

PERFORMANCE OF INLA ANALYZING BIVARIATE META-REGRESSION AND AGE-PERIOD-COHORT MODELS

Andrea Riebler & Leonhard Held

Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich,
Switzerland

INLA is a new promising technology to perform Bayesian inference without Markov chain Monte Carlo (MCMC) [1]. In this talk I would like to assess the performance of INLA for two different applications.

The first application deals with bivariate meta-analysis [2] for diagnostic studies reporting both sensitivity and specificity. Preserving the bivariate structure of the data, pairs of sensitivity and specificity are jointly analyzed. Mainly maximum likelihood (ML) approaches are thereby used for statistical inference. However, numerical problems with possible non-convergence are common. INLA provides a promising alternative. In comparison to the ML procedure SAS PROC NLMIXED it is faster and more stable while bias and mean-squared error show only small differences [3].

The second application considers age-period-cohort (APC) [4] models with smoothing priors based on second differences, which results in an improper posterior distribution. The reason here is not only the usually encountered problem of unidentifiable intercepts, which can be easily addressed by imposing sum-to-zero constraints, but also the linear dependence between age, period and cohort. To estimate identifiable posterior quantities of interest INLA is used. The results are compared with those obtained by MCMC.

References

- [1] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B*, to appear.
- [2] Reitsma, J. B., Glas A. S., Rutjes A. W. S., Scholten R. J. P. M., Bossuyt P. M., Zwinderman A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58:982-990.
- [3] Paul, M., Riebler, A., Bachmann, L., Rue, H. and Held, L. (2009). Bivariate meta-analysis with INLA: an approximate Bayesian inference approach. *Statistics in Medicine*, under review.
- [4] Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates, *Biometrics*, 39: 311324.

SPATIO-TEMPORAL DISEASE MAPPING USING INLA

Birgit Schrödle

Biostatistics Unit, Institute of Social and Preventive Medicine, University of Zurich,
Switzerland

In the presented application spatio-temporal models are used to identify unusual spatial and temporal trends within case reporting data from the Swiss federal veterinary office. Spatio-temporal models formulated in a hierarchical Bayesian framework are a very complex class of models though and parameter estimation using MCMC requires specific techniques [3]. Therefore, the INLA approach was used to obtain posterior marginals of the spatial and temporal effects and specific space-time interactions [2].

The applied spatio-temporal models will be presented briefly [1]. Furthermore, it will be shown how they are fitted using INLA and how the results can be used for data analysis. A specific requirement is that user-defined structure matrices and constraints have to be passed to the `inla` program. As a second issue the usability of tools for model comparison and criticism provided by INLA will be discussed. Finally, some aspects of practical experience when fitting spatio-temporal models using INLA will be pointed out.

References

- [1] Knorr-Held L: Bayesian modelling of inseparable space-time variation in disease risk (2000), *Statistics in Medicine* **19**, pp. 2555-2567
- [2] Rue H, Martino S, and Chopin N: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (2009), *Journal of the Royal Statistical Society: Series B*, to appear
- [3] Schmid V, Held L: Bayesian extrapolation of space-time trends in cancer registry data (2004), *Biometrics* **60**, pp. 1034-1042

APPROXIMATE BAYESIAN INFERENCE FOR SMALL AREA ESTIMATION

Virgilio Gomez-Rubio

Departamento de Matematicas, Escuela de Ingenieros Industriales - Albacete Universidad de Castilla-La Mancha, Spain
`Virgilio.Gomez@uclm.es`

Bayesian Hierarchical models have been widely used in several problems related to Small Area Estimation. Recently, Rue et al (2009) have developed methods to approximate the marginal distributions of the parameters of interest in latent Gaussian models. These approximate methods are based on the Integrated Nested Laplace Approximation (INLA). They have also developed the necessary software to fit a large set of models, including GLMs, mixed-effects models and many more.

In my talk I will discuss the opportunities and limitations of INLA in the context of Small Area Estimation. Other applications in Small Area Estimation include space-time models, ranking of areas, poverty mapping, dealing with missing observations, benchmarking and others. I will give an overview of how these problems can be tackled using the approximate marginals of the parameters of interest.

I will illustrate the applications of these methods using simulated data based on real data from Sweden.

REFERENCES:

H. Rue, S. Martino and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximation. *Journal of the Royal Statistical Society, Series B*. To appear.

Disease mapping: a tool for epidemiologists

Léa Fortunato, Linda Beale, Guangquan Li

An application for disease mapping and risk analysis, the Rapid Inquiry Facility (RIF) has been developed by the Small Area Health Statistics Unit (SAHSU) at Imperial College London, in collaboration with the US CDC Environmental Public Health Tracking Program and the EUROHEIS project. Disease mapping is a valuable method for exploring spatial patterns of health outcomes, however, when dealing with low population numbers and/or rare diseases such methods can be problematic. Both rates and SMRs become numerically unstable and typically the less populated areas will show the highest risks. One way of addressing these problems is using Bayesian smoothing methods.

The RIF allows users to carry out Empirical Bayes smoothing and to automatically link to WinBUGS or INLA to run a number of models (heterogeneity, CAR, BYM). A default protocol for disease mapping has been developed, including a sensitivity analysis of the prior distributions of the variances. Different parameters, such as the spatial fraction, will be calculated by default to have quantitative summaries of the spatial variability, as well as maps.

APPROXIMATE BAYESIAN INFERENCE FOR SURVIVAL MODELS

Rupali Akerkar

Norwegian University of Science and Technology (NTNU), Trondheim, Norway

`Rupali.Akerkar@math.ntnu.no`

In talk will discuss the use of INLA to solve Bayesian inferential problems for survival models. Examples include Weibull distributed lifetime models with frailty, and piecewise-constant baseline hazard models. I will demonstrate the results obtained with INLA and those obtained using WINBUS. The power of this approach is illustrated by reanalyzing the spatial survival model by Henderson et al (2002, JASA), which includes both semi-parametric effects of covariates, frailty and spatial effects.

ANIMAL MODEL AND INLA

Ingelin Steinsland^{1*}, Anna Marie Holand ^{1**} and Henrik Jensen¹

¹ Norwegian University of Science and Technology (NTNU), Trondheim, Norway

* `ingelins@math.ntnu.no`

** `annamaho@ntnu.no`

Quantitative genetics is the study of quantitative characters. Examples of quantitative characters are the height of a human, the length of a bird's bill or the amount of milk a cow produces. The theory is based on the assumption that the characters are determined by a large number of genes as well as by different environmental effects. An important quantitative genetic parameter is the *heritability* of a trait; how much of the phenotypic (observed trait) variance in a population that can be explained by additive genetic effects.

The key assumption in the model we use, *the animal model*, is that animal i 's trait, y_i (e.g. the length of a bird's bill) can be divided into a genetic part (or inherited part), u_i , and an individual part ϵ_i . The value of u_i is referred to as the breeding value of individual i . To calculate individual breeding values one needs data from related animals and information on how they are related, i.e. their pedigree / family tree.

In this presentation we demonstrate that the animal model is a latent Gaussian Markov Random field model, and how INLA can be used to make inference for many interesting features connected to both hyper-parameters and the latent variables.

Outline

- Approximate inference for animal models for Gaussian traits.
 - Study for Norwegian house sparrows.
 - Include sex-specific inheritance structure in a study for Swiss barn owls.
- Approximate inference for animal model for non-Gaussian traits.
 - Study of Norwegian House sparrows.
- What next, opportunities and (possible) limitations.

ELIMINATING THE PRACTICAL BOUNDARY BETWEEN MARKOV AND OTHER GAUSSIAN RANDOM FIELDS

Finn Lindgren

Centre for Mathematical Sciences, Lund, Sweden

`finn@maths.lth.se`

Gaussian random field models are used extensively in spatial and spatio-temporal statistics. Traditionally, two largely separate approaches have been used; covariance function specifications and grid-based Markov random fields. The former method is appealing in its directness, but computationally costly, whereas the latter is appealing for its computational benefits. The two approaches have coexisted without much direct links between the specifications.

In this talk I will explain how to construct a direct specification of Markov random fields approximating the Matrn family of covariance models, through the use of stochastic partial differential equations. As a simple side-effect, this model class can be expanded to fields on curved surfaces, such as a globe, as well as anisotropic, oscillating, and non-stationary fields, as illustrated with geo-statistical data. The approach also provides a link to other popular random field models, such as convolution fields and spectral representations.

BAYESX AND INLA - OPPONENTS OR PARTNERS?

Thomas Kneib

Institute of Statistics and Econometrics Georg-August-University Gottingen, Germany
Department of Statistics, Ludwig-Maximilians-University Munich, Germany
`Thomas.Kneib@stat.uni-muenchen.de`

Bayesian hierarchical models provide a rich framework for modelling complex data structures. In many cases, such hierarchical models involve conditionally Gaussian effects on the second level of the hierarchy. For example, in case of structured additive regression, the first level is constituted by the observation model (i.e. the likelihood) for the observed responses while semiparametric effects that form the predictor of the regression model are formulated in terms of conditionally Gaussian priors. The parameters of the Gaussian distribution are then typically assigned further hyperpriors that can induce different types of regularisation, such as spatial smoothness or sparseness of regression models. For a long time, Markov chain Monte Carlo simulations have been the most prominent and dominating computing scheme to derive Bayesian estimates in hierarchical models. Only recently, the interest in approximate solutions began to rise again, for example based on simple or integrated nested Laplace approximations (INLA). In this presentation, we will review both simple Laplace approximations and Markov chain Monte Carlo simulation approaches available in the Bayesian software package BayesX (<http://www.stat.uni-muenchen.de/bayesx>). For a number of examples, we will compare results obtained with BayesX with those obtained with the INLA methodology available in the corresponding R package. While we find close agreement in several situations, we also find remarkable differences in some situations that are worth further investigation and might lead to an improvement of approximate solutions. We will also sketch some extensions that can easily be treated with Markov chain Monte Carlo simulations but might be challenging when using the INLA methodology.