Some results of Composite Likelihoods

Efficiency and Higher order asymptotics

Zi Jin

Department of Statistics, University of Toronto

April 2008, The University of Warwick

1. Composite marginal likelihoods: Definition

- Low-dimensional Marginal densities (Cox & Reid, 2004).
- Univariate and bivariate distributions are often adopted, and this leads to the so-called pairwise likelihood.
- For a single vector Y, the 1st and 2nd-order log-likelihoods contribute

$$l_1(\theta; Y) = \sum_s \log f(Y_s; \theta), \quad (1)$$

$$l_2(\theta; Y) = \sum_{s>t} \log f(Y_s, Y_t; \theta) - a l_1(\theta; Y). \quad (2)$$

a is a pre-specified constant. $a = 0$: taking all possible bivariate dist'ns and leading to the pairwise log-likelihood. $a = 1/2$: taking all possible conditional dist'ns of one component given another.

2. Composite score functions and estimators

Composite score functions

For n i.i.d. vectors $Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}$ the joint composite log-likelihoods are defined by addition:

$$l_\nu(\theta; Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}) = \sum_i l_\nu(\theta; Y^{(i)}), \nu = 1, 2.$$

Composite score functions are defined in the usual way by the composite likelihood derivatives:

$$U_\nu(\theta; Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}) = \partial l_\nu(\theta; Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)})/\partial\theta, \nu = 1, 2.$$

Maximum Composite likelihood estimator

The estimating equations

$$U_\nu(\theta; Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}) = 0$$

give out the *maximum composite likelihood estimator* $\tilde{\theta}$.

3. Efficiency of Composite likelihoods

Intraclass Correlation Normal

$Y = (Y^{(1)}, \ldots, Y^{(n)})^T$, and $Y^{(1)}, \ldots, Y^{(n)} \sim i.i.d. N_q(\underset{\sim}{\mu}, \Sigma)$ where

$\Sigma = \sigma^2((1-\rho)I + \rho J))$ and $\theta = (\mu, \sigma^2, \rho)^T$.

- Full log-likelihood function of Y: $l(\theta, Y)$
- Pairwise log-likelihood function:

$$l_2(\underset{\sim}{\theta}; Y) = \sum_i \sum_{s>r}^{q} \log f(Y_r^{(i)}, Y_s^{(i)}; \underset{\sim}{\theta})$$

- Pseudo log-likelihood function: $l^* = l_2 - aql_1$

$$\hat{\theta} = \tilde{\theta} = \theta^* = (\bar{y}_{..}, \frac{S_1}{nq}, \frac{2S_2}{(q-1)S_1})^T \qquad (3)$$

$\bar{y}_{..} = \frac{1}{nq}\sum\sum_{r=1}^{q} Y_r^{(i)}$, $S_1 = \sum\sum_{r=1}^{q}(Y_r^{(i)} - \bar{y}_{..})^2$, and $S_2 = \sum\sum_{s>r}^{q}(Y_r^{(i)} - \bar{y}_{..})(Y_s^{(i)} - \bar{y}_{..})$.

3.1 Introclass Correlation Normal: Information Matrices

Question: $I(\theta) = H(\theta)J^{-1}(\theta)H(\theta)$?

Smth we need to obtain to show the equivalence:

- obtain $\dot{\Sigma}_{\sigma^2}, \ddot{\Sigma}_{\sigma^2}, \dot{\Sigma}_{\rho}, \ddot{\Sigma}_{\rho}$, derivatives of Σ with respect to σ^2 and ρ;
- $H(\theta) = E(-\frac{\partial^2 l_2}{\partial\theta\partial\theta^T})$, easy one;
- $J(\theta) = E(\frac{\partial l_2}{\partial\theta}\frac{\partial l_2}{\partial\theta^T})$, difficult one.;
- need to obtain $H(\theta), J(\theta)$ for both pairwise and pseudolikelihood.

For example, $V_{rs} = Z_r^2 + Z_s^2 - \frac{2}{\rho}Z_r Z_s$

$$\sum_{s>r}^{q}(V_{rs})^2 = \frac{1}{4}\sum_r\sum_{s\neq r}\sum_{r'}\sum_{r'\neq s'} V_{rs}V_{r's'}$$

$$= \frac{1}{4}[\sum_r\sum_{s\neq r} rssr + \sum_r\sum_{s\neq r}\sum_{s'\neq r,s} rsss' + \sum_r\sum_{s\neq r} rsrs + \sum_r\sum_{s\neq r}\sum_{s'\neq s,r} rsrs'$$
$$+\sum_r\sum_{s\neq r}\sum_{r'\neq s,r} rsr'r + \sum_r\sum_{s\neq r}\sum_{s'\neq s,r} rsr's + \sum_r\sum_{s\neq r}\sum_{r'\neq s,r}\sum_{s'\neq r,r',s} rsr's']$$

$I = HJ^{-1}H = H^*J^{-1^*}H^*$, regardless of the value of a.

3.2 Symmetric Normal

$$Y^{(i)} = (Y_1^{(i)}, \ldots, Y_q^{(i)})^T \sim N_q(\underset{\sim}{0}, V) = N_q(\underset{\sim}{0}, (1-\rho)I + \rho J)$$

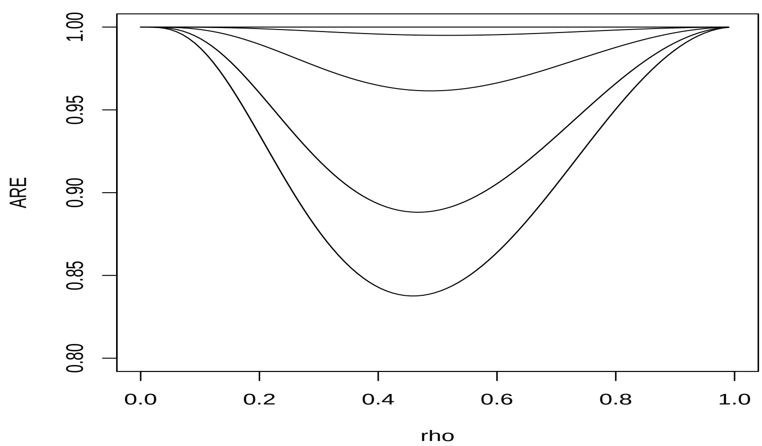


FIG 1 Ratio of asympotic variance of rhohat to rhotilde for q=2,3,5,8,10.

3.3 Unrestricted Multivariate Normal

- In Mardia 2007, conditional pairwise pseudolikelihood is defined as

$$\text{PPL} = \prod_{i=1}^{n} \prod_{j \neq k}^{q} f(y_{ji}|y_{ki}; \Sigma)$$

  and has been showed to be fully efficient for unrestricted multivariate normal distributions.

- Let $a = 1/2$, $\log CR = \log PPL$.
-

$$\begin{pmatrix} y_{ji} \\ y_{ki} \end{pmatrix} = CY \sim N(\underset{\sim}{0}, C\Sigma C')$$

  if Σ unrestricted, $\hat{\Sigma}_{jk} = \widehat{C\Sigma C'} = C\hat{\Sigma}C' = (\hat{\Sigma})_{jk}$.

3.4 AR(1)

The working full likelihood of AR(1) process is given by

$$L(a, \sigma^2) = f(x_1) \prod_{t=2}^{T} f(x_t|x_{t-1})$$

The associated log-likelihood is

$$l(a, \sigma^2) = \frac{1}{2}\log(1 - a^2) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}[S_1 + a^2 S_2 - 2aS_{12}] \quad (4)$$

where $S_1 = \sum_{t=1}^{T} x_t^2$, $S_2 = \sum_{t=2}^{T-1} x_t^2$, $S_{12} = \sum_{t=1}^{T-1} x_t x_{t+1}$.

Pairwise likelihood for AR(1)

We propose a composite likelihood formed by adjacent pairs, denoted by L_2

$$L_2 = \prod_{t=2}^{T} f(x_t, x_{t-1}) \quad (5)$$

where (X_t, X_{t-1}) follows a bivariate normal distribution with mean zero variances $\sigma^2/(1-a^2)$ and correlation a.

$$L_2(a,\sigma^2) = f(x_1)\prod_{t=2}^{T} f(x_t|x_{t-1}) \prod_{t=2}^{T} f(x_t) = L(a,\sigma^2)\prod_{t=2}^{T} f(x_t)$$

where $L(a,\sigma^2)$ is the full likelihood function. That is,

$$l_2(a,\sigma^2) = l(a,\sigma) + \sum_{t=2}^{T} f(x_t) \quad (6)$$

3.4.1 Max likelihood and pairwise likelihood estimation

MLE

$$0 = (1-\frac{1}{T})S_2\hat{a}^3 - (1-\frac{2}{T})S_{12}\hat{a}^2 - (S_2 + \frac{S_1}{T})\hat{a} + S_{12} \quad (7)$$

$$\hat{\sigma}^2 = (S_1 + \hat{a}^2 S_2 - 2\hat{a}S_{12})/T \quad (8)$$

For large T, it can be show $\hat{a} \cong S_{12}/S_2$.

MPLE

$$\tilde{a} = \frac{2S_{12}}{S_1+S_2}, \quad (9)$$

$$\tilde{\sigma}^2 = \frac{(S_1+S_2)^2 - 4S_{12}^2}{2(T-1)(S_1+S_2)}. \quad (10)$$

For large T, $\hat{a} = \tilde{a}$.

3.4.2 Simulations

Table I Maximum likelihood estimatiors and Maximum pairwise likelihood estimators for AR(1) model $X_t = 0.55X_{t-1} + \epsilon_t$, with $\epsilon \sim N(0, 0.1^2)$.

	T=200		T=100		T=20	
AR	a	σ	a	σ	a	σ
MLE	0.5651	0.0970	0.5604	0.1006	0.4978	0.0661
MPLE	0.5579	0.0096	0.5556	0.0103	0.5037	0.0046

Sample Variance

Table II Sample covariance matrices of Full and Pairwise likelihood for AR(1) model $X_t = 0.55X_{t-1} + \epsilon_t$, with $\epsilon \sim N(0, 0.1^2)$ with different lengths T.

	N=100,T=200	N=200,T=100
ratio	0.2017444	0.1973112
$\hat{I}(\hat{\theta})$	$\begin{pmatrix} 4.778e-03 & 1.186e-05 \\ 1.186e-05 & 2.318e-05 \end{pmatrix}$	$\begin{pmatrix} 6.260e-03 & -1.958e-05 \\ -1.958e-05 & 4.058e-05 \end{pmatrix}$
$\widehat{HJ^{-1}H}(\hat{\theta})$	$\begin{pmatrix} 4.796e-03 & 2.801e-06 \\ 2.801e-06 & 9.030e-07 \end{pmatrix}$	$\begin{pmatrix} 6.380e-03 & -3.087e-06 \\ -3.087e-06 & 1.568e-06 \end{pmatrix}$

	N=100,T=20	N=10,T=20
ratio	0.1999862	0.2137193
$\hat{I}(\hat{\theta})$	$\begin{pmatrix} 3.758e-02 & 4.150e-04 \\ 4.150e-04 & 2.417e-04 \end{pmatrix}$	$\begin{pmatrix} 7.102e-02 & 9.829e-04 \\ 9.829e-04 & 2.537e-04 \end{pmatrix}$
$\widehat{HJ^{-1}H}(\hat{\theta})$	$\begin{pmatrix} 3.646e-02 & 5.658e-05 \\ 5.658e-05 & 9.470e-06 \end{pmatrix}$	$\begin{pmatrix} 6.551e-02 & 1.449e-04 \\ 1.449e-04 & 1.254e-05 \end{pmatrix}$

	N=100,T=3	N=10,T=3
ratio	0.2449904	0.1826142
$\hat{I}(\hat{\theta})$	$\begin{pmatrix} 0.2730 & -0.0024 \\ -0.0024 & 0.0012 \end{pmatrix}$	$\begin{pmatrix} 1.5296e-01 & -5.0205e-03 \\ -5.0205e-03 & 1.4909e-03 \end{pmatrix}$
$\widehat{HJ^{-1}H}(\hat{\theta})$	$\begin{pmatrix} 0.2545 & 0.0004 \\ 0.0004 & 0.00006 \end{pmatrix}$	$\begin{pmatrix} 1.42323e-01 & 1.1454e-03 \\ 1.1454e-03 & 5.3282e-05 \end{pmatrix}$

Relative Efficiency

The efficiency of pairwise likelihood compared to full likelihood for AR(1) model can be obtained as

$$\left\{ \frac{|H(\theta)J^{-1}(\theta)H(\theta)|}{|I(\theta)|} \right\}^{1/2} \qquad (11)$$

where $\theta = (a, \sigma^2)'$ is a 2-dimensional parameter vector.

$$E\{(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)\} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$$

Table III Efficiency of pairwise likelihood for AR(1) model $X_t = 0.55X_{t-1} + \epsilon_t$, with $\epsilon \sim N(0, 0.1^2)$ with different lengths T.

	T=100	T=20	T=8	T=4
Efficiency	0.2296845	0.4101251	0.5005665	0.5636764

Figure 1: Efficiency against T

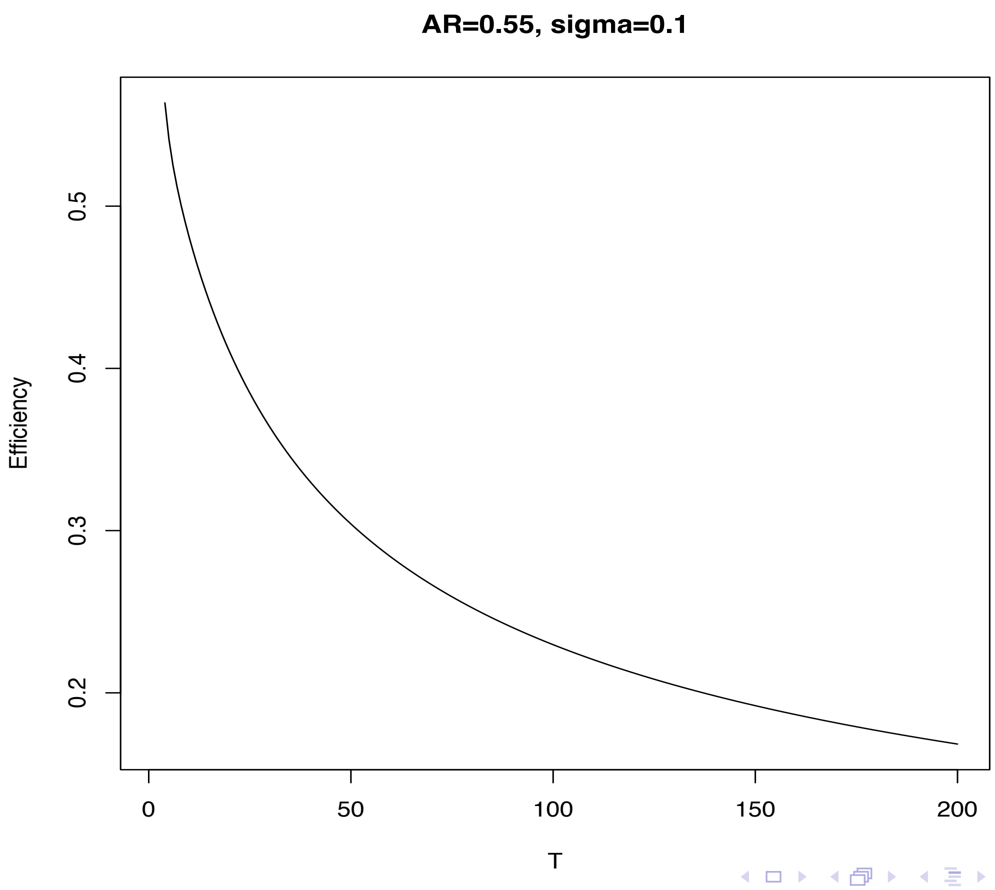


Figure 2: Given $T = 4$, Efficiency against a and σ.

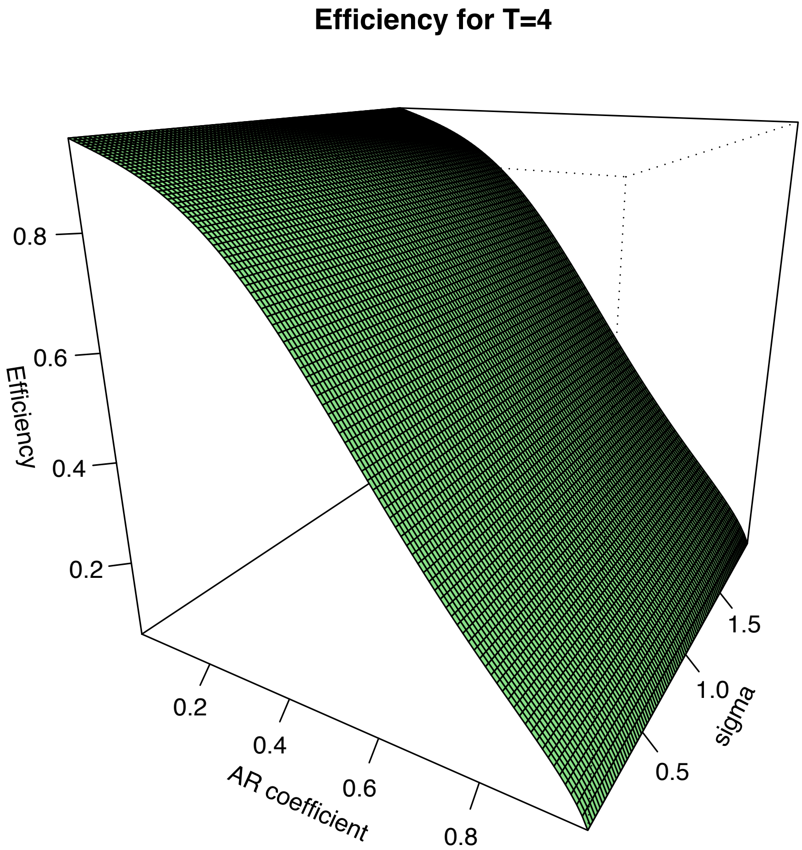
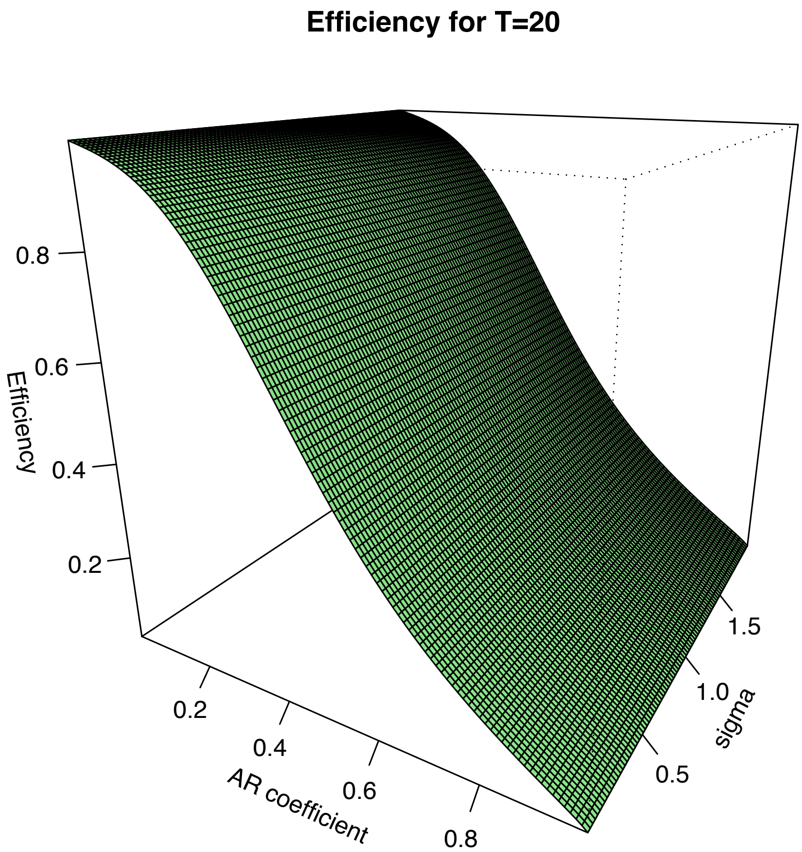


Figure 3: Given $T = 20$, Efficiency against a and σ.



4. Higher order asymptotics

- Higher accuracy inference in terms of $\tilde{r}$ under simple null hypothesis under *the first scenario*.
- Expansions for the expectation and variance of the composite signed likelihood ratio root are found to the order $n^{-3/2}$ and n^{-2}, respectively.
- Ideally we would expect that

$$E(\tilde{r}) = m(\theta) + O(n^{-3/2}), Var(\tilde{r}) = 1 + v(\theta) + O(n^{-2})$$

where $m(\theta)$ is of order $O(n^{-1})$ and $v(\theta)$ is of order $O(n^{-1})$.

- Then, the standard normal approximation to the distribution of

$$\frac{\tilde{r} - E(\tilde{r})}{\{Var(\tilde{r})\}^{1/2}}$$

has an error of order $O(n^{-3/2})$.

Expansion of $\tilde{r}^2$ with asymptotic magnitude of order $n^{-3/2}$

$$\begin{aligned}\tilde{r}^2 &= u^{rs}\tilde{l}_r\tilde{l}_s + u^{rs}u^{tu}H_{rt}\tilde{l}_s\tilde{l}_u + \frac{1}{3}u^{rs}u^{tu}u^{vw}\nu_{rtv}\tilde{l}_s\tilde{l}_u\tilde{l}_w \\ &+u^{rs}u^{tu}u^{vw}H_{rt}H_{sv}\tilde{l}_u\tilde{l}_w + \frac{1}{4}u^{rs}u^{tu}u^{vw}u^{xy}u^{zp}\nu_{rtv}\nu_{sxz}\tilde{l}_u\tilde{l}_w\tilde{l}_y\tilde{l}_p \\ &+u^{rs}u^{tu}u^{vw}u^{xy}\nu_{rtv}H_{sx}\tilde{l}_u\tilde{l}_w\tilde{l}_y + \frac{1}{3}u^{rs}u^{tu}u^{vw}H_{rtv}\tilde{l}_s\tilde{l}_u\tilde{l}_w \\ &+\frac{1}{12}u^{rs}u^{tu}u^{vw}u^{xy}\nu_{rtvx}\tilde{l}_s\tilde{l}_u\tilde{l}_w\tilde{l}_y + O_p(n^{-3/2})\end{aligned} \quad (12)$$

- The leading term $u^{rs}\tilde{l}_r\tilde{l}_s$ is the composite score statistic.
- The 1st order result of the composite likelihood ratio statistic

$$E[\tilde{r}^2] = \mu^{rs}E[\tilde{l}_r\tilde{l}_s] + O(n^{-1/2}) = \mu^{rs}\nu_{r,s} + O(n^{-1/2}).$$

Expansion of the expectation of $\tilde{r}^2/\tilde{r}$

Similar as defined in Peter McCullagh(1987, Ch.7)[10], let

$$\begin{aligned}\tilde{r}_t &= \tilde{l}_t + \frac{1}{6}u^{rs}(3H_{rt}\tilde{l}_s + u^{vw}v_{rtv}\tilde{l}_s\tilde{l}_w) \\ &+ \frac{1}{72}u^{rs}u^{vw}\Big[27H_{rt}H_{sv}\tilde{l}_w + 8u^{xy}u^{zp}v_{rtv}v_{sxz}\tilde{l}_w\tilde{l}_y\tilde{l}_p \\ &\quad + 30u^{xy}v_{rtv}H_{sx}\tilde{l}_w\tilde{l}_y + 12H_{rtv}\tilde{l}_s\tilde{l}_w + 3u^{xy}v_{rtvx}\tilde{l}_s\tilde{l}_w\tilde{l}_y\Big] \quad (13)\end{aligned}$$

It can be verified that

$$\tilde{r}^2 = \tilde{r}_t\tilde{r}_u u^{tu} + O_p(n^{-3/2}) \quad (14)$$

while

$$\begin{aligned}E(\tilde{r}_t) &= \frac{1}{6}u^{rs}\{3E(\tilde{l}_{rt}\tilde{l}_s) + u^{vw}v_{rtv}E(\tilde{l}_s\tilde{l}_w)\} + O(n^{-3/2}) \\ &= \frac{1}{6}u^{rs}\{3v_{rt,s} + u^{vw}v_{rtv}v_{s,w}\} + O(n^{-3/2}) \quad (15)\end{aligned}$$

Expansion of the expectation of $\tilde{r}^2/\tilde{r}$ with Bartlett Identities

If the second Bartlett Identity holds, equation (15) can be simplified into

$$E(\tilde{r}_t) = \frac{1}{6}u^{rs}\{3v_{rt,s} + v_{rst}\} + O(n^{-3/2}) \quad (16)$$

consistent with the results shown in Barndorff & Cox 1994.

It suffices to write the variance of $\tilde{r}$ as

$$Var(\tilde{r}) = c\{1 + v(\theta)\} + O(n^{-2})$$

where $v(\theta)$ is of order $O(n^{-1})$, computed directly with sufficient algebraic diligence.

Comments

- For certain distributions, composite marginal likelihoods may yield full MLE and are fully efficient.
- Loss of efficiency is, in general, not ignorable for curved exponential family, or restricted normal distributions.
- Though from time to time we tend to overlook the fact that composite likelihood estimators are not consistent or asymptotically normally distributed especially under the second scenario, their asymptotic behaviors are still worth of further consideration.
- Profile based inference for AR(1) model, as σ^2 can be treated as a nuisance parameter. Performance of $\tilde{r}$ and its higher order approximation.

Arnold, B.C. and Strauss, D. (1991), Pseudolikelihood estimation: some examples, *Sankhya: The Indian Journal of Statistics*, **53**, 233-243.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994), *Inference and Asymptotics*, Chap.

Besag, J.E. (1974), *Spatial interaction and the statistical analysis of lattice systems (with discussion)*, *Journal of the Royal Statistical Society*, **B 34**, 192-236.

Cox, D.R. and Reid, N. (2004), *A note on pseudolikelihood constructed from marginal densities*, *Biometrika*, **91**, 727-737.

DiCiccio, T.J. and Stern, S.E. (1994), *Constructing Approximately Standard Normal Pivots from Signed Roots of Adjusted Likelihood Ratio Statistics*, *The Scandinavian Journal of Statistics*, **21**, 447-460.

Geys, H., Molenberghs, G. and Ryan, L. M. (1999), *Pseudolikelihood Modeling of Multivariate Outcomes in Developmental Toxicology*, *Journal of the American Statistical Association*, **94**, 734-745.

- Kent, J.T. (1982), *Robust Properties of Likelihood Ratio Test*, *Biometrika*, **69**, 19-27.
- Lindsay, B.G. (1988). *Composite likelihood methods*, *Contemporary Mathematics*, **80**, 221-240.
- Mardia K.V., Hughes G. and Taylor C.C. (2007), *Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions*, *Research reports*, **STAT07-02**, Department of Statistics, University of Leeds.
- McCullagh, P. (1987), *Tensor Methods in Statistics*, Chapman and Hall.
- Severini, T.A. (2000), *Likelihood Methods in Statistics*, Oxford University Press.
- Stern, S.E.(2006), *Simple and accurate one-sided inference based on a class of M-estimators*, *Biometrika*, **93**, 973-987.
- Varin, C.(2007), *On composite marginal likelihoods*, preprint.