Statistica Sinica **21** (2011), 1-3

Introduction to Special Issue

The University of Warwick hosted a workshop on the topic of Composite Likelihood in April, 2008. The workshop was prompted by the rapidly increasing literature on the topic in recent years, the impression that the applications of composite likelihood were outpacing the available theory, and the expectation that an overview and assessment of the theory, methods, and implementation of composite likelihood methods was warranted. One of the positive outcomes of this workshop is this special issue of *Statistica Sinica* devoted to this topic.

A call for contributions was issued early in 2009, and refereed in the usual way between September 2009 and June 2010. The contributions published here cover an impressive array of contributions to the theory, methods and application of composite likelihood. Several, but by no means all, of these papers are based on talks presented at the workshop.

The issue begins with four papers that illustrate the wide scope of the area. The review paper by Varin, Reid and Firth, presents an overview of theory and application, extending the review by Varin (2008). That by Larribe and Fearnhead reviews the use of composite likelihood specifically in statistical genetics, where construction of the full joint likelihood is computationally prohibitive. Lindsay, Yi and Sun provide a quite general discussion of the construction of composite likelihood, with emphasis on the resulting score equation: their work gives considerable insight into the generality of composite likelihood, and the need to be more specific about the choice of component likelihoods in order to obtain concrete results. They also make explicit the trade-off between computational complexity and statistical efficacy. The paper by Chen on local partial likelihood illustrates the range of models that can be subsumed under the general umbrella of composite likelihood.

Several papers contribute to building a wider theory for composite likelihood inference. Pace, Salvan and Sartori suggest a simpler approximation to the distribution of the composite likelihood ratio statistic than the weighted sum of χ^2 distribution that results from the usual asymptotic theory, which makes available an alternative to Wald-type statistics for inference. In a related development, Pauli, Racugno and Ventura show that if interest lies in combining a composite likelihood with a prior, some device is needed to convert the usual distribution to a simpler χ^2_ν. Gao and Song describe the use of the EM algorithm for maximizing the composite likelihood. In their application, to a problem in genetics, the E-step is over a set of unobserved variables in a hidden Markov model. To apply their algorithm in more general missing data problems requires the rather strong assumption of missing completely at random. Molenberghs, Kenward, Verbeke and Birhanu treat the weaker missing at random case for longitudinal data, and show how to incorporate ideas of double robustness developed in the estimating equations literature to the setting of composite likelihood. He and Yi describe the use of pairwise likelihood for correlated binary data with missing observations: in their setting there is no need to make any assumptions about the distribution governing the observation process. In work motivated by multi-stage family studies in genetic epidemiology, Chen and Briollais show that an inverse probability-weighted pseudo-likelihood can be treated by the theory of composite likelihood.

Composite likelihood has been widely used for some time in applications involving longitudinal data, where typically there are a large-ish number of relatively short series, each with internal correlation over time. Longitudinal data applications are emphasized in Molenberghs et al, and in He and Yi; the first treating both marginal and conditional composite likelihoods; the latter emphasizing pairwise likelihood for binary data. Chen and Briollais use a weighted composite likelihood for time to event data; the weights are related to the sampling probability, and the components of the composite likelihood are for completely observed units; missing data is incorporated via the EM algorithm.

Three papers concentrate on time series applications. Yau and Davis give proofs of the consistency of composite likelihood estimators in the case of single long time series, with application to AR(1), MA(1), and ARFIMA(1,1,d) models. Ng, Joe, Karlis and Liu consider time series with latent autoregressive processes. Like Yau and Davis, they consider a single long time series, and show that the composite likelihood from consecutive pairs of observations is preferable to the "all possible pairs" composite likelihood. Pakel, Shephard and Sheppard also study time series: in particular a collection of time series, each following a GARCH model, but assuming some common parameters for the volatility across the series. They show that pooling information across this collection, treating the separate series as independent, but using the full joint distribution within each series, is very effective, especially for relatively short series of less than, say, 500 observations.

Composite likelihood was first suggested in the context of likelihoods for spatial processes, in Besag (1975) under the name pseudo-likelihood. In spatial models it is often more natural to use component likelihoods based on conditional distributions, and recent work by Caragea and Smith (2006, 2007) and Stein, Chi and Welty (2004) extends Besag's original approach in several directions. Composite conditional likelihood for for a type of interacting particle system called a Potts model is investigated here in Okabayashi, Johnson and Geyer.

This collection of papers reflects the breadth of theory and application areas where ideas of composite likelihood are useful and important, and we hope will encourage interest and further research in both theoretical and practical aspects.

References

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician* **24**, 179-195.

Caragea, P. and Smith, R. L. (2006). Approximate likelihoods for spatial processes. Preprint.

Caragea, P. and Smith, R. (2007). Asymptotic properties of computationally ef- ficient alternative estimators for a class of multivariate normal models. *J. Multi. Anal.* **98**, 1417-1440.

Stein, M., Chi, Z. and Welty, L. (2004). Approximating likelihoods for large spatial data sets. *J. Roy. Statist. Soc. Ser. B* **66**, 275-296.

Varin, C. (2008). On composite marginal likelihoods. *Adv. Statist. Anal.* **92**, 1128.

Guest Editors
Nancy Reid[1] and Bruce Lindsay[2]
Kung-Yee Liang[3,4]

1. University of Toronto
2. Pennsylvania State University
3. Johns Hopkins University
4. Current affiliation: National Yang-Ming University, Taiwan.