# BIOS 760 Final, 2012: Solution

1. (a) Using the change of variable $v = x - \theta$, we obtain that

$$\int_R f_\theta(x)dx = \int_\theta^\infty e^{-(x-\theta)}dx = \int_0^\infty e^{-v}dv = 1.$$

The full likelihood has the form

$$\begin{aligned}\prod_{i=1}^n f_\theta(X_i) &= \exp\left[\sum_{i=1}^n X_i\right] e^{n\theta} 1\{X_{(1)} > \theta\} \\ &= a(X_1, \ldots, X_n) b(\theta, X_{(1)}),\end{aligned}$$

and thus $X_{(1)}$ is sufficient by factorization.

(b) The density of the smallest order statistic in an i.i.d. sample with density $f_\theta(x)$, where we let $F_\theta(x)$ be the corresponding c.d.f., is

$$n f_\theta(x)(1 - F_\theta(x))^{n-1} = ne^{-(x-\theta)} \exp[-(n-1)(x-\theta)],$$

and the given form of the density follows. Now let $\gamma(x)$ be an arbitrary measurable integrable function, and note that for any $\theta \in R$,

$$\int_R \gamma(x) g_\theta(x)dx = 0 \;\Rightarrow\; \int_\theta^\infty \gamma(x) ne^{-n(x-\theta)}dx = 0,$$

which implies that

$$\int_\theta^\infty \gamma(x)e^{-nx}dx = 0 \;\Rightarrow\; \gamma(\theta)e^{-n\theta} = 0,$$

where the last equality follows from differentiating both sides of the left-hand equality with respect to $\theta$. This now implies that $\gamma(\theta) = 0$ for all $\theta \in R$. Thus $X_{(1)}$ is complete since $\gamma(x)$ was arbitrary.

(c) The expectation of $X_{(1)}$ is

$$\begin{aligned}\int_\theta^\infty xne^{-n(x-\theta)}dx &= \theta + \int_R n(x-\theta)e^{-n(x-\theta)}dx \\ &= \theta + n^{-1}\int_0^\infty ve^{-v}dv \\ &= \theta + n^{-1},\end{aligned}$$

where the second-to-last equality uses the change of variables $v = n(x - \theta)$. Thus, since $X_{(1)}$ is complete and sufficient for $\theta$, the UMVU of $\theta$ is $X_{(1)} - n^{-1}$.

2. (a) Since the log-likelihood for a single observation is $\ell(p) = \log[pU + (1-p)(1-U)]$, the corresponding score and information for the sample are

$$\dot{\ell}_n(p) = \sum_{i=1}^{n} \frac{2U_i - 1}{pU_i + (1-p)(1-U_i)}$$

and

$$I_n(p) = -\ddot{\ell}_n(p) = \sum_{i=1}^{n} \frac{(2U_i-1)^2}{[pU_i + (1-p)(1-U_i)]^2}.$$

(b) The result follows directly from the standard Newton-Raphson iteration formula for the sample log-likelihood:

$$p^{(k+1)} = p^{(k)} + \left[I_n(p^{(k)})\right]^{-1} \dot{\ell}_n(p^{(k)}).$$

(c) The form of $m_p(\delta, u)$ follows directly from the mixture structure. The marginal density for $U$ is obtained by summing over $\delta$ (corresponding to the $\Delta$ component) which yields that the marginal is $h_p(u)$. The form of the sample log-likelihood follows directly from the log-likelihood for a single observation obtained from $m_p(\delta, u)$, which is

$$\log(2) + \Delta\left[\log(p) + \log(U)\right] + (1-\Delta)\left[\log(1-p) + \log(1-U)\right].$$

(d) The expectation of $\Delta_i$, given $U_i$ at the value $p = p^{(k)}$, is the same as the probability that $\Delta_i = 1$, given $U_i$ evaluated at $p = p^{(k)}$, which is

$$\frac{m_{p^{(k)}}(1, U_i)}{\sum_{\delta=0,1} m_{p^{(k)}}(\delta, U_i)} = \frac{p^{(k)} U_i}{p^{(k)} U_i + (1-p^{(k)})(1-U_i)}.$$

(e) The expectation of the full sample log-likelihood, with individual components having the from in (c) above, given the observed data, after throwing out the constant $\log(2)$, has the form

$$\begin{aligned}
&E\left[\sum_{i=1}^{n} \Delta_i\left[\log(p) + \log(U_i)\right] + (1-\Delta_i)\left[\log(1-p) + \log(1-U_i)\right]\Bigg| Y_{obs}, p^{(k)}\right] \\
&= \sum_{i=1}^{n} E\left\{\Delta_i\left[\log(p) + \log(U_i)\right] + (1-\Delta_i)\left[\log(1-p) + \log(1-U_i)\right]\middle| U_i, p^{(k)}\right\} \\
&= \sum_{i=1}^{n} W_k(U_i)\left[\log(p) + \log(U_i)\right] + (1-W_k(U_i))\left[\log(1-p) + \log(1-U_i)\right],
\end{aligned}$$

where the second equality follows from the independence across observation. Now differentiate the last term above with respect to $p$, to obtain that the next EM update $p^{(k+1)}$ is the solution of

$$\sum_{i=1}^{n} \left[ \frac{W_k(U_i)}{p} - \frac{1 - W_k(U_i)}{1 - p} \right] = 0,$$

and the desired conclusion follows.

(f) In the EM algorithm, if $p^{(k)} = 0$, then $p^{(k+1)} = 0$ and the EM iteration is stuck at zero. For the Newton-Raphson iteration, $p^{(k+1)}$ is almost surely not zero if $p^{(k)} = 0$. This says that the EM has faulty local solutions which need to be avoided or worked around in the estimation process. The Newton-Raphson algorithm appears to not have this problem.