6.5.11. As discussed in Example 6.5.3, Z, (6.5.25), can be used as a test statistic provided we have a consistent estimator of $p_1(1-p_1)$ and $p_2(1-p_2)$ when H_0 is true. In the example, we discussed an estimator which is consistent under both H_0 and H_1. Under H_0, though, $p_1(1-p_1) = p_2(1-p_2) = p(1-p)$, where $p = p_1 = p_2$. Show that the statistic (6.5.22) is a consistent estimator of p, under H_0. Thus determine another test of H_0.

6.5.12. A machine shop that manufactures toggle levers has both a day and a night shift. A toggle lever is defective if a standard nut cannot be screwed onto the threads. Let p_1 and p_2 be the proportion of defective levers among those manufactured by the day and night shifts, respectively. We shall test the null hypothesis, $H_0 : p_1 = p_2$, against a two-sided alternative hypothesis based on two random samples, each of 1000 levers taken from the production of the respective shifts. Use the test statistic Z^* given in Example 6.5.3.

(a) Sketch a standard normal pdf illustrating the critical region having $\alpha = 0.05$.

(b) If $y_1 = 37$ and $y_2 = 53$ defectives were observed for the day and night shifts, respectively, calculate the value of the test statistic and the approximate p-value (note that this is a two-sided test). Locate the calculated test statistic on your figure in part (a) and state your conclusion. Obtain the approximate p-value of the test.

6.5.13. For the situation given in part (b) of Exercise 6.5.12, calculate the tests defined in Exercises 6.5.9 and 6.5.11. Obtain the approximate p-values of all three tests. Discuss the results.

6.6 The EM Algorithm

In practice, we are often in the situation where part of the data is missing. For example, we may be observing lifetimes of mechanical parts which have been put on test and some of these parts are still functioning when the statistical analysis is carried out. In this section, we introduce the EM algorithm, which frequently can be used in these situations to obtain maximum likelihood estimates. Our presentation is brief. For further information, the interested reader can consult the literature in this area, including the monograph by McLachlan and Krishnan (1997). Although, for convenience, we write in terms of continuous random variables, the theory in this section holds for the discrete case as well.

Suppose we consider a sample of n items, where n_1 of the items are observed, while $n_2 = n - n_1$ items are not observable. Denote the observed items by $\mathbf{X}' = (X_1, X_2, \ldots, X_{n_1})$ and unobserved items by $\mathbf{Z}' = (Z_1, Z_2, \ldots, Z_{n_2})$. Assume that the X_is are iid with pdf $f(x|\theta)$, where $\theta \in \Omega$. Assume that Z_js and the X_is are mutually independent. The conditional notation will prove useful here. Let $g(\mathbf{x}|\theta)$ denote the joint pdf of $\mathbf{X}$. Let $h(\mathbf{x}, \mathbf{z}|\theta)$ denote the joint pdf of the observed and unobserved items. Let $k(\mathbf{z}|\theta, \mathbf{x})$ denote the conditional pdf of the missing data given the observed data. By the definition of a conditional pdf, we have the identity

$$k(\mathbf{z}|\theta, \mathbf{x}) = \frac{h(\mathbf{x}, \mathbf{z}|\theta)}{g(\mathbf{x}|\theta)}. \tag{6.6.1}$$

The **observed likelihood** function is $L(\theta|\mathbf{x}) = g(\mathbf{x}|\theta)$. The **complete likelihood** function is defined by

$$L^c(\theta|\mathbf{x}, \mathbf{z}) = h(\mathbf{x}, \mathbf{z}|\theta). \tag{6.6.2}$$

Our goal is maximize the likelihood function $L(\theta|\mathbf{x})$ by using the complete likelihood $L^c(\theta|\mathbf{x}, \mathbf{z})$ in this process.

Using (6.6.1), we derive the following basic identity for an arbitrary but fixed $\theta_0 \in \Omega$:

$$\begin{aligned}
\log L(\theta|\mathbf{x}) &= \int \log L(\theta|\mathbf{x}) k(\mathbf{z}|\theta_0, \mathbf{x})\, d\mathbf{z} \\
&= \int \log g(\mathbf{x}|\theta) k(\mathbf{z}|\theta_0, \mathbf{x})\, d\mathbf{z} \\
&= \int [\log h(\mathbf{x}, \mathbf{z}|\theta) - \log k(\mathbf{z}|\theta, \mathbf{x})] k(\mathbf{z}|\theta_0, \mathbf{x})\, d\mathbf{z} \\
&= \int \log[h(\mathbf{x}, \mathbf{z}|\theta)] k(\mathbf{z}|\theta_0, \mathbf{x})\, d\mathbf{z} - \int \log[k(\mathbf{z}|\theta, \mathbf{x})] k(\mathbf{z}|\theta_0, \mathbf{x})\, d\mathbf{z} \\
&= E_{\theta_0}[\log L^c(\theta|\mathbf{x}, \mathbf{Z})|\theta_0, \mathbf{x}] - E_{\theta_0}[\log k(\mathbf{Z}|\theta, \mathbf{x})|\theta_0, \mathbf{x}], \qquad (6.6.3)
\end{aligned}$$

where the expectations are taken under the conditional pdf $k(\mathbf{z}|\theta_0, \mathbf{x})$. Define the first term on the right side of (6.6.3) to be the function

$$Q(\theta|\theta_0, \mathbf{x}) = E_{\theta_0}[\log L^c(\theta|\mathbf{x}, \mathbf{Z})|\theta_0, \mathbf{x}]. \tag{6.6.4}$$

The expectation which defines the function Q is called the E step of the EM algorithm. Recall that we want to maximize $\log L(\theta|\mathbf{x})$. As discussed below, we need only maximize $Q(\theta|\theta_0, \mathbf{x})$. This maximization is called the M step of the EM algorithm.

Denote by $\widehat{\theta}^{(0)}$ an initial estimate of θ, perhaps based on the observed likelihood. Let $\widehat{\theta}^{(1)}$ be the argument which maximizes $Q(\theta|\widehat{\theta}^{(0)}, \mathbf{x})$. This is the first-step estimate of θ. Proceeding this way, we obtain a sequence of estimates $\widehat{\theta}^{(m)}$. We formally define this algorithm as follows:

Algorithm 6.6.1 (EM Algorithm)**.** *Let $\widehat{\theta}^{(m)}$ denote the estimate on the mth step. To compute the estimate on the* $(m+1)$st *step, do*

1. *Expectation Step: Compute*

$$Q(\theta|\widehat{\theta}^{(m)}, \mathbf{x}) = E_{\widehat{\theta}^{(m)}}[\log L^c(\theta|\mathbf{x}, \mathbf{Z})|\widehat{\theta}_m, \mathbf{x}], \tag{6.6.5}$$

   where the expectation is taken under the conditional pdf $k(\mathbf{z}|\widehat{\theta}^{(m)}, \mathbf{x})$.

2. *Maximization Step: Let*

$$\widehat{\theta}^{(m+1)} = Argmax\, Q(\theta|\widehat{\theta}^{(m)}, \mathbf{x}). \tag{6.6.6}$$

Under strong assumptions, it can be shown that $\widehat{\theta}^{(m)}$ converges in probability to the maximum likelihood estimate, as $m \to \infty$. We will not show these results, but as the next theorem shows, $\widehat{\theta}^{(m+1)}$ always increases the likelihood over $\widehat{\theta}^{(m)}$.

Theorem 6.6.1. *The sequence of estimates $\widehat{\theta}^{(m)}$, defined by Algorithm 6.6.1, satisfies*

$$L(\widehat{\theta}^{(m+1)}|\mathbf{x}) \geq L(\widehat{\theta}^{(m)}|\mathbf{x}). \tag{6.6.7}$$

Proof: Because $\widehat{\theta}^{(m+1)}$ maximizes $Q(\theta|\widehat{\theta}^{(m)}, \mathbf{x})$, we have

$$Q(\widehat{\theta}^{(m+1)}|\widehat{\theta}^{(m)}, \mathbf{x}) \geq Q(\widehat{\theta}^{(m)}|\widehat{\theta}^{(m)}, \mathbf{x});$$

that is,

$$E_{\widehat{\theta}^{(m)}}[\log L^c(\widehat{\theta}^{(m+1)}|\mathbf{x}, \mathbf{Z})] \geq E_{\widehat{\theta}^{(m)}}[\log L^c(\widehat{\theta}^{(m)}|\mathbf{x}, \mathbf{Z})], \tag{6.6.8}$$

where the expectation is taken under the pdf $k(\mathbf{z}|\widehat{\theta}^{(m)}, \mathbf{x})$. By expression (6.6.3), we can complete the proof by showing that

$$E_{\widehat{\theta}^{(m)}}[\log k(\mathbf{Z}|\widehat{\theta}^{(m+1)}, \mathbf{x})] \leq E_{\widehat{\theta}^{(m)}}[\log k(\mathbf{Z}|\widehat{\theta}^{(m)}, \mathbf{x})]. \tag{6.6.9}$$

Keep in mind that these expectations are taken under the conditional pdf of $\mathbf{Z}$ given $\widehat{\theta}^{(m)}$ and $\mathbf{x}$. An application of Jensen's inequality, (1.10.5), yields

$$\begin{aligned}
E_{\widehat{\theta}^{(m)}}\left\{\log\left[\frac{k(\mathbf{Z}|\widehat{\theta}^{(m+1)}, \mathbf{x})}{k(\mathbf{Z}|\widehat{\theta}^{(m)}, \mathbf{x})}\right]\right\} &\leq \log E_{\widehat{\theta}^{(m)}}\left[\frac{k(\mathbf{Z}|\widehat{\theta}^{(m+1)}, \mathbf{x})}{k(\mathbf{Z}|\widehat{\theta}^{(m)}, \mathbf{x})}\right] \\
&= \log \int \frac{k(\mathbf{z}|\widehat{\theta}^{(m+1)}, \mathbf{x})}{k(\mathbf{z}|\widehat{\theta}^{(m)}, \mathbf{x})} k(\mathbf{z}|\widehat{\theta}^{(m)}, \mathbf{x})\, d\mathbf{z} \\
&= \log(1) = 0. \qquad (6.6.10)
\end{aligned}$$

This last result establishes (6.6.9) and, hence, finishes the proof. ∎

As an example, suppose $X_1, X_2, \ldots, X_{n_1}$ are iid with pdf $f(x - \theta)$, for $-\infty < x < \infty$, where $-\infty < \theta < \infty$. Denote the cdf of X_i by $F(x-\theta)$. Let $Z_1, Z_2, \ldots, Z_{n_2}$ denote the censored observations. For these observations, we only know that $Z_j > a$, for some a which is known, and that the Z_js are independent of the X_is. Then the observed and complete likelihoods are given by

$$\begin{aligned}
L(\theta|\mathbf{x}) &= [1 - F(a - \theta)]^{n_2} \prod_{i=1}^{n_1} f(x_i - \theta) \qquad (6.6.11) \\
L^c(\theta|\mathbf{x}, \mathbf{z}) &= \prod_{i=1}^{n_1} f(x_i - \theta) \prod_{i=1}^{n_2} f(z_i - \theta). \qquad (6.6.12)
\end{aligned}$$

By expression (6.6.1), the conditional distribution $\mathbf{Z}$ given $\mathbf{X}$ is the ratio of (6.6.12) to (6.6.11); that is,

$$\begin{aligned}
k(\mathbf{z}|\theta, \mathbf{x}) &= \frac{\prod_{i=1}^{n_1} f(x_i - \theta) \prod_{i=1}^{n_2} f(z_i - \theta)}{[1 - F(a - \theta)]^{n_2} \prod_{i=1}^{n_1} f(x_i - \theta)} \\
&= [1 - F(a - \theta)]^{-n_2} \prod_{i=1}^{n_2} f(z_i - \theta), \quad a < z_i, i = 1, \ldots, n_2. \qquad (6.6.13)
\end{aligned}$$

Thus, $\mathbf{Z}$ and $\mathbf{X}$ are independent, and $Z_1, \ldots, Z_{n_2}$ are iid with the common pdf $f(z-\theta)/[1-F(a-\theta)]$, for $z > a$. Based on these observations and expression (6.6.13), we have the following derivation:

$$\begin{aligned}
Q(\theta|\theta_0, \mathbf{x}) &= E_{\theta_0}[\log L^c(\theta|\mathbf{x}, \mathbf{Z})] \\
&= E_{\theta_0}\left[\sum_{i=1}^{n_1} \log f(x_i - \theta) + \sum_{i=1}^{n_2} \log f(Z_i - \theta)\right] \\
&= \sum_{i=1}^{n_1} \log f(x_i - \theta) + n_2 E_{\theta_0}[\log f(Z - \theta)] \\
&= \sum_{i=1}^{n_1} \log f(x_i - \theta) \\
&\quad + n_2 \int_a^\infty \log f(z - \theta) \frac{f(z - \theta_0)}{1 - F(a - \theta_0)}\, dz. \qquad (6.6.14)
\end{aligned}$$

This last result is the E step of the EM algorithm. For the M step, we need the partial derivative of $Q(\theta|\theta_0, \mathbf{x})$ with respect to θ. This is easily found to be

$$\frac{\partial Q}{\partial \theta} = -\left\{\sum_{i=1}^{n_1} \frac{f'(x_i - \theta)}{f(x_i - \theta)} + n_2 \int_a^\infty \frac{f'(z - \theta)}{f(z - \theta)} \frac{f(z - \theta_0)}{1 - F(a - \theta_0)}\, dz\right\}. \qquad (6.6.15)$$

Assuming that $\theta_0 = \widehat{\theta}_0$, the first-step EM estimate would be the value of θ, say $\widehat{\theta}^{(1)}$, which solves $\frac{\partial Q}{\partial \theta} = 0$. In the next example, we obtain the solution for a normal model.

Example 6.6.1. Assume the censoring model given above, but now assume that X has a $N(\theta, 1)$ distribution. Then $f(x) = \phi(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$. It is easy to show that $f'(x)/f(x) = -x$. Letting $\Phi(z)$ denote, as usual, the cdf of a standard normal random variable, by (6.6.15) the partial derivative of $Q(\theta|\theta_0, \mathbf{x})$ with respect to θ for this model simplifies to

$$\begin{aligned}
\frac{\partial Q}{\partial \theta} &= \sum_{i=1}^{n_1} (x_i - \theta) + n_2 \int_a^\infty (z - \theta) \frac{1}{\sqrt{2\pi}} \frac{\exp\{-(z - \theta_0)^2/2\}}{1 - \Phi(a - \theta_0)}\, dz \\
&= n_1(\overline{x} - \theta) + n_2 \int_a^\infty (z - \theta_0) \frac{1}{\sqrt{2\pi}} \frac{\exp\{-(z - \theta_0)^2/2\}}{1 - \Phi(a - \theta_0)}\, dz - n_2(\theta - \theta_0) \\
&= n_1(\overline{x} - \theta) + \frac{n_2}{1 - \Phi(a - \theta_0)} \phi(a - \theta_0) - n_2(\theta - \theta_0).
\end{aligned}$$

Solving $\partial Q/\partial \theta = 0$ for θ determines the EM step estimates. In particular, given that $\widehat{\theta}^{(m)}$ is the EM estimate on the mth step, the $(m+1)$st step estimate is

$$\widehat{\theta}^{(m+1)} = \frac{n_1}{n}\overline{x} + \frac{n_2}{n}\widehat{\theta}^{(m)} + \frac{n_2}{n} \frac{\phi(a - \widehat{\theta}^{(m)})}{1 - \Phi(a - \widehat{\theta}^{(m)})}, \qquad (6.6.16)$$

where $n = n_1 + n_2$. ■

For our second example, consider a mixture problem involving normal distributions. Suppose Y_1 has a $N(\mu_1, \sigma_1^2)$ distribution and Y_2 has a $N(\mu_2, \sigma_2^2)$ distribution. Let W be a Bernoulli random variable independent of Y_1 and Y_2 and with probability of success $\epsilon = P(W = 1)$. Suppose the random variable we observe is $X = (1 - W)Y_1 + WY_2$. In this case, the vector of parameters is given by $\boldsymbol{\theta}' = (\mu_1, \mu_2, \sigma_1, \sigma_2, \epsilon)$. As shown in Section 3.4, the pdf of the mixture random variable X is

$$f(x) = (1-\epsilon)f_1(x) + \epsilon f_2(x), \quad -\infty < x < \infty, \tag{6.6.17}$$

where $f_j(x) = \sigma_j^{-1}\phi[(x-\mu_j)/\sigma_j]$, $j = 1, 2$, and $\phi(z)$ is the pdf of a standard normal random variable. Suppose we observe a random sample $\mathbf{X}' = (X_1, X_2, \ldots, X_n)$ from this mixture distribution with pdf $f(x)$. Then the log of the likelihood function is

$$l(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log[(1-\epsilon)f_1(x_i) + \epsilon f_2(x_i)]. \tag{6.6.18}$$

In this mixture problem, the unobserved data are the random variables which identify the distribution membership. For $i = 1, 2, \ldots, n$, define the random variables

$$W_i = \begin{cases} 0 & \text{if } X_i \text{ has pdf } f_1(x) \\ 1 & \text{if } X_i \text{ has pdf } f_2(x). \end{cases}$$

These variables, of course, constitute the random sample on the Bernoulli random variable W. Accordingly, assume that $W_1, W_2, \ldots, W_n$ are iid Bernoulli random variables with probability of success ϵ. The complete likelihood function is

$$L^c(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w}) = \prod_{W_i=0} f_1(x_i) \prod_{W_i=1} f_2(x_i).$$

Hence the log of the complete likelihood function is

$$\begin{aligned} l^c(\boldsymbol{\theta}|\mathbf{x}, \mathbf{w}) &= \sum_{W_i=0} \log f_1(x_i) + \sum_{W_i=1} \log f_2(x_i) \\ &= \sum_{i=1}^{n} [(1-w_i)\log f_1(x_i) + w_i \log f_2(x_i)]. \end{aligned} \tag{6.6.19}$$

For the E step of the algorithm, we need the conditional expectation of W_i given $\mathbf{x}$ under $\boldsymbol{\theta}_0$; that is,

$$E_{\boldsymbol{\theta}_0}[W_i|\boldsymbol{\theta}_0, \mathbf{x}] = P[W_i = 1|\boldsymbol{\theta}_0, \mathbf{x}].$$

An estimate of this expectation is the likelihood of x_i being drawn from distribution $f_2(x)$, which is given by

$$\gamma_i = \frac{\widehat{\epsilon} f_{2,0}(x_i)}{(1-\widehat{\epsilon})f_{1,0}(x_i) + \widehat{\epsilon} f_{2,0}(x_i)}, \tag{6.6.20}$$

where the subscript 0 signifies that the parameters at $\boldsymbol{\theta}_0$ are being used. Expression (6.6.20) is intuitively evident; see McLachlan and Krishnan (1997) for more discussion. Replacing w_i by γ_i in expression (6.6.19), the M step of the algorithm is to maximize

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{x}) = \sum_{i=1}^{n}[(1-\gamma_i)\log f_1(x_i) + \gamma_i \log f_2(x_i)]. \tag{6.6.21}$$

This maximization is easy to obtain by taking partial derivatives of $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{x})$ with respect to the parameters. For example,

$$\frac{\partial Q}{\partial \mu_1} = \sum_{i=1}^{n}(1-\gamma_i)(-1/2\sigma_1^2)(-2)(x_i - \mu_1).$$

Setting this to 0 and solving for μ_1 yields the estimate of μ_1. The estimates of the other mean and the variances can be obtained similarly. These estimates are

$$\begin{aligned}
\widehat{\mu}_1 &= \frac{\sum_{i=1}^{n}(1-\gamma_i)x_i}{\sum_{i=1}^{n}(1-\gamma_i)} \\
\widehat{\sigma}_1^2 &= \frac{\sum_{i=1}^{n}(1-\gamma_i)(x_i - \widehat{\mu}_1)^2}{\sum_{i=1}^{n}(1-\gamma_i)} \\
\widehat{\mu}_2 &= \frac{\sum_{i=1}^{n}\gamma_i x_i}{\sum_{i=1}^{n}\gamma_i} \\
\widehat{\sigma}_2^2 &= \frac{\sum_{i=1}^{n}\gamma_i(x_i - \widehat{\mu}_2)^2}{\sum_{i=1}^{n}\gamma_i}.
\end{aligned}$$

Since γ_i is an estimate of $P[W_i = 1|\boldsymbol{\theta}_0, \mathbf{x}]$, the average $n^{-1}\sum_{i=1}^{n}\gamma_i$ is an estimate of $\epsilon = P[W_i = 1]$. This average is our estimate of $\widehat{\epsilon}$.

EXERCISES

6.6.1. Rao (page 368, 1973) considers a problem in the estimation of linkages in genetics. McLachlan and Krishnan (1997) also discuss this problem and we present their model. For our purposes, it can be described as a multinomial model with the four categories C_1, C_2, C_3, and C_4. For a sample of size n, let $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ denote the observed frequencies of the four categories. Hence, $n = \sum_{i=1}^{4} X_i$. The probability model is

C_1	C_2	C_3	C_4
$\frac{1}{2} + \frac{1}{4}\theta$	$\frac{1}{4} - \frac{1}{4}\theta$	$\frac{1}{4} - \frac{1}{4}\theta$	$\frac{1}{4}\theta$

where the parameter θ satisfies $0 \le \theta \le 1$. In this exercise, we obtain the mle of θ.

(a) Show that likelihood function is given by

$$L(\theta|\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!}\left[\frac{1}{2} + \frac{1}{4}\theta\right]^{x_1}\left[\frac{1}{4} - \frac{1}{4}\theta\right]^{x_2+x_3}\left[\frac{1}{4}\theta\right]^{x_4}. \tag{6.6.22}$$

(b) Show that the log of the likelihood function can be expressed as a constant (not involving parameters) plus the term

$$x_1 \log[2+\theta] + [x_2 + x_3] \log[1-\theta] + x_4 \log \theta.$$

(c) Obtain the partial derivative with respect to θ of the last expression, set the result to 0, and solve for the mle. (This will result in a quadratic equation which has one positive and one negative root.)

6.6.2. In this exercise, we set up an EM algorithm to determine the mle for the situation described in Exercise 6.6.1. Split category C_1 into the two subcategories C_{11} and C_{12} with probabilities 1/2 and $\theta/4$, respectively. Let Z_{11} and Z_{12} denote the respective "frequencies." Then $X_1 = Z_{11} + Z_{12}$. Of course, we cannot observe Z_{11} and Z_{12}. Let $\mathbf{Z} = (Z_{11}, Z_{12})'$.

(a) Obtain the complete likelihood $L^c(\theta|\mathbf{x}, \mathbf{z})$.

(b) Using the last result and (6.6.22), show that the conditional pmf $k(\mathbf{z}|\theta, \mathbf{x})$ is binomial with parameters x_1 and probability of success $\theta/(2+\theta)$.

(c) Obtain the E step of the EM algorithm given an initial estimate $\widehat{\theta}^{(0)}$ of θ. That is, obtain

$$Q(\theta|\widehat{\theta}^{(0)}, \mathbf{x}) = E_{\widehat{\theta}^{(0)}}[\log L^c(\theta|\mathbf{x}, \mathbf{Z})|\widehat{\theta}^{(0)}, \mathbf{x}].$$

Recall that this expectation is taken using the conditional pmf $k(\mathbf{z}|\widehat{\theta}^{(0)}, \mathbf{x})$. Keep in mind the next step; i.e., we need only terms that involve θ.

(d) For the M step of the EM algorithm, solve the equation $\partial Q(\theta|\widehat{\theta}^{(0)}, \mathbf{x})/\partial\theta = 0$. Show that the solution is

$$\widehat{\theta}^{(1)} = \frac{x_1\widehat{\theta}^{(0)} + 2x_4 + x_4\widehat{\theta}^{(0)}}{n\widehat{\theta}^{(0)} + 2(x_2 + x_3 + x_4)}. \tag{6.6.23}$$

6.6.3. For the setup of Exercise 6.6.2, show that the following estimator of θ is unbiased:

$$\widetilde{\theta} = n^{-1}(X_1 - X_2 - X_3 + X_4). \tag{6.6.24}$$

6.6.4. Rao (page 368, 1973) presents data for the situation described in Exercise 6.6.1. The observed frequencies are $\mathbf{x} = (125, 18, 20, 34)'$.

(a) Using computational packages (for example, R), with (6.6.24) as the initial estimate, write a program that obtains the stepwise EM estimates $\widehat{\theta}^{(k)}$.

(b) Using the data from Rao, compute the EM estimate of θ with your program. List the sequence of EM estimates, $\{\widehat{\theta}^k\}$, that you obtained. Did your sequence of estimates converge?

(c) Show that the mle using the likelihood approach in Exercise 6.6.1 is the positive root of the equation $197\theta^2 - 15\theta - 68 = 0$. Compare it with your EM solution. They should be the same within roundoff error.

6.6.5. Suppose $X_1, X_2, \ldots, X_{n_1}$ is a random sample from a $N(\theta, 1)$ distribution. Besides these n_1 observable items, suppose there are n_2 missing items, which we denote by $Z_1, Z_2, \ldots, Z_{n_2}$. Show that the first-step EM estimate is

$$\widehat{\theta}^{(1)} = \frac{n_1\overline{x} + n_2\widehat{\theta}^{(0)}}{n},$$

where $\widehat{\theta}^{(0)}$ is an initial estimate of θ and $n = n_1 + n_2$. Note that if $\widehat{\theta}^{(0)} = \overline{x}$, then $\widehat{\theta}^{(k)} = \overline{x}$ for all k.

6.6.6. Consider the situation described in Example 6.6.1. But suppose we have left censoring. That is, if $Z_1, Z_2, \ldots, Z_{n_2}$ are the censored items, then all we know is that each $Z_j < a$. Obtain the EM algorithm estimate of θ.

6.6.7. Suppose the following data follow the model of Example 6.6.1.

2.01	0.74	0.68	1.50^+	1.47	1.50^+	1.50^+	1.52
0.07	−0.04	−0.21	0.05	−0.09	0.67	0.14	

where the superscript $^+$ denotes that the observation was censored at 1.50. Write a computer program to obtain the EM algorithm estimate of θ.

6.6.8. The following data are observations of the random variable $X = (1-W)Y_1 + WY_2$, where W has a Bernoulli distribution with probability of success 0.70; Y_1 has a $N(100, 20^2)$ distribution; Y_2 has a $N(120, 25^2)$ distribution; W and Y_1 are independent; and W and Y_2 are independent.

119.0	96.0	146.2	138.6	143.4	98.2	124.5
114.1	136.2	136.4	184.8	79.8	151.9	114.2
145.7	95.9	97.3	136.4	109.2	103.2	

Program the EM algorithm for this mixing problem as discussed at the end of the section. Use a dotplot to obtain initial estimates of the parameters. Compute the estimates. How close are they to the true parameters?