Description the Simulated Data Set Included in the Supplementary Materials for **Bayesian Semiparametric Multivariate Density Deconvolution**

Abhra Sarkar, Debdeep Pati, Antik Chakraborty, Bani K. Mallick and Raymond J. Carroll

This file provides a description of the simulation design used to generate the simulated data set *Simulated_Data_Set.RData* included in the Supplementary Materials. In the following, we use the same notation as in Section 5 of the main paper.

- The dimension of the variable of interest $\mathbf{X}$ is $d = 4$.
- The sample size is $n = 1000$.
- For each subject, we simulated $m_i = 3$ replicates. Total number of replicates is $N = \sum_{i=1}^{n} m_i = 3000$.
- The true density of $\mathbf{X}$ was chosen to be $f_{\mathbf{X}}(\mathbf{X}) = \sum_{k=1}^{K_{\mathbf{X}}} \pi_{\mathbf{X},k}\, \mathrm{MVN}_p(\mathbf{X}|\boldsymbol{\mu}_{\mathbf{X},k}, \boldsymbol{\Sigma}_{\mathbf{X},k})$ with $p = 4$, $K_{\mathbf{X}} = 3$, $\boldsymbol{\pi}_{\mathbf{X}} = (0.25, 0.50, 0.25)^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathbf{X},1} = (0.8, 6, 4, 5)^{\mathrm{T}}$, $\boldsymbol{\mu}_{\mathbf{X},2} = (2.5, 4, 5, 6)^{\mathrm{T}}$ and $\boldsymbol{\mu}_{\mathbf{X},3} = (6, 4, 2, 4)^{\mathrm{T}}$.
- The true density of $\boldsymbol{\epsilon}$ was chosen to be $f_{\boldsymbol{\epsilon}}^{(2)}(\boldsymbol{\epsilon}) = \sum_{k=1}^{K_{\boldsymbol{\epsilon}}} \pi_{\boldsymbol{\epsilon},k}\, \mathrm{MVN}_p(\boldsymbol{\epsilon}|\boldsymbol{\mu}_{\boldsymbol{\epsilon},k}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon},k})$ with $K_{\boldsymbol{\epsilon}} = 3$, $\boldsymbol{\pi}_{\boldsymbol{\epsilon}} = (0.2, 0.6, 0.2)^{\mathrm{T}}$, $\boldsymbol{\mu}_{\boldsymbol{\epsilon},1} = (-0.3, 0, 0.3, 0)^{\mathrm{T}}$, $\boldsymbol{\mu}_{\boldsymbol{\epsilon},2} = (-0.5, 0.4, 0.5, 0)^{\mathrm{T}}$ and $\boldsymbol{\mu}_{\boldsymbol{\epsilon},3} = -(\pi_{\boldsymbol{\epsilon},1}\boldsymbol{\mu}_{\boldsymbol{\epsilon},1} + \pi_{\boldsymbol{\epsilon},2}\boldsymbol{\mu}_{\boldsymbol{\epsilon},2})/\pi_{\boldsymbol{\epsilon},3}$ so that $\sum_{k=1}^{3} \pi_{\boldsymbol{\epsilon},k}\boldsymbol{\mu}_{\boldsymbol{\epsilon},k} = \mathbf{0}$.
- For the component specific covariance matrices, we set $\boldsymbol{\Sigma}_{\mathbf{X},k} = \mathbf{D}_{\mathbf{X}}\boldsymbol{\Sigma}_{\mathbf{X},0}\mathbf{D}_{\mathbf{X}}$ for each k, where $\mathbf{D}_{\mathbf{X}} = \mathrm{diag}(0.75^{1/2}, \ldots, 0.75^{1/2})$. Similarly, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},k} = \mathbf{D}_{\boldsymbol{\epsilon}}\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},0}\mathbf{D}_{\boldsymbol{\epsilon}}$ for each k, where $\mathbf{D}_{\boldsymbol{\epsilon}} = \mathrm{diag}(0.3^{1/2}, \ldots, 0.3^{1/2})$. For each of $f_{\mathbf{X}}$ and $f_{\boldsymbol{\epsilon}}^{(2)}$, we considered the LF type of covariance structures for $\boldsymbol{\Sigma}_{\mathbf{X},0} = \{(\sigma_{ij}^{\mathbf{X},0})\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},0} = \{(\sigma_{ij}^{\boldsymbol{\epsilon},0})\}$, namely $\boldsymbol{\Sigma}_{\mathbf{X},0} = \boldsymbol{\Lambda}_{\mathbf{X}}\boldsymbol{\Lambda}_{\mathbf{X}} + \boldsymbol{\Omega}_{\mathbf{X}}$, with $\boldsymbol{\Lambda}_{\mathbf{X}} = (0.7, \ldots, 0.7)^{\mathrm{T}}$ and $\boldsymbol{\Omega}_{\mathbf{X}} = \mathrm{diag}(0.51, \ldots, 0.51)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon},0} = \boldsymbol{\Lambda}_{\boldsymbol{\epsilon}}\boldsymbol{\Lambda}_{\boldsymbol{\epsilon}} + \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$, with $\boldsymbol{\Lambda}_{\boldsymbol{\epsilon}} = (0.5, \ldots, 0.5)^{\mathrm{T}}$ and $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}} = \mathrm{diag}(0.75, \ldots, 0.75)$,
- Scale adjustments by multiplication with $\mathbf{D}_{\mathbf{X}}$ and $\mathbf{D}_{\boldsymbol{\epsilon}}$ were done so that the simulated values of each component of $\mathbf{X}$ fall essentially in the range $(-2, 6)$ and the simulated values of all components of $\boldsymbol{\epsilon}$ fall essentially in the range $(-3, 3)$.
- The true variance functions were as $s_{\ell}^2(X) = (1 + X/4)^2$ for each component ℓ.

The simulated data set contains the following variables:

- inds = an N-component vector containing identifying labels for different subjects,
- ws = an $N \times d$ matrix containing simulated true values of $\mathbf{W}_{ij}$.