Documentation for Matlab Code for "Colombian Women's Life Patterns: A Multivariate Density Regression Approach"

S. Wade* R. Piccarreta † A. Cremaschi‡ I. Antoniano-Villalobos §

November 17, 2020

In this documentation, we include a discussion on the link functions for a mixed-type response in Section 1, followed by a description on how to specify the hyperparameters in Section 2 and full details on the adaptive MCMC and SMC algorithms in Section 3, and finally, a description of how to compute several posterior and predictive quantities of interest in Section 4. We refer the interested reader to the folder `SimulationStudy`, which contains a demo for reproducing the results in the paper. `SimulatedStudyRun.m` is the main file demonstrating how to run the code on simulated data.

1 Link Functions

The code assumes a response $\mathbf{z}$ of dimension d, consisting of b age at event variables, and $d-b$ binary variables. The d-dimensional observed response $\mathbf{z}$ is linked to the d-dimensional *latent response* $\mathbf{y}$ through the functions $h_\ell(\mathbf{y}, \mathbf{x}) = z_\ell$ for $\ell = 1, \ldots, d$. These terms define constrained regions for the latent $\mathbf{y}$, such that $y_\ell \in (l_\ell, u_\ell)$ for $\ell = 1, \ldots, d$. Concretely, for

*School of Mathematics, University of Edinburgh, UK
†Department of Decision Sciences, Dondena Research Centre, Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy
‡Yale-NUS College, Singapore
§Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy

the $\ell = 1, \ldots, d - b$ age at event variables

$$z_\ell = h_\ell(\mathbf{y}, \mathbf{x}) = c_\ell(\mathbf{y}, \mathbf{x})\lfloor \exp(y_\ell) \rfloor, \quad \text{with } c_\ell(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{(0, x_1+1)}(\exp(y_\ell)),$$

and x_1 denotes the age at interview. For the $\ell = d - b + 1, \ldots, d$ binary variables,

$$z_\ell = h_\ell(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{[0,\infty)}(y_\ell).$$

In this case, the bounds required for posterior simulation are obtained from inverting $z_\ell = h_\ell(\mathbf{y}, \mathbf{x})$:

$$(l_\ell, u_\ell) = \begin{cases} (\log(x_1 + 1), \infty) & \text{for censored } z_\ell = 0 \\ (\log(z_\ell), \log(z_\ell + 1)) & \text{for uncensored } z_\ell \neq 0 \end{cases}, \text{ when } \ell = 1, \ldots, d - b \tag{1}$$

$$(l_\ell, u_\ell) = \begin{cases} (-\infty, 0) & \text{for } z_\ell = 0 \\ (0, \infty) & \text{for } z_\ell = 1 \end{cases}, \text{ when } \ell = d - b + 1, \ldots, d.$$

The model can also accommodate other response types by appropriately defining the link function that maps the d-dimensional latent response $\mathbf{y}$ to the d-dimensional observed $\mathbf{z}$. In this case, four files need to be changed:

1. `linkfunction.m`: defining the link function mapping $\mathbf{y}$ to $\mathbf{z}$,
2. `invlinkfunctions.m`: the inverse of the link function, defining bounds (l_ℓ, u_ℓ) for y_ℓ based on the observed z_ℓ, $\ell = 1, \ldots, d$,
3. lines 28-40 of `empiricalhyperparameters.m`: the initialization of the latent $\mathbf{y}$ used to set the empirical hyperparameters,
4. lines 96-113 of `AT_NWR.m`: the initialization of the latent $\mathbf{y}$ for the MCMC.

For example, in our motivating case study, we have constraints on the observed $\mathbf{z}$; specifically, the age at first child z_3 must be greater than the age at sexual debut z_1. We enforce this constraint through the link function

$$z_3 = h_3(\mathbf{y}, \mathbf{x}) = c_3(\mathbf{y}, \mathbf{x})\lfloor \exp(y_1) + \exp(y_3) \rfloor,$$

with $c_3(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{(0,x_1+1)}(\exp(y_1)+\exp(y_3))$. In this case, the bounds obtained from inverting

$z_3 = h_3(\mathbf{y}, \mathbf{x})$ are:

$$(l_3, u_3) = \begin{cases} (\log(\max[0, x_{i,1} + 1 - \exp(y_{i,1})]), \infty) & \text{for censored } z_3 = 0 \\ (\log(\max[0, z_{i,\ell} - \exp(y_{i,1})]), \log(z_{i,\ell} - \exp(y_{i,1}) + 1)) & \text{for uncensored } z_3 \neq 0 \end{cases},$$

This is an example of the more general case that we consider, when the bounds (l_ℓ, u_ℓ) may depend on $y_{\ell'}$ for $\ell' \neq \ell$.

Finally, we advise caution when initialising the latent response for the age at event variables of censored data. While one may be tempted to initialise $y_\ell = l_\ell + 1$, this may result in an extreme value for the imputed age $\exp(y_\ell) = \exp(l_\ell + 1)$. Instead, we use the initialisation $y_\ell = \log(\exp(l_\ell) + 1)$ to avoid poor, extreme values.

2 Empirical Hyperparameters

The algorithm requires specification of the hyperparameters, which are passed to the file `AT_NWR.m` through a structured array called `hyperparameters` containing:

1. (`beta0`, `Sigma0`, `U_iC`,`nu`): the hyperparameters $(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \mathbf{U}, \nu)$ of the matrix-variate Normal-Inverse Wishart prior for the regression parameters $(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$;

2. (`mu0`, `ic`, `a1`,`a2`): the hyperparameters $(\boldsymbol{\mu}_0, \mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ of the Normal-Gamma prior for covariate-dependent weight parameters $(\boldsymbol{\mu}_j, \boldsymbol{\tau}_j)$ associated to the continuous covariates;

3. `alpha_rho`: the hyperparameters $\boldsymbol{\varrho}$ of the Beta prior for covariate-dependent weight parameters $\boldsymbol{\rho}_j$ associated to the categorical covariates;

4. `M`: the hyperparameter ζ of the stick-breaking prior on v_j.

The file `empiricalhyperparameters.m` sets them empirically as follows. First, the prior parameters $(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0, \mathbf{U}, \nu)$ for $(\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$ are specified based on multivariate linear regression fit to the data. Specifically, for $\ell = 1, \ldots, b$, we set $y_{i,\ell} = (l_{i,\ell} + u_{i,\ell})/2$ and $y_{i,\ell} = \log(x_{i,1} + 2)$ for uncensored and censored observations, respectively, where the bounds $l_{i,\ell}$ and $u_{i,\ell}$ are defined in (1). Additionally, we let $y_{i,3} = -1$ for $z_{i,3} = 0$ and $y_{i,3} = 1$ for $z_{i,3} = 1$. A multivariate linear regression fit on these auxiliary responses gives estimates $\widehat{\boldsymbol{\beta}}$ of the linear coefficients and $\widehat{\boldsymbol{\Sigma}}$ of the covariance matrix. We then define

$$\mathbb{E}[\boldsymbol{\beta}_j] = \boldsymbol{\beta}_0 = \widehat{\boldsymbol{\beta}} \quad \text{and} \quad \mathbb{E}[\boldsymbol{\Sigma}_j] = \frac{1}{\nu - b - 1}\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Sigma}}.$$

Together, $\mathbf{U}$ and $\boldsymbol{\Sigma}_j$ reflect the variability of $\boldsymbol{\beta}_j$ across components, and we set $\mathbf{U}$ such that $\min(\text{diag}(\widehat{\boldsymbol{\Sigma}}))\,\mathbf{U} = g(\mathbf{X}^\intercal\mathbf{X})^{-1}$, with g being an input to the function, called `gprior` (Zellner, 1986). We further set $\nu = b + 3$, to ensure the existence of the first and second moments of $\boldsymbol{\Sigma}_j$ a-priori. Other specified hyperparameters include $\mu_{0,1} = \overline{x}_1$, $u_1 = 1/2$, $\alpha_1 = 2$, $\gamma_1 = u_1(\text{range}(x_{1:n,1})/4)^2$, $\boldsymbol{\varrho}_k = (1,1)$ for $k = p+1, \ldots, q$, and the parameters of the stick-breaking prior are $\zeta = 1$. Here $\overline{x}_1$ and $\text{range}(x_{1:n,1})$ denote the sample mean and range of $(x_{1,1}, \ldots, x_{n,1})$.

3 Posterior Inference: Further Details

In this section, we provide full details on the algorithm used for inference under the proposed model. We divide the section into two parts, concerning the MCMC algorithm for fixed truncation, and the SMC algorithm for the adaptive truncation, as reported in the paper. Code implementing the algorithm is contained in the file `AT_NWR.m`, with the adaptive MCMC in lines 314-768, followed by the SMC in lines 770-1356.

The file `AT_NWR.m` takes as an input a set of algorithm-specific parameters called `mcmcsmc` for both the MCMC and SMC. Specifically, `mcmcsmc` is a structured array containing:

1. `numbofparts`: the number of MCMC draws/number of SMC particles;
2. `start_trunc`: the initial truncation level J_0;
3. `burnin`: burnin period of the MCMC;
4. `every`: thinning factor of the MCMC;
5. `epsilon_trunc`: the discrepancy threshold defined by $\delta =$ `epsilon_trunc`$*$ `numbofparts`;
6. `numb_trunc`: the number of consecutive increments $I =$ `numb_trunc` that the discrepancy threshold must be satisfied to stop;
7. `numbofMCMC`: the small number of MCMC iterations for the rejuvenation step of the SMC;

8. top_trunc: the maximum level of truncation.

Based on our simulation studies, we suggest using a conservative guess of the number of components to set `start_trunc`. We suggest to keep the number of particles moderate, e.g. `numbofparts` = 1000, with large thinning to ensure a high ESS from the MCMC simulations. Griffin (2016) suggests the values `numbofMCMC` = 3, `epsilon_trunc` = 0.01 and `numb_trunc` = 4.

3.1 MCMC for Fixed Truncation

The first step of the adaptive truncation algorithm uses MCMC to simulate from the posterior of an approximate model with a fixed number of components J_0:

$$\mathbf{P}^n_{J_0}(\mathbf{w},\boldsymbol{\psi},\boldsymbol{\theta},\mathbf{y}|\mathbf{z},\mathbf{x}) \propto \mathbf{P}_{J_0}(\mathbf{w},\boldsymbol{\psi},\boldsymbol{\theta}) \prod_{i=1}^{n}\sum_{j=1}^{J_0} w_j(\mathbf{x}_i|\boldsymbol{\psi}_j)\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_j,\boldsymbol{\Sigma}_j)\prod_{\ell=1}^{d}\mathbb{1}_{\{z_{i,\ell}\}}(h_{i,\ell}), \quad (2)$$

with truncated covariate dependent weights

$$w_j(\mathbf{x}|\boldsymbol{\psi}_j)) = \frac{w_j g(\mathbf{x}|\boldsymbol{\psi}_j)}{\sum_{j'=1}^{J} w_{j'} g(\mathbf{x}|\boldsymbol{\psi}_{j'})}. \quad (3)$$

Additionally, we use the shorthand notation $h_{i,\ell} = h_\ell(\mathbf{y}_i,\mathbf{x}_i)$ for the functions linking the latent variables to the observed responses. The MCMC algorithm entails sampling from the full-conditionals of the parameters $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\boldsymbol{\tau}$, $\boldsymbol{\rho}$, $\mathbf{w}$, and $\mathbf{y}$. Due to lack of conjugacy, we resort to a generic Metropolis-within-Gibbs scheme to perform posterior sampling. The algorithm used here, described as Algorithm 6 in Griffin and Stephens (2013), adapts the covariance matrix in the random walk algorithm to achieve both a specified average acceptance rate ($a_0 = 0.234$) and a covariance matrix equal to $2.4^2/\mathfrak{p}$ times the covariance matrix of the posterior, $\mathfrak{p}$ being the dimension of the parameter of interest. These criteria have been shown to be optimal in many settings (Gelman et al., 1996; Roberts et al., 1997; Roberts and Rosenthal, 2001). In more detail, suppose that we want to sample a block of parameters $\boldsymbol{\phi}$ of dimension $\mathfrak{p}$ from a distribution with probability density function Q. First, we consider a transformation $t(\boldsymbol{\phi})$ that has full support on $\mathbb{R}^{\mathfrak{p}}$. At each iteration m, we propose a new value $\boldsymbol{\phi}^*$ such that:

$$\mathbf{t}^* \equiv t(\boldsymbol{\phi}^*) = t(\boldsymbol{\phi}^{m-1}) + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} \sim \mathrm{N}(0,\boldsymbol{\xi}^{m-1}). \quad (4)$$

We accept $\boldsymbol{\phi}^m = \boldsymbol{\phi}^*$ with probability equal to the minimum between 1 and the ratio:

$$a(\boldsymbol{\phi}^*, \boldsymbol{\phi}^{m-1}) = \frac{Q(\boldsymbol{\phi}^*)}{Q(\boldsymbol{\phi}^{m-1})} \frac{|\mathcal{J}_t(\boldsymbol{\phi}^{m-1})|}{|\mathcal{J}_t(\boldsymbol{\phi}^*)|}.$$

We initialize the adaptive Metropolis-Hastings (MH) algorithm in Section 4, with $\boldsymbol{\xi}^0 = \xi^0 \mathbb{I}_{\mathfrak{p}}$, where $\mathbb{I}_{\mathfrak{p}}$ denotes the identity matrix of dimension $\mathfrak{p}$. The initial value ξ^0 was calibrated for each parameter block in order to achieve reasonable initial acceptance rates. After $M_0 = 100$ iterations, we update the covariance matrix of the proposal density according to the formula:

$$\boldsymbol{\xi}^m = \frac{s^m}{m-1} \left(\sum_{m'=1}^{m} \boldsymbol{\phi}^{m'} (\boldsymbol{\phi}^{m'})^\top - \frac{1}{m} \sum_{m'=1}^{m} \boldsymbol{\phi}^{m'} \left(\sum_{m'=1}^{m} \boldsymbol{\phi}^{m'} \right)^\top \right) + s^m \epsilon \, \mathbb{I}_{\mathfrak{p}},$$

where

$$s^m = \Upsilon(\log(s^{m-1}) + m^{-0.7}(a(\boldsymbol{\phi}^*, \boldsymbol{\phi}^{m-1}) - a_0)), \quad s^0 = 2.4^2/\mathfrak{p},$$

$$\Upsilon(s) = \begin{cases} \exp(-50) & \text{if } s < -50 \\ \exp(s) & \text{if } s \in [-50, 50] \\ \exp(50) & \text{if } s > 50 \end{cases}.$$

The value $\epsilon = 0.001$ is chosen to ensure a minimum level of exploration of the parameter space.

The target distribution Q for each block of parameters corresponds to the full conditional distribution extracted from the posterior in (2). Recall that $\boldsymbol{\beta} = \boldsymbol{\beta}_{1:J_0}$, with analogous notation for $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, $\boldsymbol{\tau}$, and $\boldsymbol{\rho}$. Throughout, we make use of the subscript notation $-j$, e.g. $\boldsymbol{\beta}_{-j}$, to denote the corresponding array without the j-th entry. Details for the update of each parameter block are subsequently described.

Adaptive MH for $\boldsymbol{\beta}_j$. Lines 431-477 of `AT_NWR.m`. Each $\boldsymbol{\beta}_j$, $j = 1, \ldots, J_0$, is treated separately. In this case, a simple and convenient transformation is the vectorization $t(\boldsymbol{\beta}_j) = \mathrm{vec}(\boldsymbol{\beta}_j) \in \mathbb{R}^{\mathfrak{p}}$, with $\mathfrak{p} = (q+1)d$, so that the determinant of the Jacobian is $|\mathcal{J}_t(\boldsymbol{\beta}_j)| \equiv 1$. Therefore, the acceptance ratio $a(\boldsymbol{\beta}_j^*, \boldsymbol{\beta}_j^m)$ for the move to the MH proposal $\boldsymbol{\beta}_j^*$ from the current value depends only on the target distribution, which corresponds to the full conditional $Q(\boldsymbol{\beta}_j) = Q(\boldsymbol{\beta}_j|\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\beta}_{-j}, \boldsymbol{\Sigma}, \mathbf{x}, \mathbf{y})$ given by

$$Q(\boldsymbol{\beta}_j) \propto \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)\right]\right\} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'}),$$

where $\mathrm{tr}(\mathbf{A})$ denotes the trace of the matrix $\mathbf{A}$. Thus, the acceptance ratio is given by

$$a(\boldsymbol{\beta}_j^*, \boldsymbol{\beta}_j) = \frac{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_0)\right]\right\} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}^*, \boldsymbol{\Sigma}_{j'})}{\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)\right]\right\} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})},$$

where $\boldsymbol{\beta}_{j'}^* = \boldsymbol{\beta}_{j'}$ for $j' \neq j$. Note that when evaluating the likelihood given the proposed parameter, only the parametric mixture likelihoods $\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_j^*, \boldsymbol{\Sigma}_j)$, for $i = 1, \ldots, n$, need to be re-evaluated, while the value of the remaining terms, including the covariate dependent weights, can be recycled from the previous step.

Adaptive MH for $\boldsymbol{\Sigma}_j$. Lines 372-429 of `AT_NWR.m`. Each $\boldsymbol{\Sigma}_j$, for $j = 1, \ldots, J_0$, is treated separately. First, a transformation is proposed which is based on the vectorization of a decomposition of the matrix, $\boldsymbol{\Sigma}_j = \mathbf{L}_j \mathbf{D}_j \mathbf{L}_j^\intercal$, where $\mathbf{L}_j$ is a lower triangular matrix with unit entries on the diagonal and $\mathbf{D}_j$ is a diagonal matrix with positive entries. Specifically,

$$t(\boldsymbol{\Sigma}_j) = (\log(D_{j,1,1}), L_{j,2:d,1}, \log(D_{j,2,2}), L_{j,3:d,2}, \ldots, \log(D_{j,d-1,d-1}), L_{j,d,d-1}, \log(D_{j,d,d}))^\intercal.$$

It can be seen that $t(\boldsymbol{\Sigma}_j) \in \mathbb{R}^{\mathfrak{p}}$ for $\mathfrak{p} = d(d+1)/2$, and an inverse transformation of the proposed $\mathbf{t}^*$ in equation (4) can be found to obtain the proposed value $\boldsymbol{\Sigma}_j^*$. Specifically, the proposed matrices $\mathbf{L}_j^*$ and $\mathbf{D}_j^*$ are easily obtained as

$$\mathbf{L}_j^* = \begin{bmatrix} 1 & 0 & \ldots & & 0 \\ t_2^* & 1 & 0 & \ldots & 0 \\ \vdots & & \ddots & & \\ t_d^* & t_{2d-1}^* & & & 1 \end{bmatrix}, \quad \mathbf{D}_j^* = \begin{bmatrix} \exp(t_1^*) & 0 & \ldots & 0 \\ 0 & \exp(t_{d+1}^*) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \exp(t_{d(d+1)/2}^*) \end{bmatrix},$$

and $\boldsymbol{\Sigma}_j^* = \mathbf{L}_j^* \mathbf{D}_j^* \mathbf{L}_j^{*\intercal}$. Furthermore, it can be shown that the determinant of the Jacobian of the transformation depends only on the diagonal elements $D_{j,\ell,\ell}$ of the matrix $\mathbf{D}_j$, $|\mathcal{J}_t(\boldsymbol{\Sigma}_j)| = \prod_{\ell=1}^{d} 1/D_{j,\ell,\ell}^{d+1-\ell}$. The final element required to calculate the acceptance ratio is the full conditional distribution $Q(\mathbf{\Sigma}_j) = Q(\mathbf{\Sigma}_j|\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\beta}, \mathbf{\Sigma}_{-j}, \mathbf{x}, \mathbf{y})$ given by

$$Q(\mathbf{\Sigma}_j) \propto \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}_j^{-1}\left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0) + \mathbf{\Sigma}_0\right)\right]\right\} |\mathbf{\Sigma}_j|^{-\frac{q+\nu+d}{2}-1}$$
$$* \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \mathbf{\Sigma}_{j'}).$$

Thus, the acceptance ratio for the proposed move to $\mathbf{\Sigma}_j^*$ from the current value $\mathbf{\Sigma}_j$ is

$$a(\mathbf{\Sigma}_j^*, \mathbf{\Sigma}_j) = \frac{|\mathbf{\Sigma}_j|^{\frac{q+\nu+d}{2}+1} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}_j^{*\,-1}\left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0) + \mathbf{\Sigma}_0\right)\right]\right\}}{\left|\mathbf{\Sigma}_j^*\right|^{\frac{q+\nu+d}{2}+1} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\mathbf{\Sigma}_j^{-1}\left((\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)^\intercal \mathbf{U}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0) + \mathbf{\Sigma}_0\right)\right]\right\}}$$
$$* \prod_{\ell=1}^{d} \left(\frac{D_{j,\ell,\ell}^*}{D_{j,\ell,\ell}}\right)^{d+1-\ell} \frac{\prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \mathbf{\Sigma}_{j'}^*)}{\prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \mathbf{\Sigma}_{j'})},$$

where $\mathbf{\Sigma}_{j'}^* = \mathbf{\Sigma}_{j'}$ for $j' \neq j$. Again, when evaluating the likelihood at the proposed parameter, only the parametric mixture likelihoods $\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_j, \mathbf{\Sigma}_j^*)$, for $i = 1, \ldots, n$, need to be re-evaluated.

Adaptive MH for $\boldsymbol{\mu}_j$. Lines 528-573 of `AT_NWR.m`. Each $\boldsymbol{\mu}_j = (\mu_{j,1}, \ldots, \mu_{j,p}) \in \mathbb{R}^p$, $j = 1, \ldots, J_0$, is updated separately, and no transformation is required. Therefore, the acceptance ratio depends only on the full conditional distribution $Q(\boldsymbol{\mu}_j) = Q(\boldsymbol{\mu}_j|\mathbf{w}, \boldsymbol{\mu}_{-j}, \boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{\theta}, \mathbf{x}, \mathbf{y})$ given by

$$Q(\boldsymbol{\mu}_j) \propto \prod_{k=1}^{p} \exp\left\{-\frac{\tau_{j,k} u_k}{2}(\mu_{j,k} - \mu_{0,k})^2\right\} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\mu}_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \mathbf{\Sigma}_{j'}).$$

Here $w_{j'}(\mathbf{x}_i|\boldsymbol{\mu}_j) = w_{j'}(\mathbf{x}_i)$ denotes the truncated covariate dependent weight in (3), with dependence on $\boldsymbol{\mu}_j$ made explicit, since it is relevant for the calculation of the acceptance ratio. Specifically, we note that $w_{j'}(\mathbf{x}_i|\boldsymbol{\mu}_j)$ will depend on $\boldsymbol{\mu}_j$ for all j' through the normalizing constant, and in the case when $j' = j$ will depend on $\boldsymbol{\mu}_j$ through both the normalizing constant and the kernel in the numerator of the covariate dependent weights. Thus, the acceptance ratio for the proposed move to $\boldsymbol{\mu}_j^*$ from the current value $\boldsymbol{\mu}_j$ is

$$a(\boldsymbol{\mu}_j^*, \boldsymbol{\mu}_j) = \frac{\prod_{k=1}^p \exp\left\{-\frac{\tau_{j,k} u_k}{2}(\mu_{j,k}^* - \mu_{0,k})^2\right\} \prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\mu}_j^*) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})}{\prod_{k=1}^p \exp\left\{-\frac{\tau_{j,k} u_k}{2}(\mu_{j,k} - \mu_{0,k})^2\right\} \prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\mu}_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})},$$

where $\boldsymbol{\mu}_{j'}^* = \boldsymbol{\mu}_{j'}$ for $j' \neq j$. In this case, to efficiently evaluate the likelihood at the proposed parameter, the unnormalized covariate dependent weights $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$, for $i = 1, \dots, n$, need to be re-evaluated by multiplying by the new kernel $\prod_{k=1}^p \mathrm{N}(x_{i,k}|\mu_{j,k}^*, \tau_{j,k}^{-1})$ and dividing by the old kernel $\prod_{k=1}^p \mathrm{N}(x_{i,k}|\mu_{j,k}, \tau_{j,k}^{-1})$, and the normalizing constant of the covariate dependent weights can be efficiently recomputed by subtracting $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j)$ and adding $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$. The unnormalized covariate dependent weights for all other components and all parametric mixture likelihoods can be recycled from the previous step.

Adaptive MH for $\boldsymbol{\tau}_j$. Lines 480-526 of `AT_NWR.m`. Each $\boldsymbol{\tau}_j = (\tau_{j,1}, \dots, \tau_{j,p})$, $j = 1, \dots, J_0$, is updated separately, using a log-transformation $t(\boldsymbol{\tau}_j) = (\log(\tau_{j,1}), \dots, \log(\tau_{j,p})) \in \mathbb{R}^p$, and the determinant of the Jacobian is simply $|\mathcal{J}_t(\boldsymbol{\tau}_j)| = \prod_{k=1}^p \tau_{j,k}^{-1}$. The full conditional distribution $Q(\boldsymbol{\tau}_j) = Q(\boldsymbol{\tau}_j|\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\tau}_{-j}, \boldsymbol{\rho}, \boldsymbol{\theta}, \mathbf{x}, \mathbf{y})$ required for the calculation of the acceptance ratio is given by

$$Q(\boldsymbol{\tau}_j) \propto \prod_{k=1}^p \tau_{j,k}^{\alpha_k - 1/2} \exp\left\{-\tau_{j,k}\left[\gamma_k + \frac{u_k}{2}(\mu_{j,k} - \mu_{0,k})^2\right]\right\} \prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\tau}_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'}),$$

and, once again, the dependence $w_{j'}(\mathbf{x}_i|\boldsymbol{\tau}_j) = w_{j'}(\mathbf{x}_i)$ has been made explicit due to the relevance of this term for the calculation of the acceptance ratio. Thus, the acceptance ratio for the proposed move to $\boldsymbol{\tau}_j^*$ given the current value $\boldsymbol{\tau}_j$ is

$$a(\boldsymbol{\tau}_j^*, \boldsymbol{\tau}_j) = \frac{\prod_{k=1}^p \tau_{j,k}^{*\,\alpha_k + 1/2} \exp\left\{-\tau_{j,k}^*\left[\gamma_k + \frac{u_k}{2}(\mu_{j,k} - \mu_{0,k})^2\right]\right\}}{\prod_{k=1}^p \tau_{j,k}^{\alpha_k + 1/2} \exp\left\{-\tau_{j,k}\left[\gamma_k + \frac{u_k}{2}(\mu_{j,k} - \mu_{0,k})^2\right]\right\}} * \frac{\prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\tau}_j^*) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})}{\prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\boldsymbol{\tau}_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})},$$

where $\boldsymbol{\tau}_{j'}^* = \boldsymbol{\tau}_{j'}$ for $j' \neq j$. Again, when evaluating the likelihood at the proposed parameter, the unnormalized covariate dependent weights $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$, for $i = 1, \dots, n$, need to be re-evaluated by multiplying by the new kernel $\prod_{k=1}^p \mathrm{N}(x_{i,k}|\mu_{j,k}, \tau_{j,k}^{*\,-1})$ and dividing by the old kernel $\prod_{k=1}^{p} \mathrm{N}(x_{i,k}|\mu_{j,k}, \tau_{j,k}^{-1})$, and the normalizing constant of the covariate dependent weights are recomputed by subtracting $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j)$ and adding $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$.

Adaptive MH for $\boldsymbol{\rho}_j$. Lines 649-699 of `AT_NWR.m`. Each $\rho_{j,k}$, $j = 1, \ldots, J_0$, $k = p+1, \ldots, q$, is updated separately, using a logit transformation $t(\rho_{j,k}) = \log(\rho_{j,k}/(1-\rho_{j,k}))$, and the determinant of the Jacobian is simply $|\mathcal{J}_t(\rho_{j,k})| = [\rho_{j,k}(1-\rho_{j,k})]^{-1}$. The full conditional distribution $Q(\rho_{j,k}) = Q(\rho_{j,k}|\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\rho}_{-(j,k)}, \boldsymbol{\theta}, \mathbf{x}, \mathbf{y})$ required for the calculation of the acceptance ratio is given by

$$Q(\rho_{j,k}) \propto \rho_{j,k}^{\varrho_{j,k,1}-1}(1-\rho_{j,k})^{\varrho_{j,k,2}-1} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\rho_{j,k}) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'}),$$

and again, the dependence $w_{j'}(\mathbf{x}_i) = w_{j'}(\mathbf{x}_i|\rho_{j,k})$ becomes relevant for the calculation of the acceptance ratio. Thus, the acceptance ratio for the proposed move to $\rho_{j,k}^*$ given the current value $\rho_{j,k}$ is

$$a(\rho_{j,k}^*, \rho_{j,k}) = \frac{\rho_{j,k}^{*\,\varrho_{j,k,1}}(1-\rho_{j,k}^*)^{\varrho_{j,k,2}} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\rho_{j,k}^*) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})}{\rho_{j,k}^{\varrho_{j,k,1}}(1-\rho_{j,k})^{\varrho_{j,k,2}} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|\rho_{j,k}) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})},$$

where $\boldsymbol{\rho}_{j'}^* = \boldsymbol{\rho}_{j'}$ for $j' \neq j$ and $\rho_{j,k'}^* = \rho_{j,k'}$ for $k' \neq k$. Again, when evaluating the likelihood at the proposed parameter, the unnormalized covariate dependent weights $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$, for $i = 1, \ldots, n$, need to be re-evaluated by multiplying by the new kernel $\mathrm{Bern}(x_{i,k}|\rho_{j,k}^*)$ and dividing by the old kernel $\mathrm{Bern}(x_{i,k}|\rho_{j,k})$, and the normalizing constant of the covariate dependent weights are recomputed by subtracting $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j)$ and adding $w_j g(\mathbf{x}_i|\boldsymbol{\psi}_j^*)$.

Adaptive MH for $\mathbf{w}$. Lines 576-646 of `AT_NWR.m`. The weights $\mathbf{w} = (w_1, \ldots, w_{J_0})$ are not directly updated using the adaptive MH scheme. Rather, they are calculated according to the stick-breaking construction after the associated vector $\mathbf{v} = (v_1, \ldots, v_{J_0})$ has been updated. The adaptive MH scheme is therefore defined for each v_j, $j = 1, \ldots, J_0$, via the logit transformation $t(v_j) = \log(v_j/(1-v_j))$, with $|\mathcal{J}_t(v_j)| = [v_j(1-v_j)]^{-1}$. The full conditional distribution $Q(v_j) = Q(v_j|\mathbf{v}_{-j}, \boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{x}, \mathbf{y})$ required for the calculation of the acceptance ratio is given by

$$Q(v_j) \propto v_j^{\zeta_{j,1}-1}(1-v_j)^{\zeta_{j,2}-1} \prod_{i=1}^{n} \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|v_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'}).$$

Notice that dependence $w_{j'}(\mathbf{x}_i) = w_{j'}(\mathbf{x}_i|v_j)$ holds again for all j' due to the normalizing constant in the definition of the covariate dependent weights, but now, for all $j' \geq j$ this will also depend on v_j through the stick-break construction of w_j in the numerator of the covariate dependent weights. Thus, the acceptance ratio for the proposed move to v_j^* given the current value v_j is

$$a(v_j^*, v_j) = \frac{{v_j^*}^{\zeta_{j,1}}(1-v_j^*)^{\zeta_{j,2}} \prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|v_j^*)\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})}{v_j^{\zeta_{j,1}}(1-v_j)^{\zeta_{j,2}} \prod_{i=1}^n \sum_{j'=1}^{J_0} w_{j'}(\mathbf{x}_i|v_j)\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_{j'}, \boldsymbol{\Sigma}_{j'})},$$

where $v_{j'}^* = v_{j'}$ for $j' \neq j$. In this case, when evaluating the likelihood at the proposed parameter, the new weights can be computed as $w_j^* = w_j v_j^*/v_j$ and $w_{j'}^* = w_{j'}(1-v_j^*)/(1-v_j)$ for $j' > j$; the unnormalized covariate dependent weights $w_{j'}g(\mathbf{x}_i|\boldsymbol{\psi}_{j'}^*)$, for $i = 1, \ldots, n$ and $j' \geq j$, can be re-evaluated by multiplying by $w_{j'}^*/w_{j'}$; and the normalizing constant of the covariate dependent weights are recomputed by subtracting $\sum_{j'=j}^{J_0} w_{j'}g(\mathbf{x}_i|\boldsymbol{\psi}_{j'})$ and adding $\sum_{j'=j}^{J_0} w_{j'}g(\mathbf{x}_i|\boldsymbol{\psi}_{j'}^*)$.

Adaptive MH for y. Lines 316-364 of `AT_NWR.m`. Each latent vector $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,d})$, for $i = 1, \ldots, n$, is updated separately, and the full conditional distribution is

$$Q(\mathbf{y}_i|\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{x}_i, \mathbf{z}_i) \propto \sum_{j=1}^{J_0} w_j(\mathbf{x}|\boldsymbol{\psi}_j)\mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j) \prod_{\ell=1}^{d} \mathbb{1}_{\{z_{i,\ell}\}}(h_{i,\ell}).$$

The terms $h_\ell(\mathbf{y}_i, \mathbf{x}_i) = z_{i,\ell}$ define constrained regions for the latent $\mathbf{y}_i$, such that $y_{i,\ell} \in (l_{i,\ell}, u_{i,\ell})$, where in the general case the bounds $(l_{i,\ell}, u_{i,\ell})$ may depend on $y_{i,\ell'}$ for $\ell' \neq \ell$.

For the adaptive MH update, a logistic transformation $t(\mathbf{y}_i)$ is defined sequentially for $\ell = 1, \ldots, d$, based on the bounds $(l_{i,\ell}, u_{i,\ell})$:

$$t(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1}) = \begin{cases} \log\left(\frac{y_{i,\ell}-l_{i,\ell}}{u_{i,\ell}-y_{i,\ell}}\right) & u_{i,\ell}, l_{i,\ell} \in \mathbb{R} \\ \log(y_{i,\ell} - l_{i,\ell}) & u_{i,\ell} = \infty, l_{i,\ell} \in \mathbb{R} \\ -\log(u_{i,\ell} - y_{i,\ell}) & u_{i,\ell} \in \mathbb{R}, l_{i,\ell} = -\infty \\ y_{i,\ell} & l_{i,\ell} = -\infty, u_{i,\ell} = \infty \end{cases}.$$

From the proposed value $\mathbf{t}^*$ in equation (4), the inverse transformation can be applied to obtain the proposed $\mathbf{y}_i^*$, sequentially for $\ell = 1, \ldots, d$, as

$$y_{i,\ell}^* = \begin{cases} \frac{u_{i,\ell}^* \exp(t_\ell^*) + l_{i,\ell}}{1+\exp(t_\ell^*)} & u_{i,\ell}^*, l_{i,\ell}^* \in \mathbb{R} \\ \exp(t_\ell^*) + l_{i,\ell}^* & u_{i,\ell}^* = \infty, l_{i,\ell}^* \in \mathbb{R} \\ u_{i,\ell}^* - \exp(-t_\ell^*) & u_{i,\ell}^* \in \mathbb{R}, l_{i,\ell}^* = -\infty \\ t_\ell^* & l_{i,\ell}^* = -\infty, u_{i,\ell}^* = \infty \end{cases},$$

where the bounds may also be updated sequentially for $\ell = 1, \ldots, d$, if they depend on $y_{1:(\ell-1)}^*$, e.g. for age at first child in our application. The Jacobian matrix is lower triangular with diagonal elements given by

$$\mathcal{J}_{t,\ell,\ell}(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1}) = \begin{cases} \frac{u_{i,\ell} - l_{i,\ell}}{(y_{i,\ell} - l_{i,\ell})(u_{i,\ell} - y_{i,\ell})} & u_{i,\ell}, l_{i,\ell} \in \mathbb{R} \\ \frac{1}{y_{i,\ell} - l_{i,\ell}} & u_{i,\ell} = \infty, l_{i,\ell} \in \mathbb{R} \\ \frac{1}{u_{i,\ell} - y_{i,\ell}} & u_{i,\ell} \in \mathbb{R}, l_{i,\ell} = -\infty \\ 1 & l_{i,\ell} = -\infty, u_{i,\ell} = \infty \end{cases},$$

for $\ell = 1, \ldots, d$, and the determinant of the Jacobian is simply the product of the diagonal elements, $|\mathcal{J}_t(\mathbf{y}_i)| = \prod_{\ell=1}^d \mathcal{J}_{t,\ell,\ell}(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1})$.

Combining these terms, the acceptance ratio for the proposed move to $\mathbf{y}_i^*$ given the current value $\mathbf{y}_i$ is

$$a(\mathbf{y}_i^*, \mathbf{y}_i) = \frac{\sum_{j=1}^{J_0} w_j(\mathbf{x}|\boldsymbol{\psi}_j) \mathrm{N}_d(\mathbf{y}_i^*|\mathbf{x}_i\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)\, |\mathcal{J}_t(\mathbf{y}_i)|}{\sum_{j=1}^{J_0} w_j(\mathbf{x}|\boldsymbol{\psi}_j) \mathrm{N}_d(\mathbf{y}_i|\mathbf{x}_i\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)\, |\mathcal{J}_t(\mathbf{y}_i^*)|}.$$

In this case, only the parametric kernels $\mathrm{N}_d(\mathbf{y}_i^*|\mathbf{x}_i\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j)$, for $j = 1, \ldots, J_0$, need to be re-evaluated, while the remaining terms can be recycled from the previous step.

3.2 SMC for Adaptive Truncation

The second part of the algorithm concerns the update of the number of components of the mixture J. We follow the approach of Griffin (2016) and use M samples from the fixed truncation MCMC algorithm detailed in the previous section, in order to initialize an SMC sampler, which sequentially increases the number of components J. The algorithm is outlined in Algorithm 1. Each SMC update adds a component to the mixture until a stopping rule, based on a suitable discrepancy measure, is satisfied. In particular, we monitor the effective sample size (ESS) of the particles. As an alternative to the ESS, we also consider a discrepancy based on the conditional effective sample size (CESS), which was proposed by Zhou et al. (2016), in the context of model comparison via SMC. To use a CESS-based stopping rule, simply set `ess_type = 0` in line 826 of `AT_NWR.m`.

- Set $J = J_0$, and the initial values of the particles to $(\mathbf{w}^m_{1:J}, \boldsymbol{\psi}^m_{1:J}, \boldsymbol{\theta}^m_{1:J}, \mathbf{y}^m)$ and the unnnormalised weights $\tilde{\vartheta}^m_{J_0} = 1$ for $m = 1, \ldots, M$.
- `While` $\left[\sum_{j=J-I}^{J-1} \mathbb{1}_{[0,\delta)}\left(D(\mathbf{P}^n_j, \mathbf{P}^n_{j+1})\right)\right] < I$

  [1] Add the $(J+1)$-th additional component:
  sample $(w^m_{J+1}, \boldsymbol{\psi}^m_{J+1}, \boldsymbol{\theta}^m_{J+1})$ from $\boldsymbol{P}_0$, for $m = 1, \ldots, M$;
  compute the unnormalised weights $\tilde{\vartheta}^1_{J+1}, \ldots, \tilde{\vartheta}^M_{J+1}$ as:

  $$\tilde{\vartheta}^m_{J+1} = \tilde{\vartheta}^m_J \prod_{i=1}^{n} \frac{f_{\mathbf{P}^{J+1}_{\mathbf{x}}}\left(\mathbf{y}^m_i | \mathbf{w}^m_{1:J+1}, \boldsymbol{\psi}^m_{1:J+1}, \boldsymbol{\theta}^m_{1:J+1}\right)}{f_{\mathbf{P}^J_{\mathbf{x}}}\left(\mathbf{y}^m_i | \mathbf{w}^m_{1:J}, \boldsymbol{\psi}^m_{1:J}, \boldsymbol{\theta}^m_{1:J}\right)}.$$

  [2] Compute the effective sample size:

  $$\mathrm{ESS}_{J+1} = \frac{\left(\sum_{m=1}^{M} \tilde{\vartheta}^m_{J+1}\right)^2}{\sum_{m=1}^{M} (\tilde{\vartheta}^m_{J+1})^2}.$$

  [3] `if` $\mathrm{ESS}_{J+1} < M/2$:
  Resample the particles according to the weights $\tilde{\boldsymbol{\vartheta}}^{1:M}_{J+1}$;
  Set $\tilde{\boldsymbol{\vartheta}}^{1:M}_{J+1} = 1$;
  Run m^* MCMC updates of $(\mathbf{w}^m_{1:J+1}, \boldsymbol{\psi}^m_{1:J+1}, \boldsymbol{\theta}^m_{1:J+1}, \mathbf{y}^m)$ in parallel across $m = 1, \ldots, M$.

Algorithm 1: A sequential Monte Carlo algorithm for the normalised weight model.

4 Posterior Estimates and Predictions

The adaptive truncation algorithm `AT_NWR.m` produces weighted posterior samples, which can be used to produce various posterior and predictive quantities of interest. Letting J denote the final truncation level, the output of `AT_NWR.m` includes the particles ($\mathbf{w}^m_{1:J}$, $\boldsymbol{\theta}^m_{1:J}$, $\boldsymbol{\psi}^m_{1:J}$, $\mathbf{y}^m_{1:n}$) and unnormalized particle weights $\tilde{\vartheta}^m$, for $m = 1, \ldots, M$ (without loss of generality, we drop the subscript J), with the particles contained in the structured array `particles` and the log weights contained in the vector `logweight`. For $\ell = 1, \ldots, b$, we denote by $\tilde{Z}_\ell$ the (undiscretized) age at event, and for $\ell = b+1, \ldots, d$, we have the binary relation $Z_\ell = \mathbb{1}_{(0,\infty)}(Y_\ell)$. In the following, we describe how to utilize the weighted particles to compute posterior estimates of the (undiscretized) age at events as well as a variety of marginal and conditional predictive quantities, implemented in the functions `PredictMarginal.m`, `PredictCensorSurvival.m`, and `PredictConditional.m`.

First, we consider fitted values for the observed data points. Throughout, we indicate with ϑ^m, for $m = 1, \ldots, M$, the normalized particle weights. The posterior distribution of the ℓ-th undiscretized age at event for the i-th subject can be approximated from the weighted samples, $\tilde{z}^m_{i,\ell} := \exp(y^m_{i,\ell})$, for $m = 1, \ldots, M$. Posterior estimates of the (undiscretized) ages at events may be computed, such as the posterior mean,

$$\mathbb{E}[\tilde{Z}_{i,\ell}|\mathbf{x}, \mathbf{z}] \approx \sum_{m=1}^{M} \vartheta^m \tilde{z}^m_{i,\ell},$$

or the posterior median, approximated from the weighted samples. This may be of particular interest for censored data and useful for visualization.

Next, we consider out-of-sample prediction for a new individual with covariate values $\mathbf{x}_*$. We begin with a variety of marginal quantities that may be computed through the function `PredictMarginal.m`. Specifically, the ouputs include:

1. `zpred`: a matrix of the predictive mean of the (undiscretised) ages at event in the first b columns and the predictive mean of the binary responses in the last $d-b$ columns, with rows corresponding to the different specified covariate values $\mathbf{x}_*$;

2. `ypred`: a matrix of the predictive mean of the latent response $\mathbf{y}_*$, with rows corresponding to the different specified covariate values $\mathbf{x}_*$;

3. `zmedian`: a matrix of the predictive median of the (undiscretised) ages at event, with rows corresponding to the different specified covariate values $\mathbf{x}_*$;

4. `ymedian`: a matrix of the predictive median of latent response associated to the ages at event, with rows corresponding to the different specified covariate values $\mathbf{x}_*$;

5. `fzpred`: a matrix of the predictive density of the (undiscretised) ages at event, evaluated at a grid of (undiscretised) ages, for different covariate values $\mathbf{x}_*$;

6. `fypred`: a matrix of the predictive density of the latent response associated to the (undiscretised) ages at event, evaluated at a grid of latent values, for different covariate values $\mathbf{x}_*$.

First, the marginal predictive mean of the (undiscretised) age at event given $\mathbf{x}_*$ is:

$$\mathbb{E}[\tilde{Z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*] \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \exp\left(\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m + \frac{1}{2}\boldsymbol{\Sigma}_{j,(\ell,\ell)}^m\right), \quad (5)$$

for $\ell = 1,\dots,b$, where $\boldsymbol{\beta}_{j,(\cdot,\ell)}^m$ denotes the ℓ-th column of $\boldsymbol{\beta}$ in component j and particle m; $\boldsymbol{\Sigma}_{j,(\ell,\ell)}^m$ denotes element (ℓ,ℓ) of the matrix $\boldsymbol{\Sigma}$ in component j and particle m. For a binary response, the marginal predictive mean, or equivalently, the predictive probability of success, is computed as:

$$\mathbb{E}[Z_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*] = \mathbb{P}(Z_{*,\ell}=1|\mathbf{x},\mathbf{z},\mathbf{x}_*) = \mathbb{P}(Y_{*,\ell}>0|\mathbf{x},\mathbf{z},\mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \Phi\left(\frac{\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m}{\sqrt{\boldsymbol{\Sigma}_{j,(\ell,\ell)}^m}}\right),$$

for $\ell = b+1,\dots,d$. For the age-at-event variables, the marginal predictive density of $\tilde{Z}_{*,\ell}$ given $\mathbf{x}_*$ is given by:

$$\begin{aligned} f(\tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*) &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) f(\tilde{z}_{*,\ell}|\boldsymbol{\theta}_j^m,\mathbf{x}_*) \\ &= \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \mathrm{logN}(\tilde{z}_{*,\ell}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m, \boldsymbol{\Sigma}_{j,(\ell,\ell)}^m), \end{aligned} \quad (6)$$

for $\tilde{z}_{*,\ell} > 0$, where $\mathrm{logN}(\cdot|\mu,\sigma^2)$ denotes the log-normal density with parameters μ and σ^2. Due to the skewness of the predictive densities in our studies, we suggest the use of the predictive median, i.e. the Bayesian estimate under the absolute error loss, over the predictive mean, i.e. the Bayesian estimate of $\tilde{Z}_{*,\ell}$ under the squared error loss. The former better represents the center of the distribution in the presence of skewness and heavy tails. The marginal predictive median can be computed numerically by evaluating the marginal predictive density (6) on a sufficiently dense grid of $\tilde{z}_{*,\ell}$ values.

It is also possible to compute other quantities of interest; specifically, the function `PredictCensorSurvival.m` computes

1. `censorprob`: a matrix of the predictive probability that the age at event is greater than age at interview, with rows corresponding to the different covariate values $\mathbf{x}_*$;

2. `Szpred`: a matrix of the predictive survival curves of the (undiscretised) ages at event, evaluated at a grid of (undiscretised) ages, for different covariate values $\mathbf{x}_*$.

The marginal predictive survival function of $\tilde{Z}_{*,\ell}$ given $\mathbf{x}_*$ is:

$$\begin{aligned} S(\tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*) &= \mathbb{P}(\tilde{Z}_{*,\ell} > \tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*) = \mathbb{P}(Y_{*,\ell} > \log(\tilde{z}_{*,\ell})|\mathbf{x},\mathbf{z},\mathbf{x}_*) \\ &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \left(1 - \Phi\left(\frac{\log(\tilde{z}_{*,\ell}) - \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell)}}{\sqrt{\boldsymbol{\Sigma}^m_{j,(\ell,\ell)}}}\right)\right). \end{aligned} \tag{7}$$

In this case, the corresponding hazard function $h(\tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*) = f(\tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*)/S(\tilde{z}_{*,\ell}|\mathbf{x},\mathbf{z},\mathbf{x}_*)$ is then available. Another interesting quantity is the predictive probability that the indexed event has not yet occurred for a new individual with $x_{*,1}$ years of age, computed as:

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{*,\ell} \geq (x_{*,1}+1)|\mathbf{x},\mathbf{z},\mathbf{x}_*) &= \mathbb{P}(Y_{*,\ell} > \log(x_{*,1}+1)|\mathbf{x},\mathbf{z},\mathbf{x}_*) \\ &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \left(1 - \Phi\left(\frac{\log(x_{*,1}+1) - \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell)}}{\sqrt{\boldsymbol{\Sigma}^m_{j,(\ell,\ell)}}}\right)\right). \end{aligned} \tag{8}$$

Notice that this is simply the survival function evaluated at $\tilde{z}_{*,\ell} = x_{*,1}+1$. However, when $\tilde{z}_{*,\ell}$ changes in equation (7), we obtain the survival function given, in particular, a fixed $x_{*,1}$. When $x_{*,1}$ changes in equation (8), the conditioning event is also changing, giving place to a different function altogether. This could be interpreted as the predictive probability of censoring of the event for a new sampled individual.

Our model also recovers the joint relationship between responses, which allows inference on conditional properties. Specifically, the function `PredictConditional.m`, for the ℓ-th response conditioned on the ℓ'-th response, provides:

1. `zmean`: a matrix of the conditional mean of the ℓ-th response given different values of the ℓ'-th response and covariate values $\mathbf{x}_*$;

2. `zmedian`: a matrix of the conditional median of the ℓ-th response given different values of the ℓ'-th response and covariate values $\mathbf{x}_*$;

3. `zdens`: a matrix of the conditional density of the ℓ-th response, evaluated at a grid of values, given different values of the ℓ'-th response and covariate values $\mathbf{x}_*$.

When both ℓ and ℓ' index ages at event, the conditional predictive mean is:

$$\mathbb{E}[\tilde{Z}_{*,\ell}|\tilde{z}_{*,\ell'},\mathbf{x},\mathbf{z},\mathbf{x}_*] \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \exp(\mu^m_{j,\ell|\ell'} + \sigma^{2\,m}_{j,\ell|\ell'}/2) \frac{\mathrm{logN}(\tilde{z}_{*,\ell'}|\mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell')}, \boldsymbol{\Sigma}^m_{j,(\ell',\ell')})}{f(\tilde{z}_{*,\ell'}|\mathbf{x},\mathbf{z},\mathbf{x}_*)},$$

where

$$\mu^m_{j,\ell|\ell'} = \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell)} + \boldsymbol{\Sigma}^m_{j,(\ell,\ell')}(\boldsymbol{\Sigma}^m_{j,(\ell',\ell')})^{-1}(\log(\tilde{z}_{*,\ell'}) - \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell')}),$$
$$\sigma^{2\,m}_{j,\ell|\ell'} = \boldsymbol{\Sigma}^m_{j,(\ell,\ell)} - (\boldsymbol{\Sigma}^m_{j,(\ell,\ell')})^2(\boldsymbol{\Sigma}^m_{j,(\ell',\ell')})^{-1},$$

and the density in the denominator is the marginal predictive of equation (6). When ℓ indexes a binary response and ℓ' indexes an age at event response, the conditional predictive mean given $\tilde{z}_{*,\ell'}$ and $\mathbf{x}_*$ is:

$$\mathbb{P}(Z_{*,\ell}=1|\tilde{z}_{*,\ell'},\mathbf{x},\mathbf{z},\mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \Phi\left(\frac{\mu^m_{j,\ell|\ell'}}{\sqrt{\sigma^{2\,m}_{j,\ell|\ell'}}}\right) \frac{\mathrm{logN}(\tilde{z}_{*,\ell'}|\mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell')}, \boldsymbol{\Sigma}^m_{j,(\ell',\ell')})}{f(\tilde{z}_{*,\ell'}|\mathbf{x},\mathbf{z},\mathbf{x}_*)}.$$

When both ℓ and ℓ' index ages at event, the conditional predictive density of $\tilde{Z}_{*,\ell}$ given $\tilde{z}_{*,\ell'}$ and $\mathbf{x}_*$ takes the form:

$$f(\tilde{z}_{*,\ell}|\tilde{z}_{*,\ell'},\mathbf{x},\mathbf{z},\mathbf{x}_*) = \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \mathrm{logN}(\tilde{z}_{*,\ell}|\mu^m_{j,\ell|\ell'}, \sigma^{2\,m}_{j,\ell|\ell'}) \frac{\mathrm{logN}(\tilde{z}_{*,\ell'}|\mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,\ell')}, \boldsymbol{\Sigma}^m_{j,(\ell',\ell')})}{f(\tilde{z}_{*,\ell'}|\mathbf{x},\mathbf{z},\mathbf{x}_*)}. \quad (9)$$

The corresponding predictive medians can be obtained numerically from evaluations of this density on a dense grid of values.

References

A. Gelman, G.O. Roberts, and W.R. Gilks. Efficient Metropolis jumping rules. In J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 5*, pages 599–608. Oxford University Press, 1996.

J.E. Griffin. An adaptive truncation method for inference in Bayesian nonparametric models. *Statistics and Computing*, 26:423–441, 2016.

J.E. Griffin and D.A. Stephens. Advances in Markov chain Monte Carlo. In *Bayesian Theory and Applications*. Oxford University Press, 2013.

G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16:351–367, 2001.

G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7:110–120, 1997.

A. Zellner. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. Goel and A. Zellner, editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier, 1986.

Y. Zhou, A.M. Johansen, and J.A.D. Aston. Toward automatic model comparison: an adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726, 2016.