

### Coin tossing space

Think of a *coin toss* as a random choice from the two element set  $\{0, 1\}$ . Thus the set  $\{0, 1\}^n$  represents the set of possible outcomes of  $n$  coin tosses, and  $\Omega := \{0, 1\}^{\mathbb{N}}$ , consisting of all sequences  $(t_n)_{n \in \mathbb{N}}$ , represents the set of possible outcomes of tossing a coin infinitely many times.

If the coin is unbiased and the tosses are independent, each of the  $2^n$  outcomes in  $\{0, 1\}^n$  is equally probable. So  $\{0, 1\}^n$  is equipped with a probability measure  $\mu_n$ , which is  $2^{-n}$  times counting measure on  $\{0, 1\}^n$ . (All subsets of  $\{0, 1\}^n$  are measurable.)

---

### Compatible probability measures

Assume that  $m < n$ . Tossing a coin  $n$  times, and then ignoring all but the first  $m$  of them, is represented by a *projection map*  $\pi_{mn}: \{0, 1\}^n \rightarrow \{0, 1\}^m$ :

$$\pi_{mn}(t_1, \dots, t_n) = (t_1, \dots, t_m).$$

An event depending on  $(t_1, \dots, t_m)$  is a set  $E \subseteq \{0, 1\}^m$ . It has probability  $\mu_m(E)$ . The “same” event considered as a subset of  $\{0, 1\}^n$  is  $\pi_{mn}^{-1}(E)$ . It does in fact have the same probability, in  $\{0, 1\}^n$ , as  $E$  does, in  $\{0, 1\}^m$ :

$$\mu_n(\pi_{mn}^{-1}(E)) = \mu_m(E), \quad E \subseteq \{0, 1\}^m. \quad (1)$$

Think of this as a compatibility result: All the different probability measures  $\mu_n$  give the same result for those events they can measure.

(*Exercise:* Show this. *Hint:* Show that for each  $s \in E$  there are  $2^{n-m}$  distinct sequences  $t \in \pi_{mn}^{-1}(E)$  with  $\pi_{mn}(t) = s$ .)

---

### Lots and lots of projections

All these projection maps  $\pi_{nm}$  are compatible in the sense that

$$\pi_{km} \circ \pi_{mn} = \pi_{kn}, \quad k < m < n.$$

It is helpful to picture the finite coin tossing spaces as an infinite sequence:

$$\{0, 1\} \leftarrow \{0, 1\}^2 \leftarrow \{0, 1\}^3 \leftarrow \dots \leftarrow \{0, 1\}^n \leftarrow \{0, 1\}^{n+1} \leftarrow \dots$$

where the arrows shown are  $\pi_{12}, \pi_{23}, \dots, \pi_{n, n+1}$ . At the far right we can put the infinite coin tossing space  $\Omega$ . (Category theorists call that the *projective limit* of the above diagram, but never mind that.) There is also a projection map  $\pi_n: \Omega \rightarrow \{0, 1\}^n$ , given by  $\pi_n(t_1, t_2, \dots) = (t_1, \dots, t_n)$ . Clearly, the compatibility of projections extends to this one:

$$\pi_{mn} \circ \pi_n = \pi_m, \quad m < n. \quad (2)$$

## Finitely determined events and their probabilities

Some events in  $\Omega$  depend only on a finite number of coin tosses. (By this I mean that to determine whether  $t \in \Omega$  belongs to some event  $E$ , you only need to look at  $t_1, \dots, t_n$  for some  $n$ , and that  $n$  is independent of  $t$ .) If  $E \subseteq \{0, 1\}^{\mathbb{N}}$  then

$$\pi_n^{-1}(E) = \{t \in \Omega: (t_1, \dots, t_n) \in E\}$$

is such an event, and I expect you can convince yourself that every such “finitely determined” event has this form. (If not, consider that a *definition*.) Define

$$\mathcal{A}_n = \{\pi_n^{-1}(E) : E \subseteq \{0, 1\}^n\}, \quad n \in \mathbb{N}.$$

Then  $\mathcal{A}_n$  is a  $\sigma$ -algebra of subsets of  $\Omega$  for each  $n \in \mathbb{N}$ , and moreover  $\mathcal{A}_m \subset \mathcal{A}_n$  when  $m < n$  (for the equality  $\pi_{mn} \circ \pi_n = \pi_m$  implies  $\pi_m^{-1}(E) = \pi_n^{-1}(\pi_{nm}^{-1}(E))$  when  $E \subseteq \{0, 1\}^m$ ).

Let

$$\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{A}_n.$$

It is not hard to show that, since  $(\mathcal{A}_n)$  is an increasing sequence of algebras, the union  $\mathcal{C}$  is also an algebra of sets (do it!). We can define a function  $\iota$  on  $\mathcal{C}$  by

$$\iota(\pi_n^{-1}(E)) = \mu_n(E), \quad E \subseteq \{0, 1\}^n.$$

(*Exercise:* Show that this is well defined. *Hint:* Use (1) and (2).)

## Extending the measure

The function  $\iota$  represents the probability of *any* finitely determined event. But what about other events? How can we discuss the probability that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n t_k = \frac{1}{2},$$

for example? The set of  $t \in \Omega$  satisfying this equality is clearly not finitely determined.

What is needed is to extend  $\iota$  to a measure. It is defined on the algebra  $\mathcal{C}$ , which is *not* a  $\sigma$ -algebra. But  $\mathcal{C}$  is a semialgebra, so in order to show that we can extend  $\iota$  to a measure, we need to show that  $\iota(\emptyset) = 0$ , that  $\iota$  is finitely additive, and the appropriate subadditivity condition.

For the first property,  $\iota(\emptyset) = \mu_n(\pi_n^{-1}(\emptyset)) = \mu_n(\emptyset) = 0$ .

For the second, if  $F_1, \dots, F_m \in \mathcal{C}$  are pairwise disjoint then there is some  $n$  so that  $F_k \in \mathcal{A}_n$  for  $k = 1, \dots, m$ . Write  $F_k = \pi_n^{-1}(E_k)$ . Then the sets  $E_k$  must be pairwise disjoint, and

$$\iota\left(\bigsqcup_{k=1}^m F_k\right) = \mu_n\left(\bigsqcup_{k=1}^m E_k\right) = \sum_{k=1}^m \mu_n(E_k) = \sum_{k=1}^m \iota(F_k).$$

Finally, assume that  $F, F_k \in \mathcal{C}$  with  $F \subseteq \bigcup_{k \in \mathbb{N}} F_k$ . I *claim* that a finite number of the sets  $F_k$  will cover  $F$ , i.e.,  $F \subseteq F_1 \cup \dots \cup F_n$  for some  $n$ . If this is true, then

$$\iota(F) \leq \iota\left(\bigcup_{k=1}^n F_k\right) \leq \sum_{k=1}^n \iota(F_k) \leq \sum_{k=1}^{\infty} \iota(F_k),$$

which is the subadditivity condition we need to show.

To prove the claim, assume instead it is false. That is, we assume that each set  $A_n$ , defined by  $A_n = F \setminus (F_1 \cup \dots \cup F_n)$ , is nonempty. Note that  $A_n \in \mathcal{A}$ , and  $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ . We now define a sequence  $(u_k)$  in  $\{0, 1\}$  by induction.

First, since  $\Omega = \pi_1^{-1}\{0\} \cup \pi_1^{-1}\{1\}$ , we find  $A_n = (A_n \cap \pi_1^{-1}\{0\}) \cup (A_n \cap \pi_1^{-1}\{1\})$ , so we can select  $u_1 \in \{0, 1\}$  so that  $A_n \cap \pi_1^{-1}\{u_1\} \neq \emptyset$  for infinitely many  $n$ , and therefore for all  $n \in \mathbb{N}$ , since  $(A_n)$  is a decreasing sequence of sets.

Next, we get  $A_n \cap \pi_1^{-1}\{u_1\} = (A_n \cap \pi_2^{-1}\{u_1, 0\}) \cup (A_n \cap \pi_2^{-1}\{u_1, 1\})$ , so we can select  $u_2 \in \{0, 1\}$  so that  $A_n \cap \pi_2^{-1}\{u_1, u_2\} \neq \emptyset$  for infinitely many  $n$ , and therefore for all  $n \in \mathbb{N}$ .

Continuing this procedure, we end up with a sequence  $u = (u_k)_{k \in \mathbb{N}}$  with the property that, for each  $k$ ,  $A_n \cap \pi_k^{-1}\{u_1, \dots, u_k\} \neq \emptyset$  for all  $n \in \mathbb{N}$ .

I claim that  $u \in \bigcap_{n \in \mathbb{N}} A_n$ . To see this, merely note that  $A_n$  (for any  $n$ ) is finitely determined, i.e., there is some  $k \in \mathbb{N}$  so that  $A_n = \pi_k^{-1}E$  with  $E \subseteq \{0, 1\}^k$ . But since  $A_n \cap \pi_k^{-1}\{u_1, \dots, u_k\} \neq \emptyset$ , it follows that  $(u_1, \dots, u_k) \in E$ , and therefore  $u \in A_n$ . We have proved that  $\bigcap_{n \in \mathbb{N}} A_n \neq \emptyset$ . This contradicts the assumption that  $F \subseteq \bigcup_{k \in \mathbb{N}} F_k$ , and the claim is therefore proved.

### Probability on the coin tossing space

From the general theory, we now have a measure  $\mu$  on a  $\sigma$ -algebra  $\mathcal{A}$  extending  $\mathcal{C}$ , and hence all the algebras  $\mathcal{A}_n$ . Moreover, by construction  $\mu(\pi_n^{-1}(E)) = \mu_n(E)$  for all  $E \in \{0, 1\}^n$ , so our probability measure certainly captures our intuition about the distribution of  $n$  coin tosses.

The  $n$ th toss is represented by a function  $T_n: \Omega \rightarrow \{0, 1\}$ , defined by  $T_n(t) = t_n$ .  $T_n$  is a *stochastic variable*, i.e., a measurable function on  $\Omega$ . It is not difficult to show that the tosses  $T_n$  are stochastically independent: If  $G_1, \dots, G_n \subseteq \mathbb{R}$  are Borel sets, then

$$\mu\left(\bigcap_{k=1}^n T_k^{-1}(G_k)\right) = \prod_{k=1}^n \mu(T_k^{-1}(G_k)).$$

(Note that  $T_k^{-1}(G_k)$  is one of four sets:  $\emptyset$ ,  $\{t \in \Omega: t_k = 0\}$ ,  $\{t \in \Omega: t_k = 1\}$ , or  $\Omega$ .)

### Interlude: Pulling and pushing measures

If  $(\Lambda, \mathcal{A}, \mu)$  is a measure space and  $g: \Omega \rightarrow \Lambda$  is a given function, (where the double-headed arrow indicates that the map is onto), we can define the *pullback*  $\sigma$ -algebra  $g^*\mathcal{A}$  and measure  $g^*\mu$  on  $\Omega$ :

$$g^*\mathcal{A} = \{g^{-1}(A) : A \in \mathcal{A}\},$$
$$g^*\mu(g^{-1}(A)) = \mu(A) \quad \text{for } A \in \mathcal{A}.$$

The proof is an easy exercise. (The assumption that  $g$  is onto is important.)

Note that we used this construction above, using the projection map  $\pi_n$  to pull the probability measure  $\mu_n$  back from  $\{0, 1\}^n$  to  $\Omega$ .

We can also push a measure forward: Given a map  $g: \Omega \rightarrow \Lambda$  where  $(\Omega, \mathcal{A}, \mu)$  is a measure space, we can create the *pushforward*  $\sigma$ -algebra  $g_*\mathcal{A}$  and measure  $g_*\mu$  on  $\Lambda$ :

$$g_*\mathcal{A} = \{A \subseteq \Lambda : g^{-1}(A) \in \mathcal{A}\},$$
$$g_*\mu(A) = \mu(g^{-1}(A)) \quad \text{for } A \in g_*\mathcal{A}.$$

Again, the proof is an easy exercise.

A typical example of this is the *distribution* of a real-valued stochastic variable  $X: \Omega \rightarrow \mathbb{R}$  on a probability space  $(\Omega, \mathcal{A}, \mu)$ : Since  $X$  is by definition measurable, any interval, and therefore any Borel set, belongs to  $X_*\mathcal{A}$ . The restriction of the pushforward measure  $X_*\mu$  to the  $\sigma$ -algebra  $\mathcal{B}$  of Borel sets is called the *distribution* of  $X$ .

In particular, the *cumulative distribution function*  $F$  of  $X$  is given by

$$F(x) = \mu(\{X \leq x\}) = \mu(X^{-1}((-\infty, x])) = X_*\mu((-\infty, x])$$

– and so, the distribution of  $X$  turns out to be the Lebesgue–Stieltjes measure associated with  $F$ .

**Warning:** I skipped quite a few details in this section. The sections to follow will skip even more details. You are encouraged to try to fill in the holes yourself.

### Pushing coin tossing measure

We can map the coin tossing space  $\Omega$  onto  $[0, 1]$  via the map  $f: \Omega \rightarrow [0, 1]$  given by

$$f(t) = \sum_{k \in \mathbb{N}} t_k \cdot 2^{-k},$$

which boils down to interpreting the coin tossing sequence  $t = (t_n)_{n \in \mathbb{N}}$  as a sequence of binary digits for the number  $f(t)$ . Since the “ $k$ th toss map”  $t \mapsto t_k$  is clearly a measurable function, then so is  $f$ . Let us investigate the pushforward measure  $f_*\mu$  on  $[0, 1]$ .

A real number is called a *dyadic rational* number if it can be written as  $2^{-n}m$  for some integers  $m$  and  $n$ .

A moment’s reflection shows that, given a dyadic rational  $x = 2^{-n}m \in [0, 1]$ , the inverse image  $f^{-1}([x, x + 2^{-n}])$  is given by those  $t \in \Omega$  whose first  $n$  components correspond to the first  $n$  bits of  $x$ , plus one more sequence ending in infinitely many 1s. For example,

$$f^{-1}([\frac{3}{8}, \frac{4}{8}]) = \{t \in \Omega: (t_1, t_2, t_3) = (0, 1, 1)\} \cup \{(0, 1, 0, 1, 1, 1, \dots)\}.$$

Since any singleton set in  $\Omega$  has measure zero, it follows (not immediately, but with a modest amount of work) that

$$f_*\mu([x, y]) = y - x = \lambda([x, y])$$

for any dyadic rational  $x, y$  in  $[0, 1]$  with  $x \leq y$ , with the same result for  $(x, y)$ ,  $[x, y)$  and  $(x, y]$ . Since such intervals form a semialgebra which generates the Borel sets, the uniqueness theorem for the extension of measures allows us to conclude the following:

*The pushforward  $f_*\mu$  equals Lebesgue measure on  $[0, 1]$ .*

Strictly speaking, the uniqueness theorem only allows us to conclude that  $f_*\mu(E) = \lambda(E)$  for *Borel* sets contained in  $[0, 1]$ , but this is easily extended to Lebesgue measurable sets, since both  $\mu$  and  $\lambda$  are complete measures.

### Pushing coin tossing measure some more

We are not done pushing coin tossing measure to  $[0, 1]$ . Consider the function  $g: \Omega \rightarrow C$  given by

$$g(t) = \sum_{k \in \mathbb{N}} 2t_k \cdot 3^{-k},$$

noting that this maps  $\Omega$  onto the Cantor set  $C \subset [0, 1]$ . This map is easy to understand: Take for example  $t = (0, 1, 1, 0, 1, 1, \dots)$ :

Since  $t_1 = 0$ ,  $g(t)$  belongs to the left interval of length  $\frac{1}{3}$  left after we removed the middle of  $[0, 1]$ .

Next,  $t_2 = 1$ , so  $g(t)$  belong to the right interval of length  $\frac{1}{9}$  left over after we removed the middle of  $[0, \frac{1}{3}]$  (and  $[\frac{2}{3}, 1]$ ).

Next,  $g(t)$  belongs to the right interval of length  $\frac{1}{27}$  left over after we removed the middle of  $[\frac{2}{9}, \frac{3}{9}]$  (and three other intervals).

It should be clear that  $g$  is one to one, for as soon as we have encountered a difference between  $s$  and  $t$ , we know that  $g(s)$  and  $g(t)$  belongs to well separated intervals.

The resulting measure  $g_*\mu$  on  $C$  is sometimes referred to as the *Cantor measure*.

Its cumulative distribution function is none other than the standard Cantor function  $\psi$ . You may verify this directly, but noting that both functions take the same constant value on each of the open intervals in  $[0, 1] \setminus C$ . Another approach is to prove and then use the relation

$$\psi(g(t)) = f(t),$$

where  $f$  is the function defined in the previous paragraph.

### Infinitely many uniform variables

We return to the map  $f: \Omega \rightarrow [0, 1]$ . We can think of this map as a *stochastic variable uniformly distributed on  $[0, 1]$* .

In probability theory, one often needs infinitely many stochastically independent such variables. And cointossing space delivers!

The easiest way to do this is to notice that  $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$  is countable, so there is a one-to-one map  $c: \mathbb{N}^2 \rightarrow \mathbb{N}$ . We can use it to define  $U_n: \Omega \rightarrow [0, 1]$ :

$$U_n(t) = \sum_{k \in \mathbb{N}} t_{c(n,k)} \cdot 2^{-k}, \quad n \in \mathbb{N}.$$

There are straightforward techniques for transforming a uniformly distributed variable into a stochastic variable with any desired distribution, so you can easily generate infinitely many independent variables with a standard normal distribution, for example. And that gives you the building blocks to generate a model for Brownian motion (also known as the Wiener process). But that is a topic for a different course.