Resimulation

By James Shaheen

1. INTRODUCTION

One method of simulating random variables is to generate uniform 0 to 1 variables, then use them in the inverse of the cumulative distribution function of the random variable you want to simulate. For example, if you had U_1, a uniform 0 to 1 variable, and you wanted to use it to simulate S_1, the flip of a fair coin, you could use this function:

If $0<U_1<0.5$, then S_1=tails

If $0.5<U_1<1$, then S_1=heads

Historically it was thought that if you wanted to simulate two independent variables in this manner, you would have to generate two uniform variables. But in fact, you can simulate multiple independent discrete variables from a single uniform variable.

Example 1

> If you wanted to simulate two flips of a fair coin (S_1 and S_2), you could use this function:
>
> If $0<U_1<0.25$, then S_1=tails and S_2=tails
>
> If $0.25<U_1<0.5$, then S_1=tails and S_2=heads
>
> If $0.5<U_1<0.75$, then S_1=heads and S_2=tails
>
> If $0.75<U_1<1$, then S_1=heads and S_2=heads

By simple inspection you can see that by this method S_1 and S_2 both have a 50% chance of being heads and a 50% chance of being tails, and that they are independent from each other.

2. ADDITIONAL VARIABLES

This method can work to simulate as many discrete variables as you want.

Example 2

> If you wanted to simulate three flips of a fair coin, you could use this function:
>
> If $0<U_1<0.125$, then S_1=tails, S_2=tails, and S_3=tails

If $0.125<U_1<0.25$, then S_1=tails, S_2=tails, and S_3=heads

If $0.25<U_1<0.375$, then S_1=tails, S_2=heads, and S_3=tails

If $0.375<U_1<0.5$, then S_1=tails, S_2=heads, and S_3=heads

If $0.5<U_1<0.625$, then S_1=heads, S_2=tails, and S_3=tails

If $0.625<U_1<0.75$, then S_1=heads, S_2=tails, and S_3=heads

If $0.75<U_1<0.875$, then S_1=heads, S_2=heads, and S_3=tails

If $0.875<U_1<1$, then S_1=heads, S_2=heads, and S_3=heads

So for example, if $U_1<0.43$, then S_1=tails, S_2=heads, and S_3=heads.

3. A SIMPLER FORMULA

The method from Sections 1 and 2 gets exponentially more complicated as you simulate more variables, so it can be a pain to use, especially in cases where you're simulating something more complicated than the flip of a coin. But it can be simplified if you use the first uniform 0 to 1 variable to simulate other uniform 0 to 1 variables, and use those other uniform variables to simulate subsequent flips of a coin, or subsequent variables of whatever distribution you want to simulate.

To do this, I introduce these definitions:

Definition 1

$\text{Min}(F^{-1}(F(U_n)))$=The smallest number which, when plugged into the function F, would produce the same result that U_n produced.

Definition 2

$\text{Max}(F^{-1}(F(U_n)))$=The largest number which, when plugged into the function F, would produce the same result that U_n produced.

Example 3

Suppose you were simulating a coin flip, using this function F:

If $0<U_n<0.5$, then $F(U_n)$=tails

If $0.5<U_n<1$, then $F(U_n)$=heads

If $U_n=0.43$, then $F(U_n)$=tails. Therefore $F^{-1}(F(U_n))=F^{-1}$(tails)=anything from 0 to 0.5. Therefore

$Min(F^{-1}(F(U_n)))$=Min(anything from 0 to 0.5)=0.

$Max(F^{-1}(F(U_n)))$=Max(anything from 0 to 0.5)=0.5.

Example 4

Using the same function F from Example 3,

$Min(F^{-1}(F(0.86)))=0.5$.

$Max(F^{-1}(F(0.86)))=1$.

Using those new definitions, I introduce this recursive formula for simulating additional uniform 0 to 1 variables:

Theorem 1

$$U_{n+1}=(U_n-Min(F^{-1}(F(U_n))))/(Max(F^{-1}(F(U_n)))-Min(F^{-1}(F(U_n))))$$

Example 5

Assume we have a uniform 0 to 1 variable, U_1, equal to 0.43, and from that we want to simulate three flips of a fair coin.

We'll use the simple simulation formula:

If $0<U_n<0.5$, then $S_n=F(U_n)$=tails

If $0.5<U_n<1$, then $S_n=F(U_n)$=heads

So if $U_1=0.43$, then S_1=tails

To get U_2, we use the formula

$U_{n+1}=(U_n-Min(F^{-1}(F(U_n))))/(Max(F^{-1}(F(U_n)))-Min(F^{-1}(F(U_n))))$

$U_2=(U_1-Min(F^{-1}(F(U_1))))/(Max(F^{-1}(F(U_1)))-Min(F^{-1}(F(U_1))))$

$U_2=(0.43-Min(F^{-1}(F(0.43))))/(Max(F^{-1}(F(0.43)))-Min(F^{-1}(F(0.43))))$

$U_2=(0.43-0)/(0.5-0)=0.86$

To get S_2 we can again use our simple simulation formula:

If $0<U_n<0.5$, then $S_n=F(U_n)$=tails

If $0.5<U_n<1$, then $S_n=F(U_n)$=heads

So if $U_2=0.86$, then S_2=heads

To get U_3, we again use the recursive formula

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

$U_3=(U_2\text{-Min}(F^{-1}(F(U_2))))/(\text{Max}(F^{-1}(F(U_2)))\text{-Min}(F^{-1}(F(U_2))))$

$U_3=(0.86\text{-Min}(F^{-1}(F(0.86))))/(\text{Max}(F^{-1}(F(0.86)))\text{-Min}(F^{-1}(F(0.86))))$

$U_3=(0.86-0.5)/(1-0.5)=0.72$

And finally, we can get S_3 by using our simple simulation formula again:

If $0<U_n<0.5$, then $S_n=F(U_n)=\text{tails}$

If $0.5<U_n<1$, then $S_n=F(U_n)=\text{heads}$

So if $U_3=0.72$, then $S_3=\text{heads}$

Therefore if $U_1=0.43$, then $S_1=\text{tails}$, $S_2=\text{heads}$, and $S_3=\text{heads}$. And we were able to generate all three of those variables using just two simple formulas.

Note that neither U_2 nor S_2 are independent of U_1. But S_2 is independent of S_1, since S_2 has a 50% chance of being heads and a 50% chance of being tails regardless of whether S_1 is heads or tails.

Example 6

Assume we have a uniform 0 to 1 variable, U_1, equal to 0.29, and from that we want to simulate variables, S_1, S_2, and S_3, each of which have a 20% chance of being 0, a 50% chance of being 1, and a 30% chance of being 2.

If $0<U_n<0.2$, $S_n=F(U_n)=0$.

If $0.2<U_n<0.7$, $S_n=F(U_n)=1$.

If $0.7<U_n<1$, $S_n=F(U_n)=2$.

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

Therefore $S_1=1$.

$U_2=(0.29-0.2)/(0.7-0.2)=0.18$.

$S_2=0$.

$U_3=(0.18-0)/(0.2-0)=0.9$.

$S_3=2$.

Note that this theorem will not work with functions that have gaps in the definition of single outcomes. For example, it will work with this function:

If $0<U_1<0.5$, then S_1=tails

If $0.5<U_1<1$, then S_1=heads

But it would not work with this function, since it has a gap in the definition of S_1=tails:

If $0<U_1<0.25$, then S_1=tails

If $0.25<U_1<0.75$, then S_1=heads

If $0.75<U_1<1$, then S_1=tails

4. CONTINUOUS SIMULATED VARIABLES

In general, if you use a uniform 0 to 1 random variable to simulate a continuous variable, you will not be able to reuse it to simulate a second variable that is independent of the first simulated variable. The formula for U_{n+1} doesn't make sense if S_n is continuous.

If $F(U_n)$ is continuous, then $U_n=\mathrm{Min}(F^{-1}(F(U_n)))=\mathrm{Max}(F^{-1}(F(U_n)))$.

Therefore the formula for U_{n+1}:

$U_{n+1}=(U_n-\mathrm{Min}(F^{-1}(F(U_n))))/(\mathrm{Max}(F^{-1}(F(U_n)))-\mathrm{Min}(F^{-1}(F(U_n))))$

Comes out to

$U_{n+1}=(U_n-U_n)/(U_n-U_n)=0/0$, which is undefined.

So in general, you can only reuse a uniform 0 to 1 random variable if it was first used to simulate a discrete variable.

5. PARTIALLY DISCRETE VARIABLES

If you're using a uniform 0 to 1 random variable to simulate a variable that is partially discrete and partially continuous, then you can reuse the uniform 0 to 1 random variable if *and only if it ends up simulating the part of the variable that is discrete.*

Example 7

Assume that we want to simulate three independent variables (S_1, S_2, and S_3) that each have a 50% chance of being uniformly distributed between 0 and 0.5 and a 50% chance of being 0.5. To do this we are given the uniform 0 to 1 variables 0.6 (which we'll call U_a), 0.3 (U_b), and 0.1 (U_c), but whenever possible we should reuse those uniform variables using resimulation, so that we use as few of them as possible. To simulate S_1, S_2, and S_3, we can use the formula:

If $0<U_n<0.5$, $S_n=F(U_n)=U_n$.

If $0.5<U_n<1$, $S_n=F(U_n)=0.5$.

We need a uniform 0 to 1 variable, U_1, to simulate S_1. For lack of anything else we can use, we'll use the U_a that we were given.

Therefore $U_1=0.6$, therefore $S_1=0.5$. That's the part of the variable that is discrete, therefore we can reuse U_1 to simulate U_2.

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

Therefore

$U_2=(U_1\text{-Min}(F^{-1}(F(U_1))))/(\text{Max}(F^{-1}(F(U_1)))\text{-Min}(F^{-1}(F(U_1))))$

$U_2=(0.6\text{-Min}(F^{-1}(F(0.6))))/(\text{Max}(F^{-1}(F(0.6)))\text{-Min}(F^{-1}(F(0.6))))$

$U_2=(0.6\text{-}0.5)/(1\text{-}0.5)=0.2$.

Therefore $S_2=0.2$. That's the part of the variable that is continuous, therefore we can't reuse U_2 to simulate U_3. If we tried, we'd get:

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

$U_3=(U_2\text{-Min}(F^{-1}(F(U_2))))/(\text{Max}(F^{-1}(F(U_2)))\text{-Min}(F^{-1}(F(U_2))))$

$U_3=(0.2\text{-Min}(F^{-1}(F(0.2))))/(\text{Max}(F^{-1}(F(0.2)))\text{-Min}(F^{-1}(F(0.2))))$

$U_3=(0.2\text{-}0.2)/(0.2\text{-}0.2)=0/0$, which is undefined.

So instead, for U_3 we'll have to use U_b, the second of the uniform 0 to 1 variables we were given.

$U_3=0.3$. Therefore $S_3=0.3$. And we don't have to use U_c.

Example 8

The James Insurance Company has losses at a Poisson rate of 3 per year. 40% of the losses are for \$100; 35% are for \$1,000; and 25% are for \$10,000. Simulate how much it had in losses in one year. To do this we are given the uniform 0 to 1 variables U_a to U_k: 0.57, 0.79, 0.63, 0.02, 0.33, 0.68, 0.18, 0.94, 0.12, 0.21, and 0.95, but whenever possible we should reuse those uniform variables using resimulation, so that we use as few of them as possible.

We simulate the number of losses first, then use resimulation to simulate the amount of each loss. If losses occur at a Poisson rate, then the times between losses are exponential. So we'll start by simulating them. The CDF for an exponential distribution with □ =1/3 is:

1-e^(-3*S_n)

We can simulate exponential variables if we set that equal to U_n, and solve for S_n, which gives us:

S_n=F(U_n)=-1/3*ln(1-U_n)

We need a uniform 0 to 1 variable, U_1, to simulate S_1. For lack of anything else we can use, we'll use the U_a that we were given. Therefore U_1=0.57, therefore S_1, the time to the first loss is

=-1/3*ln(1-0.57)=0.2813.

Now we need another uniform 0 to 1 variable, U_2, to simulate S_2. S_1 was continuous, so we can't reuse U_1 to get U_2. So for lack of anything else we can use, we'll use the U_b that we were given. Therefore U_2=0.79, therefore S_2, the time from the first loss to the second loss is

=-1/3*ln(1-0.79)=0.5202.

The total time until the third loss is

=0.2813+0.5202=0.8015.

Similarly, for U_3 we'll use U_c. Therefore U_3=0.63, therefore S_3, the time from the second loss to the third loss is

=-1/3*ln(1-0.63)=0.3314.

The total time until the third loss is

=0.2813+0.5202+0.3314=1.1329.

The second loss was the last one that occurred before the end of the first year, therefore there were two losses in the year. Now we need to simulate the amount of those two losses. To do that, we need another uniform 0 to 1 variable, U_4. And here's the trick. Since S_3 was continuous, one might imagine that we can't reuse U_3 to simulate U_4. But we can.

1-S_1-S_2=1-0.2813-0.5202=0.1985. Therefore if the time from the second loss to the third loss had been less than 0.1985, there would have been (at least) three losses in the year. But as long as the time from the second loss to the third loss was greater than 0.1985, there would only have been two losses in the year. And once we know that there were only two losses in the year, it doesn't matter to us when the third loss actually occurs. It could be in year 2 or it could be in year 100. It doesn't matter. And we can reflect that by rewriting $F(U_3)$ as a partially discrete function.

The CDF for an exponential distribution with □ =1/3 is:

$1\text{-}e\hat{}(\text{-}3*S_n)=U_n$

Therefore the U_n that corresponds to a S_n of 0.1985 is:

$1\text{-}e\hat{}(\text{-}3*0.1985)=0.449$.

So we can use this partially discrete function:

If $0<U_3<0.449$, then $S_3=F(U_3)=-1/3*\ln(1-U_3)$.

If $0.449<U_3<1$, then S_3 is such that the third loss is after the end of the first year.

And since $U_3>0.449$, it falls into the discrete portion of our new function, and we can reuse U_3 to simulate U_4.

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

Therefore

$U_4=(U_3\text{-Min}(F^{-1}(F(U_3))))/(\text{Max}(F^{-1}(F(U_3)))\text{-Min}(F^{-1}(F(U_3))))$

$U_4=(0.63-0.449)/(1-0.449)=0.328$

And once we have a new uniform 0 to 1 random variable, we can simulate the size of the losses in the first year fairly easily.

If $0<U_n<0.4$, $S_n=100$.

If $0.4<U_n<0.75$, $S_n=1{,}000$.

If $0.75<U_n<1$, $S_n=10{,}000$.

$U_{n+1}=(U_n\text{-Min}(F^{-1}(F(U_n))))/(\text{Max}(F^{-1}(F(U_n)))\text{-Min}(F^{-1}(F(U_n))))$

$U_4=0.328$. Therefore the size of the first loss, $S_4=100$.

$U_5=(0.328-0)/(0.4-0)=0.82$. Therefore the size of the second loss, $S_5=10{,}000$.

Therefore the total losses are 100+10,000=10,100. All done using only the first three of the uniform random variables we were given.