

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, \dots, w_n from W with pdf

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

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 $\hat{\theta}$ of θ using the EM algorithm. We now want to estimate the variance of $\hat{\theta}$.

1. Set $\theta = (0, 1, 0.9)$ and $a = 5$. Generate $N = 1000$ samples of size $n = 100$. For each sample, compute $\hat{\theta}$ using the EM algorithm. Estimate the variance of $\hat{\theta}$.
2. We now consider one sample w_1, \dots, w_n and we wish to estimate $\text{Var}(\hat{\theta})$ from that sample, without knowing the true value of θ . We will use three methods.
 - (a) Louis' method: estimate $\hat{i}_{\mathbf{x}}(\hat{\theta})$ and $\hat{i}_{\mathbf{z}|\mathbf{y}}(\hat{\theta})$ by Monte Carlo simulation; compute an estimate of $\hat{i}_{\mathbf{y}}(\hat{\theta})$ using the missing information principle equation, and its inverse $\hat{i}_{\mathbf{y}}(\hat{\theta})^{-1}$.
 - (b) SEM algorithm: compute analytically $\hat{i}_{\mathbf{x}}(\hat{\theta})$, and estimate $\Psi'(\hat{\theta})$ using the SEM algorithm. Then, compute $\hat{i}_{\mathbf{y}}(\hat{\theta})^{-1}$ using the formula

$$\hat{i}_{\mathbf{y}}(\hat{\theta})^{-1} = \hat{i}_{\mathbf{x}}(\hat{\theta})^{-1} \left[I + \Psi'(\hat{\theta})^T (I - \Psi'(\hat{\theta})^T)^{-1} \right]. \quad (2)$$

- (c) Bootstrapping: generate $B = 1000$ bootstrap samples. Estimate $\text{Var}(\hat{\theta})$ by the sample variance of the bootstrap estimates of $\hat{\theta}$.
3. Compare the estimates of $\text{Var}(\hat{\theta})$ obtained by the different methods.