Handout: Basic Measure Theory

Topics in Probability (Janosch Ortmann)

Autumn 2013

This handout contains some basic concepts and results about measure theory that we will need in the tool. The usage of measure theory in probability is always a balancing act: on the one hand measure theory is a very powerful tool and indispensable for any rigorous treatment of our subject. On the other hand it is very technical, and it is easy to get bogged down in the details while forgetting about the probabilistic intuition.

1 Measure spaces and measurable maps

We begin this collection of facts about measure theory with a few definitions. Most of the concepts and results from this section extend to infinite measures, but we do not need this here.

1.1 Measure spaces

Definition 1.1. Let Ω be a set. A set $\mathcal{F}$ of subsets of Ω is called a *σ-algebra* if

1. $\varnothing \in \mathcal{F}$
2. $A \in \mathcal{F}$ implies $\Omega \setminus A \in \mathcal{F}$
3. If $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The pair $(\Omega, \mathcal{F})$ is also referred to as a *measure space.*

The probabilistic interpretation of this construction is that Ω is the sample space, i.e. the set of possible outcomes, say, of an experiment. Under this point of view $\mathcal{F}$ is the set of events that may take place in the experiment. More precisely, $\mathcal{F}$ represents the events about which we can obtain information. In this sense, considering a larger σ-algebra corresponds to having more information.

Examples 1.2.

1. For any set Ω the *power set* 2^{Ω} is a σ-algebra on Ω. It is finite if Ω is finite and uncountably infinite otherwise.

2. On the other hand, $\{\Omega, \varnothing\}$ is also a σ-algebra on Ω. This is the smallest σ-algebra on any set. The probabilistic interpretation is that the only events are 'something happens' and 'nothing happens'. Unsurprisingly this σ-algebra will not appear very often in our course.

3. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, which could model the rolling of a single ordinary die once. Since this is a finite set there is no problem with allowing every subset of Ω to be an event, so we will consider the measure space $(\Omega, 2^\Omega)$. For example, the set $\{1, 2, 3\}$ would correspond to the event that a number no bigger than 3 comes up. Similarly $\{2, 4, 6\}$ denotes the event that we get an even number.

4. Continuing the example of rolling a die, suppose now that a person (who didn't see the outcome) is only provided with the information whether the number 1 came up. For such an observer, the set $\{1, 2, 3\}$ would not be an event (because she could not determine whether it happened or not). The σ-algebra encoding the information available to her is $\{\varnothing, \Omega, \{1\}, \{2, 3, 4, 5, 6\}\}$.

5. Let $\Omega_N = \{H, T\}^N$, which might represent the outcome of N subsequent coin flips, where H denotes heads and T denotes tails. For example, if $N = 8$ then (H, H, T, T, T, H, H, H) would correspond to two heads, followed by three tails and then another three heads. For any $N \in \mathbb{N}$ the set Ω_N is finite and therefore we can equip it with the power set sigma algebra, $\mathcal{F} = 2^{\Omega_N}$. If we wanted to simulate an unbounded number of coin tosses we would need to be a bit more careful: see exercise 1.12 below.

Definition 1.3. Let Ω be a set and $\mathcal{A}$ a collection of subsets of Ω. The σ-algebra generated by A, denoted by $\sigma(A)$, is the smallest σ-algebra on Ω that contains $\mathcal{A}$. We can construct it explicitly by putting

$$\sigma(\mathcal{A}) = \bigcup \{\mathcal{F} \supset \mathcal{A} \colon \mathcal{F} \text{ is a } \sigma\text{-algebra on } \Omega\}.$$

The set we are taking the intersection over is non-empty (because it contains 2^Ω), therefore this is well-defined.

Remark 1.4. Notice that for any $\mathcal{A} \subset 2^\Omega$ the σ-algebra $\sigma(\mathcal{A})$ is also generated by $\{\Omega \setminus A \colon A \in \mathcal{A}\}$.

Example 1.5. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. Their *product* is given by the measurable space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, where $\mathcal{F}_1 \otimes \mathcal{F}_2$ is generated by the *measurable rectangles*:

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\left(\{A_1 \times A_2 \colon A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}\right).$$

1.2 The Borel σ-algebra

One of the most important sets in probability theory that we will want to equip with a measurable structure is the real line, or more generally $\mathbb{R}^d$. We would naturally require intervals (or, in $\mathbb{R}^d$, rectangles) to be measurable, and it turns out that the smallest σ-algebra that contains these must contain all *Borel sets*:

Definition 1.6. The σ-algebra generated by the open sets in $\mathbb{R}^d$ is called the *Borel σ-algebra* and denoted by $\mathbb{B}^d$. If $d = 1$ we omit the exponent and just write $\mathbb{B}$. Elements of $\mathbb{B}^d$ are also referred to as *Borel measurable sets* or just *Borel sets*.

The Borel σ-algebra is not the largest σ-algebra one can sensibly define on $\mathbb{R}^d$, that distinction usually goes to the σ-algebra of *Lebesgue measurable sets*. However, these are much more difficult to define and do not bring any advantage for the kind of probability theory we will discuss in this course. There are 'much more' Lebesgue measurable sets than Borel sets: the cardinality of the Lebesgue measurable sets is the same as the power set of $\mathbb{R}$, whereas the cardinality of $\mathbb{B}$ is just that of $\mathbb{R}$.

It will be convenient to note that it is not necessary to have all open sets in order to generate $\mathbb{B}$. For example, we can look at various types of intervals[1]. The proof of the following proposition is an exercise.

Proposition 1.7. *The following sets all generate* $\mathbb{B}$*:*

a) The set of closed sets in $\mathbb{R}$*,*

b) $\{(a, b) \colon a < b\}$

c) $\{(a, b] \colon a < b\}$

d) $\{[a, b) \colon a < b\}$

e) $\{[a, b] \colon a < b\}$

f) $\{(-\infty, x) \colon x \in \mathbb{R}\}$

g) $\{(-\infty, x) \colon x \in \mathbb{R}\}$

Notice that there was nothing special about $\mathbb{R}$ or $\mathbb{R}^d$ in the original definition of the Borel sets; for any topological space X we can define by $\mathcal{B}(X)$ the σ-algebra on X generated by the open sets of X. For example we can think about $\mathcal{B}(a, b)$.

We may also sometimes wish to allow the value $+\infty$ to be involved. In such a situation we will be forced to restrict to the positive reals only and we denote by $[0, \infty]$ the set of *extended positive reals*. In particular we can make this into a topological space (by the one-point compactification for those versed in topology) and talk about $\mathcal{B}[0, \infty]$. This will become relevant in the study of random variables, which we sometimes may wish to allow to attain the value ∞.

1.3 Measurable functions and random variables

Recall that for a function $f \colon X \longrightarrow Y$ and $A \subseteq Y$ we denote by $f^{-1}(A)$ the *inverse image* of A under f, i.e. the set

$$f^{-1}(A) = \{x \in X \colon f(x) \in A\}.$$

Definition 1.8. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A function $f \colon \Omega_1 \longrightarrow \Omega_2$ is said to be $\mathcal{F}_1/\mathcal{F}_2$ *measurable* if $f^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{F}_2$.

[1] A similar result holds for the higher dimensional Borel sets, but we will not need this here.

When no confusion arises we often omit reference to the σ-algebras and just talk about measurable maps from Ω_1 to Ω_2. In particular, if $\Omega_2 = \mathbb{R}^d$ it will be understood that, unless otherwise specified, we view it as equipped with the Borel σ-algebra.

We first note that, in order to check measurability, it is enough to check the condition on a generating set. The proof of this result is an exercise:

Proposition 1.9. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces, and suppose that $\mathcal{A} \subseteq \mathcal{F}_2$ generates $\mathcal{F}_2$, that is $\sigma(A) = \mathcal{F}_2$. If $f\colon \Omega_1 \longrightarrow \Omega_2$ satisfies $f^{-1}(B) \in \mathcal{F}_1$ for all $B \in \mathcal{A}$, then f is measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$.*

Note that, trivially, any such function f is measurable if $\mathcal{F}_1 = 2^{\Omega_1}$, whereas in the case $\mathcal{F}_1 = \{\varnothing, \Omega\}$ only the constant functions are measurable. Recall from above that the σ-algebra generated by a set of subsets $\mathcal{A}$ of Ω is the smallest σ-algebra containing the same information as $\mathcal{A}$. Similarly we can define the smallest σ-algebra that encodes the information given by a set of measurable maps:

Definition 1.10. Let $\mathcal{I}$ be some finite or countable indexing set and $X_j\colon \Omega_1 \longrightarrow \Omega_2,\ j \in \mathcal{I}$ be measurable maps. The *σ-algebra generated by* $\{X_j\colon j \in \mathcal{I}\}$ is given by

$$\sigma\left(X_j\colon j \in \mathcal{I}\right) = \sigma\left(\left\{X_j^{-1}(A)\colon A \in \mathcal{F}_2,\ j \in \mathcal{I}\right\}\right)$$

A very important special case (in fact, the most important one for our course) is when $\Omega_2 = \mathbb{R}^d$ and $\mathcal{F}_2 = \mathbb{B}^d$:

Definition 1.11. A measurable map $X\colon \Omega \longrightarrow \mathbb{R}^d$ is called a *random variable.*

Example 1.12. Recall Example 1.2.5 from above but suppose that we do not know in advance how many coin tosses will occur. In this case we can also consider $\Omega_\infty = \{-1, 1\}^{\mathbb{N}}$ (with -1 representing heads and 1 representing tails), which would model an infinite sequence of coin tosses. In this case we need to be a bit more careful how to define our σ algebra. One way to do this would be to define, for each $n \in \mathbb{N}$, the random variable $X_n(\omega) = \omega_n$. That is, X_n denotes the outcome of the n^{th} toss. We can then define $\mathcal{F} = \sigma(X_n\colon n \in \mathbb{N})$, which is a natural σ-algebra to consider. It contains precisely the information we can obtain about our sequence by observing a finite number of individual outcomes.

Sometimes we may wish to allow our random variables to attain the value $+\infty$. As briefly discussed above, the price to pay for this is that we have to restrict our random variables to take positive values only, and a $[0, \infty]$-valued random variable is nothing else than a $\mathcal{F}_1/\mathcal{B}[0, \infty]$ measurable function.

2 Probability measures

Having discussed the concept of a measure space we now introduce the reason for considering such spaces: measures. While we give the general definition for measures, our focus will lie on finite measures since any finite measure can be rescaled to a probability measures. Many results will continue to hold for infinite measures (or at least, 'reasonable' ones), but with additional complications that are not relevant to our course.

Definition 2.1. A (finite) *measure* on a measure space $(\Omega, \mathcal{F})$ is a map $\mu\colon \mathcal{F} \longrightarrow [0, \infty]$ such that

1. $\mu(\varnothing) = 0$,
2. for any disjoint countable collection $\{A_j \in \mathcal{F} \colon j \in \mathbb{N}\}$ of sets in $\mathcal{F}$ we have

$$\mu\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \sum_{k=1}^{\infty} \mu\left(A_k\right).$$

Definition 2.2. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is a set, called the *sample space*, $\mathcal{F}$ is a σ algebra, called the *space of events* and $\mathbb{P}$ is a *probability measure* on $(\Omega, \mathcal{F})$, that is a measure such that $\mathbb{P}(\Omega) = 1$.

We collect some elementary facts about measures. The first property is monotonicity, the second subadditivity and the last two say that we do not lose anything if we approximate a set by a sequence of sets from within or without. We say that $A_n \uparrow A$ if $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ and $A = \bigcup_{n=1}^{\infty} A_n$. Similarly $A_n \downarrow A$ if $A_1 \supseteq A_2 \subseteq A_3 \supseteq \ldots$ and $A = \bigcap_{n=1}^{\infty} A_n$.

Proposition 2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

1. *If $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$*
2. *For any countable collection $\{A_j \in \mathcal{F} \colon j \in \mathbb{N}\}$ of sets in $\mathcal{F}$ we have*

$$\mu\left(\bigcup_{k\in\mathbb{N}} A_k\right) \leq \sum_{k=1}^{\infty} \mu\left(A_k\right).$$

3. *If $A_n \uparrow A$ then $(\mu(A_n) \colon n \in \mathbb{N})$ is an increasing sequence that converges to $\mu(A)$.*
4. *If $A_n \downarrow A$ then $(\mu(A_n) \colon n \in \mathbb{N})$ is a decreasing sequence that converges to $\mu(A)$.*

Proof. These are a good exercise to do if you have never done it before. Otherwise see Theorem 1.1.1 in Durrett. □

Example 2.4. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be probability spaces. We would like to construct a (hopefully unique) probability measure $\mathbb{P}_1 \otimes \mathbb{P}_2$ on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ such that

$$\mathbb{P}_1 \otimes \mathbb{P}_2(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

Thus we have already specified our probability measure on the set of 'measurable rectangles' $\{A_1 \times A_2 \colon A_j \in \mathcal{F}_j\}$ which is easily seen to be a π-system (see below). It now follows from Proposition 2.8 below that there can be at most one such probability measure. Let us assert without proof that there always exist such a probability measure which we call the *product measure* of $\mathbb{P}_1$ and $\mathbb{P}_2$.

2.1 Existence and uniqueness of extension

One of the nice things about measures is that you do not need to specify them on every single set. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability spaces.

Definition 2.5. A collection $\mathcal{A}$ of subsets of Ω such that $\Omega \in \mathcal{A}$ and $A, B \in \mathcal{A}$ implies $A \cap B \in \mathcal{A}$ is called a π-system.

Proposition 2.6. *Let μ and ν be probability measures on $(\Omega, \mathcal{F})$. If $\mathcal{A}$ is a π-system on Ω such that $\sigma(\mathcal{A}) = \mathcal{F}$ and $\mu(A) = \nu(A)$ for every $A \in \mathcal{A}$, then μ and ν agree on $\mathcal{F}$.*

Proof. Notice that the set $\mathcal{B}$ of all sets on which μ and ν agree is a *Dynkin system*, that is closed under differences and countable disjoint unions. Therefore we need to show that the smallest Dynkin system containing a π-system $\mathcal{A}$ is equal to $\sigma(\mathcal{A})$, a result usually referred to as Dynkin's theorem. Similar to σ-algebras there always exists a smallest Dynkin system containing a collection of sets $\mathcal{A}$, denoted $\mathcal{D}(\mathcal{A})$.

First note that a Dynkin system that is also a π-system is a σ algebra, which is left as a (relatively hard) exercise. □

This means that we can a probability measure is uniquely defined by specifying its values on a π-system containing Ω. But can we always construct a probability measure by assigning some numbers to the sets in a π-system? Of course, any such function would need to have certain consistency properties. It turns out that we need a bit more measure theoretic language.

Definition 2.7. A π-system $\mathcal{S}$ on Ω is said to be a *semialgebra* on Ω if it contains $\varnothing$ and Ω and for every $A \in \mathcal{S}$ there exist disjoint $C_1, \ldots, C_n$ with $C_j \in \mathcal{S}$ and $\Omega \setminus A = \bigcup_{j=1}^{n} C_j$.

If $\mathcal{S}$ is a semialgebra on Ω then we can write any set $A \in \sigma(\mathcal{S})$ as the finite disjoint union of elements of $\mathcal{S}$. This decomposition allows us to extend a σ-additive set function on $\mathcal{S}$ to a probability measure on $\sigma(\mathcal{S})$. The proof of this result is a bit involved and we omit it.

Proposition 2.8. *Let $\mathcal{S}$ be a semialgebra on Ω and $\widetilde{\mathbb{P}}\colon \mathcal{S} \longrightarrow [0,1]$ a σ-additive function on $\mathcal{S}$ such that $\widetilde{\mathbb{P}}(\Omega) = 1$. Then there exists a unique probability measure $\mathbb{P}$ on Ω such that*

$$\mathbb{P}(S) = \sum_{j=1}^{k} \widetilde{\mathbb{P}}(S_j)$$

for all $S \in \sigma(\mathcal{S})$ with disjoint decomposition $S = \bigcup_j S_j$ of sets $S_j \in \mathcal{S}$.