https://doi.org/10.26219/heal.aueb.5740

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

PROBABILISTIC MODELS IN FINANCIAL MATHEMATICS

By

Evangelia A. Kalpinelli

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2006

https://doi.org/10.26219/heal.aueb.5740

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

PROBABILISTIC MODELS IN FINANCIAL MATHEMATICS

By

Evangelia A. Kalpinelli

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
July 2006

https://doi.org/10.26219/heal.aueb.5740

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΣΤΟΧΑΣΤΙΚΑ ΠΡΟΤΥΠΑ ΣΤΑ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΑ

Ευαγγελία Καλπινέλλη

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Ιούλιος 2006

https://doi.org/10.26219/heal.aueb.5740

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

PROBABILISTIC MODELS IN FINANCIAL MATHEMATICS

Evangelia A. Kalpinelli

Approved by the Graduate Committee

N. Frangos	M. Zazanis	E. Panas
Professor	Professor	Professor
Thesis Supervisor	Members of the Committee	

Athens, July 2006

Epameinondas Panas, Professor
Director of the Graduate Program

https://doi.org/10.26219/heal.aueb.5740

DEDICATION

TO MY PARENTS

https://doi.org/10.26219/heal.aueb.5740

ACKNOWLEDGEMENTS

Special thanks to Professors Mr. Nikolaos Frangos and Mr. Michael Zazanis for their help.

https://doi.org/10.26219/heal.aueb.5740

ABSTRACT

Evangelia Kalpinelli

PROBABILISTIC MODELS IN FINANCIAL MATHEMATICS

July 2006

The aim of a Probabilistic Model (or Probability Model) is to combine the capacity of Probability Theory to handle uncertainty with the capacity of deductive logic to exploit structure. The result is a richer and more expressive formalism with a broad range of possible application areas. The difficulty with Probabilistic Models is that they tend to multiply the computational complexities of their probabilistic and logical components.

This thesis is a brief development of the Probability Theory, placing main emphasis on the mathematical rigour and on the detailed properties of particular models rather than on general concepts. The three crucial concepts in the theory of probability, those of a random variable and of the probability distribution and the characteristic function of a random variable, are systematically developed under the view of probability spaces in the first three chapters of the present thesis. Equally developed are the Central Limit Theorem and the Theory concerning Infinitively Divisible Laws.

Having introduced the fundamentals of the Probability Theory, we then introduce some other subfields of this theory, in order both to cover all the major subfields and to keep a strong link with applications in Finance. To be quite specific, all these rather sophisticated mathematical concepts that we present, such as Martingales, Brownian Motion and Stochastic Integration, are methods that financial analysts use to describe the behaviour of markets or to derive computing methods.

Finally, it is worth to mention that Probability Theory can be applied in several other areas, like Bioinformatics, Formal Epistemology, Game Theory, Psychology etc., but the presentation of these application areas is beyond the scope of this thesis.

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

ΠΕΡΙΛΗΨΗ

Ευαγγελία Καλπινέλλη

ΣΤΟΧΑΣΤΙΚΑ ΠΡΟΤΥΠΑ ΣΤΑ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΑ

Ιούλιος 2006

Ο στόχος ενός στοχαστικού προτύπου είναι να συνδυαστεί η ικανότητα της Θεωρίας Πιθανοτήτων να αντιμετωπίσει την αβεβαιότητα με την ικανότητα της παραγωγικής λογικής να προσδιορίσει τη δομή. Το αποτέλεσμα είναι ένας πλουσιότερος και πιο εκφραστικός φορμαλισμός με ένα ευρύ φάσμα τομέων εφαρμογής. Η δυσκολία με τα στοχαστικά πρότυπα είναι ότι τείνουν να πολλαπλασιάσουν την υπολογιστική πολυπλοκότητα των στοχαστικών και ντετερμινιστικών παραγόντων τους.

Αυτή η διατριβή είναι μια συνοπτική ανάπτυξη της Θεωρίας Πιθανοτήτων, που δίνει κύρια έμφαση στη μαθηματική αυστηρότητα και στις λεπτομερείς ιδιότητες ορισμένων προτύπων παρά στις γενικές έννοιες. Οι τρεις κρίσιμες έννοιες τη Θεωρίας Πιθανοτήτων, εκείνη της τυχαίας μεταβλητής και αυτές της κατανομής πιθανότητας και της χαρακτηριστικής συνάρτησης μιας τυχαίας μεταβλητής, αναπτύσσονται συστηματικά σε χώρους πιθανότητας, στα πρώτα τρία κεφάλαια της παρούσας διατριβής. Εξίσου αναπτύσσονται και το Κεντρικό Οριακό Θεώρημα και η θεωρία σχετικά με τις Απείρως Διαιρετές Κατανομές.

Στη συνέχεια, έχοντας ήδη εισάγει τις θεμελιώδεις αρχές της Θεωρίας Πιθανοτήτων, παρουσιάζουμε ορισμένα άλλα πεδία αυτής της θεωρίας, με σκοπό και να καλύψουμε τα σημαντικότερα πεδία αλλά και για να κρατήσουμε μια ισχυρή σύνδεση με τις εφαρμογές στα χρηματοοικονομικά. Για να είμαστε αρκετά συγκεκριμένοι, όλες αυτές οι μάλλον περίπλοκες μαθηματικές έννοιες τις οποίες παρουσιάζουμε, όπως τα Martingales, η Κίνηση Brown και το Στοχαστικό Ολοκλήρωμα, είναι εκείνες που οι οικονομικοί αναλυτές χρησιμοποιούν για να περιγράψουν τη συμπεριφορά των αγορών ή για να δημιουργήσουν νέες υπολογιστικές μεθόδους.

Τελικά, αξίζει να αναφέρει ότι η Θεωρία Πιθανοτήτων μπορεί να εφαρμοστεί σε διάφορες άλλες επιστήμες, όπως η Βιοπληροφορική, η Επίσημη Επιστημολογία, η Θεωρία Παιγνίων, η Ψυχολογία κ.λπ., αλλά η παρουσίαση αυτών των τομέων εφαρμογής ξεφεύγει από τους σκοπούς αυτής της εργασίας.

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

Contents

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

Chapter 1

Probability Spaces, Measures, and Random Variables

1.1 Probability Spaces as Measurable Spaces

The standard formulation of probability theory starts with a sample space Ω. Events correspond to subsets of this space. Logic dictates that if a subset A of Ω corresponds to an event then its complement, A^c, should also correspond to an event, namely the non–occurrence of A. Similarly, if A and B are events then $A \cup B$ and $A \cap B$ should also correspond to events. Families of sets are usually called classes and from the above it should be clear that the class of all events events should be a *field* of sets.

<u>**Field:**</u> Let Ω be a set and $\mathcal{A}$ a class of subsets of Ω. $\mathcal{A}$ is a *field* if

F1. $\Omega \in \mathcal{A}$.

F2. $A, B \in \mathcal{A}$ implies that $A \cup B \in \Omega$.

F3. $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$.

Note that in view of the above definition, if $\mathcal{A}$ is a field then $\emptyset = \Omega^c \in \mathcal{A}$ and if A, B, both belong to Ω then $A \cap B = (A^c \cup B^c) \in \Omega$. Also, it follows by induction that if A_i, $i = 1, 2, \ldots, n$ belong in $\mathcal{A}$, then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i$ also belong in $\mathcal{A}$ (i.e. a field is a class of subsets of Ω closed under finite unions and intersections). Note that the *set difference* of two sets in $\mathcal{A}$, defined as $A \setminus B := A \cap B^c$ and the *symmetric difference* $A \triangle B := (A \cap B^c) \cup (A^c \cap B)$ also belong to $\mathcal{A}$.

The above framework is the adequate for the simplest situations that arise in probability theory, namely those that deal with finite sample spaces. Consider for instance the problem of casting a die. A natural choice of sample space in this case would be $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$ where ω_1 corresponds to the case where the face that lands up is 1 etc. There are six "elementary events" in this case, $\{\omega_i\}$, $i = 1, 2, \ldots, 6$ and all conceivable events are unions of these[1]. There are $2^6 = 64$ possible events, including the empty set (impossible event) and the whole space (certain event). For instance, the event that the outcome is even is $\{\omega_2, \omega_4, \omega_6\}$, while the event that the outcome is greater than or equal to 5 is $\{\omega_5, \omega_6\}$.

[1]Note that the "elementary events" are not the *elements* $\omega_1, \omega_2, \ldots$, but the *sets* $\{\omega_1\}$, $\{\omega_2\}, \ldots$. Events are always subsets of Ω.

If the sample space has a finite or a countably infinite number of elements $\{\omega_1, \omega_2, \omega_3, \ldots\}$ it is possible to think in terms of elementary events $\{\omega_i\}$. In the typical case however, when the sample space is uncountably infinite, one begins with a field of subsets of Ω. When Ω is not finite it is important to be able to extend the above considerations to *sequences of events*. In particular we wish to ensure that countable intersections and unions of events are again events and this leads us to extend the notion of the field to that of the σ–field.

σ–Field: Let S be a set and $\mathcal{F}$ a field of subsets of Ω. $\mathcal{F}$ is a *σ-field* if it also satisfies

F4. If A_i, $i = 1, 2, 3, \ldots$ belong to $\mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Again we point out that $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{F}$. Thus the countable union property, together with closure under complementation and de Morgan's laws, imply closure under countable intersections as well and a σ–field is closed under countable set operations.

The following propositions are direct consequences of the definition.

Proposition 1: Let $\mathcal{F}_i$, $i \in I$ a family of σ–fields on S, where I is an index set. Then the class $\mathcal{F} := \bigcap_i \mathcal{F}_i$ is again a σ–field.

Proposition 2: The class $\mathcal{P}(S) := \{A : A \subset S\}$, i.e. the set of all subsets of Ω is a σ–field.

Let $\mathcal{C}$ a class of subsets of Ω. The σ–field it generates is the smallest σ–field that contains all its elements i.e. the *intersection* of all the σ–fields that contain $\mathcal{C}$. We know that the family of σ–fields that contain $\mathcal{C}$ is not empty since it contains at least $\mathcal{P}(S)$, the power set of Ω.

Definition 1 *Let Ω be a set and $\mathcal{F}$ a σ-field of subsets of Ω. A probability measure defined on $(\Omega,\mathcal{F})$ is a set function $P : \mathcal{F} \to [0,1]$ such that*

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

(ii) $P(A^c) = 1 - P(A)$ *for all* $A \in \mathcal{F}$,

(iii) *for all* $A_1, A_2, A_3, \ldots \in \mathcal{F}$ *with* $A_i \cap A_j = \emptyset$ *we have* $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

1.2 Sequences of Events

Let $\{A_n\}$ be a sequence of sets belonging to $\mathcal{F}$. We say that this sequence is *increasing* if $A_n \subseteq A_{n+1}$ for all n and *decreasing* if $A_n \supseteq A_{n+1}$ for all n. The limit of a monotone sequence of events is defined as $\lim_{n\to\infty} := \bigcup_{n=1}^{\infty} A_n$ for an increasing sequence $\{A_n\}$ and $\lim_{n\to\infty} := \bigcap_{n=1}^{\infty} A_n$ for a decreasing sequence.

If $\{A_n\}_{n=1,2,\ldots}$ is an increasing sequence of events, we can write $D_n = A_n \backslash A_{n-1}$, $n = 2, 3, \ldots$, $D_1 = A_1$. Note that $D_n \in \mathcal{F}$ and $D_n \bigcap D_m = \emptyset$ when $m \neq n$. Thus $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ where the D_n's are disjoint and $P(D_n) = P(A_n) - P(A_{n-1})$, $n = 2, 3, \ldots$.

$$P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} D_n) = \sum_{n=1}^{\infty} P(D_n) = P(A_1) + \sum_{n=2}^{\infty} P(A_n) - P(A_{n-1})$$

However, the last series is telescopic and has the value $\lim_n P(A_n) - P(A_1)$. Thus

$$P(\lim_n A_n) = \lim_n P(A_n)$$

for increasing sequences. The same can be shown for decreasing sequences, hence the above equality holds for all monotonic sequences of events.

If $\{A_n\}$ is a sequence of events that is not monotonic, we define its *superior and inferior limits* as

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m, \qquad \liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

The meaning of these two events can be understood as follows: $\omega \in \limsup A_n$ or $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$ means that $\omega \in \bigcup_{m=n}^{\infty} A_m$ ***for all n*** which in turn means that for every natural number n there exists another natural $n' > n$ such that $\omega \in A_{n'}$. In other words, $\omega \in \limsup_n A_n$ if there exists a subsequence (n_k) such that $\omega \in A_{n_k}$ for every k or equivalently if ω belongs to infinitely many A_n's. We also point out that the sets $B_n := \bigcup_{m=n}^{\infty} A_m$, $n = 1, 2, \ldots$, form a decreasing sequence.

Similarly, the sequence of sets $C_n = \bigcap_{m=n}^{\infty} A_m$ is an increasing sequence of sets hence $\liminf_n A_n = \bigcup_n C_n = \lim_n C_n$. Thus $\omega \in \liminf A_n$ or $\omega \in \bigcup_{n=1}^{\infty} C_n$ if there exists a natural number n such that $\omega \in C_n$, which in turn means that $\omega \in \bigcap_{m=n}^{\infty}$, i.e. that ω belongs to *all the* A_m, *for* $m \geq n$. Hence $\liminf_n A_n$ is the set of ω that belong to *all but a finite number of the* A_n*'s.*

Theorem 1 *[Borel–Cantelli] Let $\{A_n\}$ be a sequence of events such that*

$$\sum_{n=1}^{\infty} P(A_n) < \infty . \tag{1.1}$$

Then, with probability 1, only a finite number of these events occurs.

Proof: Let Ω be the probability space and define

$$\mathbf{1}_{A_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in A_i \\ 0 & \text{if } \omega \notin A_i \end{cases} .$$

Also $\{\omega : \text{a finite number of the } A_i\text{'s occur}\} = \{\omega : \sum_{n=1}^{\infty} \mathbf{1}_{A_i}(\omega) < \infty\}$. Note however that

$$\sum_{n=1}^{\infty} P(A_n) = E \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega)$$

and hence (1.1) implies that the rhs of the above equation is finite and hence that $\sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) < \infty$ w.p. 1. ♠

An alternative proof of the Borel–Cantelli lemma (as it is widely known) goes as follows. The probability that infinitely many of the events A_n occur is precisely $P(\limsup_n A_n)$ in view of the above discussion. But $\limsup_n A_n = \lim_{n\to\infty} \bigcup_{k\geq n} A_k$, hence

$$P(\limsup_n A_n) = P(\lim_{n\to\infty} \bigcup_{k\geq n} A_k) = \lim_{n\to\infty} P(\bigcup_{k\geq n} A_k) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0,$$

the last limit being zero since the series $\sum_{k=1}^{\infty} P(A_k)$ converges by assumption.

The Borel–Cantelli lemma has the following partial converse in the case where the events A_n are independent.

Theorem 2 *[* ***Second Borel-Cantelli Lemma****] If the events A_n, $n = 1, 2, \ldots$, are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ then the probability that infinitely many of the events A_n occur is 1, i.e. $P(\limsup_n A_n) = 1$.*

Proof: It suffices to show that $\lim_n P(\bigcup_{k\geq n} A_k) = 1$ or equivalently that $\lim_n P\left(\left(\bigcup_{k\geq n} A_k\right)^c\right) = 0$. Using de Morgan's laws, $\left(\bigcup_{k\geq n} A_k\right)^c = \bigcap_{k\geq n} A_k^c$ hence $\left(\bigcup_{k\geq n} A_k\right)^c \subseteq \bigcap_{k=n}^m A_k^c$ for all $m \geq n$. Thus

$$P\left(\left({}_{k\geq n}A_k\right)^c\right) \leq P\left({}_{k=n}^{m}A_k^c\right) = \prod_{k=n}^{m} P(A_k^c) = \prod_{k=n}^{m} (1 - P(A_k)),$$

where in the next to the last equality above we have used the independence of A_n. Using the inequality $1 - x \leq e^{-x}$ which is valid for all $x \in \mathbf{R}$ we have

$$P\left(\left({}_{k\geq n}A_k\right)^c\right) \leq e^{-\sum_{k=n}^{m} P(A_k)}, \qquad \text{for all } m \geq n.$$

However, since the series $\sum_{k=1}^{\infty} P(A_k)$ diverges it follows that $\lim_{m\to\infty} \sum_{k=n}^{m} P(A_k) = \infty$ and hence, letting $m \to \infty$ we obtain

$$P\left(\left({}_{k\geq n}A_k\right)^c\right) = 0,$$

whence $P(\liminf_n A_n) = \lim_{n\to\infty} P\left({}_{k\geq n}A_k\right) = 1$. ♠

1.3 Convergence Concepts for Sequences of Random Variables

Let $\{X_n\}$ be a sequence of real random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Seeing that such a random variable is in fact a *measurable function* from Ω to $\mathbb{R}$ (we write $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ–field in $\mathbb{R}$) we realize that the issue of convergence of a sequence of random variables is the same as that of a sequence of real functions defined on an measure space.

1.3.1 Convergence in Probability and Convergence with Probability 1

Definition 2 *The sequence $\{X_n\}$ converges in probability to the random variable X if* $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0. \tag{1.2}$$

(Note that (1.2) is shorthand for the statement $\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$.) To signify that X_n converges to X in probability we often write $X_n \xrightarrow{\text{P}} X$.

Equivalently we may say that, for every $\epsilon > 0$, $\delta > 0$, there exists n_0 such that $P(|X_n - X| > \epsilon) < \delta$ for all $n \geq n_0$.

Definition 3 *The sequence $\{X_n\}$ converges to the random variable X* with probability 1 *if there exists a set Λ such that $P(\Lambda) = 0$ and for all $\omega \notin \Lambda$, $X_n(\omega) \to X(\omega)$.*

The above is *pointwise convergence* for all ω not in Λ and is usually denoted as $X_n \to X$ w.p. 1 (with probability 1) or a.s. (almost surely). Equivalently we may write $P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$ or simply $P(X_n \to X) = 1$.

In order to understand the connection between the two modes of convergence we have discussed so far let us examine closely the definition of a.s. convergence. The set on which X_n converges pointwise, i.e. the set $\{\omega : X_n(\omega) \to X(\omega)\}$ can be written as

$$\{\omega : \forall\epsilon > 0\ \exists n_0(\omega, \epsilon) \text{ such that } |X_m(\omega) - X(\omega)| < \epsilon \text{ for all } m \geq n_0(\epsilon)\}$$

https://doi.org/10.26219/heal.aueb.5740

or, equivalently,

$$\bigcap_{\epsilon>0}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{\omega : |X_m(\omega) - X(\omega)| < \epsilon\}.$$

Let $A_\epsilon := \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{\omega : |X_m(\omega) - X(\omega)| < \epsilon\}$. If $\epsilon_1 < \epsilon_2$, then $A_{\epsilon_1} \subseteq A_{\epsilon_2}$ Also, nothing is lost if we let $\epsilon = 1/k$ where $k \in \mathbb{N}$ and we can thus say that the set on which X_n converges to X is the set

$$\lim_{k\to\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{\omega : |X_m(\omega) - X(\omega)| < \frac{1}{k}\}$$

or, equivalently, $\lim_{k\to\infty}\liminf_{m\to\infty}\{|X_m - X| < 1/k\}$. Convergence with probability 1 is equivalent to the condition

$$\lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty}\{|X_m - X| > \epsilon\}\right) = 0. \tag{1.3}$$

From the above discussion we see that X_n *converges in probability to* X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \tag{1.4}$$

whereas X_n converges to 0 *with probability 1* if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(\sup_{k\geq n}|X_k - X| > \epsilon) = 0. \tag{1.5}$$

It should be clear from the above that convergence with probability 1 is stronger: it implies convergence in probability, while convergence in probability does not imply convergence w.p.1. Similarly, X_n converges w.p.1 to X iff $\lim_{n\to\infty} P(\sup_{k\geq n}|X_k - X| > \epsilon) = 0$. Convergence with probability 1 is also referred to as almost sure (abbreviated a.s.) convergence.

Before we move further, let us consider the following examples:

Example 1: Suppose that $\{X_n; n \in \mathbb{N}\}$ is a sequence of independent Bernoulli random variables with $P(X_n = 0) = 1 - \frac{1}{n}$, $P(X_n = 1) = \frac{1}{n}$. It is easy to see that X_n converges to 0 in probability. Indeed, for any $\epsilon > 0$, $P(|X_n| > \epsilon) \leq \frac{1}{n} \to 0$, and hence (1.4) is satisfied. On the other hand we can see that X_n does not converge to 0 w.p.1. Indeed, if $\epsilon \in (0, 1)$, then

$$\{\sup_{k\geq n}|X_k| > \epsilon\} = \bigcup_{k=n}^{\infty}\{X_k = 1\}$$

and hence, by de Morgan's rule

$$P(\sup_{k\geq n}|X_k| > \epsilon) = P\left(\bigcup_{k=n}^{\infty}\{X_k = 1\}\right) = 1 - P\left(\bigcap_{k=n}^{\infty}\{X_k = 0\}\right) = 1 - \prod_{k=n}^{\infty}\left(1 - \frac{1}{k}\right)$$

However, the infinite product above is equal to zero. (Actually one says that this infinite product *diverges to zero.*) Indeed,

$$\prod_{k=n}^{\infty}\left(1 - \frac{1}{k}\right) = \lim_{m\to\infty}\prod_{k=n}^{m}\left(1 - \frac{1}{k}\right) = \lim_{m\to\infty}\frac{n-1}{n}\frac{n}{n+1}\cdots\frac{m-2}{m-1}\frac{m-1}{m} = \lim_{m\to\infty}\frac{n-1}{m} = 0.$$

Hence $P(\sup_{k\geq n}|X_k| > \epsilon) = 1$ and, as a result, (1.5) is not satisfied. Let us appraise this situation: If we make n large enough we can make the probability $P(X_n = 1)$ arbitrarily close to zero, i.e. we can, in the limit be sure that $X_n = 0$. However, since $P(\bigcup_{k=n}^{\infty}\{X_k = 1\}) = 1$ we can also be sure

https://doi.org/10.26219/heal.aueb.5740

that, no mater how large we take n to be, there will be another 1 in the sequence. Put differently, the total number of 1's in the sequence is infinite with probability 1.

Example 2: Suppose now that, in the previous example, $P(X_n = 1) = \frac{1}{n^2}$. Again it is easy to see that X_n converges in probability to 0. This time we will also show that it converges to 0 w.p.1. The same calculations as above apply but this time we have

$$P(\sup_{k\geq n} |X_k| > \epsilon) = 1 - \prod_{k=n}^{\infty}\left(1 - \frac{1}{k^2}\right),$$

and

$$\begin{aligned}\prod_{k=n}^{\infty}\left(1-\frac{1}{k^2}\right) &= \lim_{m\to\infty}\prod_{k=n}^{m}\left(\frac{k^2-1}{k^2}\right)\\ &= \lim_{m\to\infty}\frac{(n-1)(n+1)}{n^2}\frac{n(n+2)}{(n+1)^2}\cdots\frac{(m-2)m}{(m-1)^2}\frac{(m-1)(m+1)}{m^2}\\ &= \lim_{m\to\infty}\frac{(n-1)(m+1)}{nm} = \frac{n-1}{n}.\end{aligned}$$

Hence

$$\lim_{n\to\infty} P(\sup_{k\geq n}|X_k| > \epsilon) = 1 - \frac{n-1}{n} = 0$$

which (according to (1.5) establishes convergence w.p.1. Unlike example 1, here we see that the total number of 1's in the sequence is finite with probability 1.

There is however a case where convergence in probability implies convergence with probability 1. Suppose that $\{Y_n\}$ converges monotonically to Y in probability. To start with the simplest case, assume that $0 \leq Y_{n+1} \leq Y_n$ for all n and $Y_n \xrightarrow{P} 0$. Because of monotonicity

$$\sup_{m\geq n} Y_m = Y_n$$

hence

$$P(\sup_{m\geq n} Y_m > \epsilon) = P(Y_n > \epsilon) \to 0$$

which implies that $Y_n \to 0$ w.p. 1.

The above result generalizes immediately to the case where either $Y_{n+1} \leq Y_n$ for all n or $Y_{n+1} \geq Y_n$ for all n and $Y_n \xrightarrow{P} Y$ by considering the sequence $\tilde{Y}_n = |Y_n - Y|$. Note that in both cases $\tilde{Y}_n$ is decreasing and by definition converges to zero in probability. Hence

$$\lim_{n\to\infty} P(\sup_{m\geq n}|Y_m - Y| > \epsilon) = \lim_{n\to\infty} P(|Y_n - Y| > \epsilon) = 0.$$

1.3.2 Convergence in the L^p sense

Let $\{X_n\}$, $n = 1, 2, \ldots$, be a sequence of real random variables such that $E|X_n|^p < \infty$ where $p \geq 1$. We say that X_n converges to X in L^p (write $X_n \xrightarrow{L^p} X$) if

$$\lim_{n\to\infty} E|X_n - X|^p = 0.$$

 https://doi.org/10.26219/heal.aueb.5740

The case $p = 2$ is of particular importance and L^2 convergence it is often referred to as *mean square* (m.s.) convergence.

It is easy to see that convergence in L^p implies convergence in probability. For this we shall need the following basic inequality (known as the Markov inequality).

Theorem 3 *Let $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ an increasing function and Y a real random variable such that $E\phi(|Y|) < \infty$. Then, for any $\alpha > 0$,*

$$P(|Y| > \alpha) \leq \frac{E\phi(|Y|)}{\phi(\alpha)}. \tag{1.6}$$

A particular choice of the function ϕ that is often useful is $\phi(x) = x^p$ with $p \geq 1$ which gives a bound on the tail of the distribution in terms of its moments.

Proof: It suffices to observe that

$$\begin{aligned} E\phi(|Y|) &= E[\phi(|Y|)\mathbf{1}(|Y| \leq \alpha)] + E[\phi(|Y|)\mathbf{1}(|Y| > \alpha)] \\ &\geq E[\phi(|Y|)\mathbf{1}(|Y| > \alpha)] \\ &\geq \phi(\alpha)E[\mathbf{1}(|Y| > \alpha)] = \phi(\alpha)P(|Y| > \alpha) \end{aligned}$$

where in the above inequalities we have used the fact that ϕ takes nonnegative values and that it is increasing. ♠

Hence, applying the above inequality with $\phi(x) = x^p$ we obtain

$$P(|X_n - X| > \epsilon) \leq \frac{E|X_n - X|^p}{\epsilon^p}. \tag{1.7}$$

If $X_n \xrightarrow{L^p} X$ then the numerator on the right hand side of (1.7) goes to 0 as $n \to \infty$, hence we have $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. Thus convergence in L^p implies convergence in probabiliy.

The relationship between convergence in L^p and a.s. convergence is more complicated. Neither one implies the other, unless certain extra conditions are satisfied.

Finally, one important result which will be used in the sequel is the following.

Theorem 4 *If a sequence of random variables $\{X_n\}$ converges to X in probability then there exists a subsequence n_k such that $X_{n_k} \to X$ w.p. 1.*

Proof: Since X_n converges in probability to X, for every k there exists n_k such that

$$P(|X_{n_k} - X| > 2^{-k}) < 2^{-k}.$$

Call A_k the event $\{\omega : |X_{n_k}(\omega) - X(\omega)| > 2^{-k}\}$. Since $\sum_{k=1}^{\infty} P(A_k) < \sum_{k=1}^{\infty} 2^{-k} < \infty$ the Borel–Cantelli theorem assures us that, with probability one, only finitely many of the A_k's will occur, i.e. that with probability 1, $|X_{n_k} - X| < 2^{-k}$ for all $k \geq k_0(\omega)$. This insures that

$$\sum_{k=1}^{\infty} |X_{n_k} - X| < \infty \qquad w.p.1$$

since the tail of the series is dominated by the convergent series $\sum_k 2^{-k}$. Thus

$$\lim_{k\to\infty} \sup_{m\geq k} |X_{n_m} - X| \leq \lim_{k\to\infty} \sum_{m\geq k} |X_{n_m} - X| = 0$$

since the series converges. ♠

https://doi.org/10.26219/heal.aueb.5740

1.4 Weak Convergence

In this section we sketch briefly (and mostly without proof) some of the most important results regarding weak convergence of distribution functions. The set up is the following: Suppose that a family of real random variables $\{X_n\}$ is given with corresponding distribution functions F_n. (It is important to note that we are not concerned at all here with the *joint statistics* of the family $\{X_n\}$, only with their marginal distributions $F_n(x) = P(X_n \leq x)$, so the random variables do not even have to be defined on the same probability space.)

Definition 4 *$\{F_n\}$ converges weakly to a distribution function F if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for each point of continuity of $F(x)$.

Weak convergence is often referred to as *convergence in distribution* and denoted by $F_n \xrightarrow{\text{d}} F$. The terminology extends to the sequence of random variables which is said to converge in distribution. One also writes $X_n \xrightarrow{\text{d}} X$.

Theorem 5 (Helly) *Let $\{F_n\}$ be an arbitrary collection of distribution functions. Then there exists a subsequence $\{F_{n_k}\}$ such that*

$$F_{n_k} \xrightarrow{\text{d}} F$$

for some distribution F.

Theorem 6 *$\{F_n\}$ converges weakly to F if and only if*

$$\lim_n \int_{\mathbb{R}} f(x)dF_n(x) = \int_{\mathbb{R}} f(x)dF(x)$$

for every bounded, continuous f.

(This is sometimes referred to as Helly's second theorem.)

As we shall see when we discuss the Central Limit Theorem later on, one of the problems that arises very often, both in practice and in theory is the following. If we have a family of distributions $\{F_n\}$ with corresponding characteristic functions f_n then,

a) If F_n converges weakly to some distribution function F can we conclude that f_n will converge to the characteristic function f of F?

b) If $f_n(t)$ converges for all t to $f(t)$, then is $f(t)$ also a characteristic function, and if it is and it corresponds to (say) the distribution F, can we conclude from this that $F_n \xrightarrow{\text{d}} F$?

The first question has an affirmative answer as one can show without much effort (essentially this follows from Helly's second theorem). The answer to the second question however is more complicated as we can see from the following example.

Let

$$F_n(x) = \begin{cases} 0 & x < -n \\ \frac{x+n}{2n} & -n \leq x < n \\ 1 & n \leq x \end{cases}$$

https://doi.org/10.26219/heal.aueb.5740

i.e. we have a family of uniform distributions on $[-n, n]$. Their ch.f.'s are

$$f_n(t) = \frac{\sin(nt)}{nt}.$$

We thus see that

$$f_n(t) \longrightarrow f(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ 1 & t = 0 \end{cases}$$

It is easy to see that the above limit is not a characteristic function (it is not continuous!). Also, in this case, $F_n(x) \to 0$ for all x so $\{F_n\}$ does not converge to a distribution function. Thus clearly it is not enough for f_n to converge.

Theorem 7 *[Convergence Theorem] Let $\{F_n\}$ be probability distributions with characteristic functions $\{f_n\}$. If*

a) $f_n(t)$ converges for every t and defines a limit function $f(t)$

b) This limit function $f(t)$ is continuous at $t = 0$

then

$\{F_n\}$ converges weakly to some distribution F with characteristic function F.

https://doi.org/10.26219/heal.aueb.5740

Chapter 2

Characteristic Functions

Let X a real random variable with distribution function F. We denote by μ the corresponding *measure* induced on the real line by F via the relationship $\mu(a,b] = F(b) - F(a)$. The characteristic function corresponding to X (or equivalently to F or μ) is

$$\phi(t) = \int_{\mathbf{R}} e^{itx} dF(x) = \int_{\mathbf{R}} e^{itx} \mu(dx) = Ee^{itX} \tag{2.1}$$

where $i = \sqrt{-1}$ is the imaginary unit and $t \in \mathbf{R}$. Thus f is a function from $\mathbf{R}$ to $\mathbf{C}$. Recalling de Moivre's formula for the complex exponential, $e^{ix} = \cos x + i \sin x$ we can also write

$$\phi(t) = \int_{\mathbf{R}} \cos(xt) dF(x) + i \int_{\mathbf{R}} \sin(xt) dF(x)$$

Suppose that the distribution function F is *symmetric*, i.e. $P(X > x) = P(X < -x)$ for every x, or equivalently $1 - F(x) = F(-x-)$. Then, taking into account the fact that $\sin x$ is an *odd* function we can see that the imaginary part of the characteristic function vanishes and we are left with

$$\phi(t) = \int_{\mathbf{R}} \cos(xt) dF(x)$$

From the above definition it is obvious that the probability distribution specifies the characteristic function. Later in this discussion we will also prove the *uniqueness theorem* which states that the characteristic function uniquely specifies the probability measure. Hence, knowledge of the characteristic function of a random variable is enough to determine its distribution. We will begin with some useful elementary results.

If $f(t)$ is a characteristic function then $f(0) = 1$. This follows by direct substitution into (2.1).

A characteristic function is *uniformly continuous*, i.e. $\forall \epsilon > 0\ \exists \delta > 0$ such that $|f(t+h) - f(t)| < \epsilon$ whenever $|h| < \delta$ for all $t \in \mathbf{R}$. Indeed,

$$\begin{aligned} |f(t+h) - f(t)| &= \left| \int_{\mathbf{R}} e^{i(t+h)x} dF(x) - \int_{\mathbf{R}} e^{itx} dF(x) \right| \\ &\leq \int_{\mathbf{R}} \left| e^{itx}(e^{ixh} - 1) \right| dF(x) = \int_{\mathbf{R}} \left| (e^{ixh} - 1) \right| dF(x) \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5740

However, $|e^{ixh} - 1| \leq 2$ and $\int_{\mathbf{R}} 2dF(x) < \infty$, hence we can appeal to the Dominated Convergence theorem to argue that $\lim_{h\to 0} \int_{\mathbf{R}} \left|(e^{ixh} - 1)\right| dF(x) = 0$. Thus the result is established.

If the characteristic function (ch. f.) of the random variable X is $f(t)$, then the ch. f. of $aX + b$ is $e^{itb} f(at)$. This follows immediately from $E[e^{it(aX+b)}] = e^{itb} E e^{i(at)X}$.

Let $f_i(t) = E[e^{itX_i}]$, $i = 1, 2$, where X_1, X_2 are *independent* random variables. Then the characteristic function of their sum is the product of the characteristic functions: $Ee^{it(X_1+X_2)} = Ee^{itX_1} Ee^{itX_2} = f_1(t) f_2(t)$. This of course generalizes to sums of independent random variables with arbitrarily many terms.

Let X, X' independent random variables *with the same distribution* and characteristic function $f(t)$. Show that $Ee^{it(X+X')} = f(t)^2$ and $Ee^{it(X-X')} = f(t)\overline{f(t)} = |f(t)|^2$. This shows that whenever $f(t)$ is a characteristic function, $|f(t)|^2$, which is always real–valued, is also a characteristic function.

Let Z be a standard normal random variable. Then its characteristic function is $f(t) = e^{-t^2/2}$. Indeed, $f(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{itx} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(tx) dx$ since the density of the standard normal is an even function. Thus differentiating with respect to t inside the integral and integrating by parts gives

$$\begin{aligned} f'(t) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} x \sin(tx) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \sin(tx) d(e^{-x^2/2}) \\ &= -\int_{-\infty}^{\infty} t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(tx) dx = -tf(t) \end{aligned}$$

Thus, $f(t)$ satisfies the differential equation

$$f'(t) = -tf(t), \qquad f(0) = 1$$

which has the solution $f(t) = e^{-t^2/2}$. From the above it follows that the charactristic function of a normal r.v. with mean μ and variance σ^2 is $e^{it\mu - t^2\sigma^2/2}$.

The exponential distribution with density e^{-x}, $x \geq 0$ has characteristic function

$$\int_0^{\infty} e^{-x} e^{itx} dx = \int_0^{\infty} e^{-x(1-it)} dx = \frac{1}{1-it}.$$

In the same way we can compute the characteristic function of the Laplace distribution with density $\frac{1}{2} e^{-|x|}$, $x \in \mathbf{R}$ as

$$\frac{1}{2}\left(\frac{1}{1+it} + \frac{1}{1-it}\right) = \frac{1}{1+t^2}$$

If $f_i(t)$ are characteristic functions (corresponding to distribution functions $F_i(x)$), $i = 1, 2, 3, \ldots$, and $p_i \geq 0$, $\sum_i p_i = 1$, then $\sum_i p_i f_i(t)$ is the characteristic function that corresponds to the distribution $\sum_i p_i F_i(x)$. This idea ofcourse extends from sums to integrals: If $F(x, a)$ is a distribution depending with a parameter a with characteristic function $f(t, a)$ and G is another distribution function, then $\int f(t, a) dG(a)$ is the characteristic function of the "mixed" distribution $\int F(x, a) dG(a)$.

The following table gives examples of distributions and the characteristic functions that correspond to them.[1]

[1]We denote by x^+ the positive part of a real number x, i.e. $x^+ = \max(0, x)$ and by x^- the negative part, $x^- = -\min(0, x)$. Thus $x = x^+ - x^-$

https://doi.org/10.26219/heal.aueb.5740

	Distribution/Density Function	Characteristic Function
1.	Deterministic: $F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$	e^{ita}
2.	Bernoulli: $F(x) = \begin{cases} 0 & x < 0 \\ q & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$	$q + pe^{it}$
3.	Uniform with density $F'(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$	$e^{it/2}\frac{\sin(t/2)}{t/2}$
4.	Standard Normal with density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$e^{-t^2/2}$
5.	Gamma with density $\frac{1}{\Gamma(\alpha)}x^{\alpha-1}e^{-x}$	$\left(\frac{1}{1-it}\right)^{\alpha}$
6.	Triangular density $F'(x) = \begin{cases} (1-\|x\|)^+ & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$	$\left(\frac{\sin(t/2)}{t/2}\right)^2$
7.	Cauchy density $\frac{1}{\pi}\frac{1}{1+x^2}$	$e^{-\|t\|}$
8.	Geometric $P(X = k) = q^{k-1}p$, $k = 1, 2, 3, \ldots$	$\frac{pe^{it}}{1-qe^{it}}$
9.	Binomial $P(X = k) = \binom{n}{k}p^k q^{n-k}$	$(q + pe^{it})^n$
10.	Poisson $P(X = k) = \frac{\alpha^k}{k!}e^{-\alpha}$ $k = 0, 1, 2 \ldots$	$e^{\alpha(e^{it}-1)}$

2.1 The Uniqueness Theorem

Before we turn to the central result regarding characteristic functions we need the following fact from analysis

$$\int_0^\infty \frac{\sin ax}{x}dx = \frac{\pi}{2}\text{sgn}(a) := \begin{cases} \frac{\pi}{2} & a > 0 \\ 0 & a = 0 \\ -\frac{\pi}{2} & a < 0 \end{cases} \tag{2.2}$$

It is enough to establish the result for $a > 0$ as the others follow from a simple change of variables. To this end, write

$$\begin{aligned} \int_0^\infty \frac{\sin ax}{x}dx &= \int_0^\infty \frac{\sin ax}{ax}d(ax) = \int_0^\infty \frac{\sin x}{x}dx \\ &= \int_0^\infty \sin x \left[\int_0^\infty e^{-ux}du\right] dx = \int_0^\infty \left[\int_0^\infty e^{-xu} \sin x dx\right] du \\ &= \int_0^\infty \frac{1}{1+u^2}du = \frac{\pi}{2} \end{aligned}$$

Theorem 8 *[Uniqueness Theorem] The characteristic function uniquely specifies the probability measure via the relationship*

$$\mu(a,b) + \frac{1}{2}\mu\{a\} + \frac{1}{2}\mu\{b\} = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-itb} - e^{-ita}}{it} f(t)dt. \tag{2.3}$$

Proof: In view of the definition of the characteristic function the integral on the right hand side of (2.3) is

$$\int_{-T}^{T} \frac{e^{-itb}-e^{-ita}}{it} \int_{\mathbf{R}} e^{itx} dF(x) dt = \int_{\mathbf{R}} \int_{-T}^{T} \frac{e^{-itb}-e^{-ita}}{it} e^{itx} dt dF(x).$$

Since for any real y

$$\frac{1}{2\pi} \int_{-T}^{T} \frac{e^{ity}}{it} dt = \frac{1}{\pi} \int_0^T \frac{e^{ity}-e^{-ity}}{2it} dt = \frac{1}{\pi} \int_0^T \frac{\sin yt}{t} dt =: S_T(y),$$

taking into account the above equation, the right hand side of (2.3) becomes

$$\lim_{T\to\infty} \int_{\mathbf{R}} \left(S_T(x-b) - S_T(x-a)\right) dF(x). \tag{2.4}$$

As we have seen from (2.2), $\lim_{T\to\infty} S_T(y) = \frac{1}{2}\mathrm{sgn}(y)$. Also, $|S_T(y)| \leq c$, hence,

$$\left| \int_{\mathbf{R}} \left(S_T(x-b) - S_T(x-a)\right) dF(x) \right| \leq \int_{\mathbf{R}} |S_T(x-b) - S_T(x-a)| \, dF(x) \leq 2c \int_{\mathbf{R}} dF(x) = 2c$$

and we can appeal to the Dominated Convergence Theorem in order to interchange the order of the limit and the integral in (2.4). Thus

$$\lim_{T\to\infty} S_T(x-b) - S_T(x-a) = \mathbf{1}(a<x<b) + \frac{1}{2}\mathbf{1}(x=a) + \frac{1}{2}\mathbf{1}(x=b)$$

and hence

$$\begin{aligned} &\lim_{T\to\infty} \int_{\mathbf{R}} \left(S_T(x-b) - S_T(x-a)\right) dF(x) \\ &\qquad = \int_{\mathbf{R}} \left[\mathbf{1}(a<x<b) + \tfrac{1}{2}\mathbf{1}(x=a) + \tfrac{1}{2}\mathbf{1}(x=b)\right] \mu(dx) \\ &\qquad = \mu(a,b) + \tfrac{1}{2}\mu\{a\} + \tfrac{1}{2}\mu\{b\}. \end{aligned}$$

♠

Theorem 9 *Suppose that the characteristic function $f(t)$ is integrable, i.e. $\int_{\mathbf{R}} |f(t)| dt < \infty$. Then the corresponding distribution function is absolutely continuous with corresponding density $p(x)$ given by*

$$p(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-itx} f(t) dt. \tag{2.5}$$

Proof: We begin with the remark that if f is integrable, then the corresponding distribution function has no atoms, i.e. $F(x) = F(x-)$. Indeed, from the uniqueness theorem we have

$$\frac{F(x)-F(x-h)}{h} = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{1-e^{iht}}{ith} e^{-itx} f(t) dt.$$

[1]The interchange of the order of the two integrals here is justified by Fubini's theorem since

$$e^{-itb} - e^{-ita} = -\int_a^b ite^{-itu} du$$

and hence

$$\left| \frac{e^{-itb}-e^{-ita}}{it} e^{itx} \right| \leq |e^{itx}| \int_a^b |e^{-itu}| du = |b-a|.$$

Since $\left|\frac{1-e^{iht}}{ith}\right| \leq 1$, the integrand is bounded by $\left|\frac{1-e^{iht}}{ith}e^{-itx}f(x)\right| \leq |f(t)|$, and since by assumption $f(t)$ is integrable, we can appeal to the Dominated Convergence Theorem to obtain

$$\begin{aligned}\lim_{h\to 0}\frac{F(x)-F(x-h)}{h} &= \lim_{h\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{1-e^{iht}}{ith}e^{-itx}f(t)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}f(t)dt.\end{aligned}$$

The above argument establishes that the left derivative of $F(x)$ exists and is given by the above expression. An identical argument shows that the right derivative also exists and equals the same quantity. This completes the proof. ♠

The behavior of the characteristic function near the origin determines the "heaviness" of the tails of the distribution. This idea is formalized in the following inequalities

Theorem 10 *[Modulus Inequalities] If we denote by $\mu[-A,A]^c$ the probability $P(|X| > A)$ for any $A > 0$, then*

$$\mu[-A,A]^c \leq \frac{2}{A}\int_{-1/A}^{1/A}[1-f(t)]dt. \tag{2.6}$$

Proof:

$$\begin{aligned}\frac{1}{2T}\int_{-T}^{T}f(t)dt &= \frac{1}{2T}\int_{-T}^{T}\int_{\mathbf{R}}e^{itx}\mu(dx)dt = \frac{1}{2T}\int_{\mathbf{R}}\mu(dx)\left(\int_{-T}^{T}e^{itx}dt\right) \\ &= \int_{\mathbf{R}}\mu(dx)\int_{0}^{T}\frac{\cos tx}{T}dt = \int_{\mathbf{R}}\frac{\sin Tx}{Tx}\mu(dx).\end{aligned} \tag{2.7}$$

Note however that

$$\left|\frac{\sin Tx}{Tx}\right| \leq \begin{cases} 1 & |x| \leq 2A \\ \frac{1}{2TA} & |x| > 2A \end{cases}$$

and hence

$$\int_{\mathbf{R}}\frac{\sin Tx}{Tx}\mu(dx) \leq \mu[-2A,2A] + \frac{1}{2TA}(1-\mu[-2A,2A]) = \left(1-\frac{1}{2TA}\right)\mu[-2A,2A] + \frac{1}{2TA}.$$

If we set $T = A^{-1}$ in the above we obtain

$$\left|\frac{A}{2}\int_{-A^{-1}}^{A^{-1}}f(t)dt\right| \leq \frac{1}{2}\mu[-2A,2A] + \frac{1}{2}.$$

From this last inequality, (2.6) follows readily. ♠

2.2 Weak Convergence

In this section we sketch briefly (and mostly without proof) some of the most important results regarding weak convergence of distribution functions. The set up is the following: Suppose that a family of random variables $\{X_n\}$ is given with corresponding distribution functions F_n. (It is important to note that we are not concerned at all here with the *joint statistics* of the family X_n, only with their marginal distributions $F_n(x) = P(X_n \leq x)$, so the random variables do not even have to be defined on the same probability space.)

https://doi.org/10.26219/heal.aueb.5740

Definition 5 *$\{F_n\}$ converges weakly to a distribution function F if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for each point of continuity of $F(x)$.

Weak convergence is often refered to as *convergence in distribution* and we write $F_n \xrightarrow{d} F$.

Theorem 11 (Helly) *Let $\{F_n\}$ be an arbitrary collection of distribution functions. Then there exists a subsequence $\{F_{n_k}\}$ such that*

$$F_{n_k} \xrightarrow{d} F$$

for some distribution F.

Theorem 12 *$\{F_n\}$ converges weakly to F if and only if*

$$\lim_n \int_{\mathbf{R}} f(x)dF_n(x) = \int_{\mathbf{R}} f(x)dF(x)$$

for every bounded, continuous f.

(This is sometimes refered to as Helly's second theorem.)

As we shall see when we discuss the Central Limit Theorem later on, one of the problems that arises very often, both in practice and in theory is the following. If we have a family of distributions $\{F_n\}$ with corresponding characteristic functions f_n then,

a) If F_n converges weakly to some distribution function F can we conclude that f_n will converge to the characteristic function f of F?

b) If $f_n(t)$ converges for all t to $f(t)$, then is $f(t)$ also a characteristic function, and if it is and it corresponds to (say) the distribution F, can we conclude from this that $F_n \xrightarrow{d} F$?

The first question has an affirmative answer as one can show without much effort (essentially this follows from Helly's second theorem). The answer to the second question however is more complicated as we can see from the following example.

Let

$$F_n(x) = \begin{cases} 0 & x < -n \\ \frac{x+n}{2n} & -n \leq x < n \\ 1 & n \leq x \end{cases}$$

i.e. we have a family of uniform distributions on $[-n, n]$. Their ch.f.'s are

$$f_n(t) = \frac{\sin(nt)}{nt}$$

We thus see that

$$f_n(t) \longrightarrow f(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ 1 & t = 0 \end{cases}$$

It is easy to see that the above limit is not a characteristic function (it is not continuous!). Also, in this case, $F_n(x) \to 0$ for all x so $\{F_n\}$ does not converge to a distribution function. Thus clearly it is not enough for f_n to converge.

Theorem 13 *[Convergence Theorem] Let $\{F_n\}$ be probability distributions with characteristic functions $\{f_n\}$. If*

a) $f_n(t)$ converges for every t and defines a limit function $f(t)$

b) This limit function $f(t)$ is continuous at $t = 0$

then

$\{F_n\}$ converges weakly to some distribution F with characteristic function f.

2.3 Positive definite functions

Definition 6 *A function $f : \mathbf{R} \to \mathbf{C}$ is* positive definite *if for every $n \in \mathbf{N}$, $t_1, t_2, \ldots, t_n \in \mathbf{R}$ and $c_1, c_2, \ldots, c_n \in \mathbf{C}$,*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i \bar{c}_j f(t_i - t_j) \geq 0, \tag{2.8}$$

where $\bar{c}$ denotes the complex conjugate of c.

(The meaning of the above inequality is that the left hand side should be real and non-negative.) Note that the positive definiteness of f is equivalent to the positive definiteness of the matrix

$$\begin{bmatrix} f(0) & f(t_1 - t_2) & f(t_1 - t_3) & \cdots & f(t_1 - t_n) \\ f(t_2 - t_1) & f(0) & f(t_2 - t_3) & \cdots & f(t_2 - t_n) \\ f(t_3 - t_1) & f(t_3 - t_2) & f(0) & \cdots & f(t_3 - t_n) \\ & & & & \\ f(t_n - t_1) & f(t_n - t_2) & f(t_n - t_3) & \cdots & f(0) \end{bmatrix} \tag{2.9}$$

for every n and $t_i \in \mathbf{R}$. Also note that we use the term positive in the weak sense following common usage (in other words, here positive means "nonnegative").

Theorem 14 *All characteristic functions are positive definite.*

Proof: We start with the remark that e^{itx} is positive definite. Indeed,

$$\sum_{i,j} c_i \bar{c}_j e^{i(t_i - t_j)} = \sum_{i,j} c_i e^{it_i x}\overline{(c_j e^{it_j x})} = \left(\sum_i c_i e^{it_i x}\right)\overline{\left(\sum_j c_j e^{it_j x}\right)} = \left|\sum_i c_i e^{it_i x}\right|^2 \geq 0.$$

To show that a characteristic function is positive definite, it is enough to mimic the above argument, interchanging summations and expectation:

$$\sum_{i,j} f(t_i - t_j) c_i \bar{c}_j = \sum_{i,j} c_i \bar{c}_j E[e^{i(t_i - t_j)X}] = E \sum_{i,j} c_i \bar{c}_j e^{i(t_i - t_j)X} = E\left|\sum_i c_i e^{it_i x}\right|^2 \geq 0$$

More interesting and far–reaching however is the fact that the converse is also true, namely that all positive definite functions $f : \mathbf{R} \longrightarrow \mathbf{C}$ are characteristic functions of some measure on the real line.

 https://doi.org/10.26219/heal.aueb.5740

This result will be established latter on. We first establish some of the properties of positive definite functions.

1. If f is positive definite, then $f(0) \geq 0$ (as before part of the assertion is that $f(0)$ is real). Indeed, (2.8) with $n = 1$ gives $c\bar{c}f(0) = |c|^2 f(0) \geq 0$.

2. $f(t) = \overline{f(-t)}$. In particular this means that a real positive definite function must be even, i.e. it must satisfy $f(t) = f(-t)$. To prove this assertion apply (2.8) with $n = 2$, $t_1 = 0$, $t_2 = t$, $c_1 = c_2 = 1$ to obtain

$$2f(0) + f(t) + f(-t) \geq 0$$

which implies[2] that $f(t) + f(-t)$ is real, hence $\Im f(t) + \Im f(-t) = 0$, or

$$\Im f(t) = -\Im f(-t) \tag{2.10}$$

If we choose $c_1 = 1$, $c_2 = i$ we obtain

$$2f(0) + if(t) - if(-t) \geq 0$$

which implies that $f(t) - f(-t)$ is pure imaginary, hence $\Re f(t) - \Re f(-t) = 0$, or

$$\Re f(t) = \Re f(-t). \tag{2.11}$$

Equations (2.10) and (2.11) together establish that

$$f(t) = \overline{f(-t)}. \tag{2.12}$$

3. $|f(t)| \leq f(0)$ for every $t \in \mathbf{R}$. To show this, take $c_1 = f(t)$, $c_2 = -|f(t)|$ to obtain $2f(0)|f(t)|^2 - 2|f(t)|^3 \geq 0$, whence the inequality follows.

4. Any positive definite function for which $f(0) = 1$, satisfies the following inequality:

$$|f(t+h) - f(t)| \leq 2|1 - f(h)|^2. \tag{2.13}$$

(The normalizing assumption $f(0) = 1$ simplifies the algebra without harming the generality of the statement.) The importance of this inequality lies in the fact that it implies that if a positive definite function is continuous at 0 then it must be continuous (and in fact uniformly continuous) on $\mathbf{R}$. We have already seen this for characteristic functions. To prove this assertion, we will use the positive definiteness of the matrix

$$\begin{bmatrix} 1 & f(-t) & f(-t-h) \\ f(t) & 1 & f(-h) \\ f(t+h) & f(h) & 1 \end{bmatrix} \tag{2.14}$$

which is obtained from (2.9) with $n = 3$, $t_1 = 0$, $t_2 = t$, $t_3 = t + h$ and $f(0) = 1$. (2.14) is positive definite if

$$1 - |f(t)|^2 \geq 0$$

$$1 + f(-t)f(-h)f(t+h) + f(t)f(h)f(-t-h) - |f(h)|^2 - |f(t)|^2 - |f(t+h)|^2 \geq 0.$$

Making use of (2.12), this last inequality can be rewritten as

$$1 + f(t)f(h)\overline{f(t+h)} + \overline{f(t)f(h)}f(t+h) - |f(t)|^2 - |f(h)|^2 - |f(t+h)|^2 \geq 0$$

[2] If $c = a + ib$ is a complex number ($a, b \in \mathbf{R}$) we denote its real part by $\Re c = a$ and its imaginary part by $\Im c = b$)

https://doi.org/10.26219/heal.aueb.5740

or

$$1+2\Re\{f(t)f(h)\overline{f(t+h)}\}-|f(t)|^2-|f(h)|^2-|f(t+h)|^2\geq 0,$$

which gives

$$|f(t+h)|^2+|f(t)|^2\leq 1-|f(h)|^2+2\Re\{f(t)f(h)\overline{f(t+h)}\}.$$

We are now ready to show (2.13)

$$\begin{aligned}|f(t)-f(t+h)|^2 &= |f(t)|^2+|f(t+h)|^2-f(t)\overline{f(t+h)}-\overline{f(t)}f(t+h)\\ &= |f(t)|^2+|f(t+h)|^2-2\Re\{f(t)\overline{f(t+h)}\}\\ &\leq 1-|f(h)|^2+2\Re\{f(t)f(h)\overline{f(t+h)}\}-2\Re\{f(t)\overline{f(t+h)}\}\\ &= 1-|f(h)|^2+2\Re\{f(t)\overline{f(t+h)}[f(h)-1]\}\\ &\leq 1-|f(h)|^2+2|1-f(h)| \qquad (2.15)\end{aligned}$$

where in this last inequality we have used the fact that

$$\begin{aligned}\Re\{f(t)\overline{f(t+h)}[f(h)-1]\} &\leq |f(t)\overline{f(t+h)}[f(t)-1]|\leq|f(t)|\,|f(t+h)|\,|1-f(t)|\\ &\leq |1-f(t)|\end{aligned}$$

(since $|f(t)|\leq f(0)=1$). Finally, note that $1-|f(h)|=|1-|f(h)||\leq|1-f(h)|$ and hence (2.15) gives

$$\begin{aligned}|f(t)-f(t+h)|^2 &\leq (1-|f(h)|)(1+|f(h)|)+2|1-f(h)|\leq|1-f(h)|\,(1+|f(h)|+2)\\ &\leq 4|1-f(h)|\end{aligned}$$

As we have seen, characteristic functions of probability measures are positive definite and positive definite functions that are continuous at zero have the same properties as characteristic functions. This is far from accidental. In fact as the next theorem shows these two classes of functions coincide.

Theorem 15 (Bochner) *Suppose that a function $f:\mathbf{R}\to\mathbf{C}$ is positive definite with $f(0)=1$ and continuous at 0. Then there exists a probability distribution F on $\mathbf{R}$ such that $f(t)=\int_{\mathbf{R}}e^{itx}dF(x)$.*

Proof: Fix $T>0$ and consider the function

$$p_T(x)=\frac{1}{T^2}\int_0^T\int_0^T f(t-s)e^{itx}e^{-isx}dsdt\geq 0. \qquad (2.16)$$

It is clear that $p_T(x)$ is real and nonnegative since the double integral is the limit of Riemann sums $\sum_j\sum_k f(t_j-s_k)e^{it_jx}e^{-is_kx}\Delta t_j\Delta s_k$ which are nonnegative by positive definiteness. Changing variables in (2.16) gives

$$p_T(x)=\int_{-T}^{T}\left(1-\frac{|t|}{T}\right)f(t)e^{-itx}dt. \qquad (2.17)$$

Remark: For any $L>0$,

$$\begin{aligned}\frac{1}{L}\int_0^L dy\int_{-y}^{y}e^{itx}dx &= \frac{1}{L}\int_0^L dy\int_0^y 2\cos tx\,dx=\frac{2}{Lt}\int_0^L\sin ty\,dy\\ &= 2\frac{1-\cos Lt}{Lt^2}\end{aligned}$$

Hence,

$$\begin{aligned}\frac{1}{L}\int_0^L dy\int_{-y}^{y} p_T(x)dx &= 2\frac{1}{2\pi}\int_{-T}^{T}\left(1-\frac{|t|}{T}\right)\frac{1-\cos Lt}{Lt^2}f(t)dt \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty}\left(1-\frac{|t|}{T}\right)^{+}\frac{1-\cos Lt}{Lt^2}f(t)dt\end{aligned}$$

Define

$$f_T(t) := \left(1-\frac{|t|}{T}\right)^{+} f(t) \tag{2.18}$$

and note that $|f_T(t)| \leq 1$. Also, note that $\left|\int_{-\infty}^{\infty}\frac{1-\cos t}{t^2}dt\right| < \infty$. Hence, by dominated convergence,

$$\lim_{L\to\infty}\frac{1}{\pi}\int_{-\infty}^{\infty} f_T(t/L)\frac{1-\cos t}{t^2}dt = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos t}{t^2}dt = 1.$$

We thus have that $\int_{-y}^{y} p_T(x)dx \uparrow 1$ as $y \uparrow \infty$. (Here we are using the following result: If g is an increasing function and $\frac{1}{X}\int_0^X g(x)dx \to a$ as $X \to \infty$, then $g(x) \to a$ as $x \to \infty$. $\int_{-y}^{y} p_T(x)dx$ must be an increasing function of y, since $p_T(x) \geq 0$ for all x. We have thus shown that $p_T(x)$ is integrable with $\int_{-\infty}^{\infty} p_T(x)dx = 1$, hence $p_T(x)$ is the probability density of some distribution. We also have

$$p_T(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1-\frac{|t|}{T}\right)^{+} f(t)e^{-itx}dx$$

Hence, for each T, $f_T(t)$ is the characteristic function of some distribution function. As $T \to \infty$,

$$f_T(t) = \left(1-\frac{|t|}{T}\right)^{+} f(t) \to f(t) \qquad \text{for all } t \in \mathbf{R}$$

and $f(t)$ is by assumption continuous at 0. Therefore the convergence theorem of the previous section guarrantees that $f(t)$ must also be a characteristic function ♠

2.4 Second Order Stationary Processes

Let $\{X_t; t \in \mathbf{R}\}$ a stochastic process with $EX_t = \mu(t)$. Define $\xi_t := X_t - \mu(t)$. The function $R(s,t) := E\xi_s\xi_t$ is called the *covariance function* of the process X. A process X is called *Gaussian* if, for any $n \in \mathbf{N}$ and any $t_1, t_2, \ldots, t_n$, $(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \sim \mathcal{N}(\mu, \Sigma)$ where

$$\mu = \begin{bmatrix}\mu_1\\ \mu_2\\ \vdots\\ \mu_n\end{bmatrix}, \qquad \Sigma = \begin{bmatrix}\mathrm{Var}(X_{t_1}) & \mathrm{Cov}(X_{t_1}, X_{t_2}) & \cdots & \mathrm{Cov}(X_{t_1}, X_{t_n})\\ \mathrm{Cov}(X_{t_2}, X_{t_1}) & \mathrm{Var}(X_{t_2}) & \cdots & \mathrm{Cov}(X_{t_1}, X_{t_n})\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{Cov}(X_{t_n}, X_{t_1}) & \mathrm{Cov}(X_{t_n}, X_{t_2}) & \cdots & \mathrm{Var}(X_{t_n})\end{bmatrix}$$

A process $\{X_t; t \in \mathbf{R}\}$ is a second order stationary process if

https://doi.org/10.26219/heal.aueb.5740

i) $EX_t = \mu$, $E(X_t - \mu)^2 = \sigma^2 < \infty$, for all $t \in \mathbf{R}$,

ii) There exists a real function $r : \mathbf{R} \longrightarrow \mathbf{R}$ such that $E(X_s - \mu)(X_t - \mu) = r(t-s)$ for all $s, t \in \mathbf{R}$,

iii) The process is mean-square continuous, i.e.

$$\lim_{h\to 0} E(X_{t+h} - X_t)^2 = 0$$

Let us now examine some of the consequences of the above properties. First, since $\mathrm{Var}(X_t) = \mathrm{Cov}(X_t, X_t)$ if follows that $r(0) = \sigma^2$. Also, $E(X_{t+h} - X_t)^2 = \sigma^2 + \sigma^2 - 2r(h) = 2(r(0) - r(h))$ and hence, property *iii)* is equivalent to the requirement that r be continuous at zero, i.e. $r(h) \to r(0)$ as $h \to 0$. Cauchy–Schwartz implies $|r(t)| \le \sigma^2\ \forall t$. Without loss of generality suppose $\sigma = 1$.

Theorem 16 *$r(t)$ is a positive definite function.*

Proof:

$$\begin{aligned} 0 &\le \left|\textstyle\sum_{j=1}^n X_{t_j} z_j\right|^2 = \overline{\left(\textstyle\sum_{j=1}^n X_{t_j} z_j\right)} \left(\textstyle\sum_{i=1}^n X_{t_i} z_i\right) \\ &= \textstyle\sum_{i=1}^n \sum_{j=1}^n z_i \bar{z}_j X_{t_i} X_{t_j}. \end{aligned}$$

Taking expectations we have

$$0 \le \sum_{i,j} z_i \bar{z}_j E X_{t_i} X_{t_j} = \sum_{i,j} r(t_j - t_i) z_i \bar{z}_j$$

This establishes that the covariance function of any stationary second order process is positive definite. Hence Bochner's theorem guarantees that there exists a uniquely determined probability measure on $\mathbf{R}$ such that

$$r(t) = \int_{-\infty}^{\infty} e^{itx} R(dx)$$

In particular, since r is an even function,

$$r(t) = 2\int_0^{\infty} \cos(tx) R(dx) \tag{2.19}$$

2.4.1 An example of a stationary second order process

Let $N(t)$ be a Poisson process with rate λ and $X(0)$ a random variable with $P(X(0) = 1) = P(X(0) = -1) = 1/2$, independent of the Poisson process. Consider the process

$$X(t) = X(0)(-1)^{N(t)} \tag{2.20}$$

Clearly, X alternates between the values 1, and -1, changing value at each Poisson point. It is easy to see that $EX(t) = EX(0)E(-1)^{N(t)} = 0$ (since $EX(0) = 0$). The covariance function is easily computed as follows:

$$EX(t)X(t+s) = E\left[X(0)^2(-1)^{N(t)+N(t+s)}\right] = E\left[(-1)^{N(t+s)-N(t)}\right]$$

Using the stationary increments property the above expectation is

$$\begin{aligned} E(-1)^{N(s)} &= \sum_{n=0}^{\infty} \frac{(\lambda s)^{2n}}{(2n)!} e^{-\lambda s} - \sum_{n=0}^{\infty} \frac{(\lambda s)^{2n+1}}{(2n+1)!} e^{-\lambda s} \\ &= \cosh(\lambda s) e^{-\lambda s} - \sinh(\lambda s) e^{-\lambda s} = e^{-2\lambda s}. \end{aligned}$$

https://doi.org/10.26219/heal.aueb.5740

Hence the covariance function is given by

$$EX(t)X(s) = r(t-s) = e^{-2\lambda|t-s|}.$$

The spectral measure can be easily computed in this case: We must have $r(t) = e^{-2\lambda|t|} = \int_{-\infty}^{\infty} e^{itx} R(dx)$ and hence

$$R(dx) = \frac{1}{2\pi\lambda}\left[\frac{1}{1+(x/2\lambda)^2}\right] dx.$$

 https://doi.org/10.26219/heal.aueb.5740

Chapter 3

The Central Limit Theorem

3.1 The Central Limit Theorem for IID Random Variables

The Central Limit Theorem for i.i.d. r.v.'s states that

Theorem 17 *If X_i, $i = 1, 2, \ldots$ are i.i.d. random variables with common mean μ and variance $\sigma^2 < \infty$, the normalized sum*

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to the standard normal.

In order to establish this, in view of the convergence theorem for characteristic functions, it is enough to show that

$$E[e^{it\left(\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}\right)}] \to e^{-t^2/2}.$$

If $f(t) = E[e^{itX_1}]$, and we assume $\mu = 0$, $\sigma = 1$ (as we may with no loss of generality) then the above relationship becomes

$$(f(t/\sqrt{n}))^n \to e^{-t^2/2} \tag{3.1}$$

To show (3.1) we begin with a set of inequalities that will also prove to be useful in the sequel. For $x \in \mathbb{R}$

$$e^{ix} - 1 = i\int_0^x e^{iy} dy$$

and taking moduli on both sides $|e^{ix} - 1| \leq |i| \int_0^x |e^i y| dy \leq |x|$ or

$$\left|e^{ix} - 1\right| \leq |x|. \tag{3.2}$$

Similarly,

$$e^{ix} - 1 - ix = \int_0^x i\left(e^{iy} - 1\right) dy$$

and again taking moduli and using (3.2) we obtain

$$\left|e^{ix} - 1 - ix\right| \leq \frac{x^2}{2}. \tag{3.3}$$

Repeating this process once more,

$$e^{ix} - 1 - ix - \frac{(ix)^2}{2} = \int_0^x i\left(e^{iy} - 1 - iy\right) dy$$

which gives

$$\left|e^{ix} - 1 - ix - \frac{(ix)^2}{2}\right| \leq \frac{|x|^3}{3!} \tag{3.4}$$

We will now combine (3.3) and (3.4) to obtain a more useful inequality:

$$\left|e^{ix} - 1 - ix - \frac{(ix)^2}{2}\right| \leq \left|e^{ix} - 1 - ix\right| + \left|\frac{(ix)^2}{2}\right| \leq \frac{x^2}{2} + \frac{x^2}{2} = x^2.$$

Hence,

$$\left|e^{ix} - 1 - ix - \frac{(ix)^2}{2}\right| \leq \min\left(x^2, \frac{|x|^3}{3!}\right). \tag{3.5}$$

We also need the following

Lemma 1 *Let c_n be a sequence of complex numbers such that $c_n \to c$ as $n \to \infty$. Then*

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c$$

We are now ready to show (3.1). Define $\alpha_n(t) := f(t/\sqrt{n}) - 1 + \frac{t^2}{2n}$. We have

$$\begin{aligned}
|\alpha_n(t)| &= \left|E\left[e^{itX/\sqrt{n}} - 1 + \frac{t^2}{2n}\right]\right| = \left|E\left[e^{itX/\sqrt{n}} - 1 - \frac{iXt}{\sqrt{n}} + \frac{(Xt)^2}{2n}\right]\right| \\
&\leq E\left|e^{itX/\sqrt{n}} - 1 - \frac{iXt}{\sqrt{n}} + \frac{(Xt)^2}{2n}\right| \\
&\leq E\left[\min\left(\frac{(Xt)^2}{n}, \frac{|Xt|^3}{3!n^{3/2}}\right)\right]
\end{aligned}$$

where to obtain the second equality we have used our assumption that $EX = 0$ and $EX^2 = 1$ and to obtain the last inequality we have used (3.5). Thus,

$$\left(f(t/\sqrt{n})\right)^n = \left(1 - \frac{1}{n}(t^2/2 - n\alpha_n(t))\right)^n.$$

Now fix t and note that

$$n|\alpha_n(t)| \leq E\left[\min\left((Xt)^2, \frac{|Xt|^3}{3!n^{1/2}}\right)\right].$$

Since $\min\left((Xt)^2, \frac{|Xt|^3}{3!n^{1/2}}\right) \leq (Xt)^2$ and $E(Xt)^2 = t^2 < \infty$, we can use dominated convergence to evaluate the limit as $n \to \infty$

$$\lim_{n\to\infty} n|\alpha_n(t)| \leq E\left[\lim_{n\to\infty} \min\left((Xt)^2, \frac{|Xt|^3}{3!n^{1/2}}\right)\right] = 0.$$

We can now appeal to the lemma with $c_n = t^2/2 + n\alpha_n(t) \to t^2/2$ and this completes the proof.

3.2 The Central Limit Theorem for Arrays of Random Variables

While the Central Limit Theorem for iid r.v.'s is in all likelihood the most often used (and misused) result of Probability, very often its assumptions are not satisfied in practice. The random variables added may not be identically distributed or they may not be independent. Here we will focus our attention to independent but not identically distributed random variables.

A natural guess would be that if $\{X_i\}$ is a sequence of independent r.v.'s with means μ_i and variances $\sigma_i^2 < \infty$ then the normalized sum

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

converges in distribution to a unit normal random variable. This will turn out to be true but not without some restrictions on *the relative sizes of the variances.*

We begin with some examples to illustrate what could go wrong. These examples are necessarily somewhat complicated but they are based on the simple trigonometric identity $\sin(2x) = 2\sin(x)\cos(x)$.

Example 1. Let X_k be $\pm\frac{1}{2^k}$, these two values taken with equal probability. Then the corresponding ch. f. is given by $f_k(t) = \cos(t/2^k)$ and the variance is $\sigma_k^2 = 4^{-k}$. Hence, setting $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \sum_{k=1}^n 4^{-k} = \frac{1}{3}(1 - 4^{-n})$, we compute the ch. f. of S_n/B_n which is given by

$$\phi_n(t) = Ee^{it(S_n/B_n)} = \prod_{k=1}^n \cos\left(\frac{t}{2^k B_n}\right).$$

Repeated application of the trigonometric identity mentioned above shows that

$$2^n \sin\left(\frac{t}{2^n B_n}\right)\phi_n(t) = \sin\left(\frac{t}{B_n}\right)$$

or

$$\phi_n(t) = \frac{\sin\left(\frac{t}{B_n}\right)}{2^n \sin\left(\frac{t}{2^n B_n}\right)}. \tag{3.6}$$

Since $B_n \to \frac{1}{\sqrt{3}}$, letting $n \to \infty$ in (3.6) gives

$$\phi_n(t) \to \frac{\sin(t\sqrt{3})}{t\sqrt{3}}.$$

We thus see that S_n/B_n converges to the uniform distribution on $[-\sqrt{3}, \sqrt{3}]$.

Example 2. Let now X_k be $\pm 2^k$ with equal probability. Then the corresponding ch. f. is given by $f_k(t) = \cos(t2^k)$ and the variance is $\sigma_k^2 = 4^k$. Again, setting $S_n = X_1 + \cdots + X_n$ and $B_n^2 = \sum_{k=1}^n 4^k = \frac{4}{3}(4^n - 1)$, we compute the ch. f. of S_n/B_n which is given by

$$\phi_n(t) = Ee^{it(S_n/B_n)} = \prod_{k=1}^n \cos\left(\frac{2^k t}{B_n}\right).$$

whence

$$2^n \sin\left(\frac{2t}{B_n}\right)\phi_n(t) = \sin\left(\frac{2^{n+1}t}{B_n}\right)$$

 https://doi.org/10.26219/heal.aueb.5740

or

$$\phi_n(t) = \frac{\sin\left(\frac{2^{n+1}t}{B_n}\right)}{2^n \sin\left(\frac{2t}{B_n}\right)}. \tag{3.7}$$

Now

$$\frac{2^{n+1}}{B_n} = \frac{2^{n+1}}{\sqrt{\frac{4}{3}(4^n - 1)}} = \frac{\sqrt{3}}{\sqrt{1 - 4^{-n}}}$$

so that $\lim_{n\to\infty} \frac{2^{n+1}}{B_n} = \sqrt{3}$. Rewrite (3.7) as

$$\phi_n(t) = \frac{\sin\left(\frac{2^{n+1}t}{B_n}\right)}{\frac{2^{n+1}t}{B_n}\left(\frac{2t}{B_n}\right)^{-1} \sin\left(\frac{2t}{B_n}\right)} \to \frac{\sin(t\sqrt{3})}{t\sqrt{3}}.$$

Hence in this case as well the limit is uniform instead of normal. These two examples are in a sense significantly different since in the first case the sum of the variances converges whereas in the second it grows rapidly.

In order to obtain the necessary and sufficient condition for normalized sums of independent but not identically distributed r.v. to converge to a normal limit it will be useful to extend the scope of our investigation to triangular arrays of r.v.'s. Consider the following scheme: Let k_n be an integer–valued sequence such that $k_n \to \infty$ as $n \to \infty$ and consider the array of random variables

$$\begin{array}{cccc} X_{11} & X_{12} & \dots & X_{1k_n} \\ X_{21} & X_{22} & \dots & X_{2k_2} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nk_n} \\ \vdots & \vdots & & \vdots \end{array}$$

We assume that all the random variables in the same row (i.e. $X_{nj}, j = 1, 2, \dots, k_n$) are independent, while we will make no assumption about the joint distribution of random variables belonging to different rows. Let $S_n = \sum_{j=1}^{k_n} X_{nj}$ be the sum of the n'th row. In its most general form, the Central Limit Problem consists in finding the conditions under which there exist real sequences a_n, b_n such that $\frac{S_n - a_n}{b_n}$ converges in distribution to a non-degenerate law. An important special case in the above scheme are triangular arrays (obtained when $k_n = n$)

$$\begin{array}{ccccc} X_{11} & & & & \\ X_{21} & X_{22} & & & \\ X_{31} & X_{32} & X_{33} & & \\ \vdots & \vdots & \vdots & & \\ X_{n1} & X_{n2} & X_{n3} & \dots & X_{nn} \\ \vdots & \vdots & \vdots & & \end{array}$$

In this section we will not examine the problem in its full generality. We will restrict our investigation to the case where all the random variables in the above array have finite variance: $\mathsf{Var}(X_{nj}) = \sigma_{nj}^2 < \infty$. Denoting the corresponding means by $\mu_{nj} = EX_{nj}$, $j = 1, 2, \dots, k_n$, and setting

$$B_n^2 := \sum_{j=1}^{k_n} \sigma_{nj}^2$$

https://doi.org/10.26219/heal.aueb.5740

we define new random variables

$$\xi_{nj} := \frac{X_{nj} - \mu_{nj}}{B_n}.$$

Thus the normalized array

$$\begin{array}{cccc} \xi_{11} & \xi_{12} & \cdots & \xi_{1k_n} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2k_2} \\ \vdots & \vdots & & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nk_n} \\ \vdots & \vdots & & \vdots \end{array} \tag{3.8}$$

has the property that all the random variables in it have zero mean and that $\sum_{j=1}^{k_n} \mathsf{Var}(\xi_{jn}) = 1$ for all n. Now we are ready to state the

Theorem 18 *[Lindeberg–Feller Central Limit Theorem] Given the normalized array (3.8), $S_n := \sum_{j=1}^{k_n} \xi_{nj}$ converges to the standard normal law if and only if*

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} E[\xi_{nj}^2; |\xi_{nj}| > \epsilon] = 0 \qquad \text{for any } \epsilon > 0. \tag{3.9}$$

Remarks Condition (3.9) is known as the Lindeberg condition. Note that if $\tilde{F}_{nj}(x) = P(\xi_{nj} \le x)$ is the distribution function of ξ_{nj}, then it can be stated equivalently as

$$\lim_{n\to\infty} \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x) = 0 \qquad \text{for any } \epsilon > 0. \tag{3.10}$$

Proof: Here we will only establish the direct half of the theorem, namely that (3.9) implies that $\phi_n(t) := Ee^{itS_n}$ converges to $e^{-it^2/2}$ which will establish convergence of S_n to the standard normal law. If we denote by $f_{nj}(t) := Ee^{it\xi_{nj}}$ the characteristic function of ξ_{nj} we note that $f_{nj}(t) - 1 = \int (e^{itx} - 1 - itx) d\tilde{F}_{nj}(x)$ (since $E\xi_{nj} = 0$). Thus, using the inequality $|e^{ix} - 1 - ix| \le \frac{x^2}{2}$, we obtain

$$|f_{nj}(t) - 1| \le \frac{t2}{2} \int x^2 d\tilde{F}_{nj}(x).$$

Let $\epsilon > 0$ be an arbitrary, positive number. Then,

$$\begin{aligned} \int x^2 d\tilde{F}_{nj}(x) &= \int_{|x|\le\epsilon} x^2 d\tilde{F}_{nj}(x) + \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x) \\ &\le \epsilon^2 + \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x) \end{aligned}$$

However, the last integral above is one of the terms in (3.10) and therefore it goes to 0 as $n \to \infty$. Thus, for n sufficiently large, it can also be made smaller than ϵ^2, for any $j \le k_n$.

We thus have (for n sufficiently large)

$$\max_{1\le j\le k_n} |f_{nj}(t) - 1| \le \epsilon^2 T^2$$

Thus, for $|t| \leq T$, $\max_{1\leq j\leq k_n} |f_{nj}(t) - 1| < 1/2$ for n sufficiently large, hence we can use the Taylor expansion for the complex logarithm

$$\log(1+z) = \sum_{n=1}^{\infty}(-1)^n \frac{z^n}{n}, \qquad |z| < 1,$$

to obtain

$$\begin{aligned}\log(\phi_n(t)) &= \sum_{j=1}^{k_n} \log f_{nj}(t) = \sum_{j=1}^{k_n} \log(1 + (f_{nj}(t) - 1)) \\ &= \sum_{j=1}^{k_n}(f_{nj}(t) - 1) + R_n \qquad (3.11)\end{aligned}$$

where

$$R_n = \sum_{j=1}^{k_n}\sum_{m=2}^{\infty} \frac{(-1)^m}{m}(f_{nj}(t) - 1)^m.$$

We can bound R_n as follows:

$$\begin{aligned}|R_n| &\leq \sum_{j=1}^{k_n}\sum_{m=2}^{\infty} \frac{1}{2}|f_{nj}(t) - 1|^m = \frac{1}{2}\sum_{j=1}^{k_n} \frac{|f_{nj}(t) - 1|^2}{1 - |f_{nj}(t) - 1|} \leq \sum_{j=1}^{k_n}|f_{nj}(t) - 1|^2 \\ &\leq \max_{1\leq j\leq k_n} |f_{nj}(t) - 1| \sum_{j=1}^{k_n}|f_{nj}(t) - 1|\end{aligned}$$

However,

$$\begin{aligned}\sum_{j=1}^{k_n}|f_{nj}(t) - 1| &= \sum_{j=1}^{k_n}\left|\int_{\mathbf{R}} (e^{itx} - 1 - itx)\, d\tilde{F}_{nj}(x)\right| \leq \sum_{j=1}^{k_n}\int_{\mathbf{R}} |e^{itx} - 1 - itx|\, d\tilde{F}_{nj}(x) \\ &\leq \frac{t^2}{2}\sum_{j=1}^{k_n}\int_{\mathbf{R}} x^2 d\tilde{F}_{nj}(x) = \frac{t^2}{2}.\end{aligned}$$

From the two inequalities above it follows that

$$|R_n| \leq \frac{t^2}{2} \max_{1\leq j\leq k_n} |f_{nj}(t) - 1|$$

and hence that

$$|R_n| \to 0 \text{ uniformly in } |t| \leq T. \qquad (3.12)$$

But

$$\sum_{j=1}^{k_n}(f_{nj} - 1) = -\frac{t^2}{2} + \rho_n \qquad (3.13)$$

$$
\begin{aligned}
\rho_n &= \frac{t^2}{2} + \sum_{j=1}^{k_n} \int_{\mathbf{R}} (e^{itx} - 1 - itx)\, d\tilde{F}_{nj}(x) \\
&= \sum_{j=1}^{k_n} \int_{|x|\le\epsilon} \left(e^{itx} - 1 - itx - \frac{(itx)^2}{2}\right) d\tilde{F}_{nj}(x) \\
&\qquad + \sum_{j=1}^{k_n} \int_{|x|>\epsilon} \left(e^{itx} - 1 - itx - \frac{(itx)^2}{2}\right) d\tilde{F}_{nj}(x)
\end{aligned}
$$

Using the inequalities (3.3) and (3.4) in the first and second term (after taking moduli) we obtain

$$
\begin{aligned}
|\rho_n| &\le \frac{|t|^3}{6} \sum_{j=1}^{k_n} \int_{|x|\le\epsilon} |x|^3 d\tilde{F}_{nj}(x) \;+\; t^2 \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x) \\
&\le \frac{|t|^3}{6}\epsilon \sum_{j=1}^{k_n} \int_{|x|\le\epsilon} x^2 d\tilde{F}_{nj}(x) \;+\; t^2 \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x) \\
&= \frac{|t|^3}{6}\epsilon + t^2\left(1 - \frac{|t|\epsilon}{6}\right) \sum_{j=1}^{k_n} \int_{|x|>\epsilon} x^2 d\tilde{F}_{nj}(x)\,,
\end{aligned}
$$

where in the last equality we have used the fact that, for a normalized array,

$$
\sum_{j=1}^{k_n} \mathsf{Var}(\xi_{nj}) = \sum_{j=1}^{k_n} \int_{\mathbf{R}} x^2 d\tilde{F}_{nj}(x) = 1.
$$

Hence $|\rho_n| \to 0$ uniformly in $|t| \le T$. This fact, together with equations (3.11), (3.12), and (3.13) shows that $\log(\phi_n(t)) \to -\frac{t^2}{2}$ and hence that S_n converges in distribution to a standard, normal random variable. ♠

Let us now return to our original problem of a sequence of independent random variables $\{X_i\}$ with distributions $F_i(x)$ and $EX_i = \mu_i$ and $\mathsf{Var}(X_i) = \sigma_i^2$. Our conjecture was that the normalized sums

$$
Z_n := \frac{S_n - \sum_{i=1}^n \mu_i}{B_n} \tag{3.14}
$$

with $S_n = X_1 + \cdots + X_n$ and $B_n^2 := \sum_{i=1}^n \sigma_i^2$ converge in distribution to $\mathcal{N}(0,1)$, provided that some condition is satisfied. This problem easily fits the preceeding framework if we set

$$
\xi_{nj} = \frac{X_j - \mu_j}{B_n}. \tag{3.15}
$$

Then the double array $\{\xi_{nj}\}_{j=1,2,\ldots,n,n=1,2,3,\ldots}$ is normalized since $\sum_{j=1}^n \mathsf{Var}(\xi_{nj}) = \sum_{j=1}^n \frac{\sigma_j^2}{B_n^2} = 1$ for all n and Z_n, as defined in (3.14) is equal to $\xi_{n1}+\xi_{n2}+\cdots+\xi_{nn}$. Thus from the Lindeberg–Feller theorem with ξ_{nj} given by (3.14) we conclude that the necessary and sufficient condition for Z_n to converge to $\mathcal{N}(0,1)$ is

$$
\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^n E\left[(X_i - \mu_i)^2; |X_i - \mu_i| > \epsilon B_n\right] = 0 \qquad \text{for any } \epsilon > 0, \tag{3.16}
$$

or equivalently

$$
\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{i=1}^n \int_{|x-\mu_i|>\epsilon B_n} (x - \mu_i)^2 dF_i(x) = 0 \qquad \text{for any } \epsilon > 0. \tag{3.17}
$$

https://doi.org/10.26219/heal.aueb.5740

Let us now return to the two examples given previously and see what went wrong.

In example 1, $B_n^2 = \frac{1}{3}(1-4^{-n})$ (and of course $\mu_i = 0$) hence, $B_n^2 < 1/3$, thus the quantity we need to examine is

$$\frac{1}{B_n^2}\sum_{i=1}^{n}\int_{|x|>\epsilon B_n} x^2 dF_i(x) \;\geq\; 3\sum_{i=1}^{n}\int_{|x|>\epsilon/2} x^2 dF_i(x).$$

Now this ought to go to zero as $n \to \infty$ for *any* ϵ, however small or large. Notice that F_i has mass only at the points $\pm 2^{-i}$, so these masses will concentrate at the origin as i becomes large. If $\epsilon < \frac{\sqrt{3}}{4}$, then at least the first term in the above sum (and perhaps some of the following) will not vanish. So,

$$\frac{1}{B_n^2}\sum_{i=1}^{n}\int_{|x|>\epsilon B_n} x^2 dF_i(x) \;\geq\; 3\int_{|x|>1/4} x^2 dF_1(x) \;=\; 3\left(\frac{1}{2}2^{-2}+\frac{1}{2}2^{-2}\right) = \frac{3}{4}$$

and hence (3.17) is not satisfied. This explains why the limiting distribution in this case was not normal.

Let us now see what happens in example 2. There $B_n^2 = \frac{4}{3}(4^n-1)$ and hence $B_n^2 < \frac{4^{n+1}}{3}$. Here F_i has mass only at the points $\pm 2^i$, and these move to infinity as i becomes large. Since $B_n < \frac{2^{n+1}}{\sqrt{3}}$, if we take $\epsilon = \frac{\sqrt{3}}{4}$ we see that

$$\int_{|x|>\epsilon B_n} x^2 dF_i(x) \;\geq\; \int_{|x|>2^{n-1}} x^2 dF_i(x) = \begin{cases} 0 & \text{if } i<n \\ \frac{1}{2}\left((2^i)^2+(-2^i)^2\right) = 4^i & \text{if } i \geq n. \end{cases}$$

Thus,

$$\frac{1}{B_n^2}\sum_{i=1}^{n}\int_{|x|>\frac{\sqrt{3}}{4}B_n} x^2 dF_i(x) \;\geq\; \frac{1}{B_n^2}\int_{|x|>2^{n-1}} x^2 dF_n(x) \;=\; \frac{3}{4^{n+1}}4^n \;=\; \frac{3}{4}$$

which means that (3.17) is again not satisfied.

Let us finally examine a third example. Suppose that X_i are independent with $P(X_i = +i) = P(X_i = -i) = 1/2$. Here we have again for convenience $\mu_i = 0$, while $\sigma_i^2 = i^2$. Again F_i is concentrated on two points that, as in example 2, move to infinity as i grows large. $B_n^2 = \sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1)$, as one can easily see by induction. Observe that B_n behaves asymptotically as $\frac{n^{3/2}}{\sqrt{3}}$, and hence, for any fixed ϵ, we will have $n/2 < \epsilon B_n$ when n is large enough. Consequently, for every ϵ there exists a n_ϵ such that $n > n_\epsilon$ implies $\int_{|x|>\epsilon B_n} x^2 dF_i(x) \leq \int_{|x|>n} x^2 dF_i(x) = 0$ for $i \leq n$ (remember that F_i has mass at $\pm i$). In other words, for n sufficiently large,

$$\frac{1}{B_n^2}\sum_{i=1}^{n}\int_{|x|>\epsilon B_n} x^2 dF_i(x) \leq \frac{1}{B_n^2}\sum_{i=1}^{n}\int_{|x|>n} x^2 dF_i(x) = 0$$

and hence (3.17) is satisfied. Thus we expect that the normalized limit in this case will be $\mathcal{N}(0,1)$. Another way to think of this example is that, if χ_i are i.i.d. random variables that take the values ± 1 with equal probability, then

$$\frac{\sum_{i=1}^n \chi_i i}{n^{3/2}} \longrightarrow \mathcal{N}(0,3).$$

(note that instead of $\sqrt{\frac{1}{3}\left(n^3+\frac{3}{2}n^2+\frac{1}{2}\right)}$ we have used $\frac{n^{3/2}}{3}$ as a normalizing constant. Why is this possible?)

Chapter 4

Infinitely Divisible Laws

Let X be a real random variable with characteristic function $\phi(t) := Ee^{itX}$ and distribution function $F(x) = P(X \leq x)$. We say that X is infinitely divisible (or, equivalently, that F, or ϕ are infinitely divisible) if, for every integer n there exist independent, identically distributed random variables X_i, $i = 1, 2, \ldots, n$, such that

$$X \stackrel{d}{=} X_1 + X_2 + \cdots + X_n$$

Equivalently we say that the characteristic function $\phi(t)$ is infinitely divisible if, for every n, there exists a *characteristic function* $\phi_n(t)$ such that

$$\phi(t) = (\phi_n(t))^n \tag{4.1}$$

A few examples will convince us that this definition is not vacuous: If X is $\mathcal{N}(0,1)$ then for each n it can be expressed as a sum of n independent normal $\mathcal{N}(0, \frac{1}{n})$ r.v's. For another example consider a Gamma distributed r.v. with shape parameter α and scale parameter β and corresponding characteristic function $\phi(t) = \left(\frac{\beta}{\beta - it}\right)^{\alpha}$. Since $\left(\frac{\beta}{\beta - it}\right)^{\alpha/n}$ is also a characteristic function (of a Gamma distribution with shape parameter α/n and scale parameter again β) we see that (4.1) is satisfied, hence the Gamma distribution is infinitely divisible.

Theorem 19 *The characteristic function of an i.d. r.v. does not vanish for any real t.*

Theorem 20 *The distribution function of a sum of independent r.v.'s having infinitely divisible distribution function is also infinitely divisible*

Proof: Let X_i, $i = 1, 2, \ldots, k$ be independent r.v.'s with infinitely divisible characteristic functions $\phi_i(t)$. Set $X = X_1 + \cdots + X_k$ and denote by $\phi(t) = Ee^{itx}$ its characteristic function. Clearly $\phi(t) = \phi_1(t)\phi_2(t)\cdots\phi_k(t)$. Since X_i is infinitely divisible, for every integer n $\phi_i^{1/n}(t)$ is also a characteristic function. Hence, since the product of characteristic functions is also a characteristic function, $\phi^{1/n}(t) = \phi_1^{1/n}(t)\phi_2^{1/n}(t)\cdots\phi_k^{1/n}(t)$ is a characteristic function and $\phi(t)$ is infinitely divisible.

Theorem 21 *The limit distribution function of a sequence of infinitely divisible distribution functions is itself infinitely divisible, i.e. if $F_n(x)$ is a sequence of infinitely divisible distribution functions such that $F_n(x) \to F(x)$ for all continuity points of the distribution function F, then F is infinitely divisible.*

 https://doi.org/10.26219/heal.aueb.5740

Proof: Let $\phi_n(t)$ denote the characteristic function of $F_n(x)$ and $\phi(t)$ the ch.f. of $F(x)$. By the convergence theorem we know that $\phi_n(t) \to \phi(t)$ for all t, and in fact this convergence is uniform in t. Since F_n is infinitely divisible, $\phi_n^{1/k}$ is also a characteristic function, and by the continuity of the square root

Theorem 22 *[Lévy–Khinchine Representation] A distribution function F with finite variance (and corresponding characteristic function $\phi(t)$) is infinitely divisible if and only if it has the representation*

$$\log \phi(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - itx\right) \frac{1}{x^2} dG(x) \tag{4.2}$$

where γ is a real constant and G a nondecreasing function of bounded variation.

Proof: Suppose that $\phi(t)$ is i.d. Then from (4.1) for any n we have

$$\log \phi(t) = n \log \phi_n(t) = n \log\left(1 + \phi_n(t) - 1\right)$$

However, for any $T > 0$, as $n \to \infty$, $\phi_n(t) \to 1$ uniformly in $|t| < T$. Hence we can write $\log\left(1 + \phi_n(t) - 1\right) = \left(\phi_n(t) - 1\right)\left(1 + \epsilon_n\right)$ where $\epsilon_n \to 0$ as $n \to \infty$ and

$$\log \phi(t) = n\left(\phi_n(t) - 1\right)\left(1 + \epsilon_n\right)$$

Denote by F_n the distribution function corresponding to the characteristic function ϕ_n. We then have

$$\phi_n(t) - 1 = \int_{\mathbf{R}} \left(e^{itx} - 1\right) dF_n(x)$$

Also,

$$n \int_{\mathbf{R}} x dF_n(x) = \int_{\mathbf{R}} x dF(x) = \gamma.$$

Hence

$$\log \phi(t) = i\gamma t + \lim_{n\to\infty} n \int_{\mathbf{R}} \left(e^{itx} - 1 - itx\right) dF_n(x) \tag{4.3}$$

Set

$$G_n(x) := n \int_{-\infty}^{x} u^2 dF_n(u)$$

Then $\{G_n(x)\}$ is a sequence of increasing functions. Also $\{G_n(\infty)\}$ is bounded. Indeed,

$$n \int_{-\infty}^{\infty} u^2 dF_n(u) = \sigma^2 + \frac{1}{n}\gamma^2$$

With these definitions,

$$\log \phi(t) = i\gamma t + \lim_{n\to\infty} \int_{\mathbf{R}} \frac{e^{itx} - 1 - itx}{x^2} dG_n(x).$$

Helly's first theorem asserts that there exists a subsequence n_k and an increasing function $G(x)$ such that $G_{n_k}(x) \to G(x)$ for all continuity points of $G(x)$. On the other hand $\left|\frac{e^{itx}-1-itx}{x^2}\right| \leq \frac{t^2}{2}$

$$\log \phi(t) = i\gamma t + \int_{\mathbf{R}} \frac{e^{itx} - 1 - itx}{x^2} dG(x).$$

Uniquness of specification: By differentiating twice the above relationship with respect to t we see that

$$-\frac{d^2}{dt^2} \log \phi(t) = \int_{\mathbf{R}} e^{itx} dG(x).$$

Hence, to a given $\phi(t)$ there corresponds a unique function $G(x)$ by the uniqueness theorem for characteristic functions.

4.1 Examples of infinitely divisible distributions

The following infinitely divisible distributions are described by means of their characteristic functions

- The deterministic distribution e^{ita}
- The Normal distribution $e^{-\sigma^2 t^2}$
- The gamma distribution $\frac{1}{(1-it)^\alpha}$, $\alpha > 0$
- The Poisson distribution $e^{-\lambda(1-e^{it})}$
- The compound Poisson distribution $e^{-\lambda(1-\psi(t))}$ where $\psi(t)$ is the characteristic function of some random variable
- The symmetric stable distribution of exponent α: $e^{-c|t|^\alpha}$ where $0 < \alpha < 2$ and $c > 0$.

In the next example we compute the characteristic measure for the gamma distribution: Starting with $\phi(t) = (1-it)^{-\alpha}$ we see that $\frac{d}{dt}\log\phi(t) = \frac{i\alpha}{1-it}$. But $\frac{1}{1-it} = \int_0^\infty e^{itx}e^{-x}dx$ and hence, integrating with respect to t, $\log\phi(t) - \log\phi(0) = i\alpha\int_0^\infty \left[\int_0^t e^{iux}du\right] e^{-x}dx = \alpha\int_0^\infty \left(e^{itx}-1\right)\frac{e^x}{x}dx$. Hence, taking into account that $\int_0^\infty e^{-x}dx = 1$, we have the representation

$$\log\phi(t) = it\alpha + \int_{\mathbf{R}} \left(e^{itx} - 1 - itx\right)\frac{1}{x^2}dG(x) \quad \text{with} \quad dG(x) = \begin{cases} 0 & x < 0 \\ \alpha x e^{-x} & x \geq 0 \end{cases}$$

This corresponds to $G(x) = \alpha\left(1 - (1+x)e^{-x}\right)$ for $x \geq 0$ and $G(x) = 0$ for $x < 0$.

Poisson random variables with non–integer values. Let N be a Poisson random variable with parameter λ: $P(N = k) = \frac{1}{k!}\lambda^k e^{-\lambda}$, $k = 0, 1, 2, \ldots$, and $a > 0$. Then the random variable $X = aN$ has distribution $\mathcal{P}(\lambda, a)$ given by $P(X = ka) = \frac{1}{k!}\lambda^k e^{-\lambda}$, $k = 0, 1, 2, \ldots$, i.e. it takes values on the integer multiples of a. Its characteristic function is

$$e^{\lambda(e^{ita}-1)}.$$

The geometric distribution. Let X be a random variable with distribution $P(X = k) = q^k p$, $k = 0, 1, 2, \ldots,$. We will show that X is infinitely divisible. Its characteristic function is given by

$$\phi(t) = E[e^{itX}] = \sum_{k=0}^\infty q^k p e^{ikt} = \frac{1-q}{1-qe^{it}}.$$

Since $0 < q < 1$, $\log(1-q)$, and $\log(1-qe^{it})$ are well defined, so

$$\log\phi(t) = \log(1-q) - \log(1-qe^{it}) = \sum_{k=1}^\infty \frac{1}{k}q^k\left(e^{ikt}-1\right).$$

Hence

$$\phi(t) = \prod_{k=1}^\infty e^{\frac{1}{k}q^k(e^{ikt}-1)} =: \prod_{k=1}^\infty \varphi_k(t),$$

where, each of the characteristic functions $\varphi_k(t)$ corresponds to a Poisson random variable

https://doi.org/10.26219/heal.aueb.5740

Chapter 5

Martingales in Discrete Time

5.1 Adapted and Predictable processes

On the probability space $(\Omega, \mathcal{F}, P)$ suppose that there has been defined an *increasing* sequence of σ–fields $\mathcal{F}_n$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all n. The family $\{\mathcal{F}_n\}$ is called a *filtration*. In practice $\mathcal{F}_n$ represents the information available at (discrete) time n.

- The process $\{X_n\}_{n\geq 0}$ is **adapted** to $\{\mathcal{F}_n\}$ if for every n X_n is measurable with respect to $\mathcal{F}_n$. We will write $X_n \in \mathcal{F}_n$.
- The process $\{X_n\}_{n\geq 0}$ is **predictable** with respect to $\{\mathcal{F}_n\}$ if for every n X_n is measurable with respect to $\mathcal{F}_{n-1}$ or symbolically $X_n \in \mathcal{F}_{n-1}$.

5.2 Stopping Times

Let T be a nonnegative, integer valued, random variable. T is a stopping time w.r.t. the filtration $\{\mathcal{F}_n\}$ iff the sequence of random variables $\mathbf{1}(T = n)$, $n = 0, 1, 2, \ldots$, is adapted to $\{\mathcal{F}_n\}$. In particular, note that if T is a stopping time then $\{\mathbf{1}(T \leq n)\}$ is also an adapted sequence, while $\{\mathbf{1}(T > n)\}$ is a *predictable* sequence. To see this, write

$$\mathbf{1}(T \leq n) = \sum_{k=0}^{n} \mathbf{1}(T = k)$$

and observe that $\mathbf{1}(T = k) \in \mathcal{F}_k \subset \mathcal{F}_n$ for $k \leq n$. This establishes that $\mathbf{1}(T \leq n) \in \mathcal{F}_n$. On the other hand $\mathbf{1}(T > n) = 1 - \mathbf{1}(T \leq n-1)$ which, in view of the above is a *predictable* sequence.

Proposition 1 *If S, T, are $\mathcal{F}_n$–stopping times then $S+T$, $S \vee T$, $S \wedge T$ are also $\mathcal{F}_n$–stopping times.*

Proof: To prove the first statement, note that

$$\mathbf{1}(S+T = n) = \sum_{k=0}^{n} \mathbf{1}(S = k)\mathbf{1}(T = n-k) \in \mathcal{F}_n.$$

The second follows from $\mathbf{1}(S \vee T \leq n) = \mathbf{1}(S \leq n)\mathbf{1}(T \leq n)$, and the fact that both $\mathbf{1}(S \leq n)$ and $\mathbf{1}(S \leq n)$ are in $\mathcal{F}_n$ since T and S are stopping times. Finally $\mathbf{1}(S \wedge T > n) = \mathbf{1}(S > n)\mathbf{1}(T > n)$.

♠

5.3 Martingales in Discrete Time

Theorem 23 *A process $\{X_n\}$ is a martingale w.r.t. the filtration $\{\mathcal{F}_n\}$ if*

- *X_n is an adapted process, i.e. $X_n \in \mathcal{F}_n$,*
- *$E|X_n| < \infty \; \forall n$,*
- *$E[X_{n+1}|\mathcal{F}_n] = X_n \; \forall n$.*

Example 1: Let $\{Y_i\}$ be independent random variables with $E|Y_i| < \infty$ for all i and consider the filtration $\mathcal{F}_n = \sigma\{Y_1, Y_2, \ldots, Y_n\}$. (This is sometimes called the natural filtration of the process.) Let $EY_i = \mu_i$. The process $X_n = \sum_{i=1}^n Y_i - \mu_i$ is an $\mathcal{F}_n$–martingale.

Example 2: Using the setup of the previous example suppose that, for all i, $\sigma_i^2 = \text{Var}(Y_i) < \infty$. The process $X_n = \left(\sum_{i=1}^n Y_i - \mu_i\right)^2 - \sum_{i=1}^n \sigma_i^2$ is an $\mathcal{F}_n$–martingale.

Example 3: Using again the same setup we assume that Y_i has distribution F_i and $\tilde{F}_i(s) := \int_{-\infty}^{\infty} e^{-sx} dF_i(x)$ is finite for s in a neighborhood of 0. Then

$$X_n := \frac{e^{-s\sum_{i=1}^n Y_i}}{\prod_{i=1}^n \tilde{F}_i(s)},$$

is an $\mathcal{F}_n$–martingale.

Example 4: Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathcal{S}$ and transition probability matrix $P(i,j)$. Also suppose that $f : \mathcal{S} \to \mathbb{R}$ be a real function. Then

$$X_n := \sum_{k=1}^n \left(f(Y_k) - \sum_{j \in \mathcal{S}} P(Y_{k-1}, j) f(j) \right)$$

is an $\mathcal{F}_n$–martingale.

Example 5: [Right Regular Sequences and Induced Martingales for Markov Chains] Let $\{Y_n\}$ be a Discrete Time Markov Chain with state space $\mathcal{S}$ and transition probability matrix $P(i,j)$. Let $f : \mathcal{S} \to \mathbb{R}$ be bounded and satisfy

$$f(i) = \sum_{j \in \mathcal{S}} P(i,j) f(j) , \quad \forall i \in \mathcal{S} .$$

Such sequences (right eigenvectors corresponding to eigenvalue 1) are called right regular sequences. Then

$$X_n = f(Y_n)$$

is a martingale.

Example 6: The above example is a special case of the following more general class of martingales. Let f be a right eigenvector corresponding to an eigenvalue λ of P, i.e.

$$\lambda f(i) = \sum_{j \in \mathcal{S}} P(i,j) f(j) , \quad \forall i \in \mathcal{S} .$$

Assuming that $E|f(Y_n)| < \infty$,

$$X_n = \lambda^{-n} f(Y_n)$$

is a martingale.

Example 7: [Likelihood Ratios] Let $\{Y_n\}$ be an i.i.d. sequence with density g. Let f be another density function. Then the process

$$X_n = \frac{f(Y_0)f(Y_1)\cdots f(Y_n)}{g(Y_0)g(Y_1)\cdots g(Y_n)}$$

is a martingale.

5.4 Submartingales, Supermartingales, and Martingale Transforms

Theorem 24 *$\{X_n\}$ is a submartingale w.r.t. $\{\mathcal{F}_n\}$ iff*

a) $X_n \in \mathcal{F}_n$

b) $E|X_n| < \infty \ \forall n$

c) $E[X_{n+1}|\mathcal{F}_n] \geq X_n, \forall n.$

$\{X_n\}$ is a ***supermartingale*** *w.r.t. $\{\mathcal{F}_n\}$ iff it satisfies a) and b) above and*

c') $E[X_{n+1}|\mathcal{F}_n] \leq X_n, \forall n.$

5.4.1 Convexity and Jensen's Inequality

A function $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is convex iff, for every $\lambda \in (0,1)$ and every $x_1, x_2 \in \mathbb{R}$

$$\phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2)$$

It can be shown that, ϕ is convex iff for every $x_0 \in \mathbb{R}$ there exists $\beta \in \mathbb{R}$ such that

$$\phi(x_0) + \beta(x - x_0) \leq \phi(x) \qquad \text{for all } x \in \mathbb{R}.$$

(This result, which is in fact true for convex functions in $\mathbb{R}^n$ is known as the *supporting hyperplane theorem.*) We are now ready to state the central result about convex functions which we shall need here:

Jensen's Inequality Let ϕ be a convex function and X a random variable with $EX < \infty$. Then

$$\phi(EX) \leq E\phi(X)\,. \tag{5.1}$$

Proof: Apply the supporting hyperplane theorem with $x_0 = EX$ to obtain

$$\phi(EX) + \beta(x - EX) \leq \phi(x) \qquad \text{for all } x \in \mathbb{R}.$$

Hence

$$\phi(EX) + \beta(X - EX) \leq \phi(X)$$

and taking expectations in the above equation establishes (5.1) since $E(X - EX) = 0$. ♠

Theorem 25 *Let $\{X_n\}$ be a martingale and g a convex function. Then $\{g(X_n)\}$ is a submartingale, provided that $E|g(X_n)| < \infty$.*

Examples: Suppose $\{X_n\}$ is a martingale. Then $\{X_n^2\}$ and $\{(X_n - a)^+\}$ are submartingales.

5.4.2 Martingale Transforms

Let $\{M_n\}$ be an $\mathcal{F}_n$–martingale and $\{C_n\}$ an $\mathcal{F}_n$–predictable process. Set $\Delta M_n = M_n - M_{n-1}$,

$$X_n = C_0 M_0 + \sum_{k=1}^{n} C_k \Delta M_k \ .$$

$\{X_n\}$ is a *Martingale Transform.* It is easy to see that martingale transforms are martingales.

Proposition 2 *Suppose $|C_n| \leq K \ \forall n$, where K is a positive real number. Then $\{X_n\}$ is a martingale.*

Proof: We first show that $E|X_n| < \infty$:

$$\begin{aligned} E|X_n| &\leq E|C_0|E|M_0| + \sum_{k=1}^{n} E|C_k|E|\Delta M_k| \\ &\leq K\left(E|M_0| + \sum_{k=1}^{n} E|M_k| + E|M_{k-1}|\right) < \infty \end{aligned}$$

($E|M_k| < \infty$ for all k since $\{M_k\}$ is a martingale.)

Next, check that $E[X_{n+1}|\mathcal{F}_n] = X_n$. Indeed,

$$X_{n+1} = X_n + C_{n+1}\left(M_{n+1} - M_n\right)$$

and, taking expectations,

$$E[X_{n+1}|\mathcal{F}_n] = E[X_n + C_{n+1}\left(M_{n+1} - M_n\right)|\mathcal{F}_n] = X_n + C_{n+1}E[M_{n+1} - M_n|\mathcal{F}_n] = 0$$

The second equality following from the fact that $X_n, C_{n+1} \in \mathcal{F}_n$, and the last from the fact that M_n is a martingale. ♠

5.5 Square–integrable martingales and orthogonality of increments

Theorem 26 *Let $\{X_n\}$ be an $\mathcal{F}_n$–martingale with $EX_n^2 < \infty$. Then, for all integers $i \leq j \leq k \leq l$,*

$$E(X_l - X_k)(X_j - X_i) = 0 \ . \tag{5.2}$$

Furthermore

$$EX_n^2 = EX_0^2 + \sum_{k=1}^{n} E(X_k - X_{k-1})^2 \ . \tag{5.3}$$

Proof: To establish (5.2) note that

$$E[(X_l - X_k)(X_j - X_i)|\mathcal{F}_k] = (X_j - X_i)E[X_l - X_k|\mathcal{F}_k] = 0 \ .$$

To show (5.3) write $X_n = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})$. Then

$$X_n^2 = X_0^2 + 2X_0\sum_{k=1}^{n}(X_k - X_{k-1}) + \sum_{k=1}^{n}(X_k - X_{k-1})^2 + 2\sum_{k=2}^{n}\sum_{j=1}^{k-1}(X_k - X_{k-1})(X_j - X_{j-1}).$$

Taking expectations and using (5.2) yields (5.3). ♠

https://doi.org/10.26219/heal.aueb.5740

5.6 The Doob–Meyer Decomposition

Theorem 27 *Let $\{X_n\}$ be a process adapted to $\mathcal{F}_n$. Then there exists an $\mathcal{F}_n$–martingale $\{M_n\}$ and an $\mathcal{F}_n$–predictable process $\{A_n\}$ with $M_0 = 0$, $A_0 = 0$ such that*

$$X_n = X_0 + M_n + A_n \ . \tag{5.4}$$

This decomposition is essentially unique in that if $X_n = X_0 + \tilde{M}_n + \tilde{A}_n$ $\forall n$, $M_n = \tilde{M}_n$, $A_n = \tilde{A}_n$ $\forall n$ (with probability 1).

If $\{X_n\}$ is a submartingale then $\{A_n\}$ is a nondecreasing process, i.e. $A_{n+1} \geq A_n$ $\forall n$ (w.p. 1).

Proof: If (5.4) is true then

$$X_{n+1} - X_n = M_{n+1} - M_n + A_{n+1} - A_n$$

hence

$$E[X_{n+1} - X_n|\mathcal{F}_n] = E[M_{n+1} - M_n|\mathcal{F}_n] + E[A_{n+1} - A_n|\mathcal{F}_n] \ .$$

Since $\{M_n\}$ is a martingale, $E[M_{n+1} - M_n|\mathcal{F}_n] = 0$. Since $\{A_n\}$ is predictable, $E[A_{n+1} - A_n|\mathcal{F}_n] = A_{n+1} - A_n$. Hence,

$$A_{n+1} = A_n + E[X_{n+1}|\mathcal{F}_n] - X_n \ .$$

Set

$$A_n = \sum_{k=1}^{n} E[X_k|\mathcal{F}_{k-1}] - X_{k-1}, \tag{5.5}$$

$$M_n = \sum_{k=1}^{n} X_k - E[X_k|\mathcal{F}_{k-1}]. \tag{5.6}$$

From (5.5) you can verify that $\{A_n\}$ is $\mathcal{F}_n$–predictable, from (5.6) that $\{M_n\}$ is an $\mathcal{F}_n$–martingale, and adding (5.5)+ (5.6) gives

$$M_n + A_n = X_n - X_0.$$

Note that if $\{X_n\}$ is a submartingale then it is $\mathcal{F}_n$–adapted and therefore the Doob-Meyer decomposition holds with A_n, M_n given by (5.5), (5.6). From (5.5) it follows that

$$A_{n+1} - A_n = E[X_{n+1}|\mathcal{F}_n] - X_n \geq 0,$$

since $\{X_n\}$ is a submartingale.

To show uniqueness, suppose we also have $X_n = X_0 + \tilde{M}_n + \tilde{A}_n$. Then $M_n + A_n = \tilde{M}_n + \tilde{A}_n$ or

$$M_n - \tilde{M}_n = \tilde{A}_n - A_n \ . \tag{5.7}$$

Taking conditional expectations we get

$$E[M_n|\mathcal{F}_{n-1}] - E[\tilde{M}_n|\mathcal{F}_{n-1}] = E[\tilde{A}_n|\mathcal{F}_{n-1}] - E[A_n|\mathcal{F}_{n-1}]$$

 https://doi.org/10.26219/heal.aueb.5740

However

$$E[M_n|\mathcal{F}_{n-1}] = M_{n-1}(\text{martingale})$$
$$E[A_n|\mathcal{F}_{n-1}] = A_n \quad (\text{predictable})$$

(The same relations hold for $\tilde{M}_n$ and $\tilde{A}_n$.) Therefore

$$M_{n-1} - \tilde{M}_{n-1} = \tilde{A}_n - A_n \,. \tag{5.8}$$

From (5.7) and (5.8) we get

$$M_{n-1} - \tilde{M}_{n-1} = M_n - \tilde{M}_n \,. \tag{5.9}$$

(5.9) holds for all n and, by induction,

$$M_n - \tilde{M}_n = M_0 - \tilde{M}_0 = 0 \,.$$

From (5.7) it follows that

$$A_n - \tilde{A}_n = 0 \,.$$

♠

Example: An application of the Doob–Meyer decomposition Let $\{X_n\}$ be an $\mathcal{F}_n$–martingale. Then $\{X_n^2\}$ is a submartingale and

$$X_n^2 = X_0^2 + A_n + M_n$$

where M_n is a martingale and A_n is predictable and are given by the expressions

$$A_n = \sum_{k=1}^{n} E[\Delta X_k^2|\mathcal{F}_{k-1}], \tag{5.10}$$

$$M_n = \sum_{k=1}^{n} X_k^2 - E[X_k^2|\mathcal{F}_{k-1}]. \tag{5.11}$$

5.6.1 Quadratic Variation of a Martingale

Let $\{X_n\}$ be an $\mathcal{F}_n$–martingale. Then $\{X_n^2\}$ is a submartingale which we can decompose into a martingale and an increasing process. This increasing process is called the *quadratic variation* of X, $\langle X\rangle$. We write

$$X_n^2 = M_n + \langle X\rangle_n \,.$$

From the Doob–Meyer decomposition we have

$$\langle X\rangle_n = X_0^2 + \sum_{k=1}^{n} E[(X_k - X_{k-1})^2|\mathcal{F}_{k-1}] \,.$$

5.7 The Optional Sampling Theorem

5.7.1 Optional Sampling Theorem for Submartingales

Let $\{X_n\}$ be a submartingale w.r.t. $\{\mathcal{F}_n\}$ and S, T, be stopping times such that $0 \le S \le T \le m$ (where m is a given integer). Then

$$EX_S \le EX_T \,. \tag{5.12}$$

Proof: Write $X_T = X_0 + (X_1 - X_0) + \cdots + (X_T - X_{T-1})$, or

$$X_T = X_0 + \sum_{k=1}^{m}(X_k - X_{k-1})\mathbf{1}(T \geq k) .$$

Similarly,

$$X_S = X_0 + \sum_{k=1}^{m}(X_k - X_{k-1})\mathbf{1}(S \geq k) .$$

Taking expectations we can write

$$EX_T = EX_0 + \sum_{k=1}^{m} E[E[(X_k - X_{k-1})\mathbf{1}(T \geq k)|\mathcal{F}_{k-1}]] .$$

Note that $\mathbf{1}(T \geq k) = 1 - \sum_{i=0}^{k-1} \mathbf{1}(T = i) \in \mathcal{F}_{k-1}$ and hence

$$E[(X_k - X_{k-1})\mathbf{1}(T \geq k)|\mathcal{F}_{k-1}] = \mathbf{1}(T \geq k)E[(X_k - X_{k-1})|\mathcal{F}_{k-1}]$$

Since $T \geq S$, $\mathbf{1}(T \geq k) \geq \mathbf{1}(S \geq k)$. Also $E[X_k - X_{k-1}|\mathcal{F}_{k-1}] \geq 0$ ($\{X_n\}$ is a submartingale). Hence

$$\mathbf{1}(T \geq k)E[(X_k - X_{k-1})|\mathcal{F}_{k-1}] \geq \mathbf{1}(S \geq k)E[(X_k - X_{k-1})|\mathcal{F}_{k-1}] .$$

From the above it follows that

$$X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1})\mathbf{1}(T \geq k)|\mathcal{F}_{k-1}] \geq X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1})\mathbf{1}(S \geq k)|\mathcal{F}_{k-1}] .$$

Taking expectations:

$$E\left[X_0 + \sum_{k=1}^{m}(X_k - X_{k-1})\mathbf{1}(T \geq k)\right] \geq E\left[X_0 + \sum_{k=1}^{m} E[(X_k - X_{k-1})\mathbf{1}(S \geq k)\right] ,$$

or

$$EX_T \geq EX_S .$$

♠

5.7.2 Doob's Maximal Inequality

Let $\{X_n\}$ be a nonnegative submartingale (i.e. $X_n \geq 0\ \forall n$). Then, $\forall \lambda > 0$, $\forall n$,

$$\lambda P\{\max_{0 \leq k \leq n} X_k > \lambda\} \leq EX_n . \tag{5.13}$$

Proof: Define the stopping time T as

$$T = \begin{cases} \min\{k : X_k > \lambda\} & \text{if} \quad \max_{0 \leq k \leq n} X_k > \lambda \\ n & \text{if} \quad \max_{0 \leq k \leq n} X_k \leq \lambda \end{cases} .$$

Notice that $\{X_T > \lambda\} = \{\max_{0 \leq k \leq n} X_k > \lambda\}$ and therefore

$$P\{X_T > \lambda\} = P\{\max_{0 \leq k \leq n} X_k > \lambda\} .$$

However, from Markov's inequality

$$\lambda P\{X_T > \lambda\} \leq EX_T ,$$

while from the Optional Sampling Theorem,

$$EX_T \leq EX_n .$$

The conclusion of the theorem follows from the above. ♠

 https://doi.org/10.26219/heal.aueb.5740

5.7.3 The Optional Sampling Theorem for Martingales

Let $\{X_n\}$ be a martingale w.r.t. $\{\mathcal{F}_n\}$. We know that $EX_n = EX_0$. If T is a stopping time, under what conditions is $EX_T = EX_0$? We start with

Lemma 2 *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathcal{F}_n\}$. Then, for all $n \geq k$,*

$$E[X_n \mathbf{1}(T = k)] = E[X_k \mathbf{1}(T = k)] \ .$$

Proof: Indeed

$$E[X_n \mathbf{1}(T = k)] = E[E[X_n \mathbf{1}(T = k)|\mathcal{F}_k]] = E[\mathbf{1}(T = k)E[X_k|\mathcal{F}_k]] = E[\mathbf{1}(T = k)X_k]$$

♠

Lemma 3 *With the assumptions of the previous lemma*

$$E[X_{T\wedge n}] = EX_0 \ .$$

Proof: We can write $X_{T\wedge n} = \sum_{k=0}^{n-1} X_k \mathbf{1}(T = k) + X_n \mathbf{1}(T \geq n)$ and taking expectations,

$$\begin{aligned} E[X_{T\wedge n}] &= \sum_{k=0}^{n-1} E[X_k \mathbf{1}(T = k)] + E[X_n \mathbf{1}(T \geq n)] \\ &= \sum_{k=0}^{n-1} E[X_n \mathbf{1}(T = k)] + E[X_n \mathbf{1}(T \geq n)] \\ &= E\left[X_n \left(\sum_{k=0}^{n-1} \mathbf{1}(T = k) + \mathbf{1}(T \geq n)\right)\right] \\ &= EX_n = EX_0 \ . \end{aligned}$$

Theorem 28 *Let $\{X_n\}$ be a martingale and T a stopping time w.r.t. $\{\mathcal{F}_n\}$. Suppose that $P(T < \infty) = 1$ and $E[\sup_k |X_{T\wedge k}|] < \infty$. Then $EX_T = EX_0$.*

Proof: From the previous lemma we have $EX_{T\wedge n} = EX_0$ $\forall n$. Since $P(T < \infty) = 1$, $\lim_{n\to\infty} X_{T\wedge n} = X_T$. Finally, $X_{T\wedge n} \leq \sup_k |X_{T\wedge k}|$. Use the Dominated Convergence Theorem to conclude that

$$\lim_{n\to\infty} E[X_{T\wedge n}] = E[\lim_{n\to\infty} X_{T\wedge n}] = EX_T \ .$$

♠

5.7.4 The Kolmogorov–Doob Inequality

Theorem 29 *Let X_n be a square–integrable martingale (i.e. $EX_n^2 < \infty$ for all n). Then*

$$P(\max_{0\leq i\leq n} |X_i| \geq \epsilon) \leq \frac{EX_n^2}{\epsilon^2} \ .$$

https://doi.org/10.26219/heal.aueb.5740

Proof: Define the sets $A_k = \{|X_i| < \epsilon,\ i \leq k\}$, $B_k = A_{k-1} \cap \{|X_k| \geq \epsilon\}$. Then $\Omega = A_n \bigcup (\bigcup_{k=0}^{n} B_k)$ and

$$EX_n^2 = \sum_{k=0}^{n} E[X_n^2 \mathbf{1}(B_k)] \ + \ E[X_n^2 \mathbf{1}(A_n)] \geq \sum_{k=0}^{n} E[X_n^2 \mathbf{1}(B_k)]$$

We have however

$$\begin{aligned} E[X_n^2 \mathbf{1}(B_k)] &= E[(X_n - X_k + X_k)^2 \mathbf{1}(B_k)] \\ &= E[(X_n - X_k)^2 \mathbf{1}(B_k)] + 2E[(X_n - X_k)X_k \mathbf{1}(B_k)] + E[X_k^2 \mathbf{1}(B_k)] \\ &\geq E[X_k^2 \mathbf{1}(B_k)] \, . \end{aligned}$$

Hence

$$EX_n^2 \geq \sum_{k=0}^{n} E[X_k^2 \mathbf{1}(B_k)] \geq \epsilon^2 \sum_{k=0}^{n} E\mathbf{1}(B_k) = \epsilon^2 E[\sum_{k=0}^{n} \mathbf{1}(B_k)] = \epsilon^2 P(\bigcup_{k=0}^{n} B_k) \, ,$$

from which the conclusion of the theorem follows immediately. ♠

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

Chapter 6

Brownian Motion

6.1 Brownian Motion

A stochastic process $\{W_t, t \geq 0\}$ is called Standard Brownian Motion if it satisfies the following three postulates

i) $P(W_0 = 0) = 1$, i.e. the process starts with probability 1 from 0 at time 0.

ii) $\{W_t, t \geq 0\}$ has continuous paths with probability 1.

iii) The increments are independent i.e. if $0 \leq t_i < t_2 < \cdots < t_k$ then $P(W_{t_i} - W_{t_{i-1}} \in H_i; i = 1, 2, \ldots, k) = \prod_{i=1}^{k} P(W_{t_i} - W_{t_{i-1}} \in H_i)$ for any (Borel) subsets H_i of $\mathbb{R}$.

iv) For $0 \leq s < t$, $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$:

$$P(W_t - W_s \in H) = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{-x^2/2(t-s)} dx$$

From the above postulates it follows that the finite dimensional distributions of the process W_t are given by

$$P(W_{t_1} \in (x_1, x_1 + dx_1), \ldots, W_{t_n} \in (x_n, x_n + dx_n)) = f(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n) dx_1 \cdots dx_n$$

with

$$\begin{aligned} f(x_1, \ldots, x_n; t_1, t_2, \ldots, t_n) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{t_1(t_2 - t_1)\cdots(t_n - t_{n-1})}} e^{-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right\}} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2} x^T \Sigma^{-1} x} \end{aligned}$$

where x^T denotes the transpose of $x = (x_1, \ldots, x_n)$ and

$$\Sigma = E\left[\begin{pmatrix} W_{t_1} \\ \vdots \\ W_{t_n} \end{pmatrix} (W_{t_1}, \ldots, W_{t_n})\right] = \begin{bmatrix} EW_{t_1}W_{t_1} & \cdots & EW_{t_1}W_{t_n} \\ \cdots & EW_{t_i}W_{t_j} & \cdots \\ EW_{t_n}W_{t_1} & \cdots & EW_{t_n}W_{t_n} \end{bmatrix} = [t_i \wedge t_j]_{i=1,\ldots,n j=1,\ldots,n}$$

is the corresponding covariance matrix, i.e. the finite dimensional distributions of brownian motion are normal. This means that brownian motion is a Gaussian process.

6.1.1 Properties of Standard Brownian Motion

1. Markov Property. Brownian motion is a Markov process with stationary transition probabilities

$$\begin{aligned} P_t(x, A) &= P(W_{t+s} \in A | W_s = x) = P(W_{t+s} - W_s \in A - x | W_s = x) \\ &= P(W_t \in A - x) = \int_{A-x} \phi(u) du \end{aligned}$$

where $A - x$ is the set $\{y - x : y \in A\}$ and $\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$.

2. Scaling Property. $\forall c > 0$ $\{\sqrt{c} W_{t/c}; t \geq 0\} \stackrel{d}{=} \{W_t; t \geq 0\}$.
Indeed, $\sqrt{c} W_{t/c}$ has continuous paths, stationary and independent increments, and the correct distribution.

3. Symmetry. $\{-W_t; t \geq 0\} \stackrel{d}{=} \{W_t; t \geq 0\}$.

4. Time reversal. $\{t W_{1/t}; t \geq 0\} \stackrel{d}{=} \{W_t; t \geq 0\}$.

6.2 Maximum of the Standard Brownian Motion

Let $M_t = \max\{W_u; 0 \leq u \leq t\}$, where as usual $\{W_t; t \geq 0\}$ is SBM (Standard Brownian Motion). For fixed t this is a nonnegative random variable, while, if we consider the process $\{M_t; t \geq 0\}$ then we have a process with $M_0 = 0$ and nondecreasing sample paths. Here we will compute the distribution of the random variable M_t using *the reflection principle*. Define $\tau := \inf\{u \geq 0 : W_u = a\}$ the first time when the brownian motion reaches the level a (note that τ is a stopping time). Now define new process, $\widehat{W}_u$, via the relationship

$$\widehat{W}_u = \begin{cases} W_u & u < \tau \\ a - (W_u - a) & u \geq \tau \end{cases}$$

In the figure below, $\widehat{W}$ is the process that is identical with W up to time τ, and after that time results from the reflection of W around the level a (the grey path in the figure). Note also that $\{M_t \geq a\} = \{\tau \leq t\}$ (i.e. the two events are the same). Now

$$P(M_t \geq a) = P(M_t \geq a, W_t \geq a) + P(M_t \geq a, W_t < a)$$

and $\{M_t \geq a\} \subset \{W_t \geq a\}$, so that $P(M_t \geq a, W_t \geq a) = P(W_t \geq a)$. Also,

$$\begin{aligned} P(M_t \geq a, W_t < a) &= P(\tau \leq t) P(W_t < a \mid \tau \leq t) \\ &= P(\tau \leq t) P(\widehat{W}_t > a \mid \tau \leq t) \\ &= P(\widehat{W}_t > a, \tau \leq t) \\ &= P(W_t > a, M_t \geq a) \\ &= P(W_t > a) \end{aligned}$$

From the above it follows that

$$P(M_t \geq a) = 2P(W_t > a) = \int_a^\infty \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx, \qquad a \geq 0. \tag{6.1}$$

Equivalently, we can say that $M_t \stackrel{d}{=} |W_t|$. This same formula gives the distribution of the time required for SBM to reach a given level $x > 0$: If we denote by $\tau_x := \inf\{t \geq 0 : W_t = x\}$ we have

$$P(\tau_x \leq t) = P(M_t \geq x) = \int_x^\infty \sqrt{\frac{2}{\pi t}} e^{-y^2/2t} dy$$

The change of variables $y = z\sqrt{t}$ transforms the above equation into

$$P(\tau_x \leq t) = \int_{x/\sqrt{t}}^\infty \sqrt{\frac{2}{\pi}} e^{-z^2/2} dz$$

and differentiating with respect to t we obtain the density function of τ_x:

$$f_{\tau_x}(t) = \frac{1}{\sqrt{2\pi t^3}} x e^{-x^2/2t}, \qquad t \geq 0.$$

(This is the inverse Gaussian distribution.)

ftbpFU5.8608in2.3289in0ptReflection principlereflection.bmp

6.3 Martingales associated with Brownian Motion

It is easy to see that standard brownian motion is a martingale. If we denote by $\mathcal{F}_s := \sigma\{W_u; 0 \leq u \leq s\}$ the history of the process up to time s then

$$E[W_t|\mathcal{F}_s] = W_s + E[W_t - W_s|\mathcal{F}_s] = W_s$$

the second term in the above equation vanishing as a result of the independent increments property.

This property, together with the *optional stopping theorem* allows us to compute probabilities of reaching boundaries. Suppose that $W_0 = x$ and let $a < x < b$. Set $\tau = \inf\{t \geq 0 : W_t = a \text{ or } b\}$. Then, by the optional stopping theorem we have

$$EW_\tau = EW_0 = x.$$

However $W_\tau = a\mathbf{1}(W_\tau = a) + b\mathbf{1}(W_\tau = b)$, and if we denote by $p_a = P(W_\tau = a)$ (and similarly for p_b) we have $ap_a + bp_b = x$ which gives (since $p_a + p_b = 1$)

$$p_a = \frac{b - x}{b - a}.$$

Similarly, one can easily show that the process $S_t = W_t^2 - t$ is also a martingale. Indeed,

$$\begin{aligned} E[W_t^2 - t|\mathcal{F}_s] &= E[(W_t - W_s)^2 + 2W_s(W_t - W_s) + W_s^2 - t|\mathcal{F}_s] \\ &= E[(W_t - W_s)^2|\mathcal{F}_s] + 2W_s E[W_t - W_s|\mathcal{F}_s] + W_s^2 - t \\ &= (t - s) + 0 + W_s^2 - t = W_s^2 - s \end{aligned}$$

With τ defined as before let us use the optional sampling theorem again. This time we obtain

$$EW_\tau^2 - E\tau = x^2$$

which gives

$$p_a a^2 + p_b b^2 - E\tau = x^2$$

or

$$\frac{(b-x)a^2+(x-a)b^2}{b-a}-x^2=E\tau$$

from which we obtain

$$E\tau = ab.$$

An important martingale associated with brownian motion is the exponential martingale. Suppose here that W_t is $BM(\mu,\sigma^2)$. Then, if θ is any real number

$$M_t := e^{\theta W_t - q(\theta)t}, \qquad \text{with } q(\theta)=\mu\theta+\frac{1}{2}\theta^2\sigma^2$$

is a martingale. Indeed,

$$E[M_t|\mathcal{F}_s] = E[e^{\theta(W_t-W_s)-q(\theta)(t-s)}|\mathcal{F}_s]M_s = M_s$$

the last equality following from the fact that $Ee^{\theta(W_t-W_s)} = e^{\mu\theta(t-s)+\frac{1}{2}\theta^2\sigma^2(t-s)}$.

We have thus seen that M_t is a martingale for any choice of θ. If we set $\theta=\theta_0=-\frac{2\mu}{\sigma^2}$ we see that $q(\theta_0)=0$ and thus the exponential martingale becomes $e^{\theta_0 W_t}$. We can use this to compute p_a and p_b (defined as before) when $\mu\neq 0$. Indeed, in this case, from the optional sampling theorem we have

$$E[e^{\theta_0 W_\tau}] = e^{\theta_0 x}$$

or

$$p_a e^{\theta_0 a} + p_b e^{\theta_0 b} = e^{\theta_0 x}$$

which gives

$$p_a = \frac{e^{\frac{2\mu}{\sigma^2}(b-x)}-1}{e^{\frac{2\mu}{\sigma^2}(b-a)}-1}.$$

The optional sampling theorem can also be used to obtain the Laplace transform of the time until we hit the boundary. Here we will assume that $\mu=0$, $\sigma=1$ which corresponds to $q(\theta)=\frac{1}{2}\theta^2$, in order to simplify the algebra. We start with

$$E[e^{\theta W_\tau - \tau q(\theta)}] = e^{\theta x}.$$

or

$$\begin{aligned} e^{\theta x} &= p_a E[e^{\theta W_\tau - \tau q(\theta)}|W_\tau = a] + p_b E[e^{\theta W_\tau - \tau q(\theta)}|W_\tau = b] \\ &= p_a e^{\theta a} E[e^{-q(\theta)\tau}|W_\tau = a] + p_b e^{\theta b} E[e^{-q(\theta)\tau}|W_\tau = b]. \end{aligned}$$

We seem to have the problem that this is one equation and we have two unknowns, $E[e^{-q(\theta)\tau}|W_\tau=a]$ and $E[e^{-q(\theta)\tau}|W_\tau=b]$ but in fact we can get around this problem by setting

$$s = q(\theta) = \frac{1}{2}\theta^2.$$

There are two solutions to this equation,

$$\theta_1 = \sqrt{2s}, \qquad \text{and} \qquad \theta_2 = -\sqrt{2s}.$$

Thus, if we set $f_a(s) = E[e^{-s\tau}; W_\tau = a]$ and $f_b(s) = E[e^{-s\tau}; W_\tau = b]$, we have

$$\begin{aligned} e^{x\sqrt{2s}} &= e^{a\sqrt{2s}} f_a(s) + e^{b\sqrt{2s}} f_b(s) \\ e^{-x\sqrt{2s}} &= e^{-a\sqrt{2s}} f_a(s) + e^{-b\sqrt{2s}} f_b(s). \end{aligned}$$

From this system we can compute $f_a(s)$, $f_b(s)$ separately, and hence also $Ee^{-s\tau} = f_a(s) + f_b(s)$. In fact, adding and subtracting the above equations we get

$$\begin{aligned} \cosh(x\sqrt{2s}) &= \cosh(a\sqrt{2s}) f_a(s) + \cosh(b\sqrt{2s}) \\ \sinh(x\sqrt{2s}) &= \sinh(a\sqrt{2s}) f_a(s) + \sinh(b\sqrt{2s}) \end{aligned}$$

or, using the fact that $\sinh(\alpha - \beta) = \sinh\alpha\cosh\beta - \cosh\alpha\sinh\beta$, we obtain

$$f_a(s)\sinh(b-a)\sqrt{2s} = \sinh(b-x)\sqrt{2s}$$

$$f_b(s)\sinh(b-a)\sqrt{2s} = \sinh(x-a)\sqrt{2s}$$

We thus have

$$f(s) = f_a(s) + b_b(s) = \frac{\sinh\left((x-a)\sqrt{2s}\right) + \sinh\left((b-x)\sqrt{2s}\right)}{\sinh\left((b-a)\sqrt{2s}\right)}$$

and using the formulas $\sinh 2\alpha = 2\sinh\alpha\cosh\alpha$, $\sinh\alpha + \sinh\beta = 2\cosh\left(\frac{\beta-\alpha}{2}\right)\sinh\left(\frac{\alpha+\beta}{2}\right)$ we obtain

$$f(s) = \frac{\cosh\left(\left(\frac{b+a}{2} - x\right)\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}$$

Since we can take $x = 0$ without loss of generality, this formula simplifies as follows

$$f(s) = \frac{\cosh\left(\frac{b+a}{2}\sqrt{2s}\right)}{\cosh\left(\frac{b-a}{2}\sqrt{2s}\right)}.$$

In particular, when $b = \ell > 0$, $a = -\ell$, then

$$f(s) = \frac{1}{\cosh\left(\ell\sqrt{2s}\right)}$$

6.4 Total and Quadratic Variation

Let f be a real function. The *total variation* of f over an interval $[0, b]$ is defined by the limit

$$Vf(0,b) = \lim_{n\to\infty} \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|$$

$t_k = \frac{k}{n}b$, $k = 0, 1, \ldots, n-1$.

Remark: If f is monotonic, $Vf(0,b) = |f(b) - f(0)|$. To give another example, suppose f is right continuous and there exist points (countably many at the most) T_i, $i = 1, 2, \ldots$, such that f is absolutely continuous on (T_i, T_{i+1}) and has jumps of size J_i at T_i. In that case

$$f(t) = f(0) + \sum_{0<T_i\le t} J_i + \int_0^t f'(u)du \tag{6.2}$$

and

$$Vf(0,b) = \int_0^b |f'(u)|du + \sum_{0<T_i\le b} |J_i| .$$

The *quadratic variation* of f is defined as

$$Qf(0,b) = \lim_{n\to\infty} \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|^2 .$$

Suppose f is absolutely continuous, i.e. $f(t) = f(0) + \int_0^t f'(u)du$. Then, from the mean value theorem,

$$Qf(0,b) = \lim_{n\to\infty} \sum_{k=0}^{n-1} |f'(t_k+\xi_k)|^2 \frac{1}{n^2}$$

where $\xi_k \in (0, \frac{1}{n})$. If $|f'(t)| \le B\ \forall t \in [0,b]$, $0 \le Qf(0,b) \le \lim_{n\to\infty} B^2 \sum_{k=0}^{n-1} \frac{1}{n^2} = 0$. If f is as in (6.2 then

$$Qf(0,b) = \sum_{0<T_i\le t} J_i^2 .$$

6.4.1 Quadratic Variation of Brownian Sample Paths

Let $W(t)$ be a standard Brownian motion. For every fixed $t>0$,

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right]^2 = t \qquad \text{w.p.1} \tag{6.3}$$

i.e. the quadratic variation of brownian paths is $QW(0,t) = t$. (The equality holds both with probability 1 and in the mean square sense.) One implication of this is that the total variation of the paths is infinite

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right| = \infty \qquad \text{w.p.1.}$$

This is a consequence of

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right| \ge \frac{\sum_{k=1}^{2^n} \left[W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right]^2}{\max_{j=1,\dots,2^n} \left|W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right|} .$$

The numerator converges to t w.p.1 as $n\to\infty$ while the denominator converges to zero since $W(t)$ is continuous (and hence uniformly continuous on bounded intervals) w.p. 1.

To show (6.3) (with convergence in the mean square sense) define

$$\delta_{k,n} := \left(W(\tfrac{k}{2^n}t) - W(\tfrac{k-1}{2^n}t)\right)^2 - \frac{t}{2^n} .$$

Note that $E\delta_{1,n} = 2^{-n}$ and $E\delta_{1,n}^2 = 3\cdot 2^{-2n}$ It suffices to show that $\sum_{k=1}^{2^n} \delta_{k,n} \overset{\text{m.s.}}{\to} 0$. The independence of brownian motion increments implies that the L^2 norm of this r.v. is $2^n E\delta_{1,n}^2 = 2^n \cdot 3 \cdot 2^{-2n} = 3\cdot 2^{-n} \to 0$.

The above calculations lead easily to the conclusion that (6.3) holds w.p. 1 as well. Fix $\epsilon > 0$ and let

$$A_n = \{\omega : |\sum_{k=1}^{2^n} \delta_{k,n}| > \epsilon\}.$$

Then, from Chebychev's inequality

$$P(A_n) \leq \frac{2t^2}{\epsilon^2 2^n}$$

and thus $\sum_{k=1}^{2^n} P(A_n) = \frac{2t^2}{\epsilon^2}$. Hence, for any given ϵ, the Borel–Cantelli theorem implies that only finitely many of the A_n occur, i.e. that, for any ϵ there exists $n_0(\epsilon)$ such that $n > n_0(\epsilon)$ implies $\sum_{k=1}^{2^n} \delta_{k,n} < \epsilon$.

6.5 Gaussian Processes

A stochastic process $\{X_t; t \in \mathbb{R}\}$ is called a Gaussian process if, for every k and every $t_1 < t_2 < \cdots < t_k$, the distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is multidimensional Gauss. It is clear that to define the joint distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ it suffices to determine the vector $(m(t_1), m(t_2), \ldots, m(t_k))$ and the covariance matrix

$$\begin{bmatrix} R(t_1,t_1) & R(t_1,t_2) & \cdots & R(t_1,t_k) \\ R(t_2,t_1) & R(t_2,t_2) & \cdots & R(t_2,t_k) \\ \vdots & \vdots & & \vdots \\ R(t_k,t_1) & R(t_k,t_2) & \cdots & R(t_k,t_k) \end{bmatrix}. \tag{6.4}$$

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

Chapter 7

Stochastic Integrals

7.1 L^2 theory of random variables

7.1.1 A brief overview of linear spaces of random variables

Consider the family, L^2, of all random variables on the probability space $(\Omega, \mathcal{F}, P)$ that have zero mean and finite second moment, i.e. for every $X \in L^2$, $EX = 0$, $EX^2 < \infty$. It is easy to see that this family is ***a linear space over*** $\mathbb{R}$ i.e. that it satisfies the axioms

L1. $0 \in L^2$

L2. If α, β are two real numbers and $X, Y \in L^2$ then $\alpha X + \beta Y \in L^2$.

A norm is a function $\|\cdot\| : L^2 \to \mathbb{R}_0^+$ from the elements of L^2 to the nonnegative reals that has the following properties

N1. (Nonnegativity) For all $x \in L^2$, $\|x\| \geq 0$,

N2. $\|x\| = 0$ iff $x = 0$,

N3. $\|\alpha x\| = |\alpha| \, \|x\|$ for all $x \in L^2$, $\alpha \in \mathbb{R}$.

N4. (Triangular inequality) $\|x + y\| \leq \|x\| + \|y\|$

An inner product is a function $(\cdot, \cdot) : L^2 \times L^2 \to \mathbb{R}$ such that

IP1. $(aX, Y) = (X, aY) = a(X, Y)$

IP2. $(X + Y, Z) = (X, Z) + (Y, Z)$

We define an *inner product* on L^2 via the relationship

$$(X, Y) := EXY \tag{7.1}$$

A linear space on which an inner product has been defined is an *inner product space*. The inner product induces a norm via the definition $\|x\| = \sqrt{(x, x)}$.

https://doi.org/10.26219/heal.aueb.5740

The elements $x_1, x_2, \ldots, x_n$ of L^2 are *linearly independent* iff

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0$$

implies $\alpha_i = 0$, $i = 1, 2, \ldots, n$.

At this point it is customary to define the *dimension* of the linear space as the maximum number of linearly independent elements of the space. In the ordinary Euclidean space $\mathbb{R}^n$ this dimension is of course n. However, since we are willing to assume that our probability space $(\Omega, \mathcal{F}, P)$ is rich enough to support sequences of independent random variables X_i, we have to dispense with the requirement that our space has finite dimension.

Note that two random variables that differ only on a set of measure 0 have to be identified here: Indeed, if $P(X = Y) = 1$ then certainly $E(X - Y)^2 = 0$, hence $\|X - Y\| = 0$ which implies that $X - Y = 0$ according to (N2). Thus when we deal with random variables in L^2 we have to think of them rather as equivalence classes.

A sequence of elements of L^2, $\{X_n\}$, is said to converge to an element of $X \in L^2$ if $\|X_n - X\| \to 0$ as $n \to \infty$. Note that this is precisely L^2 convergence for the sequence of random variables.

A sequence $\{X_n\}$ is Cauchy, if

$$\|X_n - X_m\| \to 0 \qquad \text{as } n, m \to \infty. \tag{7.2}$$

Clearly every convergent sequence is Cauchy since, if $X_n \to X$ then, using the triangular inequality (N4) we have

$$\|X_n - X_m\| \le \|X_n - X\| + \|X_m - X\|$$

and each of the two terms on the right side go to 0 as n and m go to infinity. On the other hand, a Cauchy sequence is not necessarily convergent. While (7.2) guarantees that the elements of the sequence approach each other more and more as m and n grow large, there is no guarantee that the limit this sequence is approaching is actually an element of L^2.

All Cauchy sequences are bounded, i.e. if $\{X_n\}$ is a Cauchy sequence then $\sup_n \|X_n\| < \infty$ which means that there exists $M > 0$ such that $EX_n^2 \le M$ for all $n \in \mathbb{N}$.

The space L^2 is *complete* if every Cauchy sequence of elements of L^2 converges to an element of L^2.

Theorem 30 *L^2 is complete.*

Proof. We start with a Cauchy sequence X_n of elements of L^2. According to the definition we have to show that there exists a random variable X with $EX^2 < \infty$ such that $\|X_n - X\| \to 0$ as $n \to \infty$. Choose a subsequence n_k such that $\|X_m - X_n\| < 2^{-3k/2}$ when m and n are greater than or equal to n_k. This means that $E(X_{n_{k+1}} - X_{n_k})^2 < 2^{-3k}$. Using Chebychev's inequality we have

$$P\left(\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}\right) \le \frac{E(X_{n_{k+1}} - X_{n_k})^2}{2^{-2k}} < 2^{-k}.$$

Since

$$\sum_{k=1}^{\infty} P\left(\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}\right) < \sum_{k=1}^{\infty} 2^{-k} < \infty,$$

the Borel–Cantelli lemma insures that, with probability 1, only finitely many of the inequalities $\left|X_{n_{k+1}} - X_{n_k}\right| > 2^{-k}$ are true. This is equivalent to saying that there exists some k_0 (which

 https://doi.org/10.26219/heal.aueb.5740

may depend on ω) such that, for all $k \geq k_0$, $|X_{n_{k+1}} - X_{n_k}| \leq 2^{-k}$. Hence, w.p. 1 the series $\sum_{k=1}^{\infty} |X_{n_{k+1}} - X_{n_k}|$ converges (since it is dominated by a convergent series). This in turn implies that the telescopic series $\sum_{k=1}^{\infty}(X_{n_{k+1}} - X_{n_k})$ converges absolutely and thus that $\lim_{k\to\infty} X_{n_k} =: X$ exists.

It remains to show that $EX^2 < \infty$

Finally we have to show that $\|X_n - X\| \to 0$ when $n \to \infty$. Indeed, for any given ϵ, choose n_k such that $\|X_{n_k} - X\| < \epsilon/2$ and N such that $\|X_n - X_m\| < \epsilon/2$ whenever $m \geq N, n \geq N$. Then, from the triangle inequality,

$$\|X_n - X\| \leq \|X_n - X_{n_k}\| + \|X_{n_k} - X\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

7.2 Integration with respect to functions of bounded variation

To understand the challenges involved in defining the stochastic integral we have to recall first the definition of the ordinary integral. Historically, arriving at a satisfactory definition was by no means a simple task and it has only been completed in the first two decades of the twentieth century. Suppose that F is a function of bounded variation, f a continuous function defined on a closed interval $[a, b]$, and let $t_i^n := a + \frac{i}{n}(b-a)$, $i = 0, 1, 2, \ldots, n$. Then the so-called *Riemann- Stieltjes* integral can be defined as

$$\int_a^b f(x)dF(x) := \lim_{n\to\infty} \sum_{i=0}^{n-1} f(t_i^n)\left[F(t_{i+1}^n) - F(t_i^n)\right].$$

There is of course nothing special about the equally spaced partition we have used above and in fact, one can show that any partition, $a = t_0^n < t_1^n < \cdots < t_i^n < \cdots < t_n^n = b$ of the interval $[a, b]$ will yield the same limit as $n \to \infty$, provided of course that $\max_{0\leq i\leq n-1}(t_{i+1}^n - t_i^n) \to 0$ as $n \to \infty$.

Why do we require F to be a function of bounded variation? The reason is that functions of bounded variation correspond to *signed measures*. Any increasing function F defines a measure on the real line via the relationship $\mu(a, b] = F(b) - F(a)$. Think of the measure $\mu(a, b]$ of the interval $(a, b]$ as the total mass of the interval. Since F is increasing, the mass of any interval is nonnegative and if F is absolutely continuous, i.e. if $F(x) - F(0) = \int_0^x F'(u)du$, the derivative of F correspond to the mass density.

Similarly, a function of bounded variation can be written as the difference of two increasing functions G^+ and G^-: $F(x) = G^+(x) - G^-(x)$. This representation is unique (up to an arbitrary initial value, say $G^-(-\infty) = G^+(-\infty) = 0$). Thus, if we can think of increasing functions as mass distributions on the real line, we can think of bounded variation as electrical charge distributions that can be positive in some places and negative in others. In this case the signed measure μ of the interval $(a, b]$ is the total charge of the interval (positive – negative) i.e. $\mu(a, b] = F(b) - F(a) = (G^+(b) - G^+(a)) - (G^-(b) - G^-(a))$.

The real problem that presents itself when we try to define

$$\int_0^t f_s dW_s$$

is that, since W_t has paths of infinite total variation, they do not define a (signed) measure the way a bounded variation function does, so it is not at all clear how to define the integral and what its precise meaning would be.

https://doi.org/10.26219/heal.aueb.5740

7.3 Definition of the Ito Integral

Denote by H^2 the set of all adapted processes on $(\Omega, \mathcal{F}, P)$ satisfying

$$E \int_0^t X_s^2 ds < \infty \qquad \forall t \geq 0.$$

A process X is *simple* if there exists a real sequence $\{t_k\}$, $t_k \to \infty$ as $k \to \infty$ such that

$$X(t, \omega) = X(t_{k-1}, \omega) \qquad \text{on } t \in [t_{k-1}, t_k); \qquad k = 0, 1, \ldots .$$

Also

S: The set of all simple adapted processes,

S^2: The set of all simple adapted processes in H^2,

L^2: The set of all random variables ξ in $(\Omega, \mathcal{F}, P)$ such that $E(\xi^2)^{1/2} < \infty$.

Define a norm in H^2 by means of

$$\|X\| = \left(E \int_0^t X_s^2 ds \right)^{1/2} .$$

Theorem 31 *S^2 is dense in H^2 i.e. for all $X \in H^2$ there exist simple processes $\{X_n\}$ such that*

$$\|X_n - X\| \to 0 \qquad n \to \infty$$

We will denote the stochastic integral which we are about to define by $I_t(X) := \int_0^t X_s dW_s$, $t \geq 0$. For simple processes this task is easy. We set

$$I(X) = \sum_{k=0}^{n-1} X_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \tag{7.3}$$

The stochastic integral defined above has the following two important properties.

Proposition 3 *For $X \in S^2$, $EI(X) = 0$ and $\|I(X)\| = \|X\|$.*

Proof:

$$EI(X) = \sum_{k=0}^{n-1} E \left[E \left[X_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \middle| \mathcal{F}_{t_k} \right] \right]$$

However

$$E \left[X_{t_k} \left(W_{t_{k+1}} - W_{t_k} \right) \middle| \mathcal{F}_{t_k} \right] = X_{t_k} E \left[W_{t_{k+1}} - W_{t_k} \middle| \mathcal{F}_{t_k} \right] = 0$$

whence we obtain $EI(X) = 0$. To show that the *isometry* $\|I(X)\| = \|X\|$ holds as well, i.e. that $EI(X)^2 = E \int_0^t X_s^2 ds$ we first note that

$$I^2(X) = \sum_{k=0}^{n-1} X_{t_k}^2 \left(W_{t_{k+1}} - W_{t_k} \right)^2 + 2 \sum_{j=0}^{n-2} \sum_{k=j+1}^{n-1} X_{t_j} X_{t_k} \left(W_{t_{j+1}} - W_{t_j} \right) \left(W_{t_{k+1}} - W_{t_k} \right).$$

Taking expectations on both sides of the above equation we will have to deal with two types of terms.

$$E[X_{t_k}^2\,(W_{t_{k+1}} - W_{t_k})^2] \tag{7.4}$$

and

$$E[X_{t_j}X_{t_k}\,(W_{t_{j+1}} - W_{t_j})\,(W_{t_{k+1}} - W_{t_k})] \tag{7.5}$$

Both expectations can be computed by conditioning appropriately. (7.4) becomes

$$\begin{aligned} E\Big[\,E\Big[X_{t_k}^2\,(W_{t_{k+1}} - W_{t_k})^2\Big|\,\mathcal{F}_{t_k}\Big]\Big] &= E\Big[\,X_{t_k}^2\,E\Big[(W_{t_{k+1}} - W_{t_k})^2\Big|\,\mathcal{F}_{t_k}\Big]\Big] \\ &= E\,[\,X_{t_k}^2(t_{k+1} - t_k)]\,. \end{aligned}$$

To compute (7.5) note that $j < k$ hence $j+1 \le k$ and $t_{j+1} \le t_k$, which implies that $\mathcal{F}_{t_{j+1}} \subseteq \mathcal{F}_{t_k}$. Hence the expectation in (7.5) becomes

$$\begin{aligned} &E\,[\,E\,[X_{t_j}X_{t_k}\,(W_{t_{j+1}} - W_{t_j})\,(W_{t_{k+1}} - W_{t_k})|\,\mathcal{F}_{t_k}]\,] \\ &\qquad = E\,[\,X_{t_j}X_{t_k}\,(W_{t_{j+1}} - W_{t_j})\,E\,[\,(W_{t_{k+1}} - W_{t_k})|\,\mathcal{F}_{t_k}]\,] = 0, \end{aligned}$$

the last equation following since $E\,[\,(W_{t_{k+1}} - W_{t_k})|\,\mathcal{F}_{t_k}] = 0$. Thus we have

$$EI(X)^2 = \sum_{k=0}^{n-1} E\,[X_{t_k}^2(t_{k+1} - t_k)] = E\left[\sum_{k=0}^{n-1} X_{t_k}^2(t_{k+1} - t_k)\right] = E\int_0^t X_s^2 ds \tag{7.6}$$

since X_t is a simple process. ♠

Proposition 4 *Suppose $X \in H^2$. There exists a random variable $I(X) \in L^2$, unique up to a null set, such that $I(X_n) \to I(X)$.*

Proof: Let $\{X_n\}$ be a sequence in S^2 such that $X_n \to X$. Then $\|X_n - X_m\| \to 0$ (Cauchy sequence in H^2). From the previous proposition

$$\|I(X_m) - I(X_n)\| = \|I(X_m - X_n)\| = \|X_n - X_m\| \to 0$$

Hence $\{I(X_n)\}$ is a Cauchy sequence in L^2 and, in view of the completeness of L^2, there exists a random variable $I(X) \in L^2$ such that $I(X_n) \to I(X)$. We have thus been able to define the stochastic integral $I(X)$ for arbitrary integrants $X_t \in H^2$ (not necessarily simple processes) by means of approximating them by sequences of simple processes. This definition however would not be satisfactory if the resulting limit $I(X)$ depended on the approximating sequence. In other words it is essential to establish uniqueness, i.e. to show that any other sequence of simple processes would lead to the same result. Suppose that $\{X'_n\}$ is another S^2 sequence such that $X'_n \to X$. Then

$$\|X_n - X'_n\| \le \|X_n - X\| + \|X - X'_n\| \to 0$$

where we have used the triangle inequality and the fact that both X_n and X'_n converge in H^2 to X. However, using the linearity of the integral $I(X)$ for simple functions and the isometry $\|I(Y)\| = \|Y\|$ we have established for any simple process we have

$$\|I(X_n) - I(X'_n)\| = \|I(X_n - X'_n)\| = \|X_n - X'_n\| \to 0$$

Thus,

$$\|I(X'_n) - I(X)\| \le \|I(X'_n) - I(X_n)\| + \|I(X_n) - I(X)\| \to 0\,.$$

This establishes the uniqueness of the stochastic integral $I(X)$ since it shows that $I(X'_n) \to I(X)$ in L^2. Finally, in order to show that $EI(X) = 0$ and $\|I(X)\| = \|X\|$ note that, for any sequence of random variables $\xi_n \to \xi$ in L^2, $E\xi_n \to E\xi$ and $\|\xi_n\| \to \|\xi\|$ as a consequence of the Dominated Convergence Theorem. ♠

Note that, up to this point, we have defined the Ito integral of a process X in H^2 for each t. We have not however defined $I_t(X)$ as a function of t, i.e. as a function of the upper limit of integration. This will be done presently. Let us see first an example.

We will compute explicitly $I_t(W) = \int_0^t W_s dW_s$. Since $W_t \in \mathcal{F}_t$ by assumption the integrand is an adapted process. Also $E\int_0^t W_s^2 ds = \int_0^t EW_s^2 = \int_0^t s ds = t^2/2 < \infty$, thus $W \in H^2$. Fix $t > 0$ and consider the simple functions $\{X_n\}$ defined by

$$X_n(s) = W\left(tk2^{-n}\right) \qquad \text{for } s \in \left[\tfrac{kt}{2^n}, \tfrac{(k+1)t}{2^n}\right), \qquad k = 0, 1, 2, \ldots, 2^n - 1.$$

It is easy to see that $\{X_n\}$ is a sequence of adapted processes in S^2. Also

$$\begin{aligned}\|W - X_n\| &= E\int_0^t (W_s - X_n(s))^2 ds = \int_0^t E(W_s - X_n(s))^2 ds \\ &= \sum_{k=0}^{2^n-1} \int_0^{t/2^n} E\left(W\left(tk2^{-n} + s\right) - W\left(tk2^{-n}\right)\right)^2 ds = \sum_{k=0}^{2^n-1} \int_0^{t/2^n} s ds \\ &= 2^n \cdot \frac{1}{2}\left(\frac{t}{2^n}\right)^2 = \frac{t^2}{2^{n+1}}\end{aligned}$$

Thus

$$\|W - X_n\| = \frac{t}{2^{\frac{n+1}{2}}} \to 0$$

which implies $I_t(X_n) \to I_t(W)$.

Write for simplicity $t_k = \frac{kt}{2^n}$.

$$\begin{aligned}I_t(X_n) &= \sum_{k=0}^{2^n-1} W(t_k)\left[W(t_{k+1}) - W(t_k)\right] \\ &= \frac{1}{2}\sum_{k=0}^{2^n-1} \left[W^2(t_{k+1}) - W^2(t_k)\right] - \left[W^2(t_{k+1}) + W^2(t_k) - 2W(t_k)W(t_{k+1})\right] \\ &= \frac{1}{2}\sum_{k=0}^{2^n} \left[W^2(t_{k+1}) - W^2(t_k)\right] - \frac{1}{2}\sum_{k=0}^{2^n-1} \left[W(t_{k+1}) - W(t_k)\right]^2 \\ &= \frac{1}{2}W^2(t) - \frac{1}{2}\sum_{k=0}^{2^n-1} \left[W(t_{k+1}) - W(t_k)\right]^2\end{aligned}$$

However the last term in the above string of equations is the quadratic variation of the brownian motion and it converges (in L^2) to t:

$$\sum_{k=0}^{2^n-1} \left[W(t_{k+1}) - W(t_k)\right]^2 \xrightarrow{L^2} t$$

Consequently

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t$$

https://doi.org/10.26219/heal.aueb.5740

The above explicit computation can be repeated for other type of integrands. It is akin to the evaluation of integrals in ordinary calculus via approximating sequences. In practice stochastic integrals are most often evaluated via the Ito formula (which is essentially the stochastic counterpart of the "change–of–variables" formula of ordinary calculus).

7.4 The Ito Formula

Suppose that X_s, Y_s are adapted processes in H^2 and Z_t is an *Ito process*, i.e. a process expressed as

$$Z_t = Z_0 + \int_0^t X_s dW_s + \int_0^t Y_s ds. \tag{7.7}$$

The above is also often expressed in shorthand differential form (even though it only makes symbolic sense) as

$$dZ_t = X_t dW_t + Y_t dt. \tag{7.8}$$

Theorem 32 *[Ito formula] Suppose $f : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and Z is given by (7.7). Then*

$$f(Z_t) = f(Z_0) + \int_0^t f'(Z_s) X_s dW_s + \int_0^t f'(Z_s) Y_s ds + \frac{1}{2}\int_0^t f''(Z_s) X_s^2 ds$$

In order to evaluate stochastic integrals of the form $\int_0^t f(W_s)dW_s$ we can apply the above formula with $X_s = 1$, $Y_s = 0$ which gives $Z_t = W_t$, and $F(t) = F(0) + \int_0^t f(s)ds$. Of course, f must be continuously differentiable in order for F'' to exist and be continuous. Then the Ito formula gives

$$F(W_t) = F(0) + \int_0^t f(W_s)dW_s + \frac{1}{2}\int_0^t f'(W_s)ds \tag{7.9}$$

For instance, suppose we wanted to evaluate $\int_0^t W_s dW_s$. Take $f(x) = \frac{1}{2}x^2$ and apply the above formula to obtain

$$\frac{1}{2}W_t^2 = \int_0^t W_s dW_s + \frac{1}{2}\int_0^t 1 ds$$

which gives the result we had obtained in the previous section.

Similarly, to compute $\int_0^t e^{W_s} dW_s$ take $f(x) = e^x$ to obtain

$$e^{W_t} = 1 + \int_0^t e^{W_s} dW_s + \frac{1}{2}\int_0^t e^{W_s} ds$$

whence we obtain

$$\int_0^t e^{W_s} dW_s = e^{W_t} - 1 - \frac{1}{2}\int_0^t e^{W_s} ds.$$

Note that the integral appearing on the right hand side of the above equation is an ordinary integral.

We now proceed to give the proof of (7.9).

Proof of (7.9) We shall establish this special case of the Ito formula under the additional assumption that $\int_0^t E[f(W_s)]ds < \infty$. The integral $\int_0^t f(W_s)dWs$ makes sense as an Ito integral since the process

$f(W_s)$ is adapted to $\mathcal{F}_s$ and $E\int_0^t f(W_s)dW_s = \int_0^t E[f(W_s)]ds < \infty$. Then if we set $t_k^{(n)} := \frac{kt}{2^n}$ for $k = 0, 1, 2, \ldots, 2^n - 1$ and we define an approximating sequence of simple processes via

$$X_n(s) = f(W(t_k^{(n)})) \qquad \text{when } s \in \left[t_k^{(n)}, t_{k+1}^{(n)}\right), \qquad k = 1, 2, \ldots, 2^n - 1.$$

In the sequel we shall suppress the dependence of $t_k^{(n)}$ on n and write simply t_k. The simple processes X_n are obviously adapted and belong to S^2, hence we can define their Ito integrals as

$$I(X_n) = \sum_{k=0}^{2^n-1} f(W(t_k))\,(W(t_{k+1}) - W(t_k))\,.$$

Let us use now Taylor's theorem for the function F (remember that $F' = f$) to obtain

$$F(W(t_{k+1})) - F(W(t_k)) = f(W(t_k))(W(t_{k+1}) - W(t_k)) + \frac{1}{2}f'(\xi_k)(W(t_{k+1}) - W(t_k))^2$$

where ξ_k is between $W(t_k)$ and $W(t_{k+1})$. Thus

$$\begin{aligned} I(X_n) &= \sum_{k=0}^{2^n-1} F(W(t_{k+1})) - F(W(t_k)) \;-\; \frac{1}{2}\sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1}) - W(t_k))^2 \\ &= F(W(t)) - F(0) - \frac{1}{2}\sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1}) - W(t_k))^2 \end{aligned}$$

since the first sum is telescopic. It remains to examine the limit of the second sum as $n \to \infty$. In fact we will show that it converges in L^2 to $\frac{1}{2}\int_0^t f'(W_s)ds$. To simplify the notation set

$$\begin{aligned} \Phi_n &= \sum_{k=0}^{2^n-1} f'(\xi_k)(W(t_{k+1}) - W(t_k))^2, \\ \Phi &= \int_0^t f'(W_s)ds, \\ \Psi_n &= \sum_{k=0}^{2^n-1} f'(\xi_k)(t_{k+1} - t_k). \end{aligned}$$

To show that $\Phi_n \xrightarrow{L^2} \Phi$ we must establish that $\|\Phi_n - \Phi\| \to 0$. Using the triangle inequality

$$\|\Phi_n - \Phi\| \le \|\Phi_n - \Psi_n\| + \|\Psi_n - \Phi\|.$$

Now we have

$$\|\Phi_n - \Psi_n\|^2 = E\left(\sum_{k=0}^{2^n-1} f'(\xi_k)\,\left[(W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k)\right]\right)^2$$

To ease the notation define

$$\delta_{k,n} := \left[(W(t_{k+1}) - W(t_k))^2 - (t_{k+1} - t_k)\right] \tag{7.10}$$

We have used the same quantities before, namely when we were trying to compute the quadratic variation of Brownian motion. There we had seen that

$$E\delta_{k,n} = t2^{-n} \qquad \text{and} \qquad E\delta_{k,n}^2 = 3t^2 2^{-2n}. \tag{7.11}$$

 https://doi.org/10.26219/heal.aueb.5740

Thus

$$\|\Phi_n - \Psi_n\|^2 = E\left(\sum_{k=0}^{2^n-1} f'(\xi_k)^2\delta_{k,n}^2\right) + E\left(\sum_{k=1}^{2^n-1}\sum_{l=0}^{k-1} f'(\xi_k)f'(\xi_l)\delta_{k,n}\delta_{l,n}\right)$$

At this point we note that the second expectation on the right hand side of the above equation vanishes. Also, since f' is continuous on $[0, t]$, and therefore bounded on this interval, say by M, we obtain the inequality

$$\|\Phi_n - \Psi_n\|^2 \leq \sum_{k=0}^{2^n-1} ME\delta_{k,n}^2 \leq 3t^2 2^{-2n} M 2^n = 3t^2 M 2^{-n} \to 0,$$

the second inequality following from (7.11). This completes the proof. ♠

7.5 A more general Ito formula

Often in applications the following, more general Ito formula is useful. Suppose that we are given a function $F : \mathbf{R}^2 \to \mathbf{R}$ continuously differentiable with respect of its first argument, t, and twice continuously differentiable with respect to its second argument, z. If, as before, Z_t is an Ito process, i.e.

$$dZ_t = X_t dW_t + Y_t dt,$$

then the following change of variables holds

$$\begin{aligned} dF(t, Z_t) &= \frac{\partial F}{\partial t}dt + \frac{\partial F}{\partial z}dZ_t + \frac{1}{2}\frac{\partial^2 F}{\partial z^2}(dZ_t)^2 \\ &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial z}Y_t + \frac{1}{2}\frac{\partial^2 F}{\partial z^2}X_t^2\right)dt + \frac{\partial F}{\partial z}X_t dW_t. \end{aligned} \qquad (7.12)$$

In integral form this can be written as

$$F(t, Z_t) - F(0, Z_0) = \int_0^t \left(\frac{\partial F}{\partial s} + \frac{\partial F}{\partial z}Y_s + \frac{1}{2}\frac{\partial^2 F}{\partial z^2}X_s^2\right)ds + \int_0^t \frac{\partial F}{\partial z}X_s dW_s.$$

7.6 Multidimensional version of Ito' s formula

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function and X_t and Ito process described by the equation

$$dX_t = u_t dt + v_t dW_t$$

where

$$u_t := \begin{pmatrix} u_t^1 \\ u_t^2 \\ \vdots \\ u_t^n \end{pmatrix}, \qquad v_t := \begin{bmatrix} v_t^{11} & v_t^{12} & \cdots & v_t^{1m} \\ v_t^{21} & v_t^{22} & \cdots & v_t^{2m} \\ \vdots & \vdots & & \vdots \\ v_t^{n1} & v_t^{n2} & \cdots & v_t^{nm} \end{bmatrix}, \qquad W_t := \begin{pmatrix} W_t^1 \\ W_t^2 \\ \vdots \\ W_t^m \end{pmatrix}$$

It is assumed that the processes u_t^i, v_t^{ij} are adapted and that W_t^i, $i = 1, 2, \ldots, m$ are independent standard brownian motions. The Ito formula is written symbolically as

$$df(X_t) = \nabla f dX_t + \frac{1}{2}dX_t^T H dX_t \qquad (7.13)$$

where

$$H := \begin{bmatrix} D_{11}f(X_t) & D_{12}f(X_t) & \cdots & D_{1n}f(X_t) \\ D_{21}f(X_t) & D_{22}f(X_t) & \cdots & D_{2n}f(X_t) \\ \vdots & \vdots & & \vdots \\ D_{n1}f(X_t) & D_{n2}f(X_t) & \cdots & D_{nn}f(X_t) \end{bmatrix}.$$

We thus have

$$\begin{aligned} dX_t^T H dX_t &= dW_t^T V^T H V dW_t = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{m} dW_t^i V_{ij}^T H_{jk} V_{kl} dW_t^l \\ &= dt \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} V_{ij}^T H_{jk} V_{ki} = dt \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} V_{ji} H_{jk} V_{ki} \end{aligned}$$

where in the above string of equalities we have taken into account that

$$dW_t^i dW_t^j = \delta_{ij} dt.$$

(δ_{ij}, called Kronecker's delta, is defined to be equal to 1 if $i = j$ and zero otherwise.) We can thus write (7.13) in more detailed form as

$$\begin{aligned} df(X_t) &= \sum_{i=1}^{n} D_i f(X_t) u_t^i dt + \sum_{i=1}^{n} D_i f(X_t) \sum_{j=1}^{m} v_t^{ij} dW_t^j \\ &\quad + \frac{1}{2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} v_t^{ji} D_{jk} f(X_t) v_t^{ki}\right) dt \end{aligned} \tag{7.14}$$

https://doi.org/10.26219/heal.aueb.5740

Chapter 8

Stochastic Difference Equations

We begin with the simplest stochastic difference equation. Suppose that $\{\xi_n\}$ is a sequence of independent random variables with common distribution $\mathcal{N}(0,\sigma^2)$ and define recursively the following sequence of random variables

$$X_{n+1} = \alpha X_n + \beta \xi_n, \qquad n = 0, 1, 2, \dots \tag{8.1}$$

where $X_0 \backsim \mathcal{N}(0,\sigma_0^2)$ and is assumed to be independent of the $\{\xi_n\}$. Then by repeated iteration of (8.1)

$$X_n = \alpha^n X_0 + \beta \sum_{k=1}^{n} \alpha^{n-k} \xi_{k-1} \tag{8.2}$$

In view of the independence assumptions we see that

$$\mathsf{Var}(X_n) = \alpha^{2n}\sigma_0^2 + \beta^2 \sum_{k=1}^{n} \alpha^{2(n-k)}$$

and, summing the geometric progression we obtain

$$X_n \backsim \mathcal{N}\left(0, \alpha^{2n}\sigma_0^2 + \beta^2 \frac{1-\alpha^{2n}}{1-\alpha^2}\right).$$

If $|\alpha| < 1$ then, as $n \to \infty$, $X_n \xrightarrow{d} \mathcal{N}\left(0, \frac{\beta^2}{1-\alpha^2}\right)$. If we choose $X_0 \backsim \mathcal{N}\left(0, \frac{\beta^2}{1-\alpha^2}\right)$, then $\{X_n\}$ is a stationary sequence. It is indeed easy to check that

$$\mathsf{Var}(X_n) = \alpha^{2n} \frac{\beta^2}{1-\alpha^2} + \beta^2 \frac{1-\alpha^{2n}}{1-\alpha^2} = \frac{\beta^2}{1-\alpha^2}$$

for all n. Since

$$X_{n+m} = \alpha^m X_n + \beta \sum_{l=1}^{m} \alpha^{m-l} \xi_{n+l-1},$$

we see that $E[X_{n+m}X_n] = E[X_n^2]\alpha^m$ and hence, for all n, m,

$$\mathsf{Cov}(X_n, X_{n+m}) = \frac{\beta^2}{1-\alpha^2}\alpha^m.$$

Consider now the more general linear recursion or order k

$$X_{n+k} = \alpha_1 X_{n+k-1} + \alpha_2 X_{n+k-2} + \cdots + \alpha_k X_n + \beta \xi_n$$

where $\{\xi_n\}$ is an i.i.d. sequence of normal random variables as before and $(X_0, X_1, \ldots, X_{k-1})$ is a zero mean normal random vector, independent of the sequence $\{\xi_n\}$ with given covariance matrix V_0. This can be put into vector-matrix form as follows

$$\begin{pmatrix} X_{n+1} \\ X_{n+2} \\ \vdots \\ X_{n+k-1} \\ X_{n+k} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & & 1 \\ \alpha_k & \alpha_{k-1} & \alpha_{k-2} & \cdots & \alpha_1 \end{pmatrix} \begin{pmatrix} X_n \\ X_{n+1} \\ \vdots \\ X_{n+k-2} \\ X_{n+k-1} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \beta \end{pmatrix} \xi_n.$$

Instead of studying the above stochastic difference equation we rather turn our attention to the more general case

$$X_{n+1} = AX_n + B\xi_n, \qquad n = 0, 1, 2, \ldots,$$

where X_n is a random vector in $\mathbf{R}^k$, A is a $k \times k$ matrix, B is a $k \times l$ matrix and $\{\xi_n\}$ is an i.i.d. sequence with distribution $\mathcal{N}(0, I_l)$ where I_l is the l–dimensional identity matrix. Also, $X_0 \backsim \mathcal{N}(0, V_0)$ where V_0 is a given $k \times k$ non-negative definite matrix. As in the scalar case,

$$X_n = B\xi_{n-1} + AB\xi_{n-2} + A^2 B\xi_{n-3} + \cdots + A^{n-1} B\xi_0 + A^n X_0$$

and hence the covariance matrix for X_n, $V_n := EX_n X_n^T$, is given by

$$V_n = BB^T + ABB^T A + A^2 BB^T (A^2)^T + \cdots + A^{n-1} BB^T (A^{n-1})^T + A^n V_0 (A^n)^T.$$

If $A^n \to 0$ as $n \to \infty$ (which happens if and only if the largest eigenvalue of A has modulus less than 1) then

$$V = \lim_{n \to \infty} V_n = \sum_{k=0}^{\infty} A^k BB^T (A^k)^T.$$

Since

$$\begin{aligned} V - BB^T &= ABB^T A^T + A^2 BB^T (A^2)^T + A^3 BB^T (A^3)^T + \cdots \\ &= A \left(BB^T + ABB^T A^T + A^2 BB^T (A^2)^T + \cdots \right) A^T = AVA^T \end{aligned}$$

and thus

$$V = BB^T + AVA^T. \tag{8.3}$$

If $X_0 \backsim \mathcal{N}(0, V)$ where V is given by the solution of (8.3) then $\{X_n\}$ is a stationary Gaussian process with zero mean and $EX_n X_n^T = V$ for all n, $EX_{n+m} X_n^T = A^m V$.

https://doi.org/10.26219/heal.aueb.5740

Chapter 9

Stochastic Differential Equations

9.1 Introduction

An ordinary stochastic differential equation (of the first order) is an equation of the form

$$\frac{dx}{dt} = f(x,t), \qquad x(0) = x_0. \tag{9.1}$$

We are given a function $f : \mathbb{R}^2 \to \mathbb{R}$ and an *initial condition*, x_0 and we must find a function $x(t)$ that satisfies (9.1). For instance, if $f(x,t) = \alpha x$, $(\alpha \in \mathbb{R})$ then (9.1) becomes

$$\frac{dx}{dt} = \alpha x, \qquad x(0) = x_0 \tag{9.2}$$

which has the unique solution

$$x(t) = x_0 e^{\alpha t}. \tag{9.3}$$

Similarly, if $f(x,t) = \alpha(t)x$, where $\alpha(t)$ is a piecewise continuous function then

$$\frac{dx}{dt} = \alpha(t)x, \qquad x(0) = x_0. \tag{9.4}$$

The solution of this equation can be obtained by multiplying both sides with $e^{-\int_0^t \alpha(s)ds}$ as follows:

$$e^{-\int_0^t \alpha(s)ds}\frac{dx}{dt} - x(t)\alpha(t)e^{-\int_0^t \alpha(s)ds} = 0$$

or

$$\frac{d}{dt}\left(e^{-\int_0^t \alpha(s)ds}\right) = 0$$

which gives

$$x(t)e^{-\int_0^t \alpha(s)ds} = c, \qquad \text{for all } t \geq 0$$

where c is an unspecified constant. Setting $t = 0$ above we see that $x(0) = c = x_0$ and thus the solution of (9.4) is

$$x(t) = x_0 e^{\int_0^t \alpha(s)ds}.$$

 https://doi.org/10.26219/heal.aueb.5740

Consider now the differential equation

$$\frac{dx}{dt} = \alpha x(t) + u(t) \tag{9.5}$$

where $u(t)$ is a continuous function of time. Multiplying both sides of (9.5) by $e^{-\alpha t}$ we obtain

$$e^{-\alpha t}\frac{dx}{dt} - e^{-\alpha t}\alpha x(t) = e^{-\alpha t}u(t)$$

or

$$\frac{d}{dt}\left(x(t)e^{-\alpha t}\right) = e^{-\alpha t}u(t).$$

Integrating the above equation we obtain

$$x(t)e^{-\alpha t} = c + \int_0^t e^{-\alpha s}u(s)ds$$

and setting $t = 0$ in we have $x(0) = c = x_0$, thus

$$x(t) = x_0 e^{\alpha t} + \int_0^t e^{\alpha(t-s)}u(s)ds.$$

9.2 Stochastic Differential Equations

In the same spirit we would like to study the differential equation

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t)\frac{dW_t}{dt} \tag{9.6}$$

where X_0 is given and W_t is standard brownian motion. The problem we are faced with here is that W_t has paths that are nondifferentiable, hence the way this equation is written makes no sense. This can be circumvented if we pose the problem in its integral form

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{9.7}$$

It is customary to write the above equation in differential form as follows:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t$$

9.3 Ornstein–Uhlenbeck Equation

$$dX_t = -\alpha X_t dt + \beta dW_t \tag{9.8}$$

(assume $a > 0$). Integral form

$$X_t = X_0 + \int_0^t \alpha X_s ds + \beta W_t$$

Solution

$$X_t = X_0 e^{-\alpha t} + \int_0^t \beta e^{-\alpha(t-s)}dW_s \tag{9.9}$$

 https://doi.org/10.26219/heal.aueb.5740

$$EX_t = EX_0 e^{-\alpha t}$$

$$\begin{aligned}
\mathrm{Var}(X_t) &= \mathrm{Var}(X_0)e^{-2\alpha t} + \int_0^t \beta^2 e^{-2\alpha(t-s)} ds \\
&= \mathrm{Var}(X_0)e^{-2\alpha t} + \beta^2 e^{-2\alpha t} \int_0^t e^{2\alpha s} ds \\
&= \mathrm{Var}(X_0)e^{-2\alpha t} + \frac{\beta^2}{2\alpha} e^{-2\alpha t} \left(e^{2\alpha t} - 1\right)
\end{aligned}$$

If $\mathrm{Var}(X_0) = \frac{\beta^2}{2\alpha}$ then $\mathrm{Var}(X_t) = \frac{\beta^2}{2\alpha}$.

ftbpFU3.6244in1.8697in0ptSample path of the Ornstein-Uhlenbeck process.ou-02.bmp

9.4 The Brownian Bridge Process

Suppose that W_t, $t \geq 0$, is standard brownian motion. The brownian bridge process X_t, $t \in [0,1]$, is defined (at least informally for the time being) as the process B_t, *conditional on the event* $W_1 = 0$. Based on this informal idea we can easily compute the joint distribution of the process X_t as follows. Thus, for $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$ we have that

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \stackrel{\mathrm{d}}{=} (W_{t_1}, W_{t_2}, \ldots, W_{t_n}) | W_1 = 0.$$

The right hand side of the above equation is the conditional distribution of a Gaussian vector given one of its components, hence it is again Gaussian. ftbpFU3.6253in1.8697in0ptSample path of a brownian bridgebridge.bmp

Two useful characterizations of the brownian bridge process are the following.

$$X_t = W_t - tW_1, \qquad 0 \leq t \leq 1 \tag{9.10}$$

and

$$dX_t = -\frac{1}{1-t} X_t dt + dW_t, \qquad 0 \leq t < 1, \qquad X_0 = 0. \tag{9.11}$$

It is easy to check that the process described in (9.10) is a Gaussian process which has the right mean and covariance function and therefore that it is indeed the standard brownian bridge. Regarding the SDE of (9.11) note that it can be solved through the use of an integrating factor: Indeed, $\int \frac{dt}{1-t} = -\log(1-t)$ and $e^{\int \frac{dt}{1-t}} = \frac{1}{1-t}$. Thus, multiplying the equation (9.11) by $\frac{1}{1-t}$ we obtain

$$\frac{1}{1-t} dX_t + \frac{1}{(1-t)^2} X_t dt = \frac{dW_t}{1-t}$$

or

$$d\left(\frac{1}{1-t} X_t\right) = \frac{dW_t}{1-t}$$

where we have used Ito's rule, hence

$$X_t = (1-t) \int_0^t \frac{dW_s}{1-s}.$$

In the above derivation we also took into account the initial condition $X_0 = 0$. If the process starts at time $s \in (0,1)$ at the point X_s, then the solution would have been

$$X_t = (1-t)\left(\frac{X_s}{1-s} + \int_s^t \frac{dW_u}{1-u}\right) \tag{9.12}$$

To see that this is again the standard brownian bridge it is enough to note that it is a Gaussian process with zero mean and covariance function given, when $s < t$, by (9.12) we have

$$EX_sX_t = E\left[X_s^2\frac{1-t}{1-s}\right] + E\left[X_s(1-t)\int_s^t \frac{dW_u}{1-u}\right].$$

The second integral on the right hand side of the above equation is zero by the martingale property of stochastic integrals. Thus, in view of the fact that

$$\begin{aligned} E\left[X_s^2\right] &= (1-s)^2 E\left(\int_0^s \frac{dW_u}{1-u}\right)^2 = (1-s)^2\int_0^s \frac{du}{(1-u)^2} \\ &= (1-s)^2\left(\frac{1}{1-s} - 1\right) = s(1-s), \end{aligned}$$

gives

$$EX_sX_t = s(1-t)$$

which is the correct expression for the covariance. (Note that in the above derivation we have also made use of the Ito isometry.)

9.5 Geometric Brownian Motion

Consider the following stochastic differential equation

$$dX_t = rX_tdt + \alpha X_tdW_t \tag{9.13}$$

where r and α are real constants. A solution can be obtained by applying Ito's rule on the function $\log X_t$ to obtain

$$d\log X_t = \frac{dX_t}{X_t} - \frac{1}{2X_t^2}\alpha^2X_t^2dt = \left(r - \frac{1}{2}\alpha^2\right)dt + adW_t$$

or

$$\log X_t - \log X_0 = \left(r - \frac{1}{2}\alpha^2\right)t + aW_t.$$

Hence

$$X_t = X_0e^{\left(r-\frac{1}{2}\alpha^2\right)t+\alpha W_t}. \tag{9.14}$$

Let us first consider the special case $\alpha = 1$, $r = \frac{1}{2}$, $X_0 = 1$ w.p.1. In this case we have the stochastic differential equation $dX_t = \frac{1}{2}X_tdt + X_tdW_t$ whose solution is given by

$$X_t = e^{W_t}.$$

This stochastic process is known as *geometric brownian motion.* Note that the distribution of X_t is lognormal, namely $P(X_t \leq x) = P(W_t \leq \log x)$. Differentiating the above relationship with respect to x we obtain the density $f_t(x)$ of X_t as

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} x^{-1} e^{-\frac{1}{2t}(\log x)^2}, \qquad x \geq 0.$$

In particular, the moments of X_t can be obtained easily from the moment generating function of the normal distribution, $EX_t^n = Ee^{nW_t} = e^{\frac{1}{2}n^2 t}$. For $n = 1, 2$ we obtain the first two moments,

$$EX_t = e^{\frac{1}{2}t} \qquad \text{and} \qquad \mathsf{Var}(X_t) = e^{2t} - e^t.$$

9.6 The Samuelson model for stock prices

The first successful effort to develop a mathematical model for the fluctuations of stock prices is the thesis of Bachelier (1900) who for this purpose was the first to give a description of Brownian motion. Bachelier's work was largely ignored by his contemporaries and in fact the mathematical description of Brownian motion was discovered independently by Smoluchowski (1905) and Einstein (1905). The first "modern" effort to describe stock prices using Brownian motion is the Samuelson model (1965) according to which stock prices follow a geometric Brownian motion process, i.e. are given by

$$S_t = S_0 e^{\alpha t + \beta B_t}, \qquad t \geq 0,$$

where S_0 is the price of the stock at time $t = 0$ and W_t is standard brownian motion. We point out that, as a result of the Ito rule

$$dS_t = S_0 \left(e^{\alpha t + \beta B_t} \alpha dt + e^{\alpha t + \beta B_t} \beta dB_t + \frac{1}{2} e^{\alpha t + \beta B_t} \beta^2 dt \right)$$

and hence that S_t satisfies the stochastic differential equation

$$S_t = \left(\alpha + \frac{1}{2}\beta^2 \right) S_t dt + \beta S_t dB_t.$$

If we denote the *drift* by $\mu := \alpha + \frac{1}{2}\beta^2$ and the *volatility* by $\sigma := \beta$ we can describe the Samuelson model in terms of the SDE

$$S_t = \mu S_t dt + \sigma S_t dB_t \tag{9.15}$$

which has the solution

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}, \qquad t \geq 0.$$

https://doi.org/10.26219/heal.aueb.5740

https://doi.org/10.26219/heal.aueb.5740

Bibliography

[1] Cox, D.R. and H.D. Miller. (1965) *The Theory of Stochastic Processes.* Chapman and Hall.

[2] Feller, W. (1968) *An Introduction to Probability Theory and its Applications.* Vol. I, 3rd edition, John Wiley.

[3] Karlin, S. and H.M. Taylor. (1975) *A First Course in Stochastic Processes.* 2nd edition. Academic Press.

[4] Lawer, G.F. (1995) *Introduction to Stochastic Processes.* Chapman and Hall.

[5] Norris, J.R. (1997) *Markov Chains.* Cambridge University Press.

[6] Ross, S.M. (1997) *Stochastic Models.* 7th Edition. Academic Press.

[7] Ross, S.M. (1996) *Stochastic Processes.* 2th Edition. John Wiley.

https://doi.org/10.26219/heal.aueb.5740