

Robustness and Estimating Equations in the Presence of a Nuisance Parameter¹

Robustezza e Equazioni di Stima in Presenza di Parametri di Disturbo

Luca Greco, Laura Ventura

Dipartimento di Scienze Statistiche, Università di Padova

Via C. Battisti 241, 35121 Padova

greco@stat.unipd.it, ventura@stat.unipd.it

Riassunto: In questo lavoro vengono discusse alcune equazioni di stima che forniscono stimatori robusti per un parametro di interesse in presenza di componenti di disturbo. Il metodo proposto è incentrato sull'approccio basato sulla funzione di influenza e consiste nell'applicazione di una procedura di troncamento alla funzione score di una verosimiglianza profilo generalizzata.

Keywords: B-robustness, Influence function, M-estimator, Profile likelihood.

1. Introduction

Consider a sample $y = (y_1, \dots, y_n)$ of n independent observations from a random variable Y with distribution function $F_\theta = F(y; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p > 1$. Suppose that θ is partitioned as $\theta = (\tau, \lambda)$, into a scalar parameter of interest τ and a $(p-1)$ -dimensional nuisance parameter λ . Classical inference about τ in the presence of λ is typically based on a pseudo-likelihood, i.e. a function of y and τ having properties similar to those of a likelihood function when there is no nuisance parameter. The most commonly used pseudo-loglikelihood is the profile loglikelihood $\ell_p(\tau) = \ell(\tau, \hat{\lambda}_\tau) = \sum_{i=1}^n \ell(\tau, \hat{\lambda}_\tau; y_i)$, where $\ell(\theta) = \ell(\tau, \lambda)$ denotes the usual loglikelihood for θ and $\hat{\lambda}_\tau$ is the maximum likelihood estimate (MLE) of λ for fixed τ . Other pseudo-likelihoods include the adjusted profile, the conditional and the marginal likelihood (see Severini (2000)).

In many situations of practical interest, there is no certainty that the data y come from the specified model F_θ and may in fact come from some neighborhood of the model. It is well-known that standard likelihood procedures are not robust and the need for robust statistical procedures for estimation and testing has been stressed by many authors; see Hampel *et al.* (1986) and Markatou and Ronchetti (1997). However, while robust literature offers many solutions for inference on the whole parameter θ , the situation with a nuisance parameter has been somewhat neglected. In this paper, we discuss the general problem of robust inference about τ , in the presence of λ . In particular, we obtain a robust estimator for the parameter of interest as the root of an M-type estimating equation. The followed approach is essentially based on a standard truncation argument from the theory of robust statistics (see Carroll and Ruppert (1988)). It consists in bounding a profile score-type function in order to get an estimator for τ with bounded influence function (IF). Such an estimator is called B-robust (bias-robust).

¹Partially supported by a grant from MIUR, Italy.

2. Profile and generalised profile likelihoods

To set up notation, in the following, we will denote first partial derivatives of $\ell(\tau, \lambda)$ by $\ell_\lambda = \partial\ell(\tau, \lambda)/\partial\lambda$ and $\ell_\tau = \partial\ell(\tau, \lambda)/\partial\tau$ and, with an obvious notation, second partial derivatives by $\ell_{\lambda\lambda}$, $\ell_{\lambda\tau}$ and $\ell_{\tau\tau}$.

The profile loglikelihood $\ell_p(\tau)$ is not a proper loglikelihood. In particular, the profile score $\ell'_p(\tau) = \partial\ell_p(\tau)/\partial\tau$ has not zero mean, but its expectation is of order $O(1)$. However, it has many properties as a genuine loglikelihood for τ (see Severini (2000)). First of all, the profile MLE of τ , $\hat{\tau}$, given as the solution of $\ell'_p(\tau) = \sum_{i=1}^n \ell_\tau(\tau, \hat{\lambda}_\tau; y_i) = 0$, coincides to the MLE for τ . Moreover, the profile loglikelihood ratio statistic $W_p(\tau_0) = 2(\ell_p(\hat{\tau}) - \ell_p(\tau_0))$ for testing $\tau = \tau_0$ versus $\tau \neq \tau_0$ has a standard χ^2_1 distribution. Similarly, the directed likelihood $r_p(\tau_0) = \text{sgn}(\hat{\tau} - \tau_0) \sqrt{W_p(\tau_0)}$ has a standard normal distribution. In view of this, for setting confidence regions or for testing hypothesis, $W_p(\tau)$ or $r_p(\tau)$ may be used in a standard way.

Function $\ell_p(\tau)$ can be generalised if estimates other than the MLE are used in the profile likelihood approach. Let $\tilde{\lambda}_\tau$ denotes an alternative estimate of λ for fixed τ . Provided that some weak conditions on $\tilde{\lambda}_\tau$ are satisfied (Severini (1998)), the generalised profile loglikelihood $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda}_\tau)$ has the same first-order properties as $\ell_p(\tau)$ and can be used for inference about τ . One simple way to construct $\tilde{\ell}_p(\tau)$ is to consider approximations to $\hat{\lambda}_\tau$, when its exact computation is difficult. A first approximation is given by $\tilde{\lambda}_\tau^* = \hat{\lambda} - \ell_{\lambda\lambda}(\hat{\tau}, \hat{\lambda})^{-1} \ell_{\tau\lambda}(\hat{\tau}, \hat{\lambda})(\tau - \hat{\tau})$, where $(\hat{\tau}, \hat{\lambda})$ denotes the MLE of (τ, λ) . A second closely related approximation is $\tilde{\lambda}_\tau^+ = \bar{\lambda} + \ell_{\lambda\lambda}(\bar{\tau}, \bar{\lambda})^{-1} \ell_{\lambda\tau}(\bar{\tau}, \bar{\lambda})(\tau - \bar{\tau})$, which only requires the estimate $(\bar{\tau}, \bar{\lambda})$ to be $\sqrt{n}$ -consistent. In a different way, a generalised profile loglikelihood can be obtained if estimates of the nuisance parameter of the form $\tilde{\lambda}$, i.e. which do not depend on τ , are involved. Gong and Samaniego (1981) discuss this general procedure and consider the first order asymptotic theory of the generalised MLE (GMLE) for τ based on $\ell(\tau, \tilde{\lambda})$; see also Pierce (1982), Parke (1986) and Yuan and Jennrich (2000).

The added flexibility in the choice of $\tilde{\lambda}_\tau$ or $\tilde{\lambda}$ allows for the possibility of disposing of a generalised profile loglikelihood with properties similar to those of $\ell_p(\tau)$. Furthermore, an alternative method of estimation of λ , superior to maximum likelihood in the robust sense, may be used when stability with respect to small deviations from the assumed model is required in making inference about τ . One such method of estimation could be to compute $\tilde{\lambda}_\tau^+$ with $(\bar{\tau}, \bar{\lambda})$ suitable robust M-estimator. However, this choice does not assure that the estimator $\tilde{\lambda}_\tau^+$ has a bounded IF. It becomes feasible to get an estimator for the nuisance with bounded IF if attention is restricted to an M-type estimator not depending on τ , defined by a bounded estimating function.

Let $\tilde{\lambda}$ be an appropriate estimate of λ satisfying the unbiased estimating equation

$$\Psi_\lambda(\lambda) = \sum_{i=1}^n \psi_\lambda(y_i; \lambda) = 0. \quad (1)$$

Under broad conditions, $\tilde{\lambda}$ is consistent and asymptotically normal with mean λ and variance $V_{\lambda\lambda} = M_{\lambda\lambda}^{-1} \Omega_{\lambda\lambda} (M_{\lambda\lambda}^{-1})^\top$, where $M_{\lambda\lambda} = -\int \partial\psi_\lambda(y; \lambda)/\partial\lambda^\top dF_\theta$ and $\Omega_{\lambda\lambda} = \int \psi_\lambda(y; \lambda) \psi_\lambda(y; \lambda)^\top dF_\theta$. Since the IF of the M-estimator $\tilde{\lambda}$ at a point y is given by $M_{\lambda\lambda}^{-1} \psi_\lambda(y; \lambda)$, it is B-robust at the assumed model provided that $\psi_\lambda(y; \lambda)$ is bounded.

Consider inference on the interest parameter τ , when $\tilde{\lambda}$ is used to eliminate the nuisance parameter. Let us focus on the generalised profile loglikelihood $\tilde{\ell}_p(\tau) = \ell(\tau, \tilde{\lambda})$.

The GMLE $\tilde{\tau}$ for τ is defined as the root of

$$\tilde{\ell}'_p(\tau) = \ell_\tau(\tau, \tilde{\lambda}) = \sum_{i=1}^n \ell_\tau(\tau, \tilde{\lambda}; y_i) = 0. \quad (2)$$

The IF of the corresponding estimator is $IF(y; \tilde{\tau}) = i_{\tau\tau}^{-1} \ell_\tau(\tau, \lambda; y) - i_{\tau\tau}^{-1} i_{\tau\lambda} M_{\lambda\lambda}^{-1} \psi_\lambda(y; \lambda)$, where $i_{\tau\tau} = -\int \ell_{\tau\tau}(\theta; y) dF_\theta$ and $i_{\tau\lambda} = -\int \ell_{\tau\lambda}(\theta; y) dF_\theta$. Since the IF is proportional to $\ell_\tau(\tau, \lambda; y)$, in general this estimator is not robust against outliers and influential observations. In view of this, when robustness is required, the estimating equation (2) must be modified.

3. Robust estimation of the parameter of interest

A general approach to find robust estimators consists in modifying the usual score function by an appropriate weighting function in order to achieve B-robustness. This idea has been used successfully by many authors (see Hampel *et al.* (1986), Carroll and Ruppert (1988)). An appropriate downweighting of the score functions leads to robust estimators with high efficiency at the assumed model, i.e. the optimal bias-robust estimators (OBRE). The aim of this section is to bound the generalised profile score function in (2) to obtain a B-robust estimator for τ . We focus on a technique shown in Hampel *et al.* (1986) which applies to a scalar parameter. The proposed robust estimate for τ , that is $\tilde{\tau}_r$ (RGMLE), satisfies the estimating equation

$$\Psi_\tau(\tau) = \sum_{i=1}^n \psi_\tau(y_i; \tau, \tilde{\lambda}) = \sum_{i=1}^n \left[\ell_\tau(\tau, \tilde{\lambda}; y_i) - a \right]_{-b}^b = 0, \quad (3)$$

where a is such that $\int \psi_\tau(y; \tau, \tilde{\lambda}) dF_\theta = 0$, and $b > 0$ is some constant related to the bound on the $IF(y; \tilde{\tau})$. The notation $[h(x)]_{-b}^b$ for an arbitrary function $h(x)$ means truncation at levels b and $-b$. To solve (3) in general a numerical algorithm must be used.

Some properties of $\tilde{\tau}_r$ are easy to evaluate. The RGMLE $\tilde{\tau}_r$ is found by solving for τ the system $\Psi(y; \tau, \lambda) = (\sum_{i=1}^n \psi_\tau(y_i; \tau, \lambda), \sum_{i=1}^n \psi_\lambda(y_i; \lambda))^T = 0$. The IF of the estimator $(\tilde{\tau}_r, \tilde{\lambda})$ is $M^{-1} \Psi(y; \tau, \lambda)$, where

$$M = \begin{pmatrix} M_{\tau\tau} & M_{\tau\lambda} \\ 0 & M_{\lambda\lambda} \end{pmatrix},$$

with $M_{\tau\tau} = -\int \partial \psi_\tau(y; \tau, \lambda) / \partial \tau dF_\theta$ and $M_{\tau\lambda} = -\int \partial \psi_\tau(y; \tau, \lambda) / \partial \lambda^T dF_\theta$. It is clear that it is Fisher consistent from the inclusion of a . Moreover, for the presence of b , it is B-robust. To study the asymptotic variance and normality of the estimator, the most useful references are Pierce (1982), Parke (1986) and Yuan and Jennrich (2000). The asymptotic variance matrix of $(\tilde{\tau}_r, \tilde{\lambda})$ is $V(\tau, \lambda) = M^{-1} \Omega (M^T)^{-1}$, where Ω is a block matrix, with $\Omega_{\tau\tau} = \int \psi_\tau(y; \tau, \lambda)^2 dF_\theta$ and $\Omega_{\tau\lambda} = \Omega_{\lambda\tau}^T = \int \psi_\tau(y; \tau, \lambda) \psi_\lambda(y; \lambda)^T dF_\theta$. Then, it is straightforward to obtain the expression for the asymptotic variance of $\tilde{\tau}_r$. Indeed, it corresponds to the $\tau\tau$ block of $V(\tau, \lambda)$, i.e. $V_{\tau\tau} = M_{\tau\tau}^{-2} \Omega_{\tau\tau} + 2 M_{\tau\lambda}^T \Omega_{\lambda\tau}^{-1} M_{\tau\tau}^{-1} + M_{\tau\lambda}^T \Omega_{\lambda\lambda} (M_{\tau\lambda})^{-1}$, where $M_{\tau\lambda}$ denotes the $\tau\lambda$ block of the inverse of M , given by $M_{\tau\lambda} = -M_{\tau\tau}^{-1} M_{\tau\lambda} M_{\lambda\lambda}^{-1}$. The first term of $V_{\tau\tau}$ is the asymptotic variance corresponding to τ if λ were known, whereas the remaining part reflects the cost of estimating the nuisance parameter. The former expression simplifies if $\Omega_{\lambda\tau} = 0$, as in the context of GMLE (Parke (1986)). Large-sample tests and confidence intervals for τ can be constructed using an estimate of $V_{\tau\tau}$.

4. An example

Consider a simple linear regression-scale and shape model $y = \beta_0 + \beta_1 x + \epsilon$ with errors ϵ_i distributed as an Exponential Power (EP) with parameters (η, σ, κ) . The interest parameter is the shape parameter κ , while $\lambda = (\beta_0, \beta_1, \sigma)$ is treated as nuisance. The error distribution is supposed to be centered in zero, i.e. $\eta = 0$. A robust estimate $\tilde{\lambda}$ for λ is obtained by applying the Huber proposal with rescaled MAD for regression-scale. To evaluate the performance of the proposed RGMLE for κ in terms of efficiency when the model is correctly specified, but also under small arbitrary departures from the assumed model, we simulate data from different error distributions: an EP with $(\sigma = 1, \kappa = 1.5)$, a scale contaminated model (EP2, 2% from an EP with $\sigma = 5$), a slash contaminated model (EP3, 2% from an EP divided by a Uniform on $[0, 1]$) and a location contaminated model (EP4, 2% from a Normal with expected value $\mu = 3$). The results of the Monte Carlo study (based on 10000 trials) for the MLE, the GMLE and the RGMLE for κ are given in the table below, for $n = 100$. The entries give the median and the mad of the sample distributions of each estimator. Some features emerge. Under the central model the RGMLE shows a performance similar to the MLE and the GMLE. Under the contaminated model the RGMLE exhibits its B-robustness properties, appearing less sensitive to the various deviations from the specified model.

Distribution	MLE	GMLE	RGMLE (b=2.5)	RGMLE (b=3)
EP	1.509 (0.366)	1.530 (0.336)	1.598 (0.331)	1.582 (0.332)
EP1	1.100 (0.335)	1.193 (0.321)	1.461 (0.295)	1.427 (0.290)
EP2	0.873 (0.450)	0.974 (0.471)	1.459 (0.292)	1.421 (0.284)
EP3	1.331 (0.307)	1.388 (0.289)	1.479 (0.299)	1.453 (0.294)

References

- Carroll R.J. and Ruppert D. (1988) *Transformation and Weighting in Regression*, Chapman & Hall, New York.
- Gong G. and Samaniego F. (1981) Pseudo maximum likelihood estimation: theory and applications, *The Annals of Statistics*, 9, 861–869.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986) *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York.
- Markatou M. and Ronchetti E. (1997) Robust inference: The approach based on influence functions, in: *Handbook of Statistics*, Maddala, G.S., Rao, C.R. (Eds.), North-Holland, Amsterdam, volume 15, 49–75.
- Parke W. (1986) Pseudo maximum likelihood estimation: the asymptotic distribution, *The Annals of Statistics*, 14, 355–357.
- Pierce D. (1982) The asymptotic effect of substituting estimators for parameters in certain types of statistics, *The Annals of Statistics*, 10, 475–478.
- Severini T. (1998) Likelihood functions for inference in the presence of a nuisance parameter, *Biometrika*, 85, 507–522.
- Severini T. (2000) *Likelihood Methods in Statistics*, Oxford University Press, Oxford.
- Yuan K. and Jennrich R. (2000) Estimating equations with nuisance parameters: theory and applications, *The Annals of the Institute of Statistical Mathematics*, 52, 343–350.