Data-580 Lab 3

Dhun Sheth

2023-09-25

Question 1

Part A

$$F(x) = log(x) => U = log(x) => x = e^U$$

Part B

```r
U <- runif(100000)
x <- exp(U)
```

Part C

```r
mean_x <- mean(x)
var_x <- var(x)
prob_x_less_than_2 <- mean(x < 2)
```

$E[x] = 1.7190098$
$Var[x] = 0.2429033$
$P(x < 2) = 0.69219$

Part D

$$pdf = \frac{d}{dx} cdf = \frac{d}{dx} log(x) = \frac{1}{x} = f(x)$$

$$E[x] = \int_a^b x * f(x) dx = \int_1^e 1 \ dx = [x]_1^e = e - 1 = 1.7182818$$

1.7182818

$$E[x^2] = \int_a^b x^2 * f(x) dx = \int_1^e x \ dx = [\frac{x^2}{2}]_1^e = \frac{e^2 - 1}{2} = 3.194528$$

3.194528

$$Var[x] = E[x^2] - E[x]^2 = 3.194528 \ - 1.7182818^2 = 0.2420357$$

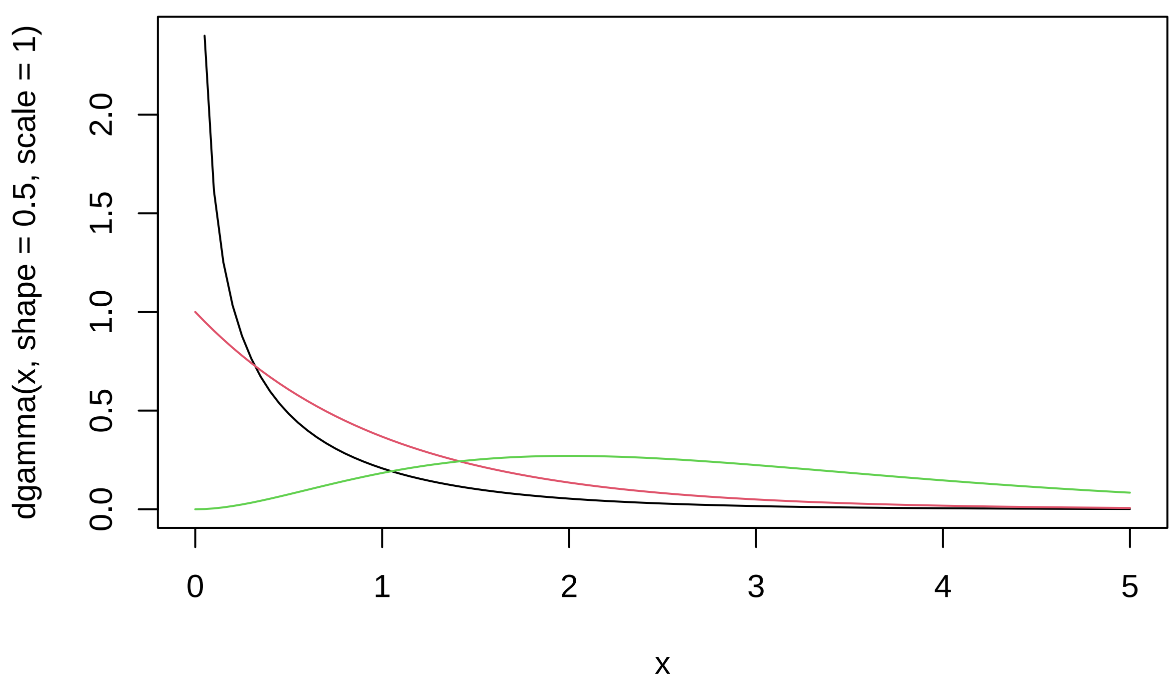
0.2420357

Comparing to the mean and variance estimate from Part C, we can see the values are very close.

Question 2

Part A

```
curve(dgamma(x, shape = 0.5, scale = 1), 0, 5)
curve(dgamma(x, shape = 1, scale = 1), 0, 5, add=TRUE, col=2)
curve(dgamma(x, shape = 3, scale = 1), 0, 5, add = TRUE, col=3)
```



Density curve with shape=0.5 and scale=1 is decreasing to 0 as x goes to infinity, and approached infinity as x goes to 0.

Density curve with shape=1 and scale=1 is decreasing to 0 as x goes to infinity, and approaches 1 as x goes to 0.

Density curve with shape=3 and scale=1 starts at 0,0 and increases, peaking at around x=2 with dgamma highest around 0.4, before decreasing to 0 as x goes to infinity.

Part B

```
prob_3 <- pgamma(3, shape=5, scale=2)
prob_7 <- pgamma(7, shape=5, scale=2)
```

$P(x <= 3) = 0.0185759$
$P(x <= 7) = 0.274555$

Part C

```
x <- rgamma(100000, shape=5, scale=2)
sim_prob_3<-mean(x<=3)
sim_prob_7<-mean(x<=7)
```

Simulated and theoretical values are close to each other.

Simulated Values:
$P(x <= 3) = 0.01937$
$P(x <= 7) = 0.27758$

Theoretical Values:
$P(x <= 3) = 0.0185759$
$P(x <= 7) = 0.274555$

Question 3

Part A

```
prob_norm_3 <- pnorm(3, mean=5, sd=2)
prob_norm_7 <- pnorm(7, mean=5, sd=2)
```

$P(x <= 3) = 0.1586553$
$P(x <= 7) = 0.8413447$

Part B

```r
x_norm <- rnorm(10000, mean=5, sd=2)
sim_norm_prob_3<-mean(x_norm<=3)
sim_norm_prob_7<-mean(x_norm<=7)
```

Simulated and theoretical values are very close to each other.

Simulated Values:
$P(x <= 3) = 0.1652$
$P(x <= 7) = 0.8461$

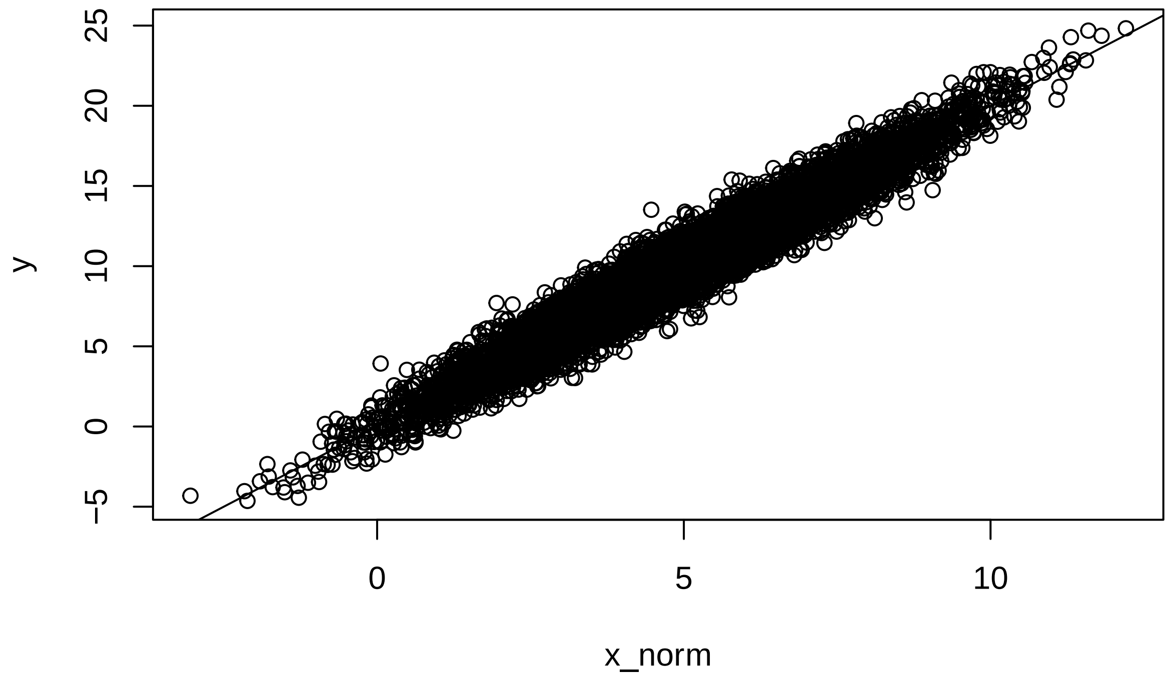
Theoretical Values:
$P(x <= 3) = 0.1586553$
$P(x <= 7) = 0.8413447$

Part C

Part i., ii., iii.

```r
z_norm <- rnorm(10000)
y <- 2*x_norm + z_norm

plot(x_norm,y)
abline(0,2)
```



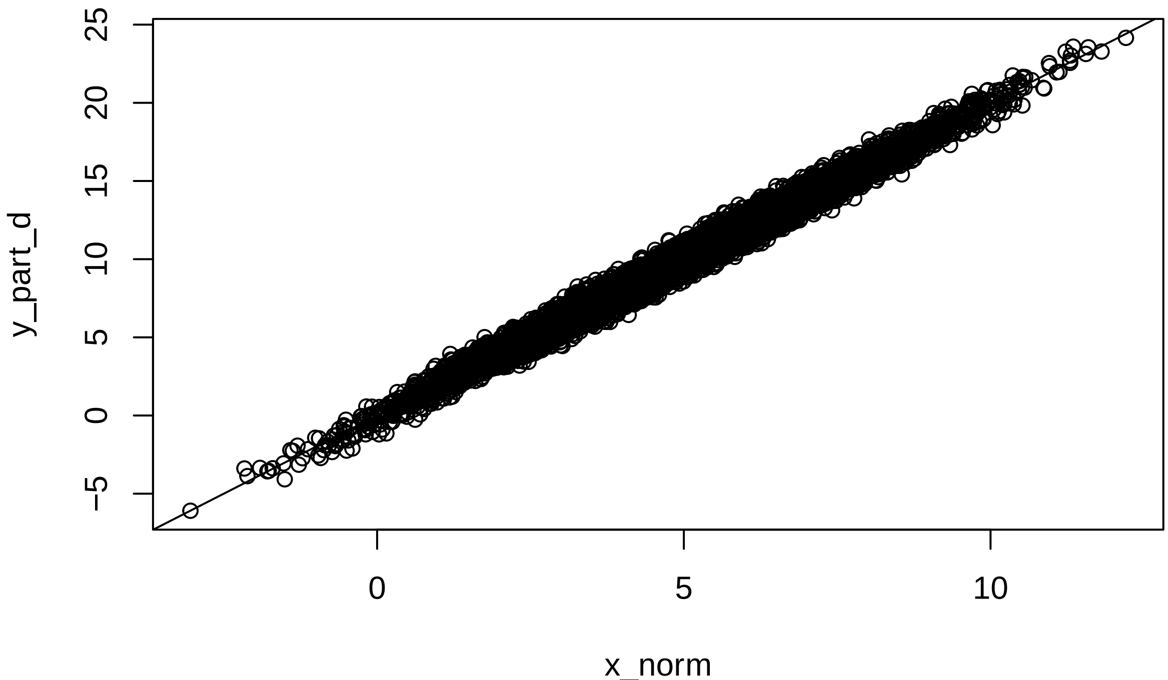
The (x,y) scatter plot forms a linear pattern which is concentrated around (5,10) and goes down in density as x moves away from 5.

Because x_norm has a standard deviation of 2, and z has a standard deviation of 1, I would expect y = 2x, and so I would expect abline(0,2) which is the line y(x)=2x to pass through the middle of scatter plot (x,y)....and this is exactly what happens.

Part D

```
z_norm_changed <- rnorm(10000, mean=0, sd=0.5)
y_part_d <- 2*x_norm + z_norm_changed

plot(x_norm,y_part_d)
abline(0,2)
```



Part E

The scatter plot with z having a standard deviation of 0.5 becomes more closer to the line y(x)=2x.

As the standard deviation of z gets larger than 1, the standard deviation of y will also get larger and the scatter plot will move away from a linear pattern matching the line y(x)=2x, and as the standard deviation of z gets smaller and closer to 0, the scatter plot will converge to the line y(x)=2x.