The University of British Columbia

Data Science 580 Modelling and Simulation I

Lab Assignment 3

1. (4 marks) Under the convention that log() represents the natural logarithm, suppose that $F(x) = \log(x)$ for $x \in [1, \mathrm{e}]$, and $F(x) = 0$, for $x \leq 1$ and $F(x) = 1$, for $x > e$. Let X be a random variable having $F(x)$ as its cumulative distribution function (cdf).

   (a) (1 mark) Find the inverse function of $F(x)$.

   (b) (1 mark) Generate 100000 random uniform variates on the interval $[0, 1]$ assigning them to an object called `U`, and apply the inverse cdf method, assigning the resulting values to an object called `X`.

   (c) (1 mark) Estimate $E[X]$ and $\mathrm{V}(X)$ using the `mean()` and `var()` functions, and estimate the probability $P(X < 2)$.

   (d) (1 mark) Find the theoretical pdf of X, and using integration, calculate $E[X]$ and compare with the estimate you obtained earlier. Also, calculate $E[X^2]$ and $\mathrm{V}(X)$. Compare with the variance estimate obtained earlier.

2. (3 marks) Gamma distributions are common in engineering models. Examples include equipment downtime and load levels for telecommunication services, meteorological rainfall, commercial insurance claims and loan defaults, where the variables are always positive and the results are skewed. The gamma distribution contains the exponential and χ^2 distributions as special cases. Like the Weibull, it has two parameters, a shape parameter and a scale parameter. It is also used for modeling survival times.

   (a) (1 mark) Explore some possible shapes of the gamma distribution using the `curve()` and `dgamma()` functions. Specifically, obtain the graphs from the following code:

   ```
   curve(dgamma(x, shape = 0.5, scale = 1), 0, 5)
   curve(dgamma(x, shape = 1, scale = 1), 0, 5, add=TRUE, col=2)
   curve(dgamma(x, shape = 3, scale = 1), 0, 5, add = TRUE, col=3)
   ```

   Comment on the shapes of each density curve.

   (b) (1 mark) Suppose X is a gamma random variable with shape parameter 5 and scale parameter 2. Use the `pgamma()` function to calculate $P(X \leq 3)$ and $P(X \leq 7)$.

   (c) (1 mark) Use the `rgamma()` function to simulate 100000 values of X, assigning them to `X`. Use the `mean()` function to estimate $P(X \leq 3)$ and $P(X \leq 7)$. Compare with the theoretical values.

3. (7 marks) Suppose X is a normal random variable with mean 5 and standard deviation 2.

   (a) (1 mark) Use the pnorm() function to calculate $\mathrm{P}(\mathrm{X} \leq 3)$ and $\mathrm{P}(\mathrm{X} \leq 7)$.

   (b) (1 mark) Use the rnorm() function to simulate 10000 values of X, assigning them to X. Use the mean() function to estimate $\mathrm{P}(\mathrm{X} \leq 3)$ and $\mathrm{P}(\mathrm{X} \leq 7)$. Compare with the theoretical values.

   (c) (3 marks) Suppose $\mathrm{Y} = 2\mathrm{X} + \mathrm{Z}$ where Z is a standard normal random variable.

   (1 mark) i. Use X and the rnorm() function to simulate 10000 Y values.

   (1 mark) ii. Use plot(X, Y) to obtain a scatterplot of the Y values, together with the corresponding X values. Briefly comment on the pattern observed.

   (1 mark) iii. Overlay the plot with a line using abline(0, 2). Does this line pass through the scatterplot as you might expect, or not?

(d) (1 mark) Repeat the previous part, but where Z has standard deviation 0.5, instead of 1.0.

(e) (1 mark) How does the scatterplot change as you change the value of the standard deviation of Z?