Short Seminar about DRPM Model

exploring the full conditionals and the structural possibilities

Federico Angelo Mor, assisted by
Profs. Alessandra Guglielmi and Alessandro Carminati

Politecnico of Milano
Thesis development

May 30, 2024

Contents

1. Introduction
2. Full conditionals derivation
   - Updating ϕ_0
   - Updating θ_t
   - Updating $\mu^\star_{jt}$
   - Updating α_t
3. Model variations
   - Distributions
   - Covariates
4. What's next

DRPM

Garritt L. Page, Fernando A. Quintana, David B. Dahl (2022)
Dependent Modeling of Temporal Sequences of Random Partitions.
Journal of Computational and Graphical Statistics, 31:2, 614-627.

The main objective of the authors was to define a spatio-temporal model capable of performing ”smooth” clusterings; a model that would favour a gentle evolution in time of the clusters, rather than rough (and therefore less interpretable) changes in them.

Their original model was just focused on time, but the authors showed how it could easily include space by re-defining the random partition model. The goal of the thesis will be to update the model to also account for covariates, deciding where and how to include them, and finally testing it on a real dataset.

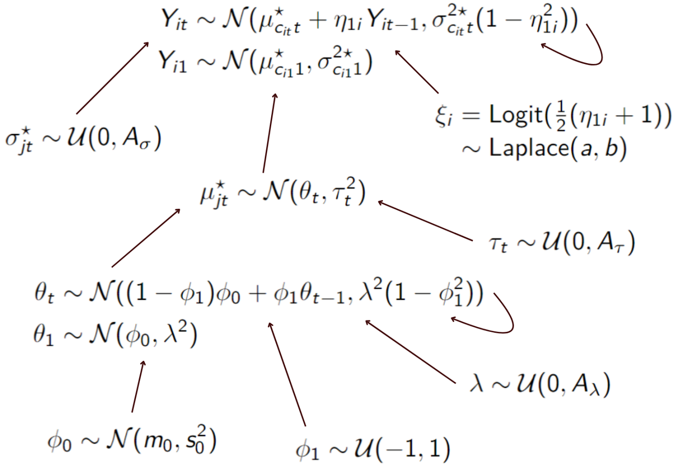
Classical derivation method

$$f(\heartsuit|\text{all the rest}) = \frac{f(\heartsuit, \text{all the rest})}{f(\text{all the rest})} \propto f(\heartsuit,\ \text{all the rest}) \propto \dots$$

$$\begin{aligned}
Y_{it}|Y_{it-1}, \boldsymbol{\mu}^{\star}_{t}, \boldsymbol{\sigma}^{2\star}_{t}, \boldsymbol{\eta}, \boldsymbol{c}_t &\overset{\text{ind}}{\sim} \mathcal{N}(\mu^{\star}_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1-\eta^{2}_{1i})) \\
&\quad i = 1, \dots, n \quad \text{and} \quad t = 2, \dots, T \\
Y_{i1} &\overset{\text{ind}}{\sim} \mathcal{N}(\mu^{\star}_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1}) \\
\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i}+1)) &\overset{\text{ind}}{\sim} \text{Laplace}(a, b) \\
(\mu^{\star}_{jt}, \sigma^{\star}_{jt}) &\overset{\text{ind}}{\sim} \mathcal{N}(\theta_t, \tau^2_t) \times \mathcal{U}(0, A_\sigma) \\
\theta_t|\theta_{t-1} &\overset{\text{ind}}{\sim} \mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi^2_1)) \\
(\theta_1, \tau_t) &\sim \mathcal{N}(\phi_0, \lambda^2) \times \mathcal{U}(0, A_\tau) \\
(\phi_0, \phi_1, \lambda) &\sim \mathcal{N}(m_0, s^2_0) \times \mathcal{U}(-1, 1) \times \mathcal{U}(0, A_\lambda) \\
\{\boldsymbol{c}_t, \dots, \boldsymbol{c}_T\} &\sim \text{tRPM}(\boldsymbol{\alpha}, M) \quad \text{with} \quad \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha)
\end{aligned}$$

Shortcut trough the model graph

(full conditional) $\propto$ (self node distribution) $\cdot$ (parent nodes distributions)



$$Y_{it} \sim \mathcal{N}(\mu^\star_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1-\eta^2_{1i}))$$

$$Y_{i1} \sim \mathcal{N}(\mu^\star_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1})$$

$$\sigma^\star_{jt} \sim \mathcal{U}(0, A_\sigma)$$

$$\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i}+1)) \sim \text{Laplace}(a, b)$$

$$\mu^\star_{jt} \sim \mathcal{N}(\theta_t, \tau^2_t)$$

$$\tau_t \sim \mathcal{U}(0, A_\tau)$$

$$\theta_t \sim \mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi_1^2))$$

$$\theta_1 \sim \mathcal{N}(\phi_0, \lambda^2)$$

$$\lambda \sim \mathcal{U}(0, A_\lambda)$$

$$\phi_0 \sim \mathcal{N}(m_0, s_0^2)$$

$$\phi_1 \sim \mathcal{U}(-1, 1)$$

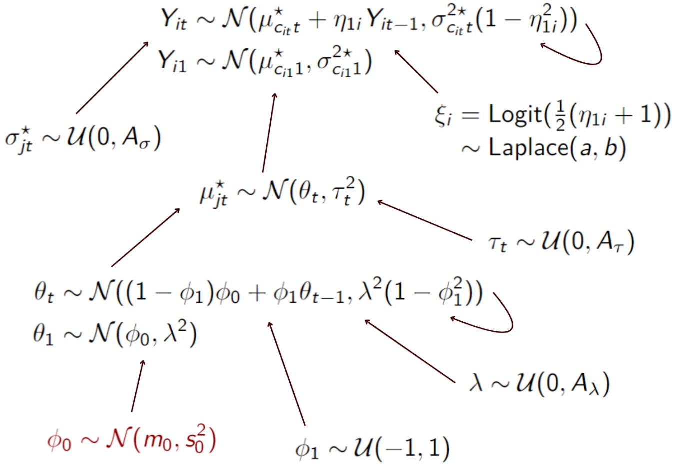
Deriving the full conditionals

For the Normal variables ϕ_0, θ_t, and $\boldsymbol{\mu}_t^\star$ the derivation is the standard one involved in Normal-Normal models, where we iterate the application of the identity

$$\sum d_i(z - c_i)^2 \propto \left(\sum d_i\right)(z - c)^2, \quad c = \frac{\sum d_i c_i}{\sum d_i}$$

The other full conditional is obtainable for the parameter α_t, related to the definition of the RPM, which is involved in a Beta-Binomial structure with the parameters γ_{it}, therefore also her derivation is quite straightforward.

Updating ϕ_0



$$Y_{it} \sim \mathcal{N}(\mu^\star_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1-\eta_{1i}^2))$$

$$Y_{i1} \sim \mathcal{N}(\mu^\star_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1})$$

$$\sigma^\star_{jt} \sim \mathcal{U}(0, A_\sigma)$$

$$\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i}+1)) \sim \text{Laplace}(a, b)$$

$$\mu^\star_{jt} \sim \mathcal{N}(\theta_t, \tau_t^2)$$

$$\tau_t \sim \mathcal{U}(0, A_\tau)$$

$$\theta_t \sim \mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi_1^2))$$

$$\theta_1 \sim \mathcal{N}(\phi_0, \lambda^2)$$

$$\lambda \sim \mathcal{U}(0, A_\lambda)$$

$$\phi_0 \sim \mathcal{N}(m_0, s_0^2)$$

$$\phi_1 \sim \mathcal{U}(-1, 1)$$

Updating ϕ_0

$$\begin{aligned}
f(\phi_0|-) &\propto f(\phi_0)\cdot f((\theta_1,\dots,\theta_T)|\phi_0,-)\\
&= \mathcal{L}_{\mathcal{N}(m_0,s_0^2)}(\phi_0)\cdot\\
&\qquad\left[\mathcal{L}_{\mathcal{N}(\phi_0,\lambda^2)}(\theta_1)\prod_{t=2}^{T}\mathcal{L}_{\mathcal{N}((1-\phi_1)\phi_0+\phi_1\theta_{t-1},\lambda^2(1-\phi_1^2))}(\theta_t)\right]\\
&\propto \exp\left\{-\frac{1}{2s_0^2}(\phi_0-m_o)^2\right\}\exp\left\{-\frac{1}{2\lambda^2}(\phi_0-\theta_1)^2\right\}\cdot\\
&\qquad\exp\left\{-\frac{1}{2\frac{\lambda^2(1-\phi_1^2)}{(T-1)(1-\phi_1)^2}}\left(\phi_0-\frac{(1-\phi_1)(\mathrm{SUM}_t)}{(T-1)(1-\phi_1)^2}\right)^2\right\}\\
&\text{where } \mathrm{SUM}_t = \sum_{t=2}^{T}\theta_t-\phi_1\theta_{t-1}
\end{aligned}$$

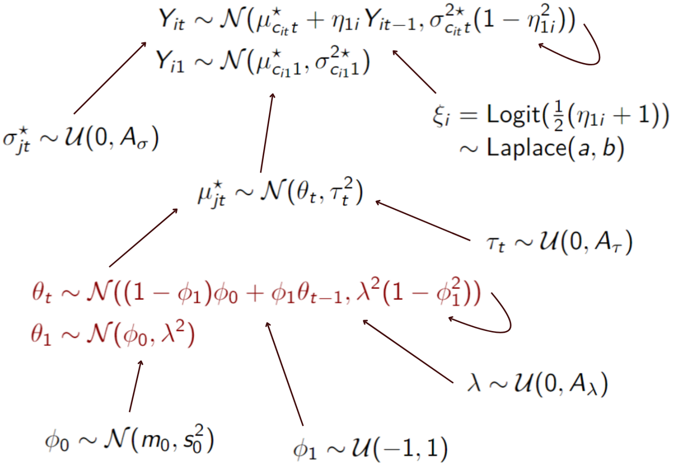
Updating ϕ_0

$$\implies f(\phi_0|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\phi_0(\text{post})}, \sigma^2_{\phi_0(\text{post})}) \text{ with}$$

$$\sigma^2_{\phi_0(\text{post})} = \frac{1}{\frac{1}{s_0^2} + \frac{1}{\lambda^2} + \frac{(T-1)(1-\phi_1)^2}{\lambda^2(1-\phi_1^2)}}$$

$$\mu_{\phi_0(\text{post})} = \sigma^2_{\phi_0(\text{post})} \left[\frac{m_0}{s_0^2} + \frac{\theta_1}{\lambda^2} + \frac{1-\phi_1}{\lambda^2(1-\phi_1^2)} \left(\sum_{t=2}^{T} \theta_t - \phi_1 \theta_{t-1} \right) \right]$$

Updating θ_t



$$Y_{it} \sim \mathcal{N}(\mu^{\star}_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1 - \eta^2_{1i}))$$

$$Y_{i1} \sim \mathcal{N}(\mu^{\star}_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1})$$

$$\sigma^{\star}_{jt} \sim \mathcal{U}(0, A_{\sigma})$$

$$\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i} + 1)) \sim \text{Laplace}(a, b)$$

$$\mu^{\star}_{jt} \sim \mathcal{N}(\theta_t, \tau^2_t)$$

$$\tau_t \sim \mathcal{U}(0, A_{\tau})$$

$$\theta_t \sim \mathcal{N}((1 - \phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1 - \phi^2_1))$$

$$\theta_1 \sim \mathcal{N}(\phi_0, \lambda^2)$$

$$\lambda \sim \mathcal{U}(0, A_{\lambda})$$

$$\phi_0 \sim \mathcal{N}(m_0, s^2_0)$$

$$\phi_1 \sim \mathcal{U}(-1, 1)$$

Updating θ_t

Due to the different law at the first time instant and to the autoregressive component, for this parameter we need to distinguish three cases:

$$
\begin{aligned}
f(\theta_t|-) &\propto f(\theta_t) f(\theta_{t+1}|\theta_t) f(\boldsymbol{\mu}_t^\star|\theta_t,\tau_t^2) && t = 1\\
f(\theta_t|-) &\propto f(\theta_t) f(\theta_{t+1}|\theta_t) f(\boldsymbol{\mu}_t^\star|\theta_t,\tau_t^2) && 1 < t < T\\
f(\theta_t|-) &\propto f(\theta_t) f(\boldsymbol{\mu}_t^\star|\theta_t,\tau_t^2) && t = T
\end{aligned}
$$

Updating θ_t for $t = T$

$$
\begin{aligned}
f(\theta_t|-) &\propto f(\theta_t) f(\boldsymbol{\mu}_t^\star, -) = f(\theta_t) \prod_{j=1}^{k_t} f(\mu_{jt}^\star | \theta_t, -) \\
&= \mathcal{L}_{\mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi_1^2))}(\theta_t) \prod_{j=1}^{k_t} \mathcal{L}_{\mathcal{N}(\theta_t, \tau_t^2)}(\mu_{jt}^\star) \\
&\propto \exp\left\{ -\frac{1}{2(\lambda^2(1-\phi_1^2))} \Big(\theta_t - ((1-\phi_1)\phi_0 + \phi_1\theta_{t-1})\Big)^2 \right\} \cdot \\
&\qquad \exp\left\{ -\frac{k_t}{2\tau_t^2} \left(\theta_t - \frac{\sum_{j=1}^{k_t} \mu_{jt}^\star}{k_t} \right) \right\}
\end{aligned}
$$

Updating θ_t for $t = T$

$$\implies f(\theta_t|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\theta_t(\text{post})}, \sigma^2_{\theta_t(\text{post})}) \text{ with}$$

$$\sigma^2_{\theta_t(\text{post})} = \frac{1}{\frac{1}{\lambda^2(1-\phi_1^2)} + \frac{k_t}{\tau_t^2}}$$

$$\mu_{\theta_t(\text{post})} = \sigma^2_{\theta_t(\text{post})} \left[\frac{\sum_{j=1}^{k_t} \mu^\star_{jt}}{\tau_t^2} + \frac{(1-\phi_1)\phi_0 + \phi_1\theta_{t-1}}{\lambda^2(1-\phi_1^2)} \right]$$

for $t = T$.

Updating θ_t for $1 < t < T$

$$
\begin{aligned}
f(\theta_t|-) &\propto \underbrace{f(\theta_t)f(\boldsymbol{\mu}_t^\star,-)}_{\text{as in the case } t=T} \; f(\theta_{t+1}|\theta_t,-) \\
&= \mathcal{L}_{\mathcal{N}(\mu_{\theta_t(\text{post})},\sigma^2_{\theta_t(\text{post})})}(\theta_t)\mathcal{L}_{\mathcal{N}((1-\phi_1)\phi_0+\phi_1\theta_t,\lambda^2(1-\phi_1^2))}(\theta_{t+1}) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2_{\theta_t(\text{post})}}\left(\theta_t-\mu_{\theta_t(\text{post})}\right)^2\right\}. \\
&\quad \exp\left\{-\frac{1}{2\frac{\lambda^2(1-\phi_1^2)}{\phi_1^2}}\left(\theta_t-\frac{\theta_{t+1}-(1-\phi_1)\phi_0}{\phi_1}\right)^2\right\}
\end{aligned}
$$

Updating θ_t for $1 < t < T$

$$\implies f(\theta_t|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\theta_t(\text{post})}, \sigma^2_{\theta_t(\text{post})}) \text{ with}$$

$$\sigma^2_{\theta_t(\text{post})} = \frac{1}{\frac{1+\phi_1^2}{\lambda^2(1-\phi_1^2)} + \frac{k_t}{\tau_t^2}}$$

$$\mu_{\theta_t(\text{post})} = \sigma^2_{\theta_t(\text{post})} \left[\frac{\sum_{j=1}^{k_t} \mu^\star_{jt}}{\tau_t^2} + \frac{\phi_1(\theta_{t-1} + \theta_{t+1}) + \phi_0(1-\phi_1)^2}{\lambda^2(1-\phi_1^2)} \right]$$

for $1 < t < T$.

Updating θ_t for $t = 1$

$$
\begin{aligned}
f(\theta_t|-) &\propto f(\theta_t) f(\theta_{t+1}|\theta_t,-) f(\boldsymbol{\mu}^\star_t|\theta_t,-) \\
&= \mathcal{L}_{\mathcal{N}(\phi_0,\lambda^2)}(\theta_t)\mathcal{L}_{\mathcal{N}((1-\phi_1)\phi_0+\phi_1\theta_{t-1},\lambda^2(1-\phi_1^2))}(\theta_{t+1}) \prod_{j=1}^{k_t} \mathcal{L}_{\mathcal{N}(\theta_t,\tau_t^2)}(\mu^\star_{jt}) \\
&\propto \exp\left\{-\frac{1}{2\lambda^2}(\theta_t-\phi_0)^2\right\} \cdot \\
&\quad \exp\left\{-\frac{1}{2\frac{\lambda^2(1-\phi_1^2)}{\phi_1^2}}\left(\theta_t - \frac{\theta_{t+1}-(1-\phi_1)\phi_0}{\phi_1}\right)^2\right\} \cdot \\
&\quad \exp\left\{-\frac{k_t}{2\tau_t^2}\left(\theta_t - \frac{\sum_{j=1}^{k_t}\mu^\star_{jt}}{k_t}\right)\right\}
\end{aligned}
$$

Updating θ_t for $t = 1$

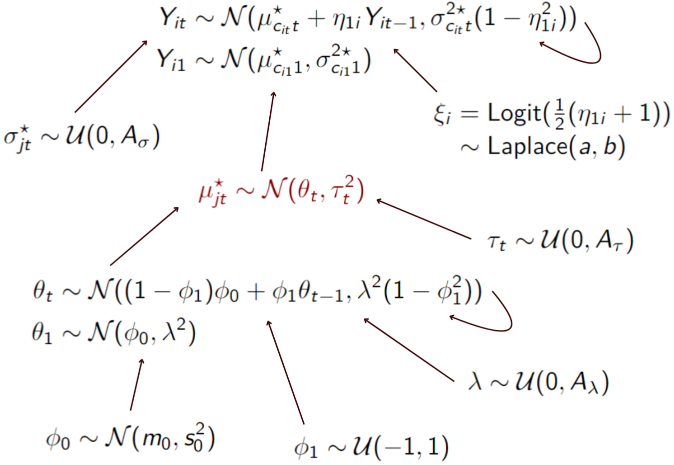
$$\implies f(\theta_t|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\theta_t(\text{post})}, \sigma^2_{\theta_t(\text{post})}) \text{ with}$$

$$\sigma^2_{\theta_t(\text{post})} = \frac{1}{\frac{1}{\lambda^2} + \frac{\phi_1^2}{\lambda^2(1-\phi_1^2)} + \frac{k_t}{\tau_t^2}}$$

$$\mu_{\theta_t(\text{post})} = \sigma^2_{\theta_t(\text{post})} \left[\frac{\phi_0}{\lambda^2} + \frac{\phi_1(\theta_{t+1} - (1-\phi_1)\phi_0)}{\lambda^2(1-\phi_1^2)} + \frac{\sum_{j=1}^{k_t} \mu^\star_{jt}}{\tau_t^2} \right]$$

for $t = 1$.

Updating $\mu^\star_{jt}$



$$Y_{it} \sim \mathcal{N}(\mu^\star_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1-\eta^2_{1i}))$$
$$Y_{i1} \sim \mathcal{N}(\mu^\star_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1})$$

$$\sigma^\star_{jt} \sim \mathcal{U}(0, A_\sigma)$$

$$\xi_i = \text{Logit}(\tfrac{1}{2}(\eta_{1i}+1)) \sim \text{Laplace}(a, b)$$

$$\mu^\star_{jt} \sim \mathcal{N}(\theta_t, \tau^2_t)$$

$$\tau_t \sim \mathcal{U}(0, A_\tau)$$

$$\theta_t \sim \mathcal{N}((1-\phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1-\phi^2_1))$$
$$\theta_1 \sim \mathcal{N}(\phi_0, \lambda^2)$$

$$\lambda \sim \mathcal{U}(0, A_\lambda)$$

$$\phi_0 \sim \mathcal{N}(m_0, s^2_0)$$

$$\phi_1 \sim \mathcal{U}(-1, 1)$$

Updating $\mu^\star_{jt}$ for $t = 1$

$$\begin{aligned}
f(\mu^\star_{jt}|-) &\propto f(\mu^\star_{jt}) f(\boldsymbol{Y}_t|-) \\
&= \mathcal{L}_{\mathcal{N}(\theta_1,\tau_t^2)}(\mu^\star_{jt}) \prod_{i\in S_{jt}} \mathcal{L}_{\mathcal{N}(\mu^\star_{jt},\sigma^{2\star}_{jt})}(Y_{i1}) \\
&\propto \exp\left\{-\frac{1}{2\tau_t^2}(\mu^\star_{jt}-\theta_t)^2\right\} \exp\left\{-\frac{1}{2\sigma^{2\star}_{jt}}\left[\sum_{i\in S_{jt}}(\mu^\star_{jt}-Y_{i1})^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_t^2}(\mu^\star_{jt}-\theta_t)^2\right\} \exp\left\{-\frac{|S_{jt}|}{2\sigma^{2\star}_{jt}}\left(\mu^\star_{jt}-\frac{\mathrm{SUM}_y}{|S_{jt}|}\right)^2\right\} \\
&\text{where } \mathrm{SUM}_y = \sum_{i\in S_{jt}} Y_{i1}
\end{aligned}$$

Updating $\mu^\star_{jt}$ for $t = 1$

$$\implies f(\mu^\star_{jt}|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\mu^\star_{jt}(\text{post})}, \sigma^2_{\mu^\star_{jt}(\text{post})}) \text{ with}$$

$$\sigma^2_{\mu^\star_{jt}(\text{post})} = \frac{1}{\frac{1}{\tau_t^2} + \frac{|S_{jt}|}{\sigma^{2\star}_{jt}}}$$

$$\mu_{\mu^\star_{jt}(\text{post})} = \sigma^2_{\mu^\star_{jt}(\text{post})} \left[\frac{\theta_t}{\tau_t^2} + \frac{\text{SUM}_y}{\sigma^{2\star}_{jt}} \right]$$

for $t = 1$.

Updating $\mu^\star_{jt}$ for $t > 1$

$$
\begin{aligned}
f(\mu^\star_{jt}|-) &\propto f(\mu^\star_{jt}) f(\boldsymbol{Y}_t|-) \\
&= \mathcal{L}_{\mathcal{N}(\theta_1,\tau_t^2)}(\mu^\star_{jt}) \prod_{i\in S_{jt}} \mathcal{L}_{\mathcal{N}(\mu^\star_{jt}+\eta_{1i}Y_{i,t-1},\sigma^{2\star}_{jt}(1-\eta^2_{1i}))}(Y_{it}) \\
&\propto \exp\left\{-\frac{1}{2\tau_t^2}(\mu^\star_{jt}-\theta_t)^2\right\} \cdot \\
&\qquad \exp\left\{-\frac{1}{2\sigma^{2\star}_{jt}}\left[\sum_{i\in S_{jt}} \frac{1}{1-\eta^2_{i1}}\left(\mu^\star_{jt}-(Y_{it}-\eta_{1i}Y_{i,t-1})\right)^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2\tau_t^2}(\mu^\star_{jt}-\theta_t)^2\right\} \exp\left\{-\frac{\text{SUM}_{e2}}{2\sigma^{2\star}_{jt}}\left(\mu^\star_{jt}-\frac{\text{SUM}_y}{\text{SUM}_{e2}}\right)^2\right\}
\end{aligned}
$$

where $\text{SUM}_y = \sum_{i\in S_{jt}} \frac{Y_{it}-\eta_{1i}Y_{i,t-1}}{1-\eta^2_{1i}}$, $\text{SUM}_{e2} = \sum_{i\in S_{jt}} \frac{1}{1-\eta^2_{1i}}$

Updating $\mu^{\star}_{jt}$ for $t > 1$

$$\implies f(\mu^{\star}_{jt}|-) \propto \text{kernel of a } \mathcal{N}(\mu_{\mu^{\star}_{jt}(\text{post})}, \sigma^2_{\mu^{\star}_{jt}(\text{post})}) \text{ with}$$

$$\sigma^2_{\mu^{\star}_{jt}(\text{post})} = \frac{1}{\frac{1}{\tau_t^2} + \frac{\text{SUM}_{e2}}{\sigma^{2\star}_{jt}}}$$

$$\mu_{\mu^{\star}_{jt}(\text{post})} = \sigma^2_{\mu^{\star}_{jt}(\text{post})}\left[\frac{\theta_t}{\tau_t^2} + \frac{\text{SUM}_y}{\sigma^{2\star}_{jt}}\right]$$

for $t > 1$.

Updating α_t

The parameter α_t operates in the definition of the distribution of the clusters. Indeed, in the model we had

$$\{\boldsymbol{c}_t, \dots, \boldsymbol{c}_T\} \sim \text{tRPM}(\boldsymbol{\alpha}, M) \quad \text{with} \quad \alpha_t \overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha)$$

where the α_t are linked to the critical parameters γ_{it}, which were the ones deciding how to reallocate units inside the clusters:

$$\gamma_{it} = \begin{cases} 1 & \text{if unit } i \text{ is not reallocated when moving from } t-1 \text{ to } t \\ & \text{(that is, when } c_{i,t-1} = c_{i,t}) \\ 0 & \text{otherwise} \end{cases}$$

The $\gamma_{it} \overset{\text{ind}}{\sim} \text{Ber}(\alpha_t)$ and so the full conditional derivation follows the classical Beta-Binomial model.

Updating α_t

If time specific α_t:

$$\begin{aligned}\alpha_t &\overset{\text{iid}}{\sim} \text{Beta}(a_\alpha, b_\alpha)\\ \boldsymbol{\gamma}_t = (\gamma_{1t}, \dots, \gamma_{nt}) &\sim \text{Bin}(n, \alpha_t)\\ \implies f(\alpha_t|-) &\sim \text{Beta}\left(a_\alpha + \sum_{i=1}^{n} \gamma_{it}, b_\alpha + n - \sum_{i=1}^{n} \gamma_{it}\right)\end{aligned}$$

If time independent α_t (i.e. $\alpha_t = \alpha \; \forall t$):

$$\begin{aligned}\alpha &\sim \text{Beta}(a_\alpha, b_\alpha)\\ \boldsymbol{\gamma} = (\gamma_{11}, \dots, \gamma_{nT}) &\sim \text{Bin}(nT, \alpha)\\ \implies f(\alpha|-) &\sim \text{Beta}\left(a_\alpha + \sum_{i=1}^{n}\sum_{t=1}^{T} \gamma_{it}, b_\alpha + nT - \sum_{i=1}^{n}\sum_{t=1}^{T} \gamma_{it}\right)\end{aligned}$$

Model variations

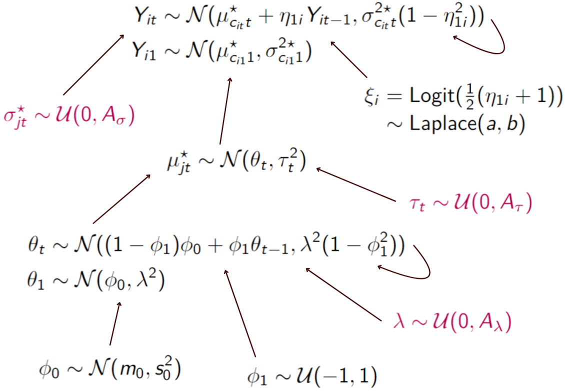
The model framework is quite adjustable: the relevant (and most complex) part is just the partition sampling scheme, while the surrounding structure is fairly flexible.

Therefore, it allows easily for possible variations or extensions, about

- changes in the distribution of some parameters;
- definition of where and how to include covariates.

Distribution choices

Some parameters are updated in the original sampling algorithm through a Metropolis step. However, we could change their distribution to recover also for them a full conditional. This may also speed up the computation.



$Y_{it} \sim \mathcal{N}(\mu^{\star}_{c_{it}t} + \eta_{1i} Y_{it-1}, \sigma^{2\star}_{c_{it}t}(1 - \eta^2_{1i}))$

$Y_{i1} \sim \mathcal{N}(\mu^{\star}_{c_{i1}1}, \sigma^{2\star}_{c_{i1}1})$

$\sigma^{\star}_{jt} \sim \mathcal{U}(0, A_\sigma)$

$\xi_i = \text{Logit}(\frac{1}{2}(\eta_{1i} + 1))$
$\sim \text{Laplace}(a, b)$

$\mu^{\star}_{jt} \sim \mathcal{N}(\theta_t, \tau^2_t)$

$\tau_t \sim \mathcal{U}(0, A_\tau)$

$\theta_t \sim \mathcal{N}((1 - \phi_1)\phi_0 + \phi_1\theta_{t-1}, \lambda^2(1 - \phi^2_1))$

$\theta_1 \sim \mathcal{N}(\phi_0, \lambda^2)$

$\lambda \sim \mathcal{U}(0, A_\lambda)$

$\phi_0 \sim \mathcal{N}(m_0, s^2_0)$

$\phi_1 \sim \mathcal{U}(-1, 1)$

Covariates modeling

The partition model proposed originally was able to generate spatial informed clusters evolving over time. The thesis goal is to extend this model to account also for covariates, mainly inside the clusters definition, by updating the EPPF as

$$P(\rho_t | M, \boldsymbol{\nu}_0, X_t) \propto \prod_{j=1}^{k_t} c(S_{jt}|M) g(\boldsymbol{s}^\star_{jt}|\boldsymbol{\nu}_0) g(X^\star_{jt})$$

but possibly also inside the likelihood of the model; for example acting on the cluster specific parameters, by moving from scalars $\mu^\star_{jt}$, which gave

$$Y_{it} \sim \mathcal{N}(\mu^\star_{c_{it},t} + \eta_{1i} Y_{i,t-1}, \dots)$$

to regression vectors $\beta^\star_{jt}$, resulting in

$$Y_{it} \sim \mathcal{N}({\beta^\star_{c_{it},t}}^T \boldsymbol{x}_{it} + \eta_{1i} Y_{i,t-1}, \dots)$$

Covariates modeling

There are several choices on how to design $g(X_{jt}^{\star})$ in

$$P(\rho_t | M, \boldsymbol{\nu}_0, X_t) \propto \prod_{j=1}^{k_t} c(S_{jt}|M) g(\boldsymbol{s}_{jt}^{\star}|\boldsymbol{\nu}_0) g(X_{jt}^{\star})$$

where

- $\rho_t = \{S_{1t}, \dots, S_{k_t t}\}$ is the partition at time t.
- M and $\boldsymbol{\nu}_0$ are parameters for the cohesion functions on the cluster $c(S_{jt}|M)$ and on the spatial part $g(\boldsymbol{s}_{jt}^{\star}|\boldsymbol{\nu}_0)$.
- $X_{jt}^{\star}$ is a $C \times |S_{jt}|$ matrix (seen as set of vectors) of values the covariates for the units in cluster j at time t.

Covariates modeling

In the case of multiple covariates, we can split $g(\cdot)$ into a product on the components:

$$g(X_{jt}^{\star}) = \prod_{c=1}^{C} g_{(c)}(\boldsymbol{x}_{(c)jt}^{\star})$$

This split possibly could allow to assign different weights on different covariates.

While in the case of a covariate independent from time we could just re-include its contribute in all time instants and provide them a specific cohesion function.

Cohesion function choices

We now list a possible set of cohesion functions, and wlog we assume to be working with a single covariate vector $\boldsymbol{x}_{jt}^{\star} = \{x_i : i \in S_{jt}\}$.

(1) Auxiliary similarity function: to deal with covariates as if they were random objects.

$$g(\boldsymbol{x}_{jt}^{\star}) = \int \underbrace{\prod_{i \in S_{jt}} q(x_i | \boldsymbol{\xi}_j^{\star})}_{\substack{\text{"likelihood" of} \\ \text{the covariates}}} q(\boldsymbol{\xi}_j^{\star}) d\boldsymbol{\xi}_j^{\star}$$

(2) Double dipper similarity: as in (1) but we use the posterior predictive distribution.

$$g(\boldsymbol{x}_{jt}^{\star}) = \int \prod_{i \in S_{jt}} q(x_i | \boldsymbol{\xi}_j^{\star}) q(\boldsymbol{\xi}_j^{\star} | \boldsymbol{x}_{jt}^{\star}) d\boldsymbol{\xi}_j^{\star}$$

Cohesion function choices

(3) Cluster variance/entropy similarity function: this allows easily to account also for categorical covariates.

$$g(\boldsymbol{x}_{jt}^{\star}) = \exp\left\{-\alpha H(\boldsymbol{x}_{jt}^{\star})\right\}$$

$$\text{continuous} \quad H(\boldsymbol{x}_{jt}^{\star}) = \frac{1}{|S_{jt}|}\sum_{l \in S_{jt}} (x_l - \bar{x}_j)^2$$

$$\text{categorical} \quad H(\boldsymbol{x}_{jt}^{\star}) = -\sum_{r=1}^{R} \hat{p}_r \log(\hat{p}_r)$$

where $\bar{x}_j$ is the mean of the values inside $\boldsymbol{x}_{jt}^{\star}$, while $\hat{p}_r$ is the proportion of values of $\boldsymbol{x}_{jt}^{\star}$ belonging to category r.

Cohesion function choices

(4) Total Grower dissimilarity: this and the following can directly work comparing the vectors of covariates.

$$g(X_{jt}^{\star}) = \exp\left\{-\alpha \sum_{l,k \in S_{jt}: l \neq k} d(\boldsymbol{x}_{lt}, \boldsymbol{x}_{kt})\right\}$$

(5) Average Grower dissimilarity:

$$g(X_{jt}^{\star}) = \exp\left\{-\frac{2\alpha}{|S_{jt}|(|S_{jt}|-1)} \sum_{l,k \in S_{jt}: l \neq k} d(\boldsymbol{x}_{lt}, \boldsymbol{x}_{kt})\right\}$$

What's next, and tentative work plan

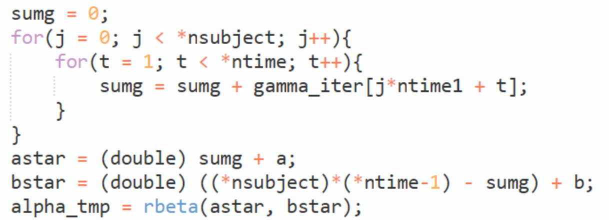
1. Finishing to understand the partition sampling method ("complex" due to the time dependence introduced by the parameters γ_{it}). july/august?
2. After that, starting to implement the model with code, choosing between Julia or C/C++. july/august/september?
3. In the end, testing it on the air pollution dataset. september/october?

As a (not-so)side note, why Julia could suit better?

- more flexible, scalable, readable, and faster (or at least with equal performance, being it also a compiled language) than C/C++.
- support from existing libraries such as Distributions, MCMCChains, Plots, or DataFrames.
- easier exporting procedure on R trough JuliaCall.

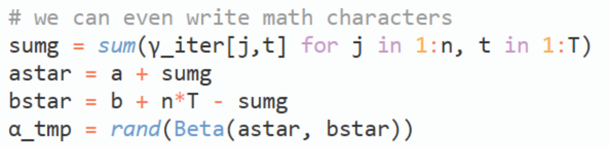
What's next, and tentative work plan

Moreover, Julia has a syntax closer to the mathematical writing.



```
sumg = 0;
for(j = 0; j < *nsubject; j++){
    for(t = 1; t < *ntime; t++){
        sumg = sumg + gamma_iter[j*ntime1 + t];
    }
}
astar = (double) sumg + a;
bstar = (double) ((*nsubject)*(*ntime-1) - sumg) + b;
alpha_tmp = rbeta(astar, bstar);
```

Figure: C/C++ section of DRPM code :/



```
# we can even write math characters
sumg = sum(γ_iter[j,t] for j in 1:n, t in 1:T)
astar = a + sumg
bstar = b + n*T - sumg
α_tmp = rand(Beta(astar, bstar))
```

Figure: How it would look like in Julia :)