My Thesis, paradoxically and a little provocatively, but nonetheless genuinely, is simply this:

PROBABILITY DOES' NOT EXISTS.

The abandonment of superstitious beliefs about the existence of Phlogiston , the Cosmic Ether, Absolute Space and Time, . . . , or Faires and Witches was an essential step along the road on scientific thinking. Probability too, if regarded as something endowed with some kind of objective existence is no less a misleading misconception, an illusory attempt to exteriorize or materialize our true probabilistic beliefs. Bruno De Finetti, Theory of Probability , a critical introductory treatment (1972).

Lecture Notes in Probability Theory Fall Semester 2015

Dario Gasbarra[1]

December 1, 2015

[1]University of Helsinki, Department of Mathematics and Statistics

Contents

Chapter 0

Getting Started: Bernoulli trials, Random Walk and Weierstrass Theorem

Consider a sequence of n stochastically independent and identically distributed biased coin tosses $Y_1, \dots, Y_n$, with head probability $x \in [0,1]$.

This means that if we denote by 0 the tail outcome and by 1 the head outcome the probability

$$\mathbb{P}_x^{(n)}\big(\{Y_1 = \omega_1, Y_2 = \omega_2, \dots, Y_n = \omega_n\}\ \big) = \prod_{i=1}^{n} \mathbb{P}_x^{(n)}\big(\{Y_i = \omega_i\}\big) = \prod_{i=1}^{n} x^{\omega_i}(1-x)^{1-\omega_i}$$
$$= x^{\sum_{i=1}^n \omega_i}(1-x)^{n-\sum_{i=1}^n \omega_i} = \mathbb{P}_x^{(n)}(\{\ \omega\}\ )$$

for every possible realization $\omega = (\omega_1, \dots, \omega_n) \in \Omega = \{0,1\}^n$, which is the space of all possible outcome sequences. Ω is finite with $\#\Omega = 2^n$.

Note that $P_x^{(n)}(\{\omega\}) \in [0,1]$, and

$$\sum_{\omega \in \Omega} \mathbb{P}_x^{(n)}(\{\ \omega\}) = 1$$

since

$$\sum_{\omega\in\Omega}\mathbb{P}_x^{(n)}(\{\,\omega\}) = \sum_{\omega_1,\dots,\omega_n\in\{0,1\}} x^{\omega_i}(1-x)^{1-\omega_i} =$$
$$\prod_{i=1}^{n}\sum_{\omega_i=0,1} x^{\omega_i}(1-x)^{\omega_i} = \big(x+(1-x)\big)^n = 1^n = 1$$

The singletons $\{\omega\}$ are the elementary events (=subsets) of Ω. Every event (=subset) $A \subseteq \Omega$ is decomposed as finite union

$$A = \bigcup_{\omega\in A}\{\,\omega\}$$

We define the additive probability of A under $\mathbb{P}_x^{(n)}$ as

$$\mathbb{P}_x^{(n)}(A) = \sum_{\omega\in A}\mathbb{P}_x^{(n)}(\{\,\omega\})$$

A random variable is understood as a function of the discrete (finite or countable) outcome space Ω, in this case we have defined $Y_i(\omega) = \omega_i$, taking values in $\{0,1\} \subset \mathbb{R}$.

Soon we will discuss about random variables on an abstract outcome space equipped with a measurable structure.

Define now $S_n = Y_1 + \cdots + Y_n$, and consider the event

$$A_{n,k} = \{S_n = k\} = \{\omega\in\Omega : \omega_1+\omega_2+\cdots+\omega_n = k\}.$$

We have

$$\mathbb{P}_x^{(n)}(S_n = k) = \sum_{\omega\in A_{n,k}}\mathbb{P}_x^{(n)}(\{\,\omega\}) = \sum_{\omega_1,\dots,\omega_n\in\{0,1\}} \mathbf{1}(\omega_1+\cdots+\omega_n = k)\mathbb{P}_x^{(n)}(\{\,\omega\})$$
$$= \binom{n}{k} x^k(1-x)^{n-k}$$

which is obtained just by counting in how many ways we can choose k elements out of a set of size n.

The expectation of a random variable $Y(\omega)$ on a discrete outcome space with respect to the probability $\mathbb{P}$ is defined as

$$\mathbb{E}_{\mathbb{P}}(Y) = \sum_{\omega\in\Omega} Y(\omega)\mathbb{P}(\{\,\omega\}) = \sum_{y\in\mathbb{R}} y P_Y(\{\,y\}\,)$$

where $P_Y(\{y\}) = \mathbb{P}(Y^{-1}\{y\}) = (\mathbb{P}\circ Y^{-1})(\{y\}) = \mathbb{P}(\{\omega : Y(\omega) = y\}\,)$ is the $\mathbb{P}$-probability of the inverse image of $y \in \mathbb{R}$.

Consider now a continuous function $f:\ [0,1] \to \mathbb{R}$, and take the expectation of $f(S_n/n)$,

$$f_n(x) := \mathbb{E}_x^{(n)}\big(f(S_n/n)\big) = \sum_{k=0}^{n} f(k/n)\mathbb{P}_x^{(n)}(S_n = k) = \sum_{k=0}^{n} f(k/n)\binom{n}{k} x^k(1-x)^{n-k},$$

which is a polynomial with respect to the head probability x.

We now illustrate a probabilistic proof due to Bernstein of Weierstrass theorem, which says, every continuous function is approximated with arbitrarily small error by a polynomial, uniformly on compact intervals.

Note that

$$|f_n(x) - f(x)| \ = \left|\mathbb{E}_x^{(n)}\big(f(S_n/n) - f(x)\big)\right| \le \sum_{k=0}^{n} \big|f(k/n) - f(x)\big|\binom{n}{k} x^k(1-x)^{n-k}$$

Since f is continuous on $[0,1]$, it is absolutely continuous and $\forall\eta > 0$ $\exists\eta > 0$ such that for $x, y \in [0,1]$

$$|x-y| \le \eta \longrightarrow |f(x) - f(y)| < \varepsilon$$

It also follows that f is bounded on the compact interval

$$\parallel f \parallel_\infty := \sup_{x\in[0,1]} |f(x)| < \infty$$

which defines the supremum norm on the space of continuous functions $C(\,[0,1] \to \mathbb{R})$.

$$\mathbb{E}_x^{(n)}\big(\big|f(S_n/n) - f(x)\big|\big) \le \varepsilon\mathbb{P}_x^{(n)}(|S_n/n - x| < \eta) + 2 \parallel f \parallel_\infty \mathbb{P}_x^{(n)}\big(|S_n/n - x| \ge \eta\big)$$
$$\le \varepsilon + \frac{2 \parallel f \parallel_\infty}{\eta^2}\mathbb{E}_x^{(n)}\big((S_n/n - x)^2\big)$$

Here we used the following property: if $Z \geq 0$ is a nonnegative random variable, $\forall \eta > 0$ we have

$$Z(\omega) \geq \eta \times \mathbf{1}(Z(\omega) > \eta), \quad \forall \omega \in \Omega$$

Since the expectation preserves the inequality, and $\mathbb{E}\big(\mathbf{1}_A\big) = \mathbb{P}(A)$, the expectation of the indicator of an event is the probability of the event, we obtain

$$\mathbb{E}(Z) \geq \eta \mathbb{P}(Z > \eta)$$

equivalently

$$\mathbb{P}(Z > \eta) \leq \frac{\mathbb{E}(Z)}{\eta}$$

This inequality is referred as Chebychev inequality, or Markov inequality. Note also that in the same way

$$\mathbb{P}(Z > \eta) = \mathbb{P}(Z^2 > \eta^2) \leq \frac{\mathbb{E}(Z^2)}{\eta^2} \tag{0.0.1}$$

and $\forall t > 0$,

$$\mathbb{P}(Z > \eta) = \mathbb{P}\big(\exp(Zt) > \exp(\eta t)\big) \leq \exp(-t\eta)\mathbb{E}\big(\exp(Zt)\big)$$

by minimizing the right hand side we get also Chernoff's inequality

$$\mathbb{P}(Z > \eta) \leq \inf_{t>0} \big\{ \exp(-t\eta)\mathbb{E}\big(\exp(Zt)\big)\big\} \, .$$

These simple inequalities are very powerful.

In our case, to use Chebychev inequality (0.0.1) we need to compute the variance:

$$\frac{1}{n^2}\mathbb{E}_x^{(n)}\big((S_n - nx)^2\big) = \frac{1}{n^2}\mathbb{E}_x^{(n)}\Bigg(\Big\{\sum_{i=1}^n (Y_i - x)\Big\}^2\Bigg)$$
$$= \frac{1}{n^2}\mathbb{E}_x^{(n)}\Bigg(\sum_{i=1}^n (Y_i - x)^2 + 2\sum_{1\leq i<j\leq n}(Y_i - x)(Y_j - x)\Big\}^2\Bigg)$$
$$= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_x^{(n)}\big((Y_i - x)^2\big) + \frac{2}{n^2}\sum_{1\leq i<j\leq n}\mathbb{E}_x^{(n)}\big(Y_i - x\big)\mathbb{E}_x^{(n)}\big(Y_j - x\big) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_x^{(n)}\big((Y_i - x)^2\big)$$
$$= \frac{n(1-x)x}{n^2} = \frac{x(1-x)}{n}$$

where we used the linearity of the expectation together with stochastic independence of Y_i, Y_j for $i \neq j$ which is the property

$$\mathbb{P}_x^{(n)}\big(\big\{Y_i = \omega_i, Y_j = \omega_j\big\}\big) = \mathbb{P}_x^{(n)}\big(\big\{Y_i = \omega_i\big\}\big)\mathbb{P}_x^{(n)}\big(\big\{Y_j = \omega_j\big\}\big) \quad \forall i \neq j,\ \omega_i, \omega_j \in \{0,1\}\ .$$

Independence implies that the expectations factorize:

$$\mathbb{E}_x^{(n)}\Big(\big(Y_i - x\big)\big(Y_j - x\big)\Big) = \mathbb{E}_x^{(n)}\big(Y_i - x\big)\mathbb{E}_x^{(n)}\big(Y_j - x\big)$$

where

$$\mathbb{E}_x^{(n)}\big(Y_i\big) = 1 \times \mathbb{P}_x^{(n)}(Y_i = 1) + 0 \times \mathbb{P}_x^{(n)}(Y_i = 0) = x\ ,$$

and

$$\mathbb{E}_x^{(n)}\big((Y_i - x)^2\big) = (1-x)^2 \times \mathbb{P}_x^{(n)}(Y_i = 1) + (0-x)^2\mathbb{P}_x^{(n)}(Y_i = 0)$$
$$= (1-x)^2 x + x^2(1-x) = (1-x)x$$

We have $\forall x \in [0,1]$,

$$\big|f_n(x) - f(x)\big| \leq \varepsilon + \frac{2 \parallel f \parallel_\infty}{n\eta^2} \sup_{x\in[0,1]} \{\, x(1-x)\}$$
$$= \varepsilon + \frac{\parallel f \parallel_\infty}{n2\eta^2} \leq 2\varepsilon$$

for $n \geq \parallel f \parallel_\infty /(2\varepsilon\eta^2)$ which means that the polynomial $f_n(x)$ is arbitrarily close to $f(x)$ under the uniform norm on the space of continuous functions $C([0,1] \to \mathbb{R})$.

This construction works for any number of steps n, and it is extended to the case with infinitely many steps. In such case the sequence of random variables $(Y_n(\omega) : n \in \mathbb{N})$ defined on the countable space $\Omega = \{\ 0,1\}^{\mathbb{N}}$, which defines a random walk on the integers as

$$X_n(\omega) = \sum_{i=1}^{n}(2Y_i(\omega) - 1) = \big(2S_n(\omega) - n\big) \in \mathbb{Z}$$

where $Y_i(\omega) = \omega_i$ as before.

Chapter 1

Kolmogorov Axioms.

In 1933, the Russian mathematician Andrey Nikolaevich Kolmogorov (1903-1987) published the book *foundations of the Theory of Probability* , which contained Axioms of modern probability theory, the construction of a probability measure on an infinite dimensional product space, as well as the general theory of conditional probability .

Definition 1.0.1 (Algebra and σ-algebra). *Let Ω an abstract set. 2^{Ω} denotes the power set of Ω, that is,the collection of all Ω subsets.*

Let $\mathcal{A} \subseteq 2^{\Omega}$ (possibly smaller) family of subsets. We say that

- *$\mathcal{A}$ is an algebra when*

  1. $\Omega \in \mathcal{A}$
  2. $A \in \mathcal{A} \Longrightarrow A^c = (\Omega \setminus A) \in \mathcal{A}$
  3. $A, B \in \mathcal{A} \Longrightarrow (A \cup B) \in \mathcal{A}$

  Since $A \cap B = \left(A^c \cup B^c\right)^c$, it follows from the definition that algebra is closed with respect finite intersections.

- *An algebra $\mathcal{A}$ is a σ-algebra when it is closed with respect to countable unions,*

$$\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{A} \Longrightarrow \left(\bigcup_{n=1}^{\infty} A_n\right) \in \mathcal{A}$$

  It follows that the σ-algebra is closed countable intersections.

- *The pair $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra of Ω subsets is called a measurable space or probability space .*
- *The elements $A \in \mathcal{A}$ of the σ-algebra are called measurable sets or events.*

Definition 1.0.2. *Consider a probability space $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra on Ω.*

- *A map $\mu : \mathcal{A} \to [0, +\infty]$ is a (positive)* **measure** *when $\mu(\emptyset) = 0$ and μ is σ-additive, which means for all sequences $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{A}$ such that $A_n \cap A_m = \emptyset \quad \forall m \neq n$,*
$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$
- *The measure μ is* **finite** *when $\mu(\Omega) < +\infty$, and it is σ-***finite** *if there exists a countable measurable sequence $\{\Omega_n : n \in \mathbb{N}\} \subseteq \mathcal{A}$ such that*
$$\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n \quad \text{and } \mu(\Omega_n) < \infty \quad \forall n.$$
- *A positive measure μ is a* **probability** *when $\mu(\Omega) = 1$.*

Remark 1.0.1. *Note that we can always normalize a finite measure μ to a probability:*
$$\mathbb{P}(A) := \frac{\mu(A)}{\mu(\Omega)}$$

Remark 1.0.2. *The power set 2^{Ω} is a σ-algebra. Is it always possible to define a probability on the probability space $(\Omega, 2^{\Omega})$, with $\mathbb{P}(B)$ defined $\forall B \subseteq \Omega$? This is OK when Ω is finite or countable. When Ω is uncountable the power set is too big. We will see a counterexample later on.*

1.1 Constructing the Probability Measure by Extension

Definition 1.1.1. *Let P a probability measure on the probability space $(\Omega, \mathcal{A})$. When the event $A \in \mathcal{A}$ has probability $P(A) = 1$ we say that A is a P-almost sure (in short a.s.) event. and the complement event A^c is P-null or P-negligible.*

Moreover when $P(A) = 1$ we can assign $P(B) = 1$ as well to all $B \supseteq A$ even when $B \notin \mathcal{A}$. Analogously if $P(A) = 0$ we assign $P(B) = 0$ to all $B \subseteq A$, also those which are not $\mathcal{A}$-measurable.

Lemma 1.1.1 (monotone convergence of measure)**.** *Given the triple $(\Omega, \mathcal{A}, \mu)$ where μ is a σ-additive measure, let $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{A}$ a non-decreasing sequence of events such that $A_n \subseteq A_{n+1} \quad \forall n$, for*

$$A = \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}, \text{ we introduce the notation } A_n \uparrow A.$$

Then $\mu(A_n) \uparrow \mu(A)$ as $n \uparrow \infty$.

Analogously when $B_n \supseteq B_{n+1} \in \mathcal{A} \quad \forall n$ and

$$B = \bigcap_{n\in\mathbb{N}} B_n \in \mathcal{A}, \text{ with the notation } B_n \downarrow B.$$

Then $\mu(B_n) \downarrow \mu(B)$.

Proof For $C_{n+1} = A_{n+1} \setminus A_n \in \mathcal{A}$.

$$A = \bigcup_{n\in\mathbb{N}} C_n \text{ where } C_i \cap C_j = \emptyset \text{ as } \quad i \neq j.$$

By the σ-additivity of P

$$P(A) = \sum_{k=1}^{\infty} P(C_k) = \lim_{n\uparrow\infty} \sum_{k=1}^{n} P(C_k) = \lim_{n\uparrow\infty} P\left(\bigcup_{k=1}^{n} C_k\right) = \lim_{n\uparrow\infty} P(A_n) \quad \square$$

The second statement follows by considering the complementary events.

Lemma 1.1.2. *The countable union of P-null events P-null. The countable union of P-almost sure events is P-almost sure.*

Problem 1.1.1. *Let $(\Omega, \mathcal{A})$ be a probability space, equipped with an additive probability measure $P : \mathcal{A} \to [0, 1]$. Show that P is σ-additive iff and only if*

$$\{B_n : n \in \mathbb{N}\} \subseteq \mathcal{A}, \quad B_n \downarrow \emptyset \quad \Longrightarrow \quad P(B_n) \downarrow 0 \text{ as } n \uparrow \infty$$

Definition 1.1.2. *For an abstract space Ω, let $\mathcal{C} \subseteq 2^{\Omega}$ a collection of Ω-subsets. We denote by $\sigma(\mathcal{C})$ the σ-algebra generated by $\mathcal{C}$, as the smallest σ-algebra of Ω containing the collection $\mathcal{C}$. meaning that if $\mathcal{A}$ is another σ-algebra on Ω containing $\mathcal{C}$, necessarily $\mathcal{A} \supseteq \sigma(\mathcal{C})$.*

Definition 1.1.3. *Let $(S, \mathcal{T})$ a topological space, where the topology $\mathcal{T}$ is the collection of all the open sets in S.*

We recall the topology axioms: The $\emptyset$ and the whole space S are open, an arbitrary union (including uncountable unions) of open sets is an open set an the finite intersection of open sets is open. Note that the countable intersection of open set does not need to be open, for example consider $\bigcap_n (-1/n, 1/n) = \{ 0\}$, which is closed but it is not open.

The **Borel** *σ-algebra $\mathcal{B}(S) := \sigma(\mathcal{T})$ is the σ-algebra generated by the open sets,*

The events in $\mathcal{B}(S)$ are called Borel sets. Borel set can be quite complicated. Because of that it is useful to define a probability P first on a smaller family of events which generates the σ-algebra $\mathcal{A}$, in such a way that the definition extends uniquely to the whole $\mathcal{A}$.

Example 1.1.1. *Let $S = \mathbb{R}$, and consider the σ-algebra generated by the half-line intervals*

$$\mathcal{A} := \sigma\{(-\infty, x] : x \in \mathbb{R}\} = \mathcal{B}(\mathbb{R})$$

Proof: since

$$(-\infty, x] = \bigcap_{n\in\mathbb{N}} (-\infty, x + n^{-1})$$

is a Borel set, it follows that $\mathcal{A} \subseteq \mathcal{B}(\mathbb{R})$. For $a < b$,

$$(a, b) = (-\infty, a]^c \cap \left(\bigcup_{n\in\mathbb{N}} (-\infty, b - n^{-1}] \right) \in \mathcal{A}$$

Let $U \subseteq \mathbb{R}$ open. $\forall q \in U \cap \mathbb{Q}$ $\exists \varepsilon_q > 0$ such that

$(q - \varepsilon_q, q + \varepsilon_q) \subseteq U$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, this implies

$$U = \bigcup_{q\in U\cap\mathbb{Q}} (q - \varepsilon_q, q + \varepsilon_q)$$

with countable union. It follows that $U \in \mathcal{S}$ and $\mathcal{B}(\mathbb{R}) = \mathcal{A}$.

Construction and extension of measure In this section we follow David Williams' book *Probability with Martingales*.

Definition 1.1.4. *A collection of Ω-subsets $\mathcal{D}$ is called a d-class (Dynkin class) if*

1. $\Omega \in \mathcal{D}$.
2. $A, B \in \mathcal{D}, A \subseteq B \Longrightarrow B \setminus A \in \mathcal{D}$.
3. $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{D}, A_n \uparrow A \Longrightarrow A \in \mathcal{D}$.

Definition 1.1.5. *A collection of Ω-subsets $\mathcal{I}$ is called π-class when it is closed with respect to finite intersections*

$$I_1, I_2 \in \mathcal{I} \Longrightarrow I_1 \cap I_2 \in \mathcal{I}$$

Problem 1.1.2. *The arbitrary intersection of d-classes is a d-class. The arbitrary intersection of π-classes is a π-class, which means*

$$d(\mathcal{C}) = \bigcap_{d\text{-classes } \mathcal{I} \supseteq \mathcal{C}} \mathcal{I}$$

is the smallest d-class which contains $\mathcal{C}$, and

$$\pi(\mathcal{C}) = \bigcap_{\pi\text{-classes } \mathcal{J} \supseteq \mathcal{C}} \mathcal{J}$$

is the smallest π-class containing $\mathcal{C}$. Moreover

$$\sigma(\mathcal{C}) = \bigcap_{\sigma\text{-algebrae } \mathcal{A} \supseteq \mathcal{C}} \mathcal{A}$$

is the smallest σ-algebra containing $\mathcal{C}$.

Lemma 1.1.3 (Dynkin).

1. *A collection of Ω-subsets $\mathcal{C}$ is a σ-algebra if and only if it is simultaneously a π-class and a d-class.*
2. *If $\mathcal{I}$ is a π-class, the smallest d-class containing $\mathcal{I}$ is a σ-algebra, meaning that $d(\mathcal{I}) = \sigma(\mathcal{I})$*

1. **Proof** Clearly a σ-algebra is simultaneously a d-class and a π-class. In the other direction we need to show that if $\mathcal{C}$ is simultaneously a d-class and a π-class it is closed under countable unions.

   Let $\{B_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$. It follows from the assumption that

$$A_n := B_1 \cup B_2 \cup \cdots \cup B_n = (B_1^c \cap B_2^c \cap \cdots \cap B_n^c)^c \in \mathcal{C}$$
$$A_n \uparrow A := \left(\bigcup_{n\in\mathbb{N}} A_n \right) = \left(\bigcup_{n\in\mathbb{N}} B_n \right) \in \mathcal{C} \quad \square$$

2. **Proof**. From the first part of the lemma it is enough to show that $d(\mathcal{I})$ is a π-class. We show first that

$$\mathcal{D}_1 := \{B \in d(\mathcal{I}) : B \cap E \in d(\mathcal{I}) \ \forall E \in \mathcal{I}\} \ , \text{ where } \mathcal{I} \subseteq \mathcal{D}_1 \subseteq d(\mathcal{I})$$

   is a d-class:

   $\Omega \in \mathcal{D}_1$, since $\Omega \in d(\mathcal{I})$, $(\Omega \cap E) = E \in \mathcal{I} \ \forall E \in \mathcal{I}$.

   When $B_1 \subseteq B_2$, $B_i \in \mathcal{D}_1, i = 1, 2$, $(B_2 \setminus B_1) \in d(\mathcal{I})$, and $\forall E \in \mathcal{I}$, $(B_2 \setminus B_1) \cap E = (B_2 \cap E) \setminus (B_1 \cap E)$. Now $(B_i \cap E) \in d(\mathcal{I}), i = 1, 2$ (by the definition of $\mathcal{D}_1$), and since $d(\mathcal{I})$ is a d-class, it follows that $(B_2 \setminus B_1) \cap E \in d(\mathcal{I}) \ \forall E \in \mathcal{I}$, and $(B_2 \setminus B_1) \in \mathcal{D}_1$.

   Similarily, when $\{B_n : n \in \mathbb{N}\} \subseteq \mathcal{D}_1$, $B_n \uparrow B = \bigcup_n B_n$ ja $E \in \mathcal{I}$,

$$B \cap E = \bigcup_n (B_n \cap E)$$

   where $(B_n \cap E) \uparrow (B \cap E)$. Now $(B_n \cap E) \in d(\mathcal{I})$ (by the definition of $\mathcal{D}_1$)) and since $d(\mathcal{I})$ is a d-class it follows that $(B \cap E) \in \mathcal{D}_1 \ \forall E \in \mathcal{I}$. In this way we have shown that $\mathcal{D}_1$ is d-class. Since $\mathcal{I} \subseteq \mathcal{D}_1 \subseteq d(\mathcal{I})$, it follows that $\mathcal{D}_1 = d(\mathcal{I})$.

   Let now

$$\mathcal{D}_2 := \{B \in d(\mathcal{I}) : B \cap A \in d(\mathcal{I}) \ \forall A \in d(\mathcal{I})\} \subseteq d(\mathcal{I})$$

   It follows also that $\mathcal{D}_2 \supseteq \mathcal{I}$, since we have shown that $E \in \mathcal{I}, B \in \mathcal{D}_1 = d(\mathcal{I})$ implies $(E \cap B) \in d(\mathcal{I})$.

By following the previous steps we can show that $\mathcal{D}_2$ is a d-class, as well which implies $\mathcal{D}_2 = d(\mathcal{I})$. It follows that the d-class $d(\mathcal{I})$ is a π-class, and from the first part of the lemma it follows that $d(\mathcal{I})$ is as well a σ-algebra. In general $d(\mathcal{I}) \subseteq \sigma(\mathcal{I})$, since $d(\mathcal{I})$ is a σ-algebra which contains $\mathcal{I}$, it follows that $d(\mathcal{I}) \supseteq \sigma(\mathcal{I})$, which means $d(\mathcal{I}) = \sigma(\mathcal{I})$ □

Exercise 1.1.1. *Prove that, if $\mathbb{P}$ is a probability measure on $\Omega = \mathbb{R}^d$ equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$, every Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ satisfies the following* **Approximation Property** *: for every $\varepsilon > 0$ there is an open set $U \subseteq \mathbb{R}^d$ and a closed set $C \subseteq \mathbb{R}^d$ such that $U \supseteq B \supseteq C$ and $\mathbb{P}(U \setminus C) \leq \varepsilon$.*

To show that consider the class

$$\mathcal{D} = \big\{ B \in \mathcal{B}(\mathbb{R}^d) \text{ which has the Approximation property } \big\} \subseteq \mathcal{B}(\mathbb{R}^d)$$

,

- *Show first that the class*

$$\mathcal{C} = \{ \, C \subseteq \mathbb{R}^d, \; C \text{ closed } \} \subset \mathcal{D}$$

  and it is a π-class (closed under intersections).

  Hint *if C is closed, let $C^\varepsilon = \{ \, y : \exists x \in C \text{ with } |x-y| < \varepsilon \} \supseteq C$.*

  Show that C^ε is open and

$$C = \bigcap_{n \in \mathbb{N}} C^{1/n}$$

  Use the σ-addivity of $\mathbb{P}$ to show that C has the Approximation propery.

- *Then show that $\mathcal{D}$ is a Dynkin class.*
- *Use Dynkin Lemma 1.1.3 to conclude that all Borel sets have the Approximation property.*

Prove also that when $B \in \mathcal{B}(\mathbb{R}^d)$ is a Borel set, $\forall \varepsilon > 0$ one can find an open set U and a compact set K with $U \geq B \geq K$ and $\mathbb{P}(U \setminus K) < \varepsilon$.

Proposition 1.1.1 (Uniqueness of the extension)**.** *Let $\mathcal{I}$ a π-class of Ω-events, $\mathcal{A} = \sigma(\mathcal{I})$. Consider two probability measures P and Q on the probability space $(\Omega, \mathcal{A})$, such that $P(\Omega) = Q(\Omega) = 1$ which coincide on $\mathcal{I}$, meaning that*

$$P(I) = Q(I) \quad \forall I \in \mathcal{I} .$$

Then necessarily $P(A) = Q(A) \quad \forall A \in \mathcal{A}$.

Proof Let

$$\mathcal{D} := \{A \in \mathcal{A} : \; Q(A) = P(A)\} \subseteq \mathcal{A}$$

We show that $\mathcal{D}$ is a d-class: $\Omega \in \mathcal{D}$ (by assumption) and when $A, B \in \mathcal{D}$, $A \subseteq B$,

$$P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$$

sinceP, Q are probabilities on $(\Omega, \mathcal{A})$.

When $A_n \uparrow A$, $A_n \in \mathcal{D}$, $A \in \mathcal{A}$, it follows from lemma (1.1.1) that

$$P(A) = \lim_n P(A_n) = \lim_n Q(A_n) = Q(A) \; ,$$

which implies $A \in \mathcal{D}$. Since $\mathcal{D}$ is a d-class, and by assumption $\mathcal{D} \supseteq \mathcal{I}$ which is a π-class, Dynkin lemma 1.1.3 implies that $\mathcal{D} \supseteq \sigma(\mathcal{I}) = \mathcal{A}$, which means $\mathcal{D} = \mathcal{A} \quad \square$.

Example 1.1.2. *For $\Omega = \mathbb{R}$, the collection*

$$\mathcal{I} := \{(-\infty, t], t \in \mathbb{R}\}$$

is a π-class, generating the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. Kun $F : \mathbb{R} \to [0,1]$, on the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we have at most one probability measure P such that $F(t) := P((-\infty, t]) \; \forall t \in \mathbb{R}$.

We will find soon the necessary and sufficient conditions for the existence of such probability P, namely that the map $F : \mathbb{R} \to [0,1]$ is non-decreasing, right-continuous at all points, and the limits at $\pm\infty$ are $\lim\limits_{t \uparrow +\infty} F(t) = 1$, $\lim\limits_{t \uparrow +\infty} F(-t) = 0$.

Eugen Dynkin(1924-2014) was a soviet Soviet and American mathematician. He has made contributions to the fields of probability and algebra, especially semisimple Lie groups, Lie algebras, and Markov processes.

Theorem 1.1.1 (Caratheodory extension). *Let $\mathcal{A}_0$ Ω an algebra of events, and $\mathcal{A} = \sigma(\mathcal{A}_0)$ the σ-algebra generated by $\mathcal{A}_0$..*

If the map $\lambda_0 : \mathcal{A}_0 \to [0, \infty]$ is σ-additive, there is a unique σ-additive measure $\lambda : \mathcal{A} \to [0, \infty]$ extensing λ_0, meaning that $\lambda(A) = \lambda_0(A) \quad \forall A \in \mathcal{A}_0$.

Proof The uniqueness of the extension was shown in lemma (1.1.1) The proof of existence is divided in several steps.

Lemma 1.1.4. *Let$\mathcal{A}_0 \subseteq 2^{\Omega}$ an event algebra, $\lambda : \mathcal{A}_0 \to [0, +\infty]$ an arbitrary map satisfying $\lambda(\emptyset) = 0$.*

We say that $L \in \mathcal{A}_0$ is a λ-set if it splits properly $\mathcal{A}_0$ in the following way:

$$\lambda(A) = \lambda(A \cap L) + \lambda(A \cap L^c), \quad \forall A \in \mathcal{A}_0$$

- *The collection of λ-events $\mathcal{L}_0$ is an algebra and $\lambda : \mathcal{L}_0 \to [0, +\infty]$ is finitely additive.*
- *If $L_1, L_2, \dots L_n \in \mathcal{L}_0$, $L_i \cap L_j = \emptyset \; \forall i \neq j$, $A \in \mathcal{A}_0$,*

$$\lambda\left(\bigcup_{k=1}^{n} (L_k \cap A)\right) = \sum_{k=1}^{n} \lambda(L_k \cap A)$$

Proof. Clearly Ω is a λ-event and the complement of a λ-event λ-event.
Let L_1, L_2 λ-events. We show that $L = (L_1 \cap L_2) \in \mathcal{L}_0$
$L^c \cap L_2 = L^c \cap L_1$, $L^c \cap L_2^c = L_2^c$. When $A \in \mathcal{A}_0$, since L_2 is a λ-event

$$\lambda(A) = \lambda(L_2 \cap A) + \lambda(L_2^c \cap A)$$
$$\lambda(L^c \cap A) = \lambda(L_2 \cap (L^c \cap A)) + \lambda(L_2^c \cap (L^c \cap A)) = \lambda(L_2 \cap L_1^c \cap A) + \lambda(L_2^c \cap A)$$
$$\lambda(L_2 \cap A) = \lambda(L_1 \cap L_2 \cap A) + \lambda(L_1^c \cap L_2 \cap A) \text{ (beacuse } L_1 \text{ is a } \lambda\text{-event)}$$

This implies

$$\begin{aligned} \lambda(A) &= \lambda(L_1 \cap L_2 \cap A) + \lambda(L_1^c \cap L_2 \cap A) + \lambda(L_2^c \cap A) \\ &= \lambda(L \cap A) + \lambda(L^c \cap A) - \lambda(L_2^c \cap A) + \lambda(L_2^c \cap A), \end{aligned}$$

which means that $L = (L_1 \cap L_2)$ is a λ-event. By taking complements we see that $(L_1 \cup L_2)$ is also a λ-event, which proves that $\mathcal{L}_0$ is an algebra.

For $L_1, L_2 \in \mathcal{L}_0$, $L_1 \cap L_2 = \emptyset$, $A \in \mathcal{A}_0$, since $L_1^c \supseteq L_2$,

$$\lambda(A \cap (L_1 \cup L_2)) = \lambda(A \cap (L_1 \cup L_2) \cap L_1) + \lambda(A \cap (L_1 \cup L_2) \cap L_1^c) = \lambda(A \cap L_1) + \lambda(A \cap L_2)$$

Finite additivity follows for $A = \Omega$ $\square$.

Definition 1.1.6. *Let $(\Omega, \mathcal{A})$ be a probability space.*

The map $\lambda : \mathcal{A} \to [0, +\infty]$ is an outer measure if

1. $\lambda(\emptyset) = 0$,
2. $A_1 \subseteq A_2 \quad , A_i \in \mathcal{A}, \ i = 1, 2, \quad \Longrightarrow \quad \lambda(A_1) \leq \lambda(A_2)$
3. *λ is σ-sub-additive, meaning that $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{A}$,*

$$\lambda\left(\bigcup_n A_n\right) \leq \sum_n \lambda(A_n)$$

Example 1.1.3. *Let $\mathcal{Q}$ a collection of probability measures on the probability space $(\Omega, \mathcal{A})$. The map $\lambda : \mathcal{A} \to [0, 1]$ given as*

$$\lambda(A) := \sup_{Q \in \mathcal{Q}} Q(A), \qquad A \in \mathcal{A},$$

is an outer measure (exercise).

Lemma 1.1.5 (Caratheodory)**.** *Let λ be an outer measure on the probability space $(\Omega, \mathcal{A})$, where $\mathcal{A}$ is a σ-algebra. The collection of λ-events*

$$\mathcal{L} = \{L \in \mathcal{A} : L \text{ is a } \lambda\text{-event } \}$$

is a σ-algebra and the map $\lambda : \mathcal{L} \to [0, +\infty]$ is a σ-additive measure.

Proof. Following lemma (1.1.4) it remains to show that for $\{L_n : n \in \mathbb{N}\} \subseteq \mathcal{L}$, $L_i \cap L_j = \emptyset \ \forall i \neq j$, it follows

$$L = \left(\bigcup_{n \in \mathbb{N}} L_n\right) \in \mathcal{L} \text{ ja } \lambda(L) = \sum_{n=1}^{\infty} \lambda(L_n) \ .$$

Since λ is σ-sub-additive, for $A \in \mathcal{A}$

$$\lambda(A) \leq \lambda(A \cap L) + \lambda(A \cap L^c)$$

Let $M_n := \bigcup_{k \leq n} L_k$. By lemma (1.1.4) it follows that $M_n \in \mathcal{L}$ (since $\mathcal{L}$ is an algebra), which means that $\forall A \in \mathcal{A}$

$$\begin{aligned} \lambda(A) &= \lambda(A \cap M_n^c) + \lambda(A \cap M_n) \geq \lambda(A \cap L^c) + \lambda(A \cap M_n) \\ &= \lambda(A \cap L^c) + \sum_{k=1}^{n} \lambda(A \cap L_k) \end{aligned}$$

since λ is σ-sub-additive in $\mathcal{A}$ and finitely additive in $\mathcal{L}$. This holds $\forall n \in \mathbb{N}$, which gives

$$\lambda(A) = \lambda(A \cap L^c) + \sum_{k=1}^{\infty} \lambda(A \cap L_k) \geq \lambda(A \cap L^c) + \lambda(A \cap L)$$

It follows by sub-additivity that $\lambda(A) = \lambda(A \cap L) + \lambda(A \cap L^c)$, which means $L \in \mathcal{L}$. σ-additivity follows by plugging in A=L:

$$\lambda(L) = \sum_{k=1}^{\infty} \lambda(L_k) \qquad \square$$

Proof of Caratheodory extension theorem (1.1.1)

For $\mathcal{G} = 2^\Omega$, we introduce the map

$$\lambda(G) := \inf_{\{F_n\}} \sum_{n \in \mathbb{N}} \lambda_0(F_n) \qquad \forall G \subseteq \Omega$$

where the infimum is taken over the event sequences $\{F_n : n \in \mathbb{N}\} \subseteq \mathcal{A}_0$ with $\bigcup_n F_n \supseteq G$.

a) The map $\lambda : 2^\Omega \to [0, +\infty]$ is an outer measure.

Proof: clearly $\lambda(\emptyset) = 0$ and λ is non-decreasing. Consider the sequence $\{G_n : n \in \mathbb{N}\} \subseteq 2^\Omega$, and an $\varepsilon > 0$. $\forall n \in \mathbb{N}$ there is an event sequence $\{F_{n,k} : k \in \mathbb{N}\} \subseteq \mathcal{A}_0$ such that

$$G_n \subseteq \bigcup_k F_{n,k} \quad \text{ja} \quad \lambda(G_n) \leq \sum_k \lambda_0(F_{n,k}) \leq \lambda(G_n) + \varepsilon 2^{-n}.$$

Let $G = \bigcup_n G_n \subseteq \bigcup_n \bigcup_k F_{n,k}$.S Since the events$\{F_{n,k} : n, k \in \mathbb{N}\}$ cover G

$$\lambda(G) \leq \sum_n \sum_k \lambda_0(F_{n,k}) \leq \sum_{n=1}^{\infty} \lambda(G_n) + \varepsilon$$

and σ-sub-addivity follows ($\varepsilon > 0$ was arbitrary).

It follows from Lemma (1.1.5) that λ is a σ-additive measure on the probability space $(\Omega, \mathcal{L})$, where

$$\mathcal{L} = \{L \subseteq \Omega : \lambda(G) = \lambda(G \cap L) + \lambda(G \cap L^c)\ \forall G \subseteq \Omega\}$$

is the σ-algebra of λ-events.

We show next that **b)** $\lambda(A) = \lambda_0(A)$ when $A \in \mathcal{A}_0$ and **c)**$\mathcal{A}_0 \subseteq \mathcal{L}$. This will imply that λ is an extension of λ_0 to the probability space $(\Omega, \mathcal{L})$ equipped with the σ-algebra $\mathcal{L} \supseteq \sigma(\mathcal{A}_0) = \mathcal{A}$.

b) When $A \in \mathcal{A}_0$, $\lambda(A) \leq \lambda_0(A)$, since obviously $A_0 \subseteq A_0 \in \mathcal{A}_0$. Consider $A \subseteq \bigcup_n A_n$ such that $A_n \in \mathcal{A}_0$. Since $\mathcal{A}_0$ is an algebra, we can assume that the events $\{A_n\}$ are non-intersecting.

Otherwise we could take $A'_1 = A_1$, and by recursion $A'_n = (A_n \setminus A'_{n-1})$. In such a way $\left(\bigcup_n A'_n\right) = \left(\bigcup_n A_n\right) \supseteq A$ and since since $\mathcal{A}_0$ is an algebra, $A'_n \in \mathcal{A}_0$.

$$\lambda_0(A) = \lambda_0\left(\bigcup_n (A \cap A_n)\right) \overset{(*)}{=} \sum_n \lambda_0(A \cap A_n) \leq \sum_n \lambda_0(A_n)$$

where the equality $(*)$ is implies by the assumption of the extension theorem, that $\lambda_0 : \mathcal{A}_0 \to [0, +\infty]$ is σ-additive on the algebra $\mathcal{A}_0$. By taking the infimum from the right hand side it follows that $\lambda_0(A) \leq \lambda(A)$.

c) Let $A \in \mathcal{A}_0$ and $G \subseteq \Omega$.

$\forall \varepsilon > 0$, there is an event sequence $\{A_n\} \subseteq \mathcal{A}_0$ such that $G \subseteq \bigcup_n A_n$ and

$$\lambda(G) + \varepsilon \geq \sum_n \lambda_0(A_n).$$

Since $A \in \mathcal{A}_0$, by the definition of λ

$$\sum_n \lambda_0(A_n) = \sum_n \lambda_0(A_n \cap A) + \sum_n \lambda_0(A_n \cap A^c) \geq \lambda(G \cap A) + \lambda(G \cap A^c)$$

because the events $\{(A_n \cap A)\}_{n\in\mathbb{N}}$ are covering $(G \cap A)$ and the events $\{(A_n \cap A^c)\}_{n\in\mathbb{N}}$ are covering $(G \cap A^C)$. Since $\varepsilon > 0$ was arbitrary,

$$\lambda(G) \geq \lambda(G \cap A) + \lambda(G \cap A^c) \tag{1.1.1}$$

On the other hand $\lambda : 2^\Omega \to [0, +\infty]$ is an outer measure, it is sub-additive, which implies equality in (1.1.1). This shows that $A \in \mathcal{L}$ □

Example: probability measure on $\mathbb{R}$ As in the example 1.1.2, let $\Omega = \mathbb{R}$ with $\mathcal{F} = \mathcal{B}(\mathbb{R})$, and consider a function $F : \mathbb{R} \to [0, 1]$ with the properties

1. F is non-decreasing
2. F is right continuous $F(t+) := \lim_{u\downarrow t} F(u) = F(t)\ \forall t \in \mathbb{R}$.
3. $F(+\infty) := \lim_{t\uparrow+\infty} F(t) = 1$, $F(-\infty) := \lim_{t\downarrow-\infty} F(t) = 0$.

Then there exists an unique probability measure $P : \mathcal{B}(\mathbb{R}) \to [0, 1]$ defined on the Borel sets such that $P((a, b]) = (F(b) - F(a))$, $a < b \in \mathbb{R}$.

Proof Uniqueness follows from Lemma (1.1.2), since $\{(a, b] : a < b \in \mathbb{R}\}$ is a π-class. Consider the collection $\mathcal{A}_0$ of events with representation

$$A = (c, \infty) \cup \bigcup_{k=1}^{m} (a_k, b_k]$$

for some $m \in \mathbb{N}, a_k \leq b_k \in \mathbb{R}$, $c \in \mathbb{R} \cup \{+\infty\}$, with $(+\infty, +\infty) = \emptyset$.

$\mathcal{A}_0$ is an algebra which generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$.

Define the map $P^{(0)} : \mathcal{A}_0 \to [0, 1]$

$$P^{(0)}(A) = \sum_{k=1}^{m} \big(F(b_k) - F(a_k)\big)\ .$$

Note that for $A \in \mathcal{A}_0$ $P^{(0)}(A)$ dows not depend on the representation of A, and the map is finitely additive on the algebra $\mathcal{A}_0$.

In order to apply Charatheodory Theorem to show that P_0 extends to a probability measure P defined on the Borel σ-algebra, we we have to show that $P^{(0)}$ is σ-additive on the algebra $\mathcal{A}_0$.

Proof by contradiction assume that this is not the case and there exists a non-increasing event sequence $A_n \in \mathcal{A}_0$, such that for some $\varepsilon > 0$. $A_n \supseteq A_{n+1}$, $P^{(0)}(A_n) \geq 4\varepsilon\ \forall n$. To check σ-additivity we need to show that osoitamme

$$\bigcap_n A_n \neq \emptyset$$

Without loss of generality we can assume that the events A_n are bounded with representation

$$A_n = \bigcup_{k=1}^{m_n} (a_k^n, b_k^n]$$

where

$$-\infty < a_k^n \leq b_k^n \leq a_{k+1}^n \leq b_{k+1}^n < +\infty \ ,$$

and let $\varepsilon > 0$. From assumption (3) it follows that $\exists z \geq 0$ such that $F(-z) < \varepsilon$ and $(1 - F(z)) < \varepsilon$.

Otherwise there would be $\varepsilon > 0$ such that $\forall z \geq 0\ F(-z) > \varepsilon$ or $F(z) < 1 - \varepsilon$, which would contradict the assumptions $F(-\infty) = 0$ and $F(+\infty) = 1$.

Since F is right continuous, there exist $x_k^n \in (a_k^n, b_k^n]$ such that

$$F(x_k^n) \leq F(a_k^n) + \varepsilon 2^{-(k+n)} \ .$$

Let

$$B'_n = \left(\bigcup_{k=1}^{m_n} (x_k^n, b_k^n] \cap [-z, z] \right) \subseteq A_n, \quad B_n = \left(\bigcap_{l \leq n} B'_l \right) \subseteq A_n$$

where $B_n \in \mathcal{A}_0$ and $B_n \supseteq B_{n+1}\ \forall n$. It follows that

$$A_n \setminus B_n = \bigcup_{l \leq n} (A_n \setminus B'_l) \subseteq \bigcup_{l \leq n} (A_l \setminus B'_l)$$

Since P_0 is finitely additive on $\mathcal{A}_0$: ,

$$P^{(0)}(A_n \setminus B_n) \leq P^{(0)}((-z,z]^c) + \sum_{l=1}^{n} P^{(0)}\big((A_l \setminus B_l') \cap (-z,z]\big)$$

$$\leq P^{(0)}((-z,z]^c) + \sum_{l=1}^{n}\sum_{k=1}^{m_l} P^{(0)}\big((a_k^l, x_k^l]\big)$$

$$= F(-z) + (1 - F(z)) + \sum_{l=1}^{n}\sum_{k=1}^{m_l}\big\{F(x_k^l) - F(a_k^l)\big\} \leq 2\varepsilon + \varepsilon\sum_{l=1}^{n} 2^{-l}\sum_{k=1}^{m_l} 2^{-k} \leq 3\varepsilon,$$

and for the geometric series we have $\sum\limits_{n=1}^{\infty} 2^{-n} = 1$. Therefore

$$P^{(0)}(B_n) = P^{(0)}(A_n) - P^{(0)}(A_n \setminus B_n) \geq (4-3)\varepsilon \quad \forall n$$

which implies $B_n \neq \emptyset$. Denote by $\bar{B}_n$ the closure of B_n, as the smallest closed set containing B_n, equivalently

$$\bar{B}_n = \bigcap_{B_n \subseteq C \text{ closed}} C$$

which is closed since the arbitrary intersection of closed sets is closed. Since $B_n \supseteq B_{n+1} \neq \emptyset$, it follows that $\bar{B}_n \supseteq \bar{B}_{n+1} \neq \emptyset$. This implies

$$\bigcap_n A_n \supseteq \bigcap_n \bar{B}_n \neq \emptyset$$

otherwise the collection $\{(\bar{B}_n)^c = (\mathbb{R} \setminus \bar{B}_n) : n \in \mathbb{N}\}$ would be an open cover of the compact interval $[-z,z]$ without finite subcovers:

$$[-z,z] \subseteq \bigcup_{n\in\mathbb{N}} (\bar{B}_n)^c$$

where $[-z,z] \supseteq \bar{B}_n \supseteq \bar{B}_{n+1} \neq \emptyset, \quad \forall n \in \mathbb{N}$ □

Example 1.1.4. *When*

$$F(x) = \begin{cases} 0 & , -\infty < x \leq 0 \\ x & , 0 < x \leq 1 \\ 1 & , 1 < x < \infty \end{cases}$$

P is the uniform probability on $[0,1]$, such that the probability an interval coincides with its length

$$\mathbb{P}_0((a,b]) = F(b) - F(a) = \min(b,1) - \max(a,0), \quad a < b.$$

To construct the Lebesgue measure we first define $\mathbb{P}_z$ on $\Omega_z = (z, z+1]$ $\forall z \in \mathbb{Z}$ as

$$\mathbb{P}_z(B) = \mathbb{P}_z(B \cap \Omega_z) = \mathbb{P}_0(B - z) \quad \forall B \in \mathcal{B}(\mathbb{R})$$

and then sum over $z \in \mathbb{Z}$ to obtain the Lebesgue measure on the whole $\mathbb{R}$

$$\lambda(B) := \sum_{z\in\mathbb{Z}} \mathbb{P}_0(B - z) = \sum_{z\in\mathbb{Z}} \mathbb{P}_z(B) = \sum_{z\in\mathbb{Z}} \lambda(B \cap \Omega_z), \quad A \in \mathcal{B}(\mathbb{R})$$

satisfying $\lambda((a,b]) = (b-a)$, $a < b$. The Lebesgue measure is not a probability but it is σ-finite and shift-invariant, meaning that

$$\lambda(A + x) = \lambda(A),\ x \in \mathbb{R},\ A \in \mathcal{B}(\mathbb{R})\ .$$

Proof for the shift-invariance *Let*

$$\mathcal{D} = \big\{ B \in \mathcal{B}(\mathbb{R}) : \lambda(B) = \lambda(x + B) \quad \forall x \in \mathbb{R} \big\}$$

Note that for any interval $(a,b]$ we have

$$\begin{aligned}
\lambda((a,b]) &= \sum_{z\in\mathbb{Z}} \mathbb{P}_z((a,b]) = \sum_{z\in\mathbb{Z}} \mathbb{P}_0((a-z, b-z]) \\
&= \sum_{z\in\mathbb{Z}} \mathbb{P}_z((a,b] \cap (z, z+1]) = \sum_{z\in\mathbb{Z}} \mathbb{P}_z((a \vee z, b \wedge (z+1)]) \\
&= (b - \lfloor b \rfloor) + \sum_{z=\lceil a \rceil}^{\lfloor b \rfloor - 1} ((z+1) - z) + (\lceil a \rceil - a) \\
&= (b - \lfloor b \rfloor) + (\lfloor b \rfloor - \lceil a \rceil) + (\lceil a \rceil - a) = b - a \\
&= (b + x) - (a + x) = \lambda((a+x, b+x]) \quad \forall x \in \mathbb{R}
\end{aligned}$$

where $x \wedge y = \min\{x,y\}$, $x \vee y = \max\{x,y\}$, $\lfloor x \rfloor = \sup\{z \in \mathbb{Z} : z \le x\}$, $\lceil x \rceil = \inf\{z \in \mathbb{Z} : z \ge x\}$. This shows that any interval $(a,b] \in \mathcal{D}$, and since the *intervals form a π-class, it is enough to show that $\mathcal{D}$ is a d-class, which implies the claim $\mathcal{D} = \mathcal{B}(\mathbb{R})$.*

Since $(\mathbb{R} + x) = \mathbb{R}$, it is clear that $\mathbb{R} \in \mathcal{D}$.

When $B \supseteq A$ and $A, B \in \mathcal{D}$, since$(A+x) \setminus (B+x) = (A \setminus B) + x$, it follows that

$$\lambda\big((B \setminus A) + x\big) = \lambda\big((B + x) \setminus (A + x)\big)$$
$$= \lambda(B + x) - \lambda(A + x) = \lambda(B) - \lambda(A) = \lambda(B \setminus A) \; .$$

Let $A_n \subseteq A_{n+1}$, where $A_n \in \mathcal{D}\; \forall n$, and $A = \bigcup_{n \in \mathbb{N}} A_n$. Note that

$$x + A = \bigcup_{n \in \mathbb{N}} (A_n + x)$$

and $(x + A_n) \uparrow (x + A)$. Since λ is σ-additive, this implies

$$\lambda(x + A) = \lambda\left(\bigcup_{n \in \mathbb{N}} (A_n + x)\right) = \lim_n \lambda(A_n + x) = \lim_n \lambda(A_n) = \lambda(A),$$

which means that $A \in \mathcal{D}$, and $\mathcal{D}$ is a d-class.

1.2 Application: product σ-algebra and product-probability

Let $(\Omega', \mathcal{F}', P')$ and $(\Omega'', \mathcal{F}'', P'')$ be probability spaces, and $\Omega = \Omega' \times \Omega''$.

We set on the product space Ω the product-σ-algebra

$$\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\big(A \times B : A \in \mathcal{F}', B \in \mathcal{F}'',\big)$$

which is generated by the product of event pairs.

Let

$$\mathcal{A}_0 = \left\{ C \subseteq \Omega : C = \bigcup_{k=1}^{m} (A_k \times B_k) \text{ where } A_k \in \mathcal{F}', B_k \in \mathcal{F}' \right\}$$

Clearly $\mathcal{A}_0$ is an algebra and $\sigma(\mathcal{A}_0) = \mathcal{F}$.

We define the additive map $\mathbb{P}^{(0)} : \mathcal{A}_0 \to [0,1]$: for

$$C = \bigcup_{k=1}^{m} (A_k \times B_k)$$

with $(A_k \times B_k) \cap (A_l \times B_l) = \emptyset$ for $l \neq k$, set

$$\mathbb{P}^{(0)}(C) = \sum_{k=1}^{m} P'(A_k) P''(B_k),$$

which does not depend on the representation of C (this is left to you to prove as an exercise). When we show that the map $\mathbb{P}^{(0)} : \mathcal{A}_0 \to [0,1]$ is also σ-additive on $\mathcal{A}_0$, it follows by Caratheodory extension theorem that the σ-additive probabilty $\mathbb{P}^{(0)}$ has an unique extension $\mathbb{P} = (P_1 \otimes P_2) : \mathcal{F}_1 \otimes \mathcal{F}_2 \to [0,1]$, such that $\mathbb{P}^{(0)}(A) = \mathbb{P}(A)$ when $A \in \mathcal{A}_0$. $(P_1 \otimes P_2)$ is the product measure on the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$.

This construction extends directly by induction to finite product spaces $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$, and also when the measures are σ-finite (exercise).

The construction of infinite dimensional product measure is also possible but it requires an additional separability assumption on Ω. Such infinite dimensional construction is known as Kolmogorov extension theorem 3.

Proof To show that $\mathbb{P}^{(0)}$ is σ-additive on the algebrassa $\mathcal{A}_0$,

we show that for an event sequence $\{C_n : n \in \mathbb{N}\} \subseteq \mathcal{A}_0$ such that $C_n \downarrow \emptyset$, it follows that $\mathbb{P}^{(0)}(C_n) \downarrow 0$. Let

$$C_n = \bigcup_{k=1}^{m_n} (A_{nk} \times B_{nk}) \ \in \mathcal{A} \tag{1.2.1}$$

with $A_{nk} \in \mathcal{F}_1$ and $B_{nk} \in \mathcal{F}_2$, such that the product sets $(A_{nk} \times B_{nk})$ $k = 1, \dots, m_n$ are disjoint (we can always find such representation, which is not unique).

Note that

- even if the representation (1.2.1) of the event $C_n \in \mathcal{A}_0$ is not unique, its product measure

$$\mathbb{P}^{(0)}(C_n) = \sum_{k=1}^{m_n} P_1(A_{nk}) P_2(B_{nk})$$

does not depend on the representation.

- By taking intersections, we can always find a representation such that $A_{nk} \cap A_{nh} = \emptyset$, $\forall h \neq k$.

and for every $\forall n, 1 \leq k \leq m_n$ there is an ℓ such that $A_{n-1,\ell} \supseteq A_{n,k}$ and $B_{n-1,\ell} \supseteq B_{n,k}$.

Consider the function $f_n : \Omega' \to [0,1]$

$$f_n(\omega') = \sum_{k=1}^{m_n} P_2(B_{nk}) \mathbf{1}_{A_{nk}}(\omega')$$

with A_{nk} disjoint. We show first that $\forall \omega' \in \Omega'$, $f_n(\omega') \downarrow 0$ as $n \uparrow \infty$. Otherwise, in case that $f_n(\omega') \not\to 0$, there would be some non-increasing event subsequences $(A_{n_\ell,k_\ell} : \ell \in \mathbb{N}) \subseteq \mathcal{F}_1$ and $(B_{n_\ell,k_\ell} : \ell \in \mathbb{N}) \subseteq \mathcal{F}_2$ such that

$$n_\ell \uparrow \infty, \quad A_{n_\ell,k_\ell} \supseteq A_{n_{\ell+1},k_{\ell+1}} \quad \text{and} \quad B_{n_\ell,k_\ell} \supseteq B_{n_{\ell+1},k_{\ell+1}} \forall \ell \in \mathbb{N},$$

and, for some $\eta > 0$,

$$\omega' \in \bigcap_{\ell \in \mathbb{N}} A_{n_\ell,k_\ell} \neq \emptyset \quad \text{and} \quad P_2(B_{n_\ell,k_\ell}) > \eta > 0$$

and since P_2 is σ-additive this implies

$$\bigcap_{n \in \mathbb{N}} B_{n_\ell,k_\ell} \neq \emptyset$$

which implies

$$\bigcap_{n \in \mathbb{N}} C_n \supseteq \big(\bigcap_{\ell \in \mathbb{N}} A_{n_\ell,k_\ell} \times B_{n_\ell,k_\ell} \big) \neq \emptyset$$

in contradiction with $\mathbf{1}_{C_n}(\omega) \downarrow 0$.

This implies that $\forall \varepsilon > 0$

$$D_{n,\varepsilon} := \big\{ \omega' \in \Omega_1 : \; f_n(\omega') > \varepsilon \big\} \downarrow \emptyset \, ,$$

where D_n is an union of some of the A_{nk} subsets. Since P_1 is σ-additive, and $D_{n,\varepsilon} \downarrow \emptyset$, it follows that $P_1(D_{n,\varepsilon}) \downarrow 0$ as $n \uparrow \infty$.

This implies that $\forall \varepsilon > 0$,

$$\begin{aligned}\mathbb{P}^{(0)}(C_n) &= \sum_{k=1}^{m_n} P_1(A_{nk}) \times P_2(B_{nk}) \\ &\le \big(P_1(D_{n,\varepsilon}) P_2(\Omega_2) + P_1(\Omega_1)\varepsilon \big) \downarrow \varepsilon, \quad \text{as } n \uparrow \infty ,\end{aligned}$$

which means that $\mathbb{P}^{(0)}(C_n) \downarrow 0 \quad \square$

Example: general probability distribution on $(\mathbb{R}^d)$ We have shown that

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d} := \underbrace{\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})}_{d\text{-times}}$$

For $s = (s_1, \dots, s_d), t = (t_1, \dots, t_d) \in \big(\mathbb{R} \cup \{ \pm\infty\}\big)^d$, denote

$$s \le t \iff s_i \le t_i,\ i = 1, \dots, d, \quad s < t \iff s_i < t_i,\ i = 1, \dots, d,$$
$$(s, t] = \big\{ x \in \mathbb{R}^d : s < x \le t \big\} = (s_1, x_1] \times \cdots \times (s_d, t_d]$$

Note that for $d > 1$, the ordering relation $\le$ is not total, not all pairs $s, t \in \mathbb{R}^d$ are comparable.

We have also seen that $\sigma\big((-\infty, q] : q \in \mathbb{Q}^d\big) = \mathcal{B}(\mathbb{R}^d)$.

Let P a probability on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Define the multivariate cumulative distribution function as

$$F(t) = \mathbb{P}\big((-\infty, t]\big) = P\big(\{ x \in \mathbb{R}^d : x_i \le t_i,\ i = 1, \dots, d\}\big),\ t \in \mathbb{R}^d$$

It follows that

1. $F(t)$ is non-decreasing with respect to the partial order relation $\le$,
2. $F(t)$ is right-continuous, which means $F(t_n) \downarrow F(t)$ as $t_n \downarrow t$,
3. $F(-\infty) := \lim_{t \downarrow -\infty} F(t) = 0$ ja $F(+\infty) := \lim_{t \uparrow +\infty} F(t) = 1$.

In the other direction, when $F : \mathbb{R}^d \to [0, 1]$ satisfies $(1), (2), (3)$, there is an unique probability $\mathbb{P}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $F(t) = \mathbb{P}((-\infty, t])\ \forall t \in \mathbb{R}^d$. The proof of 1.1 in the one-dimensional case generalized directly. If

$\mathbb{P}$ is a product measure, the multivariate cumulative distribution function factorizes as

$$F(t_1, \ldots, t_d) = F_1(t_1) \times F_2(t_2) \cdots \times F_d(t_d)$$

where

$$F_i(t_i) = F(\infty, \ldots, \infty, t_i, \infty, \ldots, \infty)$$

corresponding to probability measure $\mathbb{P}_i$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, with

$$\mathbb{P}_i(B_i) = \mathbb{P}(\mathbb{R} \times \cdots \times \mathbb{R} \times B_i \times \mathbb{R} \times \cdots \times \mathbb{R}),$$

and

$$\mathbb{P}(B_1 \times \cdots \times B_d) = P_1(B_1) \times \cdots \times P_d(B_d)$$

for $B_i \in \mathcal{B}(\mathbb{R})$, $1 \leq i \leq d$.

Chapter 2

Random Variables

Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ a map between probability spaces. We say X is a *measurable* function, and in probabilistic language, a *random variable* , when $\forall B \in \mathcal{E}$ which is measurable in the space where the random variable takes values, the inverse-image (or counter-image)

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F}$$

is measurable in the space of arguments.

Typically we will consider $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ generated by the open sets. We denote by $X \in L^0(\Omega, \mathcal{F})$ the space of $\mathbb{R}$-valued random variables, and write shortly $X \in \mathcal{F}$ when it is clear from the context that X is function of ω and not a subset of Ω.

Remark 2.0.1. *Note the analogy with the definition of a continuous function between topological spaces: let $(S_i, \mathcal{S}_i)$, $i = 1, 2$ topological spaces, where $\mathcal{S}_i$ are the collection of open sets in the respective spaces. We say that $f : (S_1, \mathcal{S}_1) \to (S_2, \mathcal{S}_2)$ is continuous if and only if for every $U \in \mathcal{S}_2$ open in the space where the function takes values, the counter-image*

$$f^{-1}(U) = \{x \in S_1 : f(x) \in U\} \in \mathcal{S}_1,$$

is open in the domain space S_1.

Lemma 2.0.1. *If $X : \Omega \mapsto E$ is a map, (not necessarily measurable)*

$$X^{-1}\Big(\bigcup_{i\in\mathcal{I}} A_i\Big) = \bigcup_{i\in\mathcal{I}} X^{-1}(A_i), \qquad X^{-1}(A^c) = \Big(X^{-1}(A)\Big)^c$$

where $A, A_i \subseteq E$ and $\mathcal{I}$ is an arbitrary index set (possibly uncountable).

Lemma 2.0.2. *Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ and $Y : (E, \mathcal{E}) \to (G, \mathcal{G})$ be measurable maps between probability spaces. Then the composition $(Y \circ X)(\omega) := Y(X(\omega))$ is measurable between $(\Omega, \mathcal{F})$ and $(G, \mathcal{G})$.*

Proof. For $B \in \mathcal{G}$,

$$(Y \circ X)^{-1}(G) = \{\omega : Y(X(\omega)) \in G\} = \{\omega : X(\omega)) \in Y^{-1}(G)\} \in \mathcal{F}$$

since $Y^{-1}(G) \in \mathcal{E}$ and X is $(\mathcal{F}, \mathcal{E})$-measurable.

Definition 2.0.1. *Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ a random variable. We define*

$$\sigma(X) = \big\{ X^{-1}(B) : B \in \mathcal{E} \big\} \subseteq \mathcal{F}$$

as the σ-algebra generated by the random variable X. It is the smallest σ-algebra on Ω such that X is measurable.

Lemma 2.0.3. *Let $f : E \to H$, a continuous function between topological spaces, and let $\mathcal{B}(E)$ and $\mathcal{B}(H)$ the corresponding Borel σ-algebrae generated by the open sets. Then*

- *the map $f : (E, \mathcal{B}(E)) \to (H, \mathcal{B}(H))$ is measurable.*
- *$\sigma(f) := f^{-1}(\mathcal{B}(E)) = \sigma\{f^{-1}(U) : \ U \subseteq H \text{ open }\}$, is the smallest σ-algebra on E which makes f on Borel-measurable.*

Prood For any open set $U \subseteq H$, since f is continuous it follows that $f^{-1}(U) \subseteq E$ is open in E. This implies

$$\sigma\{f^{-1}(U) : \ U \subseteq H \text{ open }\} \subseteq \mathcal{B}(E)$$

The collections of events

$$\{U \subseteq H \text{ open }\} \text{ and } \{f^{-1}(U) : \ U \subseteq H \text{ open }\}$$

are π-classes, since the finite intersection of open sets is open.

When $\mathcal{D}$ is a d-class in H,

$$f^{-1}(\mathcal{D}) = \left\{ f^{-1}(D) : D \in \mathcal{D} \right\}$$

is a d-class in E:

$E = f^{-1}(H) \in f^{-1}(\mathcal{D})$ because $H \in \mathcal{D}$.

For $A_1 = f^{-1}(B_1) \subseteq f^{-1}(B_2)$ where $B_i \in \mathcal{D}$, it follows that $B_1 \subseteq B_2$ and $B_2 \setminus B_1 \in \mathcal{D}$

which implies $f^{-1}(B_2) \setminus f^{-1}(B_1) = f^{-1}(B_2 \setminus B_1) \in f^{-1}(\mathcal{D})$.

Whem $A_n = f^{-1}(B_n) \in f^{-1}(\mathcal{D})$ and $A_n \uparrow A = \bigcup_{n\in\mathbb{N}} A_n$,

it follows that $A = \bigcup_{n\in\mathbb{N}} f^{-1}(B_n) = f^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \in f^{-1}(\mathcal{D})$.

Since $\mathcal{B}(H)$ is the smallest d-class which contains the π-class $\{ U \subseteq H \text{ open} \}$,

$f^{-1}(\mathcal{B}(H))$ is the smallest d-class which contains the π-class $\{f^{-1}(U) : \ U \subseteq H \text{ open} \}$, and by lemma 1.1.3 it follows that $f^{-1}(\mathcal{B}(H)) = \sigma\{f^{-1}(U) : \ U \subseteq H \text{ open} \}$ □

Corollary 2.0.1. *Let $X_i : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $i = 1, \dots, d$ $\mathbb{R}$-values random variables. Then the vector-valued map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined as $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$ is $\mathbb{R}^d$-valued random variable (which means a Borel measurable map).*

Proof. For $B_i \subseteq \mathcal{B}(\mathbb{R})$, $i = 1, \dots, d$,

$$\{\omega : X(\omega) \in B_1 \times \cdots \times B_d\} = \left(\bigcap_{i=1}^{d} \{\omega : X_i(\omega) \in B_i\}\right) \in \mathcal{F}$$

and the claim follows since $\sigma\big(B_1 \times \cdots \times B_d : B_i \in \mathcal{B}(\mathbb{R})\big) = \mathcal{B}(\mathbb{R}^d)$.

More in details, let

$$X^{-1}(B) = \left\{ \omega : X_1(\omega), \dots, X_d(\omega) \in B \right\}, \quad B \in \mathcal{B}(\mathbb{R}^d),$$
$$\mathcal{D} = \{B \in \mathcal{B}(\mathbb{R}^d) : X^{-1}(B) \in \mathcal{F}\} \supseteq \mathcal{I} = \left\{ B_1 \times B_2 \times \cdots \times B_d : \ B_i \in \mathcal{B}(\mathbb{R}) \right\}$$

where $\mathcal{D}$ is a σ-algebra and $\mathcal{I}$ is a π-class. On the other habd $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})^{\otimes d} = \mathcal{B}(\mathbb{R}^d)$, which implies that $X^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{B}(\mathbb{R}^d)$. □

Corollary 2.0.2. *Let* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *a random vector and* $f : X(\Omega) \subseteq \mathbb{R}^d \to \mathbb{R}$ *a function which is continuous on the image of* X.

Then the map $\omega \mapsto (f \circ X)(\omega) = f(X(\omega))$ *is a random variable.*

In particular, if (X, Y) *are random variables, then also* $(X + Y)$ *and* (XY) *are random variables.*

When $Y(\omega) \neq 0 \quad \forall \omega$, *then also* (X/Y) *is a random variable.*

Lemma 2.0.4. *For any sequence of random variables* $\{X_n : n \in \mathbb{N}\}$ *on a probability space* $(\Omega, \mathcal{F})$, *the map*

$$X^*(\omega) := \sup_{n \in \mathbb{N}} X_n(\omega)$$

is a random variable.

Proof $\mathcal{D} = \{B \subseteq \mathcal{B}(\mathbb{R}) : (X^*)^{-1}(B) \in \mathcal{F}\}$ is a σ-algebra. Since $\forall t \in \mathbb{R}$

$$(X^*)^{-1}((-\infty, t]) := \{\omega : X^*(\omega) \leq t\} = \bigcap_{n \in \mathbb{N}} \{\omega : X_n(\omega) \leq t\} \in \mathcal{F},$$

it follows that $(-\infty, t] \in \mathcal{D}$. Since these intervals form a π-class which generated the Borel σ-algebra, it follows that $\mathcal{D} = \mathcal{B}(\mathbb{R})$.

Corollary 2.0.3. *Also* $\liminf_n X_n(\omega)$ *and* $\limsup_n X_n(\omega)$ *are random variables (exercise).*

The next Theorem is the analogous of Dynkin lemma (1.1.3) for random variables:

Theorem 2.0.1. *(Monotone Class Theorem) Let* $\mathcal{C}$ *be a collection of* **bounded** *functions* $X : \Omega \to \mathbb{R}$. *We say that* $\mathcal{C}$ *is a* **monotone class** *when*

1. $\mathcal{C}$ *is a vector space.*

2. *The constant function* $\mathbf{1} \in \mathcal{C}$.

3. $\mathcal{C}$ *is closed with respect to monotone limits in the following sense:*

   For all sequences $\{X_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$ *such that*

   $0 \leq X_n(\omega) \leq X_{n+1}(\omega) \quad \forall \omega \in \Omega, n \in \mathbb{N},$ *and the monotone limit*
   $X(\omega) := \lim_{n \uparrow \infty} X_n(\omega)$ *is a* **bounded** *function,*

   it follows that $X \in \mathcal{C}$.

When $\mathcal{C} \supseteq \{ \mathbf{1}_A(\omega) : A \in \mathcal{I} \}$ where $\mathcal{I} \subseteq 2^\Omega$ is a π-class (closed with respect finite intersections), it follows that

$$\mathcal{C} \supseteq \left\{ \textbf{ bounded } \textit{and } \big(\Omega, \sigma(\mathcal{I})\big) \longrightarrow \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big) \textit{ measurable functions } \right\}$$

Proof. Let $\mathcal{D} = \big\{ A \subseteq \Omega : \mathbf{1}_A \in \mathcal{C} \big\}$.

It follows by the definition of monotone class that $\mathcal{D}$ is a Dynkin d-class which contains the π-class $\mathcal{I}$. By Dynkin lemma $(1.1.3)$ it follows that $\mathcal{D} \supseteq \sigma(\mathcal{I})$.

Let $X(\omega)$ $\sigma(\mathcal{I})$-measurable random variable such that

$$0 \le X(\omega) \le K < \infty \quad \forall \omega \in \Omega \,.$$

Define the sequence of maps

$$X^{(n)}(\omega) := \sum_{\ell=0}^{n-1} \frac{K\ell}{n} \, \mathbf{1}\Big\{ X(\omega) \in \Big(\frac{K\ell}{n}, \frac{K(\ell+1)}{n} \Big] \Big\} \,, \quad n \in \mathbb{N}.$$

Since $\mathcal{C}$ is a vector space, it follows that $\{X^{(n)}\} \subseteq \mathcal{C}$, and since

$$0 \le X^{(n)}(\omega) \uparrow X(\omega) \qquad \forall \omega \in \Omega$$

where by assumption $X(\omega)$ is bounded, it follows that $X(\omega) \in \mathcal{C}$ □

The next result is an example on the use of the Monotone Class Theorem.

Theorem 2.0.2. *Let kuvaus $Y : (\Omega, \mathcal{F}, P) \longrightarrow (S, \mathcal{S})$ S-arvoinen satunnaismuuttuja, esimerkiksi $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Let

$$\sigma(Y) := \big\{ Y^{-1}(A) : A \in \mathcal{S} \big\}.$$

- *$\sigma(Y)$ is the smallest σ-algebra on Ω which makes $Y(\omega)$ measurable.*
- *$\sigma(Y)$ is called the σ-algebra generated by the random variable Y.*

- *Any random variable $X : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $\sigma(Y)$-measurable if and only if there exist a Borel measurable map $f : (S, \mathcal{S}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $X(\omega) = f(Y(\omega))$.*

Proof We have already shown the implication $\Longleftarrow$ (lemma 2.0.2).
Let

$$\mathcal{C} = \{ f(Y(\omega)) : f \text{ is Borel-measurable and } |f(Y(\omega))| \text{ is bounded} \}$$

Clearly $\mathcal{C}$ is a vector space which contains the constants. Let $\{ f_n \}_{n\in\mathbb{N}}$ be a sequence of Borel-measurable functions such that

$$0 \le f_n(Y(\omega)) \le f_{n+1}(Y(\omega)) \le K < \infty \quad \forall n \in \mathbb{N}, \omega \in \Omega ,$$

and $f(y) := \limsup_n f_n(y)\ \forall y \in S$. It follows that $f : (S, \mathcal{S}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable, and

$$0 \le f_n(Y(\omega)) \uparrow f(Y(\omega)) \le K < \infty \quad \text{as } n \uparrow \infty .$$

Therefore $\mathcal{C}$ is a monotone. Since $\mathbf{1}_A(Y(\omega)) \in \mathcal{C}\ \forall A \in \mathcal{S}$, and

$$\sigma(Y) = \{ Y^{-1}(A) : A \in \mathcal{S} \}$$

it follows by the monotone class Theorem 2.0.1 that $\mathcal{C}$ contains all $\sigma(Y)$-measurable bounded $\mathbb{R}$-valued random variables.

More in general, if $X(\omega)$ is a $\sigma(Y)$-measurable $\mathbb{R}$-valued random variable which is not bounded, since $\arctan(x)$ is a bicontinuous bijection between $\mathbb{R}$ and the open interval $(-\pi/2, \pi/2)$, the random variable arctan$(X(\omega))$ on rajoitettu ja $\sigma(Y)$-measurable.

This implies that arctan$(X(\omega)) = f(Y(\omega))$ for some Borel measurable function $f(y)$, and $X(\omega) = \tan(f(Y(\omega))$ where $\tan(f(y))$ is also Borel measurable □

2.1 Probability distribution of a random variable

Definition 2.1.1. *Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a random variable, for example it could be real valued with $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and $\mathbb{P}$ a probability measure on* *the probability space* $(\Omega, \mathcal{F})$.

The push-forward (or image) measure, also referred as probability distribution of the random variable X, *is the probability*

$$P_X(B) := \big(\mathbb{P} \circ X^{-1}\big)(B) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big), \quad \forall B \in \mathcal{S}$$

Example 2.1.1. *Show that* P_X *is a probability on the image space*$(S, \mathcal{S})$.

Chapter 3

Probability on infinite dimensional product spaces and Kolmogorov extension Theorem

In the following chapters we will work with a probability space $(\Omega, \mathcal{F}, P)$ which is assumed to be rich enough to admit sequences of P-independent random variables $(X_k(\omega) : k \in \mathbb{N})$ defined on it. The construction of such spaces is a special case of Kolmogorov extension theorem.

In fact, while in for this probability theory course we will just need to construct an infinite product measure on the space of sequence, we prove at the same time a more general result, which works for any product spaces over an arbitrary index set, and it is not restricted to product measures.

The proof is based on Caratheodory extension theorem, using a compactness argument to prove σ-additivity.

Definition 3.0.2. *On a given probability space $(\Omega, \mathcal{F})$, a stochastic process is a collection of random variables $(X_t : t \in \mathbb{T})$ with index set $\mathbb{T}$, taking values $X_t(\omega) \in \mathbb{R}^d$.*

For example $\mathbb{T} = \mathbb{N}, \mathbb{R}^+, \mathbb{Q}^+, \mathbb{Z}, \mathbb{R}, \mathbb{Q}$ can be interpreted as time parameters, which could be discrete or continuous. But we can also consider $\mathbb{T} = \mathbb{R}^m$, in such case the stochastic process is also called a random field.

In the follow up we consider the case with $d = 1$ with $X_t(\omega) \in \mathbb{R}$.

Definition 3.0.3. *We say that a family of finite dimensional probability distributions (fid) on $\mathbb{R}$ with indexes in $\mathbb{T}$*

$$\Big(P_{t_1,\dots,t_n} : \mathcal{B}(\mathbb{R}^n) \to [0,1], \quad n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{T} \Big)$$

consistent *, if*

1.
$$P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = P_{t_{\pi(1)},\dots t_{\pi(n)}}\big(A_{t_{\pi(1)}} \cdots \times A_{t_{\pi(n)}}\big)$$
$$\forall n \in \mathbb{N}, A_1, \dots A_n \in \mathcal{B}(\mathbb{R}), t_1, \dots, t_n \in \mathbb{T}, \quad \forall \text{ permutations } \pi$$

2.
$$P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = P_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R})$$

We remark the set of indexes $\{\mathbb{T}\}$ in the next section is arbitrary.

Theorem 3.0.1. *(Daniell-Kolmogorov,1933) Let*

$$\Big(P_{\mathbf{t}} : \mathbf{t} \in \bigcup_{n=1}^{\infty} \mathbb{T}^n \Big)$$

a consistent family of finite dimensional distributions on $\mathbb{R}$, with indexes in an infinite but otherwise arbitrary set $\mathbb{T}$.

There exists an unique probability measure $\mathbb{P}$ on the infinite product space $\Omega = \mathbb{R}^{\mathbb{T}}$ equipped with the σ-algebra generated by the product-topology, which coincides with the σ-algebra $\sigma(\mathcal{C})$ generated by the cylinders, such that

$\forall n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{N}, B_n \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}\bigg(\Big\{ \omega \in \mathbb{R}^{\mathbb{T}} : \ (\omega_{t_1}, \dots, \omega_{t_n}) \in B_n \Big\} \bigg) = P_{t_1,\dots,t_n}(B_n) \tag{3.0.1}$$

Proof

The elements of the infinite product space $\Omega = \mathbb{R}^{\mathbb{T}}$ are maps $t \mapsto \omega_t \in \mathbb{R}$. We define the cylinders algebra $\mathcal{C}$ whose elements are the cylinders

$$C = \Big\{ \omega \in \mathbb{R}^{\mathbb{T}} : (\omega_{t_1}, \dots, \omega_{t_n}) \in B_n \Big\}$$

where $n \in \mathbb{N}, t_1, \ldots, t_n \in \mathbb{N}, B_n \in \mathcal{B}(\mathbb{R}^n)$.

Equation (3.0.1) defines a map $\mathbb{P}_0 : \mathcal{C} \to [0, 1]$.

Since the family of finite dimensionla distributions is assumed to be consistent, it follows that $\mathbb{P}_0(C)$ does not depend on the particular representation of the cylinder C and it is well defined.

Since for any pair of cylinders we can find a representation based on a common fininte set of $\mathbb{T}$-indexes, and since the finite dimensional distributions are probabilities, it is not difficult to show that $\mathbb{P}_0 : \mathcal{C} \to [0, 1]$ is finitely additive (we leave this as an exercise).

When we show that $\mathbb{P}_0$ is also σ-additive on the cylinder algebra $\mathcal{C}$, Charatheodory's extension Theorem applies and $\mathbb{P}_0$ has an unique extension to a σ-additive probability $\mathbb{P}$ defined on the σ-algebra $\sigma(C)$.

By the characterization of the σ-additivity for probability measures,we need to show the following claim:

if $\{C_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$ is a sequence of cylinders such that

$$C_n \supseteq C_{n+1} \forall n, \text{ ja } \bigcap_{n \in \mathbb{N}} C_n = \emptyset,$$

it follows that $\lim_{n \to \infty} \mathbb{P}_0(C_n) = 0$.

By contradiction, let's assume instead that there is a sequence of cylinders $\{C_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$ such that $C_n \supseteq C_{n+1}$ and $\mathbb{P}_0(C_n) \geq \varepsilon \ \forall n$ jollekin $\varepsilon > 0$, we show that $\bigcap_{n \in \mathbb{N}} C_n \neq \emptyset$.

By choosing in a suitable way the representation of the cylinders and possibly by repeating the cylinders in the sequence, or taking subsequences, Valitsemalla sopivasti sylinterien esityksiä ja mahdollisesti toistaamalla sylintereita we can always find a sequence of indexes $(t_n) \subseteq \mathbb{T}$ and another sequence of cylinders $\{D_n : n \in \mathbb{N}\}$ with representation

$$D_n = \left\{ \omega \in \mathbb{R}^{\mathbb{T}} : (\omega_{t_1}, \ldots, \omega_{t_n}) \in A_n \right\}$$

where $D_n \supseteq D_{n+1} \forall n$, $A_n \in \mathcal{B}(\mathbb{R}^n)$, $A_n \times \mathbb{R} \supseteq A_{n+1} \forall n$, and $\forall m \in \mathbb{N}$ exists n such that $D_n = C_m$.

For such sequence we have $\mathbb{P}_0(D_n) \geq \varepsilon > 0 \; \forall n$ and

$$\bigcap_{n\in\mathbb{N}} C_n = \bigcap_{n\in\mathbb{N}} D_n \ .$$

Since $P_{t_1,\dots,t_n}$ is a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, it is σ-additive, and since A_n is Borel measurable, by the Approximation Property in Exercise 1.1.1, there is a closed set $F_n \subseteq A_n$ such that

$$P_{t_1,\dots,t_n}(A_n \setminus F_n) < \varepsilon 2^{-n} \ .$$

By choosing a closed ball $\overline{B(0,r_n)}$,centered at the origin and with radius large enough, we can find also a compact $K_n = \big(F_n \cap \overline{B}(0,r_n)\big) \subseteq A_n$ such that

$$P_{t_1,\dots,t_n}(A_n \setminus K_n) = \mathbb{P}_0(D_n \setminus G_n) < \varepsilon 2^{-n}$$

where

$$G_n := \Big\{\omega \in \mathbb{R}^{\mathbb{T}} : (\omega_{t_1},\dots,\omega_{t_n}) \in K_n\Big\}$$

Since these G_n do not necessarily form anymore a non-increasing sequence, we take the intersections

$$G'_n = \bigcap_{m=1}^{n} G_m = \Big\{\omega \in \mathbb{R}^{\mathbb{T}} : (\omega_{t_1},\dots,\omega_{t_n}) \in K'_n\Big\}$$

where

$$K'_n := K_n \cap (K_{n-1} \times \mathbb{R}) \cap \dots \cap (K_1 \times \mathbb{R}^{n-1}) \subseteq K_n$$

are compact set (the intersection of a compact with a closed set is compact).

It follows that $G'_n \supseteq G'_{n+1}$, and $(K'_n \times \mathbb{R}) \supseteq K'_{n+1}$.

This implies

$$\begin{aligned}
&P_{t_1,\dots,t_n}(K_n') = \mathbb{P}_0(G_n') = \mathbb{P}_0(D_n) - \mathbb{P}_0(D_n \setminus G_n') = \\
&P_{t_1,\dots,t_n}(A_n) - P_{t_1,\dots,t_n}\left(\bigcup_{m=1}^{n} A_n \setminus (K_m \times \mathbb{R}^{(n-m)})\right) \\
&\geq P_{t_1,\dots,t_n}(A_n) - P_{t_1,\dots,t_n}\left(\bigcup_{m=1}^{n} (A_m \setminus K_m) \times \mathbb{R}^{(n-m)}\right) \\
&(\text{ since } A_m \times \mathbb{R}^{(n-m)} \supseteq A_n \text{ for } n \geq m\ ) \\
&= \mathbb{P}_0(D_n) - \mathbb{P}_0\left(\bigcup_{m=1}^{n} D_m \setminus G_m\right) \\
&\geq \mathbb{P}_0(D_n) - \sum_{m=1}^{n} \mathbb{P}_0(D_m \setminus G_m) \geq \varepsilon - \sum_{m=1}^{n} \varepsilon 2^{-m} > \frac{\varepsilon}{2} > 0
\end{aligned}$$

where we used the contradiction hypothesis $\mathbb{P}_0(D_n) > \varepsilon$.

Therefore $\forall n$, $\exists (x_1^{(n)} \dots, x_n^{(n)}) \in K_n' \neq \emptyset$.

Since the sequence G_n' is non-decreasing, it follows that $(x_1^{(n)}) \subseteq K_1' \subseteq \mathbb{R}$, $\forall n \in \mathbb{N}$.

Now K_1' on kompakti, and by the Heine-Borel Theorem it follows that seuraa there exists a convergent subsequence $x_1^{(n_l)} \to x_1^* \in K_1'$.

Also $(x_1^{(n_l)}, x_2^{(n_l)} : \ell \in \mathbb{N}) \subseteq K_2'$, and as before, there exists a further converging subsequence of the subsequence which has limit $(x_1^*, x_2^*) \in K_2' \subseteq \mathbb{R}^2$. By induction, we find a sequence $(x_n^* : n \in \mathbb{N})$ such that $(x_1^*, \dots, x_n^*) \in K_n' \subseteq \mathbb{R}^n\ \forall n \in \mathbb{N}$. The sets

$$D^* = \left\{\omega \in \mathbb{R}^{\mathbb{T}} : \omega_{t_n} = x_n^* \quad \forall n\right\} \subseteq \bigcap_{n\in\mathbb{N}} G_n' \subseteq \bigcap_{n\in\mathbb{N}} D_n = \bigcap_{n\in\mathbb{N}} C_n$$

are not empty □

Remark 3.0.1. *Kolmogorov extension Theorem can be applied not only when $\Omega_t = \mathbb{R}, \forall t \in \mathbb{T}$ but also when Ω_t are Borel probability spaces, which means that there is a measurable bijection with measurable inverse between Ω_t and some $B \in \mathcal{B}(\mathbb{R})$. This is done simply by constructing the probability first on $\mathbb{R}^{\mathbb{T}}$ using the finite dimensional distributions induced by the bijection, and then by using the measurable bijection again in the other direction to obtain the probability measure* *on $\prod_{t \in T} \Omega_t$ as the push-forward of the infinite dimensional probability measure on $\mathbb{R}^{\mathbb{T}}$. All separable metric spaces (containing a countable dense set) equipped with their Borel σ-algebra are Borel spaces.*

Chapter 4

Expectation and Integral

Definition 4.0.4. *We say that $X(\omega) : (\Omega, \mathcal{F}) \to (R, \mathcal{B}(\mathbb{R}))$ is a simple random variable when it is measurable and takes finitely many values. In such case it has representation*

$$X(\omega) = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}(\omega), \quad n \in \mathbb{N}, x_k \in \mathbb{R}, A_k \in \mathcal{F}\;;. \tag{4.0.1}$$

Note that the representation (4.0.1) *is not unique, unless we require that the sets A_k are disjoint and the values x_k are distinct.*

We denote by $\mathcal{SF}$ the class of simple random variables, and $\mathcal{SF}^+$ denotes the class of simple non-negative random variables, For $X \in \mathcal{SF}$ with representation (4.0.1) *, we define its expectation with respect to the probability $\mathbb{P}$ as*

$$E_{\mathbb{P}}(X) = \sum_{k=1}^{n} x_k \mathbb{P}(A_k) = \sum_{x \in \mathbb{R}} x \mathbb{P}(\{\omega : X(\omega) = x\}) \tag{4.0.2}$$

It is a simple exercise to check that (4.0.2) does not depend on the choice of the representation (4.0.1) of X.

Following De Finetti, the expectation $E_{\mathbb{P}}(X)$ is interpreted as the fair price of a betting contract X, which pays out x_k when the event A_k happens, in the price system where the contracts $\mathbf{1}_{A_k}$ have price $\mathbb{P}(A_k)$, $k = 1, \dots, n$.

Lemma 4.0.1. *Let $X, Y \in \mathcal{SF}$ be simple random variables and, $\lambda \in \mathbb{R}$ be deterministic.*

- *(Positivity): When* $X(\omega) \geq 0$, $\mathbb{P}$*-almost surely (outside a* $\mathbb{P}$*-null set),* $E_P(X) \geq 0$.

- *(Linearity):* $E_P(X+Y) = E_P(X) + E_P(Y)$ *and* $E_P(\lambda X) = \lambda E_P(X)$.

We extend the definition of expectation as follows:

Definition 4.0.5. *Let* $X(\omega) \geq 0\ \forall \omega \in \Omega$. *Define*

$$E_{\mathbb{P}}(X) := \sup\{E_{\mathbb{P}}(Y) : Y \in \mathcal{SF}^+,\ 0 \leq Y(\omega) \leq X(\omega) \quad \forall \omega\}$$

More in general, we decompose the random variable $X(\omega)$ *into positive and negative parts* $X(\omega) = X^+(\omega) - X^-(\omega)$, *where* $X^+(\omega) := \max\{X(\omega), 0\}$ *and* $X^-(\omega) := \max\{-X(\omega), 0\}$ *are non-negative random variables, and we define*

$$\int_\Omega X(\omega)\mathbb{P}(d\omega) := E_{\mathbb{P}}(X) := E_{\mathbb{P}}(X^+) - E_{\mathbb{P}}(X^-)\ .$$

In this way $E_P(X)$ *is well defined unless* $E_{\mathbb{P}}(X^+) = E_{\mathbb{P}}(X^-) = -\infty$.

When $E_P(X^+) = \infty$ *and* $E_P(X^-) < +\infty$, *we set* $E_P(X) = +\infty$, *and when and* $E_P(X^+) < \infty$ *and* $E_P(X^-) = \infty$, *and we set* $E_P(X) := -\infty$.

When $E_{\mathbb{P}}(|X|) = E_{\mathbb{P}}(X^+) + E_{\mathbb{P}}(X^-) < +\infty$, *then* $E_P(X)$ *is well defined and finite, we say that* X *is integrable with respect to* $\mathbb{P}$. $L^1(\Omega, \mathcal{F}, P)$ *denotes the space of* $\mathbb{P}$*-integrable random variables.*

Lemma 4.0.2. *Let* $X(\omega) \geq 0, \forall \omega$.

$$E_{\mathbb{P}}(X) = 0 \Longrightarrow \mathbb{P}(\{\omega : X(\omega) > 0\}) = 0.$$

Proof. Let $Y_n(\omega) = n^{-1}\mathbf{1}(X(\omega) > n^{-1})$. Then

$$Y_n \in \mathcal{SF}^+ \text{ and } Y_n(\omega) \leq X(\omega),\ \forall \omega.$$

By definition of the expectation it follows that

$$0 = E_{\mathbb{P}}(X) \geq E_{\mathbb{P}}(Y_n) = n^{-1}\mathbb{P}(X(\omega) > n^{-1}) \geq 0, \quad \text{and}$$
$$\mathbb{P}(X(\omega) > 0) = \mathbb{P}\left(\bigcup_n \{\omega : X(\omega) > n^{-1}\}\right) \leq \sum_n \mathbb{P}(X(\omega) > n^{-1}) = 0 \quad \square$$

Lemma 4.0.3. *When $X(\omega) \in L^0(\Omega, \mathcal{F}, \mathbb{P})$ and $B \in \mathcal{F}$ jolla $\mathbb{P}(B) = 1$, seuraa*

$$E_{\mathbb{P}}(X\mathbf{1}_B) = E_{\mathbb{P}}(X).$$

Proof: The claim holds when $X(\omega) = \mathbf{1}_A(\omega)$:

$$E_{\mathbb{P}}(\mathbf{1}_A\mathbf{1}_B) = E_{\mathbb{P}}\big(\mathbf{1}_{A\cap B}\big) = \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c) = \mathbb{P}(A),$$

since $\mathbb{P}(A \cap B^c) \leq \mathbb{P}(B^c) = 1 - \mathbb{P}(B) = 0$. By linearity it holds also for all simple random variables $X \in \mathcal{SF}$.

Let $X(\omega) \geq 0$, and $\{X_n : n \in \mathbb{N}\} \subseteq \mathcal{SF}^+$ such that $0 \leq X_n(\omega) \leq X(\omega)$ and $E_{\mathbb{P}}(X_n) \uparrow E_{\mathbb{P}}(X)$. Then, since $\mathbf{1}_B(\omega)X_n(\omega) \leq \mathbf{1}_B(\omega)X(\omega)$ and the expectation is a positive operator,

$$E_{\mathbb{P}}(\mathbf{1}_B X) \geq E_{\mathbb{P}}(\mathbf{1}_B X_n) = E_{\mathbb{P}}(X_n) \uparrow E_{\mathbb{P}}(X) \geq E_{\mathbb{P}}(\mathbf{1}_B X) \ \square$$

Corollary 4.0.1. *If $\mathbb{P}(\{\omega : X(\omega) = Y(\omega)\}) = 1$, we say that $X(\omega) = Y(\omega)$ $\mathbb{P}$-almost surely, and it follows $E_{\mathbb{P}}(X) = E_{\mathbb{P}}(Y)$.*

Remark In the class of $\mathcal{F}$-measurable random variables $L^0(\Omega, \mathcal{F}, \mathbb{P})$, we can identify the random variables X and Y when $\mathbb{P}(X = Y) = 1$.

Riemann-Stieltjes and Lebesgue-Stieltjes integrals Let $a < b \in \mathbb{R}$, $\Omega = (a, b], \mathcal{F} = \mathcal{B}((a, b])$,

$P = P_G$ where $G : (a, b] \to [0, 1]$ is non-decreasing, right continuous, such that $G(a) = 0$, $G(b) = 1$.

By Caratheodory Theorem there is an unique probability measure $P_G : \mathcal{B}((a, b]) \to [0, 1]$ such that $P_G((a, x]) = G(x) - G(a)$ for $a \leq x$.

Let $H : (a, b] \to \mathbb{R}^+$ a non-negative measurable function. We show that

$$E_{P_G}(H) = \int_a^b H(x)G(dx) \text{ (Lebesgue-Stieltjes integral)}$$

where the integral on the right size is a generalization of the Riemann-Stieltjes integral.

Definition 4.0.6. *Let* Π *a finite interval partition of the interval* $(a,b]$ *with a finite set of grid points:*

$$\Pi = (a = x_0 \le \xi_0 \le x_1 \le \xi_1 \le x_2 \le \cdots \le x_{N-1} \le \xi_{N-1} \le x_N = b), \quad n \in \mathbb{N}$$

We define the size of the partition $\Delta(\Pi) = \max_{0<i\le n}(x_i - x_{i-1})$*. The Riemannin Stieltjes integral is defined as the limit*

$$\text{(Riemann-Stieltjes)-}\int_a^b H(x)G(dx) := \lim_{\Delta(\Pi)\to 0} \sum_{k=1}^{N(\Pi)} H(\xi_i)(G(x_{i+1}) - G(x_i))$$

$$= \lim_{\Delta(\Pi)\to 0} \sum_{k=1}^{N(\Pi)} H(\xi_i) P_G((x_i, x_{i+1}]) = \lim_{\Delta(\Pi)\to 0} E_P(H^{(\Pi)}) \quad \textit{where}$$

$$H^{(\Pi)}(u) := \sum_{k=1}^{N(\Pi)} H(\xi_i)\mathbf{1}(x_i, x_{i+1}](u)$$

where the limit exists for any sequence of finite partitions $(\Pi^{(n)} : n \in \mathbb{N})$ *with* $\Delta(\Pi^{(n)}) \to 0$*, and it does not depend on the particular sequence of partitions choise of intermediate points* $\xi_i^{(n)} \in [x_i^{(n)}, x_{i+1}^{(n)}]$*.*

Proposition 4.0.1. *When* $H(x)$ *is (piecewise) continuous, and* $G(x)$ *on non-decreasing,the Riemann-Stieltjes integral* $\int\limits_a^b H(x)G(dx)$ *exists.*

Proof Let $H : [a,b] \longrightarrow \mathbb{R}$ be a continuous function, and $H^{(\Pi)}$ its simple approximation among the finite interval partition Π. By definition

$$\int_a^b H(y)G(dy) = \lim_{\Delta(\Pi)\longrightarrow 0} \int_a^b H^{(\Pi)}(y)G(dy)$$

when the limit exisis and it does not depend on the sequence of partitions.

Let $(\Pi_n : n \in \mathbb{N})$ a sequence of finite interval partitions of the interval $[a,b]$ with $\Delta(\Pi_n) \longrightarrow 0$. For the interval partitions Π_n, Π_m we have

$$\left| \int_a^b H^{(\Pi_n)}(y)G(dy) - \int_a^b H^{(\Pi_m)}(y)G(dy) \right| \le \int_a^b \left| H^{(\Pi_n)}(y) - H^{(\Pi_m)}(y) \right| G(dy)$$

$$= \sum_{i=0}^{N-1} |H(\xi_i) - H(\eta_i)|(G(x_{i+1}) - G(x_i))$$

for some $x_0 < x_1 < \cdots < x_N$ and $|\xi_i - \eta_i| \le \big(\Delta(\Pi_n) + \Delta(\Pi_m)\big)$. Since H is continuous on the compact interval $[a,b]$, it is uniformly continuous, meaning that, $\forall \varepsilon > 0$ there exists a $\delta > 0$ such that $|\xi - \eta| \le \delta \Longrightarrow |H(\xi) - H(\eta)| < \varepsilon$. This implies that, $\forall n, m$ large enough,

$$\bigg| \int_a^b H^{(\Pi_n)}(y) G(dy) - \int_a^b H^{(\Pi_m)}(y) G(dy) \bigg| \le \varepsilon \sum_{i=0}^{N-1} \big(G(x_{i+1}) - G(x_i)\big) = \varepsilon(G(b) - G(a))$$

This implies that $\left(\int\limits_a^b H^{\Pi_n}(y) G(dy) \,:\, n \in \mathbb{N} \right)$ is a Cauchy sequence which must converge to a limit.

By the same argument, if (Π'_n) is another sequence of interval partitions with $\Delta(\Pi'_n) \to 0$, by defining the partition sequence (Π''_n) such that $\Pi''_{2n} = \Pi_n$ ja $\Pi''_{2n+1} = \Pi'_n$, it follows that also $\int\limits_a^b H^{(\Pi''_n)}(y) G(dy)$ is convergent as $n \to \infty$, which means that the limit does not depend on the sequence of partitions □

Problem 4.0.1. *Show that the Riemann-Stieltjes integral $\int_a^b H(y)G(dy)$ exists when the function $H(y)$ if non-decreasing. Hint: Show first that the set of discontinuity points*

$$D := \Big\{ \, y : H(y-) := \lim_{x \uparrow y} H(x) < H(y+) = \lim_{x \downarrow y} H(x) \Big\}$$

is at most countable. Use the decomposition

$$H(x) = \left(H^c(x) - \sum_{y \le x} \Delta H(y) \right) \quad \textit{where} \quad \Delta H(y) := H(y+) - H(y-)$$

and show that $H_c(x)$ is non-decreasing and continuous. This generalizes directly to the case where $H(y) = (H^{\oplus}(y) - H^{\ominus}(y))$, $G(y) = (G^{\oplus}(y) - G^{\ominus}(y))$, with $H^{\oplus}, H^{\ominus}, G^{\oplus}$ and $G^{\ominus}$ non-decreasing.

Problem 4.0.2. *$H(x)$ on (piecewise) continuous and $G(x)$ is differentiable with continuous derivative $G'(x)$, we have*

$$\int_a^b H(x) G(dx) = \int_a^b H(x) G'(x) dx \ ,$$

where the integrals exist in Riemann-Stieltjes sense.

When the function $H(x)$ is Borel- measurable but not piecewise-continuous, there exist sequences of partitions and intermediate points such that the limit of the Riemann sums does not exists. In that case Riemann-Stieltjes integration method does not work.

In 1902 the french mathematician Henri Lebesgue in his dissertation invented a new integration approach, where the integral of the function $y = H(x)$ is approximated by partitioning the y-axis (the space where the function takes its values) instead of the x-axis (the argumnents space). Let H be a Borel-measurable function such that $H(x) \geq 0 \; \forall x$. Define the map $\alpha^{(N)} : [0, +\infty) \to [0, +\infty)$ with

$$\alpha^{(N)}(y) = \begin{cases} 0 & \text{when } 0 \leq y \leq N^{-1} \\ \left(\frac{k}{N}\right) & \text{when } kN^{-1} < y \leq (k+1)N^{-1}, \quad k \leq (N^2 - 1) \\ N & \text{when } y \geq N \end{cases} \tag{4.0.3}$$

$$\text{siis } \alpha^{(N)}(y) = \sum_{k=1}^{(N^2-1)} \left(\frac{k}{N}\right) \mathbf{1}_{\left(\frac{k}{N}, \frac{k+1}{N}\right]}(y) \quad + \; N \, \mathbf{1}_{(N,+\infty]}(y)$$

Note that the map $\alpha^{(N)}$ is continuous from the left as the indicator functions $\mathbf{1}_{(a,b]}(y)$.

We approximate from below the random variable $H(x) \geq 0$ by the simple random variable $H^{(N)}(x) := \alpha^{(N)}(H(x))$. Since

$$0 \leq H(x) - H^{(N)}(x) \leq 1/N \quad \text{ when } 0 \leq H(x) \leq N,$$

and $H^{(N)}(x) \leq H^{(N+1)}(x)$, $H^{(N)}(x) \uparrow H(x)$ is converging monotonically as $N \uparrow \infty$.

Since $H^{(N)}(x)$ is a simple random variable, its P_G-expectation is the sum

$$E_P(H^{(N)}) = \sum_{y \in \mathbb{R}} y P_G(H^{(N)} = y) =$$

$$\sum_{k=0}^{(N^2-1)} \frac{k}{N} P_G\left(\left\{x : H(x) \in \left(\frac{k}{N}, \frac{k+1}{N}\right]\right\}\right) = \int_a^b H^{(N)}(x) G(dx)$$

where the last integral is a Lebesgue-Stieltjes integral and in general it is not Riemannin integral, since the counterimages $\{ \{x : H(x) \in \big(\frac{k}{N}, \frac{k+1}{N}\big]\Big\}$ are Borel sets which are not representable as finite union of intervals. unless $H(x)$ and $H^{(N)}(x)$ are piecewise continuous.

Note also that when the function $H(x)$ is continuous in the interval $(a, b]$, the inverse map

$$H^{-1}\bigg((\frac{k}{N}, \frac{k+1}{N}] \bigg) := \bigg\{ x : H(x) \in (\frac{k}{N}, \frac{k+1}{N}] \bigg\}$$

is a finite union of invervals. In such case a finite partition of the y-axis correponds to a finite partition of the x-axis and in such case the Lebesgue and Riemann integrals coincide.

This is not always the case when $H(x)$ is just Borel-measurable.

We show in section 4.1 when that $H(x) \geq 0$ is measurable, the limit $\lim_{N\to\infty} E_{P_G}(H^{(N)})$ exists for all sequences of simple random variables such that $H^{(N)}(x) \uparrow H(x)$, and the limit is the expectation

$$E_{P_G}(H) = \int_0^\infty y \; P_G(x : H(x) \in dy) := \text{(Lebesgue-Stieltjes)-} \int_0^\infty H(x) G(dx) \;,$$

which does not depend on the choice of the approximating sequence.

4.1 Monotone Convergence Theorem.

By definition, when $0 \leq X \in \mathcal{F}$, there is a sequence of simple random variables $(X_n : n \in \mathbb{N}) \subseteq \mathcal{SF}^+$ such that $0 \leq X_n(\omega) \leq X(\omega)$ and $E(X_n) \to E(X)$.

We show that $E(X_n) \uparrow E(X)$ for all monotone sequences of random variables such that $0 \leq X_n \uparrow X$.

Lemma 4.1.1. *Let sequence $\{a_n^k : k, n \in \mathbb{N}\} \subseteq \mathbb{R}$ which on non-decreasing molempien indeksien (k, n) with respect to,*

$$a_{n-1}^k \leq a_k^n \leq a_{k+1}^n \quad \forall k, n \in \mathbb{N}$$

Then the limit as $k, n \to \infty$ exists and it does not depend on the order in which we take the limit

$$\exists a := \lim_n \big(\lim_k a_n^k\big) = \lim_k \big(\lim_n a_n^k\big),\ a \in \mathbb{R} \cup \{+\infty\}$$

Also for every subsequence $\{n(l), k(l)\}_{l\in\mathbb{N}}$ where $n(l), k(l) \to \infty$

$$\lim_{l\to\infty} a_{n(l)}^{k(l)} = a$$

Proof. By monotonicity it follows that $\forall k$ as $n \uparrow \infty$

$$\begin{array}{ccc} a_n^k & \uparrow & a_\infty^k \\ |\wedge & & |\wedge \\ a_n^{k+1} & \uparrow & a_\infty^{k+1} \end{array}$$

This implies that a_∞^k is a monotone sequence and $a' = \lim\limits_{k\to\infty} a_\infty^k$. Similarily, $\forall n$ kun $k \uparrow \infty$, $a_n^k \uparrow a_n^\infty$ and $a_n^\infty \le a_{n+1}^\infty$, which implies that the monotone limit $a'' = \lim\limits_{n\to\infty} a_n^\infty$ exists.

When $a' = a'' = +\infty$ the claim is true.

In case $a' < a'' = +\infty$, we would have a contradiction, since $\forall N, \varepsilon > 0$ there would be n, k large enough such that

$$N \le a_n^\infty \le a_n^k + \varepsilon \le a_\infty^k + \varepsilon \le a' + \varepsilon$$

When both $a', a'' < \infty$, it follows that $\forall \varepsilon > 0$ there are $\bar{n}, \bar{k}$ such that

$$a'' < a_{\bar{n}}^\infty + \varepsilon\ , \quad a_{\bar{n}}^\infty < a_{\bar{n}}^{\bar{k}} + \varepsilon\ .$$

This implies

$$a' \ge a_\infty^{\bar{k}} \ge a_{\bar{n}}^{\bar{k}} \ge a_{\bar{n}}^\infty - \varepsilon \ge a'' - 2\varepsilon$$

Since ε on arbitrary, $a' \ge a''$ follows. Similarily it follows that $a'' \ge a'$.

When $n(l), k(l) \uparrow (+\infty)$ are increasing sequences of indexes,

$$a' = \lim_{l\to\infty} a_\infty^{k(l)} \ge \lim_{l\to\infty} a_{n(l)}^{k(l)} \ge \lim_{l\to\infty} a_n^{k(l)} \quad \forall n,$$
$$a' \ge \lim_{l\to\infty} a_{n(l)}^{k(l)} \ge \lim_{n\to\infty} \lim_{k\to\infty} a_n^k = a'' = a' \ \square$$

Remark 4.1.1. *Generally speaking, this does not follow unless the sequence is monotone in the same direction with respect to both indexes. For example, for a function $g : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ always*

$$\sup_x \inf_y g(x,y) \le \inf_y \sup_x g(x,y) ,$$

but the opposite inequality does not hold without additional assumptions.

Lemma 4.1.2. *Monotone convergence Theorem for simple random variables*
Let $X, X_n \in \mathcal{SF}^+$ where $0 \le X_n(\omega) \uparrow X(\omega)$. Then $E_P(X_n) \uparrow E_P(X)$.

Consider first an indicator random variable $X(\omega) = \mathbf{1}_A(\omega)$, $A \in \mathcal{F}$. For $0 < \varepsilon < 1$, let

$$A_n^\varepsilon = \{\omega : X_n(\omega) \ge (1-\varepsilon)\} \subseteq A_{n+1}^\varepsilon \subseteq A$$

Since $X_n(\omega)$ is non-decreasing and $X_n(\omega) = 0$ when $\omega \notin A$, it follows that $A_n^\varepsilon \uparrow A$, which means that, when $\omega \in A$, also $\omega \in A_n$ for any large enough n, which means that $\forall\omega\ \exists\bar{n}(\omega)$ such that $\forall n > \bar{n}(\omega)$

$$X_n(\omega) > (\mathbf{1}_A(\omega) - \varepsilon) =$$
$$\text{, eli} \quad (1-\varepsilon) \quad \text{, kun } \omega \in A.$$

By σ-additivity, it follows that $P(A_n^\varepsilon) \uparrow P(A)$ and since $X_n(\omega) \ge (1-\varepsilon)\mathbf{1}_{A_n^\varepsilon}(\omega)$,

$$P(A) \ge E_P(X_n) \ge (1-\varepsilon)P(A_n^\varepsilon) \uparrow (1-\varepsilon)P(A)$$

follows. Since ε was arbitrary $E_P(X_n) \uparrow P(A) = E_P(X)$. Since the expectation of simple random variables is linear, the claim follows for all simple random variables $X \in \mathcal{SF}$.

Lemma 4.1.3. *Let $X(\omega) \in \mathcal{F}^+$. If $\{X_n\}, \{Y_n\} \subseteq \mathcal{SF}^+$, $X_n(\omega) \uparrow X(\omega)$ and $Y_n(\omega) \uparrow X(\omega)$ $\forall\omega$, it follows that*

$$\lim_n E_P(X_n) = \lim_n E_P(Y_n)$$

Proof. The sequence $Z^n_m(\omega) = \min\{X_n(\omega), Y_m(\omega)\}$ is non-decreasing with respect to both n, m indexes.

Since $Y_n(\omega) \uparrow X(\omega) \geq X_n(\omega) \uparrow X(\omega) \geq Y_n(\omega)$ it follows that

$$Z^n_m(\omega) \uparrow X_n(\omega) \text{ when } m \uparrow \infty, \text{ and } Z^n_m(\omega) \uparrow Y_m(\omega) \text{ when } n \uparrow \infty$$

By taking the expectation, since Z^n_m, X_n ja Y_m are simple random variables, it follows by lemma (4.1.2) that

$$E_P(Z^n_m) \uparrow E_P(X_n) \text{ when } m \uparrow \infty, \text{ and } E_P(Z^n_m) \uparrow E_P(Y_m) \text{ when } n \uparrow \infty,$$

and by lemma (4.1.1) it follows that

$$\lim_{n\to\infty} E_P(X_n) = \lim_{n\to\infty} \lim_{m\to\infty} E_P(Z^n_m) = \lim_{m\to\infty} \lim_{n\to\infty} E_P(Z^n_m) = \lim_{n\to\infty} E_P(Y_n) \ \square$$

Corollary 4.1.1. *When $X_n(\omega) \uparrow X(\omega)$ where $\{X_n\} \subseteq \mathcal{SF}^+$ ja $X \in \mathcal{F}^+$, it follows that $E_P(X_n) \uparrow E_P(X)$.*

Proof. By the definition of $E_P(Y)$ it follows that there exists $\{Y_n\} \subseteq \mathcal{SF}^+$ such that $0 \leq Y_n(\omega) \leq X(\omega)$ and $\lim_{n\to\infty} E_P(Y_n) = E_P(Y)$. Note that at this stage $Y_n(\omega)$ is not necessarily non-decreasing for all $\omega \in \Omega$.

Let $X^{(n)}(\omega) := \alpha^{(n)}(X(\omega)) \uparrow X(\omega)$ (see. 4.0.3), and define as it follows a non-decreasing sequence of random variables

$$Z_n(\omega) = \max\big\{X^{(n)}(\omega), Y_1(\omega), \dots, Y_n(\omega)\big\} \in \mathcal{SF}^+.$$

Since $Z_n(\omega) \geq X^{(n)}(\omega) \uparrow X(\omega)$ it follows also that $Z_n(\omega) \uparrow X(\omega)$.

Since $Y_n(\omega) \leq Z_n(\omega) \leq X(\omega)$ and the expectation $E_P : \mathcal{SF}^+ \to [0, +\infty]$ is a positive and linear operator

$$0 \leq E_P(Y_n) \leq E_P(Z_n) \leq E_P(X) \ ,$$

and since $\lim_n E_P(Y_n) = E_P(X)$, it follows that $E_P(Z_n) \uparrow E_P(X)$.

Now by using lemma (4.1.3) the claim follows:

$$\lim_{n\to\infty} E_P(X_n) = \lim_{n\to\infty} E_P(Z_n) = E_P(X) \ \square$$

Theorem 4.1.1. *Monotone convergence theorem.*

When $X, X_n \in \mathcal{F}^+$ such that $0 \le X_n(\omega) \uparrow X(\omega)$ P-almost surely, it follows that $E_P(X_n) \uparrow E_P(X)$ as well.

Let's assume first that $\forall \omega \in \Omega$: $0 \le X_n(\omega) \uparrow X(\omega)$. $\forall n$, we construct a sequence of simple random variable approximation by taking $X_n^{(k)}(\omega) = \alpha^{(k)}(X_n(\omega)) \in \mathcal{SF}^+$ jolla $0 \le X_n^{(k)}(\omega) \uparrow X_n(\omega)$ kun $k \uparrow \infty$. If follows from lemma (4.1.2) that $E_P(X_n^{(k)}) \uparrow E_P(X_n)$ as $k \uparrow \infty$.

Note also that, as $n \to \infty$, since the function $\alpha^{(k)}(x)$ continuous from the left, and $X_n(\omega) \uparrow X(\omega)$, it follows that

$$X_n^{(k)}(\omega) \uparrow X^{(k)}(\omega) := \alpha^{(k)}(X(\omega)) \in \mathcal{SF}^+ \ .$$

Morever, since the functions $\alpha^{(k)}(x)$ are non-decreasing, it follows also that $X_n^{(k)}(\omega) \le X_n^{(k+1)}(\omega)$. Since the expectation $E_P : \mathcal{SF}^+ \to \mathbb{R}^+ \cup [0, +\infty]$ is a positive operator, the sequence $\{E_P(X_n^{(k)}) : n, k \in \mathbb{N}\}$ is non-decreasing with respect to both indexes and by applying lemma (4.1.1) it follows that

$$\lim_{n\to\infty} E_P(X_n) = \lim_{n\to\infty} \lim_{k\to\infty} E_P(X_n^{(k)}) = \lim_{k\to\infty} \lim_{n\to\infty} E_P(X_n^{(k)}) = \lim_{k\to\infty} E_P(X^{(k)}) = E_P(X)$$

where the left hand side of the equation follows from (4.1.1).

When $0 \le X_n(\omega) \uparrow X(\omega)$ P-almost surely, there exists an event $B \in \mathcal{F}$ such that $P(B) = 1$ and $0 \le X_n(\omega)\mathbf{1}_B(\omega) \uparrow X(\omega)\mathbf{1}_B(\omega)$ $\forall \omega$. It follows from (4.0.3), (4.1.1) that

$$E_P(X_n) = E_P(X_n \mathbf{1}_B) \uparrow E_P(X\mathbf{1}_B) = E_P(X) \qquad \square$$

Corollary 4.1.2. *The expectation $E_{\mathbb{P}} : L^1(\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is a linear operator, meaning that when $a, b \in \mathbb{R}$ are deterministic,*

$$E_{\mathbb{P}}\big(aX + bY\big) = aE_{\mathbb{P}}\big(X\big) + bE_{\mathbb{P}}\big(Y\big) \ .$$

The proof is left as an exercise, by using first linearity for approximating simple random variables and applying then the monotone convergence theorem.

Lemma 4.1.4. *(Fatou lemma) When $X_n(\omega) \geq 0$ P-almost surely $\forall n \in \mathbb{N}$,*

$$0 \leq E_P\big(\liminf_n X_n\big) \leq \liminf_n E_P\big(X_n\big)$$

This follows also under the condition $X_n(\omega) \geq Z(\omega)$ $\forall n \in \mathbb{N}$ P-almost surely, where $E_P(Z^-) < +\infty$.

Proof. Let $Y_n(\omega) := \inf_{k \geq n} X_k(\omega)$, which is
$\forall \omega$, jolla $X_n(\omega) \geq Y_n(\omega) \uparrow \liminf_n X_n(\omega)$. Since by assumption $Y_n(\omega) \geq 0$ P-a.s. by the monotone convergence Theorem it follows that

$$\liminf_n E_P(X_n) \geq \liminf_n E_P(Y_n) = \lim_n E_P(Y_n) = E_P(\liminf_n X_n)$$

When $X_n(\omega) \geq Z(\omega) \geq -(Z^-(\omega))$, with $E_P(Z^-) < +\infty$,
since $(X_n(\omega) + Z^-(\omega)) \geq 0$, and by the linearity of the expectation,

$$\liminf_n E(X_n) + E_P(Z^-) = \liminf_n E_P(X_n + Z^-) \geq E_P(\liminf_n (X_n + Z^-))$$
$$= E_P(\liminf_n X_n) + E_P(Z^-) \geq 0$$

which gives the result under the assumption $E_P(Z^-) < \infty$.

Lemma 4.1.5. *(reverse-Fatou lemma) If $X_n(\omega) \leq Z(\omega) \leq Z^+(\omega)$ $\forall n \in \mathbb{N}$ P-almost surely, and $E_P(Z^+) < \infty$, it follows that*

$$\limsup_n E_P(X_n) \leq E_P(\limsup_n X_n) \leq E(Z)$$

Proof. We apply Fatou lemma to the sequence $X'_n(\omega) := -X_n(\omega) \geq -Z^+(\omega)$, and the claim follows since $(\limsup_n a_n) = -(\liminf_n (-a_n))$ □

Proposition 4.1.1. *Lebesgue dominated convergence theorem. when*

- $\lim_{n \to \infty} X_n(\omega) = X(\omega)$
- $\sup_{n \in \mathbb{N}} |X_n(\omega)| \leq Z(\omega)$ *where* $E_P(Z) < \infty$,

P-almost surely (which means $\forall \omega \in N^c$ with $P(N) = 0$), implies that $\lim_{n \to \infty} E_P(X_n) = E_P(X)$

Proof. Since $|X_n| \le Z$ ja $E_P(Z) < \infty$, both Fatou and reverse-Fatoun lemma do apply

$$E_P(X) = E_P(\liminf_n X_n) \le \liminf_n E_P(X_n) \le \limsup_n E_P(X_n) \le E_P(\limsup_n X_n) = E_P(X) \quad \Box$$

Lebesgue dominated convergence theorem applies also when we integrate against a σ-finite measure, for example in the $\mathbb{R}^d$ space eequipped with the Lebesgue measure.

Proposition 4.1.2. *Let $(\Omega, \mathcal{F})$ a measurable space equipped with a σ-finite measure $\lambda(d\omega)$, $\big(f_n(\omega)\big)_{n\in\mathbb{N}}$, $f(\omega)$ measurable functions where*

- $$\lim_{n\to\infty} f_n(\omega) = f(\omega) \quad \forall \omega \in N^c$$

  with $N \subset \Omega$ and $\mu(N) = 0$.

- *There is an integrable upper bound $g(\omega) \ge 0$ such that*

  $$|f_n(\omega)| \le g(\omega) \quad \forall \omega \in N^c,\ n \in \mathbb{N} \text{ and } \int_\Omega g(\omega)\,\lambda(d\omega) < \infty$$

Then

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\lambda(d\omega) = \int_\Omega \Big(\lim_{n\to\infty} f_n(\omega) \Big) \lambda(d\omega) = \int_\Omega f(\omega)\lambda(d\omega)$$

Proof. We can assume that $f_n(\omega) \ge 0\ \forall \omega \in \Omega, n \in \mathbb{N}$, otherwise we split as $f_n(\omega) = f_n(\omega)^+ - f(\omega)^-$ and integrate the positive and negative parts separately.

Since λ is σ-finite, there is a countable partition $(\Omega_n : n \in \mathbb{N}) \subset \mathcal{F}$ such that $\Omega_i \cap \Omega_j = \emptyset$ for $i \ne j$, $\Omega = \bigcup_{i\in\mathbb{N}} \Omega_i$ and $0 < \lambda(\Omega_i) < \infty\ \forall i \in \mathbb{N}$.

We define the sequence of probabilities

$$P_i(A) := \frac{\lambda(A \cap \Omega_i)}{\lambda(\Omega_i)}, \quad A \in \mathcal{F}$$

Note that

$$\int_\Omega g(\omega)\lambda(d\omega) = \sum_{i\in\mathbb{N}} \lambda_i(\Omega_i) \int_\Omega g(\omega) P_i(d\omega) < \infty$$

which implies $g(\omega) \in L^1(\Omega, \mathcal{F}, P_i)\ \forall i$. Lebesgue dominated konvergence theorem 4.1.1 applies to every P_i probability measure and we obtain

$$\lim_{n\to\infty} \int_0^\infty f_n(\omega) P_i(d\omega) = \int_0^\infty f(\omega) P_i(d\omega)$$

The claim follows since

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\lambda(d\omega) = \lim_{n\to\infty} \lim_{m\to\infty} \sum_{i=1}^m \int_{\Omega_i} f_n(\omega)\lambda(d\omega) = \lim_{m\to\infty} \sum_{i=1}^m \lim_{n\to\infty} \int_{\Omega_i} f_n(\omega)\lambda(d\omega)$$
$$= \lim_{m\to\infty} \sum_{i=1}^m \int_{\Omega_i} \Big(\lim_{n\to\infty} f_n(\omega)\Big)\lambda(d\omega) = \lim_{m\to\infty} \sum_{i=1}^m \int_{\Omega_i} f(\omega)\lambda(d\omega) = \int_\Omega f_n(\omega)\lambda(d\omega)$$

where we used lemma (4.1.1) for the sequence

$$\left\{ \sum_{i=1}^n \int_{\Omega_i} f_m(\omega)\lambda(d\omega) :\ m, n \in \mathbb{N} \right\} \square$$

Example 4.1.1. *Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, and let P be the uniform probability measure such that $P((a,b]) = (b-a)$ when $0 \le a \le b \le 1$.*

Let $X_n(\omega) = n\mathbf{1}(\omega \le 1/n)$.

Note that when $\omega > 0$, $X_n(\omega) \to 0$ as $n \to \infty$, since $X_n(\omega) = 0$ for $n > 1/\omega$. However, for fixed $\omega > 0$ the sequence $(X_n(\omega) : n \in \mathbb{N})$ is not monotone, and we cannot apply the monotone convergence theorem.

Since $P((0,1]) = 1$, $X_n(\omega) \to 0$ P-almost surely.

On the other hand $E_P(X_n) = nn^{-1} = 1 > 0 = E_P(\lim_n X_n)$.

Also the assumption of Lebesgue dominated convergence theorem does not apply:

$$X^*(\omega) = \sup_n X_n(\omega) = n \text{ when } \omega \in \big(\frac{1}{n+1}, \frac{1}{n}\big] \iff \omega^{-1} \in [n, n+1)$$
$$\text{which means } X^*(\omega) = \lfloor \omega^{-1} \rfloor \text{ and } (\omega^{-1} - 1) < X^*(\omega) < \omega^{-1}$$

We see that

$$E_P(X^*) > \left(\int_0^1 x^{-1} dx\right) - 1 = \int_0^1 d\log(x) - 1 = \log(1) - \log(0) - 1 = +\infty$$

which means that X^* is not integrable and the dominated convergence theorem does not apply.

Lemma 4.1.6. *Let $X(\omega) \in \mathbb{R}^d$ be a random vector, ja $Y(\omega) = g(X(\omega)) \in \mathbb{R}$ a $\sigma(X)$-measurable random variable, where $g : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a Borel measurablle function. Then*

$$\int_{\mathbb{R}^d} g(x) P_X(dx) = E_P(g(X)) = \int_\Omega g(X(\omega)) P(d\omega) = E_P(Y) = \int_{\mathbb{R}} y P_Y(dy)$$

where $P_Y(B) = P(Y^{-1}(B)) = P_X(g^{-1}(B))$ for $B \in \mathcal{B}(\mathbb{R})$.

Proof. when $g(x) = \mathbf{1}_D(x)$ for some $D \in \mathcal{B}(\mathbb{R}^d)$, the claim follows directly by the definition of the push-forward measures (or image measures) P_X and P_Y, when $g(x)$ is a simple measurable function taking finitely many values the claim follows by linearity of the expectation.

More in general assume that $g(x) \geq 0$, otherwise we first split the function as $g(x) = g^+(x) - g^-(x)$. In such case there is a sequence of measurable simple functions $g^{(N)}(x) \uparrow g(x)\ \forall x \in \mathbb{R}^d$ (kts. 4.0.3), which implies that $g^{(N)}(X(\omega)) \uparrow g(X(\omega))\ \forall \omega \in \Omega$. By the monotone convergence theorem it follows that

$$\int_\Omega g^{(N)}(X(\omega)) P(d\omega) \uparrow \int_\Omega g(X(\omega)) P(d\omega) \text{ and } \int_{\mathbb{R}^d} g^{(N)}(x) P_X(dx) \uparrow \int_{\mathbb{R}^d} g(x) P_X(dx),$$

where $\displaystyle\int_\Omega g^{(N)}(X(\omega)) P(d\omega) = \int_{\mathbb{R}^d} g^{(N)}(x) P_X(dx)$ □

Example 4.1.2. *When the cumulative distribution function $F_X(t) = P(X \leq t)$ of a $\mathbb{R}$-valued random variable X which is absolutely continuous with respect to Lebesgue measure, which means*

$$F_X(b) = F_X(a) + \int_a^b f_X(t) dt$$

for some Borel measurable function $f_X(t) \geq 0$, which is called probability density function. When the classical derivative $\frac{dF_X}{dt}(t)$ exists at all t, then it is a probability density function. More in general $\frac{dF_X}{dt}(t) = \frac{dP_X}{dt}(t) =$ is understood as the Radon Nikodym derivative of the push-forward probability measure P_X with respect to Lebesgue measure.

In such case, for every non-negative and Borel measurable test function $g(x) \geq 0$ we have

$$E_P\big(g(X)\big) = \int_\Omega g(X(\omega)) P(d\omega) = \int_{\mathbb{R}} g(x) F(dx) = \int_{\mathbb{R}} g(x) f_X(x) dx$$

For example let $X(\omega) \geq 0$ λ-exponential random variable, where $\lambda > 0$ ís a parameter, with distribution function $P(X \leq x) = F_X(x) = 1 - \exp(-\lambda x^+)$ where $\lambda > 0$, $x^+ = x \vee 0$. Let's compute:

$$E_P(X) = \int_0^\infty x F_X(dx) = \int_0^\infty x \frac{d}{dx} F_X(x) dx = \int_0^\infty x\lambda \exp(-\lambda x) dx =$$
$$\lambda \Big[x \exp(-\lambda x) \Big]_0^{+\infty} + \int_0^\infty \exp(-\lambda x) dx = 0 - 0 - \frac{1}{\lambda} \Big[\exp(-\lambda x) \Big]_0^\infty = \frac{1}{\lambda}$$

4.2 Application: change of probability measure

On a probability space $(\Omega, \mathcal{F}, P)$, consider a non-negative random variable $Z(\omega) \geq 0$ $\forall \omega \in \Omega$, with $E_P(Z) = 1$ (this implies that $P(\{\omega : Z(\omega) > 0\}) > 0$). More in general, if $E_P(Z) = \mu$, with $0 < \mu < \infty$, we can take $Z\prime = Z/\mu$ with $E_P(Z\prime) = 1$.

We define the new probability measure $Q : \mathcal{F} \to [0,1]$ as

$$Q(A) := E_P(Z\mathbf{1}_A) \quad \forall A \in \mathcal{F}$$

Q is a probability : clearly $Q(\Omega) = 1$, and finite additivity follows from the linearity of the expectation. σ-additivity follows by the monotone convergence theorem 4.1.1: when $A_n \uparrow \Omega$, also $Z(\omega)\mathbf{1}_{A_n}(\omega) \uparrow Z(\omega)$ P-almost surely, and

$$Q(A_n) = E_P\big(Z\mathbf{1}_{A_n}\big) \uparrow E_P(Z) = Q(\Omega) = 1 \ .$$

Theorem 4.2.1. *$\forall A \in \mathcal{F}$ $P(A) = 0 \implies Q(A) = 0$, and we say that Q is absolutely continuous with respect to P on $\mathcal{F}$, or P dominates Q on $\mathcal{F}$, and use the notation $Q \overset{\ll}{\mathcal{F}} P$.*

Proof. When $P(A) = 0$, $Z(\omega)\mathbf{1}_A(\omega) = 0$ P-almost surely, and $Q(A) = E_P\big(Z\mathbf{1}_A\big) = 0$.

Theorem 4.2.2. *When $X \in \mathcal{F}^+$,*

$$E_Q(X) = E_P(XZ),$$

and $X \in L^1(\Omega, \mathcal{F}, Q)$ if and only if $(XZ) \in L^1(\Omega, \mathcal{F}, P)$.

Proof. When $X(\omega) \in \mathcal{SF}^+$, the claim follows directly by using the definition and and the linearity of the expectation. When $0 \leq X \in \mathcal{F}^+$ there is an approximating sequence of simple random variables $\{X_n\} \subseteq \mathcal{SF}^+$ jolle $0 \leq X_n(\omega) \leq X(\omega)\ \forall\omega$. By appplying the Monotone convergence Theorem two times under the probability Q and the probability P it follows that $E_Q(X_n) \uparrow E_Q(X)$ and

$$E_Q(X_n) = \frac{E_P(X_n Z)}{E_P(Z)} \uparrow \frac{E_P(XZ)}{E_P(Z)} \quad \square$$

4.2.1 Likelihood ratio

We have constructed a new meare $Q \ll P$ on $\mathcal{F}$ by using a random variable $0 \leq Z \in L^1(\Omega, \mathcal{F}, P)$. It is possible to go in the opposite direction: $Q \overset{\ll}{\mathcal{F}} P$ there is a random variable $0 \leq Z(\omega) \in L^1(\Omega, \mathcal{F}, P)$ with $E_P(Z) = 1$ such that the change of measure formula $Q(A) = E_P(Z\mathbf{1}_A)$ holds $\forall A \in \mathcal{A}$.

Theorem 4.2.3. *(Radon-Nikodym Theorem) On a probability space $(\Omega, \mathcal{F})$ let P and Q be probability measures (more in general P or Q could be also σ-finite measures), such that $Q(A) = 0$ always when $A \in \mathcal{F}$ and $P(A) = 0$ (notation: $Q \overset{\mathcal{F}}{\ll} P$). In such case, there is a random variable $0 \leq Z(\omega) \in L^1(\Omega, \mathcal{F}, P)$ such that $E_P(Z) = 1$ and*

$$Q(A) = E_P(Z\mathbf{1}_A) \quad \forall A \in \mathcal{F}$$

$Z(\omega)$ is uniquely determined up to P-null events. We denote

$$Z(\omega) = \frac{dQ}{dP}(\omega)\ ,$$

which is called likelihood ratio or Radon-Nikodym derivative of Q w.r.t P on $\mathcal{F}$.

The R-N Theorem will be proved in the final chapters of the lecture notes by using martingale theory.

For random variables the change of measure formula takes the form

$$E_Q(X) = \int_\Omega X(\omega)Q(d\omega) = \int_\Omega X(\omega)\, \frac{dQ}{dP}(\omega)\ P(d\omega)$$

Definition 4.2.1. *On a probability space $(\Omega, \mathcal{F})$ the probability measures P and P' are singular (notation: $P \perp P'$), when there is an event $A \in \mathcal{F}$ such $P(A) = 0$ ja $P'(A) = 1$, (equivalently $P(A^c) = 1$ ja $P'(A^c) = 0$).*

Example 4.2.1. *On the probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{F} = \sigma(X)$, where $X(\omega)$ is a standard Gaussian random variable with $E(X) = 0, E(X^2) = 1$, and probability distribution*

$$P(X \in dx) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\frac{x^2}{2}\bigr) dx$$

We show later in the exercises that

$$\int_{\mathbb{R}} \exp\bigl(-\frac{x^2}{2}\bigr) dx = \sqrt{2\pi}.$$

Let P' another Gaussian probability distribution such that

$$P'(X_i \in dx) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\frac{(x-\mu)^2}{2}\bigr) dx$$

We compute the likelihood ratio

$$Z'(\omega) = \frac{dP'}{dP}(\omega) \quad \text{and} \quad Z(\omega) = \frac{dP}{dP'}(\omega) = \frac{1}{Z'(\omega)}$$

By the R-N theorem it follows that $Z'(\omega)$ is $\sigma(X)$-measurable, therefore there exists a Borel measurable map $z : \mathbb{R} \to \mathbb{R}^+$ such that $Z'(\omega) = z'(X(\omega))$.

Then, for all Borel measurable functions $f(x) \geq 0$

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp\bigl(-\frac{(x-\mu)^2}{2}\bigr) dx = E_{P'}(f(X)) = E_P(f(X)Z')$$
$$= E_P(f(X)z'(X)) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) z'(x) \exp\bigl(-\frac{x^2}{2}\bigr) dx \ ,$$

which implies

$$z'(x) = \exp\bigl(\mu x - \frac{1}{2}\mu^2\bigr),$$
$$Z'(\omega) = \exp\bigl(\mu X(\omega) - \frac{1}{2}\mu^2\bigr)$$

Since $E_P(Z') = 1$, it follows that

$$E_P\bigl(\exp(\mu X)\bigr) = \exp\bigl(\frac{1}{2}\mu^2\bigr)$$

4.2.2 Lebesgue decomposition

Let P and P' be probability measures on the probability space $(\Omega, \mathcal{F})$, for which we do not have $P \ll P'$ neither $P' \ll P$.

However $Q := \frac{1}{2}(P + P')$ is a probability measure which dominates both $P \ll Q$ and $P' \ll Q$.

By the Radon-Nikodym Theorem 4.2.3 it follows that both likelihood ratio

$$\zeta(\omega) := \frac{dP}{dQ}(\omega) \text{ and } \zeta'(\omega) := \frac{dP'}{dQ}(\omega) \ ,$$

exists in $L^1(\Omega, \mathcal{F}, Q)$ and are non negative with $E_Q(\zeta) = E_Q(\zeta\prime) = 1$.

Note that $\forall \omega \in \Omega$

$$\zeta(\omega) + \zeta'(\omega) = \frac{2dP}{d(P+P')}(\omega) + \frac{2dP'}{d(P+P')}(\omega) = 2\,\frac{d(P+P')}{d(P+P')}(\omega) = 2$$

and $\zeta(\omega) \geq 0, \zeta'(\omega) \geq 0$, it follows that

$$\zeta(\omega) \leq 2, \zeta'(\omega) \leq 2 \quad Q \text{ almost surely,} \quad \text{and} \quad Q\big(\{\omega : \zeta(\omega) = 0\} \cap \{\omega : \zeta'(\omega) = 0\}\big) = 0.$$

We define $\forall \omega \in \Omega$

$$Z(\omega) = \frac{dP}{dP'}(\omega) := \frac{\zeta(\omega)}{\zeta'(\omega)} \quad \text{and } Z'(\omega) = \frac{dP'}{dP}(\omega) := \frac{\zeta'(\omega)}{\zeta(\omega)} = \frac{1}{Z(\omega)}$$

where $0/0$ takes an arbitrary finite value, for example 0.

The change of measure formula generalizes as

$$E_{P'}(X) = E_P(XZ') + E_{P'}(X\mathbf{1}(\zeta = 0))$$

for $X \in \mathcal{F}^+$.

Proof

$$E_{P'}(X) = E_{P'}\big(X\{\mathbf{1}(\zeta > 0) + \mathbf{1}(\zeta = 0)\}\big) = E_Q\big(X\zeta'\mathbf{1}(\zeta > 0)\big) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big)$$
$$= E_Q\left(X\frac{\zeta'}{\zeta}\zeta\mathbf{1}(\zeta > 0)\right) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) = E_Q(XZ'\zeta) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big)$$
$$= E_P(XZ') + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) = E_P(XZ') + E_{P^\perp}(X)$$

where

$$P^{\perp}(d\omega) := \mathbf{1}(\zeta(\omega) = 0)P'(d\omega) \ ,$$

meaning that

$$P'(d\omega) = Z'(\omega)P(d\omega) + \mathbf{1}(\zeta(\omega) = 0)P'(d\omega) = Z'(\omega)P(d\omega) + P^{\perp}(d\omega)$$

P and $P^{\perp}$ are singular, because the event $A := \{\omega : \zeta(\omega) = 0\} \in \mathcal{F}$ satisfies

$$P(A) = 0 \text{ and } P^{\perp}(A) = P^{\perp}(\Omega)$$

Since $P^{\perp}(\Omega) + E_P(Z') = P'(\zeta = 0) + E_P(Z') = 1$, $P^{\perp}$ is a probability measure if and only if $P \perp P'$, (in such case $P^{\perp} = P'$). Also, $E_P(Z') \leq 1$ with $E_P(Z') = 1$ if and only if $P' \ll P$ (in such case $P^{\perp} = 0$).

Example: Conditional probability Let $B \in \mathcal{F}$ such that $P(B) > 0$, and let's change the measure by using the random variable $Z(\omega) = \mathbf{1}_B(\omega)/P(B) \geq 0$ with $E_P(Z) = 1$, obtaining

$$P(A|B) := \frac{E_P(\mathbf{1}_A \mathbf{1}_B)}{E_P(\mathbf{1}_B)} = \frac{P(A \cap B)}{P(B)} \ , \qquad A \in \mathcal{F}$$

The map $P(\ \cdot\ |B) : \mathcal{F} \to [0,1]$ is a probability measure, which is called conditional probability given the event B.

From De Finetti's point of view $P(A|B)$ is the price of the event A in a coherent pricing system P, when it is known that the event B has happened. The decomposition

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

is very useful to compute the probability of complex events:

Lemma 4.2.1. • *For a sequence of events $(A_i : i \in \mathbb{N}) \subseteq \mathcal{F}$ with $P(A_i \cap A_k) = 0$, $\forall i \neq j$, and $P(\bigcup_{i \in \mathbb{N}} A_i) = 1$, we have*

$$P(B) = \sum_{i \in \mathbb{N}} P(B|A_i)P(A_i)$$

- *When* $P(B\cap C)>0$, $P(A|B\cap C)=\frac{P(A\cap B|C)}{P(B|C)}$
-

$$P(A_1\cap A_2\cap\dots\cap A_n)=P(A_1)P(A_2|A_1)\dots P(A_n|\,A_1\cap A_2\cap\dots\cap A_{n-1})$$

We leave the proof as an exercise.

We also define the conditional expectation of a random variable $X(\omega)\geq 0$ given the event B with $P(B)>0$ as

$$E_P(X|B):=\frac{E_P\big(X\mathbf{1}_B\big)}{P(B)}=\int_\Omega X(\omega)P(d\omega|B)$$

where on the right hand side we integrate with respect to the conditional probabilility $P(d\omega\,|B)$.

Note that at this point the positivity assumption $P(B)>0$ is crucial for conditioning. How to extend the definition of conditional expectation to conditionally of an P-null event ? The answer will be given in the following chapters.

Chapter 5

Independence

Definition 5.0.2. • *On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $A, B \in \mathcal{F}$. We say that the events A and B are independent with respect to $\mathbb{P}$ (we use the notation $A \overset{\mathbb{P}}{\perp\!\!\!\perp} B$) when*

$$\text{either } \mathbb{P}(A)\mathbb{P}(B) = 0 \qquad \text{or} \qquad \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A),$$

so that the conditioning with respect to one event does not change the probability of the second event. When $\mathbb{P}(A|B) = \mathbb{P}(A)$ we have also $\mathbb{P}(B|A) = \mathbb{P}(B)$ and $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- *More in general the events in the collection $(A_j : j \in J)$ are mutually $\mathbb{P}$-independent when for every finite subset of indexes $I \subseteq J$,*

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

merkintä: $(A_j : j \in J) \overset{\mathbb{P}}{\perp\!\!\!\perp}$.

- *The random vectors $(X_j(\omega) : j \in J)$ with $X_j : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}^{d_j}, \mathcal{B}(\mathbb{R}^{d_j}))$ $j = 1, \dots, n$ are mutually $\mathbb{P}$-independent when for every finite subset $I = \{i_1, \dots, i_n\} \subseteq J$, $\forall B_i \in \mathbb{R}^{d_i}$ $i \in I$*

$$\mathbb{P}(X_{i_1} \in B_1, X_{i_2} \in B_2, \dots X_{i_n} \in B_n) = \mathbb{P}(X_{i_1} \in B_1)\mathbb{P}(X_{i_2} \in B_2) \dots \mathbb{P}(X_{i_n} \in B_n)$$

and we write $(X_j : j \in J) \overset{P}{\perp\!\!\!\perp}$.

- *The σ-algebrae $\mathcal{G}_1, \dots, \mathcal{G}_n \subseteq \mathcal{F}$ are $\mathbb{P}$-independent when $\forall A_i \in \mathcal{G}_i, i = 1, \dots, n$,*

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\dots\mathbb{P}(A_n),$$

and we write $(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_n) \overset{\mathbb{P}}{\perp\!\!\!\perp}$.

Remark 5.0.1.
- *Let $A, B \in \mathcal{F}$, when $A \overset{\mathbb{P}}{\perp\!\!\!\perp} B$, also $(A^c) \overset{\mathbb{P}}{\perp\!\!\!\perp} B$.*
- *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable map. When $(X_1, \dots, X_n) \overset{\mathbb{P}}{\perp\!\!\!\perp} Y$, it follows that $g(X_1, X_2, \dots, X_n) \overset{\mathbb{P}}{\perp\!\!\!\perp} Y$.*
- *When $A \in \mathcal{F}$ and $A \overset{\mathbb{P}}{\perp\!\!\!\perp} A$, it follows that either $\mathbb{P}(A) = 0$, or $\mathbb{P}(A) = 1$.*
- *In general from pairwise independence of a several events $A_i \overset{\mathbb{P}}{\perp\!\!\!\perp} A_j$ $\forall 1 \leq i \neq j \leq n$, mutual indepenence of the events $(A_1, A_2, \dots, A_n) \overset{\mathbb{P}}{\perp\!\!\!\perp}$ does NOT follow.*

Example 5.0.2. *Let $A, B \in \mathcal{F}$ such that $A \overset{\mathbb{P}}{\perp\!\!\!\perp} B$ ja $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$.*

Let $C = (A \cap B) \cup (A^c \cap B^c)$. Then also $C \overset{\mathbb{P}}{\perp\!\!\!\perp} A$ and $C \overset{\mathbb{P}}{\perp\!\!\!\perp} B$. You can check as an exercise that for each of the pairs $\{A, B\}$, $\{A, C\}$, $\{B, C\}$ the events are $\mathbb{P}$-independent, but as a triple $\{A, B, C\}$ are not $\mathbb{P}$-independent.

Definition 5.0.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $\mathcal{F}_i \subset \mathcal{F}$, $i = 1, \dots, n$, $\mathcal{G} \subset \mathcal{F}$, sub-$\sigma$-algebrae.*

We say that the σ-algebrae $\mathcal{F}_1, \dots, \mathcal{F}_n$ are conditionally $\mathbb{P}$-independent with respect to the $\mathcal{G}$ σ-algebra, if $\forall G \in \mathcal{G}$ with $\mathbb{P}(G) > 0$, and $A_i \in \mathcal{F}_i : i = 1, \dots, n$

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n | \, G) = \prod_{i=1}^{n} \mathbb{P}(A_i | G)$$

Lemma 5.0.2. *Random variables $X_1, \dots, X_n$ are $\mathbb{P}$-independent if and only if for any non-negative Borel measurable function $g_k : \mathbb{R} \to \mathbb{R}^+$, we have*

$$E_{\mathbb{P}}\big(g_1(X_1)\dots g_1(X_n)\big) = E_{\mathbb{P}}\big(g_1(X_1)\big)\dots E_{\mathbb{P}}\big(g_n(X_n)\big) \tag{5.0.1}$$

Proof. *The $\Longleftarrow$ implication follows by taking $g_k(x) = \mathbf{1}_{B_k}(x)$ with $B_k \in \mathcal{B}(\mathbb{R})$.*

In the other direction, if $X_1,\dots,X_n$ are $\mathbb{P}$-independent, (5.0.1) *is true when $g_k(x)=\mathbf{1}_{B_k}(x)$ with $B_k\in\mathcal{B}(\mathbb{R})$.*

When $g_k(x)=\sum_{i=1}^{m_k} y_{i,k}\mathbf{1}_{B_{i,k}}(x)$ is a simple measurable function,

$$
\begin{aligned}
E_{\mathbb{P}}\big(g_1(X_1)\dots g_n(X_n)\big)) &= E_{\mathbb{P}}\left(\prod_{k=1}^n\left\{\sum_{i=1}^{m_k} y_{i,k}\mathbf{1}_{B_{i,k}}(X_k)\right\}\right)\\
&= \sum_{i_1=1}^{m_1}\cdots\sum_{i_n=1}^{m_n} y_{i_1,1}\dots y_{i_n,n}\mathbb{P}\big(X_1\in B_{i_1,1},\dots,X_n\in B_{i_n,n}\big)\\
&= \sum_{i_1=1}^{m_1}\cdots\sum_{i_n=1}^{m_n} y_{i_1,1}\dots y_{i_n,n}\mathbb{P}\big(X_1\in B_{i_1,1}\big)\dots\mathbb{P}\big(X_n\in B_{i_n,n}\big)\\
&= \prod_{k=1}^n E_{\mathbb{P}}\left(\left\{\sum_{i=1}^{m_k} y_{i,k}\mathbf{1}_{B_{i,k}}(X_k)\right\}\right) = E_{\mathbb{P}}\big(g_1(X_1)\big)\dots E_{\mathbb{P}}\big(g_n(X_n)\big)
\end{aligned}
$$

In the general case, we construct sequences of simple Borel measurable function $\big(g_k^{(N)}(x): N\in\mathbb{N}\big)$ such that

$$0\le g_k^{(N)}(x)\uparrow g_k(x), \text{as } N\uparrow\infty,\ \forall k=1,\dots,n\ .$$

The product of these function also is also non-decreasing w.r.t. N,

$$\big(g_1^{(N)}(x_1)\dots g_n^{(N)}(x_n)\big)\uparrow\big(g_1(x_1)\dots g_n(x_n)\big),$$

and the claim follows by the monotone convergence Theorem.

5.0.3 Lovasz local lemma [1]

We show in this section how powerful results one can get by using inductively the elmentary definition of conditional probability.

When the events $A_1,\dots,A_n$ are $\mathbb{P}$-independent, also the complements are $\mathbb{P}$-independent, and

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right)=\prod_{i=1}^n(1-\mathbb{P}(A_i))$$

Lozasz local lemma give a similar lower bound when the dependence between the events is "weak enough".

[1] This section is not in the exam program

Lemma 5.0.3. *(Paul Erdös ja Lazlo Lovasz 1975) Let* $\{ A_1, \dots, A_n\} \subseteq \mathcal{F}$ *events in the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, and* (V, E) *an undirected graph with nodes* $V = \{ 1, 2, \dots, n\}$ *and edges* $E \subseteq V \times V$*.*

We assume that the edges have directions, which means that $\forall i, j \in V$ $(i, j) \in E \iff (j, i) \in E$ *and we denote* $i \sim j$*. Also we assume that there are not self-loops,* $(i, i) \notin E$ *(notation:* $i \not\sim i$*), and there exists a vector* $(x_1, \dots, x_n) \in [0, 1)^n$ *such that*

$$\mathbb{P}\left(A_i \middle| \bigcap_{j\in S} A_j^c\right) \leq x_i \prod_{j\sim i}(1 - x_j)$$

holds $\forall i \in V, S \subseteq V$ *with* $\{ i\} \cup \{j : j \sim i\} \subseteq S^c$ *and* $\mathbb{P}\left(\bigcap_{j\in S} A_j^c\right) > 0\,.$

Then, $\forall S, S' \subseteq V$ *such that* $S \cap S' = \emptyset$*,*

$$\mathbb{P}\left(\bigcap_{i\in S} A_i^c \middle| \bigcap_{j\in S'} A_j^c\right) \geq \prod_{i\in S}(1 - x_i) > 0$$

In particular, when $S' = \emptyset$*,* $\mathbb{P}\left(\bigcap_{i\in S} A_i^c\right) \geq \prod_{i\in S}(1-x_i) > 0$*, which implies* $\bigcap_{i\in S} A_i^c \neq \emptyset$*.*

Proof We use induction with respect to the total cardinality $n = |S| + |S'|$. The claim is trivial when $n = 1$, meaning that $|S| = 1, S' = \emptyset$. We show first that the claim is true when $|S| = 1$ and $S = \{i\}$ with $i \notin S'$. Let $S' = S'' \cup S'''$ where $S'' = S' \cap \{ j \sim i\}$, $S''' = S' \cap \{ j : j \not\sim i\}$. In such case

$$\mathbb{P}\left(A_i \middle| \bigcap_{j\in S'} A_j^c\right) = \frac{\mathbb{P}\left(A_i, \bigcap_{j\in S''} A_j^c \middle| \bigcap_{j\in S'''} A_j^c\right)}{\mathbb{P}\left(\bigcap_{j\in S''} A_j^c \middle| \bigcap_{j\in S'''} A_j^c\right)}$$

where

$$\mathbb{P}\left(A_i, \bigcap_{j\in S''} A_j^c \middle| \bigcap_{j\in S'''} A_j^c\right) \leq \mathbb{P}\left(A_i \middle| \bigcap_{j\in S'''} A_j^c\right) \leq x_i \prod_{j\sim i}(1 - x_j)$$

and since $|S''| + |S'''| = |S'| < 1 + |S'| = n$, by the induction hypothesis it follows that

$$\mathbb{P}\left(\bigcap_{j\in S''} A_j^c \middle| \bigcap_{j\in S'''} A_j^c\right) \geq \prod_{j\in S''}(1 - x_j) \geq \prod_{j\sim i}(1 - x_j)\ ,$$

which implies $\mathbb{P}\left(A_i^c \middle| \bigcap_{j\in S'} A_j^c\right) \geq (1-x_i)$.

When $\{ i\} \subsetneq S$ and $S' \cap S = \emptyset$, we split S as $S = \{i\} \cup (S \setminus \{ i\})$,

$$\mathbb{P}\left(\bigcap_{k\in S} A_k^c \middle| \bigcap_{j\in S'} A_j^c\right) = \mathbb{P}\left(A_i^c \middle| \bigcap_{j\in S'\cup S\setminus\{ i\}} A_j^c\right)\mathbb{P}\left(\bigcap_{k\in S\setminus\{ i\}} A_k^c \middle| \bigcap_{j\in S'} A_j^c\right)$$
$$\geq (1-x_i) \prod_{k\in S\setminus\{ i\}} (1-x_k) = \prod_{i\in S}(1-x_i) > 0 \ ,$$

where where the induction hypothesis holds since
$|S \setminus \{ i\}| + |S'| = |S| + |S'| - 1$ □

Corollary 5.0.1. *Let (V, E) be a dependence graph, where*
$\forall i \in V, S \subseteq V$ *such that* $S^c \subseteq \{ i\} \cup \{j : j \sim i\}$,
the events A_i and $\left(\bigcap_{j\in S} A_j\right)$ are $\mathbb{P}$-independent.
In other words, A_i is $\mathbb{P}$-independent from the σ-algebra $\sigma(A_j : j \neq i$ and $j \not\sim i)$.

1. *If there are $x_i \in [0, 1), i = 1, \ldots, n$ such that*

$$\mathbb{P}(A_i) \leq x_i \prod_{j\sim i}(1-x_j) \ ,$$

   it follows that $\mathbb{P}\left(\bigcap_{i\in S} A_i^c\right) \geq \prod_{i\in S}(1-x_i), \quad \forall S \subseteq V$,
   in particular $\left(\bigcap_{i\in S} A_i^c\right) \neq \emptyset$.

2. *In particular $\#\{ j : j \sim i\} \leq d$, and $\mathbb{P}(A_i) \leq \exp(-1)/(d+1), \forall i \in V$, it follows that*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i^c\right) \geq \left(1 - \frac{1}{d+1}\right)^n > 0$$

Proof

1. The Lovasz lemma assumptions follow from the $\mathbb{P}$-independence.

2. We need to find an x such that $p \le x(1-x)^d$. For $x=(1+d)^{-1}$, the first part of the assumption follows since

$$\mathbb{P}(A_i) \le \frac{\exp(-1)}{d+1} \le \frac{1}{d+1}\left(1-\frac{1}{d+1}\right)^d = x(1-x)^d \quad \square$$

where $\exp(-1) \le \Big(1-(d+1)^{-1}\Big)^d \ \square$

Example 5.0.3. *A data-network has n pairs of nodes, and between each pair of nodes p there is a set of J_p possible connections.*

Assume that $|J_p| = m$ for every pair p, and for different pairs $p \neq p'$ and for every connection $r \in J_p$ we have

$$\Big|\{r' \in J_{p'} \text{ such that the connections } r \text{ and } r' \text{ do not intersect } \}\Big| \le k$$

When k is "not too big", it is possible to connect every pair of nodes, by using connections which do not intersect each other.

Proof *For every pair of nodes p choose by sampling independently from the uniform distribution a random connnection R_p from the set J_p of possible connections. For every pair of node-pairs p, p' with $p \neq p'$ we introduce the event*

$$A_{p,p'} = \{ \text{ the random connections } R_p \text{ and } R_{p'} \text{ intersect each other } \}.$$

Clearly $\mathbb{P}(A_{p,p'}) \le k/m$. Note also that

$$A_{p,p'} \perp\!\!\!\perp \sigma\big(A_{q,q'} : q \neq p, p', \text{ and } q' \neq p, p'\big)$$

It follows that there are $d = 2(n-1)$ events $A_{p,r}$ such that $r \neq p$ and $A_{r',p'}$ jolla $r' \neq p'$, which are $\mathbb{P}$-dependent from the event $A_{p,p'}$. In other words the degree (number of neighbours) of $A_{p,p'}$ in the dependence graph of the events is given by $d = 2(n-1)$.

When

$$\mathbb{P}(A_{p,p'}) \le k/m \le \frac{\exp(-1)}{d+1} \le (d+1)^{-1}\Big(1-(d+1)^{-1}\Big)^d ,$$

which means $k \le m\exp(-1)/(2n-1)$, Lovasz local lemma applies and

$$\mathbb{P}\Big(\bigcap_{p,p'} A^c_{p,p'}\Big) > \Big(1-(d+1)^{-1}\Big)^{n(n-1)/2} > 0 ,$$

which implies $\left(\bigcap_{p,p'} A^c_{p,p'}\right) \neq \emptyset$ □

5.1 Borel Cantelli lemma

Lemma 5.1.1. *(Borel-Cantelli I).*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \implies \mathbb{P}\big(\limsup_n A_n\big) = \mathbb{P}\big(\{\omega : \omega \in A_n \textit{ for infinitely many } n\ \}\big) = 0$$

Proof.

$$\limsup_n A_n = \bigcap_n \bigcup_{k>n} A_k \subseteq \bigcup_{k>m} A_k \quad \forall m$$

Then

$$\mathbb{P}(\limsup_n A_n) \leq \sum_{k \geq m} \mathbb{P}(A_k) \to 0 \text{ as } m \to \infty \ ,$$

since the series is convergent □

Lemma 5.1.2. *Generalized Borel-Cantelli lemma I (Barndorff-Nielsen,1961)*
When $\liminf_n \mathbb{P}(A_n) = 0$ *and* $\sum_{n=1}^{\infty} \mathbb{P}(A_n \cap A^c_{n+1}) < \infty$*,*
it follows that $\mathbb{P}(\limsup_n A_n) = 0$*.*

Proof. Note that

$$\limsup_n (A_n \cap A^c_{n+1}) = (\limsup_n A_n) \cap (\limsup_n A^c_n) \ ,$$

because if $(A_n \cap A^c_{n+1})$ happens infinitely often, also A_n and A^c_n do happen infinitely often. On the other hand, if both A_n and A^c_n happen infinitely often, necessarily also $(A_n \cap A^c_{n+1})$ infinitely often, otherwise we would have either $(\liminf_n A_n) = (\limsup_n A^c_n)^c$ or $(\liminf_n A^c_n) = (\limsup A_n)^c$. Therefore

$$\mathbb{P}(\limsup_n A_n) = \mathbb{P}\big(\limsup_n (A_n \cap A^c_{n+1})\big) + \mathbb{P}\big((\limsup_n A_n) \cap (\limsup_n A^c_n)^c\big) = 0$$

since by assumption and by the first Borel Cantelli lemma 5.1.1 it follows that

$$\mathbb{P}\big(\limsup_n (A_n \cap A_{n+1}^c)\big) = 0,$$

and $\mathbb{P}((\limsup_n A_n^c)^c) = \mathbb{P}(\liminf_n A_n) \leq \liminf_n \mathbb{P}(A_n) = 0$ □.

Proposition 5.1.1. *(Kolmogorov 0-1 law) On a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, let* $\{\mathcal{F}_k : k \in \mathbb{N}\}$ *a sequence of* σ*-algebrae such that* $\mathcal{F}_k \subseteq \mathcal{F}$*. Let*

$$\mathcal{F}^\infty = \bigvee_{k \in \mathbb{N}} \mathcal{F}_k, \quad \mathcal{T}_n = \bigvee_{k \geq n} \mathcal{F}_k, \quad \mathcal{T} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_n$$

$\mathcal{T}$ *is called tail* σ*-algebra.*

When the σ*-algebrae* $\{\mathcal{F}_k : k \in \mathbb{N}\}$ *are* $\mathbb{P}$*-independent, meaning that*

$$\mathbb{P}(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_n}) = \prod_{i=1}^n \mathbb{P}(A_{k_i}) \quad \forall n, k_1, \ldots, k_n \in \mathbb{N}, A_{k_i} \in \mathcal{F}_i \ .$$

it follows that $\mathcal{T}$ *is* $\mathbb{P}$*-trivial, meaning that* $\mathbb{P}(A) \in \{0,1\}$ $\forall A \in \mathcal{T}$*.*

Proof. Let $A \in \mathcal{T}$, which means that $A \in \mathcal{T}_{n+1}$ $\forall n \in \mathbb{N}$. Since the σ-algebrae $\mathcal{F}_k$ are $\mathbb{P}$-independent, also $\mathcal{F}_1 \vee \mathcal{F}_2 \vee \mathcal{F}_n \overset{\mathbb{P}}{\perp\!\!\!\perp} \mathcal{T}_{n+1}$ and

$$A \overset{\mathbb{P}}{\perp\!\!\!\perp} \left(\bigvee_{k=1}^n \mathcal{F}_k \right) \quad \forall n \Longrightarrow A \overset{\mathbb{P}}{\perp\!\!\!\perp} \left(\bigvee_{k=1}^\infty \mathcal{F}_k \right) = \mathcal{F}_\infty \ ,$$

which means that $\mathcal{T} \overset{\mathbb{P}}{\perp\!\!\!\perp} \mathcal{F}^\infty \supseteq \mathcal{T}$. This means that $A \in \mathcal{T}$ is $\mathbb{P}$-independent from itself (!),

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A). \tag{5.1.1}$$

The only possible solutions of (5.1.1) are $\mathbb{P}(A) = 0$ and $\mathbb{P}(A) = 1$ □

Example 5.1.1. *Let* $(A_n : n \in \mathbb{N})$ *be an event sequence with* $A_n \in \mathcal{F}_n$*. We show* $(\limsup_n A_n)$ *is in the tail* σ*-algebra* $\mathcal{T}$*. In fact* $A := \limsup_n A_n = \bigcap_n B_n$ *with* $B_n = \bigcup_{k \geq n} A_k \geq \bigcup B_{n+1} \downarrow A$*, with* $B_n \in \mathcal{T}_N$ $\forall n \geq N$*.*

Since the B_n *are nested,* $A = \bigcap_{n \geq 0} B_n = \bigcap_{n \geq N} B_n \in \mathcal{T}_N$ $\forall N \in \mathbb{N}$*, therefore* $A \in \bigcap_N \mathcal{T}_N = \mathcal{T}$*.*

By taking the complement it follows also that $\liminf_n A_n = (\limsup_n A_n^c)^c \in \mathcal{T}$

Lemma 5.1.3. *(Borel-Cantelli II)*

Let $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{F}$ be a sequence $\mathbb{P}$-independent events, meaning that

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k}) \quad \forall n, i_1, \ldots, i_n \in \mathbb{N} \ .$$

Then

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Longleftrightarrow \mathbb{P}\big(\limsup_n A_n\big) = \mathbb{P}\big(\{\omega : \omega \in A_n \textit{ infinitely often } \}\big) = 1$$

Proof. Let $\mathcal{F}_n = \sigma(A_n) = \{A_n, A_n^c, \Omega, \emptyset\}$, $n \in \mathbb{N}$. Because the σ-algebrae $(\mathcal{F}_n)_{n\in\mathbb{N}}$ are $\mathbb{P}$-independent, and $(\limsup_n A_n) \in \mathcal{T}$, it is a P-trivial event, and by the Kolmogorov 0-1 law 5.1.1 $\mathbb{P}(\limsup_n A_n) \in \{0, 1\}$.

When $\mathbb{P}(\limsup_n A_n) = 1$, by the first Borel Cantelli lemma it follows that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$.

When $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$,

$$\mathbb{P}((\limsup_n A_n)^c) = \mathbb{P}(\liminf_n A_n^c) = \mathbb{P}\big(\bigcup_n \bigcap_{k\geq n} A_k^c\big) \ ,$$

and by using σ-additivity together with the inequality $(1 - x) \leq \exp(-x)$ it follows that

$$\mathbb{P}\big(\bigcap_{k\geq n} A_k^c\big) = \lim_{N\uparrow\infty} \mathbb{P}\big(\bigcap_{N\geq k\geq n} A_k^c\big) = \lim_{N\uparrow\infty} \prod_{N\geq k\geq n} (1 - \mathbb{P}(A_k)) \leq$$
$$\lim_{N\uparrow\infty} \exp\Big(-\sum_{k=n}^{N} \mathbb{P}(A_k)\Big) = \exp\Big(-\sum_{k=n}^{\infty} \mathbb{P}(A_k)\Big) = \exp(-\infty) = 0 \quad \forall n \in \mathbb{N}$$

which implies $\mathbb{P}((\limsup_n A_n)^c) = 0$ □

Remark 5.1.1. *This second Borel Cantelli lemma does not apply without the $\mathbb{P}$-independence assumption for the events. For example just take an event $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$ and set $A_n = A$ $\forall n \in \mathbb{N}$, then $\sum_n \mathbb{P}(A_n) = \infty$ but $\mathbb{P}(\limsup_n A_n) = \mathbb{P}(A) \notin \{0, 1\}$. However it is possible to generalize the lemma assuming that the $\mathbb{P}$-dependence is "weak enough".*

Lemma 5.1.4. *(Generalized Borel-Cantelli II)*

When $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *and*

$$\liminf_{n\to\infty} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{P}(A_i \cap A_j)}{\left(\sum_{i=1}^{n} \mathbb{P}(A_i)\right)^2} = c < \infty$$

ot follows that $\mathbb{P}(\limsup A_n) \geq 1/c$.

Proof. We prove this statement in a following chapter by using the Cauchy-Schwartz inequality.

Chapter 6

Stochastic convergence

6.1 A short dictionary of functional analysis

1. A topological space is a pair $(E, \mathcal{T})$, where the topology $\mathcal{T} \subseteq 2^E$ is the collections of the open sets in E. E itself and the empty set $\emptyset$ are open. A topology is closed with respect to finite intersections and arbitrary (also uncountable) unions. The complement of an open set is closed. The space E itself is both open and closed.

   Note that it is possible to equip the same spaces with different topologies. When $\mathcal{T}' \supset \mathcal{T}''$ are topologies on the same space E, we say that the $\mathcal{T}'$ is stronger than $\mathcal{T}''$.

2. Let $\{x_n : n \in \mathbb{N}\} \subseteq E$. We say that $x_n \to x$ in the topology $\mathcal{T}$ if

   $\forall U \in \mathcal{T}$ with $x \in U$, $\exists\, n_U$ such that $x_n \in U\ \forall n \geq n_U$.

3. Let $f : (E, \mathcal{T}) \to (E', \mathcal{T}')$ be a function between topological spaces. We say that f is continuous when, $\forall V \in \mathcal{T}'$, the counterimage $f^{-1}(V) = \{\, x \in E : f(x) \in V\}$ is in $\mathcal{T}$.

4. $d : E \times E \to [0, +\infty]$ is a metric (=distance) when

   - $d(e, e') = 0$ if and only if $e = e'$
   - $d(e, e') = d(e', e)$ (symmetry)

- $d(e, e'') \leq d(e, e') + d(e', e'')$ (triangle inequality)

5. A topological space $(E, \mathcal{T})$ is metric if and only if there exists a metric $d : E \times E \to [0, +\infty]$, such that the open balls in that metric generate the topology $\mathcal{T}$ in the following sense:

   $B(e, r) = \{e' : d(e, e') < r\} \in \mathcal{T} \; \forall \; e \in E, r > 0$

   and $\forall x \in U, x \in E, U \in \mathcal{T} \; \exists r > 0$ such that $x \in B(e, r) \subseteq U$.

   It is possible that different metrics generate the same topology.

   In a metric space , $\{x_n : n \in \mathbb{N}\} \subseteq E$ is a *Cauchy sequence* when $\forall \varepsilon > 0$ there is n_ε such that $d(x_n, x_m) < \varepsilon \; \forall n, m \geq n_\varepsilon$.

   We say that the metric space (E, d) is *complete* if every Cauchy sequence $\{x_n : n \in \mathbb{N}\} \subseteq E$ has a limit $x \in E$.

6. Let E be a real vector space, where we have vector addition $E \times E \to E$ with $(x, y) \mapsto (x + y)$ and scalar multiplication $\mathbb{R} \times E \to E$ with $(\lambda, x) \mapsto \lambda x$ such that when $\lambda \in \mathbb{R}$ and $x, y \in E$, also $\lambda x \in E$ and $(x + y) \in E$.

   The map $\| \cdot \| : E \to [0, +\infty)$ is a *norm* when

   i) $\| x \| = 0 \in \mathbb{R} \iff x = \mathbf{0} \in E$ ii) $\| \lambda x \| = |\lambda| \; \| x \|$

   iii) $\| x + y \| \leq \| x \| + \| y \|$.

   All normed vector spaces are metric spaces, with the distance $d(x, y) := \| x - y \|$, which generate the topology of the convergence in norm.

7. A pre-Hilbert space is a normed space where the norm is derived from a scalar product $\langle \cdot, \cdot \rangle$ where $\| x \|^2 := \langle x, x \rangle$.

   An $\mathbb{R}$-valued scalar product is symmetric $\langle x, y \rangle = \langle y, x \rangle$,

   bilinear $\langle \lambda x + \lambda' x', y \rangle = \lambda \langle x, y \rangle + \lambda' \langle x', y \rangle$, and positive $\langle x, x \rangle \geq 0$, with $\langle x, x \rangle = 0 \iff 0$.

8. A complete normed space is called Banach space and a complete pre-Hilbertin space is called Hilbert space.

Definition 6.1.1. *Let* $X(\omega), X_n(\omega), n \in \mathbb{N}$ *random variables on* $(\Omega, \mathcal{F}, \mathbb{P})$. *We say that the sequence* (X_n) *converges stochastically (or in probability) to* X, *(notation:*$X_n \xrightarrow{\mathbb{P}} X$*) when,* $\forall \varepsilon > 0$,

$$\mathbb{P}\big(\{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\big) \to 0, \quad \text{as } n \to \infty.$$

Stochastic convergence is weaker than the almost sure convergence.

Proposition 6.1.1. *1. When* $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ $\mathbb{P}$*-almost surely, also* $X_n \xrightarrow{\mathbb{P}} X$ *(in probability).*

2. *When* $X_n \xrightarrow{\mathbb{P}} X$ *(in probability), there exists deterministic subsequence* $\{n(k) : k \in \mathbb{N}\}$ *such that*

   $\lim_{k\to\infty} X_{n(k)}(\omega) = X(\omega)$ $\mathbb{P}$*-almost surely,*

3. $X_n \xrightarrow{\mathbb{P}} X$ *if and only if for every subsequence* $\{n(k)\}$ *there is a further subsequence* $\{n(k_l)\}$ *such that* $X_{n(k_l)}(\omega) \to X(\omega)$ $\mathbb{P}$*-almost surely as* $l \to \infty$.

Proof. We can asumme that $X(\omega) = 0$, otherwise we consider $X_n(\omega) = X_n(\omega) - X(\omega)$.

1. $X_n(\omega) \to 0$ $\mathbb{P}$-almost surely if and only if

$$\mathbb{P}\Bigg(\bigcap_n \bigcup_m \bigcap_{k\geq m} \{\omega : |X_k(\omega)| < n^{-1}\}\Bigg) = 1$$
$$\iff \forall n \in \mathbb{N}, \quad \mathbb{P}\Bigg(\liminf_k \{\omega : |X_k(\omega)| < n^{-1}\}\Bigg) = 1 \ .$$

   By using Fatou lemma, $\forall n$

$$1 = \mathbb{P}\Bigg(\liminf_k \{\omega : |X_k(\omega)| < n^{-1}\}\Bigg) \leq \liminf_k \mathbb{P}\Bigg(\{\omega : |X_k(\omega)| < n^{-1}\}\Bigg) = 1$$
$$\iff 0 = \limsup_k \mathbb{P}\Bigg(\{\omega : |X_k(\omega)| > n^{-1}\}\Bigg) = \lim_k \mathbb{P}\Bigg(\{\omega : |X_k(\omega)| > n^{-1}\}\Bigg) \ .$$

2. From convergence in probability it follows that there is a sequence k_n such that

$$\mathbb{P}\Bigg(\{\omega : |X_l(\omega)| > n^{-1}\}\Bigg) < 2^{-n}, \quad \forall l \geq k_n \ .$$

Since

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\{\omega : |X_{k_n}(\omega)| > n^{-1}\} \Big) < \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty$$

by using Borel-Cantelli lemma 5.1.1 it follows that

$$\begin{aligned} 0 &= \mathbb{P}\Big(\limsup_n \{\omega : |X_{k_n}(\omega)| > n^{-1}\} \Big) \\ &\geq \mathbb{P}\Big(\limsup_n \{\omega : |X_{k_n}(\omega)| > N^{-1}\} \Big) = 0 \quad \forall N \in \mathbb{N} \,, \end{aligned}$$

which implies

$$\begin{aligned} 1 &= \mathbb{P}\Big(\bigcap_N \liminf_n \{\omega : |X_{k_n}(\omega)| \leq N^{-1}\} \Big) \\ \iff & \quad X_{k_n}(\omega) \to 0 \quad \mathbb{P}\text{-almost surely.} \end{aligned}$$

3. Let $X(\omega) = 0$ and by contradiction assume that X_n does not converge to zero in probability: there would be some $\varepsilon > 0$ and a subsequence $n(k) \uparrow \infty$ as $k \uparrow \infty$ such that

$$\mathbb{P}(|X_{n(k)}| > \varepsilon) \geq \varepsilon > 0 \quad \forall k \,.$$

This gives a contradiction since by assumption where would be a subsequence $n(k_l)$ with $X_{n(k_l)}(\omega) \to 0$ $\mathbb{P}$-almost surely and in probability, such that

$$0 < \varepsilon \leq P(|X_{n(k_l)}| > \varepsilon) \to 0 \quad \text{ as } l \to \infty \quad \square$$

Remark: When X_n is convergent in probability under $\mathbb{P}$, for every subsequence (n_k) there is a further subsequence (n_{k_l}) and an event $N \subseteq \Omega$ with $\mathbb{P}(N) = 0$, such that $X_{n_{k_l}}(\omega) \to 0 \; \forall \omega \in N^c$. In general the $\mathbb{P}$-null set N depends on the subsequence (n_k). Since the collection of subsequences is not countable, we cannot use σ-additivity to conclude that there would be also a $\mathbb{P}$-null set N such that simultaneously for all subsequences (n_k) there are further subsequnces (n_{k_l}) such that $X_{n_{k_l}}(\omega) \to 0 \; \forall \omega \in N^c$. In that case $X_n(\omega)$ would be $\mathbb{P}$-almost surely convergent.

Proposition 6.1.2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function and $X_n(\omega), X(\omega) \in \mathbb{R}^d$ random vectors such that*

$$|X_n - X| \xrightarrow{\mathbb{P}} 0 \quad \textit{in probability as } n \to \infty \ ,$$

then also

$$f(X_n) \xrightarrow{P} f(X) \quad \textit{in probability as } n \to \infty \ .$$

Proof: Let $(n_k : k \in \mathbb{N})$ an arbitrary index subsequence. By the characterization (6.1.1.3), there is a further subsequence $(n_{k(\ell)} : \ell \in \mathbb{N})$ such that

$$X_{n_{k(\ell)}}(\omega) \longrightarrow X(\omega) \quad \mathbb{P}\text{-almost surely as } \ell \to \infty.$$

Because f is continuous

$$f\big(X_{n_{k(\ell)}}(\omega)\big) \longrightarrow f\big(X(\omega)\big) \quad \mathbb{P} \text{ almost surely as } \ell \to \infty$$

and the claim follows from the characterizations (6.1.1.3).

Example 6.1.1. *We show that convergence in probability is stricly weaker than almost sure convergence: Let $\Omega = (0,1]$ equipped with the Borel σ-algebra $\mathcal{F} = \mathcal{B}((0,1])$ and the uniform probability, with $\mathbb{P}((0,t]) = t$ for $t \in (0,1]$.*

We introduce the sequence of random variables

$$X_{n,k}(\omega) = \mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(\omega) \quad k = 0, 1, \dots, (2^n - 1)$$

where we order the indexes as it follows : $(n,k) \geq (m,h)$ if and only if $n > m$ or $n = m$ and $k \geq h$.

It follows that $\forall \omega \in (0,1]$, as $(n,k) \to \infty$ according to the ordering,

$$\liminf_{n,k\to\infty} X_{n,k}(\omega) = 0 \neq \limsup_{n,k\to\infty} X_{n,k}(\omega) = 1, \textit{ and}$$
$$\mathbb{P}(\{\omega : X_{n,k} > 1/2\}) = \mathbb{P}((k2^{-n}, (k+1)2^{-n}]) = 2^{-n} \to 0 \textit{ as } n \to \infty \ .$$

Problem 6.1.1. *Find for the sequence $(X_{n,k}(\omega), n \in \mathbb{N}, 0 \leq k \leq 2^n)$ a subsequence $(X_{n(l),k(l)} : l \in \mathbb{N})$ such that $X_{n(l),k(l)}(\omega) \to 0$ $\mathbb{P}$-almost surely as $l \to \infty$.*

Theorem 6.1.1. *On the space of random variables $L^0(\Omega, \mathcal{F}, \mathbb{P})$, the topology of convergence in probability is metric.*

$$X_n \xrightarrow{\mathbb{P}} X \iff d(X, X_n) \to 0, \text{ where}$$
$$d(X,Y) = d(X-Y,0) = E_{\mathbb{P}}\left(\frac{|X-Y|}{1+|X-Y|}\right) \text{ or}$$
$$d(X,Y) = d(X-Y,0) = E_{\mathbb{P}}\big(1 \wedge |X-Y|\big)$$

Proof. Assume that $X_n \xrightarrow{\mathbb{P}} X = 0$. Then, $\forall \varepsilon > 0$,

$$\frac{|X_n|}{(1+|X_n|)} \le \frac{|X_n|}{(1+|X_n|)}\mathbf{1}(|X_n| > \varepsilon) + \varepsilon \mathbf{1}(|X_n| \le \varepsilon) \le \mathbf{1}(|X_n| > \varepsilon) + \varepsilon,$$
$$d(X_n, 0) \le \mathbb{P}(|X_n| > \varepsilon) + \varepsilon < 2\varepsilon$$

for every n large enough.

For the implication in the other direction, note that the map $f(x) = x/(1+x)$ is strictly increasing with respect to x on $\mathbb{R}^+ = \{\, x \in \mathbb{R} : x > 0\}$, since it is differentiable with derivative $(f'(x) = (1+x)^{-2})$, and $\forall \varepsilon > 0$ we have

$$\frac{\varepsilon}{1+\varepsilon}\mathbf{1}(|X_n| > \varepsilon) \le \frac{|X_n|}{1+|X_n|}\mathbf{1}(|X_n| > \varepsilon) \le \frac{|X_n|}{1+|X_n|}$$
$$\frac{\varepsilon}{1+\varepsilon}P(|X_n| > \varepsilon) \le d(|X_n|, 0) \to 0 \text{ as } n \to \infty$$

To $d(X,Y)$ satisfies the triangle inequality and it is a metric:

$$\frac{|X-Y|}{1+|X-Y|} \le \frac{|X-Z| + |Z-Y|}{1+|X-Z|+|Z-Y|} \le \frac{|X-Z|}{1+|X-Z|} + \frac{|Z-Y|}{1+|Z-Y|}$$

and by taking the expectation with respect to $\mathbb{P}$ it follows that $d(X,Y) \le d(X,Z) + d(Z,Y)$ □

Convergence in L^p-norm with $p > 0$ implies stochastic probability We also discuss the connection with the convergence in $L^p(\Omega, \mathcal{F}, \mathbb{P})$.

For $p \ge 0$, we define the space

$$L^p(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X \in L^0(\Omega, \mathcal{F}, \mathbb{P}) : \| X \|_{L^p} := E_{\mathbb{P}}\big(|X|^p\big)^{1/p} < \infty \right\}$$

and we say that $X_n \stackrel{L^p}{\to} X$ if and only if $X_n - X \stackrel{L^p}{\to} 0$

$$E_P\big(|X_n - X|^p\big) \to 0$$

By Chebychev inequality, convergence in L^p is stronger than convergence in probability, since $\forall \eta > 0$

$$\mathbb{P}\big(|X_n - X| > \eta\big) = \mathbb{P}\big(|X_n - X|^p > \eta^p\big) \le \eta^{-p} E_{\mathbb{P}}\big(|X_n - X|^p\big) \to 0$$

when $X_n \stackrel{L^p}{\to} X$.

We define also the space L^p with $p = \infty$ as

$$L^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \Big\{ X \in L^0(\Omega, \mathcal{F}, \mathbb{P}) : \parallel X \parallel_{L^\infty} := \text{esssup}_{\mathbb{P}}\{|X(\omega)|\} \ < \infty \Big\}$$

where

$$\parallel X \parallel_{L^\infty} := \text{esssup}_{\mathbb{P}}\{|X(\omega)|\} = \inf\big\{ \ K \ge 0 : \mathbb{P}(\{\omega : |X(\omega)| \le K) = 1 \big\}$$

is the essential supremum of $X(\omega)$ under $\mathbb{P}$. $X \in L^\infty(\mathbb{P})$ means that X is essentially bounded w.r.t. $\mathbb{P}$, meaning that it is bounded by some constant K outside $\mathbb{P}$-null set, and the norm $\parallel X \parallel_{L^\infty}$ is the best possible constant.

Note that since the map $K \mapsto \mathbb{P}(|X| \le K)$ is right continuous, so as $K \downarrow \parallel X \parallel_{L^\infty} \mathbb{P}(|X| \le K) \downarrow \parallel X \parallel_{L^\infty}$.

We show that $\parallel X_n \parallel_{L^\infty} \to 0$ implies $X_n(\omega) \to 0$ $\mathbb{P}$-almost surely, which in turn is stronger than convergence in probability. In fact convergence in $L^{(\infty)}$-norm is equivalent to uniform convegence of $X_n(\omega)$ to zero outside a $\mathbb{P}$-null set.

Note that since the map $K \mapsto \mathbb{P}(|X| \le K)$ is right continuous, so as $K \downarrow \parallel X \parallel_{L^\infty} 1 = \mathbb{P}(|X| \le K) \downarrow \parallel X \parallel_{L^\infty}$, so that $\mathbb{P}(|X| \le \parallel X \parallel_{L^\infty}) = 1$.

When $\parallel X_n \parallel_{L^\infty} \to 0$, $\forall m \in \mathbb{N}\ \exists N = N(m) : \forall n \ge N\ \mathbb{P}(|X_n| \le 1/m) = 1$ which implies $\mathbb{P}(\limsup_n |X_n| \le 1/m) = 1$, and since the countable intersection of $\mathbb{P}$-almost sure events is a $\mathbb{P}$-almost sure event, it follows that $\limsup_n |X_n(\omega)| = 0$, $\mathbb{P}$-almost surely, which is equivalent to $X_n(\omega) \to 0$ $\mathbb{P}$-almost surely.

Note also that $\parallel X_n \parallel_{L^\infty} \to 0$ implies also $\parallel X_n \parallel_{L^\infty} \to 0$ $\forall p > 0$, since $E_{\mathbb{P}}(|X_n|^p) \le \parallel X_n \parallel^p_{L^\infty} \to 0$ as $n \to \infty$.

Proposition 6.1.3. *(Weak Law of Large Numbers) Let* $\{X_n : n \in \mathbb{N}\} \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ *a sequence of random variables with* $E(X_n) = \mu \in \mathbb{R}$ *,*

$$E_P(X_n^2) \le c < \infty \;\forall n \quad , \quad E_P(X_n X_m) = 0 \quad \textit{for } n \neq m$$

Let $S_n(\omega) = X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)$ *and* $\overline{S}_n(\omega) = n^{-1} S_n(\omega)$ *the empirical average of the sample.*

Then as $n \to \infty$*,* $\overline{S}_n \to \mu$ *in* $L^2(\mathbb{P})$*-norm and in probability.*

Proof: Letting $X_n'(\omega) = (X_n(\omega) - \mu)$ we can assume without loss of generality that $\mu = 0$. Then

$$E_P(\overline{S}_n^2) = \frac{1}{n^2}\left\{\sum_{i=1}^n E_P(X_i^2) + 2\sum_{i=1}^n \sum_{1\le j<i} E_P(X_i X_j)\right\} = \frac{1}{n^2}\sum_{i=1}^n E_P(X_i^2) \le \frac{c}{n} \to 0$$

and we have shown by using Chebychev inequality that convergence in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ norm for $p > 0$ implies convergence in probability □

Chapter 7

Changing the order of integration on a product space: Fubini Theorem

Lemma 7.0.1. *Let* $\{X_n\}_{n\in\mathbb{N}} \subseteq L^0(\Omega, \mathcal{F}, \mathbb{P})$.

1. *When* $X_n(\omega) \geq 0$ $\mathbb{P}$*-almost surely* $\forall n$*, it follows that*

$$\sum_{k=1}^{\infty} E_P(X_k) = E_P\left(\sum_{k=1}^{\infty} X_k(\omega)\right) \in [0, +\infty] \tag{7.0.1}$$

2. *When* $(X_n)_{n\in\mathbb{N}} \subseteq L^1(\mathbb{P})$ *(not necessarily non-negative),*

   and $\sum\limits_{k=1}^{\infty} E_{\mathbb{P}}(|X_k|) < \infty$ *, rhe random series*

$$S_\infty(\omega) = \sum_{k=1}^{\infty} X_k(\omega)$$

   is convergent $\mathbb{P}$*-almost surely and in* $L^1(\mathbb{P})$*-norm, and* (7.0.1) *holds.*

Proof:

1. clearly when $X_n(\omega) \geq 0$, $S_n(\omega) := \sum_{k=1}^{n} X_k(\omega) \uparrow S_\infty(\omega) \in [0, +\infty]$ $\mathbb{P}$-almost surely as $n \uparrow \infty$, and the claim follows by the monotone convergence theorem.

2. Let $Y_n = S_n(\omega) := \sum_{k=1}^n |X_k(\omega)|$, then $Y_n \uparrow Y_\infty := \sum_{k=1}^\infty |X_k(\omega)|$, and by the monotone convergence theorem 4.1.1 it follows that

$$E_{\mathbb{P}}(Y_\infty) = \sum_{k=1}^{\infty} E_{\mathbb{P}}(|X_k|)$$

therefore $Y_\infty \in L^1(P)$ and $P(Y_\infty < \infty) = 1$ $\mathbb{P}$-almost surely, and $\mathbb{P}$-almost surely $S_n(\omega) \to S_\infty(\omega)$ since the series is absolutely convergent.

By the triangle inequality it follows that

$$|S_n(\omega)| \le Y_n(\omega) \uparrow Y_\infty(\omega) < \infty$$

with $Y_\infty \in L^1(P)$, and by the dominated convergence theorem 4.1.2 it follows that the expectation series is convergent in $L^1(\mathbb{P})$ and (7.0.1) holds $\square$

Lemma 7.0.2. *Consider the two probability spaces $(\Omega_i, \mathcal{F}_i, P_i)$, i=1,2. their product space $\Omega_1 \times \Omega_2$ with the product-σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$.*

Let $X : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable defined on the product space.

Then, for every fixed $\omega_2 \in \Omega_2$, the map

$$\omega_1 \mapsto X(\omega_1, \omega_2)$$

is a $(\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ random variable, and for every fixed $\omega_1 \in \Omega_1$ the map

$$\omega_2 \mapsto X(\omega_1, \omega_2)$$

is a $(\Omega_2, \mathcal{F}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ random variable. Also the opposite implication is true:

if the map $\omega_1 \mapsto X(\omega_1, \omega_2)$ is $\mathcal{F}_1$-measurable $\forall \omega_2 \in \Omega_2$), and the map $\omega_2 \mapsto X(\omega_1, \omega_2)$ is $\mathcal{F}_2$-measurable $\forall \omega_1 \in \Omega_1$),

then the map $(\omega_1, \omega_2) \mapsto X(\omega_1, \omega_2)$ is jointly $\mathcal{F}_1 \otimes \mathcal{F}_2$-measurable.

Proof. We assume first that the random variable is bounded, $|X(\omega_1, \omega_2)| \le K$ $\forall \omega = (\omega_1, \omega_2)$ with a deterministic constant K.

Let

$$\mathcal{C} \;=\; \big\{(\mathcal{F}_1 \otimes \mathcal{F}_2)\text{-measurable and bounded random variables}$$
$$\text{for which the claim holds}\,\big\}$$

It is easy to show that $\mathcal{C}$ is a monotone class and

$$\mathbf{1}_{(A_1 \times A_2)}(\omega_1, \omega_2) = \mathbf{1}_{A_1}(\omega_1)\mathbf{1}_{A_2}(\omega_2) \in \mathcal{C},\ \forall A_i \in \mathcal{F}_i, i = 1, 2.$$

Since $\mathcal{F}_1 \times \mathcal{F}_2$ is a π-class, by the monotone class theorem it follows that $\mathcal{C}$ contains all bounded $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-measurable maps.

By using the monotone convergence theorem, the claim also to random variables X which are not necessarily bounded, since

$$X(\omega) = \lim_{K\to\infty} X^{(K)}(\omega) \quad \forall\omega = (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \qquad \text{where}$$
$$X^{(K)}(\omega) = X(\omega)\mathbf{1}(|X(\omega)| \le K) \in [-K, K]$$

and we have already shown that $X^{(K)}(\omega_1, \cdot)$ is $\mathcal{F}_2$-measurable and $X^{(K)}(\cdot, \omega_2)$ is $\mathcal{F}_1$-measurable.

In the other direction, let

$$\mathcal{C} \;=\; \Big\{ \text{functions } X(\omega_1, \omega_2) \text{ which are are bounded}$$
$$\text{and } \mathcal{F}_i\text{-measurable with respect to each coordinate } \omega_i,\, i = 1, 2 \Big\}$$

It follows that $\mathcal{C}$ is a monotone class , and $\forall A_1 \in \mathcal{F}_1, A_1 \in \mathcal{F}_2$ $\mathbf{1}_{A_1}(\omega_1) \times \mathbf{1}_{A_1}(\omega_2) \in \mathcal{C}$. By the monotone class theorem $\mathcal{C}$ contains every bounded and $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-measurable function.

Theorem 7.0.2. *(Fubini Theorem)*

Let $X : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a random variable defined on a product of probability spaces.

- *When $X(\omega_1, \omega_2) \ge 0$, define the marginal integrals as*

$$I_1^X(\omega_1) = \int_{\Omega_2} X(\omega_1, \omega_2)P_2(d\omega_2)\ , \quad I_2^X(\omega_2) = \int_{\Omega_1} X(\omega_1, \omega_2)P_1(d\omega_1)$$

(by lemma 7.0.2 and positivity it follows that these are well defined random variables on the respective probability spaces $\Omega_i, i = 1, 2$).

Then the iterated integral does not depend on the order of integration:

$$\int_{\Omega_1} I_1^X(\omega_1) P_1(d\omega_1) = \int_{\Omega_1} \left\{ \int_{\Omega_2} X(\omega_1, \omega_2) P_2(d\omega_2) \right\} P_1(d\omega_1) = \quad (7.0.2)$$

$$\int_{\Omega_2} I_2^X(\omega_2) P_1(d\omega_2) = \int_{\Omega_2} \left\{ \int_{\Omega_1} X(\omega_1, \omega_2) P_1(d\omega_1) \right\} P_2(d\omega_2) :=$$
$$\int_{\Omega_1 \times \Omega_2} X(\omega_1, \omega_2) \ (P_1 \otimes P_2)(d\omega_1 \times d\omega_2) \ .$$

The product probability $(P_1 \otimes P_2)(A)$ is well defined for $A \in (\mathcal{F}_1 \otimes \mathcal{F}_2)$ by taking $X(\omega_1, \omega_1) = \mathbf{1}_A(\omega_1, \omega_2)$.

- *More in general* (7.0.2) *holds true when*

$$\int_{\Omega_1 \times \Omega_2} |X(\omega_1, \omega_2)| \ (P_1 \otimes P_2)(d\omega_1 \times d\omega_2) < \infty$$

Proof: We introduce the class

$$\mathcal{C} = \left\{ (\mathcal{F}_1 \otimes \mathcal{F}_2)\text{-measurable and bounded random variables such that the claim holds } \right\}$$

We show that $\mathcal{C}$ is a monotone class. Clearly $\mathbf{1} \in \mathcal{C}$, and by the linearity of the expectation $(aX + bY) \in \mathcal{C}$ when $X, Y \in \mathcal{C}$ and $a, b \in \mathbb{R}$.

Let $\{X^{(n)} : n \in \mathbb{N}\} \subseteq \mathcal{C}$ be a monotone sequence of random variables in the class which approximates a bounded X from below,

$$0 \le X_n(\omega_1, \omega_2) \uparrow X(\omega_1, \omega_2) \le K < \infty \ \forall \omega_i \in \Omega_i, i = 1, 2.$$

By Lemma 7.0.2 and the monotone convergence Theorem 4.1.1 it follows that

$$I_i^{X^{(n)}}(\omega_i) \uparrow I_i^X(\omega_i) \ \in \ L^0(\Omega_i, \mathcal{F}_i) \ , \quad i = 1, 2.$$

By applying the monotone convergence Theorem once again it follows that

$$E_{P_1}(I_1^X) = \lim_{n \uparrow \infty} E_{P_1}(I_1^{X^{(n)}}) = \lim_{n \uparrow \infty} E_{P_2}(I_2^{X^{(n)}}) = E_{P_2}(I_2^X)$$

and Fubini theorem is true also for the monotone limit X, implying that $X \in \mathcal{C}$ is a monotone class.

Since $\mathbf{1}_{(A_1 \times A_2)} \in \mathcal{C} \quad \forall A_i \in \mathcal{F}_i, i = 1, 2$, and the collection of such sets form a π-class $\mathcal{F}_1 \times \mathcal{F}_2$ form a π-class, by the monotone class Theorem it follows that $\mathcal{C}$ contains all bounded and $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-measurable random variables, and by the monotone convergence Theorem it is extended to all non-negative and $(\mathcal{F}_1 \otimes \mathcal{F}_2)$-measurable random variables.

We show that the map

$$(P_1 \otimes P_2) : (\mathcal{F}_1 \otimes \mathcal{F}_2) \longrightarrow [0, 1]$$

is a σ-additive probability:

Let $\{A^{(n)} : n \in \mathbb{N}\} \subseteq (\mathcal{F}_1 \otimes \mathcal{F}_2)$ non-decreasing event sequence, with

$$A^{(n)} \subseteq A^{(n+1)} \uparrow A = \bigcup_{n \in \mathbb{N}} A^{(n)} .$$

We define the sections of these events as follows: for $\omega_1 \in \Omega_1$, let

$$A_2^{(n)}(\omega_1) := \big\{\omega_2 : (\omega_1, \omega_2) \in A^{(n)}\big\} \uparrow A_2(\omega_1) := \big\{\omega_2 : (\omega_1, \omega_2) \in A\big\} \text{ as } n \uparrow \infty .$$

$\forall \omega_1 \in \Omega_1$. Since P_2 is σ-additivie it follows that $P_2\big(A_2^{(n)}(\omega_1)\big) \uparrow P_2\big(A_2(\omega_1)\big)$, where $P_2(A_2^{(n)}(\omega_1))$ and $P_2(A_2(\omega_1))$ are $\mathcal{F}_1$-measurable random variables. By the monotone convergence Theorem it follows that

$$(P_1 \otimes P_2)(A^{(n)}) = \int_{\Omega_1} P_2\big(A_2^{(n)}(\omega_1)\big) P_1(d\omega_1) \quad \uparrow \quad \int_{\Omega_1} P_2\big(A_2(\omega_1)\big) P_1(d\omega_1) = (P_1 \otimes P_2)(A) .$$

By decomposing $X(\omega) = \big(X^+(\omega) - X^-(\omega)\big)$, and integrating separately the positive and negative parts, the claim follows as well for all $X \in L^1(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2) \quad \square$.

Corollary 7.0.1. *Fubini Theorem holds also when the probability spaces* $(\Omega_i, \mathcal{F}_i)$ *are equipped with* σ*-finite measures* μ_i*,* $i = 1, 2$*.*

Proof. when $\mu_i(d\omega_i)$ are finite measures, we can apply Fubini theorem to the normalized probability measures $P_i(d\omega_i) = \mu_i(d\omega_i)/\mu_i(\Omega_i)$ $i = 1, 2$, and then rescale back.

When μ_i are σ-finite, there are countable and measurable partitions of Ω_i, on olemassa numeroituvia measurable osituksia

$$\Omega_i = \bigcup_{i\in\mathbb{N}} \Omega_i^{(n)}, \text{ where } \big(\Omega_i^{(n)} \cap \Omega_i^{(m)}\big) = \emptyset, \ n \neq m, \quad \text{and } \mu_i(\Omega_i^{(n)}) < \infty \quad i = 1, 2$$

The claim follows when we apply Fubini to each product space $(\Omega_1^{(n)} \times \Omega_2^{(k)})$, $n, k \in \mathbb{N}$, and then take the countable sum of the integrals □

Corollary 7.0.2. *Fubini Theorem extends direclty to finite product spaces*

$$\big(\Omega_1 \times \cdots \times \Omega_d, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_d, P_1 \otimes \cdots \otimes P_d\big), \quad d \in \mathbb{N}$$

Proposition 7.0.4. *On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let*

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots X_d(\omega)) \in \mathbb{R}^d$$

be a random vector, and let $P_X(dt)$ its probability distribution on the space of values $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$:

$$P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}\big(\{\omega : X(\omega) \in B\}\big), \quad \textit{kun } B \in \mathcal{B}(\mathbb{R}^d)$$

The random variables are $X_1(\omega), \ldots, X_d(\omega)$ are independent w.r.t. $\mathbb{P}$ if and only if the distribution of X is a product of the coordindanates distributions

$$P_X = (P_{X_1} \otimes \cdots \otimes P_{X_d})$$

Proof Exercise.

Note that by Lemma 7.0.2, the map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is measurable if and only if all the coordinate maps $X_i : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $i = 1, \ldots, d$ are measurable.

Example 7.0.2. *Let X be a random variable with $\mathbb{P}(X(\omega) \geq 0) = 1$. Silloin*

$$E_P(X) = \int_\Omega X(\omega) P(d\omega) = \int_0^\infty t \, P_X(dt) = \int_0^\infty P(X > t) dt$$

Proof. since $t = \int\limits_0^\infty \mathbf{1}(s < t) ds$

$$\int_0^\infty t \, P_X(dt) = \int_0^\infty \left(\int_0^\infty \mathbf{1}(s < t) ds \right) P_X(dt)$$

Since $\mathbf{1}(s < t) \geq 0$ and the Lebesgue measure is σ-finite, Fubini Theorem applies, and we can change the order of integration:

$$= \int_0^\infty \Big(\int_0^\infty \mathbf{1}(s<t) P_X(dt) \Big) ds = \int_0^\infty P_X\big((s,+\infty)\big) ds = \int_0^\infty P(X > s) ds$$

7.0.1 Integration by parts formula

Proposition 7.0.5. *Let $F, G : \mathbb{R} \to \mathbb{R}$ non-decreasing and right-continuous functions. Then $a < b$,*

$$\int_a^b G(x) F(dx) = \int_a^b G(x-) F(dx) + \sum_{y \in (a,b]} \Delta G(y) \Delta F(y)$$
$$= F(b)G(b) - F(a)G(a) - \int_a^b F(x-) G(dx)$$

where the integrals do exist in Riemann Stieltjes sense, $F(x-) = \lim_{y \uparrow x} F(y)$ is the limit from the left. This integration by parts formula hols also when F and G have finite variation the interval $[a,b]$, meaning that $F(x) = F^+(x) - F^-(x)$, $G(x) = G^+(x) - G^-(x)$, jossa $F^\pm, G^\pm$ are non-decreasing (and right-continuous).

Proof. Note that

$$\int_a^b G(x)F(dx) = \int_{(a,b]} G(x)F(dx) = \int_{-\infty}^\infty \mathbf{1}_{(a,b]}(x) G(x) F(dx),$$

and as $x \geq a$,

$$G(x) = G(a) + \int_a^\infty \mathbf{1}(y \leq x) G(dy) \ .$$

This implies

$$\int_a^b G(x)F(dx) = \int_a^b \Big\{ G(a) + \int_a^\infty \mathbf{1}(y \leq x) G(dy) \Big\} F(dx) =$$
$$G(a)\big(F(b) - F(a)\big) + \int_a^b \Big(\int_a^b \mathbf{1}(y \leq x) G(dy) \Big) F(dx)$$

Since the measures $F(dx)$ and $G(dy)$ are finite on the compact interval $[a,b]$

and the integrand is non-negativ, Fubini Theorem applies, giving

$$= G(a)\big(F(b)-F(a)\big) + \int_a^b \biggl(\int_a^b \mathbf{1}(y \le x) F(dx) \biggr) G(dy) =$$
$$G(a)\big(F(b)-F(a)\big) + \int_a^b \big(F(b) - F(y-)\big) G(dy) =$$
$$F(b)G(b) - F(a)G(a) - \int_a^b F(y-)G(dy) \ .$$

Note that we could also write

$$\int_a^b G(x)F(dx) = \int_a^b G(x-)F(dx) + \sum_{y \in (a,b]} \Delta G(y) \Delta F(y)$$

since a non-decreasing function has at most a countable number of jumps
□

Example 7.0.3. *Let* $(x_t : t \in N)$ $(y_t : t \in N)$ *sequences, let* $\Delta x_t = x_t - x_{t-1}$ *and* $\Delta y_t = y_t - y_{t-1}$. *Then we have Abel discrete integration by parts formula:*

$$x_t y_t - x_0 y_0 = \sum_{s=1}^t x_{s-1} \Delta y_s + \sum_{s=1}^t y_{s-1} \Delta x_s + \sum_{s=1}^t \Delta x_s \Delta y_s +$$
$$= \sum_{s=1}^t x_s \Delta y_s + \sum_{s=1}^t y_{s-1} \Delta x_s = \sum_{s=1}^t x_{s-1} \Delta y_s + \sum_{s=1}^t y_s \Delta x_s$$

Proof You can imbed $\mathbb{N}$ in $\mathbb{R}$ and use the integration by parts formula on $\mathbb{R}$,
by taking $F(t) = x_0 + \sum_{n \le t} \Delta x_n$, $G(t) = y_0 + \sum_{n \le t} \Delta y_n$.

We show how integration by parts works with the Gaussian distribution:

Gaussian integration by parts and Stein equation

Proposition 7.0.6. *Consider a standard Gaussian random variable* $G(\omega)$ *with* $E(G) = 0$, $E(G^2) = 1$, *and cumulative distribution function*

$$\Phi(x) = P(G \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\bigl(-t^2/2\bigr) dt,$$

and let $f:\mathbb{R}\to\mathbb{R}$ an absolutely continuous function with

$$f(x)=f(0)+\int_0^x f'(t)dt$$

where $f':\mathbb{R}\to\mathbb{R}$ is Borel measurable and we assume that

$$E_P\big(|f'(G)|\big)<\infty$$

In such case $E_P\big(|f(G)G|\big)<\infty$, and the following Gaussian integration by parts formula (also called Stein equation) holds:

$$E_P\big(f'(G)\big)=E_P\big(f(G)G\big)$$

Proof Let

$$\phi(x):=\frac{d\Phi}{dx}(x)=\frac{1}{\sqrt{2\pi}}\exp\big(-x^2/2\big)$$

be the probability density function of G, which satisfies

$$\phi'(x)=\frac{d\phi}{dx}(x)=-\phi(x)x\ .$$

The Gaussian distribution has finite exponential moments $E_P\big(\exp(\lambda G)\big)=\exp\big(\lambda^2/2\big)<\infty$, $\forall\lambda\in\mathbb{R}$, which implies also that all polynomial moments $E_P(|G|^q)<\infty,\quad\forall q>0$, are finite. Without loss of generality we can assume that $f(0)=0$, and by Fubini Theorem it follows that

$$\begin{aligned}
&E_P\big(|f(G)G|\big)=\int_{-\infty}^{\infty}|f(x)|\,|x|\,\Phi(dx)=\int_{-\infty}^{\infty}|f(x)|\,|x|\,\phi(x)dx\\
&\le\int_{-\infty}^{0}\left(\int_x^0|f'(t)|dt\right)(-x)\,\phi(x)dx+\int_0^{\infty}\left(\int_0^x|f'(t)|dt\right)x\,\phi(x)dx\\
&=\int_{-\infty}^{0}\left(\int_{-\infty}^{t}(-x)\phi(x)dx\right)|f'(t)|dt+\int_0^{\infty}\left(\int_x^{\infty}x\,\phi(x)dx\right)|f'(t)|dt\\
&\int_{-\infty}^{\infty}\phi(t)|f'(t)|dt=E_P\big(|f'(G)|\big)<\infty
\end{aligned}$$

The integrability assumption of Fubini theorem holds and we can change the order of integration in the iterated integral:

$$E_P\big(f(G)G\big) = \int_{-\infty}^{\infty} f(x)\ x\ \Phi(dx) = \int_{-\infty}^{\infty} f(x)\ x\ \phi(x)dx$$
$$= \int_{-\infty}^{0}\left(\int_{x}^{0} f'(t)dt\right)(-x)\ \phi(x)dx + \int_{0}^{\infty}\left(\int_{0}^{x} f'(t)dt\right)x\ \phi(x)dx$$
$$= \int_{-\infty}^{0}\left(\int_{-\infty}^{t} (-x)\phi(x)dx\right)f'(t)dt + \int_{0}^{\infty}\left(\int_{x}^{\infty} x\ \phi(x)dx\right)f'(t)dt$$
$$\int_{-\infty}^{\infty} \phi(t)f'(t)dt = E_P\big(f'(G)\big) < \infty \qquad \square$$

Remark 7.0.1. *Stein equation is a very powerful tool which extends to functionals of multivariate Gaussian variables and also to functionals infinite dimensional Gaussian processes. Malliavin calculus is based on that.*

Corollary 7.0.3. *By applying the Gaussian integration by parts to $f(x) = \varphi(x)\psi(x)$ we also the generalization*

$$E_P\big(\partial\varphi(G)\psi(G)\big) = E_P\big(\varphi(G)(\psi(G)G - \partial\psi(G))\big) = E_P\big(\varphi(G)\partial^*\psi(G)\big)$$

for $\varphi(x)$, $\varphi(\psi)$ absolutely continuous with $E_P\big(\big|\partial(\varphi(G)\psi(G))\big|\big) < \infty$, where $\partial^*\psi(x) := \big(\psi(x)x - \partial\psi(x)\big)$ is the adjoint operator of the derivative operator ∂ in the space $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_G)$ equipped with the standard Gaussian probability measure.

7.0.2 Change of variable formula on the Euclidean space $\mathbb{R}^d$

In this section we follow Paul Malliavin book Integration and Probability.

Definition 7.0.2. *Let $O \subseteq \mathbb{R}^n$ open. A mapping $f = (f_1, \dots, f_n) : O \to \mathbb{R}^n$ is said to be a diffeomorphism if :*

1. *$O' = f(O)$ is open and $f : O \to O'$ is a continuous bijection with continuous inverse $g = f^{-1}$.*

2. *f and g have continuous first order partial derivatives.*

Definition 7.0.3. *The $n \times n$ Jacobian-matrix $Jf(x)$ is defined as*

$$\Big[Jf(x)\Big]_{ij} = \frac{\partial f_i}{\partial x_j}(x)$$

If $f : O \to O'$ and $g : O' \to O''$ are diffeomorphisms, the composition $g \circ f : O \to O''$ is a diffeomorphism, with Jacobian

$$\big[J(g \circ f)\big] = \big[Jg\big]\big[Jf\big] \ , \tag{7.0.3}$$

where on the right side we take the matrix product.

In particular for the inverse diffeomorphism we have $J(f^{-1}) = (Jf)^{-1}$.

Definition 7.0.4. *A partition of unity on a topological space S is a sequence of continuous functions $\varphi_n : S \to [0,1]$, $n \in \mathbb{N}$ such that*

1. *supp$(\varphi_n) = \{ x \in S : \varphi_n(x) \neq 0\}$ is compact*
2. *for each compact K, there are only finitely many n with $K \cap$ supp$(\varphi_n) \neq \emptyset$.*
3. *$\sum_{n \in \mathbb{N}} \varphi_n(x) = 1 \ \forall x \in S$.*
4. *Let $(U_\alpha : \alpha \in I)$ an open cover of S, which is a collection of open sets $U_\alpha \subseteq S$ with $\bigcup_{\alpha \in I} U_\alpha = S$. A partition of unit $(\varphi_n : n \in \mathbb{N})$ is said to be subordinate to the open cover $(U_\alpha : \alpha \in I)$,*

   if $\forall n \ \exists \alpha(n)$ with $U_{\alpha(n)} \supseteq$ supp(φ_n)

Proposition 7.0.7. *For any open cover $(U_\alpha : \alpha \in I)$ of the euclidean space $S = \mathbb{R}^d$, there is a partition of unity $(\varphi_n : n \in \mathbb{N})$ subordinate to $(U_\alpha : \alpha \in I)$*

For a proof in more general settings, see Theorem 1.4.1. in Paul Malliavin book, probability and integration.

Theorem 7.0.3. *Let $f : O \to O'$ between O and $O' \in \mathbb{R}^n$, and $\varphi : O' \to \mathbb{R}$ a continuous test function with compact support $K \subseteq O'$. We prove the change of variable formula*

$$\int_O \varphi(f(x)) \big|\det\big(Jf(x)\big)\big| dx = \int_{O'} \varphi(y) dy \ . \tag{7.0.4}$$

Note that when $f(x) = f(0) + Ax$ is linear where A is an invertible $n \times n$ matrix, the rectangle

$$V = \Delta x_1 \times \cdots \times \Delta x_n$$

maps into a parallelepiped AV with edges $A\Delta x_1, \ldots, A\Delta x_n$.

Lets denote

$$\Delta x_1 \wedge \cdots \wedge \Delta x_n := \det([\Delta x_1, \ldots, \Delta x_n])$$

The volume of the rectangle V is given by

$$|V| = |\Delta x_1 \times \cdots \times \Delta x_n| = |\det([\Delta x_1, \ldots, \Delta x_n])| = |\Delta x_1 \wedge \cdots \wedge \Delta x_n|$$

while the volume of the parallelepiped AV

$$|AV| = |A\Delta x_1 \wedge \cdots \wedge A\Delta x_n| = |\det([A\Delta x_1, \ldots, A\Delta x_n])| = |\det(A)||V|$$

and in both cases the volume is the absolute value of the wedge product of the edges.

In the linear case the Jacobian matrix $Jf(x) = A$ is constant.

We prove the change of variable formula by induction (7.0.4). When $n = 1$, the test function $\varphi(x)$ has representation

$$1 = \sum_{k=1}^{\infty} \psi_n(x)$$

where the support of ψ_n, is a small compact set . By integrating each term separately, we can assume without loss of generality that the support of $\varphi(y)$ is a small interval I where the sign of the derivative $f'(x)$ does not change when $x \in I$. By the (linear) change of variable $x \to (-x)$, this can be further reduced to the case with $f'(x) > 0$. Then the formula we have to show is

$$\int \varphi(f(x))f'(x)dx = \int \varphi(y)dy$$

Set

$$F(t) = \int_0^t \varphi(f(x))f'(x)dx, \quad G(t) = \int_{f(0)}^{f(t)} \varphi(y)dy \ .$$

Since $F'(t) = G'(t) = \varphi(f(t))f'(t)$, and $F(0) = G(0)$ it follows that $F(t) = G(t)\ \forall t$ and the change of variable formula is proved in the one-dimensional case.

For the general proof we proceed by induction, assuming that the change of variable formula holds $\forall m < n$. Let $x = (\xi, u) \in \mathbb{R}^n$, with $\xi \in \mathbb{R}$ and $u \in \mathbb{R}^{n-1}$.

Set $x' = h(x) = (\xi', u) = (f_1(\xi, u), u) \in \mathbb{R}^n$ and $g(x') = (\xi', \theta(x'))$ where $\theta(x') = \big(f_2(f_1^{-1}(\xi'), u), \dots, f_2(f_1^{-1}(\xi'), u)\big)$. By the implicit function theorem, f_1 is invertible in a neighbourhood of ξ for fixed u if

$$\frac{\partial f_1}{\partial x_1}(\xi, u) \neq 0 \ .$$

which follows when $\det Jf(x) \neq 0$. By construction, we see that $f = g \circ h$. By the property (7.0.3) of the Jacobian determinant, it is enough to prove the theorem for the maps g and h separately.

For the g map we have

$$\int_{\mathbb{R}\times\mathbb{R}^{n-1}} \varphi(\xi', \theta(x', u)) \det Jg(\xi', u) d\xi' du = \text{(Fubini)}$$
$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^{n-1}} \varphi(\xi', \theta(x', u)) \det Jg(\xi', u) du \right) d\xi' = \int_{\mathbb{R}} \left(\int_{\mathbb{R}^{n-1}} \varphi(\xi', u') du' \right) d\xi'$$

where we used the induction hypothesis and the fact that $\det J_g = \det J_{\theta(\xi')}$, since the map g does not change the first coordinate. This proves theorem for g.

For the h map, again by Fubini

$$\int_{\mathbb{R}^{n-1}\times\mathbb{R}} \varphi(f_1(\xi, u), u) \det Jh(\xi, u) du d\xi =$$
$$\int_{\mathbb{R}^{(n-1)}} \left(\int_{\mathbb{R}} \varphi(f_1(\xi, u), u) \frac{\partial f_1}{\partial \xi}(\xi, u) d\xi \right) du = \int_{\mathbb{R}^{(n-1)}} \left(\int_{\mathbb{R}} \varphi(\xi', u) d\xi' \right) du \quad \square$$

Remark The change of variable formula (7.0.4) extends to any bounded or non-negative Borel measurable test function $\varphi(x)$. For any compact rectangle of the form $R = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n]$, we can approximate pointwise the indicator $\varphi(x) = \mathbf{1}_R(x)$ by a sequence of continuous functions φ_k with $0 \leq \varphi_k(x) \uparrow \varphi(x)$ as $k \uparrow \infty$. By the monotone convergence

Theorem, (7.0.4) holds also for $\varphi(x) = \mathbf{1}_{\mathbb{R}}$, and since these rectangles generate the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$, by the monotone class Theorem (7.0.4) extends to all bounded measurable test functions φ, and for unbounded non-negative test functions $\varphi(x) \geq 0$ the result follow by again by applying the monotone convergence theorem with $0 \leq \varphi(x) \wedge K \leq \varphi(x)$, $K \uparrow \infty$.

Corollary 7.0.4. *Assume that $X(\omega) \in O \subseteq \mathbb{R}^n$, where O is open is a random vector with density $p_X(x)$, and $f : O \to O'$ is a diffeomorphism. Then $Y = f(X)$ has density on O' given by*

$$p_Y(y) = p_X(f^{-1}(y)) \big| \det\bigl(Jf(f^{-1}(y))\bigr|^{-1} \tag{7.0.5}$$

where $\bigl(Jf(f^{-1}(y))\bigr)^{-1} = Jf^{-1}(y)$.

Proof *Given a bounded Borel measurable test function $\psi(y)$, define*

$$\varphi(y) = \psi(y) p_X(f^{-1}(y)) \big| \det(Jf(f^{-1}(y)) \big|^{-1} ,$$

and apply (7.0.4) *to the test function φ, obtaining*

$$\int_{O'} \psi(y) p_X(f^{-1}(y)) \big| \det(Jf(f^{-1}(y)) \big|^{-1} dy = \int_{O'} \varphi(y) dy = \int_{O} \varphi(f(x)) \big| \det\bigl(Jf(x)\bigr) \big| dx =$$
$$\int_{O} \psi(f(x)) p_X(x) dx = E_P\bigl(\psi(f(X))\bigr) = E_P\bigl(\psi(Y)\bigr)$$

which means that the distribution of Y has density (7.0.5)*with respect to Lebesgue measure.*

Example 7.0.4. *Consider the dimensional integral*

$$I = \int_0^{\infty)} \exp\Bigl(-\frac{x^2}{2}\Bigr) dx$$

and we rewrite I^2 as a double integral

$$I^2 = \Bigl(\int_0^{\infty)} \exp\Bigl(-\frac{x^2}{2}\Bigr) dx\Bigr) \Bigl(\int_0^{\infty)} \exp\Bigl(-\frac{y^2}{2}\Bigr) dy\Bigr) \int_0^{\infty)} \int_0^{\infty)} \exp\Bigl(-\frac{x^2+y^2}{2}\Bigr) dx dy$$

We apply the chanege of variable $(x,y) \mapsto f(x,y) = \bigl(\rho(x,y), \theta(x,y)\bigr)$ where $\rho(x,y) = \sqrt{x^2+y^2}$ and $\theta(x,y) = \arctan(y/x)$

for the Jacobian we have

$$Jf(x,y)=\begin{bmatrix}\frac{\partial\rho(x,y)}{\partial x} & \frac{\partial\rho(x,y)}{\partial y}\\ \frac{\partial\theta(x,y)}{\partial x} & \frac{\partial\theta(x,y)}{\partial y}\end{bmatrix}=\begin{bmatrix}\frac{x}{\rho(x,y)} & \frac{y}{\rho(x,y)} \frac{-y}{\rho(x,y)} & \frac{x}{\rho(x,y)}\end{bmatrix}$$

with $\det Jf(x,y)=1/\rho$. Therefore we have

$$I^2=\int_0^\infty\int_0^\infty \exp\Big(-\frac{x^2+y^2}{2}\Big)dxdy=$$
$$\int_0^\infty \exp(-\frac{\rho^2}{2})\rho d\rho\int_0^{\pi/2}d\theta=\int_0^\infty \exp(-u)du\pi/2=\pi/2$$

by the change of variable $u=\rho^2/2$, which implies $I=\sqrt{\pi/2}$. By symmetry we obtiain the normalizing constant of the standard Gaussian distribution:

$$\sqrt{2\pi}=\int_{\mathbb{R}}\exp\Big(-\frac{x^2}{2}\Big)dx$$

Example 7.0.5. *Let $\xi(\omega)=(\xi_1(\omega),\dots,\xi_d(\omega))\in\mathbb{R}^d$ a Gaussian random vector, where the components $\xi_k(\omega)\in\mathbb{R}$ are P-independent and identically distributed standard-Gaussian random variables, such that $E_P(\xi_k)=0$ ja $E_P(\xi_k^2)=1$. The density of this Gaussian distribution is the product of the distribution of the coordinates:*

$$p_\xi(x)=\prod_{k=1}^d p_{\xi_k}(x_k)=(2\pi)^{-d/2}\exp\Big(-\frac{\parallel\xi\parallel^2}{2}\Big),\quad \parallel\xi\parallel^2=\sum_{k=1}^d\xi_k^2\ .$$

Let $\mu=(\mu_1,\dots,\mu_d)\in\mathbb{R}^d$ be a deterministic vector and $A=(A_{ij}:0\le i\le j\le d)$ a deterministic $d\times d$ matrix.

Let $X(\omega)=(\mu+A\xi(\omega)^\top)\in\mathbb{R}^d$. Then

- *$E_P(X)=\mu$ and $E_P(X_iX_j)-E_P(X_i)E_P(X_j)=\Sigma_{ij}$, where $\Sigma=AA^\top$.*
- *The random vector X has density w.r.t. the d-dimensional $\mathbb{R}^d$-Lebesgue measure*

$$p_X(x)=(2\pi)^{-d/2}\det(\Sigma)^{-1/2}\exp\big(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^\top\big)\qquad(7.0.6)$$

meaning that if $g : \mathbb{R}^d \to \mathbb{R}$ is a bounded measurable test function

$$E_P\Big(g(X)\Big) = E_P\big(g(\mu + A\xi^\top)\big) =$$
$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} g\big(\mu_1 + \sum_{j=1}^d A_{1j}y_j, \dots, \mu_d + \sum_{j=1}^d A_{dj}y_j\big) \prod_{j=1}^d \Big\{ \frac{1}{\sqrt{2\pi}} \exp(-y_j^2/2) \Big\} dy_1 \dots dy_d =$$
$$\int_{\mathbb{R}^d} g(x_1, \dots, x_d) p_X(x_1, \dots, x_d) dx_1 \dots dx_d$$

Proof

- We compute the density by using the change of variable formula for $x = \mu + Ay^\top$, assuming that A is invertible.

$$\begin{aligned} E_P(g(X)) = E_{P_{\xi_1} \otimes \dots \otimes P_{\xi_d}}(g(X)) &\stackrel{\perp\!\!\!\perp}{=} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} g(\mu + A\xi) dP_{\xi_1} \dots dP_{\xi_d} \\ &= \int_{\mathbb{R}^d} g(\mu + Ay) p_\xi(y) dy, \end{aligned} \tag{7.0.7}$$

where

$$p_\xi(y) = \prod_{j=1}^d p_{\xi_j}(y_j) = (2\pi)^{-d/2} \exp\left\{ -\frac{1}{2} y^\top y \right\}$$

We plug in $x = (\mu + Ay)$ in (7.0.7), where $y = A^{-1}(x - \mu)$ with Jacobian determinant $\det(A^{-1}) = 1/\det(A)$, obtaining

$$E_P(g(X)) = \int_{\mathbb{R}^d} g(x) \det(A^{-1}) p_\xi(A^{-1}(x - \mu)) dx := \int_{\mathbb{R}^d} g(x) p_X(x) dx,$$

and

$$\begin{aligned} p_X(x) &= \det(A^{-1}) \exp\left\{ -\frac{1}{2} (A^{-1}(x - \mu))^\top (A^{-1}(x - \mu)) \right\} \\ &= \det(A^{-1}) \exp\left\{ -\frac{1}{2} (x - \mu)^\top (A^\top)^{-1} A^{-1} (x - \mu) \right\} \\ &= \det(A^{-1}) \exp\left\{ -\frac{1}{2} (x - \mu)^\top (AA^\top)^{-1} (x - \mu) \right\}. \end{aligned}$$

By denoting $\Sigma := AA^\top$ which is symmetric and positive definite, with $\det(A) = \sqrt{\det(\Sigma)}$, we obtain (7.0.6), the multivariate Gaussian density of X. This distribution is characterized by its mean,

$$E_P(X) = E_P(\mu + A\xi) = \mu + AE_P(\xi) = \mu \ ,$$

and covariance matrix $\Sigma = AA^\top = E(XX^\top) - E(X)E(X)^\top$, where

$$\begin{aligned} E(XX^\top) &= E((\mu + A\xi)(\mu + A\xi)^\top) = E(\mu\mu^\top + \mu\xi^\top A^\top + A\xi\mu^\top + A\xi\xi^\top A^\top) \\ &= \mu\mu^\top + AE(\xi\xi^\top)A^\top = \mu\mu^\top + AA^\top = \mu\mu^\top + \Sigma. \end{aligned}$$

Chapter 8

Tasainen integroituvuus ja $L^1(P)$-konvergenssi

Definition 8.0.5. *Merkitään*

$$L^0(\Omega, \mathcal{F}, P) = \big\{ \ \mathbb{R}\text{-arvoiset satunnaismuuttujat todennäköisyysavaruudessa } (\Omega, \mathcal{F}, P) \ \big\}$$

jossa tarvittaessa identifioidaan X ja Y kun $X(\omega) = Y(\omega)$ P-melkein varmasti.

Kun $0 < p < \infty$, määritellään

$$L^p = L^p(\Omega, \mathcal{F}, P) = \big\{ X \in L^0(\Omega, \mathcal{F}, P) \text{ jolla } \parallel X \parallel_p < \infty \big\} \quad \text{jossa} \quad \parallel X \parallel_p = \big\{ E_P(|X|^p) \big\}^{1/p}$$

Sanomme että $X_n \overset{L^p}{\to} X$ suppenee L^p-normissa kun $E_P(|X_n - X|^p) \to 0$ kun $n \to \infty$. Määritellään myös olennainen (essential) supremum

$$\parallel X \parallel_\infty = \ P\text{-esssup} \{|X(\omega)|\} := \inf\big\{ y \in \mathbb{R} : |X(\omega)| \leq y \ P\text{-melkein varmasti} \ \big\}$$
$$L^\infty = L^\infty(\Omega, \mathcal{F}, P) = \big\{ X \in L^0(\Omega, \mathcal{F}, P) \text{ jolla } \parallel X \parallel_\infty < \infty \big\}$$

eli satunnaismuuttuja $X(\omega) \in L^\infty(P)$ jos ja vain jos on olemassa deterministinen $K < \infty$ jolla $|X(\omega)| \leq K$ P-melkein varmasti, ja s.m. on olennaisesti rajoitettu P-todennäköisyyden suhteen.

Osoitamme (myöhemmin) että $L^p(\Omega, \mathcal{F}, P)$ on Banachin avaruus (eli vektori avaruus jolla on täydellinen normi) kaikille $0 < p \leq +\infty$,

ja $L^2(\Omega, \mathcal{F}, P)$ on Hilbertin avaruus skalaaritulolla $\langle X, Y \rangle := E_P(XY)$.

Theorem 8.0.4. *Olkoon* $0 < p \leq \infty$, *ja* $\lim_{n\to\infty} \parallel X_n \parallel_p = 0$. *Seuraa että* $X_n \stackrel{P}{\to} 0$.

Tod. Kun $0 < p < +\infty$, väite seuraa **Chebychevin epäyhtälöstä** : kun $\varepsilon > 0$,

$$\varepsilon^p P(|X_n| > \varepsilon) \leq E_P(|X_n|^p) \to 0 \text{ kun } n \to \infty \;.$$

Kun $p = +\infty$, $\forall K \in \mathbb{N}$ $\exists \bar{n}$ jolla

$$P\big(\{\omega : |X_n(\omega)| \leq K^{-1}\}\big) = 1 \quad \text{ kun } n \geq \bar{n}$$

josta seuraa että

$$P\Big(\bigcap_K \bigcup_m \bigcap_{n>m} \{\omega : |X_n(\omega)| \leq K^{-1}\}\Big) = P\big(\{\omega : X_n(\omega) \to 0\}\big) = 1$$

siis $X_n \to 0$ P-melkein varmasti ja myös stokastisesti □

Proposition 8.0.8. *Lebesgue dominoidun konvegenssin lause* L^p *avaruudessa.*
Olkoon $0 < p < \infty$, *ja* $\big(X_n(\omega) : n \in \mathbb{N}\big)$, $X(\omega)$ *satunnaismuuttujat jolla*

1) $X_n \stackrel{P}{\longrightarrow} X$ *stokastisesti*
2) $Y(\omega) := \sup_{n\in\mathbb{N}}\big\{ \mid X_n(\omega)\mid \; : \; n \in \mathbb{N}\big\} \in L^p(\Omega, \mathcal{F}, P)$.

Silloin $X \in L^p(\Omega)$ *ja* $X_n \stackrel{L^p}{\longrightarrow} X$.

Tod. On olemassa alijono $(n_k)_{k\in\mathbb{N}}$ jolla $\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega)$ P-m.v.
Tästä seuraa $|X(\omega)| \leq Y(\omega)$ P-m.v. ja siksi $X \in L^p(\Omega, \mathcal{F}, P)$. Fubinin lauseesta

$$E_P\big(|X_n(\omega) - X(\omega)|^p\big) = \int_0^\infty P(|X_n - X|^p > t)dt \;.$$

Kolmion epäyhtälöstä seuraa

$$|X_n(\omega) - X(\omega)|^p \leq 2^p Y(\omega)^p \quad \text{ ja siksi } \quad P(|X_n - X|^t > t) \leq P(Y^p > t2^{-p})$$

jossa

$$\int_0^\infty P(2^p Y^p > t)dt = 2^p E_P\big(Y^p\big) < \infty$$

Määritellään funktioita

$$0 \le f_n(t) := P(|X_n - X|^p > t) \le g(t) := P\big(|Y|^p > t2^{-p}\big), \quad t \ge 0$$

Nämä funktiot ovat ei-kasvavia ja siksi mitallisia t:n suhteen. Jokaiselle $t > 0$

$$f_n(t) = P(|X_n - X|^p > t) \longrightarrow 0 \text{ kun } n \to \infty$$

koska $X_n \xrightarrow{P} X$ stokastisen konvergenssin mielessä. Koska yläraja $g(t)$ on integroituva Lebesgue mitan λ suhteen, Lebesguen dominoidun konvergenssin lause 4.1.2 astuu voimaan avaruudessa $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, ja saadaan

$$\lim_{n\to\infty} E_P\big(|X_n - X|^p\big) = \lim_{n\to\infty} \int_0^\infty f_n(t)dt = \int_0^\infty \Big(\lim_{n\to\infty} f_n(t)\Big) dt = 0 \qquad \square$$

Huomautus: Koska

$$|E_P(X) - E_P(Y)| = |E_P(X - Y)| = |E_P((X - Y)^+) - E_P((X - Y)^-)\,|$$
$$\le E_P((X - Y)^+) + E_P((X - Y)^-) = E_P(|X - Y|)$$

kun $X_n \xrightarrow{L^1(P)} X$ seuraa $E_P(X_n) \to E_P(X)$.

Lebesguen dominoidun konvergenssin lauseen ehdot L^p-konvergenssille ovat riittäviä mutta ei välttämättömiä. Haluamme karakterisoida täydellisesti L^p-konvergenssia.

Käsittelemme ensin L^1-konvergenssia. Olemme huomanneet (4.1.1) että ehdoista $X_n(\omega) \to X(\omega)$ P-melkein varmasti ja $X, X_n \in L^1(P)$ ei seuraa että $E(X_n) \to E(X)$, eikä myöskään $X_n \xrightarrow{L^1} X$. Siihen tarvitaan sen lisäksi seuraava kompaktisuuden kaltainen ehto:

Definition 8.0.6. *Satunnaismuuttujen kokoelma $\mathcal{C} \subset L^1(\Omega, \mathcal{F}, P)$ on* **tasaisesti integroituva** *P:n suhteen, kun*

$$\lim_{K\to\infty} \sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}(|X| > K)\big) = \lim_{K\to\infty} \sup_{X\in\mathcal{C}} \int_{\{\omega:|X(\omega)|>K\}} |X(\omega)| P(d\omega) = 0$$

Lemma 8.0.3.

1. *Olkoon $\mathcal{C}$ satunnaismuuttujen perhe. Jos perheelle on olemassa integroituva yläraja, eli*

   $$\sup_{X\in\mathcal{C}} |X(\omega)| \le Y(\omega) \; P \text{ melkein varmasti, jossa } Y \in L^1(P)$$

   perhe on tasaisesti integroituva.

2. *Äärellinen satunnaismuuttujen joukko $\mathcal{C} = \{X_1, X_2, \dots, X_M\} \subseteq L^1(\Omega, \mathcal{F}, P), M \in \mathbb{N}$ on tasaisesti integroituva.*

Tod. Kun

$$\sup_{X\in\mathcal{C}} |X(\omega)| \le Y(\omega) \in L^1(P)$$

seuraa $\forall X \in \mathcal{C}$

$$|X(\omega)|\mathbf{1}(|X(\omega)| > K) \le |Y(\omega)|\mathbf{1}(|Y(\omega)| > K)$$

josta seuraa

$$\sup_{X\in\mathcal{C}} E_P\Big(|X(\omega)|\mathbf{1}(|X(\omega)| > K)\Big) \le E_P\Big(|Y(\omega)|\mathbf{1}(|Y(\omega)| > K)\Big) \downarrow 0 \text{ kun } K \uparrow \infty$$

Erityisesti kun $\mathcal{C} = \{ X_1, \dots, X_n\} \subset L^1(P)$ on äärellinen joukko, se on tasaisesti integroituva koska löytyy integroituva yläraja:

$$|X_i(\omega)| \le Y(\omega) := |X_1(\omega)| + \dots + |X_n(\omega)| \in L^1 \qquad \square$$

Huomautus: Kun $\mathcal{C}$ on ylinumeroituva, kuvaus

$$|X|^*(\omega) := \sup_{X\in\mathcal{C}}\big\{|X(\omega)|\big\}, \quad \omega \in \Omega$$

ei ole välttämättä satunnaismuuttuja, ja siksi joukko

$$A := \Big\{ \omega : \sup_{X\in\mathcal{C}} |X(\omega)| \le Y(\omega)\Big\} = \bigcup_{X\in\mathcal{C}} \Big\{ \omega : |X(\omega)| \le Y(\omega)\Big\}$$

ei ole välttämättä $\mathcal{F}$-mitallinen. Silti voidaan sanoa että $P(A) = 1$ jos on olemassa B jolla $P(B) = 1$ ja $A \subseteq B \in \mathcal{F}$. Silloin $A \in \mathcal{F}^P := \sigma(\mathcal{F}, \mathcal{N}^P)$ σ-algebraan joka on täydennetty P-nolla mittaisilla joukoilla.

Proposition 8.0.9. *(Tasaisen integroituvuuden karakterisaatio) $\mathcal{C} \subseteq L^1(P)$ on tasaisesti integroituva jos ja vain jos*

$$\sup_{X\in\mathcal{C}} E_P(|X|) < \infty \quad \textit{ja} \quad \forall\varepsilon > 0 \quad \exists\, \delta : P(A) < \delta \implies \sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}_A\big) < \varepsilon$$

Tod. Olkoon $\mathcal{C} \subseteq L^1(P)$ on tasaisesti integroituva.

$$\sup_{X\in\mathcal{C}} E(|X|\,) \le M + \sup_{X\in\mathcal{C}} E_P\Bigg(|X|\mathbf{1}(|X| > M)\Bigg) < \infty$$

Tehdään vastaoletus: on olemassa tapahtumien jono $(A_k : k \in \mathbb{N})$ ja $\varepsilon > 0$ jolla $P(A_k) \le 1/k$ ja

$$\sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}_{A_k}\big) \ge \varepsilon > 0$$

Koska

$$|X(\omega)|\mathbf{1}_{A_k} \le M\mathbf{1}_{A_k}(\omega) + |X(\omega)|\mathbf{1}(|X(\omega)| > M) \ \ \forall M \in \mathbb{N},$$

seuraa

$$\begin{aligned}
&0 < \varepsilon \le \sup_{X\in\mathcal{C}} E\big(|X|\mathbf{1}_{A_k}\big) \le MP(A_k) + \sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}(|X| > M)\big) \\
&\le MP(A_k) + \varepsilon/3 \le \varepsilon 2/3
\end{aligned}$$

kun valitaan M on tarpeeksi suureksi, ja $k \ge 3M/\varepsilon$.

Toisinpäin, jos

$$\sup_{X\in\mathcal{C}} E_P\big(|X|\big) = M < \infty,$$

Chebychevin epäyhtälön avulla seuraa $\forall X \in \mathcal{C}, k \in \mathbb{N}$

$$P(|X| > k) \le M/k$$

Oletuksen mukaan $\forall\varepsilon > 0$ on olemassa δ jolla

$$E_P\big(|X|\mathbf{1}_A\big) < \varepsilon, \quad \forall X \in \mathcal{C}, \quad \forall A \text{ jolla } P(A) < \delta\ .$$

Kun $k \ge M/\delta$, seuraa $P(|X| > k) < \delta \ \ \forall X \in \mathcal{C}$, ja

$$E_P\big(|X|\mathbf{1}(|X| > k)\big) < \varepsilon, \quad \forall X \in \mathcal{C} \qquad \square$$

Corollary 8.0.5. *$X \in L^1(\Omega, \mathcal{F}, P)$, jos ja vain jos $\forall \varepsilon > 0$ on olemassa δ, jolla kun $A \in \mathcal{F}$,*

$$P(A) < \delta \Longrightarrow E_P\big(|X|\mathbf{1}_A\big) < \varepsilon$$

Lemma 8.0.4. *Olkoon satunnaismuuttujen jonot $(X_n : n \in \mathbb{N}) \subseteq L^1(P)$ ja $(Y_n : n \in \mathbb{N}) \subseteq L^1(P)$ tasaisesti integroituvia.*

Silloin myös summien jono $(X_n + Y_n : n \in \mathbb{N})$ on tasaisesti integroituva.

Tod. Huomataan että $\forall K > 0$,

$$\begin{aligned}\big\{\omega : |X_n(\omega) + Y_n(\omega)| > K\big\} &\subseteq \big\{\omega : |X_n(\omega)| + |Y_n(\omega)| > K\big\}\\ &\subseteq \big\{\omega : |X_n(\omega)| > K/2\big\} \bigcup \big\{\omega : |Y_n(\omega)| > K/2\big\},\end{aligned}$$

Siksi

$$\begin{aligned}&E_P\Big(|X_n + Y_n|\ \mathbf{1}(|X_n + Y_n| > K)\Big) \le\\ &\le E_P\Big(|X_n + Y_n|\ \mathbf{1}(|X_n| > K/2)\Big) + E_P\Big(|X_n + Y_n|\mathbf{1}(|Y_n| > K/2)\Big)\\ &\le E_P\Big(|X_n|\mathbf{1}(|X_n| > K/2)\Big) + E_P\Big(|Y_n|\mathbf{1}(|Y_n| > K/2)\Big)\\ &+ E_P\Big(|X_n|\mathbf{1}(|Y_n| > K/2)\Big) + E_P\Big(|Y_n|\mathbf{1}(|X_n| > K/2)\Big)\end{aligned}$$

Koska satunnaismuuttujen jonot (X_n) ja (Y_n) ovat tasaisesti integroituvia, väite seuraa kun osoitamme

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} E_P\Big(|X_n|\mathbf{1}(|Y_n| > K/2)\Big) = 0 \qquad (8.0.1)$$

Koska (X_n) on tasaisesti integroituva, tasaisen integroituvuuden karakterisaatiosta (8.0.9) seuraa että $\forall\varepsilon > 0$ on olemassa δ jolla

$$P(A) < \delta \implies \sup_n E_P\big(|X_n|\mathbf{1}_A\big) < \varepsilon$$

Koska (Y_n) on tasaisesti integroituva, on olemassa K jolla

$$\sup_n P\big(|\,Y_n| > K/2\big) \le \frac{2}{K} \sup_n E_P\Big(|Y_n|\mathbf{1}(|Y_n| > K/2)\Big) < \delta$$

josta seuraa väite 8.0.1 □

Theorem 8.0.5. *($L^1(P)$-konvergenssin karakterisaatio) Olkoon satunnaismyuuttujat $\{X_n : n \in \mathbb{N}\} \subseteq L^1(\Omega, \mathcal{F}, P),\ n \in \mathbb{N}$ ja $X \in L^0(\Omega, \mathcal{F})$. Silloin*

1. $X_n \stackrel{P}{\to} X$ *ja satunnaismuuttujen jono* $\{X_n : n \in \mathbb{N}\}$ *on tasaisesti integroituva,*

   jos ja vain jos

2. $X_n \stackrel{L^1}{\to} X \in L^1(P)$*, eli*

Tod. Kun $X_n \stackrel{P}{\to} X$ lauseesta (6.1.1) seuraa että on olemassa deterministinen indeksien alijono $n(k)$ jolle $X_{n(k)}(\omega) \to X(\omega)$ P-melkein varmasti.

Soveltamalla Fatoun lemmaa (4.1.4)

$$E_P(|X|) = E_P(\liminf_k |X_{n(k)}|) \le \liminf_k E_P(|X_{n(k)}|) < \infty$$

jossa tasaisen integroituvuuden oletuksesta

$$\sup_{k\in\mathbb{N}} E_P\big(|X_{n(k)}|\big) \le M + \sup_{k\in\mathbb{N}} E_P\bigg(|X_{n(k)}|\mathbf{1}(|X_{n(k)}| > M)\bigg) < \infty$$

Siis on osoitettu että $X \in L^1(P)$. Kun $K > 0$

$$\begin{aligned}
&E_P\big(|X_n - X|\big) = E_P\bigg(|X_n - X|\mathbf{1}(|X_n - X| \le K)\bigg) + E_P\bigg(|X_n - X|\mathbf{1}(|X_n - X| > K)\bigg)\\
&= \int_0^K P(|X_n - X| > t)dt - KP(|X_n - X| > K) + E_P\bigg(|X_n - X|\mathbf{1}(|X_n - X| > K)\bigg)\\
&\le \int_0^K P(|X_n - X| > t)dt + E_P\bigg(|X_n - X|\mathbf{1}(|X_n - X| > K)\bigg),
\end{aligned}$$

jossa käytettiin Fubini lauseetta:

$$\begin{aligned}
&E_P\bigg(|X_n - X|\mathbf{1}(|X_n - X| \le K)\bigg) = \int_0^\infty P\bigg(|X_n - X|\mathbf{1}(|X_n - X| \le K) > t\bigg)dt\\
&= \int_0^K P\bigg(|X_n - X|\mathbf{1}(|X_n - X| \le K) > t\bigg)dt = \int_0^K P\big(t < |X_n - X| \le K\big)dt\\
&= \int_0^K P\big(|X_n - X| > t\big)dt - \int_0^K P\big(|X_n - X| > K\big)dt\\
&= \int_0^K P\big(|X_n - X| > t\big)dt - KP\big(|X_n - X| > K\big).
\end{aligned}$$

Koska jono $(X_n : n \in \mathbb{N})$ on tasaisesti integroituva ja $X \in L^1(P)$, lemmasta (8.0.4) seuraa että jono $(| X_n - X| \ : n \in \mathbb{N})$ on tasaisesti integroituva, eli $\forall \varepsilon$ $\exists K$ jolla

$$\sup_n E_P\Big(| X_n - X| \mathbf{1}(|X_n - X| > K)\Big) < \varepsilon$$

Koska $P(|X_n - X| > t)$ on rajoitettu, ja oletetusti $\lim\limits_{n\to\infty} P(|X_n - X| > t) = 0$ $\forall t > 0$, Lebesgue dominoidun konvergenssin lauseesta seuraa

$$\lim_{n\to\infty} \int_0^K P(|X_n - X| > t) dt = 0$$

eli on olemassa N jolla $\forall n \geq N$

$$\int_0^K P(|X_n - X| > t) dt < \varepsilon$$

josta seuraa $\forall n \geq N$

$$E_P\big(|X_n - X|\big) \leq \int_0^K P(|X_n - X| > t) dt + \sup_n E_P\Big(|X_n - X|\mathbf{1}(|X_n - X| > K)\Big) \leq 2\varepsilon \ .$$

Toisinpäin, (kts. lause 8.0.4)

$$E_P(|X_n - X|) \to 0 \quad \Longrightarrow X_n \overset{P}{\to} X \ .$$

Koska $X_n = X + (X_n - X)$, jossa oletetusti $X \in L^1(P)$, lemman (8.0.4) nojalla meidän riittää todistaa että satunnaismuuttujen kokoelma

$$\big\{ \ |X_n - X| : n \in \mathbb{N}\big\}$$

on tasaisesti integroituva. Voidaan olettaa samantien $X = 0$.

Olkoon $\varepsilon > 0$ ja N jolla $\forall n > N$

$$E_P\big(|X_n|\big) < \varepsilon$$

Koska $\{ \ X_1, \ldots, X_N\} \subset L^1(P)$ on äärellinen joukko, se on tasaisesti integroituva, ja on olemassa K jolla

$$\sup_{1\leq n\leq N} E_P\big(|X_n|\mathbf{1}(|X_n > K|)\big) < \varepsilon$$

Mutta samalle K pätee myös $\forall n \geq N$

$$E_P\big(|X_n|\mathbf{1}(|X_n| > K)\big) \leq E_P(|X_n|) < \varepsilon$$

ja siksi

$$\sup_{n\in\mathbb{N}} E_P\big(|X_n|\mathbf{1}(|X_n > K|)\big) < \varepsilon \qquad \square$$

Voidaan osoittaa että tasainen integroituvuus vastaa kompaktisuuden ehtoa $L^1(P)$ avaruudessa varustettuna heikolla topologialla. Tämä ei päde vahvemmalle $L^1(P)$-normin topologialle.

Theorem 8.0.6. *(Dunford-Pettisin lause) Satunnaismuuttujen joukko $\mathcal{C} \subseteq L^1(\Omega, \mathcal{F}, P)$ on tasaisesti integroituva P-mitan suhteen jos ja vain jos on esi-kompakti (precompact) $L^1(P)$ avaruuden heikossa topologiassa: kaikille jonolle $\{X_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$ on olemassa indeksien jono $\{n(k) : k \in \mathbb{N}\}$ ja s.m. $X \in L^1(P)$ joilla*

$$\lim_{k\to\infty} E_P\big((X_{n(k)} - X)\mathbf{1}_A\big) = 0 \quad \forall A \in \mathcal{F}$$

Remark 8.0.2. *Tästä ei seura alijonon vahvempi L^1-konvergenssi*

$$E_P(|X_{n(k)} - X|) \to 0 \ .$$

Example 8.0.6. *Olkoon $G(\omega)$ standardi gaussinen satunnaismuuttuja, jolla*

$$P(G \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2)dy$$

Määritellään satunnaismuuttujen jono

$$Z_n(\omega) = \exp\big(\sqrt{n}G(\omega) - n/2\big) > 0, \qquad n \in \mathbb{N}.$$

Huomataan että

$$z_t(x) = \exp\big(xt - t^2/2\big) = \frac{\phi(x-t)}{\phi(x)} = \frac{dP_{G+t}}{dP_G}(x), \quad \textit{jossa}$$
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{x^2}{2}\big) \textit{ on G:jakauman tiheysfunktio,}$$

on satunnaismuuttjen $(G + t)$ ja G:n jakaumien välinen uskottavuusosamäärä (Radon Nikodym derivaatta), ja $Z_n(\omega) = z_{\sqrt{n}}(G(\omega))$.

Koska Gaussisen jakauman momentti-generoiva funktio on $E_P\big(\exp(tG)\big) = \exp(t^2/2)$, $\forall t \in \mathbb{R}$, *seuraa* $E_P(Z_n) = 1\ \forall n \in \mathbb{N}$.

Osoitan että $\lim_{n\to\infty} Z_n(\omega) = 0$ *P-melkein varmasti. Siitä seuraa että satunnaismuuttujen perhe* $(Z_n : n \in \mathbb{N}) \subset L^1(P)$ *ei ole tasaisesti integroituva, koska* $E_P(Z_n) = 1 \not\to 0$.

Chebychevin epäyhtälon avulla, $\forall \alpha \in \mathbb{R}$,

$$P(Z_n > \varepsilon) \le \varepsilon^{-\alpha} E_P(Z_n^\alpha) = \varepsilon^{-\alpha} E_P\Big(\exp\big(\alpha\sqrt{n}G - \alpha n/2\big)\Big) = \varepsilon^{-\alpha}\exp(-(1-\alpha)\alpha n/2)$$

kun $\alpha \in (0,1)$,

$$P(Z_n > K^{-1}) \le K^\alpha \beta^n$$

jossa $\beta = \exp((1-\alpha)\alpha/2) \in (0,1)$.

Koska geometrinen sarja $\sum_{n=0}^{\infty} \beta^n = (1-\beta)^{-1} < \infty$ *suppenee, ensimmäisestä Borel Cantelli lemmasta (5.1.1) seuraa* $\forall K \in \mathbb{N}$

$$P\Big(\limsup_n \{\, Z_n > K^{-1}\}\Big) = 0$$

ja ottaamalla numeroituvan leikkauksen komplementtijoukoista

$$P\Big(\bigcap_{K\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \bigcap_{n\ge m} \{\, Z_n \le K^{-1}\}\Big) = 1$$

eli $Z_n(\omega) \to 0$ *P-melkein varmasti* $\square$

Remark 8.0.3. *Siis ääretönulotteisessa* $L^1(P)$ *avaruudessa rajoitettu joukko ei tarvitse olla tasaisesti integroituva, eli esi-kompakti heikon topologian suhteen. Funktionaali analyysin Banach-Alaoglun lause sanoo että suljettu yksikköpallo on kompakti niin sanotun heikko-tähti (weak star) topologian suheen.*

Remark 8.0.4. *Satunnaismuuttujen kokoelman tasainen integroituuvus koskee pelkästään satunaaismuuttujen jakaumat, niiden välinen riippuuvusrakenteella ei ole yhtään roolia.*

Theorem 8.0.7. *(Leskelän ja Viholan tasaisen integroituvuuden karakterisaatio,2011) Satunnaismuuttujen kokoelma $\mathcal{C}$ on tasaisesti integroituva jos ja vain jos on olemassa $0 \le Y(\omega) \in L^1(P)$ jolla $\forall K > 0$*

$$\sup_{X \in \mathcal{C}} E_P\Big((|X| - K)^+ \Big) \le E_P\Big((Y - K)^+ \Big)$$

jossa $x^+ = x \vee 0 = x\mathbf{1}(x > 0)$.

Tod. (Todistamme nyt $\Longleftarrow$ implikaation, todistamme $\Longrightarrow$ myöhemmin, kun konveksisuuden määritelmät ovat käytössä).

Käytämme epäyhtälöä

$$x\mathbf{1}(x > K) \le 2\big(x - K/2\big)^+, \qquad K \ge 0$$

Seuraa

$$\sup_{X \in \mathcal{C}} E_P\big(|X|\mathbf{1}(|X| > K)\big) \le 2 \sup_{X \in \mathcal{C}} E_P\big((|X| - K/2)^+\big) \le 2E_P\big((Y - K/2)^+\big) \to 0$$

kun $K \to \infty$, jossa dominoidun konvergenssi lauseen oletukset ovat voimassa:

$Y(\omega) \ge (Y(\omega) - K/2)^+ \ge 0$ jossa $(Y(\omega) - K/2)^+ \to 0$ P-melkein varmasti kun $K \to \infty$, ja yläraja $Y(\omega)$ on integroituva $\square$

Remark 8.0.5. *Kun satunnaismuuttujan $Y(\omega) \ge 0$ tulkitaan osakkeen markkinaarvoksi ja $K > 0$ on deterministinen, satunnaismuuttuja $(Y(\omega) - K)^+$ kutsutaan europpalaiseksi osto-optioksi lunastushinnalla K. Option haltijalla on oikeus mutta ei velvollisuutta ostaa yhden osakkeen ennalta sovitulla hinnalla K. Option haltijan kannattaa käyttää optionsa kun osakkeen markkinahinta on suurempi kun ennalta sovittu lunastushinta, saadakseen voittonsa $(Y(\omega) - K)^+$. Kun markkinhinta on pienempi tai yhtä kuin lunastushinta, optio on arvoton.*

8.0.3 Sovellus: odotusarvon derivointi parametrin suhteen

Lemma 8.0.5. *Olkoon $X_i : (\Omega_i, \mathcal{F}_i) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $i = 1, 2$ reaaliarvoisia satunnaismuuttujia eri todennäköisyysavaruuksissa. Silloin tulo $X(\omega_1, \omega_2) = X_1(\omega_1)X_2(\omega_2)$ on satunnaismuttuja tuloavaruudessa $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$.*

Tod. Olkoon $X_i(\omega_i) \geq 0 \ \forall \omega_i \in \Omega_i$, $i = 1, 2$
(yleisemmin voidaan ensin hajottaa $X_i = (X_i^+ - X_i^-)$). Kun $t \geq 0$

$$\big\{(\omega_1, \omega_2) : \ X_1(\omega_1) X_2(\omega_2) \leq t\big\} =$$
$$\bigcup_{0<q\in\mathbb{Q}} \left(\left\{ \omega_1 : \ X_1(\omega_1) \leq \frac{t}{q} \right\} \cap \left\{ \omega_2 : \ X_2(\omega_2) \leq q \right\} \right) \ \in (\mathcal{F}_1 \otimes \mathcal{F}_2) \quad ,$$

ja Dynkinin lemmasta (1.1.3) seuraa

$$\big\{(\omega_1, \omega_2) : \ X_1(\omega_1) X_2(\omega_2) \in B\big\} \in (\mathcal{F}_1 \otimes \mathcal{F}_2) \quad \forall B \in \mathcal{B}(\mathbb{R}) \quad \square$$

Proposition 8.0.10. *Olkoon $(\Omega, \mathcal{F}, P)$ todennäköisyysavaruus jossa*

$$\{Y(t, \omega) : t \in [a, b]\} \subseteq L^1(\Omega, \mathcal{F}, P)$$

on tasaisesti integroituva satunnaismuuttujen joukko, $a < b \in \mathbb{R}$. Oletamme sen lisäksi

- *Kaikille $\omega \in \Omega$, kuvaus $t \mapsto Y(t, \omega)$ on jatkuva.*
- *Kaikille $t \in [a, b]$ kuvaus $\omega \mapsto Y(t, \omega)$ on $\mathcal{F}$-mitallinen.*

Tästä seuraa

1. *kuvaus $(t, \omega) \mapsto Y(t, \omega)$ on $(\mathcal{B}([a, b]) \otimes \mathcal{F}) \to \mathcal{B}(\mathbb{R})$ mitallinen.*
2. *kuvaus $t \mapsto E_P(Y(t))$ on jatkuva.*
3. *Satunnaiskuvauksen*

$$X(t, \omega) := \int_a^t Y(s, \omega) ds, \ t \in [a, b].$$

odotusarvo $E_P(X(t))$ on derivoituva kaikissa $t \in (a.b)$, jatkuvalla derivaatalla

$$\frac{d}{dt} E_P\big(X(t)\big) = E_P\big(Y(t)\big) = E_P\left(\frac{d}{dt} X(t)\right)$$

Tod. Määritellään tuloavaruudessa $[a,b]\times\Omega$ satunnaismuuttujen jono

$$Y^{(N)}(t,\omega)=\sum_{k=0}^{(N-1)} Y\biggl(a+(b-a)\frac{k}{N},\omega\biggr)\,\mathbf{1}\biggl(a+(b-a)\frac{k}{N}<t\le a+(b-a)\frac{(k+1)}{N}\biggr), N\in\mathbb{N}$$

Lemma (8.0.5) nojalla seuraa $Y^{(N)}$ on $(\mathcal{B}([a,b])\otimes\mathcal{F})$-mitallinen, ja jatkuvuudesta seuraa

$$\lim_{N\uparrow\infty} Y^{(N)}(t,\omega)=Y(t,\omega)\quad \forall\omega\ ,$$

siksi $Y(t,\omega)$ on myös $(\mathcal{B}([a,b])\otimes\mathcal{F})$-mitallinen.

Koska $\lim_{s\to t} Y_s(\omega)=Y_t(\omega)$ ja tasaisen integroituvuuden oletuksesta, seuraa

$$|E_P(Y_t)-E_P(Y_s)|\le E_P|Y_t-Y_s|\to 0 \quad \text{kun } s\to t.$$

Koska $\{Y_t : t\in[a,b]\}$ on tasaisesti integroituva, seuraa että $\{Y_t : t\in[a,b]\}$ on rajoitettu $L^1(P)$ normissa, eli

$$\sup_{t\in[a,b]} E_P\bigl(|Y_t|\bigr)<\infty,$$

ja siksi $|Y(t,\omega)|\in L^1\bigl([a,b]\times\Omega,\mathcal{B}([a,b])\otimes\mathcal{F},dt\otimes P(d\omega)\bigr)$. Fubinin lause soveltuu

$$E_P(X_t)=E_P\biggl(\int_a^t Y(s)ds\biggr)=\int_{[a,b]\times\Omega} Y(s,\omega)\bigl(P(d\omega)\otimes ds\bigr)=\int_a^t E_P(Y(s))ds$$

ja koska kuvaus $t\mapsto E_P(Y(t))$ on jatkuva, analyysin keskiarvon lauseesta

$$\lim_{\Delta\to 0}\Delta^{-1}\bigl\{E_P(X_{t+\Delta})-E_P(X_t)\bigr\}$$
$$=\lim_{\Delta\to 0}\Delta^{-1}\int_t^{t+\Delta} E_P(Y(s))ds=E_P(Y(t))\quad\square$$

Example 8.0.7. *Olkoon satunnaismuuttuja $X(\omega)\in\mathbb{R}$, ja $a<0<b$ jolla*

$$m_X(t):=E_P(\exp(tX))<\infty,\quad \forall t\in[a,b].$$

Olkoon $a<-\varepsilon<0<\varepsilon<b$, ja x' yhtälön $x'/\log(x')=(b-\varepsilon)$ ratkaisu.

Koska $x\exp(\varepsilon x)\le \exp(bx)$ *kun* $x\ge x'$*, seuraa*

$$E_P(X^+\exp(\varepsilon X))\le x'\exp(\varepsilon x')+E_P(\exp(bX))<\infty$$

Vastaavasti, kun $x''/\log(x'')=-(a+\varepsilon)$*, koska* $x\exp(\varepsilon x)\le\exp(-ax)$ *kun* $x\ge x''$*, seuraa*

$$E_P(X^-\exp(-\varepsilon X))\le x''\exp(\varepsilon x'')+E_P(\exp(-aX))<+\infty\ .$$

Tästä seuraa

$$|X(\omega)|\exp(tX(\omega))\le |X(\omega)|\Big\{\exp(\varepsilon X(\omega))+\exp(-\varepsilon X(\omega))\Big\}\in L^1(P)\quad \forall t\in[-\varepsilon,\varepsilon]$$

ja kokoelma

$$\Big\{X(\omega)\exp(tX(\omega)):\ t\in[-\varepsilon,\varepsilon]\Big\}\subseteq L^1(P)$$

on tasaisesti integroituva ,

$$\frac{d}{dt}m_X(t)=E_P\left(\frac{d}{dt}\exp(tX)\right)=E_P\big(X\exp(tX)\big)\quad\forall t\in(-\varepsilon,\varepsilon).$$

Eritysesti pisteessä $t=0$

$$\frac{d}{dt}m_X(0)=E_P(X)\ .$$

Koska eskponentiaali funktio kasvaa polynoomien nopeammin, $\forall n\in\mathbb{N}$*, samoin seuraa satunnaismuuttujien joukon*

$$\Big\{X^n(\omega)\exp(tX(\omega)):\ t\in[-\varepsilon,\varepsilon]\Big\}\subseteq L^1(P)$$

tasainen integroituvuus, ja

$$\frac{d^n}{dt^n}m_X(t)=E_P\left(\frac{d^n}{dt^n}\exp(tX)\right)=E_P\big(X^n\exp(tX)\big),\quad\forall t\in(-\varepsilon,\varepsilon).$$

Eritysesti pisteessä $t=0$

$$\frac{d^n m_X}{dt^n}(0)=E_P\big(X^n\big)\ .$$

Kuvaus $m_X(t)=E_P\big(\exp(tX)\big)$ *kutsutaan momentti-generoivaksi funktioksi.*

Example 8.0.8. *(Esscherin muunnos)*

Olkoon $\Theta = \{m_X(t) = E_P(\exp(tX)) < \infty\}$. *Kun* $t \in \Theta$ *määritellään mitanvaihtokaavan kautta todennäköisyysmitta*

$$P^{(t)}(A) = \frac{E_P\big(\exp(tX)\mathbf{1}_A\big)}{m_X(t)}, \quad \forall A \in \mathcal{F} .$$

Kun on olemassa $\varepsilon > 0$ *jolle* $[t-\varepsilon, t+\varepsilon] \subseteq \Theta$*, seuraa*

$$E_{P^{(t)}}(X^n) = \frac{E_P\big(X^n \exp(tX)\big)}{m_X(t)} = \frac{1}{m_X(t)} \frac{d^n m_X}{dx^n}(t), \text{ erityisesti}$$
$$E_{P^{(t)}}(X) = \frac{d}{dt} \log(m_X(t))$$

Tod. Kuten tapauksessa $t = 0$.

Chapter 9

$L^p(\Omega, \mathcal{F}, P)$ spaces

9.0.4 Inequalities

Definition 9.0.7. *Let V be a vector, for example $V = \mathbb{R}^d$. The map $g : V \to \mathbb{R} \cup \{ +\infty\}$ is convex when*

$$g\big(px + (1-p)y\big) \quad \leq \quad pg(x) + (1-p)g(y) \quad \forall x, y \in V,\ p \in [0,1]$$

Proposition 9.0.11. *The map $g : V \to \mathbb{R} \cup \{+\infty\}$ is convex*

$$\textit{epi}(g) = \big\{\ (x,r) \in V \times \mathbb{R} : r \geq g(x)\big\}$$

is a convex set, meaning that for (x,r),$(x',r') \in$ epi(g), also their convex combination is in the epigraph:

$$(x,r)p + (x',r')(1-p) = \big(xp + x'(1-p), rp + r'(1-p)\big) \in \textit{epi}(g), \quad \forall p \in [0,1].$$

Proposition 9.0.12. *The map $g : V = \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex if and only if $\forall x \in \mathbb{R}^d$*

$$g(x) = \sup\{a + b \cdot x :\ a \in \mathbb{R}, b \in \mathbb{R}^d \textit{and } g(y) \geq a + b \cdot y \quad \forall y \in \mathbb{R}\big\}$$

where $b \cdot x$ denotes the scalar product in $\mathbb{R}^d$. This means that at each point x the function admitś a tangent $y \mapsto a + b \cdot y$ with a and b depending on x which lies below $g(y)$ at all $y \in \mathbb{R}^d$.

Proposition 9.0.13. *Let $g : \mathbb{R} \to \mathbb{R}$ a convex function. Then g is continuous and it has at every point left and right derivatives*

$$\nabla g^-(t) = \lim_{r\uparrow t} \frac{g(r) - g(t)}{r - t} \le \nabla g^+(t) = \lim_{r\downarrow t} \frac{g(r) - g(t)}{r - t}$$

which are non-decreasing: $\nabla g^\pm(s) \le \nabla g^\pm(t)$ kun $s \le t$.

Proof: for $t \le s \le r$,

$$\frac{g(s) - g(t)}{s - t} \le \frac{g(r) - g(t)}{r - t},$$

since for $p = (r - s)/(r - t) \in [0, 1]$, $s = pt + (1 - p)r$ by convexity

$$g(s) - g(r) \le \big(g(t) - g(r)\big)\ p\ .$$

This implies that $\forall t$, the sequence

$$(g(t + n^{-1}) - g(t))n \quad n \in \mathbb{N}$$

is non-increasing and therefore it has a monotone limit. Since right and left derivatives $\nabla^\pm g(t)$ exist, $g(t)$ is continuous at each $t \in \mathbb{R}$. By convexity it also follows that

$$\frac{g(s) - g(t)}{s - t} \le \frac{g(r) - g(s)}{r - s}$$

when $t \le s \le r$, which implies $\nabla^+ g(t) \le \nabla^- g(r)$ for $t < r$ $\square$

Remark 9.0.6. *Since the derivatives are non-decreasing*

1. *the cardinality of the set*

$$D := \Big\{ t :\ \nabla^+ g(t) > \nabla^- g(t) \Big\}$$

is at most countable.

2. $\forall t \in \mathbb{R}$, $\delta \in [\nabla^- g(t), \nabla^+ g(t)]$

$$g(s) = g(t) + \int_t^s \nabla^\pm g(r) dr \ge g(t) + (s - t)\delta \qquad \forall s$$

It follows that $g : \mathbb{R} \to \mathbb{R}$ is convex if and only if it is absolutely continuous w.r.t. the Lebesgue measure with non-decreasing Radon-Nikodym derivative $\frac{dg}{dx}(x) \in [\nabla^- g(t), \nabla^+ g(t)]$.

Proposition 9.0.14. *(Jensen inequality) Let $X(\omega) \in \mathbb{R}$ a random variable $E_P(|X|) < \infty$ and $g : \mathbb{R} \to \mathbb{R}$ a convex map. Then*

$$g\big(E_P(X)\big) \quad \le \quad E_P\big(g(X)\big)$$

Solution. Since g is convex, it has right and left derivatives at $\mu = E_P(X)$. Then $\delta \in [\nabla^- g(\mu), \nabla^+ g(\mu)]$

$$g(X(\omega)) \ge g(E_P(X)) + \delta\big\{X(\omega) - E_P(X(\omega))\big\}$$

and the claim follows by taking expectation □

Remark Jensen inequality holds only when we integrate w.r.t. a probability measure. We have seen that when $\nu(\mathbb{R}) < +\infty$, $\int\limits_{\mathbb{R}} |x|\nu(x) < \infty$, we have

$$\frac{1}{\nu(\mathbb{R})}\int_{\mathbb{R}} g(x)\nu(dx) \ge \nu(\mathbb{R}) g\bigg(\frac{1}{\nu(\mathbb{R})}\int_{\mathbb{R}} g(x)\mu(dx)\bigg)$$

Otherwise, when $\nu(\mathbb{R}) = +\infty$ it is possible that

$$\int_{\mathbb{R}} |x|\nu(dx) = \infty \text{ ja } \int_{\mathbb{R}} |g(x)|\nu(dx) < \infty \ .$$

Lemma 9.0.6. *When $1 \le p < r$, $L^p(\Omega, \mathcal{F}, P) \supseteq L^r(\Omega, \mathcal{F}, P)$.*

Proof. Let $X \in L^r(P)$.

For $r = \infty$, $|X(\omega)|^p \le \| X \|_\infty^p$ P-almost surely and the claim follows.

For $r < \infty$, let

$$Y_n(\omega) = n \wedge |X(\omega)|^p \in L^{r/p}(P) \ .$$

Since $0 \le Y_n(\omega) \le n$, it follows that $Y_n(\omega) \in L^1(P)$.

The map $g : \mathbb{R}^+ \to \mathbb{R}^+$ with $x \mapsto g(x) = x^{r/p}$ is convex. By Jensen inequality we have

$$E_P(Y_n^{r/p}) \ge E_P(Y_n)^{r/p} \ ,$$

and by the monotone convergence Theorem, since $0 \le Y_n(\omega) \uparrow |X(\omega)|^p \ \forall\omega$, and it follows that

$$E_P(|X|^r) \ge E_P(|X|^p)^{r/p} \ .$$

We could not apply directly Jensen inequality to the random variable $|X(\omega)|^p$ because we did not know a prior whether it belonged to the space $L^{r/p}(P)$. For this reason we used the truncated random variable $\{Y_n\}$ □

Remark 9.0.7. *For these $L^p \supseteq L^r$ inclusion, it is essential that P is a probability measure or finite measure on Ω. When ν is a measure with $\nu(\Omega) = \infty$ $L^\infty(\Omega, \mathcal{F}, \nu) \not\subseteq L^1(\Omega, \mathcal{F}, \nu)$, and truncated random variables are not necessarily ν-integrable, and the $L^p \supseteq L^r$ inclustion is false. For $\nu(\Omega) < \infty$, by using the normalized probability measure $P(A) = \nu(A)/\nu(\Omega)$ we obtain*

$$\left\{ \int_\Omega |X(\omega)|^p \nu(d\omega) \right\}^{1/p} \le \nu(\Omega)^{(r-p)/(rp)} \left\{ \int_\Omega |X(\omega)|^r \nu(d\omega) \right\}^{1/r}$$

and the inclusion $L^p(\Omega, \mathcal{F}, \nu) \supseteq L^r(\Omega, \mathcal{F}, \nu)$ is true also in this case. This inequality is useless when $\nu(\Omega) = \infty$.

Proposition 9.0.15. *(Cauchy Schwartz inequality , $p = 2$)*
Let $X(\omega), Y(\omega) \in L^2(\Omega, \mathcal{F}, P)$ be square integrable random variables. Then

1. *$(X(\omega)Y(\omega)) \in L^1(\Omega, \mathcal{F}, P)$ and*

$$\| XY \|_1 = E_P(|XY|) \le \sqrt{E_P(X^2)}\sqrt{E_P(Y^2)} = \| X \|_2 \| Y \|_2$$

where the inequality is an equality if and only if $Y(\omega) = cX(\omega)$ P-almost surely for some $c \in \mathbb{R}$. When $E_P(XY) = 0$ we say that X and Y are P-orthogonal and use the notation $X \overset{P}{\perp} Y$.

2. *The triangle inequality holds:*

$$\| X + Y \|_2 \le \| X \|_2 + \| Y \|_2$$

1. Proof. Let

$$X_n(\omega) = n \wedge |X(\omega)|, \quad Y_n(\omega) = n \wedge |Y(\omega)| \,.$$

Since $0 \le X_n(\omega), Y_n(\omega) \le n \ \forall\omega$, it follows that $(X_n(\omega)Y_n(\omega)) \in L^1(P)$.

$$\forall t \in \mathbb{R}, \quad 0 \le \Big(tX_n(\omega) + Y_n(\omega) \Big)^2 = t^2 X_n(\omega)^2 + Y_n(\omega)^2 + 2tX_n(\omega)Y_n(\omega)$$

By taking expectation (which exists at least for the truncated random variables), we see that the quadratic equation

$$t^2 E_P(X_n)^2 + 2tE_P(X_nY_n) + E(Y_n^2) = 0$$

has at most one real solution, which means that the equation has a non-positive discriminant:

$$E_P(X_nY_n)^2 - E(X_n^2)E(Y_n^2) \le 0 \ .$$

Since

$$0 \le |X_n(\omega)Y_n(\omega)| \uparrow |X(\omega)Y(\omega)| \quad \forall\omega$$

it follows by the monotine convergence Theorem that

$$E_P(|XY|)^2 - E(X^2)E(Y^2) \le 0$$

2. Proof

$$E_P((X+Y)^2) = E_P(X^2) + E_P(Y^2) + 2E_P(XY)$$
$$\le E_P(X^2) + E_P(Y^2) + 2E_P(|XY|)$$
$$\le E_P(X^2) + E_P(Y^2) + 2\sqrt{E_P(X^2)}\sqrt{E_P(Y^2)} = \left\{\sqrt{E_P(X^2)} + \sqrt{E_P(Y^2)}\right\}^2 \quad \square$$

Proposition 9.0.16. *The following identitied hold on the space $L^2(\Omega, \mathcal{F}, P)$:*

1. *For $X, Y \in L^2(P)$,* $\| X+Y \|_2^2 + \| X-Y \|_2^2 = 2 \| X \|_2^2 + 2 \| Y \|_2^2$ *(parallelogram identity)*
2. $E_P(XY) = \frac{1}{4}\big(\| X+Y \|_2^2 - \| X-Y \|_2^2\big)$ *(polarisation identity)*

Proof: exercise.

Remark We can also show that if the norm $\| x \|$ of the space satisfies the parallelogram identity, the space has a scalar product (x, y) such that $\| x \|^2 = (x, x)$.

By using the Jensen inequality, the Cauchy-Schwarz extends to $L^p(\Omega, \mathcal{F}, \mu)$ spaces, where $1 \le p < \infty$ and μ is a positive measure.

Note that $X \in L^1(\mu), Y \in L^\infty(\mu)$, Since

$$|X(\omega)Y(\omega)| \leq |X(\omega)| \parallel Y \parallel_\infty$$

it follows directly that $(XY) \in L^1(\mu)$.

Proposition 9.0.17. *Let $X \in L^p(\Omega, \mathcal{F}, \mu)$ and $Y \in L^q(\Omega, \mathcal{F}, \mu)$, $1 \leq p <= \infty$ be random variable where the exponent $q = p/(p-1)$ is conjugate to p, satisfying the relation*

$$\left(q^{-1} + p^{-1}\right) = 1 \ .$$

Then

$$\int_\Omega |X(\omega)Y(\omega)|\mu(d\omega) \leq \left\{ \int_\Omega |X(\omega)|^p \mu(d\omega) \right\}^{1/p} \left\{ \int_\Omega |X(\omega)|^q \mu(d\omega) \right\}^{1/q}$$
$$= \parallel X \parallel_{L^p(\mu)} \parallel Y \parallel_{L^p(\mu)}$$

(Hölder inequality).

Also, for $X, Y \in L^p(\Omega, \mathcal{F}, \mu)$, $1 \leq p \leq \infty$

$$\parallel X + Y \parallel_{L^p(\mu)} \leq \parallel X \parallel_{L^p(\mu)} + \parallel Y \parallel_{L^p(\mu)}$$

(Minkowski inequality).

Proof. (Hölder) Let $1 < p < \infty$. We can assume without loss of generality that $X(\omega) \geq 0$ $Y(\omega) \geq 0$ and also that $E_P(|X|^p) > 0$, otherwise $X(\omega) = 0$ P-almost surely and and the inequality would follow.

Define the random variable

$$\widetilde{Y}(\omega) := \frac{Y(\omega)}{X(\omega)^{p-1}} \mathbf{1}(X(\omega) > 0) \geq 0$$

and the probability measure

$$\widetilde{P}(d\omega) = \frac{X(\omega)^p}{\parallel X \parallel^p_{L^p(\mu)}} \mu(d\omega) \ .$$

By Jensen inequality

$$\left\{ E_{\widetilde{P}}(\widetilde{Y}) \right\}^q \leq E_{\widetilde{P}}(\widetilde{Y}^q)$$

$\forall q \geq 1$, and in particular for $q = p/(p-1)$

$$\left\{ \int_\Omega \frac{Y(\omega)}{X(\omega)^{p-1}} \mathbf{1}(X(\omega) > 0) \frac{X(\omega)^p}{\| X \|^p_{L^p(\mu)}} \mu(d\omega) \right\}^q$$
$$\leq \int_\Omega \left\{ \frac{Y(\omega)}{X(\omega)^{p-1}} \mathbf{1}(X(\omega) > 0) \right\}^q \frac{X(\omega)^p}{\| X \|^p_{L^p(\mu)}} \mu(d\omega) \Longleftrightarrow$$
$$\| X \|^{-pq}_{L^p(\mu)} \left\{ \int_\Omega Y(\omega) X(\omega) \mu(d\omega) \right\}^q \leq \| X \|^{-p}_{L^p(\mu)} \int_\Omega Y(\omega)^q X(\omega)^{(q(p-1)-p)} \mu(d\omega)$$

where $q(p-1) - p = 0$. It follows that

$$\int_\Omega Y(\omega) X(\omega) \mu(d\omega) \leq \left\{ \int_\Omega Y(\omega)^q \mu(d\omega) \right\}^{1/q} \| X \|^{(1-1/q)p}_{L^p(\mu)}$$

where $(1 - 1/q)p = 1$.

Proof. (Minkowski) Note first that $\forall x, y \geq 0$,

$$(x+y)^p \leq \big(2 \max(x,y)\big)^p \leq 2^p (x^p + y^p) \ ,$$

and for $X, Y \in L^p(\Omega, \mathcal{F}, \mu)$, we have $(X + Y) \in L^p(\mu)$ as well. By Hölder inequality it follows that

$$\int_\Omega |X+Y|^p d\mu \leq \int_\Omega |X||X+Y|^{p-1} d\mu + \int_\Omega |Y||X+Y|^{p-1} d\mu$$
$$\leq \big(\| X \|_{L^p(\mu)} + \| Y \|_{L^p(\mu)}\big) \ \| \ |X+Y|^{p-1} \ \|_{L^q(\mu)}$$
$$= \big(\| X \|_{L^p(\mu)} + \| Y \|_{L^p(\mu)}\big) \ (\| X+Y \|_{L^p(\mu)})^{p/q} \ .$$

This implies

$$\| X+Y \|^{(1-1/q)p}_{L^p(\mu)} \quad \leq \quad \| X \|_{L^p(\mu)} + \| Y \|_{L^p(\mu)}$$

where $(1 - 1/q)p = 1$ □

Proposition 9.0.18. *$\forall 1 \leq p \leq \infty$, $L^p(\Omega, \mathcal{F}, P)$ is* **complete** *:*
if $\{X_n(\omega)\} \in L^p(\Omega, \mathcal{F}, P)$ is a Cauchy sequence,

meaning that $\forall \varepsilon > 0 \ \exists N_\varepsilon$ such that $\| X_n - X_m \|_{L^p(P)} < \varepsilon \ \forall n, m \geq N_\varepsilon$,

there exists a random variable $X_\infty(\omega) \in L^p(P)$ such that $\lim_{n\uparrow\infty} \| X_n - X \|_{L^p(P)}$.

This follows also when we integrate with respect to a positive measure $\mu(d\omega)$, also when $\mu(\Omega) = +\infty$.

Proof We leave as exercise the case with $p = \infty$.

Let $p < \infty$ and $\{X_n\} \subseteq L^p$ a Cauchy sequence. There is a subsequence (k_n) such that

$$\| X_r - X_s \|_p \leq 2^{-n} \quad \forall r, s \geq k_n$$

Since $(X_n - X_{k_0} : n \geq k_0)$ is also a Cauchy sequence, without loss of generality we can assume that $X_0(\omega) = 0$ and $k_0 = 0$. Consider the "telescopic sum"

$$X_{k_n}(\omega) = \sum_{m=1}^{n} (X_{k_m}(\omega) - X_{k_{m-1}}(\omega)) \ .$$

For each $\omega \in \Omega$, we construct the monotone sequence of random variables

$$Y_n(\omega) := \sum_{m=1}^{n} \big| X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \big| \uparrow Y_\infty(\omega) = \sum_{m=1}^{\infty} \big| X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \big| \in [0, \infty] \ .$$

By the Minkowski inequality (9.0.1) it follows that

$$\| Y_n \|_{L^p} \leq \sum_{m=1}^{n} \| X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \|_{L^p}$$

and by the monotone convergence Theorem

$$\| Y_\infty \|_{L^p} \leq \sum_{m=1}^{\infty} \| X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \|_{L^p} \leq \sum_{m=0}^{\infty} 2^{-m} = 2 < \infty \ .$$

This implies that P-almost surely

$$Y_\infty(\omega) = \sum_{m=1}^{\infty} \big| X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \big| < \infty,$$

which implies that P-almost surely the sequence

$$X_\infty(\omega) = \sum_{m=1}^{\infty} \big(X_{k_m}(\omega) - X_{k_{m-1}}(\omega) \big)$$

is absolutely convergent. In order to obtain a random variable $X_\infty(\omega)$ which is well defined for all $\omega \in \Omega$, we se

$$X_\infty(\omega) := \limsup_{n \to \infty} X_{k_n}(\omega), \quad \forall \omega \in \Omega \ .$$

It follows that $X(\omega)$ is a random variable and $X_{k_n}(\omega) \to X(\omega)$ P-almost surely. When $r > k_n$

$$E_P\big(|X_r - X_{k_n}|^p\big) \le 2^{-np} \qquad \Longrightarrow$$
$$E_P\big(|X_\infty - X_{k_n}|^p\big) = E_P\big(\liminf_m |X_r - X_{k_n}|^p\big) \le \liminf_n E_P\big(|X_r - X_{k_n}|^p\big) \le 2^{-np}$$

where we have used Fatou lemma. By the Minkowski inequality it follows that $X_\infty \in L^p$, $X_r \stackrel{L^p}{\to} X_\infty$ as $r \to \infty$ $\square$

9.1 $L^2(P)$-projections

Proposition 9.1.1. *Let $H \subseteq L^2(\Omega, \mathcal{F}, P)$ be a* **closed** *linear subspace, which means that if $\{X_n\} \subseteq H$ and there is an $X \in L^2(P)$ such that $\| X_n - X \|_{L^2(P)} \to 0$, necessarily $X \in H$.*

For all$X(\omega) \in L^2(\Omega, \mathcal{F}, P)$ there exists an element of H which we call orthogonal projection of X into H and denote as $Y(\omega) = (\Pi_H X)(\omega) \in H$ such that

1. $E_P((X - Y)^2) = \Delta^2 := \inf_{W \in H} E_P((X - W)^2)$,
2. $E_P((X - Y)W) = 0, \forall\, W \in H$.

The projection $\Pi_H X$ is uniquely determined up to P-null events.

Remark 9.1.1. *Recall that in the Euclidean space $\mathbb{R}^d$, the scalar product of two vectors X, Y is defined as*

$$\langle X, Y \rangle_{\mathbb{R}^d} = \sum_{\omega=1}^{d} X(\omega)Y(\omega) = d \cdot \sum_{\omega=1}^{d} X(\omega)Y(\omega)P(\{\omega\}) = E_P(XY)d$$

where $P(\omega) = 1/d$ is the uniform probability distribution on the finite probability space $\Omega = \{1, \dots, d\}$. The vectors $X, Y \in \mathbb{R}^d$ are orthogonal (notation: $X \perp_{\mathbb{R}^d} Y$), when $\langle X, Y \rangle_{\mathbb{R}^d} = 0$. This geometrical concept generalizes to the space $L^2(\Omega, \mathcal{F}, P)$ by defining the scalar product as

$$\langle X, Y \rangle_{L^2(P)} = E_P(XY) = \int_\Omega X(\omega)Y(\omega)P(d\omega).$$

When $\#\Omega=\infty$, the space $L^2(\Omega,\mathcal{F},P)$ is infinite-dimensional. Square integrable random variables $X,Y\in L^2(P)$ are orthogonal with respect to P (notation $X\perp_P Y$), when $E_P\big(XY\big)=0$.

Proof. Since $0\in H$, $\Delta^2\le E_P(X^2)<\infty$,

and by definition of infimum there exists an approximating sequence $(Y_n:n\in\mathbb{N})\subseteq H$ such that $\| X-Y_n\|_2\to\Delta$. We show that $(Y_n:n\in\mathbb{N})$ is a Cauchy sequence. Let $\varepsilon>0$ and $\bar n$ such that $n\ge\bar n$

$$\Delta^2\le E_P((X-Y_n)^2)<\Delta^2+\varepsilon.$$

We apply the parallelogram identity to the vectors $(Y_n-Y_m)/2$ and $\big(X-(Y_n+Y_m)/2\big)$, obtaining

$$2\,\|(Y_m-Y_n)/2\|_2^2=\|X-Y_m\|_2^2+\|X-Y_n\|_2^2-2\,\|X-(Y_m+Y_n)/2\|_2^2\,.$$

Since $(Y_n+Y_m)/2\in H$,

$$\|X-(Y_n+Y_m)/2\|_2^2\ge\Delta^2,$$

and for $n,m\ge\bar n$ it follows that

$$2\,\|(Y_m-Y_n)/2\|_2^2\le 2\varepsilon.$$

This implies that $(Y_n)\subseteq H$ is a Cauchy sequence in $L^2(P)$, and since $L^2(P)$ is a complete space it follows that there exists a $Y\in L^2(P)$ such that. $Y_n\stackrel{L^2}{\to}Y$. By assumption the subspace H is closed, which implies $Y\in H$.

When $W\in H\setminus\{0\}$ and $t\in\mathbb{R}$, $(Y+tW)\in H$, and $\forall t\in\mathbb{R}$

$$\begin{aligned}&\|X-Y\|_2^2\le\|X-Y-tW\|_2^2=\|X-Y\|_2^2+t^2\,\|W\|_2^2-2tE_P((X-Y)W)\\&\iff t^2\,\|W\|_2^2\ge 2tE_P((X-Y)W)\quad\forall t\in\mathbb{R}\end{aligned}$$

which implies $E_P((X-Y)W)=0$. If $\widetilde Y(\omega)\in H$ is also a projection, by taking $W=(Y-\widetilde Y)\in H$ it follows that

$$\begin{aligned}&0=E_P(XW)-E_P(XW)=E_P(YW)-E_P(\widetilde YW)=E_P((Y-\widetilde Y)W)\\&=E_P((Y-\widetilde Y)^2)\,,\end{aligned}$$

which means that $Y(\omega)=\widetilde Y(\omega)$ P-almost surely □

Lemma 9.1.1. *The L^2-projection is a linear operator: when $X, Z \in L^2(P),\ a, b \in \mathbb{R}$ and H is a closed subspace,*

$$\Pi_H(aX + bZ) = a\Pi_H X + b\Pi_H Z,$$

We leave the proof as an exercise.

Example 9.1.1. *(Projection into the linear subspace generated by random variables): Consider $Y(\omega) \in L^2(P)$ with $E_P(Y^2) > 0$, and the define the subspace*

$$H = \textit{Linearspan}\big(1, Y\big) = \{aY(\omega) + b : a, b \in \mathbb{R}\} \subseteq L^2(P).$$

Since

$$H = \{\ a(Y(\omega) - E(Y)) + b : a, b \in \mathbb{R}\}$$

$Y(\omega)$ and $(Y(\omega) - E_P(Y))$ span the same linear subspace. Without loss of generality we can assume that $E_P(Y) = 0$.

We show that H is closed in $L^2(P)$-norm: let $X_n(\omega) = (a_n Y(\omega) + b_n) \in H$ with $(a_n), (b_n) \subseteq \mathbb{R}$, such that

$$E_P\big((X_n - X)^2\big) = E_P\big((a_n Y + b_n - X)^2\big) \longrightarrow 0$$

Then (X_n) is a Cauchy sequence in $L^2(P)$, which implied that

$$E_P\big(Y^2\big)(a_n - a_m)^2 + (b_n - b_m)^2 \to 0$$

as $n, m \to \infty$, and since $E_P(Y^2) > 0$, both (a_n) and (b_n) are Cauchy sequences in $\mathbb{R}$. $\mathbb{R}$ is also complete, and there are limits $a, b \in \mathbb{R}$ with $a_n \to a$ ja $b_n \to b$. This implies that $(a_n Y + b_n) \to (aY + b)$ in $L^2(P)$-norm, with $X(\omega) = (aY(\omega) + b)$ P-almost surely, which means that $X \in H$.

Now we construct the projection of X to the subspace H, by mimizing the $L^2(P)$-distance with respect to $a, b \in \mathbb{R}$

$$E_P\bigg(\Big\{\ aY + b - X\Big\}^2\bigg) = E(X^2) + a^2 E(Y^2) + b^2 + 2abE(Y) - 2aE(XY) - 2bE(X).$$

Since we have assumed that $E(Y) = 0$,

$$\frac{\partial}{\partial a} E_P\bigg(\Big\{\ aY + b - X\Big\}^2\bigg) = 2aE(Y^2) - 2E(XY)$$
$$\frac{\partial}{\partial b} E_P\bigg(\Big\{aY + b - X\Big\}^2\bigg) = 2b - 2E(X)$$

which implies $b = E(X)$, $a = E(Y^2)/E(XY)$. Therefore, when $E(Y) = 0$, the projection of X into the linear space H spanned by the random variables Y and the constant 1 is given by

$$\Pi_H(X) = E_P(X) + \frac{E(XY)}{E(Y^2)} Y$$

More in general, by appling the result to $\widetilde{Y} = (Y - E(Y))$ we obtain

$$\Pi_H(X) = E(X) + \frac{Cov(XY)}{Var(Y^2)} (Y - E(Y))$$

Definition 9.1.1. *Let $H \subseteq L^2(P)$ be a closed subspace. We say that a subset $\mathcal{C} \subseteq H$ is total in H, when $X \in H$ and $E_P(XY) = 0 \ \forall Y \in \mathcal{C}$ implies that $X = 0$.*

Proposition 9.1.2. *Then we have the following result: assume that the sequence $(Y_n : n \in \mathbb{N}) \subset H$ is total, and that $E_P(Y_n Y_m) = \delta_{nm}$.*

Then we say that $(Y_n : n \in \mathbb{N})$ is an orthonormal basis of H, and every $X \in H$ has the expansion

$$X = \sum_{n \in \mathbb{N}} E_P(XY_n) Y_n = \ (L^2(P)) - \lim_{K \to \infty} \sum_{n=1}^{K} E_P(XY_n) Y_n \ .$$

Proof By using the bilinearity of the scalar product $\langle X, Y \rangle_{L^2(P)} = E_P(XY)$, togheter with the orthonormality of (Y_n),

$$0 \leq E_P\left(\left\{ X - \sum_{n=1}^{K} E_P(XY_n) Y_n \right\}^2 \right) =$$

$$E_P(X^2) + \sum_{k=1}^{n} E_P(XY_n)^2 - 2 \sum_{n=1}^{K} E_P(XY_n)^2 = E_P(X^2) - \sum_{n=1}^{K} E_P(XY_n)^2$$

where the approximation error is non-decreasing w.r.t. K. This implies that

$$\sum_{n=1}^{\infty} E_P(XY_n)^2 < \infty,$$

and the sequence of approximations

$$X_K = \sum_{n=1}^{K} E_P(XY_n)Y_n$$

is a Cauchy sequence in $H \subseteq L^2(P)$. Since we have assumed that H is closed, by completeness there exists $X_\infty := \sum_{n=1}^{\infty} E_P(XY_n)Y_n \ \in H$ with $\| X_K - X_\infty \|_{L(2)} = 0$. where the limit is taken $L^2(P)$ sense.

We show that necessarily $\| X - X_\infty \|_\infty = 0$.

Let $W = (X - X_\infty \|_\infty) \in H$.

We have

$$E_P(WY_n) = E_P(XY_n) - E_P(X_\infty Y_n)$$

But for $K \geq n$

$$E_P(X_\infty Y_n) = E_P(X_K Y_n) + E_P\big(Y_n(X_\infty - X_K)\big) = E_P(XY_n) + \lim_{K\to\infty} E_P\big(Y_n(X_\infty - X_K)\big)$$

where by Cauchy Schwartz

$$|E_P\big(Y_n(X_\infty - X_K)| \leq \| Y_n \|_{L^2(P)} \| X_\infty - X_K \|_{L^2(P)} \longrightarrow 0 \text{ as } K \to \infty \ \square$$

Chapter 10

Ehdollinen odotusarvo

Olkoon $\mathcal{G} \subseteq \mathcal{F}$ ali-σ-algebra.

Silloin $L^p(\Omega, \mathcal{G}, P) \subseteq L^p(\Omega, \mathcal{F}, P)$, $\forall 0 \le p \le \infty$, ja kun $p > 0$, $L^p(\Omega, \mathcal{G}, P)$ on suljettu aliavaruus.

Tod. Olkoon $\{ X_n : n \in \mathbb{N}\} \subseteq L^p(\Omega, \mathcal{G}, P)$ ja $X \in L^p(\Omega, \mathcal{F}, P)$ joilla $E_P(| X_n - X|^p) \to 0$, Seuraa että $X \Longrightarrow X_n \stackrel{P}{\to} X$ ja on olemassa alijono jolla $\{n_k\} : X_{n_k}(\omega) \to X(\omega)$ P m.v. . Olkoon $\widetilde{X}(\omega) := \liminf_k X_{n_k}(\omega) \in L^p(\Omega, \mathcal{G}, P)$. Seuraa että $X_n \stackrel{L^P}{\to} \widetilde{X}$ ja siksi $X(\omega) = \widetilde{X}(\omega)$ P-m.v.

Kun $X \in L^2(\Omega, \mathcal{F}, P)$ ja $H = L^2(\Omega, \mathcal{G}, P)$ seuraa projektio lauseesta (9.1.1) että on olemassa ortogonaalinen projektio

$$Y(\omega) = E_P(X|\mathcal{G})(\omega) := \big(\Pi_{L^2(\Omega,\mathcal{G},P)} X\big)(\omega)$$

joka kutsutaan *ehdolliseksi odotusarvoksi* jolla

- $Y \in L^2(\Omega, \mathcal{G}, P)$
- $E_P(XW) = E_P(YW) \quad \forall W \in L^2(\Omega, \mathcal{G}, P)$

Lemma 10.0.2. *Ehdollinen odotusarvo on positiivinen operaattori:*

Olkoon $0 \le X(\omega) \in L^\infty(\Omega, \mathcal{F}, P)$, $\mathcal{G} \subseteq \mathcal{F}$ *ali-σ-algebra. Silloin*

$$Y(\omega) = E_P(X|\mathcal{G})(\omega) \ge 0 \quad P\text{-}m.v.$$

Tod. Koska $L^\infty(P) \subseteq L^2(P)$ ehdollinen odotusarvo $Y(\omega)$ on olemassa L^2-projektiona. Olkoon $A = \{\omega : Y(\omega) < 0\} \in \mathcal{G}$. Koska $\mathbf{1}_A(\omega) \in L^\infty(P) \subseteq L^2(P)$,

$$0 \le E_P(X\mathbf{1}_A) = E_P(Y\mathbf{1}_A) = E_P(Y\mathbf{1}(Y<0)) = -E_P(Y^-) \le 0$$

ja väite seuraa.

Ehdollisen odotusarvon määritelmä laajennetaan $L^1(\Omega, \mathcal{F}, P)$ avaruuteen:

Theorem 10.0.1. *(Kolmogorovin määritelmä) Kun $X \in L^1(\Omega, \mathcal{F}, P)$, ja $\mathcal{G} \subseteq \mathcal{F}$ on ali-σ-algebra, on olemassa ehdollinen odotusarvo $Y(\omega) = E_P(X|\mathcal{G})(\omega) \in L^1(\Omega, \mathcal{G}, P)$ jolla $Y \in L^1(\Omega, \mathcal{G}, P)$ ja*

$$E_P(X\mathbf{1}_A) = E_P(Y\mathbf{1}_A) \quad \forall A \in \mathcal{G}$$

Ehdollinen odotusarvo on P-m.v. yksikäsitteinen.

Tod. Voidaan olettaa että $X(\omega) \ge 0\ \forall\omega$, muuten käytämme ensin hajotelmaa $X(\omega) = X(\omega)^+ - X(\omega)^-$ ja sitten määrittelemme

$$E_P(X^+|\mathcal{G})(\omega) = E_P(X^+|\mathcal{G})(\omega) - E_P(X^-|\mathcal{G})(\omega)$$

Olkoon $0 \le X_n(\omega) = X(\omega) \wedge n \uparrow X(\omega)$ kun $n \uparrow \infty$. Koska $X_n \in L^\infty$ seuraa että $X_n \in L^2(\Omega, \mathcal{F}, P)$ ja projektio lauseesta seuraa että on olemassa $Y_n \in L^2(\Omega, \mathcal{G}, P)$ jolla

$$E_P(X_n\mathbf{1}_A) = E_P(Y_n\mathbf{1}_A) \quad \forall A \in \mathcal{G}$$

Seuraa lemmasta (10.0.2) että $Y_n(\omega) \ge 0$ P-melkein varmasti.

Kun $n \ge m$ $(X_n(\omega) - X_m(\omega)) \ge 0$ josta seuraa

$$(Y_n(\omega) - Y_m(\omega)) = E_P(X_n|\mathcal{G})(\omega) - E_P(X_m|\mathcal{G})(\omega) = E_P(X_n - X_m|\mathcal{G})(\omega) \ge 0 \qquad P\text{-m.v.}$$

Olkoon $Y(\omega) = \limsup_n Y_n(\omega)$. Seuraa että $Y_n(\omega) \uparrow Y(\omega)$ P-m.v. ja monotonisen konvergenssin lauseesta, $\forall A \in \mathcal{G}$

$$E_P(X\mathbf{1}_A) = \lim_{n\uparrow\infty} E_P(X_n\mathbf{1}_A) = \lim_{n\uparrow\infty} E_P(Y_n\mathbf{1}_A) = E_P(Y\mathbf{1}_A)$$

Jos $\widetilde{Y}(\omega) \in L^1(\Omega, \mathcal{G}, P)$ toteuttaa Kolmogorovin määritelmää, koska $A = \{\omega : Y(\omega) > \widetilde{Y}(\omega)\} \in \mathcal{G}$,

$$0 \leq E_P((Y - \widetilde{Y})\mathbf{1}_A) = E_P(X\mathbf{1}_A) - E_P(X\mathbf{1}_A) = 0$$

seuraa että $Y(\omega) \leq \widetilde{Y}(\omega)$ P-m.v., samoin seuraa että $Y(\omega) \geq \widetilde{Y}(\omega)$ ja siksi $Y(\omega) = \widetilde{Y}(\omega)$ P-m.v.

Problem 10.0.1. *Osoita että (10.0.1) pätee jos ja vain jos*

$$E_P(XW) = E_P(YW) \quad \forall W \in L^\infty(\Omega, \mathcal{G}, P)$$

Problem 10.0.2. *Olkoon $\mathcal{G} = \sigma(A_1, \dots, A_n) \subseteq \mathcal{F}$ äärellisesti generoitu ali σ-algebra, jossa $\{A_1, \dots, A_n\}$ on Ω:n $\mathcal{F}$-mitallinen ositus, $A_i \in \mathcal{F}$, $\bigcup_{i=1}^n A_i = \Omega$, $A_i \cap A_j = \emptyset$ kun $i \neq j$.*

Olkoon $X(\omega) \in L^1(\Omega, \mathcal{F}, P)$. Silloin

$$E_P(X|\mathcal{G})(\omega) = \sum_{i=1}^n \frac{E_P(X\mathbf{1}_{A_i})}{P(A_i)} \mathbf{1}_{A_i}(\omega) := \sum_{i=1}^n E_P(X|A_i)\mathbf{1}_{A_i}(\omega) \quad (10.0.1)$$

jossa $E_P(X|A) = E_P(X\mathbf{1}_A)/P(A)$ on elementaarinen ehdollinen odotusarvo, joka saa mielivaltainen arvo (esimerkiksi 0) silloin kun $P(A) = 0$. Osoita että (10.0.1) toteuttaa Kolmogorovin ehdollilsen odotuarvon määritelmän. (10.0.1) yleistyy myös tapaukseen jossa mitallinen ositus $\{A_i : i \in \mathbb{N}\}$ on numeroituva, silloin

$$E_P(X|\mathcal{G})(\omega) = \sum_{i \in \mathbb{N}} \frac{E_P(X\mathbf{1}_{A_i})}{P(A_i)} \mathbf{1}_{A_i}(\omega) := \sum_{i \in \mathbb{N}} E_P(X|A_i)\mathbf{1}_{A_i}(\omega)$$

Huom: Kolmogorovin ehdollinen odotusarvo σ-algebran ehdolla $E_P(X|\mathcal{G})(\omega)$ on satunnaismuuttuja, kun elementaarinen odotusarvo tapahtuman ehdolla $E_P(X|A)$ on vakio joka on hyvin määritelty vain silloin kun $P(A) > 0$.

10.1 Ehdollinen odotusarvo Radon-Nykodim derivaattana

Olkoon $X \in L^1(\Omega, \mathcal{F}, P)$ Määritellään merkkinen mitta

$$\mu_X(A) = E_P(X\mathbf{1}_A) \quad \forall A \in \mathcal{F}$$

Huomataan että $\mu_X(A) = 0$ silloin kun $A \in \mathcal{F}$ ja $P(A) = 0$, eli $\mu \ll P$ σ-algebrassa $\mathcal{F}$, ja $X(\omega) = \frac{d\mu_X}{dP}(\omega)$ on vastaava Radon-Nikodymin derivaatta.

Olkoon $\mathcal{G} \subseteq \mathcal{F}$ ali σ-algebra. Erityisesti $\mu \ll P$ σ-algebrassa $\mathcal{G}$. Radon-Nikodymin lauseesta seuraa että on olemassa R-N derivaatta

$$Y(\omega) := \frac{d\mu_X|_{\mathcal{G}}}{dP|_{\mathcal{G}}}(\omega)$$

jossa $Y(\omega) \in L^1(\Omega, \mathcal{G}, P)$ joka toteuttaa mitanvaihtokaavaa

$$E_P(X\mathbf{1}_A) = \mu_X(A) = E_P(Y\mathbf{1}_A) \quad \forall A \in \mathcal{G}$$

Kolmogorovin määritelmästä seuraa että $Y(\omega) = E_P(X|\mathcal{G})(\omega)$.

Siis ehdollisen odotusarvon olemassa olo seuraa R-N lauseesta. Kuitenkin, koska emme ole vielä todistaneet R-N lauseetta käytimme L^2-projektiota.

10.2 Mitä voidaan sanoa kun $E_P(|X|) = \infty$?

Olkoon $0 \leq X(\omega) \in L^0(\Omega, \mathcal{F}, P)$ mutta $E_P(X) = \infty$. Myös tässä tapauksessa monotonisen konvergenssilauseen kautta seuraa että on olemassa ehdollinen odotusarvo $Y(\omega) = E_P(X|\mathcal{G})(\omega) \in [0, +\infty]$ joka on $\mathcal{G}$-mitallinen joka totetuttaa $\forall A \in \mathcal{G}$.

$$E_P(X\mathbf{1}_A) = E_P(Y\mathbf{1}_A) \in [0, +\infty]$$

Toki $Y(\omega)$ voi saada myös arvoa $+\infty$, joka tapauksessa $E_P(Y) = E_P(X) = \infty$.

Yleisemmin olkoon $X(\omega) = (X(\omega)^+ - X(\omega)^-)$, jossa $E_P(|X|) = \infty$. Silloin ehdollinen odotusarvo

$$E_P(X|\mathcal{G})(\omega) := E_P(X^+|\mathcal{G})(\omega) - E_P(X^+|\mathcal{G})(\omega) \in [-\infty, +\infty]$$

on hyvin määritelty vain joukon

$$U := \left\{ \omega : E_P(X^+|\mathcal{G})(\omega) = E_P(X^-|\mathcal{G})(\omega) = +\infty \right\}$$

ulkopuolella. Kun käy hyvin joskus $P(U) = 0$.

10.3 Ehdollisen odotusarvon ominaisuudet

1. Monotoninen konvergenssi :

   $$0 \leq X_n(\omega) \uparrow X(\omega) \Longrightarrow 0 \leq E_P(X_n|\,\mathcal{G})(\omega) \uparrow E_P(X|\,\mathcal{G})(\omega) \qquad P \text{ m.v.}$$

2. $E_P\big(E_P(X|\mathcal{G})\big) = E_P(X)$

3. Kun $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$,

   $$E_P(X|\,\mathcal{H})(\omega) = E_P\big(E_P(X|\,\mathcal{G})\big|\mathcal{H}\big)(\omega)\ P \text{ m.v.}$$

4. Jos $Y \in L^1(\Omega, \mathcal{G}, P)$, ja $X, (XY) \in L^1(\Omega, \mathcal{F}, P)$, seuraa

   $$E_P(YX|\mathcal{G})(\omega) = Y(\omega)E_P(X|\mathcal{G})(\omega)$$

5. jos σ-algebra $\mathcal{H}$ on P-riippumaton σ-algebrasta $\sigma(X) \vee \mathcal{G}$,

   $$E_P(X|\mathcal{G} \vee \mathcal{H}) = E_P(X|\mathcal{G})$$

   Tod. $\forall G \in \mathcal{G}, H \in \mathcal{H}$, seuraa

   $$E_P(X\mathbf{1}_G\mathbf{1}_H) = E_P(X\mathbf{1}_G)P(H) = E_P\big(E_P(X|\mathcal{G})\mathbf{1}_G\big)P(H) = E_P\Big(E_P(X|\mathcal{G})\mathbf{1}_G\mathbf{1}_H\Big)$$

   ja väite seuraa koska $\mathcal{G} \vee \mathcal{H} = \sigma(G \cap H : G \in \mathcal{G}, H \in \mathcal{H})$.

6. (Jensenin epäyhtälö): Kun $E_P(X|\mathcal{G})$ on hyvin määritelty ja $g : \mathbb{R} \to \mathbb{R}$ on konveksi,

   $$E_P(g(X)|\mathcal{G}) \geq g\big(E_P(X|\mathcal{G})\big)$$

   Tod.

   $$g(X(\omega)) \geq g\big(E_P(X|\mathcal{G})(\omega)\big) + \delta(\omega)\Bigg(X(\omega) - E_P(X|\mathcal{G})(\omega)\Bigg)$$

   jossa $\delta(\omega) = \nabla g_+\big(E_P(X|\mathcal{G})(\omega)\big)$ on $\mathcal{G}$-mitallinen. Kun otetaan ehdollista odotusarvoa molemmista puolesta, väite seuraa odotusarvon positiivisuudesta.

10.4 Säännöllinen ehdollinen todennäköisyys ja ytimet

Tapahtuman $A \in \mathcal{F}$ ehdollinen todennäköisyys ehdolla $\mathcal{G}$ σ-algebra on luonnollisesti

$$P\big(A\big|\mathcal{G}\big)(\omega) = E_P\big(\mathbf{1}_A\big|\mathcal{G}\big)(\omega)$$

joka on yksikäsitteinen modulo P-nolla mittaisia joukkoja. Koska ehdollinen odotusarvo on ei-negatiivinen operaattori, seuraa $P\big(A\big|\mathcal{G}\big)(\omega) \in [0,1]$ P-melkein varmasti.

Voidaanko sanoa että P-melkein varmasti, kuvaus $A \mapsto P\big(A\big|\mathcal{G}\big)(\omega) \in [0,1]$ on todennäköisyysmitta ?

Olkoon $\{A_n\} \subseteq \mathcal{F}$ tapatuhmien jono jolla $A_n \downarrow \emptyset$. Seuraa ehdollisen odotusarvon monotonisen konvergenssin lauseesta että on olemassa joukko N jolla $P(N) = 0$

$$P(A_n|\mathcal{G})(\omega) \downarrow 0 \quad \forall \omega \in N^c \tag{10.4.1}$$

Tämä joukko voi toki riippua $\{A_n\} \subseteq \mathcal{F}$ jonosta, ja kun yleisesti jonojen määrä on ylinumeroituva, ei mikään takaa että löytyy sellainen P-nolla mittainen joukko N jossa (10.4.1) pätee samaan aikaan kaikille tapahtumien jonoille joilla $A_n \downarrow \emptyset$. Siis ehdollinen todennäköisyys ei ole automaattisesti P-melkein varmasti σ-additiivinen.

Definition 10.4.1. *Olkoon* $(\Omega, \mathcal{F})$ *ja* $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *todennäköisyysavaruudet.*

Kuvaus $K : \Omega \times \widetilde{\mathcal{F}} \to [0,1]$ *on* $(\Omega, \mathcal{F}) \to (\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *todennäköisyys ydin kun*

- *kaikille kiinnitetyillle* $\omega \in \Omega$ *kuvaus* $K(\omega, \cdot) : \widetilde{\mathcal{F}} \to [0,1]$ *jossa* $\widetilde{A} \mapsto K(\omega, \widetilde{A})$ *on todennäköisyysmitta.*

- *kaikille kiinnitetyillle tapahtumille* $\widetilde{A} \in \widetilde{\mathcal{F}}$ *kuvaus* $K(\cdot, \widetilde{A}) : \Omega \to [0,1]$ *jossa* $\omega \mapsto K(\omega, \widetilde{A})$ *on* $\mathcal{F}$*-mitallinen.*

Definition 10.4.2. *Olkoon* $\widetilde{\Omega} = \Omega$ *ja* $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$, $\mathcal{G} \subseteq \mathcal{F}$.

Ehdollisella todennäköisyydellä $(\widetilde{A},\omega)\mapsto P(\widetilde{A}|\mathcal{G})(\omega)$ *jossa* $\widetilde{A}\in\widetilde{F}$ *on säännöllinen versio jos on olemassa* $(\Omega,\mathcal{G})\to(\widetilde{\Omega},\widetilde{\mathcal{F}})$ *todennäköisyys-ydin* $K(\omega,\widetilde{\mathcal{A}})$, *joka on* $\mathcal{G}$*-mitallinen* ω*:n suhteen, ja*

$$E_P(X|\mathcal{G})(\omega)=\int_{\widetilde{\Omega}}X(\widetilde{\omega})K(\omega,d\widetilde{\omega})$$

kaikille $X\in L^1(\Omega,\widetilde{\mathcal{F}},P)$

Remark 10.4.1. *määritelmässä esintyy ali-*σ *algebra* $\widetilde{\mathcal{F}}\subseteq\mathcal{F}$ *koska joskus ehdollisen todennäköisyyden säännöllinen versio on olemassa vain jollekin pienelle* σ*-algebralle eikä alkuperäiselle* σ*-algebralle* $\mathcal{F}$*. Esimerkki:* $\widetilde{\mathcal{F}}=\sigma(X)$ *jossa* X *on (*$\mathcal{F}$*-mitallinen) reaali-arvoinen satunnaismuuttuja.*

Definition 10.4.3. *Todennäköisyysavaruus* $(\Omega,\mathcal{F})$ *on Borel jos on olemassa mitallinen injektio*

$$\Psi:(\Omega,\mathcal{F})\to\big([0,1],\mathcal{B}([0,1])\big)$$

jonka käänteiskuvaus $\Psi^{-1}:\Psi(\Omega)\to\Omega$ *on myös mitallinen.*

Theorem 10.4.1. *Olkoon* $(\Omega,\mathcal{F},P)$ *todennäköisyyskolmikko, ja* $X:(\Omega,\mathcal{F})\to(\widetilde{\Omega},\widetilde{\mathcal{F}})$ *mitallinen kuvaus, jossa* $(\widetilde{\Omega},\widetilde{\mathcal{F}})$ *on Borelin avaruus, ja* $\mathcal{G}\subseteq\mathcal{F}$ *ali* σ*-algebra. On olemassa* $(\Omega,\mathcal{G})\to(\widetilde{\Omega},\widetilde{\mathcal{F}})$ *todennäköisyys ydin* $K(\cdot,\cdot)$ *joka on ehdollisen todennäköisyyden säännöllinen versio:* P *melkein varmasti,*

$$P(X\in D|\mathcal{G})(\omega):=E_P(\mathbf{1}(X\in D)|\mathcal{G})(\omega)=K(\omega,D)\quad\text{kaikille } D\in\widetilde{\mathcal{F}}$$

Todistus (Katso myös Kallenbergin kirjasta Foundations of Modern Probability, Thm 6.3, 6.4.). Oletan ensin että $\widetilde{\Omega}=\mathbb{R}$ ja $\widetilde{\mathcal{F}}=\mathcal{B}(\mathbb{R})$. Silloin $\forall q<r\in\mathbb{Q}$, ehdollisen odotusarvon positiivisuudesta seuraa

$$P\big(X\le q\big|\,\mathcal{G}\big)(\omega)\le P\big(X\le r\big|\,\mathcal{G}\big)(\omega)\quad P\text{-melkein varmasti,}\qquad(10.4.2)$$

ja koska $\mathbb{Q}\times\mathbb{Q}$ on numeroituva on olemassa $N\subset\Omega$ jolla $P(N)=0$ ja 10.4.2 pätee samaan aikaan $\forall q\le r\in\mathbb{Q},\omega\in N^c$. Määritellän ensin $\forall q\in\mathbb{Q},\omega\in\Omega$

$$K((-\infty,q],\omega):=P\big(X\le q\big|\,\mathcal{G}\big)(\omega)\ .$$

Koska kun $\omega \in N^c$ kuvaus $\mathbb{Q} \ni q \mapsto K((-\infty, q], \omega)$ on ei-vähenevä, voidaan määritellä samaan aikaan kaikille $t \in \mathbb{R}$

$$K((-\infty, t], \omega) = \begin{cases} \lim_{q \downarrow t, q \in \mathbb{Q}} K((-\infty, q], \omega) & \text{kun } \omega \in \widetilde{N}_t^c \\ Q((-\infty, t]) & \text{kun } \omega \in \widetilde{N}_t \end{cases}$$

jossa sillon kun $\omega \in N^c$, $\forall t \; K((-\infty, t], \omega)$ on olemassa samaan aika monotonisen jonon rajaarvona, ja silloin kun $\omega \notin \mathbb{N}$ määritellään $K((-\infty, t], \omega)$ mielivaltaisesti, esimerkiksi yhtä kuin ehdotonta satunnaismuuttujan jakaumaa $Q(dx) = P\big(\{\omega : X(\omega) \in dx\}\big)$.

Kuvaus $\omega \mapsto K((-\infty, t], \omega)$ on $\mathcal{G}$-mitallinen $\forall t \in \mathbb{R}$, ja $\forall \omega \in \Omega$ kuvaus $t \mapsto K((-\infty, t], \omega))$ on todennäköisyysjakauman kertymäfunktio, koska se on ei-vähenevä, oikealta jatkuva, ja toteuttaa

$$\lim_{q \downarrow \infty, q \in \mathbb{Q}} K((-\infty, q], \omega) = K(\emptyset, \omega) = 0, \; \lim_{q \uparrow \infty, q \in \mathbb{Q}} K((-\infty, q], \omega) = K(\mathbb{R}, \omega) = 1, \quad \forall \omega \in \Omega.$$

Charatheodoryn laajennuslauseesta seuraa että $\forall \omega \in \Omega$ on olemassa yksikäsitteinen σ-additiivinen laajennus $K(B, \omega)$, joka on todennäköisyysmitta $\mathcal{B}(\mathbb{R})$ σ-algebralla. Seuraavaksi osoitamme että kaikille Borel-mitallisille testi-funktioille $f(x) \geq 0$

$$\int_{\mathbb{R}} f(x) K(dx, \omega) = E_P(f(X)|\mathcal{G})(\omega) \tag{10.4.3}$$

joka tarkoittaa että todennäköisyys ydin $K(B, \omega)$ on säännollinen versio ehdollisesta todennäköisyydestä $P(X \in B|\mathcal{G})(\omega)$.

Olkoon $\mathcal{C}$ kokoelma kaikista rajoitetuista Borel mitallisista funktioista $f : \mathbb{R} \to \mathbb{R}$ jotka toteuttavat:

1. kuvaus $\omega \mapsto \int_{\mathbb{R}} f(x) K(dx, \omega)$ on $\mathcal{G}$-mitallinen
2. 10.4.3 pätee, eli $\forall G \in \mathcal{G}$

$$E_P\Big(f(X)\mathbf{1}_G\Big) = \int_{\Omega} \Big(\int_{\mathbb{R}} f(x) K(dx, \omega) \Big) \mathbf{1}_G(\omega) P(d\omega)$$

$\mathcal{C}$ on vektori avaruus joka sisältää vakiot: kun $f, g \in \mathcal{C}$ ja $\alpha, \beta \in \mathbb{R}$, koska $\forall \omega \in \Omega$ odotusarvo on lineaarinen, integraali

$$\int_{\mathbb{R}} \big(\alpha f(x) + \beta g(x)\big) K(dx, \omega) = \alpha \int_{\mathbb{R}} f(x) K(dx, \omega) + \beta \int_{\mathbb{R}} g(x)(dx, \omega)$$

on $\mathcal{G}$-mitallinen ω:n suhteen, ja $\forall G \in \mathcal{G}$

$$E_P\Big(\big(\alpha f(X) + \beta g(X)\big) \mathbf{1}_G \Big) = \alpha E_P\big(f(X)\mathbf{1}_G\big) + \beta E_P\big(g(X)\mathbf{1}_G\big) =$$
$$= \alpha \int_\Omega \Big(\int_{\mathbb{R}} f(x) K(dx,\omega) \Big) \mathbf{1}_G(\omega) P(d\omega) + \beta \int_\Omega \Big(\int_{\mathbb{R}} g(x) K(dx,\omega) \Big) \mathbf{1}_G(\omega) P(d\omega)$$
$$= \int_\Omega \int_{\mathbb{R}} \big(\alpha f(x) + \beta g(x)\big) \mathbf{1}_G(\omega) P(d\omega)$$

$\mathcal{C}$ on suljettu monotonisen rajankäynnin suhteen: Kun $0 \le f_n(x) \uparrow f(x)$ jossa $\{ f_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$, ja $f(x)$ on rajoitettu, seuraa monotonisen konvergenssin lauseesta $\forall \omega \in \Omega$

$$0 \le \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x) K(dx,\omega) = \int_{\mathbb{R}} f(x) K(dx,\omega)$$

joka on myös $\mathcal{G}$-mitallinen, ja $\forall G \in \mathcal{G}$ monotonisen konvergenssin lauseesta seuraa

$$E_P\big(f(X)\mathbf{1}_G\big) = \lim_{n\to\infty} E_P\big(f_n(X)\mathbf{1}_G\big) = \lim_{n\to\infty} \int_\Omega \Big(\int_{\mathbb{R}} f_n(x) K(dx,\omega) \Big) \mathbf{1}_G(\omega) P(d\omega)$$
$$= \int_\Omega \lim_{n\to\infty} \Big(\int_{\mathbb{R}} f_n(x) K(dx,\omega) \Big) \mathbf{1}_G(\omega) P(d\omega) = \int_\Omega \Big(\int_{\mathbb{R}} f(x) K(dx,\omega) \Big) \mathbf{1}_G(\omega) P(d\omega)$$

Koska $\mathbf{1}_{(-\infty,q]}(\cdot) \in \mathcal{C}$ ja $\mathcal{B}(\mathbb{R}) = \sigma\big((-\infty,q], q \in \mathbb{Q}]\big)$, monotonisen luokan lauseesta 2.0.1 seuraa että $\mathcal{C}$ sisältää kaikki rajoitettuja Borel-mitallisia funktioita. Monotonisen konvergenssin lauseesta seuraa että (10.4.3) pätee myös kaikille ei-negatiivisille Borel-mitallisille funktioille. Yleisemmin, kun $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ on Borelin avaruus ja

$$\Psi : (\widetilde{\Omega}, \widetilde{\mathcal{F}}) \longrightarrow ([0,1], \mathcal{B}([0,1]))$$

on mitallinen injektio jolla on mitallinen käänteiskuvaus Ψ^{-1}, kun $A \in \widetilde{\mathcal{F}}$

$$P\big(X \in A \,|\, \mathcal{G})(\omega) = P\big(\Psi(X) \in \Psi(A) \,|\, \mathcal{G})(\omega) = K(\Psi(A), \omega)$$

jossa $\Psi(A)$ on Borelin joukko, koska Ψ^{-1} on mitallinen, ja todennäköisyys ydin

$$K(B,\omega) = P\big(\Psi(X) \in B \big| \mathcal{G}\big)(\omega) \qquad P\text{-melkein varmasti } \forall B \in \mathcal{B}(0,1)$$

on ehdollisen todennäköisyyden säännölinen versio, joka on olemassa koska $\Psi(X(\omega))$ saa arvoja välissä $[0,1] \subset \mathbb{R}$ $\square$

Remark 10.4.2. *Meidän onneksi* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *ja yleisemmin kaikki separoituvat metriset avaruudet varustettuina Borelin* σ*-algebroilla ovat Borelin avaruuksia.* $\mathbb{R}^d$*-arvoisen satunnaisvektorin ehdolliset todennäköisyydet ovat aina säännöllisiä. Separoituvassa metrisessa avaruudessa* (S, d) *on olemassa tiheä numeroituva joukko* $(x_n : n \in \mathbb{N}) \subseteq S$ *eli* $\forall x \in S$ *on olemassa indeksien jono* $(m_n : n \in \mathbb{N})$ *jolla* $d(x_{m_n}, x) \downarrow 0$ *kun* $m \uparrow \infty$*. Silloin voidaan rakentaa mitallinen inijektio*

$$\Psi : \big(S, \mathcal{B}(S)\big) \longrightarrow ([0,1], \mathcal{B}([0,1]))$$

jolla on mitallinen käänteiskuvaus seuraavasti: $\forall n \in \mathbb{N}$*, olkoon*

$$m_n := \arg\min\big\{d(x, x_k) : k = 1, \dots, 2^n\big\}$$

jossa silloin kun minimi ei ole yksikäsitteinen valitaan pienin indeksi $1 \leq m_n \leq 2^n$ *jolla*

$$d\big(x, x_{m_n}\big) = \min\big\{d(x, x_k) : k = 1, \dots, 2^n\big\}$$

Indeksilla on dyadinen esitys

$$m_n = \sum_{\ell=0}^{n-1} y_\ell^{(n)} 2^\ell$$

jossa $y^{(n)} = \big(y_0^{(n)}, \dots, y_{n-1}^{(n)}\big) \in \{0,1\}^n$*. Jokainen piste* $x \in S$ *kuvautuu binaarijonoon*

$$\eta = (y_0^{(1)}, y_0^{(2)}, y_1^{(2)}, y_0^{(3)}, y_1^{(3)}, y_2^{(3)}, \dots)$$

joka vastaa lukua $\Psi(x) \in [0,1]$*. Binaarijonon* η *informaation perusteella saadaan takaisin indeksijono* $\{m_n\}$ *ja piste* $x = \lim_{n \to \infty} x_{m_n}$*. Jää osoitettavaksi (harjoitustehtävä) että sekä* Ψ *että sen käänteiskuvaus ovat mitallisia.*

10.5 Ehdollisen odotusarvon laskenta riippumattomuuden nojalla

Proposition 10.5.1. *Todennäköisyysavaruudella* $(\Omega, \mathcal{F})$*, olkoon* $\mathcal{G} \subseteq \mathcal{F}$ *ali* σ*-algebra,* $Y(\omega)$ $\mathcal{G}$*-mitallinen satunnaismuuttuja, joka saa arvot mitallisessa avaruudessa* $(S, \mathcal{S})$*, ja olkoon* $X(\omega) \in (\widetilde{S}, \widetilde{\mathcal{S}})$ *riippumaton* $\mathcal{G}$ σ*-algebrasta.*

Olkoon $f : (\widetilde{S} \times S) \to \mathbb{R}$ rajoitettu ja Borel-mitallinen kuvaus.
Silloin ehdollisella odotusarvolla on integraali-esitys

$$E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega) = E_P\big(f(X,y)\big)\Big|_{y=Y(\omega)} = \int_{\widetilde{S}} f(x,Y(\omega))P_X(dx) \quad (10.5.1)$$

jossa $P_X(B) = P(\{\omega : X(\omega) \in B\})$.

Tod. Olkoon

$$V := \big\{ f : (\widetilde{S} \times S) \to \mathbb{R} \text{ Borel mitalliset ja rajoitetut funktiot joille pätee 10.5.1 } \big\}$$

Osoitamme ensi että V on monotoninen luokka. Ehdollisen odotusarvon määritelmästä seuraa $10.5.1$ on voimassa funktiolle $f(x,y)$ jos ja vain jos $\forall G \in \mathcal{G}$

$$E_P\big(f(X,Y)\mathbf{1}_G\big) = \int_\Omega \bigg\{ \int_{\widetilde{S}} f(x,Y(\omega))P_X(dx) \bigg\} \mathbf{1}_G(\omega)P(d\omega)$$

Selvästi V on vektori avaruus koska odotusarvo on lineaarinen. Jos $\{f_n(x,y) : n \in \mathbb{N}\} \subseteq V$ ja $0 \le f_n(x,y) \uparrow f(x,y)$ jossa $f(x,y)$ on rajoitettu, seuraa monotonisen konvergenssin lauseesta että $f(x,y) \in V$.

Monotonisen luokan lauseesta seuraa että jos $\mathcal{I} \subseteq V$ on π-luokka, V sisältää kaikki rajoitettu $\sigma(\mathcal{I})$ mitalliset funktiot. Väite on osoitettu kun näytämme että 10.5.1 pätee funktiolle $f(x,y) = \mathbf{1}_B(x)\mathbf{1}_D(y)$: $\forall G \in \mathcal{G}$ riippumattomuudesta seuraa

$$\begin{aligned}
&E_P\big(\mathbf{1}_B(X)\mathbf{1}_D(Y)\mathbf{1}_G\big) = P_X(B)P(\{Y \in D\} \cap G) \\
&= \int_\Omega \bigg\{ \int_\Omega \mathbf{1}_B(X(\widetilde{\omega}))P(d\widetilde{\omega}) \bigg\} \mathbf{1}_D(Y(\omega))\mathbf{1}_G(\omega)P(d\omega) = \\
&\int_\Omega \bigg\{ \int_\Omega \mathbf{1}_B(X(\widetilde{\omega}))\mathbf{1}_D(Y(\omega))P(d\widetilde{\omega}) \bigg\} \mathbf{1}_G(\omega)P(d\omega) = \\
&\int_\Omega \bigg\{ \int_\Omega f\big(X(\widetilde{\omega}), Y(\omega)\big)P(d\widetilde{\omega}) \bigg\} \mathbf{1}_G(\omega)P(d\omega)
\end{aligned}$$

joka tarkoittaa $\mathbf{1}_B(x)\mathbf{1}_D(y) \in V$ $\square$

10.6 Ehdollisen odotusarvon laskenta mitan-vaihdon avulla: Bayesin kaava

Lemma 10.6.1. *Ehdollinen odotusarvon on itse-adjungoitu operaattori, eli kun* $X \in L^1(\Omega, \mathcal{F}, P)$, $\mathcal{G} \subseteq \mathcal{F}$ *on ali* σ*-algebra,* $\forall A \in \mathcal{F}$

$$E_P\big(X\ E_P(\mathbf{1}_A|\mathcal{G})\big) = E_P\big(E_P(X|\mathcal{G})\ E_P(\mathbf{1}_A|\mathcal{G})\big) = E_P\big(E_P(X|\mathcal{G})\ \mathbf{1}_A\big)$$

Tod. Suoraan ehdollisen odotusarvon ominaisuuksista.

Olemme esittäneet kaksi tapausta jossa osaamme laskea ehdollisia odotusarvoja: silloin kun σ-algebralla $\mathcal{G}$ on numeroituva määrä atomeja, ja riippumattomuuden nojalla lauseessa 10.5.1.

Yleisemmin voidaan joskus paluattaa ehdollisen odotusarvon laskeminen lauseen 10.5.1 tilanteeseen mitan-vaihdon avulla. Ensin esitämme mitanvaihtokaavan ehdolliselle odotusarvolle:

Theorem 10.6.1. *(Abstrakti Bayesin kaava). Todennäköisyysavaruudella* $(\Omega, \mathcal{F})$*, olkoon* $\mathcal{G} \subseteq \mathcal{F}$ *ja* $P \overset{\mathcal{F}}{\ll} Q$ *todennäköisyysmitat joilla* $Q(A) = 0 \Longrightarrow P(A) = 0$ *kun* $A \in \mathcal{F}$*.*

Radon-Nikodym lauseesta seuraa että on olemassa Radon-Nikodym derivaatta eli satunnaismuuttuja

$$0 \le Z(\omega) := \frac{dP}{dQ}(\omega)\ \in L^1(\Omega, \mathcal{F}, Q)$$

jolle odotusarvon mitanvaihtokaava on voimassa:

$$E_P(X) = E_Q(XZ) \qquad \forall X \in L^1(\Omega, \mathcal{F}, P)$$

Silloin ehdolliselle odotusarvolle pätee Bayesin kaava:

$$E_P\big(X\big|\mathcal{G}\big)(\omega) = \frac{E_Q\big(XZ\big|\mathcal{G}\big)(\omega)}{E_Q\big(Z\big|\mathcal{G}\big)(\omega)} \quad \in L^1(\Omega, \mathcal{G}, P) \tag{10.6.1}$$

Remark 10.6.1. *Olkoon*

$$N = \big\{\ \omega : E_Q(Z|\mathcal{G})(\omega)\ = 0\big\}$$

$N \in \mathcal{G}$ koska $E_Q(Z|\mathcal{G})$ on $\mathcal{G}$-mitallinen, mitan vaihto kaavasta ja ehdollisen odotusarvon määritelmästä

$$P(N) = E_Q\big(Z\mathbf{1}_N\big) = E_Q\Big(E_Q(Z|\mathcal{G})\mathbf{1}_N\Big) = E_Q\Big(E_Q(Z|\mathcal{G})\mathbf{1}\big\{\, E_Q(Z|\,\mathcal{G}) = 0\big\}\Big) = 0$$

Seuraa että P-melkein varmasti (mutta ei välttämättä Q-melkein varmasti) $E_P(Z|\mathcal{G})(\omega) > 0$, ja yhtälön (10.6.1) väsen puoli on hyvin määritelty.

Tod. Olkoon $G \in \mathcal{G}$. Mitanvaihto kaavasta odotusarvolle ja ehdollisen odotusarvon määritelmästä seuraa

$$E_P\big(X\mathbf{1}_G\big) = E_Q\big(ZX\mathbf{1}_G\big) = E_Q\big(E_Q(ZX\mathbf{1}_G|\mathcal{G})\big) = E_Q\big(E_Q(ZX|\mathcal{G})\mathbf{1_G}\big)$$
$$= E_Q\bigg(\frac{E_Q(Z|\mathcal{G})}{E_Q(Z|\mathcal{G})}E_Q(ZX|\mathcal{G})\mathbf{1}_G\bigg) = E_Q\bigg(Z\frac{E_Q(ZX|\mathcal{G})}{E_Q(Z|\mathcal{G})}\mathbf{1}_G\bigg) = E_P\bigg(\frac{E_Q(ZX|\mathcal{G})}{E_Q(Z|\mathcal{G})}\mathbf{1}_G\bigg) \quad \square$$

Example 10.6.1. *(Perinteinen Bayesin kaava) Todennäköisyysavaruudella $(\Omega, \mathcal{F})$, olkoon ja $X(\omega) \in \mathbb{R}^d, Y(\omega) \in \mathbb{R}^m$ satunnaismuuttujia joilla $\mathcal{F} = \sigma(X, Y)$, $\mathcal{G} = \sigma(Y)$.*

Olkoon $P \overset{\mathcal{F}}{\ll} Q$ todennäköisyysmitat joilla $X \overset{Q}{\perp\!\!\!\perp} Y$ ja olkoon

$$0 \le Z(\omega) := z(X(\omega), Y(\omega)) = \frac{dP}{dQ}(\omega) \ \in L^1(\Omega, \mathcal{F}, Q)$$

jollakin Borel-mitallisilla funktiolla $z(x, y) \ge 0$. Olkoon $f(x, y)$ rajoitettu Borel-mitallinen kuvaus. Bayesin kaavasta

$$E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega) = \frac{E_Q\big(f(X,Y)Z\big|\mathcal{G}\big)(\omega)}{E_Q\big(Z\big|\mathcal{G}\big)(\omega)}$$
$$= \frac{\int_\Omega f(X(\widetilde{\omega}), Y(\omega))\ z(X(\widetilde{\omega}), Y(\omega))Q(d\widetilde{\omega})}{\int_\Omega z(X(\widetilde{\omega}), Y(\omega))Q(d\widetilde{\omega})}$$
$$= \int_\Omega f(X(\widetilde{\omega}), Y(\omega))K(\omega, d\widetilde{\omega}) \quad \text{jossa}$$
$$K(\omega, d\widetilde{\omega}) = \frac{z(X(\widetilde{\omega}), Y(\omega))}{\int_\Omega\ z(X(\omega'), Y(\omega))Q(d\omega')}Q(d\widetilde{\omega})$$

on ehdollisen todennäköisyyden ydin. Voidaan myös integroida suoraan $\mathbb{R}^d$ avaruudessa jossa $X(\omega)$ saa arvoja:

$$E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega) = \frac{\int_{\mathbb{R}^d} f(x, Y(\omega))z(x, Y(\omega))Q_X(dx)}{\int_{\mathbb{R}^d} z(x, Y(\omega))Q_X(dx)} = \int_{\mathbb{R}^d} f(x, Y(\omega))k(Y(\omega), dx)$$

jossa

$$k(y,dx) = \frac{z(x,y)}{\int_{\mathbb{R}^d} z(x',y) Q_X(dx')} Q_X(dx)$$

Kun satunnaisvektorin (X,Y) jakaumalla on tiheysfunktio $(d+m)$-ulotteisen Lebesgue mitan suhteen, siis $P(X \in dx, Y \in dy) = p_{X,Y}(x,y)dxdy$, Fubini lauseesta seuraa että silloin myös marginaalijakaumilla P_X ja P_Y ovat tiheydet,

$$P(X \in dx) = p_X(x)dx = \int_{\mathbb{R}^m} p_{X,Y}(x,y)dy$$
$$P(Y \in dy) = p_Y(y)dy = \int_{\mathbb{R}^d} p_{X,Y}(x,y)dx$$

ja voidaan valita todennäköisyysavaruudeksi $\Omega = \mathbb{R}^d \times \mathbb{R}^m$ todennäköisyysmitoilla

$$Q_{X,Y}(dx,dy) := (P_X \otimes P_Y)(dx,dy) = p_X(x)p_Y(y)dxdy, \ P_{X,Y}(dx,dy) = p_{X,Y}(x,y)dxdy$$

Oletuksesta $P_{X,Y} \ll (P_X \otimes P_Y)$, seuraa että Radon Nykodim derivaatta on

$$\frac{dP_{X,Y}}{dQ_{X,Y}}(x,y) = \frac{dP_{X,Y}}{d(P_X \otimes P_Y)}(x,y) = z(x,y) = \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$$

Voidaan silloin kirjoittaa ehdollisen todennäköisyyden ytimen tiheysfunktioiden avulla

$$k(y,dx) = \frac{z(x,y)}{\int_{\mathbb{R}^d} z(x',y) P_X(dx')} P_X(dx) = \frac{p_{X,Y}(x,y)}{p_Y(y)} dx = p_{X|Y}(x|y)dx$$

jossa viimeinen yhtälö on ehdollisen tiheysfunktion määritelmä. Perinteinen Bayesin kaava on

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y|x)}{p_Y(y)} \quad .$$

10.6.1 Ehdollisen odotusarvon laskenta tuloavaruudessa

Olkoon $(\Omega, \mathcal{F}, P)$ todennäköisyyskolmikko ja $\mathcal{H} \subseteq \mathcal{F}$, $\mathcal{G} \subseteq \mathcal{F}$.

Tuloavaruudessa $(\Omega \times \Omega)$ varustettuna tulo σ-algebralla $\mathcal{H} \otimes \mathcal{G}$ määritellään Dynkinin laajennuslauseen kautta todennäköisyysmitta $\mathbb{P}$ jolla

$$\mathbb{P}(H \times G) = P(H \cap G) \quad \forall H \in \mathcal{H}, G \in \mathcal{G}$$

Proposition 10.6.1. *Olkoon* $X(\omega,\omega') \in L^1(\Omega\times\Omega, \mathcal{H}\otimes\mathcal{G}, \mathbb{P})$.

Silloin

$$\iint\limits_{\Omega\times\Omega} X(\omega,\omega')\mathbb{P}(d\omega\times d\omega') = \int_\Omega X(\omega,\omega)P(d\omega) \tag{10.6.2}$$

Tod. Kun $X(\omega,\omega') = \mathbf{1}_H(\omega)\mathbf{1}_G(\omega')$ jossa $H\in\mathcal{H}$ ja $G\in\mathcal{G}$, väite seuraa suoraan $\mathbb{P}$:n määritelmästä. Olkoon

$$V := \Big\{ X(\omega,\omega') : \text{ rajoitetut ja } \mathcal{H}\otimes\mathcal{G}\text{-mitalliset s.m. joilla (10.6.2) on voimassa } \Big\}$$

Selvästi V on vektori avaruus, ja monotonisen konvergenssin lauseesta seuraa että V on monotoninen luokka. Koska V sisältää satunnaismuuttujat $\mathbf{1}_H(\omega)\mathbf{1}_G(\omega')$ jossa $H\in\mathcal{H}$ ja $G\in\mathcal{G}$, monotonisen luokan lauseesta seuraa sisältää myös kaikki rajoitetut $\mathcal{H}\otimes\mathcal{G}$-mitalliset satunnaismuuttujat.

Yleisemmin kun X on ei-rajoitettu ja $\mathcal{H}\otimes\mathcal{G}$-mitallinen voidaan ensin hajottaa $X = (X^+ - X^-) \in L^1(\Omega\times\Omega,\mathcal{H}\otimes\mathcal{G}, P^{\otimes 2})$ satunnais-muuttujen jonolla $X_n = (X^+\wedge n) - (X^-\wedge n)$, ja käytää monotonisen konvergenssin lausetta erikseen positiivisille ja negatiivisille puolille □

Example 10.6.2. *Olkoon* $\xi(\omega),\eta(\omega)\in\mathbb{R}$ *satunnaismuuttujat* $\mathcal{H}=\sigma(\xi)$, $\mathcal{G} = \sigma(\eta)$. *Jos* $X(\omega,\omega')\geq 0$ *on* $\sigma(\xi)\otimes\sigma(\eta)$*-mitallinen, on olemassa Borel-mitallinen kuvaus* $f:(\mathbb{R}\times\mathbb{R})\to\mathbb{R}^+$ *jolla* $X(\omega,\omega') = f(\xi(\omega),\eta(\omega'))$. *Silloin*

$$\int_\Omega f(\xi(\omega),\eta(\omega))P(d\omega) = \iint\limits_{\Omega\times\Omega} f(\xi(\omega),\eta(\omega'))\mathbb{P}(d\omega,d\omega')$$

Oletamme nyt että σ-algebrat $\mathcal{H}$ ja $\mathcal{G}$ ovat P-riippumattomia eli

$$P(H\cap G) = P(H)P(G) \text{ kun } H\in\mathcal{H} \text{ ja } G\in\mathcal{G}$$

Silloin $\mathbb{P} = P\otimes P = P^{\otimes 2}$ eli

$$\mathbb{P}(H\times G) = P(H\cap G) = P(H)P(G) \quad \forall H\in\mathcal{H}, G\in\mathcal{G}$$

Tästä esityksestä seuraa suoraan että kun $G\in\mathcal{G}$, $X(\omega,\omega) = f(\xi(\omega),\eta(\omega))$,

$$E_P(X\mathbf{1}_G) = \int_\Omega X(\omega,\omega)\mathbf{1}_G(\omega)P(d\omega) = \iint\limits_{\Omega\times\Omega} X(\omega,\omega')\mathbf{1}_G(\omega')P^{\otimes 2}(d\omega,d\omega')$$

$$= \int_\Omega \bigg\{ \int_\Omega X(\omega,\omega')P(d\omega) \bigg\} \mathbf{1}_G(\omega')P(d\omega')$$

ehdollisen odotusarvon määritelmästä seuraa

$$E_P(X|\mathcal{G})(\omega') = \int_\Omega X(\omega,\omega')P(d\omega)$$

joka vastaa kaavan (10.5.1)

Yleisemmin, kun σ-algebrat $\mathcal{H}$ ja $\mathcal{G}$ eivät ole riippumattomia P-mitan suhteen, oletamme että $\mathbb{P} \ll P^{\otimes 2}$ tulo σ-algebrassa $(\mathcal{H}\otimes\mathcal{G})$, eli $\mathbb{P}(C) = 0$ kun $C \in (\mathcal{H}\otimes\mathcal{G})$ ja $P^{\otimes 2}(C) = 0$.

Seuraa Radon-Nikodymin lauseesta että on olemassa $(\mathcal{H}\otimes\mathcal{G})$-mitallinen Radon-Nikodymin derivaatta

$$0 \le Z(\omega,\omega') := \frac{d\mathbb{P}}{dP^{\otimes 2}}(\omega,\omega') \ \in \ L^1(\Omega\times\Omega, \mathcal{H}\otimes\mathcal{G}, P^{\otimes 2})$$

jolla mitan vaihdon kaava on voimassa kaikille $X \in L^1(\Omega\times\Omega,\mathcal{H}\otimes\mathcal{G},\mathbb{P})$

$$\int_\Omega X(\omega,\omega)P(d\omega) = \iint\limits_{\Omega\times\Omega} X(\omega,\omega')\mathbb{P}(d\omega,d\omega') = \iint\limits_{\Omega\times\Omega} X(\omega,\omega')Z(\omega,\omega')P^{\otimes 2}(d\omega,d\omega') =$$
$$\int_\Omega \left(\int_\Omega X(\omega,\omega')Z(\omega,\omega')P(d\omega)\right)P(d\omega')$$

Kun $G \in \mathcal{G}$ saadaan

$$\int_\Omega X(\omega,\omega)\mathbf{1}_G(\omega)P(d\omega) =$$
$$\iint\limits_{\Omega\times\Omega} X(\omega,\omega')\mathbf{1}_G(\omega')\mathbb{P}(d\omega,d\omega') = \int_\Omega\left(\int_\Omega X(\omega,\omega')Z(\omega,\omega')P(d\omega)\right)\mathbf{1}_G(\omega')P(d\omega') =$$
$$\int_\Omega Y(\omega')\mathbf{1}_G(\omega')P(d\omega'),$$

jossa

$$Y(\omega') = \int_\Omega X(\omega,\omega')Z(\omega,\omega')P(d\omega)$$

on $\mathcal{G}$-mitallinen,

$$= \iint\limits_{\Omega\times\Omega} Z(\omega,\omega')Y(\omega')\mathbf{1}_G(\omega')P^{\otimes 2}(d\omega\times d\omega') = \iint\limits_{\Omega\times\Omega} Y(\omega')\mathbf{1}_G(\omega')\mathbb{P}(d\omega,d\omega')$$
$$= \int_\Omega Y(\omega')\mathbf{1}_G(\omega')P(d\omega')$$

Kolmogorovin ehdollisen odotusarvon määritelmästä seuraa $Y(\omega') = E_P(X|\mathcal{G})(\omega')$.

Huomataan myös että

$$K(\omega, d\omega') := Z(\omega, \omega')P(d\omega')$$

on ehdollisen todennäköisyyden ydin: $\forall A \in \mathcal{H}$, kuvaus

$$\omega \mapsto K(\omega, A) := \int_A Z(\omega, \omega')P(d\omega')$$

on $\mathcal{G}$-mitallinen, ja $\int\limits_\Omega Z(\omega, \omega')P(d\omega) \equiv 1$ koska

$$\begin{aligned}
\mathbb{P}(\Omega \times G) &= P(\Omega \cap G) = P(G) = P(\Omega)P(G) = P^{\otimes 2}(\Omega \times G) \quad \forall G \in \mathcal{G} \\
\iff P(G) &= \int_\Omega \left(\int_\Omega Z(\omega, \omega')P(d\omega) \right) \mathbf{1}_G(\omega')P(d\omega') \quad \forall G \in \mathcal{G}
\end{aligned}$$

Samoin, $\int\limits_\Omega Z(\omega, \omega')P(d\omega') \equiv 1$.

Remark 10.6.2. *Tässä kappaleessa yleistettiin lause 10.5.1 ja esimerkki 10.6.1 tilanteeseen jossa $\mathcal{H}$ ja $\mathcal{G}$ ovat yleisiä ali-σ-algebrat eikä välttämättä satunnaisvektoreiden virittämiä.*

10.7 Ehdollistaminen nollamittaisiin tapahtumiin: varoitus

Olkoon $X(\omega)$ ja $Y(\omega)$ P-riippumattomia standardi gaussisia satunnaismuuttujia,

$E_P(X) = E_P(Y) = 0$, $E_P(X^2) = E_P(Y^2) = 1$. Olkoon

$$W(\omega) = (X(\omega) - Y(\omega)), \quad Z(\omega) = \mathbf{1}(Y(\omega) \neq 0)\, \frac{X(\omega)}{Y(\omega)}$$

Olkoon $N = \{\omega : \ Y(\omega) = 0\}$. $P(N) = 0$ koska Gaussinen jakauma on absoluttisesti jatkuva Lebesgue mitan suhteen jolla $\lambda(\{y\}) = 0 \ \forall y \in \mathbb{R}$, ja

$$N^c \cap \{\omega : X(\omega) = Y(\omega)\} = N^c \cap \{\omega : W(\omega) = 0\} = N^c \cap \{\omega : Z(\omega) = 1\}$$

Olkoon $f : \mathbb{R} \to \mathbb{R}$ rajoitettu Borel mitallinen kuvaus.

$$i) \qquad E_P(f(X)|\{X = Y\}) = \frac{\iint\limits_{\mathbb{R}\times\mathbb{R}} f(x)\delta_0(x-y)p_X(x)p_Y(y)dxdy}{\iint\limits_{\mathbb{R}\times\mathbb{R}} \delta_0(x-y)p_X(x)p_Y(y)dxdy}$$

$$ii) \qquad E_P(f(X)|W = 0) = \int_{\mathbb{R}} f(x)p_{X|W}(x|0)dx$$

$$iii) \qquad E_P(f(X)|Z = 1) = \int_{\mathbb{R}} f(x)p_{X|Z}(x|1)dx$$

eivät ole valttämättä samasuuruisia, vaikka

$$\{\omega : X(\omega) = Y(\omega)\} \cap \{\omega : Y(\omega) \neq 0\} = \{\omega : W(\omega) = 0\} \cap \{\omega : Y(\omega) \neq 0\}$$
$$= \{\omega : Z(\omega) = 1\} \cap \{\omega : Y(\omega) \neq 0\}.$$

Näytämme että $i) = ii) \neq iii)$.

$i)$ Perustuu tulkintaan

$$E_P(f(X)|\{X = Y\}) := \lim_{\varepsilon\downarrow 0} E_P\big(f(X)\big|\{\,|X-Y| < \varepsilon\}\big)$$

$$= \lim_{\varepsilon\downarrow 0} \frac{E_P\big(f(X)\mathbf{1}\{\,|X-Y| < \varepsilon\}\big)}{P(|X-Y| < \varepsilon)} = \lim_{\varepsilon\downarrow 0} \frac{\int\limits_{\mathbb{R}} \left(\int_{x-\varepsilon}^{x+\varepsilon} p_Y(y)dy\right) f(x)p_X(x)dx}{\int\limits_{\mathbb{R}} \left(\int_{x-\varepsilon}^{x+\varepsilon} p_Y(y)dy\right) p_X(x)dx}$$

$$= \frac{\int\limits_{\mathbb{R}} \lim\limits_{\varepsilon\downarrow 0}\left((2\varepsilon)^{-1}\int_{x-\varepsilon}^{x+\varepsilon} p_Y(y)dy\right) f(x)p_X(x)dx}{\int\limits_{\mathbb{R}} \lim\limits_{\varepsilon\downarrow 0}\left((2\varepsilon)^{-1}\int_{x-\varepsilon}^{x+\varepsilon} p_Y(y)dy\right) p_X(x)dx}$$

$$= \frac{\int\limits_{\mathbb{R}} f(x)p_Y(x)p_X(x)dx}{\int\limits_{\mathbb{R}} p_Y(x)p_X(x)dx} = \frac{\iint\limits_{\mathbb{R}\times\mathbb{R}} f(x)\delta_0(x-y)p_X(x)p_Y(y)dxdy}{\iint\limits_{\mathbb{R}\times\mathbb{R}} \delta_0(x-y)p_X(x)p_Y(y)dxdy}$$

Tässä δ_0 on Diracin delta distribuutio jolla on ominaisuus

$$\int_{\mathbb{R}} g(x)\delta_0(x)dx = g(0) = \int_{\mathbb{R}} g(x)F(dx)$$

kaikille jatkuville funktioille g, ja $F(x) = \mathbf{1}(x \geq 0)$. Diracin δ on porrasfunktion F:n derivaatta distribution mielessä.

Lasketaan:

$i)$ $$\frac{\int_{\mathbb{R}\times\mathbb{R}} f(x)\delta_0(x-y)p_X(x)p_Y(y)dxdy}{\int_{\mathbb{R}\times\mathbb{R}} \delta_0(x-y)p_X(x)p_Y(y)dxdy} = \frac{\int_{\mathbb{R}} f(x)p_X(x)p_Y(x)dx}{\int_{\mathbb{R}} p_X(x)p_Y(x)dx} = \frac{\int_{\mathbb{R}} f(x)e^{-x^2}dx}{\int_{\mathbb{R}} e^{-x^2}dx} = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} f(x)e^{-x^2}dx\ .$$

$ii)$ Bayesin kaavasta

$$p_{X|W}(x,w) = \frac{p_X(x)p_{W|X}(x,w)}{p_W(w)} = \frac{p_X(x)p_{W|X}(x,w)}{p_W(w)} = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(w-x)^2)\Big(\frac{1}{\sqrt{4\pi}}\exp(-\frac{1}{4}w^2)\Big)^{-1}$$

koska $p_{W|X}(w|x) = p_Y(w-x)$ ja W on gaussinen ja $E(W) = E(X)-E(Y) = 0$, $E(W^2) = E(X^2) + E(Y^2)$, siis

$$p_{W|X}(w|x) = \frac{1}{\sqrt{2\pi}}\exp\big(-\frac{(w-x)^2}{2}\big)$$

joka on gaussisen jakauman $\mathcal{N}(x,1)$ tiheysfunktio.

Tästä seuraa

$$\int_{\mathbb{R}} f(x)p_{X|W}(x|0)dx = \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} f(x)e^{-x^2}dx$$

joka täsmää i) arvon kanssa.

Kuitenkin

$$p_{Z|X}(z|x) = p_Y(x/z)\left|\frac{dy}{dz}\right| = \frac{1}{\sqrt{2\pi}}\exp\big(-\frac{x^2}{2z^2}\big)\ \frac{|x|}{z^2}$$

muuttujan vaihdolla $z = x/y$, ja

$$\begin{aligned}
p_Z(z) &= \int_{\mathbb{R}} p_{Z|X}(z|x)p_X(x)dx = \frac{1}{z^2 2\pi}\int_{\mathbb{R}} |x|\exp\big(-\frac{x^2}{2}(1+z^{-2})\big)dx \\
&= \frac{1}{z^2 2\pi}2\int_0^\infty r^{1/2}\exp\big(-\frac{r}{2}(1+z^{-2})\big)\frac{r^{-1/2}}{2}dr \\
&= \frac{1}{z^2 2\pi}\int_0^\infty \exp\big(-\frac{r}{2}(1+z^{-2})\big)dr = \frac{1}{z^2 2\pi}\frac{2}{(1+z^{-2})} = \frac{1}{(1+z^2)\pi}
\end{aligned}$$

muuttujan vaihdolla $r = x^2$. Täman jakauman nimi on Student-t vapausasteella 1.

Tästä seuraa Bayesin kaavalla

$$p_{X|Z}(x|z) = \frac{p_X(x)p_{Z|X}(x|z)}{p_Z(z)} = \frac{\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,\frac{1}{\sqrt{2\pi}}\exp\big(-\frac{x^2}{2z^2}\big)\,\frac{|x|}{z^2}}{(1+z^2)^{-1}\pi^{-1}}$$
$$= \frac{(1+z^2)|x|}{2}\exp\big(-\frac{x^2}{2}(1+z^{-2})\big)$$

Kun $z = 1$ saadaan $p_{X|Z}(x|1) = |x|\exp(-x^2)$ ja

$$E_P(f(X)|Z = 1) = \int_{\mathbb{R}} f(x)p_{X|Z}(x|1)dx = \int_{\mathbb{R}} f(x)|x|\exp(-x^2)dx$$

joka on eri kuin integraalien i) ii) arvo.

Nolla-mittaisilla tapahtumilla voi olla eri esityksiä eri satunnaismuuttujen avulla, ja vastaavien ehdollisten odotusarvojen pistettäiset arvot saattaavat olla eriläisiä. Tämä ei ole ristiriidassa todennäköisyysteorian kanssa koska aina voidaan vaihtaa ehdollisen odotusarvon arvot pistettäin nolla mittaisissa joukoissa.