Reference priors based on composite likelihoods
Reference prior basate sulle verosimiglianze composite

Federica Giummolè, Valentina Mameli and Laura Ventura

Abstract In this paper we propose reference priors obtained by maximizing the average $\alpha-$divergence from the posterior distribution, when the latter is computed using a composite likelihood. Composite posterior distributions have already been considered in [7] and [8], when a full likelihood for the data is too complex or even not available. The use of a curvature corrected composite posterior distribution, as in [8], allows to apply the method in [6] for maximizing the asymptotic Bayes risk associated to an $\alpha-$divergence. The result is a Jeffreys type prior that is proportional to the square root of the determinant of the Godambe information matrix.
Abstract *In questo lavoro proponiamo un metodo per ottenere una distribuzione a priori non informativa massimizzando l'$\alpha-$divergenza media dalla distribuzione a posteriori, quando quest'ultima viene calcolata a partire da una verosimiglianza composita. Delle posteriori composite sono già state proposte in [7] e [8], per trattare casi in cui la verosimiglianza completa sia difficile o addirittura impossibile da specificare. Il metodo proposto in [6] per la massimizzazione del rischio di Bayes associato ad un'$\alpha-$divergenza, si applica facilmente ad una a posteriori composita opportunamente corretta, introdotta in [8]. Il risultato è una distribuzione a priori del tipo di Jeffreys, proporzionale alla radice quadrata del determinante della matrice di informazione di Godambe.*

Key words: Composite likelihoods, Composite posterior distributions, Reference priors, $\alpha-$divergences, Godambe Information.

Federica Gummolè and Valentina Mameli
Ca'Foscari University - Venice, Departement of Environmental Sciences, Informatics and Statistics, via Torino 155, Mestre e-mail: giummole@unive.it, mameli.valentina@virgilio.it

Laura Ventura
University of Padova, Departement of Statistical Sciences, via Cesare Battisti 241, Padova e-mail: ventura@stat.unipd.it

1 Introduction

In the Bayesian setting, when the full likelihood is too complex and difficult to specify, [8] and [7] (see also references therein) propose the use of composite likelihoods to construct posterior distributions. In order to compute these composite posterior distributions, a prior for the unknown parameter of interest must be elicited. Among the most useful methods for finding objective priors, here we focus on the construction of reference priors, firstly suggested by [2]. Reference priors are based on the maximization of a distance between the posterior and the prior within an appropriate class of priors; for a review see [3] and [5]. Our purpose is to construct reference priors obtained by maximizing $\alpha-$divergences, which include as a special case the Kullback-Leibler divergence, for use in complex models.

The paper unfolds as follows. The second section reviews posterior distributions based on composite likelihoods. In particular, the composite posterior distribution obtained from a curvature corrected composite likelihood is introduced. The third section presents the construction of reference priors based on $\alpha-$divergences and suggests a method for finding reference priors when the curvature corrected composite likelihood is used. A simple example is presented in the fourth section. Finally, suggestions and comments on further developments can be found in the conclusions.

2 Posterior distributions based on composite likelihood

Let Y be a q-dimensional random vector with joint density $p(y|\theta)$, $\theta \in \Theta$, where Θ is an open subset of $\mathbb{R}^d$, and let $y = (y_1, \dots, y_n) \in \mathscr{Y}$ be a random sample of size n from the random vector Y. Given a set of measurable events $\{A_i, i \in I \subset \mathbb{N}\}$, the composite likelihood is defined as

$$L_c(\theta) = L_c(\theta; y) = \prod_{j=1}^{n} \prod_{i \in I} p(y_j \in A_i; \theta)^{w_i},$$

where w_i are positive weights, $i \in I$.

The maximum composite likelihood estimator $\hat{\theta}_c$, if it exists, maximizes the composite likelihood, namely, $\hat{\theta}_c = \arg\max_\theta L_c(\theta)$, or equivalently the composite log-likelihood $\hat{\theta}_c = \arg\max_\theta \ell_c(\theta)$, with $\ell_c(\theta) = \ell_c(\theta; y) = \log L_c(\theta)$. Under broad regularity conditions on the model, see for example [1], $\hat{\theta}_c$ is consistent and asymptotically normally distributed, with asymptotic covariance matrix given by the inverse of the Godambe information matrix $G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$, where $J(\theta) = E\left(\nabla \ell_c(Y;\theta) \nabla \ell_c(Y;\theta)^T\right)$, and $H(\theta) = E\left(\nabla^2 \ell_c(Y;\theta)\right)$; see [9]. Here ∇ and ∇^2 denote the gradient and Hessian operators, respectively.

A composite posterior distribution can be obtained by using a composite likelihood instead of the true likelihood in Bayes' formula. In particular, here we consider a composite posterior obtained from the curvature adjustment of the composite likelihood, which is defined as

$$\pi_c(\theta|y) \propto \pi(\theta)L_c(\theta^*), \tag{1}$$

with $\theta^* = \theta^*(\theta) = \hat{\theta}_c + C(\theta - \hat{\theta}_c)$, where C is a $d \times d$ fixed matrix such that $C^T H(\theta)C = G(\theta)$. A possible choice of the matrix C is given by $C = M^{-1}M_A$, with $M_A^T M_A = G$ and $M^T M = H$; see [8] and references therein.

Under regularity conditions, as $n \to \infty$, it can be shown that the composite posterior distribution (1) is, up to order $O_p(n^{-1/2})$, normally distributed with mean $\hat{\theta}_c$ and variance $K(\hat{\theta}_c)^{-1}$, i.e.

$$\pi_c(\theta|y) \dot{\sim} N_d(\hat{\theta}_c, K(\hat{\theta}_c)^{-1}), \tag{2}$$

with $K(\theta) = C^T \nabla^2 \ell_c(\theta^*)C$. Note that $K(\hat{\theta}_c)/n$ converges almost surely to $G(\theta)$ as $n \to \infty$; see [8].

3 Reference priors obtained by maximizing $\alpha-$divergences

The information present in a prior distribution may be measured in terms of divergence from the corresponding posterior: the bigger the divergence, the lower the influence of the prior on the posterior; for a review see [3] and [5]. Minimizing the information in a prior is equivalent to maximize the expected divergence D between the prior and the posterior, i.e. the functional

$$\begin{aligned} T(\pi) &= \int_{\mathscr{Y}} D_\pi(y)p(y)dy, \\ &= \int_\Theta \int_{\mathscr{Y}} [D_\pi(y)p(y|\theta)dy]\, p(\theta)d\theta, \end{aligned} \tag{3}$$

where $D_\pi(y) = D(\pi(\cdot), \pi(\cdot|y))$, with $\pi(\cdot)$ and $\pi(\cdot|y)$ denoting prior and posterior distributions for θ, respectively, and $p(\cdot)$ and $p(\cdot|\theta)$ denoting marginal and conditional distributions of Y given θ, respectively.

In particular, as a special instance of divergences between two distributions, we consider the well-known $\alpha-$divergences, defined as

$$D_\pi(y) = \frac{1}{\alpha(1-\alpha)} \int_\Theta \left\{ 1 - \left(\frac{\pi(\theta)}{\pi(\theta|y)} \right)^\alpha \right\} \pi(\theta|y)d\theta,$$

which for $\alpha \to 0$ reduces to the Kullback-Liebler divergence.

The general expected $\alpha-$divergence between the prior and the posterior in (3) can be written in the following form (see [6], formula (37))

$$T(\pi) = \frac{1 - \int_\Theta \pi^{\alpha+1}(\theta)E\,[\pi^{-\alpha}(\theta|Y)|\theta]\,d\theta}{\alpha(1-\alpha)}, \tag{4}$$

where $E(\cdot|\theta)$ denotes the conditional expectation given θ.

Here we apply the method proposed in [6] to maximize the expected $\alpha-$divergence between a prior and the composite posterior distribution defined in (1).

When $-1 < \alpha < 0$ and $0 < \alpha < 1$, using (2) and a shrinkage argument (see [5] and [6]), it can be shown that the selection of a prior $\pi(\theta)$ corresponds to the minimization with respect to $\pi(\theta)$ of the functional

$$\frac{1}{\alpha(1-\alpha)}\int \pi^{\alpha+1}(\theta)|G(\theta)|^{-\alpha/2}d\theta. \tag{5}$$

It can be proven that the prior $\pi(\theta)$ which asymptotically minimizes (5) is proportional to $|G(\theta)|^{1/2}$, i.e. the square root of the determinant of the Godambe matrix, which can be interpreted as a Jeffreys type prior.

4 Example

As an example we consider the equi-correlated multivariate normal model, for which the analytical expression of the matrix $G(\theta)$ is available; see [7].

Let Y be a $q-$dimensional random vector with mean 0 and covariance matrix Σ, with $\Sigma_{rr} = 1$ and $\Sigma_{rs} = \rho$ for $r \neq s$, with $r,s = 1,\dots,q$ and $\rho \in (-1/(q-1),1)$. A composite posterior distribution for ρ can be obtained by using a curvature adjustment of the pairwise likelihood given by [7].

The reference prior which maximizes the $\alpha-$divergence from the corresponding posterior distribution, is proportional to the square root of the determinant of the Godambe information matrix, i.e.

$$\pi(\rho) \propto |G(\rho)|^{1/2} = \left[\frac{q(q-1)(1+\rho^2)^2}{2(1-\rho)^2c(q,\rho)}\right]^{1/2},$$

with $c(q,\rho) = \left[(1-\rho)^2(3\rho^2+1)+q\rho(-3\rho^3+8\rho^2-3\rho+2)+q^2\rho^2(1-\rho)^2\right]$. A prior for the parameter ρ based on a simulated sample of size $n = 30$ and $q = 10$ is depicted in Fig. 1.

5 Conclusions

In this paper reference priors for a vector parameter based on maximizing $\alpha-$divergences are discussed in the framework of composite likelihoods. Some extensions of the proposed result can be considered. Firstly, the method can be improved to handle general pseudo-likelihoods; see for instance [10]. Secondly, for the case $\alpha \to -1$, which corresponds to the chi-square divergence, it is necessary to consider higher order terms in the asymptotic expansion of the posterior distribution. Moreover, the method can be further extended considering the class of monotone and regular

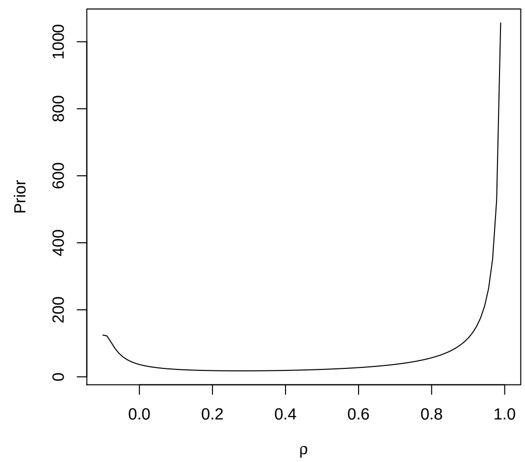

Fig. 1 Prior distribution for the parameter ρ of the equi-correlated multivariate normal model with $\mu = 0$ and $\sigma^2 = 1$ based on a simulated sample of size $n = 30$ and $q = 10$.

divergences which is a broad family of divergences asymptotically equivalent to $\alpha-$divergences; see [4]. Finally, an interesting line of research appears to be the investigation of reference priors for a parameter of interest in presence of nuisance parameters.

References

1. Barndorff-Nielsen, O.E. and Cox, D.R.: *Inference and Asymptotics*. Chapman & Hall, London (1994).
2. Bernardo, J.M.: Reference posterior distributions for Bayesian inference (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. **41**, 113-147 (1979).
3. Bernardo, J.M.: Reference analysis. In Handbook of Bayesian Statisics **25** (D.K. Dey and C.R. Rao, eds.), North-Holland, Amsterdam (2005).
4. Corcuera, J.M. and Giummolè, F.: A characterization of monotone and regular divergences. Ann. Inst. Statist. Math. **50**, 433–450 (1998).
5. Gosh, M.: Objective Priors: An Introduction for Frequentists. Stat. Sci. **26**, 187–202 (2011).
6. Liu, R., Chakrabarti, A., Samanta, T., Ghosh, J.K. and Ghosh, M.: On divergence measures leading to Jeffreys and other reference priors. Bayesian Anal. **9**, 331–370 (2014) .
7. Pauli, F., Racugno, W. and Ventura, L.: Bayesian composite marginal likelihoods. Statist. Sinica **21**, 149–164 (2011).
8. Ribatet, M., Cooley, D. and Davison, A.C.: Bayesian inference from composite likelihoods, with an application to spatial extremes. Statist. Sinica **22**, 813–845 (2012).
9. Varin, C., Reid, N. and Firth, D.: An overview of composite likelihood methods. Statist. Sinica **21**, 5–42 (2011).

10. Ventura, L., Cabras, S. and Racugno, W.: Default prior distributions from quasi- and quasi-profile likelihoods. J. Statist. Plann. Inference **140**, 2937–2942 (2010).