

Computer Practical 1

Simulation Basics

1. Transformation Methods

Recall the Box-Muller method which transforms pairs of uniformly-distributed random variables to obtain a pair of independent standard normal random variates. If

$$U_1, U_2 \stackrel{\text{iid}}{\sim} U[0, 1]$$

and

$$X_1 = \sqrt{-2 \log(U_1)} \cdot \cos(2\pi U_2)$$

$$X_2 = \sqrt{-2 \log(U_1)} \cdot \sin(2\pi U_2)$$

then $X_1, X_2 \stackrel{\text{iid}}{\sim} N(0, 1)$.

- (a) Write a function which takes as arguments two vectors ($\mathbf{U}_1, \mathbf{U}_2$) and returns the two vectors ($\mathbf{X}_1, \mathbf{X}_2$) obtained by applying the Box-Muller transform elementwise.
 - (b) The R function `runif` provides access to a PRNG. The type of PRNG can be identified using the `rngkind` command; by default it will be a Mersenne-Twister (http://en.wikipedia.org/wiki/Mersenne_twister). Generate 10,000 $U[0, 1]$ random variables using this function, convert this to two vectors of 5,000 elements in the same way as you did with the RANDU values.
2. *The Bootstrap*: This question can be answered in two ways. The more direct (and perhaps more informative, if you have the time to do so and the inclination to implement a bootstrap algorithm from scratch) is to use the `sample` function to obtain bootstrap replicates and to compute the required confidence intervals by direct means. More pragmatically, the `boot` library provides a function `boot` to obtain bootstrap samples and another, `boot.ci` which will provide various bootstrap confidence intervals.
- (a) The `nile` dataset shows the rivers. Use a histogram or other visualisation to briefly explore this data.
 - (b) What's the mean and median length of a river in this category?
 - (c) Treating the data as a simple random sample, appeal to asymptotic normality to construct a confidence interval for the mean annual flow of the Nile.
 - (d) Using the `boot::boot` function to obtain the sample and `boot::boot.ci` to obtain confidence intervals from that sample, or otherwise, obtain a *bootstrap percentile interval* for both the mean and median of the Nile's annual flow. For the median you may also wish to obtain the interval obtained by an optimistic appeal to asymptotic normality combined with a bootstrap estimate of the variance (`boot::boot.ci` will provide this).
- Note the following: `boot::boot` does the actual bootstrap resampling; it needs a function which takes two arguments to compute the statistic for each bootstrap sample, the first contains the *original* data and the second the index of the values included in a particular bootstrap resampling.
- (e) Are there any interesting qualitative differences between the various confidence intervals? How does this relate to the data?
 - (f) Are your findings stable? If you repeat the bootstrap sampling do you recover similar behaviour?
 - (g) Are there any reasons to doubt the accuracy of these confidence intervals?

3. Convergence of Sample Approximations

- (a) The `stats::ecdf` and `stats::plot.ecdf` functions compute and plot empirical distribution functions from a provided sample.
- i. Show plots of the empirical distribution function of samples of a variety of sizes ranging from 10 to 10,000 from a $U[0, 1]$ distribution. Add to your plots the distribution function of the $U[0, 1]$ distribution.
 - ii. Repeat part *i.* with a standard normal distribution.
 - iii. Repeat part *i.* with a Cauchy distribution.
- (b) For each of the three distributions considered in the previous part, determine $\sup_x |\hat{F}_n(x) - F(x)|$ for each n considered (consider the sup over *only* the sampled values of x) and plot these quantities against n . Do you notice anything interesting?

★ Indicates questions which can easily be done later if you're short of time.