

November 7, 2021

RANDOM VARIABLES II

ABSTRACT. We will discuss the existence of random variables and independent sequences of random variables.

1. UNIFORM RANDOM VARIABLES

Recall that we asserted the existence of Lebesgue measure on the unit interval; that is, by a result in measure theory, we know that there is a translation-invariant probability measure defined on the Borel subsets of the unit interval. This easily implies the existence of a random variable that is uniformly distributed on the unit interval.

Theorem 1. *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $U : \Omega \rightarrow [0, 1]$ such that U is uniformly distributed on $[0, 1]$.*

The proof of Theorem 1 is underwhelming.

Proof of Theorem 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space for Lebesgue measure, where $\Omega = [0, 1]$ is the unit interval, and \mathcal{F} is the Borel subsets of the unit interval generated by the intervals of the form $[a, b)$, and $\mathbb{P}[a, b) = b - a$ for $0 \leq a < b < 1$. Take $U : \Omega \rightarrow \Omega$ to be the identity map; that is $U(\omega) = \omega$. \square

We will see that one uniform random variable is all we need.

Theorem 2. *Let F be cdf. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \rightarrow \mathbb{R}$ with F as its cdf.*

Before we prove Theorem 2, we give a simple construction to show how Theorem 1 implies the existence of any discrete random variable.

Exercise 1.1. *Let $p : \mathbb{N} \rightarrow [0, 1]$ be a pmf. Find a (measurable) function $\phi : [0, 1) \rightarrow \mathbb{R}$ such that $\phi(U)$ has pmf p , if U is uniformly distributed in on the unit interval.*

Solution to Exercise 1.1. Let U be uniformly distributed on the unit interval. Consider the intervals

$$I_0 = [0, p_0), I_1 = [p_0, p_0 + p_1), \dots, I_{n+1} = [p_n, p_n + p_{n+1}), \dots$$

Clearly, the intervals are Borel measurable, and $\mathbb{P}(U \in I_n) = p_n$. Set

$$\phi = \sum_{n=0}^{\infty} n \mathbf{1}_{I_n}.$$

It is easy to verify that ϕ is measurable. \square

If F is a cdf, in the case that F is strictly increasing, then its inverse function is well-defined F^{-1} on $(0, 1)$. Sometimes F^{-1} is called the **quantile function**. One can verify that it is also increasing. If U is uniformly distributed on the unit interval, then since F^{-1} is increasing we have

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x); \quad (1)$$

hence $F^{-1}(U)$ has the cdf F . We should check that F^{-1} is measurable.

Lemma 3. *If $f : \mathbb{R} \rightarrow \mathbb{R}$ is a monotone function, then it is measurable.*

Proof. By taking the negative, without loss of generality, assume that f is nondecreasing. Fix $y \in \mathbb{R}$. We have

$$f^{-1}(-\infty, y) = \{x \in \mathbb{R} : f(x) < y\}.$$

Since f is nondecreasing, if $x \in f^{-1}(-\infty, y)$, then any $z \leq x$ have we $f(z) \leq f(x)$, thus we also have $z \in f^{-1}(-\infty, y)$. Hence $f^{-1}(-\infty, y)$ can be all of \mathbb{R} , the empty set, or an interval of the form $(-\infty, z)$ or $(-\infty, z]$ for $z \in \mathbb{R}$. \square

For cdf F that is not necessarily strictly increasing, we define its **generalized quantile function** via

$$F^{-1}(y) = \sup \{x \in \mathbb{R} : F(x) < y\},$$

for $y \in (0, 1)$, which is a nondecreasing and thus measurable function.

Proof of Theorem 2. We already know that there is a probability space that can house a random variable U that is uniformly distributed in the unit interval. One can verify using the right continuity of F that (1) holds for the case where we have to use the generalized quantile function. \square

Remark 1. *Note that the proof of Theorem 2 gives one way of generating random variables, if one is given a random variable that is uniformly distributed on the unit interval to begin with—in practice this is often the case, in the sense that most mathematical software has the ability to generate such pseudo random variables. However, in practice, there are usually much quicker and efficient ways to generate random variables than to use Theorem 2.*

2. INDEPENDENT SEQUENCES OF RANDOM VARIABLES

We will see that everything we want is a result of the fact that there exists one random variable that is uniformly distributed in on the unit

interval. Given $u \in [0, 1)$, we write

$$u = \sum_{i=1}^{\infty} \frac{b_i}{2^i},$$

where $b_i = b_i(u) \in \{0, 1\}$ so that the sequence $b(u) = (b_1, b_2, \dots)$ gives the binary expansion of u .

For fixed $n \geq 1$ and any binary sequence $a \in \{0, 1\}^n$ we have

$$\mathbb{P}(b_1(U) = a_1, \dots, b_n(U) = a_n) = \frac{1}{2^n},$$

since this event corresponds to an event that U is in an interval of length $\frac{1}{2^n}$. It is also not difficult to see that

$$\mathbb{P}(b_2(U) = a_2, \dots, b_n(U) = a_n) = \frac{1}{2^{n-1}},$$

since this event corresponds to the an event that U belongs to one of two disjoint intervals of length $\frac{1}{2^n}$. Moreover, if we have k remaining restrictions out the coordinates i_1, \dots, i_k , then

$$\mathbb{P}(b_{i_1}(U) = a_{i_1}, \dots, b_{i_k}(U) = a_{i_k}) = \frac{1}{2^k}.$$

Thus we have the following result.

Proposition 4. *If U is uniformly distributed on the unit interval, then $b(U)$ gives a sequence of i.i.d. Bernoulli $\frac{1}{2}$ variables.*

From Proposition 4, it is easy to see that how we can get two independent random variables that are uniformly distributed on the unit interval from one such random variable—just take odd and even bits. Moreover, we have the following corollary.

Corollary 5. *Let U be uniformly distributed on the unit interval. Let F be a cdf. There exists a measurable functions $v_i : [0, 1) \rightarrow [0, 1)$ such that*

$$v_1(U_1), v_2(U), \dots$$

is an i.i.d. sequence of random variables have F as their cdf.

Proof. By taking binary expansions, we get a an infinite sequence of i.i.d. random variables U_1, \dots, U_n, \dots that are uniformly distributed on the unit interval. By applying F^{-1} to each U_i we obtain the desired result—once you believe the following. \square

Exercise 2.1. *Show that if X and Y are independent real-valued random variables, then if g and h are Borel measurable functions, then $g(X)$ and $h(Y)$ are also independent random variables.*