Firth regression

stats methodologists meeting

20 October 2015

Bias Reduction of Maximum Likelihood Estimates: D. Firth; Biometrika (1993)

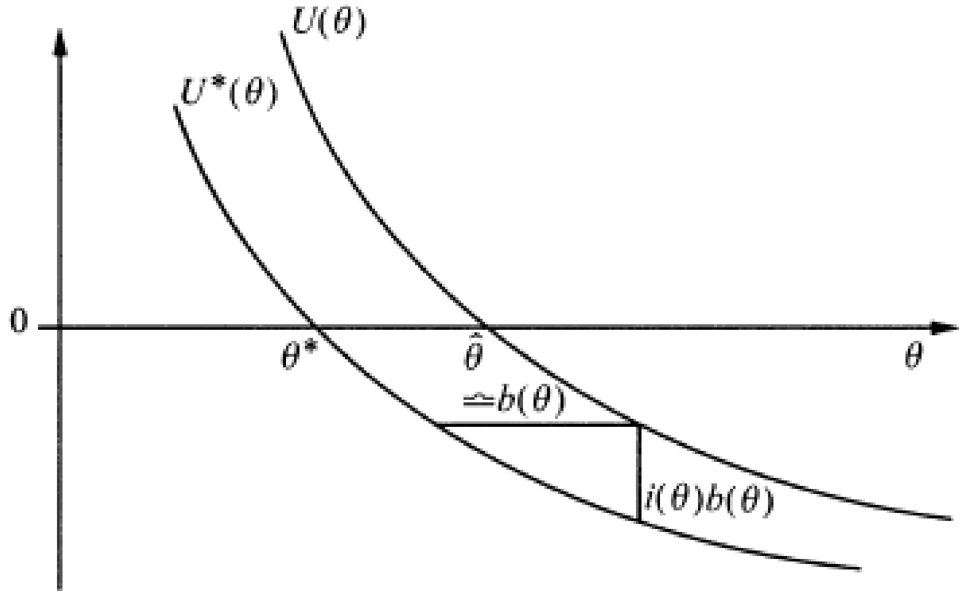


Fig. 1. Modification of the unbiased score function.

$$\nabla l(\theta) = U(\theta) = 0$$

exponential family
canonical param.

$$l(\theta) = t\theta - K(\theta)$$

$$l'(\theta) = t - K'(\theta)$$

$l(\theta) = \log L(\theta)$ is the log likelihood function

Bias Reduction of Maximum Likelihood Estimates: D. Firth; Biometrika (1993)

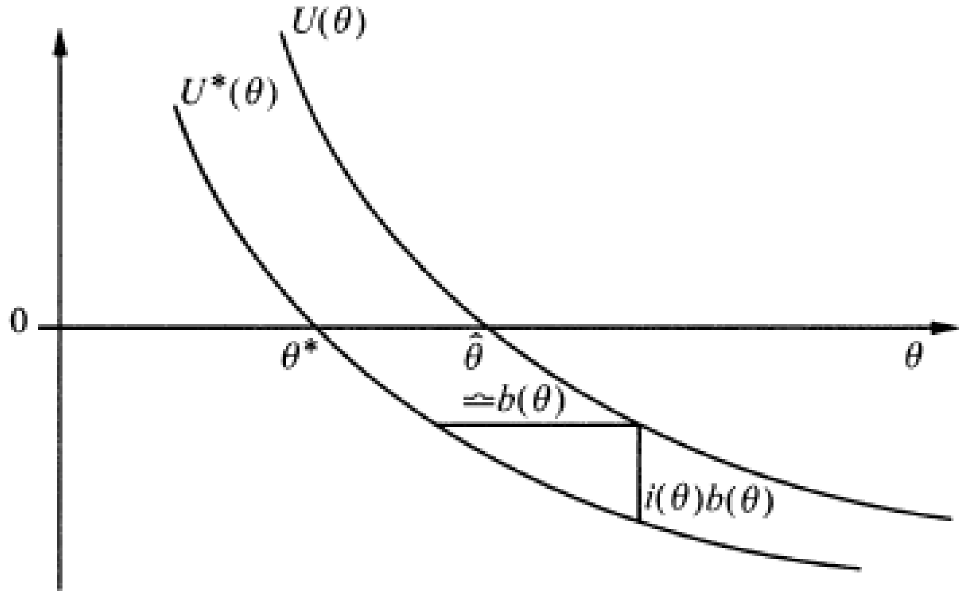


Fig. 1. Modification of the unbiased score function.

$$\nabla l(\theta) = U(\theta) = 0$$

$$U^*(\theta) = U(\theta) - i(\theta)b(\theta)$$

$i(\theta)$ is the Fisher information matrix.

derivatives are denoted by

$$U_r(\theta) = \partial l/\partial\theta^r, \quad U_{rs}(\theta) = \partial^2 l/\partial\theta^r\partial\theta^s,$$

and so on, where $\theta = (\theta^1, \ldots, \theta^p)$ is the parameter vector. The joint null cumulants are

$$\kappa_{r,s} = n^{-1}E\{U_r U_s\}, \quad \kappa_{r,s,t} = n^{-1}E\{U_r U_s U_t\}, \quad \kappa_{r,st} = n^{-1}E\{U_r U_{st}\},$$

and so on. We note here the well-known relationships

$$\kappa_{rs} + \kappa_{r,s} = 0, \quad \kappa_{rst} + \kappa_{r,st} + \kappa_{s,rt} + \kappa_{t,rs} + \kappa_{r,s,t} = 0. \tag{2·2}$$

Consider now a fairly general modification of the score function, of the form

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta),$$

in which A_r is allowed to depend on the data and is $O_p(1)$ as $n \to \infty$. Suppose that $\hat{\theta}$ and θ^* satisfy $U(\hat{\theta}) = 0$ and $U^*(\theta^*) = 0$, and write $\hat{\gamma} = n^{\frac{1}{2}}(\theta^* - \theta)$. Then, by an argument closely following that of McCullagh (1987, p. 209), based on an expansion of $U_r^*(\theta^*)$ about the true value θ, the bias of θ^* is

$$E(n^{-\frac{1}{2}}\hat{\gamma}^r) = n^{-1}\kappa^{r,s}\{-\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})/2 + \alpha_s\} + O(n^{-3/2}),$$

where $\kappa^{r,s}$ denotes the inverse of the Fisher information matrix $\kappa_{r,s}$, α_s denotes the null expectation of A_s, and the summation convention applies. The term

$$-n^{-1}\kappa^{r,s}\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})/2 = n^{-1}b_1^r(\theta)$$

is the first-order bias of $\hat{\theta}$, for example Cox & Snell (1968). The modification A_r therefore removes the first-order term if it satisfies

$$\kappa^{r,s}\alpha_s = -b_1^r + O(n^{-\frac{1}{2}}),$$

the solution to which is

$$\alpha_r = -\kappa_{r,s}b_1^s + O(n^{-\frac{1}{2}}).$$

In matrix notation, then, the vector A should be such that

$$E(A) = -i(\theta)b_1(\theta)/n + O(n^{-\frac{1}{2}}).$$

Obvious candidates for a bias-reducing choice of A are therefore $A^{(E)} = -i(\theta)b_1(\theta)/n$

If θ is the canonical parameter of an exponential family model, $\kappa_{r,st}=0$ for all r, s and t. Therefore the rth element of $A^{(E)}(\theta)$, or equivalently of $A^{(O)}(\theta)$, is given by

$$a_r=-n\kappa_{r,s}b_1^s/n=\kappa_{r,s}\kappa^{s,t}\kappa^{u,v}\kappa_{t,u,v}/2=\kappa^{u,v}\kappa_{r,u,v}/2=-\kappa^{u,v}\kappa_{ruv}/2,$$

using the identities (2·2). In matrix notation, this may be written as

$$a_r=\frac{1}{2}\operatorname{tr}\left\{i^{-1}\left(\frac{\partial i}{\partial\theta_r}\right)\right\}=\frac{\partial}{\partial\theta_r}\left\{\frac{1}{2}\log|i(\theta)|\right\}.$$

Solution of $U_r^*\equiv U_r+a_r=0$ therefore locates a stationary point of

$$l^*(\theta)=l(\theta)+\tfrac{1}{2}\log|i(\theta)|$$

or, equivalently, of the penalized likelihood function

$$L^*(\theta)=L(\theta)|i(\theta)|^{\frac{1}{2}}.$$

Solution for Logistic Regression

penalized likelihood function

$$L^*(\theta) = L(\theta)|i(\theta)|^{\frac{1}{2}}.$$

The simplest of all logistic models is that for a single binomial observation, the target parameter being $\beta = \log\{\pi/(1-\pi)\}$. The information is proportional to $\pi(1-\pi)$, so that the penalized likelihood is simply

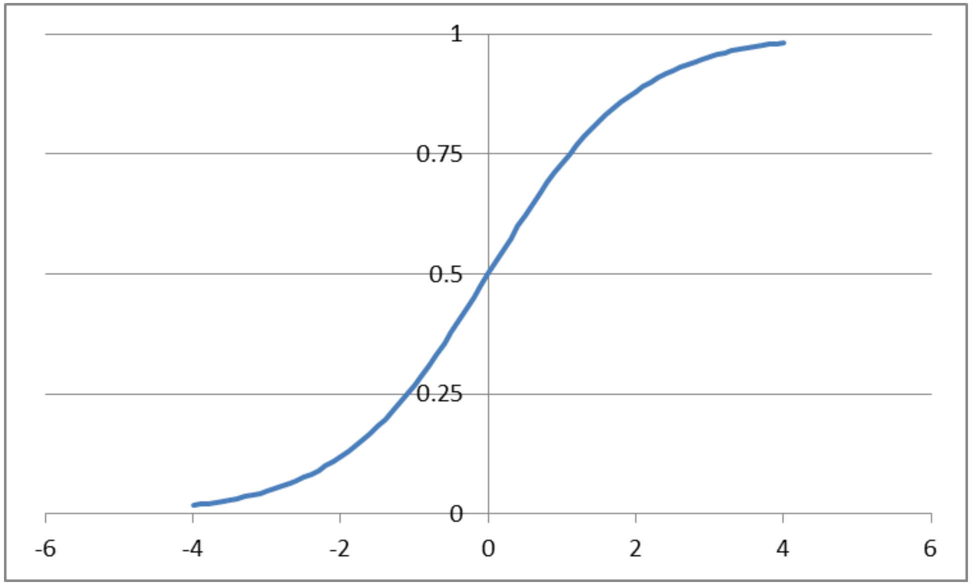
$$L^* = \pi^{y+\frac{1}{2}}(1-\pi)^{m-y+\frac{1}{2}}.$$

Maximization of L^* yields

$$\beta^* = \log\left(\frac{y+\frac{1}{2}}{m-y+\frac{1}{2}}\right),$$

shrinks estimates towards $\beta=0$

Logistic Regression: Separation Problem



$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = B_0 + B_1 X = \eta$$

$$p = 1/(1 + \exp(-\eta))$$

if X predicts outcome perfectly,
its regression coefficient => ∞

Bayesian inference: Jeffries Prior

penalized likelihood function

$$L^*(\theta) = L(\theta)|i(\theta)|^{\frac{1}{2}}.$$

Posterior = Likelihood * Prior

Binomial distribution; θ = prob. success

Jeffries Prior: $\pi_J(\theta) = I(\theta)^{\frac{1}{2}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$

Note on notation: theta in the top equation is the canonical variable, and theta in the bottom equation is the binary probability of success (p).