Computational Methods for Linear Model

Fred Wu

09 September 2021

This file [1] contains my own studies on the computations of generalised linear models using **R/C++** (particularly `Rcpp`).

1 Exponential Family

A random variable Y is said to have distribution belonging to an exponential dispersion family if it has density or mass function of the form

$$f(y_i;\theta,\phi) = \exp\left\{\frac{y_i\theta - b(\theta)}{a(\phi)} + c(y_i,\phi)\right\}.$$

where y_i is the ith observation and $\Theta = [\theta, \phi]$ is a vector of unknown parameters. then the likelihood and the log-likelihood functions are defined as

$$L(\Theta; y_i) = f(y_i;\Theta)$$

and

$$l(\Theta; y_i) = \log f(y_i;\Theta).$$

In seeking to estimate the parameters, we regard Θ as an argument of the function whilst the observed samples are considered to be fixed. However, in analysing the statistical properties of the function, we restore the random character to the observed samples, as different samples provide different estimates of the likelihood function. Therefore, it is necessary to remember that $L(\Theta; y_i) = f(y_i;\Theta)$, the likelihood is a probability density function of the observation (given the parameter values).

Derivation of $\mathbf{E}(Y_i)$ & $\mathbf{Var}(Y_i)$

For an exponential family distribution, we could compute its mean and variance with a general formula from the above equation with

$$\begin{aligned}
\mathrm{E}\left(Y_i\right) &= \mathrm{E}\left(\frac{\partial l(\Theta; y_i)}{\partial\theta}\right) + b'(\theta) \\
\mathrm{Var}\left(Y_i\right) / \left[\alpha(\phi)\right]^2 &= \mathrm{E}\left(\frac{\partial l(\Theta; y_i)}{\partial\theta}\right)^2 \\
&= \mathrm{E}\left(\frac{y_i - b'(\theta)}{a(\phi)}\right)^2.
\end{aligned}$$

[1] When compiling this document, instead of `knit` from `RStudio` interface, using `rmarkdown::render("lm.Rmd")` from the console could access Environment for pre-defined or pre-compiled objects or functions. This could be useful if you don't want `C++` codes to be compiled every time when you generate the document. In the `Rcpp` chunk below, the codes will not run but use the one that has already been compiled in the global environment.

Property of likelihood function

For a probability density function with any observed sample i, $\int f(y_i;\Theta)dy_i = \int L(\Theta;y_i) = 1$. Taking the partial derivative w.r.t θ gives rise to $\frac{\partial}{\partial\theta}\int L(\Theta;y_i)dy_i = \int \frac{\partial}{\partial\theta}L(\Theta;y_i)dy_i = 0$

The score function is the first derivative of the log-likelihood function

$$
\begin{aligned}
\frac{\partial l(\Theta;y_i)}{\partial\theta} &= \frac{\partial}{\partial\theta}\log L(\Theta;y_i) \\
&= \frac{\partial}{\partial\theta}L(\Theta;y_i)\frac{1}{L(\Theta;y_i)} \\
\frac{\partial}{\partial\theta}L(\Theta;y_i) &= \frac{\partial l(\Theta;y_i)}{\partial\theta}L(\Theta;y_i)
\end{aligned}
$$

Considering θ_0 is the true value of θ, and take the expected value on the score function over the sample space $\mathcal{Y}$

$$
\begin{aligned}
\mathrm{E}\left(\left.\frac{\partial l(\Theta;y_i)}{\partial\theta}\right|_{\theta=\theta_0}\right) &= \int_{\mathcal{Y}} \frac{\partial l(\theta_0;y_i)}{\partial\theta} f(y_i;\theta_0)dy_i \\
&= \int_{\mathcal{Y}} \frac{\partial}{\partial\theta}L(\theta_0;y_i)\frac{1}{L(\theta_0;y_i)}L(\theta_0;y_i)dy_i \\
&= 0
\end{aligned}
$$

Therefore, we derived that

$$
\mathrm{E}\left(Y_i\right) = b'\left(\theta\right)
$$

The reason to use θ_0 as a true value in the above calculation instead of θ for both likelihood and density functions is to emphasise the difference in how we interpret each one. The likelihood function is viewed as a function of θ with the observed data held fixed, but the density function is a model that describes the random behaviour of $Y_i = y_i$ when θ is fixed but unknown. For a likelihood function, the interest lies in finding a θ that the observed data is more likely to have occurred under $f(y_i;\theta_0)$. This means different θs can be used to evaluate the likelihood function. Only being evaluated at the true value θ_0 or at the same value of θ as in the population density is the expectation of the score function has mean zero.