

Extending conditional likelihood for clustered data

Estensioni della verosimiglianza condizionata per dati raggruppati

Ruggero Bellio ¹
Dipartimento di Scienze Statistiche
Università di Udine
ruggero.bellio@dss.uniud.it

Nicola Sartori ¹
Dipartimento di Scienze Statistiche
Università di Padova
sartori@stat.unipd.it

Riassunto: La verosimiglianza condizionata è un metodo inferenziale usato per la stima di un'importante classe di modelli di regressione per dati discreti divisi in gruppi. In particolare, il metodo è adatto a situazioni in cui vi siano dei parametri di interesse comuni a tutti i gruppi e dei parametri di disturbo che variano tra i gruppi. La verosimiglianza profilo modificata risulta essere una accurata approssimazione della verosimiglianza condizionata, ma presenta l'importante vantaggio di essere calcolabile per generici modelli parametrici. Il lavoro propone la verosimiglianza profilo modificata come ideale estensione della verosimiglianza condizionata e si sofferma in particolare sul caso di dati binari.

Keywords: Clustered data; Conditional Likelihood; Modified Profile Likelihood.

1. Introduction

The conditional likelihood approach is widely used for clustered data. Let Y be the response variable, described by a generalized linear model, with probability density function (McCullagh and Nelder 1989, eqn. 2.4)

$$p(y; \eta, \phi) = \exp[\{\theta(\eta) y - K(\eta)\}/a(\phi) + c(y, \phi)], \quad (1)$$

where θ , K , a and c are specific functions, ϕ is the dispersion parameter and η is the linear predictor. We observe a sample y_{ij} , $i = 1, \dots, k$, $j = 1, \dots, n_i$, and we assume that ϕ is known and the linear predictor η_{ij} is

$$\eta_{ij} = \lambda_i + x_{ij}^T \beta.$$

The interest focuses on the regression coefficients β , while $\lambda_1, \dots, \lambda_k$ are nuisance parameters.

First developed by Andersen (1970) for binary data, the conditional likelihood applies to the case of canonical link function, and it removes the cluster-specific parameters $\lambda_1, \dots, \lambda_k$ by conditioning on the sufficient statistic for them. The resulting conditional maximum estimator $\hat{\beta}_C$ of β has better properties than the unconditional estimator $\hat{\beta}$, especially when k is large compared to the size of the clusters. Further details are given in McCullagh and Nelder (1989, Chapter 7). The conditional likelihood is useful in many instances, and routines for fitting it are provided by major statistical software (Pendergast et al. 1996). The problem is that its application is limited to generalized linear model with canonical link function. In the following, we shall present a suitable extension to more general settings.

¹Il lavoro è stato svolto con finanziamenti MIUR COFIN 2001

2. The modified profile likelihood

Consider a parametric statistical model satisfying some regularity conditions (Severini 2000, Chapter 3). The modified profile likelihood (Barndorff-Nielsen 1983) approximates both conditional and marginal likelihoods when they either exist. Remarkably, the modified profile likelihood is quite effective even when neither a conditional nor a marginal likelihood exist, as it defines an estimating equation with better properties than the corresponding one defined by the ordinary profile likelihood (Severini 1998b). Its main drawback is that it is very difficult to compute outside linear exponential families or transformation models. However, recent results in the field of likelihood asymptotics have widened its applicability, and a general version is now available (Severini 1998a, 2000). If ψ is the parameter of interest, and λ the nuisance parameter, its expression is given by

$$L_M(\psi) = L_P(\psi) \frac{|j_{\lambda\lambda}(\psi, \tilde{\lambda})|^{1/2}}{|I_{\lambda\lambda}(\hat{\psi}, \hat{\lambda}; \psi, \tilde{\lambda})|} \quad (2)$$

where $(\hat{\psi}, \hat{\lambda})$ is the maximum likelihood estimate of the parameters, $\tilde{\lambda}$ is the maximum likelihood estimate of λ for a fixed value of ψ , $L_P(\psi) = L(\psi, \tilde{\lambda})$ is the profile likelihood for ψ and $j_{\lambda\lambda}$ is the $\lambda\lambda$ -block of the observed information. The factor $I_{\lambda\lambda}$ is given by

$$I_{\lambda\lambda}(\hat{\psi}, \hat{\lambda}; \psi, \tilde{\lambda}) = \text{cov}_{\psi_0, \lambda_0} \{ \ell_{\lambda}(\psi_0, \lambda_0), \ell_{\lambda}(\psi_1, \lambda_1) \} \Big|_{(\psi_0=\hat{\psi}, \lambda_0=\hat{\lambda}, \psi_1=\psi, \lambda_1=\tilde{\lambda})},$$

where ℓ_{λ} is the λ -part of the score function. In many models, the covariance required for $I_{\lambda\lambda}$ is not difficult to compute. For the case of model (1) with canonical link function, the modified profile likelihood (2) gives a very close approximation to the exact conditional likelihood, which is also available in the S-PLUS library HOA (Brazzale 1999). The important point is that the modified profile likelihood is computable also for non-canonical link functions, when there is no sufficient statistic for the nuisance parameters. The resulting estimator $\hat{\beta}_M$ has better properties than $\hat{\beta}$, in particular when k increases at a faster rate than n_i . A thorough study is given in Sartori (2001).

3. The case of binary data

Consider the particular case of (1) where Y is a binary variable, with success probability $F(\eta)$, where F is a given distribution function with corresponding density function f . The profile log-likelihood for β is given by

$$\ell_P(\beta) = \sum_{i=1}^k \ell_i(\beta, \tilde{\lambda}_i)$$

where ℓ_i is the contribution from the i -th cluster

$$\ell_i(\beta, \tilde{\lambda}_i) = \sum_{j=1}^{n_i} [y_{ij} \log F(\tilde{\eta}_{ij}) + (1 - y_{ij}) \log \{1 - F(\tilde{\eta}_{ij})\}] ,$$

with $\tilde{\eta}_{ij} = \eta_{ij}(\beta, \tilde{\lambda}_i)$. Similarly, both $|j_{\lambda\lambda}|$ and $|I_{\lambda\lambda}|$ are the product of the contributions from the i -th cluster, which are respectively given by

$$I^{(i)}(\hat{\beta}, \hat{\lambda}_i; \beta, \tilde{\lambda}_i) = \sum_{j=1}^{n_i} \frac{f(\hat{\eta}_{ij}) f(\tilde{\eta}_{ij})}{\{1 - F(\tilde{\eta}_{ij})\} F(\tilde{\eta}_{ij})},$$

$$j_{\lambda_i \lambda_i}(\beta, \tilde{\lambda}_i) = I^{(i)}(\beta, \tilde{\lambda}_i; \beta, \tilde{\lambda}_i) - \sum_{j=1}^{n_i} \{y_{ij} - F(\tilde{\eta}_{ij})\} c_{ij},$$

$$c_{ij} = \left[\frac{f'(\tilde{\eta}_{ij})}{\{1 - F(\tilde{\eta}_{ij})\} F(\tilde{\eta}_{ij})} - \frac{f^2(\tilde{\eta}_{ij}) \{1 - 2 F(\tilde{\eta}_{ij})\}}{\{1 - F(\tilde{\eta}_{ij})\}^2 F(\tilde{\eta}_{ij})^2} \right],$$

with $\hat{\eta}_{ij} = \eta_{ij}(\hat{\beta}, \hat{\lambda}_i)$. The resulting expression of $\ell_M(\beta)$ is a smooth function of β , that can be numerically maximised by one of the optimisers available within standard statistical software. The key point for a successful implementation is the availability of efficient routines for computing $\tilde{\lambda}$, that essentially requires to be able to fit a generalized linear model with an offset. For the most important link functions, such routines are widely implemented. Note that the inclusion of additional parameters in the link function, typical of many applications (e.g. Morgan 1992), can be dealt with in a straightforward way.

As an illustrative example, let us consider the crying babies data, already analysed by several authors, including the thorough analysis reported in Brazzale (1999). It is a matched-pair study concerning the crying on babies who underwent or not a treatment (lulling) on 18 different days. Namely, we have $k = 18$, n_i between 6 and 10 and a single coefficient β of interest, the treatment effect. Table 1 reports the results for logit, probit, complementary log-log and log-log link.

Likelihood Method	Treatment Estimate (s.e.)	p -value
Profile (Logit)	1.43 (0.73)	.049
Conditional (Logit)	1.26 (0.69)	.067
Mod. Profile (Logit)	1.27 (0.69)	.066
Profile (Probit)	0.80 (0.40)	.048
Mod. Profile (Probit)	0.71 (0.38)	.061
Profile (cloglog)	0.73 (0.36)	.044
Mod. Profile (cloglog)	0.62 (0.33)	.060
Profile (loglog)	1.11 (0.60)	.065
Mod. Profile (loglog)	1.03 (0.60)	.086

Table 1: Treatment estimation for crying babies data.

The numerical results confirm that L_M is an accurate approximation to the conditional likelihood, and they both provide less evidence in favour of a positive treatment effect than L_P . Similar results are obtained in the analyses with the other link functions. The important point is that the modified profile likelihood allows the statistician to select the most appropriate link function, not being confined anymore to the logit link only.

4. Ongoing research

Application of the modified profile likelihood to other models are currently under investigation. An important instance is represented by models for ordinal data with non-canonical link function, see Agresti and Natarajan (2001). Applications are also possible outside the generalized linear models specification; see, for instance, the examples considered in Lancaster (2000).

References

- Agresti, A. and Natarajan, R. (2001). Modeling clustered ordered categorical data: a survey, *Internat. Statist. Rev.* **69**, 345–372.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators, *J. R. Statist. Soc. B* **32**, 283–301.
- Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika* **70**, 343–365.
- Brazzale, A. R. (1999). Approximate conditional inference in logistic and loglinear models, *J. Comput. Graph. Statist.* **8**, 653–661.
- Lancaster, T. (2000). The incidental parameter problem since 1948, *J. Econometrics* **95**, 391–413.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (Second Edition)*, Chapman & Hall, London.
- Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*, Chapman & Hall, London.
- Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M. and Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data, *Internat. Statist. Rev.* **64**, 89–118.
- Sartori, N. (2001). Modifications to the profile likelihood in models with incidental nuisance parameters, *Submitted Manuscript*.
- Severini, T. A. (1998a). An approximation to the modified profile likelihood function, *Biometrika* **85**, 403–411.
- Severini, T. A. (1998b). Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika* **85**, 507–522.
- Severini, T. A. (2000). *Likelihood Methods in Statistics*, Oxford University Press, Oxford.