Pairwise likelihood inference for parameter-driven models

Xanthi Pedeli & Cristiano Varin

SIS 2017 @ Florence / June 30, 2017
Credits

Xanthi Pedeli

Marie Skłodowska-Curie fellow at Ca’ Foscari

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 699980

Pairwise likelihood inference
Time series of disease counts

Hepatitis A in Germany 2001-2004

Source: surveillance R package (Höhle et al., 2015)
Typical questions in disease surveillance:

- Early detection of outbreaks, without false alarms!
- Patterns? Seasonality? Trends?
- Which factors do influence disease counts?
- What is the effect of public health policies?
- What is the effect of vaccinations?
- ...
Parameter-driven models

Popular approach to extend generalized linear models to time series (Cox, 1981)

Hierarchical model specification:

- **observed process**

\[ Y_t | U_t \sim \text{GLM}(\mu_t, \gamma_t), \]

\[ \mu_t = g(x_t^\top \beta + U_t), \quad \text{var}(Y_t) = \gamma_t V(\mu_t) \]

- **latent process** $U_t$ is a stationary ARMA(p, q)

Interpretation: the linear predictor is ‘imperfect’

\[ x_t^\top \beta + U_t \]

the latent variable $U_t$ plays the role of a serially correlated error that adjusts for unobserved effects

Pairwise likelihood inference
**Intractable likelihood**

Likelihood computed integrating out the latent variables

\[
L(\theta; y) = \int \cdots \int \prod_{t=1}^{n} f(y_t|u_t; \theta)f(u_1, \ldots, u_n; \theta) \, dz_1 \cdots dz_n
\]

where \( \theta \) are the model parameters

In general, the likelihood integral has no closed-form solution:

- MCMC
- MCEM or quasi-MCEM
- importance sampling
- sequential Laplace approximations (INLA)
- composite likelihoods
- \ldots
What are composite likelihoods?

Full likelihood is intractable but low-dimensional distributions readily computed

Composite likelihood idea: combine low-dimensional terms to construct a pseudolikelihood

General setup:

- collection of marginal or conditional events \( \{A_1, \ldots, A_K\} \)
- associated component likelihoods \( L_k(\theta; y) \propto f(y \in A_k; \theta) \)

A composite likelihood is the weighted product

\[
CL(\theta; y) = \prod_{k=1}^{K} L_k(\theta; y)^{w_k}
\]

for some weights \( w_k \geq 0 \)
How it works

CLs are misspecified likelihoods constructed under the working assumption of independence between the component likelihoods.

Under suitable regularity conditions, the maximum composite likelihood estimator is consistent.

Limit distribution

\[ G(\theta)^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}(0, I_p), \]

where \( G(\theta) \) is the Godambe information,

\[ G(\theta) = H(\theta)J(\theta)^{-1}H(\theta), \]

with

\[ H(\theta) = -E\{\nabla^2 \log \text{CL}(\theta)\} \quad \text{and} \quad J(\theta) = \text{var}\{\nabla \log \text{CL}(\theta)\} \]
**Pairwise likelihood**

Pairwise likelihood is a special case of composite likelihood

Pairwise likelihood of order $d$ formed by pairs of obs separated at most $d$ units

$$L_2^{(d)}(\theta) = \prod_{t=d+1}^{n} \prod_{i=1}^{d} \log p(y_t, y_{t-i}; \theta),$$

where each component is a two-dimensional integral

$$p(y_t, y_{t-i}; \theta) = \int \int p(y_t|u_t)p(y_{t-i}|u_{t-i})p(u_t, u_{t-i}) du_t \, du_{t-i}$$

Merit: replace the intractable $n$-dimensional integral of the full likelihood with a set of bivariate integrals
How many pairs?

- Should we use all possible pairs?

Not a good idea. As the series length diverges, the number of pairs formed by independent observations also diverges making the estimator inconsistent.

- Should we use all possible correlated pairs?

Neither a good idea. We should keep at board only those pairs that are genuinely correlated, avoiding spurious effects. Otherwise, we pay in terms of efficiency loss.
Pairwise expectation-maximization

Sequence of estimators $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \ldots$.

At each step, $\hat{\theta}^{(k)}$ is defined as the maximizer of

$$Q(\theta|\hat{\theta}^{(k-1)}) = \sum_{t=d+1}^{n} \sum_{i=1}^{d} Q_{t,i}(\theta|\hat{\theta}^{(k-1)}),$$

where

$$Q_{t,i}(\theta|\hat{\theta}^{(k-1)}) = \mathbb{E}\left\{ \log p(y_t, y_{t-i}, U_t, U_{t-i}; \theta) | y_t, y_{t-i}; \hat{\theta}^{(k-1)} \right\}$$

Theoretical motivation: the algorithm guarantees the ascent-property like for the ‘standard’ EM algorithm.

Closed-form expressions for most of the model parameters in the maximization step.
Weekly counts of influenza infection (top panel) meningococcal disease (bottom panel) cases in Germany for the period 2001-2006. Source: surveillance R package (Höhle et al., 2015)
Flu and Meningitis (cont’d)

Model specification

- $Y_t$ meningococcal disease counts
- $Y_t|U_t \sim \text{Poisson}(e^{\eta_t+U_t})$
- linear predictor $\eta_t$ accounts for seasonality and association with flu,

\[ \eta_t = \beta_0 + \beta_1 \cos \left(2\pi \frac{t}{52}\right) + \beta_2 \sin \left(2\pi \frac{t}{52}\right) + \beta_3 \log(\text{flu}_t + 1) \]

transformation $\log(x + 1)$ for reducing right skewness of flu and handle zeros.

- $U_t$ assumed to be AR(1),

\[ U_t = \phi U_{t-1} + \sigma \epsilon_t, \quad |\phi| > 1 \quad \epsilon_t \overset{iid}{\sim} N(0, 1) \]
Flu and Meningitis (cont’d)

Models fitted up to order ten

Best model in terms of efficiency is that of order seven

Comparison against integrated nested Laplace using the R-INLA package (Rue et al., 2009)

<table>
<thead>
<tr>
<th></th>
<th>Poisson GLM</th>
<th>Parameter-driven</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PL ($d = 7$)</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>2.10 (0.04)</td>
<td>2.12 (0.04)</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.14 (0.03)</td>
<td>0.16 (0.02)</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.24 (0.04)</td>
<td>0.27 (0.02)</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>0.06 (0.02)</td>
<td>0.05 (0.02)</td>
</tr>
<tr>
<td>$\hat{\sigma}^2$</td>
<td>-</td>
<td>0.03 (0.01)</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>-</td>
<td>0.70 (0.09)</td>
</tr>
</tbody>
</table>

Estimates with standard errors in parentheses.
Conclusions

First step of a much wider project that will consider:

- joint modelling of *multiple diseases*
- identification of *intervention effects* (health policies, vaccination campaigns, etc.)
- spatio-temporal extensions
- combination with *sparsity methods* to handle higher-dimensions
Thanks for listening!