Background notes to course 'Stochastic Processes'
Spring 2013

always under construction as well

Abstract

Here we provide some background material for the Lecture Notes.

1 Review of definitions

Recall the following definitions: let a space E be given. A collection $\mathcal{A}$ of subsets of E is called a an *algebra* (or a *field*) if the following three conditions hold:

i) $E \in \mathcal{A}$;

ii) $A \in \mathcal{A} \Rightarrow A^c = E \setminus A \in \mathcal{A}$;

iii) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

A collection $\mathcal{A}$ is called a σ*-algebra* on E if $\mathcal{A}$ is an algebra, such that $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{A}$. Hence, an algebra is closed under taking finite unions and finite intersections, whereas a σ-algebra is closed under taking countable unions and intersections. The pair $(E, \mathcal{A})$ is called a *measurable space*, if $\mathcal{A}$ is a σ-algebra on E. Contrast this with the definition of topology.

Definition 1.1 A collection $\mathcal{T}$ of so-called open subsets of the space E is called a topology if

i) $\emptyset, E \in \mathcal{T}$;

ii) the intersection of finitely many members of $\mathcal{T}$ belongs to $\mathcal{T}$: $A_1, \ldots, A_n \in \mathcal{T}$ implies $\cap_{k=1}^{n} A_k \in \mathcal{T}$;

iii) the union of arbitrarily many members belongs to $\mathcal{T}$: $A_\alpha \in \mathcal{T}$, $\alpha \in B$, then $\cup_\alpha A_\alpha \in \mathcal{T}$.

We call E a topological space.

Let $\mathcal{C}$ be a collection of subsets of E. By $\sigma(\mathcal{C})$, the σ-algebra generated by $\mathcal{C}$, we understand the smallest σ-algebra that contains $\mathcal{C}$. The Borel-σ-algebra on the topological space E is the smallest σ-algebra that contains the open sets of E. A set $A \subset E$ is called a G_δ-set if it is a countable intersection of open sets in E. It is an F_σ-set if it is a countable union of closed (i.e. complements of open sets) sets in E.

If E is a metric space, equipped say with metric ρ, then the Borel-σ-algebra $\mathcal{B}(E)$ is the σ-algebra generated by the open sets induced by the metric ρ. A set $A \subset E$ is *open* if for each $x \in A$

there exists r_x such that $B_{r_x}(x) = \{y \mid \rho(x,y) < r_x\}$ is contained in A. If, in addition, the metric space E is *separable* (i.e. there is a countably dense subset), then $\mathcal{B}(E) = \sigma(B_q(x), q \in \mathsf{Q}, x \in$ countably dense subset of $E)$.

We use $\mathcal{B}$ to denote the Borel sets in $\mathbf{R}$.

Many statements concerning σ-algebras can be reduced to statements on certain generic collections of sets: a main one is the notion of π-system. A collection $\mathcal{A}$ of subsets of E is called a *π-system* if $\mathcal{A}$ is invariant under finite intersections, i.e. $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$.

Desirable properties of π-systems are the following (see Williams, PwM).

Lemma 1.2 i) *Let $\mathcal{I}$ be a π-system on E generating the σ-algebra $\mathcal{A}$. Suppose that μ_1, μ_2 are measures on $(E, \sigma(\mathcal{I}))$, such that $\mu_1(E) = \mu_2(E) < \infty$, and $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{I}$. Then $\mu_1(A) = \mu_2(A)$ for all $A \in (\sigma(\mathcal{I}))$;*

ii) *Let a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ be given. Let $\mathcal{I}, \mathcal{J}$ be π-systems on E with $\mathcal{A} = \sigma(\mathcal{I})$ and $\mathcal{F} = \sigma(\mathcal{J})$. Then $\mathcal{A}$ and $\mathcal{F}$ are independent if $\mathcal{I}$ and $\mathcal{J}$ are independent, that is, if $\mathsf{P}\{A \cap B\} = \mathsf{P}\{A\}\mathsf{P}\{B\}$ for all $A \in \mathcal{I}$, $B \in \mathcal{J}$.*

Suppose that we have measure spaces $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$. Then the function $X : \Omega \to E$ is called a random element if X is $\mathcal{F}/\mathcal{E}$-measurable. In other words, if $X^{-1}(A) \in \mathcal{F}$, for all $A \in \mathcal{E}$. The following lemma is helpful in checking measurability.

Lemma 1.3 *Let $(\Omega, \mathcal{F})$, $(E, \mathcal{E})$ be a measurable space. Let $\mathcal{C}$ be a collection of subsets of E, such that $\sigma(\mathcal{C}) = \mathcal{E}$. Let $X : \Omega \to E$ be a map. Then X is $\mathcal{F}/\mathcal{E}$-measurable if for $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{C}$.*

2 σ-algebra for a stochastic process

Let the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ be given, as well as the measurable space $(E, \mathcal{E})$. Let further $X = (X_t)_{t \in T}$ be a s.p. (stochastic process) with state E, for index set T. This means that X_t, $t \in T$, are all E-valued, $\mathcal{F}/\mathcal{E}$-measurable random elements. We assume that T is an interval of $\mathbf{R}$.

Given $\omega \in \Omega$, the corresponding realisation $x(\omega) = (x_t(\omega))_{t \in T}$ of the s.p. X is called a *sample path* or *trajectory* of X.

One can also view $X(\omega)$ as an E-valued function of T, i.e. there exists $f : T \to E$, such that $X(\omega) = f$, and $X_t(\omega) = f(t)$, $t \in T$. Hence, the process X takes values in the function space E^T. The question is: *how should one define a σ-algebra on E^T in some consistent way and such that we can answer questions concerning the probability of events of interest that concern the whole path?* We take the following approach.

A finite-dimensional rectangle in E^T is a set of the form

$$S = \{x \in E^T \mid x_{t_1} \in E_1, \ldots, x_{t_n} \in E_n\},$$

for $\{t_1, \ldots, t_n\} \subset I$, $E_i \in \mathcal{E}$, $i = 1, \ldots, n$, for some finite integer $n > 0$.

It is natural to view such sets as the building blocks of our σ-algebra, so we define

$$\mathcal{E}^T = \sigma\{\text{1-dimensional rectangles}\}.$$

For understanding the structure of this σ-algebra, let us recall the concept of product-σ-algebras $\mathcal{E}^n$, $n = 1, 2, \ldots, \infty$.

For n finite, this is simply the σ-algebra generated by all n-dimensional rectangles

$$\mathcal{E}^n = \sigma\{E_1 \times \cdots \times E_n \,|\, E_1, \ldots, E_n \in \mathcal{E}\}.$$

For $n = \infty$, it is the σ-algebra generated by all finite-dimensional rectangles (defined in E^∞ in a natural way):

$$\mathcal{E}^\infty = \sigma\{E_1 \times \cdots \times E_n \times E \times E \times \cdots \,|\, E_1, \ldots, E_n \in \mathcal{E}, n = 1, 2, \ldots\}.$$

Now we shall call a set $A \in \mathcal{E}^T$ a cylinder, if there exists a finite subset $\{t_1, \ldots, t_n\}$ of T and a set $B \in \mathcal{E}^n$, such that

$$A = \{x \,|\, (x_{t_1}, \ldots, x_{t_n}) \in B\}. \tag{2.1}$$

It will be called a σ-cylinder, if there exists a (at most) countable subset $\{t_1, \ldots\}$ of T and a set $B \in \mathcal{E}^\infty$, such that (2.1) holds.

We have the following characterisation.

Lemma 2.1 *$\mathcal{E}^T$ is precisely the collection of σ-cylinder sets.*

Proof. First note that the σ-cylinders are contained in the σ-algebra generated by finite-dimensional cylinders. Hence $\mathcal{E}^T$ contains all σ-cylinders. Notice that by definition $\mathcal{E}^T$ is the smallest σ-algebra containing all 1-dimensional cylinders. Hence it is the smallest σ-algebra containing all σ-cylinders. So it is sufficient to show that the σ-cylinders themselves form a σ-algebra.

First $E^T = \{x \,|\, x_t \in E\}$ for arbitrary, but fixed t. Hence it is a σ-cylinder.

Let $A = \{x \,|\, (x_{t_1}, x_{t_2}, \ldots) \in B\}$, $B \in \mathcal{E}^\infty$. Then $A^c = \{x \,|\, (x_{t_1}, x_{t_2}, \ldots) \in \mathcal{E}^\infty \setminus B\}$ is a σ-cylinder.

Finally, let $A_n = \{x \,|\, (x_{t_1^n}, x_{t_2^n}, \ldots) \in B_n\}$, $B_n \in \mathcal{E}^\infty$, $n = 1, 2, \ldots$. It is sufficient to show that the intersection is a σ-cylinder. To this end, we need to describe all A_n in terms of the same 'time' points. We therefore need to form the concatenation of the time points that define the different σ-cylinders $A_1, \ldots$. Let $S_n = \{t_k^n\}_{k=1,\ldots}$, put $S = \cup_n S_n$. S is countable.

Let $f : S \to \{1, 2, \ldots\}$ be a bijective map (representing an enumeration of the points of S). The set A_n can be defined in terms of all time points $f^{-1}(1), f^{-1}(2), \ldots$. At time points $f^{-1}(m) \notin \{t_k^n\}_k$ any path $x \in A_n$ is allowed to take any values. At time points $f^{-1}(m) \in \{t_k^n\}_k$ the values are prescribed. To formalise this, define $\tilde{B}_n \subset E^\infty$ by

$$\tilde{B}_n = \{x = (x_1, x_2, \ldots) \in E^\infty \,|\, (x_{f(t_1^n)}, x_{f(t_2^n)}, \ldots) \in B_n\}.$$

Note that $\tilde{B}_n \in \mathcal{E}^\infty$, since $B_n \in \mathcal{E}^\infty$. Then A_n can be expressed as

$$A_n = \{x \in E^T \,|\, (x_{f^{-1}(1)}, x_{f^{-1}(2)}, \ldots) \in \tilde{B}_n\}.$$

But now $\cap_n A_n = \{x \,|\, (x_{f^{-1}(1)}, x_{f^{-1}(2)}, \ldots) \in \cap_n \tilde{B}_n\}$ and is hence is a σ-cylinder.

We conclude that $\mathcal{E}^T$ must be equal to the σ-algebra of σ-cylinders. QED

Remark A stochastic process $X = (X_t)_t$, with X_t $\mathcal{F}/\mathcal{E}$-measurable, is $\mathcal{F}/\mathcal{E}^T$-measurable by construction. Hence, the induced probability measure on $(E^T, \mathcal{E}^T)$ is given by:

$$\mathsf{P}_X\{A\} = \mathsf{P}\{\omega \mid X(\omega) \in A\} = \mathsf{P}\{X^{-1}(A)\},$$

for any σ-cylinder A. This probability measure is uniquely determined by any π-system generating $\mathcal{E}^T$. A convenient π-system is given by

$$\mathcal{I} = \Big\{A \mid \exists n \in \mathbf{Z}_+, t_1 < \cdots < t_n, t_i \in T, i = 1 \ldots, n, A_1, \ldots, A_n \in \mathcal{E}, \text{ such that } A = \{x \in E^T \mid x_{t_i} \in A_i, i = 1 \ldots, n\}\Big\}.$$

Note that a random variable X_t can be obtained from X by *projection* or *co-ordinate maps*: let $S \subset T$, and let $x \in E^T$. For any $S \subseteq T$: $E^S = \{x = (x_t)_{t\in S} : x_t \in E\}$. The precise indices belonging to S do not disappear in this notation. For instance if $T = \{1,2,3,4,5\}$ and $S = \{1,5\}$, then $\mathbf{R}^S = \{x = (x_1, x_5), x_1, x_5 \in \mathbf{R}\}$. This looks artificial, but is natural in the context of *viewing the index set as the collection of observation instants of a process.*

Then the projection $\pi_S : E^T \to E^S$ is defined by

$$\pi_S(x) = \{(x_t)_{t\in S}\},$$

and $\pi_t(x) = x_t$ is simply the projection on one co-ordinate. It follows that $X_t = \pi_t X$.

Note that $\mathcal{E}^T$ is in fact defined as the smallest σ-algebra that makes all projections on one co-ordinate measurable:

$$\mathcal{E}^T = \sigma(\pi_t,\, t \in T).$$

One can view σ-algebras as the amount of information on individual states that is available to the observer: of individual states one can only observe the measurable sets to which a state belongs. By virtue of Lemma 2.1, one can only observe the values of a function in E^T in countably many time instants in T.

Corollary 2.2 *Let* $T = [0,\infty)$, $E = \mathbf{R}$, $\mathcal{E} = \mathcal{B}(\mathbf{R})$, *and* $C = \{f : T \to E \mid f \text{ continuous }\}$. *Then* $C \notin \mathcal{E}^T$.

Problem 2.1 Prove this corollary.

It is often convenient to complete a σ-algebra. Consider the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. The completion $\overline{\mathcal{F}}$ of $\mathcal{F}$ w.r.t. P is the σ-algebra generated by $\mathcal{F}$ and the sets $A \subset E$, s.t. there exists sets $A_1, A_2 \in \mathcal{F}$ with $A_1 \subset A \subset A_2$ and $\mathsf{P}\{A_2 \setminus A_1\} = 0$. Put $\mathsf{P}\{A\} = \mathsf{P}\{A_1\}$. In words, the completion of $\mathcal{F}$ is the σ-algebra obtained by adding all subsets of null sets to $\mathcal{F}$.

Problem 2.2 Show that the set of measurable real-valued functions on $T = [0,\infty)$ is not measurable, in other words, show that

$$M = \{x \in \mathbf{R}^T \mid x \text{ is } \mathcal{B}[0,\infty)/\mathcal{B} - \text{measurable}\} \notin \mathcal{B}^T.$$

Moreover, show that there is no probability measure P on $(\mathbf{R}^T, \mathcal{B}^T)$ such that $M \in \overline{\mathcal{B}^T}$, that is, such that M belongs to the completion of $\mathcal{B}^T$ w.r.t. P. Hint: consider M^c; show that no σ-cylinder can contain only measurable functions x.

σ-Algebra generated by a random variable or a stochastic process
In a similar fashion one can consider the σ-algebra generated by random variables or by a stochastic process.

Let $Y : \Omega \to E$ be an $\mathcal{F}/\mathcal{E}$-measurable random variable. Then $\mathcal{F}^Y \subset \mathcal{F}$ is the σ-algebra generated by Y. This is the minimal σ-algebra that makes Y a measurable random variable. Suppose that $\mathcal{C}$ is a π-system generating $\mathcal{E}$, then $Y^{-1}\mathcal{C}$ is a π-system generating $\mathcal{F}^Y$.

Let us consider a stochastic process $X = (X_t)_{t\in T}$ with $T = \mathbf{Z}_+$ or $T = \mathbf{R}_+$. The σ-algebra $\mathcal{F}_t^X$ is the σ-algebra generated by X upto time t. It is the minimal σ-algebra that makes the random variables X_s, $s \leq t$ measurable and so it contains each σ-algebra $\mathcal{F}_s^X$, for $s \leq t$.

How can one generate this σ-algebra? If $\mathcal{C}$ is a π-system for $\mathcal{E}$, then $\mathcal{F}_t^X = \sigma(X_s^{-1}(C),\ C \in \mathcal{C}, s \leq t\}$. Unfortunately, the collection of sets $\{X_s^{-1}(C) \,|\, C \in \mathcal{C}, s \leq t\}$ itself is not a π-system. However, the following is a π-system generating $\mathcal{F}_t^X$:

$$\mathcal{C}^X = \{X_{t_1}^{-1}(C_1) \cap \cdots \cap X_{t_n}^{-1}(C_n) \,|\, t_1 < \cdots < t_n \leq t, C_1, \ldots, C_n \in \mathcal{C}, n = 1, 2, \ldots\}.$$

This can be used in checking independence of σ-algebras, measurability issues, etc. What we see here again, is that the information contained in the minimal σ-algebra $\mathcal{F}_t^X$ concerns the behaviour of paths of X upto time t observed at at most countably many time points.

3 Measurable maps

Continuous maps between two metric spaces are always measurable with respect to the Borel-σ-algebra on these spaces.

Corollary 3.1 *Let E', E be metric spaces. Let $f : E' \to E$ be a continuous map, i.e. $f^{-1}(B)$ is open in E' for each open set $B \subset E$. Then f is $\mathcal{B}(E')/\mathcal{B}(E)$-measurable.*

Proof. The open sets of E generate $\mathcal{B}(E)$. The statement follows from Lemma 1.3. QED

In stochastic process theory right-continuous functions of the positive real line $\mathbf{R}_+$ play an important role. We want to show that such functions are measurable with respect to the Borel-σ-algebra on $\mathbf{R}_+$. We will be more specific.

Let E be a Polish space, with metric ρ say, and let $\mathcal{E}$ be the Borel-σ-algebra on E.

Lemma 3.2 *There exists a countable class $\mathcal{H}$ of continuous functions $f : E \to [0, 1]$, such that $x_n \to x$ in E iff $f(x_n) \to f(x)$ in $[0, 1]$ for all $f \in \mathcal{H}$.*

Proof. Take a countable, dense subset $y_1, y_2, \ldots$ of E. Let $f_{k,n} : E \to [0, 1]$ be a continuous function with $f_{k,n}(y) = 1$ for $\rho(y, y_k) \leq 1/n$, and $f_{k,n}(y) = 0$ for $\rho(y, y_k) \geq 2/n$. Then $\mathcal{H} = \{f_{k,n} \,|\, k, n = 1, \ldots\}$ has the desired property. QED

Lemma 3.3 *If $x : [0, \infty) \to E$ is right-continous, then x is continuous except at at most a countable collection of points.*

Proof. In view of the previous lemma, it is sufficient to show that $t \mapsto f(x_t)$ has the desired property for each $f \in \mathcal{H}$. That is, it is sufficient to consider the case $E = [0,1]$.

For $t > 0$ let

$$y_t^1 = \limsup_{s\uparrow t} x_s - \liminf_{s\uparrow t} x_s (\geq 0\,!), \quad y_t^2 = \begin{cases} |x_t - \lim_{s\uparrow t} x_s|, & \text{if } y_t^1 = 0 \\ 0, & \text{otherwise.} \end{cases}$$

If x is discontinuous at t we have either $y_t^1 > 0$ or $y_t^2 > 0$. Hence, it is sufficient to show that for any $\epsilon > 0$ the sets $A^i = \{t \,|\, y_t^i > \epsilon\}$ are at most countable (why?).

By right-continuity,

$$0 = \lim_{s\downarrow t} y_s^1 = \lim_{s\downarrow t} y_s^2, \quad t \geq 0. \tag{3.1}$$

In particular, taking $t = 0$, gives that $A^i \cap (0, \delta] = \emptyset$ for some $\delta > 0$. It follows that

$$\tau^i = \sup\{\delta > 0 \,|\, A^i \cap (0,\delta] \text{ is at most countable}\} > 0.$$

But if A^i is uncountable, we would have $\tau^i < \infty$. This implies the existence of a sequence $\tau_n^i \in A^i \cap (\tau^i, \infty)$ with $\tau_n^i \downarrow \tau^i$. Then $\limsup_{s\downarrow\tau^i} y_s^i \geq \epsilon$, contradicting (3.1). QED

Let now $X = \{X_t\}_t$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with right-continuous paths. Define for $f \in \mathcal{H}$, with $\mathcal{H}$ a convergence determining class, and y_t^i, $i = 1, 2$, defined in the above proof

$$\begin{aligned} Y_{t,f}^{(i)}(\omega) &= y_t^i(f(X_t(\omega))) \\ C_u &= \{\omega \,|\, X_t(\omega) \text{ is continuous at } t = u\} = \bigcap_{f\in\mathcal{H}} \{Y_{t,f}^{(i)} = 0, i = 1,2\}. \end{aligned}$$

It follows by right-continuity that $\liminf_{t\uparrow u} f(X_t)$, and $\limsup_{t\uparrow u} f(X_t)$ are both measurable. Hence $Y_{t,f}^{(i)}$ and C_u are as well. It makes sense to define u to be a fixed discontinuity of X if $\mathsf{P}\{C_u\} < 1$.

Corollary 3.4 *Let $X = \{X_t\}_t$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with right-continuous paths. Then X has at most countably many fixed discontinuities.*

Proof. By (3.1) and dominated convergence we have $\lim_{s\downarrow t} \mathsf{E} Y_{s,f}^{(i)} = 0$ for all t. Exactly as in the proof of the previous lemma we may conclude that $\mathsf{E} Y_{t,f}^{(i)} = 0$ except for t in an at most countable set $N_f^{(i)}$. But when $t \notin \cup_{i,f} N_f^{(i)}$, we have $Y_{t,f}^{(i)} = 0$ a.s. for all f, implying that $\mathsf{P}\{C_t\} = 1$. QED

Lemma 3.5 *If $x : [0,\infty) \to E$ is right-continuous, then x is $\mathcal{B}(\mathbf{R}_+)/\mathcal{E}$-measurable.*

Proof. It is sufficient to show for any open set $A \in \mathcal{E}$ that the set $C = \{t \,|\, x_t \in A\} \in \mathcal{B}(\mathbf{R}_+)$. Let T be the countable set of discontinuities of x. Let $t \in C \setminus T$. Then there exist $q_t^1 < t < q_t^2$, $q_t^i \in \mathsf{Q}$, such that $x_s \in A$ for $s \in (q_t^1, q_t^2)$. Let further $T \cap C = T_C$. Then $C = T_C \cup \cup_{t\in C\setminus T}(q_t^1, q_t^2)$. Since the collection of sets $\{(q_1, q_2) \,|\, q_1, q_2 \in \mathsf{Q}, q_1 < q_2\}$ is at most countable, C is the countable union of measurable sets. QED

Consider again a metric space E, with metric ρ say. Then $\rho(x,A) = \inf_{y\in A}\rho(x,y)$ is a continuous function of $x \in E$, and hence it is $\mathcal{B}(E)/\mathcal{B}(\mathbf{R})$-measurable by Corollary 3.1. Let $(\Omega,\mathcal{F},\mathsf{P})$ be a probability space, and let X be an E-valued r.v. Then $\rho(X,A)$ is $\mathcal{F}/\mathcal{B}(\mathbf{R})$-measurable and so it is a r.v.

Now, is $\rho(X,Y)$ $\mathcal{F}/\mathcal{B}(\mathbf{R})$-measurable, for Y another E-valued r.v. on Ω? This map is a composition of the map $(X,Y) : \Omega \to E^2$ and the map $\rho : E^2 \to \mathbf{R}$. The map (X,Y) is $\mathcal{F}/\mathcal{B}(E)\times\mathcal{B}(E)$-measurable, where $\mathcal{B}(E)\times\mathcal{B}(E)$ is the product-σ-algebra on E^2. Consider on E^2 the topology generated by the rectangles $A\times B$ with A,B open in E. This generates the Borel-σ-algebra $\mathcal{B}(E^2)$ of E^2. The function $\rho(.,.) : E^2 \to \mathbf{R}$ is continuous, hence $\mathcal{B}(E^2)/\mathcal{B}(\mathbf{R})$-measurable. As a consequence, for measurability of $\rho(X,Y)$, it is sufficient that $\mathcal{B}(E^2) = \mathcal{B}(E)\times\mathcal{B}(E)$. This is guaranteed if E is *separable*!

Lemma 3.6 *Let E, E' be separable, metric spaces. Then $\mathcal{B}(E)\times\mathcal{B}(E') = \mathcal{B}(E\times E')$.*

It is clear that $\mathcal{B}(E)\times\mathcal{B}(E) \subset \mathcal{B}(E^2)$. Can you define a space E and metric ρ, such that $\mathcal{B}(E^2) \neq \mathcal{B}(E)\times\mathcal{B}(E)$?

Monotone class theorems Statements about σ-algebras can often be deduced from the statement on a π-system. We have seen this with respect to measurability, and independence issues, where we have theorems asserting this.

Suppose one does not have a theorem at one's disposal. Then the idea is to show that the sets satisfying a certain condition form a monotone class or a d-system, see below for the defintion. And then one shows that the monotone class or the d-system contains a π-system generating the σ-algebra of interest. By Lemma 3.4, the d-system then contains the σ-algebra of interest and so the desired property applies to our σ-algebra.

Definition 3.7 A collection $\mathcal{S}$ of subsets of the space Ω say, is called a *d-system* or a *monotone class* if

i) $\Omega \in \mathcal{S}$;

ii) $A,B \in \mathcal{S}$ with $A \subseteq B$ implies $B \setminus A \in \mathcal{S}$.

iii) if A_n is an increasing sequence of sets in $\mathcal{S}$, then $\cup_n A_n = \lim_{n\to\infty} A_n \in \mathcal{S}$.

Lemma 3.8 *If a d-system contains a π-system, $\mathcal{I}$ say, then the d-system contains the σ-algebra $\sigma(\mathcal{I})$ generated by the π-system.*

Proof. se Appendix A.1 from Williams PwithM. QED

Part of this lemma is known as Dynkin's lemma (there many of these).

Suppose we want to deduce results on general (real-valued) measurable functions. The 'standard machinery' is the following procedure.

Standard machinery

i) Show that the desired result holds for indicator functions.

ii) Argue by linearity that this implies the result to hold for step functions, that is, finite linear combinations of indicator functions.

iii) Consider non-negative (measurable) functions. Then any such function can be approached by a non-decreasing sequence of step functions. Use monotone convergence to deduce the desired result for non-negative measurable functions.

iv) Consider general measurable functions (if appropriate) and write such functions as the difference of two non-negative functions. Apply (iii).

It is sometimes not easy to show the first step (i). Instead one would like to only consider indicator functions of elements of a π-system generating the σ-algebra. The following theorem allows to deduce results on general (real-valued) measurable functions from results on indicators of the elements of a π-system.

Theorem 3.9 ((Halmos) Monotone Class Theorem: elementary version) *Let $\mathcal{H}$ be a class of bounded functions from a set $\boldsymbol{S}$ to $\mathbf{R}$, satisfying the following conditions:*

i) *$\mathcal{H}$ is a vector space over $\mathbf{R}$ (i.e. it is an Abelian group w.r.t addition of functions, it is closed under scalar multiplication, such that $(\alpha\beta)f = \alpha(\beta f)$, $(-1)f = -f$ and $(\alpha+\beta)f = \alpha f + \beta f$, for $f \in \mathcal{H}$, $\alpha, \beta \in \mathbf{R}$);*

ii) *if f_n, $n = 1, \ldots$, is a sequence of non-negative functions in $\mathcal{H}$, such that $f_n \uparrow f$ for a bounded function f, then $f \in \mathcal{H}$;*

iii) *the constant function belongs to $\mathcal{H}$.*

If $\mathcal{H}$ contains the indicator function of every set in a π-system $\mathcal{I}$, then $\mathcal{H}$ contains every bounded $\sigma(\mathcal{I})$-measurable function.

Proof. see Williams, Diffusions, MProcesses and Martingales I. QED

We can now characterise measurable functions on the path space E^T.

Lemma 3.10 *A function $f : E^T \to \mathbf{R}$ is $\mathcal{E}^T$-measurable if and only if $f = g \circ \pi_S$ for some countable subset S of T and some $\mathcal{E}^S$-measurable function $g : E^S \to \mathbf{R}$.*

Proof. One should use the Monotone Class theorem applied to bounded functions. Then for non-negative functions, with $f = \sum_{n \geq 0} f_n$, where $f_n = f\mathbf{1}_{\{n \leq f < n+1\}}$. Notice that π_S is $\mathcal{E}^T/\mathcal{E}^S$-measurable as a consequence of Lemma 2.1. QED

4 Multivariate normal distribution

Definition. The random vector $X = (X_1, \ldots, X_n)$ has a multivariate normal distribution iff there exists a $n \times k$ matrix B, a vector $a \in \mathbf{R}^n$ and independent random variables $Z_1, \ldots, Z_k$, with $Z_i \stackrel{\mathsf{d}}{=} \mathsf{N}(0, \sigma^2(Z_i))$, such that

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = a + B \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{pmatrix}.$$

In this case, $(X_1, \ldots, X_n)$ is said to have the multivariate normal distribution $\mathsf{N}(a, \boldsymbol{\Sigma})$, where $a = (\mathsf{E}(X_1), \ldots, \mathsf{E}(X_n))^T$ and

$$\boldsymbol{\Sigma} = (\mathsf{cov}(X_i, X_j))_{1 \le i,j \le n} = B \begin{pmatrix} \sigma^2(Z_1) & & & \\ & \sigma^2(Z_2) & & \\ & & \ddots & \\ & & & \sigma^2(Z_k) \end{pmatrix} B^T. \tag{4.1}$$

Problem 4.1 Show that $a = (\mathsf{E}(X_1), \ldots, \mathsf{E}(X_n))^T$, and that the covariance matrix of X is given by (4.1).

Corollary 4.1 *Suppose that the random vector $X = (X_1, \ldots, X_n)$ has a multivariate normal distribution with mean a and covariance matrix $\boldsymbol{\Sigma}$. Let B be an $m \times n$-matrix. Then the random vector BX has a multivariate normal distribution with mean Ba and covariance matrix $B\boldsymbol{\Sigma}B^T$.*

Problem 4.1.a Prove this.

If $\det\boldsymbol{\Sigma} \neq 0$, then X is said to have a non-singular or non-degenerate multivariate normal distribution, and it has density

$$\mathsf{f}_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\{-\tfrac{1}{2}(x-a)^T \boldsymbol{\Sigma}^{-1}(x-a)\}, x = (x_1, \ldots, x_k)^T \in \mathbf{R}^n.$$

If e.g. $\sigma(X_i) = \boldsymbol{\Sigma}_{ii} = 0$, then X_i has a degenerate distribution, in other words $\mathsf{P}\{X_i = a_i\} = 1$.

The following lemma shows that sub-vectors also have a multivariate normal distribution.

Lemma 4.2 i) *The vector $(X_{\nu_1}, \ldots, X_{\nu_r})$, $\nu_1 < \ldots < \nu_r \le n$, has a $\mathsf{N}(a', \boldsymbol{\Sigma}')$ distribution, where $a' = (a_{\nu_1}, \ldots, a_{\nu_r})^T$ and $\boldsymbol{\Sigma}' = (\boldsymbol{\Sigma}_{ij})_{i,j \in \{\nu_1, \ldots, \nu_r\}}$.*

ii) *Suppose that $\boldsymbol{\Sigma}$ has a block diagonal structure, in other words, there are indices $r_0 = 0 < r_1 < \cdots < r_{m-1} < k = r_m$, such that $\boldsymbol{\Sigma}_{ij} \neq 0$ implies that $r_{s-1} + 1 \le i, j \le r_s$ for some $s \in \{1, \ldots, m\}$.*

Then the vectors $(X_1, \ldots, X_{r_1})$, $(X_{r_1+1}, \ldots, X_{r_2})$, $\ldots, (X_{r_{m-1}+1}, \ldots, X_n)$ are mutually independent.

Proof. The proof of (i) is by mere out-integrating. For the proof of (ii), note that $\mathbf{\Sigma}^{-1}$ has the same block diagonal structure as $\mathbf{\Sigma}$. Denote $a_s = (a_{r_{s-1}+1}, \ldots, a_{r_s})$, and let $\mathbf{\Sigma}_s = (\mathbf{\Sigma}_{ij})_{r_{s-1}+1 \le i,j \le r_s}$ be the s-th diagonal block of $\mathbf{\Sigma}$.

By (i) $(X_{r_{s-1}+1}, \ldots, X_{r_s}) \stackrel{\mathsf{d}}{=} \mathsf{N}(a_s, \mathbf{\Sigma}_s)$. Als note that $\det(\mathbf{\Sigma}) = \prod_{s=1}^m \det(\mathbf{\Sigma}_s)$. It then follows directly that $\mathsf{f}_X(x) = \prod_{s=1}^m \mathsf{f}_{(X_{r_{s-1}+1},\ldots,X_{r_s})}(x_{r_{s-1}+1}, \ldots, x_{r_s})$, $x \in \mathbf{R}^k$, in case of a non-singular or non-degenerate multivariate normal distribution. The proof for a singular multivariate normal distribution is similar. QED

Problem 4.2 An $n \times n$ matrix Γ is the covariance matrix of random variables $Y_1, \ldots, Y_n$ iff Γ is symmetric, non-negative definite (i.e. $x^T \Gamma x \ge 0$ for all $x \in \mathbf{R}^n$). Show this.

It is often convenient to use other characterisations of the multivariate normal distribution.

Theorem 4.3 *$X = (X_1, \ldots, X_n)$ has a multivariate normal distribution* **if and only if**

i) *for all vectors $c \in \mathbf{R}^n$, $\sum_i c_i X_i = (c, X)$ has a normal distribution;*

if and only if

(ii) *there exist a vector $a \in \mathbf{R}^n$ and a symmetric, non-negative definite $n \times n$ matrix Γ, such that for all $\theta \in \mathbf{R}^n$*

$$\mathsf{E}e^{i(\theta,X)} = e^{i(\theta,a) - \theta^T \Gamma \theta/2}.$$

Corollary 4.1 *Let $X_1, \ldots, X_n$ be independent, normally distributed random variables defined on the same probability space. Say $X_i \stackrel{\mathsf{d}}{=} \mathsf{N}(a_i, \sigma_i^2)$. Then $(X_1, \ldots, X_n)$ has a multivariate normal distribution with mean $a = (a_1, \ldots, a_n)$ and covariance matrix*

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & & 0 & \sigma_n^2 \end{pmatrix}.$$

Proof. It is a result in elementary probability that a linear combination of independent normally distributed random variables has a normal distribution. The result now follows from the previous theorem. QED

Using characteristic functions calls for inversion formulae.

Theorem 4.4 (Lévy's inversion formula (see Williams P with M))
Let $\phi(\theta) = \int_R e^{i\theta x} dF(x)$ be the characteristic function of a random variable X with distribution function F. Then for $a < b$

$$\tfrac{1}{2}(F(b) + F(b^-)) - \tfrac{1}{2}(F(a) + F(a^-)) = \lim_{T \uparrow \infty} \tfrac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \phi(\theta) d\theta.$$

Moreover, if $\int_R |\phi(\theta)|d\theta < \infty$*, then* X *has a continuous probability density function* f *with*

$$f(x) = \tfrac{1}{2\pi} \int_R e^{-i\theta x} \phi(\theta) d\theta.$$

In case of random variables assuming non-negative values only, it may be convenient to work with the Laplace transform. Let $g : [0, \infty) \to \mathbf{R}$. Then the Laplace transform $\mathcal{L}(g)$ is defined by

$$\mathcal{L}g(s) = \int_{0^-}^{\infty} e^{-st} g(t) dt,$$

with $s \in \mathbf{C}$, and very often $\mathsf{Re}(s) \leq 0$ (in the case of random variables), provided the integral exists. An inversion formula exists in this case as well.

Theorem 4.5 (Mellin's inverse formula)

$$g(t) = \tfrac{1}{2\pi i} \lim_{r \to \infty} \int_{\gamma - ir}^{\gamma + ir} e^{st} \mathcal{L}g(s) ds,$$

with $\gamma > \mathsf{Re}(\sigma)$ *for all singularities* σ *of* $\mathcal{L}g$*.*

The inverse can often be calculated by Cauchy's residue theorem.

Remark It is a consequence of both theorems for distribution functions F and G, that $\phi_F(\theta) = \phi_G(\theta)$ for all $\theta \in \mathbf{R}$ implies that $F \equiv G$. Here ϕ_F and ϕ_G denote the characteristic functions of F and G. The same conclusion holds for F and G the distributions of non-negative r.v. If $\mathcal{L}F(s) = \mathcal{L}G(s)$ for all $s \geq 0$, then $F \equiv G$.

5 Addendum on Measurability

In the proof of Lemma 1.6.11 we need that the limit of a certain sequence of measurable maps is measurable. We will show this.

Lemma 5.1 *Let* (E, d) *be a metric space and let* $\mathcal{B}(E)$ *be the Borel-*σ*-algebra of open sets compatible with* d*. Let* $(X_s^n)_{s \leq t}$*,* $n = 1, \ldots,$ *be* $(E, \mathcal{B}(E))$*-valued adapted stochastic processes on the filtered probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s \leq t}, \mathsf{P})$*, such that the map*

$$(s, \omega) \to X_s^n(\omega)$$

is measurable as a map from $([0, t] \times \Omega, \mathcal{B}([0, t]) \times \mathcal{F}_t)$ *to* $(E, \mathcal{B}(E))$*. Suppose that* $X_s^n(\omega)$ *converges to* $X_s(\omega)$*,* $n \to \infty$*. Then the map* $(s, \omega) \to X_s(\omega)$ *is measurable.*

Proof. Note that $\mathcal{B}(E) = \sigma(G \,|\, G \text{ open in } E)$. Hence it is sufficient to check that

$$X^{-1}(G) = \{(s, \omega) \,|\, X_s(\omega) \in G\} \in \mathcal{B}([0, t]) \times \mathcal{F},$$

for all G open in E. Write $G = \cup_{k=1}^{\infty} F_k$, with $F_k = \{x \in E \mid d(x, G^c) \geq k^{-1}\}$. We need the closed sets F_k to exclude that $X_s^m(\omega) \in G$ for all m but $\lim_{m\to\infty} X_s^m(\omega) \notin G$. By pointwise convergence of $X_s^m(\omega)$ to $X_s(\omega)$ we have that

$$X^{-1}(G) = \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} (X^m)^{-1}(F_k).$$

Since $(X^m)^{-1}(F_k) \in \mathcal{B}[0,t] \times \mathcal{F}$, it follows that $X^{-1}(G) \in \mathcal{B}[0,t] \times \mathcal{F}$. QED

We now give an example of a stochastic process $X = (X_t)_{0 \leq t \leq 1}$, that is not progressively measurable. Moreover, for this example we can construct a stopping time τ, for which X_τ is not $\mathcal{F}_\tau$-measurable.

Example 5.1 Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, and P is the Lebesgue-measure. Let $A \subset \Omega$ be non-measurable, i.e. $A \notin \mathcal{B}$. Put

$$X_t(\omega) = \begin{cases} t + \omega, & t \in A \\ -t - \omega, & t \notin A. \end{cases}$$

Note that $\sigma(X_t) = \mathcal{B}$, for all t. Hence, the natural filtration $\mathcal{F}_t^X = \mathcal{B}$. Further $|X_t(\omega)| = t + \omega$ is continuous in t and ω.

Now $\{(s,\omega) \mid X_s(\omega) \geq 0\} = A \times \Omega$ is not measurable. Consequently X is not progressively measurable.

Define the stopping time T by

$$T = \inf\{t \mid 2t \geq |X_t|\},$$

so that $T(\omega) = \omega$. Clearly $\{\omega : T(\omega) \leq t\} = \{\omega \leq t\} \in \mathcal{B} = \mathcal{F}_t^X$, so that indeed T is a stopping time. One has $\mathcal{F}_\tau = \mathcal{B}$. However,

$$X_{T(\omega)}(\omega) = \begin{cases} 2\omega, & \omega \in A \\ -2\omega, & \omega \notin A \end{cases}$$

and so $\{X_\tau > 0\}$ is not $\mathcal{B}$-measurable, hence not $\mathcal{F}_\tau$-measurable.

Comment on the σ-algebra $\mathcal{F}_\tau$
Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathsf{P})$ be a filtered probability space. Let τ be a stopping time. Then $\mathcal{F}_\tau$ is defined by: $\mathcal{F}_\tau = \{A \in \mathcal{F} \mid A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$.

Let X be a progressively measurable stochastic process, and let τ be a finite stopping time. Then X_τ is an $\mathcal{F}_\tau$-measurable r.v. Hence $\sigma(X_\tau) \subset \mathcal{F}_\tau$. In general $\sigma(X_\tau) \neq \mathcal{F}_\tau$!

Next, define the hitting time of x: $\tau_x = \inf\{t > 0 \mid W_t = x\}$. A direct proof of measurability of τ_x for $x \neq 0$ follows here (cf. LN Example 1.6.9). To show $\mathcal{F}/\mathcal{B}(\mathbf{R})$-measurability of, it is suffcient to show that $\tau_x^{-1}(t,\infty)$ is measurable for each t. This is because the sets $\{(t,\infty), t \in \mathbf{R}\}$ generate $\mathcal{B}(\mathbf{R})$.

For $x \neq 0$ one has

$$\begin{aligned} \{\tau_x > t\} &= \{|W_s - x| > 0, 0 \leq s \leq t\} \\ &= \cup_n \{|W_s - x| > \tfrac{1}{n}, 0 \leq s \leq t\} \\ &= \cup_n \cap_{q \in \mathsf{Q} \cap [0,t]} \{|W_q - x| > \tfrac{1}{n}\}. \end{aligned}$$

We have in fact proved that $\{\tau_x > t\} \in \sigma(W_s, s \leq t)$, hence $\{\tau_x \leq t\} \in \sigma(W_s, s \leq t)$ and so τ_x is a stopping time.

It is slightly more involved to show that $\{\tau_0 > t\}$ is a measurable event, since $W_0 = 0$, a.s. and so there does not exist a uniformly positive distance between the path identically equal to 0 and $W_s(\omega)$ on $(0,t]$.

6 Convergence of probability measures

Let $(\Omega, \mathcal{F}, \mathsf{P})$ a probability space. First of all, let μ, μ_n, $n = 1, \ldots$, be a sequence of probability measures on $(\Omega, \mathcal{F})$.

Definition 6.1 Suppose additionally that Ω is a metric space and $\mathcal{F}$ the Borel-σ-algebra. Then μ_n converges weakly to μ (in formula: $\mu_n \stackrel{w}{\to} \mu$), if $\int f d\mu_n \to \int f d\mu$ for all $f \in C(\Omega) = \{g \,|\, g \text{ continuous, bounded}\}$.

The following theorem gives equivalent criteria for weak convergence.

Theorem 6.2 (Portmanteau Theorem) *$\mu_n \stackrel{w}{\to} \mu$, $n \to \infty$, if and only if one (and hence all) of the following criteria holds:*

i) $\limsup_n \mu_n(F) \leq \mu(F)$ *for all closed sets F;*

ii) $\liminf_n \mu_n(G) \geq \mu(G)$ *for all open sets G;*

iii) $\lim_n \mu(A) = \mu(A)$ *for all sets A, such that $\mu(\delta(A)) = 0$. Here $\delta(A) = A^- - A^0$, where*

$$\begin{aligned} A^- &= \cap_{F \supset A, F \text{ closed}} F \\ A^0 &= \cup_{G \subset A, G \text{ open}} G. \end{aligned}$$

Assume that $(E, \mathcal{E})$ is a measurable space, with E separable, metric (with metric ρ) and $\mathcal{E} = \mathcal{B}(E)$ is the σ-algebra generated by the open sets. Let $X, X_1, X_2, \ldots : \Omega \to E$ be $\mathcal{F}/\mathcal{E}$-measurable random variables.

Definition 6.3 i) X_n converges a.s. to X (in formula: $X_n \stackrel{a.s.}{\to} X$), if $\mathsf{P}\{\lim_{n\to\infty} X_n = X\} = 1$;

ii) X_n converges in probability to X (in formula: $X_n \stackrel{\mathsf{P}}{\to} X$), if $\mathsf{P}\{\rho(X_n - X) \geq \epsilon\} \to 0$, $n \to \infty$, for each $\epsilon > 0$.

iii) X_n converges to X in L^1 (in formula: $X_n \stackrel{L^1}{\to} X$) if $\mathsf{E}|X_n - X| \to 0$, $n \to \infty$.

iv) X_n converges to X in distribution (in formula: $X_n \stackrel{\mathcal{D}}{\to} X$), iff $\mathsf{P}_{X_n} \stackrel{w}{\to} \mathsf{P}_X$, where P_{X_n} (P_X) is the distribution of X_n (X).

In Williams PwithM, Appendix to Chapter 13, you can find the following characterisation of convergence in probability.

Lemma 6.4 i) *$X_n \stackrel{\text{a.s.}}{\to} X$ implies $X_n \stackrel{\mathsf{P}}{\to} X$, implies $X_n \stackrel{\mathcal{D}}{\to} X$.*

ii) *$X_n \stackrel{\mathcal{D}}{\to} a$ implies $X_n \stackrel{\mathsf{P}}{\to} a$, where a is a constant (or degenerate r.v.).*

iii) *$X_n \stackrel{\mathsf{P}}{\to} X$ if and only if every subsequence $\{X_{n_k}\}_{k=1,2,\ldots}$ of $\{X_n\}_n$ contains a further subsequence $\{X_{n'_k}\}_k$ such that $X_{n'_k} \to X$, a.s.*

Suppose that X_n are i.i.d. integrable r.v.'s. Then the (Strong) Law of Large Numbers states that

$$\frac{\sum_{i=1}^n X_i}{n} \stackrel{\text{a.s.}}{\to} \mathsf{E}X_1.$$

If also $\mathsf{E}X_n^2 < \infty$, then the Central Limit Theorem states that

$$\frac{1}{\sqrt{\sigma^2(X_1)n}} \sum_{i=1}^n (X_i - \mathsf{E}X_1) \stackrel{\mathcal{D}}{\to} X^*,$$

where $X^* \stackrel{\mathsf{d}}{=} \mathsf{N}(0,1)$.

Suppose that X_n are i.i.d. random elements, defined on a common probability space, with values in a real, separable Banach space. A famous paper by Itô and Nisio contains the following surprising convergence results. Write $S_n = \sum_{i=1}^n X_i$.

Theorem 6.5 • *Equivalent are*

1. S_n, $n = 1, 2, \ldots$, *converges in probability;*
2. S_n, $n = 1, 2, \ldots$, *converges in distribution;*
3. S_n, $n = 1, 2, \ldots$, *converges a.s.*

- *If* S_n, $n = 1, 2, \ldots$, *are uniformly tight, i.e. for each* $\epsilon > 0$ *there exists a compact set* $K \subset E$, *such that*

$$\mathsf{P}\{S_n \in K\} \geq 1 - \epsilon, \quad n = 1, 2, \ldots,$$

then there exist $c_1, c_2, \ldots \in E$, *such that* $S_n - c_n$ *converges a.s.*

If the random elements are symmetric, i.e. $X_i \stackrel{\mathsf{d}}{=} -X_i$, then even more can be said. First, let E^* be the collection of bounded continuous functions on E.

Theorem 6.6 *If* X_i, $i = 1, 2, \ldots$ *are all symmetric random elements, then a.s. convergence is equivalent to uniform tightness. In particular, the following additional equaivalences hold.*

3 S_n, $n = 1, 2, \ldots$, *converges in probability;*

4 S_n, $n = 1, 2, \ldots$, *is uniformly tight;*

5 $f(S_n)$, $n = 1, 2, \ldots$, *converges in probability, for each* $f \in E^*$*;*

6 *there exists an* E*-valued random element* S, *such that* $\lim_{n\to\infty} \mathsf{E}e^{if(S_n)} \to \mathsf{E}e^{if(S)}$ *for every* $f \in E^*$.

These convergence results play an important role in the analysis of processes with independent increments.

7 Conditional expectation

Theorem 7.1 (Fundamental Theorem and Definition of Kolmogorov 1933)
Suppose we have a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. *Let* X *be a real-valued, integrable random variable, i.e.* $\mathsf{E}(|X|) < \infty$. *Let* $\mathcal{A}$ *be a sub-*σ*-algebra of* $\mathcal{F}$. *Then there exists a random variable* Y *such that*

i) Y *is* $\mathcal{A}$*-measurable;*

ii) $\mathsf{E}(|Y|) < \infty$*;*

iii) *for each* $A \in \mathcal{A}$ *we have*

$$\int_A Y d\mathsf{P} = \int_A X d\mathsf{P}.$$

If Y' *is another r.v. with properties (i,ii,iii), then* $Y' = Y$ *with probability 1, i.e.* $\mathsf{P}\{Y' = Y\} = 1$. *We call* Y *a version of the conditional expectation* $\mathsf{E}(X|\mathcal{A})$ *of* X *given* $\mathcal{A}$ *and we write* $Y = \mathsf{E}(X|\mathcal{A})$ *a.s.*

Proof. To prove existence, suppose first that $X \geq 0$. Define the measure μ on $(\Omega, \mathcal{A})$ by

$$\mu(A) = \int_A X d\mathsf{P}, \quad A \in \mathcal{A}.$$

Since X is integrable, μ is finite. Now for all $A \in \mathcal{A}$ we have that $\mathsf{P}(A) = 0$ implies $\mu(A) = 0$. In other words, μ is absolutely continuous with respect to P. By the Radon-Nikodym theorem, there exists a measurable random variable Y on $(\Omega, \mathcal{A})$ such that

$$\mu(A) = \int_A Y d\mathsf{P}, \quad A \in \mathcal{A}.$$

Clearly, Y has the desired properties. The general case follows by linearity.

As regards a.s. unicity of Y, suppose that we have two random variables Y, Y' satisfying (i,ii,iii). Let $A_n = \{Y - Y' \geq 1/n\}$, $n = 1, 2, \ldots$. Then $A_n \in \mathcal{A}$ and so

$$0 = \int_{A_n} (Y - Y') d\mathsf{P} \geq \tfrac{1}{n} \mathsf{P}\{A_n\},$$

so that $\mathsf{P}(A_n) = 0$. It follows that

$$\mathsf{P}\{Y - Y' > 0\} = \mathsf{P}\{\bigcup_{n=1}^{\infty} A_n\} \leq \sum_{n=1}^{\infty} \mathsf{P}\{A_n\} = 0.$$

Hence $Y \leq Y'$ a.s. Interchanging the roles of Y and Y' yields that $Y' \leq Y$ a.s. Thus $Y = Y'$ a.s. QED

N.B.1 Conditional expectations are random variables!
Conditional probabilities are conditional expectations: $\mathsf{P}\{X \in A \,|\mathcal{F}) = \mathsf{E}\{\mathbf{1}_{\{X \in A\}} \,|\mathcal{F}\}$. $\mathsf{E}(X \,|\, Y)$ stands for $\mathsf{E}(X \,|\, \sigma(Y))$, where Y may be a random variable, a random element, etc.

N.B.2 Suppose we have constructed an $\mathcal{A}$-measurable r.v. Z, with $\mathsf{E}(|Z|) < \infty$, such that (iii) holds for all $A \in \pi(\mathcal{A})$, i.e. (iii) holds on a π-system generating $\mathcal{A}$, containing the whole space. Then (iii) holds for all $A \in \mathcal{A}$, and so Z is a version of the conditional expectation $\mathsf{E}(X|\mathcal{A})$. This follows from the interpretation of $\int_A X d\mathsf{P}$ as a measure on $\mathcal{A}$ and the fact that two measures that are equal on a π-system, are equal on the generated σ-algebra, provided *they assign equal mass to the whole space.*

This can be used in determining conditional expectations: make a guess and check that it is the right one on a π-system.

Elementary properties The following properties follow either immediately from the definition, or from the corresponding properties of ordinary expectations.

Lemma 7.2 **i)** *If X is $\mathcal{A}$-measurable and $\mathsf{E}|X| < \infty$, then $\mathsf{E}(X \mid \mathcal{A}) = X$;*

ii) $\mathsf{E}(X \mid \{\emptyset, \Omega\}) = \mathsf{E}X$, *a.s.;*

iii) *Linearity:* $\mathsf{E}(aX + bY \mid \mathcal{A}) \overset{\text{a.s.}}{=} a\mathsf{E}(X \mid \mathcal{A}) + b\mathsf{E}(Y \mid \mathcal{A})$*;*

iv) *Positivity: if $X \geq 0$ a.s., then $\mathsf{E}(X \mid \mathcal{A}) \geq 0$;*

v) *Monotone convergence: if $X_n \uparrow X$ a.s., then $\mathsf{E}(X_n \mid \mathcal{A}) \uparrow \mathsf{E}(X \mid \mathcal{A})$ a.s.;*

vi) *Fatou: if $X_n \geq 0$ a.s., then $\mathsf{E}(\liminf X_n \mid \mathcal{A}) \leq \liminf \mathsf{E}(X_n \mid \mathcal{A})$ a.s.*

vii) *Dominated convergence: suppose that $|X_n| \leq Y$ a.s. and $\mathsf{E}Y < \infty$. Then $X_n \to X$ a.s. implies $\mathsf{E}(X_n \mid \mathcal{A}) \to \mathsf{E}(X \mid \mathcal{A})$ a.s.*

viii) *Jensen: if ϕ is a convex funcion such that $\mathsf{E}|\phi(X)| < \infty$, then $\mathsf{E}(\phi(X) \mid \mathcal{A}) \geq \phi(\mathsf{E}(X \mid \mathcal{A}))$ a.s.;*

ix) *Tower property: if $\mathcal{A} \subset \mathcal{B}$, then $\mathsf{E}(X \mid \mathcal{A}) = \mathsf{E}(\mathsf{E}(X \mid \mathcal{B}) \mid \mathcal{A})$ a.s.;*

x) *Taking out what is known: if Y is $\mathcal{A}$-measurable and bounded, then $\mathsf{E}(YX \mid \mathcal{A}) = Y\mathsf{E}(X \mid \mathcal{A})$ a.s. The same assertion holds as well if $X, Y \geq 0$ a.s. and $\mathsf{E}(XY), \mathsf{E}X < \infty$, or if $\mathsf{E}|X|^q, \mathsf{E}|Y|^p < \infty$, with $1/q + 1/p = 1$, $p, q \neq 1$;*

xi) *Role of independence: if $\mathcal{B}$ is independent of $\sigma(\sigma(X), \mathcal{A})$, then $\mathsf{E}(X \mid \sigma(\mathcal{B}, \mathcal{A})) = \mathsf{E}(X \mid \mathcal{A})$ a.s.*

Proof. Exercise! The proofs of parts (iv), (viii), (x) and (xi) are the most challenging, see Williams PwithM. QED

hier nog iets over onafhankelijkheid!!

An simple method for checking whether two conditional expectations are (a.s.) equal relies on the following lemma.

Lemma 7.2A *Suppose X, Y are r.v. defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Then $X \geq Y$, a.s. if and only if*

$$\int_F X d\mathsf{P} = \mathsf{E}\mathbf{1}_{\{F\}} X \geq \mathsf{E}\mathbf{1}_{\{F\}} Y = \int_F Y d\mathsf{P}$$

for all $F \in \mathcal{F}$.

Conditioning on σ-algebras generated by random variables Let $(\Omega_i, \mathcal{F}_i)$, $i = 1,2$ be two measure spaces. Suppose that $X : \Omega_1 \to \Omega_2$ is $\mathcal{F}_1/\mathcal{F}_2$-measurable. Then $\sigma(X) = \sigma(X^{-1}(A), A \in \mathcal{F}_2) = (X^{-1}(A), A \in \mathcal{F}_2)$, i.e. the σ-algebra generated by X is simply the collection of inverse images under X of sets in $\mathcal{F}_2$. Hence, $A \in \sigma(X)$ iff there exists a set $B \in \mathcal{F}_2$ with $X(A) = B$. The following helps to understand conditional expectations.

Theorem 7.3 *Let $(\Omega_1, \mathcal{F}_1)$ be a measure space with the property that $\mathcal{F}_1$ is the σ-algebra generated by a map g on Ω_1 with values in a measure space $(\Omega_2, \mathcal{F}_2)$, i.e. $\mathcal{F}_1 = \sigma(g)$. Then a real-valued function f on $(\Omega_1, \mathcal{F}_1)$ is measurable if and only if there exists a real-valued measurable function h on $(\Omega_2, \mathcal{F}_2)$, such that $f = h(g)$.*

Proof. If there exists such a function h, then $f^{-1}(B) = g^{-1}(h^{-1}(B)) \in \mathcal{F}_1$, by measurability of g and h. Suppose therefore that f is measurable. We have to show the existence of a function h with the above required properties.

The procedure is by going through indicator functions through elementary functions to positive functions and then to general functions f:

i) Assume that $f = \mathbf{1}_{\{A\}}$ for some set $A \in \mathcal{F}_1$. Put $B = g(A)$, then $B \in \mathcal{F}_2$. Put $h = \mathbf{1}_{\{B\}}$.

ii) Let $f = \sum_{i=1}^n a_i \mathbf{1}_{\{A_i\}}$, $A_i \in \mathcal{F}_1$. It follows from (i) that $\mathbf{1}_{\{A_i\}} = h_i(g)$ with $h_i = \mathbf{1}_{\{g(A_i)\}}$ measurableon Ω_2. But then also $h = \sum_{i=1}^n a_i h_i$ is measurable on Ω_2, and $f = h(g)$.

iii) Let $f \geq 0$. Then there is a sequence of elementary functions f_n, with $f(\omega) = \lim_{n\to\infty} f_n(\omega)$, $\omega \in \Omega_1$. By (ii) $f_n = h_n(g)$ with h_n measurable on Ω_2. It follows that $\lim_n h_n(\omega_2)$ exists for all $\omega_2 \in \{g(\omega_1) : \omega_1 \in \Omega_1\}$. Define

$$h(\omega_2) = \begin{cases} \lim_n h_n(\omega_2), & \text{if this limit exists} \\ 0, & \text{otherwise} \end{cases}$$

then h is measurable on Ω_2 and $f = h(g)$.

iv) Write $f = f^+ - f^-$ for general f, and apply (iii).

QED

Corollary 7.4 (Doob-Dynkin) *Let $Y = (Y_1, \ldots, Y_k)$, $Y_i : \Omega \to E$, be a random vector and let X be an integrable real-valued r.v., both defined on the same probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Then there exists a real-valued $\mathcal{E}^k/\mathcal{B}(\mathbf{R})$-measurable function $h : E^k \to \mathbf{R}$, such that $\mathsf{E}(X \mid Y_1, \ldots, Y_k) = h(Y_1, \ldots, Y_k)$.*

Let $Y = (Y_t)_{t\in T}$ be an $(E, \mathcal{E})$-valued random element defined on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Then there exists a measurable function $h : E^S \to \mathbf{R}$, $S = \{t_1, \ldots, \}$, such that $\mathsf{E}(X \mid Y) = h(Y_{t_1}, \ldots)$.

Proof. By Theorem 7.3, $\mathsf{E}(X \mid Y) = h(Y)$ for a measurable function $h : E^T \to \mathbf{R}$. By Lemma 3.10 $h = h' \circ \pi_S$ for a countable subset $S \subset T$ and an $\mathcal{E}^S$-measurable function $h' : E^S \to \mathbf{R}$. As a consequence, $h(Y) = h'(\pi_S(Y)) = h'(Y_{t_1}, \ldots,)$, where $S = \{t_1, \ldots\}$. QED

Suppose X is a (measurable) map from $(\Omega, \mathcal{F}, \mathsf{P})$ to $(E, \mathcal{E})$. Let $A \in \mathcal{F}$. You are accustomed to interpreting

$$\mathsf{P}(X \in B \mid A) = \frac{\mathsf{P}\Big(\{X \in B\} \cap A\Big)}{\mathsf{P}(A)}, \tag{7.1}$$

provided $\mathsf{P}(A) > 0$. Formally we get

$$\mathsf{P}(X \in B \mid A) = \mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid A),$$

which we can interpret as follows. Let $\mathcal{A} = \{\emptyset, \Omega, A, A^c\}$, then $\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})$ is $\mathcal{A}/\mathcal{E}$-measurable. Hence $\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})$ is constant on A and A^c: a.s. we have

$$\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})(\omega) = \begin{cases} c_1, & \omega \in A \\ c_2, & \omega \in A^c, \end{cases}$$

for some constants c_1 and c_2. Hence,

$$\mathsf{P}\Big(\{X \in B\} \cap A\Big) = \int_A \mathbf{1}_{\{X \in B\}} d\mathsf{P} = \int_A \mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A}) d\mathsf{P} = c_1 \mathsf{P}(A),$$

and so we obtain

$$\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})(\omega) = \frac{\mathsf{P}\Big(\{X \in B\} \cap A\Big)}{\mathsf{P}(A)},$$

if $\mathsf{P}(A) > 0$. In view of (7.1) we see that $\mathsf{P}(X \in B \mid A)$ stands for the value $\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})(\omega)$ on A.

How can we take $\mathsf{E}(\mathbf{1}_{\{X \in B\}} \mid \mathcal{A})(\omega)$ on A, if $\mathsf{P}(A) = 0$?

The following observation is useful for computations.

Lemma 7.5 *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. Let $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ be independent σ-algebras. Let $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2)$ be measure spaces. Suppose that $X : \Omega \to E_1$ is $\mathcal{F}_1$-measurable and that $Y : \Omega \to E_2$ is $\mathcal{F}_2$-measurable. Let further $f : E_1 \times E_2 \to \mathbf{R}$ be $\mathcal{E}_1 \times \mathcal{E}_2$-measurable with $\mathsf{E}|f(X, Y)| < \infty$. Then there exists an $\mathcal{E}_2$-measurable function $g : E_2 \to \mathbf{R}$, such that $\mathsf{E}(f(X, Y) \mid \mathcal{F}_2) = g(Y)$, where*

$$g(y) = \int_\Omega f(X(\omega), y) d\mathsf{P}(\omega) = \int_x f(x, y) d\mathsf{P}_X(x). \tag{7.2}$$

Proof. Define $\mathcal{H}$ as the class of bounded, measurable functions $f : E_1 \times E_2 \to \mathbf{R}$, with the property that for each $f \in \mathcal{H}$ the function $g : E_2 \to \mathbf{R}$ given by (7.2) is $\mathcal{E}_2$-measurable, with $\mathsf{E}(f(X, Y) \mid \mathcal{F}_2) = g(Y)$. It is straightforward to check that $\mathcal{H}$ is a monotone class in the sense of Theorem 3.9.

We will check that $f = \mathbf{1}_{\{B_1 \times B_2\}} \in \mathcal{H}$, for $B_1 \in \mathcal{E}_1$, $B_2 \in \mathcal{E}_2$. By virtue of Theorem 3.9, $\mathcal{H}$ then contains all bounded $\mathcal{E}_1 \times \mathcal{E}_2$-measurable functions.

Now

$$\begin{aligned} \mathsf{E}(\mathbf{1}_{\{B_1 \times B_2\}}(X, Y) \mid \mathcal{F}_2) &= \mathsf{E}(\mathbf{1}_{\{B_1\}}(X) \mathbf{1}_{\{B_2\}}(Y) \mid \mathcal{F}_2) \\ &= \mathbf{1}_{\{B_2\}}(Y) \mathsf{E}(\mathbf{1}_{\{B_1\}}(X) \mid \mathcal{F}_2) \\ &= \mathbf{1}_{\{B_2\}}(Y) \mathsf{E}(\mathbf{1}_{\{B_1\}}(X)). \end{aligned}$$

On the other hand

$$\mathsf{E}(f(X,y)\,|\,\mathcal{F}_2) = \mathbf{1}_{\{B_2\}}(y)\mathsf{E}(\mathbf{1}_{\{B_1\}}(X)\,|\,\mathcal{F}_2) = \mathbf{1}_{\{B_2\}}(y)\mathsf{E}(\mathbf{1}_{\{B_1\}}(X)).$$

Put $g(y) = \mathbf{1}_{\{B_2\}}(y)\mathsf{E}(\mathbf{1}_{\{B_1\}}(X))$, then the above implies that $g(Y)$ is a version of $\mathsf{E}(\mathbf{1}_{\{B_1\times B_2\}}(X,Y)\,|\,\mathcal{F}_2)$.

Derive the result for unbounded functions f yourself. QED

Example
Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}[0,1]$, $\mathsf{P} = \lambda$.

Let $X : [0,1] \to \mathbf{R}$ be defined by $X(\omega) = \omega^2$. X is a measurable function on $(\Omega, \mathcal{F}, \mathsf{P})$.

Let $\mathcal{F}_0 = \sigma\{[0,1/2],(1/2,1]\}$. Since $\mathsf{E}(X\,|\,\mathcal{F}_0)$ is $\mathcal{F}_0$-measurable, it must be constant on the sets $[0,1/2]$ and $(1/2,1]$. We get for $\omega \in [0,1/2]$ that

$$\frac{1}{2}\mathsf{E}(X\,|\,\mathcal{F}_0)(\omega) = \mathsf{E}\int_{[0,1/2]} \mathsf{E}(X\,|\,\mathcal{F}_0)d\mathsf{P} = \int_{[0,1/2]} X d\mathsf{P} = \frac{1}{24}, \quad a.s.$$

and for $\omega \in (1/2,1]$

$$\frac{1}{2}\mathsf{E}(X\,|\,\mathcal{F}_0)(\omega) = \int_{(1/2,1]} \mathsf{E}(X\,|\,\mathcal{F}_0) = \int_{(1/2,1]} X d\mathsf{P} = \frac{7}{24}, a.s.$$

Put

$$Z(\omega) = \begin{cases} \frac{1}{12}, & \omega \in [0,1/2] \\ \frac{7}{12}, & \omega \in (1,2,1]. \end{cases}$$

The sets $[0,1/2],(1/2,1]$ form a π-system for $\mathcal{F}_0$. Since $\int_A Z d\mathsf{P} = \int_A X d\mathsf{P}$ for the sets A in a π-system for $\mathcal{F}_0$, we have that Z is a version of $\mathsf{E}(X\,|\mathcal{F}_0)$, i.o.w. $Z = \mathsf{E}(X\,|\,\mathcal{F}_0)$ a.s.

Next let $Y : \Omega \to \mathbf{R}$ be given by $Y(\omega) = (1/2-\omega)^2$. We want to determine $\mathsf{E}(X\,|\,Y) = \mathsf{E}(X\,|\,\sigma(Y))$.

$\sigma(Y)$ is the σ-algebra generated by the π-system $\{[0,\omega]\cup[1-\omega,1]\,|\,\omega \in [0,1/2]\}$. By Corollary 7.4 there exists a $\sigma(Y)$-measurable function $h : \Omega \to \mathbf{R}$, such that $\mathsf{E}(X\,|\,Y) = h(Y)$. Since Y is constant on sets of the form $\{\omega, 1-\omega\}$, necessarily $\mathsf{E}(X\,|\,Y)$ is constant on these sets!

The easiest way to determine $\mathsf{E}(X\,|\,\sigma(Y))$ is by introducing a 'help' random variable and by subsequently applying Lemma 3.2. Say $Z : \Omega \to \{0,1\}$ is defined by $Z(\omega) = 0$ for $\omega \le 1/2$ and $Z(\omega) = 1$ for $\omega > 1/2$. Then $\sigma(Z)$ and $\sigma(Y)$ are independent, and $\sigma(Y,Z) = \mathcal{F}$ (check this). Hence $\mathsf{E}(X\,|\,Y,Z) = X$, a.s.

By Corollary 7.4 there exists a $\mathcal{B}(\mathbf{R}^+) \times \sigma(\{0\},\{1\})$-measurable function f, such that $X = f(Y,Z)$. By computation we get $f(y,z) = (-\sqrt{y}-1/2)^2\mathbf{1}_{\{z=0\}} + (\sqrt{y}+1/2)^2\mathbf{1}_{\{z=1\}}$.

By Lemma 7.5

$$g(Y) = \mathsf{E}(X\,|\,Y) = \mathsf{E}(f(Y,Z)\,|\,Y),$$

where

$$g(y) = \int_z f(y,z)d\mathsf{P}_Z(z) = \frac{1}{2}(-\sqrt{y}-1/2)^2 + \frac{1}{2}(\sqrt{y}+1/2)^2 = y + \frac{1}{4}.$$

It follows that $\mathsf{E}(X\,|\,Y) = Y + \frac{1}{4}$ a.s.

Problem 7.1 Let X and Y be two r.v. defines on the same probability space, such that $X \stackrel{\mathsf{d}}{=} \exp(\lambda)$, i.e. $\mathsf{P}\{X > t\} = e^{-\lambda t}$; $Y \geq 0$ a.s. and X and Y are independent. Show the *memoryless property of the (negative) exponential distribution*

$$\mathsf{P}\{X > t + Y \mid X > Y\} = \mathsf{P}\{X > t\}, \quad t \geq 0.$$

Problem 7.2 A rather queer example. Let $\Omega = (0,1]$. Let $\mathcal{A}$ be the σ-algebra generated by all one-point sets $\{x\}$, $x \in (0,1]$. Then $\mathcal{A} \subset \mathcal{B}(0,1]$.

i) Classify $\mathcal{A}$.

ii) Let λ be the Lebesgue measure on $((0,1], \mathcal{B}(0,1])$. Let $X : ((0,1], \mathcal{B}(0,1]) \to (\mathbf{R}, \mathcal{B})$ be any integrable r.v. Determine $\mathsf{E}(X|\mathcal{A})$. Explain heuristically.

Gambling systems A casino offers the following game consisting of n rounds. In every round t he bets $\alpha_t \geq 0$. His bet in round t may depend on his knowledge of the game's past.

The outcomes η_t, $t = 1, \ldots$ of the game are i.i.d. r.v.s with values in $\{-1, 1\}$ and $\mathsf{P}\{\eta_t = 1\} = 1/2 = \mathsf{P}\{\eta_t = -1\}$. The gambler's capital at time t is therefore $X_t = \sum_{i=1}^{t} \alpha_i \eta_i$.

A gambling statregy $\alpha_1, \alpha_2, \ldots$ is called *admissable* (or *predictable*) if α_t is $\sigma(\eta_1, \eta_2, \ldots, \eta_{t-1})$-measurable. In words this means that the gambler has no prophetic abilities. His bet at time t depends exclusively on observed past history.

Example: $\alpha_t = 1_{\eta_t > 0}$ "only bet if you will win" is *not admissible.*

Problem 7.3 By the distribution of outcomes, one has $\mathsf{E}(X_t) = 0$. Prove this.

One has $T = \min\{t | X_t \leq \alpha\}$ is a stopping time, since $\{T \leq t\} = \cup_{l=0}^{t} \{X_l \leq \alpha\}$ and

$$\{X_l \leq \alpha\} \in \sigma(\eta_1, \ldots, \eta_l\} \subset \sigma(\eta_1, \ldots, \eta_t\}, \quad l \leq t.$$

Now, $\alpha_t = \mathbf{1}_{\{T > t-1\}} = \mathbf{1}_{\{T \geq t\}} \in \sigma(\eta_1, \ldots, \eta_{t-1})$ defines an admissible gambling strategy with

$$X_t = \sum_{j=1}^{t} \alpha_j \eta_j = \sum_{j=1}^{t} \mathbf{1}_{\{T \geq j\}} \eta_j = \sum_{j=1}^{\min\{t,T\}} \eta_j = S_{\min\{t,T\}},$$

where $S_t = \sum_{j=1}^{t} \eta_j$. Hence $\mathsf{E} S_{\min\{t,T\}} = 0$ if T is a stopping time.

Hedging We have seen that the above gambling strategies cannot modify the expectation: on the average the gambler wins and loses nothing. Apart from that, which payoffs can one obtain by gambling?

We discuss a simple model for stock options. Assume that the stock price either increases by 1 or decreases by 1 every day, with probability 1/2, independently from day to day. Suppose I own α_t units of stock at time t. Then the value of my portfolio increases by $\alpha_t \eta_t$ every day (η_t are defined as in the gambling section).

Suppose the bank offers the following contract "European option": at a given time t one has the choice to buy 1 unit of stock for price C or not to buy it. C is specified in advance. Our pay-off per unit stock is $(S_t - C)^+$. In exchange, the bank receives a deterministic amount $\mathsf{E}((S_t - C)^+)$.

Can one generate the pay-off by an appropriate gambling strategy? The answer is yes, and in fact much more is true.

Lemma 7.6 *Let Y be a $\sigma(\eta_1, \ldots, \eta_n)$ measurable function. Then there is an admissible gambling strategy $\alpha_1, \ldots, \alpha_t$ such that*

$$Y - \mathsf{E}(Y) = \sum_{j=1}^{n} \alpha_j \eta_j.$$

Proof. Write $\mathcal{F}_n = \sigma(\eta_1, \ldots, \eta_n)$. Define α_j by

$$\alpha_j \eta_j = \mathsf{E}(Y|\mathcal{F}_j) - \mathsf{E}(Y|\mathcal{F}_{j-1}),$$

where $\mathcal{F}_0 = \{\emptyset, \Omega\}$. We have to show that α_j is $\mathcal{F}_{j-1}$-measurable. It is clearly $\mathcal{F}_j$-measurable and so $\alpha_j = \mathsf{E}(\alpha_j \,|\, \mathcal{F}_j)$. By Corollary 7.4 there exists a measurable function $f : \mathbf{R}^j \to \mathbf{R}$ such that $\mathsf{E}(\alpha_j|\mathcal{F}_j) = f(\eta_1, \ldots, \eta_j)$.

Problem 7.4 i) Show that $\mathsf{E}(\alpha_j \eta_j|\mathcal{F}_{j-1}) = 0$.

ii) Use (i) to show that

$$f(\eta_1, \ldots, \eta_{j-1}, 1) = f(\eta_1, \ldots, \eta_{j-1}, -1)$$

Hint: use Corollary 7.4.

iii) Explain now why α_j is $\mathcal{F}_{j-1}$-measurable. Conclude the proof of Lemma 7.6.

QED

8 Uniform integrability

Suppose that $\{X_n\}_n$ is an integrable sequence of random variables with $X_n \stackrel{a.s.}{\to} X$ for some integrable random variable X. Under what conditions does $\mathsf{E}X_n \to \mathsf{E}X$, $n \to \infty$, or, even stronger, when does $X_n \stackrel{L_1}{\to} X$?

If $X_n \le X_{n+1}$, a.s., both are true by monotone convergence. If this is not the case, we need additional properties to hold.

For an integrable random variable X, the map $A \to \int_A |X| d\mathsf{P}$ is "uniformly continuous" in the following sense.

Lemma 8.1 *Let X be an integrable random variable. Then for every $\epsilon > 0$ there exists a $\delta > 0$ such that for all $A \in \mathcal{F}$, $\mathsf{P}\{A\} \le \delta$ implies*

$$\int_A |X| d\mathsf{P} < \epsilon.$$

Proof. See for instance Williams, Pwith M. QED

Definition 8.2 Let $\mathcal{C}$ be an arbitrary collection of random variables on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. We call the collection *uniformly integrable* (UI) if to every $\epsilon > 0$ there exists a constant $K \geq 0$ such that

$$\int_{|X|>K} |X| d\mathsf{P} \leq \epsilon, \quad \text{for all } X \in \mathcal{C}.$$

The following lemma gives an important example of uniformly integrable classes.

Lemma 8.3 *Let $\mathcal{C}$ be a collection of random variables on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$.*

i) *If the class $\mathcal{C}$ is bounded in $L^p(\mathsf{P})$ for some $p > 1$, then $\mathcal{C}$ is uniformly integrable.*

ii) *If $\mathcal{C}$ is uniformly integrable, then $\mathcal{C}$ is bounded in $L^1(\mathsf{P})$.*

Proof. Exercise, see Williams, PwithM. QED

Another useful characterisation of UI is the following. It allows for instance to conclude that the sum of two UI sequences of random variables is UI.

Lemma 8.4 *Let $\mathcal{C}$ be a collection of random variables on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. $\mathcal{C}$ is UI if and only if $\mathcal{C}$ is bounded in L^1 and*

$$\sup_{A \in \mathcal{F}: \mathsf{P}\{A\} < \epsilon} \sup_{X \in \mathcal{C}} \mathsf{E}(\mathbf{1}_{\{A\}} |X|) \to 0, \quad \epsilon \to 0. \tag{8.1}$$

Proof. Let $\mathcal{C}$ be a UI collection. We have to show that for all $\delta > 0$ there exists $\epsilon > 0$ such that

$$\sup_{A \in \mathcal{F}: \mathsf{P}\{A\} < \epsilon} \sup_{X \in \mathcal{C}} \mathsf{E}(\mathbf{1}_{\{A\}} |X|) \leq \delta. \tag{8.2}$$

Notice that

$$\mathsf{E}(\mathbf{1}_{\{A\}} |X|) \leq x \mathsf{P}\{A\} + \mathsf{E}\{\mathbf{1}_{\{(x,\infty)\}}(X) |X|\}. \tag{8.3}$$

Let x be large enough, so that $\mathsf{E}\{\mathbf{1}_{\{(x,\infty)\}}(X)|X|\} \leq \delta/2$. Next, choose ϵ with $x\epsilon \leq \delta/2$. The result follows.

Boundedness in L^1 follows by putting $A = \Omega$ in (8.3) and setting x large enough, so that $\sup_{X \in \mathcal{C}} \mathsf{E}\{\mathbf{1}_{\{(x,\infty)\}}(X)|X|\} < \infty$.

For the reverse statement, assume L^1-boundedness of $\mathcal{C}$ as well as (8.1). By Chebychev's inequality

$$\sup_{X \in \mathcal{C}} \mathsf{P}\{|X| > x\} \leq \frac{1}{x} \sup_{X \in \mathcal{C}} \mathsf{E}|X| \to 0, \quad x \to \infty. \tag{8.4}$$

To show UI, we have to show that for each $\delta > 0$ there exists $x \in \mathbf{R}$ such that

$$\mathsf{E}\mathbf{1}_{\{(x,\infty)\}}(X)|X|\} \leq \delta, \quad X \in \mathcal{C}.$$

Fix $\delta > 0$. By assumption there exists $\epsilon > 0$ for which (8.2) holds. By virtue of (8.4), there exists x so that $\mathsf{P}\{X| > x\} \leq \epsilon$ for $X \in \mathcal{C}$. Put $A_X = \{X > x\}$. Then (8.4) implies that

$$\sup_X \mathsf{E}\{\mathbf{1}_{\{(x,\infty)\}}(X)|X|\} = \sup_X \mathsf{E}\{\mathbf{1}_{\{A_X\}}|X|\} \leq \sup_{A: \mathsf{P}\{A\} \leq \epsilon} \sup_X \mathsf{E}\{\mathbf{1}_{\{A\}}|X|\} \leq \delta.$$

QED

Conditional expectations give us the following important example of a uniformly integrable class.

Lemma 8.5 *Let X be an integrable random variable. Then the class*

$$\mathcal{C} = \{\mathsf{E}(X \mid \mathcal{A}) \mid \mathcal{A} \text{ is sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

is uniformly integrable.

Proof. By conditional Jensen (Lemma 7.2) $\mathsf{E}|\mathsf{E}(X \mid \mathcal{A})| \leq \mathsf{E}\mathsf{E}(|X| \mid \mathcal{A}) = \mathsf{E}|X|$. This yields L^1-boundedness of $\mathcal{C}$. Let $F \in \mathcal{F}$. Then

$$\begin{aligned} \mathsf{E}(|\mathsf{E}(X \mid \mathcal{A})\mathbf{1}_{\{F\}})| &\leq \mathsf{E}(\mathsf{E}(|X| \mid \mathcal{A})\mathbf{1}_{\{F\}}) \\ &= \mathsf{E}(\mathsf{E}(\mathsf{E}(|X| \mid \mathcal{A})\mathbf{1}_{\{F\}}|\mathcal{A})) \\ &= \mathsf{E}(\mathsf{E}(|X| \mid \mathcal{A})\mathsf{E}(\mathbf{1}_{\{F\}} \mid \mathcal{A})) = \mathsf{E}(|X|\mathsf{E}(\mathbf{1}_{\{F\}} \mid \mathcal{A})). \end{aligned}$$

Let $\delta > 0$. By virtue of Lemma 8.1 it is sufficient to show the existence of $\epsilon > 0$ such that

$$\sup_{F \in \mathcal{F}: \mathsf{P}\{F\} \leq \epsilon} \sup_{\mathcal{A}} \mathsf{E}(|X|\mathsf{E}(\mathbf{1}_{\{F\}} \mid \mathcal{A})) \leq \delta.$$

Suppose that this is not true. Then for any n there exist $F_n \in \mathcal{F}$ with $\mathsf{P}\{F_n\} \leq 1/n$ and $\mathcal{A}_n$, such that

$$\mathsf{E}(|X|\mathsf{E}(\mathbf{1}_{\{F_n\}} \mid \mathcal{A}_n)) > \delta.$$

Since $\mathsf{P}\{F_n\} \leq 1/n$, we can choose a subsequence, indexed by n again, such that $\mathsf{E}(\mathsf{E}(\mathbf{1}_{\{F_n\}} \mid \mathcal{A}_n)) = \mathsf{P}\{F_n\} \downarrow 0$ as $n \to \infty$. Hence $\mathsf{E}(\mathbf{1}_{\{F_n\}} \mid \mathcal{A}_n) \xrightarrow{\mathsf{P}} 0$ (see Theorem 8.6 below). The sequence has a subsequence converging a.s. to 0 by virtue of Lemma 6.4 (iii). By dominated convergence it follows that $\mathsf{E}(|X|\mathsf{E}(\mathbf{1}_{\{F_n\}} \mid \mathcal{A}_n)) \downarrow 0$ along this subsequence. Contradiction. QED

Uniform integrability is the necessary property for strengthening convergence in probability to convergence in L^1 (see §5).

Theorem 8.6 *Let $(X_n)_{n \in \mathbf{Z}_+}$ and X be integrable random variables. Then $X_n \xrightarrow{L^1} X$ if and only if*

i) $X_n \xrightarrow{\mathsf{P}} X$ *and*

ii) *the sequence $(X_n)_n$ is uniformly integrable.*

Proof. See Williams PwithM, §13.7. QED

9 Augmentation of a filtration

In the LN we assume 'usual conditions' in constructing right-continuous (super/sub)-martingales. This is related to making the filtration right-continuous. The following lemma gives the construction.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_t, \mathsf{P})$ be a filtered probability space. The usual augmentation is the minimal enlargement that satisfies the usual conditions (note that the usual conditions in the LN are not always the usual ones, but the essence of the construction does not change!).

Let $\mathcal{N}$ be the collection of P-null sets in the P-completion of $\mathcal{F}_\infty = \sigma(\mathcal{F}_t, t \in T)$, and put $\mathcal{G}_t = \cap_{u>t}\sigma(\mathcal{F}_u, \mathcal{N})$. Then $(\Omega, \mathcal{G}_\infty, \{\mathcal{G}_t\}_t, \mathsf{P})$ is the desired usual augmentation.

Lemma 9.1 $\mathcal{G}_t = \sigma(\mathcal{F}_{t_+}, \mathcal{N})$. *Moreover, if $t \geq 0$ and $G \in \mathcal{G}_t$, then there exists $F \in \mathcal{F}_{t_+}$, such that*

$$F \Delta G := (F \setminus G) \cup (G \setminus F) \in \mathcal{N}.$$

Problem 9.1 i) Let $(\Omega, \mathcal{A}, \mathsf{P})$ be a probability space, such that $\mathcal{A}$ is P-complete. Let $\mathcal{N}$ be the collection of P-null sets in $\mathcal{A}$. Let $\mathcal{K}$ be a sub-σ-algebra of $\mathcal{A}$. Prove that

$$\begin{aligned} \sigma(\mathcal{K}, \mathcal{N}) &= \{U \in \mathcal{A} \mid \exists K \in \mathcal{K} \text{ for which } U \Delta K \in \mathcal{N}\} \\ &= \{U \in \mathcal{A} \mid \exists K_1, K_2 \in \mathcal{K}, \text{ with } K_1 \subseteq U \subseteq K_2, \text{ and } K_2 \setminus K_1 \in \mathcal{N}\}. \end{aligned}$$

ii) Prove Lemma 9.1. Hint: use (i). Note that it amounts to proving that $\cap_{u>t}\sigma(\mathcal{F}_u, \mathcal{N}) = \sigma(\cap_{u>t}\mathcal{F}_u, \mathcal{N})$!

The problem of augmentation is that crucial properties of the processes under consideration might change. The next Lemma show that supermartingales with cadlag paths stay supermartingales (with cadlag paths) after augmentation.

Lemma 9.2 *Suppose that X is a supermartingale with cadlag paths relative to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_t, \mathsf{P})$. Then X is also a supermartingale with cadlag paths relative to the usual augmentation $(\Omega, \mathcal{G}_\infty, \{\mathcal{G}_t\}_t, \mathsf{P})$.*

Problem 9.2 Prove Lemma 9.2. You may use the previous exercise.

For Markov process theory it is useful to see that by augmentation of σ-algebras, certain measurability properties do not change intrinsically. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. Let $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra, and let $\mathcal{N}$ be collection of P-null sets.

Lemma 9.3 *Let $\mathcal{G} = \sigma(\mathcal{N}, \mathcal{A})$.*

i) *Let X be an $\mathcal{F}$-measurable, integrable random variable. Then $\mathsf{E}(X \mid \mathcal{G}) = \mathsf{E}(X \mid \mathcal{A})$, P-a.s.*

ii) *Suppose that Z is $\mathcal{G}$-measurable. Then there exist $\mathcal{A}$-measurable random variables Z_1, Z_2 with $Z_1 \leq Z \leq Z_2$.*

Problem 9.3 Prove Lemma 9.3.

10 Elements of functional analysis

Recall that a (real) vector space is called a *normed linear space*, if there exists a *norm* on V, i.e. a map $\|\cdot\| : V \to [0,\infty)$ such that

i) $\|v+w\| \le \|v\| + \|w\|$;

ii) $\|av\| = |a|\|v\|$ for all $a \in \mathbf{R}$ and $v \in V$;

iii) $\|v\| = 0$ iff $v = 0$.

A normed linear space may be regarded as a metric space, the distance between the vectors $v, w \in V$ being given by $\|v-w\|$.

If V, W are two normed linear spaces and $A : V \to W$ is a linear map, we define the norm of the operator A by

$$\|A\| = \sup\{\|Av\| \mid v \in V \text{ and } \|v\| = 1\}.$$

If $\|A\| < \infty$, we call A a *bounded linear transformation* from V to W. Observe that by construction $\|Av\| \le \|A\|\|v\|$ for all $v \in V$. A bounded linear transformation from V to $\mathbf{R}$ is called a *bounded linear functional on V.*

Hahn-Banach theorem We can now state the Hahn-Banach theorem.

Theorem 10.1 *Let W be a linear subspace of a normed linear space V and let A be a bounded linear functional on W. Then A can be extended to a bounded linear functional on V without increasing its norm.*

Proof. See for instance Rudin (1987), pp. 104–107. QED

In Chapter 3 we use the following corollary to the Hahn-Banach theorem.

Corollary 10.2 *Let W be a linear subspace of a normed linear space V. If every bounded linear functional on V that vanishes on W, vanishes on the whole space V, then $\overline{W} = V$, i.e. W is dense in V.*

Proof. Suppose that W is not dense in V. Then there exists $v \in V$, and $\epsilon > 0$, such that $\|v - w\| > \epsilon$ for all $w \in W$. Let W' be the subspace generated by W and v and define a linear functional A on W' by putting $A(w + \lambda v) = \lambda$ for $w \in W$ and $\lambda \in \mathbf{R}$. For $\lambda \neq 0$ $\|w + \lambda v\| = |\lambda|\|v - (-\lambda^{-1}w\| \ge |\lambda|\epsilon$. Hence $|A(w + \lambda v)| = |\lambda| \le \|w + \lambda v\|/\epsilon$. It follows that $\|A\| \le 1/\epsilon$, with A considered as a linear functional on W'. Hence A is bounded on W'. By the Hahn-Banach theorem, A can be extended to a bounded linear functional on V. Since A vanishes on W and $A(v) = 1$, the proof is complete. QED

Riesz representation theorem Let $E \subseteq \mathbf{R}^d$ be an arbitrary set and consider the class $C_0(E)$ of continuous functions on E that become small outside compacta. We endow $C_0(E)$ with the supremum norm

$$\|f\|_\infty = \sup_{x \in E} |f(x)|.$$

This turns $C_0(E)$ into a normed linear space, even a Banach space. The version of the Riesz representation theorem that we consider here, describes the bounded linear functionals on $C_0(E)$.

If μ is a finite Borel measure on E, then clearly the map

$$f \mapsto \int_E f d\mu$$

is a linear functional on $C_0(E)$, with norm equal to $\mu(E)$. The Riesz representation theorem states that every bounded linear functional on $C_0(E)$ can be represented as the difference of two functionals of this type.

Theorem 10.3 *Let A be a bounded linear functional on $C_0(E)$. Then there exist two finite Borel measures μ and ν such that*

$$A(f) = \int_E f d\mu - \int_E f d\nu,$$

for every $f \in C_0(E)$.

Proof. See Rudin (1987), pp. 130–132. QED

Banach-Steinhaus Theorem or Principle of Uniform Boundedness Suppose that V is a Banach space and W a normed linear space. Let T_α, $\alpha \in A$, for some index set A, be a collection of bounded linear transformations from V into W.

Theorem 10.4 *Then either there exists $M < \infty$ such that*

$$\|T_\alpha\| \leq M, \quad \alpha \in A.$$

or

$$\sup_{\alpha \in A} \|T_\alpha f\| = \infty,$$

for all f in some dense G_δ in V.

Proof. See Rudin (1986) pp. 103-104.