Stat 710: Mathematical Statistics
Lecture 11

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

We need to use various methods to derive MLE's.

Example 4.32

Let X be an observation from the hypergeometric distribution $HG(r,n,\theta-n)$ (Table 1.1, page 18) with known r, n, and an unknown $\theta = n+1, n+2, \ldots$.

In this case, the likelihood function is defined on integers and the method of using the likelihood equation is certainly not applicable.

Note that

$$\frac{\ell(\theta)}{\ell(\theta-1)} = \frac{(\theta-r)(\theta-n)}{\theta(\theta-n-r+x)},$$

which is larger than 1 if and only if $\theta < rn/x$ and is smaller than 1 if and only if $\theta > rn/x$.

Thus, $\ell(\theta)$ has a maximum $\theta =$ the integer part of rn/x, which is the MLE of θ.

We need to use various methods to derive MLE's.

Example 4.32

Let X be an observation from the hypergeometric distribution $HG(r,n,\theta-n)$ (Table 1.1, page 18) with known r, n, and an unknown $\theta = n+1, n+2, \ldots$.

In this case, the likelihood function is defined on integers and the method of using the likelihood equation is certainly not applicable.

Note that

$$\frac{\ell(\theta)}{\ell(\theta-1)} = \frac{(\theta-r)(\theta-n)}{\theta(\theta-n-r+x)},$$

which is larger than 1 if and only if $\theta < rn/x$ and is smaller than 1 if and only if $\theta > rn/x$.

Thus, $\ell(\theta)$ has a maximum $\theta =$ the integer part of rn/x, which is the MLE of θ.

In applications, MLE's typically do not have analytic forms and some numerical methods have to be used to compute MLE's.
But first, we may verify whether an MLE exists and whether it is unique

Example 4.33

Let $X_1, \dots, X_n$ be i.i.d. from the gamma distribution $\Gamma(\alpha, \gamma)$ with unknown $\alpha > 0$ and $\gamma > 0$.
The log-likelihood function is

$$\log \ell(\theta) = -n\alpha \log \gamma - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \log x_i - \frac{1}{\gamma} \sum_{i=1}^n x_i$$

and the likelihood equation is

$$-n \log \gamma - \frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^n \log x_i = 0$$

and

$$-\frac{n\alpha}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^n x_i = 0.$$

In applications, MLE's typically do not have analytic forms and some numerical methods have to be used to compute MLE's.
But first, we may verify whether an MLE exists and whether it is unique

Example 4.33

Let $X_1,...,X_n$ be i.i.d. from the gamma distribution $\Gamma(\alpha,\gamma)$ with unknown $\alpha > 0$ and $\gamma > 0$.
The log-likelihood function is

$$\log \ell(\theta) = -n\alpha \log \gamma - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \frac{1}{\gamma} \sum_{i=1}^{n} x_i$$

and the likelihood equation is

$$-n \log \gamma - \frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} \log x_i = 0$$

and

$$-\frac{n\alpha}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^{n} x_i = 0.$$

Example 4.33 (continued)

The second equation yields $\gamma = \bar{x}/\alpha$.
Substituting $\gamma = \bar{x}/\alpha$ into the first equation we obtain that

$$\log\alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \frac{1}{n}\sum_{i=1}^{n}\log x_i - \log\bar{x} = 0.$$

This equation does not have an explicit solution.
A numerical method has to be applied to compute the MLE for any given observations $x_1, ..., x_n$.

We now show that a solution exists a.s. and it is the unique MLE.
Define

$$h(\alpha) = \log\alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + Y - \log\bar{X},$$

where $Y = n^{-1}\sum_{i=1}^{n}\log X_i$
We show that $h(\alpha) = 0$ has a solution a.s. and it is the unique MLE.
Let C be the Euler constant defined as

$$C = \lim_{m\to\infty}\left(\sum_{k=0}^{m-1}\frac{1}{k+1} - \log m\right).$$

From calculus,

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = -C + \sum_{k=0}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+\alpha}\right)$$

and

$$\frac{d}{d\alpha}\left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right] = \sum_{k=0}^{\infty}\frac{1}{(k+\alpha)^2},$$

Then

$$\begin{aligned}
h'(\alpha) &= \frac{1}{\alpha} - \sum_{k=0}^{\infty}\frac{1}{(k+\alpha)^2} \\
&< \frac{1}{\alpha} - \sum_{k=0}^{\infty}\left(\frac{1}{k+\alpha} - \frac{1}{k+1+\alpha}\right) \\
&= \frac{1}{\alpha} + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - \frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} \\
&= \frac{1}{\alpha} - \frac{d}{d\alpha}\log\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
&= \frac{1}{\alpha} - \frac{d}{d\alpha}\log\alpha = 0.
\end{aligned}$$

Example 4.33 (continued)

Hence, $h(\alpha)$ is decreasing.

Also, it follows from the last two equalities of the previous expression that, for $m = 2, 3, ...,$

$$\frac{\Gamma'(m)}{\Gamma(m)} = \frac{1}{m-1} + \frac{1}{m-2} + \cdots + 1 + \frac{\Gamma'(1)}{\Gamma(1)} = \sum_{k=0}^{m-2} \frac{1}{k+1} - C.$$

Therefore,

$$\lim_{m\to\infty} \left[\log m - \frac{\Gamma'(m)}{\Gamma(m)} \right] = \lim_{m\to\infty} \left[\log m - \sum_{k=0}^{m-2} \frac{1}{k+1} + C \right] = 0$$

by the definition of C.

Hence, $\lim_{\alpha\to\infty} h(\alpha) = Y - \log \bar{X}$, which is negative by Jensen's inequality when X_i's are not all the same.

Thus, $\lim_{\alpha\to\infty} h(\alpha) < 0$ a.s.

Example 4.33 (continued)

Since

$$\begin{aligned}
\lim_{\alpha\to 0}\left[\log\alpha - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right] &= \lim_{\alpha\to 0}\left[\log\alpha + C - \sum_{k=0}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+\alpha}\right)\right] \\
&= \lim_{\alpha\to 0}\left[\log\alpha + C + \frac{1}{\alpha} - 1 + \sum_{k=1}^{\infty}\frac{1-\alpha}{(k+1)(k+\alpha)}\right] \\
&= \lim_{\alpha\to 0}\left(\log\alpha + \frac{1}{\alpha}\right) + C - 1 + \sum_{k=1}^{\infty}\frac{1}{(k+1)k} \\
&= \infty,
\end{aligned}$$

we have $\lim_{\alpha\to 0} h(\alpha) = \infty$.

Since h is continuous and decreasing, $h(\alpha) = 0$ has a unique solution a.s.

Thus, the likelihood equations have a unique solution a.s., which is the MLE of θ.

The Newton-Raphson method

A commonly used numerical method is the Newton-Raphson iteration method, which repeatedly computes

$$\widehat{\theta}^{(t+1)} = \widehat{\theta}^{(t)} - \left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau} \bigg|_{\theta=\widehat{\theta}^{(t)}} \right]^{-1} \frac{\partial \log \ell(\theta)}{\partial\theta} \bigg|_{\theta=\widehat{\theta}^{(t)}},$$

$t = 0, 1, ...$, where $\widehat{\theta}^{(0)}$ is an initial value and $\partial^2 \log \ell(\theta)/\partial\theta\partial\theta^\tau$ is assumed of full rank for every $\theta \in \Theta$.

If, at each iteration, we replace

$$\left[\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau} \bigg|_{\theta=\widehat{\theta}^{(t)}} \right]^{-1}$$

by

$$\left[\left\{ E\left(\frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau} \right) \right\} \bigg|_{\theta=\widehat{\theta}^{(t)}} \right]^{-1},$$

where the expectation is taken under P_θ, then the method is known as the Fisher-scoring method.

If the iteration converges, then $\widehat{\theta}^{(\infty)}$ or $\widehat{\theta}^{(t)}$ with a sufficiently large t is a numerical approximation to a solution of the likelihood equation.

Example 4.33 (continued)

In Example 4.33, let

$$s(\theta) = \frac{\partial \log \ell(\theta)}{\partial \theta} = n\left(-\log\gamma - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + Y, -\frac{\alpha}{\gamma} + \frac{\bar{X}}{\gamma^2}\right),$$

$$R(\theta) = \frac{\partial^2 \log \ell(\theta)}{\partial\theta\partial\theta^\tau} = n\begin{pmatrix} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right]^2 - \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} & -\frac{1}{\gamma} \\ -\frac{1}{\gamma} & \frac{\alpha}{\gamma^2} - \frac{2\bar{X}}{\gamma^3} \end{pmatrix},$$

and

$$F(\theta) = E[R(\theta)] = n\begin{pmatrix} \left[\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right]^2 - \frac{\Gamma''(\alpha)}{\Gamma(\alpha)} & -\frac{1}{\gamma} \\ -\frac{1}{\gamma} & -\frac{\alpha}{\gamma^2} \end{pmatrix}.$$

Then the Newton-Raphson iteration equation is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [R(\hat{\theta}^{(k)})]^{-1} s(\hat{\theta}^{(k)}), \quad k = 0, 1, 2, ...$$

and the Fisher-scoring iteration equation is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [F(\hat{\theta}^{(k)})]^{-1} s(\hat{\theta}^{(k)}), \quad k = 0, 1, 2, ...$$

Example 4.34 (Application of MCMC)

Let X be a random k-vector from P_θ with the following p.d.f. w.r.t. a σ-finite measure ν:

$$f_\theta(x) = \int f_\theta(x,y) d\nu(y),$$

where $f_\theta(x,y)$ is a joint p.d.f. w.r.t. $\nu \times \nu$.
This type of distribution is called a *mixture* distribution.
Thus, the likelihood $\ell(\theta) = f_\theta(x)$ involves a k-dimensional integral.
In many cases this integral has to be computed in order to compute an MLE of θ.
Let $\tilde{\ell}_m(\theta)$ be the MCMC approximation to $\ell(\theta)$ based on one of the MCMC methods described in §4.1.4 and a Markov chain of length m.
Under the some conditions (Theorem 4.4), $\tilde{\ell}_m(\theta) \to_{a.s.} \ell(\theta)$ for every fixed θ and x.
Suppose that, for each m, there exists $\tilde{\theta}_m$ that maximizes $\tilde{\ell}_m(\theta)$ over $\theta \in \Theta$.
Geyer (1994) studies the convergence of $\tilde{\theta}_m$ to an MLE.