Problem 7.16. (i) Denote the set of parameter by $\theta = (\pi, \mu_0, \mu_1, \sigma^2)$. The likelihood function is

$$L(\theta; Y) = \prod_{i=1}^{n} p(x_i; \pi) p(y_i \mid x_i; \mu_0, \mu_1, \sigma^2) \propto \prod_{i=1}^{n} \pi^{x_i}(1 - \pi)^{1-x_i}\sigma^{-1}exp[-\frac{(y_i - x_i\mu_1 - (1-x_i)\mu_0)^2}{2\sigma^2}].$$

Then the log-likelihood is

$$l(\theta; Y) \propto \sum_{i=1}^{n} x_i \log \frac{\pi}{1-\pi} - (n/2) \log \sigma^2 - \sum_{i=1}^{n} \frac{(y_i - x_i\mu_1 - (1-x_i)\mu_0)^2}{2\sigma^2}.$$

With complete data, the sufficient statistics are:

$$\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i y_i, \sum_{i=1}^{n} y_i, \text{ and } \sum_{i=1}^{n} y_i^2.$$

The MLE of $\theta$, $\hat{\theta}$, has components

$$\hat{\pi} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \hat{\mu_1} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i}, \quad \hat{\mu_0} = \frac{\sum_{i=1}^{n} (1-x_i) y_i}{\sum_{i=1}^{n} (1-x_i)}.$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i\hat{\mu_1} - (1-x_i)\hat{\mu_0})^2.$$

The marginal mean and variance of $Y$ can be derived easily

$$\mu_Y = E(Y) = E(E(Y \mid X)) = E(\mu_1 X + (1-X)\mu_0) = \pi \mu_1 + (1-\pi)\mu_0.$$  

$$\sigma_Y^2 = Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X)) = \sigma^2 + Var(\mu_1 X + (1-X)\mu_0) = \sigma^2 + (\mu_1 - \mu_0)^2 Var(X) = \sigma^2 + (\mu_1 - \mu_0)^2 \pi (1-\pi).$$

(ii) Without loss of generality, suppose the last $n - r$ values of $Y$ are missing. At E-step of the $(k+1)$th iteration, let $\theta^{(k)}$ be the current estimate. The complete-data sufficient statistics can be updated via:

$$E(y_i \mid x_i, \theta^{(k)}) = x_i\mu_1^{(k)} + (1-x_i)\mu_0^{(k)}.$$  

$$E(y_i^2 \mid x_i, \theta^{(k)}) = \sigma^{(k)2} + \{x_i\mu_1^{(k)} + (1-x_i)\mu_0^{(k)}\}^2, \quad \text{for } i \in \{r+1, \ldots, n\}.$$
The M-step can be carried as in part (i). Then the marginal mean and variance of $Y$ can be derived from $\hat{\theta}$.

(iii) The generating of posterior sample is similar to Example 6.16 on page 114 of the textbook. Without missing values, draw

$$(\pi | X, Y) \sim Beta\left(\sum_{i=1}^{n} x_i + \frac{3}{2}, \sum_{i=1}^{n} (1 - x_i) + \frac{3}{2}\right),$$

$$(\sigma^2 | X, Y) \sim Inv - \chi^2(n - 2, s^2),$$

$$(\mu_1 | \sigma^2, X, Y) \sim N\left(\hat{\mu}_1, \frac{\sigma^2}{\sum_{i=1}^{n} x_i}\right),$$

$$(\mu_0 | \sigma^2, X, Y) \sim N\left(\hat{\mu}_0, \frac{\sigma^2}{\sum_{i=1}^{n} (1 - x_i)}\right),$$

where $s^2 = n\hat{\sigma}^2/(n - 2)$ and $\mu_0$ and $\mu_1$ are drawn independently given $\sigma^2$.

With missing values, only the posterior distribution of the regression parameters are changed accordingly.

**Problem 8.6.** With $f(Y; \theta) \propto exp\{\theta S(Y) - a(\theta)\}$ and complete data, we know that $E(S(Y)) = a'(\theta)$ and the MLE is a solution of $S(Y) = a'(\theta)$. With missing data, the sufficient statistics $S(Y)$ are updated in the E-step. Then simply apply the procedure for calculating MLE under complete data.

**Problem 8.7.** When $g(t) = t^{-1}$, $Y$ belong to the exponential family. The complete-data sufficient statistics are: $\sum_{i=1}^{n} x_{ij}y_i$ and $\sum_{i=1}^{n} \log y_i$, $j = 1, \ldots, J$.

**Problem 8.8.** When $k$ is known, the complete-data sufficient statistics are $\sum_{i=1}^{n} x_{ij}y_i$, $j = 1, \ldots, J$; $\beta$ is a set of natural parameters as in Problem 8.6., then the M-step solve the following equation system:

$$E\left(\sum_{i=1}^{n} x_{ij}y_i \mid Y_{obs}, \beta^{(k)}\right) = E\left(\sum_{i=1}^{n} x_{ij}y_i\right) = \sum_{i=1}^{n} x_{ij}\mu_i = \sum_{i=1}^{n} x_{ij}(\sum_{h=1}^{J} x_{ih}/\beta_h)^{-1}, \quad j \in \{1, \ldots, J\};$$

where $E(\sum_{i=1}^{n} x_{ij}y_i \mid Y_{obs}, \beta^{(k)})$ should be calculated based on the censoring mechanism. For example, if $y_i$ is missing, then $E(y_i \mid y_i \text{ missing}) = E(y_i \mid y_i > c) = \frac{\int_{c}^{\infty} y_i p(y_i | x_i; \beta) dy_i}{\int_{c}^{\infty} p(y_i | x_i; \beta) dy_i}$. 

When $k$ is unknown, then similarly the M-step solve the above equations plus the following equation:

$$E(\sum_{i=1}^{n} \log y_i \mid Y_{obs}, \beta^{(k)}) = E(\sum_{i=1}^{n} \log y_i).$$

**Problem 8.12.** When the observed data is $(Y_1 + Y_3)$ with distribution $Bin(163 + 34, 1 - \theta/4)$, there is no need to use EM to estimate $\theta$. Simply $\hat{\theta} = 4 \frac{34}{197}$ based on standard results for a binomial distribution. The variance can be estimated accordingly. Apparently the variance estimate under for the collapsed data is larger than the one provided in the textbook because the data are collapsed in a way that partial information regarding $\theta$ is lost.