Flexible and Efficient Ordinal Regression with Bayesian Nonparametrics

Jizhou Kang

May 31st, 2024

Department of Statistics, UC Santa Cruz

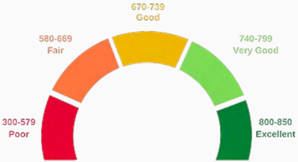
Ph.D. Oral Defense

Background

- Ordinal data are widely encountered in many fields:

Background

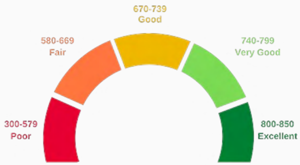
- Ordinal data are widely encountered in many fields:



Personal credit rating

Background

- Ordinal data are widely encountered in many fields:

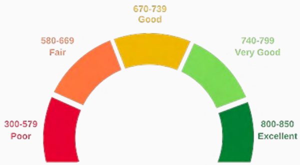


Personal credit rating

Symptom of insomnia

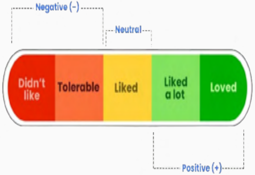
Background

- Ordinal data are widely encountered in many fields:



Personal credit rating

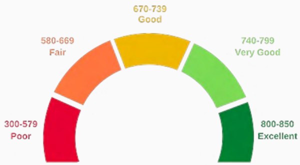
Symptom of insomnia



Attitude to a bill

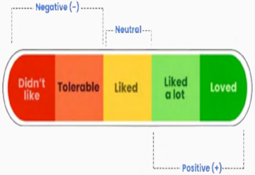
Background

- Ordinal data are widely encountered in many fields:



Personal credit rating

Symptom of insomnia

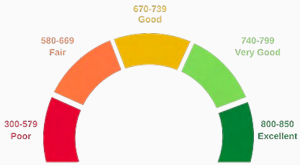


Attitude to a bill

- Ordinal responses together with covariates form up the ordinal regression problem;

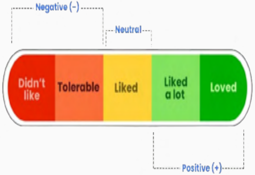
Background

- Ordinal data are widely encountered in many fields:



Personal credit rating

Symptom of insomnia



Attitude to a bill

- Ordinal responses together with covariates form up the ordinal regression problem;
- Study settings: cross-sectional and longitudinal studies.

Objectives

- Flexible and efficient ordinal regression modeling:
  - Flexible: allow general forms for ordinal response distribution, ordinal regression relationship, and temporal dependence in longitudinal settings;
  - Efficient: demand fewer computational resources and less tuning sophistication;

Objectives

- Flexible and efficient ordinal regression modeling:
  - Flexible: allow general forms for ordinal response distribution, ordinal regression relationship, and temporal dependence in longitudinal settings;
  - Efficient: demand fewer computational resources and less tuning sophistication;

- Challenges:
  - Incorporate the ordinal discrete nature of the responses;
  - Balance between model flexibility and implementation efficiency;
  - Features of a specific problem: heterogeneous effects, missingness, overdispersion, etc.;

Objectives

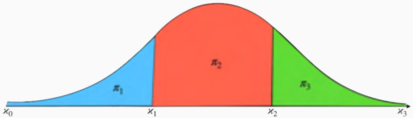
- Flexible and efficient ordinal regression modeling:
  - Flexible: allow general forms for ordinal response distribution, ordinal regression relationship, and temporal dependence in longitudinal settings;
  - Efficient: demand fewer computational resources and less tuning sophistication;
- Challenges:
  - Incorporate the ordinal discrete nature of the responses;
  - Balance between model flexibility and implementation efficiency;
  - Features of a specific problem: heterogeneous effects, missingness, overdispersion, etc.;
- Use Bayesian nonparametrics (BNP)! Flexibility is a designed virtue of BNP models, and efficient implementation techniques have been developed for BNP models.

Latent variable models for ordinal responses

- Modeling a univariate ordinal response Y with C categories. Encode the response as $\mathbf{Y} = (Y_1, \cdots, Y_C)$, such that $Y = j$ if-f $Y_j = 1$ and $Y_k = 0$, for any $k \neq j$;
- It is typically assumed $\mathbf{Y} \sim Mult(1, \pi_1, \cdots, \pi_C)$. The remaining task is how to acknowledge the order of the categories;

Latent variable models for ordinal responses

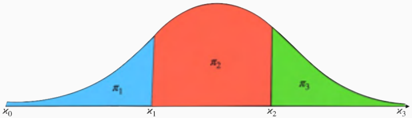
- Modeling a univariate ordinal response Y with C categories. Encode the response as $\mathbf{Y} = (Y_1, \cdots, Y_C)$, such that $Y = j$ if-f $Y_j = 1$ and $Y_k = 0$, for any $k \neq j$;
- It is typically assumed $\mathbf{Y} \sim Mult(1, \pi_1, \cdots, \pi_C)$. The remaining task is how to acknowledge the order of the categories;
- Cumulative link model (McCullagh, 1980):
  - A single latent continuous variable Z;
  - Cut-off points:
    $-\infty = \varkappa_0 < \varkappa_1 < \ldots < \varkappa_{C-1} < \varkappa_C = \infty$;
  - $Y = j$ if-f $Z \in (\varkappa_{j-1}, \varkappa_j]$.



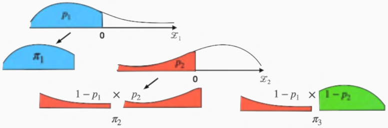
Latent variable models for ordinal responses

- Modeling a univariate ordinal response Y with C categories. Encode the response as $\mathbf{Y} = (Y_1, \cdots, Y_C)$, such that $Y = j$ if-f $Y_j = 1$ and $Y_k = 0$, for any $k \neq j$;
- It is typically assumed $\mathbf{Y} \sim Mult(1, \pi_1, \cdots, \pi_C)$. The remaining task is how to acknowledge the order of the categories;

- Cumulative link model (McCullagh, 1980):
  - A single latent continuous variable Z;
  - Cut-off points: $-\infty = \varkappa_0 < \varkappa_1 < \ldots < \varkappa_{C-1} < \varkappa_C = \infty$;
  - $Y = j$ if-f $Z \in (\varkappa_{j-1}, \varkappa_j]$.



- Sequential model (Tutz, 1991):
  - $C - 1$ latent continuous variable $(\mathcal{Z}_1, \cdots, \mathcal{Z}_{C-1})$;
  - Binary split for each $\mathcal{Z}_j$;
  - $Y = j$ if-f $\mathcal{Z}_j > 0$ and $\mathcal{Z}_k \leq 0$, $k = 1, \cdots, j - 1$.



Which structure should we use?

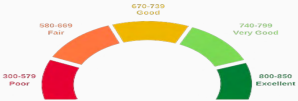
- The two approaches are compatible in modeling ordinal responses. There are scenarios where the two model formulations are equivalent;

Which structure should we use?

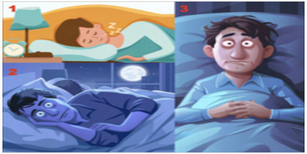
- The two approaches are compatible in modeling ordinal responses. There are scenarios where the two model formulations are equivalent;

- Choose a specific structure:

- Cumulative link model:
  - Determine response scale:

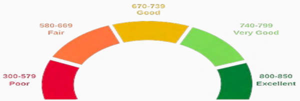


- Sequential model:
  - Inference for the covariate effect on conditional probability:

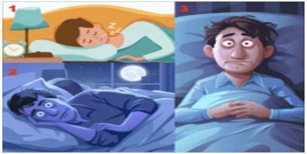


Which structure should we use?

- The two approaches are compatible in modeling ordinal responses. There are scenarios where the two model formulations are equivalent;
- Choose a specific structure:
- Cumulative link model:
    - Determine response scale:



    - Incorporate certain prior beliefs;
    - Study theoretical properties.
- Sequential model:
    - Inference for the covariate effect on conditional probability:



    - Flexible model for the ordinal regression relationship;
    - Efficient implementation through parallel computing.

Outline

- Project 1: Structured Mixture of Continuation-ratio Logits Models for Ordinal Regression
- Project 2: Bayesian Nonparametric Methods for Risk Assessment in Developmental Toxicity Studies with Ordinal Responses

Outline

- Project 1: Structured Mixture of Continuation-ratio Logits Models for Ordinal Regression
- Project 2: Bayesian Nonparametric Methods for Risk Assessment in Developmental Toxicity Studies with Ordinal Responses

---

- Project 3: Flexible Bayesian Modeling for Longitudinal Binary and Ordinal Responses
- Project 4: A Case Study: Estimating Maturity of Sheepshead Minnows

Models for Cross-sectional Ordinal Regression

Motivation

- In ordinal regression problems, flexible inference methods need to capture the covariate-response relationship, as well as incorporate the ordinal discrete nature of the responses;
- The covariate-response relationship is depicted by the probability response curves;

Motivation

- In ordinal regression problems, flexible inference methods need to capture the covariate-response relationship, as well as incorporate the ordinal discrete nature of the responses;
- The covariate-response relationship is depicted by the probability response curves;
- Challenges:
    - Non-standard response distributions;
    - Non-standard regression relationships between the ordinal response categories and covariates;
    - In terms of computation, the proposed method should have tractable inference algorithm.

Summary of existing literature

	Cumulative link model	Sequential model
Parametric model	Probit regression (Albert & Chib, 1993)	Logits regression family (Agresti, 2013), continuation-ratio logits models (Tutz, 1991)
Semiparametric model	Relaxing normality assumption (Newton et al., 1996), linearity assumption (Mukhopadhyay & Gelfand, 1997), or both (Chib & Greenberg, 2010)	Replacing systematic component with Gaussian process (Linderman et al., 2015), Adding random effects term with DP prior to the systematic component (Tang & Duan, 2012)
Nonparametric model	Bayesian density estimation for the joint distribution of covariates and responses, for categorical variables (Shahbaba & Neal, 2009; Dunson & Bhattacharya, 2010), and ordinal variables (DeYoreo & Kottas, 2018)	Common-atoms DDP model for specific type of problems (Fronczyk & Kottas, 2013)

General modeling framework

- The general logit stick-breaking process (LSBP) model for ordinal regression:

$$\mathbf{Y} | G_{\mathbf{x}} \sim \int K(\mathbf{Y} | \mathbf{m}, \theta) dG_{\mathbf{x}}(\theta), \quad G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_{\ell}(\mathbf{x}) \delta_{\boldsymbol{\theta}_{\ell}(\mathbf{x})}$$

General modeling framework

- The general logit stick-breaking process (LSBP) model for ordinal regression:

$$\mathbf{Y}|G_{\mathbf{x}} \sim \int K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta})dG_{\mathbf{x}}(\boldsymbol{\theta}), \quad G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x})\delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$$

  - Continuation-ratio logits as the kernel: $K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}(\mathbf{x})) = Bin(Y_1 \mid m_1, \varphi(\theta_1(\mathbf{x}))) \cdots Bin(Y_{C-1} \mid m_{C-1}, \varphi(\theta_{C-1}(\mathbf{x})))$, where $\varphi(x) = e^x/(e^x + 1)$;

General modeling framework

- The general logit stick-breaking process (LSBP) model for ordinal regression:

$$\mathbf{Y}|G_{\mathbf{x}} \sim \int K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}) dG_{\mathbf{x}}(\boldsymbol{\theta}), \quad G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x}) \delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$$

  - Continuation-ratio logits as the kernel: $K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}(\mathbf{x})) = Bin(Y_1 \mid m_1, \varphi(\theta_1(\mathbf{x}))) \cdots Bin(Y_{C-1} \mid m_{C-1}, \varphi(\theta_{C-1}(\mathbf{x})))$, where $\varphi(x) = e^x/(e^x + 1)$;

  - The weights $\omega_\ell(\mathbf{x})$ are generated by LSBP: $\omega_1(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_1)$ and $\omega_\ell(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_\ell) \prod_{h=1}^{\ell-1}(1 - \varphi(\mathbf{x}^T\gamma_h))$, $\ell = 2, 3, \cdots$;

General modeling framework

- The general logit stick-breaking process (LSBP) model for ordinal regression:

$$\mathbf{Y}|G_{\mathbf{x}} \sim \int K(\mathbf{Y}|\mathbf{m},\boldsymbol{\theta})dG_{\mathbf{x}}(\boldsymbol{\theta}), \quad G_{\mathbf{x}} = \sum_{\ell=1}^{\infty}\omega_\ell(\mathbf{x})\delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$$

  - Continuation-ratio logits as the kernel: $K(\mathbf{Y}|\mathbf{m},\boldsymbol{\theta}(\mathbf{x})) = Bin(Y_1 \mid m_1, \varphi(\theta_1(\mathbf{x})))\cdots Bin(Y_{C-1} \mid m_{C-1}, \varphi(\theta_{C-1}(\mathbf{x})))$, where $\varphi(x) = e^x/(e^x+1)$;

  - The weights $\omega_\ell(\mathbf{x})$ are generated by LSBP: $\omega_1(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_1)$ and $\omega_\ell(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_\ell)\prod_{h=1}^{\ell-1}(1-\varphi(\mathbf{x}^T\gamma_h))$, $\ell = 2, 3, \cdots$;

  - The atoms, $\boldsymbol{\theta}_\ell(\mathbf{x}) = \{\theta_{j\ell}(\mathbf{x}) : j = 1, \cdots, C-1\}$, have linear regression structure, $\theta_{j\ell}(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_{j\ell} \overset{ind.}{\sim} N(\mathbf{x}^T\boldsymbol{\mu}_j, \mathbf{x}^T\Sigma_j\mathbf{x})$, and are independent across ℓ;

General modeling framework

- The general logit stick-breaking process (LSBP) model for ordinal regression:

$$\mathbf{Y}|G_{\mathbf{x}} \sim \int K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}) dG_{\mathbf{x}}(\boldsymbol{\theta}), \quad G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x}) \delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$$

  - Continuation-ratio logits as the kernel: $K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}(\mathbf{x})) = Bin(Y_1 \mid m_1, \varphi(\theta_1(\mathbf{x}))) \cdots Bin(Y_{C-1} \mid m_{C-1}, \varphi(\theta_{C-1}(\mathbf{x})))$, where $\varphi(x) = e^x/(e^x + 1)$;
  - The weights $\omega_\ell(\mathbf{x})$ are generated by LSBP: $\omega_1(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_1)$ and $\omega_\ell(\mathbf{x}) = \varphi(\mathbf{x}^T\gamma_\ell) \prod_{h=1}^{\ell-1}(1 - \varphi(\mathbf{x}^T\gamma_h))$, $\ell = 2, 3, \cdots$;
  - The atoms, $\boldsymbol{\theta}_\ell(\mathbf{x}) = \{\theta_{j\ell}(\mathbf{x}) : j = 1, \cdots, C-1\}$, have linear regression structure, $\theta_{j\ell}(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}_{j\ell} \overset{ind.}{\sim} N(\mathbf{x}^T\boldsymbol{\mu}_j, \mathbf{x}^T\Sigma_j\mathbf{x})$, and are independent across ℓ;
  - Hyperprior: $\gamma_\ell \overset{i.i.d.}{\sim} N(\gamma_0, \Gamma_0)$ and $\boldsymbol{\mu}_j|\Sigma_j \overset{ind.}{\sim} N(\boldsymbol{\mu}_j|\boldsymbol{\mu}_{0j}, \Sigma_j/\kappa_{0j})$, $\Sigma_j \overset{ind.}{\sim} IW(\Sigma_j|\nu_{0j}, \Lambda_{0j}^{-1})$.

Model structure illustration

- The continuation-ratio logits structure offers a sequential mechanism to allocate the ordinal response Y:

Generate $\mathcal{H}_1 \sim Bern(\Delta_1), \Delta_1 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_1), \boldsymbol{\beta}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$

$\mathcal{H}_1 = 1$ $\quad$ $\mathcal{H}_1 = 0$

$Y = 1$ $\quad$ Generate $\mathcal{H}_2 \sim Bern(\Delta_2), \Delta_2 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_2), \boldsymbol{\beta}_2 \sim N(\boldsymbol{\mu}_2, \Sigma_2)$

$\mathcal{H}_2 = 1$ $\quad$ $\mathcal{H}_2 = 0$

$Y = 2$ $\quad$ Generate $\mathcal{H}_3 \sim Bern(\Delta_3), \Delta_3 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_3), \boldsymbol{\beta}_3 \sim N(\boldsymbol{\mu}_3, \Sigma_3)$

$\mathcal{H}_3 = 1$ $\quad$ $\mathcal{H}_3 = 0$

$Y = 3$ $\quad$ $\cdots$

Model structure illustration

- The continuation-ratio logits structure offers a sequential mechanism to allocate the ordinal response Y:

Generate $\mathcal{H}_1 \sim Bern(\Delta_1), \Delta_1 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_1), \boldsymbol{\beta}_1 \sim N(\boldsymbol{\mu}_1, \Sigma_1)$

$\mathcal{H}_1 = 1$ $\mathcal{H}_1 = 0$

$Y = 1$ Generate $\mathcal{H}_2 \sim Bern(\Delta_2), \Delta_2 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_2), \boldsymbol{\beta}_2 \sim N(\boldsymbol{\mu}_2, \Sigma_2)$

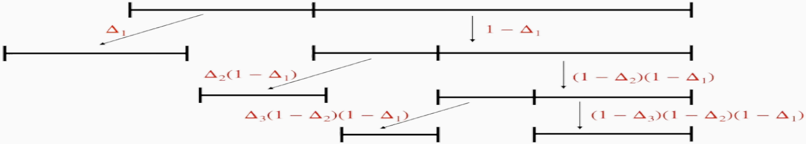
$\mathcal{H}_2 = 1$ $\mathcal{H}_2 = 0$

$Y = 2$ Generate $\mathcal{H}_3 \sim Bern(\Delta_3), \Delta_3 = \varphi(\mathbf{x}^T \boldsymbol{\beta}_3), \boldsymbol{\beta}_3 \sim N(\boldsymbol{\mu}_3, \Sigma_3)$

$\mathcal{H}_3 = 1$ $\mathcal{H}_3 = 0$

$Y = 3$ $\cdots$

- The stick-breaking weights are also determined by it.



Simplified models

- **Common-weights**: defining the weights through the stick-breaking process that defines DP, we obtain the common-weights model:
  - $G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$;
  - $\eta_\ell \overset{i.i.d.}{\sim} Beta(1, \alpha)$, $\omega_1 = \eta_1$ and $\omega_\ell = \eta_\ell \prod_{h=1}^{\ell-1}(1 - \eta_\ell)$, for $\ell = 2, 3, \cdots$;
  - $\boldsymbol{\theta}_\ell(\mathbf{x})$ are defined as in the general model;

Simplified models

- **Common-weights**: defining the weights through the stick-breaking process that defines DP, we obtain the common-weights model:
  - $G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\boldsymbol{\theta}_\ell(\mathbf{x})}$;
  - $\eta_\ell \overset{i.i.d.}{\sim} Beta(1, \alpha)$, $\omega_1 = \eta_1$ and $\omega_\ell = \eta_\ell \prod_{h=1}^{\ell-1}(1 - \eta_\ell)$, for $\ell = 2, 3, \cdots$;
  - $\boldsymbol{\theta}_\ell(\mathbf{x})$ are defined as in the general model;

- **Common-atoms**: we formulate the common-atoms model:
  - $G_{\mathbf{x}} = \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x}) \delta_{\boldsymbol{\theta}_\ell}$;
  - $\theta_{j\ell} | \mu_j, \sigma_j^2 \overset{ind.}{\sim} N(\mu_j, \sigma_j^2)$;
  - $\omega_\ell(\mathbf{x})$ are determined as in the general model.

Flexible ordinal regression relationships

Model properties

The general model allows flexible estimate of the probability response curves.

1. Marginal regression relationships:

$$Pr(\mathbf{Y} = j | G_{\mathbf{x}}) = \sum_{\ell=1}^{\infty} \omega_{\ell}(\mathbf{x}) \left\{ \varphi(\theta_{j\ell}(\mathbf{x})) \prod_{k=1}^{j-1} [1 - \varphi(\theta_{k\ell}(\mathbf{x}))] \right\}$$

2. Conditional regression relationships:

$$Pr(\mathbf{Y} = j | \mathbf{Y} \geq j, G_{\mathbf{x}}) = \sum_{\ell=1}^{\infty} w_{j\ell}(\mathbf{x}) \left\{ \varphi(\theta_{j\ell}(\mathbf{x})) \right\}$$

where $w_{j\ell}(\mathbf{x}) = \frac{\omega_{\ell}(\mathbf{x}) \prod_{k=1}^{j-1} [1 - \varphi(\theta_{k\ell}(\mathbf{x}))]}{\sum_{\ell=1}^{\infty} \omega_{\ell}(\mathbf{x}) \prod_{k=1}^{j-1} [1 - \varphi(\theta_{k\ell}(\mathbf{x}))]}$.

- Both have a weighted sum form with locally adjustable weights.

Formal assessment of model flexibility

Full Kullback-Leibler (KL) support of the proposed model

Denote by $\mathcal{P}_{\mathbf{x}}$ the prior induced by the mixture model, and consider $\{p^0_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$, a generic collection of covariate-dependent probabilities for an ordinal response with C categories. Assume that the probability of each response category is strictly positive. Then, the mass functions $\{p^0_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ are in the KL support of $\mathcal{P}_{\mathbf{x}}$.

Formal assessment of model flexibility

Full Kullback-Leibler (KL) support of the proposed model

Denote by $\mathcal{P}_{\mathbf{x}}$ the prior induced by the mixture model, and consider $\{p^0_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$, a generic collection of covariate-dependent probabilities for an ordinal response with C categories. Assume that the probability of each response category is strictly positive. Then, the mass functions $\{p^0_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ are in the KL support of $\mathcal{P}_{\mathbf{x}}$.

- **Sketch of the proof:**
  - Leverage the existing results regarding the KL support of prior for continuous responses (Barrientos et al. 2012);
  - Formulate the ordinal LSBP mixture model in terms of latent continuous responses;
  - Find the connection between the KL support of the prior for continuous responses and the induced prior for categorical outcomes arising from discretizing the continuous responses.

Model implementation

- **Prior specification:**
  - Conjugate hyperprior, $\gamma_\ell \overset{i.i.d.}{\sim} N(\gamma_0, \Gamma_0)$, $\boldsymbol{\mu}_j|\Sigma_j \overset{ind.}{\sim} N(\boldsymbol{\mu}_{0j}, \Sigma_j/\kappa_{0j})$, $\Sigma_j \overset{ind.}{\sim} IW(\nu_{0j}, \Lambda_{0j}^{-1})$;
  - "Baseline choice" of hyperparameter: $\boldsymbol{\mu}_{0j} = \gamma_0 = \mathbf{0}_p$, $\Sigma_j = \Gamma_0 = 10^2\mathbf{I}_p$, and $\kappa_{0j} = \nu_{0j} = p + 2$;
  - Under the baseline prior, $E(Pr(\mathbf{Y} = j|G_{\mathbf{x}})) \equiv 2^{-j}$, $j = 1, \cdots, C-1$, and $E(Pr(\mathbf{Y} = C|G_{\mathbf{x}})) \equiv 2^{-(C-1)}$;

Model implementation

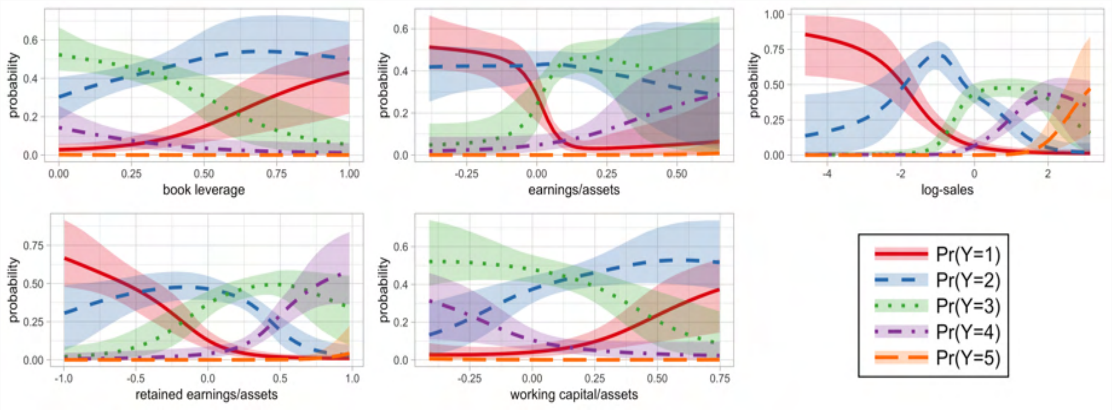
- **Prior specification:**
  - Conjugate hyperprior, $\gamma_\ell \overset{i.i.d.}{\sim} N(\gamma_0, \Gamma_0)$, $\boldsymbol{\mu}_j|\Sigma_j \overset{ind.}{\sim} N(\boldsymbol{\mu}_{0j}, \Sigma_j/\kappa_{0j})$, $\Sigma_j \overset{ind.}{\sim} IW(\nu_{0j}, \Lambda_{0j}^{-1})$;
  - "Baseline choice" of hyperparameter: $\boldsymbol{\mu}_{0j} = \gamma_0 = \mathbf{0}_p$, $\Sigma_j = \Gamma_0 = 10^2\mathbf{I}_p$, and $\kappa_{0j} = \nu_{0j} = p + 2$;
  - Under the baseline prior, $E(Pr(\mathbf{Y} = j|G_\mathbf{x})) \equiv 2^{-j}$, $j = 1, \cdots, C-1$, and $E(Pr(\mathbf{Y} = C|G_\mathbf{x})) \equiv 2^{-(C-1)}$;

- **Posterior inference:**
  - Blocked Gibbs sampler: truncating $G_\mathbf{x}$ at a large enough level L and introducing latent configuration variable $\mathcal{L}_i$ for $i = 1, \cdots, n$;
  - Pólya-Gamma augmentation: introduce two groups of Pólya-Gamma latent variables for the weights and atoms;
  - Same structure for the weights and atoms: same sampling strategy;
  - All model parameters can be sampled via Gibbs sampling.

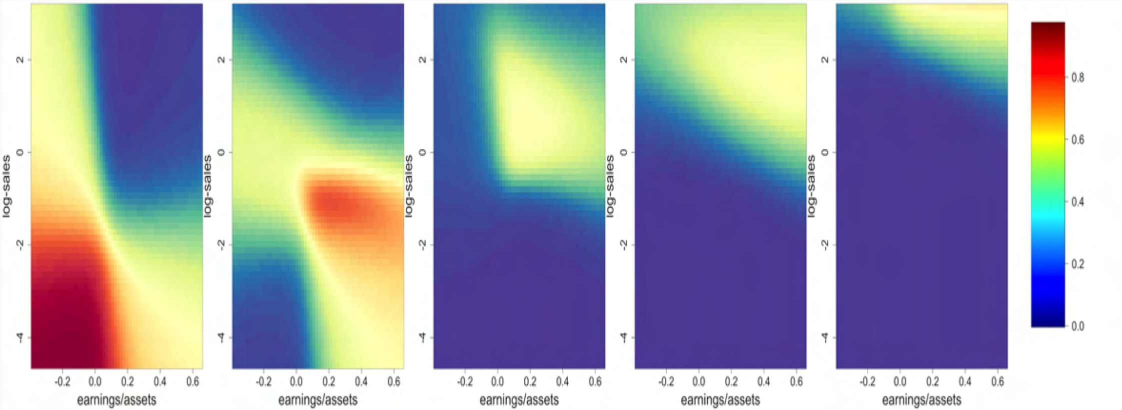
Real Data Application (Credit Ratings of U.S. Companies)

- Standard and Poor's (S&P) credit ratings for 921 U.S. firms;
- For each firm, a credit rating on a seven-point ordinal scale is available, along with five characteristics;
- Combined the first two categories and the last two categories to produce an ordinal response with five levels;
- The covariates are: (1) book leverage X_1, (2) earnings before interest and taxes divided by total assets X_2, (3) standardized log-sales X_3, (4) retained earnings divided by total assets X_4, (5) working capital divided by total assets X_5;
- Quantities of interest: the first and second order marginal probability curves $Pr(\mathbf{Y} = j | G_{\mathbf{x}}; \mathbf{x}_{\mathbf{s}})$ for $j = 1, \cdots, 5$ and $\mathbf{s} \in \{1, \cdots, 5\}$.

First order marginal probability curves



Second order marginal probability surfaces



Summary and Discussion

- We propose a unified toolbox for ordinal regression by directly modeling the discrete response distribution. The virtues of the proposed models rely on the following key ingredients:
  1. Continuation-ratio logits representation;
  2. Pólya-Gamma data augmentation technique;
  3. Logit stick-breaking process prior;
- Some practical suggestions in picking the model:
  1. Common-weights model: the most parsimonious formulation with practically sufficient flexibility;
  2. Common-atoms model: a more appropriate choice when expecting complicated covariate-response relationships;
  3. General model: the most versatile structure, benefits especially in applications involving sufficiently large amounts of data and non-standard regression relationships.

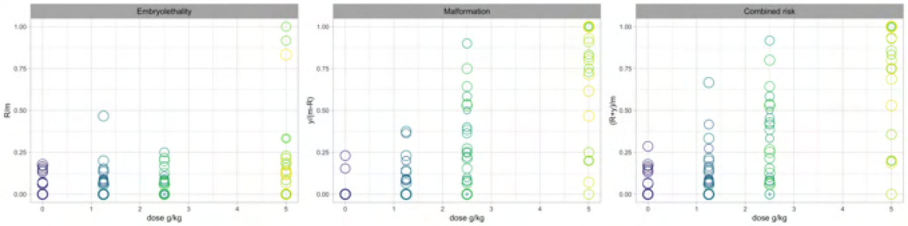
Models for Ordinal Regression with Heterogeneous Responses

Data structure from developmental toxicity study

- Data at dose levels, x_d, $d = 1, \cdots, N$, including a control group (dose$= 0$);
- n_d pregnant laboratory animals (dams) at dose level x_d;
- For the i-th dam at dose x_d:
  - m_{di}: number of implants;
  - R_{di}: number of resorptions and prenatal deaths;
  - y_{di}: number of live pups with a malformation;
- The ordinal responses are $\mathbf{Y}_{di} = (R_{di}, y_{di}, m_{di} - R_{di} - y_{di})$, which can be equivalently encoded by standard ordinal responses $\{\tilde{\mathbf{Y}}_{diq} = (\tilde{R}_{diq}, \tilde{y}_{diq}, 1 - \tilde{R}_{diq} - \tilde{y}_{diq})\}$, for $q = 1, \cdots, m_{di}$, such that $\mathbf{Y}_{di} = \sum_{q=1}^{m_{di}} \tilde{\mathbf{Y}}_{diq}$;
- Focus on the dose-response curves of the clustered categorical endpoints, embryolethality $D(x) = \Pr(\tilde{R} = 1 \mid x)$, fetal malformation for live pups $M(x) = \Pr(\tilde{y} = 1 \mid \tilde{R} = 0, x)$, and combined negative outcomes $r(x) = \Pr(\tilde{R} = 1 \,\text{or}\, \tilde{y} = 1 \mid x)$.

Motivating example

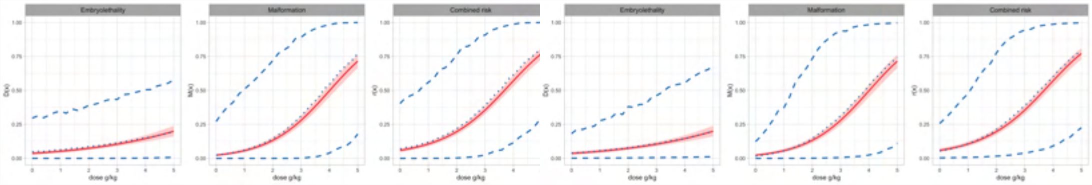
- Data from a development toxicity study that evaluates the toxic effects of ethylene glycol (EG), using pregnant rats;



- Features of the data:
  - an overall increasing trend, with no obvious parametric form to model it;
  - vast variability in the responses, of which the magnitude also differs across dose levels;
  - a potentially different dose-response relationship for non-viable fetuses and malformed pups.

Parametric continuous mixture models

- Beta-Binomial distribution $BB(m, \theta, \lambda)$, Logistic-Normal-Binomial distribution $LNB(m, \theta, \sigma^2)$;
- Postulating the sequential mechanism of the ordinal responses, we have
    - "CR-BB" model: $(R, y) \mid m, \theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \boldsymbol{\lambda} \sim BB(R \mid m, \theta_1(\mathbf{x}), \lambda_1) BB(y \mid m - R, \theta_2(\mathbf{x}), \lambda_2)$;
    - "CR-LNB" model: $(R, y) \mid m, \theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \boldsymbol{\sigma}^2 \sim LNB(R \mid m, \theta_1(\mathbf{x}), \sigma_1^2) LNB(y \mid m - R, \theta_2(\mathbf{x}), \sigma_2^2)$;
- Posterior inference for the dose-response curves:



"CR-BB model" "CR-LNB model"

Nonparametric discrete mixture models

- The general model proposed in the last section can be applied here ("Gen-Bin" model):

$$(R, y) \mid m, G_{\mathbf{x}} \sim \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x})\, Bin(R \mid m, \varphi(\theta_{1\ell}(\mathbf{x})))Bin(y \mid m - R, \varphi(\theta_{2\ell}(\mathbf{x})));$$

- The common-weights model is also applicable ("CW-Bin" model), while the common-atoms model is not because its induced prior expected dose-response curves cannot have monotone shape;
- Based on the results for probability response curves, the induced dose-response curves under these models have flexible shapes;
- We can also establish a positive intracluster correlation result, demonstrating that the model enables overdispersion.

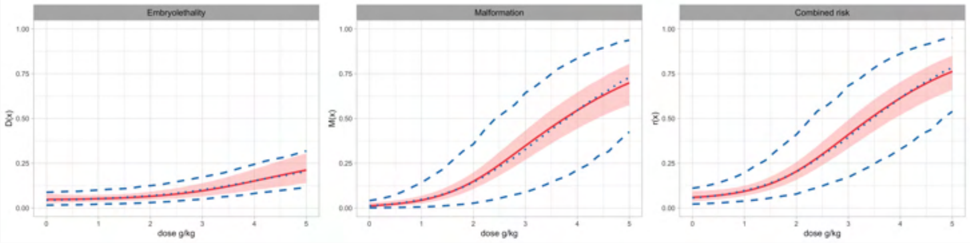
Nonparametric mixture models with overdispersed kernel

- We consider a combination of these two types of mixture models, which potentially combines the advantage of these two types of models;
- We adopt the LNB model as the kernel, which is then encapsulated in the general nonparametric mixing structure ("Gen-LNB" model)

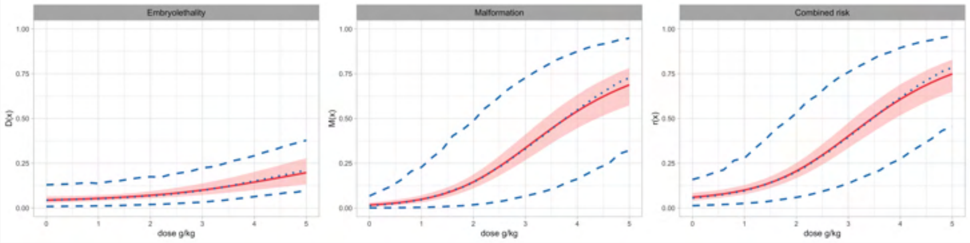
$$(R, y) \mid m, G_{\mathbf{x}}, \boldsymbol{\sigma}^2 \sim \sum_{\ell=1}^{\infty} \omega_\ell(\mathbf{x})\, LNB(R \mid m, \theta_{1\ell}(\mathbf{x}), \sigma_1^2) LNB(y \mid m - R, \theta_{2\ell}(\mathbf{x}), \sigma_2^2);$$

- The mixing structure is inherited from the "Gen-Bin" model, rendering the computational techniques developed for it readily adaptable here. We do not use BB as the kernel because it breaks the developed Gibbs sampling scheme;
- Similarly, we can formulate the model with a common-weights mixing structure and the product of LNB kernel ("CW-LNB" model).

Posterior estimate of dose-response curves



“CW-Bin” model vs “CW-LNB” model



“Gen-Bin” model vs “Gen-LNB” model.

Model comparison

- We perform model comparison based on posterior predictive loss (PPL) and interval score (IS);
- We use one randomly chosen sample comprising roughly 20% of the data as the test set, denoted by $\{(m'_{di}, R'_{di}, y'_{di}) : d = 1, \ldots, N, i = 1, \ldots, n'_d\}$;
- Fitting the model to the reduced data, and obtain posterior predictive samples at each observed dose level, denoted as m^*_d, R^*_d, and y^*_d;
- The criteria are defined separately for embryolethality (R/m), malformation $(y/(m-R))$, and combined risk $(R+y)/m$. Using embryolethality as an example:
  - PPL, goodness-of-fit: $G(\mathcal{M}) = \sum_{d=1}^{N} \sum_{i=1}^{n'_d} \{R'_{di}/m'_{di} - \mathrm{E}(R^*_d/m^*_d \mid \text{data})\}$;
  - PPL, penalty: $P(\mathcal{M}) = \sum_{d=1}^{N} n'_d \mathrm{Var}(R^*_d/m^*_d \mid \text{data})$;
  - IS: $S(\mathcal{M}) = \sum_{d=1}^{N} \sum_{i=1}^{n'_d} \{(u^e_d - l^e_d) + \frac{2}{\alpha}(l^e_d - \frac{R'_{di}}{m'_{di}})\mathbf{1}(\frac{R'_{di}}{m'_{di}} < l^e_d) + \frac{2}{\alpha}(\frac{R'_{di}}{m'_{di}} - u^e_d)\mathbf{1}(\frac{R'_{di}}{m'_{di}} > u^e_d)\}$, where l^e_d and u^e_d denote the limits of the 95% posterior predictive credible interval.

Summary of model comparison results

Endpoint	Criterion	"CW-Bin"	"CW-LNB"	"Gen-Bin"	"Gen-LNB"
Embryolethality	$G(\mathcal{M})$	0.72	0.72	0.71	0.72
	$P(\mathcal{M})$	0.56	0.53	0.45	0.58
	$S(\mathcal{M})$	20.73	18.45	20.46	18.73
Malformation	$G(\mathcal{M})$	1.34	1.39	1.33	1.36
	$P(\mathcal{M})$	1.18	1.10	0.95	1.17
	$S(\mathcal{M})$	16.07	14.97	16.81	14.93
Combined risk	$G(\mathcal{M})$	1.46	1.50	1.43	1.49
	$P(\mathcal{M})$	1.08	1.01	0.89	1.03
	$S(\mathcal{M})$	25.84	23.50	27.11	20.91

Remarks

- The parametric continuous mixture models fail in providing reliable uncertainty quantification for the dose-response curves;
- Contrarily, nonparametric discrete mixture models, with enhanced flexibility, offer rich inference for the response distributions and for the dose-response curves;
- The key advantage of incorporating overdispersed kernel within a nonparametric mixture model lies in improved posterior predictive interval estimation;
- The modeling approaches examined here are directly applicable in other areas, which may involve more ordered categories and/or more covariates.

Models for Longitudinal Binary and Ordinal Responses

Motivation and objectives

- Focus on longitudinal studies with binary outcome, then extend the method to deal with longitudinal studies with ordinal outcome;

Motivation and objectives

- Focus on longitudinal studies with binary outcome, then extend the method to deal with longitudinal studies with ordinal outcome;

- The main quantities of interests in such a study are:
  - the probability response curve;
  - the lead-lag correlations among repeated measurements;

Motivation and objectives

- Focus on longitudinal studies with binary outcome, then extend the method to deal with longitudinal studies with ordinal outcome;
- The main quantities of interests in such a study are:
  - the probability response curve;
  - the lead-lag correlations among repeated measurements;
- Motivating application: ecological momentary assessment (EMA) studies, which involve the repeated measuring of people's current thoughts, emotions, behavior, and physiological states, in their natural environment. Non-response is inevitable.

A taxonomic review of models

- Marginal models: Molenberghs and Verbeke (2006);
- Conditional models: Di Lucca et al. (2013), DeYoreo and Kottas (2018);
- Subject-specific models:
  - Continuous: Ghosh and Hanson (2010); Quintana et al. (2016);
  - Binary: Jara et al. (2007); Tang and Duan (2012);
  - Mixed-scale: Kunihama et al. (2019);
- Functional data analysis tools: functional principal components analysis Van Der Linde (2009); Matuk et al. (2022).

Hierarchical formulation of the proposed model

- Adopt a functional data perspective, treating each observed data vector $\mathbf{Y}_i$ as the evaluation of trajectory $Y_i(\tau)$ on grid $\boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i})^\top$, for $i = 1, \cdots, n$;

Hierarchical formulation of the proposed model

- Adopt a functional data perspective, treating each observed data vector $\mathbf{Y}_i$ as the evaluation of trajectory $Y_i(\tau)$ on grid $\boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i})^\top$, for $i = 1, \cdots, n$;
- At the observed data level, we assume:

$$Y_i(\tau_{it}) \mid Z_i(\tau_{it}), \epsilon_{it} \overset{ind.}{\sim} Bin(1, \varphi(Z_i(\tau_{it}) + \epsilon_{it})), \quad t = 1, \cdots, T_i, \quad i = 1, \cdots, n,$$

where $\varphi(x) = \exp(x)/\{1 + \exp(x)\}$, and the error terms $\epsilon_{it} \mid \sigma_\epsilon^2 \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$;

Hierarchical formulation of the proposed model

- Adopt a functional data perspective, treating each observed data vector $\mathbf{Y}_i$ as the evaluation of trajectory $Y_i(\tau)$ on grid $\boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i})^\top$, for $i = 1, \cdots, n$;
- At the observed data level, we assume:

$$Y_i(\tau_{it}) \mid Z_i(\tau_{it}), \epsilon_{it} \overset{ind.}{\sim} Bin(1, \varphi(Z_i(\tau_{it}) + \epsilon_{it})), \quad t = 1, \cdots, T_i, \quad i = 1, \cdots, n,$$

  where $\varphi(x) = \exp(x)/\{1 + \exp(x)\}$, and the error terms $\epsilon_{it} \mid \sigma_\epsilon^2 \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$;
- The main building block for the model construction is a hierarchical Gaussian process prior for $Z_i(\cdot)$, which we termed the signal process.

$$Z_i \mid \mu, \Sigma \overset{i.i.d.}{\sim} GP(\mu, \Sigma), \quad \mu \mid \Sigma \sim GP(\mu_0, \Sigma/\kappa), \quad \Sigma \sim IWP(\nu, \Psi_\phi).$$

  Specifically we set $\kappa = (\nu - 3)^{-1}$;

Hierarchical formulation of the proposed model

- Adopt a functional data perspective, treating each observed data vector $\mathbf{Y}_i$ as the evaluation of trajectory $Y_i(\tau)$ on grid $\boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i})^\top$, for $i = 1, \cdots, n$;
- At the observed data level, we assume:

$$Y_i(\tau_{it}) \mid Z_i(\tau_{it}), \epsilon_{it} \overset{ind.}{\sim} Bin(1, \varphi(Z_i(\tau_{it}) + \epsilon_{it})), \quad t = 1, \cdots, T_i, \quad i = 1, \cdots, n,$$

  where $\varphi(x) = \exp(x)/\{1 + \exp(x)\}$, and the error terms $\epsilon_{it} \mid \sigma_\epsilon^2 \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$;
- The main building block for the model construction is a hierarchical Gaussian process prior for $Z_i(\cdot)$, which we termed the signal process.

$$Z_i \mid \mu, \Sigma \overset{i.i.d.}{\sim} GP(\mu, \Sigma), \quad \mu \mid \Sigma \sim GP(\mu_0, \Sigma/\kappa), \quad \Sigma \sim IWP(\nu, \Psi_\phi).$$

  Specifically we set $\kappa = (\nu - 3)^{-1}$;
- We use an Inverse-Wishart process (IWP) prior for the covariance kernel. It is defined such that, on any finite grid $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)$ the projection $\Sigma(\boldsymbol{\tau}, \boldsymbol{\tau})$ follows $IW(\nu, \Psi_\phi(\boldsymbol{\tau}, \boldsymbol{\tau}))$. Here, $\Psi_\phi(\cdot, \cdot)$ is a non-negative definite function with parameters ϕ.

Model properties

Proposition

Under the proposed model formulation, the signal process $Z(\tau)$ follows marginally a student-t process (TP). That is, for a generic grid vector $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)^\top$, $\mathbf{Z}_{\boldsymbol{\tau}} = Z(\boldsymbol{\tau}) \sim MVT(\nu, \boldsymbol{\mu}_{0\boldsymbol{\tau}}, \boldsymbol{\Psi}_{\boldsymbol{\tau},\boldsymbol{\tau}})$, where $\boldsymbol{\mu}_{0\boldsymbol{\tau}} = \mu_0(\boldsymbol{\tau})$, and $\boldsymbol{\Psi}_{\boldsymbol{\tau},\boldsymbol{\tau}} = \Psi_\phi(\boldsymbol{\tau}, \boldsymbol{\tau})$;

- TP is closed under marginalization. We can utilize the analytical form of the TP predictive distribution to develop a predictive inference scheme that resembles that of GP-based models. It is particularly useful in posterior inference;
- We can study the local behavior, such as smoothness, of the signal process trajectories by modeling them as TP;
- Modeling as TP facilitates the interpretation of the degrees of freedom parameter ν. It controls how heavy tailed the process is.

Highlights of the MCMC algorithm

- Recall that under unbalanced setting, the grid vectors for each subject $\boldsymbol{\tau}_i$ are different. We consider pooled grid $\boldsymbol{\tau} = \cup_{i=1}^{n} \boldsymbol{\tau}_i$;
- Let $\tilde{\mathbf{Z}}_i = Z_i(\boldsymbol{\tau})$, $\mathbf{Z}_i = Z_i(\boldsymbol{\tau}_i)$, and $\mathbf{Z}_i^* = \tilde{\mathbf{Z}}_i \setminus \mathbf{Z}_i$;
- Factorizing the prior of $\tilde{\mathbf{Z}}_i$ as $p(\tilde{\mathbf{Z}}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\mathbf{Z}_i^* \mid \mathbf{Z}_i, \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\mathbf{Z}_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$. In a MCMC iteration, we first sample $\mathbf{Z}_i$, then conditioning on $\mathbf{Z}_i$ to sample $\mathbf{Z}_i^*$ (GP-based predictive sampling);
- Binary response to continuous latent process with errors, $Y_i(\tau_{it}) \mid Z_i(\tau_{it}), \epsilon_{it} \overset{ind.}{\sim} Bin(1, \varphi(Z_i(\tau_{it}) + \epsilon_{it}))$, reminds us the Pólya-Gamma technique;
- All model parameters can be sampled via Gibbs sampling, with standard full condition distributions.

Prediction and uncertainty

- We can make predictions on any time grid. Consider predicting $\mathbf{Z}_i^+ = Z_i(\boldsymbol{\tau}^+)$, where $\boldsymbol{\tau}^+ \supset \boldsymbol{\tau}$ is a finer grid. Let $\check{\boldsymbol{\tau}} = \boldsymbol{\tau}^+ \setminus \boldsymbol{\tau}$ and $\check{\mathbf{Z}}_i = Z_i(\check{\boldsymbol{\tau}})$;
- We have the joint distribution:

$$\begin{pmatrix} \tilde{\mathbf{Z}}_i \\ \check{\mathbf{Z}}_i \end{pmatrix} \sim MVT\left(\nu, \begin{pmatrix} \boldsymbol{\mu}_{0\boldsymbol{\tau}} \\ \boldsymbol{\mu}_{0\check{\boldsymbol{\tau}}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\tau},\boldsymbol{\tau}} & \boldsymbol{\Psi}_{\boldsymbol{\tau},\check{\boldsymbol{\tau}}} \\ \boldsymbol{\Psi}_{\check{\boldsymbol{\tau}},\boldsymbol{\tau}} & \boldsymbol{\Psi}_{\check{\boldsymbol{\tau}},\check{\boldsymbol{\tau}}} \end{pmatrix}\right),$$

  and the prediction for $\check{\mathbf{Z}}_i$ are made based on the conditional distribution:

$$\check{\mathbf{Z}}_i \mid \tilde{\mathbf{Z}}_i \sim MVT\left(\nu + |\boldsymbol{\tau}|, \check{\boldsymbol{\mu}}_{i\check{\boldsymbol{\tau}}}, \frac{\nu + S_{i\boldsymbol{\tau}} - 2}{\nu + |\boldsymbol{\tau}| - 2}\check{\boldsymbol{\Psi}}_{\check{\boldsymbol{\tau}},\check{\boldsymbol{\tau}}}\right);$$

- For an in-sample subject, we first predict $Z_i(\boldsymbol{\tau}_i^*)$ conditioning on $Z_i(\boldsymbol{\tau}_i)$ by the GP predictive distribution, and next predict $Z_i(\check{\boldsymbol{\tau}})$ conditioning on $Z_i(\boldsymbol{\tau}_i)$ and $Z_i(\boldsymbol{\tau}_i^*)$ by the TP predictive distribution;
- TP is scaling the predictive covariance by a factor that is related to the prediction error on observed grid, which can adjust the predictive covariance at unobserved grid points.

Model extension to deal with longitudinal ordinal responses

- Suppose the observation on subject i at time τ_{it}, denoted by Y_{it}, takes C possible categories;
- We encode the response as a vector with binary entries $\mathbf{Y}_{it} = (Y_{i1t} \cdots, Y_{iCt})$, such that $Y_{it} = j$ is equivalent to $Y_{ijt} = 1$ and $Y_{ikt} = 0$ for any $k \neq j$;
- We assume a multinomial response distribution for $\mathbf{Y}_{it}$, factorized in terms of binomial distributions (continuation-ratio logits),

$$Mult(\mathbf{Y}_{it} \mid m_{it}, \omega_{i1t}, \cdots, \omega_{iCt}) = \prod_{j=1}^{C-1} Bin(Y_{ijt} \mid m_{ijt}, \varphi(Z_{ijt} + \epsilon_{ijt}))$$

  where $m_{it} = \sum_{j=1}^{C} Y_{ijt} \equiv 1$, $m_{i1t} = m_{it}$, and $m_{ijt} = m_{it} - \sum_{k=1}^{j-1} Y_{ikt}$;
- We adopt the proposed hierarchical GP-IWP modeling framework on $\{Z_{ijt}\}$ separately.

Sequential treatment of ordinal response and its practical implication

- The continuation-ratio logits structure offers a sequential mechanism to allocate the ordinal response Y_{it};

Generate $\mathcal{H}_{i1t} \sim Bern(\Delta_{i1t}), \Delta_{it1} = \varphi(Z_{i1t} + \epsilon_{i1t})$

$\mathcal{H}_{i1t} = 1$ $\quad$ $\mathcal{H}_{i1t} = 0$

$Y_{it} = 1$ $\quad$ Generate $\mathcal{H}_{i2t} \sim Bern(\Delta_{i2t}), \Delta_{i2t} = \varphi(Z_{i2t} + \epsilon_{i2t})$

$\mathcal{H}_{i2t} = 1$ $\quad$ $\mathcal{H}_{i2t} = 0$

$Y_{it} = 2$ $\quad$ Generate $\mathcal{H}_{i3t} \sim Bern(\Delta_{i3t}), \Delta_{i3t} = \varphi(Z_{i3t} + \epsilon_{i3t})$

$\mathcal{H}_{i3t} = 1$ $\quad$ $\mathcal{H}_{i3t} = 0$

$Y_{it} = 3$ $\quad$ $\cdots$

- We can re-organize the original data set containing longitudinal ordinal responses to create $C - 1$ data sets with longitudinal binary outcomes. Then, fit the proposed model for binary responses parallelly on the $C - 1$ re-organized data sets.

Real data application (Studentlife study)

- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.

Real data application (Studentlife study)

- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;

Real data application (Studentlife study)

- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;
- The data were collected by 48 students over a 10-week term at Dartmouth College;

Real data application (Studentlife study)

- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;
- The data were collected by 48 students over a 10-week term at Dartmouth College;
- We focus on the subset of the data that corresponds to assess the students' emotional status using the Photographic Affect Meter (PAM);

Real data application (Studentlife study)

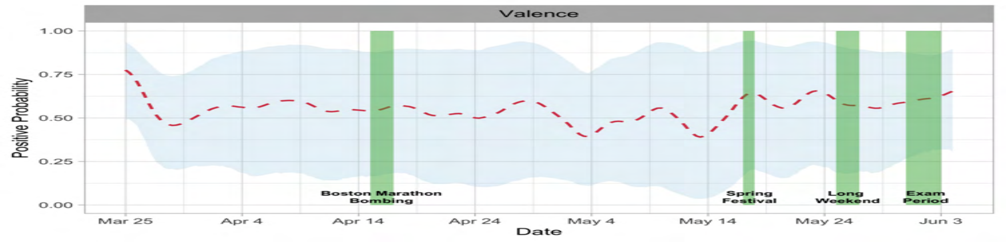
- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;
- The data were collected by 48 students over a 10-week term at Dartmouth College;
- We focus on the subset of the data that corresponds to assess the students' emotional status using the Photographic Affect Meter (PAM);
- The outcome of PAM have two attributes, valence and arousal. Each of them are integer scores from -2 to 2 (excluding 0). We start from dichotomize them by their sign, representing the positive values by 1;

Real data application (Studentlife study)

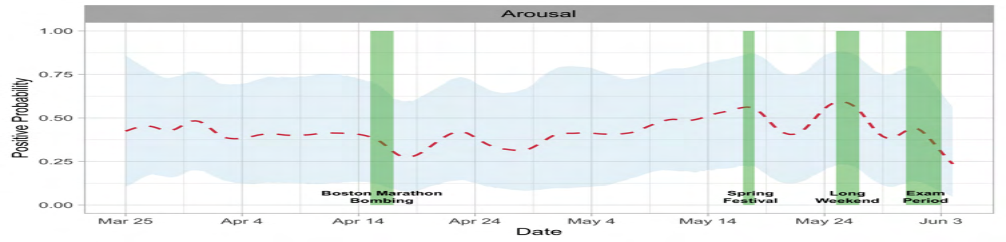
- Ecological momentary assessment (EMA) studies, which involve repeated measurements of subjects on uncommon grids; Missing values are typically considered to be a nuisance.
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;
- The data were collected by 48 students over a 10-week term at Dartmouth College;
- We focus on the subset of the data that corresponds to assess the students' emotional status using the Photographic Affect Meter (PAM);
- The outcome of PAM have two attributes, valence and arousal. Each of them are integer scores from -2 to 2 (excluding 0). We start from dichotomize them by their sign, representing the positive values by 1;
- Objective: analyzing the change of valence and arousal responses to evaluate students' affects as the term progresses.

Binary response case: result

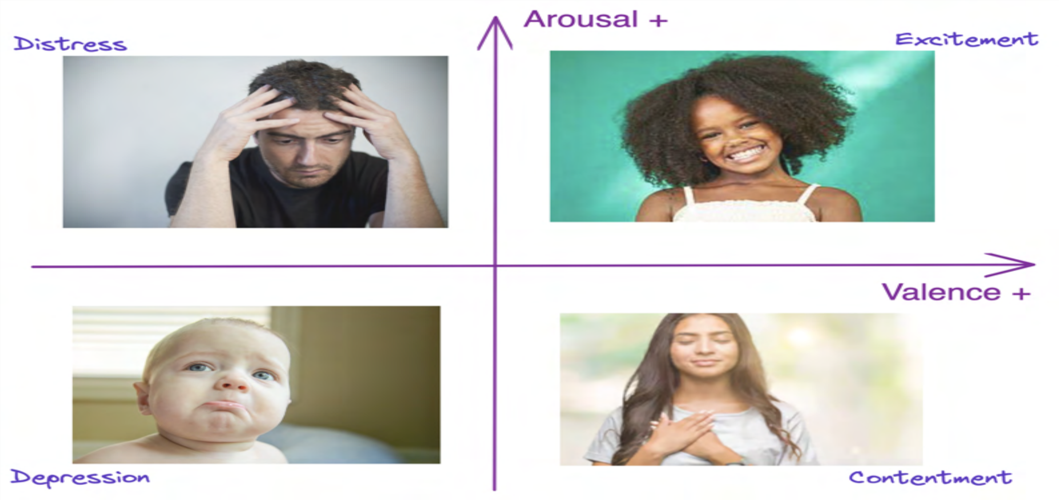
- Valence:



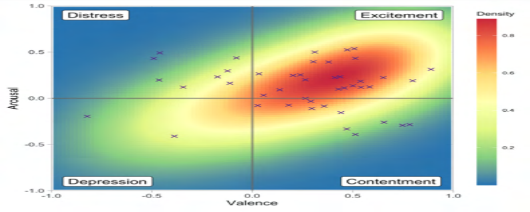
- Arousal:



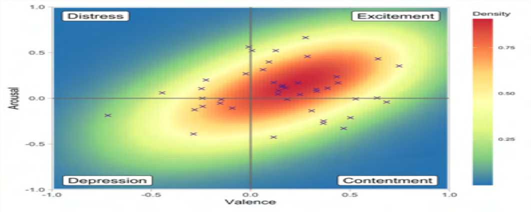
The mood coordinate space



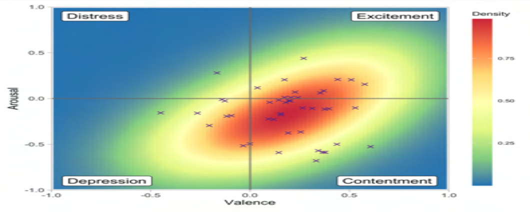
Categorizing emotional status



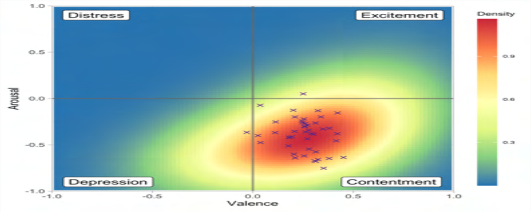
Green Key



Memorial Day

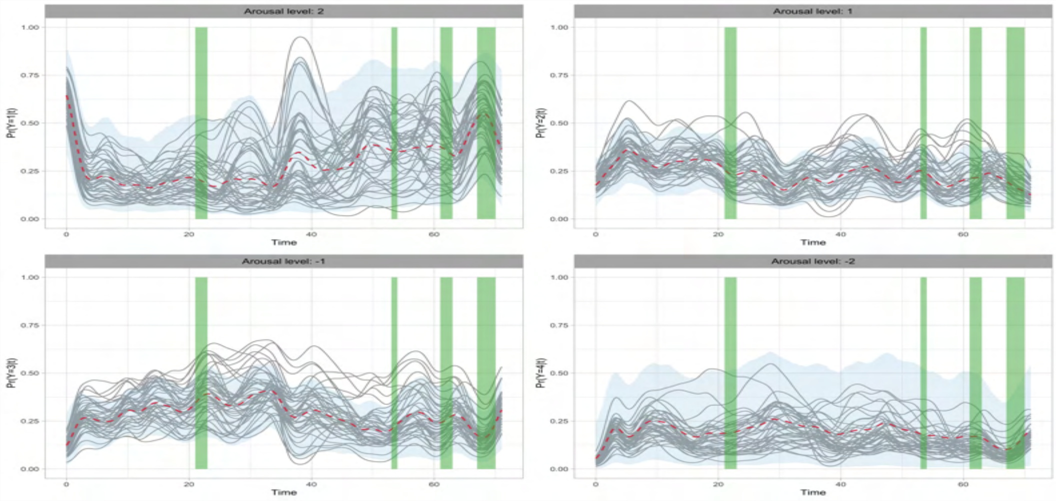


Final Exams Begin



Final Exams End

Four levels ordinal arousal score data



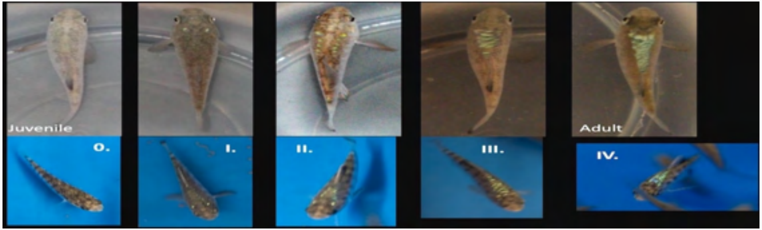
Summary of contributions

- We model the mean and covariance jointly and nonparametrically, avoiding potential biases caused by a pre-specified model structure;
- The model unifies the toolbox for balanced and unbalanced longitudinal studies;
- The model encourages borrowing of strength, preserving systematic patterns that are common across all subject responses;
- We develop a computationally efficient posterior simulation method by taking advantage of conditional conjugacy;
- The model can be extended to deal with ordinal responses with a moderate to large number of categories.

Models for Estimating Maturity of Sheepshead Minnows

Data structure

- Data from a longitudinal study consisting of the maturity status of sheepshead minnows under pre-determined experiment conditions;
- Three categorical experiment conditions (parent temperature (26 or 32), offspring temperature (26 or 32), and exposure time (7, 30 or 45)) split data into 12 groups;
- For each fish, we have observations at eight equally spaced time points;
- Ordinal response is the color stage, indicating maturity status; We use a binary version (immature vs mature).

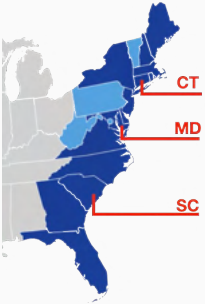


Objectives

- Estimate differences in trends in maturity across the treatment combinations;

Objectives

- Estimate differences in trends in maturity across the treatment combinations;
- Investiagte the relationship between transgenerational plasticity (TGP) and environment predictability;



- TGP occurs when phenotypes are shaped by parent and offspring environments;
- The fish are collected from three locations, Connecticut (CT), Maryland (MD) and South Carolina (SC);
- US east coast exhibits a latitudinal gradient in thermal predictability; Location with higher latitude corresponds to smaller thermal predictability;
- By theory, TGP has a positive relationship with thermal predictability; We are expected to show TGP decreases with increasing latitude.

Main methodology

- Let $\mathbf{Y}_{gi}$ denote the observed binary maturity status sequence at grid $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_T)^\top$ of the i-th subject in g-th group;
- At the observed data level, we assume

$$Y_{git} \mid Z_{git}, \boldsymbol{\epsilon}_{git} \overset{ind.}{\sim} Bin(1, \varphi(Z_{git} + \epsilon_{git})),\ t = 1, \ldots, T,\ i = 1, \ldots, n_g,\ g = 1, \ldots, G;$$

where the error term $\epsilon_{git} \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$;

- We assume Z_{git} is the evaluation of a continuous signal process $Z_{gi}(\tau)$ at time t;
- Model continuous signal process $Z_{gi}(\tau)$ through GP:

$$Z_{gi}(\tau) | \mu_g(\tau), \Sigma_g(\tau, \tau) \overset{i.i.d.}{\sim} GP(\mu_g(\tau), \Sigma_g(\tau, \tau)), \quad i = 1, \ldots, n_g,\ g = 1, \ldots, G;$$

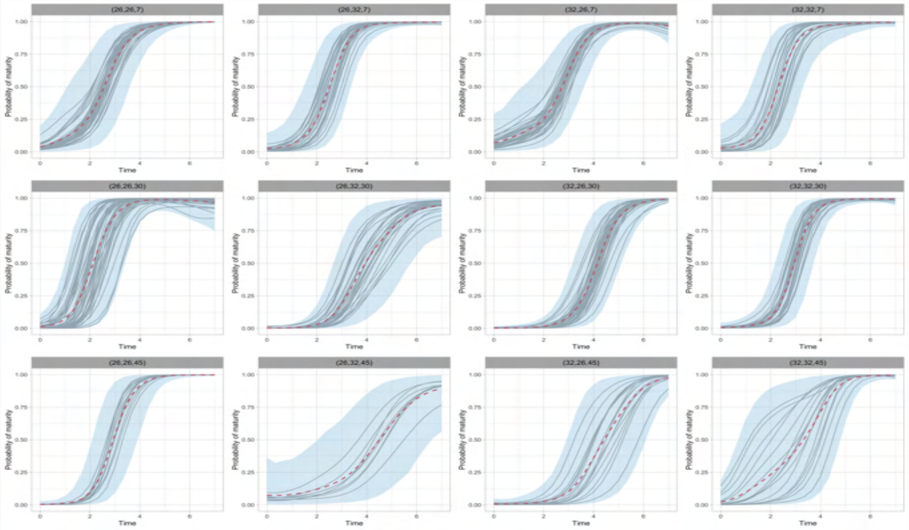
- Joint hierarchical nonparametric prior for the mean and covariance function of the GP:

$$\begin{aligned} \mu_g(\tau) | \Sigma_g(\tau, \tau), \mu_{0g}(\tau), \nu_g &\overset{ind.}{\sim} GP(\mu_{0g}(\tau), (\nu_g - 3)\Sigma_g(\tau, \tau)), \\ \Sigma_g(\tau, \tau) | \nu_g, \Psi_{\sigma_g^2, \rho_g}(\tau, \tau) &\overset{ind.}{\sim} IWP(\nu_g, \Psi_{\sigma_g^2, \rho_g}(\tau, \tau)). \end{aligned}$$

Prior specification

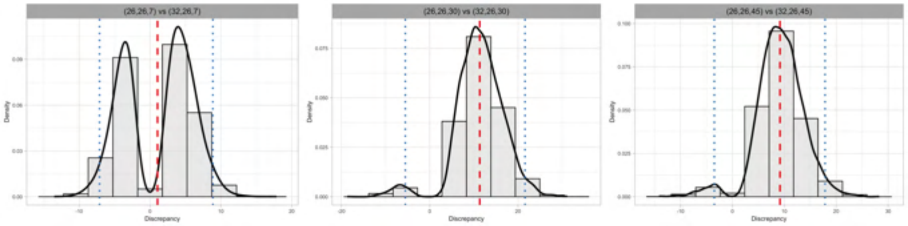
- We further assume $\mu_{0g}(\tau) \equiv \mu_{0g}$, and specify the covariance kernel of IWP as Matérn covariance kernel with smoothness $5/2$;
- Denote the aforementioned joint prior for the mean and covariance function as $JP(\mu_{0g}, \sigma_g^2, \rho_g, \nu_{0g})$;
- We seek to introduce an appropriate level of dependence across groups through prior placed on $\{\mu_{0g}, \sigma_g^2, \rho_g, \nu_{0g} : g = 1, \cdots, 12\}$;
- The best option, selected by multiple model comparison criteria, is
  - we assume $\mu_{0g} = \mathbf{x}_g^\top \boldsymbol{\alpha}$, where $\mathbf{x}_g$ is a vector of indicators for each group. We further place a shrinkage prior on $\boldsymbol{\alpha}$;
  - We assume conditionally independent scale parameters σ_g^2, i.e., $\sigma_g^2 \mid \theta \sim Gamma(a_\sigma, a_\sigma \theta^{-1})$, and $\theta \sim IG(a_\theta, b_\theta)$;
  - We assume a common smoothness parameter shared by groups, i.e., $\rho_g \equiv \rho \sim Unif(a_\rho, b_\rho)$;
  - we assume a common degrees of freedom parameter shared by groups, i.e., $\nu_g \equiv \nu \sim Unif(a_\nu, b_\nu)$.

Posterior estimate of maturity probability

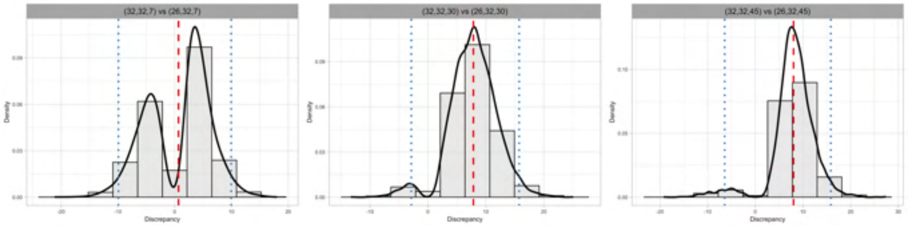


Comparison of treatment effect on maturity

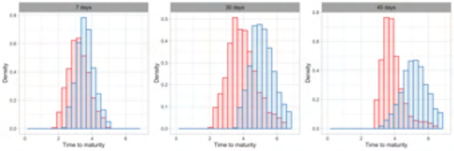
- Relative effect between groups with offspring temperature (OT) 26:



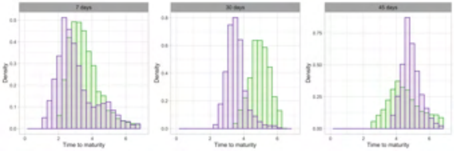
- Relative effect between groups with offspring temperature (OT) 32:



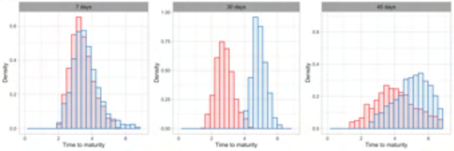
Thermal TGP and latitude



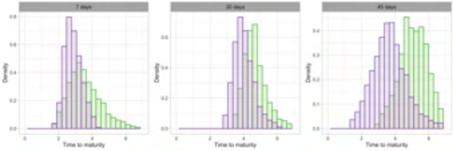
CT, with offspring temperature 26



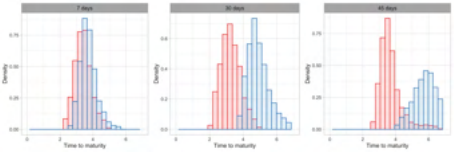
CT, with offspring temperature 32



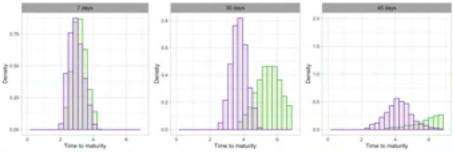
MD, with offspring temperature 26



MD, with offspring temperature 32



SC, with offspring temperature 26



SC, with offspring temperature 32

Concluding Remarks

Conclusions

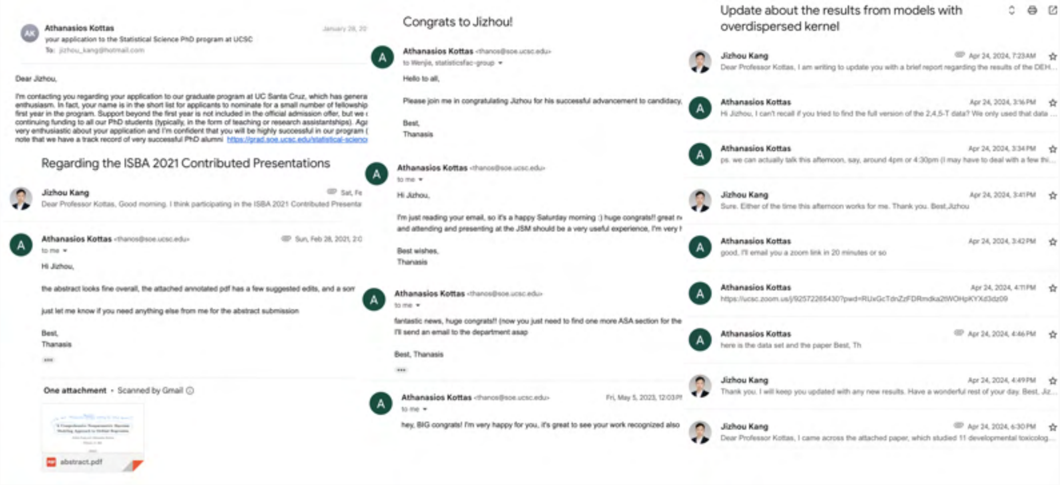
- We have developed a suite of statistical models for ordinal regression;
- Key of the model: the continuation-ratio factorization;
- Possible future works:
  - The structural similarity between nonparametric priors for discrete distributions and models for categorical data boost new models for categorical data analysis;
  - Extensions of the proposed models, enhancing flexibility and keeping efficiency;
  - Scale up inference in the big data era: variational inference algorithms.

Scholarly articles from dissertation research

- Kang, J. and Kottas, A. (2022+), "Structured Mixture of Continuation-ratio Logits Models for Ordinal Regression", *arXiv:2211.04034*, (revised, under review);
- Kang, J. and Kottas, A. (2023+), "Flexible Bayesian Modeling for Longitudinal Binary and Ordinal Responses", *arXiv:2307.00224*, (submitted, under review);
- Kang, J. and Kottas, A. (2024+), "Bayesian Nonparametric Risk Assessment in Developmental Toxicity Studies with Ordinal Responses", (in preparation);
- Kang, J., Kottas, A., Lee, W. and Munch, S. (2024+), "Bayesian Modeling of Repeated Ordinal Responses Collected Under Different Treatments: An application to estimating maturity of Sheepshead Minnows", (in preparation).

Acknowledgments

Acknowledgments


Athanasios Kottas
your application to the Statistical Science PhD program at UCSC
To: jizhou_kang@hotmail.com

Dear Jizhou,

I'm contacting you regarding your application to our graduate program at UC Santa Cruz, which has genera enthusiasm. In fact, your name is in the short list for applicants to nominate for a small number of fellowship first year in the program. Support beyond the first year is not included in the official admission offer, but we continuing funding to all our PhD students (typically, in the form of teaching or research assistantships). Ag very enthusiastic about your application and I'm confident that you will be highly successful in our program note that we have a track record of very successful PhD alumni https://grad.soe.ucsc.edu/statistical-scienc

Regarding the ISBA 2021 Contributed Presentations

Jizhou Kang
Dear Professor Kottas, Good morning. I think participating in the ISBA 2021 Contributed Presenta

Athanasios Kottas <thanos@soe.ucsc.edu> Sun, Feb 28, 2021, 2:0
to me
Hi Jizhou,
the abstract looks fine overall, the attached annotated pdf has a few suggested edits, and a sor
just let me know if you need anything else from me for the abstract submission
Best,
Thanasis

One attachment • Scanned by Gmail
abstract.pdf

Congrats to Jizhou!

Athanasios Kottas <thanos@soe.ucsc.edu>
to Wenjie, statisticsfac-group
Hello to all,
Please join me in congratulating Jizhou for his successful advancement to candidacy.
Best,
Thanasis

Athanasios Kottas <thanos@soe.ucsc.edu>
to me
Hi Jizhou,
I'm just reading your email, so it's a happy Saturday morning :) huge congrats!! great n and attending and presenting at the JSM should be a very useful experience, I'm very
Best wishes,
Thanasis

Athanasios Kottas <thanos@soe.ucsc.edu>
to me
fantastic news, huge congrats!! (now you just need to find one more ASA section for the I'll send an email to the department asap
Best, Thanasis

Athanasios Kottas <thanos@soe.ucsc.edu> Fri, May 5, 2023, 12:03 PM
to me
hey, BIG congrats! I'm very happy for you, it's great to see your work recognized also

Update about the results from models with overdispersed kernel

Jizhou Kang — Apr 24, 2024, 7:23 AM
Dear Professor Kottas, I am writing to update you with a brief report regarding the results of the DEH...

Athanasios Kottas — Apr 24, 2024, 3:16 PM
Hi Jizhou, I can't recall if you tried to find the full version of the 2,4,5-T data? We only used that data ...

Athanasios Kottas — Apr 24, 2024, 3:34 PM
ps. we can actually talk this afternoon, say, around 4pm or 4:30pm (I may have to deal with a few thi...

Jizhou Kang — Apr 24, 2024, 3:41 PM
Sure. Either of the time this afternoon works for me. Thank you. Best,Jizhou

Athanasios Kottas — Apr 24, 2024, 3:42 PM
good, I'll email you a zoom link in 20 minutes or so

Athanasios Kottas — Apr 24, 2024, 4:11 PM
https://ucsc.zoom.us/j/92572265430?pwd=RUxGcTdnZzFDRmdka2tWOHpKYXd3dz09

Athanasios Kottas — Apr 24, 2024, 4:46 PM
here is the data set and the paper Best, Th

Jizhou Kang — Apr 24, 2024, 4:49 PM
Thank you. I will keep you updated with any new results. Have a wonderful rest of your day. Best, Jiz...

Jizhou Kang — Apr 24, 2024, 6:30 PM
Dear Professor Kottas, I came across the attached paper, which studied 11 developmental toxicolog...

MANY THANKS!

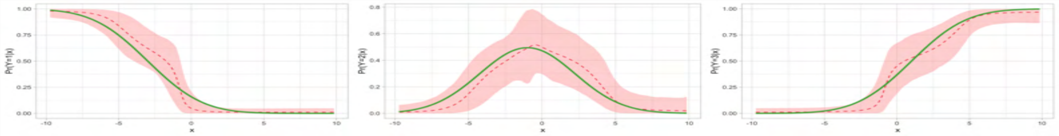
I am happy to answer any questions.

© Jizhou Kang (jkang37@ucsc.edu)

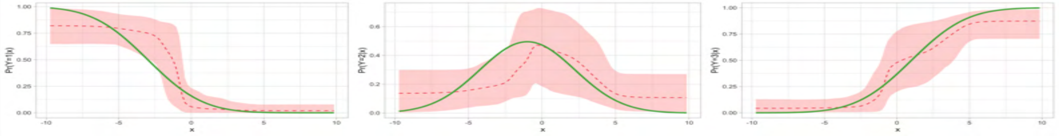
Synthetic data examples

- In both experiments, n pairs of ordinal response and covariate $(\mathbf{Y}_i, x_i)$ are generated, where $x_i \overset{i.i.d.}{\sim} Unif(x_i| -10, 10)$ such that with the intercept, the covariate vector is $\mathbf{x}_i = (1, x_i)^T$;
- **First experiment**: We generate $n = 100$ responses by first sampling a latent continuous variable $\tilde{y}_i$ from normal distribution, then discretizing $\tilde{y}_i$ with cut-off points to get the ordinal response $\mathbf{Y}_i$;
- **Second experiment**:
  - We generate data from $\mathbf{Y} \sim \sum_{k=1}^{3} \omega_k(\mathbf{x}) K(\mathbf{Y}|\mathbf{m}, \boldsymbol{\theta}_k(\mathbf{x}))$;
  - The true probability response curves have nonstandard shape.
  - Perform the experiment with $n = 800$ simulated data.

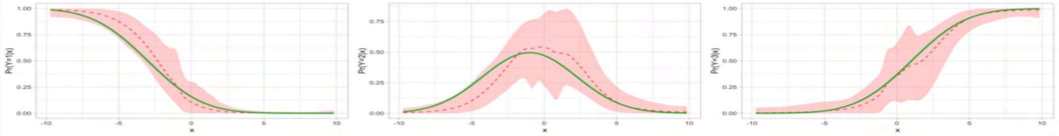
First experiment result (baseline prior)



Common-weights model

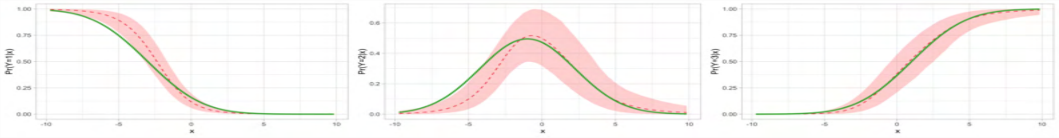


Common-atoms model

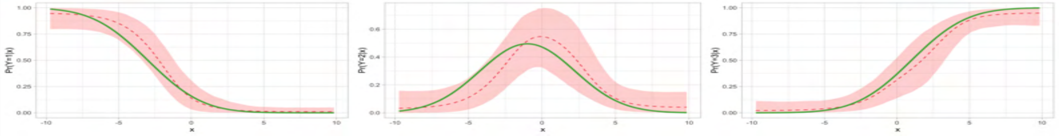


General model

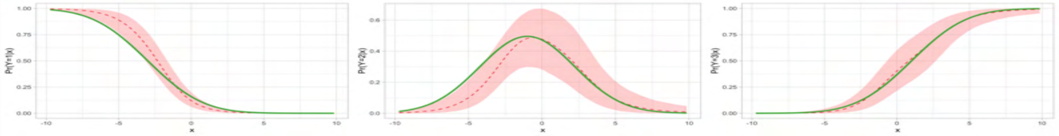
First experiment result (specified prior)



Common-weights model

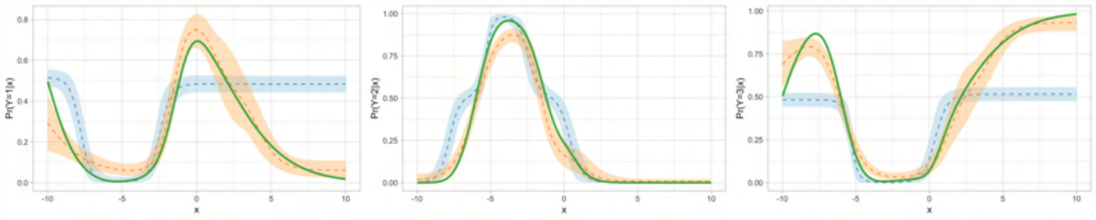


Common-atoms model

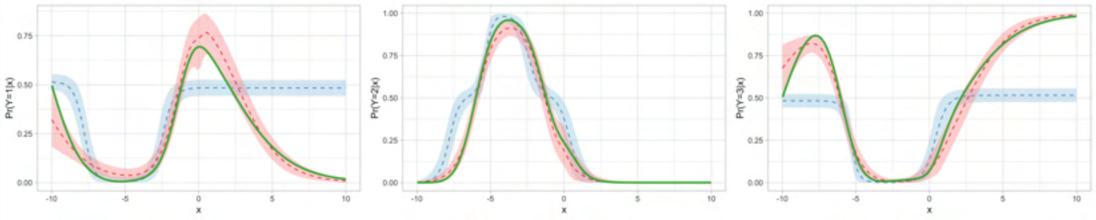


General model

Second experiment result

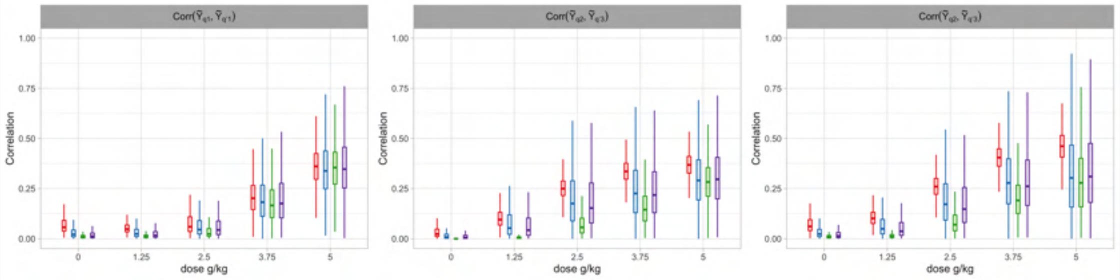


Common-weights and common-atoms models



Common-weights and general models

Posterior distributions of the intracluster correlations



- The correlations depict an overall increasing trend with toxin levels;
- The intracluster correlation at the new dose level indicates a smooth borrowing of strength across dose levels;
- The correlation distribution from models with overdispersed kernel spread a wider range.

General settings

- In both experiments, we simulate longitudinal binary responses from:

$$Y_i(\boldsymbol{\tau}_i) \mid \mathcal{Z}_i(\boldsymbol{\tau}_i) \overset{ind.}{\sim} Bin(1, \eta(\mathcal{Z}_i(\boldsymbol{\tau}_i))), \quad \boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i}), \quad i = 1, \cdots, n,$$

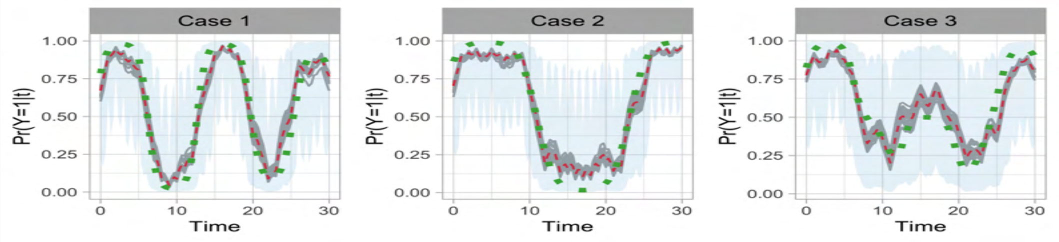
$$\mathcal{Z}_i(\boldsymbol{\tau}_i) = f(\boldsymbol{\tau}_i) + \boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i \quad \boldsymbol{\epsilon}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}),$$

  a generic data generating process with:

  - $\eta(\cdot)$: a generic link function mapping $\mathbb{R}$ to $(0, 1)$;
  - $f(\tau)$: a generic signal function of time;
  - $\boldsymbol{\omega}_i$: a realization from a mean 0 continuous process that depicts the temporal covariance within the i-th subject.

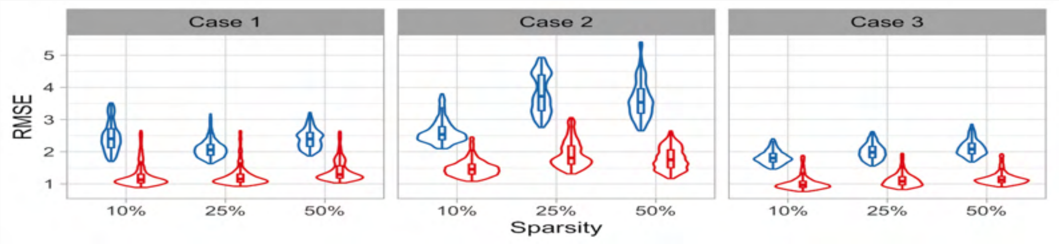
First set of experiments: result

- Focusing on the performance in recovering the fluctuation of the temporal trend;
- We simulate data with different link function, signal function, and temporal covariance structure combinations;
- To enforce an unbalanced study design, we randomly drop out a proportion of the simulated data. We consider different choices of drop out proportions.



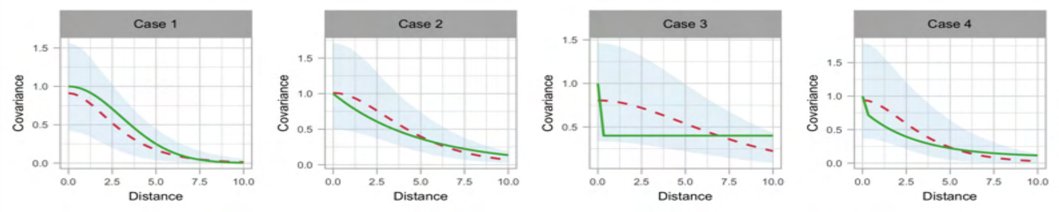
First set of experiments: comparison

- We compare the proposed model with its simplified backbone;
- Instead of modeling the mean function μ through a GP, we consider modeling it parametricly by $\mu(\tau) \equiv \mu_0$, and $\mu_0 \sim N(a_\mu, b_\mu)$;
- For criterion, we use the rooted mean square error (RMSE) between the model estimated signal process and the truth.



Second set of experiments: result

- Focusing on the performance in the within subject covariance structure;
- We simulate data with a number of possible choices for ω_i;
- None of these choices imply covariance structures that are in the same form as the covariance kernel used in the proposed model.



Second set of experiments: comparison

- We consider an alternative, simplified modeling approach, instead of modeling the covariance function nonparametricly, we assume a covariance kernel of certain parametric form;
- Specifically, $Z_i \overset{i.i.d.}{\sim} GP(\mu, \Psi_\phi)$, $\mu \sim GP(\mu_0, \Psi_\phi/\kappa)$, with parametric Ψ_ϕ;
- We compute the 2-Wasserstein distance between the model estimated distribution of ω_i and the truth.