STA410 Assignment 1

Zikun Chen
1001117882

September 6, 2018

Question 1

(a)

$$
\begin{aligned}
LHS &= \sum_{i=1}^{k+1}(x_i - \bar{x}_{k+1})^2 \\
&= \sum_{i=1}^{k+1} x_i^2 - \bar{x}_{k+1}^2(k+1) \\
&= \sum_{i=1}^{k+1} x_i^2 - (k+1)(\frac{k\bar{x}_k + x_{k+1}}{k+1})^2 && \text{from the given recursive relationship} \\
&= \sum_{i=1}^{k+1} x_i^2 - \frac{k^2\bar{x}_k^2 + 2k\bar{x}_k x_{k+1} + x_{k+1}^2}{k+1} \\
&= \sum_{i=1}^{k} x_i^2 + x_{k+1}^2 - \frac{k^2\bar{x}_k^2 + 2k\bar{x}_k x_{k+1} + x_{k+1}^2}{k+1} \\
&= \sum_{i=1}^{k} x_i^2 + \frac{kx_{k+1}^2 + x_{k+1}^2 - k^2\bar{x}_k^2 - 2k\bar{x}_k x_{k+1} - x_{k+1}^2}{k+1} \\
&= \sum_{i=1}^{k} x_i^2 + \frac{k}{k+1}(x_{k+1}^2 - k^2\bar{x}_k^2 - 2\bar{x}_k x_{k+1}) \\
&= \sum_{i=1}^{k} x_i^2 + \frac{k}{k+1}(x_{k+1}^2 - k^2\bar{x}_k^2 - 2\bar{x}_k x_{k+1} + \bar{x}_k^2 - \bar{x}_k^2) \\
&= \sum_{i=1}^{k} x_i^2 - k\bar{x}_k + \frac{k}{k+1}(x_{k+1}^2 - 2\bar{x}_k x_{k+1} + \bar{x}_k^2) \\
&= \sum_{i=1}^{k}(x_i - \bar{x}_k)^2 + \frac{k}{k+1}(x_{k+1} - \bar{x}_k)^2 = RHS
\end{aligned}
$$

(b)

$$
\begin{aligned}
RHS &= \sum_{i=1}^{n}(x_i - x_0)^2 - n(x_0 - \bar{x})^2 \\
&= \sum_{i=1}^{n}(x_i^2 - 2x_ix_0 + x_0^2) - n(x_0^2 - 2x_0\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - 2x_0\sum_{i=1}^{n} x_i + nx_0^2 - nx_0^2 + 2nx_0\bar{x} - n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2nx_0\bar{x} + 2nx_0\bar{x} - n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})^2 = LHS
\end{aligned}
$$

(c) Since on the RHS of the part b) equation, both values $A = \sum_{i=1}^{n}(x_i - x_0)^2$ and $B = n(x_0 - \bar{x})^2$ can be fairly large, which would lead to catastrophic cancellation, we want to avoid this by choosing x_0 as close to $\bar{x}$ so that B will be relatively small compared to A.

We can choose x_0 to be the sample mean of a small sample from the data set, which is possibly large in size. We know that for many distributions, sample mean is a very good estimator or approximation of the population mean. Furthermore, if the data are assumed to be random and not ordered in any way, we can just take the first few observations and treat them as our sample for calculation.

No matter what the sample size is, it will be smaller than n, and adding n numbers together and taking the average requires O(n) floating point operations, which satisfies the requirement.

Question 2

(a)

$$\begin{aligned}
g(s) &= E(S^Y) \\
&= \sum_{i=1}^{\infty} s^i P(Y=i) \\
&= \sum_{i=1}^{\infty} s^i \sum_{j=1}^{\infty} P(Y=i|N=j)P(N=j) \\
&= \sum_{i=1}^{\infty} s^i P(Y=i|N=j) \sum_{j=1}^{\infty} P(N=j) \\
&= \sum_{i=1}^{\infty} s^i P(Y=i|N=j) \sum_{j=1}^{\infty} P(N=j) \\
&= \sum_{i=1}^{\infty} s^i P(\sum_{k=1}^{j} X_k = i) \sum_{j=1}^{\infty} P(N=j) \\
&= \phi_{S_j}(s) \sum_{j=1}^{\infty} P(N=j) && \text{where } S_j = X_1 + \cdots + X_j \\
&= [\phi_X(s)]^j \sum_{j=1}^{\infty} P(N=j) && \text{since } X\text{'s are i.i.d. RVs} \\
&= \sum_{j=1}^{\infty} P(N=j)[\phi_X(s)]^j \\
&= \phi_Y(\phi_X(s)) = \exp(-\lambda(1-\phi_X(s)))
\end{aligned}$$

(b) We want to show that $P(Y \geq ml) < P(N \geq m)$

Let's first show $P(Y < ml) \geq P(N < m)$

If $N < m$, i.e. the number of X's is smaller than m, and X can only take integer values from 0 to l, the sum of X's, which is Y, cannot reach or exceed ml, $P(Y < ml)$.

So event A: $N < m$ implies event B: $Y < ml$

$A \subseteq B \implies P(A) \leq P(B)$

$1 - P(N < m) < 1 - P(Y < ml)$

Therefore, $P(Y \geq ml) < P(N \geq m) \leq \epsilon$

(c) We want to choose M such that:

$$P(Y \geq M) \leq \frac{exp(-\lambda(1-\phi(s)))}{s^M} = \epsilon$$

Taking ln on each side:

$$
\begin{aligned}
-\lambda(1-\phi(s)) - Mln(s) &= ln(\epsilon) \\
Mln(s) &= -ln(\epsilon) - \lambda(1-\phi(s)) \\
M &= \frac{-ln(\epsilon) - \lambda(1-\phi(s))}{ln(s)}
\end{aligned}
$$

To simplify the calculations, we want to take as few points as possible for FFT. Therefore, we want to minimize M.

Take:

$$M = \inf_{s>1} \frac{-ln(\epsilon) - \lambda(1-\phi(s))}{ln(s)}$$

(d) R code:

```
lambda = 7
epsilon = 10E-5

# PDF for X
p <- function(x) {
  p <- 0
  if (x >= 0 & x <=6){
    p <- (4-abs(3-x))/16
  }
  return(p)}

# PGF for X
phi <- function(s) {
  phi = 0
  for (x in 0:6) {
    phi <- phi + p(x) * s^x
  }
  return(phi)}

# function for M
M <- function(s) {
  (-log(epsilon) - lambda*(1-phi(s)))/log(s)
}

# fit a curve for M and look for the minimum, which turns out to be around 71.5
curve(M, from=1.3, to=1.5, xname="s", ylab="M", main = "Plot for Bound M")
M = round(optimize(M, interval=c(1, 2))$objective)

# Fast Fourier Transformation with Bound M
x=c(0:M-1)
```

```
px <- sapply(x, p)
pjhat <- fft(px)
gj <- exp(-lambda * (1-pjhat))
py <- Re(fft(gj, inv=T)/M)

# Plot the Distribution of Y
plot(x, py, main = "Probability Distribution of Y", xlab = "y", ylab = "P(Y = y)")
```

M turns out to take minimum integer value around 72.

Graph for M between when $s_k = 1.5$:

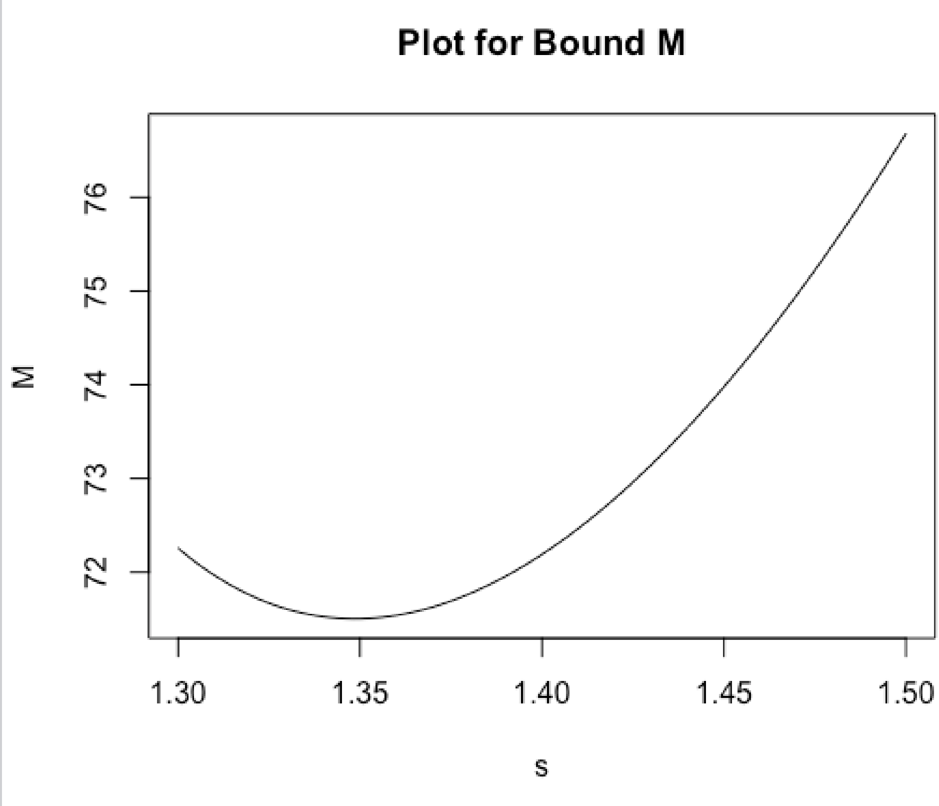
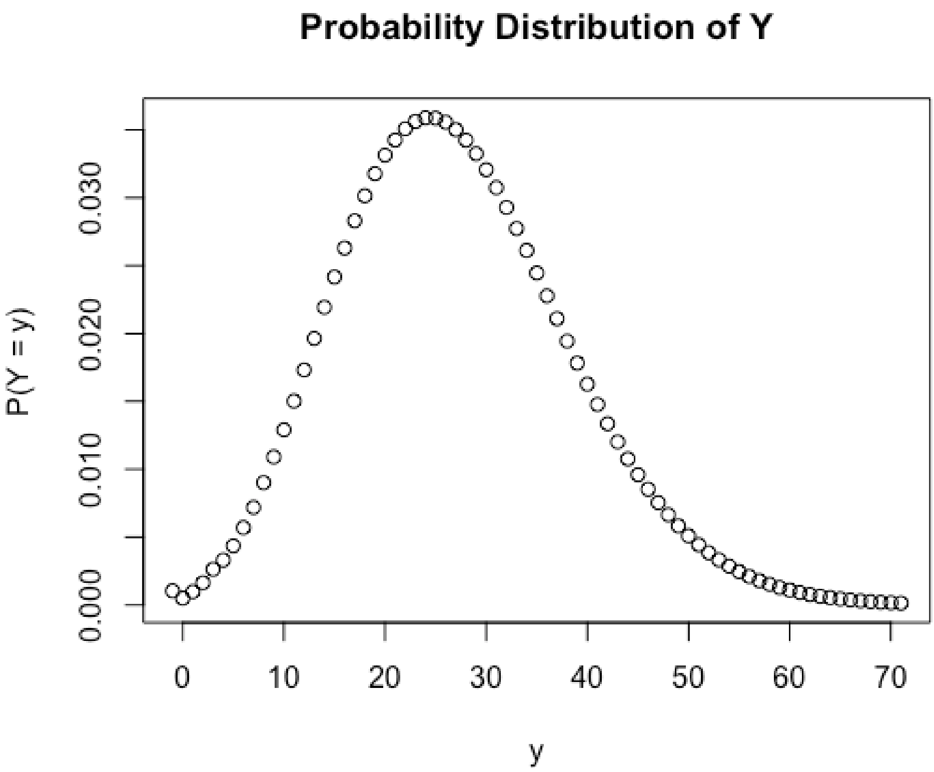


Figure 1: s is in between 1.3 and 1.5, curve fitted graph

Distribution of Y:



Probability Distribution of Y
P(Y = y)
0.000
0.010
0.020
0.030
0
10
20
30
40
50
60
70
y