Empirical Composite Likelihoods

Nicola Lunardon, Francesco Pauli, Laura Ventura

Dept. of Statistics, University of Padova, Italy

email: ventura@stat.unipd.it

Outline

- Composite likelihoods may be useful for approximating likelihood based inference when the full likelihood is too complex to deal with.
- Stemming from a misspecified model, the asymptotic distribution of the composite likelihood ratio statistic departs from the familiar standard chi-square asymptotic distribution.
- Several adjustments have been proposed in the literature, which all require the elements of the Godambe information.
- This paper proposes and discusses a computationally and theoretically attractive approach based on the derivation of an empirical likelihood function from the composite score.
- For the special case of the pairwise likelihood, our proposal can allow reference to the usual asymptotic chi-square distribution.

Composite likelihoods

- Consider independent observations y_i of a random vector $Y_i = (Y_{i1}, \ldots, Y_{iq})$, $i = 1, \ldots, n$, with $Y_i \sim f(y_i; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$, $y_i \in \mathcal{Y}$.
- In some situations it may be difficult to evaluate $f(y; \theta)$ and thus the full likelihood $L(\theta)$.
- However, suppose it may be possible to compute likelihood contributions $L_k(\theta; y_i) = L(\theta; A_k(y_i))$, for the events $A_k(y_i)$, $k = 1, \ldots, K$, on $\mathcal{Y}$.
- The composite likelihood is then defined as (Lindsay 1988, Varin et al 2010)

$$cL(\theta; y) = \prod_{i=1}^{n} \prod_{k=1}^{K} L_k(\theta; y_i)^{w_k}$$

  with w_k positive weights.
- Let $c\ell(\theta) = \log cL(\theta; y)$ be the composite loglikelihood and let $cU(\theta)$ be the composite score function $(\partial/\partial\theta)c\ell(\theta)$.

An example: The pairwise likelihood

- When the events $A_k(y_i)$ are defined in terms of pairs of observations (y_{ir}, y_{is}) from the bivariate marginal density $f(y_{ir}, y_{is}; \theta)$, the pairwise likelihood is obtained (Cox Reid 2004)

$$pL(\theta; y) = \prod_{i=1}^{n} \prod_{r=1}^{q-1} \prod_{s=r+1}^{q} f(y_{ir}, y_{is}; \theta)$$

- The pairwise loglikelihood is

$$p\ell(\theta; y) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \log f(y_{ir}, y_{is}; \theta)$$

- The pairwise score function is

$$pU(\theta; y) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \frac{\partial}{\partial \theta} \log f(y_{ir}, y_{is}; \theta)$$

- The validity of inference on θ using $cL(\theta; y)$ can be justified invoking the theory of unbiased estimating functions.
- Indeed, $cU(\theta; y)$ is still an unbiased estimating function, since it is a linear combination of valid score functions.
- The composite MLE $\hat{\theta}_c$ is consistent and approximately normal with mean θ and variance

$$V(\theta) = H(\theta)^{-1} J(\theta) H(\theta)^{-1}$$

with $H(\theta) = E(-\partial cU(\theta)/\partial\theta^{\mathsf{T}})$ and $J(\theta) = E(cU(\theta) cU(\theta)^{\mathsf{T}})$.
- Matrix $G(\theta) = V(\theta)^{-1}$ is the Godambe information.

- The asymptotic distribution of the Wald-type statistic $cw_w(\theta) = (\hat{\theta}_c - \theta)^\mathsf{T} G(\theta)(\hat{\theta}_c - \theta)$ is χ^2_d. The same result holds for the score-type statistic $cw_s(\theta) = cU(\theta)^\mathsf{T} J(\theta)^{-1} cU(\theta)$.
- Let $cw(\theta) = 2(c\ell(\hat{\theta}_c) - c\ell(\theta))$ be the composite likelihood ratio statistic.
- Its asymptotic null distribution is

$$cw(\theta) \dot{\sim} \sum_{i=1}^{d} \lambda_i Z_i^2$$

  with Z_i^2 independent χ^2_1 random variables and λ_i eigenvalues of $H(\theta)^{-1} J(\theta)$.
- All the above results extend to the case of partial interest about ψ, with $\theta = (\psi, \lambda)$.

- Wald-type statistics lack invariance under reparameterization and force confidence regions to have an elliptical shape.
- Score-type statistics seem to suffer from numerical instability (Molenberghs Verbeke 2005, Ch. 9).
- Under this respect, a likelihood ratio type statistic would be more appealing.
- However, its approximate $\sum \lambda_i Z_i^2$ distribution departs from the familiar pivot result. This calls for adjustments in order to obtain the standard χ_d^2 distribution:
  - For $d = 1$, most proposed adjustments agree and lead to the exact asymptotic reference.
  - For $d > 1$, some adjustments are not parameterization invariant or only match some moments of the asymptotic reference.
- All the adjustments require the evaluation of $H(\theta)$ and $J(\theta)$.

- Simple adjustments are based on moments conditions:
  1. First order moment matching gives $cw_1(\theta) = cw(\theta)/\tilde{\lambda}$, with $\tilde{\lambda} = \sum \lambda_i/d = tr(H(\theta)^{-1}J(\theta))/d$, with a χ^2_d approximate null distribution.
  2. First and second order moment matching gives the Satterthwaite (1946) adjustment $cw_2(\theta) = cw(\theta)/\kappa$, with a χ^2_ν approximate null distribution, where $\kappa = \sum \lambda_i^2/\sum \lambda_i$ and $\nu = (\sum \lambda_i)^2/(\sum \lambda_i^2)$.
  3. Matching of moments up to higher order are available (see Lindsay et al 2000).
- Chandler and Bate (2007) propose a vertical scaling of $cw(\theta)$ giving $cw_{cb}(\theta) = cw(\theta)cw_w(\theta)/(\hat{\theta}_c - \theta)^\top H(\hat{\theta}_c)(\hat{\theta}_c - \theta)$ having χ^2_d null distribution, but which is not parameterization invariant.
- Pace et al (2010) propose the parameterization invariant scaling $cw_{inv}(\theta) = cw(\theta)cw_s(\theta)/cU(\theta)^\top H(\theta)^{-1}cU(\theta)$ also having the usual asymptotic null distribution.

Empirical likelihood from the composite score function

- We can define an empirical likelihood based on a general unbiased estimating equation for $\theta \in \mathbb{R}^d$:

$$\eta(y;\theta) = \frac{1}{m}\sum_{j=1}^{m} \eta_j(\mathcal{Y}_j;\theta) = 0 \ , \quad \text{with } \mathcal{Y}_j \subset \mathcal{Y}$$

- The empirical likelihood is defined as (Owen 2001)

$$L_e(\theta) = \frac{1}{m}\prod_{j=1}^{m} \frac{1}{(1+\lambda^{\mathsf{T}}\eta_j(\mathcal{Y}_j;\theta))}$$

where the Lagrangian multiplier λ satisfies $(1/m)\sum_{j=1}^{m} \frac{\eta_j(\mathcal{Y}_j;\theta)}{(1+\lambda^{\mathsf{T}}\eta_j(\mathcal{Y}_j;\theta))} = 0$

- The empirical likelihood ratio statistic for θ derived from $\eta(y;\theta)$ is

$$w_e(\theta) = 2\sum_{j=1}^{m} \log(1+\lambda^{\mathsf{T}}\eta(\mathcal{Y}_j;\theta))$$

- The empirical composite likelihood ratio statistic derived from $\eta(y;\theta) = cU(\theta)$ is

$$cw_e(\theta) = 2\sum_{k=1}^{K} \log(1 + \lambda^{\mathsf{T}} cU(\theta; A_k))$$

- Under suitable conditions (see Adimari and Guolo 2010) it can be shown that:
  1. When $d = 1$, $cw_e(\theta)/\tilde{\lambda} \dot{\sim} \chi^2_1$.
  2. When $d > 1$, the asymptotic null distribution of $cw_{e1}(\theta) = cw_e(\theta)/\tilde{\lambda}$ can be approximated with a χ^2_d (as for $cw_1(\theta)$).
- These results hold also for the pairwise score function

$$pU(\theta) = \sum_{k=1}^{K} pU(\theta; A_k)$$

with $K = nq(q-1)/2$, obtaining $pw_{e1}(\theta)$.

- Let us focus on the pairwise likelihood function.
- The pairwise score function with $K = n$ can be written with

$$pU(\theta; y_i) = \sum_{s=1}^{q-1} \sum_{r=s+1}^{q} \frac{\partial}{\partial \theta} \log f(y_{is}, y_{ir}; \theta)$$

- The pairwise empirical likelihood ratio is

$$pw_e(\theta) = 2 \sum_{i=1}^{n} \log(1 + \lambda^{\mathsf{T}} pU(\theta; y_i))$$

- In this situation, we have $pw_e(\theta) \dot{\sim} \chi^2_d$ (the proof follows from Adimari and Guolo 2010).

Simulation results

- One-way normal-theory random effects model:
  $Y_{ir} = \mu + \xi_i + \epsilon_{ir}$, $i = 1, \ldots, n$, $r = 1, \ldots, q$, and ξ_i and ϵ_{ir} independently normally distributed with zero mean and variances σ_ξ^2 and σ_ϵ^2.
- The problem can be reformulated by writing Y_i as a multivariate normal with components having mean μ and variance $\sigma^2 = \sigma_\xi^2 + \sigma_\epsilon^2$, and with correlation $\rho = \sigma_\xi^2/(\sigma_\xi^2 + \sigma_\epsilon^2)$ between any two components of the same vector.
- This example has been chosen so that we can easily do closed form calculations both of complete and pairwise likelihood quantities, and not for direct interest in the application of composite likelihood.
- The special case with $\mu = 0$, $\sigma^2 = 1$ and $\theta = \rho$ has been treated in detail by Cox Reid (2004).
- Here interest on inference about $\theta = (\mu, \sigma^2, \rho)$.

- The pairwise likelihood is

$$p\ell(\theta) = -\frac{nq(q-1)}{2}\log\sigma^2 - \frac{nq(q-1)}{4}\log(1-\rho^2)$$

$$-\frac{q-1+\rho}{2\sigma^2(1-\rho^2)}SS_W - \frac{q(q-1)SS_B + nq(q-1)(\bar{y}-\mu)^2}{2\sigma^2(1+\rho)}$$

with $SS_W = \sum_{i=1}^n (\bar{y}_i - \bar{y})^2$ and $SS_B = \sum_{i=1}^n \sum_{r=1}^q (y_{ir} - \bar{y}i)^2$.

- For this model the pairwise MLE coincides with the full MLE Mardia et al (2009). Moreover, $pU(\theta) = J(\theta)H(\theta)^{-1}U(\theta)$, so that $G(\theta) = i(\theta)$. As a consequence the Wald and the score statistics based on the full likelihood coincide with those based on the pairwise likelihood.
- We run a simulation experiment with three values of ρ (from a moderate to a strong correlation). We computed the empirical coverages of confidence regions based on several statistics.

$q = 30$	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
$n = 15$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$w(\theta)$	0.891	0.943	0.987	0.889	0.941	0.987	0.888	0.941	0.987
$pw_1(\theta)$	0.838	0.890	0.949	0.839	0.892	0.952	0.845	0.899	0.959
$pw_2(\theta)$	0.865	0.919	0.972	0.863	0.919	0.972	0.869	0.924	0.976
$pw_w(\theta)$	0.809	0.860	0.924	0.776	0.831	0.900	0.715	0.767	0.837
$pw_s(\theta)$	0.906	0.947	0.983	0.906	0.947	0.983	0.905	0.948	0.983
$pw_{cb}(\theta)$	0.831	0.884	0.944	0.820	0.876	0.941	0.762	0.818	0.891
$pw_{inv}(\theta)$	0.907	0.953	0.989	0.898	0.948	0.989	0.890	0.941	0.986
$pw_{e1}(\theta)$	0.904	0.953	0.990	0.907	0.949	0.989	0.848	0.871	0.880
$pw_e(\theta)$	0.886	0.930	0.976	0.884	0.935	0.949	0.856	0.870	0.888
$pw_{e,inv}(\theta)$	0.955	0.988	0.988	0.892	0.926	0.946	0.820	0.846	0.865
$n = 30$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$w(\theta)$	0.892	0.944	0.987	0.896	0.944	0.988	0.894	0.945	0.988
$pw_1(\theta)$	0.855	0.905	0.961	0.855	0.906	0.967	0.868	0.919	0.974
$pw_2(\theta)$	0.882	0.931	0.980	0.879	0.933	0.982	0.891	0.940	0.985
$pw_w(\theta)$	0.850	0.900	0.955	0.824	0.879	0.941	0.709	0.763	0.831
$pw_s(\theta)$	0.901	0.947	0.986	0.902	0.947	0.984	0.902	0.948	0.985
$pw_{cb}(\theta)$	0.861	0.914	0.967	0.852	0.908	0.963	0.743	0.796	0.869
$pw_{inv}(\theta)$	0.900	0.949	0.989	0.898	0.947	0.989	0.893	0.942	0.986
$pw_{e1}(\theta)$	0.900	0.950	0.990	0.900	0.946	0.976	0.871	0.923	0.958
$pw_e(\theta)$	0.815	0.876	0.937	0.826	0.883	0.941	0.855	0.903	0.951
$pw_{e,inv}(\theta)$	0.903	0.952	0.988	0.891	0.925	0.952	0.869	0.920	0.951

- Correlated binary outcomes: Multivariate probit model with logistic marginal and constant cluster sizes.
- The pairwise likelihood is

$$p\ell(\theta) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \log Pr(Y_{ir} = y_{ir}, Y_{is} = y_{is}; \theta)$$

with $Pr(Y_{ir} = 1, Y_{is} = 1; \theta) = \Phi_2(x_{ir}\beta/\sigma, x_{is}\beta/\sigma; \rho)$.

- Pairwise likelihood inference is much simpler than full likelihood inference since it involves only bivariate normal integrals
- Here interest on inference about $\theta = (\beta, \rho)$, with $\sigma = 1$.
- We run a simulation experiment with three values of ρ. We computed the empirical coverages of confidence regions based on several statistics.

	$q=3$		$q=6$		$q=10$	
n	50	80	50	80	50	80
				$\rho=0.25$		
$pw_1(\theta)$	0.935	0.941	0.925	0.931	0.919	0.930
$pw_{e1}(\theta)$	0.934	0.942	0.934	0.937	0.927	0.933
$pw_e(\theta)$	0.908	0.933	0.913	0.933	0.914	0.932
				$\rho=0.50$		
$pw_1(\theta)$	0.934	0.943	0.931	0.937	0.916	0.928
$pw_{e1}(\theta)$	0.932	0.943	0.935	0.939	0.922	0.932
$pw_e(\theta)$	0.921	0.934	0.924	0.932	0.921	0.940
				$\rho=0.50$		
$pw_1(\theta)$	0.925	0.934	0.931	0.938	0.920	0.925
$pw_{e1}(\theta)$	0.916	0.932	0.932	0.940	0.923	0.930
$pw_e(\theta)$	0.898	0.922	0.925	0.934	0.924	0.935

Concluding remarks

- The proposed statistic show reasonable coverage performances and are in general accurate.
- For large q, $pw_e(\theta)$ appears preferable to $pw_{e1}(\theta)$.
- Also moment matching and Pace et al adjustments perform well but they all require the evaluation of the matrices $H(\theta)$ and $J(\theta)$. The estimation/approximation of $H(\theta)$ and $J(\theta)$ is an open issue (see Varin et al 2010).
- Bayesian application of the empirical composite likelihood is under investigation, following Lazar (2003) and Pauli et al (2010).

Some references

- Adimari, Guolo (2010). To appear in *Statist. Meth. and Appl.*.
- Chanler, Bate (2007). *Biometrika*, 94, 167-183.
- Cox, Reid (2004). *Biometrika*, 91, 729-737.
- Lazar (2003). *Biometrika*, 90, 319-326.
- Lindsay (1988). *Biometrika*, 69, 19-27.
- Lindsay, Pilla, Basak (2000). *Ann. Inst. Statist. Math.*, 52, 215-230.
- Mardia, Kent, Hughes, Taylor (2009). *Biometrika*, 96, 975-982.
- Molenberghs, Verbeke (2005). *Springer*, New York.
- Owen (2001). *Chapman and Hall*, London.
- Pace, Salvan, Sartori (2010). To appear in *Stat. Sinica*, special issue on composite likelihood.
- Pauli, Racugno, Ventura (2010). To appear in *Stat. Sinica*, special issue on composite likelihood.
- Satterthwaite (1946). *Biometrics Bull.*, 2, 110-114.
- Varin, Reid, Firth (2010). To appear in *Stat. Sinica*, special issue on composite likelihood.