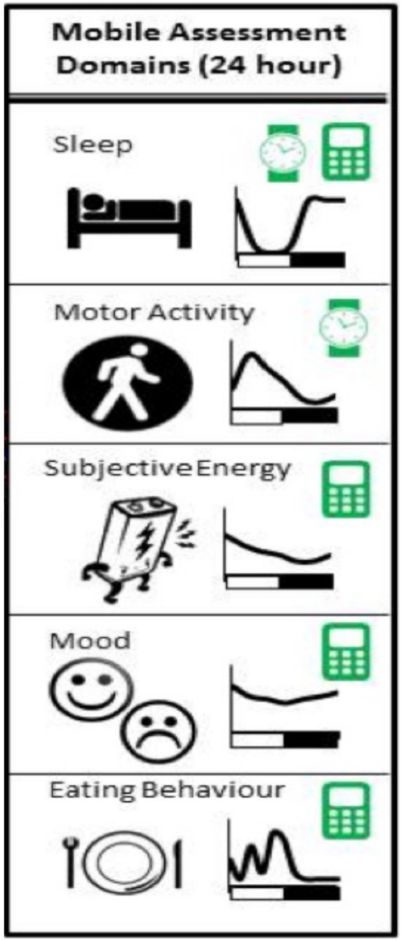
Modelling of mixed type intensive longitudinal data via Semiparametric Gaussian Copula and its application to real-time mobile monitoring of daily health behaviours

Debangan Dey
4th Year PhD Student
Department of Biostatistics

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Electronic Diaries (EMA)

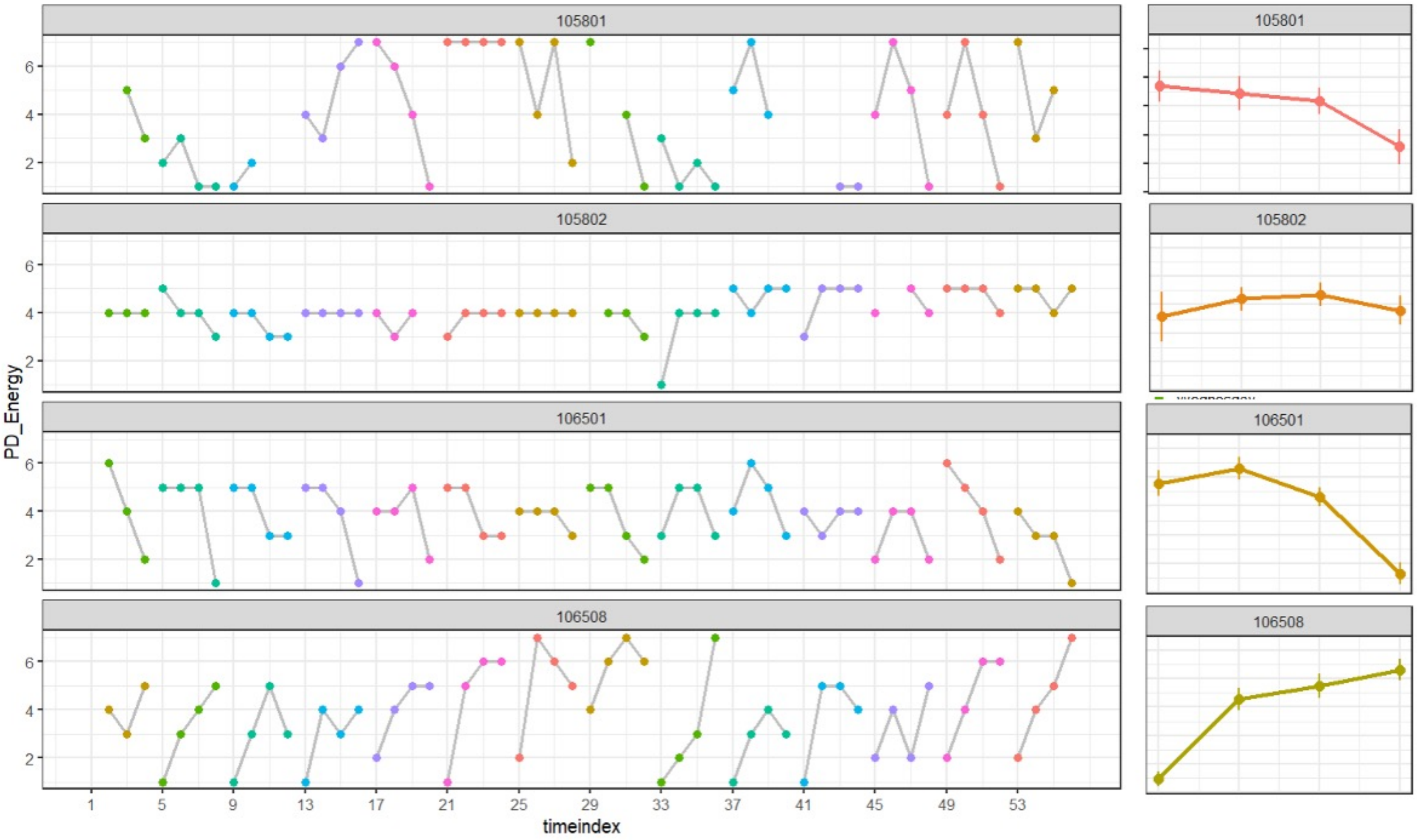


- Real-time self-reports of mood, energy, stress, pain-level, anxiety, headache recorded through smartphones.
- Objectively recorded physical activity and sleep through smartwatches.
- Intensive longitudinal data.
- Different measurement scales (binary, ordinal, truncated, continuous, categorical).
- Differences in subjective interpretation of scales.

NIMH Family Study

- A nested case-control design of 499 adults with cases being subjects with different mood disorders.
- An actigraphy device worn on the nondominant wrist plus EMA 4 times per day for 2 weeks.

Trajectories of Energy (scale 1-7)



Case Study: Migraine

Headache incidence

Sleep, mood, medical history

History of mood/anxiety disorder
History of migraine

Self-reported sleep quality (1-7)
Actigraphy measured sleep duration
Actigraphy measured sleep midpoint

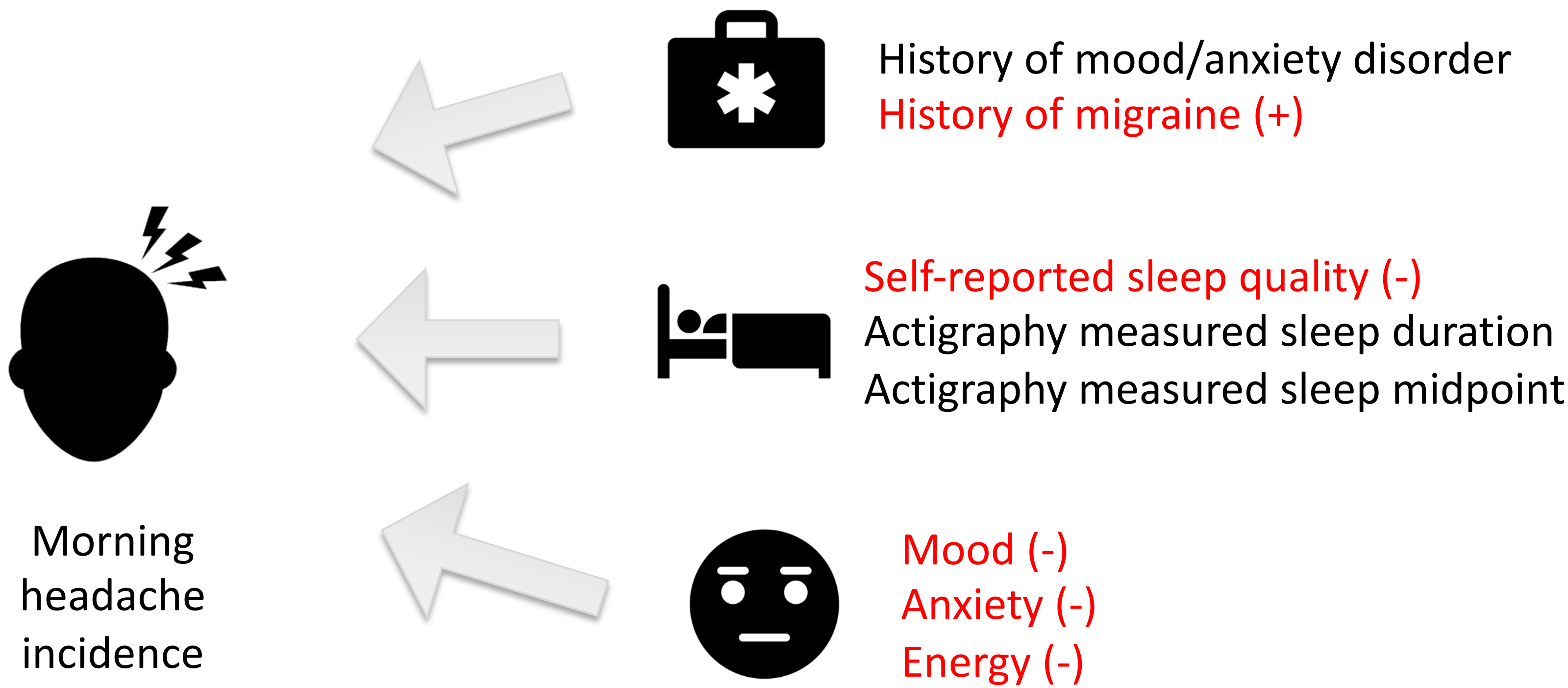
Mood
Anxiety
Energy

Morning headache incidence

Time-dependent covariates

Significant associations

Linear mixed effects model

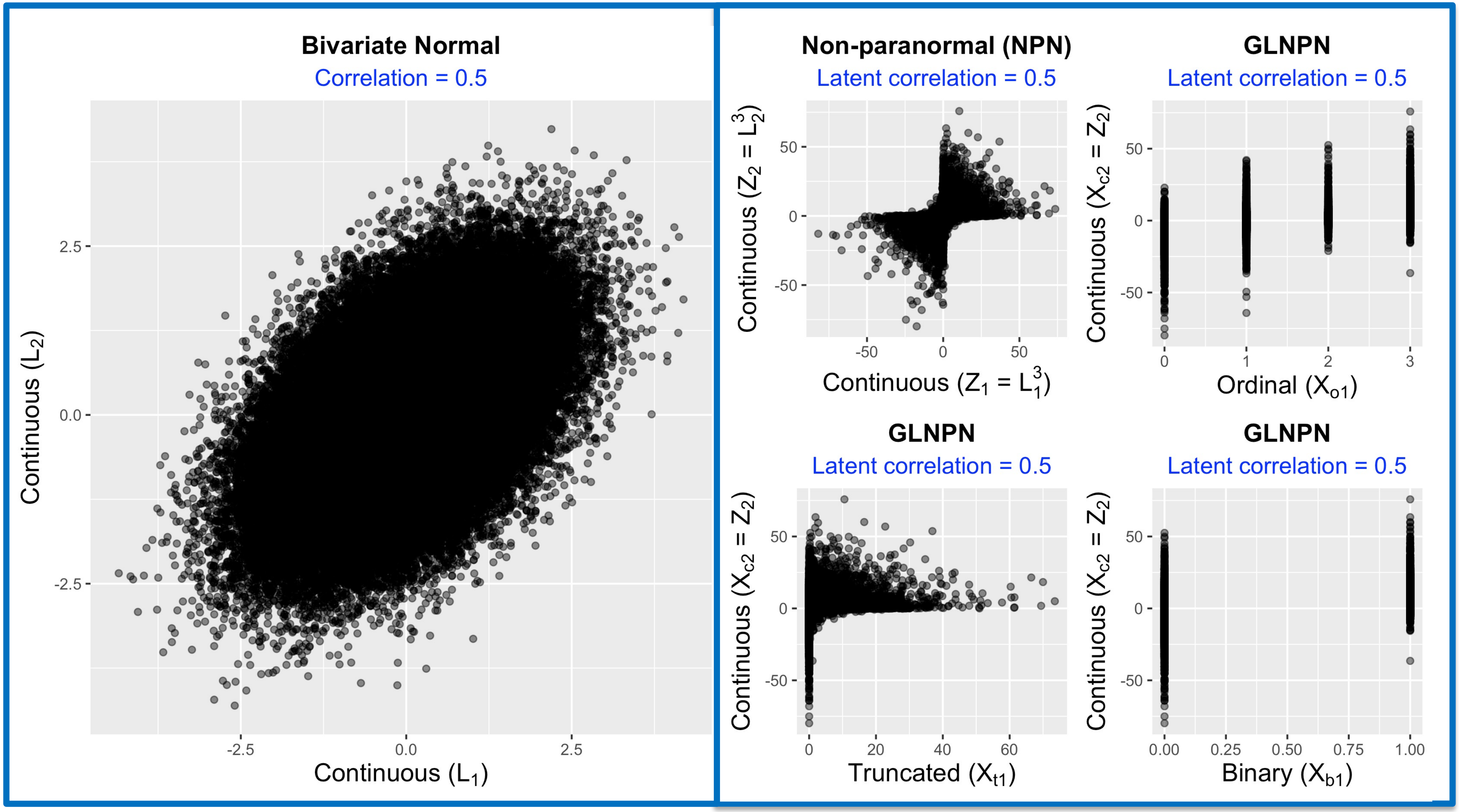


Challenges

- Self-reported mood/sleep variables need to be treated as ordinal variables rather than continuous.
- How can we tackle different subject-specific scales for EMA variables?
- Can we build a joint modeling framework for all our binary, ordinal and continuous variables?

Semi-parametric Gaussian Copula

Illustration



Latent → Observed

Semi-parametric Gaussian Copula/ Non-paranormal Distribution (NPN)

Observed variables are monotone transformations (f) of jointly standardized correlated normal latent variables ($N_p(0, \Sigma)$).

Generalized Latent NPN (GLNPN)

Observed variables are truncated, categorized or binarized version of monotone transformations (f) of jointly standardized correlated normal latent variables ($N_p(0, \Sigma)$).

Joint model

- Time-points: $t_1, t_2, \ldots, t_m$
- Time-varying outcome (e.g. Mood): $Y_i(t_1), \ldots, Y_i(t_m)$
- Time-varying covariate (e.g. Physical activity): $X_i(t_1), \ldots, X_i(t_m)$
- $(Y_i(t_1), \ldots, Y_i(t_m), X_i(t_1), \ldots, X_i(t_m)) \sim GLNPN\ (0, \Sigma, f)$

Regression and PCA

- Σ is assumed to be cross-correlation matrix from known functional covariance Kernels of Gaussian processes.
- Function-on-function Regression coefficients can be estimated from conditional distribution derived from the latent smooth correlation matrix.
- We can also perform functional PCA on the latent space for dimension reduction.

Advantages

- Takes care of mixed type of variables representing subject-specific heterogeneous scales,
- Identifies within-day patterns of mode-specific and domain-specific behavioral measures;
- Evaluates cross-domain inter-relationships to characterize mode-specific and multi-modal dynamic behavioral phenotypes,
- Develops individualized prediction models for dynamic prediction of adverse short-term health and behavioral events.

THANK YOU!

Questions?