Problems on simulating random variables and vectors

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. The time in days, T, that a manufacturing system is out of operation every time it breaks down has cumulative distribution function

$$F_T(t) = \begin{cases} 1-(2/t)^3 & \text{if } t>2 \\ 0 & \text{otherwise} \end{cases} .$$

   Apply the inverse transform method in order to write a function that simulates from the cdf above.

2. Random variable T follows a Pareto distribution with scale parameter 2 and shape parameter 3. We can use its density mass function as instrumental dmf in order to simulate from a Pareto model whose tail is not as heavy. Consider now a random variable S (Pareto with scale parameter 2 and shape parameter 4) with cdf

$$F_S(t) = \begin{cases} 1-(2/t)^4 & \text{if } t>2 \\ 0 & \text{otherwise} \end{cases} .$$

   a. Determine M such that $f_S(t) \le M f_T(t)$, where f_S and f_T are the density mass functions of S and T.
   b. Build an acceptance-rejection algorithm to simulate from F_S.
   c. Use your simulation algorithm to check that $\mathbb{E}[S] = 8/3 = 2.6667$. In order to do so, build a 99% CI algorithm on it out of $MC = 10000$ observations.

3. In Hossack, Pollard and Zehnwirth, CUP (2003) we find the table below with claim counts and claim amounts

Claim amount (euros)	Counts
0-50	1728
50-100	1346
100-200	1869
200-400	1822
400-800	1056
——	——
Total	7821

Estimate the value of the cdf of variable *claim amount* at values $\{0, 50, 100, 200, 400, 800\}$ and use linear interpolation to approximate the cdf. Generate 10000 observations from such a distribution and represent them in a histogran with bins $[0, 50]$, $[50, 100]$, $[100, 200]$, $[200, 400]$ and $[400, 800]$. Run the chi-square test with those $k = 5$ levels.

4. A rv X follows a truncated normal distribution in the interval $[a, b]$ with parameters μ and σ, $X \sim \text{TN}(\mu, \sigma; a, b)$, if its density mass function is given by

$$f_X(x) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}, \quad a \le x \le b,$$

   where $\phi(\cdot)$ is the dmf of a standard normal distribution and $\Phi(\cdot)$ its cdf. Build three functions to simulate observations of a $\text{TN}(0, 1; -1, 1)$ truncated normal random variable by means of
   a. A rejection algorithm whose candidate function is the dmf of a standard normal distrtibution

b. A rejection algorithm whose candidate function is the dmf of a $\mathrm{U}(-1, 1)$ distribution.
c. A inverse transform technique. and compare their behaviour in terms of efficiency.

5. Alice, Bob, Charly, and Dave share an apartment that has only one bathroom. They have decided that every morning from 7 to 8 each one of them can use the bathroom during 15 minutes. The bath turns (1st turn 7 to 7:15, 2nd turn 7:15 to 7:30, 3rd turn 7:30 to 7:45, 4th turn 7:45 to 8) must be completely random, with the single restriction that Alice must have one of the first turns, since she must leave home at 7:50. They have decided to proceed as follows: they select (at random) who takes the 1st turn, then they select at random who takes the 2nd turn, if Alice is not taking any of them, she takes the third and whoever has not yet used the bathroom at 7:45 takes the fourth. If Alice has taken one of the two first turns, they select that random who takes the 3rd turn, and the other faltmate takes the 4th turn.
   a. Is that method fair in the sense that all allowed permutations have the same probability?, why?
   b. Build a fair algorithm to distribute the turns (with the given restriction).
6. Consider the sepal length of the 50 iris setosa plants of dataset `iris`. Build a bivariate normal random vector whose first component is formed by those observations (notice that last week we checked that it is normally distributed) and the second component has mean 5, standard deviation 0.3 and the correlation is 0.6.