

Bayesian Isotonic Estimation for Exponential Family and Beyond

Jayanta Kumar Pal and Dipak K.Dey

Technical Report #2008-1
February 25, 2008

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Bayesian Isotonic Estimation for Exponential Family and Beyond

Jayanta Kumar Pal and Dipak K. Dey

SAMSI, NC and University of Connecticut

Abstract

In the restricted parameter estimation, the use of exponential family have been introduced to include applications from several scientific studies. The MLE based approach or the smoothing type estimators have been studied using monotone link functions. In this paper, we introduce Bayesian techniques to investigate such methods in a general scenario, with illustrations to special examples such as binomial, Poisson etc. The conjugate priors in the exponential family problem helps us to obtain posterior distributions with similar expressions. The log-concavity of the posterior densities allow us to use adaptive rejection sampling for the individual draws. An MCMC method involving Gibbs sampler is developed to sample from that posterior which yields credible regions for the parameters. We modify our method to include change-point estimation as well, where the underlying parameter curve has some known or unknown change-point. Finally, the method is extended to semi-parametric models, where the link function consists of a monotone function of one particular co-variate and a linear model on the other co-variate. The method is shown to be applicable to some simulated data from the binomial and Poisson family to corroborate our findings. We also apply our technique to estimate the current immunization status for the Rubella virus in a study conducted on Austrian males.

Keywords : Bayesian methods, Generalized Linear models, Gibbs sampler, Shape restricted estimation

1 Introduction

Shape restrictions in the parameter space have generated considerable interest among the statisticians in the last few decades. Monotonicity is one such restriction that arises naturally

in the context of reliability, renewal theory, epidemiology, dose-response study etc. For example, the unknown chances of survival of a disease increases or decreases with a particular treatment or clinical measurement. Techniques have been developed in non-parametric curve estimation, or density estimation likewise in such estimation problems. This can be also applied to functional estimation, such as the distribution function, hazard function, reliability function, survival function etc where the curve under consideration is a monotone function of the underlying parameters.

The main focus of these methods are based on least square estimators, primarily related to the normal theory. To make these methods applicable to a wider class of problems, use of generalized linear models has recently become popular. Though the modeling remains similar, the least square techniques are not readily extendable, as the squared error loss is not the most appropriate one in many cases.

Bayesian methods, on the other hand, has the potential to be used more extensively in such restricted parameter space problems. Since the unknown parameters of the model (that can be the mean, intercept, slope) are related to each other by known inequalities, priors can be defined to incorporate those restrictions. The structure of an exponential family also enables us to confine ourselves to conjugate class of priors. Further, we use a simple prior using Dirichlet distribution which takes care of the monotonicity, first introduced by [1]. The main advantage of using the Bayesian method lies in the construction of not only the credible intervals for the unknown parameters, but also the credible region for a number of related quantities including the functional curves. The joint posterior distribution of the parameters can be constructed, using simple MCMC technique such as a Gibbs sampler.

Another interesting estimation problem that can be addressed using the methodology described in this paper is the change-point estimation problem. Wu et al, in [6] uses the isotonic regression to motivate this. Consider an example where a drug is administered to patients at a specific time-point, rather than gradually over days. The goal is to investigate whether that has affected a sudden change in the response. The prior distribution mentioned above can be appropriately modified to reflect such a scenario and the posterior credible region constructed based on that.

2 Restricted parameters in Generalized linear models

As a generalization of the normal theory, we start with the one-parameter exponential family in its canonical form. In this section, we develop our methodology in presence of covariates, under the assumption that the underlying model is a generalized linear model. The conditional distribution of the response x given \mathbf{z} has the form of an exponential family, with the canonical link function $\eta(\mathbf{z}) \equiv \beta_0 + \beta_1 z_1 + \dots + \beta_p z_p$. The underlying form of the exponential family is given by,

$$f(x|\eta) = \exp[x\eta - b(\eta)]h(x), \quad (1)$$

where the density of x given \mathbf{z} is parametrized by $\eta(\mathbf{z})$. We assume that the link function $\eta(\mathbf{z})$ is not fixed, but has a monotone pattern. To this end, the data \mathcal{X} consists of k groups of measurements, with n_i observations $(x_{i1}, \mathbf{z}_{i1}), \dots, (x_{in_i}, \mathbf{z}_{in_i})$ from the i th group ($i = 1, \dots, k$), where x_{ij} 's are response and \mathbf{z}_{ij} 's are p dimensional covariates. The conditional density of x_{ij} on \mathbf{z}_{ij} is given by $f(x, \eta_i(\mathbf{z}_{ij}))$, with respect to some σ -finite measure μ (Lebesgue measure or counting measure in the common examples). The functions η_i are linear, i.e. $\eta_i(\mathbf{z}) \equiv \beta_{0i} + \beta_{1i} z_1 + \dots + \beta_{pi} z_p$. Examples include the normal family (with unknown mean but known variance), binomial distribution and Poisson distribution. In some of the known examples, we need a re-parametrization of the problem to bring it to the canonical form. We start with the most general form and illustrate the methodology in the special cases later. Additionally, though the data consists of independent observations, they are collected in an order (specified through the groups) that affects the relationship between x and \mathbf{z} in a monotone way. Here, the parameters $\beta_i = (\beta_{0i}, \dots, \beta_{pi})$ are assumed to change monotonically. We will proceed by modeling these parameters in a manner that incorporates the shape restrictions. The ordering can be due to some temporal nature of the data, stratified by the cohorts, or some ordinal variable that affects the influence of the covariates on the response.

For simplicity in the computations, we restrict ourselves to $p = 1$, and examine the effect of change in a single parameter in this paper. Our proposed method can be easily modified for more covariates, and varying them simultaneously. The focus is on the scope of such restrictions, rather than technicalities.

The model under consideration is as follows. Let $z \sim p(z)$, β_0 and β_1 are unknown parameters representing the intercept and slope of the regression respectively. Given z , the density of x (with respect to Lebesgue measure or counting measure), is given by $f(x, \beta_0 + \beta_1 z)$, as in (1). Therefore, the joint density of the pair (x, z) is given by

$$p(z) \exp[x(\beta_0 + \beta_1 z) - b(\beta_0 + \beta_1 z)]h(x).$$

We will discuss the two cases, viz. the slope is increasing (or decreasing) over time, or the intercept is changing monotonically.

2.1 Changing intercept

In this model, we assume that, the baseline effect on the response is seen to be greater or less with the ordering of the groups. We assume that the intercept of the model β_0 changes monotonically, i.e., $\beta_{01} \leq \beta_{02} \leq \dots \leq \beta_{0k}$ or $\beta_{01} \geq \beta_{02} \geq \dots \geq \beta_{0k}$. Moreover, we assume that the slope β_1 remain fixed. For notational simplicity, we will denote $\sum_{i=1}^k \sum_{j=1}^{n_i}$ and $\prod_{i=1}^k \prod_{j=1}^{n_i}$ as $\sum_{i,j}$ and $\prod_{i,j}$ respectively. The likelihood of the data \mathcal{X} is given by

$$l(\mathcal{X}|\beta_{01}, \beta_{02}, \dots, \beta_{0n}, \beta_1) = \prod_{i,j} [p(z_{ij})h(x_{ij})] \exp\left[\sum_{i=1}^k \beta_{0i} r_i + \beta_1 \sum_{i=1}^k s_i - \sum_{i,j} b(\beta_{0i} + \beta_1 z_{ij})\right],$$

where $r_i = \sum_{j=1}^{n_i} x_{ij}$, and $s_i = \sum_{j=1}^{n_i} x_{ij} z_{ij}$. The parameters need to be assigned with a prior that incorporates the underlying shape restriction, and uses the structure of the exponential family through conjugate priors as well. To incorporate the shape restriction in the parameter space, but still retain its generality, we introduce a characterization for origin and scale of the intercepts as follows. Let $\beta_{0i} = \beta_0 + \rho \phi_i$ where β_0 denotes the centering of the intercepts, ρ affects the spread and $\Phi = (\phi_1, \dots, \phi_k)$ is responsible for the shape restriction. The sign of ρ indicates whether the trend is increasing or decreasing. The monotonicity of β_{0i} 's is preserved by imposing a similar restriction on Φ , namely $-1 \leq \phi_1 \leq \dots \leq \phi_k \leq 1$. To make the best use of the exponential family structure, we will consider the conjugate priors of the one-parameter exponential family to have a simple form

of the posterior. For a discussion of such conjugate priors, see [3]. Namely, we assign the priors as,

$$\begin{aligned}\pi(\beta_0) &\propto \exp[\alpha_0\beta_0 - \gamma_0b(\beta_0)], \\ \pi(\beta_1) &\propto \exp[\alpha_1\beta_1 - \gamma_1b(\beta_1)], \\ \pi(\rho) &\propto \exp[\alpha_2\rho - \gamma_2b(\rho)],\end{aligned}$$

with hyper-parameters $\alpha_i, \gamma_i, i = 0, 1, 2$. The prior for Φ is assigned as follows : first we restrict $\phi_1 = \phi_2 = -1$ and $\phi_{k-1} = \phi_k = 1$. For the intermediate values, we need to include a wide class of possibilities and Dirichlet distribution is best suited for our cause. We define a Dirichlet distribution with $k - 3$ co-ordinates with hyper-parameters $\alpha_3, \dots, \alpha_{k-1} \geq 1$, namely,

$$f(p_3, \dots, p_{k-1} | \alpha_3, \dots, \alpha_{k-1}) \propto \prod_{i=3}^{k-1} p_i^{\alpha_i - 1}$$

in the $k - 3$ dimensional simplex $\{(p_3, \dots, p_{k-1}) \in [0, 1]^{k-3} : \sum_{j=3}^{k-1} p_j = 1\}$. The ϕ 's are then constructed as $\phi_i = 2 \sum_{j=3}^i p_j - 1$ for $i = 3, \dots, k - 2$. This will ensure that Φ is non-decreasing and bounded in $[-1, 1]$. As a consequence, $\{\beta_{0i}\}$ will then be bounded between $\theta - \rho$ and $\theta + \rho$, and monotonically non-decreasing or non-increasing. This prior with monotonicity restrictions has been introduced by Alvarez and Dey, [1] and will be used in this paper extensively. We will denote it by π_{mon} throughout.

The complete data likelihood can be written as

$$l(\mathcal{X} | \beta_0, \beta_1, \rho, \Phi) \propto \exp[\beta_0 \sum_{i=1}^k r_i + \rho \sum_{i=1}^k \phi_i r_i + \beta_1 \sum_{i=1}^k s_i - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij})].$$

Therefore, the joint likelihood for the data and the parameters becomes,

$$\begin{aligned}l(\mathcal{X}, \beta_0, \beta_1, \rho, \Phi) \\ \propto \exp[\beta_0(\alpha_0 + \sum_{i=1}^k r_i) + \rho(\alpha_2 + \sum_{i=1}^k \phi_i r_i) + \beta_1(\alpha_1 + \sum_{i=1}^k s_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij}) \\ - \gamma_0 b(\beta_0) - \gamma_1 b(\beta_1) - \gamma_2 b(\beta_2)] \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1},\end{aligned}$$

using the fact that the prior on Φ takes the form

$$\pi_{mon}(\Phi) \propto \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}$$

on the set $-1 = \phi_1 = \phi_2 \leq \phi_3 \dots \leq \phi_{k-2} \leq \phi_{k-1} = \phi_k = 1$. We denote this set, where π_{mon} puts its entire mass, as Ω_k . The role of the hyper-parameters α_i , $i = 3, \dots, k-1$ is to incorporate further knowledge or belief about the shape of the parameters beyond monotonicity. For a detailed discussion, see [1]. For example, the change-point problem can be investigated using $\alpha_i = 1$ for all $i \neq \tau$ and $\alpha_\tau = 10$, where τ is the time-point where the change is assumed to take place. If we like to impose conditions like convexity or concavity, that can be incorporated using increasing or decreasing sequences of α_i 's. In the absence of any such knowledge, the simple uniform prior ($\alpha_i = 1$ for all i) can be used. Since the volume of Ω_k is bounded, this will be a proper prior on that set.

To construct credible regions for the parameters β_0, β_1 , we need to compute the posterior distributions of the parameters β_0, β_1, ρ and Φ jointly, i.e., $\pi(\beta_0, \beta_1, \rho, \Phi | \mathcal{X})$, and sample from that distribution. That seems to be difficult in practice and we need to resort to Markov Chain Monte Carlo techniques such as Gibbs sampler. To do this, we need to compute the conditional posterior distributions of the parameters individually, and construct a chain using the conditional distributions for the transition from one sample to another, starting from some initial value of the parameters. As we can see, the conditional posterior distributions have a relatively simple form that can be used to construct the sample, after the burn-in of a first few hundreds of samples. Observe that,

$$\begin{aligned} \pi(\beta_0 | \mathcal{X}, \beta_1, \rho, \Phi) &\propto \exp[\beta_0(\alpha_0 + \sum_{i=1}^k r_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij}) - \gamma_0 b(\beta_0)], \\ \pi(\beta_1 | \mathcal{X}, \beta_0, \rho, \Phi) &\propto \exp[\beta_1(\alpha_1 + \sum_{i=1}^k s_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij}) - \gamma_1 b(\beta_1)], \\ \pi(\rho | \mathcal{X}, \beta_0, \beta_1, \Phi) &\propto \exp[\rho(\alpha_2 + \sum_{i=1}^k \phi_i r_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij}) - \gamma_2 b(\rho)], \\ \pi(\Phi | \mathcal{X}, \beta_0, \beta_1, \rho) &\propto \exp[\rho \sum_{i=1}^k \phi_i r_i - \sum_{i,j} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij})] \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}. \end{aligned}$$

The conditionals of β_0, β_1 and ρ have a similar form of the underlying exponential family, an artifact of our choice of conjugate priors. Observe that, the densities are log-concave in nature (since b'' is a positive function), and we can use Adaptive Rejection Sampling (ARS henceforth) to sample from

them. See [5] for an illustration of ARS. The conditional density of Φ can be sampled as follows. Given the rest of the parameters and the data \mathcal{X} , we can write,

$$\pi(\Phi|\mathcal{X}, \beta_0, \beta_1, \rho) \propto \prod_{i=3}^{k-1} \left\{ \exp[\rho\phi_i r_i - \sum_{j=1}^{n_i} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij})] (\phi_i - \phi_{i-1})^{\alpha_i - 1} \right\}$$

on the set Ω_k . Fixing $\phi_1 = \phi_2 = -1$, $\phi_{k-1} = \phi_k = 1$, the conditional density of ϕ_i given $\phi_j, j \neq i$ is

$$\pi(\phi_i|\mathcal{X}, \beta_0, \beta_1, \rho, \{\phi_j, j \neq i\}) \propto \exp[\rho\phi_i r_i - \sum_{j=1}^{n_i} b(\beta_0 + \rho\phi_i + \beta_1 z_{ij})] (\phi_i - \phi_{i-1})^{\alpha_i - 1} (\phi_{i+1} - \phi_i)^{\alpha_{i+1} - 1},$$

defined on the set $\phi_{i-1} \leq \phi_i \leq \phi_{i+1}$ for $i = 3, \dots, k-2$. This is a log-concave density, since α_i 's are all larger than or equal to 1, and ARS can be applied to sample from this. In each step, we choose a random permutation of the numbers $\{3, \dots, k-2\}$ and sample ϕ_i conditional on the existing values, in that particular random order. If we consider specific members of the exponential family, such as binomial, Poisson or normal etc., the above expressions will be more specific and easier to work with. We include such illustrations in Section 4.

2.2 Changing slope

The effect of the covariate, measured by β_1 can also change monotonically or at a specific time-point. This can be interpreted as a strengthening or weakening of the effect of the covariate on the response. Here, we assume that β_0 remain fixed, but β_{1i} 's follow an increasing or decreasing trend. To model this, we again introduce three parameters β_1, ρ, Φ such that $\beta_{1i} = \beta_1 + \rho\phi_i$, representing the centering, spread and the shape-restrictions. The same priors are used on $\beta_0, \beta_1, \rho, \Phi$ to keep the computations simple. The likelihood now changes to,

$$l(\mathcal{X}|\beta_0, \beta_1, \rho, \Phi) \propto \exp[\beta_0 \sum_{i=1}^k r_i + \rho \sum_{i=1}^k \phi_i s_i + \beta_1 \sum_{i=1}^k s_i - \sum_{i,j} b(\beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij})].$$

As we have used the same prior, the computation of the posterior is similar to the earlier one. The actual posterior is still difficult to sample from and we use a similar Gibbs' sampler to obtain the

posterior samples. The conditional posteriors are now given by,

$$\begin{aligned}\pi(\beta_0|\mathcal{X}, \beta_1, \rho, \Phi) &\propto \exp[\beta_0(\alpha_0 + \sum_{i=1}^k r_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij}) - \gamma_0 b(\beta_0)], \\ \pi(\beta_1|\mathcal{X}, \beta_0, \rho, \Phi) &\propto \exp[\beta_1(\alpha_1 + \sum_{i=1}^k s_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij}) - \gamma_1 b(\beta_1)], \\ \pi(\rho|\mathcal{X}, \beta_0, \beta_1, \Phi) &\propto \exp[\rho(\alpha_2 + \sum_{i=1}^k \phi_i s_i) - \sum_{i,j} b(\beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij}) - \gamma_2 b(\rho)], \\ \pi(\Phi|\mathcal{X}, \beta_0, \beta_1, \rho) &\propto \exp[\rho \sum_{i=1}^k \phi_i s_i - \sum_{i,j} b(\beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij})] \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}.\end{aligned}$$

As shown before, the first three densities are log-concave in nature, and ARS can be used to sample from the posteriors. Now, Φ needs to be sampled in a random order using the conditional distributions for its components, which is log-concave, and therefore ARS will be used again.

2.3 Initialization

To initiate the Gibbs' sampler, we need to specify some starting values for the parameters $(\beta_0, \beta_1, \rho, \Phi)$ in both the models involving changing intercept and changing slope. For that, we use whatever information we have about the ϕ_i 's and use an increasing function with range $[-1, 1]$ lying in Ω_k which reflects that information. For example, if we have reason to believe that there is a strong change-point, then we take a sigmoidal type function. In absence of any information, we can use the linear function $\phi_i = \frac{2i-(k+1)}{k-3}$ truncated to $[-1, 1]$. To get starting values of the other parameters, we observe that, $\eta_i(z_{ij})$ is a linear function of ϕ_i and z_{ij} simultaneously. (For the GLM with changing intercept, $\eta_i(z_{ij}) = \beta_0 + \rho\phi_i + \beta_1 z_{ij}$ and for the GLM with changing slope, $\eta_i(z_{ij}) = \beta_0 + \rho\phi_i z_{ij} + \beta_1 z_{ij}$). Therefore, we use standard GLM techniques to obtain the MLE-s for (β_0, β_1, ρ) using ϕ_i 's and z_{ij} 's as covariates. For example, in case of binomial family and logit link function, a simple logistic regression routine can be used to initialize the sampler.

Another initialization technique that do not require such assumption about ϕ can be used for the mean function estimation (without the covariate). To see this, we note that MLE of η_i in such a model also minimizes the least square problem $\sum_{i=1}^k n_i (r_i/n_i - b'(\eta_i))^2$. Since b' is a monotone

function of η , we can use the Pool-Adjacent-Violator-Algorithm (PAVA) to the values r_i/n_i -s with weights n_i and then use $(b')^{-1}$ to get estimates of $\eta_i = \theta + \rho\phi_i$. Finally, the estimates of θ , ρ and ϕ_i 's are individually recovered using the fact that $\phi_1 = -1$ and $\phi_k = 1$. To keep the Φ in the set Ω_k , we also set $\phi_2 = -1$ and $\phi_{k-1} = 1$. In some examples, inverting the function $b'(\cdot)$ may turn out to be problematic. The binomial family is one such example, where we will sometimes get $r_i = 0$ or n_i , yielding $b'(\eta_i) = 0$ or 1 at the end-points. We prefer using small continuity corrections which will not affect the final estimates, e.g., instead of using $\hat{p}_i = r_i/n_i$ for the PAVA, we may use $\tilde{p}_i = (r_i + .5)/(n_i + 1)$.

3 Application to wider class of problems

The Bayesian methodology applied to the shape-restricted problems can be extended beyond the simple generalized linear models structure. In this Section, we discuss such extensions, ranging from semi-parametric models to change-point analysis. We also discuss how the methods can be modified to compute credible intervals for mean and other functional curves under the shape-restrictions in the one-parameter exponential family.

3.1 Extension to semi-parametric models

The generalized linear model with canonical link function has scope beyond the shape restricted parameter estimation. The class of semi-parametric models with a functional component and a linear component is one such direction. To motivate this, we consider a regression problem where $(x_i, w_i, \mathbf{z}_i), i = 1, \dots, n$ are the observations, and we seek to build a model of x using future values of the variables w and \mathbf{z} . Here, though both the response x and the covariate w are scalar, the predictor \mathbf{z} can be of several dimensions. This is analogous to the structure of analysis of covariance, where the covariates are either quantitative or categorical. Here, we assume that the covariate w is not categorical. Instead, it is either quantitative (discrete or continuous) or an ordinal one. However, its effect on x need not be linear or have any specific parametric form. Rather, we restrict ourselves to the assumption of monotonicity, i.e., the effect of w on the response x is either increasing or decreasing.

The conditional distribution of x given the covariates w and \mathbf{z} is a one-parameter exponential family as (1), with link function

$$\eta(w, \mathbf{z}) \equiv \psi(w) + \beta^t \mathbf{z},$$

where ψ is assumed to be monotone. We consider the cases of continuous and discrete (and ordinal) w separately.

Case 1: continuous w .

Let the joint density of (w, \mathbf{z}) be denoted as $g(w, \mathbf{z})$. Then, the likelihood of the data can be written as,

$$l(\mathcal{X}|\psi, \beta) = \exp\left[\sum_{i=1}^n x_i \{\psi(w_i) + \beta^t \mathbf{z}_i\} - \sum_{i=1}^n b(\psi(w_i) + \beta^t \mathbf{z}_i)\right] \prod_{i=1}^n g(w_i, \mathbf{z}_i).$$

Defining the order statistics of w_i as $w_{(1)} < \dots < w_{(n)}$ (since w is continuous, ties occur with probability zero) and the corresponding realizations of x_i and \mathbf{z}_i as $x_{(i)}$ and $\mathbf{z}_{(i)}$ respectively, we can rewrite the likelihood as,

$$l(\mathcal{X}|\psi, \beta) \propto \exp\left[\sum_{i=1}^n x_{(i)} \{\psi_i + \beta^t \mathbf{z}_{(i)}\} - \sum_{i=1}^n b(\psi_i + \beta^t \mathbf{z}_{(i)})\right] \prod_{i=1}^n g(w_i, \mathbf{z}_i),$$

where $\psi_i = \psi(w_{(i)})$. The monotonicity of the function ψ ensures that $\psi_1 \leq \dots \leq \psi_n$. As we find ourselves now in a more familiar set-up, we model $\psi_i = \theta + \rho\phi_i$ as before, and introduce conjugate priors on (θ, ρ, β) . The prior for ϕ_i 's are again π_{mon} on $-1 = \phi_1 = \phi_2 \leq \dots \leq \phi_{n-2} \leq \phi_{n-1} = \phi_n = 1$ to reflect the ordering. Using the hyper-parameters $\alpha_i, i = 1, \dots, n-1$, γ_1, γ_2 and $\delta_i, \nu_i, i = 1, \dots, p$ (where p is the dimension of \mathbf{z}) we define the prior distribution as

$$l(\theta, \rho, \beta, \Phi) \propto \exp[\alpha_1 \theta + \alpha_2 \rho + \sum_{l=1}^p \delta_l \beta_l - \gamma_1 b(\theta) - \gamma_2 b(\rho) - \sum_{l=1}^p \nu_l b(\beta_l)] \prod_{i=3}^{n-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}.$$

The posterior conditional distributions of the parameters are then given as,

$$\begin{aligned}
\pi(\theta|\mathcal{X}, \beta, \rho, \Phi) &\propto \exp[\theta(\alpha_1 + \sum_{i=1}^n x_i) - \sum_{i=1}^n b(\theta + \rho\phi_i + \beta^t \mathbf{z}_i) - \gamma_1 b(\theta)], \\
\pi(\rho|\mathcal{X}, \beta, \theta, \Phi) &\propto \exp[\rho(\alpha_2 + \sum_{i=1}^n \phi_i x_i) - \sum_{i=1}^n b(\theta + \rho\phi_i + \beta^t \mathbf{z}_i) - \gamma_2 b(\rho)], \\
\pi(\beta_l|\mathcal{X}, \theta, \rho, \Phi, \{\beta_k, k \neq l\}) &\propto \exp[\beta_l(\delta_l + \sum_{i=1}^n z_{il} x_i) - \sum_{i=1}^n b(\theta + \rho\phi_i + \beta^t \mathbf{z}_i) - \nu_l b(\beta_l)], \\
\pi(\Phi|\mathcal{X}, \beta, \theta, \rho) &\propto \exp[\rho \sum_{i=1}^n \phi_i x_i - \sum_{i=1}^n b(\theta + \rho\phi_i + \beta^t \mathbf{z}_i)] \prod_{i=3}^{n-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}.
\end{aligned}$$

Since the goal is to obtain credible regions for ψ and β , or the function η in a combined sense, we need to draw samples from the posterior distribution, computed above. The Gibbs sampler with an ARS scheme needs to be used for the posterior distribution.

Case 2: discrete or ordinal w .

Since w is not categorical, we can order the possible values of w in a sequence, namely w^1, \dots, w^k (ordinal variables readily yield such an order). Even if the range of w is infinite, we only observe finitely many of them, and make inference on that finite set of values. To start with, we group the observations according to different values of w , i.e., $C_i = \{(x_l, \mathbf{z}_l) : w_l = w^i\}$. Namely, we form k groups with fixed values of w and rename the observations as $(x_{ij}, \mathbf{z}_{ij})$ for $j = 1, \dots, n_i$ where n_i is the cardinality of C_i . The likelihood is now,

$$\begin{aligned}
l(\mathcal{X}|\psi, \beta) &= \prod_i P(w = w^i)^{n_i} \prod_{i,j} f_{\mathbf{z}|w}(\mathbf{z}_{ij}|w = w^i) \exp[\sum_{i,j} x_{ij}(\psi(w^i) + \beta^t \mathbf{z}_{ij}) - \sum_{i,j} b(\psi(w^i) + \beta^t \mathbf{z}_{ij})] \\
&\propto \exp[\sum_{i=1}^k \psi_i r_i + \beta^t \sum_{i=1}^k \mathbf{s}_i - \sum_{i,j} b(\psi_i + \beta^t \mathbf{z}_{ij})].
\end{aligned}$$

where $r_i = \sum_{j=1}^{n_i} x_{ij}$, $\mathbf{s}_i = \sum_{j=1}^{n_i} \mathbf{z}_{ij}$, $\psi_i = \psi(w^i)$, satisfying the order $\psi_1 \leq \dots \leq \psi_k$ because of the monotonicity. Now we can apply our method from the GLM scenario as in the last Section and obtain the posterior using conjugate priors. The sampling will be done using ARS techniques.

3.2 Estimation of mean and functional curves

In this section, we focus on the simpler scenario, where one tries to estimate the mean curve of an exponential family in the absence of covariates. Examples include finding the chances of contracting a particular disease over a period of few years, modeling the mean annual temperature in a region etc. The data comes in the form of the realizations of a single variable in k different groups or time-points, namely $\mathcal{X} = (x_{11}, x_{12}, \dots, x_{1n_1}, x_{21}, \dots, x_{kn_k})$. The x_{ij} 's are assumed to be independent observations from the exponential family of (1), with unknown parameters (η_1, \dots, η_k) , based on which group it is from. The likelihood of the data \mathcal{X} given the parameters $\Theta = (\eta_1, \dots, \eta_k)$ is then

$$L(\mathcal{X}|\Theta) = \exp\left[\sum_{i=1}^k (r_i \eta_i - n_i b(\eta_i))\right] \prod_{i,j} h(x_{ij}).$$

where $r_i = \sum_{j=1}^{n_i} x_{ij}$. Observe that r_1, \dots, r_k are the sufficient statistics. Now we concentrate on the assumption of having a natural ordering of the parameters, namely $\eta_1 \leq \dots \leq \eta_k$ or $\eta_1 \geq \dots \geq \eta_k$. This is a special case of the earlier formulation with $z_{ij} = 0$ and our interest focuses on the parameters $\beta_{0i} = \eta_i$. The mean of the observations are given by $E(x_{ij}) = b'(\eta_i)$, an increasing function of the η_i 's. In essence, the assumption of monotonicity in the parameter is same as the assumption of monotonicity of the mean curve. The shape restriction is done through the parametrization as before : $\eta_i = \theta + \rho \phi_i$, such that θ, ρ denote the centering and spread, and $\Phi = (\phi_1, \dots, \phi_k) \in \Omega_k$ imposes the monotonicity. The priors are taken as specified, e.g.,

$$\begin{aligned} \pi(\theta) &\propto \exp[\theta \alpha_1 - \gamma_1 b(\theta)], \\ \pi(\rho) &\propto \exp[\rho \alpha_2 - \gamma_2 b(\rho)], \\ \pi(\Phi) &\propto \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}. \end{aligned}$$

The goal is to obtain posterior samples of $\Theta = (\eta_1, \dots, \eta_k)$, which will lead to credible regions around the mean curve as well. Our goal, after observing the data \mathcal{X} will center on finding credible regions for the vector $\Theta = \theta \mathbf{1} + \rho \Phi$. Therefore, we need to sample from the posterior distribution $\pi(\theta, \rho, \Phi | \mathcal{X})$, and construct a credible band of, say 95% level of confidence for the true Θ . As it is difficult to sample from the actual posterior, we need to use a Gibbs sampler on the three

parameters (or set of parameters). Observe that, the conditional posteriors can be simplified now as,

$$\begin{aligned}\pi(\theta|\mathcal{X}, \rho, \Phi) &\propto \exp[\theta(\alpha_1 + \sum_{i=1}^k r_i) - \sum_{i=1}^k n_i b(\theta + \rho\phi_i) - \gamma_1 b(\theta)], \\ \pi(\rho|\mathcal{X}, \theta, \Phi) &\propto \exp[\rho(\alpha_2 + \sum_{i=1}^k \phi_i r_i) - \sum_{i=1}^k n_i b(\theta + \rho\phi_i) - \gamma_2 b(\rho)], \\ \pi(\Phi|\mathcal{X}, \theta, \rho) &\propto \exp[\rho \sum_{i=1}^k \phi_i r_i - \sum_{i=1}^k n_i b(\theta + \rho\phi_i)] \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}.\end{aligned}$$

Again the sampling can be done using ARS as mentioned earlier.

Apart from the mean curve, the exponential family also leads to monotonicity in the functionals such as the distribution function, survival function, hazard function and so on. We consider the data without the co-variate and parametrized by the single parameter described above.

Distribution function and the survival function: Under the monotonicity assumption, the cumulative distribution curves of the observations, namely

$$F(x; \eta) = \int_{-\infty}^x \exp(t\eta - b(\eta))h(t)d\mu(t)$$

which is a decreasing function of η for any fixed x . i.e., the monotonicity of the parameters η_i will imply that the distribution functions are monotonically arranged as well. Once we can obtain a posterior sample of the parameters, we can construct credible bands for the distribution function curves simultaneously. Since the survival function or the reliability function $\bar{F}(t)$ is given by $1 - F(t)$, this is an increasing function of the η 's, and we can obtain credible regions for them as well.

Hazard function: The hazard function of the exponential family is given by,

$$H(t; \eta) = \frac{f(t, \eta)}{\bar{F}(t; \eta)} = \frac{\exp(t\eta - b(\eta))h(t)}{\int_t^{\infty} \exp(x\eta - b(\eta))h(x)d\mu(x)},$$

which is also a decreasing function of η . We can also construct credible regions for these curves.

The construction of those curves are particularly meaningful in the change point problems, where the curves are expected to be seen in two different groups or clusters, one above the other. The monotonicity will ensure that the curves will not intersect each other, and each sample will consist of a pile of non-overlapping curves, with the two clusters significantly apart.

3.3 Change-point problem

As mentioned earlier, the formulation of the monotone estimation retains the scope for the change-point analysis for the underlying parametric curve. Moreover, the structure is incorporated through the hyper-parameters $\alpha_3, \dots, \alpha_{k-1}$, making the modification quite general and applicable to all types of scenarios described.

In the change-point estimation, the parameters θ_i , β_{0i} , β_{1i} or ψ_i (depending on the problem we are looking at), are assumed to have an abrupt or sudden increase (or decrease in some cases) at a specific point τ . Since $\phi_1 \leq \dots \leq \phi_k$ are the parameters which reflect that change, the quantity $\phi_\tau - \phi_{\tau-1}$ is assumed to be significantly larger than the other increments. This can be incorporated in the prior distribution by taking α_τ much larger than the other α_i 's, e.g., $\alpha_i = \alpha$ for $i \neq \tau$ and $\alpha_\tau = \tilde{\alpha} \gg \alpha$, say. The subsequent analysis reflects the change-point scenario accordingly.

In the scenario where τ is unknown, i.e., we know that the parameter curve has a change point but do not know its location, we include τ as a hyper-parameter as well. Its prior distribution is taken as a discrete uniform over the set $\{3, \dots, k-1\}$. At each step, we need to compute the posterior probabilities of $\tau = i$ for each candidate values and sample from that discrete distribution before the next step. Then, we proceed with that particular value of τ for the next step in the Gibbs sampler. Observe that, the posterior probability (when $\alpha_i = \alpha$ for $i \neq \tau$ and $\alpha_\tau = \tilde{\alpha} \gg \alpha$) of $\tau = r$ is

$$\frac{(\phi_r - \phi_{r-1})^{\tilde{\alpha}-\alpha}}{\sum_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\tilde{\alpha}-\alpha}}.$$

This discrete distribution on $\{3, \dots, k-1\}$ is sampled in each step. As an artifact, we end up with posterior samples for the change-point τ as well, whose location can be of interest from scientific perspective.

This methods can be accordingly modified to the case of multiple change-points (locations known or unknown). Let there be l such change-points, namely $\tau_1 < \tau_2 \dots < \tau_l$. The hyper-parameters $\alpha_{\tau_1}, \dots, \alpha_{\tau_l}$ will then be changed to larger values than the other α_i 's. For the unknown location, we put another hierarchial prior on $\{\tau_1, \dots, \tau_l\}$ that is uniform on $\{3, \dots, k-1\}$, subject to their orderings. The posterior probabilities are then computed to sample from. We leave the detail computation for the sake of brevity.

4 Illustration

In this section, we illustrate the methods for specific members of the exponential family, such as binomial, Poisson, normal using either the generalized linear model with changing slope or intercept, or the mean curve estimation.

Binomial family : Here the density of an observation x with respect to the counting measure on $\{0, 1\}$ takes the form

$$f(x, \eta) = p^x(1 - p)^{1-x} = \exp(x\eta - b(\eta))h(x)$$

with $\eta = \log(p/1 - p)$, $h(x) = 1$ and $b(\eta) = -\log(1 - p) = \log(1 + e^\eta)$. We consider the problem of changing intercepts with one covariate under the current set-up. The priors of β_0 , β_1 and ρ can be chosen from the appropriate beta distribution, with the canonical transformation mentioned above. The prior for Φ is taken as the generic prior π_{mon} . The posterior densities are then given as,

$$\begin{aligned} \pi(\beta_0 | \mathcal{X}, \beta_1, \rho, \Phi) &\propto e^{\beta_0(\alpha_0 + \sum_{i=1}^k r_i)} (1 + e^{\beta_0})^{-\gamma_0} \prod_{i,j} (1 + e^{\beta_0 + \rho\phi_i + \beta_1 z_{ij}})^{-1}, \\ \pi(\beta_1 | \mathcal{X}, \beta_0, \rho, \Phi) &\propto e^{\beta_1(\alpha_1 + \sum_{i=1}^k s_i)} (1 + e^{\beta_1})^{-\gamma_1} \prod_{i,j} (1 + e^{\beta_0 + \rho\phi_i + \beta_1 z_{ij}})^{-1}, \\ \pi(\rho | \mathcal{X}, \beta_0, \beta_1, \Phi) &\propto e^{\rho(\alpha_2 + \sum_{i=1}^k r_i \phi_i)} (1 + e^\rho)^{-\gamma_2} \prod_{i,j} (1 + e^{\beta_0 + \rho\phi_i + \beta_1 z_{ij}})^{-1}, \\ \pi(\Phi | \mathcal{X}, \beta_0, \beta_1, \rho) &\propto \prod_{i=3}^{k-1} \left\{ e^{\rho\phi_i r_i} \prod_{j=1}^{n_i} (1 + e^{\beta_0 + \rho\phi_i + \beta_1 z_{ij}})^{-1} \right\} \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}. \end{aligned}$$

The first three densities are sampled using the ARS technique, since they are all log-concave densities. The conditional distribution of Φ needs to be sampled in a random permutation using the log-concave density

$$\pi(\phi_i | \mathcal{X}, \beta_0, \beta_1, \rho, \{\phi_j, j \neq i\}) \propto e^{\rho\phi_i r_i} \prod_{j=1}^{n_i} (1 + e^{\beta_0 + \rho\phi_i + \beta_1 z_{ij}})^{-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1} (\phi_{i+1} - \phi_i)^{\alpha_{i+1} - 1},$$

making use of the ARS.

The case of changing slopes have only a few adjustments to make from the above computation.

Complete conditional distributions are written as follows:

$$\begin{aligned}
\pi(\beta_0|\mathcal{X}, \beta_1, \rho, \Phi) &\propto e^{\beta_0(\alpha_0 + \sum_{i=1}^k r_i)} (1 + e^{\beta_0})^{-\gamma_0} \prod_{i,j} (1 + e^{\beta_0 + \beta_1 z_{ij} + \rho \phi_i z_{ij}})^{-1}, \\
\pi(\beta_1|\mathcal{X}, \beta_0, \rho, \Phi) &\propto e^{\beta_1(\alpha_1 + \sum_{i=1}^k s_i)} (1 + e^{\beta_1})^{-\gamma_1} \prod_{i,j} (1 + e^{\beta_0 + \beta_1 z_{ij} + \rho \phi_i z_{ij}})^{-1}, \\
\pi(\rho|\mathcal{X}, \beta_0, \beta_1, \Phi) &\propto e^{\rho(\alpha_2 + \sum_{i=1}^k s_i \phi_i)} (1 + e^{\rho})^{-\gamma_2} \prod_{i,j} (1 + e^{\beta_0 + \beta_1 z_{ij} + \rho \phi_i z_{ij}})^{-1}, \\
\pi(\Phi|\mathcal{X}, \beta_0, \beta_1, \rho) &\propto \prod_{i=3}^{k-1} e^{\rho \phi_i s_i} \prod_{j=1}^{n_i} (1 + e^{\beta_0 + \beta_1 z_{ij} + \rho \phi_i z_{ij}})^{-1} \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1}.
\end{aligned}$$

Sampling from the posterior will be done similarly.

Case 2: Poisson family

The probability density function of the observation x with respect to the counting measure on $\{0, 1, 2, \dots\}$ is given by

$$f(x, \eta) = e^{-\lambda} \frac{\lambda^x}{x!} = \exp(x\eta - b(\eta)),$$

where $\eta = \log \lambda$, $h(x) = x!$ and $b(\eta) = e^\eta$. We will illustrate this distribution with mean curve estimation. The GLM can be described likewise. Observe that, we can identify the problem as $r_i \sim Poi(n_i \eta_i)$, for $i = 1, \dots, k$, using the sufficiency of the sums r_1, \dots, r_k .

Writing $\eta_i = \theta + \rho \phi_i$, we look for the posterior distribution of θ, ρ, Φ . The conjugate priors for θ and ρ come from the Gamma family with appropriate parameters and the prior for Φ is taken as the prior π_{mon} . The posterior can be computed as follows:

$$\begin{aligned}
\pi(\theta|\rho, \Phi, \mathcal{X}) &\propto \exp(\theta(\sum_{i=1}^k r_i + \alpha_1) - \sum_{i=1}^k n_i e^{\theta + \rho \phi_i} - \beta_1 e^\theta), \\
\pi(\rho|\theta, \Phi, \mathcal{X}) &\propto \exp(\rho(\sum_{i=1}^k r_i \phi_i + \alpha_2) - \sum_{i=1}^k n_i e^{\theta + \rho \phi_i} - \beta_2 e^\rho).
\end{aligned}$$

Both this densities are log-concave, and sampled using ARS technique. Finally,

$$\pi(\Phi|\theta, \rho, \mathcal{X}) \propto \prod_{i=3}^{k-2} e^{\rho \phi_i r_i - n_i e^{\theta + \rho \phi_i}} \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i - 1},$$

truncated in the set $-1 \leq \phi_3 \leq \dots \leq \phi_{k-2} \leq 1$. Therefore, the conditional density of ϕ_i given $\theta, \rho, \mathcal{X}$ and $\phi_j, j \neq i$ is given by

$$\pi(\phi_i|\theta, \rho, \mathcal{X}, \{\phi_j, j \neq i\}) \propto e^{\rho \phi_i r_i - n_i e^{\theta + \rho \phi_i}} (\phi_i - \phi_{i-1})^{\alpha_i - 1} (\phi_{i+1} - \phi_i)^{\alpha_{i+1} - 1}$$

defined on the set $\phi_{i-1} \leq \phi_i \leq \phi_{i+1}$ for $i = 3, \dots, k-2$. Making use of the log-concavity of the density, we can easily sample the complete posterior using the Gibbs sampler and ARS. For the individual components, we use a random permutation to sample from the conditionals.

Case 3 : Normal family

The probability density function of the observation x with respect to the Lebesgue measure is given by

$$f(x, \eta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - \mu)^2\right] = \exp(x\eta - b(\eta)),$$

where $\eta = \mu$, $h(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ and $b(\eta) = \frac{1}{2}\eta^2$. Again, we discuss this family with respect to mean curve estimation only. The conjugate priors come from the Gaussian family and the prior for Φ remains the same. The posterior can be computed as follows:

$$\begin{aligned} \pi(\theta|\rho, \Phi, \mathcal{X}) &\propto \exp\left(\theta\left(\sum_{i=1}^k r_i + \alpha_1\right) - \frac{1}{2}\sum_{i=1}^k n_i(\theta + \rho\phi_i)^2 - \frac{1}{2}\beta_1\theta^2\right) \\ &\sim N\left(\frac{\sum_{i=1}^k (r_i - \rho n_i \phi_i) + \alpha_1}{\sum_{i=1}^k n_i + \beta_1}, \frac{1}{\sum_{i=1}^k n_i + \beta_1}\right). \end{aligned}$$

Next,

$$\begin{aligned} \pi(\rho|\theta, \Phi, \mathcal{X}) &\propto \exp\left(\rho\left(\sum_{i=1}^k r_i \phi_i + \alpha_2\right) - \frac{1}{2}\sum_{i=1}^k n_i(\theta + \rho\phi_i)^2 - \frac{1}{2}\beta_2\rho^2\right) \\ &\sim N\left(\frac{\sum_{i=1}^k (r_i - \theta n_i)\phi_i + \alpha_2}{\sum_{i=1}^k n_i \phi_i^2 + \beta_2}, \frac{1}{\sum_{i=1}^k n_i \phi_i^2 + \beta_2}\right). \end{aligned}$$

Finally,

$$\begin{aligned} \pi(\Phi|\theta, \rho, \mathcal{X}) &\propto \prod_{i=3}^{k-1} \exp\left(\rho\phi_i r_i - \frac{1}{2}n_i(\theta + \rho\phi_i)^2\right) \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i-1} \\ &= \prod_{i=3}^{k-1} N\left(\frac{r_i - \theta n_i}{\rho n_i}, \frac{1}{n_i \rho^2}\right) \prod_{i=3}^{k-1} (\phi_i - \phi_{i-1})^{\alpha_i-1}, \end{aligned}$$

defined on the set $\phi_3 \leq \dots \leq \phi_{k-1}$. The conditional density of ϕ_i given $\theta, \rho, \mathcal{X}$ and $\phi_j, j \neq i$ is given by

$$\pi(\phi_i|\theta, \rho, \mathcal{X}, \{\phi_j, j \neq i\}) \propto N\left(\frac{r_i - \theta n_i}{\rho n_i}, \frac{1}{n_i \rho^2}\right) (\phi_i - \phi_{i-1})^{\alpha_i-1} (\phi_{i+1} - \phi_i)^{\alpha_{i+1}-1}$$

truncated to the set $\phi_{i-1} \leq \phi_i \leq \phi_{i+1}$ for $i = 3, \dots, k-1$. Hence, we can easily sample the complete posterior using the Gibbs sampler.

5 Simulation and Data Analysis

5.1 Application to simulated data

The first simulation is from the simple binomial family without covariates. We fixed the number of groups as $k = 20$, and simulated n_1, \dots, n_k from a Poisson distribution with mean 10. Nevertheless, n_i 's are treated as fixed group sizes in the analysis. Next, we generated η_i 's as the order statistics from a Gaussian distribution with mean 0 and variance 3. Let $p_i = e^{\eta_i}/(1 + e^{\eta_i})$ be our unknown set of parameters. The data is finally obtained as $r_i \sim \text{bin}(n_i, p_i)$. The values of η_i are $-3.23, -2.55, -1.61, -1.50, -1.37, -0.07, 0.02, 0.12, 0.31, 0.37, 0.80, 0.83, 0.98, 1.09, 1.23, 1.57, 2.32, 2.52, 2.55, 3.34$ giving the true values $\theta = .0579$ and $\rho = 3.2857$. The true values of ϕ_i are computed as $(\eta_i - \theta)/\rho$.

The hyper-parameters are specified as $\alpha_i = 1, i = 1, \dots, k - 1$ and $\gamma_1 = \gamma_2 = 2$, to make the complete prior distribution as non-informative. To run the MCMC, we initialize the chain as follows. Let $\tilde{p}_i = (r_i + .5)/(n_i + 1)$. We compute the initial values $\tilde{\eta}_i$'s as the isotonic minimizer of $\sum_{i=1}^k (n_i + 1) \left(\frac{r_i + .5}{n_i + 1} - b'(\eta_i) \right)^2$, computed using the PAVA.

We ran the Gibbs sampler for 10000 iterations, and left out the first 5000 as burn-in samples. We also decided to include only 1000 of the remaining chain, using only 1 and discarding 4 from every 5 consecutive iterations. This is done to reduce the effect of any autocorrelation present in the sample. The sampling MCMC distribution of θ and ρ are shown in Figure 2. The sampling distribution of Φ are shown as 1000 continuous curves. To get an idea of the sampling distribution, we checked out the sampling distribution for $\hat{\phi}_{12}$, where the true value is $\phi_{12} = .2352$. Figure 2 includes the sampling distribution of the estimator $\hat{\phi}_{12}$. It is observed that all three posterior distributions resemble the Gaussian distribution. Moreover, all of them have a small bias. Next, we implement our method on a Poisson family with co-variates where the slope of the regression curves are increasing with a known change-point. To this end, we have $k = 50$ groups, each having $n_i = 10$ observations. First, we simulate $z_{ij}, j = 1, \dots, 10, i = 1, \dots, 50$ from a normal distribution with mean 0 and variance 10. The response observations, x_{ij} are then sampled from independently from Poisson distributions, with mean $\exp[\beta_0 + \beta_{1i}z_{ij}]$. Also, the true values are taken as $\beta_0 = 0$

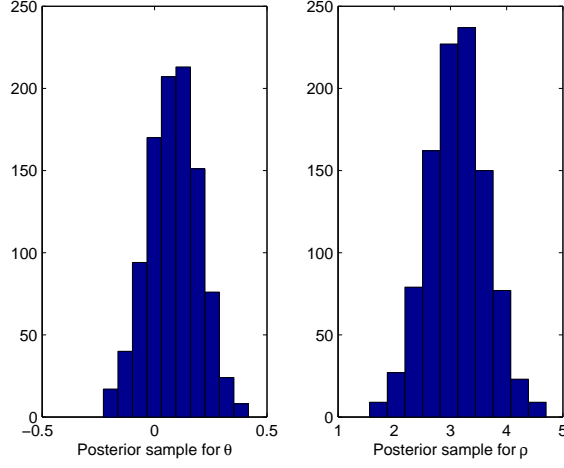


Fig. 1: *The sampling distribution for θ and ρ for binomial data without covariates*

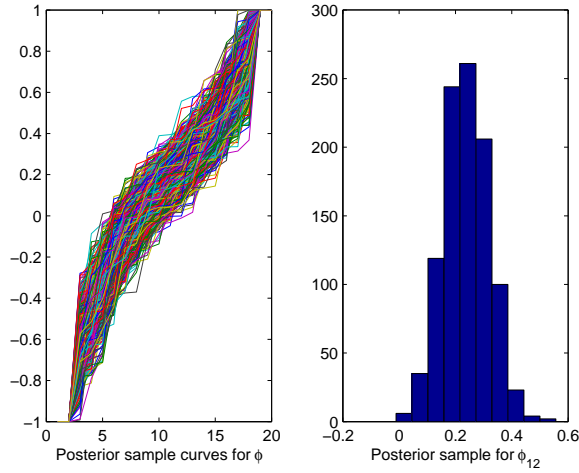


Fig. 2: *The sampling distribution for ϕ for binomial data without covariates*

and $\beta_{1i} = \sigma\left(\frac{i-25.5}{24.5}\right)$, where $\sigma(t)$ is the sigmoid function $1/(1 + e^{-10t})$ between $[-1, 1]$. This is an increasing function with a known change point at 0, meaning $\beta_{11} \leq \beta_{12} \dots \leq \beta_{1n}$ with a sharp change-point at $i = 26$. The hyper-parameters are taken as before, i.e., $\alpha_i = 1$ and $\gamma_i = 2$ to indicate non-informative priors. To incorporate the change-point structure, only α_{26} is changed to 11. Since, $\beta_{11} = \sigma(-1) = 1/(1 + e^{10})$ and $\beta_{1k} = \sigma(1) = e^{10}/(1 + e^{10})$, we have the true values of the

parameters as $\beta_1 = .5$ and $\rho = (e^{10} - 1)/2(1 + e^{10}) = .5$ as well. Figure 3 illustrates the simulation results. The top panel has the posterior distributions for β_0 (true value 0) and β_1 (true value .5), and the bottom panel has the posterior distributions for ρ (true value .5) and 100 posterior samples for ϕ .

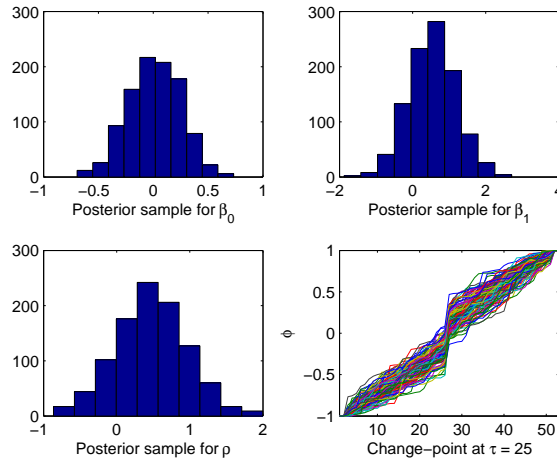


Fig. 3: *The posterior distribution for parameters β_0 , β_1 , ρ and Φ in the Poisson family simulation with co-variates and known change-points*

5.2 Application to real data

To illustrate the methodology in real situations, we consider a data set concerning 230 Austrian males with their age as co-variate. The goal is to investigate the age-specific immunization intensity of Rubella sero-prevalence. The subjects were tested at the Institute of Virology in Vienna. The data has been analyzed from the frequentist approach by [4].

Let T_i denote the age of each individual. For grouping purposes, we treat T_i 's as integers only, with multiple observations available for each year. The age ranges from 0 year to 76 years. The status we observe at the current age is $X_i = 1_{W_i \leq T_i}$, where W_i denote the unobserved immunization age. The regression function is therefore $P(X_i = 1|T_i) = F(T_i)$, where F is the cumulative Distribution Function of W . This is a binomial family model, commonly referred as Current

Status Model in Survival Analysis, with F being an increasing function.

The analysis in [4] hints that the function F possibly has a change-point. In our analysis, we look for an estimate of the change point as well, using the uniform prior on the range $(0, 30)$. This is a modification of the method described earlier. We assume that the change-point is actually before age 30, since the earlier analysis suggests so without any doubt.

We present the 95% credible intervals for the regression function from age 0-28 years in Table 1. The frequentist results can be compared by referring to [4]. In Figure 4 we present the credible intervals over time and the posterior distribution for the change-point.

age	posterior mean	95% C.I.
2	0.469	.380-.582
4	0.472	.382-.584
6	0.491	.386-.642
8	0.494	.389-.703
10	0.574	.402-.792
12	0.587	.416-.864
14	0.827	.583-.931
16	0.869	.686-.967
18	0.871	.741-.969
20	0.882	.741-.972
22	0.886	.752-.985
24	0.924	.802-.989
26	0.941	.841-.997
28	0.944	.843-1.000

Table 1: *Posterior mean and 95% credible intervals for the immunization status for Rubella virus, between ages 0-28 years.*

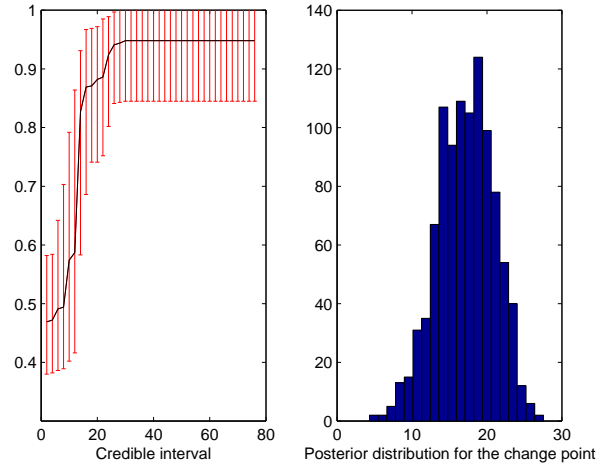


Fig. 4: *95% credible interval for the immunization status in the Rubella data set, and posterior distribution for the change-point*

6 Conclusion

The implementation for Bayesian techniques in shape-restricted problems is fairly new and has the potential of being widely applicable. The intrinsic smoothness of the Bayesian methodology counters the discontinuity issues of the monotone estimates, without the subjectivity of a bandwidth. The method also gets rid of spiking problem, by imposing the estimator to be flat at the two ends. Moreover, prior knowledge such as convexity/concavity or the presence of change-points can be incorporated through the hyper-parameters, without having too many assumptions or constraints.

The generality of the method can be viewed from the wide range of scenarios where it is applied. The linear models with changing slope or intercepts, semi-parametric models with monotone link function, curve estimation etc, all such different situations can be handled simultaneously, an advantage over the frequentist methods that needs to be modified for different situations.

The exponential family structure along with the conjugate priors enable us to work with the easy-to-sample posterior, due to its log-concavity. Moreover, the basic problem can be extended to include other heteroscedastic distributions as well, where the mean and variance need not be

related. The two-parameter exponential family, for example, can be treated in a similar manner, and that creates the scope of future extensions.

Finally, as shown in the real data analysis, the posterior samples readily produce credible intervals, and we can circumnavigate the problem of finding the distribution of the estimators etc (large sample or small sample). This is a natural advantage shared by the Bayesian techniques.

References

- [1] Alvarez, E. E. and Dey, D. K. (2008). Bayesian Isotonic Change-point Analysis. *Annals of the Institute of Statistical Mathematics*, **59** .
- [2] Banerjee, M. (2007). Likelihood based responses for monotone response models. *Annals of Statistics*, **35**, 931-956.
- [3] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Annals of Statistics*, **7**, 269-281.
- [4] Pal, J. K. and Banerjee, M. (2008). Estimation of smooth link functions in monotone response models. *Journal of Statistical Planning and Inference*, doi : 10.1016/j.jspi.2007.12.007
- [5] Wild, P. and Gilks, W. R. (1993). Algorithm AS 287 : Adaptive Rejection Sampling from Log-Concave density estimation. *Applied Statistics*, **42**, 701-709.
- [6] Wu, W. B., Woodroffe, M. B. and Mentz, G. (2001). Isotonic Estimation : Another look at the change-point problem. *Biometrika*, **88**, 793-804.