STAT 111, Week 2 - Uniform Random Variables

Leia Donaway — Swarthmore College

January 30, 2025

This week we are working from the Continuous Random Variables Week 2 guidelines which are on Moodle. The first part is Uniform random variables, which you can read more about in Blitzstein 5.2 and 5.3.

We should first recognize that, as humans, we measure things discretely, in units that shrink to a level of specificity that we can work with. To this end, we don't tend to think of things as continuous variables, since we measure at discrete points. It is useful and convenient for us to think about the notion of spreading a discrete variable out such that it is feasibly continuous so that we can use calculus on it. Conceptually, when we think about a continuous uniform random variable, we have a nice random number generator. We'll explore that process and some results surrounding this discovery as set up for the other ideas that come out of continuous random variables.

a) Give the probability mass function (pmf) for a discrete Uniform variable X, defined on the integers $1, .., n$. Sketch the probability histogram for $n = 6$ (e.g., for rolling a six-sided die). Note how, with bars of width 1.0, the probability of any subset of the sample space (e.g., odd numbers) is equal to the area of the bars corresponding to those outcomes. This means that the pmf is also a density function.

pmf $\mathbb{P}(X = k) = \frac{1}{n}$, where $k = 1, ..., n$. A probability histogram for eg. $n = 6$ looks like a histogram 1 through 6 with bar width 1 and height $\frac{1}{6}$. Note how the width of bar $*$ height of bar = the probabilty of the outcome corresponding to the bar, which makes this also a density function.

b) Define $U = X/n$, with $U \in (0, 1]$ for all n. To have the area of the bars correspond to probability for U, divide the pmf by the width of each bar to get the density function $f_u(u)$. Describe the limit of this probability density function (pdf) and of this discrete random variable as $n \to \infty$. Define the continuous Uniform random variable $U \sim Unif(0, 1)$ and the more general $Unif(a, b)$ distribution.

$U = \frac{X}{n} = X * \frac{1}{n}$, so $U \in \{\frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n} = 1\} \implies U \in (0, 1]\ \forall n$

In order to get the density function $f_u(u)$, we again need the area of the bars of our histogram to be equal to the probability of the outcome they correlate with (U). We should redraw the histogram after redistributing the area such that now we divide the pmf by the width of each bar. Generally, this will look like

$f_u(u) = \frac{\mathbb{P}(\mathbb{U}=\approxeq)}{width} = \frac{\frac{1}{n}}{\frac{1}{n}} = 1, 0 < u \leq 1$, as $n \to \infty$, $f_u(u) = \begin{cases} 1 & 0 < u \leq 1 \\ 0 & \text{otherwise} \end{cases}$, a Uniform(0,1)

distribution.
Our 6-sided die example will now have a histogram with bar width of $\frac{1}{6}$ and bar height of 1. As $n \to \infty$, $U \sim$ Unif(0,1).

More generally, $U \sim$ Unif(a,b) has pdf $f_u(u) = \begin{cases} \frac{1}{b-a} & a \le u \le b \\ 0 & \text{otherwise} \end{cases}$

c) Let F be the cdf for a continuous random variable, with F a monotone increasing function (briefly explain the other possibilities). Let $G(u)$ be the inverse function, so that $G(u) = x$ if $F(x) = u$. For a random variable X that has cdf F, show that $U = F(X)$ follows a $Unif(0, 1)$ distribution. Give intuition for why this makes sense.

We specify that F is monotone increasing so that we know we are working with a continuous random variable. The cdf would have to include portions of discreteness to be non-changing at any point. Consider what it'd be like to sketch this.

Our purpose here is to demonstrate the universality of the uniform by doing a transformation in both directions. Blitzstein Hwang 5.3 explains further, the following will be proof and a brief explaination of the intuition behind why this works.
The definitions of $F(x)$ as the cdf of X means that $\mathbb{P}(X \le x) = F(x)$. We are applying the cdf on the CRV U, as if is just a function.
$U = F(x)$, so $f_u(u) = \mathbb{P}(U \le u) = \mathbb{P}(F(x) \le u) = \mathbb{P}(G(F(x)) \le G(u)) = \mathbb{P}(x \le G(U)) = F(G(u)) = u$.
We have $F(X)$ $\sim$Unif(0,1). Why does this make sense? $F(X)$, as the cdf of X is the probability that X is less than or equal to the given value u. When we map X through its own CDF we effectively rescale it (like in part b) where the values are al between zero and one, as it is a probability space. Think about a cdf graph. The area under the curve increases, 0 to 1, at some point being equal to .5. This is the median, where half of outcomes fall above and half of outcomes fall below. Consider drawing this and thinking about the chance of randomly choosing an outcome above the 3rd quartile, for example.

d) Reversing the calculation from part d, suppose $U \sim Unif(0, 1)$ and define the transformed random variable $X = G(U)$. Show that X has cdf F. This is called the "inverse-cdf method" for random number generation. As an example, for $U \sim Unif(0, 1)$ and a constant $\lambda > 0$, show that $X = log(U)/\lambda$ follows an Exponential(λ) distribution with cdf $F(x) = 1e\lambda x$ for $x > 0$. Compare a large sample generated using this method to a sample generated using rgamma. Compare the moments and plot the sorted values of each sample against each other to see how close they come to the line $y = x$.

We know from part c that if $U \sim$ Unif(0, 1) and $X = F1(U)$, then X is an r.v. with CDF F. Here, we are asked to show part two of the Universality of the Uniform Theorem: Let X be an r.v. with CDF F. Then $F(X) \sim$ Unif(0, 1). One can imagine the inverse of the process we took in part c, it is left as an exercise for the reader. We can work out an example like the Exponential.
Starting with the cdf, solve for x in terms of u. $F(x) = 1 - e^{-\lambda x}$, $x > 0$
$u = 1 - e^{-\lambda x}$
$e^{-\lambda x} = 1 - u$
$-\lambda x = \log(1 - u)$
$x = \frac{\log(1-u)}{-\lambda}$. We know that log(u) and log(1-u) are equivalent and both Unif(0,1).

$x = \frac{\log(1-u)}{-\lambda}$ ∼Exponential(λ)
Finally, running the following code in R will allow you to perform the large sample comparison with rgamma. This will compare our method to the one the computer uses with the rgamma command (they are the same method, in fact). You should get something very close to $y = x$, with some variation at the end being caused by a few influential points.

```
x=rgamma(10000,1,.5)
u=runif(10000)
x_2=-2*log(u)
plot(sort(x),sort(x_2),pch=".")
abline(0,1,col="red")
```