Frailty Factors and Time-dependent Hazards in Modelling Ear Infections in Children Using BASSIST

Mervi Eerola[1], Heikki Mannila[2] and Marko Salmenkivi[3]

[1] Rolf Nevanlinna Institute, University of Helsinki, P.O. Box 4, FIN-00014 University of Helsinki, Finland,
Mervi.Eerola@rni.helsinki.fi
[2] Department of Computer Science, University of Helsinki, P.O. Box 26, FIN-00014 University of Helsinki, Finland,
Heikki.Mannila@cs.helsinki.fi
[3] Department of Computer Science, University of Helsinki, P.O. Box 26, FIN-00014 University of Helsinki, Finland,
Marko.Salmenkivi@cs.helsinki.fi

1 Introduction

The BASSIST system is a general purpose tool for MCMC sampling for intensity models. The system allows the user to specify an intensity model in a high-level language. The model is used to generate a simulation program that uses the Metropolis-Hastings algorithm to obtain the desired samples. In contrast to BUGS (Spiegelhalter *et al.*, 1996), BASSIST contains several primitives that are suited for modelling event data, including piecewise constant functions etc.

In this paper we describe the use of the BASSIST system when modelling the occurrences of middle ear infection (acute otitis media, AOM). Previous modelling on AOM by nonparametric Bayesian methods has been done in Andreev & Arjas (1998).

Typically, the infections occur several times to some particularly infection prone children. This recurrence causes dependence among the observations which can partly be modelled by observed information on the individuals but the risk is also related to the immunological development of a child. This process can of course not be fully observed. Moreover, there are anatomic and unobserved genetic differences which are presumably important risk determinants. Common to all these factors is that they are time and age dependent, perhaps in a very complex way. This suggests that there are several sources of unobserved heterogeneity which should be accounted for in the modelling.

In this paper we will show some basic models to illustrate the use of BASSIST when modelling the common baseline intensity nonparametrically and then adding more structure by observed time-dependent risk factors and individual specific unobserved random variables, commonly called frailties in survival analysis. The analysis is fully Bayesian and uses the flexible ideas of variable dimension MCMC methods suggested by Green (1995).

2 The data

In the pilot phase of a large vaccination trial on ear infections caused by Streptococcus pneumoniae, 329 children in the Tampere area in Finland were followed from the age of 2 to 24 months. The aim was to study the natural course of the disease in small children without vaccination.

In addition to the random occurrences of acute otitis media episodes (AOM's), changes in several risk factors were recorded repeatedly in health visits at the ages of 2, 3, 4, 5, 6, 9, 12, 15, 18, and 24 months. For illustration purposes, we will only consider here the effects of two well known risk factors: weaning and day care attendance.

3 Steps in the modelling

3.1 Homogeneous model

We started by modelling the age-dependent intensity $\lambda(t)$ by a piecewise constant function $f(t) = \sum_{j=1}^{N} 1_{\{j<t\leq j+1\}}\lambda_j$. The levels of the function, the jump times and the number of pieces were random variables with the following priors (hyper constants could easily be replaced by an additional hierarchy level):

- number of pieces $N \sim$ Poisson(10)
- levels of functions $\lambda_j \sim$ Uniform(0.0000001, 0.15)
- jump times $l_j \sim$ Uniform(1.5,24.9)

3.2 Observed history

We proceeded by adding observed history in terms of time-dependent indicator functions of weaning and day care multiplicatively into the model. In order to distinguish between risk factors for first and recurrent infections, we used the parameterisation in Eerola (1989) to model the infection intensity for individual i as

$$\lambda_i(t) = f(t)Y_i(t)\beta_i(t)\gamma_i(t)\alpha_i(t)\beta_i^*(t)\gamma_i^*(t)$$

where $Y_i(t)$ is an indicator function of being at risk at time t.

The time-dependent functions of weaning at T_W and daycare entrance at T_D, $\beta_i(t) = 1_{\{T_{W,i}\geq t\}}\beta$ and $\gamma_i(t) = 1_{\{T_{D,i}\leq t\}}\gamma$, indicate the relative risk of these factors compared to the baseline $f(t)$ in that the values $\beta, \gamma < 1$ correspond to protective effect and the values > 1 increased risk of AOM. More commonly, $\beta = exp(\tilde{\beta})$, as in the Cox model, but since we have shown the posterior results for β we write the intensity model in terms of it.

The indicator function $\alpha_i(t) = 1_{\{T_{1,i}+30\leq t\}}\alpha$ is used to distinguish between those who have had at least one infection already. The parameters for recurrent episodes are estimated among these individuals.

The functions $\beta_i^*(t) = 1_{\{T_{1,i}+30\leq t\}}1_{\{T_{W,i}\geq t\}}\beta^*$ (for $\gamma_i^*(t)$ respectively) are interactions between the covariates and the onset at $T_{1,i}$ where $T_{1,i}$ is the time of first infection for i. They correspond to the additional effect of the covariates on recurrent infections compared to the effect on the first infection. The total relative risk for recurrences is therefore $\beta \cdot \beta^*$. The lag time of 30 days is a convention to define distinct ear infection episodes.

In this model the baseline intensity for the first and subsequent episodes is the same, which seems to be a plausible assumption for recurrent events of the same nature. The prior distributions used for $\alpha, \beta, \beta^*, \gamma$ and γ^* were Gamma(2,2) in which case the prior mean of the relative risk is 1.

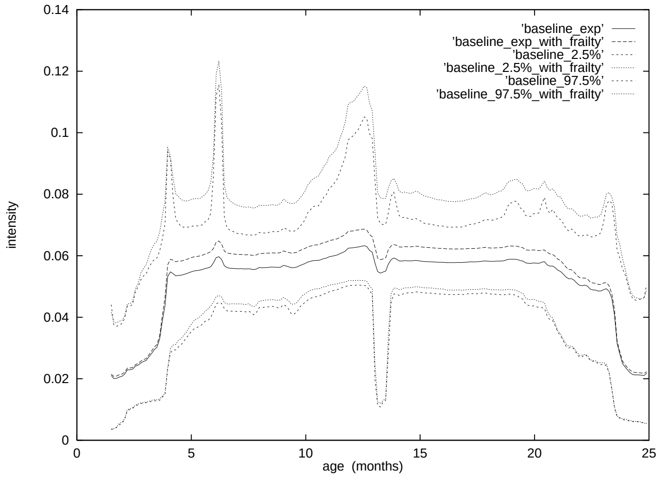

Fig. 1. Posterior baseline intensity curves $f(t)$ for the two models with no observed history: crude hazard only and unobserved frailties added (mean and 95 %-credibility-intervals)

3.3 Unobserved frailties

Finally, we added an individual constant frailty factor to account for the additional dependence due to possible genetic and anatomic variations among the children. The effect of the frailty is assumed to be multiplicative. For the case of no observed history, we get the model

$$\lambda_i(t) = f(t)\zeta_i,$$

where ζ_i is an unobserved frailty factor. When taking the observed history into account we get a model of the form

$$\lambda_i(t) = f(t)\zeta_i Y_i(t)\beta_i(t)\gamma_i(t)\alpha_i(t)\beta_i^*(t)\gamma_i^*(t).$$

The frailties are assumed to be a random sample from a common distribution Gamma(ν, η). We also make the usual assumption that the prior mean $E(\zeta|\nu,\eta) = 1$ which in the Gamma distribution means that $\nu = \eta$. It is of interest to let the hyperparameter ν also be random which allows us to consider the variability in the data in the light of the accumulating observations. The prior of ν was Uniform(0.00001,100).

The joint posterior distribution of the parameters $\theta = (\lambda, k, l, \alpha, \beta, \gamma, \zeta, \nu)$ is then (by assuming independence of the parameters)

$$p(\lambda, k, l, \alpha, \beta, \gamma, \zeta, \nu|data) = p(\lambda)p(k)p(l)p(\alpha)p(\beta)p(\gamma)p(\zeta|\nu)p(\nu)L(data|\theta)$$

where the Poisson likelihood for the data is

$$L(data|\theta) = \prod_{i=1}^{329}\prod_{k=1}^{n_i} \lambda_i(T_{i,k})exp(-\int_0^\infty \lambda_i(s)ds).$$

4 The BASSIST system

Seminal work in software tools for Bayesian analysis was done by the BUGS group (Spiegelhalter *et al.*, 1996), who have created a very interesting tool for describing hierarchical models and performing MCMC simulations for obtaining estimates of posterior distributions. However, BUGS as such is not suitable for time-dependent data.

Thus we designed a new system, BASSIST, which contains several primitives suited for modelling event data, including piecewise constant functions etc. The system allows the user to specify an intensity model in a high-level language. For reasons of brevity, we omit most of the details in this paper; see Arjas *et al.* (1996) for a description of the first version of the program and Toivonen *et al.* (1998) for the new version.

Briefly, BASSIST is given a high-level description of a full probability model. The model is used to generate a simulation program that uses the Metropolis-Hastings algorithm to obtain the desired samples.

Below is a simple example of how the BASSIST description language is used to describe a homogenous model with unobserved frailties. We omit the details of the description language.

```
child(  var bdate;
        var sex;
        var frailty ~ dgamma(3,3);
 )


event_sequence infections(
    child;
    var time ~ Poisson process (child.frailty * f(time));
)

var piecewise constant function f(
        var start; var end;
        var pieces ~ dpoisson(10);
        piece(
           var start ~ dunif(1.5, 24.9);
           var level ~ dunif(0.0000001, 0.15); );
);
```

The BASSIST system was used to perform MCMC simulations of the models described above. Convergence was monitored by using the CODA package (Best *et al.*, 1995). Typically, the simulations took about 50 000 iterations (we picked up parameter values from every tenth sweep) before they passed the tests. For the case of the models with frailties and a hyperparameter for them, we needed about 500 000 iterations. On a 166 MHz Pentium, such a run took 15 hours.

5 Empirical results

The piecewise constant posterior baseline curves show that the risk of AOM's increases rapidly around the age of 5-6 months and another peak arises around the age of 12 months (Figure 1). We expect to be able to 'explain' this peak by adding information on day care because the maternal leave in Finland is 12 months long.

In Figure 1 baseline intensity is higher for the model with frailty factors. The reason is that the children with relatively many infections - and thus higher frailty factors - are absent from the risk set more frequently than those with few infections. Consequently, the average of frailty factors in the risk set is smaller than 1 during almost the whole time period and causes the baseline intensity to rise in proportion.

The recurrent infections tend to occur at the end of the follow-up which suggests that different factors might be responsible for their occurrence. Figure 2 indicates that day care loses significance for recurrent infections because the interaction parameter γ^* is clearly below 1. Since weaning was modelled as having the value 1 as long as breast feeding continues, the posterior mode of β^* around 2 indicates that the risk for recurrent AOM's increases considerably when weaning has occurred (Figure 2).

It is obvious that part of the proneness to ear infections can be captured by modelling the previous infections; when including the indicator of the first infection $1_{\{T_{1,i} \leq t\}}$ in the model the distribution of the frailty hyperparameter η is very skewed with a long tail (Figure 3). For the sake of comparison, Figure 4 shows the clearer differences between the frailty factors in the model with no observed history.

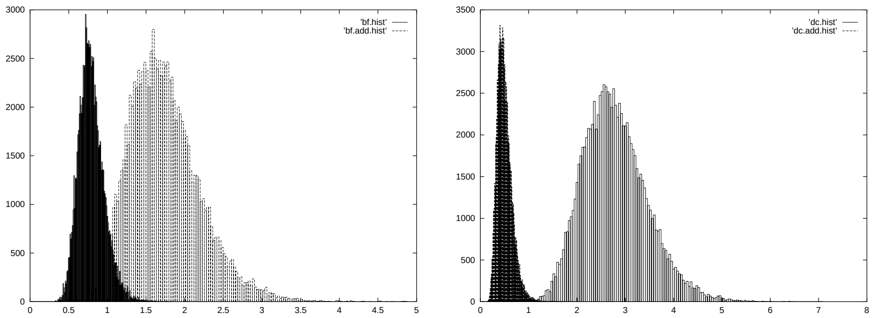

Fig. 2. The posterior distributions of the parameters β (dark), and β^* on the left side, γ and γ^* (dark) on the right side

Acknowledgements

We thank the Pnc study group from the National Public Health Institute (KTL), Finland, for letting us use the acute otitis media data.

References

Andreev, A. & Arjas, E. (1998). Acute Middle Ear Infection in Small Children: A Bayesian Analysis. In *Longitudinal Data Analysis (LIDA)*, in press. Boston: Kluwer Academic Publisher.

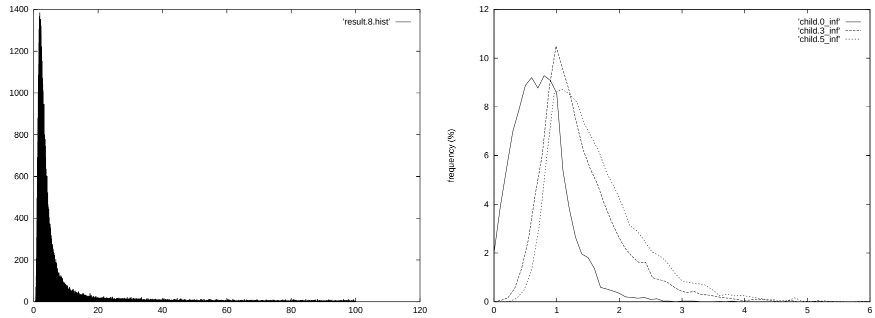

Fig. 3. The posterior distributions of the frailty hyperparameter η and the frailty factors ζ_i for children with 0,3 and 5 infections, respectively. The model with observed history

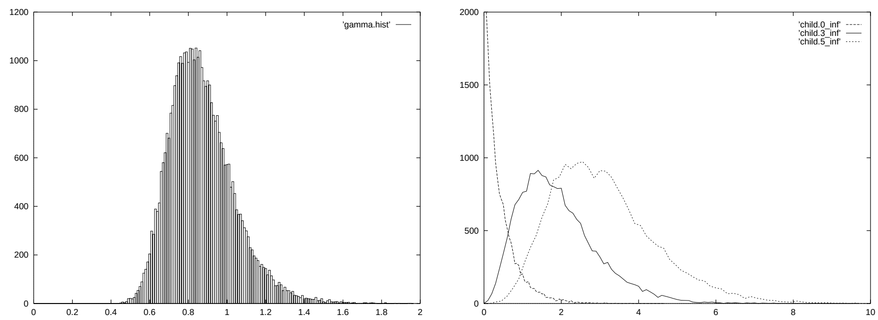

Fig. 4. The posterior distributions of the frailty hyperparameter η and the frailty factors ζ_i for children with 0,3 and 5 infections, respectively. The model with no observed history

Arjas, E., Mannila, H., Salmenkivi, M., Suramo, R. & Toivonen, H.(1996). BASS : Bayesian Analyzer of Event Sequences. In: *COMPSTAT'96 Proceedings in Computational Statistics* (ed. A. Prat), 199-204. Heidelberg: Physica-Verlag.

Best, N., Cowles, M. & Vines, K. (1995). *CODA. Convergence Diagnosis and Output Analysis Software for Gibbs sampling output. Version 0.30.* MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Eerola, M. (1989). Repeatable events in event-history analysis; the effect of childhood separation on future mental health. *Bull. Int. Statist. Inst.,* **LIII**(1), 213-225.

Green, P (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrica,* **82**, 4, 711-732.

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996). *BUGS, Bayesian inference Using Gibbs Sampling Manual, Version 0.50.* MRC Biostatistics Unit, Institute of Public Health, Cambridge.

Toivonen, H., Mannila, H., Salmenkivi, M. & Laakso, K. (1998). BASSIST - a tool for MCMC simulation of statistical models. In: *EUROSIM'98 Proceedings of 3rd International Congress of the Federation of EUROpean SIMulation Societies*, in press.