Logistic-Beta Processes for Modeling Dependent Random Probabilities with Beta Marginals

Presenter: Changwoo Lee (Texas A&M)

joint work with Alessandro Zito (Harvard), Huiyan Sang (Texas A&M), and David Dunson (Duke)

https://arxiv.org/abs/2402.07048

Stat café, Department of Statistics, Texas A&M University
March 25, 2024

Beta distributions and BNP models

$$V \sim \text{Beta}(a, b), \quad p(v) = \frac{1}{B(a,b)} v^{a-1}(1-v)^{b-1}, \ v \in (0,1)$$

- **Beta distribution** for modeling random probabilities / ratios
- Natural interpretation of parameters, Conjugacy with binomial, negative binomial, ...
- Key component in many **Bayesian nonparametric (BNP) models**.
  - (Ex) Dirichlet process (DP) mixture model [Ferguson, 1973, Lo, 1984, Sethuraman, 1994],

$$f(y) = \sum_{h=1}^{\infty} \pi_h \mathcal{K}(y; \theta_h)$$

$$\pi_h = V_h \prod_{l<h}(1 - V_l), \quad V_h \overset{\text{iid}}{\sim} \text{Beta}(1, b), \quad \theta_h \overset{\text{iid}}{\sim} G_0 \quad , h = 1, 2, \ldots$$

Dependent BNP models

- Common recipe to build dependent BNP models: Replacing independent components with **stochastic processes** indexed by covariate $x \in \mathcal{X}$
- (Ex) Dependent DP mixture with covariate-dependent weights [MacEachern, 1999]

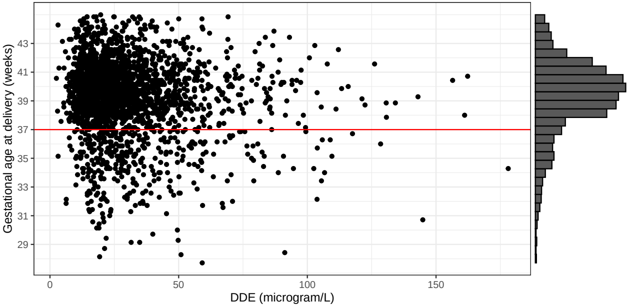
$$f(y) = \sum_{h=1}^{\infty} \left\{ V_h \prod_{l<h} (1 - V_l) \right\} \mathcal{K}(y; \theta_h) \quad \Longrightarrow \quad f(y \mid x) = \sum_{h=1}^{\infty} \left\{ V_h(x) \prod_{l<h} (1 - V_l(x)) \right\} \mathcal{K}(y; \theta_h)$$

$$V_h \stackrel{\text{iid}}{\sim} \text{Beta}(1, b), \quad h = 1, 2, \ldots \qquad V_h(x) \stackrel{\text{iid}}{\sim} \text{“beta process”}, \quad h = 1, 2, \ldots$$

- Stochastic process extension of beta plays an important role in many BNP models
  Examples: dependent Pólya tree [Trippa et al., 2011], dependent IBP [Perrone et al., 2017]
- Dependent DP application example: **Bayesian density regression**

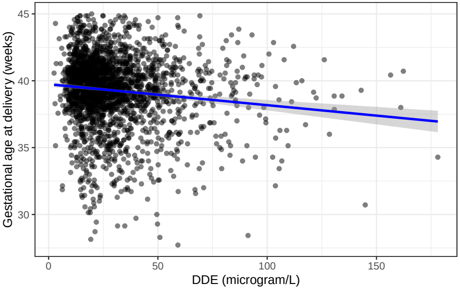
Example: Bayesian density regression (1)

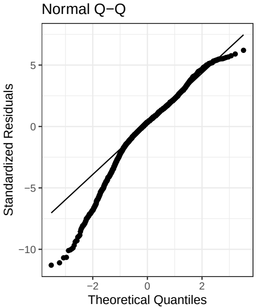
- Probabilistic modeling of conditional density $f(y \mid x)$ with uncertainty quantification
- (Ex) Health outcomes Y (GAD) and exposure X (DDE, metabolite of pesticide DDT)
  - Conditional prob. of preterm birth given DDE exposure level $\mathbb{P}(Y < 37 \mid X = x)$?



Example: Bayesian density regression (2)

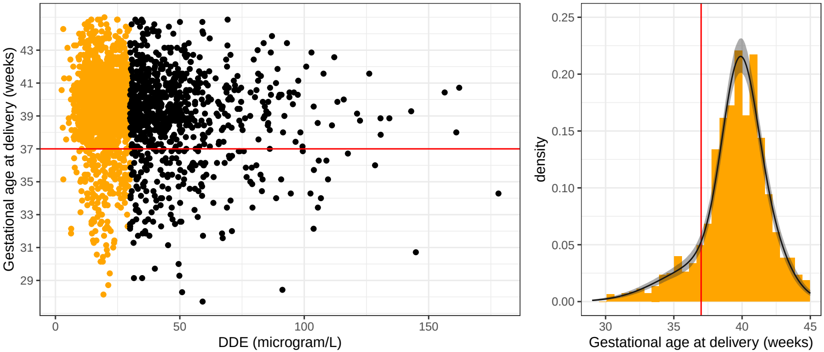
- Normal linear model assumes $f(y \mid x)$ follows normal distribution
- Too restrictive & assumptions do not meet in practice
- Bayesian nonparametric models offer highly flexible specifications





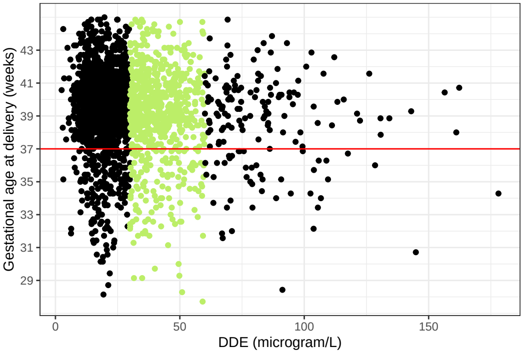
Example: Bayesian density regression (3)

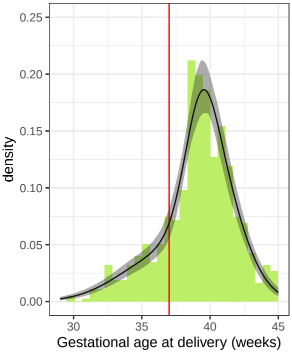
- Fitting DP mixture model for different subsets of data



Example: Bayesian density regression (3)

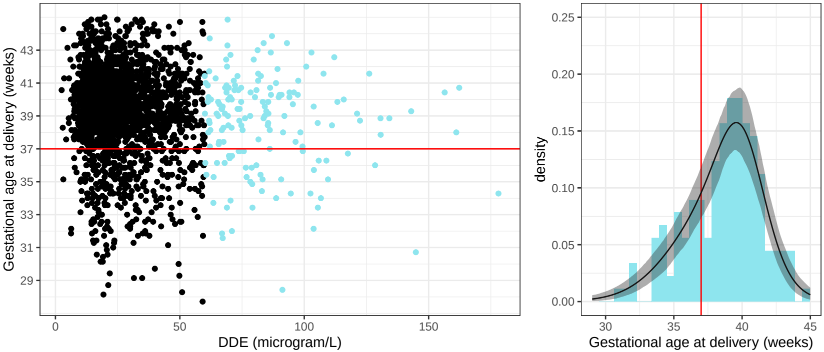
- Fitting DP mixture model for different subsets of data





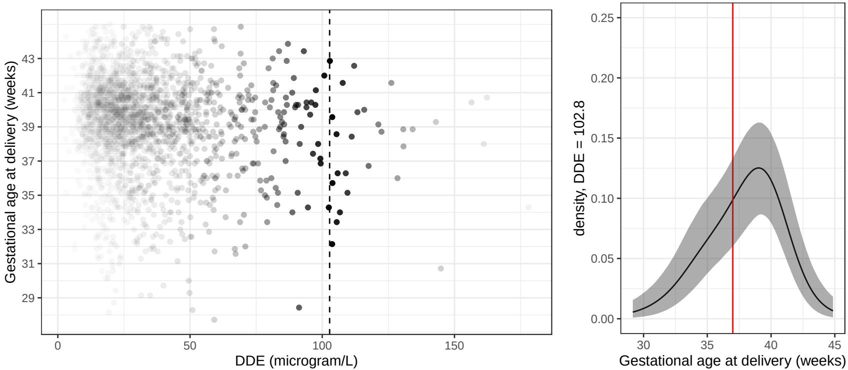
Example: Bayesian density regression (3)

- Fitting DP mixture model for different subsets of data



Example: Bayesian density regression (4)

- **Dependent Dirichlet process (DDP)** mixture model [MacEachern, 1999]
- Model conditional density $f(y \mid x)$, borrowing information across x



Dependent DP mixture model

- **Depenedent DP mixture** model with weights & atoms both depend on x

$$f(y \mid x) = \sum_{h=1}^{\infty} \left\{ V_h(x) \prod_{l<h} (1 - V_l(x)) \right\} \mathrm{N}(y; \mu_h(x), \tau_h^{-1})$$

$$V_h(\cdot) \overset{\text{iid}}{\sim} \text{“Beta process” with Beta(1,b) marginal}, \quad h = 1, 2, \ldots$$

$$\mu_h(x) = \beta_{0h} + \beta_{1h} x, \quad h = 1, 2, \ldots$$

- **Marginally DP** at each x is a key for preserve interpretability & properties
- Nontrivial “beta process”, often **highly challenging posterior computation**

Logistic-beta process

Three desired properties of "beta process":

(I) Accommodate **broad dependence structure**, both discrete and continuous $x \in \mathcal{X}$

(II) Allow **wide range of strengths** of dependence, from perfect to possibly negative

(III) Facilitate **efficient posterior inference** algorithms

We propose **Logistic-beta process**, whose logistic transformation $x \mapsto 1/(1+e^{-x})$ leads to stochastic process with common beta marginals that satisfies (I) - (III)

Logistic-beta process

Three desired properties of "beta process":

(I) Accommodate **broad dependence structure**, both discrete and continuous $x \in \mathcal{X}$

(II) Allow **wide range of strengths** of dependence, from perfect to possibly negative

(III) Facilitate **efficient posterior inference** algorithms

We propose **Logistic-beta process**, whose logistic transformation $x \mapsto 1/(1+e^{-x})$ leads to stochastic process with common beta marginals that satisfies (I) - (III)

Logistic-beta process

Three desired properties of "beta process":

(I) Accommodate **broad dependence structure**, both discrete and continuous $x \in \mathcal{X}$

(II) Allow **wide range of strengths** of dependence, from perfect to possibly negative

(III) Facilitate **efficient posterior inference** algorithms

We propose **Logistic-beta process**, whose logistic transformation $x \mapsto 1/(1+e^{-x})$ leads to stochastic process with common beta marginals that satisfies (I) - (III)

Logistic-beta process

Three desired properties of "beta process":

(I) Accommodate **broad dependence structure**, both discrete and continuous $x \in \mathcal{X}$

(II) Allow **wide range of strengths** of dependence, from perfect to possibly negative

(III) Facilitate **efficient posterior inference** algorithms

We propose **Logistic-beta process**, whose logistic transformation $x \mapsto 1/(1+e^{-x})$ leads to stochastic process with common beta marginals that satisfies (I) - (III)

Logistic-beta distribution

- Start from **univariate logistic-beta** (LB) distribution
  - Also called type IV generalized logistic or Fisher's z distribution (up to location-scale)

- We say $\eta \sim \mathrm{LB}_1(a, b)$ with shape parameters $a, b > 0$ if

$$\mathrm{LB}_1(\eta; a, b) = \frac{1}{B(a, b)} \left(\frac{1}{1 + e^{-\eta}} \right)^a \left(\frac{e^{-\eta}}{1 + e^{-\eta}} \right)^b, \quad \eta \in \mathbb{R}$$

- When $a = b = 1$, it becomes standard logistic distribution

- Applying logistic transformation $\sigma(x) = 1/(1 + e^{-x})$ gives $\sigma(\eta) \sim \mathrm{Beta}(a, b)$

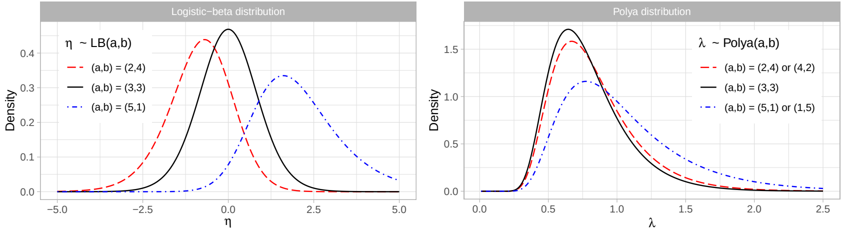
- In other words, if $\pi \sim \mathrm{Beta}(a, b)$, then $\mathrm{logit}(\pi) = \log(\pi/(1 - \pi)) \sim \mathrm{LB}_1(a, b)$

Logistic-beta distribution

- **Normal variance-mean mixture** representation of LB [Barndorff-Nielsen et al., 1982]

$$\mathrm{LB}_1(\eta; a, b) = \int_0^\infty \mathrm{N}\left(\eta; 0.5\lambda(a-b), \lambda\right) \mathrm{Polya}(\lambda; a, b) \mathrm{d}\lambda$$

- We say $\lambda \sim \mathrm{Polya}(a, b)$ if $\lambda \stackrel{\mathrm{d}}{=} \sum_{k=0}^{\infty} 2\epsilon_k / \{(k+a)(k+b)\}, \quad \epsilon_k \stackrel{\text{iid}}{\sim} \mathrm{Exp}(1)$



Multivariate Logistic-beta

- Multivariate extension with normal variance-mean mixture.
- We say $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top \sim \mathrm{LB}_n(a, b, \mathbf{R})$ with correlation matrix $\mathbf{R}_{n \times n}$ if

$$\begin{aligned} \boldsymbol{\eta} \mid \lambda &\sim \mathrm{N}_n\left(0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R}\right), \\ \lambda &\sim \mathrm{Polya}(a, b) \end{aligned}$$

- Since $\mathbf{R}$ has a unit diagonal, each component of $\boldsymbol{\eta}$ marginally follows $\mathrm{LB}_1(a, b)$
- Logistic transformation $\eta_i \mapsto \sigma(\eta_i)$ gives **multivariate beta** with $\mathrm{Beta}(a, b)$ marginals.
- Correlation matrix $\mathbf{R}$ controls dependence
- Briefly mentioned in [Barndorff-Nielsen et al., 1982], density function (complicated) is studied by [Grigelionis, 2008], but with no connection to beta distribution

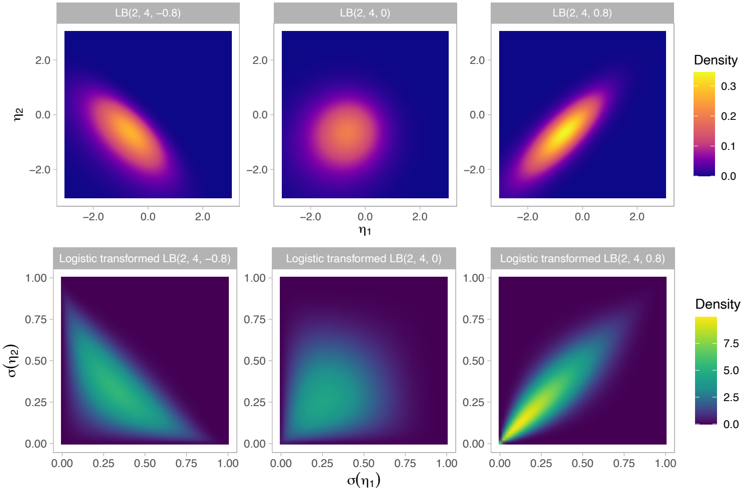


Figure: (Top) Density of $\boldsymbol{\eta} \sim \mathrm{LB}_2(a=2, b=4, \mathbf{R})$ with $R_{12} \in \{-0.8, 0, 0.8\}$. (Bottom) Density of $\sigma(\boldsymbol{\eta})$.

Multivariate logistic-beta

- Covariance (and correlation) is simply a linear function of R_{ij}

$$\operatorname{cov}(\eta_i, \eta_j) = \begin{cases} 2\psi'(a)R_{ij}, & \text{if } a = b, \\ \psi'(a) + \psi'(b) + 2(R_{ij} - 1)\{\psi(a) - \psi(b)\}/(a - b), & \text{if } a \neq b, \end{cases}$$

where $\psi(x), \psi'(x)$ are digamma and trigamma functions.

- $R_{ij} = R_{ji} = 0$ does not imply $\eta_i \perp \eta_j$.
- If $a = b$ (symmetric), correlation has a full range $[-1, 1]$.
- If $a \neq b$ (asymmetric), the range of $\operatorname{corr}(\eta_i, \eta_j)$ is

$$\operatorname{Range}(\operatorname{corr}(\eta_i, \eta_j)) = \Big[\underbrace{1 - \frac{4(\psi(a) - \psi(b))}{(a-b)(\psi'(a) + \psi'(b))}}_{\text{Nontrivial lower bound}}, 1\Big]$$

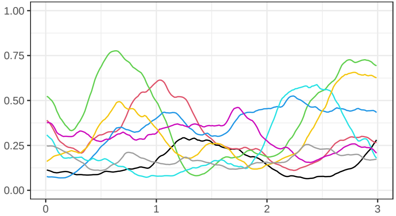
- Different from the minimal correlation (> -1) from Fréchet lower bound copula

Logistic-beta process

- Correlation kernel $R : \mathcal{X} \times \mathcal{X} \to [-1, 1]$ with $R(x, x) = 1$
- We say $\eta(\cdot)$ follows **logistic-beta process**, denoted as $\eta \sim \mathrm{LBP}(a, b, R)$, if every finite collection $\boldsymbol{\eta}$ follows logistic-beta with a, b, and $\mathbf{R}_{n\times n}$ with $R_{ij} = R(x_i, x_j)$
- Logistic transformation $\sigma(\eta(x))$ has $\mathrm{Beta}(a, b)$ marginals
- (Example 1) Discrete time indices $\mathcal{X} = \{1, 2, \ldots, \}$: $R(x, x') = \rho^{|x-x'|}$ with $|\rho| < 1$.
- (Example 2) Continuous spatial domain $\mathcal{X} = \mathbb{R}^2$: Matérn correlation kernel with range and smoothness parameters. Example with $(a, b) = (2, 4)$, $\varrho = 0.3$, $\nu = 1.5$, $\mathcal{X} = [0, 3]$

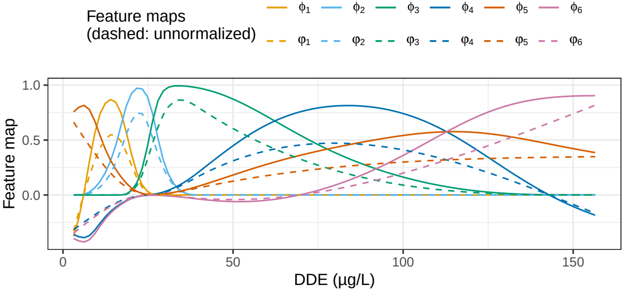
Logistic-beta process

- Correlation kernel $R : \mathcal{X} \times \mathcal{X} \to [-1, 1]$ with $R(x, x) = 1$
- We say $\eta(\cdot)$ follows **logistic-beta process**, denoted as $\eta \sim \mathrm{LBP}(a, b, R)$, if every finite collection $\boldsymbol{\eta}$ follows logistic-beta with a, b, and $\mathbf{R}_{n \times n}$ with $R_{ij} = R(x_i, x_j)$
- Logistic transformation $\sigma(\eta(x))$ has $\mathrm{Beta}(a, b)$ marginals
- (Example 1) Discrete time indices $\mathcal{X} = \{1, 2, \dots, \}$: $R(x, x') = \rho^{|x - x'|}$ with $|\rho| < 1$.
- (Example 2) Continuous spatial domain $\mathcal{X} = \mathbb{R}^2$: Matérn correlation kernel with range and smoothness parameters. Example with $(a, b) = (2, 4)$, $\varrho = 0.3$, $\nu = 1.5$, $\mathcal{X} = [0, 3]$



Logistic-beta process with normalized feature map kernel

- Correlation kernel R with normalized feature map (normalized basis)
- $\phi : \mathcal{X} \to \mathbb{R}^q$ such that $\|\phi(x)\|_2^2 = 1$, define $R(x, x') = \langle \phi(x), \phi(x') \rangle$
- Example: normalized spline basis functions with $q = 6$.



Logistic-beta process with normalized feature map kernel

- Two different representations of $\eta \sim \mathrm{LBP}(a, b, R)$ with $R(x, x') = \langle \phi(x), \phi(x') \rangle$
- **Hierarchical representation** of n-dimensional realization $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} = 0.5\lambda(a-b)\mathbf{1}_n + \sqrt{\lambda}\boldsymbol{\Phi}\boldsymbol{\gamma}, \quad \lambda \sim \mathrm{Polya}(a, b), \quad \boldsymbol{\gamma} \sim \mathrm{N}_q(\mathbf{0}, \mathbf{I}_q)$$

where $\boldsymbol{\Phi}_{n\times q}$ be a basis matrix with ith row $\phi(x_i)$.

- Conditioned on λ, $\boldsymbol{\eta}$ is parameterized by normal coefficients $\boldsymbol{\gamma} \sim \mathrm{N}_q(\mathbf{0}, \mathbf{I}_q)$
- Clear dimension reduction from n to q, based on q-dimensional feature map

Logistic-beta process with normalized feature map kernel

- Two different representations of $\eta \sim \mathrm{LBP}(a, b, R)$ with $R(x, x') = \langle \phi(x), \phi(x') \rangle$
- **Linear predictor representation** of n-dimensional realization $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} = \{\psi(a) - \psi(b)\}(\mathbf{1}_n - \boldsymbol{\Phi}\mathbf{1}_q) + \boldsymbol{\Phi}\boldsymbol{\beta}, \quad \boldsymbol{\beta} \sim \mathrm{LB}_q(a, b, \mathbf{I}_q)$$

- Assume Bernoulli response model $z(x_i) \overset{\text{ind}}{\sim} \mathrm{Ber}(\sigma\{\eta(x_i)\})$
- With logit link, $\boldsymbol{\eta}$ is a linear predictor with LB coefficients $\boldsymbol{\beta}$ with fixed varying intercept
- Resembles basis function representation of GP (“weight-space view”)

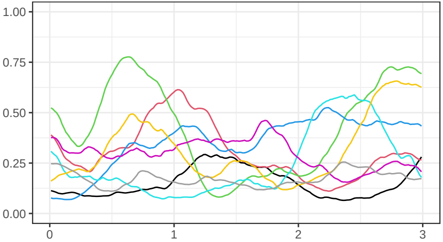
Latent LBP model

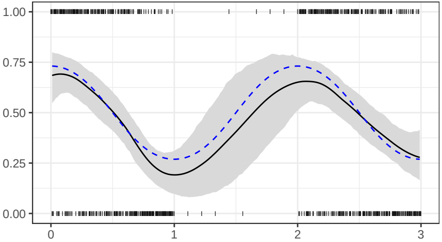
- Latent LBP model for binary data

$$z(x_i) \overset{\text{ind}}{\sim} \text{Ber}(\sigma\{\eta(x_i)\}), \quad i = 1, \ldots, n,$$
$$\eta \sim \text{LBP}(a, b, R)$$

- Lead to marginally $\text{Beta}(a, b)$ prior on success probabilities $\mathbb{P}(z(x) = 1)$

- Goal: Infer $\eta(x^*)$ at arbitrary x^*, with **efficient posterior computation**





Posterior computation strategies

- Posterior: $p(\boldsymbol{\eta} \mid \mathbf{z}) \propto p(\boldsymbol{\eta}) \prod_{i=1}^{n} p(z_i \mid \eta_i) = p(\boldsymbol{\eta}) \prod_{i=1}^{n} \exp(z_i \eta_i)/(1 + \exp(\eta_i))$
- Pólya-Gamma(PG) augmentation [Polson et al., 2013] $\rightarrow$ **conditionally normal likelihood**

$$z_i \overset{\text{ind}}{\sim} \text{Ber}(\sigma(\eta_i))$$

$$\mathcal{L}(\eta) = \prod_{i=1}^{n} \frac{\exp(z_i \eta_i)}{1 + \exp(\eta_i)} \quad \longrightarrow$$

$$\mathcal{L}_{\text{aug}}(\eta, \omega) = \prod_{i=1}^{n} 0.5 e^{(z_i - 0.5)\eta_i - \omega_i \eta_i^2/2} \text{PG}(\omega_i; 1, 0)$$

$$\omega_i \mid \eta_i, z_i \overset{\text{ind}}{\sim} \text{PG}(1, \eta_i), \quad i = 1, \ldots, n$$

$$\eta_i \mid \omega_i, z_i \overset{\text{ind}}{\sim} \text{N}((z_i - 0.5)\omega_i^{-1}, \omega_i^{-1}), \quad i = 1, \ldots, n$$

- LBP as a normal variance-mean mixture $\rightarrow$ **conditionally normal prior**

$$\boldsymbol{\eta} \sim \text{LB}_n(a, b, \mathbf{R}) \quad \longrightarrow$$

$$\boldsymbol{\eta} \mid \lambda \sim \text{N}_n(0.5\lambda(a - b)\mathbf{1}_n, \lambda \mathbf{R})$$

$$\lambda \sim \text{Polya}(a, b)$$

- **Normal-normal conjugacy** for full conditional $p(\boldsymbol{\eta} \mid \omega, \lambda, \mathbf{z}) \propto \mathcal{L}_{\text{aug}}(\boldsymbol{\eta} \mid \omega, \mathbf{z}) p(\boldsymbol{\eta} \mid \lambda)$

Posterior computation strategies

- Posterior: $p(\boldsymbol{\eta} \mid \mathbf{z}) \propto p(\boldsymbol{\eta}) \prod_{i=1}^{n} p(z_i \mid \eta_i) = p(\boldsymbol{\eta}) \prod_{i=1}^{n} \exp(z_i \eta_i)/(1 + \exp(\eta_i))$
- Pólya-Gamma(PG) augmentation [Polson et al., 2013] $\rightarrow$ **conditionally normal likelihood**

$$z_i \stackrel{\text{ind}}{\sim} \text{Ber}(\sigma(\eta_i))$$

$$\mathcal{L}(\boldsymbol{\eta}) = \prod_{i=1}^{n} \frac{\exp(z_i \eta_i)}{1 + \exp(\eta_i)} \longrightarrow$$

$$\mathcal{L}_{\text{aug}}(\boldsymbol{\eta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} 0.5 e^{(z_i - 0.5)\eta_i - \omega_i \eta_i^2/2} \text{PG}(\omega_i; 1, 0)$$

$$\omega_i \mid \eta_i, z_i \stackrel{\text{ind}}{\sim} \text{PG}(1, \eta_i), \quad i = 1, \ldots, n$$

$$\eta_i \mid \omega_i, z_i \stackrel{\text{ind}}{\sim} \text{N}((z_i - 0.5)\omega_i^{-1}, \omega_i^{-1}), \quad i = 1, \ldots, n$$

- LBP as a normal variance-mean mixture $\rightarrow$ **conditionally normal prior**

$$\eta \sim \text{LB}_n(a, b, \mathbf{R}) \longrightarrow$$

$$\eta \mid \lambda \sim \text{N}_n(0.5\lambda(a - b)\mathbf{1}_n, \lambda\mathbf{R})$$

$$\lambda \sim \text{Polya}(a, b)$$

- **Normal-normal conjugacy** for full conditional $p(\eta \mid \omega, \lambda, \mathbf{z}) \propto \mathcal{L}_{\text{aug}}(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \mathbf{z}) p(\boldsymbol{\eta} \mid \lambda)$

Posterior computation strategies

- Posterior: $p(\boldsymbol{\eta} \mid \mathbf{z}) \propto p(\boldsymbol{\eta}) \prod_{i=1}^{n} p(z_i \mid \eta_i) = p(\boldsymbol{\eta}) \prod_{i=1}^{n} \exp(z_i \eta_i)/(1 + \exp(\eta_i))$
- Pólya-Gamma(PG) augmentation [Polson et al., 2013] $\rightarrow$ **conditionally normal likelihood**

$$z_i \overset{\text{ind}}{\sim} \text{Ber}(\sigma(\eta_i))$$

$$\mathcal{L}(\boldsymbol{\eta}) = \prod_{i=1}^{n} \frac{\exp(z_i \eta_i)}{1 + \exp(\eta_i)} \quad \longrightarrow$$

$$\mathcal{L}_{\text{aug}}(\boldsymbol{\eta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} 0.5 e^{(z_i - 0.5)\eta_i - \omega_i \eta_i^2/2} \text{PG}(\omega_i; 1, 0)$$

$$\omega_i \mid \eta_i, z_i \overset{\text{ind}}{\sim} \text{PG}(1, \eta_i), \quad i = 1, \ldots, n$$

$$\eta_i \mid \omega_i, z_i \overset{\text{ind}}{\sim} \text{N}((z_i - 0.5)\omega_i^{-1}, \omega_i^{-1}), \quad i = 1, \ldots, n$$

- LBP as a normal variance-mean mixture $\rightarrow$ **conditionally normal prior**

$$\boldsymbol{\eta} \sim \text{LB}_n(a, b, \mathbf{R}) \quad \longrightarrow$$

$$\boldsymbol{\eta} \mid \lambda \sim \text{N}_n(0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R})$$

$$\lambda \sim \text{Polya}(a, b)$$

- **Normal-normal conjugacy** for full conditional $p(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z}) \propto \mathcal{L}_{\text{aug}}(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \mathbf{z}) p(\boldsymbol{\eta} \mid \lambda)$

Posterior computation strategies

- Posterior: $p(\boldsymbol{\eta} \mid \mathbf{z}) \propto p(\boldsymbol{\eta}) \prod_{i=1}^{n} p(z_i \mid \eta_i) = p(\boldsymbol{\eta}) \prod_{i=1}^{n} \exp(z_i \eta_i)/(1+\exp(\eta_i))$
- Pólya-Gamma(PG) augmentation [Polson et al., 2013] $\rightarrow$ **conditionally normal likelihood**

$$z_i \overset{\text{ind}}{\sim} \text{Ber}(\sigma(\eta_i))$$

$$\mathcal{L}(\boldsymbol{\eta}) = \prod_{i=1}^{n} \frac{\exp(z_i \eta_i)}{1+\exp(\eta_i)} \qquad \longrightarrow$$

$$\mathcal{L}_{\text{aug}}(\boldsymbol{\eta}, \boldsymbol{\omega}) = \prod_{i=1}^{n} 0.5 e^{(z_i - 0.5)\eta_i - \omega_i \eta_i^2/2} \text{PG}(\omega_i; 1, 0)$$

$$\omega_i \mid \eta_i, z_i \overset{\text{ind}}{\sim} \text{PG}(1, \eta_i), \quad i = 1, \ldots, n$$

$$\eta_i \mid \omega_i, z_i \overset{\text{ind}}{\sim} \text{N}((z_i - 0.5)\omega_i^{-1}, \omega_i^{-1}), \quad i = 1, \ldots, n$$

- LBP as a normal variance-mean mixture $\rightarrow$ **conditionally normal prior**

$$\boldsymbol{\eta} \sim \text{LB}_n(a, b, \mathbf{R}) \qquad \longrightarrow$$

$$\boldsymbol{\eta} \mid \lambda \sim \text{N}_n(0.5\lambda(a-b)\mathbf{1}_n, \lambda \mathbf{R})$$

$$\lambda \sim \text{Polya}(a, b)$$

- **Normal-normal conjugacy** for full conditional $p(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z}) \propto \mathcal{L}_{\text{aug}}(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \mathbf{z}) p(\boldsymbol{\eta} \mid \lambda)$

Posterior computation strategies

- One cycle of Gibbs sampler, update $\omega \to \lambda \to \boldsymbol{\eta}$ from full conditionals

---

Step 1. Update the Pólya-Gamma auxiliary variables from $p(\boldsymbol{\omega} \mid \boldsymbol{\eta}, \mathbf{z})$,

$$\omega_i \mid \boldsymbol{\eta} \overset{\text{ind}}{\sim} \text{PG}(1, \eta(x_i)), \quad i = 1, \ldots, n$$

Step 2. Update the Pólya mixing parameter from $p(\lambda \mid \boldsymbol{\eta}, \boldsymbol{\omega}, \mathbf{z})$,

$$p(\lambda \mid \boldsymbol{\omega}, \mathbf{z}) \propto \text{Polya}(\lambda; a, b)\text{N}_n\left(\boldsymbol{\eta}; 0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R}\right)$$

Step 3. Update the latent LBP parameters from $p(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z})$,

$$\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z} \sim \text{N}_n\left((\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1}((\mathbf{z} - 0.5\mathbf{1}_n) + 0.5(a-b)\mathbf{R}^{-1}\mathbf{1}_n), (\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1}\right)$$

- Can we do better?

Posterior computation strategies

- One cycle of **blocked Gibbs sampler**, update $\omega \to (\lambda, \boldsymbol{\eta})$
- Blocking possible due to normal-normal conjugacy

Step 1. Update the Pólya-Gamma auxiliary variables from $p(\boldsymbol{\omega} \mid \boldsymbol{\eta}, \mathbf{z})$,

$$\omega_i \mid \boldsymbol{\eta} \overset{\text{ind}}{\sim} \text{PG}(1, \eta(x_i)), \quad i = 1, \ldots, n$$

Step 2. Update the Pólya mixing parameter from $p(\lambda \mid \boldsymbol{\omega}, \mathbf{z}) = \int p(\lambda, \boldsymbol{\eta} \mid \boldsymbol{\omega}, \mathbf{z}) d\boldsymbol{\eta}$,

$$p(\lambda \mid \boldsymbol{\omega}, \mathbf{z}) \propto \text{Polya}(\lambda; a, b) \text{N}_n \left(\boldsymbol{\Omega}^{-1}(\mathbf{z} - 0.5\mathbf{1}_n); 0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R} + \boldsymbol{\Omega}^{-1} \right)$$

Step 3. Update the latent LBP parameters from $p(\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z})$,

$$\boldsymbol{\eta} \mid \boldsymbol{\omega}, \lambda, \mathbf{z} \sim \text{N}_n \left((\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1}((\mathbf{z} - 0.5\mathbf{1}_n) + 0.5(a-b)\mathbf{R}^{-1}\mathbf{1}_n), (\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1} \right)$$

Posterior computation strategies (detail 1)

Step 2. Update the Pólya mixing parameter from $p(\lambda \mid \boldsymbol{\omega}, \mathbf{z}) = \int p(\lambda, \boldsymbol{\eta} \mid \boldsymbol{\omega}, \mathbf{z}) d\boldsymbol{\eta}$,

$$p(\lambda \mid \boldsymbol{\omega}, \mathbf{z}) \propto \text{Polya}(\lambda; a, b) \text{N}_n \left(\boldsymbol{\Omega}^{-1}(\mathbf{z} - 0.5\mathbf{1}_n); 0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R} + \boldsymbol{\Omega}^{-1} \right)$$

- The Pólya density ▸ more satisfies the following identity for $a + b = a' + b'$:

$$\text{Polya}(\lambda; a', b') = B(a, b) B(a', b')^{-1} \exp\{\lambda(ab - a'b')/2\} \text{Polya}(\lambda; a, b),$$

- **Adaptive Pólya proposal** scheme for Step 2: selecting the proposal (e.g. for independent M-H) as $\text{Polya}(a', b')$, where (a', b') are adaptively chosen with fixed sum.
- Avoids Pólya density evaluation as they cancel out in the acceptance ratio.

Posterior computation strategies (detail 2)

Step 2. Update the Pólya mixing parameter from $p(\lambda \,|\, \boldsymbol{\omega}, \mathbf{z}) = \int p(\lambda, \boldsymbol{\eta} \,|\, \boldsymbol{\omega}, \mathbf{z}) d\boldsymbol{\eta}$,

$$p(\lambda \,|\, \boldsymbol{\omega}, \mathbf{z}) \propto \text{Polya}(\lambda; a, b) \text{N}_n \left(\boldsymbol{\Omega}^{-1}(\mathbf{z} - 0.5\mathbf{1}_n); 0.5\lambda(a-b)\mathbf{1}_n, \lambda\mathbf{R} + \boldsymbol{\Omega}^{-1} \right)$$

Step 3. Update the latent LBP parameters from $p(\boldsymbol{\eta} \,|\, \boldsymbol{\omega}, \lambda, \mathbf{z})$,

$$\boldsymbol{\eta} \,|\, \boldsymbol{\omega}, \lambda, \mathbf{z} \sim \text{N}_n \left((\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1}((\mathbf{z} - 0.5\mathbf{1}_n) + 0.5(a-b)\mathbf{R}^{-1}\mathbf{1}_n), (\boldsymbol{\Omega} + \lambda^{-1}\mathbf{R}^{-1})^{-1} \right)$$

- GP computation strategies preserving marginal variances can be seamlessly applied
- (Ex1) Low-rank (normalized feature map), where Step 3 becomes

$$\boldsymbol{\gamma} \,|\, \boldsymbol{\omega}, \lambda, \mathbf{z} \sim \text{N}_q((\mathbf{I}_q + \lambda\boldsymbol{\Phi}^\top\boldsymbol{\Omega}\boldsymbol{\Phi})^{-1}\sqrt{\lambda}\boldsymbol{\Phi}^\top((\mathbf{z} - 0.5\mathbf{1}_n) - 0.5\lambda(a-b)\boldsymbol{\omega}), (\mathbf{I}_q + \lambda\boldsymbol{\Phi}^\top\boldsymbol{\Omega}\boldsymbol{\Phi})^{-1})$$

- (Ex2) Modified predictive process [Finley et al., 2009], low-rank + diag, Woodbury formula

Posterior, copula-based model

- Compare with copula-based models, e.g. Gaussian copula

$$\begin{aligned} z(x_i) &\overset{\text{ind}}{\sim} \text{Ber}(F_B^{-1}(F_Z(\zeta_i))), \quad i = 1, \ldots, n, \\ \zeta &\sim \text{N}_n(0, R) \end{aligned}$$

  where F_Z is cdf of standard normal and F_B is cdf of $\text{Beta}(a, b)$.

- Lead to marginally $\text{Beta}(a, b)$ prior on success probabilities $\mathbb{P}(z(x) = 1)$, but,

$$p(\boldsymbol{\zeta} \mid \mathbf{z}) \propto \text{N}_n(\boldsymbol{\zeta}; 0, \mathbf{R}) \prod_{i=1}^{n} [F_B^{-1}(F_Z(\zeta_i))]^{z_i} [1 - F_B^{-1}(F_Z(\zeta_i))]^{1-z_i}$$

- Posterior computation of ζ is a nightmare, even in this simple Bernoulli model

Logistic-beta DDP mixture model

- **Logistic-beta dependent DP** mixture model with weights & atoms both depend on x

$$f(y \mid x) = \sum_{h=1}^{\infty} \left\{ V_h(x) \prod_{l<h} (1 - V_l(x)) \right\} \mathrm{N}(y; \mu_h(x), \tau_h^{-1})$$

$$V_h(\cdot) \overset{\text{iid}}{\sim} \text{“Beta process” with Beta(1,b) marginal}, \quad h = 1, 2, \ldots$$

$$\mu_h(x) = \beta_{0h} + \beta_{1h} x, \quad h = 1, 2, \ldots$$

with priors on atom parameters $\beta_{0h}, \beta_{1h}, \tau_h$, independently across h

- **Rich dependence** structure on any domain $x \in \mathcal{X}$ through correlation kernel R
- **Efficient posterior computation** exploiting conditional conjugacy

Logistic-beta DDP mixture model

- **Logistic-beta dependent DP** mixture model with weights & atoms both depend on x

$$f(y \mid x) = \sum_{h=1}^{\infty} \left\{ V_h(x) \prod_{l<h} (1 - V_l(x)) \right\} \mathrm{N}(y; \mu_h(x), \tau_h^{-1})$$

$$\mathrm{logit}(V_h(\cdot)) \overset{\mathrm{iid}}{\sim} \mathrm{LBP}(1, b, R), \quad h = 1, 2, \ldots$$

$$\mu_h(x) = \beta_{0h} + \beta_{1h} x, \quad h = 1, 2, \ldots$$

with priors on atom parameters $\beta_{0h}, \beta_{1h}, \tau_h$, independently across h

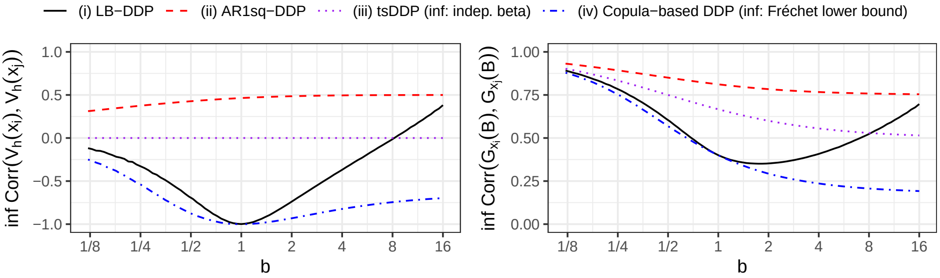
- **Rich dependence** structure on any domain $x \in \mathcal{X}$ through correlation kernel R
- **Efficient posterior computation** exploiting conditional conjugacy

Properties of logistic-beta DDP ▸ details

- Logistic-beta DDP: collection of dependent random probability measures $\{G_x : x \in \mathcal{X}\}$

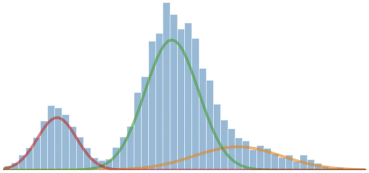
$$G_x(\cdot) = \sum_{h=1}^{\infty} \left(\sigma\{\eta_h(x)\} \prod_{l<h} [1 - \sigma\{\eta_l(x)\}] \right) \delta_{\theta_h}(\cdot), \quad \eta_h \overset{\text{iid}}{\sim} \text{LBP}(1, b, R), \theta_h \overset{\text{iid}}{\sim} G_0$$

- We study **range** of corr$(G_{x_i}(B), G_{x_j}(B))$ for Borel set B, $x_i \neq x_j$ with independent atoms



LB-DDP mixture posterior computation

$$f(y \mid x) = \sum_{h=1}^{\infty} \pi_h(x) \mathrm{N}(y; \mu_h(x), \tau_h^{-1})$$



Sample-specific assignment prob. $\{\pi_h(x_i)\}$ is

$$\mathbb{P}(s_i = h) = V_h(x_i) \prod_{l<h} (1 - V_l(x_i)), h \in \{1, 2, 3, \dots\}$$

Continuation-ratio scheme [Tutz, 1991]

$$\begin{aligned} V_1(x_i) &= \mathbb{P}(s_i = 1) \\ V_2(x_i) &= \mathbb{P}(s_i = 2 \mid s_i \geq 2) \\ V_3(x_i) &= \mathbb{P}(s_i = 3 \mid s_i \geq 3) \\ &\vdots \end{aligned}$$

Equivalent to a series of latent LBP models

$$\begin{aligned} 1(s_i = 1) &\overset{\text{ind}}{\sim} \mathrm{Ber}(\sigma(\eta_1(x_i))), \; i : s_i \geq 1 \\ 1(s_i = 2) &\overset{\text{ind}}{\sim} \mathrm{Ber}(\sigma(\eta_2(x_i))), \; i : s_i \geq 2 \\ 1(s_i = 3) &\overset{\text{ind}}{\sim} \mathrm{Ber}(\sigma(\eta_3(x_i))), \; i : s_i \geq 3 \\ &\vdots \\ \eta_h &\overset{\text{iid}}{\sim} \mathrm{LBP}(1, b, R), \; h = 1, 2, \dots \end{aligned}$$

LB-DDP mixture posterior computation

- One cycle of **blocked Gibbs sampler** (with truncation level H):

Step 1: For $i = 1, \ldots, n$, update component allocations $s_i \in \{1, \ldots, H\}$.

Step 2: **for** $h = 1, \ldots, H-1$ **do**

2-1. Update Pólya-gamma variables from $p(\boldsymbol{\omega}_h \mid \boldsymbol{\gamma}, \lambda, \mathbf{s})$, which is Pólya-Gamma,

2.2. Update the Pólya mixing parameter λ_h from $p(\lambda_h \mid \boldsymbol{\omega}_h, \mathbf{s}) = \int p(\lambda_h, \boldsymbol{\gamma}_h \mid \boldsymbol{\omega}_h, \mathbf{s}) d\boldsymbol{\gamma}_h$,

2.3. Update $\boldsymbol{\gamma}_h$ from $p(\boldsymbol{\gamma}_h \mid \lambda_h, \boldsymbol{\omega}_h, \mathbf{s})$, which is multivariate normal

Step 3: For $h = 1, \ldots, H$, update component-specific parameters

$$p(\theta_h \mid -) \propto G^0(\theta_h) \prod_{i:s_i=h} \mathcal{K}(y_i \mid \theta_h).$$

Similar to logit stick-breaking process [Rigon and Durante, 2021], with only step (2.2) added.

Latent LBP simulation studies

- Latent LBP model for binary data, beta marginal success probabilities

$$z(x_i) \overset{\text{ind}}{\sim} \text{Ber}(\sigma\{\eta(x_i)\}), \quad i = 1, \ldots, n,$$
$$\eta \sim \text{LBP}(a, b, R)$$

- Spatial domain $\mathcal{X} = [0, 1]^2$, $n_{\text{train}} = 400$, $n_{\text{test}} = 100$, Matérn correlation, fixed $\nu = 1.5$
- Data generation with (1) Latent LBP, (2) Gaussian copula, range $\varrho \in \{0.1, 0.2, 0.4\}$
- Aim 1: Assess **benefits of posterior inference strategies** involved in LBP
- Aim 2: Compare **predictive performance and computational advantages** of LBP.

Latent LBP simulation studies

- Aim 1: Assess **benefits of posterior inference strategies** involved in LBP

Data generation	Latent LBP algorithm		ESS	ESS/sec	Acc. rate (%)
Latent LBP $\varrho = 0.1$	Blocked	Adapted	**245.08 (12.86)**	**3.35 (0.18)**	**54.28 (1.03)**
		Non-adapted	196.35 (11.28)	2.69 (0.15)	49.39 (1.39)
	Non-blocked	Adapted	7.89 (0.35)	0.14 (0.01)	13.60 (0.26)
		Non-adapted	7.13 (0.37)	0.13 (0.01)	12.44 (0.31)
Latent LBP $\varrho = 0.2$	Blocked	Adapted	**257.01 (16.32)**	**2.89 (0.18)**	**62.26 (1.12)**
		Non-Adapted	247.83 (16.53)	2.82 (0.19)	57.60 (1.47)
	Non-blocked	Adapted	7.31 (0.32)	0.11 (0.00)	13.42 (0.22)
		Non-adapted	6.58 (0.32)	0.10 (0.00)	12.48 (0.31)
Latent LBP $\varrho = 0.4$	Blocked	Adapted	**368.26 (17.54)**	**4.12 (0.20)**	**66.12 (0.98)**
		Non-adapted	328.72 (19.00)	3.67 (0.21)	61.45 (1.32)
	Non-blocked	Adapted	6.40 (0.29)	0.09 (0.00)	13.01 (0.22)
		Non-adapted	6.56 (0.32)	0.10 (0.00)	12.88 (0.31)

ESS: effective sample size, ESS/sec: effective sampling rate, higher the better

Latent LBP simulation studies

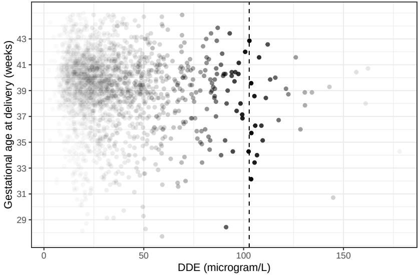
- Aim 2: Compare **predictive performance and computational advantages** of LBP.

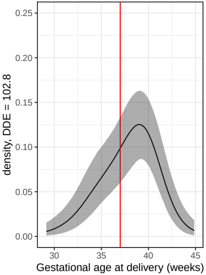
Data generation	Model	RMSE $\times 100$		mean CRPS $\times 100$		mESS/sec
		training	test	training	test	
Gauss. copula $\varrho = 0.1$	Latent LBP	11.93 (0.14)	12.32 (0.17)	6.59 (0.10)	6.80 (0.11)	**21.11 (0.21)**
	Gauss. copula	**11.82 (0.13)**	**12.24 (0.16)**	**6.48 (0.09)**	**6.71 (0.11)**	0.48 (0.01)
Gauss. copula $\varrho = 0.2$	Latent LBP	8.67 (0.15)	8.80 (0.16)	4.78 (0.10)	4.85 (0.10)	**17.58 (0.16)**
	Gauss. copula	**8.61 (0.16)**	**8.75 (0.17)**	**4.74 (0.10)**	**4.82 (0.11)**	0.40 (0.01)
Gauss. copula $\varrho = 0.4$	Latent LBP	6.11 (0.16)	6.14 (0.16)	3.39 (0.10)	3.41 (0.10)	**17.41 (0.17)**
	Gauss. copula	**6.10 (0.16)**	**6.13 (0.16)**	**3.38 (0.10)**	**3.40 (0.10)**	0.47 (0.01)

RMSE, CRPS: lower the better, mESS/sec: higher the better
Note: LBP results are based on a misspecified model. (see more)

Real data analysis

- No publicly available software for DDP with x-dependent weights, continuous x.
- Compare with logit stick-breaking process (LSBP) [Ren et al., 2011, Rigon and Durante, 2021] under similar settings. ▸ more LSBP is not DDP but computation is fast & tractable.
- Analyze **preterm birth probabilities** based on two subgroups of data (smoking Y/N).
- MCMC with $35,000$ iterations run in < 10 mins in personal laptop (Apple M1), $n \approx 1000$





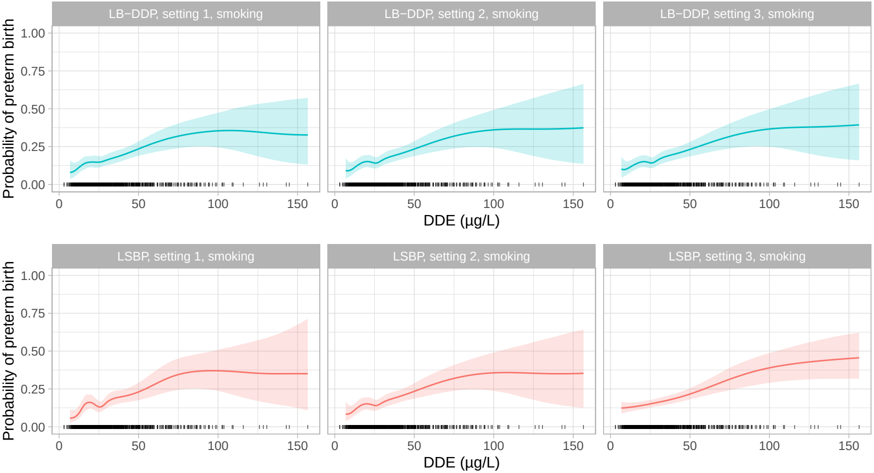


Figure: Estimated probability of preterm birth with 95% credible intervals for the smoking group, under three different hyperparameter settings for LB-DDP $b \in \{0.2, 1, 2\}$ and LSBP mixture models.

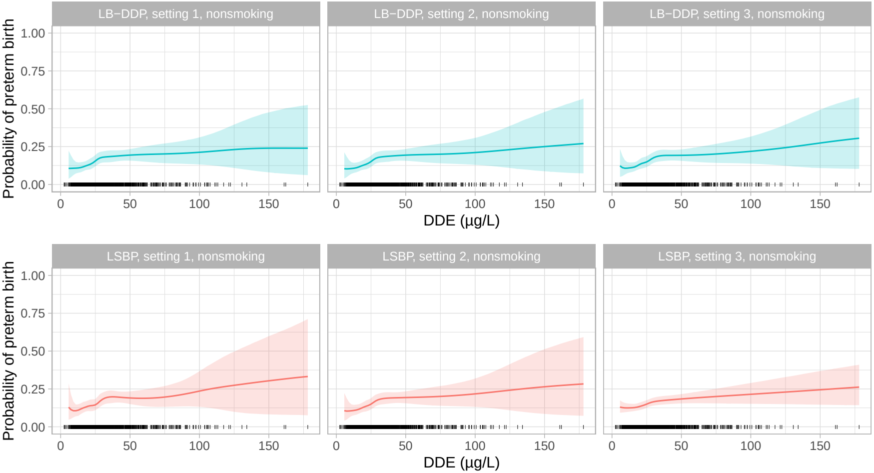


Figure: Estimated probability of preterm birth with 95% credible intervals for the nonsmoking group, under three different hyperparameter settings for LB-DDP $b \in \{0.2, 1, 2\}$ and LSBP mixture models.

Conclusion

- **Logistic-beta process** (LBP) for modeling dependent beta random probabilities
  - (I) Accommodate broad dependence structure, both discrete and continuous ✓
  - (II) Allow a wide range of dependence strengths, perfect correlation to possibly negative ✓
  - (III) Facilitate efficient posterior inference algorithm ✓
- We apply LBP to DDP mixture models, where dependent beta plays crucial role
- LB-DDP has **full flexibility** & clear interpretation & hyperparameter robustness & **computational tractability**, a rare feature with x-dependent weights [Wade et al., 2023]
- Application to x-dependent clustering & other dependent BNP models and beyond

Thank you!

References I

Barndorff-Nielsen, O., Kent, J., and Sørensen, M. (1982).
Normal variance-mean mixtures and z distributions.
International Statistical Review, 50(2):145–159.

Barrientos, A. F., Jara, A., and Quintana, F. A. (2012).
On the support of MacEachern's dependent Dirichlet processes and extensions.
Bayesian Analysis, 7(2):277–310.

DeYoreo, M. and Kottas, A. (2018).
Modeling for dynamic ordinal regression relationships: An application to estimating maturity of rockfish in California.
Journal of the American Statistical Association, 113(521):68–80.

Ferguson, T. S. (1973).
A Bayesian analysis of some nonparametric problems.
The Annals of Statistics, 1(2):209–230.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009).
Improving the performance of predictive process modeling for large datasets.
Computational Statistics & Data Analysis, 53(8):2873–2884.

References II

Grigelionis, B. (2008).
On Pólya mixtures of multivariate Gaussian distributions.
Statistics & Probability Letters, 78(12):1459–1465.

Lo, A. Y. (1984).
On a class of Bayesian nonparametric estimates: I. density estimates.
The Annals of Statistics, 12(1):351–357.

Longnecker, M. P., Klebanoff, M. A., Zhou, H., and Brock, J. W. (2001).
Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth.
The Lancet, 358(9276):110–114.

MacEachern, S. N. (1999).
Dependent nonparametric processes.
In *ASA Proceedings of the Section on Bayesian Statistical Science*, volume 1, pages 50–55.

MacEachern, S. N. (2000).
Dependent Drichlet processes.
Technical report, Department of Statistics, The Ohio State University.

References III

Nieto-Barajas, L. E., Müller, P., Ji, Y., Lu, Y., and Mills, G. B. (2012).
A time-series DDP for functional proteomics profiles.
Biometrics, 68(3):859–868.

Perrone, V., Jenkins, P. A., Spano, D., and Teh, Y. W. (2017).
Poisson random fields for dynamic feature models.
Journal of Machine Learning Research, 18:1–45.

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

Ren, L., Du, L., Carin, L., and Dunson, D. B. (2011).
Logistic stick-breaking process.
Journal of Machine Learning Research, 12(1):203–239.

Rigon, T. and Durante, D. (2021).
Tractable Bayesian density regression via logit stick-breaking priors.
Journal of Statistical Planning and Inference, 211:131–142.

References IV

Sethuraman, J. (1994).
A constructive definition of Dirichlet priors.
Statistica Sinica, 4(2):639–650.

Trippa, L., Müller, P., and Johnson, W. (2011).
The multivariate beta process and an extension of the Polya tree model.
Biometrika, 98(1):17–34.

Tutz, G. (1991).
Sequential models in categorical regression.
Computational Statistics & Data Analysis, 11(3):275–295.

Wade, S., Inacio, V., and Petrone, S. (2023).
Bayesian dependent mixture models: A predictive comparison and survey.
arXiv preprint arXiv:2307.16298.

Properties of LB-DDP ◂ back

Theorem (correlation range)

Consider a single-atoms LB-DDP $\{G_x : x \in \mathcal{X}\}$ *with concentration parameter* b *and correlation kernel* R*, where its atoms are i.i.d. from a nonatomic base measure. Let* $\mu(x_i, x_j) = \mathbb{E}[\sigma\{\eta(x_i)\}\sigma\{\eta(x_j)\}]$ *with* $\eta \sim \mathrm{LBP}(1, b, R)$*. Then, for any Borel set* B*,*

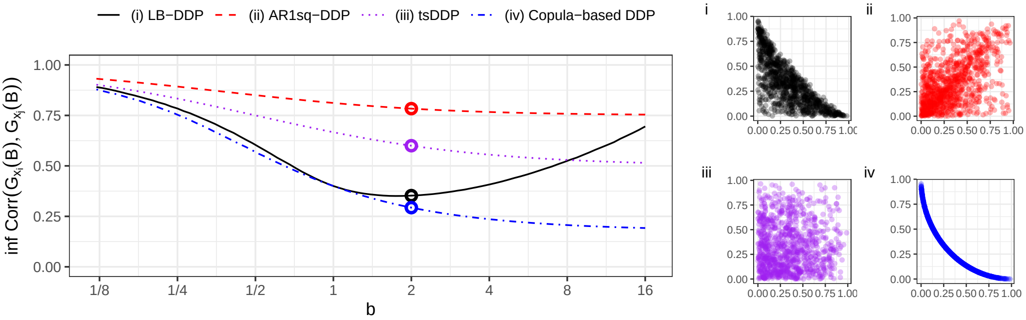
$$\textit{corr}(G_{x_i}(B), G_{x_j}(B)) = \frac{(1+b)^2}{2/\mu(x_i, x_j) - (1+b)}.$$

Theorem (full weak support, [Barrientos et al., 2012])

Consider an LB-DDP with an atom process $\{\theta(x) : x \in \mathcal{X}\}$ *with support* Θ *that can be represented with a collection of copulas with positive density w.r.t. Lebesgue measure. Then, LB-DDP has full weak support, i.e. the topological support of LB-DDP coincides with the space of collections of all probability measures with support* Θ *indexed by* $\mathcal{X}$*.*

Properties of LB-DDP back

- AR1sq-DDP [DeYoreo and Kottas, 2018]: smallest range, due to shared component
- tsDDP [Nieto-Barajas et al., 2012]: infimum attained at independent stick-breaking ratios
- Copula-based DDP [MacEachern, 2000]: infimum attained at Fréchet lower bound



Pólya density

back

- Density function of $\mathrm{Polya}(a,b)$:

$$\mathrm{Polya}(\lambda; a, b) = \sum_{k=0}^{\infty} \binom{-(a+b)}{k} \frac{k+(a+b)/2}{B(a,b)} \exp\left\{-\frac{(k+a)(k+b)}{2}\lambda\right\}, \qquad (1)$$

where $\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k(k-1)\cdots 1}$ for any $x \in \mathbb{R}$ and $k \in \{1, 2, \ldots, \}$, with provision $\binom{x}{0} = 1$.

- Alternating series, evaluation becomes unstable, especially near the origin

Latent LBP simulation studies ◂ back

Assessing the predictive performance of latent LBP under the correctly specified model

Data generation	Model	RMSE $\times 100$		mean CRPS $\times 100$		mESS/sec
		training	test	training	test	
Latent LBP	Latent LBP	**11.59 (0.15)**	**12.17 (0.20)**	**6.31 (0.10)**	**6.62 (0.12)**	**21.27 (0.24)**
$\varrho = 0.1$	Gauss. copula	11.66 (0.15)	12.19 (0.20)	6.35 (0.09)	6.65 (0.12)	0.48 (0.01)
Latent LBP	Latent LBP	**8.54 (0.18)**	**8.73 (0.19)**	**4.67 (0.11)**	**4.77 (0.11)**	**17.87 (0.16)**
$\varrho = 0.2$	Gauss. copula	8.59 (0.17)	8.76 (0.18)	4.70 (0.10)	4.79 (0.11)	0.44 (0.01)
Latent LBP	Latent LBP	**6.12 (0.19)**	**6.16 (0.19)**	**3.41 (0.11)**	**3.43 (0.11)**	**17.48 (0.16)**
$\varrho = 0.4$	Gauss. copula	6.15 (0.17)	6.19 (0.18)	3.43 (0.10)	3.44 (0.11)	0.47 (0.01)

Real data analysis settings ◂ back

- Collaborative perinatal project data, publicly available [Longnecker et al., 2001]
- Collected between 1959-1966, smoking group $n = 1023$, non-smoking $n = 1290$.
- Data are standardized, $\beta_{0h}, \beta_{1h} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$, $\tau_h \overset{\text{iid}}{\sim} \mathrm{Gamma}(1, 1)$
- Both LB-DDP, LSBP used normalized natural spline basis with 6 degrees of freedom
- 35,000 MCMC iteration took 7 mins for LB-DDP, 5 mins for LSBP
  - Setting 1: $b = 0.2$ (LB-DDP), $\sigma_\alpha^2 = 100$ (LSBP), co-clustering prob. ≈ 0.84
  - Setting 2: $b = 1$ (LB-DDP), $\sigma_\alpha^2 = \pi^2/3$ (LSBP), co-clustering prob. ≈ 0.5
  - Setting 3: $b = 2$ (LB-DDP), $\sigma_\alpha^2 = 0.2^2$ (LSBP), co-clustering prob. ≈ 0.33