Flexible Bayesian Modeling for Longitudinal Binary and Ordinal Responses

Jizhou Kang

Dec 7th, 2023

Department of Statistics, UC Santa Cruz

Joint work with Prof. Athanasios Kottas

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;
- Flexible modeling made easy:
  - Flexible: allow general forms for the probability response curves and the temporal correlation structure;
  - Easy: clear prior specification strategy, efficient posterior inference algorithm, easy to work with expressions for quantities of interest.

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;
- A specific example: *Studentlife study*. The study integrates automatic sensing data and EMA components to probe students' mental health status;

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;
- 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;

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;
- 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);

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;
- 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;

Motivating application

- Ecological momentary assessment (EMA) studies, which involve the repeated measuing of people's current thoughts, emotions, behavior, and physiological states, in their natural environment;
- 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.

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).

Proposed Methodology

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

- 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 $\mathcal{H}_{i3t} \sim Bern(\Delta_{i3t}), \Delta_{i3t} = \varphi(Z_{i3t} + \epsilon_{i3t})$

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

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

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

$\mathcal{H}_{i1t} = 1$: $Y_{it} = 1$

$\mathcal{H}_{i1t} = 0$: Generate $\mathcal{H}_{i2t} \sim Bern(\Delta_{i2t}), \Delta_{i2t} = \varphi(Z_{i2t} + \epsilon_{i2t})$

$\mathcal{H}_{i2t} = 1$: $Y_{it} = 2$

$\mathcal{H}_{i2t} = 0$: Generate $\mathcal{H}_{i3t} \sim Bern(\Delta_{i3t}), \Delta_{i3t} = \varphi(Z_{i3t} + \epsilon_{i3t})$

$\mathcal{H}_{i3t} = 1$: $Y_{it} = 3$

$\mathcal{H}_{i3t} = 0$: $\cdots$

- 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.

Synthetic Data Examples

General settings

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

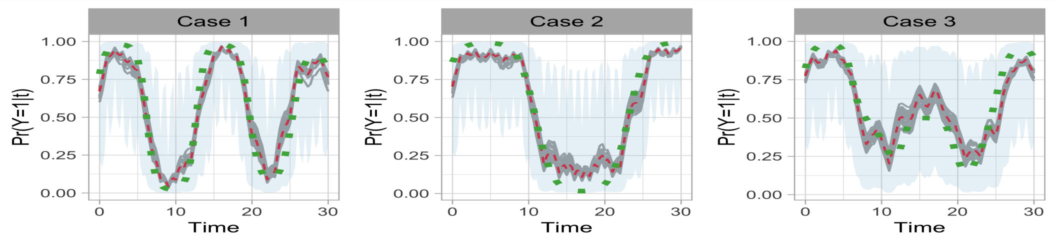
$$Y_i(\boldsymbol{\tau}_i) \mid \mathcal{Z}_i(\boldsymbol{\tau}_i) \overset{ind.}{\sim} Bin(1, \eta(\mathcal{Z}_i(\boldsymbol{\tau}_i))), \quad \boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{iT_i}), \quad i = 1, \cdots, n,$$
$$\mathcal{Z}_i(\boldsymbol{\tau}_i) = f(\boldsymbol{\tau}_i) + \boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i \quad \boldsymbol{\epsilon}_i \overset{i.i.d.}{\sim} N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}),$$

  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.

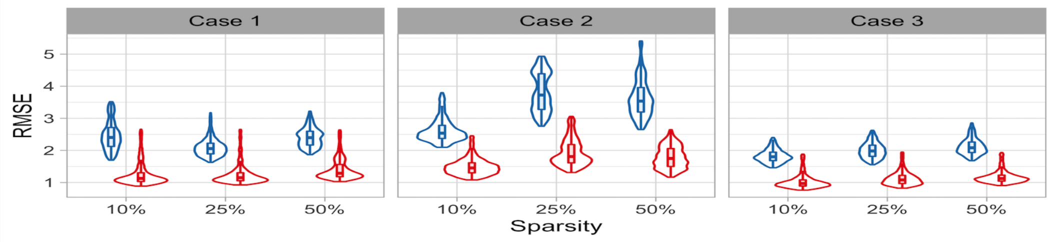
First set of experiments: result

- 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.



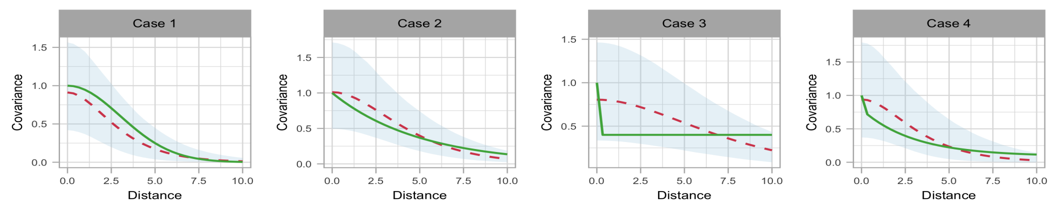
First set of experiments: comparison

- 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.



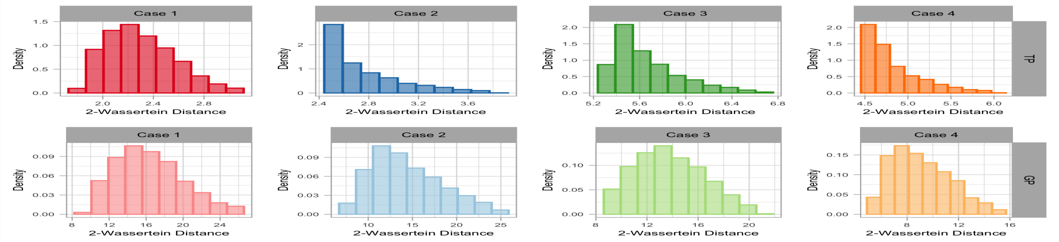
Second set of experiments: result

- 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

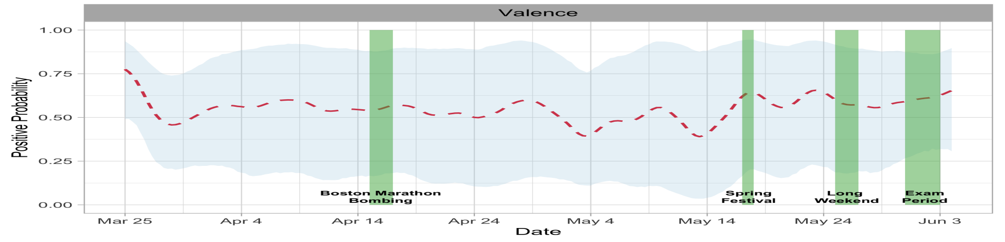
- We consider an alternative, simplified modeling approach, instead of modeling the covariance function nonparametricly, we assume a covariance kernel of certain parametric form;
- Specifically, $Z_i \overset{i.i.d.}{\sim} GP(\mu, \Psi_\phi)$, $\mu \sim GP(\mu_0, \Psi_\phi/\kappa)$, with parametric Ψ_ϕ;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



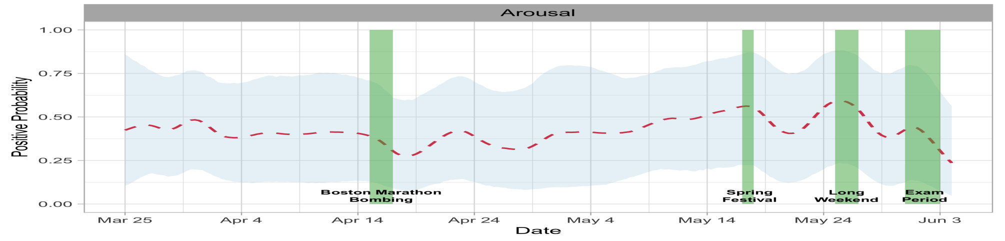
Real Data Application (Studentlife study)

Binary response case: result

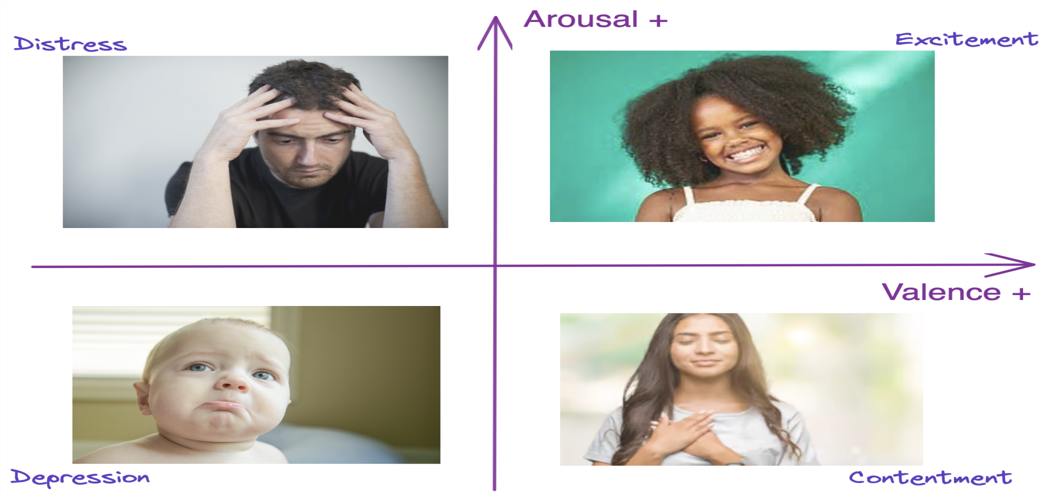
- Valence:



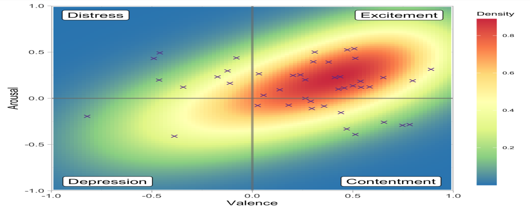
- Arousal:



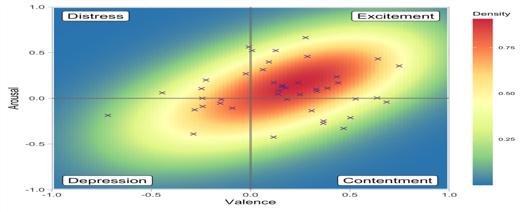
The mood coordinate space



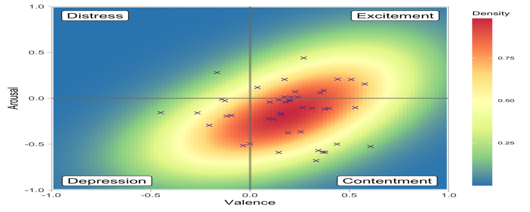
Categorizing emotional status



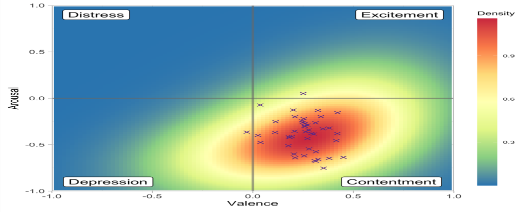
Green Key



Memorial Day

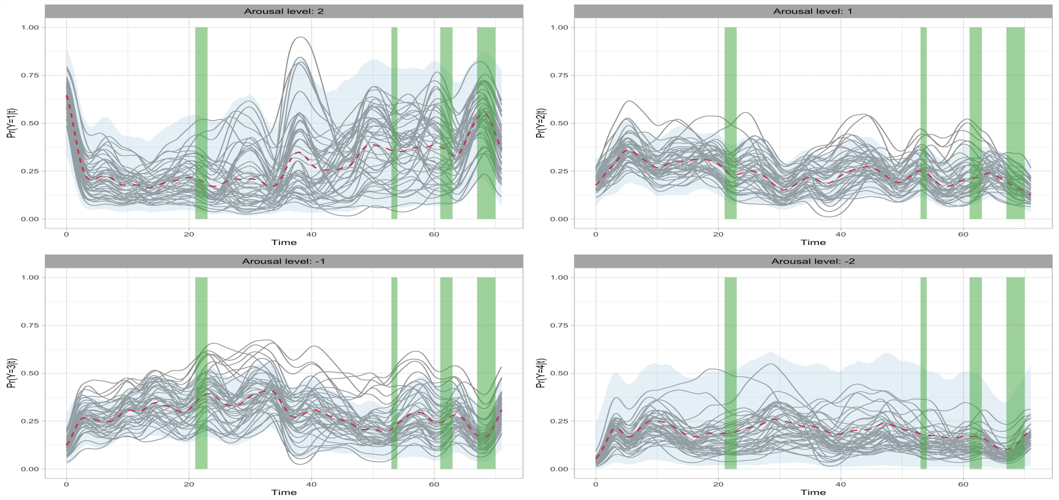


Final Exams Begin



Final Exams End

Four levels ordinal arousal score data



Concluding Remarks

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.

MANY THANKS!

I am happy to answer any questions.

Jizhou Kang and Athanasios Kottas (2023+), "Flexible Bayesian Modeling for Longitudinal Binary and Ordinal Responses", arXiv: 2307.00224.