ORIGINAL PAPER

Bayesian nonparametric quantile mixed-effects models via regularization using Gaussian process priors

Yuta Tanabe[1] · Yuko Araki[2] · Masahiro Kinoshita[3] · Hisayoshi Okamura[4] · Sachiko Iwata[5] · Osuke Iwata[5]

Abstract

In this study, we proposed using Bayesian nonparametric quantile mixed-effects models (BNQMs) to estimate the nonlinear structure of quantiles in hierarchical data. Assuming that a nonlinear function representing a phenomenon of interest cannot be specified in advance, a BNQM can estimate the nonlinear function of quantile features using the basis expansion method. Furthermore, BNQMs adjust the smoothness to prevent overfitting by regularization. We also proposed a Bayesian regularization method using Gaussian process priors for the coefficient parameters of the basis functions, and showed that the problem of overfitting can be reduced when the number of basis functions is excessive for the complexity of the nonlinear structure. Although computational cost is often a problem in quantile regression modeling, BNQMs ensure the computational cost is not too high using a fully Bayesian method. Using numerical experiments, we showed that the proposed model can estimate nonlinear structures of quantiles from hierarchical data more accurately than the comparison models in terms of mean squared error. Finally, to determine the cortisol circadian rhythm in infants, we applied a BNQM to longitudinal data of urinary cortisol concentration collected at

✉ Yuko Araki
  yaraki@tohoku.ac.jp

1   Graduate School of Integrated Science and Technology, Shizuoka University, 3-5-1 Johoku, Naka-ku, Hamamatsu City, Shizuoka 432-8011, Japan

2   Graduate School of Information Sciences, Tohoku University, 6-3-09 Aramaki, Aza, Aoba-ku, Sendai, Miyagi 980-8579, Japan

3   Department of Pediatrics and Child Health, Kurume University School of Medicine, 67 Asahimachi, Kurume City, Fukuoka 830-0011, Japan

4   Cognitive and Molecular Research Institute of Brain Diseases, Kurume University School of Medicine, 67 Asahimachi, Kurume City, Fukuoka 830-0011, Japan

5   Center for Human Development and Family Science, Department of Neonatology and Pediatrics, Nagoya City University Graduate School of Medical Sciences, 1 Kawasumi, Mizuho-cho, Mizuho-ku, Nagoya City, Aichi 467-8601, Japan

Kurume University. The result suggested that infants have a bimodal cortisol circadian rhythm before their biological rhythms are established.

1 Introduction

Regression models are an important statistical tool for capturing the average effects of covariance on a response variable. However, when the distribution of the response variable dependent on the covariates is not symmetric or there exist outliers, the mean regression is not appropriate for representing the phenomenon. In fact, real data analysis has suggested that the effects from covariates on the response variable may not be homogeneous when there are individuals with outlying values of the response variable. Quantile regression (Koenker and Bassett 1978) is a regression method that enables estimation of multiple quantiles. Because of the robustness of quantiles, quantile regression is known to be robust against outliers in the response variable and asymmetry in the response variable distribution. Furthermore, since any quantile can be used as a target of estimation, it is possible to choose a quintile for analysis based on the objective of the particular research and to compare the estimation results of different quantiles. In other words, quantile regression can be used to perform a more diversified analysis than mean regression. Because of these advantages, quantile regression has been attracting attention in recent years. In this paper, we focus on quantile regression modeling of hierarchical data with a nonlinear structure. Note that while the proposed model in this paper is based on longitudinal data, it can be applied to hierarchical data, since longitudinal data are a type of hierarchical data.

Linear mixed-effects models (Laird and Ware 1982) have been a common choice for the analysis of hierarchical data, including longitudinal data. Mixed-effects models are regression models that can capture correlations in observed values within individuals and variations between individuals. The predictor of mixed-effects models is represented by the sum of two terms, a fixed-effects term and a random-effects term. The fixed-effects term formulates the overall trend due to an unknown parameter common to all individuals and the random-effects term formulates the variation among individuals due to an unknown parameter that differs between individuals. Using these two terms, mixed-effects models can simultaneously estimate both the overall transition trend of a response variable and the transitions specific to each individual, taking into account inter-individual variability.

Although linear mixed-effects models estimate conditional expectations of response variables, studies analyzing hierarchical data have also been conducted in the framework of quantile regression. For example, linear quantile mixed-effects models assuming a univariate random effect and multivariate random effects have been proposed in Geraci and Bottai (2007) and Geraci and Bottai (2014), respectively. These models perform quantile regression in the framework of linear mixed-effects models

and inherit the advantages of these models while using arbitrary quantiles instead of the mean value as the estimation targets.

Furthermore, Galarza et al. (2020) and Geraci (2019b) have proposed nonlinear quantile mixed-effects models for hierarchical data with nonlinear structures. These models are based on nonlinear mixed-effects models such as Lindstrom and Bates (1990) and include the assumption that a nonlinear function that adequately models the phenomenon of interest is already known. For this reason, Galarza et al. (2020) used a three-parameter logistic model (Pinheiro and Bates 1995), a well-known growth curve model, as a predictor of their nonlinear quantile mixed-effects models, and Geraci (2019b) used a bi-exponential model (Pinheiro and Bates 1995), which is a pharmacokinetic model, as a predictor of his nonlinear quantile mixed-effects models.

However, there are not many situations in which the models are known for the phenomena of interest. For example, our research studies the infant circadian rhythm of cortisol secretion using longitudinal data. Since the rhythm of cortisol secretion during infancy is unknown, it is impossible to determine the nonlinear function as a predictor of the model in advance.

The purpose of this study was to develop a model that can estimate the transition of quantile points of a response variable from longitudinal data with a nonlinear structure when the nonlinear function cannot be identified in advance. To achieve this goal, we consider a basis expansion method for both fixed- and random-effects terms in quantile mixed-effects models.

Since models with basis expansions can be expressed in the same form as a linear model, they can be treated as a special case of the linear quantile mixed-effects model, such as those proposed by Geraci and Bottai (2007) and Geraci and Bottai (2014). In practice, however, the computational cost of this approach is high. We first need to determine both the number of basis functions for the fixed-effects term and the number of basis functions for the random-effects term, and it is much more expensive to determine the combination of two elements than just one element. Furthermore, smoothing using basis expansions requires determining the value of the regularization parameters, which prevent overfitting and adjust the smoothness of the function.

To overcome these issues, we proposed the Bayesian nonparametric quantile mixed-effects models (BNQMs), which solve these problems by using Bayesian inference. In BNQMs, Bayesian regularization can be performed by assuming a specific distribution of the prior for the coefficient parameters of the basis function. The Markov chain Monte Carlo (MCMC) method is used in BNQMs to estimate the posterior of unknown parameters. In addition, the priors or hyperpriors are assumed hierarchically for the parameters of the priors for the coefficient parameters, and regularization parameters can be estimated simultaneously by the MCMC method. Such a method is called a fully Bayesian method. With this fully Bayesian method, it is no longer necessary to repeat point estimations to determine the optimum value from the candidate values of the regularization parameter, by comparison, using an information criterion.

In this study, we also proposed the use of a Gaussian process (GP) prior for the coefficient parameter vectors of the basis functions. We showed that the smoothness can be appropriately adjusted when the number of basis functions is excessive using a GP prior for Bayesian regularization. Based on this, we can fix the number of basis functions, and the smoothness of the regression curve can be adjusted by optimizing

the hyperparameters of the GP, thus reducing the computational cost. In addition, since BNQMs use the full Bayesian method, these hyperparameters can also be estimated by the MCMC method.

A clear advantage of this approach over existing methods, including the aforementioned Galarza et al. (2020) and Geraci (2019b), as well as the additive and semiparametric models proposed by Yue and Rue (2011), Waldmann et al. (2013), Fenske et al. (2013), and Geraci (2019a), is that the proposed regularization method was developed by focusing more on the problem of using basis function expansions for random-effects terms. Here, the usefulness of the proposed regularization is confirmed by numerical experiments.

The remainder of the article is organized as follows. Section 2 explains the details of BNQMs. We then report our numerical experiments on BNQMs in Section 3. After showing the usefulness of such models in numerical experiments, an application example of a BNQM is shown by analyzing infant cortisol data in Section 4. Finally, this article is concluded in Section 5.

2 Bayesian nonparametric mixed-effects models

2.1 Model

We take the observation of the i-th individual at the j-th measurement time t_{ij}, $\{(t_{ij}, y_{ij}); i = 1, ..., N, j = 1, ..., n_i\}$. Then, the τ-th quantile of y_{ij} at t_{ij} can be modeled as

$$Q_{y_{ij}}(\tau|t_{ij}) = \sum_{k=1}^{p} \beta_{\tau k}\phi_k(t_{ij}) + \sum_{l=1}^{q} b_{\tau i l}\psi_l(t_{ij})$$
$$= \boldsymbol{\beta}_\tau^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_{\tau i}^\top \boldsymbol{\psi}(t_{ij}), \tag{1}$$

where $\tau \in (0, 1)$, $Q_{y_{ij}}(\cdot) \equiv F^{-1}(\cdot)$, and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))^\top$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \ldots, \psi_q(t))^\top$ are vectors of the basis functions in the fixed- and random-effects terms, respectively, $\boldsymbol{\beta}_\tau = (\beta_{\tau 1}, \ldots, \beta_{\tau p})^\top$ is a $p \times 1$ coefficient parameter vector of $\boldsymbol{\phi}(t)$, and $\boldsymbol{b}_{\tau i} = (b_{\tau i 1}, \ldots, b_{\tau i q})^\top$ is a $q \times 1$ coefficient parameter vector of $\boldsymbol{\psi}(t)$, where $\boldsymbol{b}_{\tau i} \sim N(\boldsymbol{0}, \boldsymbol{\Gamma}_\tau)$ is assumed. Here, $\boldsymbol{\Gamma}_\tau$ is a $q \times q$ positive-definite covariance matrix. In the present study, Spline-based Gaussian basis functions (Kawano and Konishi 2007) were used in the numerical experiments and cortisol data analysis. For these basis functions, equidistant points corresponding to the knots of the cubic B-spline (de Boor 2001) are the centers.

Equation (1) can be rewritten in the form of a linear quantile mixed-effects model:

$$Q_{\boldsymbol{y}_i}(\tau|\boldsymbol{t}_i) = \boldsymbol{\Phi}_i \boldsymbol{\beta}_\tau + \boldsymbol{\Psi}_i \boldsymbol{b}_{\tau i}, \tag{2}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$, and $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^\top$. Here, $\boldsymbol{\Phi}_i = (\boldsymbol{\phi}(t_{i1}), \ldots, \boldsymbol{\phi}(t_{in_i}))^\top$ is an $n_i \times p$ design matrix in the fixed-effects term and $\boldsymbol{\Psi}_i =$

$(\boldsymbol{\psi}(t_{i1}), \ldots, \boldsymbol{\psi}(t_{in_i}))^\top$ is an $n_i \times q$ design matrix in the random-effects term. For simplicity of notation, we will omit the subscript τ in the remainder of the paper.

2.2 Estimation

2.2.1 Likelihood function

In BNQMs, we consider Bayesian estimation of the unknown parameters of Eq. (1) using the MCMC method. Yu and Moyeed (2001) proposed a method to express likelihood using an asymmetric Laplace distribution as a Bayesian approach in quantile regression. We assume that the conditional distribution of y_{ij} is an asymmetric Laplace distribution, $y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma \sim AL(\mu_{ij}, \sigma, \tau)$; therefore, its probability density function can be written as

$$p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij} - \mu_{ij}}{\sigma}\right)\right\}, \tag{3}$$

where $\mu_{ij} = \boldsymbol{\beta}^\top\boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_i^\top\boldsymbol{\psi}(t_{ij})$ is the location parameter, $-\infty < \mu_{ij} < \infty, \sigma > 0$ is the scale parameter, $0 < \tau < 1$ is the skewness parameter, and $\rho_\tau(u) = u(\tau - I(u < 0))$ is the loss function with indicator function $I(\cdot)$.

Furthermore, Kozumi and Kobayashi (2011) proposed the following location-scale mixture representation of the asymmetric Laplace distribution:

$$\begin{aligned}
y_{ij}|v_{ij} &\sim N(\mu_{ij} + a\sigma v_{ij}, \delta^2\sigma^2 v_{ij}), \\
v_{ij} &\sim \text{Exp}(1), \\
a &= \frac{1 - 2\tau}{\tau(1 - \tau)}, \\
\delta^2 &= \frac{2}{\tau(1 - \tau)},
\end{aligned} \tag{4}$$

where Exp(1) denotes an exponential distribution with mean 1 and $\mu_{ij} = \boldsymbol{\phi}(t_{ij})^\top\boldsymbol{w} + \boldsymbol{\psi}(t_{ij})^\top\boldsymbol{b}_i$. This representation, which uses a normal distribution and an exponential distribution hierarchically, improves the sampling efficiency of the MCMC method.

Based on these settings, the conditional distribution of y_{ij} in the BNQM can be written as follows:

$$p(y_{ij}|\boldsymbol{\beta}, \boldsymbol{b}_i, v_{ij}, \sigma) = N(y_{ij}|\mu_{ij} + a\sigma v_{ij}, \delta^2\sigma^2 v_{ij}), \tag{5}$$

where $N(y_{ij}|\mu_{ij} + a\sigma v_{ij}, \delta^2\sigma^2 v_{ij})$ is a probability density function of the normal distribution with mean $\mu_{ij} + a\sigma v_{ij}$ and variance $\delta^2\sigma^2 v_{ij}$. From equation (5), the likelihood function can be written as

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{v}, \sigma) = \prod_{i=1}^{N}\prod_{j=1}^{n_i} N(y_{ij}|\mu_{ij} + \alpha\sigma v_{ij}, \delta^2\sigma^2 v_{ij}), \tag{6}$$

where $\boldsymbol{b} = (\boldsymbol{b}_1^\top, \cdots, \boldsymbol{b}_N^\top)^\top$ and $\boldsymbol{v} = (v_{11}, \cdots, v_{1n_1}, \cdots, v_{N1}, \cdots, v_{Nn_N})^\top$.

2.2.2 GP prior regularization for basis expansion method

In the basis expansion method, regularization is important for adjusting the smoothness of the estimated function. In the Bayesian framework, regularization can be achieved using a specific distribution as a prior. In this study, we proposed using a GP prior for the regularization of the coefficients of the basis functions. We will call this Bayesian regularization method "GP prior regularization". The proposed GP prior regularization method is explained in this section and the advantages of using a GP prior for regularization in the basis expansion method were demonstrated in a simple numerical experiment, detailed later.

In the Bayesian framework, the regularization can be realized by setting a specific prior for the coefficient parameters. For a Bayesian regularization for a p-dimensional coefficient parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$, a ridge regularization (l_2 regularization) assuming the following multivariate normal distribution for $\boldsymbol{\beta}$ is often performed:

$$\boldsymbol{\beta} \sim N_p(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}). \tag{7}$$

In this case, the hyperparameter σ_β^2 is used to regulate the norm of $\boldsymbol{\beta}$ to prevent overfitting.

Consider the basis expansion method expressed by the following equation:

$$\begin{aligned} f(t) &= \beta_1 \phi_1(t) + \cdots + \beta_p \phi_p(t) \\ &= \boldsymbol{\beta}^\top \boldsymbol{\phi}(t). \end{aligned} \tag{8}$$

In our treatment, we focused on the fact that the parameters $\beta_k, \beta_l (k, l = 1, \ldots, p)$ are assumed to be independent of each other and we developed a method to set the prior. Specifically, we considered that the smoothness of the estimation curve could be more effectively adjusted using a prior that satisfies the assumption that the value of the coefficient parameter is closer when the location of the corresponding basis function is closer. For this, we assume a GP prior for the coefficient parameters of the basis function, which can be written as

$$\beta_k \sim GP(0, k(s_k, s_l)), k, l = 1, \ldots, p, \tag{9}$$

where $k(s_k, s_l)$ is a kernel function. In the GP regression, kernel functions determine the similarity of the outputs, and Eq. (9) defines the similarity (covariance) between β_k and β_l. (If $k = l$, the kernel function defines the variance of β_k.)

We propose using the center of the basis function s_k as the input of the kernel function, which satisfies the assumption that the coefficient parameter is closer when the corresponding basis function is closer. In this study, we use the following RBF kernel as $k(s_k, s_l)$:

$$k(s_k, s_l | \alpha, \rho) = \alpha^2 \exp \left\{ -\frac{1}{2\rho^2}(s_k - s_l)^2 \right\}, \tag{10}$$

where α and ρ are hyperparameters ($\alpha > 0$, $\rho > 0$), and the scale (magnitude of amplitude) and smoothness of the function $f(t)$ estimated by the basis expansion method expressed as equation (8) are adjusted.

Equation (9) is equivalent to assuming the following multivariate normal distribution as the prior for $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{K}), \tag{11}$$

where $\boldsymbol{K}$ is the $p \times p$-dimensional covariance matrix and its (k, l) element is determined by $k(s_k, s_l)$.

This method addresses the problem of overfitting when the number of basis functions is too large for the complexity of the nonlinear structures. To demonstrate this advantage, we conducted a simple numerical experiment. Figures 1 and 2 compare the fitting results for quantile regression using the basis expansion method when ridge regularization expressed by Eq. (7) and GP prior regularization expressed by Eq. (9) are used, with a change in the number of basis functions. The results show that the smoothness can be optimized even if the number of basis functions is excessive when a GP prior is assumed. In this method, the numbers of basis functions in the fixed-effects term and the random-effects term are unified to a fixed value p. The fact that it is not necessary to determine the combination of the number of basis functions based on an information criterion drastically reduces the computational cost.

2.2.3 Prior setting of BNQM

Here, we summarize the prior setting of BNQMs. Since BNQMs are mixed-effects models, we assume a GP prior for the coefficient parameters of the basis function in the fixed-effects and random-effects terms, expressed as

$$\begin{aligned}
\boldsymbol{\beta} | \alpha_f, \rho_f &\sim N(\mathbf{0}, \boldsymbol{G}), \\
\boldsymbol{b}_i | \alpha_r, \rho_r &\sim N(\mathbf{0}, \boldsymbol{\Gamma}),
\end{aligned} \tag{12}$$

where $\boldsymbol{G}$ and $\boldsymbol{\Gamma}$ are the $p \times p$-dimensional covariance matrices and the (k, l) elements of $\boldsymbol{G}$ and $\boldsymbol{\Gamma}$ are given by the following RBF kernels:

$$\begin{aligned}
k_f(s_k, s_l | \alpha_f, \rho_f) &= \alpha_f{}^2 \exp\left\{ -\frac{1}{2\rho_f{}^2}(s_k - s_l)^2 \right\}, \\
k_r(s_k, s_l | \alpha_r, \rho_r) &= \alpha_r{}^2 \exp\left\{ -\frac{1}{2\rho_r{}^2}(s_k - s_l)^2 \right\},
\end{aligned} \tag{13}$$

in which α_f, α_r, ρ_f, and ρ_r are hyperparameters ($\alpha_f, \alpha_r, \rho_f, \rho_r > 0$). Note that α_f and α_r adjust the scale of the regression curve and ρ_f and ρ_r adjust the smoothness of the regression curve. Furthermore, in this study, we estimate the posterior of these hyperparameters by the MCMC method, assuming the following priors for these hyperparameters:

$$\alpha_f \sim N_+(0, \sigma_\alpha^2), \alpha_r \sim N_+(0, \sigma_\alpha^2),$$
$$\rho_f \sim IG(g_a, g_b), \rho_r \sim IG(g_a, g_b), \tag{14}$$

where $N_+(0, \sigma_\alpha^2)$ is the half-normal distribution and $IG(g_a, g_b)$ is the inverse gamma distribution. In addition, we assume a half-normal distribution $N_+(0, c^2)$ for the prior for σ, which is a scale parameter of the asymmetric Laplace distribution in this study.

2.3 Posterior

Using the likelihood of Sect. 2.2.1 and the prior of Sect. 2.2.3, the posterior of all unknown parameters in BNQMs can be expressed by Bayes' theorem as follows:

$$p(\boldsymbol{\beta}, \boldsymbol{b}, \alpha_f, \alpha_r, \rho_f, \rho_r, \sigma, \boldsymbol{v}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{b}, \sigma, \boldsymbol{v}) \times$$
$$p(\boldsymbol{\beta}|\alpha_f, \rho_f)p(\boldsymbol{b}|\alpha_r, \rho_r) \times$$
$$p(\alpha_f)p(\alpha_r)p(\rho_f)p(\rho_r)p(\sigma)p(\boldsymbol{v}), \tag{15}$$

where $p(\boldsymbol{b}|\alpha_f, \rho_f) = \prod_{i=1}^N p(\boldsymbol{b}_i|\alpha_r, \rho_r)$.

In BNQMs, sampling is performed with the posterior distribution of equation (15) by the MCMC method. Specifically, the MCMC method uses an algorithm called the No-U-Turn sampler (NUTS; Hoffman and Gelman 2014), which is an extension of the Hamiltonian Monte Carlo (HMC) method (Duane et al. 1987). This algorithm has been implemented as a standard in the probabilistic programming language Stan (Stan Development Team 2020), and BNQMs have been implemented using Stan. The details of HMC and NUTS are described in appendix B. In this study, posterior mean values were used as point estimates of the unknown parameters, and the lower 2.5% and upper 2.5% points of the posterior were used to estimate each 95% credible interval.

3 Numerical experiments

In this section, we evaluate the usefulness of BNQMs through numerical experiments. The target quantiles of this numerical experiment are set to be $\tau = (0.10, 0.25, 0.50, 0.75, 0.90)$. The purpose of the proposed model is to appropriately capture the nonlinear transitions of quantiles. To evaluate whether BNQMs achieve this purpose, two types of data structures are considered:

(S1) Symmetric distribution of y given t.
(S2) Asymmetric distribution of y given t.

For S1, the error is generated from the standard normal distribution, while for S2, the error is generated from the chi-square distribution with three degrees of freedom to reproduce each situation. In S1 and S2, the observation y_{ij} at the j-th measurement point of the i-th subject is generated according to the following corresponding equations ($i = 1, \cdots, 10, j = 1, \ldots, 50$):
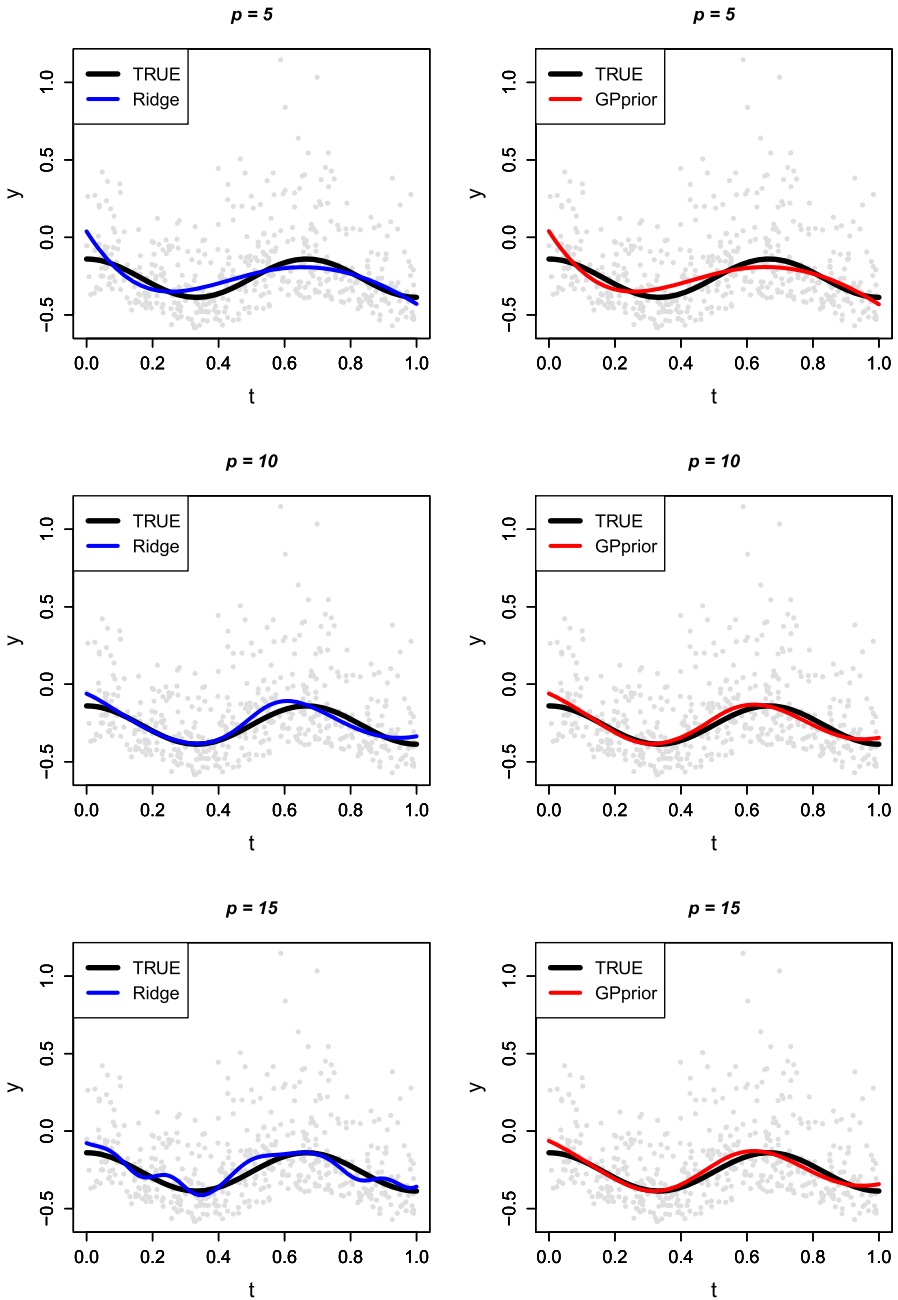
Fig. 1 Comparison of fitting by nonparametric quantile regression with Bayesian ridge regularization and GP prior regularization performed. Here, p is the number of basis functions
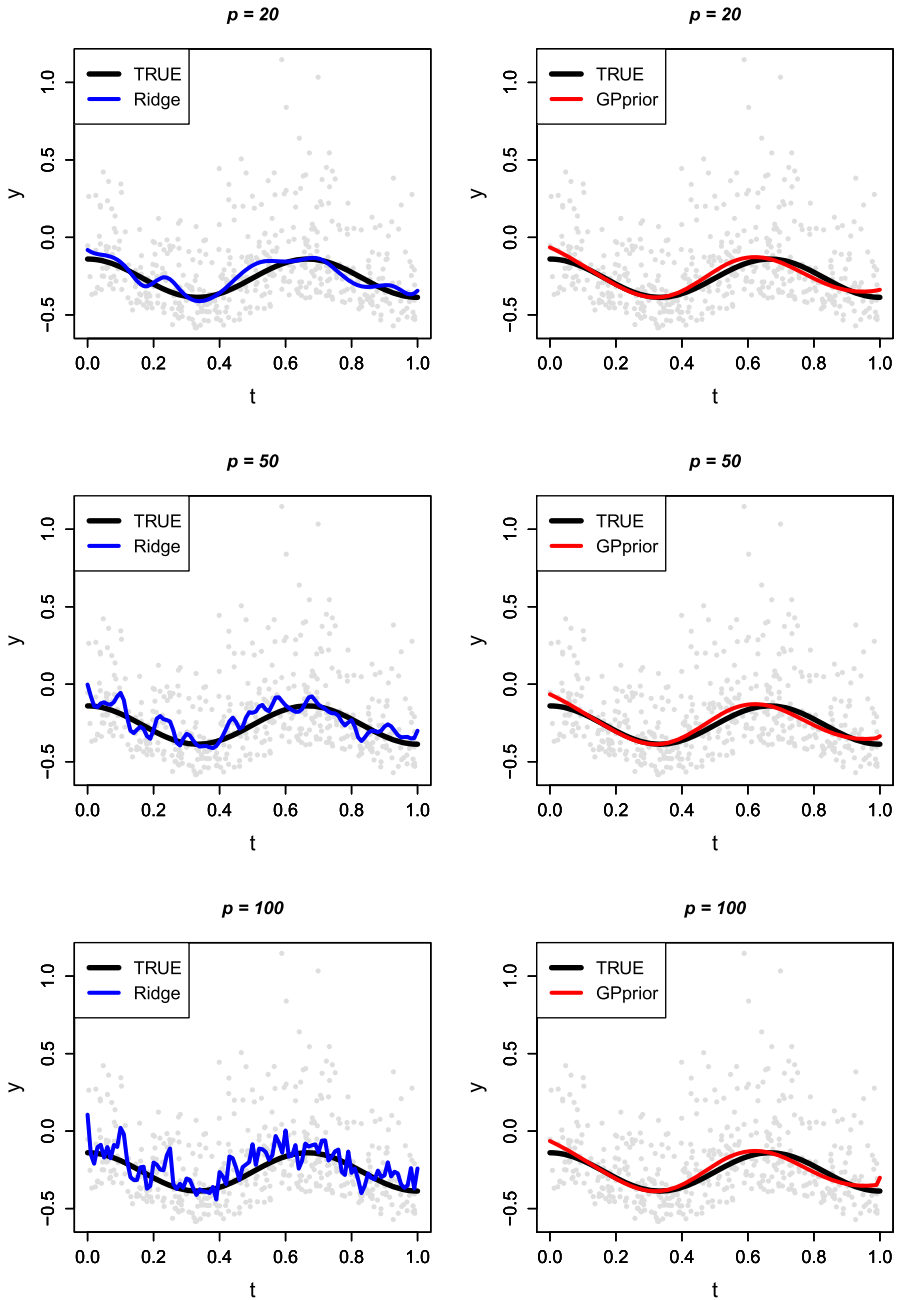
Fig. 2 Comparison of fitting by nonparametric quantile regression with Bayesian ridge regularization and GP prior regularization performed (continuation of Fig. 1). Here, p is the number of basis functions

(S1)

$$y_{ij} = 0.1 \cos(3\pi t_{ij}) + b_{i1} \sin(2\pi t_{ij}) + \sigma \epsilon_{ij},$$
$$\sigma = 0.2 \exp\{-\sin(2\pi t_{ij}) + 0.01\},$$
$$\epsilon_{ij} \sim N(0, 1), \tag{16}$$

(S2)

$$y_{ij} = 0.1 \cos(3\pi t_{ij}) - 0.5 + b_{i1} \sin(2\pi t_{ij}) + \sigma \epsilon_{ij},$$
$$\sigma = 0.01 \cos(3\pi t_{ij}) + 0.1,$$
$$\epsilon_{ij} \sim \chi^2(3), \tag{17}$$

where the measurement time t_{ij} is generated from the uniform distribution $U(0, 1)$ and the random-effect b_{i1} is generated from $N(0, 0.2^2)$.

From Eqs. (16) and (17), 100 sets of simulation data are generated. The data for each structure type are shown in Fig. 3, in which the red curves represent the true quantile structures ($\tau = 0.10, 0.25, 0.50, 0.75, 0.90$). As shown in Takeuchi et al. (2006), the true quantile structure $Q^*_{y_{ij}}(\tau|t_{ij})$ can be expressed as a product of the standard deviation σ and the inverse of the cumulative distribution function of the error $F^{-1}(\tau)$. From this, the true fixed effects structures of the τ-th quantile in S1 and S2 are given by the following equations:

(S1)

$$Q^*_{y_{ij}}(\tau|t_{ij}) = 0.1 \cos(3\pi t_{ij}) + 0.2 \exp\{-\sin(2\pi t_{ij}) + 0.01\} F^{-1}(\tau), \tag{18}$$

(S2)

$$Q^*_{y_{ij}}(\tau|t_{ij}) = 0.1 \cos(3\pi t_{ij}) - 0.5 + \{0.01 \cos(3\pi t_{ij}) + 0.1\} F^{-1}(\tau). \tag{19}$$

We compared the estimation accuracy of the proposed BNQMs-GP with BNQMs-Ridge and the Bayesian nonparametric quantile regression model: BNQ-GP.

BNQMs-GP and BNQMs-Ridge are the same BNQMs with different regularization methods. By comparing these two models, we verify the usefulness of our proposed regularization using GPprior. In order to show that the method is robust in the case of an excessive number of basis functions, we prepared cases where the numbers of basis functions are $p = 15$ and $p = 30$ for the fixed-effects term and the random-effects term, respectively. Although the number of basis functions 30 is considered excessive for the nonlinear structure of the simulation data (S1, S2) that we have prepared, we will investigate the accuracy of curve fitting in this situation.

Unlike the proposed BNQMs-GP, the BNQ-GP is not a mixed-effects model. From the comparison of the two models, it was confirmed that the proposed model constructed within the framework of the mixed-effects model worked better for hierarchical data with large individual differences.
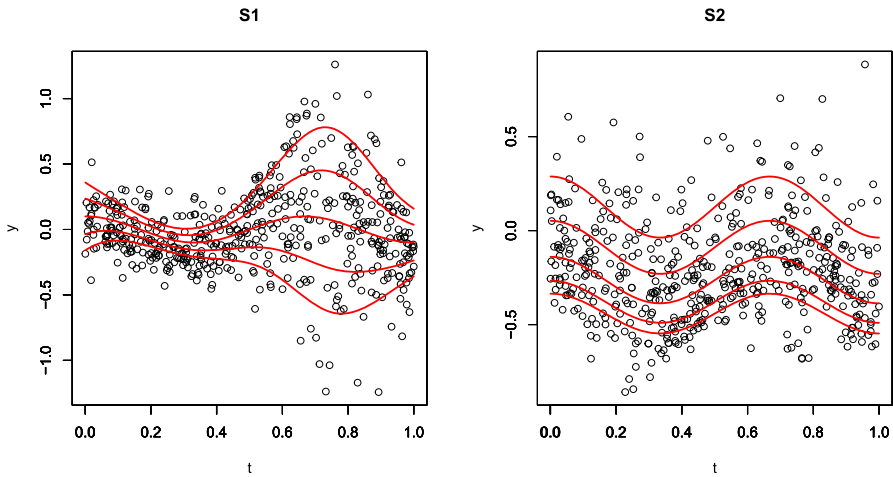
Fig. 3 True quantile structures (fixed-effects) for S1 (left) and S2 (right) expressed by Eqs. (18) and (19), respectively. The red curves represent the true structures when $\tau = 0.90, 0.75, 0.50, 0.25, 0.10$, from top to bottom

(BNQMs-GP)

$$
\begin{aligned}
Q_{y_{ij}}(\tau | t_{ij}) &= \boldsymbol{\beta}^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_i^\top \boldsymbol{\psi}(t_{ij}), \\
\beta_k &\sim GP(0, k_f(s_k, s_l)), k = 1, \ldots, p, \\
b_{ik} &\sim GP(0, k_r(s_k, s_l)), \\
k_f(s_k, s_l | \alpha_f, \rho_f) &= \alpha_f{}^2 \exp\left\{ -\frac{1}{2\rho_f{}^2}(s_k - s_l)^2 \right\}, \\
k_r(s_k, s_l | \alpha_r, \rho_r) &= \alpha_r{}^2 \exp\left\{ -\frac{1}{2\rho_r{}^2}(s_k - s_l)^2 \right\}.
\end{aligned}
\tag{20}
$$

(BNQMs-Ridge)

$$
\begin{aligned}
Q_{y_{ij}}(\tau | t_{ij}) &= \boldsymbol{\beta}^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_i^\top \boldsymbol{\psi}(t_{ij}), \\
\beta_k &\sim N(0, \lambda_\beta), k = 1, \ldots, p, \\
b_{ik} &\sim N(0, \lambda_b), \\
\lambda_\beta &\sim N(0, 1) \\
\lambda_b &\sim N(0, 1)
\end{aligned}
\tag{21}
$$

(BNQ-GP)

$$
\begin{aligned}
Q_{y_{ij}}(\tau | t_{ij}) &= \boldsymbol{\beta}^\top \boldsymbol{\phi}(t_{ij}), \\
\beta_k &\sim GP(0, k(s_k, s_l)), k = 1, \ldots, p, \\
k(s_k, s_l | \alpha, \rho) &= \alpha^2 \exp\left\{ -\frac{1}{2\rho^2}(s_k - s_l)^2 \right\}.
\end{aligned}
\tag{22}
$$

Here, $p = 15, 30$, $\alpha_f \sim N_+(0, 1)$, $\alpha_r \sim N_+(0, 1)$, $\rho_f \sim IG(6.28, 1.35)$, and $\rho_r \sim IG(6.28, 1.35)$. The method for setting these hyperpriors is described in detail in A. As the weakly informative prior for the scale parameter σ of the asymmetric Laplace distribution, the half-normal distribution $\sigma \sim N_+(0, 1)$ was used. The estimation accuracy of each model is evaluated using the mean and standard deviation of the mean square error (MSE) and the coverage probability of 95% credible intervals. Here, MSE is expressed by the following equation:

$$\frac{1}{Nm} \sum_{i=1}^{N} \sum_{j=1}^{m} \{Q^*_{y_{ij}}(\tau|t_{ij}) - \hat{Q}_{y_{ij}}(\tau|t_{ij})\}^2, \, N = 10, m = 50, \tag{23}$$

where $\hat{Q}_{y_{ij}}(\tau|t_{ij})$ is the estimated value of the fixed-effects term of each comparison model and $Q^*_{y_{ij}}(\tau|t_{ij})$ is the true value. Note that the coverage probability of 95% credible intervals is the percentage of the 95% credible interval that contained $Q^*_{y_{ij}}(\tau|t_{ij})$.

Tables 1 and 2 show the mean and standard deviation (SD) of 100 MSEs and the coverage probabilities obtained by fitting comparison models. A small MSE indicates an accurate estimation and a small SD indicates a stable estimation. For both S1 and S2, BNQMs-GP shows smaller mean values of MSEs at all quantiles than BNQMs-Ridge and BNQ-GP. These results show that the estimation accuracy of BNQMs, which are mixed-effects models, is higher than BNQ-GP, which is not a mixed-effects model. For BNQMs, the variation due to individual differences could be regarded as a random-effects term by the mixed-effects modeling, and thus a more accurate and stable estimate could be obtained. Since BNQ-GP is not a mixed-effects model, it cannot account for variations due to individual differences, resulting in large variations in estimated values. These results show that BNQMs can estimate the nonlinear structure of quantiles from hierarchical data more accurately than a non-mixed-effects model. Furthermore, it can be seen that the accuracy of BNQMs-Ridge drops significantly when the number of basis functions is p=30, while BNQMs-GP shows almost no change. This indicates that GPprior can be used to adjust the smoothness of the function appropriately when the number of basis functions is excessive, even though this is a complex task of performing quantile regression in a mixed-effects model.

The coverage probability was also the closest to 95% for BNQM-GP. However, there were cases where the coverage probability was lower than 95%, and this occurred at the edges of the distribution. This may be due to the characteristics of quantile regression. Note that Yang et al. (2016) argue that interval estimation may not work well if confidence intervals are obtained directly from the posterior distribution with ALD as the working likelihood, because the posterior inference is not asymptotically valid. At this time, we do not deal with this existing underlying problem and consider it as future work, but more reliable interval estimation may be possible, for example, using bootstrap confidence intervals. As a result, we were also able to further assert the good points of the proposed method through this additional experiment. That is, the proposed method can maintain the accuracy of estimation at the edges even when the number of basis functions is excessive. When the number of basis functions is excessive, the regularization by Ridge makes the estimation quite unstable and the

Table 1 Mean ($\times 10^{-3}$) and SD ($\times 10^{-3}$) of MSE, and coverage probability in S1

		BNQMs-GP		BNQMs-Ridge		BNQ-GP	
	p	MSE	Coverage	MSE	Coverage	MSE	Coverage
$\tau = 0.10$	15	5.18 (4.42)	91.0%	8.56 (6.03)	88.7%	11.6 (8.30)	52.1%
	30	5.75 (4.65)	89.8%	19.5 (11.2)	82.6%	11.8 (8.25)	42.8%
$\tau = 0.25$	15	3.08 (3.00)	94.1%	10.1 (2.96)	92.1%	5.20 (5.07)	64.8%
	30	3.02 (2.90)	94.1%	6.98 (4.45)	90.9%	5.24 (5.04)	64.8%
$\tau = 0.50$	15	2.57 (3.11)	93.0%	2.71 (2.44)	91.7%	3.39 (4.10)	72.8%
	30	2.63 (3.07)	93.3%	3.07 (2.17)	91.1%	3.42 (4.13)	73.4%
$\tau = 0.75$	15	3.74 (4.10)	92.0%	5.13 (4.43)	90.6%	5.60 (5.31)	63.7%
	30	3.81 (4.14)	91.8%	7.68 (4.90)	91.2%	5.64 (5.28)	64.0%
$\tau = 0.90$	15	5.25 (5.41)	91.6%	8.41 (6.36)	89.1%	10.4 (6.38)	53.0%
	30	5.80 (6.05)	90.7%	19.0 (10.4)	82.3%	10.7 (6.51)	53.1%

Table 2 Mean ($\times 10^{-3}$) and SD ($\times 10^{-3}$) of MSE, and coverage probability in S2

		BNQMs-GP		BNQMs-Ridge		BNQ-GP	
	p	MSE	Coverage	MSE	Coverage	MSE	Coverage
$\tau = 0.10$	15	1.97 (2.15)	94.8%	2.60 (2.19)	91.8%	8.44 (6.08)	42.5%
	30	2.02 (2.21)	94.5%	4.02 (2.31)	92.4%	8.46(6.06)	42.8%
$\tau = 0.25$	15	2.17 (2.19)	94.4%	2.83 (2.17)	92.2%	2.95 (2.83)	64.9%
	30	2.21 (2.24)	94.1%	3.65 (2.34)	93.1%	2.96(2.85)	64.7%
$\tau = 0.50$	15	2.56 (2.15)	94.6%	3.59 (2.18)	92.1%	2.97 (2.78)	71.5%
	30	2.60 (2.18)	94.6%	5.54 (2.43)	92.1%	2.98(2.79)	71.2%
$\tau = 0.75$	15	3.40 (2.46)	95.0%	4.44 (2.37)	93.0%	4.87 (3.83)	69.0%
	30	3.46 (2.47)	94.9%	5.97 (2.52)	91.2%	4.92(3.86)	69.9%
$\tau = 0.90$	15	6.54 (3.48)	95.9%	8.81 (3.84)	91.9%	8.27 (5.66)	65.1%
	30	7.61 (3.93)	90.4%	15.1 (4.29)	78.8%	8.73(6.03)	66.3%

coverage immediately becomes worse at once. On the other hand, the proposed method was able to suppress the decline of the coverage at the edges of the distribution. This is a value merit of the proposed method that is not found shared with other methods.

4 Application

In this section, we present a real data analysis using the proposed model. To clarify the cortisol secretion circadian rhythm in infants, longitudinal data of urinary cortisol concentrations were collected at Kurume University School of Medicine. The sample size of these data was 455 and the subjects comprised 12 infants aged 31 to 124 days after birth (approximately 1–4 months of age) whose urinary cortisol levels were repeatedly measured over several days. The data are color-coded in Fig. 6 by subject.
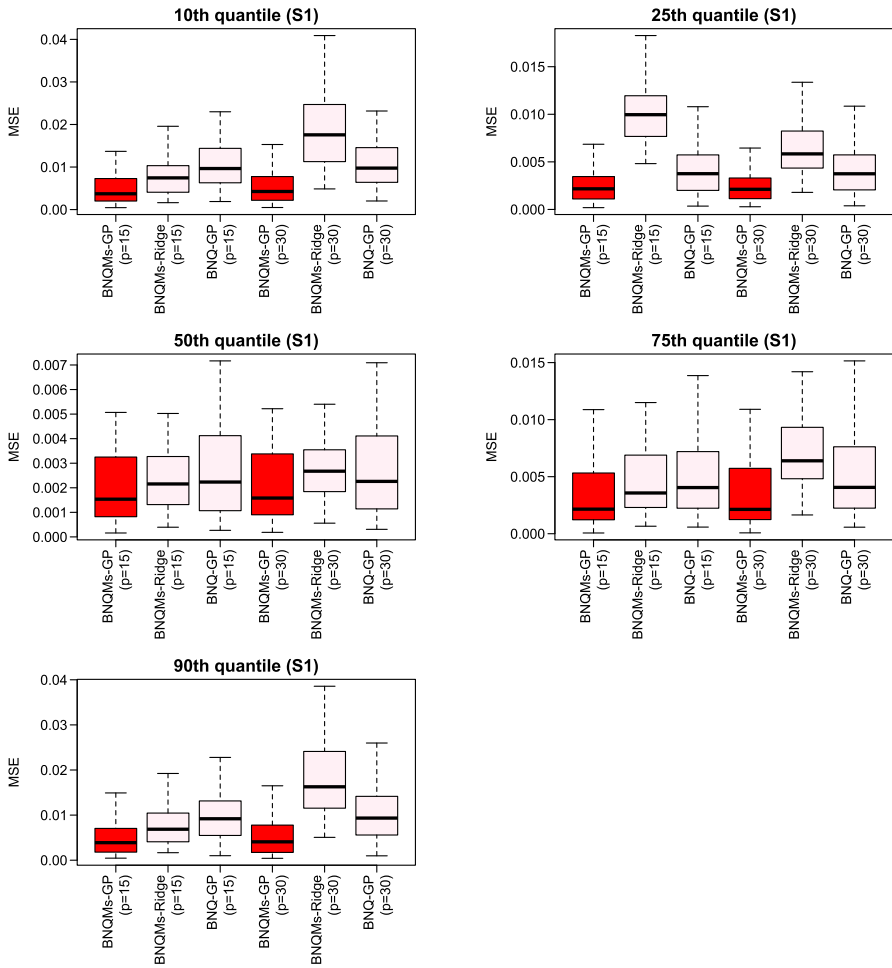
Fig. 4 Boxplots of 100 MSEs for S1

Since the number of measurement points per day and the measurement interval vary, and are sparse for each individual, a mixed-effects model is considered to be suitable. Although the secretion rhythm of cortisol is known to have a nonlinear structure, the secretion rhythm of infants has not been clarified (de Weerth et al. 2003), so a nonlinear equation cannot be assigned to the model in advance. Therefore, the proposed model, which can estimate the nonlinear structure from the data by basis expansions, is useful in this situation. The role of a BNQM in an application to cortisol data is to estimate the overall nonlinear structure (namely, the circadian rhythm) as a fixed-effects term while capturing fluctuations due to individual differences as a random-effects term. In addition, the proposed model estimates the quantile transitions of cortisol values. This has two main advantages. The first advantage is that the model can appropriately handle the problem that the distribution of cortisol concentration is asymmetric. There
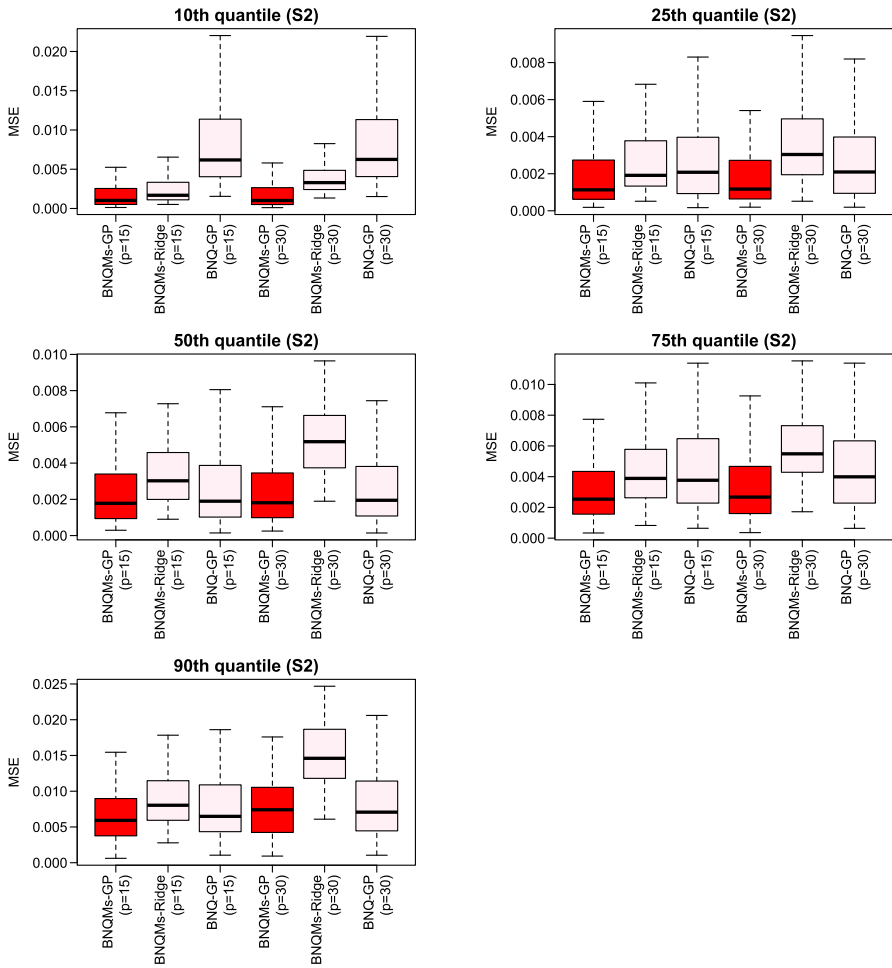
Fig. 5 Boxplots of 100 MSEs for S2

was a small number of individuals with extremely high cortisol concentrations in these data. In other words, the distribution of cortisol concentrations has a long right tail. The mean is greatly influenced by these extreme values, but quantile regression can instead focus on the transition of the median (50th quantile), which is less sensitive. The second advantage is that it is possible to capture the difference in rhythm based on the cortisol concentration itself. It is possible to compare the estimation results of multiple quantiles, and we expect to obtain useful information that could not be obtained by looking at the average. In this study, the 10th, 25th, 50th, 75th, and 90th quantiles were analyzed, and the rhythm differences between high, middle, and low cortisol levels were considered.
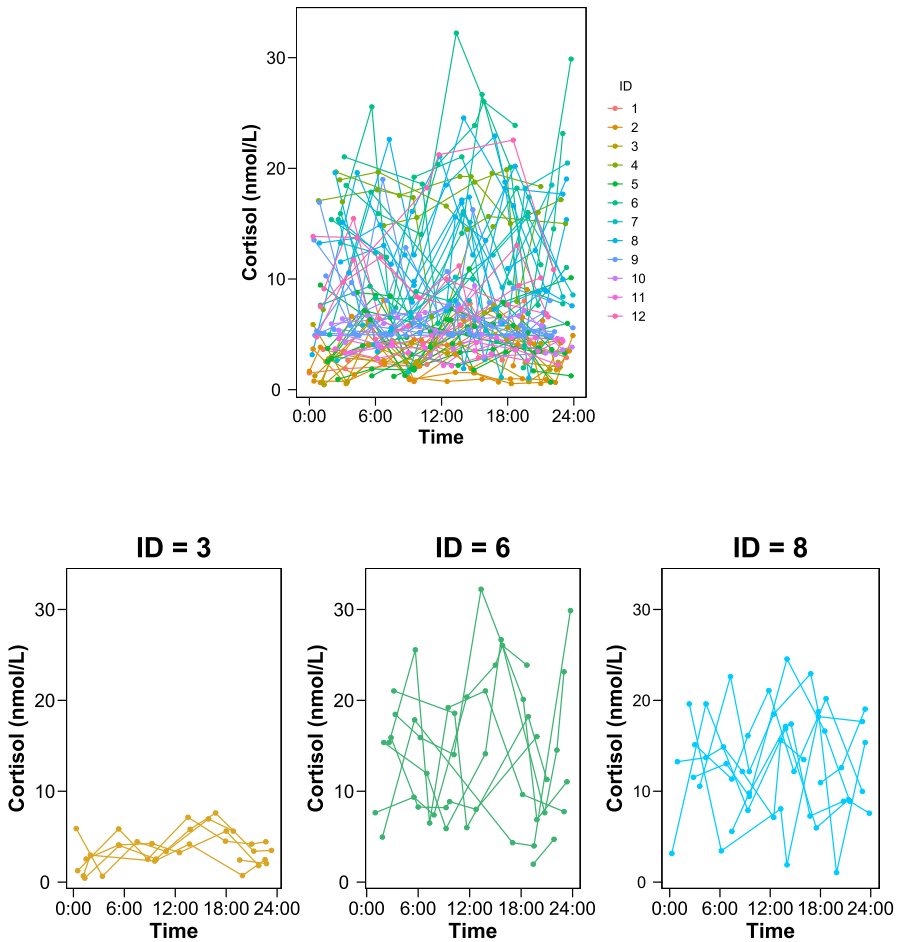
Fig. 6 Line graph of cortisol secretion over time with the lines color-coded by subject. The upper graph shows all of the infants, and the three lower graphs show individual infants as examples. Some patterns, such as ID3, stay at a low value, while others, such as ID6 and ID8, are highly variable

The settings of the BNQM used in the analysis are as follows:

$$Q_{y_{ij}}(\tau|t_{ij}) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(t_{ij}) + \boldsymbol{b}_i^\top \boldsymbol{\psi}(t_{ij}),$$
$$\beta_k \sim GP(0, k_f(s_k, s_l)), k = 1, \ldots, p,$$
$$b_{ik} \sim GP(0, k_r(s_k, s_l)),$$
$$k_f(s_k, s_l|\alpha_f, \rho_f) = {\alpha_f}^2 \exp\left\{-\frac{1}{2{\rho_f}^2}(s_k - s_l)^2\right\},$$
$$k_r(s_k, s_l|\alpha_r, \rho_r) = {\alpha_r}^2 \exp\left\{-\frac{1}{2{\rho_r}^2}(s_k - s_l)^2\right\}, \tag{24}$$
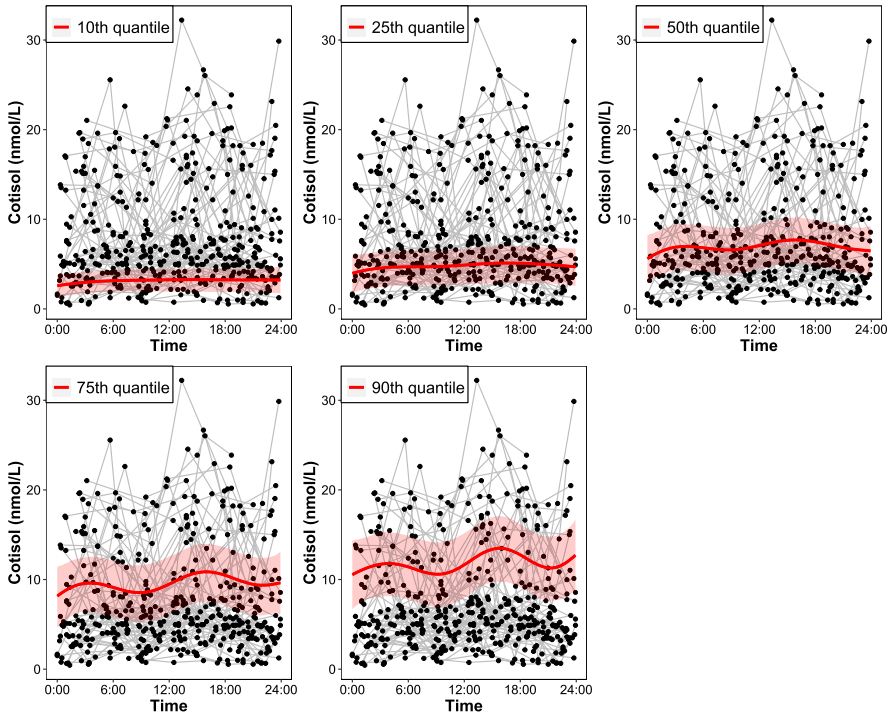
Fig. 7 Estimation curves and 95% credible intervals

where the response variable y_{ij} is the urinary cortisol concentration at the j-th measurement point of the i-th subject, and t_{ij} is the corresponding measurement time. Note that y_{ij} and t_{ij} were normalized to have a maximum value of 1 and a minimum value of 0 before they were analyzed. Specifically, the normalized value x_i^* is expressed by the following formula, where x is the data to be normalized and x_i is the i-th value of x:

$$x_i^* = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \tag{25}$$

Then, the result was determined by converting inversely. Again, $p = 15$, $\alpha_f \sim N_+(0, 1)$, $\alpha_r \sim N_+(0, 1)$, $\rho_f \sim IG(6.28, 1.35)$, and $\rho_r \sim IG(6.28, 1.35)$, as described in appendix A. As the weakly informative prior for the scale parameter σ of the asymmetric Laplace distribution, the half-normal distribution $\sigma \sim N_+(0, 1)$ was used.

Cortisol has a role as a biomarker of biological rhythms (Weitzman et al. 1971). After a biological rhythm has been established, the secretion rhythm of cortisol is monophasic, with the highest level in the morning and the lowest level around midnight (Krieger et al. 1971). However, the subjects in this study were infants between 31 and 124 days old, which is an ambiguous period when biological rhythms are beginning to

be established (Kinoshita et al. 2016; Kidd et al. 2005). Therefore, we analyzed their circadian rhythms using the proposed model and considered the results.

Figure 7 shows the 10th, 25th, 50th, 75th, and 90th quantile curves estimated by the BNQM. Furthermore, the 95% credible interval for the estimated curve of each quantile was determined and is included in Fig. 7. As shown, cortisol secretion tends to increase twice, in the morning and evening. This suggests that the cortisol secretion rhythm of the infants in this study was bimodal. This result could be obtained due to the use of the basis expansion method. In addition, the range of variation increases with the quantile. In other words, at higher cortisol levels, the variation in cortisol level is greater. Identifying this trend is a contribution of quantile regression.

5 Conclusion

In this study, we proposed Bayesian nonparametric quantile mixed-effects models (BNQM), which enable quantile regression considering the hierarchical structure of data by mixed-effects modeling. Furthermore, using a basis expansion method for both the fixed-effects term and random-effects term, it is possible to deal with cases in which the nonlinear structure differs greatly between subjects.

In the mixed-effects model based on the basis expansion method, it is necessary to optimize the combination of the number of basis functions and the regularization parameters, which is usually difficult from the viewpoint of computational complexity. Therefore, for BNQMs, we proposed using a new Bayesian regularization method, in which the coefficient parameters of the basis functions are assumed to follow a Gaussian process (GP) prior. Using this regularization method, BNQMs can appropriately adjust the smoothness of the regression curves even when the number of basis functions is too large for the complexity of the nonlinear structure. Using BNQMs can also reduce the burden of selecting the number of basis functions. Moreover, because BNQMs are based on a fully Bayesian method, they do not require repeated point estimation for determining the optimum value from candidate values of the hyperparameter of the GP prior by comparison using an information criterion. The proposed method also has the advantage that a credible interval can be estimated directly, instead of the confidence interval being approximated by the bootstrap method, which is an iterative method.

The performance of BNQMs was evaluated by a Monte Carlo simulation. The proposed BNQMs showed the highest estimation accuracy for each data structure and were shown to be useful as a quantile regression technique in hierarchical data with a nonlinear structure. Then, a BNQM was applied to longitudinal data of cortisol in infants. The results suggested that the cortisol secretion rhythm in infancy is bimodal, and the magnitude of the amplitude increases as the cortisol level itself increases.

Finally, we discuss future work. In this study, we adopted the asymmetric Laplace distribution as the working likelihood because it is the most commonly used. However, studies that include distributions of the SKD family, mentioned in Galarza et al. (2017) and Wichitaksorn et al. (2014), as a working likelihood option would be useful for further improving the estimation accuracy of the model. Also, in this study, the credible interval was obtained directly from the posterior distribution, whereas

it is inherently ideal to adjust the posterior variance for cases where the asymmetric Laplace distribution is incorrectly specified, as argued by Yang et al. (2016), because the posterior inference would not be asymptotically valid. Discussions of the choice of working likelihoods and the associated asymptotic validity of the posterior distribution such as these are beyond the scope of the present paper, but certainly an interesting topic for future research. The next step could also be to go beyond simple curve fitting and extend the method to additive or semi-parametric models with more terms added to the current model in order to adjust for confounding.

A Setting of hyperpriors

Here, we describe the method of setting the prior for α and ρ, which are parameters of a GP prior, based on Gelman et al. (2020) and Betancourt (2020).

A.1 Prior predictive checks

In this study, when using a GP prior for the coefficient parameter vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)^\top$ of the basis function $\boldsymbol{\phi}(t) = (\phi_1(t), \cdots, \phi_m(t))^\top$, the following RBF kernel was used as the kernel function:

$$k(s_k, s_l | \alpha, \rho) = \alpha^2 \exp\left\{-\frac{1}{2\rho^2}(s_k - s_l)^2\right\}. \tag{26}$$

By assuming the priors of the hyperparameters α and ρ ($\alpha, \rho > 0$) defined in the RBF kernel, the posteriors of α and ρ can be estimated by the MCMC method.

The hyperparameter α determines the amplitude of the sampled function $f(t)$, and ρ determines the smoothness of $f(t)$, where $f(t) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(t)$. Note, however, that it is not $f(t)$ but $\boldsymbol{\beta}$ that is sampled directly from the GP prior. Figure 8 shows the change in amplitude of the function $f(t)$ due to changes in the value of α, and Fig. 9 shows the change in smoothness of the function $f(t)$ due to changes in ρ. For these prior predictive checks, spline-based Gaussian basis functions are used as basis functions, and the number of basis functions is $m = 15$. Each sample $t_i (i = 1, \cdots, 500)$ generated from the uniform distribution $U(0, 1)$ was used as the input of the basis functions.

The results of Figs. 8 and 9 can be used to select a prior for hyperparameters in a GP, and Gelman et al. (2020) notes that these prior predictive checks constitute a useful method for understanding the effects of priors.

Here, the above setting of the range of input t of $0 \leq t \leq 1$ corresponds to the range of time points in the numerical experiment in Sect. 3 and the analysis of infant cortisol data in Sect. 4. In other words, the prior predictive check of this section is the prior predictive check for the numerical experiment in Sect. 3 and the analysis of infant cortisol data in Sect. 4. (In the analysis of infant cortisol data, normalization

processing was performed in advance so that data with a maximum value of 1 and minimum value 0 were obtained.)

A.2 Select of prior for α

In this study, we used the following prior for α:

$$\alpha \sim N_+(0, \sigma_\alpha^2),$$

where $N_+(0, \sigma_\alpha^2)$ is a half-normal distribution. In particular, we used $\sigma_\alpha = 1$ in the numerical experiment (Sect. 3) and the analysis of infant cortisol data (Sect. 4). Figure 10 shows the probability density function of $\alpha \sim N_+(0, 1)$.

From Fig. 10, it can be inferred that the prior $\alpha \sim N_+(0, 1)$ gives the prior information that the value of α will be approximately in the range less than 2. From the amplitudes of the estimated curves for the case of $0 < \alpha < 2$ in Fig. 8, we consider that this is a weakly informative assumption that is appropriate for samples of numerical experiments and infant cortisol data (after normalization).

A.3 Select of prior for ρ

In this study, we used the following prior for ρ, referring to Betancourt (2020):

$$\rho \sim IG(g_a, g_b), \tag{27}$$

where $IG(g_a, g_b)$ is the inverse gamma distribution. In the GP, if the value of ρ is too small, overfitting occurs, and if the value of ρ is too large, non-identifiability occurs. Therefore, the inverse gamma distribution is suitable because it can suppress both the upper and lower limits of ρ. In particular, the Monte Carlo simulation in Sect. 3 and the analysis of infant cortisol data in Sect. 4 used $g_a = 6.28$ and $g_b = 1.35$, respectively, and the method for setting these values is described in the following, based on Betancourt (2020). The prior predictive check shown in Fig. 9 confirms the smoothness of the function for various values of ρ, and we can use this prior predictive check to determine the lower l and upper u limits for each value of ρ. In this study, $l = 0.1$ and $u = 0.7$ were used. Betancourt (2020) expressed l as the lower limit and u as the upper limit using the lower probability and upper probability as

$$\int_0^{l=0.1} p(\rho|g_a, g_b)\mathrm{d}\rho = 0.01,$$
$$\int_{u=0.7}^{\infty} p(\rho|g_a, g_b)\mathrm{d}\rho = 0.01. \tag{28}$$

The parameters simultaneously satisfying these two conditions are defined as g_a and g_b. By solving this optimization problem of the simultaneous equations, $g_a \approx 6.28$ and $g_b \approx 1.35$ are obtained.
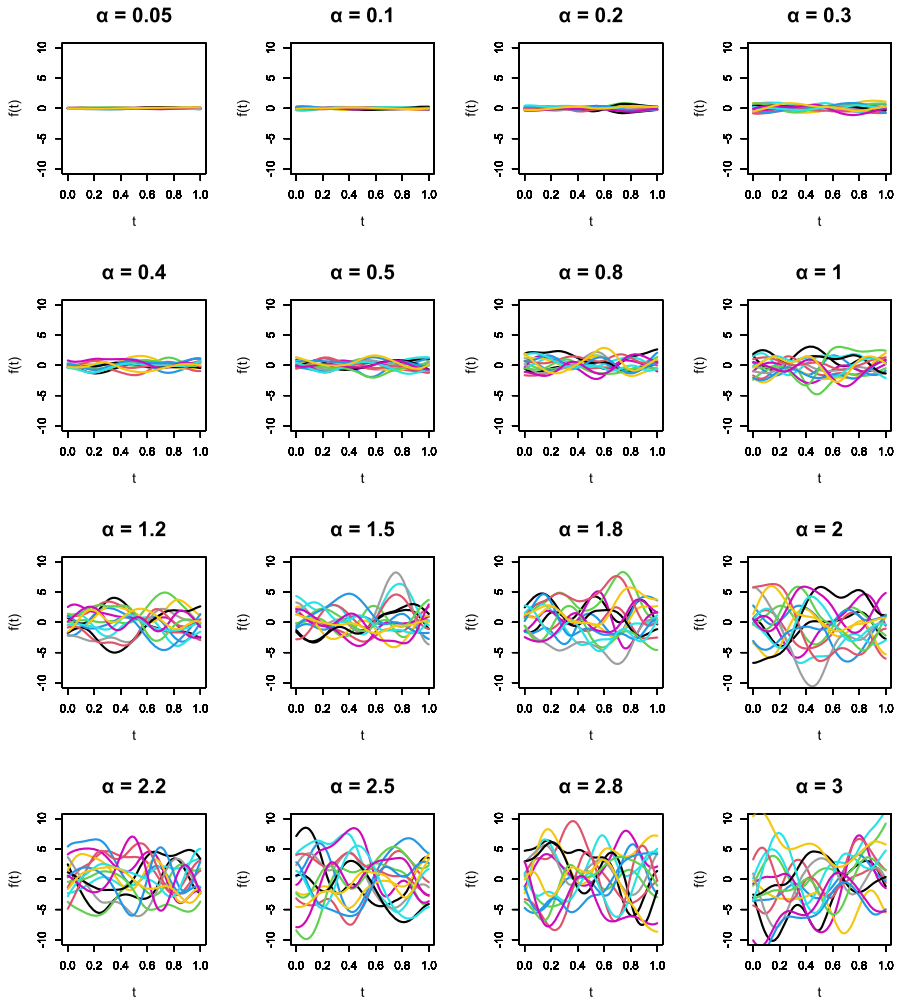
Fig. 8 Change in the sampled functions $f(t) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(t)$ for different values of α for $\rho = 0.1$

B HMC and NUTS

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{b}^\top, \alpha_f, \alpha_r, \rho_f, \rho_r, \sigma, \boldsymbol{v}^\top)^\top$ be the vector of the unknown parameters in a BNQM (see equation (15)). Using $\boldsymbol{\theta}$, the posterior can be rewritten as $p(\boldsymbol{\theta}|\boldsymbol{y})$. It is necessary to estimate the unknown parameter vector $\boldsymbol{\theta}$ in the posterior inference. Here, we summarize the algorithm of HMC and NUTS when there is a d-dimensional unknown parameter $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)^\top$ for any Bayesian model, based on Hoffman and Gelman (2014) and the Stan reference manual (Stan Development Team 2020).
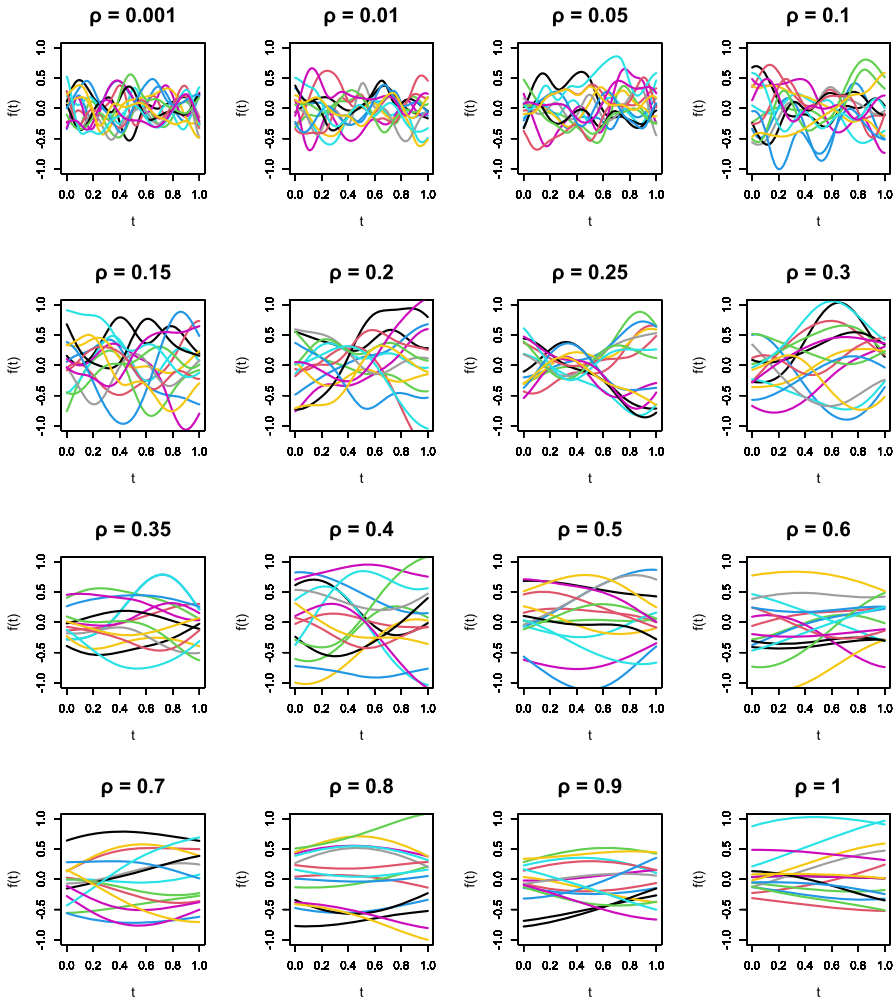
Fig. 9 Change in the sampled functions $f(t) = \boldsymbol{\beta}^\top \boldsymbol{\phi}(t)$ for different values of ρ for $\alpha = 0.2$

B.1 HMC

Step 1: Initial setting

Setting for $t = 1$. Initialize the parameters $\boldsymbol{\theta}^{(1)}$ and set ϵ, L, $\boldsymbol{\Sigma}$. (Here, ϵ and L are the width of one small discrete transition and the number of repetitions in the leapfrog integrator (step 3), respectively, and $\boldsymbol{\Sigma}$ is the covariance matrix of the multivariate normal distribution used to generate random samples.)

Step 2: Random sample generation

Draw $\boldsymbol{\rho}^{(t)}$ from the following d-dimensional multivariate normal distribution:

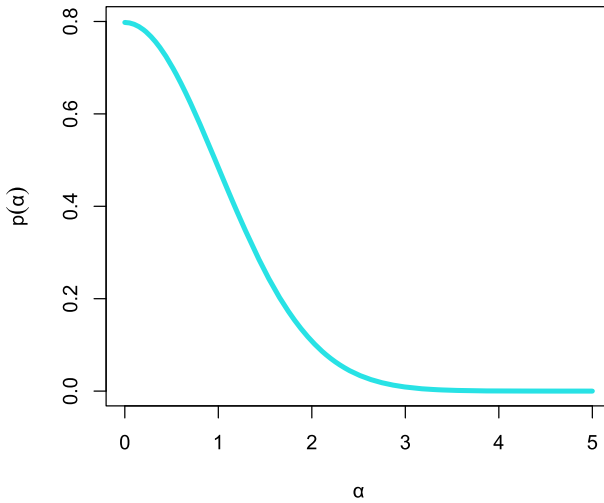$$\boldsymbol{\rho}^{(t)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}).$$

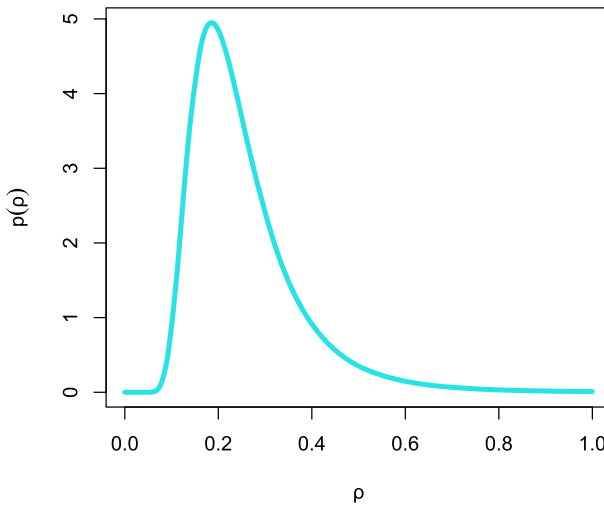**Fig. 10** Half-normal distribution $N_+(0, 1)$

Fig. 11 Inverse gamma distribution $IG(6.28, 1.35)$

Step 3: Leapfrog Integrator

Set $\rho = \rho^{(t)}, \theta = \theta^{(t)}$ and repeat the following updates L times:

$$\rho = \rho - \frac{\epsilon}{2} \frac{\partial - \log p(\theta|y)}{\partial \theta},$$
$$\theta = \theta + \epsilon \Sigma \rho$$
$$\rho = \rho - \frac{\epsilon}{2} \frac{\partial - \log p(\theta|y)}{\partial \theta}.$$

Then, denote the final θ and ρt by θ^* and ρ^*, respectively.

Step 4: Metropolis accept step

> Accept the candidate ($\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$) with the following probability and otherwise maintain the current state ($\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$):

$$\min(1, r),$$

> where $r = \exp\{\log(\boldsymbol{\rho}, \boldsymbol{\theta}) - \log(\boldsymbol{\rho}^*, \boldsymbol{\theta}^*)\}$.

Step 5: Determine whether to continue HMC

> If $t = T$ (where T is the number of HMC iterations.), end sampling; otherwise set $t = t + 1$ and return to step 2.

B.2 NUTS

The HMC algorithm in the previous section has parameters ϵ, L, and $\boldsymbol{\Sigma}$, which need to be set and affect the sampling efficiency. The advantage of the HMC algorithm is that the average transition distance can be increased.

For the same L, increasing ϵ increases the transition distance in the leapfrog integrator but decreases the acceptance rate. Increasing L and decreasing ϵ increases the transition distance and acceptance rate but increases the computational cost. Depending on the value of L, the transition may make a U-turn, resulting in a shorter travel distance. For such situations, Hoffman and Gelman (2014) proposed using the NUTS algorithm, an extension of the HMC algorithm. Their proposed algorithm uses half the squared distance between the current parameter $\boldsymbol{\theta}$ and the candidate point $\boldsymbol{\theta}^*$ to determine whether a transition makes a U-turn:

$$\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta})^\top(\boldsymbol{\theta}^* - \boldsymbol{\theta}).$$

Specifically, the criterion that the first derivative with respect to time t of half the squared distance becomes less than 0 (meaning that half the squared distance does not increase even if the number of updates L is increased) is used:

$$\frac{d}{dt}\frac{1}{2}(\boldsymbol{\theta}^* - \boldsymbol{\theta})^\top(\boldsymbol{\theta}^* - \boldsymbol{\theta}) = (\boldsymbol{\theta}^* - \boldsymbol{\theta})^\top\frac{d}{dt}(\boldsymbol{\theta}^* - \boldsymbol{\theta}) = (\boldsymbol{\theta}^* - \boldsymbol{\theta})^\top\boldsymbol{\rho}.$$

Thus, the NUTS algorithm can automatically set L. For more details on the algorithm, see Hoffman and Gelman (2014) and the Stan reference manual (Stan Development Team 2020).

References

Betancourt, M. (2020). Robust gaussian process modeling. https://betanalpha.github.io/assets/case_studies/gaussian_processes.html. Accessed on 30 Jan 2021

de Boor, C. (2001) A practical guide to splines; rev. ed. Applied Mathematical Sciences, Springer, Berlin. https://cds.cern.ch/record/1428148

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B, 195*(2), 216–222.

Fenske, N., Fahrmeir, L., Hothorn, T., Rzehak, P., & Höhle, M. (2013). Boosting structured additive quantile regression for longitudinal childhood obesity data. *The International Journal of Biostatistics, 9*(1), 1–18.

Galarza, C., Lachos Davila, V., Barbosa Cabral, C., & Castro Cepero, L. (2017). Robust quantile regression using a generalized class of skewed distributions. *Statistics, 6*(1), 113–130.

Galarza, C. E., Castro, L. M., Louzada, F., & Lachos, V. H. (2020). Quantile regression for nonlinear mixed effects models: A likelihood based perspective. *Statistical Papers, 61*(3), 1281–1307.

Gelman, A., Vehtari, A., Simpson, D., Margossian, CC., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, PC., & Modrák, M. (2020) Bayesian workflow. arXiv preprint arXiv:2011.01808

Geraci, M. (2019). Additive quantile regression for clustered data with an application to children's physical activity. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 68*(4), 1071–1089.

Geraci, M. (2019). Modelling and estimation of nonlinear quantile regression with clustered data. *Computational Statistics & Data Analysis, 136,* 30–46.

Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics, 8*(1), 140–154.

Geraci, M., & Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing, 24*(3), 461–479.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.

Kawano, S., & Konishi, S. (2007). Nonlinear regression modeling via regularized gaussian basis functions. *Bulletin of Informatics and Cybernetics, 39,* 83.

Kidd, S., Midgley, P., Nicol, M., Smith, J., & McIntosh, N. (2005). Lack of adult-type salivary cortisol circadian rhythm in hospitalized preterm infants. *Hormone Research in Pædiatrics, 64*(1), 20–27.

Kinoshita, M., Iwata, S., Okamura, H., Saikusa, M., Hara, N., Urata, C., et al. (2016). Paradoxical diurnal cortisol changes in neonates suggesting preservation of foetal adrenal rhythms. *Scientific Reports, 6,* 35553.

Koenker, R. & Bassett, Jr G. (1978) Regression quantiles. *Econometrica: Journal of the Econometric Society*, *46*(1), 33–50.

Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation, 81*(11), 1565–1578.

Krieger, D. T., Allen, W., Rizzo, F., & Krieger, H. P. (1971). Characterization of the normal temporal pattern of plasma corticosteroid levels. *The Journal of Clinical Endocrinology & Metabolism, 32*(2), 266–284.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics, 46*(3), 673–687.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics, 4*(1), 12–35.

Stan Development Team (2020) Stan modeling language users guide and reference manual, version 2.25.0. http://mc-stan.org/

Takeuchi, I., Le, Q. V., Sears, T. D., & Smola, A. J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research, 7*(Jul), 1231–1264.

Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., & Flexeder, C. (2013). Bayesian semiparametric additive quantile regression. *Statistical Modelling, 13*(3), 223–252.

de Weerth, C., Zijl, R. H., & Buitelaar, J. K. (2003). Development of cortisol circadian rhythm in infancy. *Early Human Development, 73*(1–2), 39–52.

Weitzman, E. D., Fukushima, D., Nogeire, C., Roffwarg, H., Gallagher, T. F., & Hellman, L. (1971). Twenty-four hour pattern of the episodic secretion of cortisol in normal subjects. *The Journal of Clinical Endocrinology & Metabolism, 33*(1), 14–22.

Wichitaksorn, N., Choy, S. B., & Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *Canadian Journal of Statistics, 42*(4), 579–596.

Yang, Y., Wang, H. J., & He, X. (2016). Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *International Statistical Review, 84*(3), 327–344.

Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters, 54*(4), 437–447.

Yue, Y. R., & Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics & Data Analysis, 55*(1), 84–96.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.