Working Paper Series, N. 4, November 2015

On penalised likelihood and bias reduction

Nicola Lunardon
Department of Statistical Sciences
University of Padova
Italy

Abstract: We propose a likelihood function endowed with a penalisation that reduces the bias of the maximum likelihood estimator in regular parametric models. The penalisation hinges on the first two derivatives of the log likelihood and can be computed numerically. The asymptotic properties and the sensitivity to nuisance parameters of the penalised likelihood and derived quantities are addressed. In models for stratified data in a two-index asymptotic setting, the bias of the penalised profile score function is found to be equivalent to the bias of a modified profile score function.

Keywords: Bartlett identity, Bias reduction, Incidental parameter, Jeffreys prior, Modified profile likelihood, Penalised likelihood, Profile likelihood, Profile score bias, Stratified data, Two-index asymptotics

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

1 Introduction **1**

2 Penalised likelihood function **2**

3 Sensitivity to the nuisance component **2**
3.1 Two-index asymptotic setting 3

4 Monte Carlo simulation **4**
4.1 Beta regression 4
4.2 Gamma samples 5

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Nicola Lunardon
tel: +39 049 827 4124
lunardon@stat.unipd.it
http://homes.stat.unipd.it/nicolalunardon/

On penalised likelihood and bias reduction

Nicola Lunardon
Department of Statistical Sciences
University of Padova
Italy

Abstract: We propose a likelihood function endowed with a penalisation that reduces the bias of the maximum likelihood estimator in regular parametric models. The penalisation hinges on the first two derivatives of the log likelihood and can be computed numerically. The asymptotic properties and the sensitivity to nuisance parameters of the penalised likelihood and derived quantities are addressed. In models for stratified data in a two-index asymptotic setting, the bias of the penalised profile score function is found to be equivalent to the bias of a modified profile score function.

Keywords: Bartlett identity, Bias reduction, Incidental parameter, Jeffreys prior, Modified profile likelihood, Penalised likelihood, Profile likelihood, Profile score bias, Stratified data, Two-index asymptotics

1 Introduction

Consider a regular model $\{f(y;\theta) : \theta \in \mathbb{R}^p, y \in \mathbb{R}^d\}$. The log likelihood, which we will just call the likelihood, and score functions based on n independent observations $y_1, \dots, y_n$ are

$$\ell(\theta) = \sum_{i=1}^{n} \ell^i(\theta) = \sum_{i=1}^{n} \log f(\theta; y_i), \quad \ell_\theta(\theta) = \sum_{i=1}^{n} \ell^i_\theta(\theta) = \sum_{i=1}^{n} \partial \ell^i(\theta)/\partial\theta,$$

while $j(\theta)$ and $i(\theta)$ are the observed and expected information. The maximum likelihood estimate $\hat\theta$ is either defined as the maximiser of the likelihood or as the root of the score equation $\ell_\theta(\theta) = 0$. The bias of the maximum likelihood estimator is

$$E_\theta(\hat\theta - \theta) = \alpha(\theta) + O(n^{-2}),$$

with $\alpha(\theta) = O(n^{-1})$ a p-dimensional vector whose expression is given, for instance, in Barndorff-Nielsen and Cox (1994, formula 5·26). In the sequel we are concerned with the approach to bias prevention, i.e., a methodology that adds a penalisation to either the score function or the likelihood to define a new estimator having bias of order n^{-2}.

Firth (1993) introduced the approach to bias prevention based on the penalised score function

$$\ell_\theta(\theta) - v(\theta)\alpha(\theta), \tag{1}$$

where $v(\theta)$ can be either $j(\theta)$ or $i(\theta)$. Although this route to bias prevention is general and has proven to be effective, it poses computational and conceptual challenges. On the one hand, the computation of (1) is possible only when the analytic expression of $\alpha(\theta)$ is available (Kosmidis, 2014a). A substantial effort is therefore being done to elicit the functional form of the penalised score function for large classes of models (Firth, 1993; Kosmidis and Firth, 2009; Kosmidis, 2014b). On the other hand, when $p > 1$ the approach does not necessarily relate to a likelihood function because it is not guaranteed the existence of the penalised likelihood, i.e., the primitive of (1); exceptions are given in Kosmidis and Firth (2009, § 4·2) and when θ is the canonical parameter in exponential family models (Firth, 1993, § 3·1).

In this paper we propose a route to bias prevention based on a penalised likelihood function which is available irrespective of p. The function is obtained through a fully empirical penalisation that depends on the first two derivatives of the likelihood function. The penalisation can be computed numerically and thus allow the routinely use of the proposed methodology.

2 Penalised likelihood function

Let $\theta = (\psi, \lambda)$ with ψ and λ an interest and a nuisance component of dimension p_0 and $q = p - p_0$, respectively. The proposed penalised likelihood function for θ is

$$\bar{\ell}(\theta) = \ell(\theta) + \log |j(\theta)|^{1/2} - \log |k(\theta)|^{1/2} = \ell(\theta) + \Delta(\theta), \tag{2}$$

with $k(\theta) = \sum_{i=1}^{n} \ell_\theta^i(\theta)\{\ell_\theta^i(\theta)\}^{\mathrm{T}}$, whereas the profile version for ψ is $\bar{\ell}\{\bar{\theta}(\psi)\}$, with $\bar{\theta}(\psi)$ the maximiser of $\bar{\ell}(\theta)$ for fixed ψ. We stress that $\bar{\theta}(\bar{\psi}) = \bar{\theta} = (\bar{\psi}, \bar{\lambda})$ is the global maximiser of both functions. Our formulation to bias prevention is related to that of Firth (1993) because

$$\partial\Delta(\theta)/\partial\theta = -v(\theta)\alpha(\theta) + O_p(n^{-1/2}).$$

The bias of $\bar{\theta}$ is $O(n^{-2})$ and, when θ is the true parameter value, the limiting distributions of the penalised profile likelihood ratio statistic and of its directed version

$$\bar{W}\{\bar{\theta}(\psi)\} = 2\left[\bar{\ell}\{\bar{\theta}(\bar{\psi})\} - \bar{\ell}\{\bar{\theta}(\psi)\}\right], \quad \bar{r}\{\bar{\theta}(\psi)\} = \mathrm{sgn}(\bar{\psi} - \psi)\bar{W}\{\bar{\theta}(\psi)\}^{1/2},$$

are chi-square with p_0 degrees of freedom and standard normal, respectively. The conditions and proof of these results, and of any additional result herein enclosed concerning the proposed penalised likelihood function and related quantities, are in a Supplementary Material file not included and available from the Author.

Example 1 (Comparison with likelihood penalised via Jeffreys prior) *Consider a full exponential family of order p with canonical parameter θ. Firth (1993) showed that the primitive function of the penalised score function* (1) *is the likelihood penalised via Jeffreys prior, i.e.,*

$$\ell(\theta) + \log |i(\theta)|^{1/2} = \ell(\theta) + \log |j(\theta)|^{1/2}.$$

The penalised likelihood $\bar{\ell}(\theta)$ differs from the function above by $-\log |k(\theta)|^{1/2}$. In the Supplementary Material it is shown that this term is in fact unnecessary and does not invalidate the result for the bias of $\bar{\theta}$. Firth's penalised likelihood is simpler to compute, thus recommended.

3 Sensitivity to the nuisance component

Let $\hat{\theta}(\psi)$ be maximiser of $\ell(\theta)$ for fixed ψ or the root of $\ell_\lambda(\theta) = 0$ and $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$, with $\ell_\lambda(\theta) = \partial\ell(\theta)/\partial\lambda$. The profile likelihood $\ell\{\hat{\theta}(\psi)\}$ is not a proper likelihood because estimation of the nuisance component entails the failure of Bartlett identities. Failure of the first identity implies that the profile score function $\ell_\psi\{\hat\theta(\psi)\}$ is biased, i.e.,

$$E_\theta\big[\ell_\psi\{\hat\theta(\psi)\}\big] = \tau(\theta) + O(n^{-1}) = O(1),$$

with $\ell_\psi(\theta) = \partial\ell(\theta)/\partial\psi$ and $\tau(\theta) = O(1)$ a p_0-dimensional vector whose expression is given in McCullagh and Tibshirani (1990, Appendix A). Despite that the profile score bias does not bear on the consistency of $\hat\psi$, inferential procedures might behave poorly when the ratio n/q is moderate or small. Penalisation of the profile likelihood is a long-standing procedure to cope with this problem. The penalisation reduces the bias of the resulting profile score function to $O(n^{-1})$. Some examples of these functions, referred to as modified profile likelihoods, are given in Barndorff-Nielsen (1983), Cox and Reid (1987), and DiCiccio and Stern (1993).

Inference for ψ based on $\bar\ell\{\bar\theta(\psi)\}$ is impinged on by the nuisance component as the penalisation reduces the bias of the maximum likelihood estimator rather than the profile score bias. The penalised profile score function $\bar\ell_\psi\{\bar\theta(\psi)$ has bias

$$E_\theta\big[\bar\ell_\psi\{\bar\theta(\psi)\}\big] = \tau(\theta) - \{i^{\psi\psi}(\theta)\}^{-1}\alpha_\psi(\theta) + O(n^{-1}) = O(1), \tag{3}$$

with $\bar\ell_\psi(\theta) = \partial\bar\ell(\theta)/\partial\psi$, and $i^{\psi\psi}(\theta)$ and $\alpha_\psi(\theta)$ the block of $i(\theta)^{-1}$ and the component of $\alpha(\theta)$ pertaining to ψ, respectively. The penalisation in $\bar\ell\{\bar\theta(\psi)\}$ is also effective in reducing the bias of $\bar\ell_\psi\{\bar\theta(\psi)\}$ to $O(n^{-1})$ only when the bias of the profile score function and of the maximum likelihood estimator are related.

Example 2 (Simultaneous bias reduction) *Let y_1 and y_2 be samples of size n from independent exponential random variables with expectation ψ/λ and $\psi\lambda$, respectively. The profile score function is $\ell_\psi\{\hat\theta(\psi)\} = i_{\psi\psi}\{\hat\theta(\psi)\}(\hat\psi - \psi)$ and $\{i^{\psi\psi}(\theta)\}^{-1} = i_{\psi\psi}(\theta)$ because $i_{\psi\lambda}(\theta) = 0$. From (3) follows $E_\theta\big[\bar\ell_\psi\{\bar\theta(\psi)\}\big] = O(n^{-1})$.*

3.1 Two-index asymptotic setting

Consider the model for independent stratified observations $\{f(y;\psi,\lambda_j),\ j = 1,\dots,q\}$ as in Sartori (2003). The likelihood and score functions for $\theta = (\psi,\lambda_1,\dots,\lambda_q)$ based on $n = \sum_{j=1}^q m_j$ observations $y_{jk} = (y_{j1},\dots,y_{jm_j})$ are

$$\ell(\theta) = \sum_{j=1}^{q}\sum_{k=1}^{m_j}\ell^{jk}(\theta), \quad \ell_\theta(\theta) = \sum_{j=1}^{q}\ell_\theta^{j}(\theta) = \sum_{j=1}^{q}\sum_{k=1}^{m_j}\ell_\theta^{jk}(\theta) = \sum_{j=1}^{q}\sum_{k=1}^{m_j}\partial\ell^{jk}(\theta)/\partial\theta,$$

with $\ell^{jk}(\theta) = \log f(y_{jk};\psi,\lambda_j)$. In this framework both the dimension of the nuisance component or, equivalently, the number of strata q and the observations in each stratum m_j might diverge. The goal is to study how the structure of the sequence $\{q, m_j\}$ affects inferential procedures for ψ. In the following we assume that $m_j = m$, i.e., $n = qm$. Furthermore, we highlight that $E_\theta(\hat\theta - \theta) = \alpha(\theta) + O(m^{-2}) = O(m^{-1})$ and that $k(\theta) = \sum_{j=1}^q\sum_{k=1}^m \ell_\theta^{jk}(\theta)\{\ell_\theta^{jk}(\theta)\}^{\mathrm{T}}$.

The discussion for the profile likelihood provided earlier applies verbatim when q is fixed. Conversely, the bias of the profile score function accumulates across strata and is

$$E_\theta\big[\ell_\psi\{\hat\theta(\psi)\}\big] = \sum_{j=1}^{q} E\big[\ell_\psi^{j}\{\hat\theta(\psi)\}\big] = \sum_{j=1}^{q}\{\tau^j(\theta) + O(m^{-1})\} = O(q).$$

Sartori (2003) showed that this result modifies the first-order theory for $\hat{\psi}$ because $\hat{\psi} - \psi = O_p(n^{-1/2})$ if $q/m = o(1)$, $\hat{\psi} - \psi = O_p(m^{-1})$ otherwise, i.e., the maximum likelihood estimator is root-n consistent only when the sample size within each stratum increases faster than the number of strata. Modified profile likelihoods are less sensitive to the dimension of the nuisance component as the bias of the modified profile score function is $O(q/m)$ and the corresponding estimator $\hat{\psi}_{\mathrm{M}}$ satisfies $\hat{\psi}_{\mathrm{M}} - \psi = O_p(n^{-1/2})$ if $q/m^3 = o(1)$, $\hat{\psi}_{\mathrm{M}} - \psi = O_p(m^{-2})$ otherwise. Analogue conditions on q and m bear on the asymptotic theory of likelihood ratio-type statistics, and directed versions, derived from profile and modified profile likelihoods; see Sartori (2003).

The bias of the penalised profile score function is

$$E_\theta\left[\bar{\ell}_\psi\{\bar{\theta}(\psi)\}\right] = \sum_{j=1}^{q}\{\tau^j(\theta) + O(m^{-1})\} - \{i^{\psi\psi}(\theta)\}^{-1}\alpha_\psi(\theta) = \max\{O(1), O(q/m)\}. \tag{4}$$

Technical details, addressed in the Supplementary Material, prevent the derivation of convergence rates for $\bar{\psi}$ as those outlined for $\hat{\psi}$ and $\hat{\psi}_{\mathrm{M}}$. Nonetheless the comparison of profile, modified profile, and penalised profile score function bias leads to the following remarks. First, the proposed penalised profile likelihood $\bar{\ell}\{\bar{\theta}(\psi)\}$ is preferable to the profile likelihood as the penalisation eliminates most of the profile score bias while being effective in reducing the bias of the maximum likelihood estimator. Second, unless $q/m = o(1)$, the penalised profile likelihood safeguards inference for ψ against the nuisance component as a modified profile likelihood and can therefore be used whenever the latter function is difficult to compute.

Example 3 (Neyman-Scott problem revisited) *Let y_{jk} be a sample from independent normal random variables with means λ_j and variance ψ, $j = 1, \ldots, q$, $k = 1, \ldots, m$. For fixed m, the maximum likelihood estimator for ψ is inconsistent as discussed, for instance, in McCullagh and Tibshirani (1990, § 3·1). The use of the penalised profile likelihood leads to a consistent estimator because* (4) *gives $E_\theta\left[\bar{\ell}_\psi\{\bar{\theta}(\psi)\}\right] = 0$, being $\ell\{\hat{\theta}(\psi)\} = mq(\hat{\psi}-\psi)/(2\psi^2)$, $\{i^{\psi\psi}(\theta)\}^{-1} = mq/(2\psi^2)$, and $E_\theta(\hat{\psi} - \psi) = -\psi/m$. Our result is consistent with the findings in Firth (1993, § 4·5).*

4 Monte Carlo simulation

4.1 Beta regression

To show that the approach to bias prevention based on (1) and (2) are equivalent, we consider an illustrative example concerning a generalised linear model with beta responses (Kosmidis and Firth, 2009). Suppose that $y_1, \ldots, y_n$ is a sample from independent beta random variables with expectation μ_i and variance $\mu_i(1 - \mu_i)/(1 + \phi)$, $i = 1, \ldots, n$. The mean of the ith unit is related to a set of covariates through the link function $g(\cdot)$, i.e., $\mu_i = g^{-1}(\beta^{\mathrm{T}} x_i)$, where β is the vector of regression coefficients, $\phi > 0$ the precision parameter, and $x_i = (1, x_{i1}, \cdots, x_{ip^*})^{\mathrm{T}}$, with $p^* = p - 2$. Let $\theta = (\beta, \phi)$ be the parameter of interest.

Simulation are run by setting $n = 25, 75$, $\theta = (0{\cdot}1, 0, -0{\cdot}1, 0{\cdot}2, 2)$, $g(t) = e^t/(1 + e^t)$, and by generating x_{i1}, x_{i2}, and x_{i3} from the following random variables: Poisson with mean 2, exponential with unit mean, and standard normal, respectively. The penalisation in the proposed likelihood (2) is computed with numerical derivative routines, whereas the penalised score function (1) with the R function betareg which allows to set $\nu(\theta)$ equal to either $j(\theta)$ or $i(\theta)$; the function is in the betareg package (Gruen *et al.*, 2012; R Core Team, 2013).

In Table 1 we report Monte Carlo estimates of bias and variance of estimators based on 40 000 replicates. The result of estimators derived from the proposed penalised likelihood and from the penalised score are similar, indicating that the approaches are equivalent. Since the bias of the maximum likelihood estimator for the regression coefficient is in fact negligible, the effectiveness of bias prevention strategies and their equivalence is better seen by focusing on the results for the estimator of the precision parameter.

Table 1: Beta regression. Monte Carlo estimates of bias and variance (in parentheses) of estimators derived from the likelihood (Lik.), penalised likelihood (Pen. lik.), and penalised score function (Pen. score) computed with $j(\theta)$ and $i(\theta)$. All entries are multiplied by 100

n		Lik.	Pen. lik.	Pen. score $j(\theta)$	Pen. score $i(\theta)$
25	β_1	0.28 (22.61)	0.07 (20.13)	0.12 (20.91)	−0.08 (20.30)
	β_2	−0.03 (3.47)	0.02 (3.07)	−0.01 (3.30)	−0.01 (3.19)
	β_3	−0.16 (3.74)	0.23 (3.31)	−0.11 (3.80)	0.12 (3.61)
	β_4	0.89 (12.28)	−0.21 (11.05)	0.12 (11.50)	−0.17 (11.15)
	ϕ	58.69 (59.80)	3.48 (35.93)	−1.48 (34.17)	−2.00 (33.99)
75	β_1	−0.06 (6.72)	−0.03 (6.50)	0.01 (6.59)	−0.02 (6.56)
	β_2	0.01 (0.80)	0.01 (0.78)	−0.01 (0.79)	0.01 (0.79)
	β_3	−0.01 (1.99)	0.02 (1.93)	−0.04 (2.01)	−0.01 (2.00)
	β_4	0.20 (1.72)	−0.06 (1.67)	0.06 (1.70)	0.02 (1.69)
	ϕ	16.11 (10.14)	−0.13 (8.57)	−0.36 (8.59)	−0.39 (8.58)

4.2 Gamma samples

The purpose of this example is to compare the sensitivity of inferential procedures derived from $\bar{\ell}\{\bar{\theta}(\psi)\}$ with those derived from profile likelihood and modified profile likelihood. Let y_{jk} be a sample from independent gamma random variables with shape and scale parameters ψ and $1/\lambda_j$, respectively, $j = 1, \ldots, q$, $k = 1, \ldots, m$. In Sartori (2003) are provided expressions for the profile and Barndorff-Nielsen's modified profile likelihood. Since the considered model is full exponential with canonical parameter $\theta = (\psi, \lambda_1, \ldots, \lambda_q)$, we can also consider the profile version of Firth's penalised likelihood given in Example 1.

Simulation are run for several values of m and q and by setting $\psi = 1$ and λ_j generated from a chi-squared variate with 10 degrees of freedom, $j = 1, \ldots, q$. Results for quantities related to $\bar{\ell}\{\bar{\theta}(\psi)\}$ are given for $q \leq 30$ because for these values computations are feasible.

In Table 2 we report Monte Carlo estimates of bias and variance of estimators based on 20 000 replicates; we denoted by $\hat{\psi}_{\mathrm{F}}$ the estimator derived from Firth's penalised profile likelihood. The maximum likelihood estimator is strongly biased even when $m \geq q$. The comparison of estimators $\hat{\psi}_{\mathrm{M}}$, $\hat{\psi}_{\mathrm{F}}$, and $\bar{\psi}$ highlights that the bias prevention approach provides results which are comparable to the ones obtained from a modified profile likelihood. The results for $\hat{\psi}_{\mathrm{F}}$ and $\bar{\psi}$ are close, confirming the claim made in Example 1.

In Figure 1 are depicted the Monte Carlo estimates of distribution functions of the directed versions of the profile, r, modified profile, r_{M}, Firth's penalised profile, r_{F}, likelihood ratio and $\bar{r}$. As expected, the distribution function of all statistics converge to the standard normal distribution for fixed q as m increases. Contrary to the distribution function of r the ones of r_{F} and $\bar{r}$ are stable when m is fixed and q increases as the distribution function of r_{M}.

Table 2: Gamma samples. Monte Carlo estimates of bias and variance (in parentheses) of estimators. All entries are multiplied by 100

m		$q=5$	$q=15$	$q=30$	$q=100$	$q=400$
5	$\hat{\psi}_{\mathrm{M}}$	9.59 (11.33)	3.41 (2.81)	2.02 (1.33)	1.01 (0.38)	0.72 (0.09)
	$\hat{\psi}$	30.58 (17.20)	22.89 (4.23)	21.17 (2.00)	19.91 (0.57)	19.55 (0.14)
	$\hat{\psi}_{\mathrm{F}}$	−4.21 (9.26)	−3.97 (2.63)	−3.87 (1.29)	−3.87 (0.37)	−3.85 (0.09)
	$\bar{\psi}$	2.25 (9.91)	−0.47 (2.65)	−0.89 (1.28)		
10	$\hat{\psi}_{\mathrm{M}}$	3.82 (3.94)	1.49 (1.19)	0.79 (0.57)	0.33 (0.17)	0.19 (0.04)
	$\hat{\psi}$	12.70 (4.78)	10.12 (1.45)	9.34 (0.69)	8.83 (0.21)	8.68 (0.05)
	$\hat{\psi}_{\mathrm{F}}$	−0.94 (3.60)	−0.67 (1.16)	−0.74 (0.56)	−0.76 (0.17)	−0.77 (0.04)
	$\bar{\psi}$	−0.14 (3.64)	−0.28 (1.16)	−0.42 (0.56)		
15	$\hat{\psi}_{\mathrm{M}}$	2.51 (2.42)	0.90 (0.75)	0.49 (0.37)	0.18 (0.11)	0.09 (0.03)
	$\hat{\psi}$	8.16 (2.75)	6.44 (0.86)	6.00 (0.42)	5.67 (0.13)	5.57 (0.03)
	$\hat{\psi}_{\mathrm{F}}$	−0.32 (2.29)	−0.28 (0.74)	−0.29 (0.37)	−0.32 (0.11)	−0.32 (0.03)
	$\bar{\psi}$	−0.07 (2.30)	−0.17 (0.74)	−0.20 (0.37)		

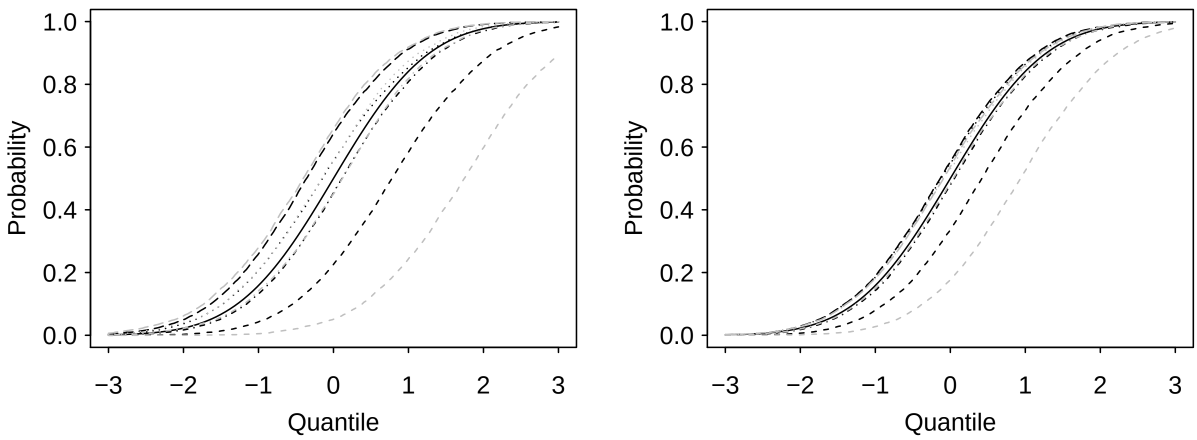

Figure 1: Gamma samples. Monte Carlo estimates of distribution functions. Short-dashed line, r, dotted-dashed line, r_{M}, long-dashed line r_{F}, dotted line, $\bar{r}$, solid line, standard normal. Left panel, $m=5$, $q=5$ (black), $q=30$ (gray); right panel, $m=15$, $q=5$ (black), $q=30$ (gray)

References

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. B*, **49**, 1–39.

DiCiccio, T. J. and Stern, S. E. (1993). *An adjustment to profile likelihood based on observed information*. Stanford University. Department of Statistics.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

Gruen, B., Kosmidis, I., and Zeileis, A. (2012). Extended beta regression in R: Shaken, stirred, mixed, and partitioned. *J. Statist. Soft.*, **48**, 1–25.

Kosmidis, I. (2014a). Bias in parametric estimation: reduction and useful side-effects. *WIREs: Comput. Stat.*, **6**, 185–196.

Kosmidis, I. (2014b). Improved estimation in cumulative link models. *J. Roy. Statist. Soc. B*, **76**, 169–196.

Kosmidis, I. and Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika*, **96**, 793–804.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc. B*, **52**, 325–344.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**, 533–549.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITA DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy