Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning

Yiye JIANG

Institut de Mathématiques de Bordeaux

Multivariate distributional time series

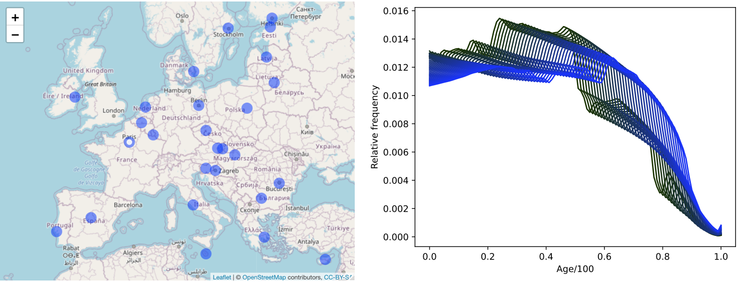
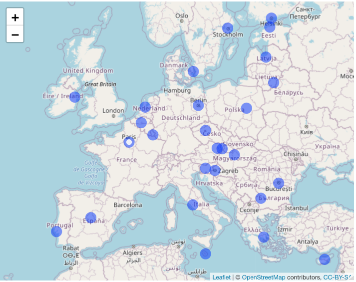


Figure 1: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).*

Multivariate distributional time series



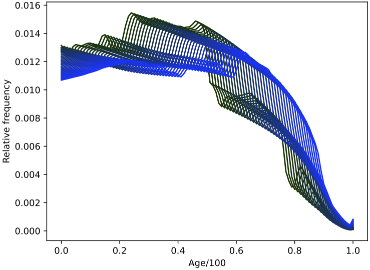


Figure 2: *Observations of the age distributions across European union countries over years* 1995 *to* 2035 *(projected).* On the right are the observations $(\boldsymbol{\mu}_{it})_t \in \mathcal{P}([0,1])$ along time recorded at i = France. Lighter curves correspond to more recent years.

Vector auto-regressive model

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series.
Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise,

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ is the regression operation.

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in IR$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ is the regression operation.

$$\textit{Extension}: \quad \boldsymbol{x}_{it} \in IR \longrightarrow \boldsymbol{\mu}_{it} \in \mathcal{W}_2(IR).$$

Vector auto-regressive model

Let $\boldsymbol{x}_{it} \in I\!R$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$, a multivariate time series. Assume $\mathbb{E}\boldsymbol{x}_{it} = u_i$ exists and time invariant. The VAR model of order 1 writes as

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j) + \boldsymbol{\epsilon}_{it},$$

where $\boldsymbol{\epsilon}_{it}$ is a white noise, and $\sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j)$ is the regression operation.

$$\textit{Extension}: \quad \boldsymbol{x}_{it} \in I\!R \longrightarrow \boldsymbol{\mu}_{it} \in \mathcal{W}_2(I\!R).$$

$$\pi \overset{\mathcal{W}_2(I\!R^d)}{\Longleftrightarrow} T$$

$$\overset{d=1}{\Longleftrightarrow} F_2^{-1} \circ F_1 \overset{\text{fix ref}}{\Longleftrightarrow} \text{functions} \nearrow$$

Related work: Univariate Wasserstein AR model

Chen et al. (2021); Zhang et al. (2021); Zhu and Müller (2021) extended the univariate AR model

$$\boldsymbol{x}_t - u = \alpha(\boldsymbol{x}_{t-1} - u) + \boldsymbol{\epsilon}_t,$$

by interpreting the regression operation from the geometric point of view.

Related work: Univariate Wasserstein AR model

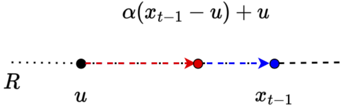


Figure 3: *Geometric interpretation of regression dependency.*

Related work: Univariate Wasserstein AR model

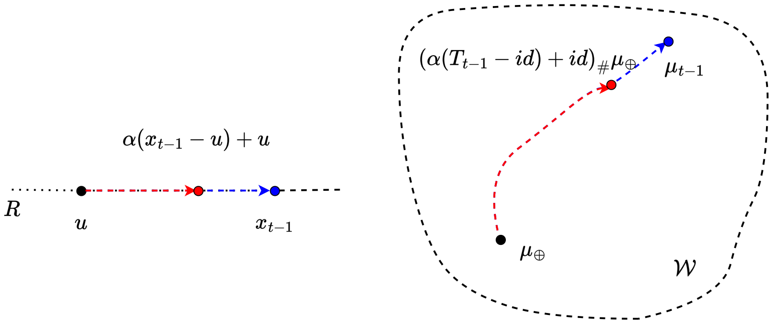


Figure 4: *Geometric interpretation of regression dependency.* $\mu_{\oplus}$ is the time-invariant Fréchet mean of $\boldsymbol{\mu}_t$, T_{t-1} is the optimal transport map which pushforwards $\mu_{\oplus}$ to $\boldsymbol{\mu}_{t-1}$.

Multivariate Wasserstein AR model

Construction of univariate regression operation (ignoring the noise)

$$\boldsymbol{x}_t = u + \alpha(\boldsymbol{x}_{t-1} - u) \Longrightarrow \boldsymbol{\mu}_t = \mathrm{Exp}_{\mu_\oplus}(\alpha(\boldsymbol{T}_{t-1} - id))$$

Multivariate Wasserstein AR model

Construction of univariate regression operation (ignoring the noise)

$$\boldsymbol{x}_t = u + \alpha(\boldsymbol{x}_{t-1} - u) \Longrightarrow \boldsymbol{\mu}_t = \mathrm{Exp}_{\mu_\oplus}(\alpha(\boldsymbol{T}_{t-1} - id))$$

Multivariate regression operation

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j),$$

Multivariate Wasserstein AR model

Construction of univariate regression operation (ignoring the noise)

$$\boldsymbol{x}_t = u + \alpha(\boldsymbol{x}_{t-1} - u) \Longrightarrow \boldsymbol{\mu}_t = \mathrm{Exp}_{\mu_\oplus}\left(\alpha(\boldsymbol{T}_{t-1} - id)\right)$$

Multivariate regression operation

$$\boldsymbol{x}_{it} - u_i = \sum_{j=1}^{N} A_{ij}(\boldsymbol{x}_{j,t-1} - u_j),$$

$\Longleftrightarrow$

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \quad \xrightarrow{\text{ref pt}} \mathbb{E}\tilde{\boldsymbol{x}}_{it} = 0, \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^{N} A_{ij}\tilde{\boldsymbol{x}}_{jt}, \end{cases}$$

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} = ? \xrightarrow{\text{ref pt}} \mathbb{E}_{\oplus} \widetilde{\boldsymbol{\mu}}_{it} = c \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{\boldsymbol{x}}_{jt}, \end{cases}$$

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} =? \xrightarrow{\text{ref pt}} \mathbb{E}_{\oplus} \widetilde{\boldsymbol{\mu}}_{it} = c \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^{N} A_{ij} \tilde{\boldsymbol{x}}_{jt}, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} = \operatorname{Exp}_c \left(\sum_{j=1}^{N} A_{ij} (\widetilde{\boldsymbol{T}}_{i,t-1} - id) \right) \end{cases}$$

where $\widetilde{\boldsymbol{T}}_{i,t-1} = \widetilde{\boldsymbol{F}}_{i,t-1}^{-1} \circ F_c$.

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} =? \xrightarrow{\text{ref pt}} \mathbb{E}_\oplus \widetilde{\boldsymbol{\mu}}_{it} = c \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^N A_{ij} \tilde{\boldsymbol{x}}_{jt}, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} = \text{Exp}_c \left(\sum_{j=1}^N A_{ij} (\widetilde{\boldsymbol{T}}_{i,t-1} - id) \right) \end{cases}$$

where $\widetilde{\boldsymbol{T}}_{i,t-1} = \widetilde{\boldsymbol{F}}_{i,t-1}^{-1} \circ F_c$.

Center a random measure to Lebesgue mean

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{F}_{i,t}^{-1} \ominus F_{i,\oplus}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},$$

where $\boldsymbol{F}_{i,t}^{-1}, F_{i,\oplus}^{-1}$ et $\widetilde{\boldsymbol{F}}_{i,t}^{-1}$ are respectively quantile functions of $\boldsymbol{\mu}_{it}, \mathbb{E}_\oplus \boldsymbol{\mu}_{it}$, and $\widetilde{\boldsymbol{\mu}}_{it}$, all $^{-1}$ are the left continuous inverse.

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} =? \xrightarrow{\text{ref pt}} \mathbb{E}_\oplus \widetilde{\boldsymbol{\mu}}_{it} = c \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^N A_{ij} \tilde{\boldsymbol{x}}_{jt}, \Longrightarrow \widetilde{\boldsymbol{\mu}}_{it} = \text{Exp}_c \left(\sum_{j=1}^N A_{ij} (\widetilde{\boldsymbol{T}}_{i,t-1} - id) \right) \end{cases}$$

where $\widetilde{\boldsymbol{T}}_{i,t-1} = \widetilde{\boldsymbol{F}}_{i,t-1}^{-1} \circ F_c$.

Center a random measure to Lebesgue mean

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{F}_{i,t}^{-1} \ominus F_{i,\oplus}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},$$

where $\boldsymbol{F}_{i,t}^{-1}, F_{i,\oplus}^{-1}$ et $\widetilde{\boldsymbol{F}}_{i,t}^{-1}$ are respectively quantile functions of $\boldsymbol{\mu}_{it}, \mathbb{E}_\oplus \boldsymbol{\mu}_{it}$, and $\widetilde{\boldsymbol{\mu}}_{it}$, all $^{-1}$ are the left continuous inverse. The difference operation $\ominus$ is proposed in Zhu and Müller (2021).

$$\begin{cases} \text{Center} & \tilde{\boldsymbol{x}}_{it} = \boldsymbol{x}_{it} - u_i, \Longrightarrow \tilde{\boldsymbol{\mu}}_{it} =? \xrightarrow{\text{ref pt}} \mathbb{E}_\oplus \tilde{\boldsymbol{\mu}}_{it} = c \\ \text{Push} & \tilde{\boldsymbol{x}}_{it} = \sum_{j=1}^N A_{ij} \tilde{\boldsymbol{x}}_{jt}, \Longrightarrow \tilde{\boldsymbol{\mu}}_{it} = \text{Exp}_c \left(\sum_{j=1}^N A_{ij} (\tilde{\boldsymbol{T}}_{i,t-1} - id) \right) \end{cases}$$

where $\tilde{\boldsymbol{T}}_{i,t-1} = \tilde{\boldsymbol{F}}_{i,t-1}^{-1} \circ F_c$.

Center a random measure to Lebesgue mean

$$\tilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{F}_{i,t}^{-1} \ominus F_{i,\oplus}^{-1} := \boldsymbol{F}_{i,t}^{-1} \circ (F_{i,\oplus}^{-1})^{-1},$$

where $\boldsymbol{F}_{i,t}^{-1}, F_{i,\oplus}^{-1}$ et $\tilde{\boldsymbol{F}}_{i,t}^{-1}$ are respectively quantile functions of $\boldsymbol{\mu}_{it}, \mathbb{E}_\oplus \boldsymbol{\mu}_{it}$, and $\tilde{\boldsymbol{\mu}}_{it}$, all $^{-1}$ are the left continuous inverse. The difference operation $\ominus$ is proposed in Zhu and Müller (2021).

Assumption

All $\boldsymbol{\mu}_{it}$, $t \in \mathbb{Z}$, $i = 1, \ldots, N$ are supported on $[0, 1]$.

Regression operation (igonoring the noise)

$$\widetilde{\boldsymbol{\mu}}_{it} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - id)\right)$$

Regression operation (igonoring the noise)

$$\tilde{\boldsymbol{\mu}}_{it} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{F}}_{i,t-1} - id)\right)$$

Regression operation (igonoring the noise)

$$\widetilde{\boldsymbol{\mu}}_{it} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - id)\right)$$

A tractable in estimation:
$\forall \gamma \; a.c. \in \mathcal{W}$, $\mathrm{Exp}_\gamma |_{\mathrm{Log}_\gamma \mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_\gamma \mathcal{W}$ to $\mathcal{W}$, with the inverse map Log_γ.
$\forall g \in \mathrm{Tan}_\gamma$, $g \in \mathrm{Log}_\gamma \mathcal{W} \iff g + id$ is non-decreasing γ-a.e, Bigot et al. (2017).

Regression operation (igonoring the noise)

$$\tilde{\boldsymbol{\mu}}_{it} = \mathrm{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{F}}_{i,t-1} - id)\right)$$

A tractable in estimation:
$\forall \gamma$ $a.c. \in \mathcal{W}$, $\mathrm{Exp}_\gamma |_{\mathrm{Log}_\gamma \mathcal{W}}$ is an isometric homeomorphism from $\mathrm{Log}_\gamma \mathcal{W}$ to $\mathcal{W}$, with the inverse map Log_γ.
$\forall g \in \mathrm{Tan}_\gamma$, $g \in \mathrm{Log}_\gamma \mathcal{W} \iff g + id$ is non-decreasing γ-a.e, Bigot et al. (2017).

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

Wasserstein multivariate AR Model

$$\widetilde{\boldsymbol{\mu}}_{it} = \epsilon_{it} \# \operatorname{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\widetilde{\boldsymbol{F}}_{i,t-1} - id)\right),$$

where $\{\epsilon_{it}\}_{i,t}$ are i.i.d. random increasing functions, ϵ_{it} is almost surely independent of $\boldsymbol{\mu}_{j,t-1}$, $i,j = 1, \ldots, N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\epsilon_{it}(x)\right] = x,\ x \in [0,1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

Wasserstein multivariate AR Model

$$\tilde{\boldsymbol{\mu}}_{it} = \boldsymbol{\epsilon}_{it} \# \operatorname{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij}(\tilde{\boldsymbol{F}}_{i,t-1} - id)\right),$$

where $\{\boldsymbol{\epsilon}_{it}\}_{i,t}$ are i.i.d. random increasing functions, $\boldsymbol{\epsilon}_{it}$ is almost surely independent of $\boldsymbol{\mu}_{j,t-1}$, $i, j = 1, \ldots, N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\boldsymbol{\epsilon}_{it}(x)\right] = x,\ x \in [0,1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ *and* $0 \leqslant A_{ij} \leqslant 1$.

Quantile function representation

$$\tilde{\boldsymbol{F}}_{i,t}^{-1} = \boldsymbol{\epsilon}_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij}\left(\tilde{\boldsymbol{F}}_{j,t-1}^{-1} - id\right) + id\right],$$

Wasserstein multivariate AR Model

$$\tilde{\mu}_{it} = \epsilon_{it} \# \operatorname{Exp}_{Leb}\left(\sum_{j=1}^{N} A_{ij} (\tilde{F}_{i,t-1} - id) \right),$$

where $\{\epsilon_{it}\}_{i,t}$ are i.i.d. random increasing functions, ϵ_{it} is almost surely independent of $\mu_{j,t-1}$, $i,j = 1,\dots,N$, for all $t \in \mathbb{Z}$, and

$$\mathbb{E}\left[\epsilon_{it}(x)\right] = x,\ x \in [0,1].$$

Assumption

$\sum_{j=1}^{N} A_{ij} \leqslant 1$ and $0 \leqslant A_{ij} \leqslant 1$.

Quantile function representation

$$\tilde{F}^{-1}_{i,t} = \epsilon_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\tilde{F}^{-1}_{j,t-1} - id \right) + id \right], \quad A \iff \mathcal{G}$$

Iterated random function (IRF) system

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) + id \right], \qquad (1)$$

Admissible as a TS model: existence, uniqueness and stationarity (additionally Hilbert space).

Iterated random function (IRF) system

$$\widetilde{\boldsymbol{F}}_{i,t}^{-1} = \epsilon_{i,t} \circ \left[\sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) + id \right],$$

Admissible as a TS model: existence, uniqueness and stationarity (additionally Hilbert space).

Consider the product metric space

$$(\mathcal{X}\,,\, d) := (\mathcal{T}\,,\, \|\cdot\|_{Leb})^{\otimes N}\,,$$

where $\mathcal{T} := \mathrm{Log}_{Leb}\,\mathcal{W} + id$ is the space of all quantile functions of $\mathcal{W}$, equipped with the norm $\|\cdot\|_{Leb}$ in the tangent space at the Lebesgue measure. Thus, for any $X = (X_i)_{i=1}^N, Y = (Y_i)_{i=1}^N \in \mathcal{X}$

$$d(X, Y) := \sqrt{\sum_{i=1}^{N} \|X_i - Y_i\|_{Leb}^2}.$$

By Wu and Shao (2004), IRF system in a complete, separable metric space

$$\text{exp decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \xrightarrow{\text{add str}} \text{stationarity.}$$

By Wu and Shao (2004), IRF system in a complete, separable metric space

$$\text{exp decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \xrightarrow{\text{add str}} \text{stationarity.}$$

Assumption

Contraction of the regression operator (at exp decay rate)

1. $\mathbb{E}\left[\epsilon_{i,t}(x) - \epsilon_{i,t}(y)\right]^2 \leqslant L^2(x-y)^2,\ \forall x, y \in [0,1],\ t \in \mathbb{Z},\ i = 1, \ldots, N,$
2. $\|A\|_2 < \frac{1}{L}$.

By Wu and Shao (2004), IRF system in a complete, separable metric space

$$\text{exp decay rate} \rightarrow \text{stability} \rightarrow \text{existence} \xrightarrow{\text{add str}} \text{stationarity}.$$

Assumption

Contraction of the regression operator (at exp decay rate)

1. $\mathbb{E}\left[\epsilon_{i,t}(x) - \epsilon_{i,t}(y)\right]^2 \leqslant L^2(x-y)^2, \ \forall x, y \in [0,1], \ t \in \mathbb{Z}, \ i = 1, \ldots, N,$
2. $\|A\|_2 < \frac{1}{L}$.

Theorem

Under Assumptions N-simplex and contraction, the IRF system (1) *almost surely admits a solution* $X_t, \ t \in \mathbb{Z}$, *with* $X_t \overset{d}{=} \pi, \ \forall \ t \in \mathbb{Z}$. *Moreover, if there exists another solution* $\boldsymbol{S}_t, \ t \in \mathbb{Z}$, *then for all* $t \in \mathbb{Z}$

$$X_t \overset{d}{=} \boldsymbol{S}_t, \quad \textit{almost surely}.$$

(X, d) with d the induced metric of inner product

$$\langle X, Y \rangle = \sum_{i=1}^{N} \langle X_i, Y_i \rangle_{Leb} \,.$$

(X, d) with d the induced metric of inner product

$$\langle X, Y \rangle = \sum_{i=1}^{N} \langle X_i, Y_i \rangle_{Leb} .$$

A random process $\{V_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is said to be stationary if the following properties are satisfied.

1. $\mathbb{E} \|V_t\|^2 < \infty$
2. The Hilbert mean $U := \mathbb{E}[V_t]$ does not depend on t.
3. The auto-covariance operators defined as

$$\mathcal{G}_{t,t-h}(V) := \mathbb{E} \langle V_t - U, V \rangle (V_{t-h} - U) , \quad V \in \mathcal{H},$$

do not depend on t, that is $\mathcal{G}_{t,t-h}(V) = \mathcal{G}_{0,-h}(V)$ for all t.

(X, d) with d the induced metric of inner product

$$\langle X, Y \rangle = \sum_{i=1}^{N} \langle X_i, Y_i \rangle_{Leb} .$$

A random process $\{V_t\}_t$ in a separable Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is said to be stationary if the following properties are satisfied.

1. $\mathbb{E} \| V_t \|^2 < \infty$
2. The Hilbert mean $U := \mathbb{E}[V_t]$ does not depend on t.
3. The auto-covariance operators defined as

$$\mathcal{G}_{t,t-h}(V) := \mathbb{E} \langle V_t - U, V \rangle (V_{t-h} - U) , \quad V \in \mathcal{H},$$

   do not depend on t, that is $\mathcal{G}_{t,t-h}(V) = \mathcal{G}_{0,-h}(V)$ for all t.

Theorem

The unique solution given in Theorem 1 is stationary as a random process in $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ in the sense of Definition above.

Constrained least-square estimation

$$\widetilde{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) - id \right\|_{Leb}^2 ,$$

where B_+^1 is the constraint set of N-simplex.

Constrained least-square estimation

$$\widetilde{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_+^1}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widetilde{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widetilde{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) - id \right\|_{Leb}^{2},$$

where B_+^1 is the constraint set of N-simplex.

Population mean is also an unknown parameter, we estimate as

$$\boldsymbol{F}_{\bar{\mu}_i}^{-1} = \frac{1}{T} \sum_{t=1}^{T} F_{\boldsymbol{\mu}_{i,t}}^{-1},$$

and center $\boldsymbol{\mu}_{i,t}$ by $\boldsymbol{F}_{\bar{\mu}_i}^{-1}$ with difference $\ominus$

$$\widehat{\boldsymbol{F}}_{i,t}^{-1} := \boldsymbol{F}_{i,t}^{-1} \ominus \boldsymbol{F}_{\bar{\mu}_i}^{-1} = \boldsymbol{F}_{i,t}^{-1} \circ \boldsymbol{F}_{\bar{\mu}_i}.$$

Constrained least-square estimation

$$\widehat{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B_{+}^{1}}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widehat{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) - id \right\|_{Leb}^{2}, \quad (1)$$

The optimization problem (1) can be solved by the accelerated projected gradient descent (Parikh and Boyd, 2014, Chapter 4.3). The projection onto B_{+}^{1} is given in Thai et al. (2015).

Constrained least-square estimation

$$\widehat{\boldsymbol{A}}_{i:} = \underset{A_{i:} \in B^1_+}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \left\| \widehat{\boldsymbol{F}}_{i,t}^{-1} - \sum_{j=1}^{N} A_{ij} \left(\widehat{\boldsymbol{F}}_{j,t-1}^{-1} - id \right) - id \right\|_{Leb}^2 ,$$

The optimization problem (1) can be solved by the accelerated projected gradient descent (Parikh and Boyd, 2014, Chapter 4.3). The projection onto B^1_+ is given in Thai et al. (2015).

Note that the N-simplex constraint promotes the sparsity in $\widehat{\boldsymbol{A}}$.

Constrained least-square estimation

Theorem

Assume[a] *the transformed sequence* $\widetilde{\boldsymbol{F}}_t^{-1}$, $t = 0, 1, \ldots, T$ *checks Model* (1) *with Assumption N-simplex true. Suppose additionally* $\widetilde{\boldsymbol{F}}_0^{-1} \overset{d}{=} \pi$ *with* π *the stationary distribution defined in Theorem 1. Given Assumption contraction of regression operation holds true. Then given the true coefficient A satisfies Assumption N-simplex, we have*

$$\widehat{\boldsymbol{A}} - A \xrightarrow{p} 0.$$

[a] Complete statement of theorem sees Jiang (2022)

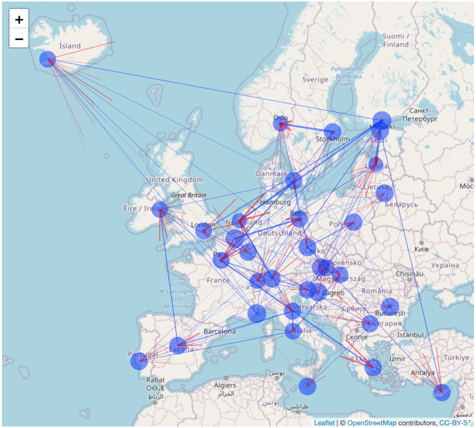


Figure 5: *Inferred age structure graph.* The non-zero coefficients A_{ij} are represented by the weighted directed edges from node j to node i. Thicker arrow corresponds to larger weights. The blue circles around nodes represent the weights of self-loop.

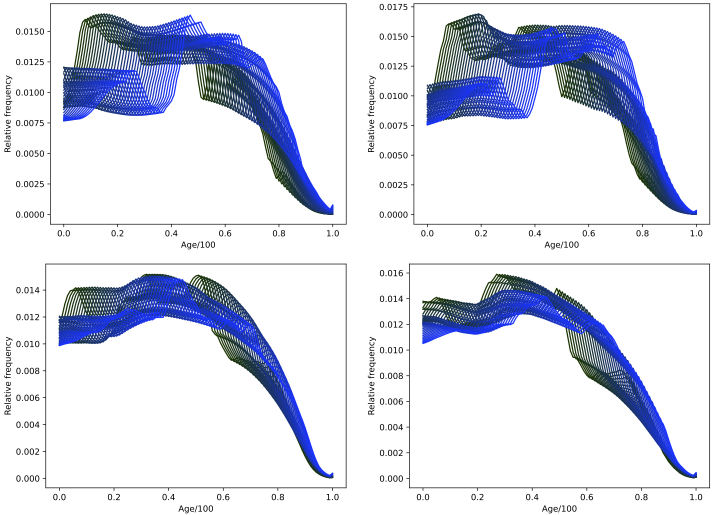


Figure 6: *Evolution of age structure from* 1996 *to* 2036 *(projected). Estonia (top left), Latvia(top right), Sweden (bottom left) versus Norway (bottom right).*

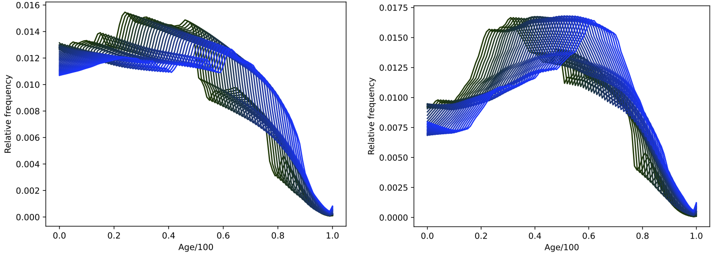


Figure 7: *Evolution of age structure from* 1996 *to* 2036 *(projected) of France (left) versus Italy (right).*

	From	To
1	Estonia	Latvia
2	Sweden	Norway
3	Belgium	Germany
4	Finland	Netherlands
5	France	Greece

Table 1: *Top* 5 *edges with the largest weights excluding all the self-loops, based on the data from* 1996 *to* 2036 *(projected).*

L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

J. Bigot, R. Gouet, T. Klein, and A. López. Geodesic pca in the wasserstein space by convex pca. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):1–26, 2017.

Y. Chen, Z. Lin, and H.-G. Müller. Wasserstein regression. *Journal of the American Statistical Association*, pages 1–14, 2021.

Y. Jiang. Wasserstein multivariate auto-regressive models for modeling distributional time series and its application in graph learning. *stat*, 1050:12, 2022.

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

J. Thai, C. Wu, A. Pozdnukhov, and A. Bayen. Projected sub-gradient with ℓ_1 or simplex constraints via isotonic regression. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 2031–2036. IEEE, 2015.

W. B. Wu and X. Shao. Limit theorems for iterated random functions. *Journal of Applied Probability*, 41(2):425–436, 2004.

C. Zhang, P. Kokoszka, and A. Petersen. Wasserstein autoregressive models for density time series. *Journal of Time Series Analysis*, 2021.

C. Zhu and H.-G. Müller. Autoregressive optimal transport models. *arXiv preprint arXiv:2105.05439*, 2021.

The space $\mathcal{W} := \mathcal{W}_2(IR)$ has a pseudo-Riemannian structure (Ambrosio et al., 2008). Let $\gamma \in \mathcal{W}$ be an absolutely continuous measure, the tangent space at γ is defined as

$$\mathrm{Tan}_\gamma = \overline{\{t(F_\mu^{-1} \circ F_\gamma - id) : \mu \in \mathcal{W},\ t > 0\}}^{\mathcal{L}^2_\gamma(IR)},$$

Definition

The exponential map $\mathrm{Exp}_\gamma : \mathrm{Tan}_\gamma \to \mathcal{W}$ is defined as

$$\mathrm{Exp}_\gamma\, g = (g + id)\#\gamma.$$

Definition

The logarithmic map $\mathrm{Log}_\gamma : \mathcal{W} \to \mathrm{Tan}_\gamma$ is defined as

$$\mathrm{Log}_\gamma\, \mu = F_\mu^{-1} \circ F_\gamma - id.$$

Related work: Univariate Wasserstein AR model

Describe this regression relationship with

AR model of optimal transport (Zhu and Müller, 2021):

$$\boldsymbol{T}_{t+1} = \boldsymbol{\epsilon}_t \circ (\alpha(\boldsymbol{T}_t - id) + id), \quad 0 < \alpha < 1$$

AR model of tangent vector (Zhang et al., 2021):

$$\boldsymbol{T}_{t+1} - id = \alpha(\boldsymbol{T}_t - id) + \boldsymbol{\epsilon}_t, \qquad 0 < |\alpha| < 1,$$

Tangent vector with regression operator (Chen et al., 2021)

$$\boldsymbol{T}_{t+1} - id = \Gamma(\boldsymbol{T}_t - id) + \boldsymbol{\epsilon}_t, \quad \Gamma : Log_{\mu_\oplus}(\mathcal{W}) \to Log_{\mu_\oplus}(\mathcal{W})$$

the model in tangent space than is the ordinary AR model for functional TS in Hilbert space, expect the log image issue