

## **Simulation**

Suppose that  $X, Y, X, \dots$  are independent random variables whose distributions we know. We want to find out the distributions of simple functions of these, like  $X^2$ ,  $X + Y$ ,  $X - Y$ ,  $XYZ$ ,  $\dots$ . We can use Theorem 7 to find the distribution of  $X^2$ , and Theorems 5 and 6 give us the distribution of  $X + Y$  in special cases, but what about the others?

If the theory is too hard, then we can *simulate*: generate large numbers of values of  $X, Y, Z, \dots$  that have the appearance of being independent realizations of those random variables. Calculate the relevant function of those values (e.g. calculate  $x - y$  etc.): the empirical distribution of the data should then be a good approximation to the distribution we are trying to find.

How do we simulate? For  $\text{Bin}(30, \frac{1}{2})$ , we can simulate a single value by tossing a fair coin 30 times. However, most random variables cannot be simulated by physical experiments.

Calculators and computers usually have a “random number generator”, which produces values that appear to be independent realizations of the random variable  $U$ , where  $U \sim \text{uniform}[0, 1]$ .

If  $X$  is discrete, with values  $x_1 < x_2 < \dots < x_i < \dots$ , put  $p_i = \mathbb{P}(X = x_i)$  and  $r_i = \mathbb{P}(X \leq x_i)$ . Then

$$p_i = \mathbb{P}(X \leq x_i) - \mathbb{P}(X \leq x_{i-1}) = r_i - r_{i-1}.$$

Because  $U$  is uniform on  $[0, 1]$ , we know that

$$\mathbb{P}(r_{i-1} < U \leq r_i) = r_i - r_{i-1} = p_i = \mathbb{P}(X = x_i).$$

Hence we obtain the simulated value of  $X$  from the simulated value of  $U$  by putting

$$X = x_i \quad \text{if} \quad r_{i-1} < U \leq r_i.$$

**Example** Let  $X \sim \text{Poisson}(4)$ . Table 2 of the *New Cambridge Statistical Tables* [1] gives the cumulative distribution function of  $X$  starting as follows:

|                        |        |        |        |        |        |        |        |     |
|------------------------|--------|--------|--------|--------|--------|--------|--------|-----|
| $i$                    | 0      | 1      | 2      | 3      | 4      | 5      | 6      | ... |
| $\mathbb{P}(X \leq i)$ | 0.0183 | 0.0916 | 0.2381 | 0.4335 | 0.6288 | 0.7851 | 0.8893 | ... |

So the following five values of  $U$  give corresponding values of  $X$  as follows.

|     |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|
| $U$ | 0.1018 | 0.7365 | 0.3248 | 0.3485 | 0.8685 |
| $X$ | 2      | 5      | 3      | 3      | 6      |

If  $X$  is continuous, we do a similar trick using the cumulative distribution function  $F$  of  $X$ . Plot the values of  $X$  along the horizontal axis. The graph of  $F$  lies above this, taking values in the vertical direction between 0 and 1. Let  $u$  be a random value in  $[0, 1]$ ; this defines the point  $(0, u)$  on the vertical axis; draw a horizontal line through this point until the line cuts the graph of  $F$ , then drop a vertical line from here to meet the horizontal axis at the point  $(x, 0)$ . We take  $x$  as our value of  $X$ . The next theorem tells us that our simulated  $X$  does indeed have the right distribution.

**Theorem 14** Let  $F$  be the cumulative distribution function of a continuous random variable. Then  $F$  is monotonic increasing, so it has an inverse function  $F^{-1}: [0, 1] \rightarrow \mathbb{R}$ . Let  $U \sim \text{uniform}[0, 1]$  and put  $X = F^{-1}(U)$ . Then the cumulative distribution function of  $X$  is  $F$ .

**Proof** We have  $U \sim \text{uniform}[0, 1]$ , so

$$\mathbb{P}(U \leq u) = \begin{cases} 0 & \text{if } u < 0 \\ u & \text{if } 0 \leq u \leq 1 \\ 1 & \text{if } u > 1 \end{cases}$$

Therefore

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(F^{-1}(U) \leq x) \\ &= \mathbb{P}(F(F^{-1}(U)) \leq F(x)), && \text{because } F \text{ is monotonic increasing,} \\ &= \mathbb{P}(U \leq F(x)) \\ &= F(x), && \text{because } 0 \leq F(x) \leq 1. \quad \blacksquare \end{aligned}$$

**Example** Let  $X \sim \text{Exp}(2)$ . Then

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-2x} & \text{if } x \geq 0 \end{cases}$$

If  $u = F(x) = 1 - e^{-2x}$  then  $1 - u = e^{-2x}$ . Now,  $1 - u$  is positive, so that we can take natural logarithms of both sides, which gives  $\ln(1 - u) = -2x$ . Therefore  $x = -[\ln(1 - u)]/2$ . This tells us that

$$F^{-1}(u) = -\frac{\ln(1 - u)}{2}.$$

Applying this function  $F^{-1}$  to the same five values of  $U$  as before gives us the following five values of  $X$ .

|     |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|
| $U$ | 0.1018 | 0.7365 | 0.3248 | 0.3485 | 0.8685 |
| $X$ | 0.0537 | 0.6669 | 0.1964 | 0.2142 | 1.0144 |

---

[1] D. V. Lindley and W. F. Scott, *New Cambridge Statistical Tables*, Cambridge University Press.