Stat 710: Mathematical Statistics
Lecture 12

Jun Shao

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

Lecture 12: MLE in generalized linear models (GLM) and Quasi-MLE

MLE in exponential families

Suppose that X has a distribution from a natural exponential family so that the likelihood function is

$$\ell(\eta) = \exp\{\eta^\tau T(x) - \zeta(\eta)\} h(x),$$

where $\eta \in \Xi$ is a vector of unknown parameters.
The likelihood equation is then

$$\frac{\partial \log \ell(\eta)}{\partial \eta} = T(x) - \frac{\partial \zeta(\eta)}{\partial \eta} = 0,$$

which has a unique solution $T(x) = \partial \zeta(\eta)/\partial \eta$, assuming that $T(x)$ is in the range of $\partial \zeta(\eta)/\partial \eta$.
Note that

$$\frac{\partial^2 \log \ell(\eta)}{\partial \eta \partial \eta^\tau} = -\frac{\partial^2 \zeta(\eta)}{\partial \eta \partial \eta^\tau} = -\,\mathrm{Var}(T)$$

(see the proof of Proposition 3.2).

MLE in exponential families

Since $\mathrm{Var}(T)$ is positive definite, $-\log \ell(\eta)$ is convex in η and $T(x)$ is the unique MLE of the parameter $\mu(\eta) = \partial \zeta(\eta)/\partial \eta$.
Also, the function $\mu(\eta)$ is one-to-one so that μ^{-1} exists.
By Definition 4.3, the MLE of η is $\widehat{\eta} = \mu^{-1}(T(x))$.

If the distribution of X is in a general exponential family and the likelihood function is

$$\ell(\theta) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\} h(x),$$

then the MLE of θ is $\widehat{\theta} = \eta^{-1}(\widehat{\eta})$, if η^{-1} exists and $\widehat{\eta}$ is in the range of $\eta(\theta)$.
Of course, $\widehat{\theta}$ is also the solution of the likelihood equation

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{\partial \eta(\theta)}{\partial \theta} T(x) - \frac{\partial \xi(\theta)}{\partial \theta} = 0.$$

The results for exponential families lead to an estimation method in a class of models that have very wide applications.

MLE in exponential families

Since $\mathrm{Var}(T)$ is positive definite, $-\log \ell(\eta)$ is convex in η and $T(x)$ is the unique MLE of the parameter $\mu(\eta) = \partial \zeta(\eta)/\partial \eta$.
Also, the function $\mu(\eta)$ is one-to-one so that μ^{-1} exists.
By Definition 4.3, the MLE of η is $\widehat{\eta} = \mu^{-1}(T(x))$.

If the distribution of X is in a general exponential family and the likelihood function is

$$\ell(\theta) = \exp\{[\eta(\theta)]^\tau T(x) - \xi(\theta)\} h(x),$$

then the MLE of θ is $\widehat{\theta} = \eta^{-1}(\widehat{\eta})$, if η^{-1} exists and $\widehat{\eta}$ is in the range of $\eta(\theta)$.
Of course, $\widehat{\theta}$ is also the solution of the likelihood equation

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \frac{\partial \eta(\theta)}{\partial \theta} T(x) - \frac{\partial \xi(\theta)}{\partial \theta} = 0.$$

The results for exponential families lead to an estimation method in a class of models that have very wide applications.

GLM

The GLM is a generalization of the normal linear model discussed in §3.3.1-§3.3.2.

The GLM is useful since it covers situations where the relationship between $E(X_i)$ and Z_i is nonlinear and/or X_i's are discrete.

The structure of a GLM

The sample $X = (X_1, ..., X_n)$ has independent X_i's and X_i has the p.d.f.

$$\exp\left\{\frac{\eta_i x_i - \zeta(\eta_i)}{\phi_i}\right\} h(x_i, \phi_i), \qquad i = 1, ..., n,$$

w.r.t. a σ-finite measure ν, where η_i and ϕ_i are unknown, $\phi_i > 0$,

$$\eta_i \in \Xi = \left\{\eta : \ 0 < \int h(x, \phi) e^{\eta x/\phi} d\nu(x) < \infty\right\} \subset \mathscr{R}$$

for all i, ζ and h are known functions, and $\zeta''(\eta) > 0$ is assumed for all $\eta \in \Xi^\circ$, the interior of Ξ.

Note that the p.d.f. belongs to an exponential family if ϕ_i is known.

As a consequence,

$$E(X_i) = \zeta'(\eta_i) \quad \text{and} \quad \mathrm{Var}(X_i) = \phi_i \zeta''(\eta_i), \qquad i = 1, ..., n.$$

GLM

The GLM is a generalization of the normal linear model discussed in §3.3.1-§3.3.2.
The GLM is useful since it covers situations where the relationship between $E(X_i)$ and Z_i is nonlinear and/or X_i's are discrete.

The structure of a GLM

The sample $X = (X_1,...,X_n)$ has independent X_i's and X_i has the p.d.f.

$$\exp\left\{\frac{\eta_i x_i - \zeta(\eta_i)}{\phi_i}\right\} h(x_i,\phi_i), \qquad i = 1,...,n,$$

w.r.t. a σ-finite measure ν, where η_i and ϕ_i are unknown, $\phi_i > 0$,

$$\eta_i \in \Xi = \left\{\eta : \ 0 < \textstyle\int h(x,\phi)e^{\eta x/\phi} d\nu(x) < \infty\right\} \subset \mathscr{R}$$

for all i, ζ and h are known functions, and $\zeta''(\eta) > 0$ is assumed for all $\eta \in \Xi^\circ$, the interior of Ξ.
Note that the p.d.f. belongs to an exponential family if ϕ_i is known.
As a consequence,

$$E(X_i) = \zeta'(\eta_i) \quad \text{and} \quad \mathrm{Var}(X_i) = \phi_i \zeta''(\eta_i), \qquad i = 1,...,n.$$

The structure of a GLM

Define $\mu(\eta) = \zeta'(\eta)$.
It is assumed that η_i is related to Z_i, the ith value of a p-vector of covariates, through

$$g(\mu(\eta_i)) = \beta^\tau Z_i, \qquad i = 1, ..., n,$$

where β is a p-vector of unknown parameters and g, called a *link function*, is a known one-to-one, third-order continuously differentiable function on $\{\mu(\eta) : \eta \in \Xi^\circ\}$.
If $\mu = g^{-1}$, then $\eta_i = \beta^\tau Z_i$ and g is called the *canonical* or *natural* link function.
If g is not canonical, we assume that $\frac{d}{d\eta}(g \circ \mu)(\eta) \neq 0$ for all η.
In a GLM, the parameter of interest is β.
We assume that the range of β is
$B = \{\beta : (g \circ \mu)^{-1}(\beta^\tau z) \in \Xi^\circ \text{ for all } z \in \mathscr{Z}\}$,
where $\mathscr{Z}$ is the range of Z_i's.
ϕ_i's are called *dispersion* parameters and are considered to be nuisance parameters.

MLE in GLM

An MLE of β in a GLM is considered under assumption

$$\phi_i = \phi/t_i, \qquad i = 1,...,n,$$

with an unknown $\phi > 0$ and known positive t_i's.
Let $\theta = (\beta, \phi)$ and $\psi = (g \circ \mu)^{-1}$.

$$\log \ell(\theta) = \sum_{i=1}^{n} \left[\log h(x_i, \phi/t_i) + \frac{\psi(\beta^\tau Z_i)x_i - \zeta(\psi(\beta^\tau Z_i))}{\phi/t_i} \right]$$

$$\frac{\partial \log \ell(\theta)}{\partial \beta} = \frac{1}{\phi} \sum_{i=1}^{n} \{[x_i - \mu(\psi(\beta^\tau Z_i))]\psi'(\beta^\tau Z_i)t_i Z_i\} = 0$$

$$\frac{\partial \log \ell(\theta)}{\partial \phi} = \sum_{i=1}^{n} \left\{ \frac{\partial \log h(x_i, \phi/t_i)}{\partial \phi} - \frac{t_i[\psi(\beta^\tau Z_i)x_i - \zeta(\psi(\beta^\tau Z_i))]}{\phi^2} \right\} = 0.$$

From the first likelihood equation, an MLE of β, if it exists, can be obtained without estimating ϕ.
The second likelihood equation, however, is usually difficult to solve.
Some other estimators of ϕ are suggested by various researchers; see, for example, McCullagh and Nelder (1989).

MLE in GLM

Suppose that there is a solution $\widehat{\beta} \in B$ to the likelihood equation.

$$\mathrm{Var}\left(\frac{\partial \log \ell(\theta)}{\partial \beta}\right) = M_n(\beta)/\phi, \quad \frac{\partial^2 \log \ell(\theta)}{\partial \beta \partial \beta^\tau} = [R_n(\beta) - M_n(\beta)]/\phi.$$

where

$$M_n(\beta) = \sum_{i=1}^n [\psi'(\beta^\tau Z_i)]^2 \zeta''(\psi(\beta^\tau Z_i)) t_i Z_i Z_i^\tau$$

$$R_n(\beta) = \sum_{i=1}^n [x_i - \mu(\psi(\beta^\tau Z_i))] \psi''(\beta^\tau Z_i) t_i Z_i Z_i^\tau.$$

Consider first the simple case of canonical g, $\psi'' \equiv 0$ and $R_n \equiv 0$.
If $M_n(\beta)$ is positive definite for all β, then $-\log \ell(\theta)$ is strictly convex in β for any fixed ϕ and, therefore, $\widehat{\beta}$ is the unique MLE of β.
For noncanonical g, $R_n(\beta) \neq 0$ and $\widehat{\beta}$ is not necessarily an MLE.
If $R_n(\beta)$ is dominated by $M_n(\beta)$, i.e.,

$$[M_n(\beta)]^{-1/2} R_n(\beta) [M_n(\beta)]^{-1/2} \to 0$$

in some sense, then $-\log \ell(\theta)$ is convex and $\widehat{\beta}$ is an MLE for large n.
See more details in the proof of Theorem 4.18 in §4.5.2.

Computation of MLE in GLM

In a GLM, an MLE $\widehat{\beta}$ usually does not have an analytic form.
A numerical method such as the Newton-Raphson or the Fisher-scoring method has to be applied.
Using the Newton-Raphson method, we have the following iteration procedure:

$$\widehat{\beta}^{(t+1)} = \widehat{\beta}^{(t)} - [R_n(\widehat{\beta}^{(t)}) - M_n(\widehat{\beta}^{(t)})]^{-1} s_n(\widehat{\beta}^{(t)}), \qquad t = 0, 1, ...,$$

where $s_n(\beta) = \phi \partial \log \ell(\theta)/\partial \beta$.
Note that $E[R_n(\beta)] = 0$ if β is the true parameter value and x_i is replaced by X_i.
This means that the Fisher-scoring method uses the following iteration procedure:

$$\widehat{\beta}^{(t+1)} = \widehat{\beta}^{(t)} + [M_n(\widehat{\beta}^{(t)})]^{-1} s_n(\widehat{\beta}^{(t)}), \qquad t = 0, 1, ...$$

If the canonical link is used, then the two methods are identical.

Example 4.36

Consider the GLM with $\zeta(\eta) = \eta^2/2$, $\eta \in \mathscr{R}$.
If g is the canonical link, then the model is the same as a linear model with independent ε_i's distributed as $N(0, \phi_i)$.
If $\phi_i \equiv \phi$, then the likelihood equation is exactly the same as the normal equation in §3.3.1.
If Z is of full rank, then $M_n(\beta) = Z^\tau Z$ is positive definite.
Thus, the LSE $\widehat{\beta}$ in a normal linear model is the unique MLE of β.
Suppose now that g is noncanonical but $\phi_i \equiv \phi$.
Then the model reduces to the one with independent X_i's and

$$X_i = N\left(g^{-1}(\beta^\tau Z_i), \phi\right), \qquad i = 1, ..., n.$$

This type of model is called a *nonlinear regression model* (with normal errors) and an MLE of β under this model is also called a nonlinear LSE, since maximizing the log-likelihood is equivalent to minimizing the sum of squares $\sum_{i=1}^n [X_i - g^{-1}(\beta^\tau Z_i)]^2$.
Under certain conditions the matrix $R_n(\beta)$ is dominated by $M_n(\beta)$ and an MLE of β exists.

Example 4.37 (The Poisson model)

Consider the GLM with $\zeta(\eta) = e^\eta$, $\eta \in \mathscr{R}$, $\phi_i = \phi/t_i$.
If $\phi_i = 1$, then X_i has the Poisson distribution with mean e^{η_i}.
Under the canonical link $g(t) = \log t$,

$$M_n(\beta) = \sum_{i=1}^n e^{\beta^\tau Z_i} t_i Z_i Z_i^\tau,$$

which is positive definite if $\inf_i e^{\beta^\tau Z_i} > 0$ and the matrix $(\sqrt{t_1}Z_1, ..., \sqrt{t_n}Z_n)$ is of full rank.
There is one noncanonical link that deserves attention.
Suppose that we choose a link function so that $[\psi'(t)]^2\zeta''(\psi(t)) \equiv 1$.
Then $M_n(\beta) \equiv \sum_{i=1}^n t_i Z_i Z_i^\tau$ does not depend on β.
In §4.5.2 it is shown that the asymptotic variance of the MLE $\widehat{\beta}$ is $\phi[M_n(\beta)]^{-1}$.
The fact that $M_n(\beta)$ does not depend on β makes the estimation of the asymptotic variance (and, thus, statistical inference) easy.
Under the Poisson model, $\zeta''(t) = e^t$ and, therefore, we need to solve the differential equation $[\psi'(t)]^2 e^{\psi(t)} = 1$.
A solution is $\psi(t) = 2\log(t/2)$ and the link $g(\mu) = 2\sqrt{\mu}$.

Quasi-MLE

If assumption ϕ_i is arbitrary, or the distribution assumption on X_i does not hold (e.g., X_i is longitudinal), but

$$E(X_i) = \zeta'(\eta_i) \quad \text{and} \quad \mathrm{Var}(X_i) = \phi_i \zeta''(\eta_i), \qquad i = 1, ..., n.$$

and

$$g(\mu(\eta_i)) = \beta^\tau Z_i, \qquad i = 1, ..., n,$$

still hold, and we estimate β by solving equation

$$\sum_{i=1}^{n} \left\{ [x_i - \mu(\psi(\beta^\tau Z_i))] \psi'(\beta^\tau Z_i) t_i Z_i \right\} = 0$$

then the resulting estimator is called a quasi-MLE.

This method is also called the method of generalized estimating equations (GEE).

Quasi-MLE or GEE has some good asymptotic properties.

They are efficient if the GEE is a likelihood equation, and is robust if it is not.