STA410 Assignment 4

Zikun Chen
1001117882

December 5, 2017

Question 1

(a)

$$RHS = \theta_1 + \sum_{i=2}^{k} \phi_i = \theta_1 + \sum_{i=2}^{k} (\theta_i - \theta_{i-1})$$

$$= \theta_1 + \sum_{i=2}^{k} \theta_i - \sum_{i=2}^{k} \theta_{i-1} = \theta_1 + \sum_{i=2}^{k} \theta_i - \sum_{j=1}^{k-1} \theta_j$$

$$= \theta_1 + (\sum_{i=2}^{k-1} \theta_i + \theta_k) - (\theta_1 + \sum_{j=2}^{k-1} \theta_j) = \theta_k$$

(b) Substitute the equality from part a) into the objective function:

$$f(\theta_1, \phi_1, \cdots, \phi_n) = (y_1 - \theta_1)^2 + \sum_{i=2}^{n} [y_i - (\theta_1 + \sum_{j=2}^{i} \phi_j)]^2 + \lambda \sum_{i=1}^{n} |\phi_i|$$

Take partial derivative with respect to θ_1 and set it to 0:

$$\frac{\partial}{\partial \theta_1} = -2(y_1 - \theta_1) + \sum_{i=2}^{n} -2[y_i - \theta_1 - \sum_{j=2}^{i} \phi_j] = 0$$

$$\theta_1 = y_1 + \sum_{i=2}^{n} (y_i - \theta_1 - \sum_{j=2}^{i} \phi_j)$$

$$\theta_1 = y_1 + \sum_{i=2}^{n} (y_i - \sum_{j=2}^{i} \phi_j) - \sum_{i=2}^{n} \theta_1$$

$$\theta_1 = y_1 + \sum_{i=2}^{n} (y_i - \sum_{j=2}^{i} \phi_j) - (n-1)\theta_1$$

$$n\theta_1 = y_1 + \sum_{i=2}^{n}(y_i - \sum_{j=2}^{i}\phi_j)$$

$$\theta_1 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \sum_{j=2}^{i}\phi_j)$$

(c) Objective function has a differentiable part g_d and a non-differentiable part g_{nd}

$$f(\theta_1, \phi_1, \cdots, \phi_n) = g_d(\theta_1, \phi_1, \cdots, \phi_n) + g_{nd}(\phi_1, \cdots, \phi_n)$$
$$= (y_1 - \theta_1)^2 + \sum_{i=2}^{n}[y_i - (\theta_1 + \sum_{j=2}^{i}\phi_j)]^2 + \lambda\sum_{i=2}^{n}|\phi_i|$$

Therefore, the sub-gradient with respect to ϕ_j is:

$$\frac{\partial g_d}{\partial \phi_j} + \lambda\partial|\phi_j|$$

$$\partial|\phi_j| = \{v : |y| \geq |\phi_j| + v(y - \phi_j)\forall y\} = \begin{cases} 1 & \text{if } \phi_j > 0 \\ [-1, 1] & \text{if } \phi_j = 0 \\ -1 & \text{if } \phi_j < 0 \end{cases}$$

Since the inner index k is from 2 to i, for a particular j, ϕ_j only exists in the terms where $i \geq j$, so i starts from j when we take the partial derivative:

$$\frac{\partial g_d}{\partial \phi_j} = \sum_{i=j}^{n} -2(y_i - \theta_i - \sum_{k=2}^{i}\phi_k)$$
$$= -2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2}^{i}\phi_k)$$
$$= -2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k - \phi_j)$$
$$= -2[\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) - \sum_{i=j}^{n}\phi_j]$$
$$= 2(n - j + 1)\phi_j - 2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k)$$

So the sub-gradient with respect to ϕ_j is:

$$2(n - j + 1)\phi_j - 2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) + \lambda\partial|\phi_j|$$

If $\phi_j = 0$, set the sub-gradient to 0 gives:

$$2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) = \lambda\partial|\phi_j|$$

And we know, when $\phi_j = 0$:

$$-1 \leq \partial|\phi_j| \leq 1$$

$$\Rightarrow -\lambda \leq \lambda\partial|\phi_j| = 2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) \leq \lambda$$

$$\Rightarrow |\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k)| \leq \frac{\lambda}{2}$$

If $\phi_j > 0$, $\partial|\phi_j| = 1$, set the sub-gradient to 0 gives:

$$2(n-j+1)\phi_j - 2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) + \lambda = 0$$

$$\Rightarrow (n-j+1)\phi_j = \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) - \frac{\lambda}{2}$$

$$\Rightarrow \phi_j = \frac{1}{n-j+1}\{\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) - \frac{\lambda}{2}\} > 0$$

$$\Rightarrow \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) - \frac{\lambda}{2} > 0 \Rightarrow \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) > \frac{\lambda}{2}$$

If $\phi_j < 0$, $\partial|\phi_j| = -1$, set the sub-gradient to 0 gives:

$$2(n-j+1)\phi_j - 2\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) - \lambda = 0$$

$$\Rightarrow (n-j+1)\phi_j = \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) + \frac{\lambda}{2}$$

$$\Rightarrow \phi_j = \frac{1}{n-j+1}\{\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i}\phi_k) + \frac{\lambda}{2}\} < 0$$

$$\Rightarrow \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i} \phi_k) + \frac{\lambda}{2} < 0 \Rightarrow \sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i} \phi_k) < -\frac{\lambda}{2}$$

Therefore, the objective funciton is minimized over ϕ_j for fixed values of θ_1 and $\{\phi_j : 2 \leq i \neq j \leq n\}$ at

$$\phi_j = 0 \text{ if } |\sum_{i=j}^{n}(y_i - \theta_i - \sum_{k=2;k\neq j}^{i} \phi_k)| \leq \frac{\lambda}{2}$$

$$\phi_j = \frac{1}{n-j+1}\{\sum_{i=j}^{n}(y_i-\theta_i-\sum_{k=2;k\neq j}^{i} \phi_k)-\frac{\lambda}{2}\} \text{ if } \sum_{i=j}^{n}(y_i-\theta_i-\sum_{k=2;k\neq j}^{i} \phi_k) > \frac{\lambda}{2}$$

$$\phi_j = \frac{1}{n-j+1}\{\sum_{i=j}^{n}(y_i-\theta_i-\sum_{k=2;k\neq j}^{i} \phi_k)+\frac{\lambda}{2}\} \text{ if } \sum_{i=j}^{n}(y_i-\theta_i-\sum_{k=2;k\neq j}^{i} \phi_k) < -\frac{\lambda}{2}$$

(d) **Program:**

```
coor.des <- function(y,lambda,theta1,phis,eps=1.e-6) {
  n <- length(y)
  lambda2 <- lambda/2
  phis <- c(NA, phis)
  # initial objective funciton value
  term2 <- 0
  for (i in 2:n) {
    term2 <- term2 + (y[i]-theta1-sum(phis[2:i]))^2
  }
  new.obj <- (y[1]-theta1)^2 + term2 + lambda*sum(abs(phis[2:n]))
  no.conv <- T
  iter <- 0
  while (no.conv) {
    obj <- new.obj
    # update theta1
    theta1 <- y[1]
    for (i in 2:n) {
      theta1 <- theta1 + y[i]-sum(phis[2:i])
    }
    theta1 <- theta1/n
    # update phi's
    term2 <- 0
    for (j in 2:n) {
      sumj <- 0
      for (i in j:n) {
        sumj <- sumj + y[i]-theta1-sum(phis[2:i][-j])
      }
```

```
      if (abs(sumj)<=lambda2) {
        phis[j] <- 0
      }
      else if (abs(sumj)>lambda2){
        phis[j] <- (sumj-lambda2)/(n-j+1)
      }
      else {
        phis[j] <- (sumj+lambda2)/(n-j+1)
      }
      term2 <- term2 + (y[j]-theta1-sum(phis[2:j]))^2
    }
    iter <- iter + 1
    # compute the new objective function value
    new.obj <- (y[1]-theta1)^2 + term2 + lambda*sum(abs(phis[2:n]))
    if (abs(obj-new.obj)<eps) no.conv <- F
  }
  r <- list(theta1=theta1, phis=phis, iter=iter)
  r
}

lambda <- 0
y <- c(rep(0,250),rep(1,250),rep(0,50),rep(1,450)) + rnorm(1000,0,0.1)
# using the result of the seidal function from A2 for initial estimates
thetas <- seidel(y, lambda)$theta
theta1 <-thetas[1]
n <- length(thetas)
phis <- thetas[2:n] - thetas[1:n-1]

coor.des(y,lambda,theta1,phis)
```

Question 2

(a)

$$lnL = \sum_{i=1}^{n}\sum_{k=1}^{m}[u_{ik}ln(f_k(x_i)) + u_{ik}ln(\lambda_k)]$$

$$ln(f_k(x_i)) = ln(\frac{1}{\sqrt{2\pi\sigma_k^2}}e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}) = -\frac{(x_i-\mu_k)^2}{2\sigma_k^2} - ln(\sqrt{2\pi\sigma_k^2})$$

$$\frac{\partial ln(f_k(x_i))}{\partial\mu_k} = \frac{(x_i-\mu_k)}{\sigma_k^2}$$

$$\frac{\partial ln(f_k(x_i))}{\partial\sigma_k^2} = \frac{(x_i-\mu_k)^2}{2(\sigma_k^2)^2} - \frac{1}{\sqrt{2\pi\sigma_k^2}}\frac{\sqrt{2\pi}}{2\sqrt{\sigma_k^2}} = \frac{(x_i-\mu_k)^2}{2(\sigma_k^2)^2} - \frac{1}{2\sigma_k^2}$$

$$\frac{\partial lnL}{\partial \mu_k} = \sum_{i=1}^{n} u_{ik} \frac{(x_i - \hat{\mu_k})}{\sigma_k^2} = 0 \Rightarrow \sum_{i=1}^{n} u_{ik} x_i - \hat{\mu_k} \sum_{i=1}^{n} u_{ik} = 0$$

$$\hat{\mu_k} = (\sum_{i=1}^{n} u_{ik})^{-1} \sum_{i=1}^{n} u_{ik} x_i$$

$$\frac{\partial lnL}{\partial \sigma_k^2} = \sum_{i=1}^{n} u_{ik} (\frac{(x_i - \hat{\mu_k})^2}{2(\hat{\sigma_k^2})^2} - \frac{1}{2\hat{\sigma_k^2}}) = 0 \Rightarrow \sum_{i=1}^{n} u_{ik}(x_i - \hat{\mu_k})^2 - \hat{\sigma_k^2} \sum_{i=1}^{n} u_{ik} = 0$$

$$\hat{\sigma_k^2} = (\sum_{i=1}^{n} u_{ik})^{-1} \sum_{i=1}^{n} u_{ik}(x_i - \hat{\mu_k})^2$$

For $\hat{\lambda_k}$, we use Lagrange Multiplier α since $\sum_{k=1}^{m} \lambda_k = 1$:

$$lnL = \sum_{i=1}^{n} \sum_{k=1}^{m} [u_{ik} ln(f_k(x_i)) + u_{ik} ln(\lambda_k)] + \alpha(1 - \sum_{k=1}^{m} \lambda_k)$$

$$\frac{\partial lnL}{\partial \lambda_k} = \frac{1}{\hat{\lambda_k}} \sum_{i=1}^{n} u_{ik} - \alpha = 0 \Rightarrow \hat{\lambda_k} = \frac{1}{\alpha} \sum_{i=1}^{n} u_{ik}$$

$$\frac{\partial lnL}{\partial \alpha} = 0 \Rightarrow \sum_{k=1}^{m} \hat{\lambda_k} = 1$$

$$\Rightarrow \sum_{k=1}^{m} (\frac{1}{\alpha} \sum_{i=1}^{n} u_{ik}) = 1 \Rightarrow \frac{1}{\alpha} \sum_{i=1}^{n} \sum_{k=1}^{m} u_{ik} = \frac{1}{\alpha} \sum_{i=1}^{n} 1 = \frac{n}{\alpha} = 1$$

Therefore, $\alpha = n$ and $\hat{\lambda_k} = \frac{1}{n} \sum_{i=1}^{n} u_{ik}$

(b), (c) **Program:**

```
normalmixture <- function(x,k,mu,sigma,lambda,eps=1e-6,max.iter=500) {
  n <- length(x)
  x <- sort(x)
  vars <- sigma^2
  means <- mu
  lam <- lambda/sum(lambda)  # guarantee that lambdas sum to 1
  delta <- matrix(rep(0,n*k),ncol=k)
  # initial deltas
  for (i in 1:n) {
    xi <- x[i]
    for (j in 1:k) {
      mj <- means[j]
      varj <- vars[j]
      denom <- 0
      for (u in 1:k) {
```

```
      mu <- means[u]
      varu <- vars[u]
      denom <- denom + lam[u]*dnorm(xi,mu,sqrt(varu))
    }
    delta[i,j] <- lam[j]*dnorm(xi,mj,sqrt(varj))/denom
  }
}
# initial log likelihood value
new.loglik <- 0
s <- rep(0, n)
for (j in 1:k) {
  s <- s + lam[j]*dnorm(x, means[j], sqrt(vars[j]))
}
new.loglik <- sum(log(s))

iter <- 1
no.conv <- T
while (no.conv && iter <= max.iter) {
  loglik <- new.loglik
  # compute updates of deltas
  for (i in 1:n) {
    xi <- x[i]
    for (j in 1:k) {
      mj <- means[j]
      varj <- vars[j]
      denom <- 0
      for (u in 1:k) {
        mu <- means[u]
        varu <- vars[u]
        denom <- denom + lam[u]*dnorm(xi,mu,sqrt(varu))
      }
      delta[i,j] <- lam[j]*dnorm(xi,mj,sqrt(varj))/denom
    }
  }
  # compute updated estimates of means, variances, and probabilities
  for (j in 1:k) {
    deltaj <- as.vector(delta[,j])
    sum_dj <- sum(deltaj)
    lambda[j] <- sum_dj/n
    means[j] <- sum(x*deltaj)/sum_dj
    vars[j] <- sum((x-means[j])^2*deltaj)/sum_dj
  }
  lam <- lambda/sum(lambda)
  iter <- iter + 1

  # compute log likelihood of the normal mixture
```

```
    new.loglik <- 0
    s <- rep(0, n)
    for (j in 1:k) {
      s <- s + lam[j]*dnorm(x, means[j], sqrt(vars[j]))
    }
    new.loglik <- sum(log(s))
    if (abs(new.loglik-loglik)<eps) no.conv <- F
  }
  r <- list(mu=means,var=vars,delta=delta,lambda=lam,loglik=new.loglik,iter=iter)
  r
}

# getting data from txt file
setwd("/Users/zikunchen/Desktop")
stamp <- read.table("stamp.txt", fill = TRUE)
stamp <- unname(unlist(stamp))
stamp <- stamp[!is.na(stamp)]

# 5 modes
k <- 5
# initial estimate based on graph
plot(density(stamp,bw=0.0026))
lambda <- rep(1/k, k)
mu <- c(0.079, 0.09, 0.1, 0.11, 0.12)
sigma <- c(0.0026, 0.0026, 0.0026, 0.0026, 0.0026)
r5 <- normalmixture(stamp,k,mu,sigma,lambda)

# 6 modes
k <- 6
# initial estimate based on graph
plot(density(stamp,bw=0.0024))
lambda <- rep(1/k, k)
mu <- c(0.079, 0.09, 0.1, 0.11, 0.12, 0.13)
sigma <- c(0.0024, 0.0024, 0.0024, 0.0024, 0.0024, 0.0024)
r6 <- normalmixture(stamp,k,mu,sigma,lambda)

# 7 modes
k <- 7
# initial estimate based on graph
plot(density(stamp,bw=0.0015))
lambda <- rep(1/k, k)
mu <- c(0.071, 0.08, 0.09, 0.1, 0.11, 0.12, 0.124)
sigma <- c(0.0015, 0.0015, 0.0015, 0.0015, 0.0015, 0.0015, 0.0015)
r7 <- normalmixture(stamp,k,mu,sigma,lambda)
```

(d) I prefer the 7-mode model.

Maximized log-likelihood for mode 5, 6, and 7 is 1503.211, 1507.341 and 1531.271 respectively.

Let's first compare the 5-mode and 6-mode models using likelihood ratio test.

If the 5-mode model is true, then then twice the difference in the maximized log likelihoods has approximately a χ^2 distribution with 3 degrees of freedom (since the 6-mode model has two additional parameters compared to the 5-mode model).

The likelihood ratio test statistic then is -2(1503.211 - 1507.341) = 8.26 and the p-value is approximately $P(\chi^2 > 8.26) = 0.041$; this means that there is significant evidence that the 6-mode mixture model is superior.

Now let's look at 6-mode vs. 7-mode.

Assume that the 6-mode model is better, the distribution of the likelihood is still chi-square with degree of freedom 3.

The likelihood ratio test statistic then is -2(1507.341 - 1531.271) = 47.86

$P(\chi^2 > 47.86) < 0.00001$ means that there is overwhelming evidence that the 7-mode model is better than 6-mode model.

Therefore, 7-mode is the best model according to likelihood ratio test.