Supplementary Materials for "A Generalized Gaussian Process Model for Computer Experiments with Binary Time Series"

S1 Algorithm: Estimation of $(\boldsymbol{\beta}, \boldsymbol{\omega})$

Set initial values $\boldsymbol{\omega} = (\sigma^2, \boldsymbol{\theta}) = \mathbf{1}_{d+1}, \boldsymbol{\beta} = \mathbf{1}_m, p_{it} = 1$, and set $\tilde{\eta}_{it} = \log \frac{p_{it}}{1-p_{it}} + \frac{y_{it}-p_{it}}{p_{it}(1-p_{it})}$ for each i and t.
repeat
    repeat
        Set $\boldsymbol{W}$ as an $N \times N$ diagonal matrix with diagonal elements $W_{it} = p_{it}(1-p_{it})$
        Set $\boldsymbol{V} = \boldsymbol{W}^{-1} + \sigma^2(\boldsymbol{R}_{\boldsymbol{\theta}} \otimes I_T)$
        Update $\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\tilde{\boldsymbol{\eta}}$
        Set $\boldsymbol{Z} = \sigma^2(\boldsymbol{R}_{\boldsymbol{\theta}} \otimes I_T)\boldsymbol{V}^{-1}(\tilde{\boldsymbol{\eta}} - \boldsymbol{X}'\boldsymbol{\beta})$
        Update $p_{it} = \left(\frac{\exp\{\boldsymbol{X}'\boldsymbol{\beta}+\boldsymbol{Z}\}}{\mathbf{1}_N+\exp\{\boldsymbol{X}'\boldsymbol{\beta}+\boldsymbol{Z}\}}\right)_{it}$ and $\tilde{\eta}_{it} = \log \frac{p_{it}}{1-p_{it}} + \frac{y_{it}-p_{it}}{p_{it}(1-p_{it})}$ for each i and t
    until $\{\tilde{\eta}_{it}\}_{it}$ converges
    Update $\boldsymbol{\omega} = \arg\min_{\boldsymbol{\omega}} L(\boldsymbol{\omega})$, where $L(\boldsymbol{\omega})$ is the negative log-likelihood function (12)
    Update $(\sigma^2, \boldsymbol{\theta}) = \boldsymbol{\omega}$
until $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ converge
Return $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$

S2 Assumptions

1. The parameter $\boldsymbol{\beta}$ belongs to an open set $B \subseteq \mathbb{R}^m$ and the parameter $\boldsymbol{\omega}$ belongs to an open set $\Omega \subseteq \mathbb{R}^{d+1}$.

2. The model matrix X_{it} lies almost surely in a nonrandom compact subset of $\mathbb{R}^m$ such that $Pr(\sum_{i=1}^n \sum_{t=1}^T X_{it}'X_{it} > 0) = 1$.

For any matrix A, define $\|A\| \equiv \sqrt{\mathrm{tr}(A'A)}$; for the covariance matrix $\boldsymbol{V}(\boldsymbol{\omega})$, define $V_i(\boldsymbol{\omega}) \equiv \partial \boldsymbol{V}(\boldsymbol{\omega})/\partial \omega_i$ and $V_{ij}(\boldsymbol{\omega}) \equiv \partial \boldsymbol{V}(\boldsymbol{\omega})/\partial \omega_i \partial \omega_j$; for $\boldsymbol{\omega} \in \Omega$, denote $\xrightarrow{u}$ as uniform convergence of nonrandom functions over compact subsets of Ω.

3. $J_N(\boldsymbol{\omega})P_N(\boldsymbol{\omega})^{-1} \xrightarrow{d} W(\boldsymbol{\omega})$ for some nonsingular $W(\boldsymbol{\omega})$, which is continuous in $\boldsymbol{\omega}$, where $P_N(\boldsymbol{\omega}) = \mathrm{diag}(\|\Pi(\boldsymbol{\omega})V_1(\boldsymbol{\omega})\|, \ldots, \|\Pi(\boldsymbol{\omega})V_{d+1}(\boldsymbol{\omega})\|)$ and $\Pi(\boldsymbol{\omega}) = \boldsymbol{V}(\boldsymbol{\omega})^{-1} - \boldsymbol{V}(\boldsymbol{\omega})^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\omega})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\omega})^{-1}$.

4. If there exists a sequence $\{r_N\}_{N\geq 1}$ with $\limsup_{N\to\infty} r_N/N \leq 1-\delta$, for some $\delta \in (0,1)$, such that for any compact subset $K \subseteq \Omega$, there exist constants $0 < C_1(K) < \infty$ and $C_2(K) > 0$ such that

$$\lim \sup_{N\to\infty} \max\{|\lambda_N|, |\lambda_N^i|, |\lambda_N^{ij}| : 1 \leq i, j \leq k\} < C_1(K) < \infty$$

and

$$\lim \sup_{N\to\infty} \min\{|\lambda_1|, |\lambda_{r_N}^i| : 1 \leq i \leq k\} > C_2(K) > 0,$$

uniformly in $\boldsymbol{\omega} \in K$, where $|\lambda_1| \leq \ldots \leq |\lambda_N|$ are the absolute eigenvalues of $\boldsymbol{V}(\boldsymbol{\omega})$, $|\lambda_1^i| \leq \ldots \leq |\lambda_N^i|$ are the absolute eigenvalues of $V_i(\boldsymbol{\omega})$, and $|\lambda_1^{ij}| \leq \ldots \leq |\lambda_N^{ij}|$ are the absolute eigenvalues of $V_{ij}(\boldsymbol{\omega})$.

Assumption 2 holds when the row vectors of $\boldsymbol{X}$ are linear independent. Thus, if only the linear effect is considered in the mean function, then orthogonal designs or orthogonal array-based designs, such as OA-based Latin hypercube designs (Tang, 1993), can be chosen for sampling schemes. The conditions for Assumption 4 can be referred to Cressie and Lahiri (1996), in which the checkable conditions for rectangular lattice of data sites and irregularly located data sites are given. For instance, for rectangular lattice of data sites, with certain correlation functions, a sufficient condition is choosing data locations whose minimum distance is sufficiently large. More details can be seen in Cressie and Lahiri (1996). Thus, space-filling designs, such as Latin hypercube designs (McKay et al., 1979) and maximin distance designs (Johnson et al., 1990), can be chosen for sampling schemes.

S3 Proof of Theorem 3.1

The model (4) can be seen as a binary time series model with random effects by multiplying an identity matrix on $\boldsymbol{Z}$, that is,

$$\text{logit}(\boldsymbol{p}) = \boldsymbol{X}\boldsymbol{\beta} + I_N \boldsymbol{Z}, \quad \boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_N, \Sigma(\boldsymbol{\omega})),$$

where I_N and $\boldsymbol{Z}$ are viewed as the model matrix and coefficients of random effects, respectively. Therefore, if the variance-covariance parameters are given, inference of $\boldsymbol{\beta}$ is a special case of the binary time series model with random effects in Hung et al. (2008). Therefore, following Theorem 1 in Hung et al. (2008), the score function $S_N(\boldsymbol{\beta}, \boldsymbol{\omega})$ is asymptotically normally distributed.

S4 Proof of Theorem 3.3

According to Breslow and Clayton (1993), one can view the inference on the variance-variance component as an iterative procedure for the linear mixed model

$$\tilde{\boldsymbol{\eta}} = \boldsymbol{X}\boldsymbol{\beta} + I_N \boldsymbol{Z} + \epsilon, \quad \epsilon \sim \mathcal{N}(\boldsymbol{0}_N, \boldsymbol{W}^{-1})$$

with the iterative weight $\boldsymbol{W}^{-1}$. Thus, it is a special case of the Gaussian general linear model in Cressie and Lahiri (1993) with response vector $\tilde{\boldsymbol{\eta}}$ and variance-covariance component $\Sigma(\boldsymbol{\omega}) + \boldsymbol{W}^{-1}$ with parameters $\boldsymbol{\omega}$. Since the asymptotic distribution of REML estimators for the variance-covariance parameters has been shown in Cressie and Lahiri (1993) for a Gaussian general linear model, the result directly follows as a special case of Corollary 3.3 in Cressie and Lahiri (1993). Note that Assumption 4 in the supplementary material S2 implies the conditions for Corollary 3.3 in Cressie and Lahiri (1993). See the proof of Theorem 2.2 in Cressie and Lahiri (1996).

S5 Proof of Lemma 4.1

We start the proof by deriving the conditional distribution from a simple model (1) (without time-series), and then extend the result to prove Lemma 4.1. First, a definition and a lemma about multivariate log-normal distribution are in order.

Definition S5.1. Suppose $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$ has a multivariate normal distribution with mean $\boldsymbol{\mu}_n$ and covariance variance $\boldsymbol{\Sigma}_{n\times n}$. Then $\boldsymbol{b} = \exp\{\boldsymbol{\xi}\}$ has a *multivariate log-normal distribution*. Denote it as $\boldsymbol{b} \sim \mathcal{LN}(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_{n\times n})$.

Lemma S5.1. *Suppose $\boldsymbol{b}^n$ and b_{n+1} have a multivariate log-normal distribution*

$$\begin{pmatrix} \boldsymbol{b}^n \\ b_{n+1} \end{pmatrix} \sim \mathcal{LN}\left(\begin{pmatrix} \boldsymbol{\mu}^n \\ \mu_{n+1} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{n\times n} & \boldsymbol{r} \\ \boldsymbol{r}' & \sigma^2_{n+1} \end{bmatrix} \right).$$

The conditional distribution of b_{n+1} given $\boldsymbol{b}^n$ is $b_{n+1}|\boldsymbol{b}^n \sim \mathcal{LN}(\mu^*, v^*)$, where $\mu^* = \mu_{n+1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}(\log \boldsymbol{b}^n - \boldsymbol{\mu}^n)$ and $v^* = \sigma^2_{n+1} - \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{r}$.

Proof. Using transformation of a standard normal distribution, one can show that the joint probability density function of the multivariate log-normal distribution $\boldsymbol{b}^n$ is

$$g_{\boldsymbol{b}^n}(b_1, \ldots, b_n) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{n\times n}|^{1/2}} \frac{1}{\prod_{i=1}^n b_i} \exp\{-\frac{1}{2}\left(\log \boldsymbol{b}^n - \boldsymbol{\mu}_n\right)' \boldsymbol{\Sigma}^{-1}_{n\times n}\left(\log \boldsymbol{b}^n - \boldsymbol{\mu}_n\right)\}.$$

Denote $\boldsymbol{b}^{n+1} = (b_1, \ldots, b_n, b_{n+1})$, $\boldsymbol{\mu}^{n+1} = (\mu_1, \ldots, \mu_n, \mu_{n+1})$ and

$$\boldsymbol{\Sigma}_{(n+1)\times(n+1)} = \begin{bmatrix} \boldsymbol{\Sigma}_{n\times n} & \boldsymbol{r} \\ \boldsymbol{r}' & \sigma_{n+1} \end{bmatrix}.$$

Then, the conditional probability density function of b_{n+1} given $\boldsymbol{b}^n$ can be derived as

$$\begin{aligned} g_{b_{n+1}|\boldsymbol{b}^n}(b_{n+1}|\boldsymbol{b}^n) &\propto g(b_1,\ldots,b_n,b_{n+1}) \\ &\propto \frac{1}{b_{n+1}}\exp\{-\frac{1}{2}\left(\log \boldsymbol{b}^{n+1}-\boldsymbol{\mu}_{n+1}\right)'\boldsymbol{\Sigma}^{-1}_{(n+1)\times(n+1)}\left(\log \boldsymbol{b}^{n+1}-\boldsymbol{\mu}_{n+1}\right)\}. \end{aligned}$$

Let $\boldsymbol{a}_1 = \log \boldsymbol{b}^n - \boldsymbol{\mu}^n$ and $\boldsymbol{a}_2 = \log b^{n+1} - \mu^{n+1}$. Applying the partitioned matrix inverse results (page 99 of Harville (1997)) gives

$$\begin{aligned} &\left(\log \boldsymbol{b}^{n+1}-\boldsymbol{\mu}^{n+1}\right)'\boldsymbol{\Sigma}^{-1}_{(n+1)\times(n+1)}\left(\log \boldsymbol{b}^{n+1}-\boldsymbol{\mu}^{n+1}\right) \\ =&\begin{bmatrix}\boldsymbol{a}_1' & \boldsymbol{a}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{n\times n} & \boldsymbol{r} \\ \boldsymbol{r}' & \sigma_{n+1}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{a}_1 \\ \boldsymbol{a}_2\end{bmatrix} \\ =&(\boldsymbol{a}_2-\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1)'\sigma^{-1}_{22\cdot 1}(\boldsymbol{a}_2-\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1)+\boldsymbol{a}_1'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1 \\ =&(\boldsymbol{a}_2-\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1)^2/\sigma_{22\cdot 1}+\boldsymbol{a}_1'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1, \end{aligned}$$

where $\sigma_{22\cdot 1} = \sigma^2_{n+1} - \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{r}$ and is a real number.

Thus, the conditional probability density function of b_{n+1} given $\boldsymbol{b}^n$ can be simplified as

$$\begin{aligned} g_{b_{n+1}|\boldsymbol{b}^n}(b_{n+1}|\boldsymbol{b}^n) &\propto \frac{1}{b_{n+1}}\exp\{-\frac{1}{2\sigma_{22\cdot 1}}(\boldsymbol{a}_2-\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1)^2-\frac{1}{2}\boldsymbol{a}_1'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1\} \\ &\propto \frac{1}{b_{n+1}}\exp\{-\frac{1}{2\sigma_{22\cdot 1}}(\boldsymbol{a}_2-\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{a}_1)^2\} \\ &= \frac{1}{b_{n+1}}\exp\{-\frac{1}{2\sigma_{22\cdot 1}}\left(\log b_{n+1}-(\mu_{n+1}+\boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}(\log \boldsymbol{b}^n-\boldsymbol{\mu}^n))\right)^2\}. \end{aligned}$$

Therefore, according to the probability density function of a log-normal distribution, we have $b_{n+1}|\boldsymbol{b}^n \sim \mathcal{LN}(\mu^*, v^*)$, where $\mu^* = \mu_{n+1} + \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}(\log \boldsymbol{b}^n - \boldsymbol{\mu}^n)$ and $v^* = \sigma_{22\cdot 1} = \sigma^2_{n+1} - \boldsymbol{r}'\boldsymbol{\Sigma}^{-1}_{n\times n}\boldsymbol{r}$. □

Lemma S5.2. *Consider the model (1) (without time-series), given* $(p(\mathbf{x}_1),\ldots,p(\mathbf{x}_n))' = \boldsymbol{p}^n$, *the conditional distribution of* $p(\mathbf{x}_{n+1})$ *is a logit-normal distribution, that is,* $p(\mathbf{x}_{n+1})|\boldsymbol{p}^n \sim$

Logitnormal$(m(\boldsymbol{p}^n), v(\boldsymbol{p}^n))$ *with*

$$m(\boldsymbol{p}^n) = \mu(\mathbf{x}_{n+1}) + \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}(\log\frac{\boldsymbol{p}^n}{\mathbf{1}-\boldsymbol{p}^n} - \boldsymbol{\mu}^n) \quad \textit{and} \quad v(\boldsymbol{p}^n) = \sigma^2(1 - \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{r}_{\boldsymbol{\theta}}),$$

where $\boldsymbol{\mu}^n = (\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_n))'$, $\mu(\mathbf{x}_i) = \alpha_0 + \mathbf{x}_i'\boldsymbol{\alpha}$, $\boldsymbol{r}_{\boldsymbol{\theta}} = (R_{\boldsymbol{\theta}}(\mathbf{x}_{n+1}, \mathbf{x}_1), \ldots, R_{\boldsymbol{\theta}}(\mathbf{x}_{n+1}, \mathbf{x}_n))'$, *and* $\boldsymbol{R}_{\boldsymbol{\theta}} = \{R_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)\}$.

Proof. Let $\eta_i = \mu(\mathbf{x}_i) + Z(\mathbf{x}_i)$ and $b_i = \exp\{\eta_i\} = p(\mathbf{x}_i)/(1 - p(\mathbf{x}_i))$ for $i = 1, \ldots, n+1$. Since $(\eta_1, \ldots, \eta_n, \eta_{n+1})' \sim \mathcal{N}(\boldsymbol{\mu}^{n+1}, \sigma^2\boldsymbol{R}_{\boldsymbol{\theta}}^*)$, where $\boldsymbol{\mu}^{n+1} = ((\boldsymbol{\mu}^n)', \mu(\mathbf{x}_{n+1}))'$ and

$$\boldsymbol{R}_{\boldsymbol{\theta}}^* = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{\theta}} & \boldsymbol{r}_{\boldsymbol{\theta}} \\ \boldsymbol{r}_{\boldsymbol{\theta}}' & 1 \end{bmatrix},$$

we have $(b_1, \ldots, b_n, b_{n+1})' \sim \mathcal{LN}(\boldsymbol{\mu}^{n+1}, \sigma^2\boldsymbol{R}_{\boldsymbol{\theta}}^*)$ by Definition S5.1. Thus, using Jacobian of the transformation and Lemma S5.1, we have

$$\begin{aligned}
&g_{p(\mathbf{x}_{n+1})|p(\mathbf{x}_1),\ldots,p(\mathbf{x}_n)}(p_{n+1}|p_1, \ldots, p_n) \\
=&g_{b_{n+1}|b_1,\ldots,b_n}(\frac{p_{n+1}}{1-p_{n+1}}|\frac{p_1}{1-p_1}, \ldots, \frac{p_n}{1-p_n})\frac{1}{(1-p_{n+1})^2} \\
\propto&\frac{1-p_{n+1}}{p_{n+1}}\exp\{-\frac{\left(\log\frac{p_{n+1}}{1-p_{n+1}} - (\mu(\mathbf{x}_{n+1}) + \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}(\log\frac{\boldsymbol{p}^n}{\mathbf{1}-\boldsymbol{p}^n} - \boldsymbol{\mu}^n))\right)^2}{2\sigma^2(1 - \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{r}_{\boldsymbol{\theta}})}\}\frac{1}{(1-p_{n+1})^2} \\
\propto&\frac{1}{p_{n+1}(1-p_{n+1})}\exp\{-\frac{\left(\log\frac{p_{n+1}}{1-p_{n+1}} - (\mu(\mathbf{x}_{n+1}) + \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}(\log\frac{\boldsymbol{p}^n}{\mathbf{1}-\boldsymbol{p}^n} - \boldsymbol{\mu}^n))\right)^2}{2\sigma^2(1 - \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{r}_{\boldsymbol{\theta}})}\}.
\end{aligned}$$

Therefore, according to the probability density function of a logit-normal distribution, we have $p(\mathbf{x}_{n+1})|\boldsymbol{p}^n \sim$ *Logitnormal*$(m(\boldsymbol{p}^n), v(\boldsymbol{p}^n))$. □

Similarly, the result of Lemma S5.2 can be extended to the general model (3). Given $\boldsymbol{Y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T', y_{n+1,1}, \ldots, y_{n+1,s-1})'$, at a fixed time-step s, $p_s(\mathbf{x}_i)$ can be seen to have the model (1) with mean function $\mu(\mathbf{x}_i, \boldsymbol{Y}) = \sum_{r=1}^R \varphi_r y_{i,s-r} + \alpha_0 + \mathbf{x}_i'\boldsymbol{\alpha} + \sum_{l=1}^L \boldsymbol{\gamma}_l\mathbf{x}_i y_{i,s-l}$. Thus,

by Lemma S5.2, denote $\boldsymbol{p}_s = (p_s(\mathbf{x}_1), \ldots, p_s(\mathbf{x}_n))'$, we have

$$p_s(\mathbf{x}_{n+1})|\boldsymbol{p}_s, \boldsymbol{Y} \sim \mathit{Logitnormal}(m(\boldsymbol{p}_s, \boldsymbol{Y}), v(\boldsymbol{p}_s, \boldsymbol{Y})),$$

where $m(\boldsymbol{p}_s, \boldsymbol{Y}) = \mu(\mathbf{x}_{n+1}, \boldsymbol{Y}) + \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}(\log \frac{\boldsymbol{p}_s}{\mathbf{1}-\boldsymbol{p}_s} - \boldsymbol{\mu}^n)$, $\boldsymbol{\mu}^n = (\mu(\mathbf{x}_1, \boldsymbol{Y}), \ldots, \mu(\mathbf{x}_n, \boldsymbol{Y}))'$, and $v(\boldsymbol{p}_s, \boldsymbol{Y}) = \sigma^2(1 - \boldsymbol{r}_{\boldsymbol{\theta}}'\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}\boldsymbol{r}_{\boldsymbol{\theta}})$. By the fact that $Z_t(\mathbf{x})$ is independent over time, which implies $p_s(\mathbf{x})$ is independent of $p_t(\mathbf{x})$ for any $t \neq s$, $p_s(\mathbf{x}_{n+1})|D_{n+1,s}$ and $p_s(\mathbf{x}_{n+1})|\boldsymbol{p}_s, \boldsymbol{Y}$ have the same distribution. So, $p_s(\mathbf{x}_{n+1})|D_{n+1,s} \sim \mathit{Logitnormal}(m(D_{n+1,s}), v(D_{n+1,s}))$, where $m(D_{n+1,s}) = m(\boldsymbol{p}_s, \boldsymbol{Y})$ and $v(D_{n+1,s}) = v(\boldsymbol{p}_s, \boldsymbol{Y})$.

S6 Proof of Theorem 4.3

(i) First, one can show that if $(p_s(\mathbf{x}_{n+1}), D_{n+1,s})$ has a joint distribution for which the conditional mean of $p_s(\mathbf{x}_{n+1})$ given $D_{n+1,s}$ exists, then $\mathbb{E}\left[p(\mathbf{x}_{n+1})|D_{n+1,s}\right]$ is the minimum mean squared error predictor of $p(\mathbf{x}_{n+1})$. See Theorem 3.2.1 in Santner et al. (2003). Thus, by the result of Lemma 4.1, we have the conditional mean $\mathbb{E}\left[p(\mathbf{x}_{n+1})|D_{n+1,s}\right] = \kappa(m(D_{n+1,s}), v(D_{n+1,s}))$ with variance $\mathbb{V}\left[p(\mathbf{x}_{n+1})|D_{n+1,s}\right] = \tau(m(D_{n+1,s}), v(D_{n+1,s}))$.
(ii) If $\mathbf{x}_{n+1} = \mathbf{x}_i$ for $i = 1, \ldots, n$, then $m(D_{n+1,s}) = \log(p_s(\mathbf{x}_i)/(1-p_s(\mathbf{x}_i)))$ and $v(D_{n+1,s}) = 0$, which implies that

$$\kappa(m(D_{n+1,s}), 0) = \exp\{m(D_{n+1,s})\}/(1 + \exp\{m(D_{n+1,s})\}) = p_s(\mathbf{x}_i)$$

and $\tau(m(D_{n+1,s}), 0) = 0$ by using transformation of a normal distribution. Thus, by Theorem 4.3 (i), we have $\mathbb{E}\left[p_s(\mathbf{x}_{n+1})|D_{n+1,s}\right] = p_s(\mathbf{x}_i)$ and $\mathbb{V}\left[p_s(\mathbf{x}_{n+1})|D_{n+1,s}\right] = 0$.
(iii) Let $X \sim \mathcal{N}(m(D_{n+1,s}), v(D_{n+1,s}))$, $P = \exp\{X\}/(1 + \exp\{X\})$, which has the distribution $\mathit{Logitnormal}(m(D_{n+1,s}), v(D_{n+1,s}))$, and $Q(q; D_{n+1,s})$ be the q-th quantile of P. Consider the function $f(x) = \log(x/(1-x))$. The derivative is $f'(x) = 1/(x(1-x))$. Thus,

for $0 < x < 1$ the derivative is positive and the $f(x)$ function is increasing in x. Then,

$$
\begin{aligned}
& Pr\left\{P > Q(q; D_{n+1,s})\right\} = q \\
\Leftrightarrow & Pr\left\{\frac{\exp\{X\}}{1+\exp\{X\}} > Q(q; D_{n+1,s})\right\} = q \\
\Leftrightarrow & Pr\left\{f(\frac{\exp\{X\}}{1+\exp\{X\}}) > f(Q(q; D_{n+1,s}))\right\} = q \\
\Leftrightarrow & Pr\left\{X > \log\frac{Q(q; D_{n+1,s})}{1-Q(q; D_{n+1,s})}\right\} = q
\end{aligned}
$$

$$
\begin{aligned}
\Leftrightarrow & Pr\left\{\frac{X - m(D_{n+1,s})}{\sqrt{v(D_{n+1,s})}} > \frac{1}{\sqrt{v(D_{n+1,s})}}\left(\log\frac{Q(q; D_{n+1,s})}{1-Q(q; D_{n+1,s})} - m(D_{n+1,s})\right)\right\} = q \\
\Leftrightarrow & \frac{1}{\sqrt{v(D_{n+1,s})}}\left(\log\frac{Q(q; D_{n+1,s})}{1-Q(q; D_{n+1,s})} - m(D_{n+1,s})\right) = z_q \\
\Leftrightarrow & Q(q; D_{n+1,s}) = \frac{\exp\{m(D_{n+1,s}) + z_q\sqrt{v(D_{n+1,s})}\}}{1+\exp\{m(D_{n+1,s}) + z_q\sqrt{v(D_{n+1,s})}\}}.
\end{aligned}
$$

S7 Metropolis-Hastings Algorithm and Approximation for Theorem 4.4

The Metropolis-Hastings (MH) algorithm for generating random samples from $\boldsymbol{p}|\boldsymbol{Y}$ is given as follows.

for $j = 1$ to J **do**
  Set $N_s = nT + s - 1$.
  Start with a zero vector $\boldsymbol{p}$ of size N_s.
  for $k = 1$ to N_s **do**
    Generate a random value p_k^* from $Logitnormal(m(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}), v(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}))$.

Generate an uniform random variable $U \sim Unif(0,1)$.
if $U < \min\{1, \frac{f(y_k|p_k^*)}{f(y_k|p_k)}\}$ **then**
Set $\boldsymbol{p} = (p_1, \ldots, p_k^*, \ldots, p_{N_s})$.
Set $\boldsymbol{p}^{(j)} = \boldsymbol{p}$
Return $\{\boldsymbol{p}^{(j)}\}_{j=1,\ldots,J}$.

In the algorithm, we first sample a value for the k-th component p_k from the conditional distribution of p_k given $p_j, y_j, j \neq k$, which is $Logitnormal(m(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}), v(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}))$, where

$$m(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}) = \mu_t(\mathbf{x}_i) - \sum_{k \neq j} \frac{Q_{kj}}{Q_{kk}} \left(\log \frac{p_k}{1-p_k} - \mu_t(\mathbf{x}_i) \right), \quad v(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}) = \frac{\sigma^2}{Q_{kk}},$$

in which $\mu_t(\mathbf{x}_i) = \sum_{r=1}^{R} \varphi_r y_{i,t-r} + \mathbf{x}_i' \boldsymbol{\alpha} + \sum_{l=1}^{L} \boldsymbol{\gamma}_l \mathbf{x}_i y_{i,t-l}$ and Q_{kj} is the (k,j)-element of $\boldsymbol{R}_{\boldsymbol{\theta}}^{-1}$. Similar to Zhang (2002), we use the single-component MH algorithm, that is, to update only a single component at each iteration. Moreover, the proposed distribution $f(p_k)$ is used for the single MH algorithm, so that the probability of accepting a new p_k^* is the minimum of 1 and $\frac{f(p_k^*|y_k)f(p_k)}{f(p_k|y_k)f(p_k^*)} \left(= \frac{f(y_k|p_k^*)}{f(y_k|p_k)} \right)$.

Based on the samples $\{\boldsymbol{p}^{(j)}\}_{j=1,\ldots,J}$, the mean, variance, and q-quantile of $p_s(\mathbf{x}_{n+1})|\boldsymbol{Y}$ can be respectively approximated by

$$\frac{1}{J} \sum_{j=1}^{J} \kappa(m(\boldsymbol{p}^{(j)}, \boldsymbol{Y}), v(\boldsymbol{p}^{(j)}, \boldsymbol{Y})),$$

$$\frac{1}{J} \sum_{j=1}^{J} \tau(m(\boldsymbol{p}^{(j)}, \boldsymbol{Y}), v(\boldsymbol{p}^{(j)}, \boldsymbol{Y})) + \frac{1}{J-1} \sum_{j=1}^{J} \left[\kappa(m(\boldsymbol{p}^{(j)}, \boldsymbol{Y}), v(\boldsymbol{p}^{(j)}, \boldsymbol{Y}))^2 - \frac{1}{J} \sum_{j=1}^{J} \kappa(m(\boldsymbol{p}^{(j)}, \boldsymbol{Y}), v(\boldsymbol{p}^{(j)}, \boldsymbol{Y})) \right],$$

and the q-quantile of $\{p_s^{(j)}\}_{j=1}^{J}$, where $p_s^{(j)}$ is generated from $Logitnormal(m(\boldsymbol{p}^{(j)}, \boldsymbol{Y}), v(\boldsymbol{p}^{(j)}, \boldsymbol{Y}))$.

Similarly, the distribution of $y_s(\mathbf{x}_{n+1})|\boldsymbol{Y}$ can be approximated by the sample distribution of $\{y_s^{(j)}\}_{j=1}^J$, where $y_s^{(j)}$ is generated from a Bernoulli distribution with probability $p_s^{(j)}$.

S8 Algorithm: Dynamic Binary Emulator

for $j = 1$ to J **do**
  Set $N = nT$.
  Start with a zero vector $\boldsymbol{p}$ of size N.
  for $i = 1$ to N **do**
    Generate a random value p_k^* from $Logitnormal(m(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}), v(\boldsymbol{p}_{-k}, \mathbf{y}_{-k}))$.
    Generate an uniform random variable $U \sim Unif(0, 1)$.
    if $U < \min\{1, \frac{f(y_k|p_k^*)}{f(y_k|p_k)}\}$ **then**
      Set $\boldsymbol{p} = (p_1, \ldots, p_k^*, \ldots, p_N)$.
  Set $\boldsymbol{p}_{n+1} = \boldsymbol{p}, \boldsymbol{Y}_{n+1} = \boldsymbol{Y}$, zero vectors $\boldsymbol{p}_{\text{new}}$ and $\mathbf{y}_{\text{new}}$ of size T.
  for $t = 1$ to T **do**
    Given $D_{n+1,t} = \{\boldsymbol{p}_{n+1}, \boldsymbol{Y}_{n+1}\}$, draw a sample $p_t(\mathbf{x}_{n+1})$ from $Logitnormal(m(D_{n+1,t}), v(D_{n+1,t}))$, and then draw a sample $y_t(\mathbf{x}_{n+1})$ from a Bernoulli distribution with parameter $p_t(\mathbf{x}_{n+1})$.
    Update $\boldsymbol{p}_{n+1} = (\boldsymbol{p}'_{n+1}, p_t(\mathbf{x}_{n+1}))'$, $\boldsymbol{Y}_{n+1} = (\boldsymbol{Y}'_{n+1}, y_t(\mathbf{x}_{n+1}))'$, $(\boldsymbol{p}_{\text{new}})_t = p_t(\mathbf{x}_{n+1})$, and $(\mathbf{y}_{\text{new}})_t = y_t(\mathbf{x}_{n+1})$.
  Set $\boldsymbol{p}_{\text{new}}^{(j)} = \boldsymbol{p}_{\text{new}}$ and $\mathbf{y}_{\text{new}}^{(j)} = \mathbf{y}_{\text{new}}$.
Take pointwise median from $\{\boldsymbol{p}_{\text{new}}^{(j)}\}_{j=1,\ldots,J}$ and $\{\mathbf{y}_{\text{new}}^{(j)}\}_{j=1,\ldots,J}$.

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25.

Cressie, N. and Lahiri, S. N. (1993). The asymptotic distribution of reml estimators. *Journal of Multivariate Analysis*, 45(2):217–233.

Cressie, N. and Lahiri, S. N. (1996). Asymptotics for reml estimation of spatial covariance parameters. *Journal of Statistical Planning and Inference*, 50(3):327–341.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*, volume 157. Springer.

Hung, Y., Zarnitsyna, V., Zhang, Y., Zhu, C., and Wu, C. F. J. (2008). Binary time series modeling with application to adhesion frequency experiments. *Journal of the American Statistical Association*, 103(483).

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer New York.

Tang, B. (1993). Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. *Biometrics*, 58(1):129–136.