Extra Exercises on Programming with R

1. A Galton board consists of a vertical board with series of peg. When a ball is released on top of the Galton board, the ball will bounce left or right at each peg and the final position of the ball is recorded. Figure 2 shows an example of a Galton board.

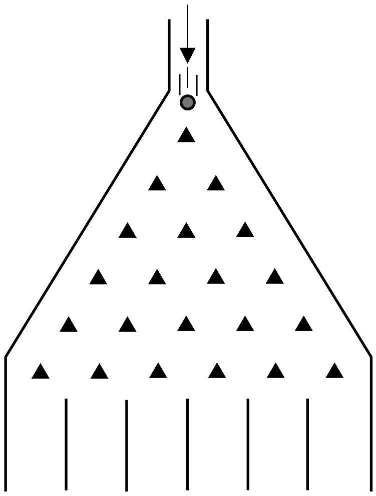

Figure 2. An illustration of a Galton board.

Let X denotes the position of a ball in the Galton board. Let $X = 0$ be the starting position of the ball when it is released at the top. At each step or level of pegs, the ball will either bounce left of right with equal probability. If the ball bounces left, then the deduct 1 from X, i.e., $X_{new} = X_{old} – 1$. If the ball bounces right, then add 1 to X, i.e., $X_{new} = X_{old} – 1$. In this question, we are interested in the final value of X after k number of steps.

a) Write a function in R to generate a value of X after k steps. The function takes input k, which is the number of steps, and returns the value of X.

b) Using the function written in question (a), write a function in R to generate n samples of X after k steps. The function takes input n and k, which are the number of samples and the number of steps, respectively. The function returns a vector of X of length n.

c) Using the function written in question (b), estimate the mean and variance for X using $k = 20$. Estimate the same values for $k = 30$.

d) In a game of chance, a ball is dropped into the Galton board with 20 levels of pegs. The game costs \$10 to play but if the position of the ball, X, where X is as defined previously, is greater than or equal to 8, then the player wins \$100.

i) Modify the function written in question (b) such that the function now has a list as the output, which consists of the vector X as well as the returns for each X. Note that if $X < 8$, then the return is \$0, whereas if $X \geq 8$, then the return is \$100.

ii) Using the function written in question (d) (i), estimate the profit or loss for playing the game.

2. According to Chebyshev's theorem, for any numerical data set, at least $1 - \frac{1}{k^2}$ of the data lie within k standard deviations of the mean, i.e. within the interval with endpoints $\bar{x} \pm ks$ for samples or $\mu \pm k\sigma$ for populations, where k is any positive whole number that is greater than 1.

a) Build an R function that, given any sample data set X and any positive whole number k which is greater than 1, measures the proportion of data that lies within $\bar{x} \pm ks$ where $\bar{x}$ and s are the mean and standard deviation of sample X respectively.

b) By making use of the function in (a), check for the validity of the Chebyshev's theorem for 100 samples of random numbers which follows a standard normal distribution of length 10 000.

c) Check that the Chebyshev's theorem is also applicable to samples that are not normally distributed. You are free to use any other distribution.