Hacettepe Journal of Mathematics and Statistics
Volume 35 (1) (2006), 91 – 102

SPECIFICATION OF HYPER-PARAMETERS FOR NORMAL PRIOR DISTRIBUTIONS INDUCED ON LOG-LINEAR PARAMETERS

Haydar Demirhan* and Canan Hamurkaroglu*

Received 07 : 10 : 2004 : Accepted 07 : 06 : 2006

Abstract

In this paper, the specification of hyper-parameters of the prior distribution of log-linear parameters is taken into account. Determination of a prior for the log-linear parameters is considered. Some approaches are given to specify the covariance matrix of the prior distribution, which reflects our degree of belief in the prior information. A new approach is proposed to specify the dispersion parameter of the prior distribution. An example is given to clarify the argued matters and a sensitivity analysis is also conducted.

Keywords: Log-linear model, Cauchy-tail distribution, Cholesky decomposition, Gibbs sampling, Metropolis-Hastings algorithm, Bayesian estimation.

1. Introduction

Log-linear models are widely used for the analysis of contingency tables. When one has *a priori* information about the subject before taking a sample, one may use it along with the information that comes from the sample. Bayesian methods allow this information to be included in the estimation process. There are two main approaches for the Bayesian analysis of the log-linear models; the first one is to use a Dirichlet distribution and the second is to use a normal distribution as the prior distribution of the log-linear parameters. The Dirichlet approach has the advantage of allowing a convenient factorization of the likelihood through the identification of cliques within the undertaken model graph. Knuiman&Speed [8] stated that the hyper-Dirichlet distribution is the most tractable choice for multinomial distributed data, however, the researcher restricts himself to decomposable models when using a hyper-Dirichlet distribution as prior. In addition, the necessity of specifying a very large number of hyper-parameters adds an additional level of complexity when a hyper-Dirichlet distribution is used as prior [6]. The second approach is to use a multivariate normal (MVN) distribution as prior distribution [2, 6, 8, 9]. This prior rids us of the decomposable model restriction. The disadvantages

*Department of Statistics, Hacettepe University, Beytepe, 06800, Ankara, Turkey.

of a MVN prior are that a convenient decomposition of the likelihood could not be obtained, and the log-linear parameter updates require global calculations rather than local [6].

Leighty&Johnson [9] discuss Bayesian approaches for the log-linear parameters with a MVN prior induced on the log-linear parameters. Leighty& Johnson [9] use a Cauchy-tail prior for the precision parameter, which reflects one's degree of belief in the prior. Chen&Dunson [1] use a Cholesky decomposition to re-parameterize the covariance matrix of prior distribution of one of the generalized linear mixed model components.

In this paper, the prior specification for the Bayesian estimation of the parameters of a log-linear model is considered. Joint posterior distributions are reached using various prior distributions. To obtain posterior estimates of the log-linear parameters, Markov chain Monte Carlo methods (MZMC) are used. Firstly, the approach of Leighty&Johnson [9] is described, and then it is proposed that the Cholesky decomposition of the covariance matrix of prior distribution can be used to specify a prior distribution for the dispersion parameter of the prior of log-linear parameters. In addition, full conditional posterior distribution of one of the log-linear parameters given the others is derived, when it can be found analytically, and derivation of the mean vector and covariance matrix of the joint posterior distribution of the log-linear parameters is given.

In section 2, basic notations for log-linear models are given, in section 3 determination of the prior distribution and its hyper-parameters is mentioned. Section 4 gives joint posterior densities associated with the log-linear parameters. Finally, a numerical example is presented as an application of the approaches discussed in section 5.

2. The log-linear model and notation

In this paper, the notations given in King&Brooks [6] are used. The set of sources, where the data come from, is denoted by S. The number of elements in a set is denoted by $|\cdot|$, so each source is labelled such that $S = \{S_\zeta : \zeta = 1, \ldots, |S|\}$. The set of levels for source S_ζ is K_ζ, for $\zeta = 1, \ldots, |S|$. The cells of a contingency table can be represented by the set $K = K_1 \times \cdots \times K_{|S|}$, so the cells are indexed by $\boldsymbol{k} \in$K. Expected cell counts and observed cell counts are denoted by $\mathrm{n}_{\boldsymbol{k}}$ and $y_{\boldsymbol{k}}$ for $\boldsymbol{k} \in K$, respectively.

To construct models other than the saturated model, the set of subsets of S, $\wp(S) = \{s : s \subseteq S\}$ is defined. Then, to represent a log-linear model, the index, $m \subseteq \wp(S)$ is used, where m lists the log-linear terms presented in the model. Each element of the model, m is included in a set c such that $c \in m \subseteq \wp(S)$. The constant term of the log-linear model is represented by the inclusion of the empty set in $\wp(S)$. The set $\boldsymbol{M}^c$ contains all possible combinations of the levels of sources included in c. In general, the highest level is not included among the elements of $\boldsymbol{M}^c$. Thus the set $\boldsymbol{M}^c$ is $\{\boldsymbol{m}_1^c, \ldots, \boldsymbol{m}_{|\boldsymbol{M}^c|}^c\}$. Then the log-linear model vector for each $c \in m \subseteq \wp(S)$ is

$$\boldsymbol{\beta}^c = \{\beta_{\boldsymbol{m}_1}^c, \beta_{\boldsymbol{m}_2}^c, \ldots, \beta_{\boldsymbol{m}_{|\boldsymbol{M}^c|}}^c\}.$$

Thus, the log-linear parameter vector for the model m is

$$\boldsymbol{\beta}^m = \{(\boldsymbol{\beta}^{c_1})^T, (\boldsymbol{\beta}^{c_2})^T, \ldots, (\boldsymbol{\beta}^{c_{|m|}})^T\}.$$

The design matrix or model matrix corresponding to the model $m \subseteq \wp(S)$ is denoted by $\boldsymbol{X}_m$. Using this design matrix and the parameter vector, the log-linear model is represented as follows:

$$\log \boldsymbol{n} = \boldsymbol{X}_m \boldsymbol{\beta}_m. \tag{1}$$

In addition, the likelihood of the log-linear parameters can be approximated by the asymptotic normal distribution of the maximum likelihood estimator (MLE), $\boldsymbol{b}$, that is

$$\ell(\boldsymbol{\beta}_m|\boldsymbol{b}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{b}-\boldsymbol{\beta}_m)^T \boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b}-\boldsymbol{\beta}_m)\right\}. \tag{2}$$

More detailed notations for the elements of the design matrix, order of parameters and cells, and examples are given in King&Brooks [6, 7] and Demirhan [3].

3. Prior specification

In this section, a normal prior is induced on the log-linear parameters and specification of hyper-parameters of the prior is considered. When the MVN distribution is taken for the log-linear parameters as prior,

$$\boldsymbol{\beta}_m \sim N(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m),$$

for $m \subseteq \wp(S)$, the probability density function (pdf) of $\boldsymbol{\beta}_m$ is

$$p(\boldsymbol{\beta}_m) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)^T \boldsymbol{\Sigma}_m^{-1}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)\right\}, -\infty < \beta_m^i \in \boldsymbol{\beta}_m < \infty.$$

After specifying the family of the prior distribution, its parameters and hyper-parameters are determined. The mean vector of the log-linear parameters follows a proper uniform distribution. Determination of the prior for the covariance matrix, $\boldsymbol{\Sigma}_m$, can be done in two ways.

3.1. The Approach of Leighty&Johnson. In the approach of Leighty&Johnson [9], a prior distribution for $\boldsymbol{\Sigma}_m$ is specified in two stages. In the first stage, the covariance matrix of the prior distribution is taken as,

$$\boldsymbol{\Sigma}_m = \alpha \boldsymbol{C}_m = \alpha c \boldsymbol{I}_m,$$

where $\boldsymbol{I}_m$ is the identity matrix of dimension $p = \dim(\boldsymbol{\beta}_m)$, and $c = p/\mathrm{tr}(\boldsymbol{V}_{\boldsymbol{b}}^{-1})$, where $\boldsymbol{V}_{\boldsymbol{b}}^{-1}$ is the inverse of the covariance matrix of MLEs [9]. In this case, for $m \subseteq \wp(S)$,

$$p(\boldsymbol{\beta}_m|\alpha) \propto \alpha^{-p/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)^T \frac{1}{\alpha}\boldsymbol{C}_m^{-1}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)\right\}, \tag{3}$$

where $\boldsymbol{\mu}_m$ follows a proper uniform distribution. The distribution of the general precision parameter α is given by the second stage prior.

It is taken that $\tau = 1/(1+\alpha)$ and that $\tau \sim \text{uniform}(0,1)$, to make calculations easier. Values of τ represent the degree of our belief in the prior. Leonard [10] and Leighty&Johnson [9] state that values of this precision parameter close to zero, represent disbelief. According to Dellaportas&Forster [2] values of this precision parameter close to zero give a vague prior and values close to one give an improper prior.

If τ is distributed as uniform$(0,1)$, then the distribution of α is a Cauchy-tail prior with pdf,

$$p(\alpha) = \frac{1}{(1+\alpha)^2},\ \alpha \geq 0,$$

which can be obtained by a transformation of variables using the distribution of τ, and the pdf given in Eq. (3) is expressed in terms of τ as follows:

$$p(\boldsymbol{\beta}_m|\tau) \propto \left[\frac{\tau}{1-\tau}\right]^{p/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)^T \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)\right\}.$$

The distribution of α can be taken as inverse-gamma, however if the variances of the parameters are small, inferences become too sensitive to changes occurring in the values of parameters of the distribution. Using log-uniform or uniform makes the tails of the

distribution more flat, and also taking log-uniform causes an improper posterior, and when the number of levels of the sources is small, a uniform prior results in an improper posterior, besides, it brings a small bias when the variance tends to infinity. The reasons for our choosing a Cauchy-tail prior are that it helps to identify a reasonable prior mean for τ, and our analysis is intended to be used when little prior information is available [4, 6, 9].

3.2. An approach based on the Cholesky decomposition. In the second approach, the Cholesky decomposition of $\mathbf{\Sigma}_m$ is used and then prior distributions are induced on the elements appearing as a result of the decomposition. Let the Cholesky decomposition of $\mathbf{\Sigma}_m$ be

$$\mathbf{\Sigma}_m = \mathbf{\Psi}(\mathbf{\Psi})^T.$$

When $\mathbf{\Psi} = \mathbf{\Lambda\Gamma}$, the decomposition becomes

$$\mathbf{\Sigma}_m = \mathbf{\Lambda\Gamma\Gamma}^T\mathbf{\Lambda},$$

where $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ and for $i = 1, \ldots p$ and $j = i+1, \ldots, p$, $\mathbf{\Gamma} = (\gamma_{ij})$. Here and hereafter p is the number of parameters included in the model and is given by

$$1 + \sum_{c \in m} \prod_{\zeta : S_\zeta \in c} (|K_\zeta| - 1).$$

Using this decomposition it is assumed that $\mathbf{\Sigma}_m$ is defined uniquely, that $\lambda_i > 0$, and that $\mathbf{\Gamma}$ is a lower triangular matrix. Let ω_{ij} denote the (i,j)th element of $\mathbf{\Sigma}_m$, whose dimension is $p \times p$. For $i = 1, \ldots, p$, $j = 1, \ldots, p$; ω_{ii} and ω_{ij} are found as follows:

$$\omega_{ii} = \lambda_i^2(1 + \sum_{j=1}^{i-1} \gamma_{ij}^2),$$

and

$$\omega_{ij} = \lambda_i\lambda_j(\gamma_{ij} + \sum_{r=1}^{i-1} \gamma_{ir}\gamma_{jr}).$$

The non-zero elements of $\mathbf{\Gamma}$ represent prior information on covariances between the log-linear parameters. When the log-linear parameters are assumed *a priori* to be mutually independent from each other, $\mathbf{\Sigma}_m$ reduces to the form: $\mathbf{\Sigma}_m = \text{diag}(\lambda_1^2, \ldots, \lambda_p^2)$.

Then, prior distributions can be placed on each λ_i, for $i = 1, \ldots, p$. It is proposed that when it is assumed that λ_i are independently and identically distributed for all i, a Cauchy-tail prior could be induced on each λ_i, using the transformation, $\xi_i = 1/(1+\lambda_i)$. In this case, the distribution of ξ_i is uniform$(0,1)$. Choosing the distribution of ξ_i as uniform, prior belief in the prior could be expressed more easily because it has a closed form. Here, the values of ξ_i close to zero represent disbelief and the values close to one represent a strong belief in the prior information. Furthermore, by using our approach, precision of the prior belief can be represented for each log-linear parameter separately. At the same time, we can utilize the advantages of using the Cauchy-tail distribution mentioned previously.

4. Posterior inferences

When the prior information is combined with the information coming from the sample, a posterior distribution appears. The posterior distribution differs with the choice of prior. Two different approaches are described for the specification of prior the distribution of the hyper-parameter, $\mathbf{\Sigma}_m$ in §3.1 and §3.2. Thus, the posterior distribution will vary according to the changes in the specification of the prior of $\mathbf{\Sigma}_m$. If the prior is specified

by the first approach, a posterior analysis could be made for $\boldsymbol{\beta}_m$, conditioning on the relative precision parameter τ or the joint posterior density of $\boldsymbol{\beta}_m$, and τ can be used for the posterior analysis. When the proposed procedure is used to specify the prior for $\boldsymbol{\beta}_m$, the posterior distribution of $\boldsymbol{\beta}_m$ given λ_i or the joint posterior distribution of λ_i and $\beta_{\boldsymbol{m}_r}^{c_\ell} \in \boldsymbol{\beta}_m$ could be used for the posterior analysis, where, $c_\ell \in \wp(S)$, $\ell = 1, \ldots, |m|$ and $r = 1, \ldots, |\boldsymbol{M}^c|$.

Posterior densities are given for various cases and the full conditional distributions of some cases are derived when it is possible to find a known distributional form for the relevant full conditional distribution. If such a form is found, Gibbs sampling algorithm can be used for marginal posterior inferences; otherwise, the Metropolis-Hastings (M-H) algorithm, which is described in general in A.1, is a way to obtain them.

4.1. Posterior densities when the approach of Leighty&Johnson is used.

4.1.1. *Posterior density of $\boldsymbol{\beta}_m$ given τ.* The posterior distribution of $\boldsymbol{\beta}_m$ given τ is as follows:

$$\begin{aligned} p(\boldsymbol{\beta}_m|\boldsymbol{b},\tau) &\propto p(\boldsymbol{\beta}_m|\tau)\ell(\boldsymbol{\beta}_m|\boldsymbol{b}) \\ &\propto \left[\frac{\tau}{1-\tau}\right]^{p/2} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)^T \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)\right\} \\ &\qquad \times \exp\left\{-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{\beta}_m)^T \boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b} - \boldsymbol{\beta}_m)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right\}. \end{aligned}$$

Here, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ are the mean vector and covariance matrix of the joint distribution of the log-linear parameters, and can be obtained as

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left[\boldsymbol{V}_{\boldsymbol{b}}^{-1} + \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}\right]^{-1} \tag{4}$$

and

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left[\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}\boldsymbol{\mu}_m\right], \tag{5}$$

respectively. The derivation of (4) and (5) is given in A.2.

4.1.2. *Joint posterior density of $\boldsymbol{\beta}_m$ and τ.* The joint posterior density of $\boldsymbol{\beta}_m$ and τ can be perceived such as

$$\begin{aligned} p(\boldsymbol{\beta}_m,\tau|\boldsymbol{b}) &\propto p(\tau|\boldsymbol{b})p(\boldsymbol{\beta}_m|\tau,\boldsymbol{b}) \\ &\propto p(\tau)\ell(\tau|\boldsymbol{b})p(\boldsymbol{\beta}_m|\tau)\ell(\boldsymbol{\beta}_m|\boldsymbol{b}) \\ &\propto \ell(\tau|\boldsymbol{b})p(\boldsymbol{\beta}_m|\tau)\ell(\boldsymbol{\beta}_m|\boldsymbol{b}). \end{aligned} \tag{6}$$

Leighty&Johnson [9] take $\ell(\tau|\boldsymbol{b})$ proportional to the MVN distribution of $\boldsymbol{b}$, that is

$$\begin{aligned} \ell(\tau|\boldsymbol{b}) \propto \det&\left(\boldsymbol{V}_{\boldsymbol{b}} + \frac{1-\tau}{\tau}\boldsymbol{C}_m\right)^{-1/2} \\ &\times \exp\left\{-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{\mu}_m)^T\left(\boldsymbol{V}_{\boldsymbol{b}} + \frac{1-\tau}{\tau}\boldsymbol{C}_m\right)^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_m)\right\}. \end{aligned} \tag{7}$$

From Eqs. (6) and (7), $p(\boldsymbol{\beta}_m, \tau|\boldsymbol{b})$ is found as follows:

$$\begin{aligned} p(\boldsymbol{\beta}_m, \tau|\boldsymbol{b}) \propto \det\!\left(\boldsymbol{V}_{\boldsymbol{b}} + \frac{1-\tau}{\tau}\boldsymbol{C}_m\right)^{-1/2} \left(\frac{\tau}{1-\tau}\right)^{p/2} \\ \times \exp\!\left\{-\frac{1}{2}\left[(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)^T \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)\right.\right. \\ \times (\boldsymbol{b} - \boldsymbol{\beta}_m)^T \boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b} - \boldsymbol{\beta}_m) \\ \left.\left. \times (\boldsymbol{b} - \boldsymbol{\mu}_m)^T \left(\boldsymbol{V}_{\boldsymbol{b}} + \frac{1-\tau}{\tau}\boldsymbol{C}_m\right)^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_m)\right]\right\}. \end{aligned}$$

4.2. Posterior densities when the Cholesky decomposition of $\boldsymbol{\Sigma}_m$ is used. In this subsection, posterior distributions are studied by dividing the subject into two cases. The first case is that of a Cauchy-tail prior on λ_i when ξ_i is given, the second is with the same prior but when ξ_i is not given.

4.2.1. *Posterior density of $\boldsymbol{\beta}_m$ given $\boldsymbol{\xi}$ with Cauchy-tail prior on each λ_i.* As stated in § 3.2, ξ_i is used instead of λ_i to make the calculations easier. Then,

$$\begin{aligned} p(\boldsymbol{\beta}_m|\boldsymbol{b}, \boldsymbol{\xi}) &\propto p(\boldsymbol{\beta}_m|\xi_i)\ell(\boldsymbol{\beta}_m|\boldsymbol{b}) \\ &\propto \left[\prod_{i=1}^{p} \frac{\xi_i^2}{(1-\xi_i)^2}\right]^{p/2} \exp\!\left\{-\frac{1}{2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)^T \operatorname{diag}\!\left(\frac{\xi_i^2}{(1-\xi_i)^2}\right)(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)\right\} \\ &\qquad \times \exp\!\left\{-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{\beta}_m)^T \boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b} - \boldsymbol{\beta}_m)\right\} \\ &\propto \exp\!\left\{-\frac{1}{2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right\}. \end{aligned} \tag{8}$$

The covariance matrix and the mean vector of $p(\boldsymbol{\beta}_m|\boldsymbol{b}, \xi_i)$ is found as follows:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left[\boldsymbol{V}_{\boldsymbol{b}}^{-1} + \operatorname{diag}\!\left(\frac{\xi_i^2}{(1-\xi_i)^2}\right)\right]^{-1} \tag{9}$$

and

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left[\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \operatorname{diag}\!\left(\frac{\xi_i^2}{(1-\xi_i)^2}\right)\boldsymbol{\mu}_m\right]. \tag{10}$$

The derivation of (9) and (10) is mentioned in A.2.

4.2.2. *Joint posterior density of $\boldsymbol{\beta}_m$ and $\boldsymbol{\xi}$ with Cauchy-tail prior on each λ_i.* The joint posterior density of $\boldsymbol{\beta}_m$ and $\boldsymbol{\xi}$ can be obtained as

$$\begin{aligned} p(\boldsymbol{\beta}_m, \boldsymbol{\xi}|\boldsymbol{b}) &\propto p(\boldsymbol{\xi}|\boldsymbol{b})p(\boldsymbol{\beta}_m|\boldsymbol{\xi}, \boldsymbol{b}) \\ &\propto \ell(\boldsymbol{\xi}|\boldsymbol{b})p(\boldsymbol{\beta}_m|\boldsymbol{\xi})\ell(\boldsymbol{\beta}_m|\boldsymbol{b}), \end{aligned}$$

where $\ell(\boldsymbol{\xi}|\boldsymbol{b})$ can be taken proportional to the MVN distribution of $\boldsymbol{b}$. This argument is similar that of §4.1.2., but using the covariance matrix of the MVN distribution. Then, $\ell(\boldsymbol{\xi}|\boldsymbol{b})$ is taken as follows:

$$\begin{aligned} \ell(\boldsymbol{\xi}|\boldsymbol{b}) \propto \det\!\left[\boldsymbol{V}_{\boldsymbol{b}} + \operatorname{diag}\!\left(\frac{(1-\xi_i)^2}{\xi_i^2}\right)\right]^{-1/2} \\ \times \exp\!\left\{-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{\mu}_m)^T \left[\boldsymbol{V}_{\boldsymbol{b}} + \operatorname{diag}\!\left(\frac{(1-\xi_i)^2}{\xi_i^2}\right)\right]^{-1}(\boldsymbol{b} - \boldsymbol{\mu}_m)\right\}. \end{aligned}$$

Then,

$$
\begin{aligned}
p(\boldsymbol{\beta}_m, \boldsymbol{\xi}|\boldsymbol{b}) \propto \det\Big[\boldsymbol{V}_{\boldsymbol{b}} + \operatorname{diag}\Big(\frac{(1-\xi_i)^2}{\xi_i^2}\Big)\Big]^{-1/2}\Big[\prod_{i=1}^{p}\frac{\xi_i^2}{(1-\xi_i)^2}\Big]^{p/2} \\
\times \exp\Big\{-\frac{1}{2}\Big[(\boldsymbol{b}-\boldsymbol{\mu}_m)^T\Big[\boldsymbol{V}_{\boldsymbol{b}} + \operatorname{diag}\Big(\frac{(1-\xi_i)^2}{\xi_i^2}\Big)\Big]^{-1}(\boldsymbol{b}-\boldsymbol{\mu}_m) \\
\times (\boldsymbol{\beta}_m-\boldsymbol{\mu}_m)^T\operatorname{diag}\Big(\frac{\xi_i^2}{(1-\xi_i)^2}\Big)(\boldsymbol{\beta}_m-\boldsymbol{\mu}_m) \\
\times (\boldsymbol{b}-\boldsymbol{\beta}_m)^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b}-\boldsymbol{\beta}_m)\Big]\Big\}.
\end{aligned} \tag{11}
$$

4.3. Derivation of full conditional posterior distributions. For simplicity, the log-linear parameter vector, $\boldsymbol{\beta}_m$,

$$
\Big(\beta^{c1}_{\boldsymbol{m}_1}, \ldots, \beta^{c1}_{\boldsymbol{m}_{|M^{c1}|}}, \beta^{c2}_{\boldsymbol{m}_1}, \ldots, \beta^{c2}_{\boldsymbol{m}_{|M^{c2}|}}, \ldots, \beta^{c_{|m|}}_{\boldsymbol{m}_1}, \ldots, \beta^{c_{|m|}}_{\boldsymbol{m}_{\left|M^{c_{|m|}}\right|}}\Big)
$$

is denoted by

$$
\big(\beta_1, \beta_2 \ldots, \beta_p\big).
$$

In this section, the full conditional posterior distribution of β_i given $\boldsymbol{\beta}_{-i}$ and τ or $\boldsymbol{\xi}$ with a Cauchy-tail prior on each λ_i is derived. Here, $\boldsymbol{\beta}_{-i}$ denotes the vector that contains the elements of $\boldsymbol{\beta}$, but the i th element of it. The derivation is the same for both of the situations; however the notations convenient for the case when τ is given are used. For the other case, $\boldsymbol{\xi}$ is substituted in place of τ in the Eqs. (12), (13) and (14); and Eqs. (9) and (10) are used to represent $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, respectively.

Because τ and $\boldsymbol{\beta}_{-i}$ are given, $p(\beta_i|\boldsymbol{\beta}_{-i}, \tau, \boldsymbol{b})$ is a function of only β_i. Then from A.2,

$$
p(\boldsymbol{\beta}_m|\tau, \boldsymbol{b}) \propto \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_m - 2\boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\Big]\Big\}. \tag{12}
$$

Let φ_{ij} denote the (i,j) th element of the matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ for $i = 1, \ldots, p$ and $j = 1, \ldots, p$. Then

$$
\begin{aligned}
p(\beta_i|\boldsymbol{\beta}_{-i}, \tau, \boldsymbol{b}) &\propto \exp\Big\{-\frac{1}{2}\Big[\beta_i^2\varphi_{ii} + 2\beta_i\Big(\mu_{\beta_i}\varphi_i - \sum_{i\neq j}\varphi_{ij}(\beta_j - \mu_{\beta_i})\Big) + E\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}\Big[\beta_i^2\varphi_{ii} + 2\beta_i\big(\mu_{\beta_i}\varphi_i - D\big)\Big) + E\Big\}
\end{aligned} \tag{13}
$$

after excluding the constant E and once adding and once subtracting the constant,

$$
\frac{1}{2\varphi_{ii}^{-1}}\Big(\mu_{\beta_i} - \frac{D}{\varphi_{ii}}\Big)^2,
$$

$p(\beta_i|\boldsymbol{\beta}_{-i}, \tau, \boldsymbol{b})$ is obtained as

$$
p(\beta_i|\boldsymbol{\beta}_{-i}, \tau, \boldsymbol{b}) \propto \exp\Big\{-\frac{1}{2\varphi_{ii}^{-1}}\Big[\beta_i - \Big(\mu_{\beta_i} - \frac{D}{\varphi_{ii}}\Big)\Big]^2\Big\}, -\infty < \beta_i < \infty. \tag{14}
$$

The constant D is defined in matrix notation as:

$$
D = (\boldsymbol{\beta}_{-i} - \boldsymbol{\mu}_{-i})^T\boldsymbol{\eta}
$$

where $\boldsymbol{\beta}_{-i}$ and $\boldsymbol{\mu}_{-i}$ are vectors including all the elements of $\boldsymbol{\beta}_m$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ except for the i th element. The vector $\boldsymbol{\eta}$ contains elements of the k th row or k th column of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ other than the i th.

In conclusion, on each pass of the Gibbs sampling algorithm a sample point is generated from the normal distribution with mean $\left(\mu_{\beta_i} - D/\varphi_{ii}\right)$ and variance $1/\varphi_{ii}$ [3].

For the cases given in §4.1.2 and 4.2.2, the M-H algorithm can be used to find posterior estimates of the log-linear parameters.

5. A numerical example

A $4 \times 2 \times 2$ table containing information about 1154 individuals, whose death was caused by skin cancer, was used. See Table 1. The data were collected by the US National Center for Health Statistics and taken from Nazaret [11].

First of all the MLEs of the log-linear parameters were obtained. Then a vague prior was specified for the log-linear parameters by the approach of Leighty&Johnson [9], and then by using the proposed approach based on the Cholesky decomposition.

Table 1. Skin Cancer Data (Observed Counts).

	Melanoma			Non-Melanoma		
	Male	Female	Total	Male	Female	Total
Lips	0	1	1	47	34	81
Eyelids	1	4	5	75	79	154
Ears	3	3	6	106	39	145
Other	24	19	43	360	359	719
Total	38	27	55	588	511	1099

The three sources are the sites of the body (SB) where the cancer originally developed, sex (SX) and the type of cancer (TC). Following the notation given in § 2, $|S| = 3$, S_1 is SB, S_2 is SX and S_3 is TC, so $K_1 = \{1, 2, 3, 4\}$, $K_2 = K_3 = \{1, 2\}$ and

$$\begin{aligned} K = \{&(1,1,1),(2,1,1),(3,1,1),(4,1,1),(1,2,1),(2,2,1),(3,2,1),\\ &(4,2,1),(1,1,2),(2,1,2),(3,1,2),(4,1,2),(1,2,2),(2,2,2),\\ &(3,2,2),(4,2,2)\}. \end{aligned}$$

For this example, $\wp(S) = \{\emptyset, \{S_1\}, \{S_2\}, \{S_3\}, \{S_1, S_2\}, \{S_1, S_3\}, \{S_2, S_3\}, \{S_1, S_2, S_3\}\}$, and model concerned is

$$m = \{\{S_1\}, \{S_2\}, \{S_3\}, \{S_1, S_2\}\}.$$

Then the log-linear model is the same as (1) with

$$\boldsymbol{\beta}_m = (\boldsymbol{\beta}^{c_1}, \boldsymbol{\beta}^{c_2}, \boldsymbol{\beta}^{c_3}, \boldsymbol{\beta}^{c_4})$$

and

$$\boldsymbol{\beta}^{c_1} = (\beta_1^{c_1}, \beta_2^{c_1}, \beta_3^{c_1}),\ \boldsymbol{\beta}^{c_2} = (\beta_1^{c_2}), \boldsymbol{\beta}^{c_3} = (\beta_1^{c_3}),\ \boldsymbol{\beta}^{c_4} = (\beta_{11}^{c_4}, \beta_{21}^{c_4}, \beta_{31}^{c_4}).$$

The MLEs of the log-linear parameters were obtained using the Newton-Raphson method. The vector containing the MLEs of the log-linear parameters was found to be

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_m = \big(&-1.750985113, -0.008052855, -0.381914929, 0.19729938,\\ &-3.824214123, -0.219551426, -0.28684926, 0.695668566\big). \end{aligned}$$

5.1. Analyses by the approach of Leighty&Johnson. In this subsection, posterior analyses for $\boldsymbol{\beta}_m$ are carried on when τ is given and when it is not. The approach of Leighty&Johnson [9] is used to determine the prior distributions.

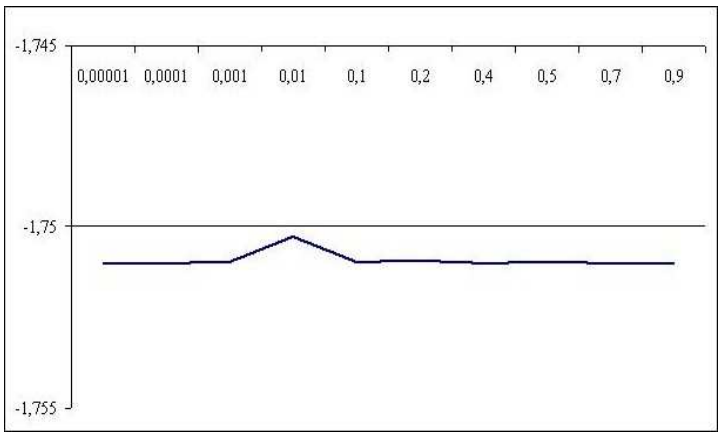
All elements of the mean vector of the prior of log-linear parameters were taken as the average of the MLEs, which was equal to -0.69732497. To specify a vague prior, τ was taken as 10^{-4} and α was obtained as 10^4. Then, $\boldsymbol{\Sigma}_m$ was found as a diagonal matrix with the elements 10.43736952 on the main diagonal.

The likelihood function, used for the analysis, was given by (2). The full conditional posterior distribution of § 4.3 was used to implement the Gibbs sampling. The total number of iterations was 200000, 3750 of which were discarded as burn-in. To reduce the autocorrelation of the Gibbs sequence, a record was made at the end of each 750 cycles. Convergence of the Gibbs sequence was tested by a formal method, namely Geweke's modified z-test, which was introduced by Geweke [5]. Test statistics for all parameters were close to zero, thus it is concluded that there is not enough evidence to conclude that the convergence has not been achieved. Posterior estimates of the log-linear parameters were given by the vector,

$$\widetilde{\boldsymbol{\beta}}_m = \big(-1.68554084, -0.06951882, -0.40279793, 0.17844316, \\ -3.79476511, -0.22285423, -0.27682759, 0.66529336\big).$$

The marginal posterior distributions of $\beta_1^{c1}, \beta_2^{c1}, \beta_3^{c1}$ and β_{21}^{c4} were symmetric and sharper than the normal distribution. Others were not symmetric. Figure 1 presents the results of the sensitivity analysis for β_m^{c1}, conducted by taking the values of τ as 0.00001, 0.0001, 0.001, 0.001, 0.1, 0.2, 0.4, 0.5, 0.7 and 0.9. It can be concluded from Figure 1 that the posterior estimates of $\boldsymbol{\beta}_m$ are insensitive to the changes occurring in τ. For the other elements of $\boldsymbol{\beta}_m$ the same conclusion holds, thus figures for these are omitted.

Figure 1. Posterior estimates of β_m^{c1} versus τ.



When τ is also a random variable, the joint posterior distribution of Eq. (8) is used for the posterior analysis of $\boldsymbol{\beta}_m$ and τ. To obtain the marginal posterior distributions of the log-linear parameters and posterior estimates of them, the M-H algorithm was used. 10000 iterations were made and a record was taken at the end of each 10 cycles. The first 1000 iterations were discarded to reduce the autocorrelation. Geweke's modified z-test indicated convergence. The posterior estimate of τ was found to be 0.002338203. The

marginal posterior distribution of τ was approximately a uniform distribution. Posterior estimates of the log-linear parameters were obtained as follows:

$$\widetilde{\boldsymbol{\beta}}_m = \big(-1.750811502, -0.007962351, -0.381661032, \\ 0.197719215, -3.823869793, -0.2193746, -0.286279317, 0.696276107\big).$$

The marginal distributions corresponding to the levels of SB were approximately symmetric and similar to the normal distribution. The marginal distributions of $\beta_1^{c_1}$, $\beta_2^{c_1}$, $\beta_3^{c_1}$, $\beta_1^{c_2}$, $\beta_{11}^{c_4}$, $\beta_{21}^{c_4}$ and $\beta_{31}^{c_4}$ were similar; they were symmetric and sharper than the normal distribution.

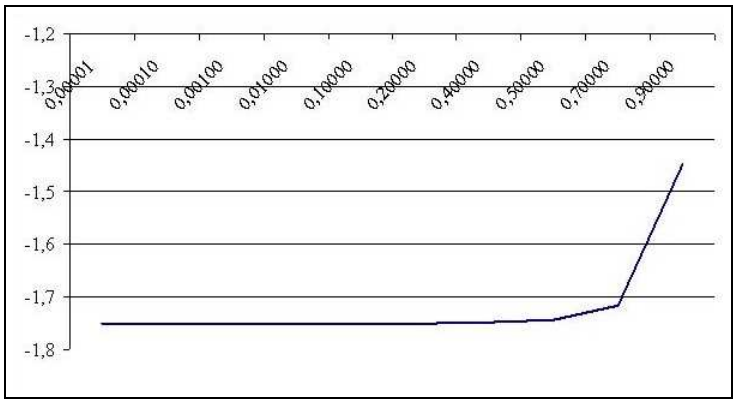
5.2. Analyses when the Cholesky decomposition is used. Firstly, posterior estimates of the elements of β_m given $\boldsymbol{\xi}$ were obtained. Again all elements of the mean vector of the prior of the log-linear parameters were taken to give -0.69732497, and after the aforementioned convenient modifications, the full conditional posterior distribution of § 4.3 was used to implement the Gibbs sampling. Implementation conditions were the same as those in § 5.1. The value of ξ_i was taken as 0.0001 for $i = 1, \ldots, 5$ and 0.01 for $i = 6, 7, 8$. Geweke's modified z-test results indicated that convergence had been achieved. Thus the posterior estimates of the log-linear parameters were found to be

$$\widetilde{\boldsymbol{\beta}}_m = \big(-1.750952825, -0.008036772, -0.381929293, 0.197446479, \\ -3.824182879, -0.219367624, -0.286845778, 0.695666129\big).$$

The marginal distributions of $\beta_1^{c_1}$, $\beta_3^{c_1}$, $\beta_{21}^{c_4}$ and $\beta_{31}^{c_4}$ were symmetric and normal-like distributions. Other marginal distributions were also similar but they were not symmetric or normal-like. To investigate the sensitivity of the posterior estimates to the chosen values of the elements of $\boldsymbol{\xi}$, ξ_i were taken as ξ and values of ξ chosen as 0.00001, 0.0001, 0.001, 0.001, 0.1, 0.2, 0.4, 0.5, 0.7 and 0.9. Figure 2 shows the changes in $\widetilde{\beta}_m^{c_1}$ for various values of ξ. There is not any major change in the estimated values of $\beta_m^{c_1}$ while $\xi < 0.5$, therefore, the posterior estimates are sensitive to the values of ξ when $\xi > 0.5$.

Figures for the other elements of $\widetilde{\boldsymbol{\beta}}_m$ are almost the same as those of $\widetilde{\beta}_m^{c_1}$, and thus are omitted here.

Figure 2. Posterior estimates of $\beta_m^{c_1}$ versus $\boldsymbol{\xi}$.



Secondly, a posterior analysis of $\boldsymbol{\beta}_m$ and $\boldsymbol{\xi}$ was considered for this example. Eq. (11) was used as the joint posterior distribution of $\boldsymbol{\beta}_m$ and $\boldsymbol{\xi}$. Again, the M-H algorithm was used to obtain marginal posterior distributions of the elements of the log-linear parameter vector. Implementation conditions were the same as for the case of the posterior analysis

of $\boldsymbol{\beta}_m$ and τ given in § 5.1. After deciding that convergence had taken place by using Geweke's modified z-test, the posterior estimate of $\boldsymbol{\xi}$ was found to be

$$\widetilde{\boldsymbol{\xi}} = \big(0.000097, 0.000098, 0.000097, 0.000097, \\ 0.000097, 0.000978, 0.000974, 0.000972\big),$$

and the posterior estimate of the log-linear parameter vector was obtained as

$$\widetilde{\boldsymbol{\beta}}_m = \big(-1.750751651, -0.007971609, -0.382000615, 0.197287176, \\ -3.824238483, -0.219585117, -0.286828099, 0.695709563\big).$$

The marginal distributions of the elements of $\boldsymbol{\xi}$ were all approximately uniform, and the marginal distributions of the elements of the log-linear parameter vector had an approximately normal distribution.

6. Discussion

In this paper, a Bayesian estimation of the log-linear parameters has been discussed for various prior distributions. An approach has been proposed for specifying a prior distribution for the dispersion parameter of the prior distribution of log-linear parameters. The proposed approach is more flexible than the approach of Leighty&Johnson [9], because in the proposed approach the user can be represent his degree of belief in the prior information on each log-linear parameter separately. Thus, it results in more reliable Bayesian inferences. Furthermore, when Figure 1 and Figure 2 of § 5 are compared, the Bayesian estimates obtained by using our approach are seen to be less sensitive for the indifference cases.

Appendices

A.1 Implementation of the Metropolis-Hastings Algorithm

Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ denote the interested parameter vector, where p is the dimension of the vector. At each step of the algorithm, let the interested element of $\boldsymbol{\beta}$ be β_i, for $i = 1, \ldots, p$. The algorithm starts with the generation of a candidate point from a proposal distribution, $q(\boldsymbol{\beta},\ \boldsymbol{\beta}^*)$. Here, $\boldsymbol{\beta}_{-i} = \boldsymbol{\beta}^*_{-i}$ for $\boldsymbol{\beta}_{-i} = (\beta_j, j \neq i)$ and $\boldsymbol{\beta}^*_{-i} = (\beta^*_j, j \neq i)$. After generating the candidate point from $q(\boldsymbol{\beta}, \boldsymbol{\beta}^*)$, $\boldsymbol{\beta}^* = (\beta_1, \ldots, \beta_{i-1}, \beta^*_i, \beta_{i+1}, \ldots, \beta_p)$, the generated value is accepted with the probability of α that

$$\alpha(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = min\left[1, \frac{p(\boldsymbol{\beta}^*)q(\boldsymbol{\beta}^*, \boldsymbol{\beta})}{p(\boldsymbol{\beta})q(\boldsymbol{\beta}, \boldsymbol{\beta}^*)}\right].$$

If the candidate point is not accepted, the process is restarted by the generation of a new candidate point.

A.2 Joint posterior distribution of the log-linear parameters given τ or given $\boldsymbol{\lambda}$

Derivation of the parameters of the joint posterior distribution of the log-linear parameters given τ is as follows:

$$p(\boldsymbol{\beta}_m|\boldsymbol{b}, \tau) \propto p(\boldsymbol{\beta}_m|\tau)\ell(\boldsymbol{\beta}_m|\boldsymbol{b}) \\ \propto \left[\frac{\tau}{1-\tau}\right]^{p/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)^T \frac{\tau}{1-\tau}\boldsymbol{C}_m^{-1}(\boldsymbol{\beta}_m - \boldsymbol{\mu}_m)\right\} \\ \times \exp\left\{-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{\beta}_m)^T \boldsymbol{V}_{\boldsymbol{b}}^{-1}(\boldsymbol{b} - \boldsymbol{\beta}_m)\right\}.$$

When $\boldsymbol{\Sigma}_m = \big(\tau/1-\tau\big)^{p/2}$,

$$\begin{aligned} p(\boldsymbol{\beta}_m|\boldsymbol{b},\tau) &\propto \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{b}^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} - \boldsymbol{b}^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{\beta}_m - \boldsymbol{\beta}_m^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \boldsymbol{\beta}_m^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{\beta}_m \\ &\qquad + \boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\beta}_m - \boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m - \boldsymbol{\mu}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\beta}_m + \boldsymbol{\mu}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m\Bigg]\Bigg\} \\ &\propto \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{\beta}_m^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{\beta}_m + \boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\beta}_m - \boldsymbol{b}^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{\beta}_m + \boldsymbol{\mu}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\beta}_m \\ &\qquad - \boldsymbol{\beta}_m^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} - \boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m + \boldsymbol{b}^T\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \boldsymbol{\mu}_m^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m\Bigg]\Bigg\} \\ &\propto \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{\beta}_m^T\big(\boldsymbol{V}_{\boldsymbol{b}}^{-1} + \boldsymbol{\Sigma}_m^{-1}\big)\boldsymbol{\beta} + \boldsymbol{\beta}_m^T\big(\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m\big) \\ &\qquad - \big(\boldsymbol{b}^T\boldsymbol{V}_{\boldsymbol{b}}^{-1} + \boldsymbol{\mu}_m^T\boldsymbol{\Sigma}_m^{-1}\big)\boldsymbol{\beta}_m\Bigg]\Bigg\}. \end{aligned}$$

Letting $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \big(\boldsymbol{V}_{\boldsymbol{b}}^{-1} + \boldsymbol{\Sigma}_m^{-1}\big)^{-1}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\big(\boldsymbol{V}_{\boldsymbol{b}}^{-1}\boldsymbol{b} + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m\big)$,

$$p(\boldsymbol{\beta}_m|\boldsymbol{b},\tau) \propto \exp\Bigg\{-\frac{1}{2}\Bigg[\boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_m - \boldsymbol{\beta}_m^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}} - \boldsymbol{\mu}_{\boldsymbol{\beta}}^T\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_m\Bigg]\Bigg\}.$$

When $\boldsymbol{\lambda}$ is given instead of τ, $\boldsymbol{\Sigma}_m$ is taken as $\operatorname{diag}\Big(\frac{\xi_i^2}{(1-\xi_i)^2}\Big)$, and no changes occur in the rest of the derivation process [3].

References

[1] Chen, Z. and Dunson, D. B. *Random effects selection in linear mixed models*, Biometrics **59**, 762–769, 2003.

[2] Dellaportas, P. and Forster, J. J. *Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models*, Biometrika **86**, 615–633, 1999.

[3] Demirhan, H. *Bayesian Estimation of The Parameters and Expected Cell Counts in Logarithmic Linear Models* (Unpublished M.Sc. thesis, Hacettepe University, Institute of Natural Sciences, Ankara, 2004).

[4] Gelman, A. *Prior distributions for variance parameters in hierarchical models*, www. stat. columbia.edu/ gelman/ research/ unpublished/tau5.pdf, Visit date: February 2004.

[5] Geweke, J. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion)*, Bayesian Statistics **4**, 169–193, Oxford University Press, Oxford, 1992.

[6] King, R. and Brooks, S. P. *Prior induction in log-linear models for general contingency table analysis*, The Annals of Statistics **29** (3), 715–747, 2001.

[7] King, R. and Brooks, S. P. *On the analysis of population size*, Biometrika **88** (2), 317–336, 2001.

[8] Knuiman, M. W. and Speed, T. P. *Incorporating prior information into the analysis of contingency tables*, Biometrics **44**, 1061–1071, 1988.

[9] Leighty, R. M. and Johnson, W.J. *A Bayesian log-linear model analysis of categorical data*, Journal of Official Statistics **6** (2), 133–155, 1990.

[10] Lenonard, T. *Bayesian estimation methods for two-way contingency tables*, Journal of Royal Statistical Society, Ser. B **37**, 23–37, 1975.

[11] Nazaret, W. *Bayesian log-linear estimates for three-way contingency tables*, Biometrika **74**, 401–410, 1987.