4.3 Universality of the Uniform

Unit 4: Continuous Random Variables

Adapted from Blitzstein-Hwang Chapter 5.

In this section, we will discuss a remarkable property of the Uniform distribution: given a $\textbf{Unif}(0,1)$ r.v., we can construct an r.v. with *any continuous distribution we want.* Conversely, given an r.v. with an arbitrary continuous distribution, we can create a $\textbf{Unif}(0,1)$ r.v. We call this the *universality of the Uniform*, because it tells us the Uniform is a universal starting point for building r.v.s with other distributions. Universality of the Uniform also goes by many other names, such as the *probability integral transform*, *inverse transform sampling*, the *quantile transformation*, and even the *fundamental theorem of simulation*.

To keep the proofs simple, we will state the universality of the Uniform for a case where we know that the inverse of the desired CDF exists. More generally, similar ideas can be used to simulate a random draw from *any* desired CDF, as a function of a $\textbf{Unif}(0,1)$ r.v.

> THEOREM 4.3.1 (UNIVERSALITY OF THE UNIFORM).
>
> Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from $(0,1)$ to $\mathbb{R}$. We then have the following results.
>
> 1. Let $U \sim \textbf{Unif}(0,1)$ and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.
> 2. Let X be an r.v. with CDF F. Then $F(X) \sim \textbf{Unif}(0,1)$.

Let's make sure we understand what each part of the theorem is saying. The first part of the theorem says that if we start with $U \sim \textbf{Unif}(0,1)$ and a CDF F, then we can create an r.v. whose CDF is F by plugging U into the inverse CDF F^{-1}. Since F^{-1} is a function (known as the *quantile function*), U is a random variable, and a function of a random variable is a random variable, $F^{-1}(U)$ is a random variable; universality of the Uniform says its CDF is F.

The second part of the theorem goes in the reverse direction, starting from an r.v. X whose CDF is F and then creating a $\textbf{Unif}(0,1)$ r.v. Again, F is a function, X is a random variable, and a function of a random variable is a random variable, so $F(X)$ is a random variable. Since any CDF is between 0 and 1 everywhere, $F(X)$ must take values between 0 and 1. Universality of the Uniform says that the distribution of $F(X)$ is

Uniform on $(0, 1)$.

> **WARNING 4.3.2.**
>
> The second part of universality of the Uniform involves plugging a random variable X *into its own CDF* F. This may seem strangely self-referential, but it makes sense because F is just a function (that satisfies the properties of a valid CDF), so $F(X)$ is a function of a random variable and hence is itself a random variable. There is a potential notational confusion, however: $F(x) = P(X \le x)$ by definition, but it would be incorrect to say "$F(X) = P(X \le X) = 1$". Rather, we should first find an expression for the CDF as a function of x, then replace x with X to obtain a random variable. For example, if the CDF of X is $F(x) = 1 - e^{-x}$ for $x > 0$, then $F(X) = 1 - e^{-X}$.

Proof

1. Let $U \sim \textbf{Unif}(0, 1)$ and $X = F^{-1}(U)$. For all real x,

$$P(X \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x),$$

so the CDF of X is F, as claimed. For the last equality, we used the fact that $P(U \le u) = u$ for $u \in (0, 1)$.

2. Let X have CDF F, and find the CDF of $Y = F(X)$. Since Y takes values in $(0, 1)$, $P(Y \le y)$ equals 0 for $y \le 0$ and equals 1 for $y \ge 1$. For $y \in (0, 1)$,

$$P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Thus Y has the $\textbf{Unif}(0, 1)$ CDF.

To gain more insight into what the quantile function F^{-1} and universality of the Uniform mean, let's consider an example that is familiar to millions of students: percentiles on an exam.

Example 4.3.3 (Percentiles).

A large number of students take a certain exam, graded on a scale from 0 to 100. Let X be the score of a random student. Continuous distributions are easier to deal with here, so let's approximate the discrete distribution of scores using a continuous distribution. Suppose that X is continuous, with a CDF F that is strictly increasing on $(0, 100)$. In reality, there are only finitely many students and only finitely many possible scores, but a continuous distribution may be a good approximation. Suppose that the median score on the exam is 60, i.e., half of the students score above 60 and the other half score below 60 (a convenient aspect of assuming a continuous distribution is that we don't need to worry about how many students had scores *equal* to 60). That is, $F(60) = 1/2$, or, equivalently, $F^{-1}(1/2) = 60$.

If Jimmy scores a 72 on the exam, then his *percentile* is the fraction of students who score below a 72. This is $F(72)$, which is some number in $(1/2, 1)$ since 72 is above the median. In general, a student with score x has percentile $F(x)$. Going the other way, if we start with a percentile, say 0.95, then $F^{-1}(0.95)$ is the score that has that percentile. A percentile is also called a *quantile*, which is why F^{-1} is called the quantile function. The function F converts scores to quantiles, and the function F^{-1} converts quantiles to scores.

The strange operation of plugging X into its own CDF now has a natural interpretation: $F(X)$ is the percentile attained by a random student. It often happens that the distribution of scores on an exam looks very non-Uniform. For example, there is no reason to think that 10% of the scores are between 70 and 80, even though $(70, 80)$ covers 10% of the range of possible scores.

On the other hand, the distribution of *percentiles* of the students *is* Uniform: the universality property says that $F(X) \sim \textbf{Unif}(0, 1).$ For example, 50% of the students have a percentile of at least 0.5. Universality of the Uniform is expressing the fact that 10% of the students have a percentile between 0 and 0.1, 10% have a percentile between 0.1 and 0.2, 10% have a percentile between 0.2 and 0.3, and so on---a fact that is clear from the definition of percentile.

To illustrate universality of the Uniform, we will apply it to the Logistic distribution.

Example 4.3.4 (Universality with Logistic).

The Logistic CDF is

$$F(x) = \frac{e^x}{1+e^x}, \quad x \in \mathbb{R}.$$

Suppose we have $U \sim \textbf{Unif}(0, 1)$ and wish to generate a Logistic r.v. Part 1 of the universality property says that $F^{-1}(U) \sim \textbf{Logistic}$, so we first invert the CDF to get F^{-1}:

$$F^{-1}(u) = \log\left(\frac{u}{1-u}\right).$$

Then we plug in U for u:

$$F^{-1}(U) = \log\left(\frac{U}{1-U}\right).$$

Therefore $\log\left(\frac{U}{1-U}\right) \sim \textbf{Logistic}$. Conversely, Part 2 of the universality property states that if $X \sim \textbf{Logistic}$, then

$$F(X) = \frac{e^X}{1+e^X} \sim \textbf{Unif}(0, 1).$$