Statistica Sinica Preprint No: SS-14-151R2	
Title	A caveat on the robustness of composite likelihood estimators: the case of mis-specified random effect distribution
Manuscript ID	SS-14-151R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2014.151
Complete List of Authors	Helen Ogden
Corresponding Author	Helen Ogden
E-mail	h.ogden@warwick.ac.uk

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

A CAVEAT ON THE ROBUSTNESS OF COMPOSITE LIKELIHOOD ESTIMATORS: THE CASE OF A MIS-SPECIFIED RANDOM EFFECT DISTRIBUTION

Helen E. Ogden

University of Warwick

Abstract: Composite likelihoods are a class of alternatives to the full likelihood which may be used for inference in many situations where the likelihood itself is intractable. A composite likelihood estimator will be robust to certain types of model misspecification, since it may be computed without the need to specify the full distribution of the response. This potential for increased robustness has been widely discussed in recent years, and is considered a secondary motivation for the use of composite likelihood. The purpose of this paper is to show that there are some situations in which a composite likelihood estimator may actually suffer a loss of robustness compared to the maximum likelihood estimator. We demonstrate this in the case of a generalized linear mixed model under misspecification of the random-effect distribution. As the amount of information available on each random effect increases, we show that the maximum likelihood estimator remains consistent under such misspecification, but various marginal composite likelihood estimators are inconsistent. We conclude that composite likelihood estimators cannot in general be claimed to be more robust than the maximum likelihood estimator.

Key words and phrases: Consistency; Generalized linear mixed model; Laplace approximation; Pairwise interactions

1. Introduction

Suppose we observe independent samples $y^{(1)}, \ldots, y^{(r)}$, where each $y^{(i)} = (y_1^{(i)}, \ldots, y_m^{(i)})$ is assumed to be an independent sample from a model depending on an unknown parameter θ. The likelihood

$$L(\theta) = \prod_{i=1}^{r} L(\theta; y^{(i)})$$

is sometimes difficult to compute, and composite likelihoods (Lindsay (1988)) provide a class of alternatives for conducting inference about θ in such circum-

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

stances. A marginal composite likelihood

$$L^C(\theta) = \prod_{i=1}^{r} \prod_{s \in S} L(\theta; y_s^{(i)})^{w_s}$$

is formed by taking a product of component likelihoods, each of which is the likelihood given some subset of the data y_s, where w_s is a weight assigned to component s. A review of composite likelihoods and their many uses is given by Varin, Reid, and Firth (2011).

If the model is correctly specified, the composite likelihood estimator is consistent as $r \to \infty$, provided that the parameter of interest remains identifiable, although the estimator typically has a higher asymptotic variance than the maximum likelihood estimator. There is a hope that some compensation for this loss of efficiency can be provided by an increased robustness of the composite likelihood estimator to misspecification of the model. Lindsay, Yi and Sun (2011) motivate the use of composite likelihoods in this way, stating that 'Compared to the traditional likelihood method, the composite likelihood method may be less efficient, but it could be a lot computationally faster to implement and be more robust to model misspecification.'

This notion is motivated by the fact that it is not necessary to specify the full distribution of the response in order to be able to compute a composite likelihood. If the marginal distributions of Y_s for each subset $s \in S$ are correctly specified, then the corresponding estimator of θ is consistent as $r \to \infty$, even if the full model is misspecified. The maximum likelihood estimator need not be consistent in such a setting, since the likelihood relies on the full, misspecified, distribution of Y. Xu and Reid (2011) discuss this type of robustness in some detail, and provide a formal proof of the consistency of the composite likelihood estimator in this setting.

In some situations the relevant marginal distributions themselves may be misspecified, in which case the marginal composite likelihood estimator no longer retains this robustness property. However, in that case the full distribution of Y must also be incorrect, so the maximum likelihood estimator need not be robust to this misspecification either. From this, it is tempting to conclude that a marginal composite likelihood estimator must always be at least as robust to model misspecification as the full likelihood estimator. We show that this is not

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

the case.

To see this, we consider the impact of incorrectly specifying the random-effects distribution in a generalized linear mixed model. Generalized linear mixed models are a widely-used class of models, but one in which the likelihood is often difficult to compute. Because of this intractability, composite likelihood methods have been proposed as alternatives to full-likelihood inference.

In Section 3, we consider a class of generalized linear mixed models with simple nested structure, in which the likelihood is tractable. In this setting, we obtain some asymptotic results to show that under misspecification of the random-effect distribution the maximum likelihood estimator is consistent, while various composite likelihood estimators are inconsistent.

In Section 4, we consider a class of generalized linear mixed models with intractable likelihoods, in which each observation involves a pair of random effects. We compare the pairwise likelihood estimator proposed by Bellio and Varin (2005) with an estimator found by maximizing the Laplace approximation to the likelihood, in terms of the limiting value of each estimator under misspecification of the random-effect distribution. When only a small amount of information is available on each random effect, we find that the asymptotic bias in the pairwise likelihood estimator is smaller than that of Laplace estimator. However, as the amount of information available on each random effect increases, the magnitude of the asymptotic bias in the Laplace estimator decreases towards zero, but the asymptotic bias in the pairwise likelihood estimator remains fixed away from zero, in agreement with the asymptotic results of Section 3.

2. Generalized linear mixed models

In a generalized linear mixed model, the distribution of $Y^{(i)}$ takes exponential family form, with distribution determined by the linear predictor

$$\eta^{(i)} = X^{(i)}\beta + Z(\sigma)u^{(i)},$$

where X is the design matrix for the fixed effects, $u^{(i)} = (u_1^{(i)}, \dots u_n^{(i)})$ are random effects, where each $u_j^{(i)}$ is assumed to have independent standard normal distribution, and $Z(\sigma)$ is a design matrix for the random effects, whose entries may depend on a parameter σ.

The addition of random effects often makes the model more realistic, but the

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

likelihood of $\theta = (\beta, \sigma)$ given each $y^{(i)}$ is

$$L(\theta; y^{(i)}) = \int_{\mathbb{R}^n} f(y^{(i)} | \eta^{(i)} = X^{(i)}\beta + Z(\sigma)u^{(i)}) \prod_{j=1}^{n} \phi(u_j^{(i)}) du^{(i)},$$

where $\phi(.)$ is the standard normal density, and it might be difficult to approximate this n-dimensional integral well. Moreover, there is often no good reason to suppose that the random effects are normally distributed, so inferences made should be checked for sensitivity to this assumption.

There is a large literature on the impact of random-effects misspecification on the maximum likelihood estimator. A detailed review is provided by McCulloch and Neuhaus (2011), who conclude that the asymptotic bias due to such misspecification may be expected to be small in many cases, and highlight some situations where it may be of more concern.

In situations where the likelihood is intractable, it is common to replace the true likelihood with a Laplace approximation, and to use the approximated likelihood for inference (Pinheiro and Bates (1995)). A disadvantage of this approach is that the resulting estimator need not be consistent as the amount of data increases, even if the model is correctly specified.

Composite likelihood estimators are consistent under correct model specification, albeit with some loss of efficiency compared to the maximum likelihood estimator. Bellio and Varin (2005) construct a pairwise likelihood for inference in these models, with a contribution from each pair $(y_j^{(i)}, y_k^{(i)})$ which are dependent under the model. Thus, defining t_j as the set of non-zero elements in the jth row of $Z(\sigma)$, a pairwise likelihood may be defined by

$$L^{\text{pair}}(\theta; y) = \prod_{i=1}^{r} \prod_{(j,k): t_j \cap t_k \neq \emptyset} L(\theta; y_j^{(i)}, y_k^{(i)}).$$

Bellio and Varin (2005) focus on those cases in which each t_j contains two elements, although they note that essentially the same method could be applied more generally. Each term $L(\theta; y_j^{(i)}, y_k^{(i)})$ can be written as an integral over the $|t_j \cup t_k|$ random effects involved in the pair of observations. In the case where there are two random effects involved in each observation, this is a three-dimensional integral that is relatively easy to approximate well.

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

3. Robustness in a two-level model

3.1. A two-level model

Taking $n = 1$ in the generalized linear mixed model, and writing $x_j^{(i)}$ for the jth column of $X^{(i)}$, we obtain the two-level model $\eta_j^{(i)} = \beta^T x_j^{(i)} + \sigma u^{(i)}$. Here the $u^{(i)}$ are independent (scalar) random effects, which are assumed to be drawn from a $N(0, 1)$ distribution. Since $n = 1$, the likelihood is just a product of one-dimensional integrals. For ease of notation, we further assume that $x_j^{(i)} = x^{(i)}$ is constant across j, so that we can write $\eta_j^{(i)} = \eta^{(i)} = \beta^T x^{(i)} + \sigma u^{(i)}$; our results also hold for the more general case.

We are interested in the robustness of estimators of β to deviations from the assumed random-effects distribution. We suppose that each $u^{(i)}$ is drawn independently from a non-normal distribution with mean zero and variance one and that, for some parameter values $(\beta, \sigma) = (\beta_0, \sigma_0)$, the rest of the model is correctly specified.

Under this misspecification, we are interested in the limit β_m^∞ of the maximum likelihood estimator of β as $r \to \infty$, and how this limit varies with m. In particular, we show that as $m \to \infty$, $\beta_m^\infty \to \beta_0$, so that if r and m simultaneously tend to infinity, the maximum likelihood estimator is consistent.

3.2. Consistency of the maximum likelihood estimator

The likelihood for $\theta = (\beta, \sigma)$ is given by $L(\theta; y) = \prod_{i=1}^r L(\theta; y^{(i)})$, where

$$L(\theta; y^{(i)}) = \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^m f(y_j^{(i)} \mid \eta^{(i)} = \beta^T x^{(i)} + \sigma u^{(i)}) \right\} \phi(u^{(i)}) du^{(i)}.$$

We write $\ell_i(\theta; y^{(i)}) = \log L(\theta; y^{(i)})$ for the contribution to the log-likelihood from $y^{(i)}$, and $s_i(\theta; y) = \nabla_\theta \ell_i(\theta; y)$ for the corresponding score function. In the case that m is fixed and $r \to \infty$, the results of White (1982) show that the maximum likelihood estimator $\hat{\theta}_m^r = (\hat{\beta}_m^r, \hat{\sigma}_m^r)$ of θ converges to the value θ_m^∞ which solves $\bar{s}(\theta) = E\left\{s_i(\theta, Y^{(i)})\right\} = 0$, where the expectation is taken over the true distribution of $Y^{(i)}$ and the covariates $x^{(i)}$.

Intuitively, for large m, it should be possible to obtain an estimate of the value of each linear predictor $\eta^{(i)}$ from the data $y^{(i)}$ that is close to the true value $\eta_0^{(i)}$. This means that for sufficiently large m, inference given the data y should

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

in probability as $m \to \infty$, so

$$\begin{aligned}\ell_i(\theta_1; y^{(i)}) - \ell_i(\theta_2; y^{(i)}) &= \log g_i(\eta_0^{(i)} \mid y^{(i)}, \theta_1) - \log g_i(\eta_0^{(i)} \mid y^{(i)}, \theta_2) + o(1) \\ &= \log\left\{\frac{1}{\sigma_1}\phi\left(\frac{\eta_0^{(i)} - \beta_1^T x^{(i)}}{\sigma_1}\right)\right\} - \log\left\{\frac{1}{\sigma_2}\phi\left(\frac{\eta_0^{(i)} - \beta_2^T x^{(i)}}{\sigma_2}\right)\right\} + o(1) \\ &= \ell_i(\theta_1; \eta_0^{(i)}) - \ell_i(\theta_2; \eta_0^{(i)}) + o(1).\end{aligned}$$

Letting $\theta_1 = \theta$, $\theta_2 = \theta + h$ and considering the limit as $h \to 0$, we therefore have $s_i(\theta; y^{(i)}) = s_i(\theta; \eta^{(i)}) + o(1)$, as claimed. □

References

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12**, 171-178.

Bellio, R. and Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling* **5**, 217-227.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics* **80**, 221-239

Lindsay, B. G., Yi, G. Y., and Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statist. Sinica* **21**, 71-105.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science* **26**, 388-402.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comput. Graph. Stat.* **4**, 12-35.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5-42.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1-25.

Statistica Sinica: Preprint
doi:10.5705/ss.2014.151

Xu, X. and Reid, N. (2011). On the robustness of maximum composite likelihood estimate. *J. Statist. Plan. Infer.* **141**, 3047-3054.

Yi, G. Y. and Reid, N. (2010). A note on mis-specified estimating functions. *Statist. Sinica* **20**, 1749-1769.

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK
E-mail: h.ogden@warwick.ac.uk