Annotated list of references for the talk "Likelihood inference in complex models"

Nancy Reid

January 19, 2006

This list is unlikely to be complete or completely accurate; comments welcome.

Andersen, E.W. (2004). Composite likelihood and two-stage estimation in family studies. *Biostatistics* **5**, 15–30.
Models for joint survival times of family members are constructed using copulas, and estimation of correlation parameters via pairwise likelihood is suggested. See also Parner (2001).

Anderson, J.A. and Pemberton, J.D. (1985). The grouped continuous model for multivariate ordered categorical variables and covariate adjustment. *Biometrics* **41**, 875–85.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. R. Statist. Soc.* B, **34**, 192–236.
Introduction of "pseudo-likelihood" for spatial (auto-normal and auto-logistic) processes, formed by compounding conditional distribution of y_i given its neighbours, over all observations y_i.

Bellio, R. and Varin, C. (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling* **5**, 217–227.
`http://homes.stat.unipd.it/sammy/?page=PUBBLICAZIONI&lang=EN`
Generalized linear model with crossed random effects, of the form

$$f\{EY_{ij} \mid u_i v_j)\} = x_{ij}^T\beta + u_i + v_j$$

where $u_i \sim N(0, \sigma_u^2)$ and $v_j \sim N(0, \sigma_v^2)$. Illustration on McCullagh and Nelder's salamander data. The full joint likelihood requires integration over $R^{q_1} \times R^{q_2}$, whereas the pairwise likelihood requires only three-dimensional integrals. A simulation based approach is needed to estimate $E(\ell_2'\ell_2'^T)$, which in turn is needed for the variance of the pairwise likelihood estimator. Emphasis on estimation of variance components.

Cox, D.R. and Reid, N. (2004). A note on pseudo-likelihood constructed from marginal densities. *Biometrika* **91**, 792–737.
Use of a linear combination of pairwise likelihood and likelihood from the marginal densities in the form

$$\ell_2(\theta) = \sum_{i=1}^{n}[\sum_{s<r} \log f_{sr}(y_{is}, y_{ir}) - aq \sum_{s} \log f_s(y_i s)]$$

with a a constant to be chosen. Summary of asymptotic behaviour of the point estimator that solves $\ell_2(\tilde{\theta}) = 0$, as $n \to \infty$ and also as $q \to \infty$. Illustration with symmetric multivariate normal and dichotomized multivariate normal.

Fearnhead, P. (2003). Consistency of estimators of the population-scaled recombination rate. *Th. Pop. Biol.* **64**, 65–79.
Discussion of the pseudo-likelihood estimators in the context of population genetics. In particular the pairwise estimator of McVean et al. (2002) is not consistent. The composite likelihood estimator of Fearnhead and Donnelly (2002) is consistent. See also Hudson (2001).

Fearnhead, P. and Donnelly, P. (2002). Approximate likelihood methods for estimating local recombination rates (with Discussion). *J. R. Statist. Soc.* B **64**, 657–680.

Geys, H., Molenberghs, G. and Ryan, L. (1997). Pseudo-likelihood inference for clustered binary data. *Comm. Statist. Th. Meth.* **26**, 2743–2767.

Geys, H., Molenberghs, G., Lipsitz, S.R. (1998). A note on the comparison of pseudo-likelihood and generalized estimating equations for marginal odds ratio models. *J. Statist. Comp. Sim.* **62**, 45–72.

Heagerty, P.J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *J. Am. Statist. Assoc.* **93**, 1099–111.
Binary spatial data modelled by thresholding a gaussian random field. Pairwise score function constructed as

$$U_{HL}(\theta) = \frac{1}{W_n} \sum_{r,s} w_{rs} U_{rs}(\theta)$$

where $U_{rs}(\theta)$ is the score function from the joint density of (y_r, y_s), and the weights are to be chosen. A penalized score function is also considered. Resampling suggested

for the estimation of $E(U_{HL}U_{HJ}^T)$ needed. Simulations verify the method is efficient. Illustrated on gypsy moth deforestation data.

Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika* **90**, 355–66.
Observations y_{ir} are modelled as Poisson with mean $Z_{ir}\exp(\alpha_r + \beta x_i)$, with $Z_i = (Z_{i1}, \dots Z_{ip})$ is modelled as a multivariate Gamma, thus introducing correlation via these frailty components or random effects. Pairwise likelihood method confirmed by simulation and illustrated on an analgesia experiment.

Hjort, N.L. and Varin, C. (2005). ML, PL, QL in Markov chain models. preprint
Detailed consideration of the properties of estimators from full maximum likelihood, pseudo-likelihood (composed of conditional probabilities) and quasi-likelihood (composed of marginal (pairwise) probabilities) in the case of n observations on a stationary Markov chain. Illustration on a number of special models. Investigation of the methods under model failure.

Hudson, R.R. (2001). Two-locus sampling distributions and their application. *Genetics* **159**, 1805–1817.

Kuk, A.Y.C. & Nott, D.J. (2000). A pairwise likelihood approach to analyzing correlated binary data. *Stat. Prob. Lett.* **47**, 329–35.
Correlated binary data for clustered data (y_{is}) is constructed by assuming a logistic regression for the marginal distributions of y_{is}, and a logistic regression model for the association between any two observations y_{is}, y_{ir} in the same cluster. The pairwise likelihood is formed by adding over all clusters and all possible pairings within a cluster. There is a discussion of the merits of weighting the sum over clusters inversely to the cluster size; this is recommended for accurate estimation of the marginal parameters, but not the parameters in the correlation. Calculations of relative efficiency relative to full maximum likelihood estimation is presented.

LeCessie, S. & van Houwelingen, J.C. (1994). Logistic regression for correlated binary data. *Appl. Statist.* **43**, 95–108.
Binary data has marginally a logistic regression, and a correlation parameter is introduced either by dichotomizing a latent normal variable, or by an odds ratio formulation as in Kuk and Nott (2000). Initially only clusters of size two are considered; in a generalization to larger clusters the pairwise likelihood is used. Weighting inversely proportial to cluster size is proposed. Application to a large Danish study on preterm infants, where observations on multiple births generate dependence.

de Leon, A.R. (2005). Pairwise likelihood approach to grouped continuous model and its extension. *Stat. Prob. Lett.*, **75**, 49–57.

Models discrete data obtained by thresholding continuous multivariate data at a series of fixed cut points, and thus includes the dichotomized multivariate normal as a special case, extending a model proposed in Anderson and Pemberton (1985). Evaluation of pairwise maximum likelihood estimation by simulation. The suggestion is made that with a combination of grouped and continuous data the pairwise likelihood for the grouped data could be combined with the full likelihood for the continuous data.

Liang, G. and Yu, B. (2003). Maximum pseudo-likelihood estimation in network tomography. *IEEE Trans. Sig. Proc.* **51**, 2043–2053.
Application of pairwise likelihood estimation in the problem of estimating features of a complex network. Bin has also pointed out that pairwise likelihood and generalizations is used in the machine learning literature.

Lindsay, B.L. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes*, Ed. N.U. Prabhu, pp. 221–239. Providence: American Mathematical Society.
Study of the asymptotic properties of inference based on any type of likelihood formed by 'composing' terms that are individually likelihood functions. This includes Besag's pseudo-likelihood and pairwise likelihood as special cases. Extended discussion of the efficiency of composite likelihood, and improvement of efficiency through weighting schemes.

McVean, G., Awadalla, P. & Fearnhead, P. (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–41.

Nott, D.J. and Rydén, T. (1999). Pairwise likelihood methods for inference in image models. *Biometrika* **86**, 661–76.
Spatial binary data modelled as a thresholded a Gaussian random field; attributes the idea to an unpublished paper in N.L. Hjort and Heaguerty and Lele (1998). Seems to be the first use of the term 'pairwise' for the compounding of all bivariate marginals. Discusses weighting contributions of different pairs according to spatial distance. Simulation estimate of the variance of the score function.

Parner, E.T. (2001). A composite likelihood approach to multivariate survival data. *Scand. J. Statist.* **28**, 295–302.
Bivariate survival data is generated by assuming that the hazards for failure times T_1 and T_2 are related through a common frailty, or random effect, assumed to follow a gamma distribution. The likelihood for a pair of observations is modelled to included potential censoring. The parameters are estimated using pairwise likelihood. The method is illustrate on an adoption study.

Renard, D. Molenberghs, G. and Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Comp. Stat. Data. Anal.* **44**, 649–667.
Binary data with probit, instead of logit, link:

$$p(y_{ir} = 1 \mid x_{ir}, b_i) = \Phi(x_{ir}^T \beta + z_{ir}^T b_i)$$

where the random effects $b_i \sim N(0, \sigma_b^2)$. Pairwise likelihood is used to estimate mean parameters and covariance parameters. The asymptotic theory is similar to that in Kuk and Nott (2000). Pairwise likelihood is compared to maximum likelihood and to penalized quasi-likelihood; it is found to be both robust and efficient. See also Geys et al. (1997, 1998).

Rannalla, B. and Slatkin, M. (2000). Methods for multipoint disease mapping using linkage disequilibrium. *Genetic Epid.* **19**(Suppl.1), S71–S77.
Discussion of methods of constructing composite likelihood in place of the exact likelihood when the observations are vectors of multilocus haplotypes.

Varin, C., Host, G. and Skare, O. (2005). Pairwise likelihood inference in spatial generalized linear mixed models. *Comp. Statist. Data. Anal.* **49**, 1173-1191.
An EM algorithm is presented for maximizing the pairwise likelihood function for spatial data. A resampling method is used to estimate the variance of the score function. Extensive simulation studies again show that the pairwise likelihood is reasonably efficient.

Varin, C. and Vidoni P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92**, 519–528.
Proposes the use of the composite likelihood function to derive model selection criteria. Illustrated on time series of counts. Both pairwise likelihood and a likelihood involving trivariate distributions are considered.

Varin, C. and Vidoni, P. (2005). Pairwise likelihood inference for general state space models. preprint.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canad. J. Statist.* **33**, 335–356.