Solution keys of ASSIGNMENT 2   Due on Mar. 19, 2004

---

1. Since the likelihood is the same as

$$y_i \sim p 1_{(y_i=0)} + (1-p)\text{Pois}(y_i; \lambda) \text{ for } i = 1, \dots, n$$

where 1 is an indicator function, the loglikelihood will be given by

$$\ell = n_0 \log(p + (1-p)e^{-\lambda}) + (n - n_0)\log(1-p) + \sum_{k=1}^{6} k n_k \log \lambda - (n - n_0)\lambda$$

where $n_k = \sum_{i=1}^{n} 1_{(y_i=k)}$. Then, Newton-Raphson and the method of scoring will be computed as

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\partial \ell}{\partial \theta}\left(\frac{\partial^2 \ell}{\partial \theta^2}\right)^{-1} \text{ and}$$

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\partial \ell}{\partial \theta}\left(\text{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]\right)^{-1},$$

respectively where $\theta = (p, \lambda)$. Both methods give us the estimates of p and λ such as **0.6151** and **1.0378**, and the corresponding standard error estimates are **0.0134** and **0.0392**, respectively. The convergence rate of Newton-Raphson can be faster than the method of Scoring, but the latter is numerically more stable.

2.

(a) Because the observation that is censored is only known up to the tail probability, the likelihood is given by

$$L(\mu|y) = \left(\prod_{j:\delta_j=0} \frac{1}{\mu} e^{-c_j/\mu}\right)\left(\prod_{j:\delta_j=1} e^{-c_j/\mu}\right).$$

Thus, the MLE $\hat{\mu}$ will be $\sum_j c_j/n_0$ where δ_j is equal to 1 if censored, and $n_0 = \sum_j(1 - \delta_j)$.

(b) Intuitively the missing data will be $Y_{\text{mis}} = \{w_j; \forall_j \text{ s.t. } \delta_j = 1\}$. Since the augmented-data loglikelihood is linear in the missing data, the EM algorithm is given by

E-step: $\hat{w}_j \equiv \text{E}[w_j|\mu^{(t)}, y] = c_j + \mu^{(t)}$ by the memoryless property.

M-step: $\mu^{(t+1)} = \left(\sum_{j:\delta_j=0} c_j + \sum_{j:\delta_j=1} \hat{w}_j\right)/n.$

(c) You can show it by proving the second derivative is always less than 0, or by a mapping such that $\text{M}(\mu^{(t)}) = n_0\hat{\mu}/n + n_1\mu^{(t)}/n$ where $n_1 = \sum_j \delta_j$.

(d) The rate of convergence is given by $\text{DM} = I - I_{\text{obs}}(I_{\text{aug}})^{-1}$, so in this case $\text{DM} = n_1/n$.

3.

(a) The joint density is given by

$$p(y_i, z_{i1}, \ldots, z_{iK}) = \prod_{k=1}^{K} \left\{ \pi_k f(y_i | \mu_k, \sigma_k^2) \right\}^{z_{ik}}.$$

(b) Because the loglikelihood is linear in z_{ik}, the E-step consists of computing $\mathrm{E}[z_{ik} | y_i, \theta^{(t)}]$ for each i and k, which results in the relative probability of belonging to the component k; refer to the code.

(c) This is a simple calculation to solve the first order conditions; refer to the code.

(d) The EM code is as follows:

```
leuk <- read.table("c:/Splus/Data/leuk.txt",header=T)
E.step <- function(){
  for(i in 1:N){
    for(j in 1:K)
      temp[j] <- prob[j]*dnorm(data[i],mu[j],sig[j])
    miss[i,] <<- temp/sum(temp)
  }
}
M.step <- function(){
  for(j in 1:K){
    prob[j] <<- sum(miss[,j])/N
    mu[j] <<- sum(miss[,j]*data)/sum(miss[,j])
    sig[j] <<- sqrt(sum(miss[,j]*(data-mu[j])^2)/sum(miss[,j]))
  }
}
loglike <- function(){
  temp <- 0
  for(j in 1:K)
    temp <- temp + prob[j]*dnorm(data,mu[j],sig[j])
  return(sum(log(temp)))
}
em.iter <- function(){
  E.step()
  M.step()
  iter <<- iter+1
  em.draws[iter,] <<- c(mu,sig,prob,loglike())
}
data <- leuk[,1]
K <- 3 # number of poplutations
n.iters <- 10000
N <- length(data)
miss <- matrix(0,ncol=K,nrow=N)
mu <- sig <- temp <- 1:K
prob <- rep(1/K,K)
```

Table 1: The EM outputs.

		Gene 1	Gene 2	Gene 3
k = 1	μ_1	0.4165	-1.0915	0.3267
	σ_1	0.9005	0.5158	1.6517
	p_1	1.0000	1.0000	1.0000
	$\ell(\hat{\theta})$	-94.6148	-54.4912	-138.2925
	BIC	197.7829	117.5357	285.1383
k = 2	μ_1	0.0521	0.2159	-1.2217
	μ_2	2.4375	-1.2580	1.3246
	σ_1	0.2593	0.5392	0.1353
	σ_2	0.3374	0.1320	1.3905
	p_1	0.8472	0.1130	0.3919
	p_2	0.1528	0.8870	0.6081
	$\ell(\hat{\theta})$	-38.6473	7.6485	-102.2423
	BIC	**98.6779**	**6.0863**	225.8679
k = 3	μ_1	0.0521	0.2207	-1.2230
	μ_2	2.1262	-1.0498	2.9815
	μ_3	2.6715	-1.2900	0.5739
	σ_1	0.2593	0.5345	0.1346
	σ_2	0.2698	0.0341	0.2380
	σ_3	0.1305	0.1106	0.9945
	p_1	0.8472	0.1126	0.3905
	p_2	0.0656	0.1189	0.1888
	p_3	0.0872	0.7686	0.4207
	$\ell(\hat{\theta})$	-36.8716	10.6041	-91.9861
	BIC	107.9565	13.0051	**218.1855**

```
iter <- 0
log.diff <- diff <- 100
em.draws <- matrix(0,ncol=3*K+1,nrow=n.iters)

while(log.diff > 1.e-5 || diff > 1.e-5){
  pre.loglike <- loglike()
  pre.param <- c(mu,sig,prob)
  em.iter()
  log.diff <- loglike() - pre.loglike
  diff <- sqrt(sum((pre.param-c(mu,sig,prob))^2))
  print(c(iter,loglike()))
}
em.draws <- em.draws[1:iter,]
```

(e) The outputs are given by Table 1.

Table 2: The Observed Information.

		π	μ_1	μ_2	σ_1^2	σ_2^2
Louis's method	π	556.2548	-0.0007	-0.0012	-0.0031	0.0107
	μ_1	-0.0007	907.4966	-0.0016	-0.0043	0.0145
	μ_2	-0.0012	-0.0016	96.627	-0.0073	0.0259
	σ_1^2	-0.0031	-0.0043	-0.0073	6750.4061	0.066
	σ_2^2	0.0107	0.0145	0.0259	0.066	424.1894
SEM	π	555.8483	-0.5326	-0.2408	-2.6693	-0.3567
	μ_1	-0.5326	906.843	-0.37	-2.7597	-0.4816
	μ_2	-0.2408	-0.37	96.5585	-2.5275	-0.1917
	σ_1^2	-2.6693	-2.7597	-2.5275	6745.5834	-2.5964
	σ_2^2	-0.3567	-0.4816	-0.1917	-2.5964	424.1162

(f) Based on Table 1, the estimated number of classes for Gene 1, 2, and 3 is given by 2 classes, 2 classes, and 3 classes, respectively.

(g) Since the augmented-data loglikelihood is linear in the missing data, the Louis' method will be the same as $I_{\mathrm{obs}} = I_{\mathrm{aug}} - \mathrm{Var}(\ell'|Y_{\mathrm{obs}}, \theta)$ and the SEM as $I_{\mathrm{obs}} = (I - \mathrm{DM})I_{\mathrm{aug}}$. The corresponding I_{obs} is thus given by Table 2.