Example of Simulation

Let us check to see how a calculation of probabilties and expectations which can be found as theoretical quantities by calculus can also be found by computer simulation.

Suppose it is desired to find, for a random variable X **with specified density equal to** $f_X(x) = 3x^2/26, \ \ 1 \le x \le 3$ **, the two quantities (a)** $P(1.5 \le X \le 2.5)$**, and (b)** $E(X^2 - 5X)$**.**

First, by calculus, we get the 'theoretical' answers

$$P(1.5 \le X \le 2.5) \;=\; \int_{1.5}^{2.5} \frac{3x^2}{26}\,dx \;=\; \left(\frac{x^3}{26}\right)\Big|_{1.5}^{2.5} \;=\; \frac{2.5^3 - 1.5^3}{26} \;=\; 0.4711$$

and

$$E(X^2 - 5X) \;=\; \int_1^3 (x^2 - 5x)\,\frac{3x^2}{26}\,dx \;=\; \left(\frac{3x^5}{130} - \frac{15x^4}{104}\right)\Big|_1^3 \;=\; -5.9538$$

Our next step is to calculate the same quantities by using the intuitive relative-frequency and numerical-average interpretations respectively of probabilities and expectations, using artificial or simuated data. We will construct independent and identically distributed random variables $X_1, X_2, \ldots, X_{10000}$ with the specified density f_X above, using the idea we learned in class, from the psudo-random uniform variables $U_1, \ldots, \ldots, U_{10000}$ which we can readily generate on the computer. As we learned in class, the idea is to take the inverse of the distribution function F_X and evaluate it successively at each of the generated variates U_i. So our first step still involves calculus: for $1 \le x \le 3$,

$$F_X(x) \;=\; \int_1^x \frac{3w^2}{26}\,dw \;=\; (x^3 - 1)/26$$

Therefore, solving $y = F_X(x) = (x^3 - 1)/26$ for y, we have

$$F_X^{-1}(y) \;=\; \left(26y \,+\, 1\right)^{1/3}$$

Now the generated variables X_i are calculated very quickly as $X_i = (26U_i + 1)^{1/3}$. I generated and stored a batch of 10000 random variable values X_i in this way.

The simulated probability is now a relative frequency:

$$10^{-4} \cdot \sum_{i=1}^{10000} I_{[1.5 \le X_i \le 2.5]} \;=\; 0.4706$$

and the expectation is approximated, or 'estimated' in statistical language, by

$$10^{-4} \cdot \sum_{i=1}^{10000} (X_i^2 - 5X_i) \;=\; -5.9450$$

Of course, the left-hand sides of each of the last two displayed equations are still random variables, so if I generated a further (independent and identically distributed) batch of 10000 variables X_i^*, $i = 1, 2, \ldots, 10000$, and recalculated the averages

$$10^{-4} \cdot \sum_{i=1}^{10000} I_{[1.5 \le X_i^* \le 2.5]} \quad , \qquad 10^{-4} \cdot \sum_{i=1}^{10000} ((X_i^*)^2 - 5X_i^*)$$

I should get results very close, but not identical, to those obtained before. In fact, when I did this using ten subsequent batches of 10000 variables each, I got the following answers:

	Probability Average (a)	Expectation Average (b)
Run 1	0.4708	-5.9600
2	0.4681	-5.9539
3	0.4694	-5.9573
4	0.4616	-5.9620
5	0.4784	-5.9579
6	0.4716	-5.9615
7	0.4644	-5.9528
8	0.4650	-5.9502
9	0.4696	-5.9580
10	0.4752	-5.9653