The University of British Columbia

Data Science 580 Modelling and Simulation I

Lab Assignment 2

Instructions: The TA will demonstrate exercises 6 through 9. You are expected to submit answers to exercises 1 through 5. Please use a png, pdf or html file for submission.

1. Suppose X is a uniform random variable on the interval $[a, b]$. Use calculus to compute the following expected values:
   (a) $\mu = E[X]$.
   (b) $E[(X - \mu)^3]$. (This gives a measure of skewness. Positive values indicate right-skewness and negative values indicate left-skewness, and 0 indicates symmetry.)
   (c) $E[(X - \mu)^4]$. (This gives a measure of kurtosis, or heavy-tailed-ness in the distribution.)

2. Simulate 100000 random uniforms on the interval $[0, 1]$, assigning them to an object called `X`.
   (a) Compute the mean of `X` and compare with the theoretical value.
   (b) Compute the mean of `(X-mean(X))^3` and compare with the theoretical value.

3. Simulate 10000 binomial pseudorandom numbers with parameters 20 and 0.3, assigning them to a vector called `binsim`. Let X be a binomial(20, 0.3) random variable. Use the simulated numbers to estimate
   (a) $P(X \leq 5)$.
   (b) $P(X = 5)$.
   (c) $E[X]$.
   (d) $\mathrm{Var}(X)$.

4. Estimate the mean and variance of a Poisson random variable whose mean is 7.2 by simulating 10000 Poisson pseudorandom numbers. Compare with the theoretical values.

5. Simulate vectors of 10000 pseudorandom Poisson variates with mean 5, 25, 125 and 625, assigning the results to `P1`, `P2`, `P3` and `P4`, respectively.
   (a) Estimate $E[X]$, where X is Poisson with rates $\lambda = 5, 25, 125$ and 625.
   (b) The $\log(1+X)$ transformation is typically used to convert count data that contains zero values, such as the COVID-19 infectious count. Estimate $E[\log(1+X)]$, where X is Poisson with rates $\lambda = 5, 25, 125$ and 625.
   (c) To analyze a series of hourly frequencies of ventricular premature depolarizations (VPDs), we used the natural logarithmic transformation to take into account the statistical assumptions of normality and homogeneity of variance of the data. However, the square root transformation may be more appropriate in this case. Estimate $E[\sqrt{X}]$, where X is Poisson with rates $\lambda = 5, 25, 125$ and 625.

6. The following function simulates `N` random samples of uniformly distributed numbers, each of size `n`, returning the average of each sample.

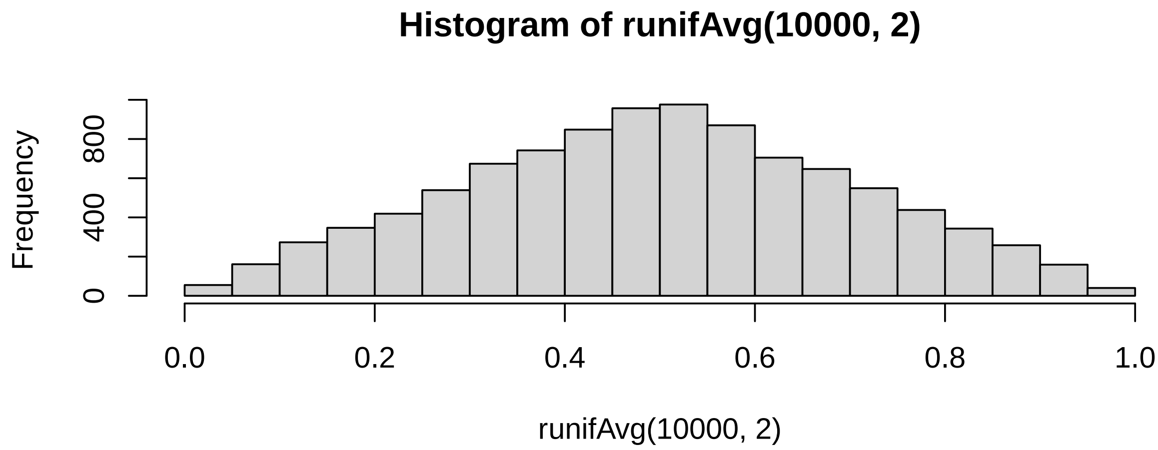

Figure 1:

```r
runifAvg <- function(N, n) {
  X <- matrix(runif(N*n),N,n)
  rowMeans(X)
}
```

For example, the averages of 5 independent samples of size 2 are

```r
runifAvg(5,2)

## [1] 0.5149873 0.5373508 0.4378053 0.4321671 0.4910189
```

A histogram of the averages of 10000 independent samples of size 2 is obtained from

```r
hist(runifAvg(10000,2))
```

and pictured in Figure 1.

Copy the function and paste it into an R session and construct histograms of the averages of 10000 independent samples of size 3, 6, 10, 15, 30 and 100. What do you observe as the sample size increases?

7. Replace `runif(N*n)` with `rbinom(N*n, 1, 0.25)` in `runifAvg`.

   Paste the resulting function into an R session and construct histograms of the averages of 10000 independent samples of size 2, 3, 6, 10, 15, 30 and 100. What do you observe as the sample size increases? Do you think a larger sample is needed to achieve a stable 'limit' in this case than in the previous case?

8. Simulate a Poisson process with rate $\lambda = 1.5$ on the interval $[0, T] = [0, 100000]$ using the `rpois` and `runif` functions. The process points should be assigned to the vector `PoissonPoints`.

9. Use the `cut()` function to help count the numbers of points of the Poisson process in each of intervals $[0, 1], [1, 2], \ldots, [99999, 100000]$.

Use the `table()` function to estimate the probability distribution and compare with the Poisson distribution which has rate $\lambda = 1.5$.