Stat 111 PSET 04

Luke Kolar

1a)

By the Delta method, if $g(x) = \sqrt{x}$ and $(g'(x))^2 = \frac{1}{4x}$ then:

$$\begin{aligned}
\sqrt{n}\{S - \mathrm{sd}(Y_1)\} &\xrightarrow{d} N(0, g'[\mathrm{Var}(Y_1)]^2 \times \mathrm{Var}((Y_1 - \mathrm{E}[Y_1])^2)) \\
\sqrt{n}\{S - \mathrm{sd}(Y_1)\} &\xrightarrow{d} N(0, \frac{1}{4\mathrm{Var}(Y_1)} \times \mathrm{Var}((Y_1 - \mathrm{E}[Y_1])^2)) \\
\sqrt{n}\{S - \mathrm{sd}(Y_1)\} &\xrightarrow{d} N(0, \frac{\mathrm{Var}((Y_1 - \mathrm{E}[Y_1])^2)}{4\mathrm{Var}(Y_1)})
\end{aligned}$$

Therefore:

$$\mathrm{AsyVar} = \tfrac{\mathrm{Var}((Y_1 - \mathrm{E}[Y_1])^2)}{4\mathrm{Var}(Y_1)}$$

1b)

We have standard error formula $\mathrm{SE}(\hat{\theta}) = \sqrt{\mathrm{Var}(\hat{\theta})}$, so:

$$\begin{aligned}
\mathrm{SE}(S) &= \sqrt{\frac{\mathrm{AsyVar}}{n}} \\
&= \sqrt{\frac{\mathrm{Var}((Y_1 - \mathrm{E}[Y_1])^2)}{4n\mathrm{Var}(Y_1)}}
\end{aligned}$$

1c)

To estimate AsyVar we can break our result into two parts, with α being our numerator and β as the denominator [note: I am unsure if these symbols are "off-limits"; apologies if so]. We have:

$$\begin{aligned}
\alpha &= \mathrm{Var}((Y_1 - \mathrm{E}[\overline{Y}])^2) \\
\beta &= 4\mathrm{Var}(Y_1)
\end{aligned}$$

We need to simplify α:

$$
\begin{aligned}
\alpha &= \mathrm{Var}((Y_1 - \mathrm{E}[\overline{Y}])^2) \\
&= \mathrm{E}[(Y_1 - \mathrm{E}[\overline{Y}])^4] - \mathrm{E}[(Y_1 - \mathrm{E}[\overline{Y}])^2]^2 \\
&= \mathrm{E}[(Y_1 - \mathrm{E}[\overline{Y}]^4] - \mathrm{Var}(Y_1)^2
\end{aligned}
$$

Now we use Law of Large Numbers to estimate:

$$
\begin{aligned}
\hat{\alpha} &= \frac{\sum_{i=1}^{n}(Y_i - \mathrm{E}[\overline{Y}])^4}{n} - S^4 \\
\hat{\beta} &= 4S^2
\end{aligned}
$$

So we can consistently estimate AsyVar with:

$$
\mathrm{As\hat{y}Var} = \sqrt{\tfrac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathrm{E}[\overline{Y}])^4 - S^4}{4S^2}}
$$

2a)

$$
\begin{gathered}
p_j = \tfrac{\exp(\theta x_j)}{1+\exp(\theta x_j)} \\
p_j(1 + \exp(\theta x_j)) = \exp(\theta x_j) \\
p_j + p_j(\exp(\theta x_j)) = \exp(\theta x_j) \\
p_j = \exp(\theta x_j) - p_j(\exp(\theta x_j)) \\
\tfrac{p_j}{1-p_j} = \exp(\theta x_j) \\
\log(\tfrac{p_j}{1-p_j}) = \theta x_j
\end{gathered}
$$

2b)

We have likelihood function:

$$
\begin{aligned}
L(\theta; \mathbf{y}|\mathbf{x}) &= \prod_{i=1}^{n}(p_i^{y_i})(1-p_i)^{(1-y_i)} \\
&= \prod_{i=1}^{n}(p_i^{y_i})\frac{1-p_i}{(1-p_i)^{y_i}} \\
&= \prod_{i=1}^{n}(\frac{p_i}{1-p_i})^{y_i}(1-p_i) \\
&= \prod_{i=1}^{n}\exp(\theta y_i x_i)(1-p_i)
\end{aligned}
$$

So log-likelihood is:

$$
\begin{aligned}
\log L(\theta; \mathbf{y}|\mathbf{x}) &= \sum_{i=1}^{n} (\theta y_i x_i + \log(1 - p_i)) \\
&= \sum_{i=1}^{n} (\theta y_i x_i + \log(1 - \frac{\exp(\theta x_i)}{1 + \exp(\theta x_i)})) \\
&= \sum_{i=1}^{n} (\theta y_i x_i + \log(\frac{1}{1 + \exp(\theta x_i)})) \\
&= \sum_{i=1}^{n} (\theta y_i x_i + \log(1) - \log(1 + \exp(\theta x_i))) \\
&= \sum_{i=1}^{n} (\theta y_i x_i - \log(1 + \exp(\theta x_i)))
\end{aligned}
$$

2c)

$$
\begin{aligned}
s(\theta; \mathbf{y}|\mathbf{x}) &= \frac{\partial \log L}{\partial \theta} \\
&= \sum_{i=1}^{n} (\frac{\partial}{\partial \theta} \theta y_i x_i - \frac{\partial}{\partial \theta} \log(1 + \exp(\theta x_i))) \\
&= \sum_{i=1}^{n} (y_i x_i - \frac{x_i \exp(\theta x_i)}{1 + \exp(\theta x_i)}) \\
&= \sum_{i=1}^{n} x_i (y_i - \frac{\exp(\theta x_i)}{1 + \exp(\theta x_i)}) \\
&= \sum_{i=1}^{n} x_i (y_i - p_i)
\end{aligned}
$$

2d)

$$
\begin{aligned}
s'(\theta; \mathbf{y}|\mathbf{x}) &= \frac{\partial^2 \log L}{\partial \theta^2} \\
&= \sum_{i=1}^{n} x_i (\frac{\partial}{\partial \theta} y_i - \frac{\partial}{\partial \theta} \frac{\exp(\theta x_i)}{1 + \exp(\theta x_i)}) \\
&= -\sum_{i=1}^{n} x_i (\frac{\partial}{\partial \theta} \frac{\exp(\theta x_i)}{1 + \exp(\theta x_i)}) \\
&= -\sum_{i=1}^{n} x_i (\frac{x_i \exp(\theta x_i)}{1 + \exp(\theta x_i)} + \frac{-x_i \exp(\theta x_i)^2}{(1 + \exp(\theta x_i))^2}) \\
&= -\sum_{i=1}^{n} x_i (x_i p_i - x_i p_i) \\
&= -\sum_{i=1}^{n} x_i^2 p_i (1 - p_i)
\end{aligned}
$$

2e)

As long as the predictors $(x_1, ..., x_n)$ aren't all zero, the function is **concave** because its second derivative found in part d is always less than zero.

2f)

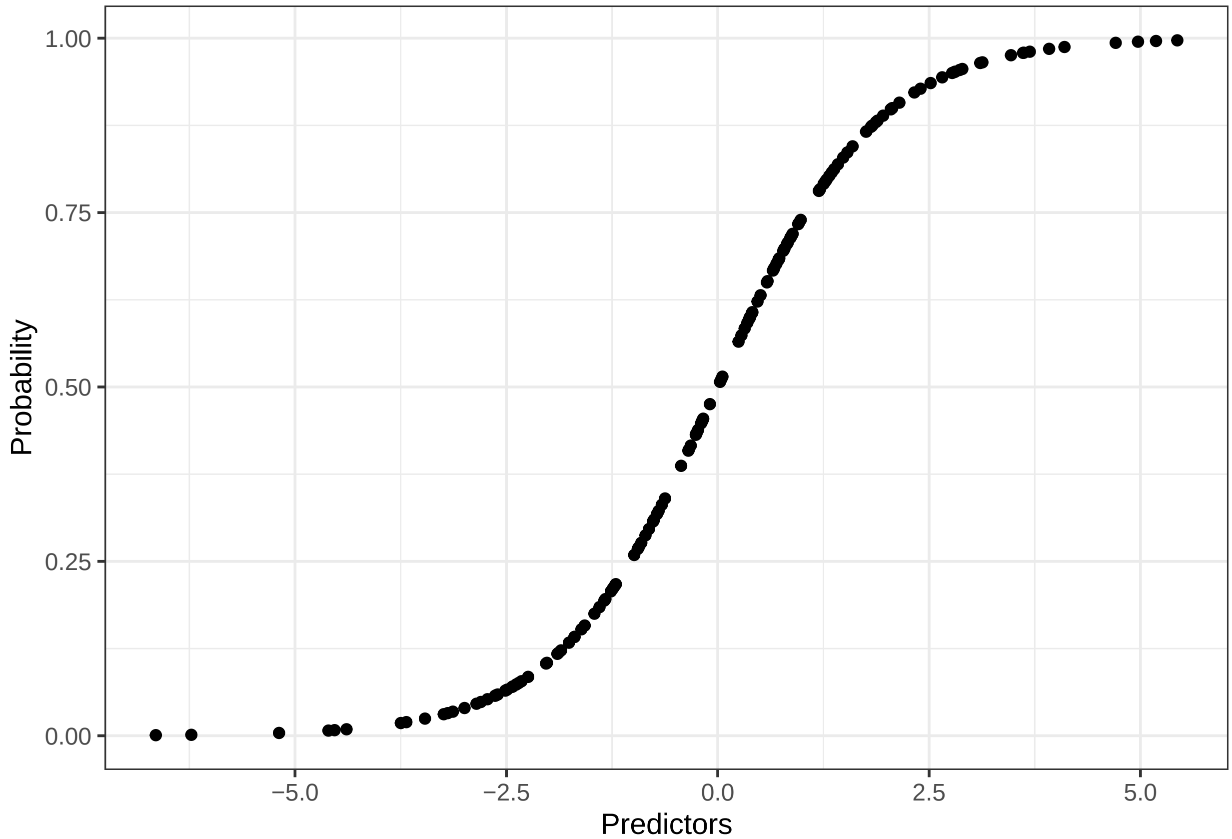
```
data <- read.csv("predictBinary.csv")

newtons_method <- function(num) {

  df <- data.frame(rep = rep(0, num),
                   MLE_hat = rep(0, num))

  for (i in 1:num){
    ifelse(i == 1, theta <- 0, theta <- new_theta)

    prob <- exp(theta * data$X) / (1 + exp(theta * data$X))
    score <- sum(data$X * (data$Y - prob))
    score_deriv <- (-sum((data$X)^(2) * prob * (1 - prob)))

    new_theta <- theta + score / (-score_deriv)

    df$rep[i] = i
    df$MLE_hat[i] = new_theta
  }

  return(df)
}

head(newtons_method(15)$MLE_hat, 7)
```

```
## [1] 0.5752882 0.8947623 1.0404841 1.0627315 1.0631630 1.0631632 1.0631632
```

Newton's Method converges to an MLE $\hat{\theta}$ approximation of **1.0631632** by the sixth iteration.

2g)

```
data$prob <- exp(1.0631632 * data$X) / (1 + exp(1.0631632 * data$X))

ggplot(data, aes(x = X, y = prob))+ geom_point() +
  labs(main = "Fitted probabilities for predictors",
       x = "Predictors", y = "Probability") + theme_bw()
```



The slope is positive at all points. It appears to have an inflection point at $p = 0.5$.

2h)

We know $\text{SE}(\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta})}$, and page 111 of the lecture notes tells us that variance of an asymptotic distribution of $\hat{\theta}$ is 1 over the Fisher information. So:

$$\text{SE}(\hat{\theta}) = \sqrt{\frac{1}{\mathcal{I}(\hat{\theta})}}$$

```
data$prob <- exp(1.0631632 * data$X) / (1 + exp(1.0631632 * data$X))
score <- sum(data$X * (data$Y - data$prob))
score_deriv <- (sum((data$X)^(2) * data$prob * (1 - data$prob)))

1 / (score_deriv^0.5)
```

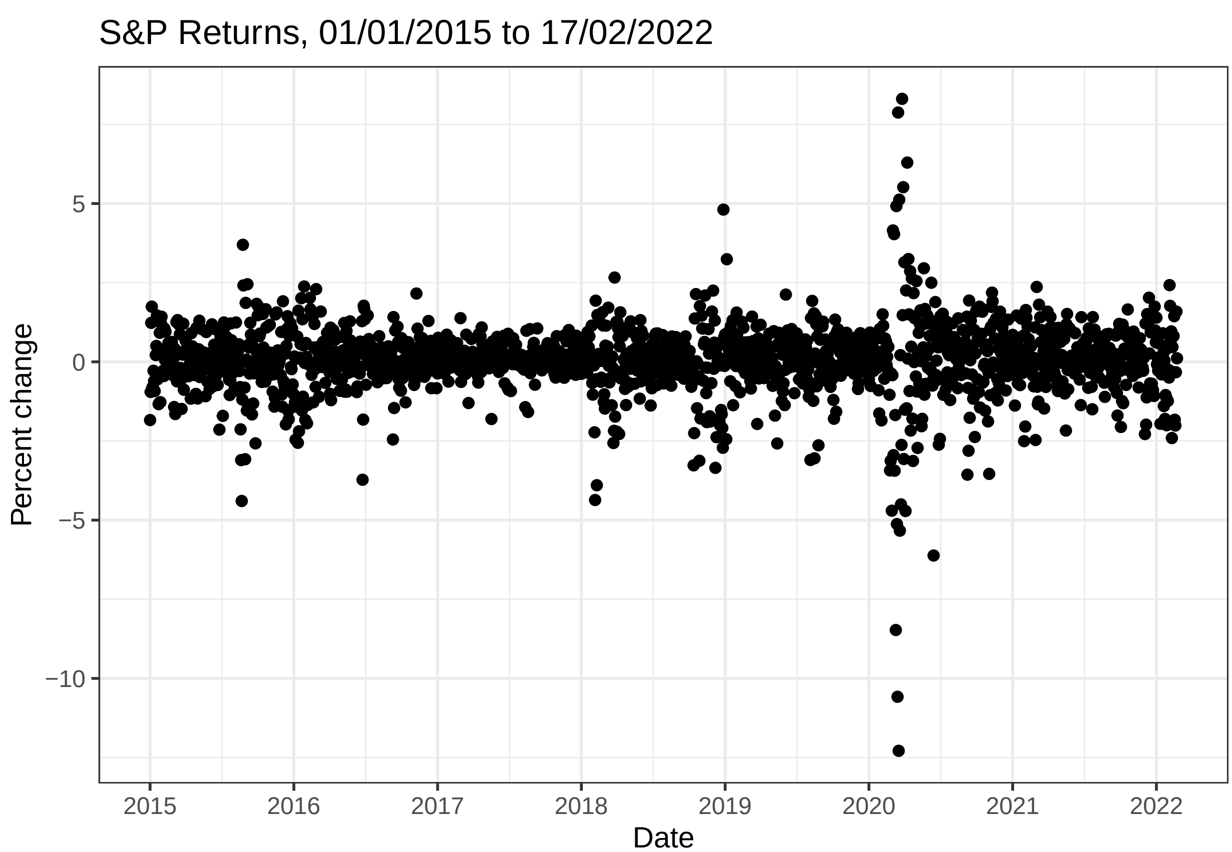
```
## [1] 0.1687642
```

Our estimate is **0.1687642**.

3a)

```
sp <- read.csv("SP500time.csv")

year_days <- function(year){sp %>% filter()}


ggplot(sp, aes(x = X, y = Y)) + geom_point() +
  xlim(1, 1794) +
  scale_x_continuous(breaks = c(1, 1 + 251*seq(1:7)),
                     labels = c(2015:2022)) +
  labs(title = "S&P Returns, 01/01/2015 to 17/02/2022",
       x = "Date", y = "Percent change") +
  theme_bw()
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```



3b)

i)

Per example 4.8.3 in the textbook, we have:

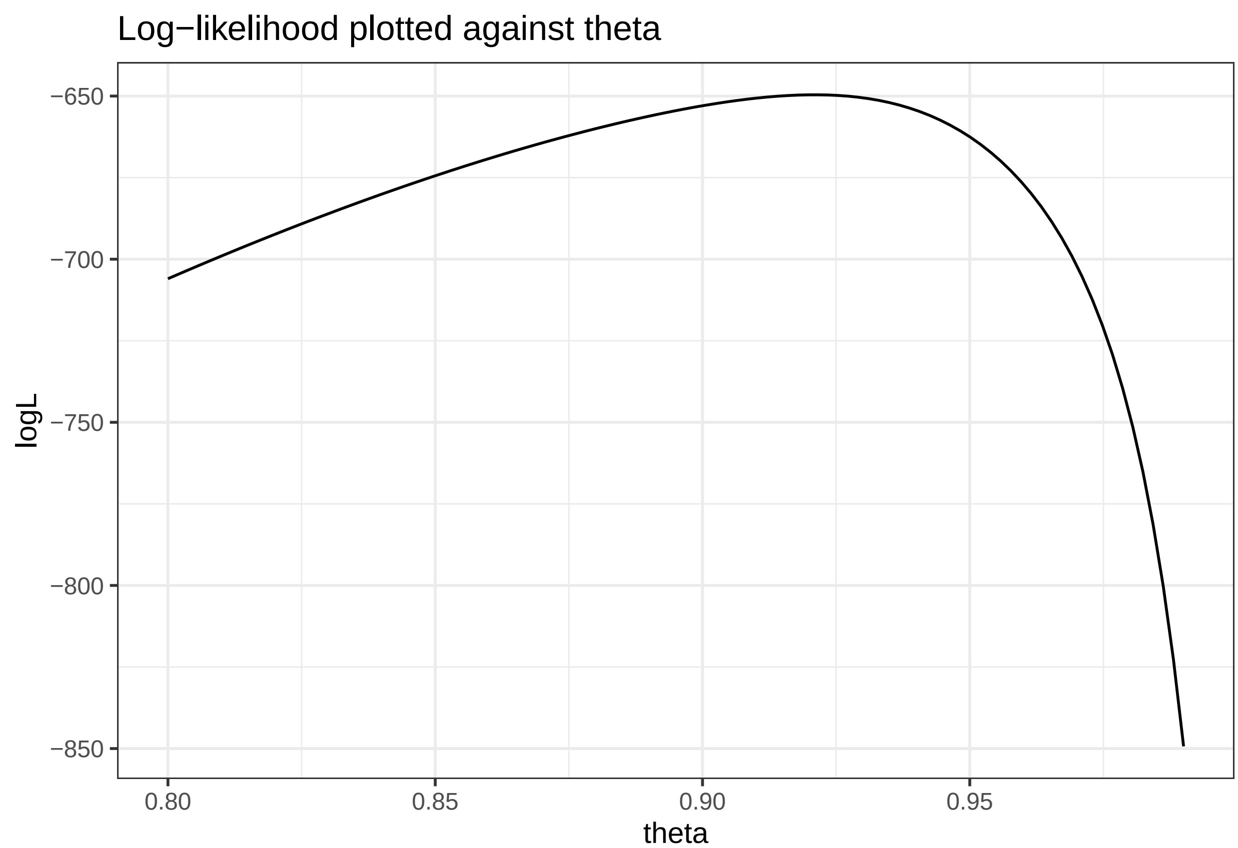
$$
\begin{aligned}
\log L(\theta) &= \log f(y_2, ..., y_n | y_1, \theta) \\
&= \log P(Y_2 = y_2, ..., Y_n = y_n | Y_1 = y_1, \theta) \\
&= \log \textstyle\sum_{i=2}^{n} P(Y_i = y_i | Y_{i-1} = y_{i-1}, ..., Y_1 = y_1, \theta) \\
&= \log \textstyle\sum_{i=2}^{n} P(Y_i = y_i | Y_{i-1} = y_{i-1}, ..., Y_1 = y_1, \theta)
\end{aligned}
$$

Now we use the normal distribution PDF:

$$
\begin{aligned}
\log L(\theta) &= \sum_{i=2}^{n} (\log(\frac{1}{\sigma_i \sqrt{2\pi}}) - y_i^2 / 2\sigma_i^2) \\
&= c + \sum_{i=2}^{n} (\log(\frac{1}{\sigma_i}) - y_i^2 / 2\sigma_i^2)
\end{aligned}
$$

ii)

```
logL_df <- data.frame(logL = rep(0, nrow(sp)),
                      var = rep(0, nrow(sp)))

logL_sum_df <- data.frame(theta = rep(0, 101),
                          sum = rep(0, 101))

for (j in (1:101)){
  theta = 0.8 + 0.19*((j - 1) / 100)
  logL_sum_df$theta[j] = theta
  var = 1
  logL_df$var[1] = var

  for (i in 2:nrow(sp)){
    var = theta*var + (1 - theta)*((sp$Y[i-1])^2)
    log_funct = log(1 / (var^0.5)) - ((sp$Y[i])^2 / (2*var))

    logL_df$var[i] = var
    logL_df$logL[i] = log_funct

  }
  logL_sum_df$sum[j] = sum(logL_df$logL)
  is.MLE = ifelse(j == 1, -1000, ifelse(is.MLE < logL_sum_df$sum[j],
                                        logL_sum_df$sum[j], is.MLE))
}

logL_sum_df %>% ggplot(aes(x = theta, y = sum)) +
  geom_line()  + theme_bw() +
  labs(title = "Log-likelihood plotted against theta",
       y = "logL")
```



iii)

```
logL_sum_df %>%
  filter(is.MLE == sum) %>%
  select(theta) %>% pull()
```

```
## [1] 0.9216
```

Using the variable I called in the last line to my `for` loop from part ii, we can find the MLE: **0.9216**.

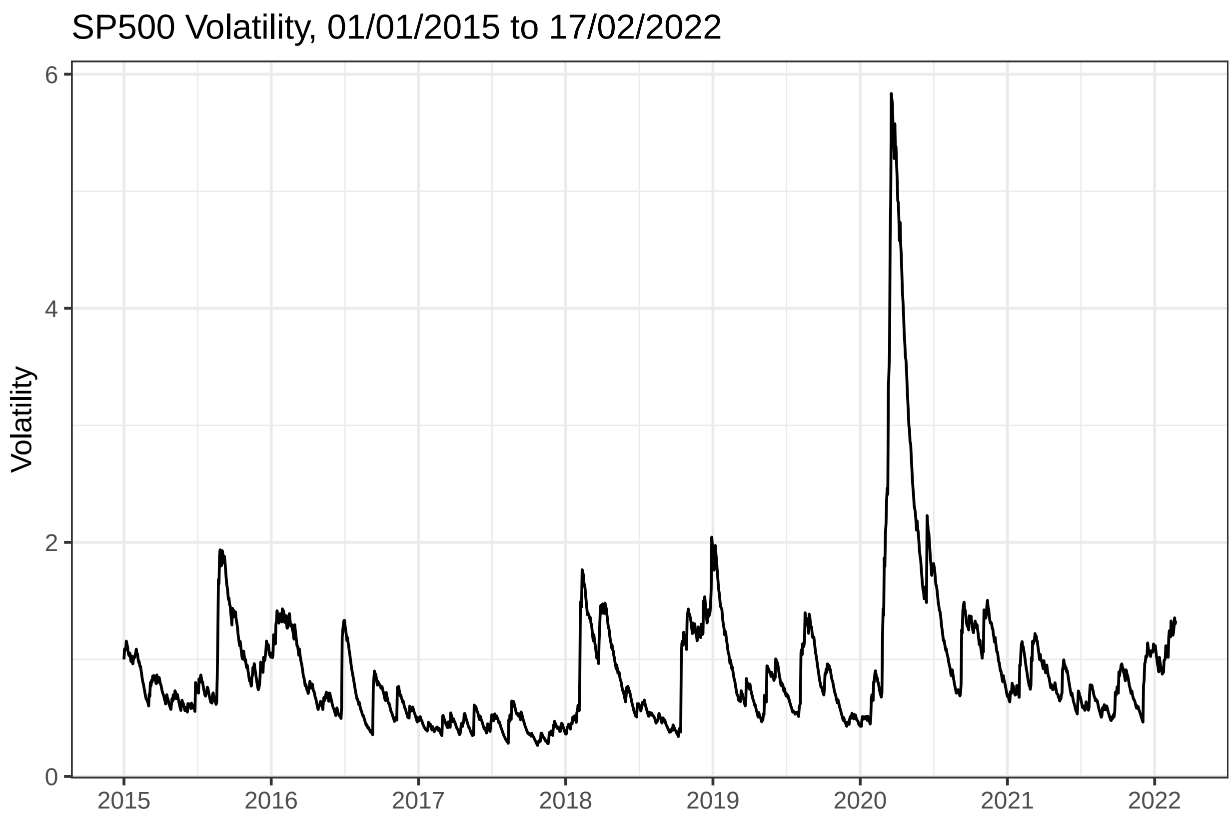
3c)

i)

```
sp$vol[1] = 1

for (i in 2:nrow(sp)){
  ifelse(i == 2, var <- 1, var <- var)

  var = 0.9216*var + (1 - 0.9216)*((sp$Y[i-1])^2)

  sp$vol[i] = var^0.5
}
```

```r
ggplot(sp, aes(x = X, y = vol)) + geom_line() +
  xlim(1, 1794) +
  scale_x_continuous(breaks = c(1, 1 + 251*seq(1:7)),
                     labels = c(2015:2022)) +
  labs(x = "", y = "Volatility",
       title = "SP500 Volatility, 01/01/2015 to 17/02/2022") + theme_bw()
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```



ii)

```r
quantile(sp$vol, prob = c(0.01, 0.5, 0.99))
```

```
##        1%       50%       99%
## 0.3069453 0.7504895 4.5467073
```

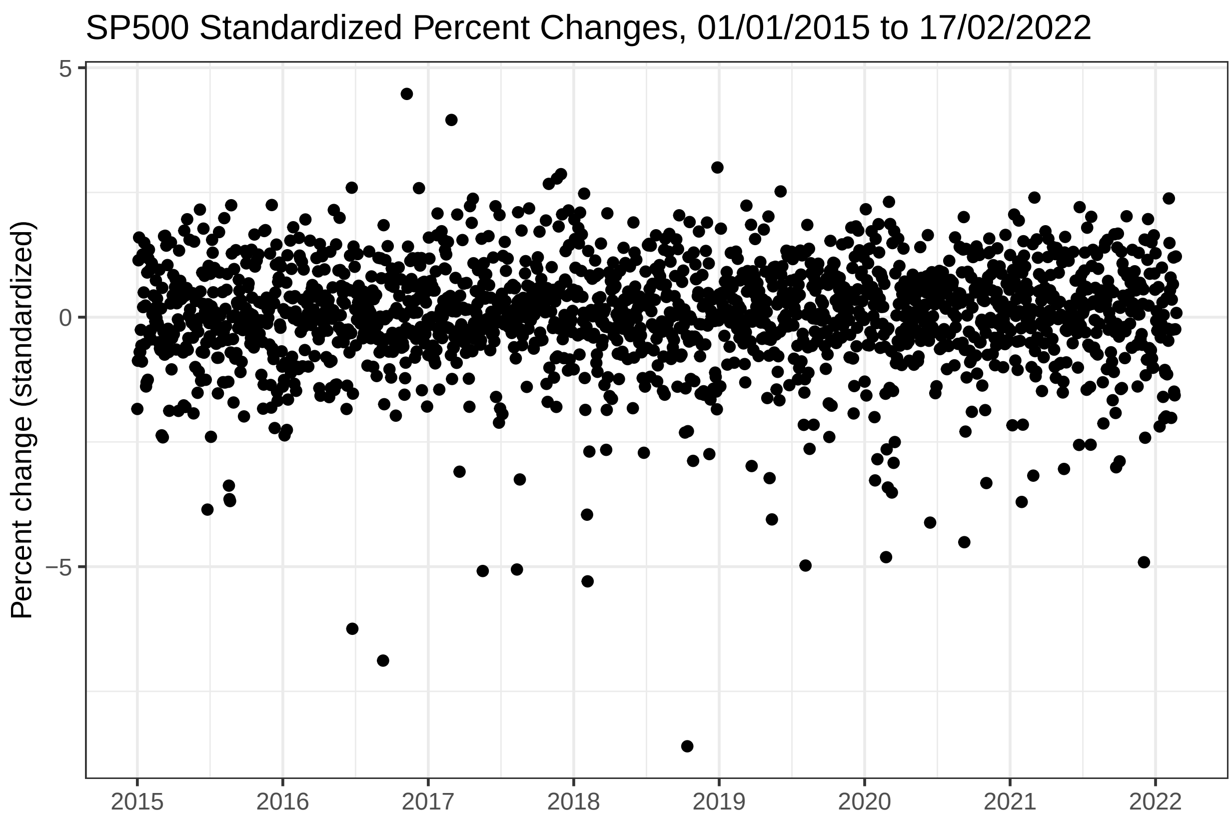
Quantile estimates printed above.

iii)

```r
ggplot(sp, aes(x = X, y = Y/vol)) + geom_point() +
  xlim(1, 1794) +
  scale_x_continuous(breaks = c(1, 1 + 251*seq(1:7)),
```

```
                    labels = c(2015:2022)) +
  labs(x = "", y = "Percent change (standardized)",
       title = "SP500 Standardized Percent Changes, 01/01/2015 to 17/02/2022") + theme_bw()
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```



iv)

The standardized model seems to have removed the large swings in the graph that let to higher variance - particularly around the onset of the pandemic in 2020.

4a)

We have likelihood function:

$$L(p; y_1, ..., y_n) = \prod_{i=1}^{n} (p^{y_i})(1-p)^{(1-y_i)}$$

So log-likelihood is:

$$\log L(p; y_1, ..., y_n) = \sum_{i=1}^{n} (y_i \log p) + \sum_{i=1}^{n} [(1-y_i)\log(1-p)]$$

4b)

$$\begin{aligned} s(p) &= \frac{\partial \log L}{\partial p} \\ &= \frac{1}{p}\sum_{i=1}^{n}(y_i) - \frac{1}{1-p}\sum_{i=1}^{n}(1-y_i) \\ &= \frac{n\overline{y}}{p} - \frac{n-n\overline{y}}{1-p} \end{aligned}$$

4c)

We set score equal to zero:

$$\begin{aligned} 0 &= \frac{n\overline{y}}{p} - \frac{n-n\overline{y}}{1-p} \\ \frac{n-n\overline{y}}{1-p} &= \frac{n\overline{y}}{p} \\ p(1-p)\times\frac{n-n\overline{y}}{1-p} &= \frac{n\overline{y}}{p}\times p(1-p) \\ np - np\overline{y} &= n\overline{y} - np\overline{y} \end{aligned}$$

Our MLE of p is:

$$\hat{p} = \overline{y}$$

4d)

We solve using the information equality (pg. 107 of Stat 111 notes):

$$\begin{aligned} \mathcal{I}_Y &= -\mathrm{E}[s'(p; Y_1, ..., Y_n)] \\ &= -\mathrm{E}[-\frac{n\overline{Y}}{p^2} - \frac{n-n\overline{Y}}{(1-p)^2}] \\ &= \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} \\ &= \frac{n}{p} + \frac{n}{(1-p)} \\ &= \frac{n}{p(1-p)} \end{aligned}$$

4e)

Because the random variables are i.i.d., we just divide by n:

$$\mathcal{I}_{Y_1} = \frac{1}{p(1-p)}$$

4f)

$$
\begin{aligned}
D_{KL}(F_{\mathbf{Y}|p^*}||F_{\mathbf{Y}|p}) &= \mathrm{E}[\log L(p^*;Y_1,...,Y_n)] - \mathrm{E}[\log L(p;Y_1,...,Y_n)] \\
&= n[p^*\log(p^*) + (1-p^*)\log(1-p^*) - p^*\log(p) - (1-p^*)\log(1-p)] \\
&= n[p^*\log(\frac{p^*}{p}) + (1-p^*)\log(\frac{1-p^*}{1-p})]
\end{aligned}
$$

4g)

We set the derivative of our previous answer equal to zero and solve:

$$
\begin{aligned}
0 = \frac{\partial D_{KL}(F_{\mathbf{Y}|p^*}||F_{\mathbf{Y}|p})}{\partial p} \\
= n[\frac{-p^*}{p} + \frac{1-p^*}{1-p}]
\end{aligned}
$$

We see that when $p = p^*$, the derivative equals zero. But we also need to check the second derivative to make sure this is a local minimum.

$$
\begin{aligned}
0 &< \frac{\partial^2 D_{KL}(F_{\mathbf{Y}|p^*}||F_{\mathbf{Y}|p})}{\partial p^2} \\
0 &< n[\frac{p^*}{p^2} + \frac{1-p^*}{(1-p)^2}]
\end{aligned}
$$

The second derivative is positive for all p, so the function is convex and $p = p^*$ minimizes.

4h)

We have:

$$D_{KL}(F_{\mathrm{Y}_1|p^*}||F_{\mathrm{Y}_1|p}) = p^*\log(\tfrac{p^*}{p}) + (1-p^*)\log(\tfrac{1-p^*}{1-p})$$

This was part of our solution to part f. This works by linearity because we take expected value.