Pseudo-Likelihoods for Bayesian Inference

Laura Ventura and Walter Racugno

Abstract

The interplay between Bayesian and frequentist inference can play a remarkable role in order to address some theoretical and computational drawbacks, due to the complexity or misspecification of the model, or to the presence of many nuisance parameters. In this respect, in this paper we review the properties and applications of the so-called pseudo-posterior distributions, i.e., posterior distributions derived from the combination of a pseudo-likelihood function with suitable prior information. In particular, we illustrate the various notions of pseudo-likelihood highlighting their use in the Bayesian framework. Moreover, we show the simple but effective application of pseudo-posterior distributions in three challenging examples.

1 Introduction

In the presence of models with complicated dependence structures, of multidimensional nuisance parameters, or of model misspecifications, both frequentist and Bayesian inference may encounter some theoretical and computational difficulties. Indeed, in these situations the original likelihood function may be intractable or computationally cumbersome. In order to take into proper account of such difficulties,

L. Ventura (✉)
Department of Statistics, University of Padova, Padova, Italy
e-mail: ventura@stat.unipd.it

W. Racugno
Department of Mathematics and Informatics, University of Cagliari, Cagliari, Italy
e-mail: racugno@unica.it

T. Di Battista et al. (eds.), *Topics on Methodological and Applied Statistical Inference*, Studies in Theoretical and Applied Statistics,
DOI 10.1007/978-3-319-44093-4_19

it is possible to consider surrogates of the original likelihood, which produce the wide class of the so-called *pseudo-likelihoods*; see, for instance, [55, Chap. 4], [71, Chaps. 8 and 9], and [76], and references therein.

The aim of this paper is to review the properties and to illustrate some applications of the so-called pseudo-posterior distributions, i.e., distributions derived from the combination of a pseudo-likelihood function with suitable prior information. It is a Bayesian non-orthodox procedure widely used in the recent statistical literature and theoretically motivated in several papers; see, among others [4,11,12,17,19–21,30,34,36,40,46,51,58,60,63,67–69,73,77–79,81], and references therein.

The outline of the paper is as follows. Section 2 gives a brief review on pseudo-likelihood functions. Section 3 introduces the notion of pseudo-posterior distribution, discusses the choice of the prior and the validation of a pseudo-posterior distribution, also through first and higher-order asymptotic results. In Sect. 4 we illustrate the calculation of pseudo-posterior distributions using a one-way random effects model with heteroscedastic error variances, the Cox proportional hazards model, and a multilevel probit model. Finally, some concluding remarks close the paper.

2 Notion of Pseudo-Likelihood

Let $y = (y_1, \ldots, y_n)$ be a random sample of size n from a statistical model with parameter space Θ, not necessarily finite-dimensional. Let $\tau = \tau(\theta)$, with $\tau \in T \subseteq \mathbb{R}^k$, $k \geq 1$, be the parameter of interest. The more complex is the component complementary to τ in θ, then the more useful is the possibility of basing inference on a likelihood function which depends on τ only.

Let us denote with $L_{ps}(\tau) = L_{ps}(\tau; y)$ a pseudo-likelihood function for τ, that is a function of the data y which depends only on the parameter of interest τ and which behaves, in some respects, as it were a genuine likelihood. This means that, under mild regularity conditions, $L_{ps}(\tau)$ has unbiased score function, the pseudo-maximum likelihood estimator $\hat{\tau}_{ps}$ is consistent and asymptotically normal, and the pseudo-likelihood ratio test $W_{ps}(\tau) = 2(\ell_{ps}(\hat{\tau}_{ps}) - \ell_{ps}(\tau))$, with $\ell_{ps}(\tau) = \log L_{ps}(\tau)$, has null asymptotic χ^2_k distribution. Some well-known examples of pseudo-likelihood functions are the marginal, the conditional, the profile, the approximate conditional, the modified profile, the integrated, the partial, the quasi, the empirical, the weighted, the composite and the pairwise likelihood. For reviews on pseudo-likelihood functions see, e.g., [55, Chap. 4], [71, Chaps. 8 and 9], and [76], and references therein.

There are several reasons for introducing a pseudo-likelihood function for inference on τ. Here we propose a possible taxonomy of pseudo-likelihoods based on three main classes.

1. Elimination of nuisance parameters. Consider a parametric model with density function $p(y; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p > 1$, and write $\theta = (\tau, \lambda)$, where the nuisance parameter λ is of dimension $p - k$. For inference on τ, pseudo-likelihoods based on a statistical model defined as a reduction of the original model are the *marginal* and the *conditional likelihoods* [71, see Chap. 8]. However, they are available essentially only in exponential and in group families. Outside of these cases, one simple and general way of obtaining a pseudo-likelihood for τ is to replace the nuisance parameter λ with its maximum likelihood estimate (MLE) for fixed τ, i.e., $\hat{\lambda}_\tau$, in the original likelihood $L(\tau, \lambda)$. The corresponding function $L_p(\tau) = L(\tau, \hat{\lambda}_\tau)$ is the well-known *profile likelihood*. It is not a genuine likelihood and its behavior may not be entirely satisfactory, especially when the dimension of λ is large. Various modifications of $L_p(\tau)$ have been proposed, starting from the *approximate conditional likelihood* of [24], which is based on the choice of an orthogonal parameterization, to the various proposals of *modified profile likelihoods*, which require notions about higher-order asymptotic methods (see [71, Chap. 9]). All the available modifications of the profile likelihood are equivalent to the second order and share the common feature of reducing the score bias to $O(n^{-1})$ (see, e.g., [56]). A further approach that can be applied generally for the elimination of nuisance parameters is to average the likelihood function $L(\tau, \lambda)$ with respect to a "weight" function $\pi(\lambda)$ on λ, in order to define the *integrated likelihood function* $L_I(\tau) = \int L(\tau, \lambda)\, \pi(\lambda)\, d\lambda$ (see [71, Chap. 8], [10]).

2. Semi or nonparametric models. The *quasi-likelihood* (see [2,6,8,48]) is a pseudo-likelihood function associated to a semi parametric model specified in terms of first (and sometimes second) order moments of a particular unbiased estimating function. Instead, the *empirical likelihood* [54] was introduced to deal with inference problems on k-dimensional smooth functionals in nonparametric models. The study of these pseudo-likelihoods, when derived from M-estimators, has been investigated in [1,3,54]. When robustness with respect to influential observations or to model misspecifications is of interest, also the *weighted likelihood* can be considered (see, e.g., [38,47]), which is a pseudo-likelihood defined through a set of weights which are supposed to opportunely down-weight likelihood single term components.

3. Complex models. The class of *composite likelihoods* (see [76], and references therein) is useful when the fully specified likelihood is computationally cumbersome as well as when a fully specified model is out of reach. This class contains the ordinary likelihood, as well as many other interesting alternatives, such as the *Besag pseudo-likelihood* [13], the *m-order likelihood* for stationary processes [5], the *approximate likelihood* of [74], and the *composite marginal likelihood* and the *pairwise likelihood* [26], constructed from marginal densities. Also the *partial likelihood* [22,23], introduced for inference about the regression coefficients in the proportional hazards model, may be considered a member of this class.

Finally, we remark that since the 1970s numerous other pseudo-likelihoods have been considered. Some of these are: the pseudo-likelihood of [35], where nuisance parameters are eliminated by means of a simple plug-in estimate; the *bootstrap likelihood* [28,29], which is in the spirit of empirical likelihood; the *dual likelihood* [53], which associates a likelihood to a martingale estimating equation; the *projected likelihood* [49,82] for semi parametric models; the *penalized likelihood* [25,37] for an infinite-dimensional parameter of interest such as a density or a regression function;

the various instances of *predictive likelihood* [14,16]; the *h-likelihood* [44,45,50], that is a hierarchical likelihood, for inferences from random effect models.

3 Pseudo-Posterior Distributions

Assuming a prior distribution $\pi(\tau)$ on τ and treating $L_{ps}(\tau)$ as an ordinary likelihood, from a purely formal expression of Bayes' theorem we obtain

$$\pi_{ps}(\tau|y) \propto \pi(\tau)\ L_{ps}(\tau)\ . \tag{1}$$

The posterior distribution $\pi_{ps}(\tau|y)$ is obtained "miming" the Bayesian procedure and thus is called *pseudo-posterior*. In general, Bayesian inferential procedures based on pseudo-likelihoods are called *hybrid*, or *quasi* or *pseudo* Bayesian methods.

When basing inference on τ on the pseudo-posterior distribution $\pi_{ps}(\tau|y)$, three issues need to be addressed

(a) the choice of the suitable pseudo-likelihood $L_{ps}(\tau)$;
(b) the choice of the prior $\pi(\tau)$;
(c) the validation of inference based on $\pi_{ps}(\tau|y)$.

Section 3.1 focuses on the choice of the pseudo-likelihood to be used in (1), which depends on the model and the objectives of the analysis. Section 3.2 reviews the results on the choice of the prior. Finally, Sect. 3.3 discusses the validation of a pseudo-posterior distribution, both numerically and through asymptotic results.

3.1 Areas of Application of Pseudo-Posterior Distributions

Although (1) cannot always be considered as orthodox in a Bayesian setting, the use of alternative likelihoods is nowadays widely shared, and several papers focus on the Bayesian application of some well-known pseudo-likelihoods. Of course, the choice of the pseudo-likelihood to be used in (1) depends on the objectives of the analysis. A possible classification of the main areas of applications of the pseudo-posterior $\pi_{ps}(\tau|y)$ may be based on the following five classes.

Elimination of nuisance parameters. When $\theta = (\tau, \lambda)$ and only inference on τ is of interest, the marginal, the conditional, the modified profile, and the approximate conditional likelihoods can be used in (1). Note that the use of these pseudo-likelihoods in $\pi_{ps}(\tau|y)$ has the advantages of avoiding the elicitation on the nuisance parameter λ and of the computation of a multidimensional integral necessary to compute the marginal posterior distribution for τ. Moreover, these pseudo-likelihood functions $L_{ps}(\tau)$ have an orthodox Bayesian interpretation. This means that they are equivalent to a suitable integrated likelihood, of the form $L_I(\tau) = \int L(\tau,\lambda)\,\pi(\lambda|\tau)\,d\lambda$, for a specific conditional prior $\pi(\lambda|\tau)$ (see, e.g., [57,70]). As a further remark, note that the pseudo-posterior distribution $\pi_{ps}(\tau|y)$ is a genuine posterior distribution when using in (1) the modified profile likelihood with the corresponding matching prior (see [77,81]) or in non-normal regression-scale models, in which there is no loss of information about τ when using a pseudo-posterior distribution derived from a marginal likelihood (see [60]). For Bayesian applications of the marginal, the conditional, the modified profile, and of the approximate conditional likelihoods see, among others [10–12,17,19,20,32,34,51,60,64,70,77,79–81], and references therein.

Semi or nonparametric models. When dealing with semi parametric or nonparametric statistical models, for Bayesian inference on τ the quasi and the empirical likelihoods can be used. Note that the use of these pseudo-likelihoods in $\pi_{ps}(\tau|y)$ has the advantages of requiring the elicitation of the prior only on the parameter of interest τ. For applications of these pseudo-likelihoods for Bayesian inference see [42,46,60,68,78], and references therein.

Robustness. When robustness with respect to outliers, influential observations or model misspecifications is required, the quasi, the empirical and the weighted likelihoods can be used to obtain resistant pseudo-posterior distributions. Indeed, the occurrence of anomalous values can seriously alter the shape of the ordinary likelihood function and then lead to ordinary posterior distributions far from those one would obtain without these data inadequacies, as illustrated in [4,36,78].

Complex models. The composite and pairwise likelihoods deal with complex statistical models, for which the ordinary likelihood and thus the ordinary posterior distribution are impractical to compute or even analytically unknown. The use of these pseudo-likelihood in Bayesian inference has been discussed in [58,63,65,67,73].

Proportional hazards model. In the Bayesian framework, the use of the partial likelihood to derive a posterior distribution on the regression parameters of the Cox model has the advantage of avoiding the specification of a prior process on the unknown baseline cumulative hazard function. For the use of this pseudo-likelihood in Bayesian inference, see, among others [21,39,40,67,69].

3.2 Choice of the Prior

The choice of the prior distribution on τ in (1) involves the same problems typical of the standard Bayesian perspective. In particular, this occurs both when the elicitation of a proper prior distribution is required and when using default prior distributions that are often improper. For instance, the choice of parametric priors in $\pi_{ps}(\tau|y)$ has been considered in several papers (see, e.g., [4,36,40,42,58,60,67,73]).

Non-informative priors have been considered by [21,58,60]. Ventura et al. [78] discuss how to modify the Jeffreys' prior to yield a default prior for τ to be used with a general pseudo-likelihood $L_{ps}(\tau)$. It is shown that the Jeffreys-type prior for τ associated to $L_{ps}(\tau)$ is given by

$$\pi^J_{ps}(\tau) \propto \sqrt{|i_{ps}(\tau)|}\ , \qquad (2)$$

where $i_{ps}(\tau)$ is the pseudo-expected information matrix, i.e.,

$$i_{ps}(\tau) = E(-\partial^2 \ell_{ps}(\tau)/\partial\tau\partial\tau^{\mathsf{T}}) .$$

This means that a parametrization invariant prior distribution for τ, derived from a pseudo-likelihood function, is proportional to the square root of the determinant of the pseudo-expected information.

The other prominent studied default priors are the matching priors, designed to produce Bayesian credible sets which are optimal frequentist confidence sets in a certain asymptotic sense (see, e.g., [27]). The use of matching priors has been widely discussed in (1) with $L_{ps}(\tau)$ denoting a marginal, conditional or modified profile likelihood for a scalar parameter of interest τ; see, e.g., [17, 19, 51, 61, 64, 77, 79–81]. For instance, when using the modified profile likelihood, the corresponding matching prior is (see [77]),

$$\pi_{mp}(\tau) \propto i_{\tau\tau.\lambda}(\tau, \hat{\lambda}_\tau)^{1/2} , \tag{3}$$

with $i_{\tau\tau.\lambda}(\tau, \lambda) = i_{\tau\tau}(\tau, \lambda) - i_{\tau\lambda}(\tau, \lambda) i_{\lambda\lambda}(\tau, \lambda)^{-1} i_{\lambda\tau}(\tau, \lambda)$ partial information, and $i_{\tau\tau}(\tau, \lambda)$, $i_{\tau\lambda}(\tau, \lambda)$, $i_{\lambda\lambda}(\tau, \lambda)$ and $i_{\lambda\tau}(\tau, \lambda)$ blocks of the expected Fisher information from the genuine likelihood $L(\tau, \lambda)$.

3.3 Validation of the Pseudo-Posterior Distribution

The pseudo-posterior distribution $\pi_{ps}(\tau|y)$ calls for its validation for Bayesian inference. At the current state, a general finite-sample theory for pseudo-posterior distributions is not available, and every single problem has to be examined.

For the pseudo-posterior distributions listed in Sect. 3.1, the validation may be based on asymptotic results. In particular, paralleling the results for the full posterior distribution and under standard regularity conditions, it can be shown that (see [36, 42, 58])

$$\pi_{ps}(\tau|y) \mathrel{\dot\sim} N_k\left(\hat{\tau}_{ps}, j_{ps}(\hat{\tau}_{ps})^{-1}\right) , \tag{4}$$

where $j_{ps}(\hat{\tau}_{ps})$ is the pseudo-observed information evaluated at the pseudo-MLE. An asymptotically equivalent normal approximation is $\pi_{ps}(\tau|y) \mathrel{\dot\sim} N_k\left(\tilde{\tau}_{ps}, \tilde{j}_{ps}(\tilde{\tau}_{ps})^{-1}\right)$, where $\tilde{\tau}_{ps}$ is the pseudo-posterior mode and $\tilde{j}_{ps}(\tilde{\tau}_{ps}) = -(\partial \log L_{ps}(\tau))/(\partial\tau\partial\tau^{\mathsf{T}})|_{\tau=\tilde{\tau}_{ps}}$. Moreover, paralleling results for the full posterior distribution, also a higher-order tail area approximation can be derived for a scalar parameter of interest τ (see [67]). In particular, it holds

$$\int_{\tau_0}^{\infty} \pi_{ps}(\tau|y)\, d\tau \mathrel{\ddot=} \Phi(r^*_{ps}(\tau_0)) , \tag{5}$$

where $\Phi(\cdot)$ is the standard normal distribution function and

$$r^*_{ps}(\tau) = r_{ps}(\tau) + r_{ps}(\tau)^{-1} \log b(r_{ps}(\tau)) ,$$

with

$$r_{ps}(\tau) = sign(\hat{\tau}_{ps} - \tau)[2(\ell_{ps}(\hat{\tau}_{ps}) - \ell_{ps}(\tau))]^{1/2}$$

pseudo-signed likelihood root and

$$b(r_{ps}(\tau)) = j_{ps}(\hat{\tau}_{ps})^{1/2} \frac{r_{ps}(\tau)}{\ell'_{ps}(\tau)} \frac{\pi(\tau)}{\pi(\hat{\tau}_{ps})} .$$

The symbol "$\stackrel{..}{=}$" in (5) indicates that the approximation holds with error of order $O(n^{-3/2})$. From a practical point of view, the tail area approximation (5) can be used to compute posterior quantiles of τ, or equi-tailed credible intervals as $\{\tau : |r^*_{ps}(\tau)| \leq z_{1-\alpha/2}\}$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. Moreover, it can be used to approximate posterior moments or highest posterior density (HPD) credible intervals when using the HOTA algorithm (see [64,67]). The HOTA algorithm is essentially an inverse transform sampling method, which gives independent samples from the pseudo-posterior distribution.

A numerical possibility for a finite-sample validation of Bayesian inference based on $\pi_{ps}(\tau|y)$ is to use the procedure by Mohanan-Boos (1992). These authors discuss a criterion for evaluating whether or not an alternative likelihood can be used for Bayesian inference and, to this end, they introduce a definition of validity, based on the coverage properties of posterior credible sets. In practice, they compute the statistic $H = \int_{-\infty}^{\tau} \pi_{ps}(t|y)\, dt$, which corresponds to posterior coverage set functions of the form $(-\infty, t^\alpha]$, where t^α is the αth percentile of the pseudo-posterior distribution. They assume that $\pi_{ps}(\tau|y)$ is valid by coverage if H is uniformly distributed in $(0, 1)$. Validity of Bayesian inference for the empirical likelihood was assessed in [42], for the quasi-likelihood in [36], and for the weighted likelihood in [4].

4 Three Examples of Pseudo-Posterior Distributions

In this section we illustrate the calculation of pseudo-posterior distributions in three illustrative examples based on: the modified profile likelihood in a one-way random effects model with heteroscedastic error variances, the partial likelihood in the Cox proportional hazards model, and the composite likelihood in a multilevel probit model. It is argued that pseudo-posterior distributions have an important role to play in Bayesian statistics.

4.1 Elimination of Nuisance Parameters with Matching Priors

Let $\theta = (\tau, \lambda)$, with τ scalar parameter of interest and λ multidimensional nuisance parameter. Bayesian inference on τ is based on the marginal posterior distribution

$$\pi_m(\tau|y) = \int \pi(\theta|y)\, d\lambda = \frac{\int \pi(\tau,\lambda) L(\tau,\lambda)\, d\lambda}{\int \pi(\tau,\lambda) L(\tau,\lambda)\, d\lambda d\tau} . \qquad (6)$$

The computation of (6) may present some difficulties. First of all, it requires the elicitation on both ψ and λ. Second, it requires a multidimensional numerical integration.

These drawbacks can be avoided when using the class of matching priors in $\pi_m(\tau|y)$. In this case, the marginal posterior distribution can be written as (see, e.g., [81], and references therein)

$$\pi_m(\tau|y) \propto \pi_{mp}(\tau) L_{mp}(\tau) \ , \tag{7}$$

where $\pi_{mp}(\tau)$ is the matching prior (3), and $L_{mp}(\tau) = L_p(\tau) M(\tau)$ is the modified profile likelihood for τ with $M(\tau)$ suitable defined correction term. The advantages of (7) are that: (1) no elicitation on the nuisance parameter λ is required; (2) no numerical integration or MCMC simulation is needed; (3) accurate Bayesian inference even for small sample sizes. Moreover, it can routinely be applied in practice using results from likelihood asymptotics and the `R` package bundle `hoa` (see [81]).

Accurate tail probabilities from (7) can be computed using the third-order approximation (5), which reduces to (see also [80])

$$\int_{\tau_0}^{\infty} \pi_m(\tau|y)\, d\tau \ \ddot{=}\ \Phi(r_p^*(\tau_0)) \ , \tag{8}$$

where

$$r_{ps}^*(\tau) = r_{ps}(\tau) + r_{ps}(\tau)^{-1} \log \frac{q(\tau)}{r_p(\tau)}$$

is the modified directed profile likelihood of [7], with

$$q(\tau) = \ell_p'(\tau) \frac{i_{\tau\tau.\lambda}(\hat{\tau}, \hat{\lambda})^{1/2}}{i_{\tau\tau.\lambda}(\tau, \hat{\lambda}_\tau)^{1/2}} \frac{1}{M(\tau)} \ .$$

The prior $\pi_{mp}(\tau)$ is also a strong matching prior [33] since a frequentist p-value coincides with a Bayesian posterior survivor probability. Moreover, note that the equi-tailed credible interval $\{\mu : |r_p^*(\tau)| \leq z_{1-\alpha/2}\}$ for τ derived from (8) coincides with an accurate higher-order likelihood-based confidence interval for τ with approximate level $(1-\alpha)$. Therefore, this credible interval is also a likelihood-based confidence interval for τ, with accurate frequentist coverage.

In order to illustrate the use of (7), consider inference for the consensus mean in inter-laboratory studies. The analysis of data from inter-laboratory studies has received attention over the past several years, and it deals with the one-way random effects model with heteroscedastic error variances; see, among others [72], and references therein. Let us assume that there are m laboratories, with n_j observations at the j-th laboratory, for $j = 1, \ldots, m$. The model is

$$y_{ij} = \tau + \tau_j + \varepsilon_{ij} \ , \qquad i = 1, \ldots, n_j \ , \quad j = 1, \ldots, m \ , \tag{9}$$

where y_{ij} denotes the i-th observation at the j-th laboratory, and τ_j and ε_{ij} are independent random variables with distribution $\tau_j \sim N(0, \sigma^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_j^2)$, respectively. The parameter of interest is the consensus mean τ, which is also the mean of the y_{ij}, $i = 1, \ldots, n_j$ and $j = 1, \ldots, m$. The remaining $(m+1)$ parameters of the model, i.e., within-laboratory variances $(\sigma_1^2, \ldots, \sigma_m^2)$ and between laboratory variability σ^2, are nuisance parameters. Consider the marginal posterior distribution for τ based on the matching prior $\pi_{mp}(\tau)$. With respect to a standard Bayesian approach (see, e.g., [75]), it does not require the elicitation on the nuisance parameter

$\lambda = (\sigma^2, \sigma_1^2, \ldots, \sigma_m^2)$ and it enables us to perform simple and accurate Bayesian inference also when m and/or the n_j, $j = 1, \ldots, m$, are small. The log likelihood function for τ and $\lambda = (\sigma^2, \sigma_1^2, \ldots, \sigma_m^2)$ from model (9) is given by

$$\ell(\tau, \lambda) = -\frac{1}{2} \sum_{j=1}^{m} \left((n_j - 1) \log \sigma_j^2 - \log \rho_j + \rho_j (\bar{y}_j - \tau)^2 + \frac{(n_j - 1)s_j^2}{\sigma_j^2} \right),$$

with $\rho_j = 1/(\sigma^2 + \sigma_j^2/n_j)$, $\bar{y}_j = \sum_{i=1}^{n_j} y_{ij}/n_j$ and $s_j^2 = \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2/(n_j - 1)$, for $j = 1, \ldots, m$. Starting from $\ell(\tau, \lambda)$, all the quantities involved in (7) are given in [72], which discuss higher-order frequentist confidence intervals for τ. In particular, the matching prior of τ is given by

$$\pi_{mp}(\tau) \propto \sqrt{\sum_{j=1}^{m} \frac{1}{\hat{\sigma}_\tau^2 + \hat{\sigma}_{j\tau}^2/n_j}},$$

with $\hat{\sigma}_\tau^2$ and $\hat{\sigma}_{j\tau}^2$ partial MLEs of σ^2 and σ_j^2, $j = 1, \ldots, m$, for fixed τ. Note that to compute (7), the HOTA simulation scheme can be used [64].

Let us consider the study involving nine laboratories carried out by the Nutrient Composition Laboratory of the US Department of Agriculture. The objective was to validate a proposed simple nonenzymatic gravimetric method for determining total dietary fiber in some foods. Six samples (apple, apricots, cabbage, carrots, onions, and soy fiber) were sent in blind duplicates to the participating laboratories. The data on fiber in apples were analyzed by [75], using non informative priors. For this example, $m = 9$ and the number of measurements n_j made by the jth laboratory is 2, for $j = 1, \ldots, 9$. The posterior distributions for τ are illustrated in Fig. 1, and the credible intervals for the consensus mean and some summary statistics are given in the following table:

	mean (sd)	median	0.95 equi-tailed	0.95 HPD
$\pi_{mp}(\tau\|y)$	12.91 (0.27)	12.93	(12.35,13.46)	(12.33,13.43)
$\pi_m^{vr}(\tau\|y)$	12.87 (0.66)	12.90	(12.19,13.61)	(12.19,13.61)
(4)	12.91 (0.22)	12.91	(12.47,13.34)	(12.47,13.34)

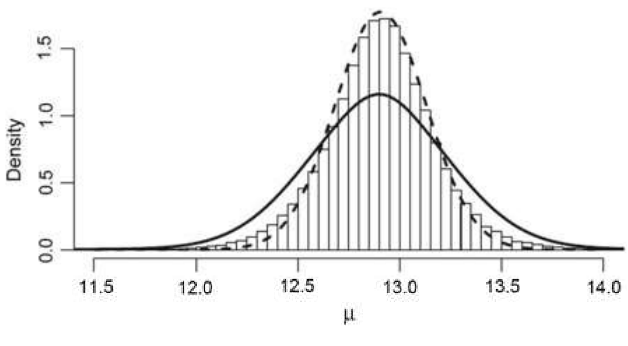

Fig. 1 HOTA posterior distribution (histogram), $\pi_m^{vr}(\tau|y)$ (*solid*) and first-order approximation (4) (*dashed*) for the mean dietary fiber in apples

The overall computation time was 1 s. The dashed curve in Fig. 1 is the first-order approximation (4), while the solid curve is the marginal posterior $\pi_m^{vr}(\tau|y)$ for τ discussed in [75]. This posterior is based on the independent priors $\pi(\tau) \propto 1$, $\pi(\sigma_j) \propto 1/\sigma_j$, $j = 1, \ldots, m$, and $\pi(\sigma) \propto 1$. Note that the first-order 95 % equi-tailed credible interval appears unsuitable since it is too short owing to a poor normal approximation to the posterior distribution (see also [15]).

4.2 Inference on the Cox Proportional Hazards Model

The Cox proportional hazards model [22, 23] is widely used for semiparametric survival data modeling. In its simplest form the failure times $T_1, \ldots, T_n$, for n independent individuals, have hazard functions $h(t; x_i) = h_0(t) \exp\{x_i^T \beta\}$, where $\beta = (\beta_1, \ldots, \beta_p)$ is a vector of unknown regression parameters, x_i is a $(p \times 1)$ vector of covariates for the ith individual, $i = 1 \ldots, n$, and $h_0(t)$ is the baseline hazard function. Suppose that the failure time is subject to right-censoring by a mechanism independent of their values and uninformative about their distribution. The data are n pairs (t_i, δ_i), where t_i denotes the observed lifetimes for the ith individual and δ_i is an indicator of the survival status, with $d_i = 1$ if t_i is a failure time (uncensored) and $d_i = 0$ if t_i represents a right-censored value, that is if $T_i > t_i$, $i = 1, \ldots, n$. The partial likelihood for β is given by

$$L_P(\beta) = \prod_{i=1}^{m} \frac{e^{x_i^T \beta}}{\sum_{j \in \mathscr{R}(t_{(i)})} e^{x_j^T \beta}} \,, \tag{10}$$

where $t_{(i)}$ is the ordered failure time, $\mathscr{R}(t_{(i)})$ is the risk set comprising those individuals at risk at time $t_{(i)}$, $i = 1, \ldots, n$, and $m = \sum_i \delta_i$.

In the Bayesian framework prior opinion should be modeled through a prior process on the baseline cumulative hazard function and a prior density $\pi(\beta)$ on the regression parameters, since both $h_0(t)$ and β are unknown. To avoid issues related to the elicitation on $h_0(t)$, in practice the partial likelihood (10) can be used directly to derive the pseudo-posterior distribution

$$\tilde{\pi}_P(\beta|y) \propto \pi(\beta)\, L_P(\beta); \tag{11}$$

see [39, 40, 69], and references therein, for various Bayesian applications of (11). Suppose it is of interest to focus on the scalar parameter β_j, i.e., the jth component of β. Let then $\beta = (\psi, \lambda)$, with $\psi = \beta_j$ the parameter of interest and $\lambda = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_p)$ the $(p-1)$-dimensional nuisance parameter. Non-informative priors on β, such as $\pi(\beta) \propto 1$ (see, e.g., [21]) or vague normal priors (see, e.g., [40]), can be considered.

Let us consider a real dataset concerning a clinical study on malignant mesothelioma (MM) [31]; this example is discussed in [67]. The dataset reports censored survival times for $n = 109$ and the type of malignant mesothelioma, i.e., type epithelioid, biphasic, or sarcomatoid. The partial likelihood (10) is thus a function of $\beta = (\beta_1, \beta_2)$.

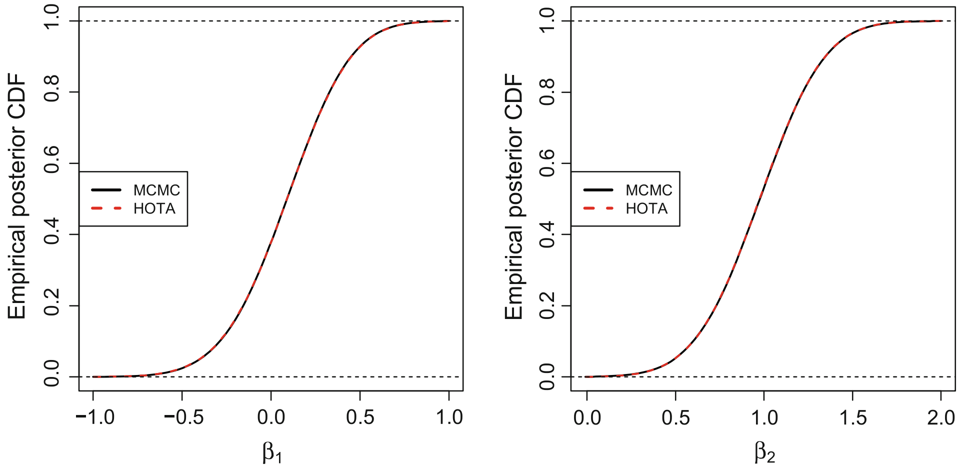

Fig. 2 Marginal posterior distributions for β_1 and β_2 computed with HOTA and MCMC for the Cox regression model

The marginal partial posterior distributions for β_1 and β_2 can be computed both using the HOTA algorithm based on higher-order approximations or with MCMC, both based on 10^4 simulations and a non-informative prior on β. A graphical comparison of the two cumulative distribution functions is given in Fig. 2, whereas numerical comparisons are reported in the following table:

Method		**Mean**	**Std. Dev**	$Q_{0.025}$	**Median**	$Q_{0.975}$	**0.95 HPD**
HOTA	β_1	0.084	0.291	−0.501	0.089	0.641	(−0.480, 0.656)
HOTA	β_2	0.974	0.291	0.396	0.976	1.540	(0.415, 1.557)
MCMC	β_1	0.084	0.291	−0.501	0.089	0.640	(−0.488, 0.644)
MCMC	β_2	0.975	0.292	0.397	0.976	1.541	(0.395, 1.541)

The results indicate that the MCMC and the HOTA algorithm give virtually indistinguishable results. MCMC is run for a large number of simulations and the usual convergence checks and post processing tasks are applied (e.g., thinning, burn-in, etc.), whereas HOTA is very simple to implement in this example since it is available at little additional computational cost over simple first-order approximations. Moreover, HOTA gives independent samples at a negligible computational cost and it can be used for quick prior sensitivity analyses [62], since it is possible to easily assess the effect of different priors on marginal posterior distributions, given the same Monte Carlo error. This is not generally true for MCMC or importance sampling methods, which in general have to be tuned for the specific model and prior.

4.3 Correlated Binary Data

The pairwise likelihood is particularly useful for modeling correlated binary outcomes, as discussed in [43]. This kind of data arise, e.g., in the context of repeated measurements on the same subject, where a maximum likelihood analysis involves multivariate integrals whose dimension equals the cluster sizes.

Let us focus on a multilevel probit model with constant cluster sizes. In particular, let S_i be a latent q-variate normal with mean $\gamma_i = X_i\beta/\sigma$, with β unknown regression coefficient, σ known scale parameter and X_i design matrix for unit i, and covariance matrix Σ, with $\Sigma_{hh} = \sigma^2$, $\Sigma_{hk} = \sigma^2\rho$, $h \neq k$, $i = 1, \ldots, n$. Then, the observed y_{ih} is equal to 1 if $S_{ih} > 0$, and 0 otherwise, for $h = 1, \ldots, q$.

The full likelihood is cumbersome since it entails calculation of multiple integrals of the multivariate normal distribution. On the other hand, the pairwise log likelihood is (see, e.g., [41,43])

$$p\ell(\beta, \rho) = \sum_{i=1}^{n}\sum_{h=1}^{q-1}\sum_{k=h+1}^{q} \log P(Y_{ih} = y_{ih}, Y_{ik} = y_{ik}; \beta, \rho) \, , \tag{12}$$

where $P(Y_{ih} = 1, Y_{ik} = 1; \beta, \rho) = \Phi_2(\gamma_{ih}, \gamma_{ik}; \rho)$ denotes the standard bivariate normal distribution function with correlation coefficient ρ, and $\gamma_{ih} = x_{ih}\beta/\sigma$ is the component h of γ^i $(i = 1, \ldots, n, h, k = 1, \ldots, q)$. Pairwise likelihood inference is much simpler than using the full likelihood since it involves only bivariate normal integrals.

In principle, the pairwise likelihood can be used directly in the Bayes' theorem as it is a genuine likelihood, giving [73]

$$\pi_{p\ell}(\beta, \rho|y) \propto \pi(\beta, \rho) \exp(p\ell(\beta, \rho)) \, .$$

However, [58] suggest to combine a calibrated version of the pairwise likelihood with the prior, obtaining the calibrated posterior

$$\pi^{c}_{p\ell}(\beta, \rho|y) \propto \pi(\beta, \rho) \exp(c\, p\ell(\beta, \rho)) \, , \tag{13}$$

with c suitable constant (see formula (2.3) in [58]). The calibration is necessary in order to alleviate the inefficiency of composite likelihood methods. Moreover, the use of $\pi^{c}_{p\ell}(\beta, \rho|y)$ recovers, approximately, the asymptotic properties of the pairwise posterior. Examples of $\pi_{p\ell}(\beta, \rho|y)$ and of $\pi^{c}_{p\ell}(\beta, \rho|y)$ are discussed also in [63,65].

Let us consider an example in [65], which discuss the use of the pairwise likelihood function in Approximate Bayesian Computation (ABC) methods. The data have been generated with $\beta_0 = \rho = 0.5$ and $\beta_1 = \sigma = 1$, and with $n = 50$ and $q = 7$, where β_0 is the intercept and β_1 the coefficient of a covariate, which has been generated from a $U(-1, 1)$. For the parameter $\theta = (\beta_0, \beta_1, \kappa)$, with $\kappa = logit((\rho(q-1)+1)/q)$, a normal prior $N(0, 45)^3$ is assumed.

The marginal pairwise posteriors for ρ, β_0 and β_1, derived from the calibrated and non-calibrated pairwise posteriors, are illustrated in Fig. 3. For the purposes of comparison we report also an MCMC approximation of the posterior based on the full likelihood. Clearly, the non-calibrated pairwise posterior is quite different from the target (MCMC), whereas the calibrated pairwise posterior behaves much better.

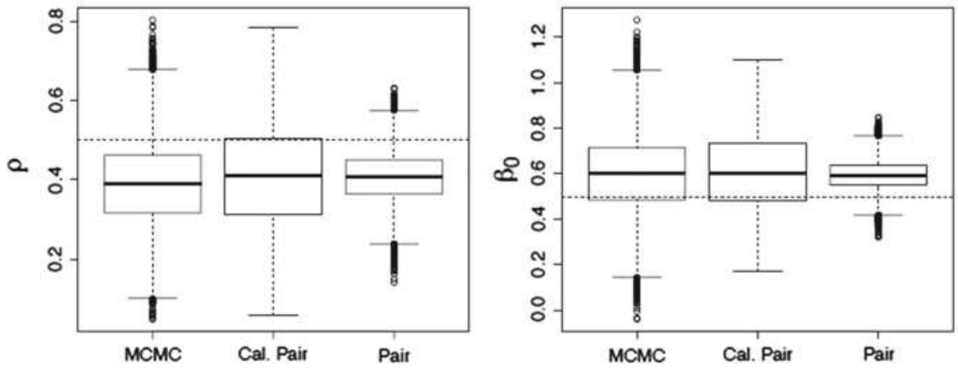

Fig. 3 Correlated binary data: Calibrated pairwise posterior (Cal. Pair) compared with the pairwise (Pair) and the exact (MCMC) posteriors. The *horizontal lines* represent the true parameter values

5 Final Remarks

Posterior distributions based on suitable pseudo-likelihoods have been proved useful for Bayesian inferences on a parameter of interest in several contexts (see also [9]). A first notable situation arises when elimination of a nuisance parameter is of interest. In this case the use of a pseudo-likelihood allows to avoid the elicitation of the prior of the nuisance parameter and the computation of a multidimensional integral in the integrated likelihood. A second striking situation is when the ordinary likelihood, and thus the corresponding posterior distribution, is difficult or even impractical to compute. In this respect, the use of a pseudo-posterior distribution based on the partial and the composite likelihoods may be particularly useful to deal with complex models.

Finally, we note that the interplay between Bayesian and likelihood procedures is still lively and opens to new research topics. A first instance refers to the use of composite likelihood score functions as summary statistics in Approximate Bayesian Computation (ABC) in order to obtain accurate approximations to the posterior distribution in complex models [65]. Moreover, also scoring rules, that generalize the proper and the composite likelihoods, can be used for developing posterior distributions using ABC methods (see the preliminary results in [66]). Finally, in [18] it is shown how higher-order approximations and matching priors are useful to derive an accurate approximation of the measure of evidence for the full Bayesian significance test introduced by [59] for precise hypotheses.

Acknowledgments This work was supported by a grant from the University of Padua (Progetti di Ricerca di Ateneo 2013) and by the grant *Progetto di Ricerca di Base, Legge Regionale Sardegna. 7/2007-2012*.

References

1. Adimari, G.: On the empirical likelihood ratio for smooth functions of M-functionals. Scand. J. Stat. **24**, 47–59 (1997)
2. Adimari, G., Ventura, L.: Quasi-profile log likelihoods for unbiased estimating functions. Ann. Inst. Stat. Math. **54**, 235–244 (2002a)
3. Adimari, G., Ventura, L.: Quasi-likelihood from M-estimators: a numerical comparison with empirical likelihood. Stat. Methods Appl. **11**, 175–185 (2002b)
4. Agostinelli, C., Greco, L.: A weighted strategy to handle likelihood uncertainty in Bayesian inference. Comput. Stat. **28**, 319–339 (2013)
5. Azzalini, A.: Maximum likelihood of order m for stationary stochastic processes. Biometrika **70**, 381–367 (1983)
6. Barndorff-Nielsen, O.E.: Quasi profile and directed likelihoods from estimating functions. Ann. Insti. Stat. Math. **47**, 461–464 (1995)
7. Barndorff-Nielsen, O.E., Chamberlin, S.R.: Stable and invariant adjusted directed likelihoods. Biometrika **81**, 485–499 (1994)
8. Bellio, R., Greco, L., Ventura, L.: Adjusted quasi-profile likelihoods from estimating functions. J. Stat. Plan. Inference **138**, 3059–3068 (2008)
9. Berger, J.O., Bayarri, S.: The interplay between Bayesian and frequentist inference. Stat. Sci. **19**, 58–80 (2004)
10. Berger, J.O., Liseo, B., Wolpert, R.: Integrated likelihood methods for eliminating nuisance parameters. Stat. Sci. **14**, 1–28 (1999)
11. Bertolino, F., Racugno, W.: Analysis of the linear correlation coefficient using pseudo-likelihoods. J. Italian Stat. Soc. **1**, 33–50 (1992)
12. Bertolino, F., Racugno, W.: Robust Bayesian analysis of variance and the x2-test by using marginal likelihoods. The Statistician **43**, 191–201 (1994)
13. Besag, J.E.: Spatial interaction and the statistical analysis of lattice systems (with discussion). J. R. Stat. Soc. B **34**, 192–236 (1974)
14. Bjornstad, J.F.: Predictive likelihood: a review (with discussion). Stat. Sci. **5**, 242–265 (1990)
15. Brazzale, A.R., Davison, A.C., Reid, N.: Applied asymptotics. Case Studies in Small- Sample Statistics. Cambridge University Press, Cambridge (2007)
16. Butler, R.W.: Approximate predictive pivots and densities. Biometrika **76**, 489–501 (1989)
17. Cabras, S., Castellanos, M.E., Racugno, W., Ventura, L.: A matching prior for the shape parameter of the skew-normal distribution. Scand. J. Stat. **39**, 236–247 (2012)
18. Cabras, S., Racugno, W., Ventura, L.: Higher-order asymptotic computation of Bayesian significance tests for precise null hypotheses in the presence of nuisance parameters. J. Stat. Comput. Simul. (to appear) (2014)
19. Chang, H., Mukerjee, R.: Probability matching property of adjusted likelihoods. Stat. Probab. Lett. **76**, 838–842 (2006)
20. Chang, H., Kim, B.H., Mukerjee, R.: Bayesian and frequentist confidence intervals via adjusted likelihoods under prior specification on the interest parameter. Statistics **43**, 203–211 (2009)
21. Chen, M., Ibrahim, J., Shao, Q.: Posterior propriety and computation for the Cox regression model with applications to missing covariates. Biometrika **93**, 791–807 (2006)
22. Cox, D.R.: Regression models and life tables (with discussion). J. R. Stat. Soc. B **34**, 187–200 (1972)
23. Cox, D.T.: Partial likelihood. Biometrika **62**, 269–276 (1975)
24. Cox, D.R., Reid, N.: Orthogonal parameters and approximate conditional inference (with discussion). J. R. Stat. Soc. B. **49**, 1–39 (1987)
25. Cox, D.D., O'Sullivan, F.: Asymptotic analysis of penalized likelihood and related estimators. Ann. Stat. **18**, 1676–1695 (1990)
26. Cox, D.R., Reid, N.: A note on pseudo likelihood constructed from marginal densities. Biometrika **91**, 729–737 (2004)

27. Datta, G.S., Mukerjee, R.: Probability Matching Priors: Higher Order Asymptotics. Springer, Berlin (2004)
28. Davison, A.C., Hinkley, D.V., Worton, B.J.: Bootstrap likelihood. Biometrika **79**, 113–130 (1992)
29. Davison, A.C., Hinkley, D.V., Worton, B.J.: Accurate and efficient construction of bootstrap likelihoods. Stat. Comput. **5**, 257–264 (1995)
30. Efron, B.: Bayes and likelihood calculations from confidence intervals. Biometrika **80**, 3–26 (1993)
31. Epithelial-mesenchymal transition in malignant mesothelioma: Fassina, A., Cappellesso, R., Guzzardo, V., Dalla Via, L., Piccolo, S., Ventura, L., Fassan, M. Mod. Pathol. **25**, 86–99 (2012)
32. Fraser, D.A.S., Reid, N.: Bayes posteriors for scalar interest parameters. Bayesian Stat. **5**, 581–585 (1996)
33. Fraser, D.A.S., Reid, N.: Strong matching of frequentist and Bayesian parametric inference. J. Stat. Plan. Inference **103**, 263–285 (2002)
34. Direct Bayes for interest parameters: Fraser, D.A.S., Reid, N., Wong, A., Yun Yi, G. Valencia **7**, 529–533 (2003)
35. Gong, G., Samaniego, F.J.: Pseudo maximum likelihood estimation: theory and applications. Ann. Stat. **89**, 861–869 (1981)
36. Greco, L., Racugno, W., Ventura, L.: Robust likelihood functions in Bayesian inference. J. Stat. Plan. Inference **138**, 1258–1270 (2008)
37. Green, P.J.: Penalized likelihood for general semi parametric regression models. Int. Stat. Rev. **55**, 245–260 (1987)
38. Hu, F., Zidek, J.V.: The weighted likelihood. Can. J. Stat. **30**, 347–371 (2002)
39. Ibrahim, J., Chen, M., Sinha, D.: Bayesian Survival Analysis. Springer, New York (2002)
40. Kim, Y., Kim, D.: Bayesian partial likelihood approach for tied observations. J. Stat. Plan. Inference **139**, 469–477 (2009)
41. Kuk, A., Nott, D.: A pairwise likelihood approach to analyzing correlated binary data. Stat. Probab. Lett. **47**, 329–335 (2000)
42. Lazar, N.A.: Bayes empirical likelihood. Biometrika **90**, 319–326 (2003)
43. Le Cessie, S., van Houwelingen, J.C.: Logistic regression for correlated binary data. Appl. Stat. **43**, 95–108 (1994)
44. Lee, Y., Nelder, J.A.: Hierarchical generalized linear models (with discussion). J. R. Stat. Soc. B. **58**, 619–678 (1996)
45. Lee, Y., Nelder, J.A., Noh, M.: H-likelihood: problems and solutions. Stat. Comput. **17**, 49–55 (2007)
46. Lin, L.: Quasi Bayesian likelihood. Stat. Methodol. **3**, 444–455 (2006)
47. Markatou, M., Basu, A., Lindsay, B.G.: Weighted likelihood estimating equations with a bootstrap root search. J. Am. Stat. Assoc. **93**, 740–750 (1998)
48. McCullagh, P.: Quasi-likelihood functions. Ann. Stat. **11**, 59–67 (1983)
49. McLeish, D.L., Small, C.G.: A projected likelihood function for semiparametric models. Biometrika **79**, 93–102 (1992)
50. Meng, X.L.: Decoding the H-likelihood. Stat. Sci. **24**, 280–293 (2009)
51. Min, X., Sun, D.: A matching prior based on the modified profile likelihood in a generalized Weibull stress-strength model. Can. J. Stat. **41**, 83–97 (2013)
52. Monahan, J. F., Boos, D. D.: Proper likelihoods for Bayesian analysis. Biometrika. Oxford University Press (OUP), **79**(2), 271–278 (1992). http://dx.doi.org/10.1093/biomet/79.2.271
53. Mykland, P.A.: Dual likelihood. Ann. Stat. **23**, 396–421 (1995)
54. Owen, A.B.: Empirical Likelihood. Chapman & Hall, London (2001)
55. Pace, L., Salvan, A.: Principles of Statistical Inference. World Scientific, Singapore (1997)
56. Pace, L., Salvan, A.: Adjustments of the profile likelihood from a new perspective. J. Stat. Plan. Inference **136**, 3554–3564 (2006)

57. Pace, L., Salvan, A., Ventura, L.: Likelihood based discrimination between separate scale and regression models. J. Stat. Plan. Inference **136**, 3539–3553 (2006)
58. Pauli, F., Racugno, W., Ventura, L.: Bayesian composite marginal likelihoods. Stat. Sin. **21**, 149–164 (2011)
59. Pereira, C., Stern, J.: Evidence and credibility: full Bayesian significance test for precise hypotheses. Entropy **1**, 99–110 (1999)
60. Racugno, W., Salvan, A., Ventura, L.: Bayesian analysis in regression models using pseudo-likelihoods. Commun. Stat. Theory Methods **39**, 3444–3455 (2010)
61. Reid, N., Mukerjee, R., Fraser, D.A.S.: Some aspects of matching priors. Lect. Notes-Monogr. Ser. **42**, 31–43 (2003)
62. Reid, N., Sun, Y.: Assessing sensitivity to priors using higher order approximations. Commun. Stat. Theory Methods **39**, 1373–1386 (2010)
63. Ribatet, M., Cooley, D., Davison, A.: Bayesian inference from composite likelihoods, with an application to spatial extremes. Stat. Sin. **22**, 813–845 (2012)
64. Ruli, E., Sartori, N., Ventura, L.: Marginal posterior simulation via higher-order tail area approximations. Bayesian Anal. **9**, 129–146 (2014)
65. Ruli, E., Sartori, N., Ventura, L.: Approximate Bayesian computation with composite score functions (submitted)
66. Ruli, E., Sartori, N., Ventura, L.: Approximate Bayesian computation with proper scoring rules. Atti della XLVII Riunione Scientifica della SIS. Cagliari, 11-13 giugno 2014, 1–6 (2014)
67. Ruli, E., Ventura, L.: Higher-order Bayesian approximations for pseudo-posterior distributions. Commun. Stat. Simul. Comput. (to appear)
68. Schennach, S.M.: Bayesian exponentially tilted empirical likelihood. Biometrika **92**, 31–46 (2005)
69. Sinha, D., Ibrahim, J., Chen, M.: A Bayesian justification of Cox's partial likelihood. Biometrika **90**, 629–641 (2003)
70. Severini, T.A.: On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. Stat. Sin. **9**, 713–724 (1999)
71. Severini, T.A.: Likelihood Methods in Statistics. Oxford University Press, New York (2000)
72. Sharma, G., Mathew, T.: Higher order inference for the consensus mean in inter-laboratory studies. Biom. J. **53**, 128–136 (2011)
73. Smith, E.L., Stephenson, A.G.: An extended Gaussian max-stable process model for spatial extremes. J. Stat. Plan. Inference **139**, 1266–1275 (2009)
74. Stein, M.L.: Approximating likelihoods for large spatial data sets. J. R. Stat. Soc. B **66**, 275–296 (2004)
75. Vangel, M.G., Rukhin, A.L.: Maximum likelihood analysis for heteroscedastic one-way random effects ANOVA in interlaboratory studies. Biometrics **55**, 129–136 (1999)
76. Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Stat. Sin. **21**, 5–42 (2011)
77. Ventura, L., Cabras, S., Racugno, W.: Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. J. Am. Stat. Assoc. **104**, 768–774 (2009)
78. Ventura, L., Cabras, S., Racugno, W.: Default prior distributions from quasi- and quasi-profile likelihoods. J. Stat. Plan. Inference **140**, 2937–2942 (2010)
79. Ventura, L., Racugno, W.: Recent advances on Bayesian inference for $P(X < Y)$. Bayesian Anal. **6**, 411–428 (2011)
80. Ventura, L., Reid, N.: Approximate Bayesian computation with modified log likelihood ratios. Metron **7**, 231–245 (2014)
81. Ventura, L., Sartori, N., Racugno, W.: Objective Bayesian higher-order asymptotics in models with nuisance parameters. Comput. Stat. Data Anal. **60**, 90–96 (2013)
82. Wang, J.: Quadratic artificial likelihood functions using estimating functions. Scand. J. Stat. **33**, 379–390 (2006)