1\cdot dP_\xi + \int_{|x|>1} |x|^\beta dP_\xi \le \int_{|x|\le 1} dP_\xi + \int_{|x|>1} |x|^\alpha dP_\xi \le \int_{\mathbb{R}} dP_\xi + \int_{\mathbb{R}} |x|^\alpha dP_\xi = 1 + E|x|^\alpha < \infty$$ ▨

Markov's Inequality: Let $\xi\ge 0$, $a>0$, $E\xi<\infty$. Then $P(\xi\ge a) \le \frac{E\xi}{a}$.

☐ $$\xi \ge \xi I(\xi\ge a) \ge a\, I(\xi \ge a) \Rightarrow E\xi \ge E(a\, I(\xi\ge a)) = a P(\xi\ge a) \Rightarrow P(\xi \ge a) \le \frac{E\xi}{a}$$ ▨

Jensen's Inequality: Let the Borel function $g(x)$ be convex downward and $E|\xi|<\infty$. Then $g(E\xi) \le Eg(\xi)$.

☐ $\forall x_0 \in \mathbb{R}\ \exists \lambda(x_0) \in \mathbb{R}: \forall x \in \mathbb{R}\ \ g(x) \ge g(x_0) + \lambda(x_0)(x-x_0)$. Let $x=\xi$, $x_0 = E\xi$. Therefore



$$g(\xi) \ge g(E\xi) + \lambda(E\xi)(\xi - E\xi) \Rightarrow Eg(\xi) \ge E[g(E\xi) + \lambda E\xi(\xi-E\xi)] = g(E\xi) + \overbrace{E[\lambda E\xi(\xi - E\xi)]}^{=0} \Rightarrow Eg(\xi) \ge g(E\xi)$$ ▨

Notice that $Var\,\xi = E\xi^2 - (E\xi)^2 \ge 0 \Rightarrow E\xi^2 \ge (E\xi)^2$.

Convergence of Random Variables

Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of random variables. Let ξ be a random variable.

(1) **Almost Sure Convergence**: ξ_n is said to converge almost surely to ξ ($\xi_n \xrightarrow{a.s.} \xi$) if

$$P(\{\omega : \xi_n(\omega) \to \xi(\omega)\}) = 1 \quad \text{as } n \to \infty.$$

Example: $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}(\mathbb{R}) \cap [0,1]$, P - Classical Lebesgue measure. Let $\xi_n(\omega) = \omega + \omega^n$ $\forall \omega \in \Omega$ and $n = 1, 2, \dots$
Then $\xi_n \xrightarrow{a.s.} \xi \sim U([0,1])$.
We can define $\xi(\omega) = \omega$. We have $\xi_n(\omega) = \omega + \omega^n \to \omega = \xi(\omega)$ $\forall \omega \in [0,1)$. Notice that $\xi_n(1) = 2 \not\to 1 = \xi(1)$, but $P(\{1\}) = 0$.

(2) **Convergence in Probability**: ξ_n is said to converge in probability to ξ ($\xi_n \xrightarrow{P} \xi$) if

$$\forall \varepsilon > 0 \quad P(\{\omega : |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}) \to 0 \quad \text{as } n \to \infty.$$

Example: $\xi_n \sim Exp(n)$. $\xi_n \xrightarrow{P} 0$.

(3) **Convergence in Distribution**: ξ_n is said to converge in distribution to ξ ($\xi_n \xrightarrow{d} \xi$) if
$\forall f : \mathbb{R} \to \mathbb{R}$ bounded and continuous, it holds that $Ef(\xi_n) \to Ef(\xi)$ as $n \to \infty$.

Theorem (Alexandrov's Theorem): $\xi_n \xrightarrow{d} \xi \iff \forall B \in \mathcal{B}(\mathbb{R}) : \mathbb{P}_\xi(\partial B) = 0 \quad \mathbb{P}_{\xi_n}(B) \to \mathbb{P}_\xi(B) \iff$
$\iff \forall x \in \mathbb{R} : F_\xi$ is continuous in x $\quad F_{\xi_n}(x) \to F_\xi(x)$.

Example: $\xi_n \sim Bin(n, \frac{\lambda}{n})$, $n > \lambda$. Then $\xi_n \xrightarrow{d} \eta \sim Pois(\lambda)$

(4) **Convergence in Mean**: Let $p \geq 1$. ξ_n is said to converge in mean (or in L_p) to ξ ($\xi_n \xrightarrow{L_p} \xi$) if

$$E|\xi_n - \xi|^p \to 0 \quad \text{as } n \to \infty.$$

Example: $\xi_n \sim U([0, 1/n])$. Then $\xi_n \xrightarrow{L_p} 0$.

Theorem:

$$\begin{tikzcd}
a.s. \arrow[dr, Rightarrow] & & \\
& P \arrow[r, Rightarrow] & d \\
L_p \arrow[ur, Rightarrow] & &
\end{tikzcd}$$

☐ (1) $\boxed{a.s. \Rightarrow P}$ Let $P(\xi_n \to \xi) = 1$. $A = \{\omega : \xi_n(\omega) \to \xi(\omega)\} = \{\omega : \forall \varepsilon > 0 \ \exists n \ \forall k \geq n \ |\xi_k - \xi| < \varepsilon\}$

$A_\varepsilon = \{\omega : \exists n \ \forall k \geq n \ |\xi_k - \xi| \leq \varepsilon\} \supset A \Rightarrow P(A_\varepsilon) = 1.$

$B_n = \{\omega : |\xi_n - \xi| < \varepsilon\} \supset \tilde{B}_n = \{\omega : \forall k \geq n \ |\xi_k - \xi| \leq \varepsilon\}$

$$\forall \varepsilon > 0 \qquad A_\varepsilon = \bigcup_n \underbrace{\tilde{B}_n}_{\text{increasing}} \Rightarrow \overbrace{P(A_\varepsilon)}^{\overset{1}{\shortparallel}} = P\Big(\bigcup_n \tilde{B}_n\Big) = \lim_{n \to +\infty} P(\tilde{B}_n) = 1 \Rightarrow \lim_{n \to \infty} P(B_n) = 1 \Rightarrow \lim_{n \to \infty} P(\{\omega : |\xi_n - \xi| \geq \varepsilon\}) = 0.$$

(2) $\boxed{L_p \Rightarrow P}$ Let $E|\xi_n-\xi|^p \to 0$. Then $P(|\xi_n-\xi|>\varepsilon) \underset{\overset{\uparrow}{p>0}}{=} P(|\xi_n-\xi|^p>\varepsilon^p) \le \frac{E|\xi_n-\xi|^p}{\varepsilon^p} \to 0.$

(3) $\boxed{P \Rightarrow d}$ Let $\forall \varepsilon>0$ $P(|\xi_n-\xi|>\varepsilon) \to 0$. Let f be continuous and bounded $(\exists C\ |f|<C)$.

We need: $Ef(\xi_n) \to Ef(\xi)$

Fact 1) Consider $N: P(|\xi|>N) \le \frac{\varepsilon}{100C}$

Let's show why such N exists: Let $\tilde{\xi}_n = I(|\xi|<n)\xi$. Notice that $\tilde{\xi}_n \xrightarrow{a.s.} \xi$, then $\tilde{\xi}_n \xrightarrow{P} \xi$

Therefore, $P(|\tilde{\xi}_n-\xi|>\varepsilon) \to 0$. So, $P(\{\omega: |\tilde{\xi}_n-\xi|>\varepsilon\}) = P(\{\omega: |\xi|>n\}) \to 0$.

Fact 2) In the interval $[-2N, 2N]$ $f(x)$ is absolutely continuous (since f is continuous) and then $\exists \delta: 0<\delta<N$ s.t. $\forall x,y \in [-2N,2N]$ that $|x-y|<\delta$ it holds that $|f(x)-f(y)| < \frac{\varepsilon}{100}$

Fact 3) Since $\xi_n \xrightarrow{P} \xi$, then $\exists N_2: \forall n \ge N_2$ $P(|\xi_n-\xi| \ge \delta) \le \frac{\varepsilon}{100C}$

Now, let's introduce the following partition of Ω: $\begin{cases} A_1 = I(|\xi|>N), \\ A_2 = I(|\xi| \le N, |\xi_n-\xi| \ge \delta), \\ A_3 = I(|\xi| \le N, |\xi_n-\xi|<\delta). \end{cases}$

Let's proceed: Let $n > N_2$, then $|Ef(\xi_n) - Ef(\xi)| = |E(f(\xi_n)-f(\xi))| \le E|f(\xi_n)-f(\xi)| =$

$$= E\underbrace{|f(\xi_n)-f(\xi)|}_{<2C} A_1 + E\underbrace{|f(\xi_n)-f(\xi)|}_{2C} A_2 + E\underbrace{|f(\xi_n)-f(\xi)|}_{<\varepsilon/100} A_3 \le$$

$$\le 2C\,P(|\xi|>N) + 2C\,P(|\xi_n-\xi| \ge \delta) + \frac{\varepsilon}{100} \le 2C \cdot \frac{\varepsilon}{100C} + 2C\frac{\varepsilon}{100C} + \frac{\varepsilon}{100} < \varepsilon$$

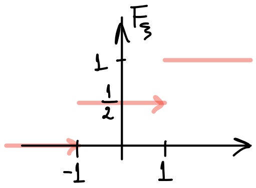
▨

Now, let's have some examples to show that it's not possible to draw more arrows in the statement of the theorem.

Example 1: $d \not\Rightarrow P$

□ Consider the sequence $\xi_1, \xi_2, \ldots$ s.t. $\xi_1 = \xi_2 = \ldots$ and $P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}$. Let $\xi = -\xi_1$, then $F_{\xi_1} = F_{\xi}$

Therefore $\xi_n \xrightarrow{d} \xi$, but $P(|\xi - \xi_n| > 1) = 1$ meaning that $\xi_n \not\xrightarrow{P} \xi$.

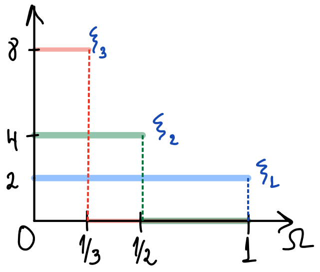


▨

Example 2: a.s. $\not\Rightarrow$ L_p

□

$\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}(\mathbb{R}) \cap [0,1]$, $\mathbb{P}$-classical Lebesgue measure. Let $\xi_n = 2^n I([0, \frac{1}{n}))$



Then, $\xi_n \xrightarrow{a.s.} 0$ since $\forall \omega \in (0,1)$ $\exists n_0$ $\forall n \geq n_0$ $\xi_n(\omega) = 0$.
Notice that $P(\{\omega : \xi_n(\omega) \not\to 0\}) = P(\{0\}) = 0$.

On the other hand, $E|\xi_n - 0|^p = E|\xi_n|^p = 2^{np} \frac{1}{n} \to +\infty$ meaning that $\xi_n \not\xrightarrow{L_p} 0$.

▨

Example 3: $L_p \not\Rightarrow$ a.s.

□

$\Omega = [0,1)$, $\mathbb{P}$-classical Lebesgue measure. Let $\xi_1 = I[0, \frac{1}{2})$, $\xi_2 = I[\frac{1}{2}, 1)$, $\xi_3 = I[0, \frac{1}{4})$, ...

Then $E|\xi_n|^p = P(\xi_n = 1) \to 0$, that is $\xi_n \xrightarrow{L_p} 0$.

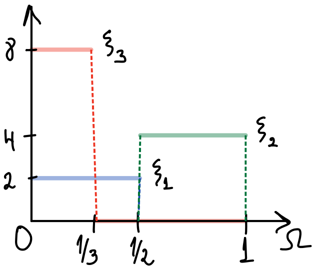
But $\forall \omega \in \Omega$ $\begin{cases} \exists \text{ infinitely many } n \text{ s.t. } \xi_n(\omega) = 1 \\ \exists \text{ infinitely many } n \text{ s.t. } \xi_n(\omega) = 0 \end{cases}$, therefore $\xi_n \not\xrightarrow{a.s.} 0$.

▨

Example 4: $P \not\Rightarrow$ a.s. $\wedge$ $P \not\Rightarrow L_p$

□

Let's combine examples 2 and 3. $\Omega = [0,1)$, $\mathbb{P}$-classical Lebesgue measure.



Let $\xi_1 = 2^1 I[0, 1/2)$, $\xi_2 = 2^2 I[1/2, 1)$, $\xi_3 = 2^3 I[0, 1/4)$, ...

Then $\xi_n \xrightarrow{P} 0$, since $\forall \varepsilon > 0$ $P(|\xi_n - 0| > \varepsilon) = P(\xi_n = 2^n) \to 0$.

But $\xi_n \not\xrightarrow{a.s.} 0$, since $\forall \omega \in \Omega$ $\begin{cases} \exists \text{ infinitely many } n \text{ s.t. } \xi_n(\omega) = \text{const} \\ \exists \text{ infinitely many } n \text{ s.t. } \xi_n(\omega) = 0 \end{cases}$

But $\xi_n \not\xrightarrow{L_p} 0$, since $E|\xi_n|^p = 2^{np} p(\xi_n = 2^n) < 2^{np} \frac{1}{2^{\lfloor \log_2 n \rfloor - 1}} < 2^{np} \frac{4}{n} \longrightarrow +\infty$.

▨

Limit Theorems

Theorem (Poisson Limit Theorem): $\xi \sim Bin(n,p)$, $pn \to \lambda > 0$. Then $\forall k \in \mathbb{N} \cup \{0\}$ $P(\xi = k) \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}$.

Moivre-Laplace Local Theorem: Let $X_n \sim Bin(n,p)$, $\varphi(n) = \bar{o}(n^{2/3})$. Then

$$\sup_{\substack{k \in \mathbb{Z}_+ \\ |k-np| \leq \varphi(n)}} \left| \frac{P(X_n = k)}{\frac{1}{\sqrt{2\pi n p(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}} - 1 \right| \longrightarrow 0 \quad \text{as } n \to +\infty$$

Moivre-Laplace Integral Theorem: Let $X_n \sim Bin(n,p)$. Then

$$\sup_{-\infty \leq a < b \leq +\infty} \left| P\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} < b\right) - \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \longrightarrow 0 \quad \text{as } n \to +\infty.$$

Weak Law of Large Numbers (Chebyshev's Form): Let $\xi_1, \xi_2, \ldots$ be i.i.d.r.v. s.t. $E\xi_1 = a$, $Var\,\xi_1 = \sigma^2$. Let $S_n = \xi_1 + \ldots + \xi_n$. Then $\forall \varepsilon > 0$ $P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) \longrightarrow 0$ as $n \to +\infty$.

General Form of WLLN: Let $\xi_1, \xi_2, \ldots$ be s.t. $\forall i \neq j$ $cov(\xi_i, \xi_j) = 0$ and $\exists C \in \mathbb{R}$ s.t. $\forall i$ $Var\,\xi_i \leq C$. Then $\forall\, \omega(n) \to +\infty$ $P\left(\left|S_n - ES_n\right| > \sqrt{n}\,\omega(n)\right) \longrightarrow 0$ as $n \to +\infty$.

Strong Law of Large Numbers

Theorem 1 (Kolmogorov's): Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables s.t. $\forall i$ $E\xi_i^2 < \infty$. Let there be positive numbers b_n s.t. $b_n \uparrow \infty$ and $\sum_{i=1}^{\infty} \frac{Var\,\xi_i}{b_i^2} < \infty$. Then

$$\frac{S_n - ES_n}{b_n} \xrightarrow{a.s.} 0.$$

Theorem 2 (Kolmogorov's): Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d.r.v. with $E|\xi_1| < \infty$. Then $\frac{S_n - ES_n}{n} \xrightarrow{a.s.} 0$.

Theorem (Central Limit Theorem): Let $\xi_1, \xi_2, \ldots$ be i.i.d.r.v. with $E\xi_1 = a$, $Var\,\xi_1 = \sigma^2$. Then

$$\frac{S_n - na}{\sqrt{n\sigma^2}} \xrightarrow{d} \eta \sim \mathcal{N}(0,1).$$

Theorem (Law of the Iterated Logarithm): Let ξ_n be a sequence of i.i.d.r.v. with $E\xi_1 = 0$ and $E\xi_i^2 = \sigma^2 > 0$. Then

$$P\left(\varlimsup_{n \to \infty} \frac{S_n}{\sqrt{2\sigma^2 n \ln \ln n}} = 1\right) = 1.$$

Conditional Expectation

Def: Let ξ be a random variable in $(\Omega, \mathcal{F}, P)$. Let $A \in \mathcal{F}$ s.t. $P(A) > 0$. Then the conditional expected value of ξ under the condition A is given by $E(\xi|A) := \frac{E(\xi I_A)}{P(A)}$.

Let ξ be a random variable in $(\Omega, \mathcal{F}, P)$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. Then the conditional expectation $E(\xi|\mathcal{G})$ is a random variable in $(\Omega, \mathcal{F}, P)$ s.t.

(1) $E(\xi|\mathcal{G})$ is $\mathcal{G}$-measurable, that is $\forall B \in \mathcal{B}(\mathbb{R})$ $\{\omega : \xi(\omega) \in B\} \in \mathcal{G}$

(2) $\forall A \in \mathcal{G}$ $E(\xi I_A) = E(E(\xi|\mathcal{G}) I_A)$

Proposition: $\Omega = \bigsqcup_{i=1}^{\infty} B_i$, $B_i \in \mathcal{F}$, $\mathcal{G} = \sigma(\{B_1, B_2, \dots\})$, $P(B_i) > 0$. Then $E(\xi|\mathcal{G}) = \sum_{i=1}^{\infty} \frac{E(\xi I_{B_i})}{P(B_i)} I(\omega \in B_i)$

Proposition: If $E|\xi| < \infty$, then $E(\xi|\mathcal{G})$ exists and it is unique almost surely.

Theorem (Radon-Nikodým Theorem): Let ν be an absolutely continuous measure withe respect to P. Then $\exists!$ (P-almost surely) function $f : \Omega \to \mathbb{R}$ $\mathcal{G}$-measurable s.t. $\forall A \in \mathcal{G}$ $\nu(A) = \int_A f \, dP$.

Def: $E(\xi|\eta) := E(\xi|\mathcal{F}_\eta)$

Properties of Conditional Expectation:

(1) If ξ is $\mathcal{F}$-measurable, then $E(\xi|\mathcal{F}) = \xi$

(2) If $\mathcal{F}_\xi \perp\!\!\!\perp \mathcal{G}$, then $E(\xi|\mathcal{G}) = E\xi$

(3) If $\mathcal{G}_1 \subset \mathcal{G}_2$, then $E(E(\xi|\mathcal{G}_1)|\mathcal{G}_2) = E(\xi|\mathcal{G}_1) = E(E(\xi|\mathcal{G}_2)|\mathcal{G}_1)$

(4) $E(E(\xi|\mathcal{G})) = E\xi$

(5) If $\xi_1 \le \xi_2$ almost surely, then $E(\xi_1|\mathcal{G}) \le E(\xi_2|\mathcal{G})$ almost surely.

(6) $E(c_1\xi_1 + c_2\xi_2 | \mathcal{G}) = c_1 E(\xi_1|\mathcal{G}) + c_2 E(\xi_2|\mathcal{G})$

(7) $|E(\xi|\mathcal{G})| \le E(|\xi||\mathcal{G})$

(8) If $P(\xi_n \to \xi) = 1$ and $\forall n \in \mathbb{N}$ $|\xi_n| \le \eta$, $E\eta < \infty$, then $P(E(\xi_n|\mathcal{G}) \to E(\xi|\mathcal{G})) = 1$

(9) If η is $\mathcal{G}$-measurable, then $E(\xi\eta|\mathcal{G}) = \eta E(\xi|\mathcal{G})$

Let η be a discrete random variable s.t. $X_\eta = \{y_i\}_{i=1}^{\infty}$. Then $E(\xi|\eta) = \sum_{i=1}^{\infty} E(\xi|\eta=y_i)\, I(\eta=y_i)$,
where $E(\xi|\eta=y_i) = \frac{E(\xi I(\eta=y_i))}{P(\eta=y_i)}$. Moreover, if ξ is discrete, then $E(\xi|\eta=y_i) = \sum_{x} x\, P(\xi=x|\eta=y_i)$.

Conditional Probability Distribution

Let $\varphi(y) := E(\xi|\eta=y)$ be a Borel function $\mathbb{R}\to\mathbb{R}$ s.t. $\forall B \in \mathcal{B}(\mathbb{R})$ $E(\xi I(\eta\in B)) = \int_B \varphi(y)\, dP_\eta$

Proposition: $E(\xi|\eta) = \varphi(\eta)$.

Def (Conditional Distribution): $\mathbb{P}_\xi(B|\eta=y) = E(I(\xi\in B)\,|\,\eta=y)$

Def (Conditional PDF): If $\exists\, p_{\xi|\eta}(x|y) : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ s.t. $\forall B\in\mathcal{B}(\mathbb{R})$ $\mathbb{P}_\xi(B|\eta=y) = \int_B p_{\xi|\eta}(x|y)\,dx$, then $p_{\xi|\eta}$ is called a conditional PDF.

Proposition: Let (ξ,η) be an absolutely continuous random vector with density $p_{(\xi,\eta)}(x,y)$. Then $\exists$ a conditional PDF of $\mathbb{P}_\xi(B|\eta=y)$, and it is computed (almost surely) by

$$p_{\xi|\eta}(x|y) = \begin{cases} \frac{p_{(\xi,\eta)}(x,y)}{p_\eta(y)}, & p_\eta(y)\neq 0 \\ 0, & p_\eta(y)=0 \end{cases}$$

Proposition: Let $f:\mathbb{R}\to\mathbb{R}$ be a Borel function s.t. $E|f(\xi)|<\infty$. If $\exists\, p_{\xi|\eta}$, then $E(f(\xi)|\eta=y) = \int_{\mathbb{R}} f(x)\, p_{\xi|\eta}(x|y)\,dx$

Characteristic Functions

Def: Let ξ be a random variable. Then the characteristic function of ξ $\varphi_\xi : \mathbb{R} \to \mathbb{C}$ is given by $\varphi_\xi(t) = Ee^{i\xi t}$

Def: Let ξ be a random vector. Then the characteristic function of ξ $\varphi_\xi : \mathbb{R}^n \to \mathbb{C}$ is given by $\varphi_\xi(t) = Ee^{i\langle \xi, t\rangle}$

Properties of the Characteristic Function

(1) $|\varphi_\xi(t)| \leq 1$, $\varphi_\xi(0) = 1$

(2) $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$

(3) $\forall t \in \mathbb{R}$ $\varphi_\xi(t) \in \mathbb{R} \iff \xi \overset{d}{=} -\xi$

(4) $\varphi_\xi(t)$ is uniformly continuous in $\mathbb{R}$.

(5) $\xi \perp\!\!\!\perp \eta \Rightarrow \varphi_{\xi+\eta}(t) = \varphi_\xi(t) \cdot \varphi_\eta(t)$

(6) $\varphi_{a\xi+b}(t) = e^{ibt}\varphi_\xi(at)$

☐ (1) $|\varphi_\xi(t)| = |Ee^{i\xi t}| \leq E|e^{i\xi t}| = 1.$

(2) $\varphi_\xi(-t) = Ee^{i\xi(-t)} = E\cos\xi t - iE\sin\xi t = \overline{\varphi_\xi(t)}.$

(3) $(\Rightarrow)$ Let $\varphi_\xi(t) \in \mathbb{R}$ $\forall t \in \mathbb{R}$. Then $\begin{cases} \varphi_{-\xi}(t) = Ee^{-i\xi t} = \varphi_\xi(-t) = E\cos\xi t - i\cancelto{0}{E\sin\xi t} = E\cos\xi t \\ \varphi_\xi(t) = E\cos\xi t \end{cases}$

Therefore $\xi \overset{d}{=} -\xi$ due to uniqueness theorem.

$(\Leftarrow)$ Let $\xi \overset{d}{=} -\xi$. Then $\varphi_\xi = \varphi_{-\xi}$. Therefore $E(\cancel{\cos\xi t} + i\sin\xi t) = E(\cancel{\cos\xi t} - i\sin\xi t) \Rightarrow E\sin\xi t = 0.$

(4) Let $t > s$. Then $|\varphi(t) - \varphi(s)| = |Ee^{i\xi t} - Ee^{i\xi s}| = |Ee^{i\xi s}(e^{i\xi(t-s)} - 1)| \leq E|e^{i\xi s}(e^{i\xi(t-s)} - 1)|$

$$\overset{|z_1z_2| = |z_1||z_2|,\ z_{1,2}\in\mathbb{C}}{=} E\underbrace{|e^{i\xi s}|}_{1}|e^{i\xi(t-s)} - 1| = E|e^{i\xi(t-s)} + 1| = E|\underbrace{\cos\xi(t-s) + i\sin\xi(t-s) - 1}_{a+ib}| =$$

$$= E\sqrt{(\cos\xi(t-s) - 1)^2 + (\sin\xi(t-s))^2} = E\sqrt{2 - 2\cos\xi(t-s)} \underset{\sin^2 x = \frac{1-\cos 2x}{2}}{=} E\sqrt{4\sin^2\left(\frac{\xi(t-s)}{2}\right)} =$$

$$= E\,2\left|\sin\left(\frac{\xi(t-s)}{2}\right)\right|$$

If $t - s \to 0$, then $\frac{(t-s)\overset{\text{const}}{\overbrace{\xi(\omega)}}}{2} \to 0$ $\forall \omega \in \Omega$. So we can write that $\frac{(t-s)\xi}{2} \xrightarrow{\text{a.s.}} 0$, therefore $2\left|\sin^2\left(\frac{\xi(t-s)}{2}\right)\right| \xrightarrow{\text{a.s.}} 0$, then by the Dominated Convergence Theorem we get $E\,2\left|\sin\left(\frac{\xi(t-s)}{2}\right)\right| \to 0.$

We can conclude that $\forall \varepsilon > 0$ $\exists \delta > 0$ s.t. if $|t - s| < \delta$ then $E\,2\left|\sin\left(\frac{\xi(t-s)}{2}\right)\right| < \varepsilon.$

(5) [Convolution] $\varphi_{\xi+\eta}(t) = Ee^{i(\xi+\eta)t} = Ee^{i\xi t}e^{i\eta t} = E(\cos\xi t + i\sin\xi t)(\cos\eta t + i\sin\eta t)$

$= E(\cos\xi t\cos\eta t + i\cos\xi t\sin\eta t + i\sin\xi t\cos\eta t + i^2\sin\xi t\sin\eta t)$

$= E\cos\xi t\,E\cos\eta t + iE\cos\xi t\,E\sin\eta t + iE\sin\xi t\,E\cos\eta t + iE\sin\xi t \cdot iE\sin\eta t$

$= Ee^{i\xi t}Ee^{i\eta t} = \varphi_\xi(t)\varphi_\eta(t).$

(6) $\varphi_{a\xi+b}(t) = E e^{i(a\xi+b)t} = E e^{ia\xi t} e^{ibt} = e^{ibt} \varphi_\xi(at)$.

Theorem (Uniqueness): Let ξ, η be random vectors. Then $\xi \overset{d}{=} \eta \Leftrightarrow \varphi_\xi = \varphi_\eta$.

Theorem (on Independence): Random variables $\xi_1, \dots, \xi_n$ are independent $\Leftrightarrow \varphi_{(\xi_1,\dots,\xi_n)}(t_1,\dots,t_n) = \prod_{j=1}^{n} \varphi_{\xi_j}(t_j) \ \forall t_1,\dots,t_n \in \mathbb{R}$.

□ (⇒) Let $\xi_1, \dots, \xi_n$ be independent. Then $\varphi_{(\xi_1,\dots,\xi_n)}(t_1,\dots t_n) = E e^{i(\xi_1 t_1 + \dots + \xi_n t_n)} = E e^{i\xi_1 t_1} \cdot \ldots \cdot e^{i\xi_n t_n} =$

$= E e^{i\xi_1 t_1} \cdot \ldots \cdot E e^{i\xi_n t_n} = \varphi_{\xi_1}(t_1) \cdot \ldots \cdot \varphi_{\xi_n}(t_n)$.

similar to the proof of Convolution

(⇐) $\forall t_1,\dots,t_n \in \mathbb{R}$ let $\varphi_{(\xi_1,\dots,\xi_n)}(t_1,\dots,t_n) = \prod_{j=1}^{n} \varphi_{\xi_j}(t_j)$.

Consider the random vector $\eta = (\eta_1, \dots, \eta_n)$ s.t. $\eta_1, \dots, \eta_n$ are independent and $\forall j \ \eta_j \overset{d}{=} \xi_j$

Random vector η exists since we can construct the following CDF: $F(x_1,\dots,x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n)$,

and by Kolmogorov's Theorem $\exists \eta$ s.t. $F(x_1,\dots,x_n)$ is its CDF.

Then, due to the first part of the proof, we get $\varphi_\eta(x_1,\dots,x_n) = \prod_{j=1}^{n} \varphi_{\eta_j}(x_j) = \prod_{j=1}^{n} \varphi_{\xi_i}(x_j) = \varphi_{(\xi_1,\dots,\xi_n)}(x_1,\dots,x_n)$

and by Uniqueness theorem it follows $(\eta_1,\dots,\eta_n) \overset{d}{=} (\xi_1,\dots,\xi_n) \Rightarrow \eta_1,\dots,\eta_n$ are independent. ▨

Characteristic Function of Some Distributions:

(1) $U(\{1,\dots,n\})$, $\varphi_\xi(t) = \dfrac{e^{it} - e^{i(n+1)t}}{n(1-e^{it})}$

(2) $Ber(p)$, $\varphi_\xi(t) = (1-p) + pe^{it}$

(3) $Bin(n,p)$, $\varphi_\xi(t) = (1-p+pe^{it})^n$

(4) $Geom(p)$, $\varphi_\xi(t) = \dfrac{pe^{it}}{1-e^{it}(1-p)}$

(5) $Pois(\lambda)$, $\varphi_\xi(t) = e^{\lambda(e^{it}-1)}$

(6) $U([a,b])$, $\varphi_\xi(t) = \dfrac{e^{itb} - e^{ita}}{it(b-a)}$

(7) $Exp(\lambda)$, $\varphi_\xi(t) = \dfrac{\lambda}{\lambda - it}$

(8) $\mathcal{N}(a,\sigma^2)$, $\varphi_\xi(t) = e^{iat - \frac{t^2\sigma^2}{2}}$

(9) $Gamma(\alpha,\lambda)$, $\varphi_\xi(t) = \left(1 - \dfrac{it}{\lambda}\right)^{-\alpha}$

(10) $Laplace(\mu,\theta)$, $\varphi_\xi(t) = \dfrac{e^{i\mu t}}{1+\theta^2 t^2}$

(11) $Cauchy(\theta)$, $\varphi_\xi(t) = e^{-\theta|t|}$

Theorem (On the Series Expansion): If $E|\xi|^n < \infty$, then $\forall t \in \mathbb{R}$ $\forall r \leq n$ $\exists \varphi_\xi^{(r)}(t)$ and

(1) $E\xi^r = \dfrac{\varphi_\xi^{(r)}(0)}{i^r}$

(2) $\varphi_\xi(t) = \sum\limits_{r=0}^{n} \dfrac{(it)^r}{r!} E\xi^r + \dfrac{(it)^n}{n!}\varepsilon_n(t)$, where $|\varepsilon_n(t)| \leq 3E|\xi|^n$ and $\varepsilon_n(t) \to 0$ as $t \to 0$

(3) If $\exists \varphi_\xi^{(2n)}(0)$ and is finite, then $E\xi^{2n} < \infty$.

(4) If $\forall n \geq 1$ $E|\xi|^n < \infty$ and $\overline{\lim\limits_{n\to\infty}} \dfrac{(E|\xi|^n)^{1/n}}{n} = C < \infty$. Then $\forall t: |t| < \frac{1}{C}$ $\varphi_\xi(t) = \sum\limits_{r=0}^{\infty} \dfrac{(it)^r}{r!} E\xi^r$

Theorem (Inversion Formula): (1) $\forall a<b$: F_ξ is continuous in a and in b $F_\xi(b) - F_\xi(a) = \lim\limits_{C\to\infty} \dfrac{1}{2\pi}\displaystyle\int_{-C}^{C} \dfrac{e^{-ita} - e^{-itb}}{it}\varphi_\xi(t)\,dt$.

(2) If $\displaystyle\int_{-\infty}^{+\infty} |\varphi_\xi(t)|\,dt < \infty$, then $\exists p_\xi(x)$ and $p_\xi(x) = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty} e^{-itx}\varphi_\xi(t)\,dt$.

Theorem (Lévy's Continuity Theorem): (1) If $\xi_n \xrightarrow{d} \xi$, then $\varphi_{\xi_n}(t) \to \varphi_\xi(t)$ $\forall t \in \mathbb{R}$.

(2) If $\varphi_{\xi_n}(t) \to \varphi(t)$ $\forall t \in \mathbb{R}$, φ is continuous in 0. Then $\exists \xi$ s.t. $\xi_n \xrightarrow{d} \xi$ and $\varphi(t) = \varphi_\xi(t)$ $\forall t \in \mathbb{R}$.

☐ (1) $\varphi_{\xi_n}(t) = E\underbrace{\cos\xi_n t}_{f(\xi_n)} + iE\underbrace{\sin\xi_n t}_{g(\xi_n)} \longrightarrow E\cos\xi t + iE\sin\xi t = \varphi_\xi(t)$ $\forall t \in \mathbb{R}$.

where f, g - continuous and bounded.

▨

Theorem (Central Limit Theorem): Let $\xi_1, \xi_2, \ldots$ be i.i.d.r.v. with $E\xi_1 = a$, $Var\,\xi_1 = \sigma^2$. Then

$$\frac{S_n - na}{\sqrt{n\sigma^2}} \xrightarrow{d} \eta \sim N(0,1).$$

☐ Let $\tilde{\xi}_i = \dfrac{\xi_i - a}{\sqrt{\sigma^2}}$, $E\hat{\xi}_i = 0$, $Var\,\hat{\xi}_i = 1$. Then $\dfrac{S_n - na}{\sqrt{n\sigma^2}} = \left[\dfrac{(\xi_1 - a)}{\sqrt{\sigma^2}} + \ldots + \dfrac{(\xi_n - a)}{\sqrt{\sigma^2}}\right]\dfrac{1}{\sqrt{n}} = \dfrac{\tilde{\xi}_1 + \ldots + \tilde{\xi}_n}{\sqrt{n}}$

$$\varphi_{\frac{S_n - na}{\sqrt{n\sigma^2}}}(t) = \varphi_{\frac{\hat{\xi}_1 + \ldots + \hat{\xi}_n}{\sqrt{n}}}(t) = \prod_{j=1}^{n} \varphi_{\frac{\hat{\xi}_j}{\sqrt{n}}}(t) = \prod_{j=1}^{n} \varphi_{\tilde{\xi}_j}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{\hat{\xi}_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + i\left(\frac{t}{\sqrt{n}}\right)E\hat{\xi}_1 - \frac{1}{2}\left(\frac{t}{\sqrt{n}}\right)^2 E\tilde{\xi}_1^2 + \bar{o}\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right]^n$$

$$= \left[1 - \frac{t^2}{2n} + \bar{o}\left(\frac{t^2}{n}\right)\right]^n = e^{n\ln\left(1 - \frac{t^2}{2n} + \bar{o}\left(\frac{1}{n}\right)\right)} = e^{-\frac{t^2}{2} + \bar{o}(1)} \longrightarrow e^{-\frac{t^2}{2}} = \varphi_\eta(t),\ \eta \sim N(0,1).$$

Therefore by Levy's continuity Th. $\dfrac{S_n - na}{\sqrt{n\sigma^2}} \xrightarrow{d} \eta$

▨

WLLN (Khinchin's Form): Let $\xi_1, \xi_2, \dots$ be a sequence of i.i.d.r.v. with $E\xi_1 = a$. Then $\frac{S_n}{n} \xrightarrow{P} a$.

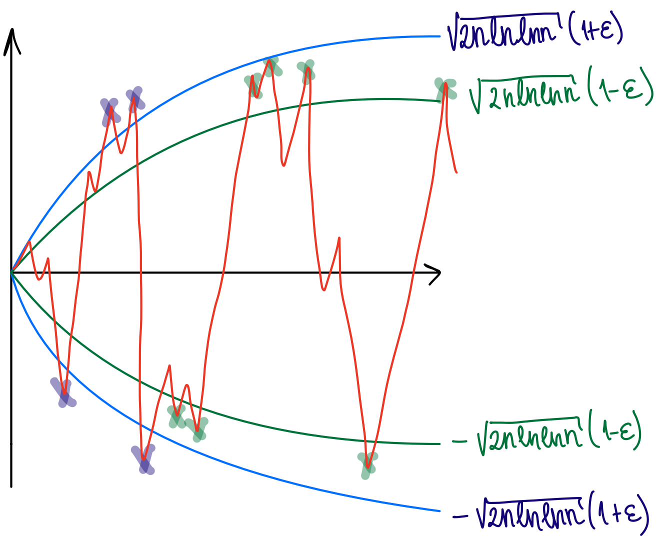
□ $$\varphi_{\frac{S_n}{n}}(t) = \varphi_{\frac{\xi_1+\dots+\xi_n}{n}}(t) = \left[\varphi_{\xi_1}\left(\frac{t}{n}\right)\right]^n = \left[1 + i\left(\frac{t}{n}\right)E\xi_1 + \bar{o}\left(\frac{t}{n}\right)\right]^n = e^{n\ln\left(1+\frac{ita}{n}+\bar{o}\left(\frac{1}{n}\right)\right)} = e^{iat+\bar{o}(1)} \to e^{iat}$$

where $e^{iat} = \varphi_\eta(t)$ and $\eta = a$ a.s. Then by Levy's continuity Th. $\frac{S_n}{n} \xrightarrow{d} a \Rightarrow \frac{S_n}{n} \xrightarrow{P} a$ ▨

WLLN	SLLN
Th.1 (Chebyshev's Form): Let $\xi_1, \xi_2, \dots$ be i.r.v. s.t. $\forall i\ E\xi_i^2 < \infty$. Then $\frac{S_n - ES_n}{n} \xrightarrow{P} 0$	**Th.1 (Kolmogorov's):** Let $\xi_1, \xi_2, \dots$ be i.r.v. s.t. $\forall i\ E\xi_i^2 < \infty$. Let there be positive numbers b_n s.t. $b_n \uparrow +\infty$ and $\sum_{i=1}^{\infty} \frac{Var\,\xi_i}{b_n^2} < \infty$. Then $\frac{S_n - ES_n}{b_n} \xrightarrow{a.s.} 0$
Th.2 (Khinchin's Form): Let $\xi_1, \xi_2, \dots$ be i.i.d.r.v. s.t. $E\lvert\xi_1\rvert < \infty$. Then $\frac{S_n}{n} \xrightarrow{P} E\xi_1$.	**Th.2 (Kolmogorov's):** Let $\xi_1, \xi_2, \dots$ be i.i.d.r.v. s.t. $E\lvert\xi_1\rvert < \infty$. Then $\frac{S_n}{n} \xrightarrow{a.s.} E\xi_1$.

Theorem (Law of the Iterated Logarithm): Let ξ_n be a sequence of i.i.d.r.v. with $E\xi_i = 0$ and $E\xi_i^2 = \sigma^2 > 0$. Then

$$P\left(\overline{\lim_{n\to+\infty}} \frac{|S_n|}{\sqrt{2\sigma^2 n \ln\ln n}} = 1\right) = 1$$



This law gives us a "boundary of operation" of the limit Th:

SLLN: $\forall \omega_n \uparrow +\infty \quad \frac{S_n}{\sqrt{n\ln\ln n}\,\omega_n} \xrightarrow{a.s.} 0$

WLLN: $\forall \omega_n \uparrow +\infty \quad \frac{S_n}{\sqrt{n}\,\omega_n} \xrightarrow{P} 0$

CLT: $\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$

Gaussian Vectors

It is better to define gaussian vectors using their characteristic functions, since the density of these vectors is not always defined. (This happens when Σ is not invertible)

Def: ξ is called a Gaussian random vector if $\varphi_\xi(t) = e^{i\langle a,t\rangle - \frac{1}{2}\langle \Sigma t,t\rangle}$, where $a \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$ is a symmetric and positive semi-definite matrix.

Proposition: ξ is a Gaussian random vector $\Leftrightarrow$ $\exists$ independent $\eta_1, \dots, \eta_m \sim N(0,1)$, $A \in \mathbb{R}^{n\times m}$, $b \in \mathbb{R}^n$ s.t. $\xi = A\begin{pmatrix}\eta_1\\ \vdots \\ \eta_m\end{pmatrix} + b \Leftrightarrow$

$\Leftrightarrow \forall \lambda \in \mathbb{R}^n \ \langle \lambda, \xi\rangle \sim N(a_\lambda, \sigma_\lambda^2)$.

☐ (1 ⇒ 2) Let's consider the case when $\det \Sigma \neq 0$. (so Σ is invertible, in general, of course, this doesn't always hold)

First, let's find the required η_i.

$$\varphi_\xi(t) = E e^{i\langle \xi,t\rangle} = e^{i\langle a,t\rangle - \frac{1}{2}\langle \Sigma t,t\rangle} \Rightarrow \left[\begin{array}{l} e^{-i\langle a,t\rangle} \underbrace{\varphi_\xi(t)}_{Ee^{i\langle \xi,t\rangle}} = e^{-\frac{1}{2}\langle \Sigma t,t\rangle} \\ e^{-i\langle a,t\rangle} E e^{i\langle \xi,t\rangle} = E e^{i\langle \xi - a,t\rangle} = \varphi_{\xi-a}(t) \end{array}\right. \Rightarrow \varphi_{\xi-a}(t) = e^{-\frac{1}{2}\langle \Sigma t,t\rangle}$$

- Now, we must "convert" Σ into the identity matrix, in order to get $e^{-\frac{1}{2}\langle t,t\rangle}$

Since $\det \Sigma \neq 0$ and Σ is symmetric and positive semi-definite then $\Sigma = C^T D C$, where

T is orthogonal $(C^T = C^{-1})$ and D is diagonal matrix without zeros in the diagonal. (Since $\det \Sigma > 0$, so Σ is actually positive definite)

Then

$$\langle \Sigma t,t\rangle = \langle C^T D C\, t, t\rangle = (C^T D C\, t)^T t = (C^T \sqrt{D}\sqrt{D}\, C t)^T t = (\sqrt{D}\, C t)^T (C^T \sqrt{D})^T t = (\sqrt{D}\, C t)^T (\sqrt{D}\, C t)$$

Therefore,

$$\varphi_{\xi-a}(t) = e^{-\frac{1}{2}\langle \Sigma t,t\rangle} = e^{-\frac{1}{2}\langle \sqrt{D}\, C t, \sqrt{D}\, C t\rangle} \Rightarrow \varphi_{\xi-a}\left((\sqrt{D}\, C)^{-1} t\right) = e^{-\frac{1}{2}\langle t,t\rangle} = \varphi_\eta(t)$$

I need a new variable t' s.t. when I replace it, we obtain $e^{-\frac{1}{2}\langle t,t\rangle}$

$t' = (\sqrt{D}\, C)^{-1} t$

Where $\eta = \begin{pmatrix}\eta_1 \\ \vdots \\ \eta_n\end{pmatrix}$ and $\eta_1, \dots, \eta_n$ are independent with $N(0,1)$.

Finally $\varphi_\eta(t) = \varphi_{\xi-a}\left((\sqrt{D^{-1}}C)^{-1}t\right) = Ee^{i\langle \xi-a, C^T\sqrt{D^{-1}}t\rangle} = E\,e^{i\langle \underbrace{\sqrt{D^{-1}}C}_{(C^T\sqrt{D^{-1}})^T}(\xi-a),\, t\rangle} = \varphi_{\sqrt{D^{-1}}C(\xi-a)}(t)$

That is $\underbrace{\sqrt{D^{-1}}C(\xi-a)}_{\text{Let } \tilde{\eta}} \overset{d}{=} \eta \Rightarrow \boxed{\xi = a + (\sqrt{D^{-1}}C)^{-1}\,\tilde{\eta}}$

$(2 \Rightarrow 3)$ $\xi = A\eta + b$. Let $\lambda \in \mathbb{R}^n$

Then $\langle \lambda, \xi \rangle = \lambda^T\xi = \lambda^T A\eta + \lambda^T b \sim \mathcal{N}\left(\lambda^T b, \sum_{i=1}^{m}(\lambda^T a_i)^2\right)$

$(3 \Rightarrow 1)$ Let $\forall \lambda \in \mathbb{R}^n$ $\langle \lambda, \xi \rangle \sim \mathcal{N}(a_\lambda, \sigma_\lambda^2)$. Then

$$\varphi_\xi(\lambda) = Ee^{i\langle \lambda, \xi \rangle} = \varphi_{\underbrace{\langle \lambda, \xi \rangle}_{\text{random variable}}}(1) = e^{ia_\lambda - \frac{1}{2}\sigma_\lambda^2} = e^{i\langle a, \lambda \rangle - \frac{1}{2}\langle \Sigma\lambda, \lambda \rangle}$$

where $a_\lambda = E\langle \lambda, \xi \rangle = \langle \lambda, a \rangle$ and $a = (E\xi_1, \dots, E\xi_n)$

$$\sigma_\lambda^2 = Var\langle \lambda, \xi \rangle = Var(\lambda_1\xi_1 + \dots + \lambda_n\xi_n) = cov(\lambda_1\xi_1 + \dots + \lambda_n\xi_n, \lambda_1\xi_1 + \dots + \lambda_n\xi_n) = \sum_{j_1, j_2} \lambda_{j_1}\lambda_{j_2} cov(\xi_{j_1}, \xi_{j_2})$$

$= \langle \Sigma\lambda, \lambda \rangle \geq 0$. (so Σ is positive semi-definite) and Σ - cov matrix.

▨

Corollary: $a = (E\xi_1, \dots, E\xi_n)$ is called the mean vector and Σ is the covariance matrix of ξ.

Proposition: Let $\xi = (\xi_1, \dots, \xi_n)$ be a Gaussian random vector. $\xi_1, \dots, \xi_n$ are independent $\Leftrightarrow$ $cov(\xi_i, \xi_j) = 0$ $\forall i \neq j$.

☐ If $\forall i \neq j$ $cov(\xi_i, \xi_j) = 0$, then Σ-diagonal matrix.

$$\varphi_\xi(t) = e^{i\langle a, t \rangle - \frac{1}{2}\langle \Sigma t, t \rangle} = e^{i\sum a_j x_j - \frac{1}{2}\sum \sigma_{jj}^2 x_j^2} = \prod_{j=1}^{n} e^{ia_j x_j - \frac{1}{2}\sigma_{jj}^2 x_j^2} = \prod_{j=1}^{n} \varphi_{\xi_j}(x_j)$$

Then $\xi_1, \dots, \xi_n$ are independent random variables

▨

If $\det\Sigma \neq 0$, then for $x \in \mathbb{R}^n$ the PDF of a n-dimensional Gaussian vector is defined as

$$\boxed{p_\xi(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} \exp\left\{-\frac{1}{2}(x-a)^T\Sigma^{-1}(x-a)\right\}}$$

Convergence of random vectors

Let $\xi_1, \xi_2, \dots$ be a random vector with k components.

(1) **Almost Sure Convergence:** ξ_n is said to converge almost surely to ξ ($\xi_n \xrightarrow{a.s.} \xi$) if
$P(\{\omega : \xi_n(\omega) \to \xi(\omega)\}) = 1$ as $n \to \infty$.

(2) **Convergence in Probability:** ξ_n is said to converge in probability to ξ ($\xi_n \xrightarrow{P} \xi$) if
$\forall \varepsilon > 0$ $P(\{\omega : \|\xi_n(\omega) - \xi(\omega)\|_2 > \varepsilon\}) \to 0$ as $n \to \infty$.

(3) **Convergence in Mean:** Let $p \geq 1$. ξ_n is said to converge in mean (or in L_p) to ξ ($\xi_n \xrightarrow{L_p} \xi$) if

$$E\|\xi_n - \xi\|_p^p \longrightarrow 0 \text{ as } n \to \infty, \text{where } \|\xi_n - \xi\|_p = \left(\sum_{i=1}^{k} (\xi_n^i - \xi^i)^p\right)^{1/p}$$

(4) **Convergence in Distribution:** ξ_n is said to converge in distribution to ξ ($\xi_n \xrightarrow{d} \xi$) if
$\forall f : \mathbb{R}^k \to \mathbb{R}$ bounded and continuous, it holds that $Ef(\xi_n) \longrightarrow Ef(\xi)$ as $n \to \infty$.

Theorem:

$$\begin{tikzcd}
a.s. \arrow[dr, Rightarrow] & & \\
 & P \arrow[r, Rightarrow] & d \\
L_p \arrow[ur, Rightarrow] & &
\end{tikzcd}$$

Proposition: (1) $\xi_n \xrightarrow{a.s.} \xi \iff \xi_n^i \xrightarrow{a.s.} \xi^i \quad \forall i \in \{1, \dots, k\}$

(2) $\xi_n \xrightarrow{P} \xi \iff \xi_n^i \xrightarrow{P} \xi^i \quad \forall i \in \{1, \dots, k\}$

(3) $\xi_n \xrightarrow{L_p} \xi \iff \xi_n^i \xrightarrow{L_p} \xi^i \quad \forall i \in \{1, \dots, k\}$

(4) $\xi_n \xrightarrow{d} \xi \implies \xi_n^i \xrightarrow{d} \xi^i \quad \forall i \in \{1, \dots, k\}$

Example for (4): Let ξ, η be i.i.d.r.v. Let $\{\xi_n\}_{n=1}^{+\infty}$ and $\{\eta_n\}_{n=1}^{+\infty}$ be s.t. $\xi_n = \xi$ and $\eta_n = \eta$ $\forall n \in \mathbb{N}$.
Clearly $\xi_n \xrightarrow{d} \xi$ and $\eta_n \xrightarrow{d} \xi$, but $(\xi_n, \eta_n) \not\xrightarrow{d} (\xi, \xi)$, since actually $(\xi_n, \eta_n) \xrightarrow{d} (\xi, \eta)$ and $(\xi, \xi) \overset{d}{\neq} (\xi, \eta)$ (all values for (ξ, ξ) are on the line $y = x$, while for (ξ, η) that's not the case).

Theorem (SLLN for vectors): Let $\xi_1, \xi_2, \dots$ be i.i.d.r.v. s.t. the expectation of all components is finite and $E\xi_1 = a$. Then

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow{a.s.} a$$

Theorem (Multidimensional CLT): Let $\xi_1, \xi_2, \dots$ be i.i.d. random vectors s.t. $E\xi_1 = a$ and Σ is the covariance matrix of ξ_1. Then

$$\frac{S_n - na}{\sqrt{n}} \xrightarrow{d} \eta \sim \mathcal{N}(0, \Sigma)$$

References

[1] Probability (Graduate Texts in Mathematics) 2nd Edition by Albert N. Shiryaev.

[2] Basics of Probability Theory by Zhukovskii M.E. and Rodyonov I.V. (*in Russian*)

[3] Course in Probability Theory and Statistics by Sevastyanov B.A. (*in Russian*)

[4] Introduction To Probability by Joseph K. Blitzstein and Jessica Hwang.