Covariance Partition Priors: A Bayesian Approach to Simultaneous Covariance Estimation for Longitudinal Data

Jeremy Gaskins
Department of Bioinformatics and Biostatistics
University of Louisville

Joint work with Mike Daniels (University of Texas)

June 3, 2014
SRCOS Summer Research Conference

The Problem

Assume that we have M groups of mean-zero, longitudinal data. Groups may be defined by differing treatments and/or baseline covariates.

$$\mathbf{Y}_{mi} \sim \mathrm{N}_T\left(\mathbf{0}, \boldsymbol{\Sigma}_m\right), \quad i = 1, \ldots, n_m; \;\; m = 1, \ldots, M$$

How can we model/estimate the set of covariance matrices $\{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M\}$?

The Problem

Assume that we have M groups of mean-zero, longitudinal data. Groups may be defined by differing treatments and/or baseline covariates.

$$\mathbf{Y}_{mi} \sim \mathrm{N}_T\left(\mathbf{0}, \boldsymbol{\Sigma}_m\right), \quad i = 1, \ldots, n_m; \; m = 1, \ldots, M$$

How can we model/estimate the set of covariance matrices $\{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M\}$?

- Assume Equality → If untrue, this can lead to invalid inference of mean effects.
- Model Each Separately → Inefficient if there is common structure, especially if n_m are small or T is large.

We want a prior for $\{\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M\}$ that finds the middle ground between the two by sharing information across groups to improve estimation efficiency.

The Modified Cholesky Parametrization

We use the modified Cholesky parametrization for each $\mathbf{\Sigma}_m$ to avoid the positive-definite constraint.

$$\mathbf{\Sigma}_m^{-1} = \mathbf{T}_m \mathbf{G}_m \mathbf{T}_m^\top$$

- $\mathbf{T}_m$ is an upper-triangular matrix with 1's on the diagonal and $T(T-1)/2$ unconstrained elements $-\phi_{m;jt}$ $(j < t)$ called the *generalized autoregressive parameters* (GARPs).
- $\mathbf{G}_m = \text{diag}\{\gamma_{m1}^{-1}, \ldots, \gamma_{mT}^{-1}\}$, where $\gamma_{mt} > 0$ are the T *innovation variances* (IVs).

The GARPs and IVs are interpreted as regression coefficients and variances of the sequential regressions.

(Pourahmadi, 1999 and 2000)

The Modified Cholesky Parametrization, (2)

For $\mathbf{Y} \sim \mathrm{N}_T\left(\mathbf{0}, \mathbf{\Sigma} = (\mathbf{T}_m \mathbf{G}_m \mathbf{T}_m^\top)^{-1}\right)$, we can factor the distribution as

$$f(Y_1)\, f(Y_2|y_1)\, f(Y_3|y_1, y_2) \cdots f(Y_T|y_1, \ldots, y_{T-1}).$$

The t-th sequential distribution is

$$Y_t|y_1, \ldots, y_{t-1} \sim N\left(\sum_{j<t} \phi_{m;jt} y_j, \gamma_{mt}\right).$$

Normal and inverse gamma are conjugate distributions for the GARPs and IVs.

The Goal

Our goal is to develop a prior distribution for the sets of GARPs and IVs that borrows strength across the M groups and encourages a lower-dimensional structure on $\boldsymbol{\Sigma}_m$ that is natural for longitudinal data.

The Goal

Our goal is to develop a prior distribution for the sets of GARPs and IVs that borrows strength across the M groups and encourages a lower-dimensional structure on $\boldsymbol{\Sigma}_m$ that is natural for longitudinal data.

We do this by allowing the sequential distributions to be equal across groups, that is,

$$f_m(Y_t | y_1, \ldots, y_{t-1}) = f_{m'}(Y_t | y_1, \ldots, y_{t-1})$$

by partitioning the group labels for each t. Hence, $\phi_{m;t} = \phi_{m';t}$ ($\phi_{m;t} = (\phi_{m;1t}, \ldots, \phi_{m;t-1,t})^\top$) and $\gamma_{mt} = \gamma_{m't}$. We also favor a sparse structure for $\mathbf{T}$ by some shrinking the GARPs toward zero.

Covariance Partition Prior

Notation:

- $\mathcal{M} = \{1, \ldots, M\}$ is the collection of group labels.
- $\mathcal{P} = \{S_1, \ldots, S_d\}$ is a partition of $\mathcal{M}$ with degree d.
- The sets $S_1, \ldots, S_d$ are non-empty and mutually exclusive, with $\mathcal{M}$ as the union.
- Ω is the collection of all possible partitions $\mathcal{P}$.
- The M-th Bell Number B_M is the cardinality of Ω.

The groups in the same set of $\mathcal{P}_t = \{S_{1t}, \ldots, S_{d_t t}\}$ have the same dependence parameters for the t-th sequential distribution $f(Y_t | y_1, \ldots, y_{t-1})$.

We need to specify a joint prior distribution $\pi(\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_T)$ for the set of partitions that will vary smoothly in time.

Markov Chain of Partitions

To that end, we introduce a Markov chain on partitions.

- Markov Property: $\text{pr}(\mathcal{P}_{t+1}|\mathcal{P}_1, \ldots, \mathcal{P}_t) = \text{pr}(\mathcal{P}_{t+1}|\mathcal{P}_t)$
- Distance between partitions:

$$d(\mathcal{P}_1, \mathcal{P}_2) = 2|\mathcal{P}_1 \cap \mathcal{P}_2| - |\mathcal{P}_1| - |\mathcal{P}_2|$$

$$\mathcal{P}_1 \cap \mathcal{P}_2 = \{S : S \neq \emptyset;\ S = S_1 \cap S_2 \text{ for some } S_1 \in \mathcal{P}_1, S_2 \in \mathcal{P}_2\}$$

This distance is the minimum number of merges and splits of the sets of $\mathcal{P}_1$ to obtain $\mathcal{P}_2$ (Day, 1981).

- *Closeness* between partitions:

$$c_q(\mathcal{P}_1, \mathcal{P}_2) = \frac{1}{1 + \{d(\mathcal{P}_1, \mathcal{P}_2)\}^q},$$

$q \geq 0$ is a smoothness parameter.

Markov Chain of Partitions, (2)

- *Attractiveness* of partition $\mathcal{P}$:

$$a_q(\mathcal{P}) = \frac{1}{B_M} \sum_{\mathcal{P}' \in \Omega} c_q(\mathcal{P}, \mathcal{P}')$$

- Transition probability:

$$\mathrm{pr}(\mathcal{P}_{t+1} | \mathcal{P}_t) = \frac{c_q(\mathcal{P}_t, \mathcal{P}_{t+1})}{B_M \, a_q(\mathcal{P}_t)}$$

- The stationary probability is proportional to the attractiveness.

$$\mathrm{pr}(\mathcal{P}_t) = \frac{a_q(\mathcal{P}_t)}{B_M \, A_q}, \quad (A_q = B_M^{-1} \sum_{\mathcal{P} \in \Omega} a_q(\mathcal{P})).$$

- We specify the distribution of the initial partition $\mathcal{P}_1$ to be the stationary probability, so that the marginal distribution of all partitions is the set of stationary probabilities.

Markov Chain of Partitions, (3)

Distribution of the partition process:

$$\begin{aligned}\pi(\mathcal{P}_1,\ldots,\mathcal{P}_T) &= \text{pr}(\mathcal{P}_1)\prod_{t=2}^{T}\text{pr}(\mathcal{P}_t|\mathcal{P}_{t-1}) \\ &= \frac{a_q(\mathcal{P}_1)}{B_M\,A_q}\prod_{t=2}^{T}\frac{c_q(\mathcal{P}_{t-1},\mathcal{P}_t)}{B_M\,a_q(\mathcal{P}_{t-1})}\end{aligned}$$

What about q?

Clearly, the prior depends on the choice of smoothing parameter $q \geq 0$.

1. $q = 0$: $c_0(\mathcal{P}, \mathcal{P}') = 1/2$ since $d^0 = 1$. Hence, $\text{pr}(\mathcal{P}_{t+1}|\mathcal{P}_t) = 1/B_M$. This is the *independent-uniforms prior*.
2. q large: $c_q(\mathcal{P}, \mathcal{P}') \approx 0$ if $d(\mathcal{P}, \mathcal{P}') > 1$. Hence, moves that require more than one merge or split are practically impossible.

What about q?

Clearly, the prior depends on the choice of smoothing parameter $q \geq 0$.

1. $q = 0$: $c_0(\mathcal{P}, \mathcal{P}') = 1/2$ since $d^0 = 1$. Hence, $\text{pr}(\mathcal{P}_{t+1}|\mathcal{P}_t) = 1/B_M$. This is the *independent-uniforms prior*.
2. q large: $c_q(\mathcal{P}, \mathcal{P}') \approx 0$ if $d(\mathcal{P}, \mathcal{P}') > 1$. Hence, moves that require more than one merge or split are practically impossible.

We choose q to be a uniform random variable with discrete support

$$\mathcal{Q} = \{0, 0.025, 0.05, 0.075, \ldots, \bar{Q}\},$$

where $\bar{Q}$ is some maximum value of Q such that $a_{\bar{Q}}(\mathcal{P}) \approx a_\infty(\mathcal{P})$.

Prior on Cholesky Parameters

For each set $S_{it} \in \mathcal{P}_t$, we associate the parameters $(\tilde{\phi}_{it}, \tilde{\gamma}_{it})$ so that $(\phi_{m;t}, \gamma_{mt}) = (\tilde{\phi}_{it}, \tilde{\gamma}_{it})$ for all $m \in S_{it}$.

In addition to finding equalities across groups, we seek sparsity in the $\mathbf{T}_m$ matrices. Under multivariate normality, $\phi_{m;jt} = 0$ implies Y_j and Y_t are independent given $Y_1, \ldots, Y_{t-1}$.

We will incorporate a shrinkage prior on $\phi_{m;jt}$ that shrinks the regression coefficient toward zero with more aggressive shrinking for the higher lag $|t - j|$ GARPs by adapting the Bayesian lasso (Park and Casella, 2008).

Prior Distributions

Conditional on the partitions, the Cholesky parameters for set S_{it} are drawn as follows.

$$\begin{aligned}
\delta_{ijt}|\mathcal{P}_t &\sim \text{Exp}\left(\frac{1}{2}\xi_0^2|t-j|^2\right) \quad (j=1,\ldots,t-1), \\
\tilde{\gamma}_{it} \mid \mathcal{P}_t &\sim \text{InvGamma}(\lambda_1,\lambda_2), \\
\tilde{\phi}_{it} \mid \tilde{\gamma}_{it}, \boldsymbol{\Delta}_{it}, \mathcal{P}_t &\sim \text{N}_{t-1}(\mathbf{0}, \tilde{\gamma}_{it}\boldsymbol{\Delta}_{it}), \\
\boldsymbol{\Delta}_{it} &= \text{diag}\left(\delta_{i1t},\ldots,\delta_{i,t-1,t}\right)
\end{aligned}$$

As $|t-j|$ increases, δ_{ijt} is more concentrated near zero, so $\tilde{\phi}_{i;jt}$ is aggressively shrunk toward zero. ξ_0^2 controls the overall rate of shrinkage.

Prior Distributions

Conditional on the partitions, the Cholesky parameters for set S_{it} are drawn as follows.

$$\begin{aligned}
\delta_{ijt}|\mathcal{P}_t &\sim \text{Exp}\left(\frac{1}{2}\xi_0^2|t-j|^2\right) \quad (j=1,\ldots,t-1), \\
\tilde{\gamma}_{it} \mid \mathcal{P}_t &\sim \text{InvGamma}(\lambda_1,\lambda_2), \\
\tilde{\phi}_{it} \mid \tilde{\gamma}_{it}, \boldsymbol{\Delta}_{it}, \mathcal{P}_t &\sim \text{N}_{t-1}(\mathbf{0}, \tilde{\gamma}_{it}\boldsymbol{\Delta}_{it}), \\
\boldsymbol{\Delta}_{it} &= \text{diag}\left(\delta_{i1t},\ldots,\delta_{i,t-1,t}\right)
\end{aligned}$$

As $|t-j|$ increases, δ_{ijt} is more concentrated near zero, so $\tilde{\phi}_{i;jt}$ is aggressively shrunk toward zero. ξ_0^2 controls the overall rate of shrinkage.

Priors for hyperparameters: $\xi_0^2, \lambda_1, \lambda_2 \sim$ ind. Gamma$(1,1)$

Sampling Strategy

Inference under the Covariance Partition Priors requires an MCMC posterior sample. Let $C_t = \{(\phi_{mt}, \gamma_{mt})\}_{m=1}^{M}$, $\Delta_t = \{\boldsymbol{\Delta}_{it}\}_{i=1}^{d_t}$, $H = \{q, \xi_0^2, \lambda_1, \lambda_2\}$, and $\mathcal{Y}$ be the data.

Updating the parameters for the t-th sequential distributions requires sampling from

$$\begin{aligned}&\left[\mathcal{P}_t, C_t, \Delta_t \,|\, \mathcal{P}_{(-t)}, C_{(-t)}, \Delta_{(-t)}, H, \mathcal{Y}\right] \\ &\quad = \quad [\mathcal{P}_t, \Delta_t \,|\, \mathcal{P}_{t-1}, \mathcal{P}_{t+1}, H, \mathcal{Y}] \times [C_t \,|\, \mathcal{P}_t, \Delta_t, H, \mathcal{Y}]\end{aligned}$$

After we draw $(\mathcal{P}_t, \Delta_t)$ from the first conditional, sampling from $[C_t \,|\, \mathcal{P}_t, \Delta_t, H, \mathcal{Y}] = \prod_{i=1}^{d_t} \left[\tilde{\phi}_{it}, \tilde{\gamma}_{it} \,|\, S_{it}, \boldsymbol{\Delta}_{it}, H, \mathcal{Y}\right]$ can be done conjugately from the InvGamma and Normal distributions.

Sample Partition and Shrinkage: $[\mathcal{P}_t, \Delta_t \,|\, \mathcal{P}_{t-1}, \mathcal{P}_{t+1}, H, \mathcal{Y}]$

- Propose a candidate partition $\mathcal{P}_t^\star = \{S_{1t}^\star, \ldots, S_{d_t^\star t}^\star\}$ using a mixture of biased random walk (relocate one group) and split/merge steps.
- Propose new shrinkage factors $\mathbf{\Delta}_t^\star \sim g(\cdot|\mathcal{P}_t, \mathcal{P}_t^\star)$, where g is the conditional distribution for $\mathbf{\Delta}$ given a pooled estimator for the GARPs/IV.
- Calculate the marginal likelihood $\mathcal{L}(Y_{mi}, m \in S_{it}|\mathbf{\Delta}_{it})$ of the all sets:

$$\int \left\{ \prod_{m \in S_{it}} \prod_{i=1}^{n_m} f(Y_{mit}|Y_{mi1}, \ldots, Y_{mi,t-1}; \tilde{\phi}, \tilde{\gamma}) \right\} \pi(\tilde{\phi}|\tilde{\gamma}, \mathbf{\Delta}_{it}) \pi(\tilde{\gamma}) \, \mathrm{d}\tilde{\phi} \, \mathrm{d}\tilde{\gamma}.$$

- Accept the move from $(\mathcal{P}_t, \mathbf{\Delta}_t)$ to $(\mathcal{P}_t^\star, \mathbf{\Delta}_t^\star)$ with Metropolis-Hastings probability $\min\{1, \alpha\}$, where α is

$$\frac{\mathrm{pr}(\mathcal{P}_t^\star|\mathcal{P}_{t-1})\,\mathrm{pr}(\mathcal{P}_{t+1}|\mathcal{P}_t^\star)}{\mathrm{pr}(\mathcal{P}_t|\mathcal{P}_{t-1})\,\mathrm{pr}(\mathcal{P}_{t+1}|\mathcal{P}_t)} \frac{\prod_{i=1}^{d_t^\star} \mathcal{L}(Y_{mi}, m \in S_{it}^\star|\mathbf{\Delta}_{it}^\star)\pi(\mathbf{\Delta}_{it}^\star)}{\prod_{i=1}^{d_t} \mathcal{L}(Y_{mi}, m \in S_{it}|\mathbf{\Delta}_{it})\pi(\mathbf{\Delta}_{it})} \frac{\mathrm{pr}(\mathcal{P}_t^\star \to \mathcal{P}_t)\prod_{i=1}^{d_t} g(\mathbf{\Delta}_{it}|\mathcal{P}_t, \mathcal{P}_t^\star)}{\mathrm{pr}(\mathcal{P}_t \to \mathcal{P}_t^\star)\prod_{i=1}^{d_t^\star} g(\mathbf{\Delta}_{it}^\star|\mathcal{P}_t, \mathcal{P}_t^\star)}.$$

Additional Sampling Comments

- We also develop a step that jointly updates a neighborhood of partitions $\mathcal{P}_t, \ldots, \mathcal{P}_{t+k}$ along with their shrinkage factors.
- The hyperparameter ξ_0^2 is sampled conjugately and λ_1, λ_2 through slice sampling (Neal, 2003).
- q is updated through a Metropolis-Hasting step that requires $a_q(\mathcal{P}_t)$ and $a_{q^\star}(\mathcal{P}_t)$. Since $\mathcal{Q}$ is discrete, we compute $a_q(\mathcal{P})$ for all $q \in \mathcal{Q}$ and $\mathcal{P} \in \Omega$ before hand and look it up during each MH step.
- Because $\boldsymbol{\Delta}_{it}$ is only updated if changes to the set $S_{it} \in \mathcal{P}_t$ are accepted in the partition update move, we include an additional step that conjugately samples the shrinkage factors (inverse Gaussian) to speed mixing.

Depression Study Data

We consider the effectiveness of our covariance partition priors on modeling a data set from a $T = 17$ week depression study (Thase et al., 1997). The study wanted to consider whether pharmocotheraphy was more effective in reducing depression than psychotherapy alone.

Previous analyses have shown that severity of baseline depression symptoms influences the rate and variability of improvement.

We analyze the data with $M = 8$ groups defined by drug treatment or control, the initial severity (high or moderate), and gender. Sample sizes range from 28 to 73 patients, with group 8 at 193 patients.

We will use a quadratic mean function with group-specific regression coefficients with a flat prior. We assume all missing data is ignorable.

Comparison Methods and Model Selection

Priors on partitions:

- Covariance partition prior: $\pi(\mathcal{P}_1, \ldots, \mathcal{P}_T)$ with random q
- Independent-uniforms prior: $q = 0$
- Group pooling: $\mathcal{P}_t = \mathcal{P}_{\text{pool}} = \{\mathcal{M}\}$ for all t

  Assumes a common covariance matrix for all groups
- Group independence: $\mathcal{P}_t = \mathcal{P}_{\text{ind}} = \{\,\{1\}, \cdots, \{M\}\,\}$ for all t

  Assumes an independent covariance matrix for each group

Additionally, we consider both the Cholesky shrinkage model and a non-sparse version with $\mathbf{\Delta} = \sigma^2\mathbf{I}$ with $\sigma^2 \sim \text{InvGamma}(0.1, 0.1)$.

Comparison Methods and Model Selection

Priors on partitions:

- Covariance partition prior: $\pi(\mathcal{P}_1, \ldots, \mathcal{P}_T)$ with random q
- Independent-uniforms prior: $q = 0$
- Group pooling: $\mathcal{P}_t = \mathcal{P}_{\text{pool}} = \{\mathcal{M}\}$ for all t

  Assumes a common covariance matrix for all groups
- Group independence: $\mathcal{P}_t = \mathcal{P}_{\text{ind}} = \{\,\{1\}, \cdots, \{M\}\,\}$ for all t

  Assumes an independent covariance matrix for each group

Additionally, we consider both the Cholesky shrinkage model and a non-sparse version with $\boldsymbol{\Delta} = \sigma^2 \mathbf{I}$ with $\sigma^2 \sim \text{InvGamma}(0.1, 0.1)$.

We compare the model fits between priors using the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002).

Model Selection

$$\mathrm{DIC} = \mathit{Dev} + 2p_D$$
$$\mathit{Dev} = \mathit{Dev}(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}^{-1}} \mid \mathbf{y}_{\text{obs}}) = -2\text{loglik}(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}^{-1}} \mid \mathbf{y}_{\text{obs}})$$
$$p_D = \mathrm{E}_{\text{post}}\{\mathit{Dev}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{y}_{\text{obs}})\} - \mathit{Dev}$$

Model Selection

$$\text{DIC} = Dev + 2p_D$$

$$Dev = Dev(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}^{-1}} \mid \mathbf{y}_{\text{obs}}) = -2\text{loglik}(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}^{-1}} \mid \mathbf{y}_{\text{obs}})$$

$$p_D = \text{E}_{\text{post}}\{Dev(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{y}_{\text{obs}})\} - Dev$$

Partition Prior	Cholesky Prior	Dev	p_D	DIC
cov. partition prior	shrinkage	39,020	513	40,046
cov. partition prior	non-shrink	38,987	544	40,074
indep.-uniforms prior	shrinkage	38,934	595	40,123
indep.-uniforms prior	non-shrink	38,884	643	40,170
$\mathcal{P}_t = \mathcal{P}_{\text{ind}}$	shrinkage	38,658	814	40,286
$\mathcal{P}_t = \mathcal{P}_{\text{pool}}$	non-shrink	39,879	213	40,306
$\mathcal{P}_t = \mathcal{P}_{\text{pool}}$	shrinkage	39,908	200	40,309
$\mathcal{P}_t = \mathcal{P}_{\text{ind}}$	non-shrink	38,549	912	40,374

Posterior Partition Structure

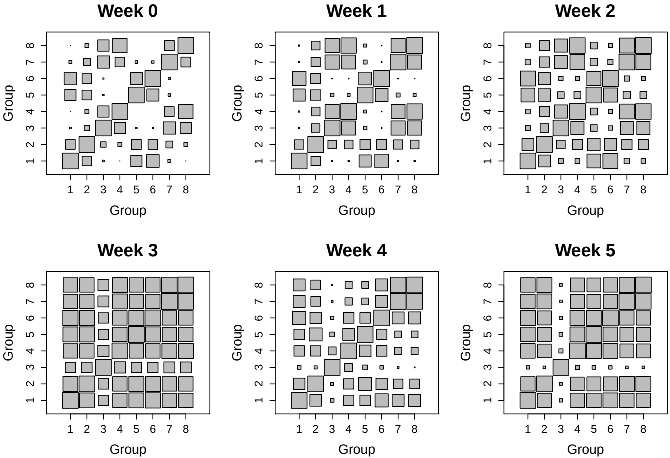

Figure: The posterior probability that m_1 and m_2 are in the same set of the partition $\mathcal{P}_t$, that is, $\text{pr}(\exists S_{it} \in \mathcal{P}_t : \{m_1, m_2\} \in S_{it})$.

Posterior Partition Structure, (2)

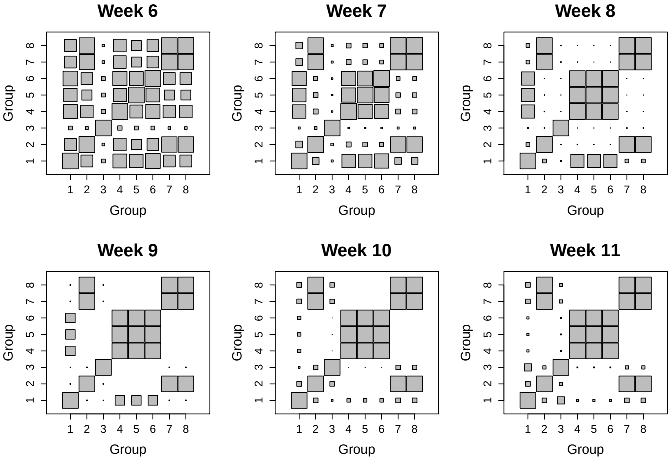

Figure: The posterior probability that m_1 and m_2 are in the same set of the partition $\mathcal{P}_t$, that is, $\text{pr}(\exists S_{it} \in \mathcal{P}_t : \{m_1, m_2\} \in S_{it})$.

Posterior Partition Structure, (3)

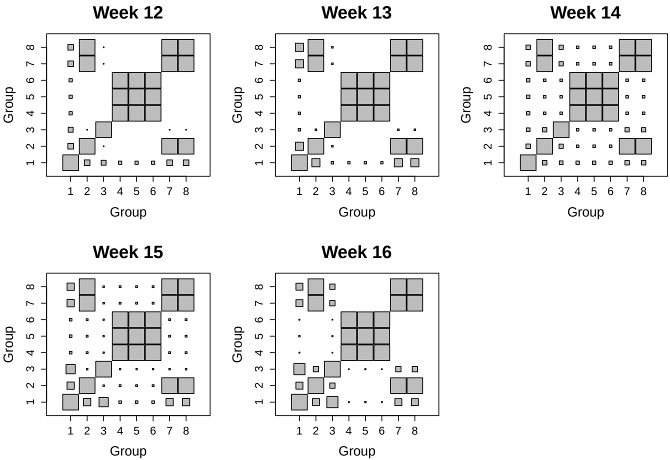

Figure: The posterior probability that m_1 and m_2 are in the same set of the partition $\mathcal{P}_t$, that is, $\text{pr}(\exists S_{it} \in \mathcal{P}_t : \{m_1, m_2\} \in S_{it})$.

Conclusions

- Covariance Partition Priors provide a new approach to parsimoniously estimating multiple longitudinal covariance matrices.
- They consider a large model space by stochastically searching for common structure on the sequential regressions across groups by introducing a Markov chain on partitions.
- Within groups (or sets of a partition), the sequential distributions have sparse structure due to a shrinkage prior.
- Empirical evidence shows good performance of our methodology in both simulation (not shown) and real data situations.
- The posterior partition structures may be interpretable and provide new insight to practitioners and subject-matter experts.

References

- Day, W.H.E. (1981) The complexity of computing metric distances between partitions. *Mathematical Social Sciences*, 1:269–287.
- Neal, R.M. (2003) Slice sampling. *The Annals of Statistics*, 31:705–767.
- Park, T., and Casella, G. (2008) The Bayesian lasso. *JASA*, 103:681–686.
- Pourahmadi, M. (1999) Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86:677-690.
- Pourahmadi, M. (2000) Maximum likelihood estimation of general linear models for multivariate normal covariance matrix. *Biometrika*, 87:425–435.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *JRSSB*, 64:583–639.
- Thase, M., Greenhouse, J., Frank, E., Reynolds, C., Pilkonis, P., Hurley, K., Grochocinski, V., and Kupfer, D. (1997) Treatment of major depression with psychotherapy or psychotherapy- pharmacotherapy combinations. *Archives of General Psychiatry*, 54:1009–1015.