The measure-theoretic parts

Márton Balázs* and Bálint Tóth*

October 13, 2014

This little write-up is part of important foundations of probability that were left out of the unit Probability 1 due to lack of time and prerequisites. Here we include measure-theoretic definitions and a few simple propositions that could not be included in Probability 1. If you are after the well-posedness and rationale of those definitions, you can find them in any course or book on elementary measure theory.

1 Sample space and events

Probability deals with random experiments, the outcomes of which are modelled by a *sample space* Ω. Events are special subsets of the set Ω. When Ω is finite or countably infinite, then we can consider all of its subsets without problems. However, if Ω is uncountable, then we need measure theory to guide us along the subsets. We make some of the most basic definitions, those needed to properly build up probability, of measure theory below.

Definition 1 *Let $\mathcal{F}$ be a set of subsets of Ω, that is, $\mathcal{F}\subseteq\mathcal{P}(\Omega)$. Then $\mathcal{F}$ is called a* σ-algebra, *if*

- *$\Omega\in\mathcal{F}$, that is, the full sample space is included in $\mathcal{F}$,*
- *$\forall E\in\mathcal{F}\ E^c:=\Omega-E\in\mathcal{F}$, that is, $\mathcal{F}$ is closed for taking complements,*
- *for any countably many sets $E_1, E_2, \dots$ in $\mathcal{F}$, $\bigcup_{i=1}^{\infty}E_i\in\mathcal{F}$, that is, $\mathcal{F}$ is closed for taking countable unions.*

Elements of the σ-algebra are usually called measurable sets.

From now on, $\mathcal{F}$ will always denote a σ-algebra of Ω, and its elements E (the measurable sets) will be called *events*. When Ω is finite or countable, we almost always work with $\mathcal{F}=\mathcal{P}(\Omega)=$ the *power set* of Ω, being the set of all subsets of Ω. When $\Omega=\mathbb{R}$, or an interval of thereof, then we usually consider the Borel σ-algebra (the one generated by open intervals). Thus, in this case, not all subsets of the sample space can be considered events, only those *measurable* w.r.t. the Borel σ-algebra. We shall not exploit the theory of σ-algebras in detail, we only use the following simple facts that follow immediately from the definition:

- For all finitely many $E_1, E_2, \dots, E_n$ in $\mathcal{F}$, $\bigcup_{i=1}^{n}E_i\in\mathcal{F}$, thus $\mathcal{F}$ is also closed for taking finite unions.
- It is also closed for taking finite or countably infinite intersections.
- $\emptyset\in\mathcal{F}$, the empty set (or null event) is contained.

Many detailed down-to-earth examples have been provided in Probability 1. We proceed here with the important notion of limits of events.

Definition 2 *A sequence $\{E_i\}_{i=1}^{\infty}$ of sets in $\mathcal{F}$ is called* increasing, *if for all $i\ge 1$ we have $E_i\subseteq E_{i+1}$. In this case $E_i=\bigcup_{j=1}^{i}E_j$, and we define the limit $\lim_{i\to\infty}E_i=\bigcup_{i=1}^{\infty}E_i$. Due to the definition of σ-algebras, this will again be an element of the σ-algebra. Similarly, a sequence $\{E_i\}_{i=1}^{\infty}$ in $\mathcal{F}$ is* decreasing, *if for all $i\ge1$ we have $E_i\supseteq E_{i+1}$. Then $E_i=\bigcap_{j=1}^{i}E_j$, and we define the limit $\lim_{i\to\infty}E_i=\bigcap_{i=1}^{\infty}E_i$, again an element of the σ-algebra.*

*University of Bristol / Budapest University of Technology and Economics

2 Probability

We are now ready for a proper definition of probability.

Definition 3 *A function $\mathbf{P}:\mathcal{F}\to\mathbb{R}$ is called a* probability, *or* probability measure, *if it is*

- **non-negative:**
$$\mathbf{P}\{E\}\ge 0\qquad\forall E\in\mathcal{F},$$
- **countably additive** *(or σ-additive)***:** *for all $E_1, E_2, \dots$ finitely or countably infinitely many* **mutually exclusive** *events,*
$$\mathbf{P}\Big\{\bigcup_i E_i\Big\}=\sum_i\mathbf{P}\{E_i\},$$
- **normed to 1:**
$$\mathbf{P}\{\Omega\}=1.$$

In this case the triplet $(\Omega,\mathcal{F},\mathbf{P})$ is called a probability space.

The first two properties make up for the definition of a *measure*. It is the third property that in particular results in a probability measure. Several elementary propositions and examples have been included in Probability 1. One that was left out from there is the following:

Proposition 4 (continuity of probability) *Let $\{E_i\}_i$ be an increasing or decreasing sequence of events. Then the limit on the left hand-side below exists, and*
$$\lim_{i\to\infty}\mathbf{P}\{E_i\}=\mathbf{P}\{\lim_{i\to\infty}E_i\}.$$

Proof. Consider the case of increasing events first. Let
$$F_1:=E_1,\qquad F_i:=E_i-E_{i-1}\qquad i>1.$$
These events are mutually exclusive,
$$\bigcup_{j=1}^{i}F_j=\bigcup_{j=1}^{i}E_j=E_i\qquad\text{and}\qquad\bigcup_{j=1}^{\infty}F_j=\bigcup_{j=1}^{\infty}E_j=\lim_{j\to\infty}E_j.$$
Therefore
$$\lim_{i\to\infty}\mathbf{P}\{E_i\}=\lim_{i\to\infty}\mathbf{P}\Big\{\bigcup_{j=1}^{i}F_j\Big\}=\lim_{i\to\infty}\sum_{j=1}^{i}\mathbf{P}\{F_j\}=\sum_{j=1}^{\infty}\mathbf{P}\{F_j\}=\mathbf{P}\Big\{\bigcup_{j=1}^{\infty}F_j\Big\}=\mathbf{P}\{\lim_{j\to\infty}E_j\}.$$
For decreasing events $\{E_i^c\}_i$ is increasing, and
$$\lim_{i\to\infty}\mathbf{P}\{E_i\}=1-\lim_{i\to\infty}\mathbf{P}\{E_i^c\}=1-\mathbf{P}\{\lim_{i\to\infty}E_i^c\}=1-\mathbf{P}\Big\{\bigcup_{i=1}^{\infty}E_i^c\Big\}$$
$$=1-\mathbf{P}\Big\{\Big(\bigcap_{i=1}^{\infty}E_i\Big)^c\Big\}=\mathbf{P}\Big\{\bigcap_{i=1}^{\infty}E_i\Big\}=\mathbf{P}\{\lim_{i\to\infty}E_i\}.$$
□

3 Conditional probability

Conditional probability has been motivated and its properties investigated in Probability 1. Here we only repeat its definition and notice one of its simple properties.

Definition 5 *Let F be an event of positive probability. The* conditional probability of the event E given the condition F *is defined as*
$$\mathbf{P}\{E\,|\,F\}:=\frac{\mathbf{P}\{E\cap F\}}{\mathbf{P}\{F\}}.$$

Proposition 6 *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space, and* $F \in \mathcal{F}$ *an event of positive probability. Then* $\mathbf{P}\{\cdot \,|\, F\}$ *is a probability measure:* $\mathbf{P}\{\cdot \,|\, F\} : \mathcal{F} \to \mathbb{R}$ *is a set function that is*

- **non-negative:**
$$\mathbf{P}\{E \,|\, F\} \geq 0 \qquad \forall E \in \mathcal{F},$$

- **countably additive** *(or σ-additive)***:** *for all* $E_1, E_2, \dots$ *finitely or countably infinitely many* **mutually exclusive** *events*
$$\mathbf{P}\Big\{\bigcup_i E_i \,\Big|\, F\Big\} = \sum_i \mathbf{P}\{E_i \,|\, F\},$$

- **normed to 1:**
$$\mathbf{P}\{\Omega \,|\, F\} = 1.$$

Proof. Non-negativity is a trivial consequence of the definition. For countable additivity just notice that $\{E_i \cap F\}_i$ is also a sequence of mutually exclusive events:

$$\mathbf{P}\Big\{\bigcup_i E_i \,\Big|\, F\Big\} = \frac{\mathbf{P}\{(\bigcup_i E_i) \cap F\}}{\mathbf{P}\{F\}} = \frac{\mathbf{P}\{\bigcup_i (E_i \cap F)\}}{\mathbf{P}\{F\}} = \frac{\sum_i \mathbf{P}\{E_i \cap F\}}{\mathbf{P}\{F\}} = \sum_i \mathbf{P}\{E_i \,|\, F\}.$$

The third property is also easy:
$$\mathbf{P}\{\Omega \,|\, F\} = \frac{\mathbf{P}\{\Omega \cap F\}}{\mathbf{P}\{F\}} = \frac{\mathbf{P}\{F\}}{\mathbf{P}\{F\}} = 1.$$

□

Corollary 7 *All properties of* $\mathbf{P}\{\cdot\}$ *remain valid for* $\mathbf{P}\{\cdot \,|\, F\}$, unless one changes the condition F.

4 Random variables

Random variables have also been treated in Probability 1. Here we make a proper definition and explore some properties of the distribution function which needs continuity of probability from above.

Definition 8 *Fix the probability space* $(\Omega, \mathcal{F}, \mathbf{P})$. *A* random variable *is a measurable function* $X : \Omega \to \mathbb{R}$, *where* $\mathbb{R}$ *is endowed with the Borel σ-algebra.*

Often we just write X for a random variable and omit the function notation $X(\omega)$, *Measurability* of a function between two sets endowed with σ algebras means that the pre-image of measurable sets is measurable. In our case, the pre-image of a Borel set in $\mathbb{R}$ is an event in $\mathcal{F}$. This make perfect sense: almost any question we might be interested in regarding a random variable is of the form $X \in B$ with B in the Borel σ-algebra as this latter contains all countable unions and intersections of intervals. Then, the probability of such an event

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

is meaningful as by measurability $X^{-1}\{B\} \in \mathcal{F}$ is an event and as such it does have a probability.

Recall the definition
$$F : \mathbb{R} \to [0, 1] \qquad ; \qquad x \mapsto F(x) = \mathbf{P}\{X \leq x\}$$

of the cumulative distribution function of a random variable. We now give a proper proof of its limits and right-continuity.

Proposition 9 *Let F be the distribution function of the random variable X. Then*

1. *F is non-decreasing,*
2. $\lim_{x\to\infty} F(x) = 1$,
3. $\lim_{x\to-\infty} F(x) = 0$,
4. *F is continuous from the right.*

Vice versa: any function F with the above properties is a cumulative distribution function. There is a sample space and a random variable on it that realises this distribution function.

Proof. We prove here the four properties only.

1. Let $x < y$ be two fixed real numbers. Then as events $\{X \leq x\} \subseteq \{X \leq y\}$, thus $F(x) = \mathbf{P}\{X \leq x\} \leq \mathbf{P}\{X \leq y\} = F(y)$.

2. By the monotonicity of F and Proposition 4,
$$\lim_{x\to\infty} F(x) = \lim_{n\to\infty} F(n) = \lim_{n\to\infty} \mathbf{P}\{X \leq n\} = \mathbf{P}\{\lim_{n\to\infty}\{X \leq n\}\} = \mathbf{P}\Big\{\bigcup_{n>0}\{X \leq n\}\Big\} = \mathbf{P}\{\Omega\} = 1.$$

3. Similarly,
$$\lim_{x\to-\infty} F(x) = \lim_{n\to-\infty} F(n) = \lim_{n\to-\infty} \mathbf{P}\{X \leq n\} = \mathbf{P}\{\lim_{n\to-\infty}\{X \leq n\}\} = \mathbf{P}\Big\{\bigcap_{n<0}\{X \leq n\}\Big\} = \mathbf{P}\{\emptyset\} = 0.$$

4. Let $y \in \mathbb{R}$ be fixed. Then using monotonicity of F and Proposition 4 again,
$$\begin{aligned}\lim_{x\searrow y} F(x) &= \lim_{n\to\infty} F\Big(y + \frac{1}{n}\Big) = \lim_{n\to\infty} \mathbf{P}\Big\{X \leq y + \frac{1}{n}\Big\}\\ &= \mathbf{P}\Big\{\lim_{n\to\infty}\Big\{X \leq y + \frac{1}{n}\Big\}\Big\} = \mathbf{P}\Big\{\bigcap_{n>0}\Big\{X \leq y + \frac{1}{n}\Big\}\Big\} = \mathbf{P}\{X \leq y\} = F(y).\end{aligned}$$

□

We refer to Probability 1 for the definition of discrete and absolutely continuous random variables, and mention Lebesgue's Decomposition Theorem without proof:

Theorem 10 (Lebesgue's Decomposition Theorem) *Let F be a distribution function. Then*

$$F = F_{\textit{absolutely continuous}} + F_{\textit{discrete}} + F_{\textit{singular}},$$

where

- $F_{\textit{absolutely continuous}}$ *is an* absolutely continuous *function, that is, there exists a function $h \geq 0$ with*
$$F_{\textit{absolutely continuous}}(a) = \int_{-\infty}^{a} h(x)\,\mathrm{d}x;$$

- $F_{\textit{discrete}}$ *is a piecewise constant, right-continuous non-decreasing function;*

- $F_{\textit{singular}}$ *is a continuous, non-decreasing function which is singular to the Lebesgue measure. This is to say that its derivative is Lebesgue-almost everywhere zero. An example is the* Cantor function *which is the infinite composition of*
$$G(x) = \begin{cases} \frac{3}{2}x, & 0 \leq x \leq \frac{1}{3},\\ \frac{1}{2}, & \frac{1}{3} \leq x \leq \frac{2}{3},\\ \frac{3}{2}x - \frac{1}{2}, & \frac{2}{3} \leq x \leq 1 \end{cases}$$
with itself.

Notice that in general the above $F_\bullet$ terms are not distribution functions on their own (their limits do not necessarily satisfy Proposition 9).