STA457 HW1

Depeng Ye 1002079500

30/09/2019

Problem 1

a)

Since all r_i are i.i.d. random variables that follow $N(\mu, \sigma^2)$.
By the additivity of Normality, $r_t(3) = r_t + r_{t-1} + r_{t-2}$ should follow Normal distribution as well with the mean of $E[r_t(3)] = E[r_t + r_{t-1} + r_{t-2}] = E[r_t] + E[r_{t-1}] + E[r_{t-2}] = \mu + \mu + \mu = 3\mu$, and variance of $Var[r_t(3)] = Var[r_t + r_{t-1} + r_{t-2}] = Var[r_t] + Var[r_{t-1}] + Var[r_{t-2}] = \sigma^2 + \sigma^2 + \sigma^2 = 3\sigma^2$.
Hence $r_t(3) \sim N(3\mu, 3\sigma^2)$.

b)

$$
\begin{aligned}
Cov(r_t(k),\ r_t(k+l)) &= Cov(r_t + r_{t-1} + \cdots + r_{t-k+1}, r_t(k+l)) \\
&= Cov(r_t,\ r_t(k+l)) + Cov(r_{t-1},\ r_t(k+l)) + \cdots + Cov(r_{t-k+1},\ r_t(k+l)) \\
&= Cov(r_t,\ r_t) + Cov(r_t,\ r_{t-1}) + \cdots + Cov(r_t,\ r_{t-k-l+1}) + Cov(r_{t-1},\ r_t) + \\
&\quad Cov(r_{t-1},\ r_{t-1}) + \cdots + Cov(r_{t-1},\ r_{t-k-l+1}) + \cdots \\
&= Var(r_t) + Var(r_{t-1}) + \cdots + Var(r_{t-k+1}) \\
&= k\sigma^2
\end{aligned}
$$

Problem 2

a)

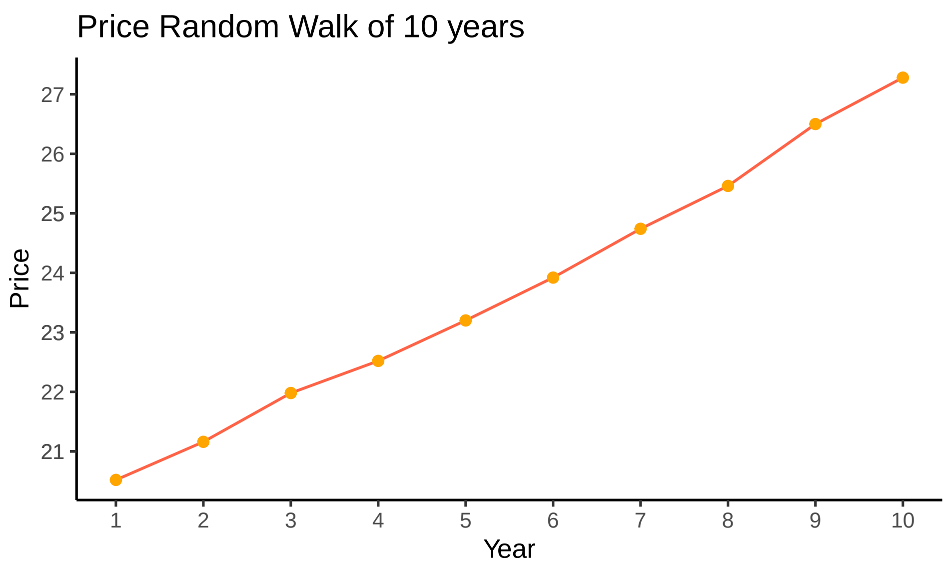
Simulating the annual price of the first 10 years using rnorm() function as random error and the results are shown in the table below.

```r
set.seed(2)
p0 <- 20
years <- 10
mu <- 0.03
sigma <- 0.005
error <- rnorm(10, mu, sigma)
current_price <- data.frame(exp(log(p0) + cumsum(error)))
year <- data.frame(c('P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8', 'P9', 'P10'))
full_list <- cbind(year,round(current_price, digits = 2))
colnames(full_list) <- c("Year", "Price")
knitr::kable(full_list, digits = 2, caption = "Price of the asset for nest 10 years",
             col.names = c('year', 'price'))
```

Table 1: Price of the asset for nest 10 years

year	price
P1	20.52
P2	21.16
P3	21.98
P4	22.52
P5	23.20
P6	23.92
P7	24.74
P8	25.46
P9	26.50
P10	27.28

The plot is shown as follows:



b)

Simulate P_{10} 2000 times using rnorm() function as random error. The results are not shown because they will take too much of space. The expected value is calculated using mean() fucniton.

```
P_10_vec = c()
for (i in 1 : 2000){
  e = rnorm(10, mu, sigma)
  P_10 = exp(log(p0) + sum(e))
  P_10_vec = append(P_10_vec, P_10)
  i = i + 1
}
print(mean(P_10_vec), digits = 2)
```

```
## [1] 27
```

The expected value of P_{10} is $E[P_{10}] = 27$.

c)

My simulated result in a) ($P_{10} = 27.28$) and slightly larger than the actual expected value of $E[P_{10}] = 27$ in part b).

The reason is that the effect of the random error is reduced by the large number of data set. When there is 2000 estimates of P_{10}, the mean of the random error used (rnorm) tend to become 0. Therefore, the expected value of P_{10} is more accurate in part b).

Problem 3

When compared with the examples shown in the lecture, exchanging horizontal and vertical axes will invert the convexity of the curve in Q-Q plot. Namely, the concave pattern in lecture's plot will be a convex pattern in this question and vice versa.

a)

The convex pattern in the plot required in this question interprets a right-skewed distribution of sample. For example, Chi-square distribution.

b)

The concave pattern in the plot required in this question interprets a left-skewed distribution of sample.

c)

The convex-concave pattern in the plot required in this question interprets a light tailed pattern in the sample distribution.

d)

The concave-convex pattern in the plot required in this question interprets a heavy tailed pattern in the sample distribution.

Problem 4

a)

Because $X_1, X_2, \ldots$ is a lognormal geometric random walk with parameter (μ, σ^2), we have $X_2 = X_0 \exp(r_1 + r_2)$. It is easy to interpret from the above equation that

$$
\begin{aligned}
& X_2 = X_0 \exp(r_1 + r_2) \\
\Rightarrow & \frac{X_2}{X_0} = \exp(r_1 + r_2) \\
\Rightarrow & \log(\frac{X_2}{X_0}) = r_1 + r_2
\end{aligned}
$$

which implies $\log \frac{X_2}{X_0}$ follows $N(2\mu, 2\sigma^2)$. Because $(r_1 + r_2) \sim N(2\mu, 2\sigma^2)$ form Problem 1.
Try to solve this question based on Standard Normal distribution.

$$\begin{aligned}
P(X_2 > 1.5X_0) &= P(\frac{X_2}{X_0} > 1.5) \\
&= P(\log \frac{X_2}{X_0} > \log 1.5) \\
&= P(\frac{\log \frac{X_2}{X_0} - 2\mu}{\sqrt{2}\sigma} > \frac{\log 1.5 - 2\mu}{\sqrt{2}\sigma}) \\
&= P(Z > \frac{\log 1.5 - 2\mu}{\sqrt{2}\sigma}) \\
&= 1 - P(Z < \frac{\log 1.5 - 2\mu}{\sqrt{2}\sigma})
\end{aligned}$$

Hence,

$$P(X_2 > 1.5X_0) = 1 - \Phi(Z < \frac{\log 1.5 - 2\mu}{\sqrt{2}\sigma})$$

where Φ is the cumulative density funciton of standardized normal distribution.

b)

Want to find the 0.8 quantile of X_k for all k, need to first know the distribution of X_k.
Because $X_1, X_2, X_3, \ldots$ are a lognormal geometric random walk with parameters (μ, σ^2) and $X_k = X_0 \exp(r_1 + r_2 + \cdots + r_k)$. With similar interpretation as part a), should have $\log \frac{X_k}{X_0} \sim N(k\mu, k\sigma^2)$
Assume that the 0.8 quantile of X_k is Q, by definition of quantile:

$$P(X_k < Q) = 0.8$$

Take log on both sides and then standardize the left side:

$$\begin{aligned}
&P(X_k < Q) = 0.8 \\
&\Rightarrow P(\log X_k < \log Q) = 0.8 \\
&\Rightarrow P(\frac{\log X_k - k\mu}{\sqrt{k}\sigma} < \frac{\log Q - k\mu}{\sqrt{k}\sigma}) = 0.8
\end{aligned}$$

Notice that only $\log \frac{X_k}{X_0}$ follows normal distribution, try to substract $\frac{\log X_0}{\sqrt{k}\sigma}$ from both sides of the inequality inside of $P()$:

$$\begin{aligned}
&\Rightarrow P(\frac{\log X_k - \log X_0 - k\mu}{\sqrt{k}\sigma} < \frac{\log Q - \log X_0 - k\mu}{\sqrt{k}\sigma}) = 0.8 \\
&\Rightarrow P(\frac{\log \frac{X_k}{X_0} - k\mu}{\sqrt{k}\sigma} < \frac{\log Q - \log X_0 - k\mu}{\sqrt{k}\sigma}) = 0.8
\end{aligned}$$

where $\frac{\log \frac{X_k}{X_0} - k\mu}{\sqrt{k}\sigma}$ is the standardized Z of $\log \frac{X_k}{X_0}$, in this way we can now check the Z-table.
By checking the Z-table,

$$
\begin{aligned}
\frac{\log Q - \log X_0 - k\mu}{\sqrt{k}\sigma} &= 0.84162 \\
\log Q - \log X_0 - k\mu &= 0.84162\sqrt{k}\sigma \\
Q &= \exp(\log X_0 + k\mu + 0.84162\sqrt{k}\sigma) \\
Q &= X_0 + \exp(k\mu + 0.84162\sqrt{k}\sigma)
\end{aligned}
$$

Hence, the 0.8 quantile of X_k is $Q = X_0 + \exp(k\mu + 0.84162\sqrt{k}\sigma)$.

c)

Want to find $E[X_k^2]$ as a function of k, need the second moment density function of $N(k\mu, k\sigma^2)$ given that we know $\log \frac{X_k}{X_0} \sim N(k\mu, k\sigma^2)$ i.e. $\log X_k \sim N(\log X_0 + k\mu, k\sigma^2)$.
From the distribution we have the density function:

$$
f(x_k) = \frac{1}{\sqrt{2\pi k}\sigma x_k} \exp\left(- \frac{(\log x_k - \log x_0 - k\mu)^2}{2k\sigma^2} \right)
$$

Considering the second moment, the expectation of X_k^2 is :

$$
E[X_k^2] = \int_{-\infty}^{\infty} X_k^2 \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{(\log x_k - \log x_0 - k\mu)^2}{2k\sigma^2} \right) d(\sum_{i=1}^{k} r_i)
$$

Now consider $X_k = X_0 \exp(r_k + r_{k-1} + \cdots + r_2 + r_1)$. Using the notation in Problem 1 we have: $r_k(k) = r_k + r_{k-1} + \cdots + r_2 + r_1$.
Then X_k can be written as $X_0 \exp(r_k(k))$, and $\log X_k = \log X_0 + r_k(k)$. \

$$
\begin{aligned}
E[X_k^2] &= \int_{-\infty}^{\infty} X_k^2 \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{(\log x_k - \log x_0 - k\mu)^2}{2k\sigma^2} \right) d(\sum_{i=1}^{k} r_i) \\
&= \int_{-\infty}^{\infty} X_0^2 \exp(r_k(k))^2 \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{[\log x_0 + r_k(k) - (\log x_0 + k\mu)]^2}{2k\sigma^2} \right) dr_k(k) \\
&= X_0^2 \exp(2k\mu + \frac{4k\sigma^2}{2}) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{[r_k(k) - (k\mu + 2k\sigma^2)]^2}{2k\sigma^2} \right) dr_k(k)
\end{aligned}
$$

Notice that inside the integral is the pdf of normal distribution $N(k\mu + 2k\sigma^2, k\sigma^2)$, therefore the integral should be equal to 1.
Thus the expectation of X_k^2 is :

$$
E[X_k^2] = X_0^2 e^{2k\mu + \frac{4k\sigma^2}{2}} = X_0^2 e^{2k\mu + 2k\sigma^2}
$$

d)

To calculate the variance of X_k, use the formula $Var(X_k) = E[X_k^2] - E[X_k]^2$. We have already calcuated $E[X_k^2]$ in part c), and now only need to find $E[X_k]$.

Similarly, use the pdf method to find expectation fo X_k:

$$
\begin{aligned}
E[X_k] &= \int_{-\infty}^{\infty} X_k \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{(\log x_k - \log x_0 - k\mu)^2}{2k\sigma^2} \right) d(\sum_{i=1}^{k} r_i) \\
&= \int_{-\infty}^{\infty} X_0 \exp(r_k(k)) \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{[\log x_0 + r_k(k) - (\log x_0 + k\mu)]^2}{2k\sigma^2} \right) dr_k(k) \\
&= X_0 \exp\left(k\mu + \frac{k\sigma^2}{2} \right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi k}\sigma} \exp\left(- \frac{[r_k(k) - (k\mu + k\sigma^2)]^2}{2k\sigma^2} \right) dr_k(k) \\
&= X_0 e^{k\mu + \frac{k\sigma^2}{2}}
\end{aligned}
$$

Hence together with te result from part c):

$$
\begin{aligned}
Var(X_k) &= E[X_k^2] - E[X_k]^2 \\
&= X_0^2 e^{2k\mu + 2k\sigma^2} - (X_0 e^{k\mu + \frac{k\sigma^2}{2}})^2 \\
&= X_0^2 \big(e^{2k\mu + 2k\sigma^2} - e^{2k\mu + k\sigma^2} \big) \\
&= X_0^2 e^{2k\mu + k\sigma^2} (e^{k\sigma^2} - 1)
\end{aligned}
$$

Problem 5

Want to find the MLE of σ^2 when μ is known with $Y_1, Y_2, \ldots, Y_n$ i.i.d. $N(\mu, \sigma^2)$.
The pdf of Y_i is $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\big(-\frac{(y-\mu)^2}{2\sigma^2}\big)$.
First find the likelihood:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(y_i)$$

The find the log likelihood:

$$\begin{aligned} l(\mu, \sigma^2) &= \log(L(\mu, \sigma^2)) \\ &= \sum_{i=1}^{n} \log f(y_i) \\ &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 \end{aligned}$$

MLE of σ^2 given that μ is known:
Let:

$$0 = \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\Rightarrow \widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \mu)^2$$

Hence, the MLE of σ^2 when μ is known is: $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \mu)^2$.

■

Problem 6

a)

Want to find the mean adn variance of $Z = YX_1 + (1-Y)X_2$.
First do the expectation part:

$$\begin{aligned} E[Z] &= E[YX_1 + (1-Y)X_2] \\ &= E[YX_1] + E[(1-Y)X_2] \\ &= E[Y]E[X_1] + E[1-Y]E[X_2] \\ &= E[Y]E[X_1] + (E[1] - E[Y])E[X_2] \end{aligned}$$

Since X_1 and X_2 follows $N(0, \sigma^2)$ and Y follows $Bernoulli(p)$. $E[X_1] = E[X_2] = 0$, $E[Y] = p$.
The above equation becomes:

$$E[Z] = p \times 0 + (1-p) \times 0 = 0$$

Then consider $Var(Z)$:

$$\begin{aligned} Var(Z) &= Var(YX_1 + (1-Y)X_2) \\ &= Var(YX_1) + Var((1-Y)X_2) + 2Cov(YX_1, (1-Y)X_2) \end{aligned}$$

because X_1, X_2 and Y are independent, with the formula $Var(XY) = E[X]^2Var(Y) + E[Y]^2Var(X) + Var(X)Var(Y)$:

$$\begin{aligned} Var(Z) = {} & E[X_1]^2Var(Y) + E[Y]^2Var(X_1) + Var(Y)Var(X_1) + E[X_2]^2Var(Y) + E[1-Y]^2Var(X_2) + \\ & Var(X_2)Var(Y) + 2Cov(YX_1, (1-Y)X_2) \\ = {} & 0 + p^2\sigma_1^2 + p(1-p)\sigma_1^2 + 0 + (1-p)^2\sigma_2^2 + \sigma_2^2 p(1-p) + 2Cov(YX_1, (1-Y)X_2) \end{aligned}$$

Consider the Covariance using the formula $Cov(X,Y) = E[XY] - E[X]E[Y]$.

$$\begin{aligned} Cov(YX_1, (1-Y)X_2) &= E[YX_1(1-Y)X_2] - E[YX_1]E[(1-Y)X_2] \\ &= E[Y]E[X_1]E[X_2] - E[Y^2]E[X_1]E[X_2] - E[Y]E[X_1]E[X_2](1-E[Y]) \\ &= 0 \end{aligned}$$

Plug back in to Variance of Z:

$$\begin{aligned} Var(Z) &= p^2\sigma_1^2 + p(1-p)\sigma_1^2 + (1-p)^2\sigma_2^2 + \sigma_2^2 p(1-p) \\ &= \sigma_1^2 p + \sigma_2^2(1-p) \end{aligned}$$

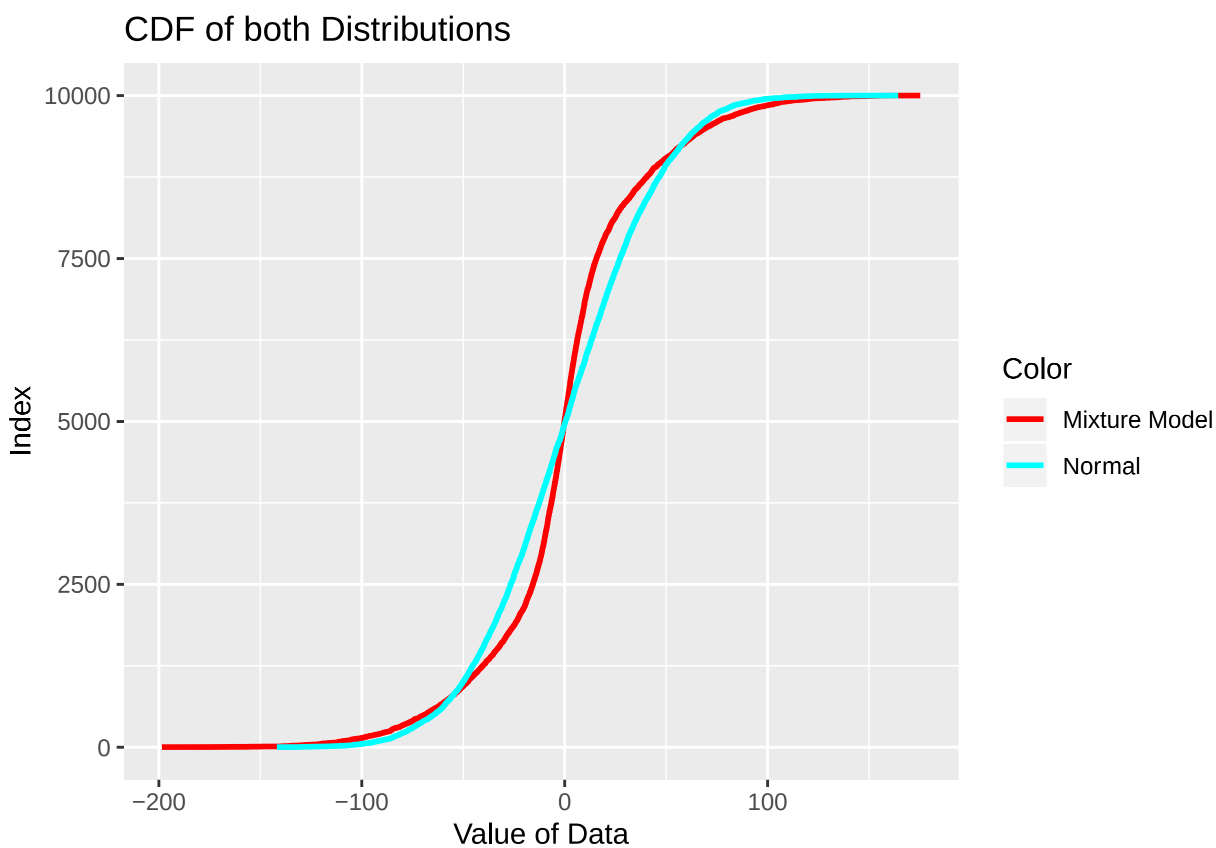
Thus, the expectation and variance of Z are: $E[Z] = 0$ and $Var(Z) = \sigma_1^2 p + \sigma_2^2(1-p)$.

b)

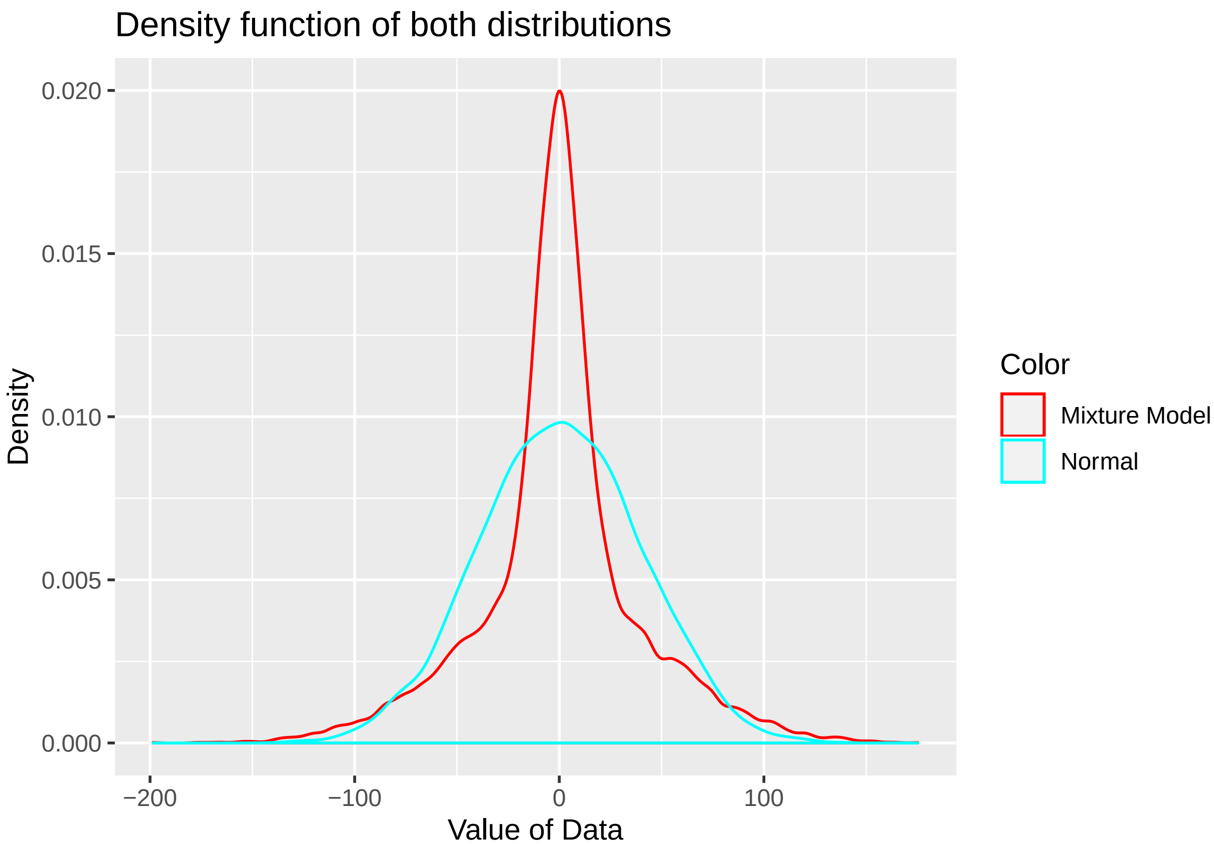
Use rnorm() and rbern() function to generate some data that follows the pattern of Z and a normal distribution with same mean and variance with Z.
Then plot the probability density function and the cumulative distribution of both distributions:

```r
set.seed(7)
X_1 = rnorm(10000, 0, 10)
X_2 = rnorm(10000, 0, 50)
Y = rbern(10000, 0.4)
Z_norm = rnorm(10000, 0, sqrt(10^2*0.4 + 50^2 * 0.6))
Z = Y*X_1 + X_2 - Y*X_2
Z = sort(Z)
Z = data.frame(cbind(Index = c(1:10000), Z = Z))
Z_norm = sort(Z_norm)
Z_norm = data.frame(Z_norm = Z_norm)
Z_Data = cbind(Z, Z_norm)

ggplot() +
  geom_line(data = Z_Data, aes(x = Z, y = Index, col = 'Mixture Model'), size = 1) +
  geom_line(data = Z_Data, aes(x = Z_norm, y = Index, col = 'Normal'), size = 1) +
  labs(y = "Index", x = "Value of Data", title = "CDF of both Distributions") +
  scale_color_manual(name = "Color", values = c("Mixture Model" = "red", "Normal" = "cyan"))
```

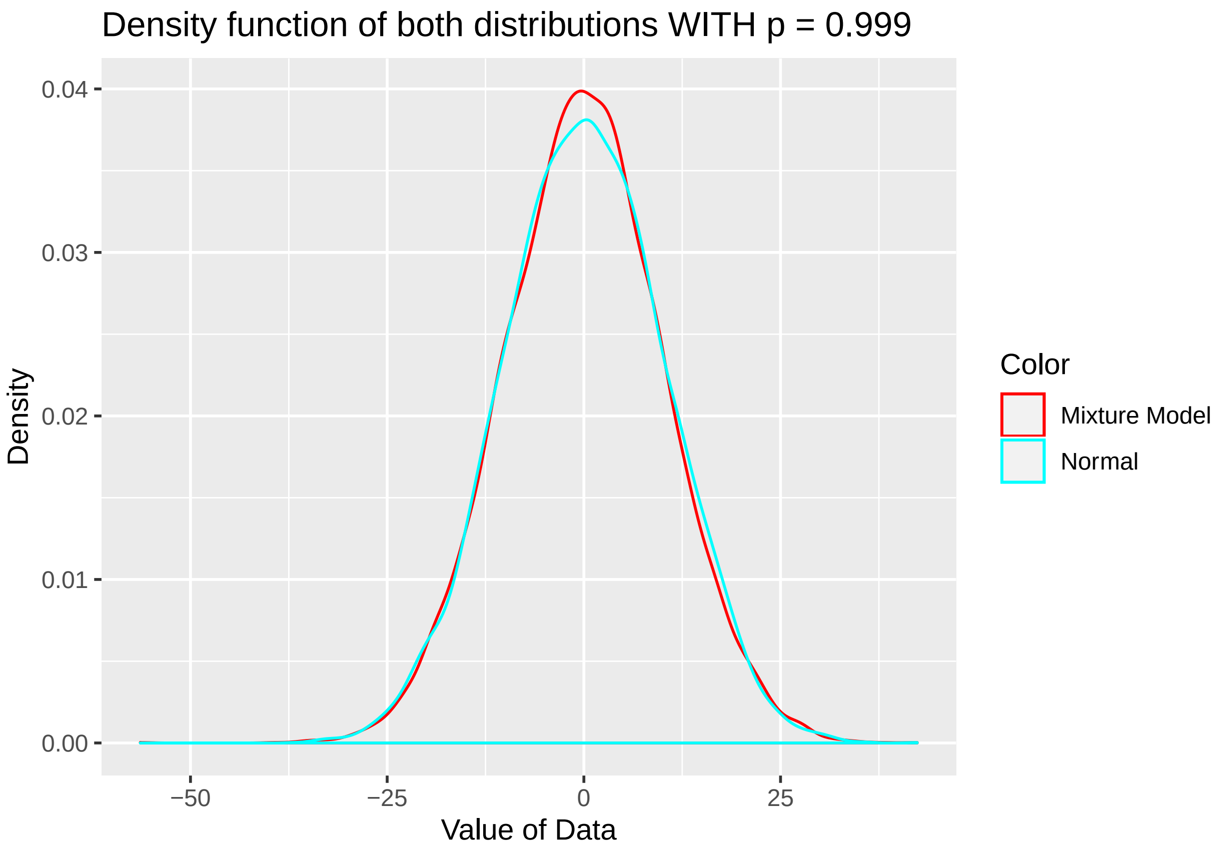


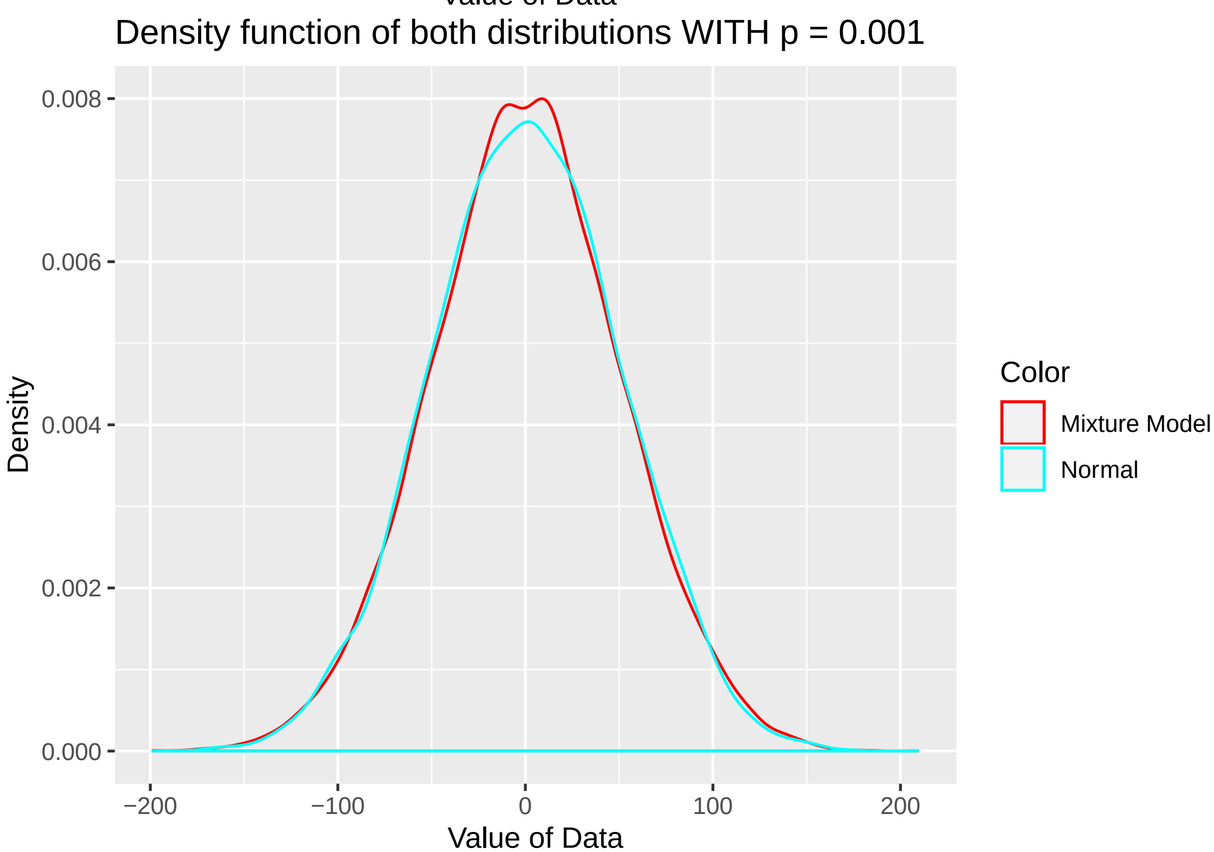

```r
ggplot() +
  geom_density(aes(x = Z_Data$Z, col = 'Mixture Model')) +
  geom_density(aes(x = Z_Data$Z_norm, col = "Normal")) +
  scale_color_manual(name = "Color", values = c("Mixture Model" = "red", "Normal" = "cyan")) +
  labs(y = "Density", x = "Value of Data", title = "Density function of both distributions")
```



It is easy to notice from the above graphs that the Mixture Model has heavier tail than the normal distribution with same mean and variance.

I have tries to change the value of p from 0.001 to 0.999. I have noticed that whenever the value p is closer to the end point of interval$(0, 1)$, the two plots are getting closer to each other. i.e. that misture model can be regarded as a normal distribution, but to be honest it is still heavy tailed. Example of $p = 0.999$ and $p = 0.001$ graphs are as follows:


Density function of both distributions WITH p = 0.999
0.04
0.03
0.02
0.01
0.00
Density
−50
−25
0
25
Value of Data
Color
Mixture Model
Normal


Density function of both distributions WITH p = 0.001
0.008
0.006
0.004
0.002
0.000
Density
−200
−100
0
100
200
Value of Data
Color
Mixture Model
Normal