1. Derive the E-M algorithm for right-censored normal data with known variance, say $\sigma^2 = 1$. Consider Y_i's that are i.i.d. from a $N(\theta, 1)$, $i = 1, 2 \cdots, n$. We observe $(x_1, x_2, \cdots, x_n)$ and $(\delta_1, \delta_2, \cdots, \delta_n)$, 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 $\{Z_i : \delta_i = 0\}$

(a) Derive the complete log likelihood, $l(\theta|Y)$.
(b) Show the conditional expectation

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

and

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

where $\phi()$ and $\Phi()$ are pdf and CDF of standard normal.
(c) Derive the E-step and the M-step using part (a) and (b). Give the updating equation.
(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. You may use the following code as a start.

```
library(MASS)
VAs=subset(VA, prior==0)
attach(VAs)
ltime=log(stime)/sd(log(stime))
```

2. Do Problem 4.2 on textbook page 121-123. I have the question printed out below.

TABLE 4.2 Frequencies of respondents reporting numbers of risky sexual encounters; see Problem 4.2.

Encounters, i	0	1	2	3	4	5	6	7	8
Frequency, n_i	379	299	222	145	109	95	73	59	45
Encounters, i	9	10	11	12	13	14	15	16	
Frequency, n_i	30	24	12	4	2	0	1	1	

Jamshidian and Jennrich elaborate on inverse-secant updating and discuss the more complex BFGS approach [344]. These authors also provide a useful survey of a variety of EM acceleration approaches and a comparison of effectiveness. Some of their approaches converge faster on examples than does the approach described above. In a related paper, they present a conjugate gradient acceleration of EM [343].

PROBLEMS

4.1. Recall the peppered moth analysis introduced in Example 4.2. In the field, it is quite difficult to distinguish the *insularia* and *typica* phenotypes due to variations in wing color and mottle. In addition to the 622 moths mentioned in the example, suppose the sample collected by the researchers actually included $n_{\mathrm{U}} = 578$ more moths that were known to be *insularia* or *typical* but whose exact phenotypes could not be determined.

a. Derive the EM algorithm for maximum likelihood estimation of p_{C}, p_{I}, and p_{I} for this modified problem having observed data n_{C}, n_{I}, n_{T}, and n_{U} as given above.

b. Apply the algorithm to find the MLEs.

c. Estimate the standard errors and pairwise correlations for $\hat{p}_{\mathrm{C}}$, $\hat{p}_{\mathrm{I}}$, and $\hat{p}_{\mathrm{I}}$ using the SEM algorithm.

d. Estimate the standard errors and pairwise correlations for $\hat{p}_{\mathrm{C}}$, $\hat{p}_{\mathrm{I}}$, and $\hat{p}_{\mathrm{I}}$ by bootstrapping.

e. Implement the EM gradient algorithm for these data. Experiment with step halving to ensure ascent and with other step scalings that may speed convergence.

f. Implement Aitken accelerated EM for these data. Use step halving.

g. Implement quasi-Newton EM for these data. Compare performance with and without step halving.

h. Compare the effectiveness and efficiency of the standard EM algorithm and the three variants in (e), (f), and (g). Use step halving to ensure ascent with the three variants. Base your comparison on a variety of starting points. Create a graph analogous to Figure 4.3.

4.2. Epidemiologists are interested in studying the sexual behavior of individuals at risk for HIV infection. Suppose 1500 gay men were surveyed and each was asked how many risky sexual encounters he had in the previous 30 days. Let n_i denote the number of respondents reporting i encounters, for $i = 1, \ldots, 16$. Table 4.2 summarizes the responses.

These data are poorly fitted by a Poisson model. It is more realistic to assume that the respondents comprise three groups. First, there is a group of people who, for whatever reason, report zero risky encounters even if this is not true. Suppose a respondent has probability α of belonging to this group.

With probability β, a respondent belongs to a second group representing typical behavior. Such people respond truthfully, and their numbers of risky encounters are assumed to follow a Poisson(μ) distribution.

Finally, with probability $1 - \alpha - \beta$, a respondent belongs to a high-risk group. Such people respond truthfully, and their numbers of risky encounters are assumed to follow a Poisson(λ) distribution.

The parameters in the model are α, β, μ, and λ. At the tth iteration of EM, we use $\boldsymbol{\theta}^{(t)} = (\alpha^{(t)}, \beta^{(t)}, \mu^{(t)}, \lambda^{(t)})$ to denote the current parameter values. The likelihood of the observed data is given by

$$L(\boldsymbol{\theta}|n_0, \ldots, n_{16}) \propto \prod_{i=0}^{16} \left[\frac{\pi_i(\boldsymbol{\theta})}{i!}\right]^{n_i}, \tag{4.81}$$

where

$$\pi_i(\boldsymbol{\theta}) = \alpha 1_{\{i=0\}} + \beta\mu^i \exp\{-\mu\} + (1 - \alpha - \beta)\lambda^i \exp\{-\lambda\} \tag{4.82}$$

for $i = 1, \ldots, 16$.

The observed data are $n_0, \ldots, n_{16}$. The complete data may be construed to be $n_{z,0}, n_{t,0}, \ldots, n_{t,16}$, and $n_{p,0}, \ldots, n_{p,16}$, where $n_{k,i}$ denotes the number of respondents in group k reporting i risky encounters and $k = z, t$, and p correspond to the zero, typical, and promiscuous groups, respectively. Thus, $n_0 = n_{z,0} + n_{t,0} + n_{p,0}$ and $n_i = n_{t,i} + n_{p,i}$ for $i = 1, \ldots, 16$. Let $N = \sum_{i=0}^{16} n_i = 1500$.

Define

$$z_0(\boldsymbol{\theta}) = \frac{\alpha}{\pi_0(\boldsymbol{\theta})}, \tag{4.83}$$

$$t_i(\boldsymbol{\theta}) = \frac{\beta\mu^i \exp\{-\mu\}}{\pi_i(\boldsymbol{\theta})}, \tag{4.84}$$

$$p_i(\boldsymbol{\theta}) = \frac{(1 - \alpha - \beta)\lambda^i \exp\{-\lambda\}}{\pi_i(\boldsymbol{\theta})} \tag{4.85}$$

for $i = 0, \ldots, 16$. These correspond to probabilities that respondents with i risky encounters belong to the various groups.

a. Show that the EM algorithm provides the following updates:

$$\alpha^{(t+1)} = \frac{n_0 z_0(\boldsymbol{\theta}^{(t)})}{N}, \tag{4.86}$$

$$\beta^{(t+1)} = \sum_{i=0}^{16} \frac{n_i t_i(\boldsymbol{\theta}^{(t)})}{N}, \tag{4.87}$$

$$\mu^{(t+1)} = \frac{\sum_{i=0}^{16} i\, n_i t_i(\boldsymbol{\theta}^{(t)})}{\sum_{i=0}^{16} n_i t_i(\boldsymbol{\theta}^{(t)})}, \tag{4.88}$$

$$\lambda^{(t+1)} = \frac{\sum_{i=0}^{16} i\, n_i p_i(\boldsymbol{\theta}^{(t)})}{\sum_{i=0}^{16} n_i p_i(\boldsymbol{\theta}^{(t)})}. \tag{4.89}$$

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

c. Estimate the standard errors and pairwise correlations of your parameter estimates, using any available method.

4.3. The website for this book contains 50 trivariate data points drawn from the $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Some data points have missing values in one or more coordinates. Only 27 of the 50 observations are complete.

a. Derive the EM algorithm for joint maximum likelihood estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It is easiest to recall that the multivariate normal density is in the exponential family.

b. Determine the MLEs from a suitable starting point. Investigate the performance of the algorithm, and comment on your results.

c. Consider Bayesian inference for $\boldsymbol{\mu}$ when

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.6 & 1.2 \\ 0.6 & 0.5 & 0.5 \\ 1.2 & 0.5 & 3.0 \end{pmatrix}$$

is known. Assume independent priors for the three elements of $\boldsymbol{\mu}$. Specifically, let the jth prior be

$$f(\mu_j) = \frac{\exp\{-(\mu_j - \alpha_j)/\beta_j\}}{\beta_j\left[1 + \exp\{-(\mu_j - \alpha_j)/\beta_j\}\right]^2},$$

where $(\alpha_1, \alpha_2, \alpha_3) = (2, 4, 6)$ and $\beta_j = 2$ for $j = 1, 2, 3$. Comment on difficulties that would be faced in implementing a standard EM algorithm for estimating the posterior mode for $\boldsymbol{\mu}$. Implement a gradient EM algorithm, and evaluate its performance.

d. Suppose that $\boldsymbol{\Sigma}$ is unknown in part (c) and that an improper uniform prior is adopted, that is, $f(\Sigma) \propto 1$ for all positive definite $\boldsymbol{\Sigma}$. Discuss ideas for how to estimate the posterior mode for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

4.4. Suppose we observe lifetimes for 14 gear couplings in certain mining equipment, as given in Table 4.3 (in years). Some of these data are right censored because the equipment was replaced before the gear coupling failed. The censored data are in parentheses; the actual lifetimes for these components may be viewed as missing.

Model these data with the Weibull distribution, having density function $f(x) = abx^{b-1}\exp\{-ax^b\}$ for $x > 0$ and parameters a and b. Recall that Problem 2.3 in Chapter 2 provides more details about such models. Construct an EM algorithm to estimate a and b. Since the Q function involves expectations that are analytically unavailable, adopt the MCEM strategy where necessary. Also, optimization of Q cannot be completed analytically. Therefore, incorporate the ECM strategy of conditionally maximizing with respect to each parameter separately, applying a one-dimensional