The University of British Columbia

Computer Science/Data Science 405/505 Modelling and Simulation

Assignment 2 – Solution

Exercises

1. Use 1000 pseudorandom numbers generated by the `runif()` function to simulate values from the probability distribution of a random variable X which takes the value 0, with 50% probability, 1 with 30% probability and 2, with 20% probability. To do this exercise, undertake the following steps.

   (a) First, verify for yourself that the cumulative distribution function for the random variable X takes the value 0.5 at 0, 0.8 at 1, and 1.0 at 2. In other words, $P(X \leq 0) = 0.5$, $P(X \leq 1) = 0.8$ and $P(X \leq 2) = 1.0$.
   $P(X \leq 0) = P(X = 0) = 0.5$.
   $P(X \leq 1) = P(X = 0) + P(X = 1) = .5 + .3 = 0.8$.
   $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 1.0$.

   (b) Next, write a function called `pX`, as in Example 4.6 of the textbook, which takes an argument `x` and returns the value of $P(X \leq x)$.

```
pX <- function(x)  {
    return(c(.5, .8, 1)[x+1])
}
```

   (c) Now, imitate Example 4.7 of the textbook to write a function `rX` which takes `n` as an argument and returns a vector of length n consisting of random numbers that follow the distribution of X.

```
rX <- function(n) {
    U <- runif(n)
    X <- numeric(n)
    for (x in 0:1) {
        X[U >= pX(x)] <- x + 1
    }
    return(X)
}
```

   (d) Assign output from the function `rX(1000)` to an object called `myX`.

```
myX <- rX(1000)
```

   (e) Use the `table()` function to count the numbers of 0's, 1's and 2's in the vector `myX`. How do these counts differ from what you would have expected?

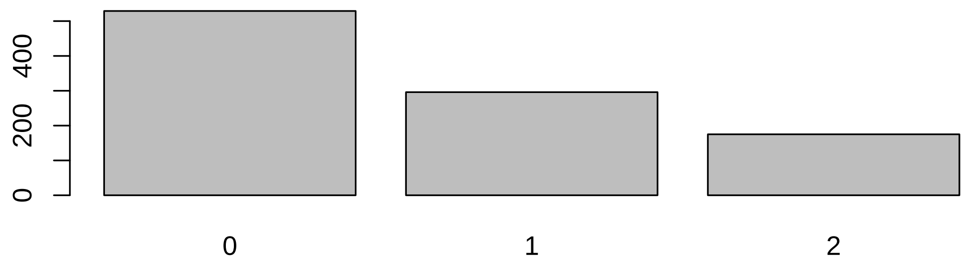
```
table(myX)

## myX
##   0   1   2
## 529 296 175
```

   The observed counts are slightly different from what we would expect: 500, 300, 200.

(f) Construct a bar plot which displays the observed distribution of values from `myX`.

```
barplot(table(myX))
```



2. 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) $\text{Var}(X)$.

```
set.seed(366360) # use this to get the output below
X<-rbinom(10000,20,0.3)
```

(a)
```
mean(X<=5)

## [1] 0.4166
```

(b)
```
mean(X==5)

## [1] 0.1819
```

(c)
```
mean(X)

## [1] 6.0084
```

(d)
```
var(X)

## [1] 4.139543
```

3. 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) Use these simulated datasets to estimate $E[X]$, $E[\sqrt{X}]$, $\text{Var}(\sqrt{X})$ and $\text{Var}(X)$, where X is Poisson with rates $\lambda = 5, 25, 125$ and 625.

```
P1<-rpois(10000,5)
P2<-rpois(10000,25)
P3<-rpois(10000,125)
P4<-rpois(10000,625)
```

```
mean(P1)

## [1] 5.0074

mean(sqrt(P1))

## [1] 2.171536

var(sqrt(P1))

## [1] 0.2918597

var(P1)

## [1] 5.132458

mean(P2)

## [1] 24.9912

mean(sqrt(P2))

## [1] 4.973686

var(sqrt(P2))

## [1] 0.2536767

var(P2)

## [1] 25.07423

mean(P3)

## [1] 124.9927

mean(sqrt(P3))

## [1] 11.16874

var(sqrt(P3))

## [1] 0.2520104

var(P3)

## [1] 125.6108

mean(P4)

## [1] 625.4838

mean(sqrt(P4))
```

```
## [1] 25.00479
var(sqrt(P4))
## [1] 0.2440839
var(P4)
## [1] 610.3722
```

(b) Noting that the variance of X increases with the mean of X, when X is a Poisson random variable, what is the effect of taking a square root of X on the relationship between the variance and the mean? (Statisticians often take square roots of count data to 'stabilize the variance'; based on what you saw in (a), can you explain what this means?)

After taking the square root of the measurements, the mean still increases as a function of rate, but the variance appears to be nearly constant, around 0.24 to 0.26.