MAT 653: Statistical Simulation

Instructor: Dr. Wei Li Scribe: Meng He

Oct 21th , 2021

EM Algorithms

E-step

$$Q(\theta \mid \theta^{(m)}, \overrightarrow{X}) = E_{\theta^{(m)}}\big[\log L^c(\theta \mid \overrightarrow{X}, \overrightarrow{Z}) \mid \overrightarrow{X}\big]$$

where $\overrightarrow{X}$ is observable, $\overrightarrow{Z}$ is unobservable; $E_{\theta^{(m)}}$ is the conditional expectation of Z given X, where the conditional density is given by $f(\overrightarrow{Z} \mid \overrightarrow{X}, \theta^{(m)})$;
$\log L^c(\theta \mid \overrightarrow{X}, \overrightarrow{Z})$ is complete data likelihood function.

M-step

maximize $\theta \to Q(\theta \mid \theta^{(m)}, \overrightarrow{X})$
let $\theta^{(m+1)} \leftarrow \arg\max_{\theta} \quad Q(\theta \mid \theta^{(m)}, \overrightarrow{X})$

Two component mixture of normals

Let $X_i \overset{iid}{\sim} \frac{1}{4}N(\mu_1, 1) + \frac{3}{4}N(\mu_2, 1)$, $\theta = (\mu_1, \mu_2)$
log-likelihood function:

$$\log L(\theta \mid \overrightarrow{X}) = \sum_{i=1}^{n} \log\big(\frac{1}{4}f(X_i \mid \mu_1) + \frac{3}{4}f(X_i \mid \mu_2)\big)$$

where $L(\theta \mid \overrightarrow{X}) = \prod_{i=1}^{n} f(X_i \mid \mu_1, \mu_2)$
The p.d.f of X_i is:

$$f(X_i \mid \mu_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i - \mu_j)^2}{2}}$$

Define some latent variables (binary):

$$P(Z_i = 1) = \frac{1}{4}, P(Z_i = 0) = \frac{3}{4}$$

where $Z_i \perp (W_{1i}, W_{2i})$

If $W_1 \sim f_1$, $W_2 \sim f_2$, $Z \sim Bernoulli(p)$ and $Z \perp (W_1, W_2)$, Z, W_1, W_2 are independent.

$$X = ZW_1 + (1 - Z)W_2$$

Complete data likelihood function $(\overrightarrow{X}, \overrightarrow{Z})$:

$$
\begin{aligned}
L^c = P(\overrightarrow{X}, \overrightarrow{Z}) &= \prod_{i=1}^{n} P(X_i, Z_i) \\
&= \prod_{i=1}^{n} \big[P(X_i \mid Z_i = 1) P(Z_i = 1) Z\big]^{Z_i} \big[P(X_i \mid Z_i = 0) P(Z_i = 0)\big]^{1-Z_i}
\end{aligned}
$$

So the log-likelihood function:

$$
\log\left(L^c(\theta \mid \overrightarrow{X}, \overrightarrow{Z})\right) = \sum_{i=1}^{n} \left[Z_i \log f(X_i \mid \mu_1) + Z_i \log \frac{1}{4} + (1 - Z_i) \log f(X_i \mid \mu_2) + (1 - Z_i) \log \frac{3}{4}\right]
$$

Let $\theta^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)})$

$$
\begin{aligned}
Q(\theta \mid \theta^{(0)}, \overrightarrow{X}) &= E_{\theta^{(0)}}\big[\log L^c(\theta \mid \overrightarrow{X}, \overrightarrow{Z}) \mid \overrightarrow{X}\big] \\
&= -\sum_{i=1}^{n} \Big\{\frac{1}{2} E_{\theta^{(0)}}[Z_i (X_i - \mu_1)^2 \mid X_i] + \frac{1}{2} E_{\theta^{(0)}}[(1 - Z_i)(X_i - \mu_2)^2 \mid X_i]\Big\} \\
&\quad + \sum_{i=1}^{n} \Big\{\big(\log(\frac{1}{4})\big) E_{\theta^{(0)}}(Z_i \mid X_i) - \big(\log(\frac{3}{4})\big) E_{\theta^{(0)}}(1 - Z_i \mid X_i)\Big\} + C
\end{aligned}
$$

where

$$
E_{\theta^{(0)}}[Z_i (X_i - \mu_1)^2 \mid X_i] = (X_i - \mu_1)^2 E_{\theta^{(0)}}[Z_i \mid X_i]
$$

we can calculate

$$
E[Z \mid X] = \frac{P(X \mid Z = 1) P(Z = 1)}{P(X = x \mid Z = 1) P(Z = 1) + P(X = x \mid Z = 0) P(Z = 0)}
$$

Proof:

$$
\begin{aligned}
E[Z \mid X] &= P(Z = 1 \mid X) \\
&= \frac{P(Z = 1, X)}{P(X = x)} \\
&= \frac{P(X \mid Z = 1) P(Z = 1)}{P(X = x)} \\
&= \frac{P(X \mid Z = 1) P(Z = 1)}{P(X = x \mid Z = 1) P(Z = 1) + P(X = x \mid Z = 0) P(Z = 0)}
\end{aligned}
$$

So we can calculate $E_{\theta^{(0)}}[Z_i (X_i - \mu_1)^2 \mid X_i]$:

$$
\begin{aligned}
E_{\theta^{(0)}}[Z_i (X_i - \mu_1)^2 \mid X_i] &= \frac{\frac{1}{4} f_1(X_i \mid \mu_1^{(0)})}{\frac{1}{4} f_1(X_i \mid \mu_1^{(0)}) + \frac{3}{4} f_2(X_i \mid \mu_2^{(0)})} \\
&= \alpha_i^{(0)}(X_i; \mu_1^{(0)}, \mu_2^{(0)})
\end{aligned}
$$

And $Q(\theta \mid \theta^{(0)}, \overrightarrow{X})$ can be expressed:

$$
Q(\theta \mid \theta^{(0)}, \overrightarrow{X}) = -\sum_{i=1}^{n} \Big\{\frac{1}{2} (X_i - \mu_1)^2 \alpha_i^{(0)}(X_i; \mu_1^{(0)}, \mu_2^{(0)}) + \frac{1}{2} (X_i - \mu_2)^2 (1 - \alpha_i^{(0)}(X_i; \mu_1^{(0)}, \mu_2^{(0)}))\Big\} + const
$$

Set the derivatives w.r.t. the θ to zero function:

$$\begin{cases} \frac{\partial Q(\theta|\theta^{(0)},\overline{X})}{\partial \mu_1} \overset{set}{=} 0 \\ \frac{\partial Q(\theta|\theta^{(0)},\overline{X})}{\partial \mu_2} \overset{set}{=} 0 \end{cases}$$

we can calculate the solutions $\mu_1^{(1)}$ and $\mu_2^{(1)}$

$$\begin{cases} \mu_1^{(1)} = \frac{\sum_{i=1}^{n} \alpha_i^{(0)}(X_i;\mu_1^{(0)},\mu_2^{(0)})X_i}{\sum_{i=1}^{n} \alpha_i^{(0)}(X_i;\mu_1^{(0)},\mu_2^{(0)})} \\ \mu_2^{(1)} = \frac{\sum_{i=1}^{n} \left[1-\alpha_i^{(0)}(X_i;\mu_1^{(0)},\mu_2^{(0)})\right]X_i}{\sum_{i=1}^{n} \left[1-\alpha_i^{(0)}(X_i;\mu_1^{(0)},\mu_2^{(0)})\right]} \end{cases}$$

Right censored data (R.C example 5.13 and 5.14)

Let $X_i \overset{iid}{\sim} f(x-\theta)$, where f is density function of $N(0,1)$, F be the CDF of $N(0,1)$, so $X_i \sim N(\theta,1)$
The Goal is to estimate θ. However, X_i are not fully observed. They are right censored. The actual observation are Y_i.
Let (1) Y_i is observed

$$Y_i = \begin{cases} a \quad if\ X_i \geq a \\ X_i \quad if\ X_i \leq a \end{cases}$$

where a is fixed.
It implies $Y_i = \min(X_i, a)$.
(2)$\delta_i = I(X_i \leq a)$ be the indicator for non-censoring, or equivalently, for observing the actual X_i;
(3)n be sample size.

Assume (1)$(Y_1, Y_2, \dots, Y_m)$ all less than a,
(2)$(Y_{m+1}, Y_{m+2}, \dots, Y_n)$ all equal than a.

So the observed data likelihood function:

$$\begin{aligned} L(\theta \mid Y_1, Y_2, \dots, Y_n, \delta_1, \delta_2, \dots, \delta_n) &= \prod_{i=1}^{n} \left(f(Y_i-\theta)\right)^{\delta_i} \left(P(Y_i = a)\right)^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left(f(Y_i-\theta)\right)^{\delta_i} \left(1 - P(X_i \leq a)\right)^{1-\delta_i} \\ &= \prod_{i=1}^{n} \left(f(Y_i-\theta)\right)^{\delta_i} \left(1 - F(a-\theta)\right)^{1-\delta_i} \end{aligned}$$

Since $(Y_1, Y_2, \dots, Y_m)$ are uncensored, $(Y_{m+1}, Y_{m+2}, \dots, Y_n)$ are censored.
Let $\overline{Z}$ be the vector of the unobservable $X_{m+1}, X_{m+2}, \dots, X_n$.
The complete data likelihood function:

$$L^c(\theta \mid Y_1, Y_2, \dots, Y_m; \overline{Z}) = \prod_{i=1}^{m} f(Y_i-\theta) \prod_{i=m+1}^{n} f(X_i-\theta)$$

The log-likelihood function:

$$\log L^c(\theta \mid Y_1, Y_2, \dots, Y_m; \overrightarrow{Z}) = \sum_{i=1}^{m} \log f(Y_i - \theta) + \sum_{i=m+1}^{n} \log f(X_i - \theta)$$

$Q(\theta \mid \theta^{(0)}, \overrightarrow{X})$ can be expressed:

$$\begin{aligned} Q(\theta \mid \theta^{(0)}; Y_1, \dots, Y_n, \delta_1, \dots, \delta_n) &= E_{\theta^{(0)}}\left[\log L^c(\theta \mid X_1, \dots, X_n) \mid Y_1, \dots, Y_n, \delta_1, \dots, \delta_n\right] \\ &= -\frac{1}{2}\sum_{i=1}^{m}(Y_i - \theta)^2 - \frac{1}{2}\sum_{i=m+1}^{n} E_{\theta^{(0)}}\left[(Z_i - \theta)^2 \mid Y_1, \dots, Y_n, \delta_1, \dots, \delta_n\right] \end{aligned}$$

For those $i = m+1, \dots, n$, $\delta_i = 0$, So

$$E_{\theta^{(0)}}\left[(Z_i - \theta)^2 \mid Y_i, \delta_i\right] = E_{\theta^{(0)}}\left[(X_i - \theta)^2 \mid X_i \geq a\right]$$

We can calculate $\theta^{(1)}$ by:

$$\frac{\partial Q(\theta \mid \theta^{(0)}; Y_1, \dots, Y_n, \delta_1, \dots, \delta_n)}{\partial \theta} \overset{set}{=} 0$$

Which can be expressed:

$$\sum_{i=1}^{m}(Y_i - \theta) + \sum_{i=m+1}^{n} E_{\theta^{(0)}}\left[(X_i - \theta) \mid X_i \geq a\right] \overset{set}{=} 0$$

$$\sum_{i=1}^{m} Y_i - m\theta + \sum_{i=m+1}^{n} E_{\theta^{(0)}}[X_i \mid X_i \geq a] - (n-m)\theta \overset{set}{=} 0$$

we can calculate the solution $\theta^{(1)}$:

$$\theta^{(1)} = \frac{\sum_{i=1}^{m} Y_i + (n-m)E_{\theta^{(0)}}[X_i \mid X_i \geq a]}{n}$$

Finally, we can show that

$$E_{\theta^{(0)}}[X_i \mid X_i \geq a] = \theta^{(0)} + \frac{\phi(a - \theta^{(0)})}{1 - \Phi(a - \theta^{(0)})}$$

Proof: The conditional density of X_i given $X_i \geq a$, $X_i \sim f(x - \theta)$, $f \sim N(0,1)$

$$f(X_i = x \mid X_i \geq a) = \frac{\phi(x - \theta)}{1 - \Phi(a - \theta)} \qquad x > a$$

The last thing is to compute the expectation.

$$
\begin{aligned}
E_{\theta^{(0)}}[X_i \mid X_i \geq a] &= \int_a^\infty x \frac{\phi(x-\theta^{(0)})}{1-\Phi(a-\theta^{(0)})}\,dx \\
&= \int_a^\infty (\theta^{(0)} + x - \theta^{(0)}) \frac{\phi(x-\theta^{(0)})}{1-\Phi(a-\theta^{(0)})}\,dx \\
&= \theta^{(0)} + \int_a^\infty (x-\theta^{(0)}) \frac{\phi(x-\theta^{(0)})}{1-\Phi(a-\theta^{(0)})}\,dx \\
&\overset{w=x-\theta^{(0)}}{=} \theta^{(0)} + \frac{\int_{a-\theta^{(0)}}^\infty w\phi(w)\,dw}{1-\Phi(a-\theta^{(0)})} \\
&= \theta^{(0)} + \frac{-\phi(w)\Big|_{a-\theta^{(0)}}^\infty}{1-\Phi(a-\theta^{(0)})} \\
&= \theta^{(0)} + \frac{\phi(a-\theta^{(0)})}{1-\Phi(a-\theta^{(0)})}
\end{aligned}
$$