Posterior Computation

The likelihood for the logistic regression model is analytically inconvenient and not amenable to easy posterior inference. A naive MCMC strategy to fit logistic models is via the use of Metropolis-Hastings algorithm. Such methods, however, suffer from slow mixing unless efficient problem-specific proposal distributions are used. Efficient auxiliary variable augmentation strategies have been proposed to address these issues (Holmes & Held, 2006; Polson, Scott, & Windle, 2013). In this work we employed scheme proposed in (Polson et al., 2013), introducing Pòlya-Gamma auxiliary variables $\omega_{it} \sim PG\{n_i(t),\ \mu(t) + \xi(t)\,\eta_i\}$. Conditionally on these variables, the likelihood is normal, the likelihood contribution of the i^{th} individual being

$$L_i(\boldsymbol{\beta}, \eta_i, \boldsymbol{\gamma}, \boldsymbol{\omega}) \propto \prod_{t=1}^{T} \left\{ e^{\kappa_{it}(\boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{x}_{it}\boldsymbol{\gamma}\eta_i)}\ e^{-\frac{\omega_{it}}{2}(\boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{x}_{it}\boldsymbol{\gamma}\eta_i)^2} \right\},$$

where $\kappa_{it} = y_i(t) - \frac{1}{2}n_i(t)$. We denote with $\boldsymbol{x}_{it}$ the J B-spline functions evaluated at the t^{th} point for individual i. Given this likelihood, all full conditional distributions can be obtained in closed form. It is thus straightforward to perform posterior inference using a Gibbs sampling algorithm (Gelfand & Smith, 1990; Geman & Geman, 1984) comprising the following steps.

1. For the spline coefficients $\boldsymbol{\beta}$ of the latent mean function $\mu(t)$,

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \sigma_\beta^2, \eta_1, \dots, \eta_n \sim MVN_J\left\{ \left(\frac{P}{\sigma_\beta^2} + \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X} \right)^{-1} \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{r}, \left(\frac{P}{\sigma_\beta^2} + \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X} \right)^{-1} \right\},$$

where $\boldsymbol{\Omega} = diag\{\omega_{it}\}$ and the vector $\boldsymbol{r} = \{r_{it}\}_{i=1,t=1}^{n,T}$, $r_{it} = \frac{\kappa_{it}}{\omega_{it}} - \boldsymbol{x}_{it}^T\,\boldsymbol{\gamma}\,\eta_i$ and $\kappa_{it} = y_i(t) - \frac{1}{2}n_i(t)$.

2. For the smoothness inducing penalty parameter σ_β^2,

$$\sigma_\beta^2 \mid \boldsymbol{\beta} \sim Inv - Ga\left(a_\beta + \frac{J}{2}, b_\beta + \frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{P}\boldsymbol{\beta} \right).$$

3. For the spline coefficients $\boldsymbol{\gamma}$ of the latent mean function $\xi(t)$,

$$\boldsymbol{\gamma} \mid \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\omega}, \sigma_\gamma^2, \eta_1, \dots, \eta_n \sim MVN_J\left\{ \left(\frac{P}{\sigma_\gamma^2} + \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X} \right)^{-1} \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{r}', \left(\frac{P}{\sigma_\gamma^2} + \boldsymbol{X}^T\boldsymbol{\Omega}\boldsymbol{X} \right)^{-1} \right\},$$

where $\boldsymbol{\Omega} = diag\{\omega_{it}\}$ and the vector $\boldsymbol{r}' = \{r'_{it}\}_{i=1,t=1}^{n,T}$, $r'_{it} = \frac{\kappa_{it}}{\eta_i\,\omega_{it}} - \frac{\boldsymbol{x}_{it}^T\boldsymbol{\beta}}{\eta_i}$ and $\kappa_{it} = y_i(t) - \frac{1}{2}n_i(t)$.

4. For the smoothness inducing penalty parameter σ_γ^2,

$$\sigma_\gamma^2 \mid \boldsymbol{\gamma} \sim Inv - Ga\left(a_\gamma + \frac{J}{2}, b_\gamma + \frac{1}{2}\boldsymbol{\gamma}^T\boldsymbol{P}\boldsymbol{\gamma} \right).$$

5. For the random effects distributions η_i,

$$\eta_i \mid \boldsymbol{y}, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\eta^2 \sim Normal\left[\frac{\sum_{t=1}^{T} \kappa_{it}\, \boldsymbol{x}_{it}^T\, \boldsymbol{\gamma} - \boldsymbol{\omega}_{it}\, \boldsymbol{x}_{it}^T\, \boldsymbol{\gamma}\, \boldsymbol{x}_{it}^T\, \boldsymbol{\beta}}{\frac{1}{\sigma_\eta^2} + \sum_{i=1}^{T} \omega_{it}\, (\boldsymbol{x}_{it}^T\, \boldsymbol{\gamma})^2}, \left\{\frac{1}{\sigma_\eta^2} + \sum_{i=1}^{T} \omega_{it}\, (\boldsymbol{x}_{it}^T\, \boldsymbol{\gamma})^2\right\}^{-1}\right].$$

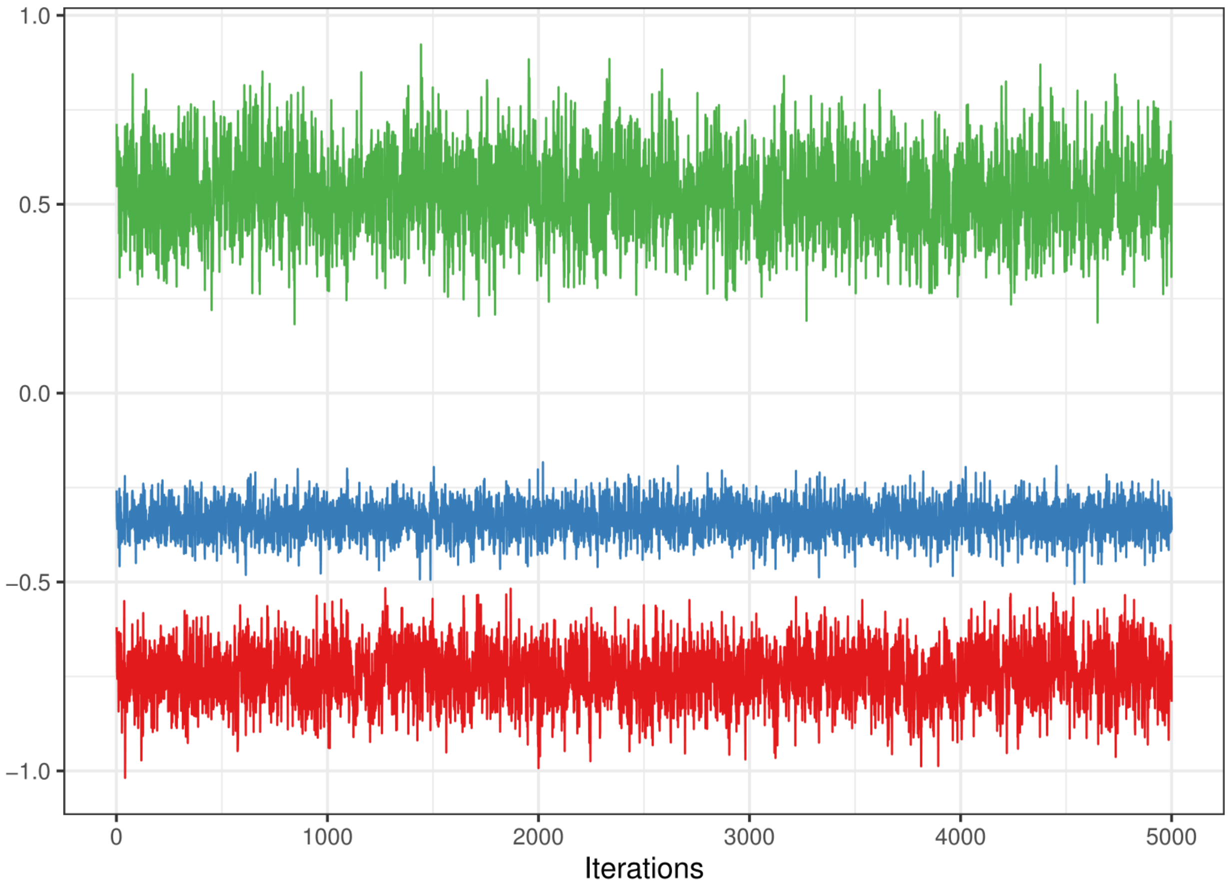
6. For the random variance term σ_η^2,

$$\sigma_\eta^2 \mid \eta_1, \dots \eta_n \sim Inv - Ga\left(\frac{a_\eta + n}{2}, \frac{a_\eta + \sum_{i=1}^{n} \eta_i^2}{2}\right).$$

Convergence Diagnostics

This section presents some MCMC convergence diagnostics for the Gibbs sampler described in the main manuscript. The results presented here are for the speech learning data set. Diagnostics for the simulation experiments were similar and hence omitted. Figure S.1 shows the trace plots of the sampled values of three individual level curves at the same time point $t_0 = 1$. The Geweke test (Geweke, 1992) for stationarity of the chains, which formally compares the means of the first and last part of a Markov chain (by default the first 10% and the last 50%), is also performed. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution. The test indicated that convergence was satisfactory.

Supplemental Material S1. Traceplots of the sampled values of three individual curves $\mu_i(t_0)$, (denoted by the different colors) at the same time point $t_0 = 1$.



Additional Comments on Implementation Issues

It is possible to reformulate the proposed spline based as

$$\ln \frac{\pi_i(t)}{1-\pi_i(t)} = \boldsymbol{B}_{q,J}^T(t)\,\boldsymbol{\beta} + \boldsymbol{B}_{q,J}^T(t)\,\boldsymbol{\gamma}\,\eta_i,$$

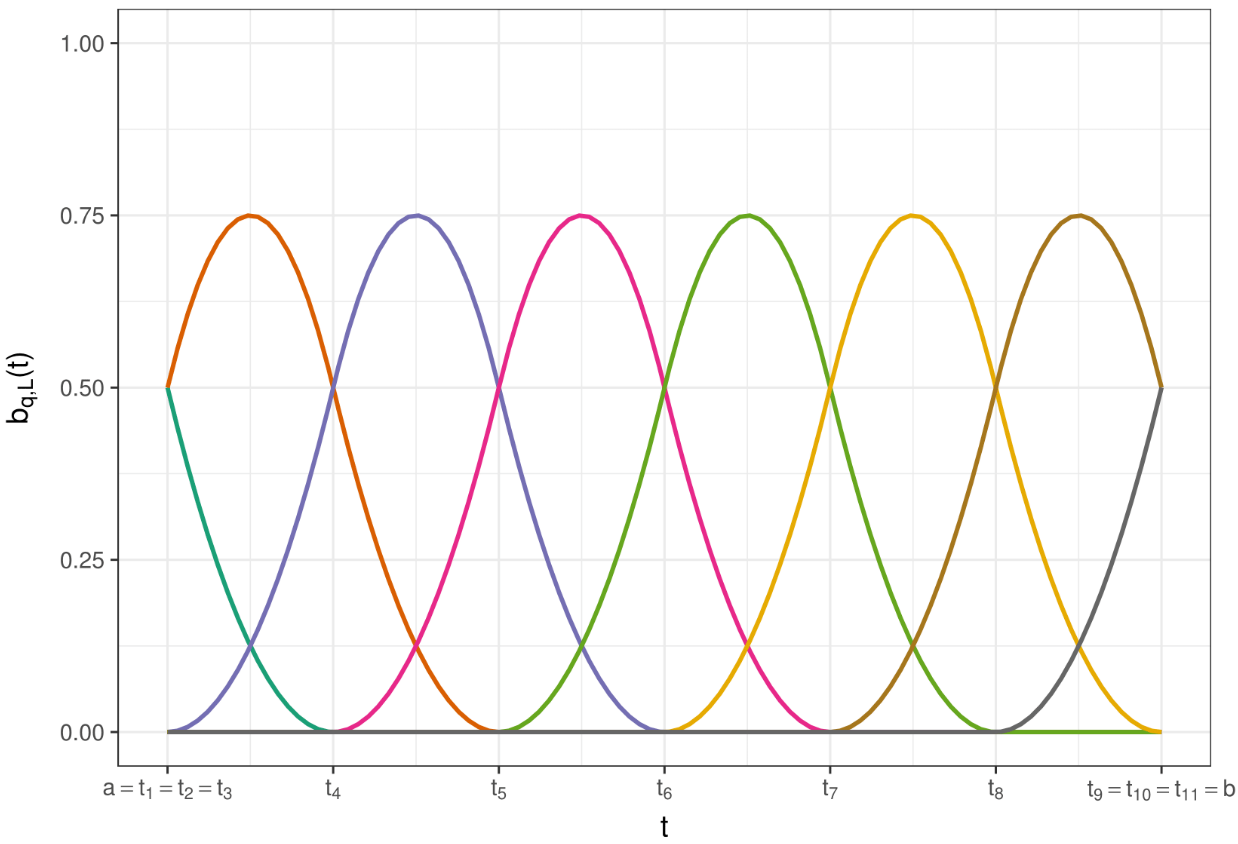
which can be identified with a traditional logistic linear mixed model as

$$\ln \frac{\pi_i(t)}{1-\pi_i(t)} = \boldsymbol{X}(t)^T\,\boldsymbol{\beta} + \boldsymbol{Z}(t)^T\,\boldsymbol{u}_i,$$

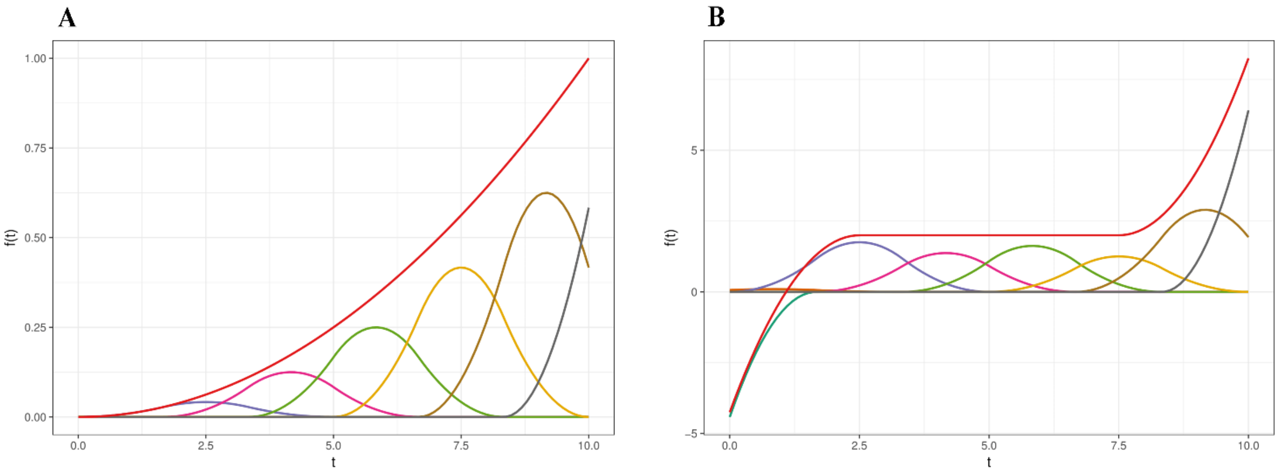
where $\boldsymbol{X}(t)^T = \boldsymbol{B}_{q,J}^T(t)$ are covariates associated with fixed effects parameters $\boldsymbol{\beta}$, and $\boldsymbol{Z}(t)^T = \boldsymbol{B}_{q,J}^T(t)$ are covariates associated to the random effects $\boldsymbol{u}_i = \boldsymbol{\gamma}\,\eta_i$. The model could then be fitted using the `glmer` package in `R`. As also mentioned in the main paper, our attempts fit such complex models with `glmer` was fraught with serious convergence issues.

Additional Figures

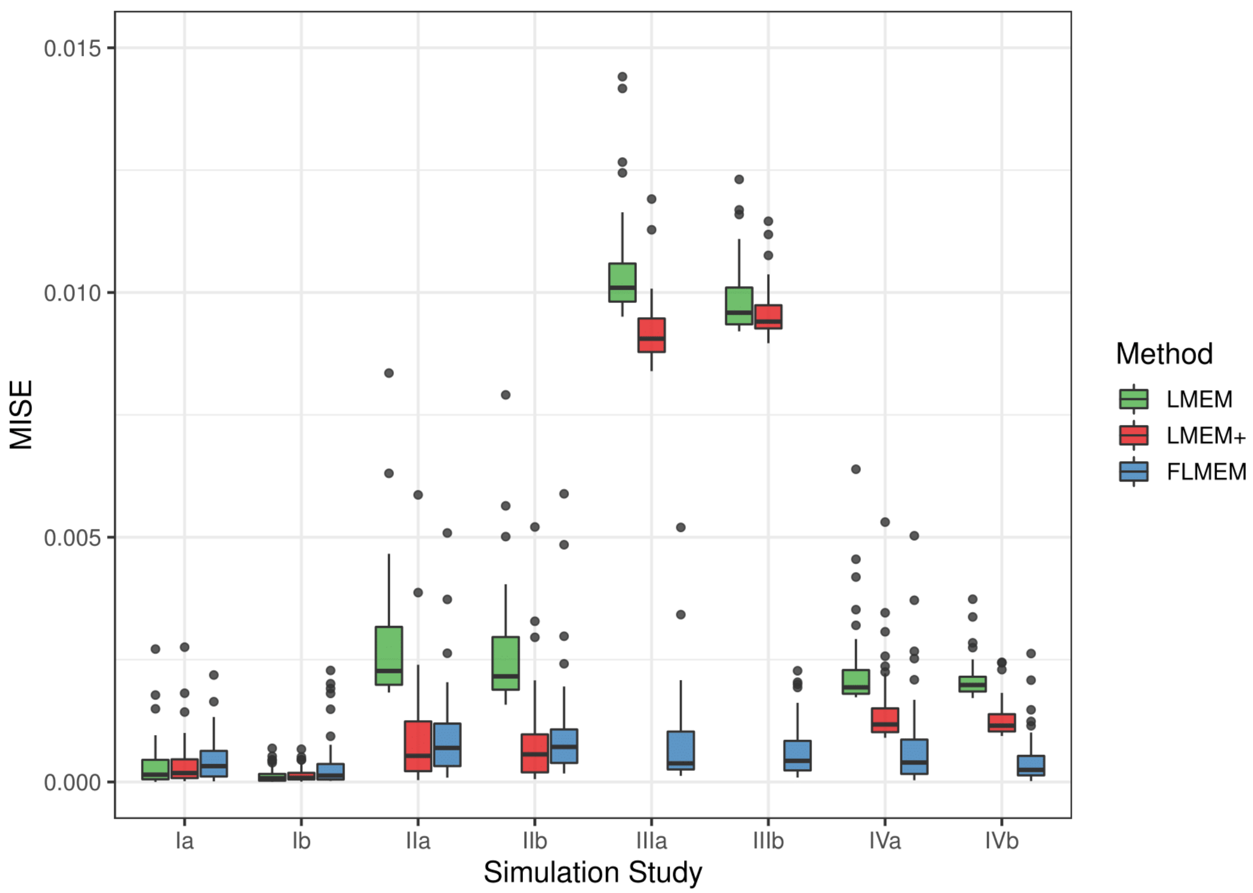
Supplemental Material S2. Plot of 8 quadratic ($q = 2$) B-splines on $[a, b]$ defined using 11 knot points that divide $[a, b]$ into $K = 6$ equal subintervals.



Supplemental Material S3. Flexibility of spline mixtures. The red curve is the function produced by the weighted sum of the spline bases of Supplemental Material S2, weighted by coefficients $\boldsymbol{\beta}$. The weighted splines are shown in other varied colors.



Supplemental Material S4. Distribution of the mean integrated squared errors (MISEs) between the true and the estimated population function $\pi(t)$ estimated by the three models under different simulation scenarios.



References

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*(410), 398–409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F.M. Smith (Eds.), *Bayesian Statistics 4* (pp. 169–193). Oxford, England: Oxford University Press.

Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, *1*(1), 145–168.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349.