STA457 Problem Set 1

Brian Lee 1002750855

1 Problem 1

Proof. a) First, recall that if $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent, then

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

To see a brief proof of this, we can just observe that

$$\begin{aligned}
\mathbb{E}\left(e^{t(X+Y)}\right) &= \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) && \text{since } X \text{ and } Y \text{ are independent} \\
&= \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right) \exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right) \\
&= \exp\left((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right)
\end{aligned}$$

Since the MGF determines the distribution, we have that $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Therefore, since $r_i \sim N(\mu, \sigma^2)$, it follows that

$$r_t(3) = r_t + r_{t-1} + r_{t-2} \sim N(3\mu, 3\sigma^2)$$

b) To make sense of the above, we have that t, k, l are positive integers such that $t > k + l$, and consider

$$\begin{aligned}
\operatorname{Cov}\left(r_t(k), r_t(k+l)\right) &= \operatorname{Cov}\left(\sum_{i=0}^{k} r_{t-i}, \sum_{j=0}^{k+l} r_{t-j}\right) \\
&= \sum_{i=0}^{k}\sum_{j=0}^{k+1} \operatorname{Cov}\left(r_{t-i}, r_{t-j}\right) && \text{multilinearity of covariance} \\
&= \sum_{i=0}^{k} \operatorname{Cov}\left(r_{t-i}, r_{t-i}\right) && \text{only diagonals remain by independence} \\
&= \sum_{i=0}^{k} \operatorname{Var}(r_{t-i}) \\
&= (k+1)\sigma^2
\end{aligned}$$

□

2 Problem 2

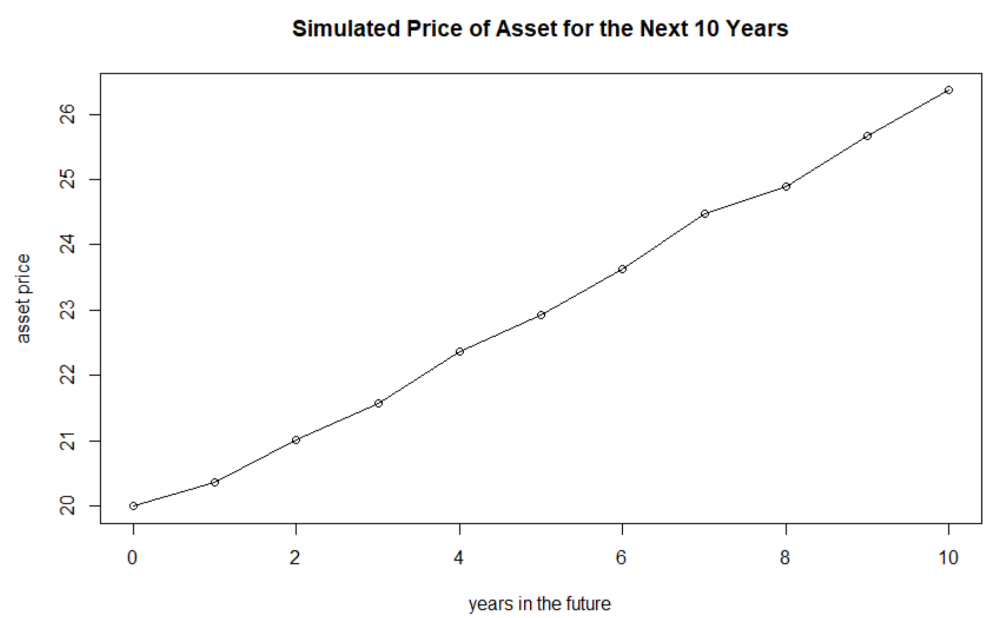
(a) We simulated the annual price of the asset for the next 10 years using R; the R code used to simulate and plot the result is as below:

```r
x <- 1:10
y <- 0:10
y[1]=20
for (val in 1:10) {
x[val] <- rnorm(1, mean=0.03, sd=0.005)
}
tot <- sum(x)
for (ct in 1:10) {
        list_ctr = 1:ct
        value = 20*exp(sum(x[list_ctr]))
        y[ct+1] = value
}
print(y)
x_list = 0:10
plot(x_list, y, type = 'o', main="Simulated Price of Asset
        for the Next 10 Years", xlab="years in the future", ylab="asset price")
```

The prices the code outputs (for one particular simulation) are summarized in the table below

t	P_t
0	20
1	20.3772
2	21.0192
3	21.5708
4	22.3739
5	22.9269
6	23.6355
7	24.4855
8	24.887
9	25.6585
10	26.3635

And the result, plotted against time is:



Despite time being discrete here, I chose to insert lines connecting the data points to illustrate the different rates of growth between time periods.

(b) The code used to simulate P_{10} 200 times and estimate the mean (via the law of large numbers, since we're taking a large enough sample size such that the sample mean approximates the actual mean) is as below:

```
values = 1:2000
for (ct in 1:2000) {
samp = rnorm(10, mean=0.03, sd=0.005)
tot = sum(samp)
values[ct] = 20*exp(tot)
}
print(mean(values))
```

The output which approximates $\mathbb{E}(P_{10})$ was 27.02212.

(c) Note that $P_{10} = 20\exp(r_1+\ldots+r_t)$, where $\exp(r_1+\ldots+r_t)$ is Lognormal$(0.3, 0.025)$ distributed since r_i are i.i.d. $N(0.03, 0.05^2)$ hence $\mu = 0.03 \cdot 10 = 0.3$ and $\sigma^2 = 0.05^2 \cdot 10 = 0.025$. Since a Lognormal$(\mu, \sigma^2)$ distribution has expectation:

$$\exp\left(\mu + \frac{\sigma^2}{2}\right)$$

we get that

$$\mathbb{E}(P_{10}) = 20\mathbb{E}(\exp(r_1 + \ldots + r_t)) = 20\exp\left(0.3 + \frac{0.025}{2}\right) = 20\exp(0.30125) = 27.03094...$$

We see that the actual expected value is quite similar to the approximation we got in part (b).

3 Problem 3

Consider a convex pattern on a normal plot where sample quantiles are on the vertical axis, and suppose you can represent the pattern approximately as the graph of some convex function f. If we assume that our pattern is well behaved enough such that f is also a sufficiently regular function (continuous, bijective), which along with convexity would imply that f has an inverse which is concave. If so, the graph $(y, f(y))$ is equivalent to the graph $(x, f^{-1}(x))$ with the axis swapped. Obviously, the reverse is true with concave functions, and their inverse being convex under the right conditions. Keeping this in mind:

(a) We can see that a convex pattern in our modified QQ plot corresponds to a concave pattern in a normal QQ plot, hence indicates right skewness.

(b) Similarly with (a), the concave pattern in the modified QQ plot corresponds to a convex pattern in a normal QQ plot, hence indicates left skewness.

(c) We can proceed similarly as the other cases, except we just consider a convex-concave pattern as graphs of two separate functions that agree continuously, f_1 which is convex and f_2 which is concave (i.e. if x_0 is the point at which the second derivative changes sign, we define the function to the left of x_0 as f_1 which is convex, and the function to the right of x_0 as f_2, which is concave). Using what we have above, we get that a convex-concave pattern in our modified QQ plot corresponds to a concave-convex pattern in a normal QQ plot, hence indicate light tails

(d) Similarly as (c), a concave-convex pattern corresponds to a convex-concave pattern in a normal QQ plot, hence indicates heavy tails.

4 Problem 4

(a) First, we note that

$$\mathbb{P}(X_2 > 1.5X_0) = \mathbb{P}(\exp(r_1 + r_2) > 1.5)$$

Since r_1 and r_2 are i.i.d. $N(\mu, \sigma^2)$, $r_1 + r_2 \sim N(2\mu, 2\sigma^2)$ thus $\exp(r_1 + r_2)$ is Lognormal$(2\mu, 2\sigma^2)$ distributed, hence

$$\mathbb{P}(\exp(r_1 + r_2) > 1.5) = \int_{\frac{3}{2}}^{\infty} \frac{1}{2x\sigma\sqrt{\pi}} \exp\left(-\frac{(\log x - 2\mu)^2}{4\sigma^2}\right) = \frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{\log\frac{3}{2} - 2\mu}{2\sigma}\right)$$

Hence

$$\mathbb{P}(X_2 > 1.5X_0) = \frac{1}{2}\left(1 - \operatorname{erf}\left(\frac{\log\frac{3}{2} - 2\mu}{2\sigma}\right)\right)$$

(b) Let q_k be the 0.8 quantile of X_k, i.e. q_k that solves

$$\mathbb{P}(X_k \le q_k) = 0.8$$

Now, let $S_k = r_1 + \ldots + r_k$, and we know that $S_k \sim N(k\mu, k\sigma^2)$ since r_i are i.i.d, thus

$$\begin{aligned}
\mathbb{P}(X_k \le q_k) &= \mathbb{P}\left(\exp(r_1 + \ldots + r_k) \le \frac{q_k}{X_0}\right) \\
&= \mathbb{P}\left(S_k \le \log\frac{q_k}{X_0}\right) \\
&= \mathbb{P}\left(\frac{S_k - k\mu}{k\sigma^2} \le \frac{a_k - k\mu}{k\sigma^2}\right) && \text{where } a_k = \log\frac{q_k}{X_0} \\
&= \mathbb{P}\left(Z \le \frac{a_k - k\mu}{k\sigma^2}\right) \\
&= 0.8
\end{aligned}$$

Where $Z \sim N(0, 1)$ is the standard normal distribution; by using R or using a table of values, we can see that

$$\frac{a_k - k\mu}{k\sigma^2} \simeq 0.84 \;\Rightarrow a_k \simeq (0.84)k\sigma^2 + k\mu$$

This gives

$$e^{a_k} = \exp\left((0.84)k\sigma^2 + k\mu\right) = \frac{q_k}{X_0}$$

This gives us that

$$q_k = X_0 \exp\left((0.84)k\sigma^2 + k\mu\right)$$

(c) and (d): Note that $X_k = X_0 \exp(r_1 + \ldots + r_k)$ where $r_1 + \ldots + r_k = S_k \sim N(k\mu, k\sigma^2)$, and $\exp(S_k)$ is Lognormal$(k\mu, k\sigma^2)$ distributed. This gives us that:

$$\operatorname{Var}(X_k) = X_0^2 \operatorname{Var}(\exp(S_k)) = X_0^2(e^{k\sigma^2} - 1)\exp\left(2k\mu + k\sigma^2\right)$$

and

$$\mathbb{E}(X_k) = X_0 \cdot \mathbb{E}(\exp(S_k)) = X_0 \exp\left(k\mu + \frac{k\sigma^2}{2}\right)$$

hence

$$\mathbb{E}(X_k^2) = \operatorname{Var}(X_k) + \mathbb{E}(X_k)^2 = X_0^2\left[(e^{k\sigma^2} - 1)\exp\left(2k\mu + k\sigma^2\right) + \exp\left(2k\mu + k\sigma^2\right)\right] = X_0^2 \exp(2k\mu + 2k\sigma^2)$$

5 Problem 5

We seek to optimize (maximize) the log-likelihood function of $Y_1, \ldots, Y_n$ with respect to σ^2. We have that the density function of Y_i is:

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

Hence the likelihood function, since the Y_i are independent, is the product of the densities:

$$L(\mu, \sigma^2 : y_1, \ldots, y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2\right)$$

Thus the log-likelihood is

$$l(\mu, \sigma^2; y_1, \ldots, y_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

Differentiating with respect to σ^2, we get:

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2} l(\mu, \sigma^2; y_1, \ldots, y_n) &= \frac{\partial}{\partial\sigma^2}\left(-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right) \\
&= 0 - \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2 \\
&= -\frac{1}{2\sigma^2}\left(n - \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right) \\
&= 0
\end{aligned}$$

Assuming $\sigma^2 > 0$, it must be that $n = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$, hence a critical point is at

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

Hence

$$\widehat{\sigma^2}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^2$$

6 Problem 6

(a) Using that X_1, X_2 and Y are independent,

$$\begin{aligned}
\mathbb{E}(Z) &= \mathbb{E}(YX_1 + (1-Y)X_2) \\
&= \mathbb{E}(Y)\mathbb{E}(X_1) + \mathbb{E}(X_2) - \mathbb{E}(Y)\mathbb{E}(X_2) \\
&= p \cdot 0 + 0 + p \cdot 0 \\
&= 0
\end{aligned}$$

Hence $\mathbb{E}(Z) = 0$. Also,

$$\begin{aligned}
\text{Var}(Z) &= \text{Var}(YX_1) + \text{Var}(X_2) + \text{Var}(YX_2) \\
&= (\sigma_Y^2 + \mu_Y^2)(\sigma_{X_1}^2 + \mu_{X_1}^2) - \mu_Y^2\mu_{X_1}^2 + \sigma_2^2 + (\sigma_Y^2 + \mu_Y^2)(\sigma_{X_2}^2 + \mu_{X_2}^2) - \mu_Y^2\mu_{X_2}^2 \\
&= (p(1-p) + p^2)\sigma_1^2 + \sigma_2^2 + (p(1-p) + p^2)\sigma_2^2 \\
&= p(\sigma_1^2 + \sigma_2^2) + \sigma_2^2
\end{aligned}$$

Hence $\text{Var}(Z) = p(\sigma_1^2 + \sigma_2^2) + \sigma_2^2$.

(b) We compute the kurtosis of $\mathbb{Z}$: since $\mathbb{E}(Z) = 0$, we compute

$$\begin{aligned}
\mathbb{E}(Z^4) &= \mathbb{E}\left[(YX_1 + (1-Y)X_2)^4\right] \\
&= \mathbb{E}\left[Y^4X_1^4 + 4Y^3(1-Y)X_1^3X_2 + 6Y^2(1-Y)^2X_1^2X_2^2 + 4Y(1-Y)^3X_1X_2^3 + (1-Y)^4X_2^4\right] \\
&= \mathbb{E}(Y^4)\mathbb{E}(X_1^4) + 4\mathbb{E}(Y^3)\mathbb{E}(1-Y)\mathbb{E}(X_1^3)\mathbb{E}(X_2) + 6\mathbb{E}(Y^2)\mathbb{E}((1-Y)^2)\mathbb{E}(X_1^2)\mathbb{E}(X_2^2) \\
&\quad + 4\mathbb{E}(Y)\mathbb{E}((1-Y)^3)\mathbb{E}(X_1)\mathbb{E}(X_2^3) + \mathbb{E}((1-Y)^4)\mathbb{E}(X_2^4) \\
&= 3p\sigma_1^4 + 0 + 6p(1-p)\sigma_1^2\sigma_2^2 + 0 + 3(1-p)\sigma_2^4 \\
&= 3(p\sigma_1^4 + 2p(1-p)\sigma_1^2\sigma_2^2 + (1-p)\sigma_2^4)
\end{aligned}$$

Combining this with our computation of the variance above, we get that

$$\mathrm{Kurt}(Z) = \frac{E(Z^4)}{\mathrm{Var}(Z)^2} = 3\frac{p\sigma_1^4 + 2p(1-p)\sigma_1^2\sigma_2^2 + (1-p)\sigma_2^4}{(p(\sigma_1^2+\sigma_2^2)+\sigma_2^2)^2}$$

Since the kurtosis of a normal random variable with the same random variable is equal to 3, if the kurtosis is greater than 3, it is heavy tailed and has a heavier tail than a normal distribution of the same variance; we can determine when this happens

$$\frac{p\sigma_1^4 + 2p(1-p)\sigma_1^2\sigma_2^2 + (1-p)\sigma_2^4}{(p(\sigma_1^2+\sigma_2^2)+\sigma_2^2)^2} > 1$$

this is equivalent to the condition

$$(p-1)\sigma_1^4 + 4p\sigma_1^2\sigma_2^2 + (p+3)\sigma_2^4 < 0$$

Thus, for values of p such that the above condition is true, Z is heavier tailed than the normal distribution with the same variance. On the other hand, if

$$(p-1)\sigma_1^4 + 4p\sigma_1^2\sigma_2^2 + (p+3)\sigma_2^4 > 0$$

this implies that

$$\frac{p\sigma_1^4 + 2p(1-p)\sigma_1^2\sigma_2^2 + (1-p)\sigma_2^4}{(p(\sigma_1^2+\sigma_2^2)+\sigma_2^2)^2} < 1$$

hence which would mean that $\mathrm{Kurt}(Z) < 3$, thus if p satisfies the above condition then Z is lighter tailed than a normal distribution with the same variance.
To try and interpret these results, let's consider some specific values of p; for p very close to 1, the above 'condition' becomes approximately

$$4\sigma_1^2\sigma_2^2 + 4\sigma_2^2 > 0$$

indicating that Z is lighter tailed than the normal distribution of the same variance. If p is very close to 1, then X_1 will dominate over X_2 in the mixture distribution hence the distribution will behave almost like $X_1 \sim N(0, \sigma_1)$, which is lighter tailed than $N(0, \sigma_1 + 2\sigma_2)$ distribution which is a normal distribution with approximately the same variance as Z.
In general, looking at the expression $(p-1)\sigma_1^4 + 4p\sigma_1^2\sigma_2^2 + (p+3)\sigma_2^4$, we see that it will be positive unless x_1 dominates greatly over x_2, with a small enough value of p. Noting that this expression being negative implies heavy-tailedness, which corresponds to the mixture being more concentrated towards the one with higher variance (lower p value and X_1 dominating X_2).