Computational Statistics
EM algorithm: variance estimation

We consider again the problem of estimating the parameters in a mixture of a normal distribution $\mathcal{N}(\mu, \sigma)$ and a uniform distribution $\mathcal{U}([-a, a])$, where a is a known constant. The observed data are an iid sample $w_1, \ldots, w_n$ from W with pdf

$$g(w; \theta) = \pi \phi(w; \mu, \sigma) + (1 - \pi)c, \tag{1}$$

where $\phi(\cdot; \mu, \sigma)$ is the normal pdf, $c = (2a)^{-1}$, π is the proportion of the normal distribution in the mixture and $\theta = (\mu, \sigma, \pi)^T$ is the vector of parameters. Typically, the uniform distribution corresponds to outliers in the data. The proportion of outliers in the population is then $1 - \pi$.

We have seen how to find the MLE $\widehat{\theta}$ of θ using the EM algorithm. We know want to estimate the variance of $\widehat{\theta}$.

1. Set $\theta = (0, 1, 0.9)$ and $a = 5$. Generate $N = 1000$ samples of size $n = 100$. For each sample, compute $\widehat{\theta}$ using the EM algorithm. Estimate the variance of $\widehat{\theta}$.

2. We now consider one sample $w_1, \ldots, w_n$ and we wish to estimate $\mathrm{Var}(\widehat{\theta})$ from that sample, without knowing the true value of θ. We will use two methods.

   (a) Louis' method: compute $\widehat{i}_{\mathbf{x}}(\widehat{\theta})$ and estimate $\widehat{i}_{\mathbf{z}|\mathbf{y}}(\widehat{\theta})$ by Monte Carlo simulation; compute an estimate of $\widehat{i}_{\mathbf{y}}(\widehat{\theta})$ using the missing information principle equation, and its inverse $\widehat{i}_{\mathbf{y}}(\widehat{\theta})^{-1}$.

   (b) Bootstrapping: generate $B = 1000$ bootstrap samples. Estimate $\mathrm{Var}(\widehat{\theta})$ by the sample variance of the bootstrap estimates of θ.

3. Compare the estimates of $\mathrm{Var}(\widehat{\theta})$ obtained by the different methods.