CHILD MORTALITY USING BAYESIAN SEMI-PARAMETRIC DISCRETE-TIME SURVIVAL MODEL

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

Abstract

The Bayesian Approach offers the viable and rigorous solution, though there is also the added benefit of providing much-needed uncertainty and probability assessments in non-linear and non-Gaussian situations in a valid and rigorous way. Mortality and its various determinants have been traditionally studied in a regression modeling framework. Initial studies mostly used the usual linear regression models which, however, are not appropriate in situations where the mortality information is given by a binary indicator of death or alive. Binary regression models (logit and probit) are, therefore, a logical alternatives. There are, however, problems, with logit and probit models, namely, that they do not take into consideration the information on the survival time. Hence, most studies now utilize the survival analysis techniques. Recently, Fahrmeir and co-researchers at the LMU Munich have proposed a Bayesian Geo-Additive modeling framework which encompasses most of the known regression models and improves upon their shortcomings. The proposed model is also called Bayesian semi-parametric structured regression model..

Key words: Bayesian Semi-Parametric Discrete-Time Survival Model, Bayesian Model Comparison, Markov Chain Monte Carlo (MCMC), Smoothness Priors.

1. Introduction

There has been much recent interest in Bayesian inference for generalized additive and related models. The increasing popularity of Bayesian methods for these and other model classes is mainly caused by the introduction of Markov Chain Monte Carlo (MCMC) simulation techniques which allow realistic modeling of complex problems. In the last two decades, great advances in technology and industry have produced various high throughput measurement instruments. Large scale measurements have been collected in different areas in industry and technology, in the form of data as curves (e.g., thermal diffusivity measurements), data array and images (e.g., DNA microarray experiments), and spatial observations of space-time systems (e.g., satellite images), etc. Efficient use of these great resources raises a host of new methodological and experimental design issues in statistics, such as: how to assess and calibrate the instruments, how to combine data from heterogeneous sources, how to do the error propagation in high dimensional and non-linear input-output systems, how to combine model uncertainty and prior information (e.g., type B uncertainty) with experimental results, how to design multi-stage experiments used in automated testing and other areas. The Bayesian Approach offers a systematic and flexible approach to these problems. By adopting an objective or non-informative prior, the Bayesian Approach produces estimates and uncertainty measures comparable to the classical approach.

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

2. Bayesian Semi-Parametric Discrete-Time Survival Model

Survival time is not recorded continuously but is only known to lie within a month or, generally, a time interval. Data of this kind are known as interval censored. Since many ties occur, these data cause problems when continuous-time models are used. Instead of continuous time one observes the discrete time T with values $t = 1, \cdots, k$, where $T = t$ denotes death in month or interval t, and k is the last observation interval. Suppose x_{it} is a vector of covariates up to month t, then the discrete hazard function is defined as:

$$\lambda(t|x_{it}) = P(T = t|T \geq t, x_{it}) \ldots\ldots. (1)$$

It is the conditional probability of death in month t given that the child has reached month t. The associated discrete survivor function is given by:

$$S(t|x_{it}) = P(T > t|x_{it}) = \prod_{i=1}^{k}\big(1 - \lambda(t|x_{it})\big) \ldots\ldots. (2)$$

Survival information on each child is recorded as $\left(t_i, \delta_i, x^*_{it_i}\right)$; i is the number of observed child, t_i is the observed lifetime in months, δ_i is the survival indicator with $\delta_i = 1$ if child i is dead and $\delta_i = 0$ if it is still alive. Thus for $\delta_i = 1$, t_i is the age of the child at death, and for $\delta_i = 0$, t_i is the current age of the child at interview. $x^*_{it_i} = (x_{it}, t = 1, \cdots, t_i)$ is the observed covariate sequence. Discrete time survival models can be cast into the framework of binary regression models by defining binary event indicators $y_{it}, t = 1, \cdots, T$ with

$$y_{it} = \begin{cases} 1: if\ t = t_i\ and\ \delta_i = 1 \\ 0: if\ t < t_i \end{cases} \ldots\ldots (3)$$

The hazard function for child i can then be written as a binary response model

$$P(y_{it} = 1|x_{it}) = h(\eta_{it}) \ldots\ldots\ldots\ldots (4)$$

Where x_{it} are the covariate processes for child i. h is an appropriate response or link function, and the predictor η_{it} is a function of the covariates. Common choices for such survival models are the grouped Cox model, logit or probit models. We shall consider here the probit model. The usual model is then

$$P(y_{it} = 1|\eta_{it}) = \phi(\eta_{it}) \ldots\ldots\ldots\ldots\ldots (5)$$

With partially linear predictor

$$\eta_{it} = f_0(t) + x'_{it}\gamma \ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The baseline hazard effect $f_0(t), t = 1,2, \cdots$ is an unknown, usually non-linear function of t to be estimated from the data. Treating the effects $f_0(t), t = 1,2, \cdots$ as separate parameters usually gives either very unstable estimates or may even lead to divergence of the estimation procedure. In a purely parametric framework the baseline hazard is therefore often modeled by a few dummy variables dividing the time-axis t into a number of relatively small segments or by some low order polynomial. In general it is difficult to correctly specify such parametric functional forms for the baseline effects in advance. Non-parametric modeling based on some qualitative smoothness restriction offers a more flexible solution to explore unknown dynamic patterns in $f_0(t)$. Then (6) can be regarded as the basic form of a semi-parametric predictor where the effects γ of covariates are fixed and time-constant.

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

In many applications, the restriction to constant covariate effects is not realistic. Rather, the effect of some covariates may vary over time. Therefore, the predictor is generalized to

$$\eta_{it} = f_0(t) + \sum_{j=1}^{p} f_j(t)x_{ij} + w_{it}'\gamma \quad \ldots\ldots\ldots(7)$$

Here w_{it}consists of covariates with time constant effect γ, the smooth functions $f_j(t)x_{ij}$ are non-linear effects of covariates x_j, $j = 1, \cdots, p$.

3. Prior Distribution

In Bayesian inference, the unknown functions f_j, $j = 1, \ldots, p$, the fixed effects parameters $\gamma = (\gamma_1, \ldots, \gamma_r)'$as well as the variance parameter τ^2 are considered as random variables and have to be supplemented by appropriate prior assumptions. In the absence of any prior knowledge we assume independent diffuse priors $\gamma_j \propto$ constant, $j = 1, \ldots, r$ for the parameters of fixed effects. Another common choice is highly dispersed Gaussian priors. Several alternatives are currently available for a smoothness prior of the unknown function $f_j, j = 1, \cdots, p$**.**Among others, these are random walk priors (Fahrmeir and Lang, 2001a), Bayesian smoothing splines (Hastie and Tibshirani, 2000) and Bayesian P-splines (Lang and Brezger, 2004), are the most known.

The General Form of the Priors

Suppose that $f = \big(f(1), \ldots, f(n)\big)'$ is the vector of corresponding function evaluations at observed values of X. We express the vector f as the matrix product of a (deterministic, non-random) design matrix X and a vector of unknown regression parameters β, i.e.

$$f = X'\beta$$

Then, the general form of the prior for β is

$$\beta|\tau^2 \propto exp\left(-\frac{1}{2\tau^2}\beta' K\beta\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

Where K, is a penalty matrix that penalizes too abrupt jumps between neighboring parameters. In most cases K will be rank deficient and therefore the prior for β improper. This implies that $\beta|\tau^2$ follows a partially improper Gaussian prior $\beta|\tau^2 \sim N(0, \tau^2 K^-)$ where K^- is a generalized inverse of the penalty matrix K. In the frequentist approach the smoothing parameter is equivalent to the variance parameter τ^2 which controls the tradeoff between flexibility and smoothness. In order to be able to estimate the "smoothing parameter" τ^2 simultaneously with β, a highly dispersed but proper hyper-prior is assigned to τ^2.The proper prior for τ^2 is required to obtain a proper posterior for β (Hobert and Casella, 1996).We choose an inverse gamma distribution with hyper-parameters a and b, i.e.

$$\tau^2 \sim IG(a, b)$$

With probability density function given by

$$p(\tau^2|a, b) = (\tau^2)^{-a-1} exp\left(\frac{b}{\tau^2}\right)$$

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

Common choices for $a\ and\ b$ are $a = 1\ and\ b = 0.005 (b = 0.0005)$.Alternatively, one may take $a = b = 0.001$. Brezger and Lang (2006) also suggest a general structure of the priors as

$$p(\beta_j|\tau_j^2) \propto \frac{1}{(\tau_j^2)^{rank(K_j)/2}} exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right)$$

Where K_j is a penalty matrix that shrinks parameters towards zero or penalizes too abrupt jumps between neighboring parameters. In most cases K_j will be rank deficient and therefore the prior for β_j is partially improper.

- **Priors for Fixed Effects**

As indicated above for the parameter vector γ of fixed effects we choose a diffuse prior

$$\gamma_j \propto \text{ Constant}, \qquad j = 1, \dots, r$$

Another choice would be to work with a multivariate Gaussian distribution $\gamma \sim N(\gamma_0, \Sigma_{\gamma 0})$.

- **Priors for Time Varying Effects under Linear Setup**

Priors for the unknown functions, $f_1, \dots, f_p$ depend on the type of the covariates and on prior beliefs about the smoothness of f_j. In the following we express the vector of function evaluations $f_j = \left(f_j(x_{ij}), \dots, f_j(x_{ij})\right)'$ of a function f_j as the matrix product of a design matrix X_j and a vector of unknown parameters $\beta_j s'$, i.e.

$$f_j = X_j' \beta_j$$

Then, we obtain the predictor which takes linear form as

$$\eta = X_1\beta_1 + \cdots + X_p\beta_p + W'\gamma$$

Where W corresponds to the usual design matrix for categorical covariates. For the variance parameter τ^2 , the inverse gamma with hyperparamters $a\ and\ b$ assumed as discussed above. On the other hand if the unknown functions $f_j, j = 1, \dots, p$ assumes a smooth nonlinear function, then a random walk priors or Bayesian P-spline can be considered as discussed in the next section.

- **First and Second Order Random Walk Priors**

Let us consider the case of a time scale or continuous covariate x with equally-spaced observations $x_i, i = 1, \dots, m$, $m \leq n$. then, $x_{(1)}, \dots, x_{(m)}$defines the ordered sequence of distinct covariate values. Here m denotes the number of different observations for x in the data set. A common approach in dynamic or state space models is to estimate one parameter $f(t)$ for each distinct $x(t)$; i.e., define, $f(t) = f\left(x_{(t)}\right)$ and let $f = \left(f(1), \dots, f(m)\right)'$ denote the vector of function evaluation. Then a first order random walk prior for f is defined by

$$f(t) = f(t-1) + u(t) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (9)$$

A second order random walk is given by

$$f(t) = 2f(t-1) - f(t-2) + u(t) \dots\dots\dots\dots\dots (10)$$

With Gaussian errors $u(t) \sim N(0, \tau^2)$ and diffuse priors $f(1) \propto$ const, or $f(1)$ and $f(2) \propto$ const, for initial values, respectively. Both specifications act as smoothness prior that penalize too rough functions f. A first-order random

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

walk penalizes too abrupt jumps $f(t)-f(t-1)$ between successive states and a second-order random walk penalizes deviations from the linear trend $2f(t-1)-f(t-2)$. In addition, the variance τ^2controls the degree of smoothness of f. Thus the conditional prior distribution of $f\ (t)$ given its immediate past $f(t-1)$ is given by:

$$f_t|f_{t-1},\tau^2 \sim N(f_{t-1},\tau^2) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (11)$$

Moreover, random walk priors may be equivalently defined in a more symmetric form by specifying the conditional distributions of function $f(t)$ given its left and right neighbors. That means, we can write the prior in $(9\ and\ 10)$ in general form as

$$f|\tau^2 \propto exp\left(-\frac{1}{2\tau^2}f'Kf\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (12)$$

The penalty matrix is of the form $K=D'D$ .where D is a first or second order difference matrix. For example, for a random walk of first order the penalty matrix is given by:

$$K=\begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & -1 & 2 & -1 & \\ & & & -1 & 1 & \end{pmatrix}$$

Here the design matrix K is the penalty matrix that penalizes too abrupt jumps between neighboring parameters. More often, K is not full rank and this implies that $f|\tau^2$ follows a partially improper Gaussian prior.

$$f|\tau^2 \sim N(0,\tau^2K^-)$$

Where K^- is a generalized inverse of the penalty matrix K

For the case of non-equally spaced observations, random walk or autoregressive priors have to be modified to account for non-equal distances $\delta_t=X(t)-X(t-1))$ between observations. Random walks of first order are now specified by:

$$f(t)=f(t-1)+u(t) \dots\dots\dots\dots\dots (13)$$

Where: $u(t)\sim N(0,\delta_t\tau^2)$

i. e. , by adjusting from τ^2 to $\delta_t\tau^2$

Similarly random walks of second order is given by

$$f(t)=\left(1+\frac{\delta_t}{\delta_t-1}\right)f(t-1)-\left(\frac{\delta_t}{\delta_t-1}\right)f(t-2)+u(t) \dots\dots\dots (14)$$

Where $u(t)\sim N(0,w_t\tau^2)$ and w_t is an appropriate weight. Several possibilities are conceivable for the weights. See (Fahrmeir and Lang, 2001) for a discussion.

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

- **Bayesian P-splines**

Any smoother depends heavily on the choice of smoothing parameter, and for P-spline in a mixed (fixed and time varying) framework. A closely related approach for time scale or continuous covariates is based on the P-splines approach introduced by Eilers and Marx (1996). The basic assumption of this approach is that the unknown function f_j can be approximated by a spline of degree l with equally spaced knots $x_{min} = \xi_0 < \xi_1 < \cdots < \xi_{r-1} < \xi_r = x_{max}$ within the domain of x_j .The domain from x_{min} to x_{max} can be divided into n' equal intervals by $n' + 1$ knots. Each interval will be covered by $l + 1$ B-splines of degree l.The total number of knots for construction of the B-splines will be $n' + 2l + 1$ .The number of B-splines in the regression is $n' + l$. It is well known that such a spline can be written in terms of a linear combination of $M = r + l$ B-splines basis functions B_j, i.e.

$$f_j(x_{ij}) = \sum_{p=1}^{M} \beta_j B_j\,(x) \ldots\ldots\ldots\ldots (15)$$

The basis functions B_j are defined locally in the sense that they are nonzero only on a domain spanned by $2 + l$ knots. The $n * M$ design matrix x_j for P-splines is more intricate than the case of random walk priors. Each row $\boldsymbol{i}$ contains the value of the B-spline basis functions evaluated at x_i hence $x_j(i,p) = B_{jp}(x_{ij})$. In accordance with the properties of B-splines, each row X has $M + 1$ non-zero values. As for the number of knots, Eilers and Marx (1996) recommended the number of inner knots to range between 20 and 40 and introduced a penalization of the differences between regression coefficients of adjacent B-spline basis functions in order to generate a smoothing effect.

4. Posterior Inference

When performing Bayesian inference, all inferential conclusions are based on the posterior of the model. In an empirical Bayes approach to structured additive regression, no hyper-prior are assigned to the hyper-parameters, i.e. the variances τ_j^2 are treated as fixed. In this case, the specific form of the posterior depends only on the parameterization of the regression terms in the model. Then, we use Markov Chain Monte Carlo (MCMC) simulations to draw samples from the posterior and statistical inference is done by means of Markov Chain Monte Carlo techniques in a full Bayesian setting. Now we restrict the presentation to models with predictor (7) **.**Full Bayesian inference is based on the entire posterior distribution. Let α be the vector of all unknown parameters, and then the posterior is given by

$$p(\alpha|y) \propto L(y, \beta_1, \tau_1, \beta_2, \tau_2, \cdots, \beta_p, \tau_p, \gamma) \prod_{j=1}^{p} p(\beta_j|\tau_j^2) p(\tau_j^2) \ldots (8)$$

$$\propto L(y, \beta_1, \tau_1, \beta_2, \tau_2, \cdots, \beta_p, \tau_p, \gamma) \prod_{j=1}^{p} \frac{1}{(\tau_j^2)^{rk(K_j)/2}} exp\left(-\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \prod_{j=1}^{p} (\tau_j^2)^{-a_j-1} exp\left(-\frac{b_j}{\tau_j^2}\right)$$

Bayesian inference via MCMC is based on updating full conditionals of single parameters or blocks of parameters, given the rest and the data. For Gaussian models, Gibbs sampling with so-called multi move steps can be applied. For non-Gaussian responses Gibbs sampling is no longer feasible and Metropolis Hastings (MH) algorithms based

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

on iteratively weighted least squares (IWLS) proposals are needed. More detail can be found in Rue (2001) or Fahrmeir and Lang (2001).

In a fully Bayesian approach, parameter estimates are generated by drawing random samples from the posterior (8) via MCMC simulation techniques. The variance parameters τ_j^2 can be estimated simultaneously with the regression coefficients β_j by assigning additional hyper-priors to them. The most common assumption is, that the τ_j^2 are independently inverse gamma distributed, i.e.$\tau_j^2 \sim IG(a_j, b_j)$, with hyper-parameters a_j and b_j specified a priori. A standard choice is to use $a_j = b_j = 0.001$. In some data situations (for small sample sizes), the estimated non-linear functions f_j may depend considerably on the particular choice of hyper-parameters. It is therefore good practice to estimate all models under consideration using a (small) number of different choices for a_j and b_j to assess the dependence of results on minor changes in the prior assumptions. The full conditionals for the variance parameters $\tau_j^2$are inverse gamma with parameters $a_j = a_j + 0.5rank(K_j)$and $b_j = b_j + 0.5\beta_j K \beta_j$, and updating can be done by simple Gibbs steps, drawing random numbers directly from the inverse gamma densities. Convergence of the Markov chains to their stationary distributions is assessed by inspecting sampling paths and autocorrelation functions of sampled parameters. In the majority of cases, however, the IWLS updating scheme has excellent mixing properties and convergence problems do not occur.

For Gaussian prior distributions, efficient proposal densities for exponential family and hazard regression can be derived based on iteratively weighted least squares (IWLS) proposals as introduced by Gamerman (1997) in the context of random effects models(Brezger and Lang, 2006 for exponential family regression and Hennerfeind, Brezger and Fahrmeir, 2006 for hazard regression). Since the density of the Gaussian prior for the regression coefficients is differentiable, the corresponding full conditionals can be approximated with a Taylor series expansion. The general idea of IWLS proposals is then to obtain a Gaussian proposal by matching the mode and the curvature of the full conditional based on the Taylor expansion. This proposal has two advantages: Firstly, it can be used with multivariate coefficient vectors to take correlations into account in the proposals and, secondly, it automatically adapts to the form of the full conditional thereby avoiding manual tuning of the proposal densities. Both sampling schemes gave rather coherent results; we rely hereon for updating the parameters in an MCMC sampler, we use a Metropolis Hastings (MH) algorithm based on iteratively weighted least squares (IWLS) proposals, implemented in BayesX.

5. Model Selection

A widely used statistic for comparing models in a Bayesian framework is the Deviance Information Criterion. The deviance information criterion (DIC) is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion, also known as the Schwarz criterion). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Like AIC and BIC it is an asymptotic approximation as the sample size

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

becomes large. It is only valid when the posterior distribution is approximately multivariate normal. Define the deviance as $D(\theta) = -2logp(y|\theta) + C$, where y are the data, θ are the unknown parameters of the model and $p(y|\theta)$ is the likelihood function. C is a constant that cancels out in all calculations that compare different models and which therefore does not need to be known. The expectation $\overline{D} = E[D(\theta)]$ is a measure of how well the model fits the data; the larger this is, the worse the fit. The effective number of parameters of the model is computed as $\mathrm{pD} = \overline{\mathrm{D}} - \mathrm{D}(\bar{\theta})$, where $\bar{\theta}$ is the expectation of θ. The larger this is, the better it is for the model to fit the data. The deviance information criterion is calculated as

$$DIC = pD + \overline{D}$$

The idea is that models with smaller DIC should be preferred to models with larger DIC. Models are penalized both by the value of $\overline{D}$, which favors a good fit, but also (in common with AIC and BIC) by the effective number of parameters pD . Since $\overline{D}$ will decrease as the number of parameters in a model increases, the pD term compensates for this effect by favoring models with a smaller number of parameters. The advantage of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over θ, which is not readily available from the MCMC simulation. But to calculate DIC, simply compute $\overline{D}$ as the average of $D(\theta)$ over the samples of θ, and $D(\bar{\theta})$ as the value of D evaluated at the average of the samples of θ . Then the DIC follows directly from these approximations.

6. Sensitivity Analysis

Investigation of sensitivity to the choice of hyper-prior must be considered since the performance of the selected models can be sensitive to the choice of the variance components priors. Therefore consider alternative specifications, and carry out sensitivity of the selected models assuming an IG with scale and shape parameters a and b respectively. Assume the four alternatives a=0.5, b=0.0005; a=1, b=0.005; a=0.001, b=0.001; a=0.01, b=0.01. The first specification was suggested by Kelsall and Wakefield (1999), for modelling the precision of the spatial effects in an MRF model. The second alternative was proposed in Besag and Kopparberg (1995).The remaining two priors with equal scale and shape parameters, especially $a = b = 0.001$, have often been used as standard choice on the variances of random effects. Re-running MCMC simulations based on these specifications, using the selected models.

7. Conclusion

Traditional parametric duration models are not flexible enough for the identification and separation of cohort and period effects. Without any rather informative prior knowledge about specific forms of non-linear effects, a very large number of parameters have to be introduced, making estimation either very unreliable or even impossible due to divergence or non-existence of estimates. In this situation, non- or semi-parametric approaches that do not assume certain parametric forms of various non-linear and temporal effects are needed. Bayesian Approaches which allow a flexible framework for realistically complex models. These models allow us to analyze usual linear effects of

Author: Tesfaye Abera Jimma, Department of Mathematical and Statistical Science Ethiopian Civil Service University, Tel: +251; 911-679202/+251-912-152361; E-mail: Tesfaye4god@gmail.com; P.O. Box 57935, Addis Ababa, Ethiopia.

categorical covariates, multiple time scales and non-linear effects of continuous covariates within a unified semi-parametric Bayesian framework for modeling and inference.

References

[1] Adebayo, B. (2005). Spatial Modelling of Initiation and Duration of Breastfeeding: Analysis of Breastfeeding Behavior in Malawi. *World Health and Population vol. 10 No.3 2009*.

[2] Belitz, C., Hübner, J., Klasen, S. and Lang, S. (2007): Determinants of the Socioeconomic and spatial pattern of under nutrition by sex in India: A geo-additive semi parametric regression approach. Working paper, Department of Statistics, University of Mϋnich

[3] Brezger, A. and Lang, S. (2006). Generalized Structured Additive Regression Based on Bayesian P-Splines. Computational Statistics and Data Analysis, 50, 967-991.

[4] Eilers, H.C. and Marx, B.D. (1996). Flexible Smoothing with B-splines and Penalties .*Statistical Science*, 11(2): 89-121.

[5] Kandala, N., Fahrmeir, L., Klasen, S. and Priebe, J. (2007). Geo-additive models of Childhood Under nutrition in three Sub-Saharan African Countries, Clinical Sciences Research Institute, Warwick Medical School, University of Warwick, CV2 2DX, and Coventry, UK.

[6] Kelsall, J.E. and Wakefield, J.C. (1999): Bayesian models for spatially correlated disease and exposure data. Oxford University Press, Oxford; 1999:151.

[7] Nelder, J.A. and Wedderburn R.W.M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A* 135, 370-384.

[8] Pawitan, Y., Krishna M., Michiels, S. and Ploner, A.(2005). Bias in the estimation of false discovery rate in microarray studies. *Bioinformatics, 21, 3865 -3872.*

[9] Spiegel halter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002): BayesianMeasures of Model Complexity and Fit. *Journal of the Royal Statistical Society B*, 65, 583-639.