ST 314 Data Analysis 2

Ivan Timothy Halim

January 21, 2019

Part 1

Random Variable 1

The best model for random variable 1 is the Exponential Distribution because times are more likely to be close to 0 and less likely as they get further away from 0.

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt$$

$\mu = 30$ $\qquad \lambda = \frac{1}{\mu} = \frac{1}{30}$

Random Variable 2

We should use the Uniform Distribution because any time between 30 and 90 has an equal likelihood.

$$F(x) = \int_a^x \frac{1}{b-a} dt$$

$a = 30$
$b = 90$

Random Variable 3

We should use the Normal Distribution because the distribution of cylinder lengths is symmetrical, where lengths are more likely to be close to the mean rather than further away from the mean.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\mu = 3.25$
$\sigma = 0.003$

Part 2

a. 90^{th} percentile:

$p(X < 356) = p(Z < z) = 0.9$

$z = 1.28$ (by z-table)

10^{th} percentile:

$p(X < 150) = p(Z < z) = 0.1$

$z = -1.28$ (by z-table)

b. $\mu = \frac{356+150}{2} = 253$

$$\sigma = \frac{x-\mu}{z} = \frac{356-253}{1.28} = 80.37$$

c. $1 - p(X < a) = 0.03$

$p(X < a) = p(Z < z) = 0.97$

$z = 1.88$ (by z-table)

$$z = \frac{x-\mu}{\sigma} = \frac{a-253}{80.37} = 1.88$$

$a = 404.16$

Part 3

a.

i. $f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$

$\alpha = 2$

$\beta = 7$

$\Gamma(\alpha) = (\alpha - 1)! = (2-1)! = 1$

$$f(x) = \frac{1}{49} x e^{-\frac{x}{7}}$$

ii. $\mu = \alpha\beta = 14$

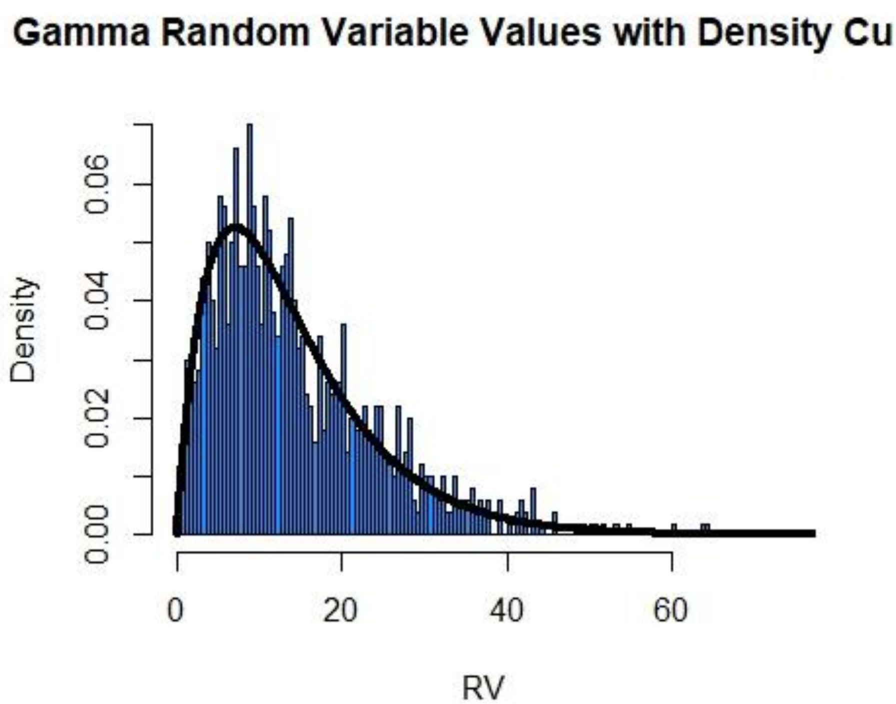
iii. $p(X < 4) = \int_0^4 \frac{1}{49} x e^{-\frac{x}{7}}\, dx = uv - \int v\, du$

$u = x \qquad dv = \frac{1}{49} e^{-\frac{x}{7}} dx$

$du = dx \qquad v = -\frac{1}{7} e^{-\frac{x}{7}}$

$$= -\frac{1}{7} x e^{-\frac{x}{7}} - \int_0^4 -\frac{1}{7} e^{-\frac{x}{7}}\, dx = 0.1126$$

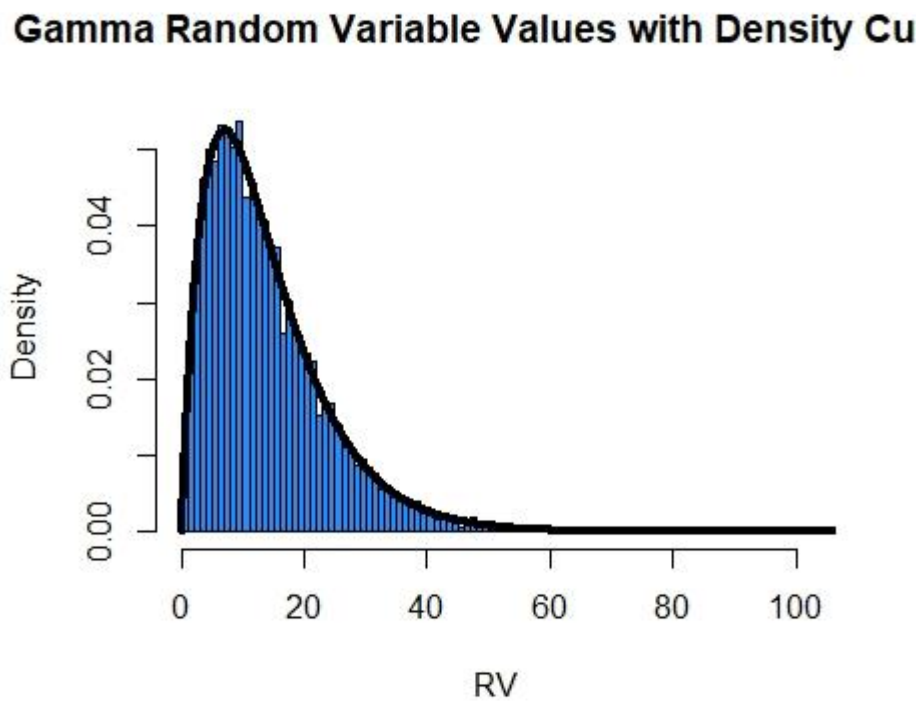
b. The overall shape of the graph resembles the probability density curve although there are some points that lie outside of the density curve.



c. $Actual\ proportion = 0.107$

$$Error = \frac{0.107 - 0.1126}{0.1126} \times 100\% = 4.96\%$$

d. Compared to the previous graph, this graph is a lot smoother and fits very nicely with the probability density curve.

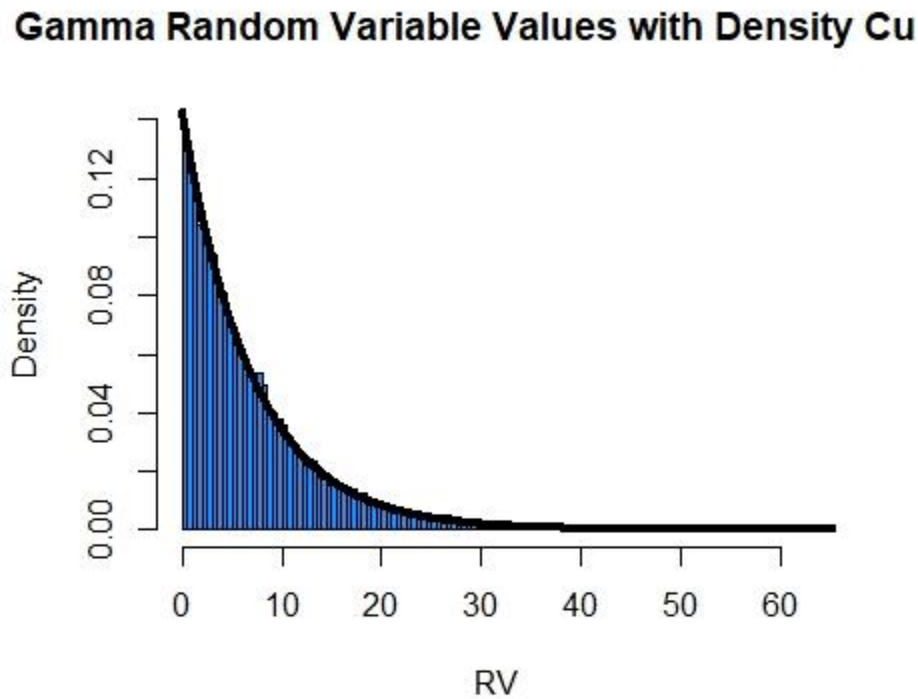


e. $Actual\ proportion\ =\ 0.1173$

$$Error\ = \frac{0.1173 - 0.1126}{0.1126} \times\ 100\%\ =\ 4.19\%$$

Increasing the number of observations improves the accuracy of the model.

f. The shape of the graph resembles the Exponential Distribution.



g. This is an Exponential Distribution. The probability density function is:

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt$$

$$\alpha = 1$$
$$\beta = 7$$
$$\mu = \alpha\beta = 7$$
$$\lambda = \frac{1}{\mu} = \frac{1}{7}$$

$$F(x) = \int_0^x \frac{1}{7} e^{-\frac{t}{7}} dt$$

h. Increasing the alpha (α) moves the maximum value to the right.
Increasing the beta (β) increases the spread of the graph.