The University of British Columbia

Data Science 580 Modelling and Simulation I

Lab Assignment 4

1. Let X be a N(3, 5) normal random variable with mean 3 and standard deviation 5. Simulate $m \times n$ pseudorandom numbers, assigning them to a matrix M with $m = 10000$ rows and $n = 100$ columns. Calculate the sample mean denoted as X_n in each column and use all the sample means to estimate $P(3 < X_n < 4)$. Compare it with the theoretical value.

```
m <- 10000; n <- 100; sigma <- 5
# m samples of size n:
M <- matrix(rnorm(m*n, mean = 3, sd = sigma), nrow=n)
xbar <- apply(M, 2,mean)
mean(xbar<4)-mean(xbar<2)

## [1] 0.953

1-2*pnorm((2-3)/sqrt(25/n))

## [1] 0.9544997
```

   (a) Estimate $P(X_n > 4.5)$ and compare it with the theoretical value.

   (b) Are the sample mean X_n and sample standard deviation S_n independent? Plot the sample mean against the sample standard deviation

   ```
   sigmahat=apply(M, 2,sd)
   ```

   (c) Verify if $E(X_n/S_n) = E(X_n)E(1/S_n)$ is correct.

2. Autoregressive models may attempt to predict the future price of a stock based on its past performance.

   (a) Simulate 100 observations for an AR(1) process with parameters .2, with a standard deviation of 1 (the default).

   (b) Obtain the acf plot. Comment on the pattern of autocorrelations.

   (c) Repeat (a) and (b) for 100 observations of an AR(1) process with parameter -.8. Comment on any differences that you see between the two different AR(1) processes that you simulated.

3. (a) Simulate 100 observations on the predictive model, using `set.seed(1333)`,

$$y = 2 + 3x + \varepsilon$$

   for the cases where (i) ε is white noise with standard deviation 1 (ii) ε is AR(1) with parameter 0.8 and standard deviation 1. Use `x <- seq(0, 1, length=100)` as the sequence of x values. Store the first result in `y0` and the second in `y1`.

   (b) Run the regressions

```r
y0.lm <- lm(y0 ~ x)
y1.lm <- lm(y1 ~ x)
```

and determine the regression coefficients for each fitted model using `coef(y0.lm)` and `coef(y1.lm)`. Write out the fitted model in each case.

(c) Use each of your models to predict a new observation at $x = 0.8$. What is the true expected value of such a value? Which of the two predictions is closer to this value?

(d) Comment on the possible effects of dependence in the noise term on the accuracy of regression models, based on the observation above.