Project 3

Benjamin Draves

Exercise 1

(a)

Suppose that we have the following hierachial model

$$X_i | Z_i \overset{iid}{\sim} \text{Po}(Z_i \lambda_d + (1 - Z_i)\lambda_c)$$

$$Z_i \overset{iid}{\sim} \text{Bern}(\pi)$$

and we look to use the EM-algorithm to estimate the parameter $\theta = (\lambda_c, \lambda_d, \pi)$. We begin by deriving the E-step by first finding the full log-likelihood then the expected log-likelihood Q.

$$\mathcal{L}(\mathbf{X}, \mathbf{Z}) = \prod_{i=1}^{n} \mathbb{P}(X_i, Z_i) = \prod_{i=1}^{n} \mathbb{P}(X_i | Z_i)\mathbb{P}(Z_i) = \prod_{i=1}^{n} p(X_i; \lambda_c)^{Z_i} p(X_i; \lambda_d)^{1-Z_i} \pi^{Z_i}(1 - \pi)^{1-Z_i}$$

$$\updownarrow(\mathbf{X}, \mathbf{Z}) = \sum_{i=1}^{n} \left\{ Z_i \log(p(X_i; \lambda_c)) + (1 - Z_i)\log(p(X_i; \lambda_d)) + Z_i \log(\pi) + (1 - Z_i)\log(1 - \pi) \right\}$$

$$= \sum_{i=1}^{n} \left\{ Z_i \log(\pi p(X_i; \lambda_c)) + (1 - Z_i)\log((1 - \pi)p(X_i; \lambda_d)) \right\}$$

Using this form of the log-likelihood, we can find the conditional expectation $Z | X; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)}$. Before we do this however, we begin by defining $\alpha_i^{(t)} = \mathbb{E}_{Z|X;\pi^{(t)},\lambda_d^{(t)},\lambda_c^{(t)}}[Z_i] = \mathbb{P}(Z_i = 1 | X_i; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})$.

$$\alpha_i^{(t)} = \mathbb{P}(Z_i = 1 | X_i; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})$$

$$= \frac{\mathbb{P}(Z_i = 1, X_i; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})}{\mathbb{P}(X_i)}$$

$$= \frac{\mathbb{P}(X_i | Z_i = 1; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})\mathbb{P}(Z_i = 1 | \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})}{\mathbb{P}(X_i | Z_i = 1; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})\mathbb{P}(Z_i = 1 | \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)}) + \mathbb{P}(X_i | Z_i = 0; \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})\mathbb{P}(Z_i = 0 | \pi^{(t)}, \lambda_d^{(t)}, \lambda_c^{(t)})}$$

$$= \frac{\pi^{(t)} p(X_i; \lambda_d^{(t)})}{\pi^{(t)} p(X_i; \lambda_d^{(t)}) + (1 - \pi^{(t)})p(X_i; \lambda_c^{(t)})}$$

Using this, we can find our Q as follows

$$Q = \sum_{i=1}^{n} \left\{ \alpha_i^{(t)} \log(\pi p(X_i; \lambda_c)) + (1 - \alpha_i^{(t)})\log((1 - \pi)p(X_i; \lambda_d)) \right\}$$

(b)

Using the Q that we developed in the previous porition of the problem, we can determine our maximization step by differentiating Q. We do so for each parameter to develop the three updating steps. First we note that $\frac{d}{d\lambda}\log(p(X,\lambda)) = \frac{X-\lambda}{\lambda}$

$$\frac{d}{d\lambda_d}Q = \sum_{i=1}^{n}(1-\alpha_i^{(t)})\frac{d}{d\lambda_d}\log(p(X_i;\lambda_d)) = \sum_{i=1}^{n}(1-\alpha_i^{(t)})\left(\frac{X_i-\lambda_d}{\lambda_d}\right)$$

Now setting equal to 0 and solving for λ_d we get our update $\lambda_d^{(t+1)}$.

$$\sum_{i=1}^{n}(1-\alpha_i^{(t)})\frac{X_i}{\lambda_d^{(t+1)}} - \sum_{i=1}^{n}(1-\alpha_i^{(t)}) = 0$$

$$\frac{1}{\lambda_d^{(t+1)}}\sum_{i=1}^{n}(1-\alpha_i^{(t)})X_i = \sum_{i=1}^{n}(1-\alpha_i^{(t)})$$

$$\lambda_d^{(t+1)} = \frac{\sum_{i=1}^{n}(1-\alpha_i^{(t)})X_i}{\sum_{i=1}^{n}(1-\alpha_i^{(t)})}$$

Following an idential procedure for λ_c we have the following

$$\frac{d}{d\lambda_c}Q = \sum_{i=1}^{n}\alpha_i^{(t)}\frac{d}{d\lambda_c}\log(p(X_i;\lambda_c)) = \sum_{i=1}^{n}\alpha_i^{(t)}\left(\frac{X_i-\lambda_c}{\lambda_c}\right)$$

Now setting equal to 0 and solving for λ_c we get our update $\lambda_c^{(t+1)}$.

$$\sum_{i=1}^{n}\alpha_i^{(t)}\frac{X_i}{\lambda_c^{(t+1)}} - \sum_{i=1}^{n}\alpha_i^{(t)} = 0$$

$$\frac{1}{\lambda_c^{(t+1)}}\sum_{i=1}^{n}\alpha_i^{(t)}X_i = \sum_{i=1}^{n}\alpha_i^{(t)}$$

$$\lambda_c^{(t+1)} = \frac{\sum_{i=1}^{n}\alpha_i^{(t)}X_i}{\sum_{i=1}^{n}\alpha_i^{(t)}}$$

Lastly, we have our update for $\pi^{(t+1)}$

$$\frac{d}{d\pi}Q = \sum_{i=1}^{n}\alpha_i^{(t)}\frac{d}{d\pi}\log(\pi) + (1-\alpha_i^{(t)})\frac{d}{d\pi}\log(1-\pi)$$

$$= \sum_{i=1}^{n}\alpha_i^{(t)}\frac{1}{\pi} - (1-\alpha_i^{(t)})\frac{1}{1-\pi}$$

Now setting equal to 0 and solving for π we achieve our update for $\pi^{(t+1)}$.

$$\frac{1}{\pi^{(t+1)}} \sum_{i=1}^{n} \alpha_i^{(t)} = \frac{1}{1 - \pi^{(t+1)}} \sum_{i=1}^{n} (1 - \alpha_i^{(t)})$$

$$(1 - \pi^{(t+1)}) \sum_{i=1}^{n} \alpha_i^{(t)} = \pi^{(t+1)} \sum_{i=1}^{n} (1 - \alpha_i^{(t)})$$

$$\sum_{i=1}^{n} \alpha_i^{(t)} - \pi^{(t+1)} \sum_{i=1}^{n} \alpha_i^{(t)} = n\pi^{(t+1)} - \pi^{(t+1)} \sum_{i=1}^{n} \alpha_i^{(t)}$$

$$\sum_{i=1}^{n} \alpha_i^{(t)} = n\pi^{(t+1)}$$

$$\pi^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \alpha_i^{(t)}$$

(c)

```r
# read in data
i <- 1:20
X <- c(2, 0, 0, 1, 3, 0, 1, 6, 2, 0, 1, 0, 2, 0, 8, 0, 1, 3,
    2, 0)

# initialize inputs
pi0 <- 0.5
lam_c0 <- sum(X * (X < mean(X)))/sum(X < mean(X))
lam_d0 <- sum(X * (X > mean(X)))/sum(X > mean(X))

# define EM function
em <- function(pi, lam_c, lam_d) {

    # set up stopping criterion
    precision <- 1e-07
    diff <- Inf
    Q <- Inf

    while (diff > precision) {

        # update estimates
        alpha <- (pi * dpois(X, lam_d, log = FALSE))/(pi * dpois(X,
            lam_d, log = FALSE) + (1 - pi) * dpois(X, lam_c,
            log = FALSE))
        lam_c <- sum((1 - alpha) * X)/sum(1 - alpha)
        lam_d <- sum(alpha * X)/sum(alpha)
        pi <- mean(alpha)

        # update Q and stopping condition
        Q_new <- sum(alpha * (dpois(X, lam_d, log = TRUE) + log(pi)) +
            (1 - alpha) * (dpois(X, lam_c, log = TRUE) + log(1 -
                pi)))
        diff <- abs(Q_new - Q)
```

```
        Q <- Q_new
    }
    # return estimates
    return(list(pi = pi, lambdac = lam_c, lambdad = lam_d))

}
res <- em(pi0, lam_c0, lam_d0)
res
```

```
## $pi
## [1] 0.1244548
##
## $lambdac
## [1] 0.9581722
##
## $lambdad
## [1] 6.115288
```

(d)

Using Bayes Rule, we can calculate the following posterior probabilities.

$$\mathbb{P}(Z_i = 1 | X_i = x) = \frac{\hat{\pi}\text{Po}(X_i; \hat{\lambda}_d)}{\hat{\pi}\text{Po}(X_i; \hat{\lambda}_d) + (1 - \hat{\pi})\text{Po}(X_i; \hat{\lambda}_c)}$$

Using this, we can calculate the probability that the intersection is a black spot given the number of observed accidents as follows.

```
pi_hat <- res[[1]]
lam_c_hat <- res[[2]]
lam_d_hat <- res[[3]]

prob_danger <- function(x) (dpois(x, lam_d_hat) * pi_hat)/((dpois(x,
    lam_d_hat) * pi_hat) + (dpois(x, lam_c_hat) * (1 - pi_hat)))

c(prob_danger(X[1]), prob_danger(X[5]))
```
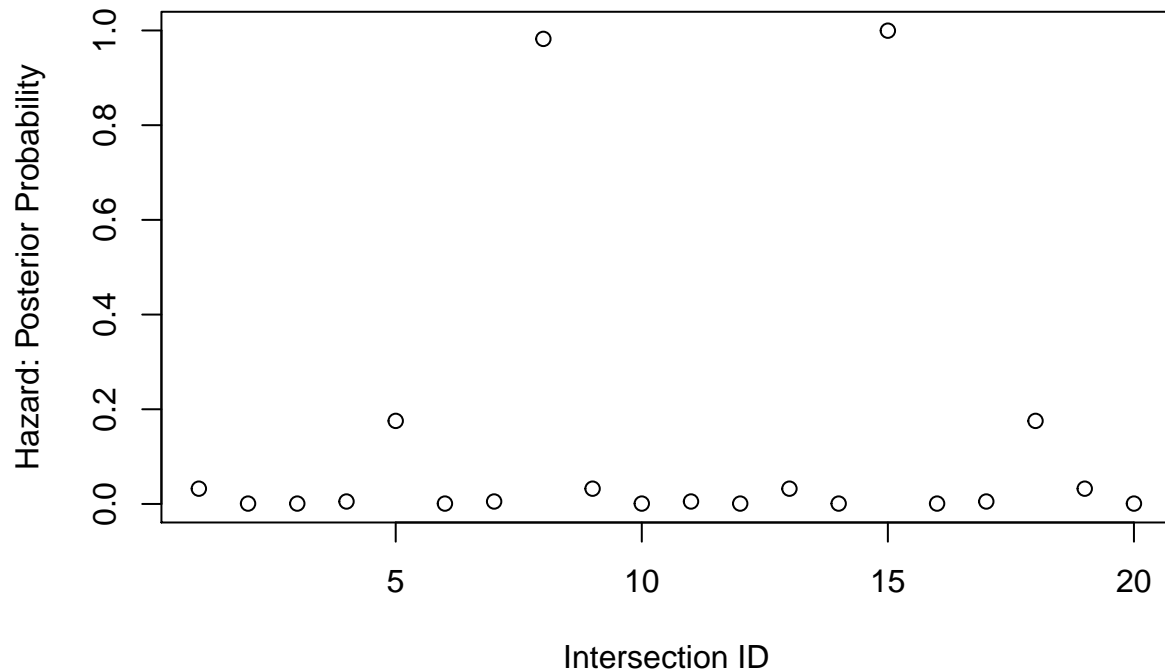
```
## [1] 0.03226482 0.17545325
```

```
hazard <- as.numeric(prob_danger(X))
plot(hazard, ylab = "Hazard: Posterior Probability", xlab = "Intersection ID")
```

Here we see that the first intersection has posterior probability of 0.0322648236 of being dangerous and the fifth has probability 0.17545325 The plot here shows each posterior probability. Clearly, we flag intersections 8 and 15 as possible black spots.

(e)

```
res <- em(pi0, lam_d0, lam_c0)
pi_hat <- res[[1]]
lam_c_hat <- res[[2]]
lam_d_hat <- res[[3]]

hazard <- as.numeric(prob_danger(X))
plot(hazard, ylab = "Hazard: Posterior Probability", xlab = "Intersection ID")
```
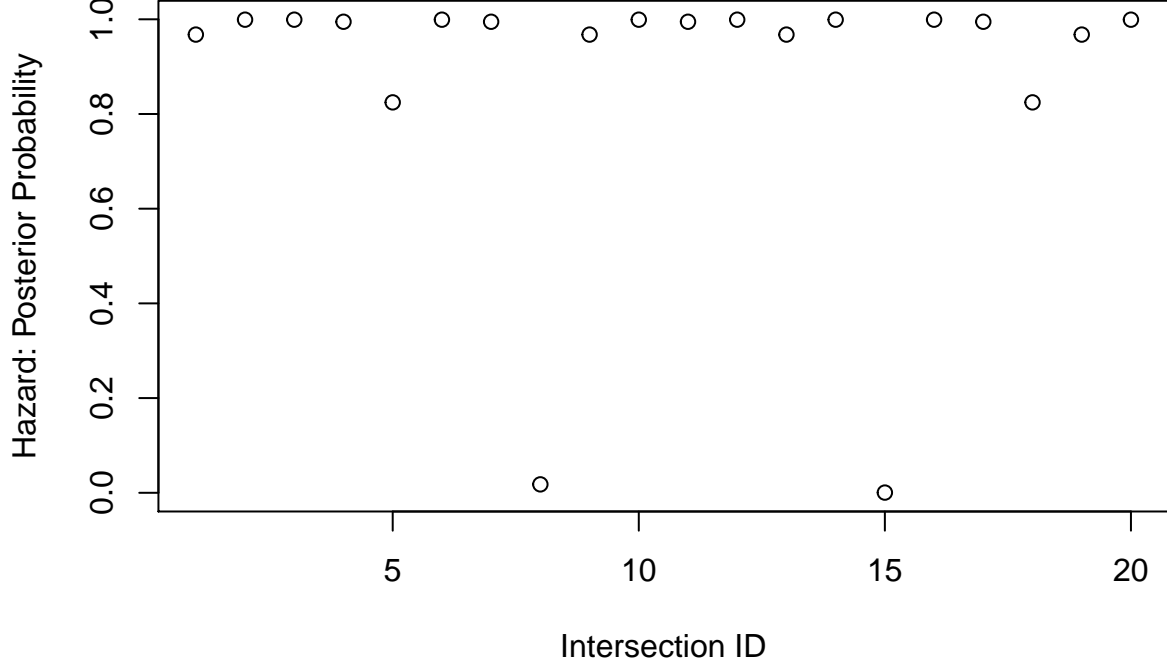
Here we note that the EM algorithm doesn't identify which event is a success in the latent bernoulli random variable. In the previous instance it labeld a success as a danger intersection where here it labeled the safe intersections as a success. Extending this reparameterization, with $X_i | Z_i \sim \text{Po}(Z_i \lambda_d + (1 - Z_i) \lambda_c)$, we see that if $Z_i = 1$ if the intersection is safe, then λ_c is the rate at which accidents occur at a dangerous intersection. Therefore, we see that it is reasonable to expect that $\lambda_c = 6.115288$, the previous rate of accidents at a dangerous intersection.

Exercise 2

(a)

Suppose that we have the following hierachial model

$$Z \sim DU(1, n - 1)$$
$$X_j \sim N(\mu_1 I(j \leq Z) + \mu_2 I(j > Z), \sigma^2)$$

here we look to use the EM algorithm to find estimate of $\theta = (\mu_1, \mu_2, \sigma^2)$. We begin by deriving the E-step by finding the full log-likelihood then the expected (conditional) log-likelihood Q.

$$\mathcal{L}(\mathbf{X}, \mathbf{Z}) = \prod_{j=1}^{n} \mathbb{P}(X_j | Z) \mathbb{P}(Z) \propto \prod_{j=1}^{n} \phi(\mu_1, \sigma^2)^{I(j \leq Z)} \phi(\mu_2, \sigma^2)^{I(j > Z)}$$

$$\updownarrow(\mathbf{X}, \mathbf{Z}) \propto \sum_{j=1}^{n} I(j \leq Z) \left(-\frac{1}{2} \log(\sigma^2) - \frac{(X_j - \mu_1)^2}{2\sigma^2} \right) + I(j > Z) \left(-\frac{1}{2} \log(\sigma^2) - \frac{(X_j - \mu_2)^2}{2\sigma^2} \right)$$

$$= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} \left(I(j \leq Z)(X_j - \mu_1)^2 + I(j > Z)(X_j - \mu_2)^2 \right)$$

Now, we define the Q by taking the conditional expecation $\mathbb{E}_{Z|X,\theta}$

$$Q(\theta;\theta^{(t)}) = -\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(\mathbb{E}_{Z|X,\theta^{(t)}}[I(j \leq Z)](X_j - \mu_1)^2 + \mathbb{E}_{Z|X,\theta^{(t)}}[I(j > Z)](X_j - \mu_2)^2\right)$$

We can also express this quantity in vector form as follows

$$Q(\theta;\theta^{(t)}) = -\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(\mathbb{E}_{Z|X,\theta^{(t)}}[I(j \leq Z)](X_j - \mu_1)^2 + \mathbb{E}_{Z|X,\theta^{(t)}}[I(j > Z)](X_j - \mu_2)^2\right)$$

$$= -\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n-1}\sum_{j=1}^{n}\left(\mathbb{E}_{Z|X,\theta^{(t)}}[I(j \leq Z)](X_j - \mu_1)^2\right.$$

$$\left. + \mathbb{E}_{Z|X,\theta^{(t)}}[I(j > Z)](X_j - \mu_2)^2\right)\mathbb{P}(Z = i)$$

$$= -\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n-1}(X - \mu_{(i)})^T(X - \mu_{(i)})\mathbb{P}(Z = i)$$

$$= -\frac{n}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n-1}(X - \mu_{(i)})^T(X - \mu_{(i)})\mathbb{E}_{Z|X,\theta^{(t)}}(I(Z = i))$$

(b)

Now, we turn to the M - step by first defining the followign quantity.

$$\pi_i^{(t)} = \mathbb{E}_{Z|X;\theta^{(t)}}(I(Z = i)) = \mathbb{P}(Z = i|X;\theta^{(i)}) = \frac{\mathbb{P}(Z = i, X)}{\mathbb{P}(X)}$$

$$= \frac{\mathbb{P}(X|Z = i)\mathbb{P}(Z = i)}{\sum_{i=k}^{n-1}\mathbb{P}(X|Z = k)\mathbb{P}(Z = k)}$$

$$= \frac{1/(n-1)}{1/(n-1)}\frac{\exp[(X - \mu_{(i)})^T(X - \mu_{(i)})/2\sigma^2]}{\sum_{k=1}^{n-1}\exp[(X - \mu_{(k)})^T(X - \mu_{(k)})/2\sigma^2]}$$

$$= \frac{\exp[(X - \mu_{(i)})^T(X - \mu_{(i)})/2\sigma^2]}{\sum_{k=1}^{n-1}\exp[(X - \mu_{(k)})^T(X - \mu_{(k)})/2\sigma^2]}$$

(c)

Using the Q value and the $\pi_i^{(t)}$ quanties derived above, we can derive the update steps.

$$\frac{dQ}{d\mu_1} = -\frac{1}{2\sigma^{(t)2}}\sum_{j=1}^{n}\mathbb{E}_{Z|X;\theta^{(t)}}(I(j \leq Z))\frac{d}{d\mu_1}(X_j - \mu_1^{(t)})^2$$

$$= -\frac{1}{2\sigma^{(t)2}}\sum_{j=1}^{n}\mathbb{E}_{Z|X;\theta^{(t)}}(I(j \leq Z))[2(X_j - \mu_1^{(t)})]$$

Setting this quantity equal to zero, we can solve for the update as follows

$$\mu_1^{(t+1)} \sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j \le Z)) = \sum_{j=1}^{n} X_j \mathbb{E}_{Z|X;\theta^{(t)}}(I(j \le Z))$$

$$\mu_1^{(t+1)} = \frac{\sum_{j=1}^{n} X_j \mathbb{E}_{Z|X;\theta^{(t)}}(I(j \le Z))}{\sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j \le Z))}$$

Calculating the same quanities for μ_2 we have the following

$$\frac{dQ}{d\mu_2} = -\frac{1}{2\sigma^{(t)2}} \sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z)) \frac{d}{d\mu_2}(X_j - \mu_2^{(t)})^2$$

$$= -\frac{1}{2\sigma^{(t)2}} \sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z))[2(X_j - \mu_2^{(t)})]$$

Setting this quantity equal to zero, we can solve for the update as follows

$$\mu_2^{(t+1)} \sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z)) = \sum_{j=1}^{n} X_j \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z))$$

$$\mu_2^{(t+1)} = \frac{\sum_{j=1}^{n} X_j \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z))}{\sum_{j=1}^{n} \mathbb{E}_{Z|X;\theta^{(t)}}(I(j > Z))}$$

Lastly we derive the updates for $\sigma^{2,(t+1)}$.

$$\frac{dQ}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n-1} (X - \mu_{(i)})^T (X - \mu_{(i)}) \mathbb{E}_{Z|X;\theta^{(t)}}(I(Z = i))$$

$$= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n-1} \pi_i^{(t)} (X - \mu_{(i)})^T (X - \mu_{(i)})$$

Therefore, setting equal to zero and solving for $\sigma^{2,(t+1)}$

$$\frac{n}{2\sigma^{2,(t+1)}} = \frac{1}{2(\sigma^2)^{2,(t+1)}} \sum_{i=1}^{n-1} \pi_i^{(t)} (X - \mu_{(i)})^T (X - \mu_{(i)})$$

$$\sigma^{2,(t+1)} = \frac{1}{n} \sum_{i=1}^{n-1} \pi_i^{(t)} (X - \mu_{(i)})^T (X - \mu_{(i)})$$

(d)

```r
# set up LSE functions
eps <- local({
    epsilon <- 1
    function() {
        while (1 + epsilon/2 > 1) epsilon <<- epsilon/2
        epsilon
    }
})
logeps <- log(eps()/2)
log1pe <- function(x) {
    l <- ifelse(x > 0, x, 0)
    x <- ifelse(x > 0, -x, x)
    ifelse(x < logeps, l, l + log(1 + exp(x)))
}
lse2 <- function(x, y) {
    m <- max(x, y)
    d <- -abs(x - y)
    ifelse(d < log(eps()), m, m + log1pe(d))
}
lse <- function(x) Reduce(lse2, x)

# set up helper functions
muit <- function(mu1, mu2, n, i) mu1 * (1:n <= i) + mu2 * (1:n >
    i)

# alpha = (X-mu)^T(X-mu)/2sigma^2
alpha <- function(mu1, mu2, sigma2, X, i) {
    mu <- muit(mu1, mu2, length(X), i)
    -sum((X - mu)^2/(2 * sigma2))
}

# log pi functions
logpi <- function(mu1, mu2, sigma2, X) {
    wrapper <- function(i) alpha(mu1, mu2, sigma2, X, i)
    alphas <- sapply(1:(length(X) - 1), wrapper)
    alphas - lse(alphas)
}

em <- function(mu1, mu2, sigma2, X) {

    # set up stopping criterion
    precision <- 1e-05
    diff <- Inf
    Q <- 0

    while (diff > precision) {

        # update logpis
        log.pi <- logpi(mu1, mu2, sigma2, X)

        # update wrapper functions
        EZ_greater <- function(j) sum(exp(log.pi[1:(j - 1)])),
```

```r
        na.rm = TRUE)
    EZ_less <- function(j) sum(exp(log.pi[(j):(length(X) -
        1)]), na.rm = TRUE)

    # update estimates
    mu1 <- sum(sapply(1:(length(X) - 1), EZ_less) * X[-n],
        na.rm = TRUE)/sum(sapply(1:(length(X) - 1), EZ_less),
        na.rm = TRUE)
    mu2 <- sum(sapply(2:length(X), EZ_greater) * X[-1], na.rm = TRUE)/sum(sapply(2:length(X),
        EZ_greater), na.rm = TRUE)

    # update sigma
    sigma_wrapper <- function(j) {
        mu <- muit(mu1, mu2, length(X), j)
        sum((X - mu)^2)
    }

    sigma2 <- 1/length(X) * sum(exp(logpi(mu1, mu2, sigma2,
        X)) * sapply(1:(length(X) - 1), sigma_wrapper))

    # update Q and stopping condition
    Q_new <- -length(X)/2 * log(sigma2) - 1/(2 * sigma2) *
        (sum(sapply(1:length(X), EZ_less) * (X - mu1)^2,
            na.rm = TRUE) + sum(sapply(1:length(X), EZ_greater) *
            (X - mu2)^2, na.rm = TRUE))
    diff <- abs(Q_new - Q)
    Q <- Q_new
}



    return(list(mu1 = mu1, mu2 = mu2, sigma2 = sigma2, pi = exp(logpi(mu1,
        mu2, sigma2, X))))

}
```

```r
# read in data
data(Nile)

# initialize inputs
mid <- round(length(Nile)/2)
n <- length(Nile)
mu1 <- mean(Nile[1:mid])
mu2 <- mean(Nile[(mid + 1):n])
sigma2 <- 1/n * (sum((Nile[1:mid] - mu1)^2) + sum((Nile[(mid +
    1):n] - mu2)^2))

res <- em(mu1, mu2, sigma2, Nile)

c(res[[1]], res[[2]], res[[3]])
```
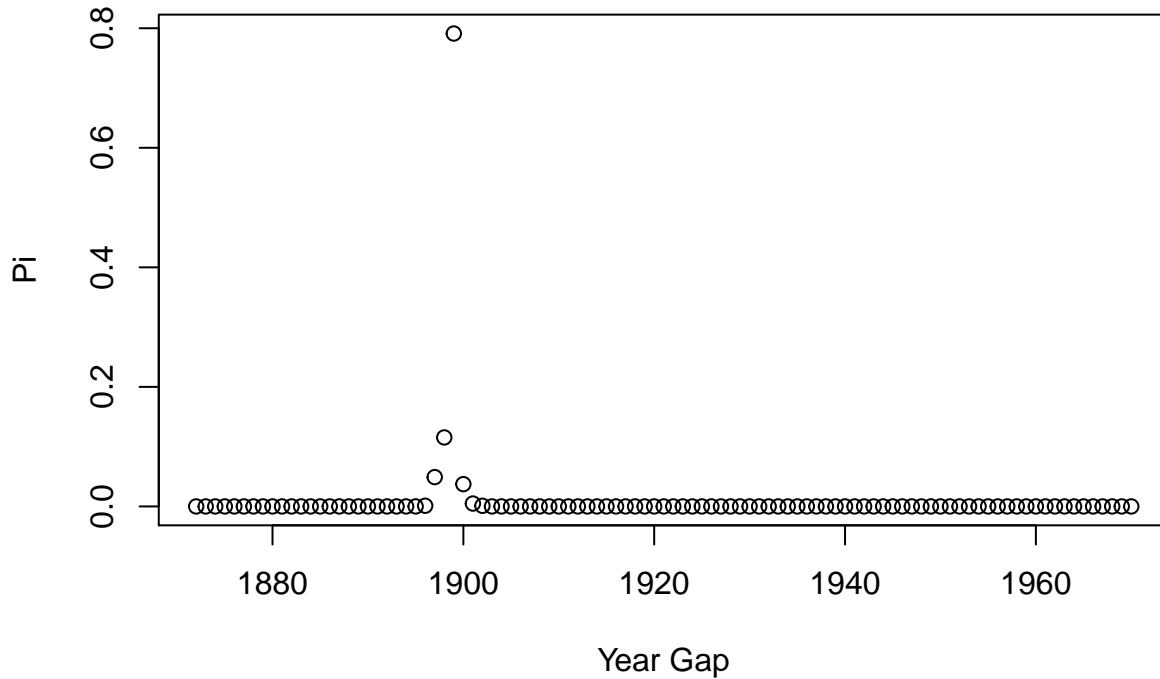
```
## [1]   1097.2884    850.7235 16140.2826
```

```r
plot(time(Nile)[-1], res[[4]], ylab = "Pi", xlab = "Year Gap")
```


Year Gap

Here we see that the estimates of $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2) = (1097.2884, 850.7235, 16140.2826)$. To visualize these results consider the following two plots.

```r
changepoint <- which.max(res[[4]])
center <- c(rep(res[[1]], changepoint), rep(res[[2]], length(Nile) -
    changepoint))
lower <- center - sqrt(res[[3]])
upper <- center + sqrt(res[[3]])

plot(Nile, xlab = "Year", ylab = "Flow")
points(time(Nile), center, col = "red", type = "l", lty = 1)
points(time(Nile), lower, col = "red", type = "l", lty = 2)
points(time(Nile), upper, col = "red", type = "l", lty = 2)
```