Measures and random variables

What measure theory do we need to know? Enough to be able to compute measures of interesting events, and expectations of interesting random variables (we haven't defined these yet).

1 σ-fields

Definition 1.1 (σ-field)**.** *A σ-field on a space Ω is a subset $\mathcal{F}$ of the set of all subsets of Ω, that is closed under countable set operations (unions, intersections and complements). In particular, $\mathcal{F}$ is a σ-field if*

(i) $\emptyset, \Omega \in \mathcal{F}$*;*

(ii) $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$*;*

(iii) $A_1, A_2, \ldots \in \mathcal{F}$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$*.*

Elements of $\mathcal{F}$ are called **measurable sets**.

Example 1.2. The power set (set of all subsets) of Ω, denoted 2^{Ω}, is a σ-field.

Exercise 1. *What are all the σ-fields on the three-element space $\Omega := \{A, B, C\}$? Can you generalize this to any finite set?*

Example 1.3. Let $A \in \mathcal{F}$ if and only if A is countable or A^c is countable. This σ-field separates points, yet we will see it is actually so small as to be trivial for most purposes.

Example 1.4. An important σ-field on any topological space is the **Borel** σ-field, which is defined to be the smallest σ-field containing all the open sets.

We often define a σ-field to be the smallest σ-field containing a family $\mathcal{A}$, which is well defined because the intersection of all σ-fields containing $\mathcal{A}$, if there are any, is a σ-field. We denote this $\sigma(\mathcal{A})$, pronounced, "the σ-field generated by $\mathcal{A}$." Intuitively, this is $\mathcal{A}$ together with countable unions and complements of things in $\mathcal{A}$, unions and complements of those, etc. This is actually a correct recursive definition of $\sigma(\mathcal{A})$ if one remembers to recurse transfinitely. This is spelled out (quite briefly) in G. Folland "Real Analysis", page 39, in the notes on Section 2. For a more detailed discussion, see Section 1F of Moschovakis "Descriptive Set Theory". In the case of the real numbers, the Borel subsets are usually denoted $\mathcal{B}$.

The purpose of σ-fields is twofold. (1) They are needed to define measures and unconditional expectations. (2) They are needed to specify what information is visible to us when we take conditional expectations.

The first of these is more or less a technicality. It would be nice if all sets were measurable, that is, if $\mathcal{F}$ were the set of all subsets of Ω. When Ω is countable, indeed $\mathcal{F}$ is usually taken to be all subsets. However, when Ω is the real line or a subset thereof, and one is dealing with Lebesgue measure (see below), a classical result is that one cannot define a translation invariant measure of the σ-field $\mathcal{F}$ of all subsets of a real interval without contradicting the axiom of choice. In this course we will never run into any non-measurable subsets of reals. Therefore, the word "measurable" will usually appear in parentheses. This indicates that you can omit the word when reading the sentence to yourself, thereby simplifying the cognitive task of understanding the rest of it.

For the second purpose, when we get to conditional expectations, we will be more careful about σ-fields and measurability.

2 Measures

Definition 2.1 (measure space). *A measure space is a pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ is a σ-field on Ω.*

Definition 2.2 (measure). *A* **measure** *on $(\Omega, \mathcal{F})$ is a function μ on $\mathcal{F}$ taking values in the* **extended reals** $\mathbb{R}^* := \mathbb{R} \cup \{+\infty\}$ *such that (1) for all $A \in \mathcal{F}$, $\mu(A) \geq 0 = \mu(\emptyset)$ and (2) for all sequences $\{A_j : j \geq 1\}$ of disjoint sets in $\mathcal{F}$,*

$$\mu\left(\bigcup_j A_j\right) = \sum_j \mu(A_j)\,.$$

If, in addition, $\mu(\Omega) = 1$, then μ is a probability measure. [Usually will then use notation $\mathbb{P}$ rather than μ.]

Example 2.3. For any $x \in \Omega$, let δ_x denote the measure for which $\delta_x(A) = \mathbf{1}_{x\in A}$ where $\mathbf{1}_{x\in A}$ denotes the **indicator function** that is 1 when $x \in A$ and 0 when $x \notin A$. This measure is sometimes referred to as the **point mass at** x.

Example 2.4. Let $\Omega = \{x_1, x_2, \ldots\}$ be countable and let $p_1, p_2, \ldots$ be nonnegative real numbers. Define

$$\mu(A) := \sum_{k: x_k \in A} p_k \,.$$

Then μ is a measure. If $\sum_{k=1}^{\infty} p_k = 1$ then μ is a probability measure. The set of measures on $(\Omega, \mathcal{F})$ is closed under positive linear combinations. In this case $\mu = \sum_k p_k \delta_{x_k}$.

Exercise 2. *Let $\Omega = \{A, B\}$. For each possible σ-field $\mathcal{F}$ on Ω, describe all possible probability measures on $(\Omega, \mathcal{F})$.*

Definition 2.5 (probability space)**.** *A* **probability space** *is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where $(\Omega, \mathcal{F})$ is a measure space and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure.*

Definition 2.6 (measurable function)**.** *A function $f : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is said to be measurable if $A \in \mathcal{S}$ implies $f^{-1}(A) \in \mathcal{F}$.*

Exercise 3. *True or false? If $f : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $g : (S, \mathcal{S}) \to (\Xi, \mathcal{G})$ are measurable then $g \circ f$ is measurable.*

3 Properties and examples

Some general properties of measures are as follows.

1. **monotonicity**. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$.
2. **subadditivity**. If $A_m \in \mathcal{F}$ for each m then
$$\mu\Big(\bigcup_{m=1}^{\infty} A_m\Big) \leq \sum_{m=1}^{\infty} \mu(A_m)\,.$$
3. **continuity from below**. If $A_n \uparrow A$, that is, if $A_1 \subseteq A_2 \subseteq \cdots$ and $\bigcup_{m=1}^{\infty} A_m = A$, with each $A_m \in \mathcal{F}$, then
$$\mu(A) = \lim_{m\to\infty} \mu(A_m)\,.$$

4. **continuity from above**. If $A_n \downarrow A$, that is, if $A_1 \supseteq A_2 \supseteq \cdots$ and $\bigcap_{m=1}^{\infty} A_m = A$, with each $A_m \in \mathcal{F}$, and if μ is a finite measure, then

$$\mu(A) = \lim_{m \to \infty} \mu(A_m)\,.$$

Exercise 4. *Prove these four properties.*

Example 3.1 (uniform measure on a finite space)**.** Let Ω be a finite set and let $\mathcal{F}$ be all subsets of Ω. Define $\mathbb{P}(A) = |A|/|\Omega|$. Then $\mathbb{P}$ is called the **uniform measure** on $(\Omega, \mathcal{F})$. Examples of probability spaces endowed with the uniform measure are an ideal coin flip coin (here $\Omega = \{H, T\}$) and an ideal roll of a die (here $\Omega = [6]$)[1]. Thus, the phrase "Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space corresponding to an ideal six-sided die roll" is taken to mean, let $\Omega = [6], \mathcal{F} = 2^{\Omega}$ and $\mathbb{P}$ be the uniform measure.

How do we get some more nontrivial measures? There are two theorems that will help us build more complicated measures. The first is Carathéodary's extension theorem.

We say $\mathcal{A}$ is an **algebra of sets** if $\mathcal{A}$ is closed under finite unions and complements.

Theorem 3.2 (Carathéodary)**.** *Suppose μ is defined on an algebra $\mathcal{A}$ and is countably additive: for any countable collections of disjoint sets $\{B_k\}$,*

$$\mu(\bigcup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} \mu(B_k)$$

provided that all the sets are in $\mathcal{A}$. Then, provided also that Ω is the countable union of sets in $\mathcal{A}$ of finite measure, there is a unique extension of μ to $\sigma(\mathcal{A})$. □

Existence is the harder part of this theorem and requires going through (in some fashion) the construction of outer measure.

Example 3.3. Lebesgue measure on the Borel subsets of the unit interval. Here, and throughout I will use the notation $\mathbf{m}$ for Lebesgue measure. Construction of Lebesgue measure is often the capstone of a beginning real analysis course. A sketch is as follows. We define $\mathbf{m}$ on partial domains as follows.

1. For the interval from a to b, whether open, half-open or closed, define $\mathbf{m}$ to be equal to $b - a$.

[1]Notation throughout is that $[n]$ denotes the set of integers $\{1, \ldots, n\}$.

2. Extend **m** additively to the algebra $\mathcal{A}$ of finite unions of intervals. This requires proving that **m** is well defined: it has the same value no matter which way we decompose a set S into a finite union of intervals.

3. Show that **m** is countably additive on $\mathcal{A}$. Apply Carath'eodary's extension theorem to show that **m** extends uniquely to $\sigma(\mathcal{A}) = \mathcal{B}$.

Exercise 5. *Use the propeties of measures to evaluate the Lebesgue measure of the set* $\mathbb{Q} \cap [0,1]$ *of rational numbers in the unit interval.*

The classical **Cantor set** is the set of all numbers in $[0,1]$ with no ones in their ternary expansion. It is also characterized as the decreasing intersection of the sets C_n, where C_n is what is left, starting with the unit interval, and at each time step $1, \ldots, n$ removing the middle third of each interval that remains.

Exercise 6. *Use the propeties of measures to evaluate the Lebesgue measure of the the classical Cantor set.*

Define a "Fat Cantor set" C' defined as follows. Let $C_0 := [0,1]$. For each $n \geq 1$, inductively define C'_n to be the result of writing C'_{n-1} as the disjoint union of closed intervals J and removing the middle $1/(n+1)^2$ portion of each such interval J, always removing an open interval so the resulting set remains closed. Let C' denote the decreasing intersection of the sets C'_n.

Exercise 7. *Use the propeties of measures to evaluate the Lebesgue measure of the Fat Cantor set* C'.

4 Products

Finite product spaces and measures

A useful lemma for constructing product measures is:

Theorem 4.1 (Dynkin's π-λ Theorem)**.** *Suppose $\mathcal{A}$ is closed under intersection and is a subset of $\mathcal{B}$. If $\mathcal{B}$ contains the whole space, Ω and is closed under included differences ($G \setminus H$ where $H \subseteq G$) and increasing unions, then $\mathcal{B}$ contains any σ-field containing $\mathcal{A}$.* □

Now we define the product $(\Omega, \mathcal{F}, \mathbb{P})$ of $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ with $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ as follows. Define $\Omega = \Omega_1 \times \Omega_2$ to be the Cartesian product. Define $\mathcal{F}$ to be the product σ-field, the smallest σ-field containing $A \times B$ for every $A \in \mathcal{F}_1, B \in \mathcal{F}_2$. The following theorem them defines the product measure $\mathbb{P}$.

Theorem 4.2. *There is a unique measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that for every $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$,*

$$\mathbb{P}(A \times B) = \mathbb{P}_1(A) \cdot \mathbb{P}_2(B) .$$

Existence is the first of many technical proofs we will skip.

Proof of uniqueness: Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures on $\mathcal{F}$ satisfying $\mathbb{P}(A \times B) = \mathbb{P}'(A \times B) = \mathbb{P}_1(A) \cdot \mathbb{P}_2(B)$ for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. The set $\mathcal{A}$ of generalized rectangles $A \times B$ for $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ is closed under intersection. Let $\mathcal{B}$ be the set of $G \in \mathcal{F}$ such that $\mathbb{P}(G) = \mathbb{P}'(G)$. Clearly $\mathcal{B}$ contains Ω and is closed under included differences. We have also seen that $\mathcal{B}$ is closed under increasing limits. It follows from Dynkin's π-λ Theorem that $\mathcal{B}$ contains the σ-field generated by all rectangles, which is $\mathcal{F}$. $\square$

Infinite product spaces and measures

Let $\Omega = \prod_j \Omega_j$ be the space of infinite sequence with the jth element coming from Ω_j. The infinite product σ-field $\mathcal{F} = \prod_{j=1}^{\infty} \mathcal{F}_j$ is defined to be the smallest σ-field containing all *finitely determined rectangles*, that is, all sets of the form $\prod_{j=1}^{m} A_j \times \prod_{j=m+1}^{\infty} \Omega_j$.

For technical reasons we will never speak of again, when constructing infinite products we will restrict to **nice** spaces, meaning spaces that are isomorphic to the $([0,1], \mathcal{B})$ or some subset thereof in the category of measure spaces. This space is isomorphic to a countable power of itself, and in fact most spaces we would ever dream of are nice.

The following result allows us to construct infinite product measures.

Theorem 4.3 (Kolmogorov's extension theorem)**.** *Given nice measure spaces $(\Omega_j, \mathcal{F}_j)$, for $j = 1, 2, 3, \ldots$, suppose for each $N \geq 1$ a measure μ_N is defined on $\prod_{j=1}^{N} \Omega_j$ in such a way that for any $j < N$ and any $A \in \prod_{i=1}^{j} \mathcal{F}_j$,*

$$\mu_N \left(A \times \prod_{i=j+1}^{N} \Omega_i \right) = \mu_j(A) .$$

Then there is a measure μ on $\prod_{j=1}^{\infty} \mathcal{F}_j$ such that for all j and all $A \in \prod_{i=1}^{j} \mathcal{F}_j$,

$$\mu \left(A \times \prod_{i=j+1}^{\infty} \Omega_i \right) = \mu_j(A) .$$

Exercise 8. *Use Dynkin's π-λ theorem to prove that the infinite product measure μ is uniquely defined on the product measure space $\prod_{j=1}^{\infty}(\Omega_j, \mathcal{F}_j)$.*

Here are some very important examples of infinite product measures. In the next section we will discuss interpretations of these as coin flips, random variables, etc. The first one will be used to model infinitely many coin flips. We will use zeros and ones rather than heads and tails, so that we can also view this as infinite binary sequences, or paths in an infinite binary tree.

Example 4.4 (binary sequences)**.** Let $\Omega_0 := \{0, 1\}, \mathcal{F}_0 := 2^{\Omega_0}$ and let $\mathbb{P}_0 := \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ be the uniform measure on the set Ω_0. Let $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_0, \mathcal{F}_0, \mathbb{P}_0)^\infty$. Noting that the finite product $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)^N$ is precisely uniform over the 2^N binary sequences of length N, we can reasonably interpret this infinite product measure as "uniform measure on infinite binary sequences". Here and subsequently, we use quotation marks to indicate non-mathematical interpretations.

Example 4.5 (the infinite dimensional hypercube)**.** Let $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0) := ([0, 1], \mathcal{B}, \mathbf{m})$. Let $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega_0, \mathcal{F}_0, \mathbb{P}_0)^\infty$. Noting that $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)^N$ is N-dimensional Lebesgue measure on the hypercube $[0, 1]^N$, we can consider this measure $\mathbb{P}$ to be "Lebesgue measure on $[0, 1]^\infty$".

5 Random variables and probability models

We repeat two definitions from before and add another.

Definition 5.1. *A* **random variable** *is a measurable map from Ω to $\mathbb{R}$. Sometimes we generalize this to S-valued random variables, which are measurable maps from $(\Omega, \mathcal{F})$ to a space $(S, \mathcal{S})$. An* **event** *is a measurable subset on Ω, that is, an element of $\mathcal{F}$. The* ***indicator function*** $\mathbf{1}_A$ *of an event A is the function that is 1 if $\omega \in A$ and zero otherwise.*

Probability modeling is about associating probability spaces and random variables to real life situations described in words. The model must contain enough random variables to correspond to all quantities you need to discuss, having all the properties suggested by the verbal description. There are always many ways to do this. The choice is dictated by transparency (what's easiest to understand), tractability (what gives you the means to compute) and generality (which models you can keep around and use for future scenarios).

Example 5.2 (a coin flip)**.** The most economical way to model a coin flip is with a two-element space we have seen before: $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\{H, T\}, 2^\Omega, \frac{1}{2}\delta_H + \frac{1}{2}\delta_T\right)$. The function f mapping H to 1 and T to 0 is a random variable that counts the number of heads. The event $\{H\}$ is the event whose non-mathematical interpretation would be "when you flipped the coin you got HEADS". Up to symmetries, these are the only nontrivial random variable and event you can define on this space. The choice of this space to model a coin flip scores high on transparency – the notation makes it blindingly obvious what's going on – but low on generalizability: we can't extend this model to anything more than a single coin flip.

We won't discuss the philosophy of probability in this course, except to say that we interpret probability spaces in the **multiple universe** paradigm. Each $\omega \in \Omega$ represents a possible state of the universe and the value $f(\omega)$ of a random variable $f : \Omega \to \mathbb{R}$ is the way f turned out in this particular universe. Thus, the HEADS-count random variable f in the example turns out to be 1 in the H state of the universe and 0 in the T state.

A commonly used notation for events evokes this interpretation. An event G usually corresponds to a set of $\omega \in \Omega$ for which the random variables that have been defined have certain properties or relations. [Physicists: think of the random variables as observables and events as defined in terms of observables.] Therefore, an event $G \in \mathcal{F}$ is usually of the form $\{\omega : X(\omega) \in A\}$. We abbreivate this to the notation $\{X \in A\}$. The curly braces are no longer set brackets, rather they are an abberviation for "the set of all states of the universe such that ...".

Laws

Let $X : \Omega \to \mathbb{R}$ be a random variable. The *distribution* or *law* of X is the measure on $\mathbb{R}$ which is the "push-forward of $\mathbb{P}$ by X". Specifically, the law of X is the measure μ defined by $\mu(A) = \mathbb{P}(X^{-1}(A))$. Most of the time, when we define random variables on a space Ω, we care only about their laws, including the joint laws, which are the push-forwards of $\mathbb{P}$ under the map $(X_1, \dots X_k : \Omega \to \mathbb{R}^k)$.

Example 5.3. Let Ω be $[0,1)$ with Lebesgue measure and let X_1 be the first binary bit function, so X_1 is 1 on $[1/2,1)$ and 0 on $[0,1/2)$. Then μ is the measure giving mass 1/2 to the point 0 and 1/2 to the point 1.

This is another probability space that can be used to model a single coin flip. The random variable X_1 counts the number of heads. It is less transparent in this model than in the previous model that we have modeled an ideal coin flip because there's a lot of irrelevant information in the model: if we observe HEADS, we don't really care which element of $[1/2,1)$ is the true state of the universe. The law of X_1 in this model is the same as the law of f in the previous model. Because our mission is only to model the HEADS count of a single coin flip, the two are equally good models.

The second model, though less transparent, is more generalizable. Let $X_2 = \mathbf{1}_{[1/4,1/2)\cup[3/4,1)}$ be the second binary bit. Then the pair (X_1, X_2) can be used to model two independent coin flips, with X_1 telling us whether the first coin was HEADS and X_2 telling us about the second. We will discuss independence in detail soon, but for now the point is, the first space is just big enough to model a single coin flip whereas the second space has room for further experiments. In fact, letting X_n denote the n^{th} binary bit, the second probability space is big enough to model an entire infinite sequence of independent fair coin flips. As a private exercise, you might try to verify that the law of $(X_1, X_2, \ldots)$ on this probability space is the product measure $\left(\frac{1}{2}\delta_1 + \frac{1}{2}\delta_0\right)^{\infty}$ constructed in Example 4.4.

Exercise 9. *Make the most transparent probability model you can for the following scenario.*

> *Two independent 6-sided dice are rolled. Their sum is a random variable, X. We wish to compute the probability of the event $\{X \text{ is prime}\}$. Your model should define $\Omega, \mathcal{F}, \mathbb{P}$ and any random variables you will use to arrive at an event interpretable as "the sum of the two dice produced a prime number".*

Equality in distribution

If X and Y have the same law, we write $X \stackrel{\mathcal{D}}{=} Y$. Note that it is not necessary for X and Y to be defined on the same probability space in order for this to make sense. It is simply a fact about their laws. By contrast, the statement $X = Y$ means that X and Y are random variables on the same probability space and that their values are the same at each point of Ω. To make things more confusing, one also often considers the event $\{X = Y\}$. This is the subset of Ω at which the functions X and Y agree. So for example, with the variables X_1 and X_2 above, we have $X_1 \stackrel{\mathcal{D}}{=} X_2$, but not $X_1 = X_2$, while the event $\{X_1 = X_2\}$ (which is shorthand for $\{\omega \in \Omega : X_1(\omega) = X_2(\omega)\}$) is the set $[0, 1/4) \cup [3/4, 1)$ (both variables being 0 on the first interval and 1 on the second). Thus we may compute $\mathbb{P}(X_1 = X_2) = 1/2$.

Exercise 10. *Consider the uniform measure on the symmetric group, which is Example 3.1 where $\Omega = S_n$. Let $M(\pi)$ count the number of (circular) local maxima of π, that is, the number of $j \in [n]$ such that $\pi(j) > \max\{\pi(j+1), \pi(j-1)\}$ where $j+1$ and $j-1$ are computed modulo n (in other words, 1 and n are neighbors). Analogously, let $m(\pi)$ denote the number of (circular) local minima. For each random variable listed here, state whether it is equal to M, equal in distribution to M but not equal to M, or not equal in distribution.*

(a) $M(\pi^{-1})$

(b) $M(\pi^2)$

(c) $M(\pi \cdot (123 \cdots n))$ *where* $(123 \cdots n)$ *denotes the* n *cycle*

(d) $M(\pi \cdot (23))$ *where* (23) *denotes the transposition that transposes 2 and 3*

(e) $m(\pi)$*, defined analogously to* M *except counting minima instead of maxima.*

Distribution functions

Say that a function $F : \mathbb{R} \to [0, 1]$ is a *distribution function* if is is nondecreasing, has limits 0 and 1 at $-\infty$ and $+\infty$ respectively, and is right-continuous: $F(x) = \lim_{y \downarrow x} F(y)$. If X is a random variable, its distribution function is the function F_X defined by $F_X(x) = \mathbb{P}(X \leq x)$. Note that F_X really depends only on the law of X. For example, let X_2 be the second binary bit: $X_2 = 1$ on $[1/4, 1/2) \cup [3/4, 1)$ and 0 otherwise. Then X_2 also has the fair coin-flip law giving mass 1/2 to 0 and 1/2 to 1, so $F_{X_1} = F_{X_2} = 0$ for $x < 0$, 1/2 for $x \in [0, 1)$ and 1 for $x \geq 1$.

It is a theorem (Theorem 1.2.2 of Durrett) that the map $X \mapsto F_X$ induces a one-to-one correspondence between laws and distribution functions. When a distribution function F is continuous, we say the corresponding law is *non-atomic*, since it means that $\mu(\{x\}) = 0$

for each x (μ has no "atoms"). If F is differentiable, with $F'(x) = g(x)$, we say that μ is **absolutely continuous** with respect to Lebesgue measure and has **density** g. For by definition of the correspondence, in this case,

$$\mu((a,b]) = F(b) - F(a) = \int_a^b g(x)\,dx\,.$$

This is an Riemann integral from freshman calculus. When we have constructed the Lebesgue integral, it will follow that for every measurable set A,

$$\mu(A) = \int_A g(x)\,dx := \int \mathbf{1}_A g(x)\,dx\,.$$

In order for g to be a density for some probability distribution, it is necessary and sufficient that $g \geq 0$ and $\int g(x)\,dx = 1$. Undergraduate probability courses define continuous random variables via their densities. As you will see, defining them via CDF is more general: all random variables have CDF's, whereas only some have densities.

Exercise 11. *Prove the (hopefully already familiar) change of variables formula: if X is a random variable with density f and $\phi : \mathbb{R} \to \mathbb{R}$ is increasing, differentiable, with differentiable inverse, then $\phi(X)$ is a random variable with density g, where*

$$g(x) = \frac{f(\phi^{-1}(x))}{\phi'(\phi^{-1}(x))} \qquad \textit{or equivalently} \qquad g(\phi(x)) = \frac{f(x)}{\phi'(x)}\,. \tag{5.1}$$

In the prerequisites, we assumed familiarity with several families of random variables: geometric, binomial, Poisson, uniform, exponential and normal. Here are some further examples of probability distributions on the real numbers and their CDF's.

Example 5.4 (point mass at 0)**.** Define $\mu(S) = 1$ if S contains the point 0 and $\mu(S) = 0$ otherwise. Thus all the mass of μ is on the point 0. The CDF for this probability distribution is the Heaviside function, defined by $F(x) = 1$ for $x \geq 0$ and $F(x) = 0$ for $x < 0$. The measure μ is often denoted by δ_0. In general, δ_x denotes the pointmass at x. Another notation for the Heaviside function is $\mathbf{1}_{[0,\infty)}$. N.B.: the indicator function notation is used for set membership even when it is not defining a random variable.

Example 5.5 (an atomic measure supported on the rationals, whose CDF has dense discontinuities)**.** Let $q_1, q_2, \ldots$ be an enumeration of the rational numbers and let μ be the measure $\sum_{n=1}^{\infty} 2^{-n} \delta_{q_n}$. Pause for a minute to parse this definition! This multiplication and summing is at the level of measures, not at the level of the points q_n. The interpretation is that μ gives mass 1/2 to q_1, mass 1/4 to q_2, and so on. Its CDF is defined by

$$F(x) = \sum_{n=1}^{\infty} 2^{-n} \mathbf{1}_{q_n \le x} \, .$$

Undergraduate probability courses typically consider two types of probability measures: **atomic measures** which can be written as weighted sums of pointmass measures δ_x, and **absoultely continuous** measures which have densities. It is possible for a measure to be neither: the CDF can be continuous without being differentiable. Here is one such example.

Example 5.6 (uniform distribution on the Cantor set)**.** Recall the classical Cantor set from Exercise 6. We will define a function F and then show that it may be interpreted as the CDF for a distribution that is uniform on the Cantor set, C. First we define F to be zero on $(-\infty, 0]$ and one on $[1, \infty)$. Next, if the ternary expansion of a point $x \in [0,1] \setminus C$ is given by $a_1 a_2 a_3 \cdots$, and if j is the least value such that $a_j = 1$, then $F(x)$ is defined to be the dyadic rational number whose binary expansion is $a_1' a_2' \cdots a_{j-1}' 1$ where $a_i' = a_i/2$. It is not hard to see that F is nondecreasing and continuous. It follows that F extends to a continuous function on the whole interval $[0,1]$.

Let us see why this is the CDF for the "uniform distribution on C". Clearly F is constant on each of the "middle third" intervals $a_1 \cdots a_j$ where indicates that any continuation is allowed. The interpretation is that the measure μ whose CDF is F gives no mass to any of the middle third intervals. The complement of C is the countable union of these intervals, hence $\mu(C^c) = 0$, in other words, μ is supported on C. As for the uniformity, we see that $F(x) = 1/2$ for any $x \in [1/3, 2/3]$, so μ give mass 1/2 to $C \cap [0, 1/3]$. Likewise, we see that μ gives measure 2^{-n} to each of the 2^n subsets of C given by specifying the first n ternary digits. This is why we interpret μ as uniform on C.

Exercise 12. *Draw the CDF for the uniform Cantor set measure.*

Random number generation

You are not going to have a special random number generator for each distribution you need to generate samples from. Instead, most likely, you have a $U[0,1]$ generator. How do we get from this to generating the X we want with distribution function F? Let U be uniform on $[0,1]$ and let F^{-1} denote the inverse,

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}\,.$$

Claim: $X := F^{-1}(U)$ has $F_X = F$. Proof: the event that $F^{-1}(U) \leq x$ is the same as the event $U \leq F(x)$. [why? if $U \leq F(x)$ then x is in the inf defining $F^{-1}(U)$, so $F^{-1}(U) \leq x$, while conversely, if things arbitrarily close to x from above are in the inf, then $F(x) \geq U$ as well, by right continuity of F.] Thus $P(X \leq x) = P(U \leq F(x)) = F(x)$.

Example 5.7. If X is exponential then $F_X = 1 - \exp(-t)$ so $F_X^{-1} = -\log(1-t)$. Therefore, if U denotes a random variable uniform on $[0,1]$ then $-\log(1-U)$ is exponentially distributed.

Example 5.8. Suppose we wish to generate a random variable whose law is uniform on the Cantor set. Letting F denote the Cantor CDF from Example 5.6, we see that $F(02000220...)$ in ternary is 01000110... in binary, so F^{-1} replaces binary zeros and ones with ternary zeros and twos respectively. If U is uniform on $[0,1]$ then its binary digits are IID fair coin flips. Therefore, we may generate a random variable X with uniform Cantor distribution by generating a sequence of IID fair coin flips and setting the n^{th} ternary digit of X to be 0 if the n^{th} coin is HEADS and 2 if the n^{th} coin is TAILS.

Example 5.9 (discontinuities at the rationals)**.** Let U be a random variable uniform on $[0,1]$ and let X be the least real number so that $\sum_{y \leq x} q(y)$ is at least U. Then X has the distribution of Example 5.5.

I'm not a computer scientist but here's a little advice on using random number generators when running simulations that I picked up along the way. Hash functions such as the Rinjdael cipher tend to produce better results. They are often faster and they save space because they are random access: you don't have to store the results of your computations in some big array because you can recompute on the fly. If you do nevertheless use an off-the-shelf random number generator, make sure it's one that passes benchmark tests (some of the linear recursions with small numbers of bits do not) and that it does not use physical random input such as clock data. Also, make a habit of recording what seeds produced what data so you can go back and reproduce if you have to.

6 Independence

We will be talking about the joint distributions of collections of random variables. Just as a random variable X has a law (with CDF F_X), a finite or infinite collection of random variables $\{X_1, X_2, \ldots\}$ has a joint law $\mathcal{L}$ which is a probability distribution on $\mathbb{R}^n$ if the collection has n variables, $n \le \infty$. Formally, this is defined by

$$\mathcal{L}(A) = \mathbb{P}(\{\omega : (X_1(\omega), X_2(\omega), \ldots) \in A\}) \,.$$

Definition 6.1 (independence)**.**

1. **Independence for two events**: *Events A and B are independent iff $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. No implications one way or the other about causality!*

2. **Independence for two random variables**: *Random variables X and Y are independent iff $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ for all A, B.*

3. **Joint independence**: *Collections of n variables (definition of "jointly independent") by requiring*

$$\mathbb{P}(X_j \in A_j, j = 1, \ldots, n) = \prod_{j=1}^{n} \mathbb{P}(X_j \in A_j)$$

for all sequences $A_1, \ldots, A_n$ of measurable sets.

Independence of events is the same as independence of their indicator random variables, so there is no need for a separate definition for joint independence of events.

An easy consequence of Dynkin's π-λ theorem is that it is sufficient, when establishing joint independence of real random variables, to show that

$$\mathbb{P}(X_1 \le x_1, \ldots, X_n \le x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \le x_j)$$

for all real n-tuples $x_1, \ldots, x_n$. Equivalently, in the obvious notation,

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \,.$$

If the n-vector $(X_1, \ldots, X_n)$ has a density over $\mathbb{R}^n$, this is also the same as

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \,.$$

A very useful lemma (you can prove it as an exercise if you want) is:

Lemma 6.2. *Functions of disjoint groups of independent variables are independent. More formally, if $X_{i,j}$ are jointly independent and $f_i : \mathbb{R}^{m_i} \to \mathbb{R}$ are (measurable) functions, then $\{f_i(X_{i,1}, \ldots, X_{i,m_i})\}$ are independent. Thus for example if X, Y, Z, W are independent, then $X + Y$ is independent of $Z \cdot W$.*

Independence is one reason that product measures are so important. The proof of the following proposition is just to write out definitions.

Proposition 6.3. *The variables $X_1, \ldots, X_n$ are jointly independent if and only if the law of the random vector $(X_1, \ldots, X_n)$ is $\mathcal{L}(X_1) \times \cdots \times \mathcal{L}(X_n)$.* □

Exercise 13. *On the probability space $([0,1], \mathcal{B}, \mathbf{m})$, consider the events $G := [0, 1/2)$, $H := [0, 1/4) \cup [1/2, 3/4)$ and $K := [0, 1/4) \cup [3/4, 1)$. Are these three events independent?*

If X is a random variable on $(\Omega, \mathcal{F})$, we extend the notation $\sigma(\mathcal{A})$ by defining $\sigma(X)$ to be the σ-field generated by all sets of the form $X^{-1}(B)$, where B is a Borel set of real numbers. This is the same as the σ-field generated by sets of the form $X^{-1}(J)$ where J is an interval.

Exercise 14. *Prove that*

$$\sigma(\{X^{-1}(B) : B \text{ is a Borel set}\}) = \sigma(\{X^{-1}(B) : B \text{ is an interval}\}) . \qquad (*)$$

Hints:
1. Equality of sets is the same as $\subseteq$ and $\supseteq$ together; which one is nontrivial?
2. What do you know about the set of A such that $X^{-1}(A)$ is an element of the right-hand side of (*) ?