1. On probability and notations

We will shortly review some of the concepts from probability theory and we will also introduce some notations we will be using.

Behind all the randomness there is (possibly huge) black box which is called a *probability space*. This is the triple $(\Omega, \mathscr{F}, \mathbf{P})$.

The set Ω is the set of elementary events. What it really is, is irrelavant and most of the time it is just an infinite set Ω. The set $\mathscr{F}$ is a subset of all the subsets of elementary events $\mathscr{P}(\Omega)$ which contains the so called *events*. In practice, one can usually think that all the possible subsets of the set Ω are events, but in general this not usually the case.

In general, there are too many "elementary events" so that all the combinations could be though as events. The following describes what the events are:

1.1. **Defininining properties.**

- the set Ω is a *sure* event
- if A is an event, then $A^C := \Omega \setminus A$ is also an event (so called *complementary event*)
- if $\{ A_k : \ k = 0, 1, 2, \dots \}$ are events then their union
  $$\bigcup_{k=0}^{\infty} A_k = \{A_k \text{ happens for some } k = 0, 1, 2, \dots \}$$
  is an event
- if $\{ A_k : \ k = 0, 1, 2, \dots \}$ are events, then their intersection
  $$\bigcap_{k=0}^{\infty} A_k = \{A_k \text{ happens for every } k = 0, 1, 2, \dots \}$$
  is an event.

Since we will need the concepts of the general measure theory as well, we define

1.2. **Definition.** Let $S \neq \emptyset$ be a set and $\mathscr{G}$ a family of sets with the properties as above when Ω is replaced with the set S and $\mathscr{G}$. This family $\mathscr{G}$ is called *σ-algebra*. The pair $(S, \mathscr{G})$ is called *measurable space*.

The set of all events $\mathscr{F}$ is therefore always a σ-algebra.

1.3. **Definition.** When $\mathscr{F}$ is a σ-algebra and $\mathscr{G} \subset \mathscr{F}$ is a subset such that $\mathscr{G}$ is also a σ-algebra, then the set $\mathscr{G}$ is called *the sub-σ-algebra*.

1.4. **Notation.** If $\mathscr{C} \subset \mathscr{F}$ is *any subset*, then the smallest sub-σ-algebra $\mathscr{G} \supset \mathscr{C}$ is called *the σ-algebra generated by $\mathscr{C}$* and we will denote it by $\sigma(\mathscr{C}) := \mathscr{G}$.

The probability $\mathbf{P}$ assigns to every event a number (the *probability*) from the closed interval $[0,1]$. The probability has the following properties

1.5. **Defininining properties.**

- the probability of the sure event Ω is $\mathbf{P}(\Omega) = 1$ and the probability of the *impossible* event $\emptyset$ is $\mathbf{P}(\emptyset) = 0$.
- if A is an event, the probability of the complementary event $A^C := \Omega \setminus A$ is $\mathbf{P}\left(A^C\right) = 1 - \mathbf{P}(A)$ and
- if $(A_k)_{k\in\mathbb{N}}$ are disjoint events then

$$\mathbf{P}(A_k \text{ happens for some } k \in \mathbb{N}) = \sum_{k\in\mathbb{N}} \mathbf{P}(A_k)$$

We note that *the impossible event* $\emptyset = \Omega^C$ has probability 0 and the probability of the complementary event already follows from the other two conditions. The third condition is called σ*-additivity*.

Therefore, the *measure* on a measurable space is the generalisation of this

1.6. **Definition.** Let $(S, \mathscr{G})$ be a measurable space. The mapping $\mu : \mathscr{G} \to \mathbb{R}_+$ is a *measure*, if $\mu(\emptyset) = 0$ and μ is σ-additive. The triplet $(S, \mathscr{G}, \mu)$ is called *measure space* and if $\mu(S) = 1$, it is called *probability space*.

We still recall the concepts of random variables and conditional probability. First the random variables.

Random variable X is (nearly) arbitrary mapping from probability space to *state space* S, but the nearly means some restrictions. In general, we need to have some regularity in the state space or the regular events. Then the condition is: if $A \subset S$ any regular event, then the set $\{X \in A\}$ has to be an event in probability space Ω.

1.7. **Definition.** Let $(S_j, \mathscr{G}_j)$, for $j = 1, 2$, be two measurable spaces. The function $f : S_1 \to S_2$ is *measurable*, if $\{f \in U\} \in \mathscr{G}_1$ for every $U \in \mathscr{G}_2$.

Usually we know the σ-algebra of the state space and most of the time it will be the *Borel sets*[10], and therefore, we will not usually explicitly say the events in state space. On the other hand, the space we are mapping from will quite likely have multiple different σ-algebras, so we usually express this by saying that f is $\mathscr{G}_1$-measurable.

1.8. **Definition.** Let $(S, \mathcal{G})$ be a measurable space. We say that X is S-valued *random variable*, when the mapping $X : \Omega \to S$ is $\mathscr{F}$-measurable.

[10]In topological space the *Borel sets* $\mathscr{B}(S)$, is the σ-algebra generated by the open sets

Next we need the expectation and conditional expectation. When random variable X has a countable state space $S = \{i_0, i_1, \dots\} \subset \mathbb{R}_+$ then we can define *the expectation* $\mathbf{E}\, X$ *of the random variable* X as a number (or possibly as ∞)

$$\mathbf{E}\, X := \sum_{k=0}^{\infty} i_k \mathbf{P}\,(\, X = i_k\,)\,. \tag{1.9}$$

In general, the state space $S \subset \mathbb{C}$ is an uncountable subset of complex numbers. This is more of *Todennäköisyysteoria* or the *Mitta- ja integraali* course material, since in general expectation is just an integral with respect to the probability measure $\mathbf{P}$.

1.10. **Known fact.** If $f\colon S \to \mathbb{R}_+$ is measurable and bounded, then for each $\varepsilon > 0$ there is a measurable function $f_\varepsilon\colon S \to \mathbb{R}_+$, such that $0 \leq f - f_\varepsilon \leq \varepsilon$ and the state space of f_ε is finite[11].

This simple fact is enough for the definition of the integral with respect to measure.

1.11. **Definition.** Let $(S, \mathscr{G}, \mu)$ be a measure space. Let $f\colon S \to \mathbb{R}_+$ be measurable. Then the *integral of* f *with respect to measure* μ is

$$\int_S f(x)\mu(\,\mathrm{d}x) := \sup\{ \int_S g(x)\mu(\,\mathrm{d}x) : \; g \leq f \text{ and } g \text{ is simple } \}.$$

If Ω is the probability space and X is an $\mathbb{R}_+$-values random variable, then *the expectation of the random variable* X is

$$\mathbf{E}\, X := \int_\Omega X(\omega)\mathbf{P}\,(\,\mathrm{d}\omega\,)$$

By linearity, the integration can be extended to complex valued functions.

We will be using the notation called *Iverson brackets*[12]. Since the indicator functions are used in so many occasions, a simple, clear and consistent notation is needed.

1.12. **Notation.** The Iverson brackets means the map from statements to numbers $\{0, 1\}$:

$$[\,\text{statement}\,] := \begin{cases} 1, & \text{if the statement is true,} \\ 0, & \text{if the statement is not true.} \end{cases}$$

We will generalise this to random events A

[11]A function with finite state space is called *simple*

[12]after Kenneth Eugene Iverson, the source for this notation is Donald Erwin Knuth's *The Art of Computer Progamming, Vol I*

1.13. **Notation.** If A is an event, then $[\,A\,]$ is the random variable such that

$$[\,A\,](\omega) := [\,\omega \in A\,] = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases}$$

We used this notation in the introduction and as another example of this notation

$$\mathbf{E}\,[\,A\,] = 0 \times \mathbf{P}\,(\,[\,A\,] = 0\,) + 1 \times \mathbf{P}\,(\,[\,A\,] = 1\,) = \mathbf{P}\,(\,A\,)\,, \tag{1.14}$$

since $\{[\,A\,] = 1\} = A$.

In this fast review we still need the concepts of *conditional probability and expectation*.

1.15. **Notation.** We will denote the conditional probability of the event A given the event B as

$$\mathbf{P}\,(\,A \,|\, B\,) := \frac{\mathbf{P}\,(\,A \cap B\,)}{\mathbf{P}\,(\,B\,)}$$

This simple form of conditional probability has the same properties as the usual probability, so it leads to *conditional expectation* given an event B.

1.16. **Notation.** Suppose X is a real valued random variable. Then

$$\mathbf{E}\,(X \,|\, B) := \int_{\mathbb{R}} x\,\mathbf{P}\,(\,X \in \mathrm{d}x \,|\, B\,)$$

This can be easily generalised to conditional expectation given a *simple* random variable Y. If

$$Y = \sum_y y[\,Y = y\,]$$

then we can think that

$$[\,Y = y\,]\mathbf{E}\,(X \,|\, Y) := [\,Y = y\,]\mathbf{E}\,(X \,|\, Y = y) \tag{1.17}$$

Summing over the state space gives then

$$\mathbf{E}\,(X \,|\, Y) = \sum_y [\,Y = y\,]\mathbf{E}\,(X \,|\, Y) = \sum_y [\,Y = y\,]\mathbf{E}\,(X \,|\, Y = y)$$

We note that now the conditional expectation is a random variable as well.

Conditioning with respect to finite σ-algebra is still easy, since we can write the equation (1.17) as

$$[\,B\,]\mathbf{E}\,(X \,|\, \mathscr{G}) := [\,B\,]\mathbf{E}\,(X \,|\, B) \tag{1.18}$$

for each event $B \in \mathscr{G}$. Defining the conditional expectation in this way reveals that $\mathbf{E}\,(X \,|\, \mathscr{G})$ is $\mathscr{G}$-measurable.

The general case for general σ-algebra $\mathscr{G}$ is somewhat trickier, since now we have to consider the case $0/0$.

For this, we notice that the finite case can be recovered also from the following formula (1.19)

$$\mathbf{E}\left([\,B\,]\mathbf{E}\,\left(X\,|\,\mathscr{G}\right)\right)=\mathbf{E}\left([\,B\,]\mathbf{E}\,\left(X\,|\,B\right)\right)=\mathbf{E}\left([\,B\,]X\right), \tag{1.19}$$

for every event $B\in\mathscr{G}$. This formula (1.19) follows from the formula (1.18) by taking expectations and this equation has a unique $\mathscr{G}$-measurable solution. This method generalises easily. Therefore, we define

1.20. **Definition.** Let X be a complex value random variable, with $\mathbf{E}\,|X|\,<\infty$ and let $\mathscr{G}\subset\mathscr{F}$ be some sub-σ-algebra. We say that *conditional expectation of the random variable X given $\mathscr{G}$* is a random variable $\mathbf{E}\,(X\,|\,\mathscr{G})$, that is $\mathscr{G}$-measurable, $\mathbf{E}\,|\mathbf{E}\,(X\,|\,\mathscr{G})\,|\,<\infty$ and which solves the equation (1.19) for every $B\in\mathscr{G}$.

The existance and the uniqueness of the conditional expectation is a non-trivial thing. This follows from the Radon–Nikodym's Theorem.

With the conditional probability we can define *independence*.

1.21. **Definition.** We say that a family of events $\{\,A_\lambda:\ \lambda\in I\,\}$ is *independent*, if for each finite subset $\{\lambda_0,\ldots,\lambda_d\}\subset I$ holds

$$\mathbf{P}\left(A_{\lambda_d}\,|\,A_{\lambda_0}A_{\lambda_1}\ldots A_{\lambda_{d-1}}\right)=\mathbf{P}\left(A_{\lambda_d}\right).$$

We say that the family of random variables $\{\,X_\lambda:\ \lambda\in I\,\}$ is *independent*, if for every family $\{\,B_\lambda:\ \lambda\in I\,\}$ in the state space, the corresponding family

$$\{\,\{X_\lambda\in B_\lambda\}:\ \lambda\in I\,\}$$

is independent.

1.22. **Notation.** We say that something holds *almost surely* or *a.s.*, if probability for this to hold is 1.

Later on when we condition with respect to a general random variable X, we use the defintion

$$\mathbf{E}\,(Y\,|\,X)\,:=\mathbf{E}\,(Y\,|\,\sigma(X))\,,$$

where

$$\sigma(X):=\sigma\{\,\{X\in A\}:\ A\text{ is a measurable set }\}$$

is the σ-algebra generated by the random variable X.

We still need to define stochastic process.

1.23. **Definition.** Let $T\neq\emptyset$ and let $(X_t;t\in T)$ be a family of S-valued random variables. We call this family S-valued *stochastic processes*.

As we notice, the set T has no restrictions. However, we will usually assume the following.

1.24. **Assumption.** *Set of times* T is either $T = \alpha\mathbb{N}$ for some $\alpha > 0$ or $T \subset \mathbb{R}$ is an interval. If $t \in T$, then t is called *time instance.* If $T = \alpha\mathbb{N}$, we say the time is *discrete*, otherwise the time is *continuous*.

Sometimes using indices for the time instances is not very readable. Therefore, we use the following convention.

1.25. **Notation.** We may freely denote the random variable at the time instance $t \in T$ either by X_t or $X(t)$.

1.26. **Assumption.** (1) If S is countable, then $\mathscr{S}$ is always $\mathscr{P}(S)$.

(2) In practise, the state space S is of following form

$$S = \begin{cases} \{0, 1, \dots, d\}, \\ \mathbb{N} := \{0, 1, \dots\}, \\ \mathbb{Z} := \{\dots, -2, -1, 0, 1, 2, \dots\}, \\ D \subset \mathbb{R}^d, \text{ kun } d \in \mathbb{N}_+ := \mathbb{N} \setminus \{0\} \end{cases}$$

(3) If $S = D \subset \mathbb{R}^d$, then S is a Borel set (but usually even open or closed) and $\mathscr{S} = \mathscr{B}(S)$.

Therefore, a stochastic process is just an *arbitratry family of time dependent* random variables with state space S.

1.27. *Remark.* As a last remark, if we give a suitable σ-algebra for the set S^T of all mappings from the set T to the set S (the so called product-σ-algebra), then $(X(t))$ is a stochastic process iff $X(\omega) := t \mapsto X(t, \omega)$ is S^T-valued random variable. The value $X(\omega)$ of the random variable from the set of times T to the state space S is usually called either as *the realisation* of the stochastic process. Also we will call it as the *path*.

And it is vital to note, that since the product set S^T contains *all* the functions and note just regular mappings there is no reason to assume that the paths would be continuous or even measurable functions. This is one of the things we need to think in the sequel.