An analysis of Japanese liver cancer mortality data with Bayesian age–period–cohort models

Wataru Sakamoto *

Abstract — **Age–period–cohort (APC) models have been widely used in the analysis of incidence and mortality data. Bayesian APC models, in which multivariate Gaussian priors are incorporated on age, period and cohort effects, can evade the identifiability problem. Inference with integrated nested Laplace approximations (INLA) has recently been a useful tool. An application of the Bayesian APC models with INLA to Japanese liver cancer mortality data is illustrated, in which a sudden change of the cohort effect was revealed.**

Keyword: *information criteria; integrated nested Laplace approximation; Gaussian Markov random field*

1 Introduction

Cancer has been the leading cause of death for Japanese people. To analyze cancer incidence and mortality data and to predict mortality rate in the future should be essential to establishing an evidence-based target in cancer control strategy.

Age–period–cohort (APC) models have been widely used in epidemiology, such as in the analysis of incidence and mortality data. However, classical APC models have the identifiability problem (e.g. Kupper *et al.*, 1985), because of the identity: *period* = *cohort* + *age*, and some approaches to overcome the identifiability problem have been proposed. One way is to consider linear constraints among parameters or parsimonious parametrization (e.g. Carstensen, 2007). R packages *Epi* (Hills *et al.*, 2014) and *apc* (Nielsen, 2015) are available for the APC analysis.

Another approach to evade the identifiability problem is to put regularization constraints to the variation in each component of age, period and cohort. This corresponds to a Bayesian approach in which multivariate Gaussian priors are incorporated on age, period and cohort effects. Nakamura (1986) considered a Bayesian APC model, in which successive parameters on age, period and cohort effects were assumed to change gradually, and smoothing parameters were estimated by an empirical Bayes method. Bayesian hierarchical approaches to the APC model have also been studied; see Schmid and Held (2007) (and references therein), in which Markov chain Monte Carlo (MCMC) algorithm was used for inference. Riebler *et al.* (2012) used integrated nested Laplace approximations (INLA) (Rue *et al.*, 2009) for inference in Bayesian APC models, which enable fast computation and have equivalent performance with MCMC.

In this paper we apply Bayesian APC models to Japanese liver cancer mortality data, and illustrate usefulness of the Bayesian APC models and inference with INLA.

2 Bayesian age–period–cohort models

Let y_{ijk} be the number of deaths or incidences and n_{ijk} be the populations, observed for the i-th age group ($i = 1, \ldots, I$), j-th period group ($j = 1, \ldots, J$) and k-th cohort group ($k = 1, \ldots, K$). The following identities hold: $\text{Age}_i + \text{Cohort}_k = \text{Period}_j$. Consider a Poisson model

$$y_{ijk} \sim \text{Po}(n_{ijk}\lambda_{ijk}), \quad i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, K, \tag{2.1}$$

where λ_{ijk} is the unknown relative risk. The APC model has the following form on the linear predictor η_{ijk}:

$$\log(\lambda_{ijk}) \equiv \eta_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k, \tag{2.2}$$

*Faculty of Environmental Science and Technology, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan, E-mail: w-sakamoto@okayama-u.ac.jp

where μ, α_j, β_j and γ_k are unknown constants representing mean, age, group and cohort effects, respectively.

In Bayesian analysis of the APC model (2.2), we incorporate priors such as follows on each of the age, period and cohort effects: for example, on age effect $\{\alpha_i\}$,

- Random walk prior of first order (RW1): $p(\alpha_i) \propto 1$ marginally, and
$$\alpha_i - \alpha_{i-1} \sim \mathrm{N}(0, \theta_\alpha^{-1}), \quad \text{i.e.,} \quad \alpha_i|\alpha_{i-1} \sim \mathrm{N}(\alpha_{i-1}, \theta_\alpha^{-1}), \quad i = 2, \ldots, I,$$
which assumes a constant trend over time. The precision parameter θ_α controls variation.
- Random walk prior of first order (RW2): $p(\alpha_i) \propto 1$ marginally, and
$$\alpha_i - 2\alpha_{i-1} + \alpha_{i-2} \sim \mathrm{N}(0, \theta_\alpha^{-1}), \quad \text{i.e.,} \quad \alpha_i|\alpha_{i-1}, \alpha_{i-2} \sim \mathrm{N}(2\alpha_{i-1} - \alpha_{i-2}, \theta_\alpha^{-1}), \quad i = 3, \ldots, I,$$
which assumes a linear trend over time. The precision parameter θ_α controls smoothness.

These priors play a role of solving the identifiability problem (Schmid and Held, 2007). Priors on period and cohort effects $\{\beta_j\}$ and $\{\gamma_k\}$ are incorporated similarly, with precision parameters θ_β and θ_γ, respectively. Moreover, hyper-priors such as gamma distributions are introduced to the precision parameters. For the constant term μ in (2.2), we usually consider a non-informative flat prior.

3 Bayesian inference

3.1 Latent Gaussian models

The Bayesian APC model described in Section 2 belongs to the class of latent Gaussian models (Rue *et al.*, 2009), which is characterized as follows:

- $\boldsymbol{y} = \{y_{ijk}\}$ is a vector of observations, each of which follows independently the distribution $p(y_{ijk}|\eta_{ijk}, \boldsymbol{\theta})$ for a given latent variable, η_{ijk} in the APC model, as described next.
- $\boldsymbol{z} = \{\eta_{ijk}, \mu, \alpha_i, \beta_j, \gamma_k\}$ is a high-dimensional vector of latent variables, of which the prior forms an intrinsic Gaussian Markov random field (GMRF) (Rue and Held, 2005), which follows the multivariate normal distribution with mean vector $\mathbf{0}$ and sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$. Let $p(\boldsymbol{z}|\boldsymbol{\theta})$ be its density.
- $\boldsymbol{\theta} = (\theta_\alpha, \theta_\beta, \theta_\gamma)$ is a hyper-parameter vector. Let $\pi(\boldsymbol{\theta})$ be a hyper-prior on $\boldsymbol{\theta}$, which is usually non-normal.

The joint posterior density of $(\boldsymbol{z}, \boldsymbol{\theta})$ becomes, via the Bayes' theorem,
$$\pi(\boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \prod_{i,j,k} p(y_{ijk}|\eta_{ijk}, \boldsymbol{\theta}) \cdot p(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

3.2 Integrated nested Laplace approximation

The latent Gaussian model described above is regarded as a hierarchical Bayes model, and we are mainly interested in predicting each latent variable, say z_i, that is, obtaining its marginal posterior distribution:
$$p(z_i|\boldsymbol{y}) = \iint \pi(\boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{z}_{-i}\, d\boldsymbol{\theta}. \tag{3.1}$$
The marginalization in (3.1) requires high-dimensional integration. The integrated nested Laplace approximation (INLA), proposed by Rue *et al.* (2009), enables fast computation without posterior sampling, using the Laplace approximation of the posterior distribution of the sparse latent variables, which is normally approximated, and using numerical integration on the distribution of hyper-parameters, which is non-normal. It has been shown that the INLA has comparable estimation performance to the Markov chain Monte Carlo method (Rue *et al.*, 2009; Riebler *et al.*, 2012).

An R package `INLA` (`http://www.r-inla.org/`) (Martino and Rue, 2010) for Bayesian analysis with the latent Gaussian models and the INLA is available, and we used it through our study.

4 Analysis of Japanese liver cancer mortality data

Japanese cancer mortality data are based on Vital Statistics Japan (Ministry of Health, Labour and Welfare), and can be available from the website of Cancer Registry and Statistics, Cancer Information Service, National Cancer Center Japan: `http://ganjoho.jp/reg_stat/statistics/dl/index.html`. The populations and the numbers of liver cancer death, separated by gender, are recorded every year from 1958 to 2014 for each of 5-year age categories. The cohort category has 5-year length and is shifted year-by-year. Analyses of the Japan liver cancer data were also conducted by Kamo *et al.* (2011), Tonda *et al.* (2015) and so on.

We considered the Bayesian APC models described in Section 2 and conducted Bayesian inference with INLA as in Section 3. The random walk priors of order 1 or 2 (RW1 or RW2) were incorporated for each of age, period and cohort effects. The deviance information criterion (Spiegelhalter et al., 2002; Gelman *et al.*, 2014)

$$\mathrm{DIC} = -2\log p(\boldsymbol{y}|\hat{\boldsymbol{z}}_B, \hat{\boldsymbol{\theta}}_B) + 2d_{\mathrm{DIC}}, \tag{4.1}$$

was computed for each combination of the priors, where $\hat{\boldsymbol{z}}_B$ and $\hat{\boldsymbol{\theta}}_B$ are Bayes estimates of $\boldsymbol{z}$ and $\boldsymbol{\theta}$, respectively, and d_{DIC} is the effective number of parameters. Bayesian inference with INLA and computation of the DIC value were conducted by the R function `inla` of the package `INLA`.

The DIC values for some combinations of the priors are listed in Table 1. For the male data set, the model with RW1 priors on age, period and cohort effects gave the smallest DIC value, while for the female data set, the model with RW2 priors on age and period effects and RW1 priors on cohort effect gave the smallest DIC value, The model with RW2 priors on age, period and cohort effects fitted poorly for both male and female data sets, so DIC values are not listed.

Figure 1 shows posterior means and 95% credible intervals of age, period and cohort effects in the model giving smallest DIC, fitted to male and female data. The age effect was non-monotonic; the risk was higher in infancy, getting lower until 15 years old, and getting higher with age. On the cohort effect, the risk was changed suddenly and became highest around the birth year of 1930. The sudden change would explain the reason why the model with RW2 prior for the cohort effect was poor, because RW2 was too smooth. The period effect was smaller than the age and cohort effects.

It would be possible to provide prediction in future periods by extrapolation using the fitted Bayesian APC models.

5 Concluding remarks

We illustrated an application of the Bayesian age–period–cohort models to Japanese liver cancer mortality data. A sudden change of the cohort effect was suggested by selecting a model with RW1 prior. In our future works, we need to compare the performance of INLA with some other computation methods MCMC and the empirical Bayes method with simple Laplace approximation in the Bayesian APC models. Also, we should apply the described method to incidence and mortality data other than liver cancer. Moreover, we intend to consider a Bayesian APC model with region effects using cancer data separated by prefecture.

Table 1: DIC values for some combinations of the priors

$\{\alpha_i\}$	$\{\beta_j\}$	$\{\gamma_k\}$	Male	Female
RW1	RW1	RW1	12050.15	8955.228
RW2	RW1	RW1	12050.56	8957.157
RW2	RW2	RW1	12052.43	8938.019
RW2	RW1	RW2	12054.91	8953.035

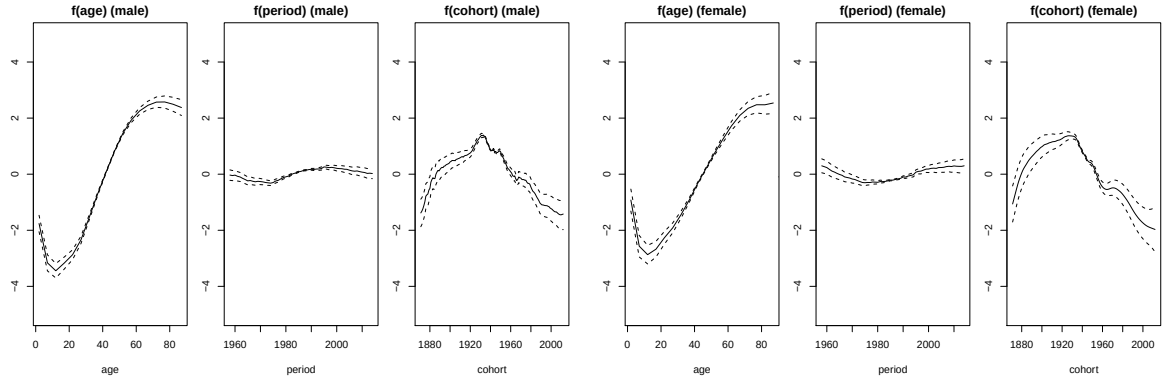

Figure 1: Posterior means and 95% credible intervals of age, period and cohort effects.

Acknowledgment

This research was partly supported by JSPS KAKENHI Grant No. 26330042.

References

[1] Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, **26**, 3018–3045.

[2] Gelman, A., Hwang, J. and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, **24**, 997–1016.

[3] Hills, M., Plummer, M. and Carstensen, B. (2014). A short introduction to R for epidemiology (Version 4). `http://bendixcarstensen.com/Epi/R-intro.pdf`

[4] Kamo, K., Tonda, T. and Satoh, K. (2011). Cancer mortality risk visualization on age-period plane. *Proceedings of the Institute of Statistical Mathematics*, **59**, 217–237 (in Japanese).

[5] Kupper, L. L., Janis, J. M., Karmous, A. and Greenberg, B. G. (1985). Statistical age–period–cohort analysis: a review and critique. *Journal of Chronic Diseases*, **38**, 831–836.

[6] Martino, S. and Rue, H. (2010). Implementing approximate Bayesian inference using integrated nested Laplace approximation: a manual for the inla program. Department of Mathematical Sciences, NTNU, Norway.

[7] Nakamura, T. (1986). Bayesian cohort models for general cohort table analyses. *Annals ofthe Institute of Statistical Mathematics*. **38**, 353–370.

[8] Nielsen, B. (2015). apc: An R package for age–period–cohort analysis. *The R Journal*, **7**, 52–64.

[9] Riebler, A., Held, L. and Rue, H. (2012). Estimation and extrapolation of time trends in registry data-borrowing strength from related population. *The Annals of Applied Statistics*, **6**, 304–333.

[10] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications.* Chapman and Hall.

[11] 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*, **71**, 319–392.

[12] Schmid, V. J. and Held, L. (2007). Bayesian age–period–cohort modeling and prediction - BAMP. *Journal of Statistical Software*, **21**, 1–15.

[13] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

[14] Tonda, T., Satoh, K. and Kamo, K. (2015). Detecting a local cohort effect for cancer mortality data using a varying coefficient model. *Journal of Epidemiology*, **25**, 639–646.