Statistics: Homework 1

Due on Aug 5, 2014

Instructor: Rados Radoicic 6:00 pm

Weiyi Chen

Problem 1

The standard Student's t distribution

The standard Student's t distribution is a special case of Student's t distribution. By first explaining this special case, the exposition of the more general case is greatly facilitated. The standard Student's t distribution is characterized as follows:
Definition: We say that X has a standard Student's t distribution with n degrees of freedom if its probability density function is:

$$f_X(x) = c(1+\frac{x^2}{n})^{-(n+1)/2} \tag{1}$$

where c is a constant:

$$c = \frac{1}{\sqrt{n}}\frac{1}{B(\frac{n}{2},\frac{1}{2})} \tag{2}$$

where $B()$ is a beta function.
Proposition: The probability density function of X can be written as:e

$$f_X(x) = \int_0^\infty f_{X|Z=z}(x)f_Z(z)dz \tag{3}$$

where $f_{X|Z=z}(x)$ is the probability density function of a normal distribution with mean 0 and variance $\sigma^2 = \frac{1}{2}$, $f_Z(z)$ is the probability density function of a Gamma random variable with parameters n and $h=1$. This is obvious. If X is a zero-mean normal random variable with variance $1/z$, conditional on $Z=z$, then we can think of X as a ratio:

$$X = \frac{Y}{\sqrt{Z}} \tag{4}$$

where Y has a standard normal distribution, Z has a Gamma distribution and Y and Z are independent.
Expected value: The expected value of a standard Student's t random variable X is well-defined only for $n>1$ and it is equal to:

$$E[X] = 0 \tag{5}$$

It follows from the fact that the density function is symmetric around 0:

$$\begin{aligned} E[X] &= \int_{-\infty}^0 xf_X(x)dx + \int_0^\infty xf_X(x)dx \\ &= -\int_0^\infty xf_X(x)dx + \int_0^\infty xf_X(x)dx \\ &= 0 \end{aligned} \tag{6}$$

by exchanging the bounds of integraton and $f_X(x) = f_X(x)$. The above integrals are finite (and so the expected value is well-defined) only if $n>1$, because

$$\begin{aligned} \int_0^\infty &= \lim_{u\to\infty}\int_0^u xc(1+\frac{x^2}{n})^{-(n+1)/2}dx \\ &= c\int_0^u \left[-\frac{n}{n-1}(1+\frac{x^2}{n})^{-(n-1)/2}\right]_0^u \\ &= \frac{-cn}{n-1}\left[\lim_{u\to\infty}(1+\frac{u^2}{n})^{-\frac{1}{2}(n-1)}\right] \end{aligned} \tag{7}$$

and the above limit is finite only if $n>1$.
Variance: The variance of a standard Student's t random variable X is well-defined only for $n>2$ and it is equal to:

$$Var[X] = \frac{n}{n-2} \tag{8}$$

It can be derived thanks to the usual variance formula ($Var(X) = E[X^2] - (E[X])^2$) and to the integral representation of the Beta function:

$$\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= 2\int_0^{\infty} x^2 f_X(x)dx \\
&= 2c\int_0^{\infty} x^2(1+\frac{x^2}{n})^{-(n+1)/2}dx \\
&= 2c\int_0^{\infty} n(1+t)^{-n/2-1/2}\frac{\sqrt{n}}{2}\frac{1}{\sqrt{t}}dt \\
&= cn^{3/2}\int_0^{\infty} t^{3/2-1}(1+t)^{-3/2-(n/2-1)}dt \\
&= cn^{3/2}B(\frac{3}{2},\frac{n}{2}-1) \\
&= \frac{1}{\sqrt{n}}\frac{1}{B(\frac{n}{2},\frac{1}{2})}n^{3/2}B(\frac{3}{2},\frac{n}{2}-1) \\
&= n\frac{\Gamma(n/2+1/2)}{\Gamma(n/2)\Gamma(1/2)}\frac{\Gamma(1/2+1)\Gamma(n/2-1)}{n/2+1/2} \\
&= n\frac{\Gamma(1/2)\frac{1}{2}\Gamma(n/2)\frac{2}{n-2}}{\Gamma(1/2)\Gamma(n/2)} \\
&= \frac{n}{n-2}
\end{aligned} \tag{9}$$

Therefore using $(E[X])^2 = 0$,

$$Var[X] = E[X^2] - (E[X])^2 = \frac{n}{n-2} \tag{10}$$

From the above derivation, it should be clear that the variance is well-defined only when $n > 2$. Otherwise, if $n \leq 2$, the above improper integrals do not converge (and the Beta function is not well-defined).

Student's t distribution in general

Definition: We say that X has a Student's t distribution with mean μ, scale σ^2 and n degrees of freedom if its probability density function is:

$$f_X(x) = \frac{c}{\sigma}(1+\frac{(x-\mu)^2}{n\sigma^2})^{-\frac{n+1}{2}} \tag{11}$$

where c is a constant:

$$c = \frac{1}{\sqrt{n}B(n/2,1/2)} \tag{12}$$

and $B()$ is the Beta function. A random variable X which has a t distribution with mean μ, scale σ^2 and n degrees of freedom is just a linear function of a standard Student's t random variable:

$$X = \mu + \sigma Z \tag{13}$$

where Z is a random variable having a standard t distribution.
Expectation: The expected value of a Student's t random variable X is well-defined only for $n > 1$ and it is equal to:

$$E[X] = \mu \tag{14}$$

It is an immediate consequence of the fact that $X = \mu + \sigma Z$ and the linearity of the expected value:

$$E[X] = E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu + \sigma 0 = \mu \tag{15}$$

As we have seen above, $E[Z]$ is well-defined only for $n > 1$ and, as a consequence, also $E[X]$ is well-defined only for $n > 1$.

Variance: The variance of a Student's t random variable X is well-defined only for $n > 2$ and it is equal to:

$$Var[X] = \frac{n}{n-2}\sigma^2 \tag{16}$$

It can be derived using the formula for the variance of affine transformations on $X = \mu + \sigma Z$:

$$Var[X] = Var[\mu + \sigma Z] = \sigma^2 Var[Z] = \sigma^2 \frac{n}{n-2} \tag{17}$$

As we have seen above, $Var[Z]$ is well-defined only for $n > 2$ and, as a consequence, also $Var[X]$ is well-defined only for $n > 2$.

Problem 2

Gamma mgf

The gamma pdf is given

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \tag{18}$$

for $0 < x < \infty, \alpha > 0, \beta > 0$, where $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}dt$ denotes the gamma function. The mgf is

$$\begin{aligned}
M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta} \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/(\frac{\beta}{1-\beta t})} \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha)(\frac{\beta}{1-\beta t})^\alpha \\
&= (\frac{1}{1-\beta t})^\alpha
\end{aligned} \tag{19}$$

for $t < \frac{1}{\beta}$. If $t \geq \frac{1}{\beta}$ then the quantity $1 - \beta t$ is nonpositive and the integral is infinite. Thus, the mgf of the gamma distribution exists only if $t < 1/\beta$.

Gamma skewness and kurtosis

Gamma mgf giving the logarithmic moment-generating function as

$$\begin{aligned}
\mu_0 &= R(t) = -\alpha \ln(1-\beta t) \\
\mu_1 &= R'(t) = \frac{\alpha\beta}{1-\beta t} \\
\mu_2 &= R''(t) = \frac{\alpha\beta^2}{(1-\beta t)^2} \\
\mu_3 &= R^{(3)}(t) = \frac{2\alpha\beta^3}{(1-\beta t)^3} \\
\mu_4 &= R^{(4)}(t) = \frac{6\alpha\beta^4}{(1-\beta t)^4}
\end{aligned} \tag{20}$$

The skewness and kurtosis are then

$$\begin{aligned}
\gamma_1 &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{2}{\sqrt{\alpha}} \\
\gamma_2 &= \frac{\mu_4}{\mu_2^2} = \frac{6}{\alpha} + 3
\end{aligned} \tag{21}$$

Problem 3

(a)

We have two Normal distributions $N_1 \sim N(0,1)$ and $N_2 \sim N(0,10)$ with the same mean (which without loss of generality is set to zero) but with different standard deviations. we form a mixture, M, of these two different distributions, i.e. a distribution where the probability of drawing from the first distribution is $p_1 = 95\%$ and from the second distribution is $p_2 = 5\%$. The kurtosis of the mixture is as follows:

$$\frac{3(p_1\sigma_1^4 + p_2\sigma_2^4)}{(p_1\sigma_1^2 + p_2\sigma_2^2)^2} - 3 = \frac{3(0.95 \times 1 + 0.05 \times 100)}{(0.95 \times 1 + 0.05 \times 10)^2} - 3 = 5.489893 \tag{22}$$

(b)

Actually I have used that formula as a conclusion in my part(a). I will show how to derive it in this part. Let $X_1, ..., X_n$ denote random variables from the n component distributions, and let X denote a random variable from the mixture distribution. Then, for any function $H()$ for which $\mathrm{E}[H(X_i)]$ exists, and assuming that the component densities $p_i(x)$ exist (to distinguish, I use w to denote weights here.)

$$\mathrm{E}[H(X)] = \int_{-\infty}^{\infty} H(x) \sum_{i=1}^{n} w_i p_i(x)\, dx \tag{23}$$

$$= \sum_{i=1}^{n} w_i \int_{-\infty}^{\infty} p_i(x) H(x)\, dx = \sum_{i=1}^{n} w_i\, \mathrm{E}[H(X_i)]. \tag{24}$$

The relation,

$$\mathrm{E}[H(X)] = \sum_{i=1}^{n} w_i\, \mathrm{E}[H(X_i)], \tag{25}$$

holds more generally.
It is a trivial matter to note that the jth moment about zero (i.e. choosing $H(x) = x^j$) is simply a weighted average of the jth moments of the components. Moments about the mean $H(x) = (x\mu)^j$ involve a binomial expansion:

$$\mathrm{E}[(X-\mu)^j] = \sum_{i=1}^{n} w_i\, \mathrm{E}[(X_i - \mu_i + \mu_i - \mu)^j] \tag{26}$$

$$= \sum_{i=1}^{n} \sum_{k=0}^{j} \left(\begin{array}{c} j \\ k \end{array} \right) (\mu_i - \mu)^{j-k} w_i\, \mathrm{E}[(X_i - \mu_i)^k], \tag{27}$$

where μ_i denotes the mean of the ith component.
In case of a mixture of one-dimensional normal distributions with weights w_i, means μ_i and variances σ_i^2, the total mean and variance will be:

$$\mathrm{E}[X] = \mu = \sum_{i=1}^{n} w_i \mu_i, \tag{28}$$

$$\mathrm{E}[(X-\mu)^2] = \sigma^2 = \sum_{i=1}^{n} w_i((\mu_i - \mu)^2 + \sigma_i^2) \tag{29}$$

$$\tag{30}$$

Problem 3 [(b)] continued on next page...

Similarly we are able to derive $\mathrm{E}[(X-\mu)^3], \mathrm{E}[(X-\mu)^4]$ and the kurtosis is

$$\gamma_2 = \frac{\mathrm{E}[(X-\mu)^4]}{(\mathrm{E}[(X-\mu)^2])^2} - 3 \tag{31}$$

$$= \frac{1}{\sigma^4}\sum_{i=1}^{n} w_i[3\sigma_i^4 + 6(\mu_i-\mu)^2\sigma_i^2 + (\mu_i-\mu)^4] - 3 \tag{32}$$

In our problem, there are only two components with $p_1 = p, p_2 = 1-p$, $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1, \sigma_2 = \sigma$ as parameters, then

$$\gamma_2 = \frac{3[p+(1-p)\sigma^4]}{[p+(1-p)\sigma^2]^2} - 3 \tag{33}$$

(c)

We can rewrite γ_2 as

$$\gamma_2 = \frac{3(1-p)\sigma^4 + 3p}{(1-p)^2\sigma^4 + 2(1-p)p\sigma^2 + p^2} - 3 \tag{34}$$

Then we take $\sigma = 1/(1-p)$ (my intuition is to make p very close to 1 and σ very large)

$$\gamma_2 = \frac{3(1-p) + 3p(1-p)^4}{(1-p)^2 + 2(1-p)^3p + p^2(1-p)^4} - 3 \tag{35}$$

When $p \to 1$, or when $q = 1-p \to 0$,

$$\lim_{p\to 1}\gamma_2 = \lim_{q\to 0}\frac{3q + 3(1-q)q^4}{q^2 + 2(1-q)q^3 + (1-q)^2q^4} - 3 = \lim_{q\to 0}\frac{3}{q} - 3 \to \infty \tag{36}$$

which means kurtosis can be arbitrarily large.
To find values of σ and p so that the kurtosis is 10000, we can just let $\sigma = 1/(1-p)$

$$\gamma_2 = \frac{3(1-p) + 3p(1-p)^4}{(1-p)^2 + 2(1-p)^3p + p^2(1-p)^4} - 3 = 10000 \tag{37}$$

which generates

$$p \approx 0.9997, \sigma \approx 3334.3 \tag{38}$$

(d)

In this part, our goal is to construct p and σ, to reach a kurtosis at least M. Again we can continue using the same constrution as did in part(c), take $p = p_0 + \epsilon$ ($\epsilon \in (0, 1-p_0)$) and $\sigma = 1/(1-p)$, we still have the kurtosis as

$$\gamma_2 = \frac{3(1-p) + 3p(1-p)^4}{(1-p)^2 + 2(1-p)^3p + p^2(1-p)^4} - 3 \tag{39}$$

No matter p_0 how close to 1, that is $p_0 \to 1$, we also have $p = p_0 + \epsilon \to 1$, then we let

$$\lim_{p\to 1}\gamma_2 = 3/(1-p) - 3 \geq M \tag{40}$$

which generates

$$p \geq \frac{M}{M+3} \tag{41}$$

If we take

$$p = \max\{\frac{M}{M+3}, p_0\}, \sigma = \frac{1}{1-p} \tag{42}$$

we are able to get a kurtosis at least M.

Problem 4

(A)

For a Poisson distribution,

$$f(x_1, x_2, \ldots, x_n|\lambda) = \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{x_1!x_2!\ldots x_n!} \tag{43}$$

$$\ln f = -n\lambda + (\ln\lambda)\sum x_i - \ln(\prod x_i!) \tag{44}$$

$$\frac{d(\ln f)}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0 \tag{45}$$

$$\hat{\lambda} = \frac{\sum x_i}{n} \tag{46}$$

It is easy to verify that

$$\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda) = -\frac{x}{\lambda^2} \tag{47}$$

Therefore the Fisher information is

$$I(\lambda) = -E[-\frac{X}{\lambda^2}] = \frac{1}{\lambda} \tag{48}$$

The asymptotic normality of MLE is

$$\sqrt{n}(\hat{\lambda} - \lambda_0) \to N(0, \lambda_0) \tag{49}$$

(B)

For a normal distribution $N(\mu, \sigma^2)$ where σ^2 is known. The likelihood function is

$$L(x_1, x_2, \ldots, x_n|\mu) = (2\pi\sigma^2)^{-n/2}\exp(-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2) \tag{50}$$

The log-likelihood function is

$$l(x_1, x_2, \ldots, x_n|\mu) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_j - \mu)^2 \tag{51}$$

The partial derivative of the log-likelihood with respect to the mean is

$$\frac{\partial}{\partial\mu}l(x_1, x_2, \ldots, x_n|\mu) = \frac{1}{\sigma^2}(\sum_{j=1}^{n}x_j - n\mu) = 0 \tag{52}$$

Therefore the maximum likelihood estimator of the mean is

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n}x_j \tag{53}$$

It is easy to verify that

$$\frac{\partial^2}{\partial\mu^2}l(x|\mu) = -\frac{1}{\sigma^2} \tag{54}$$

Therefore the Fisher information is

$$I(\mu) = -E[-\frac{1}{\sigma^2}] = \frac{1}{\sigma^2} \tag{55}$$

The asymptotic normality of MLE is

$$\sqrt{n}(\hat{\mu} - \mu_0) \to N(0, \sigma^2) \tag{56}$$

Problem 4 continued on next page...

(C)

For a normal distribution $N(\mu, \sigma^2)$ where $\mu = 0$ is known. The likelihood function is

$$L(x_1, x_2, \dots, x_n|\sigma) = (2\pi\sigma^2)^{-n/2} \exp(-\frac{1}{2\sigma^2} \sum_{j=1}^{n} x_j^2) \tag{57}$$

The log-likelihood function is

$$l(x_1, x_2, \dots, x_n|\sigma) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} x_j^2 \tag{58}$$

The partial derivative of the log-likelihood with respect to the variance is

$$\frac{\partial}{\partial\sigma^2} l(x_1, x_2, \dots, x_n|\sigma) = \frac{1}{2\sigma^2}(\frac{1}{\sigma^2} \sum_{j=1}^{n} x_j^2 - n) \tag{59}$$

Therefore the maximum likelihood estimator of the variance is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} x_j^2 \tag{60}$$

It is easy to verify that

$$\frac{\partial^2}{(\partial\sigma^2)^2} l(x|\sigma) = -\frac{x^2}{\sigma^6} + \frac{1}{2\sigma^4} \tag{61}$$

Therefore the Fisher information is

$$I(\sigma^2) = -E[-\frac{X^2}{\sigma^6} + \frac{1}{2\sigma^4}] = \frac{1}{2\sigma^4} \tag{62}$$

The asymptotic normality of MLE is

$$\sqrt{n}(\hat{\sigma}^2 - \sigma_0^2) \to N(0, 2\sigma_0^4) \tag{63}$$

Problem 5

MLE of b

Uniform distribution has p.d.f.

$$f(x|b) = \begin{cases} \frac{1}{b-a}, & a \le b \\ 0, & \text{otherwise} \end{cases} \tag{64}$$

First note that $f(x|b) = 1/(b-a)$, for $a \le x \le b$ and 0 elsewhere. Then it is easy to see that the likelihood function is given by

$$L(b|x) = \prod_{i=1}^{n} \frac{1}{b-a} I(X_i \in [a, b]) = (b-a)^{-n} I(max(X_1, \dots, X_n) \le b) \tag{65}$$

Here the indicator function $I(A)$ equals to 1 if event A happens and 0 otherwise. What the indicator above means is that the likelihood will be equal to 0 if at least one of the factors is 0 and this will happen if at least one observation X_i will fall outside of the allowed interval $[a, b]$. Another way to say it is that the maximum among observations will exceed b, i.e.

$$L(b) = 0 \text{ if } b < \max(X_1, X_2, \dots, X_n) \tag{66}$$

Problem 5 [MLE of b] continued on next page. . .

and

$$L(b) = \frac{1}{(b-a)^n} \text{ if } b \geq \max(X_1, X_2, \dots, X_n) \tag{67}$$

which is a decreasing function in the second part. Therefore $\hat{b} = \max(X_1, \dots, X_n)$ is the MLE.

MLE of a

Follow the same step in the previous part and using the property of symmetry, MLE of a is

$$\hat{a} = \min(X_1, \dots, X_n) \tag{68}$$

MLE of mean μ

According to the invariance property of MLE, $\hat{a}$ and $\hat{b}$ are the MLE of a and b, then $g(\hat{a}, \hat{b})$ is the MLE of $g(a, b)$ for any one-to-one function g, respectively for a and b. So for the mean μ, we will let

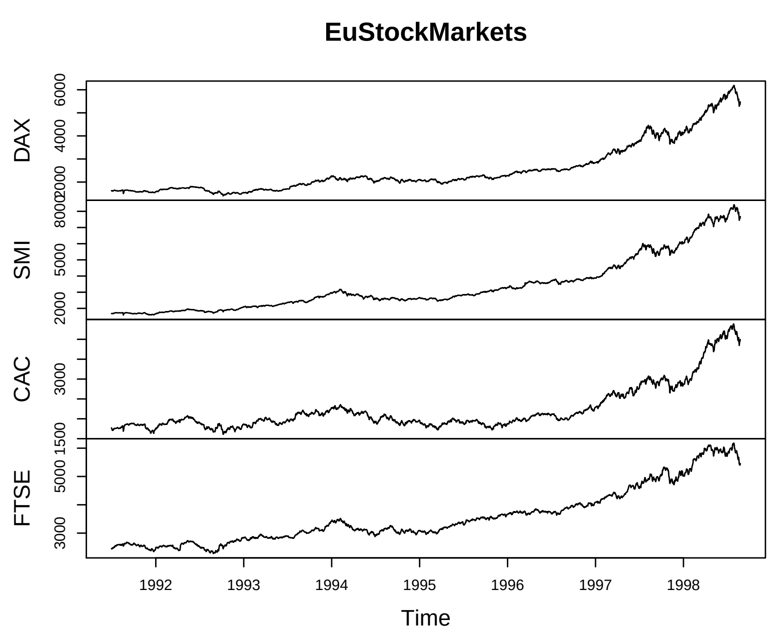
$$g(a, b) = \frac{a+b}{2} \tag{69}$$

Therefore, MLE of μ is

$$\hat{\mu} = g(\hat{a}, \hat{b}) = \frac{\hat{a} + \hat{b}}{2} = \frac{\max(X_1, X_2, \dots, X_n) + \min(X_1, \dots, X_n)}{2} \tag{70}$$

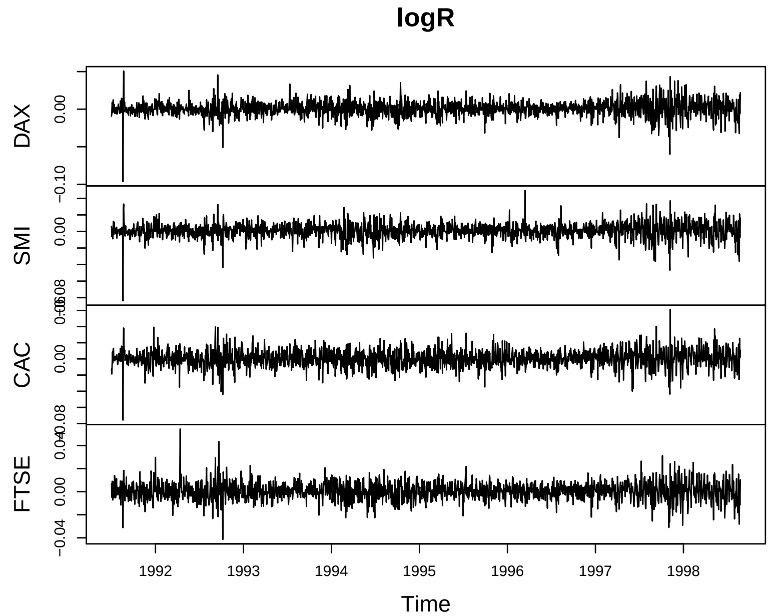
Problem 6

Problem 1



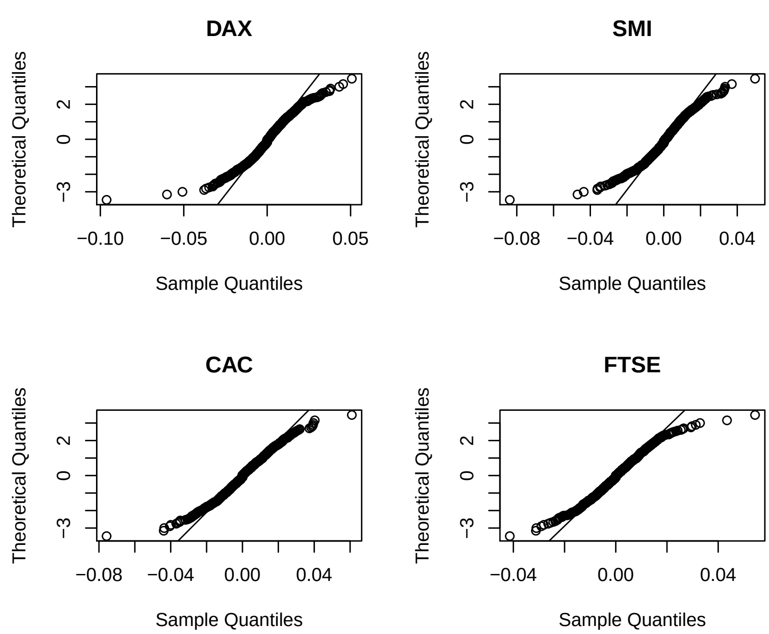
The series does not look stationary. The fluctuations does not seem to be of constant size, the volatility becomes larger and larger when time passes.

Problem 2



The series looks stationary. The fluctuations seems to be of constant size.

Problem 3



The marginal distribution of all series are symmetric and their tails appear heavier than a normal distribution. This is the printed out results

```
    Shapiro-Wilk normality test

data: logR[, i]
```

```
W = 0.9538, p-value < 2.2e-16


    Shapiro-Wilk normality test

data:  logR[, i]
W = 0.9554, p-value < 2.2e-16


    Shapiro-Wilk normality test

data:  logR[, i]
W = 0.982, p-value = 1.574e-14


    Shapiro-Wilk normality test

data:  logR[, i]
W = 0.9799, p-value = 1.754e-15
```

All the Shapiro-Wilk tests reject the null hypothesis of normality with extremely small p-values.

I don't have time to accomplish the left tests, sorry.