Second Year R Coursework Outline

This is a sample of some coursework produced in 2019 for *MAS2602:Computing for Mathematics and Statistics,* a second-year undergraduate module at Newcastle University.

Question 1

Requires use of Monte Carlo simulation to estimate the value of an integral.

Question 2

Given the CDF of a random variable, create a function which creates a sample from its distribution. *d* is a given parameter from which to generate a sample.

Question 3

Create a function to sample from the log-normal distribution and investigate its properties.

MAS2901: R Assignment

Joseph Marks

Thursday 21st November

Question 1

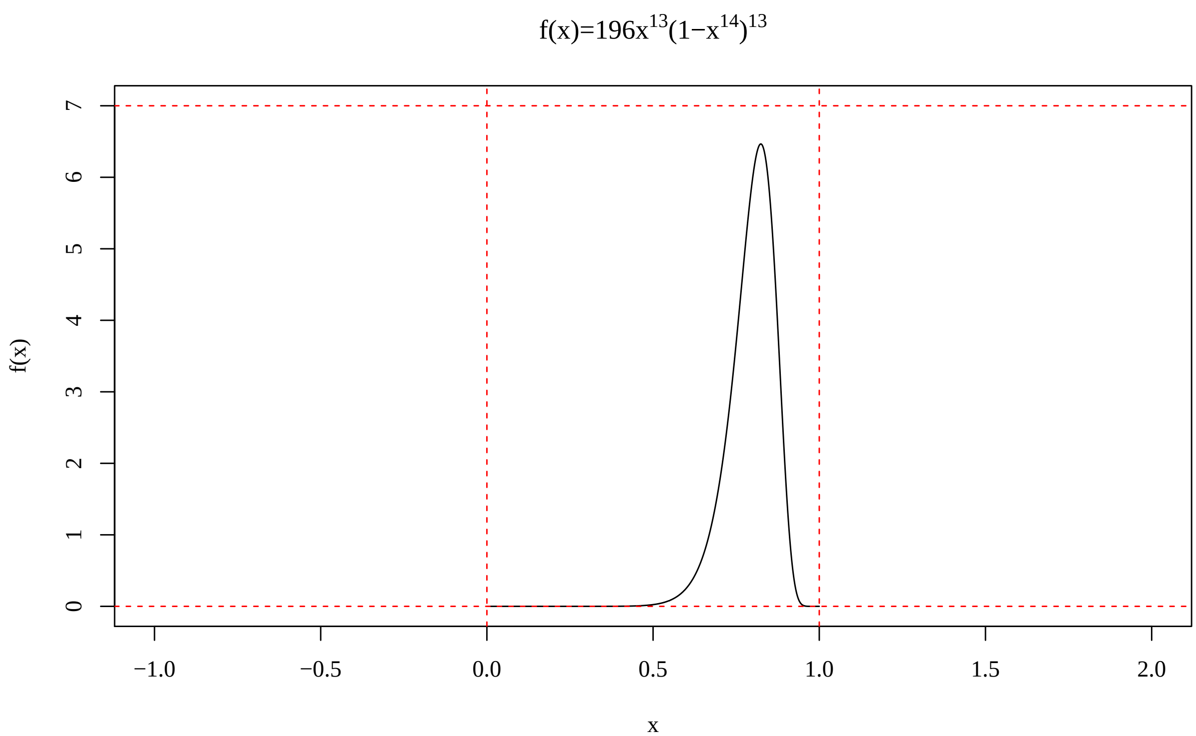
Part A

```r
set.seed(8063804)
a = sample(3:14, 1)
b = sample(3:14, 1)
```

Output: $a = 14$ and $b = 14$.

$\therefore f(x) = 14^2 x^{13}(1 - x^{14})^{13} = 196x^{13}(1 - x^{14})^{13}$

Part B



Part C

```
N = 1000000 #Number of simulations
no_of_hits = 0 #Set hit counter to zero

for (i in 1:N){
  x1 = runif(1, 0, 1) #Generate random x co-ordinate
  y1 = runif(1, 0, 7) #Generate random y co-ordinate
  f_x1 = 196*(x1^13)*((1-x1^14)^13) #Evaluate f(x) at generated values

  if (y1<f_x1){ #Find if generated y value is below function
    no_of_hits = no_of_hits + 1 #Increment hit counter
    }
}

P = no_of_hits/N #Find proportion of points below function
area_under_curve = P*7 #Product of P and area of sample grid
```

With `set.seed(8063804)` this outputs an area under the curve equal to $0.997661 \approx 1$.

In context, this suggests $\displaystyle\int_0^1 196x^{13}(1-x^{14})^{13}\mathrm{d}x \approx 0.997661$.

Question 2

We suppose X is a random variable with PDF:

$$f_X(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}} \text{ for } x > 0.$$

Part A

The CDF of X is given by:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)\mathrm{dt} \\ &= \int_0^x \frac{t}{\sigma^2}e^{-\frac{t^2}{2\sigma^2}}\mathrm{dt} \end{aligned}$$

We introduce the substitution $u = \frac{t^2}{2\sigma^2} \iff \sigma^2\mathrm{du} = t\mathrm{dt}$. This transforms the integral to:

$$\begin{aligned} F_X(x) &= \int_0^{x_u} e^{-u}\mathrm{du} \\ &= -\left[e^{-u}\right]_0^{x_u} \\ &= -\left[e^{-\frac{t^2}{2\sigma^2}}\right]_0^x \end{aligned}$$

Therefore $F_X(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$

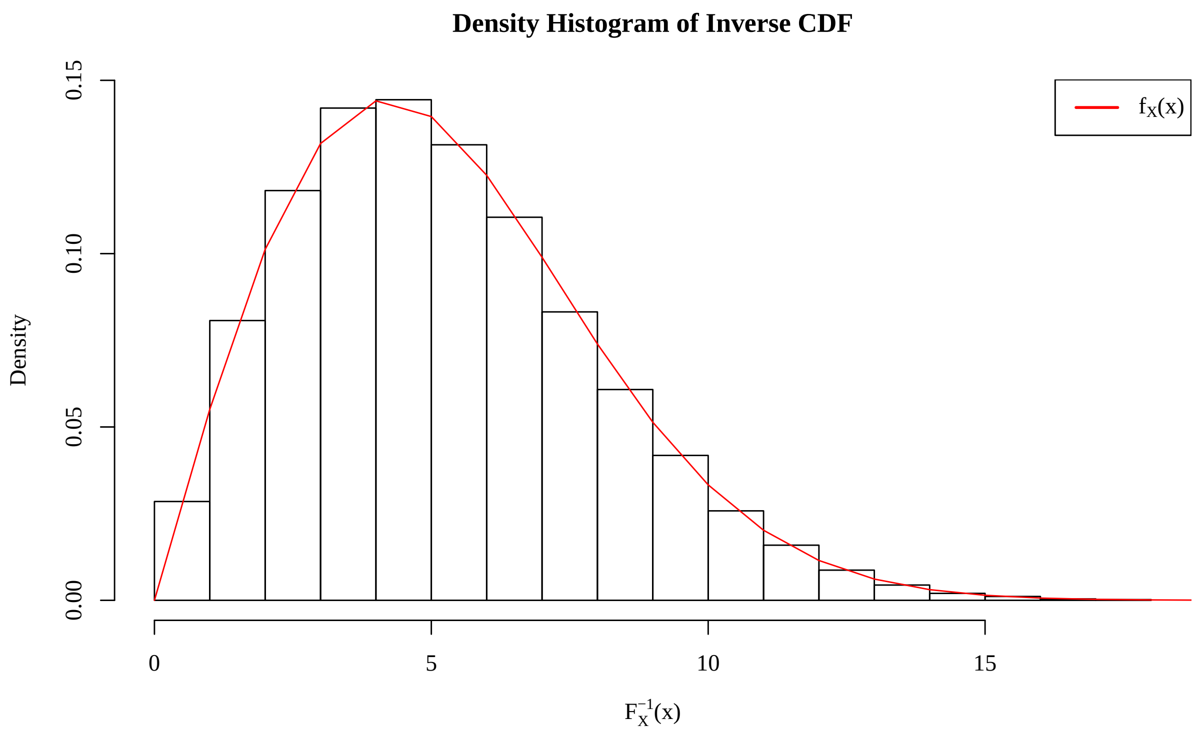
Part B

We note that $F_x^{-1}(x) = \sqrt{-2\sigma^2 \log(1-x)}$

```
inv.cdf = function(N,k){
  X = 1:N #Create a vector size N
  for(i in 1:N){ #Go through each element of vector
    U = runif(1) #Generate values from uniform distribution
    X[i] = sqrt(-2*(k^2)*log(1-U)) #Set index of vector with inverse CDF value
    }
  return(X)
}
```

Part C

We set $d = 8 \implies \sigma = 4.2$. We plot `inv.cdf(10000, 4.2)` and produce:



Part D

see Part C.

Part E

This PDF, $f(x)$, is the definition of the Rayleigh distribution.

Part F

The mean of the Rayleigh distribution is given as:

$$\sigma\sqrt{\frac{\pi}{2}}.$$

The variance of the Rayleigh distribution is given as:

$$\sigma^2\frac{4-\pi}{2}.$$

We compare our results by running the following code:

```
X_out = inv.cdf(10000, sigma)
cat("Expected Mean:", sigma*sqrt(pi/2), "Simulated Mean: ", mean(X_out),
"\nExpected Var:", (sigma^2)*((4-pi)/2), "Simulated Var: ", var(X_out))
```

Which produces:

Expected Mean: 5.263919   Simulated Mean: 5.263012
Expected Var: 7.571153   Simulated Var: 7.457969

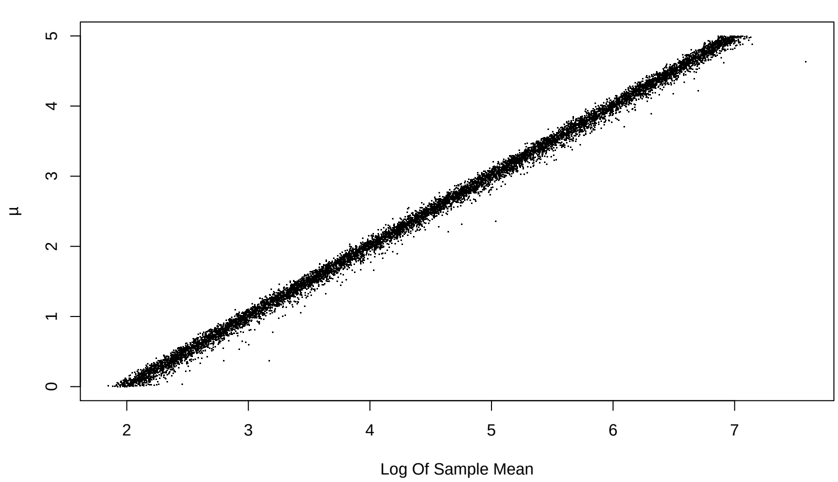
The expected and simulated values are very close, as we would hope.

Question 3

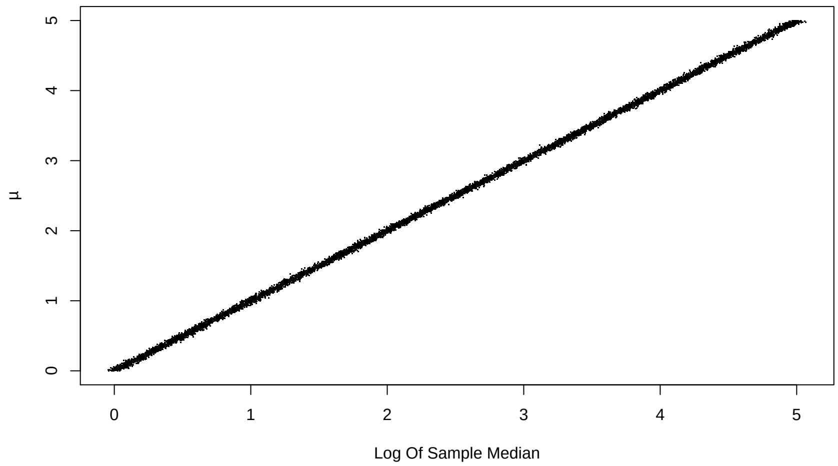
Part A

```
samp_LN = function(N, mu, sigma){
  X = rnorm(N, mu, sigma) #Generate values from normal distribution
  Y = exp(X) #Transform to log-normal
  return(Y)
}
```

Part B



Part C



Part D

Given some fixed value of sigma, it appears the median and means are both functions of the form $\exp(f(\mu))$.