Computational Statistics - STAT 575 - HW #2

Alex Towell (atowell@siue.edu)

Contents

Problem 1

Derive the E-M algorithm for right-censored normal data with known variance, say $2 = 1$. Consider Y_i's that are i.i.d. from a $N(\theta, 1)$, $i = 1, 2 \ldots, n$. We observe $(x_1, \ldots, x_n)$ and $(\delta_1, \ldots, \delta_n)$, where $x_i = \min(y_i, c)$, and $\delta_i = I(y_i < c)$. Let C be the total number of censored (incomplete) observations. We denote the missing data as $\{Z_i : \delta_i = 0\}$.

Part (a)

Derive the complete log-likelihood, $\mathrm{l}(\theta | Y)$.

The unobserved random variates $\{Y_i\}$ are i.i.d. normally distributed,

$$Y_i \sim \mathrm{f}_{\mathrm{Y_i}}(y | \theta)$$

where

$$\mathrm{f}_{\mathrm{Y_i}}(y | \theta) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \theta)^2\right).$$

The likelihood function is therefore

$$\mathrm{L}(\theta | \{y_i\}) = \prod_{i=1}^{n} (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y_i - \theta)^2\right) \tag{1}$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left(-\sum_{i=1}^{n} \frac{1}{2}(y_i - \theta)^2\right). \tag{2}$$

Taking the logarithm of L,

$$\begin{aligned}
\ell(\theta|\{y_i\}) &= \log \mathrm{L}(\theta|\{y_i\}) && (3)\\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2 && (4)\\
&= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}y_i^2 + \theta\sum_{i=1}^{n}y_i - \frac{n}{2}\theta^2. && (5)
\end{aligned}$$

Anticipating that we will be maximizing the complete log-likelihood with respect to θ, we put any terms that are not a function of θ into k, obtaining the result

$$\ell(\theta|\{y_i\}) = k + \theta\sum_{i=1}^{n}y_i - \frac{n}{2}\theta^2.$$

Part (b)

Show the conditional expectation

$$\mathrm{E}(Y|x,\delta=1,\theta^{(t)}) = x$$

and

$$\mathrm{E}(Y|x,\delta=0,\theta^{(t)}) = \mathrm{E}(Y|Y>x) = \theta^{(t)} + \frac{\phi(x-\mu)}{1-\Phi(x-\mu)}$$

where ϕ and Φ are pdf and cdf of standard normal.

The distribution of Y given $\delta = 1$, is uncensored and therefore it is given that Y realized the value x. Since the expectation of a constant x is x, that means $\mathrm{E}(Y|Y=x) = x$.

If $\delta = 0$, Y is censored, i.e., $Y > x$. To take its expectation, we first need to derive the conditional conditional distribution of Y given $Y > x$ and $\theta^{(t)}$.

The probability $\Pr(Y \le y|Y > x)$ is given by

$$\Pr(Y \le y|Y>x) = \Pr(x < Y \le y)/\Pr(Y > x)$$

which may be rewritten as

$$\Pr(Y \le y|Y>x) = \frac{F_Y(y|\theta^{(t)}) - F_Y(x|\theta^{(t)})}{1 - F_Y(y|\theta^{(t)})}.$$

where $F_{Y|\theta^{(t)}}$ is the cdf of the normal distribution with $\sigma = 1$ and $\mu = \theta^{(t)}$.

We may rewrite $F_{Y|\theta^{(t)}}$ in terms of the standard normal,

$$F_Y(y|\theta^{(t)}) = \Phi(y-\theta^{(t)}),$$

and thus we may rewrite the conditional distribution of $Y|Y > x$ as

$$\Pr(Y \le y|Y>x) = \frac{\Phi(y-\theta^{(t)}) - \Phi(x-\theta^{(t)})}{1-\Phi(x-\theta^{(t)})}$$

and thus after further simplifying, we obtain the cdf of $Y|x$,

$$F_{Y|x}(y|\theta^{(t)}) = 1 - \frac{1-\Phi(y-\theta^{(t)})}{1-\Phi(x-\theta^{(t)})}$$

which has a density given by

$$f_Y(y|x,\theta^{(t)}) = \frac{\phi(y-\theta^{(t)})}{1-\Phi(x-\theta^{(t)})}I(y>x).$$

The expectation of $Y|(x, \theta^{(t)})$ is given by

$$
\begin{aligned}
\mathrm{E}(Y|x, \theta^{(t)}) &= \int_x^\infty y f_Y(y|x, \theta^{(t)}) dy && (6) \\
&= \int_x^\infty y \left(\frac{\phi(y - \theta^{(t)})}{1 - \Phi(x - \theta^{(t)})} \right) dy && (7) \\
&= \frac{1}{1 - \Phi(x - \theta^{(t)})} \int_x^\infty y \phi(y - \theta^{(t)}) dy. && (8)
\end{aligned}
$$

Analytically, this is a tricky integration problem. Certainly, it would be trivial to numerically integrate this to obtain a solution, but we seek a closed-form solution.

I searched online, and discovered an interesting way to tackle this integration problem.

Let f and F respectively denote the pdf and cdf of the normally distributed Y. Then,

$$\frac{df}{dy} = -(y - \theta) f(y)$$

and

$$\int_a^b \frac{df}{dy} dy = f(b) - f(a).$$

Then,

$$
\begin{aligned}
\mathrm{E}(Y|x, \theta^{(t)}) &= \frac{1}{1 - F(x)} \int_x^\infty y f(y) dy && (9) \\
&= -\frac{1}{1 - F(x)} \int_x^\infty -(y - \theta^{(t)}) f(y) dy + \frac{\theta^{(t)}}{1 - F(x)} \int_x^\infty f(y) dy && (10) \\
&= -\frac{1}{1 - F(x)} \int_x^\infty \frac{df}{dy} dy + \frac{\theta^{(t)}}{1 - F(x)} (1 - F(x)) && (11) \\
&= -\frac{1}{1 - F(x)} \left(f(\infty) - f(x) \right) + \theta^{(t)} && (12) \\
&= \frac{f(x)}{1 - F(x)} + \theta^{(t)}. && (13)
\end{aligned}
$$

We may rewrite the last line as

$$\mathrm{E}(Y|x, \theta^{(t)}) = \theta^{(t)} + \frac{\phi(x - \theta^{(t)})}{1 - \Phi(x - \theta^{(t)})}.$$

Part (c)

Derive the E-step and M-step using parts (a) and (b). Give the updating equation.

E-step

The E-step entails taking the conditional expectation of the complete log-likelihood function $\ell(\theta|\{Y_i\})$ given the observed data $\{x_i\}$ and $\{\delta_i\}$.

$$
\begin{aligned}
Q(\theta|\theta^{(t)}) &= \mathrm{E}_{Y_i|x_i,\delta_i}(\ell(\theta|\{Y_i\}) && (14)\\
&= \mathrm{E}_{Y_i|x_i,\delta_i}\left(k + \theta\sum_{i=1}^{n} Y_i - \frac{n}{2}\theta^2\right) && (15)\\
&= k - \frac{n}{2}\theta^2 + \theta\sum_{i=1}^{n}\mathrm{E}_{Y_i|x_i,\delta_i}(Y_i). && (16)
\end{aligned}
$$

We have already solved the expectation of Y_i given x_i and δ_i. We rewrite Q by substituting $\mathrm{E}(Y_i|x_i,\delta_i)$ with its previously found solution,

$$
Q(\theta|\theta^{(t)}) = k - \frac{n}{2}\theta^2 + \theta\sum_{i=1}^{n}\delta_i x_i + (1-\delta_i)\left(\theta^{(t)} + \frac{\phi(x_i-\theta^{(t)})}{1-\Phi(x_i-\theta^{(t)})}\right).
$$

Letting $C = \sum_{i=1}^{n}(1-\delta_i)$, $R := \sum_{i=1}^{n}\delta_i x_i$, and separating out all terms that are independent of $\theta^{(t)}$,

$$
Q(\theta|\theta^{(t)}) = k - \frac{n}{2}\theta^2 + C\theta\theta^{(t)} + R\theta + \theta\sum_{i=1}^{n}\frac{(1-\delta_i)\phi(x_i-\theta^{(t)})}{1-\Phi(x_i-\theta^{(t)})}.
$$

M-step

We wish to solve

$$
\theta^{(t+1)} = \arg\max_\theta Q(\theta|\theta^{(t)}).
$$

by solving

$$
\left.\frac{dQ(\theta|\theta^{(t)})}{d\theta}\right|_{\theta=\theta^{(t+1)}} = 0,
$$

which may be written as

$$
-n\theta^{(t+1)} + C\theta^{(t)} + R + \sum_{i=1}^{n}\frac{(1-\delta_i)\phi(x_i-\theta^{(t)})}{1-\Phi(x_i-\theta^{(t)})} = 0.
$$

Solving for $\theta^{(t+1)}$ obtains the updating equation

$$
\theta^{(t+1)} = \frac{R}{n} + \frac{C}{n}\theta^{(t)} + \frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_i)\phi(x_i-\theta^{(t)})}{1-\Phi(x_i-\theta^{(t)})}.
$$

where

$$
R := \sum_{i=1}^{n}\delta_i x_i
$$

and

$$
C := \sum_{i=1}^{n}(1-\delta_i).
$$

Part (d)

Use your algorithm on the V.A. data to find the MLE of μ. Take the log of the event times first and standardize by sample standard deviation. You may simply use the censored data sample mean as your starting value.

In the following R code, we implement the updating equation derived in the previous step. We encapsulate the procedure into a function that takes its arguments in the form of a censored set, uncensored set, starting value ($\theta^{(1)}$), and an ϵ value to control stopping condition.

```r
# assuming the uncensored and censored data are distributed normally,
# we use the EM algorithm to derive an estimator given censored and uncensored
# data.
mean_normal_censored_estimator_em <- function(uncensored,censored,theta,eps=1e-6,debug=T)
{
  dev <- sd(log(c(uncensored,censored)))
  censored <- log(censored) / dev
  uncensored <- log(uncensored) / dev
  theta <- log(theta) / dev

  n <- length(censored) + length(uncensored)
  C <- length(censored)
  R <- sum(uncensored)

  s <- function(theta)
  {
    sum <- 0
    for (i in 1:C)
    {
      num <- dnorm(censored[i],mean=theta,sd=1)
      denom <- 1-pnorm(censored[i],mean=theta,sd=1)
      sum <- sum + (num / denom)
    }
    sum
  }

  i <- 1
  repeat
  {
    theta.new <- R/n + C/n * theta + (1/n)*s(theta)
    if (debug==T) { cat("theta[", i, "] =",theta,", theta[", i+1, "] =",theta.new,"\n") }
    if (abs(theta.new - theta) < eps)
    {
      theta <- theta.new * dev
      theta <- exp(theta)
      return(theta)
    }
    i <- i + 1
    theta <- theta.new
  }
}
```

We apply this procedure to the indicated data set.

```r
library(MASS) # has VA data
VAs <- subset(VA,prior==0)
censored <- VAs$status == 0
censored_xs <- VAs[censored,c("stime")]
uncensored_xs <- VAs[!censored,c("stime")]

mu <- mean(uncensored_xs)
cat("mean of the uncensored sample is ", mu, ".")
```

```
## mean of the uncensored sample is  112.1648 .
```

```r
sol <- mean_normal_censored_estimator_em(uncensored_xs,censored_xs,mu)
```

```
## theta[ 1 ] = 3.857928 , theta[ 2 ] = 3.424258
## theta[ 2 ] = 3.424258 , theta[ 3 ] = 3.415443
## theta[ 3 ] = 3.415443 , theta[ 4 ] = 3.415286
## theta[ 4 ] = 3.415286 , theta[ 5 ] = 3.415283
## theta[ 5 ] = 3.415283 , theta[ 6 ] = 3.415283
```

```r
sol
```

```
## [1] 65.2625
```

We see that our estimate of θ is $\hat{\theta} = 65.2624985$. (The θ before transforming it to the appropriate scale was 3.415283.)

This mean is somewhat lower than anticipated, which makes me suspect something is wrong with my updating equation. If I have the time, I will revisit it.

Problem 2

Part (a)

There are $N = 1500$ gay men in the survey sample where X_i denotes the i-th persons response to the number of risky sexual encounters he had in the previous 30 days. Thus, we observe a sample $\vec{X} = (X_1, X_2, \ldots, X_N)$.

We assume there are 3 groups in the population, denoted by $z = 1$, $t = 2$, and $p = 3$. Group 1 members report 0 risky sexual encounters regardless of the truth where the probability of being a member of group 1 is denoted by α,

Group 2 members accurately report risky sexual encounters and represent typical behavior where the probability of being a member of group 2 is denoted by β. We assume this group's number of sexual encounters follows a poisson with mean μ.

Group 3 members accurately report risky sexual encounters and represent high-risk behavior where the probability of being a member of group 3 is $\gamma = 1 - \alpha - \beta$. We assume this group's number of sexual encounters follows a poisson with mean λ.

This represents a finite mixture model with a pdf

$$X_i \sim \mathrm{f}(x|\vec{\theta}) = \alpha I(x = 0) + \beta \,\mathrm{POI}(x|\mu) + (1 - \alpha - \beta)\,\mathrm{POI}(x|\lambda)$$

with a parameter vector

$$\vec{\theta} = (\alpha, \beta, \mu, \lambda)'.$$

Let the uncertain group that the i-th person belongs to be denoted by Z_i. If we observe group membership data, $X_i | Z_i = z_i$, then

$$X_i | Z_i = 1 \sim I(x = 0), \tag{17}$$
$$X_i | Z_i = 2 \sim \mathrm{POI}(\mu), \tag{18}$$
$$X_i | Z_i = 3 \sim \mathrm{POI}(\lambda), \tag{19}$$

where

$$Z_i \sim f_{Z_i}(z_i|\vec{\theta}) = \Pr(Z_i = z_i) = \begin{cases} \alpha & z_i = 1, \\ \beta & z_i = 2, \\ \gamma = 1 - \alpha - \beta & z_i = 3, \end{cases}$$

and thus

$$\mathrm{f}_{X_i, Z_i}(x_i, z_i|\vec{\theta}) = \alpha I(z_i = 1) + \beta\,\mathrm{POI}(\mu) I(z_i = 2) + (1 - \alpha - \beta)\,\mathrm{POI}(\lambda) I(z_i = 3).$$

The *complete* likelihood function is thus given by

$$\mathcal{L}(\vec{\theta}|\vec{X},\vec{Z}) = \prod_{i=1}^{N} \mathrm{f}_{X_i,Z_i}(x_i, z_i|\vec{\theta}),$$

which may be rewritten as

$$\mathcal{L}(\vec{\theta}|\vec{X},\vec{Z}) = \left(\prod_{\{i|z_i=1\}} \alpha I(x_i = 0)\right)\left(\prod_{\{i|z_i=2\}} \beta\frac{\mu^{x_i}e^{-\mu}}{x_i!}\right)\left(\prod_{\{i|z_i=3\}} \gamma\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}\right).$$

We wish to rewrite this so that the data is explicitly represented. First, we do the transformation

$$\mathcal{L}(\vec{\theta}|\vec{X},\vec{Z}) = \left(\prod_{\{i|z_i=1,x_i=0\}} \alpha\right)\prod_{k=0}^{16}\left(\prod_{\{i|z_i=2,x_i=k\}} \beta\frac{\mu^{k}e^{-\mu}}{k!}\right)\prod_{k=0}^{16}\left(\prod_{\{i|z_i=3,x_i=k\}} \gamma\frac{\lambda^{k}e^{-\lambda}}{k!}\right).$$

We let $n_{a,b}$ denote the (unobserved) cardinality of $\{i|z_i = a, x_i = b\}$, thus

$$\mathcal{L}(\vec{\theta}|\{n_{j,k}\}) = \alpha^{n_{1,0}}\prod_{k=0}^{16}\beta^{n_{2,k}}\frac{\mu^{kn_{2,k}}e^{-\mu n_{2,k}}}{(k!)^{n_{2,k}}}\prod_{k=0}^{16}\gamma^{n_{3,k}}\frac{\lambda^{kn_{3,k}}e^{-\lambda n_{3,k}}}{(k!)^{n_{3,k}}}$$

is the complete likelihood. The complete log-likelihood is thus

$$\ell(\vec{\theta}|\{n_{j,k}\}) = n_{1,0}\log\alpha + \sum_{k=0}^{16}\log\left(\beta^{n_{2,k}}\frac{\mu^{kn_{2,k}}e^{-\mu n_{2,k}}}{(k!)^{n_{2,k}}}\right) + \sum_{k=0}^{16}\log\left(\gamma^{n_{3,k}}\frac{\lambda^{kn_{3,k}}e^{-\lambda n_{3,k}}}{(k!)^{n_{3,k}}}\right)$$

which simplies to

$$\begin{aligned}\ell(\vec{\theta}|\{n_{j,k}\}) = n_{1,0}\log\alpha + \sum_{k=0}^{16} & n_{2,k}(\log\beta + k\log\mu - \mu - \log k!) + \\ & n_{3,k}(\log\gamma + k\log\lambda - \lambda - \log k!).\end{aligned} \tag{20}$$

Anticipating taking $\frac{d\ell}{d\vec{\theta}}$ to solve for the maximum of the log-likelihood, we remove any terms that are not a function of $\vec{\theta}$, resulting in the kernel

$$\ell(\vec{\theta}|\{n_{j,k}\}) = n_{1,0}\log\alpha + \sum_{k=0}^{16}\left\{n_{2,k}(\log\beta + k\log\mu - \mu) + n_{3,k}(\log\gamma + k\log\lambda - \lambda)\right\}.$$

E-step

The conditional expectation to solve in the EM algorithm is given by

$$Q(\vec{\theta}|\vec{\theta}^{(t)}) = \mathrm{E}(\ell(\vec{\theta}))$$

where $\{n_{k,j}\}$ are random and $\{n_j\}$ and $\vec{\theta}^{(t)}$ are given. We rewrite this as

$$Q(\vec{\theta}|\vec{\theta}^{(t)}) = \mathrm{E}\left(n_{1,0}\log\alpha + \sum_{k=0}^{16}\left\{n_{2,k}(\log\beta + k\log\mu - \mu) + n_{3,k}(\log\gamma + k\log\lambda - \lambda)\right\}\right).$$

Using the linearity of expectations, we rewrite the above to

$$Q(\vec{\theta}|\vec{\theta}^{(t)}) = \mathrm{E}(n_{1,0})\log\alpha + \sum_{k=0}^{16}\left\{\mathrm{E}(n_{2,k})(\log\beta + k\log\mu - \mu) + \mathrm{E}(n_{3,k})(\log\gamma + k\log\lambda - \lambda)\right\}$$

given $\{n_j\}$ and $\theta^{(t)}$.

Consider $\mathrm{E}\big(n_{2,k}|\{n_j\}, \theta^{(t)}\big)$. To solve this expectation, we must first derive the distribution of $n_{2,k}$.

Suppose $x_j = k$, then probability that the j-th person belongs to group 2 is given by

$$\Pr(Z_j = 2|x_j = k) = \Pr(Z_j = 2)\Pr(x_j = k|Z_j = 2)/\Pr(x_j = k).$$

We note that $\Pr(x_j = k)$ is equivalent to $\pi_k(\vec{\theta})$, $\Pr(Z_j = 2)$ is the definition of β, and $\Pr(x_j = k|Z_j = 2)$ is $\mathrm{f}_{X_j|Z_j}(k|Z_j = 2) = \mathrm{POI}(k|\mu)$.

Making the substitutions yields the result

$$t_k(\vec{\theta}) = \Pr(Z_j = 2|x_j = k) = \beta\,\mathrm{POI}(k|\mu)/\pi_k(\vec{\theta}).$$

Assuming $\{X_i\}$ are i.i.d., observe that $k \neq 0$, the distribution of $n_{2,k}$ given n_k, $\theta^{(t)}$ is binomial distributed with a probability of success $t_k(\vec{\theta}^{(t)})$. Thus,

$$\mathrm{E}(n_{2,k}) = n_k t_k(\vec{\theta}^{(t)}).$$

The same logic holds for $n_{3,k}$ and $n_{1,0}$, and thus

$$\mathrm{E}(n_{3,k}) = n_k p_k(\vec{\theta}^{(t)})$$

and

$$\mathrm{E}(n_{1,0}) = n_0 z_0(\vec{\theta}^{(t)}),$$

which means

$$Q(\vec{\theta}|\vec{\theta}^{(t)}) = n_0 z_0(\vec{\theta}^{(t)})\log\alpha + \sum_{k=0}^{16}\Big\{n_k t_k(\vec{\theta}^{(t)})(\log\beta + k\log\mu - \mu) + n_k p_k(\vec{\theta}^{(t)})(\log\gamma + k\log\lambda - \lambda)\Big\}$$

M-step

We wish to solve

$$\vec{\theta}^{(t+1)} = \arg\max_{\vec{\theta}} Q(\vec{\theta}|\vec{\theta}^{(t)}).$$

by solving

$$\nabla Q(\vec{\theta}|\vec{\theta}^{(t)})\Big|_{\vec{\theta}=\vec{\theta}^{(t+1)}} = \vec{0}.$$

We use the Lagrangian to impose the restriction $\alpha + \beta + \gamma = 1$, thus we seek to perform the constrained maximization of

$$Q_l(\vec{\theta}, c|\vec{\theta}^{(t)}) = Q(\vec{\theta}|\vec{\theta}^{(t)}) + c(1 - \alpha - \beta - \gamma).$$

Thus, when we solve for α,

$$\frac{\partial Q_l}{\partial\alpha} = \frac{n_0 z_0(\theta^{(t)})}{\alpha} - c = 0,$$

we get the result

$$\alpha^{(t+1)} = \frac{1}{c} n_0 z_0(\theta^{(t)}).$$

Similar results hold for β and γ, obtaining

$$\beta^{(t+1)} = \frac{1}{c}\sum_{k=0}^{16} n_k t_k(\theta^{(t)}).$$

and

$$\gamma^{(t+1)} = \frac{1}{c} \sum_{k=0}^{16} n_k p_k(\theta^{(t)}).$$

This does not look too promising until we realize that

$$n_0 z_0(\theta^{(t)}) + \sum_{k=0}^{16} n_k t_k(\theta^{(t)}) + \sum_{k=0}^{16} n_k p_k(\theta^{(t)}) = N.$$

Thus, $c(\alpha^{(t)} + \beta^{(t)} + \gamma^{(t)}) = N$, which means $c = N$ since $\alpha^{(t)} + \beta^{(t)} + \gamma^{(t)} = 1$. Making this substitution obtains the result

$$\alpha^{(t+1)} = \frac{1}{N} n_0 z_0(\theta^{(t)}) \tag{21}$$

$$\beta^{(t+1)} = \frac{1}{N} \sum_{k=0}^{16} n_k t_k(\theta^{(t)}) \tag{22}$$

$$\gamma^{(t+1)} = \frac{1}{N} \sum_{k=0}^{16} n_k p_k(\theta^{(t)}). \tag{23}$$

Solving an estimator for μ at iteration $(t+1)$,

$$\left.\frac{\partial Q_l}{\partial \mu}\right|_{\mu=\mu^{(t+1}} = 0 \tag{24}$$

$$\sum_{k=0}^{16} n_k t_k(\theta^{(t)})(k/\mu^{(t+1)} - 1) = 0 \tag{25}$$

$$\frac{1}{\mu^{(t+1)}} \sum_{k=0}^{16} n_k t_k(\theta^{(t)}) k = \sum_{k=0}^{16} n_k t_k(\theta^{(t)}) \tag{26}$$

$$\mu^{(t+1)} = \frac{\sum_{k=0}^{16} k n_k t_k(\theta^{(t)})}{\sum_{k=0}^{16} n_k t_k(\theta^{(t)})}. \tag{27}$$

The same derivation essentially follows for λ, and thus

$$\lambda^{(t+1)} = \frac{\sum_{k=0}^{16} k n_k p_k(\theta^{(t)})}{\sum_{k=0}^{16} n_k p_k(\theta^{(t)})}.$$

Part (b)

Estimate the parameters of the model, using the observed data.

```
# we observe n = (n0,n1,...,n16)
ns <- c(379,299,222,145,109,95,73,59,45,30,24,12,4,2,0,1,1)
N <- sum(ns)

# theta := (alpha, beta, mu, lambda)'
# note that there is an implicit parameter gamma s.t.
# alpha + beta + gamma = 1
# the initial value assumes each category z, t, or p
# is equally probable, and so we let
#     (alpha^(0),beta^(0)) = (1/3,1/3)
```

```r
# and mu^(0) and lambda^(0) are just arbitrarily chosen to be 2 and 3,
# with the insight that group 3 is more risky than group 2.
theta <- c(1/3,1/3,2,3)

# theta := (alpha, beta, mu, lambda)
Pi <- function(i,theta)
{
  res <- 0
  if (i == 0)
    res <- theta[1]

  res <- res + theta[2] * theta[3]^i * exp(-theta[3])
  res <- res + (1 - theta[1] - theta[2]) * theta[4]^i * exp(-theta[4])
  res
}

z0 <- function(theta)
{
  theta[1] / Pi(0,theta)
}

t <- function(i,theta)
{
  theta[2] * theta[3]^i * exp(-theta[3]) / Pi(i,theta)
}

p <- function(i,theta)
{
  (1-theta[1] - theta[2]) * theta[4]^i * exp(-theta[4]) / Pi(i,theta)
}

# update algorithm, based on EM algorithm
update <- function(theta,ns)
{
  # note: n0 := ns[1] instead of ns[0] since R does not use zero-based indexes
  alpha <- ns[1] * z0(theta) / N
  beta <- 0
  mu_num <- 0
  mu_denom <- 0

  lam_num <- 0
  lam_denom <- 0

  for (i in 0:16)
  {
    ti <- t(i,theta)
    pi <- p(i,theta)

    beta <- beta + ns[i+1] * ti

    mu_num <- mu_num + i * ns[i+1] * ti
    mu_denom <- mu_denom + ns[i+1] * ti
```

```
    lam_num <- lam_num + i * ns[i+1] * pi
    lam_denom <- lam_denom + ns[i+1] * pi
  }

  beta <- beta / N
  mu <- mu_num / mu_denom
  lam <- lam_num / lam_denom

  c(alpha,beta,mu,lam)
}

em <- function(theta,ns,steps=10000,debug=T)
{
  for(i in 1:steps)
  {
    theta = update(theta,ns)
    if (debug==T)
    {
      if (i %% 1000 == 0) { cat("iteration =",i," theta = (",theta,")'\n") }
    }
  }
  theta
}

# solution theta = (alpha, beta, mu, lambda)
sol <- em(theta,ns,10000,T)
```

```
## iteration = 1000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 2000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 3000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 4000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 5000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 6000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 7000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 8000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 9000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
## iteration = 10000  theta = ( 0.1221661 0.5625419 1.467475 5.938889 )'
```

We see that the solution is $0.1221661, 0.5625419, 1.4674746, 5.9388889$.

Part (c)

Estimate the standard errors and pairwise correlations of your parameters, using any available method.

We have chosen to use the Bootstrap method.

```
# ns = (379,299,222,145,109,95,73,59,45,30,24,12,4,2,0,1,1)
# 379 responded 0 encounters
# 299 responded 1 encounters
# 222 responded 2 encounters
# ...
# 1 responded 16 encounters
#
# to resample, we resample from the data set that includes each
# persons response, as determined by ns.
```

```r
data <- NULL
for (i in 1:length(ns))
{
  data <- append(data,rep((i-1),ns[i]))
}

make_into_counts <- function(data)
{
  ns <- NULL
  for (i in 0:16)
  {
    ni <- data[data == i]
    l <-length(ni)
    ns <- append(ns,l)
  }
  ns
}

m <- 1000 # bootstrap replicates
steps <- 500
theta.bs <- em(theta,ns,steps,F)
thetas <- rbind(theta.bs)
for (i in 2:m)
{
  indices <- sample(N,N,replace=T)
  resampled <- make_into_counts(data[indices])
  theta.bs <- em(theta,resampled,steps,F)
  thetas <- rbind(thetas,theta.bs)
  if (i %% 100 == 0) { cat("iteration", i, ": ", theta.bs, "\n") }
}
```

```
## iteration 100 :  0.1216909 0.5421939 1.438922 5.775278
## iteration 200 :  0.1499283 0.5680229 1.634296 6.170384
## iteration 300 :  0.1394451 0.5792493 1.618992 6.387598
## iteration 400 :  0.1367241 0.5662286 1.584978 6.274667
## iteration 500 :  0.1300427 0.5689354 1.450772 6.209751
## iteration 600 :  0.1306883 0.5345558 1.382276 5.687018
## iteration 700 :  0.1334102 0.5792411 1.736552 6.222815
## iteration 800 :  0.1048649 0.5736968 1.42188 6.09429
## iteration 900 :  0.1155012 0.5519168 1.407184 6.15593
## iteration 1000 :  0.1244258 0.549874 1.483944 5.897094
```

```r
cov.bs <- cov(thetas)
cor.bs <- cor(thetas)
```

The Bootstrap estimator of the covariance matrix is given by

```
##                [,1]          [,2]         [,3]        [,4]
## [1,]  0.0004531987 -2.053855e-04 1.838140e-03 0.001991739
## [2,] -0.0002053855  4.784477e-04 6.370216e-05 0.001249046
## [3,]  0.0018381396  6.370216e-05 1.354867e-02 0.015906119
## [4,]  0.0019917392  1.249046e-03 1.590612e-02 0.041218913
```

and the correlation matrix is given by

```
##             [,1]        [,2]        [,3]       [,4]
```

```
## [1,]  1.0000000 -0.44107073 0.74179825 0.4608291
## [2,] -0.4410707  1.00000000 0.02502007 0.2812632
## [3,]  0.7417982  0.02502007 1.00000000 0.6730814
## [4,]  0.4608291  0.28126321 0.67308140 1.0000000
```