A weakly informative prior for resonance frequencies

Marnix Van Soom & Bart de Boer

MaxEnt 2021

VRIJE UNIVERSITEIT BRUSSEL

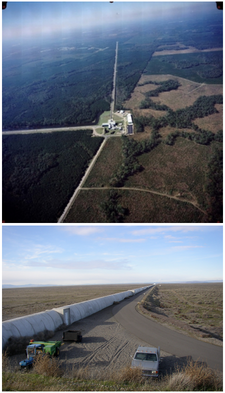
Resonance frequencies in the natural sciences

Resonance frequencies in the natural sciences

[Wilson+ 2014]

Resonance frequencies in the natural sciences

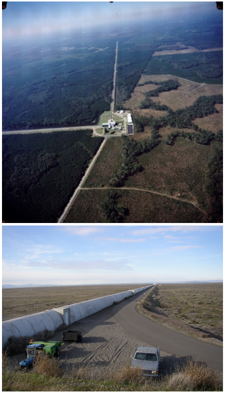
[Wilson+ 2014]



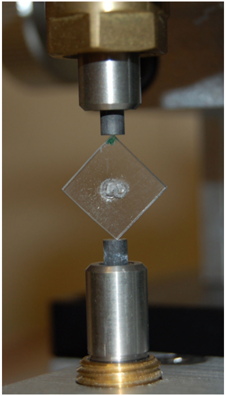
[Littenberg and Cornish 2015]

Resonance frequencies in the natural sciences

[Wilson+ 2014]



[Littenberg and Cornish 2015]



[Xu+ 2019]

Measurement model

Measurement model

$$p(D, \boldsymbol{x})$$

Measurement model

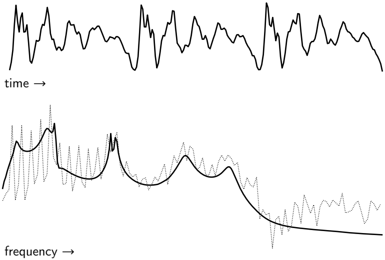
$$p(D, \boldsymbol{x})$$

$D = \text{data}$

Measurement model

$$p(D, \boldsymbol{x})$$

$$D = \text{data}$$
$$\boldsymbol{x} = \{x_1 \cdots x_K\}$$
$$= K \text{ resonance frequencies}$$



Measurement model

$$p(D, \boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\ \pi(\boldsymbol{x})$$

Measurement model

$$p(D, \boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\ \pi(\boldsymbol{x})$$

- Assume $\mathscr{L}(\boldsymbol{x})$ is given

Measurement model

$$p(D, \boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\ \pi(\boldsymbol{x})$$

- Assume $\mathscr{L}(\boldsymbol{x})$ is given
- **CHOOSE $\pi(\boldsymbol{x})$ SUBJECT TO LIMITED PRIOR INFORMATION**

Measurement model

$$p(D, \boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\ \pi(\boldsymbol{x})$$

- Assume $\mathscr{L}(\boldsymbol{x})$ is given
- **CHOOSE $\pi(\boldsymbol{x})$ SUBJECT TO LIMITED PRIOR INFORMATION**
  - Don't know prior estimates $\hat{\boldsymbol{x}}$

Measurement model

$$p(D,\boldsymbol{x}) = \mathscr{L}(\boldsymbol{x})\ \pi(\boldsymbol{x})$$

- Assume $\mathscr{L}(\boldsymbol{x})$ is given
- CHOOSE $\pi(\boldsymbol{x})$ SUBJECT TO LIMITED PRIOR INFORMATION
  - Don't know prior estimates $\hat{\boldsymbol{x}}$
  - Don't know K

$$Z(K) = \int \mathrm{d}^K\boldsymbol{x}\,\mathscr{L}(\boldsymbol{x})\pi(\boldsymbol{x})$$

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

- Most common choice in literature

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

- Most common choice in literature
- K iid $x_k \in [a,b]$ with *global* bounds

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

- Most common choice in literature
- K iid $x_k \in [a,b]$ with *global* bounds
- **Label switching problem** ↯

First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

- Most common choice in literature
- K iid $x_k \in [a,b]$ with *global* bounds
- **Label switching problem** ↯
  - Exchange symmetry

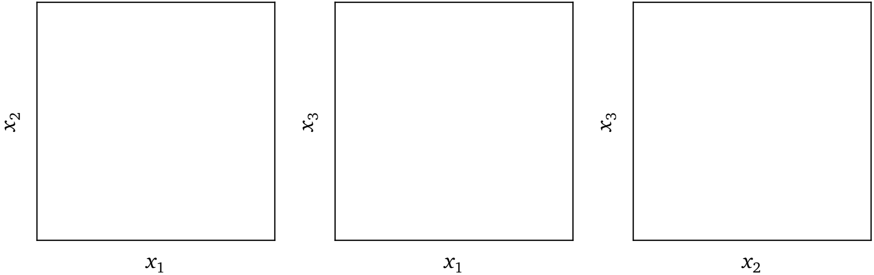
First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$

- Most common choice in literature
- K iid $x_k \in [a,b]$ with *global* bounds
- **Label switching problem** ↯
  - Exchange symmetry
  - Frustrates calculating $Z(K)$ [Celeux+ 2018]

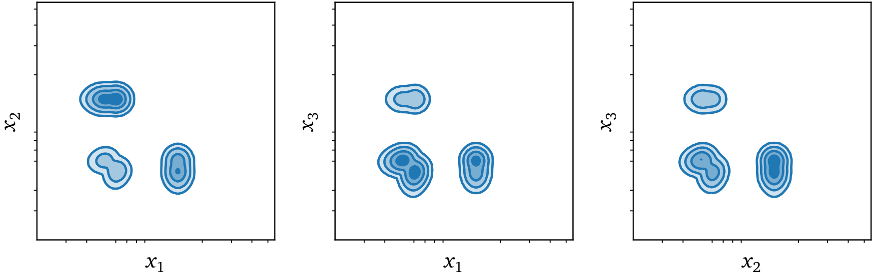
First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$



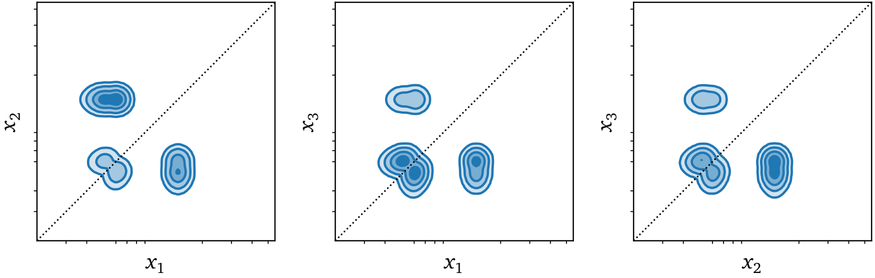
First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$



First candidate: $\pi_1(\boldsymbol{x}|\cdot)$

$$\pi_1(\boldsymbol{x}|a,b) = \prod_{k=1}^{K} h(x_k|a,b)$$



Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

$$\pi_2(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b}) = \prod_{k=1}^{K} h(x_k|a_k, b_k)$$

Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

$$\pi_2(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b}) = \prod_{k=1}^{K} h(x_k|a_k, b_k)$$

- K independent $x_k \in [a_k, b_k]$ with *local* bounds $(\boldsymbol{a}, \boldsymbol{b})$

Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

$$\pi_2(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b}) = \prod_{k=1}^{K} h(x_k|a_k,b_k)$$

- K independent $x_k \in [a_k, b_k]$ with *local* bounds $(\boldsymbol{a},\boldsymbol{b})$
- **Multiplet problem** $\lightning$

Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

$$\pi_2(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b}) = \prod_{k=1}^{K} h(x_k|a_k,b_k)$$

- K independent $x_k \in [a_k, b_k]$ with *local* bounds $(\boldsymbol{a},\boldsymbol{b})$
- **Multiplet problem** ↯
  - Need overlap for close frequencies

Second candidate: $\pi_2(\boldsymbol{x}|\cdot)$

$$\pi_2(\boldsymbol{x}|\boldsymbol{a},\boldsymbol{b}) = \prod_{k=1}^{K} h(x_k|a_k,b_k)$$

- K independent $x_k \in [a_k, b_k]$ with *local* bounds $(\boldsymbol{a},\boldsymbol{b})$
- **Multiplet problem** $\lightning$
  - Need overlap for close frequencies
  - But this brings back the label switching problem

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k)$$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k)$$

- **Maximum entropy distribution**

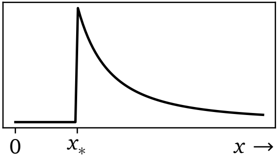
Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k)$$

- **Maximum entropy distribution**

- Chain of coupled *Pareto distributions*

$$\mathsf{Pareto}(x|x_*, \lambda) = \frac{\lambda}{x}\left(\frac{x_*}{x}\right)^{\lambda} \qquad (x_* \leq x)$$



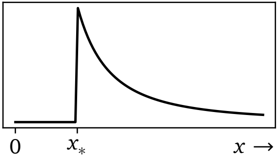
Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$

- **Maximum entropy distribution**

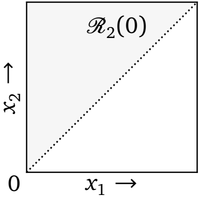
- Chain of coupled *Pareto distributions*

$$\mathsf{Pareto}(x|x_*, \lambda) = \frac{\lambda}{x}\left(\frac{x_*}{x}\right)^{\lambda} \qquad (x_* \leq x)$$



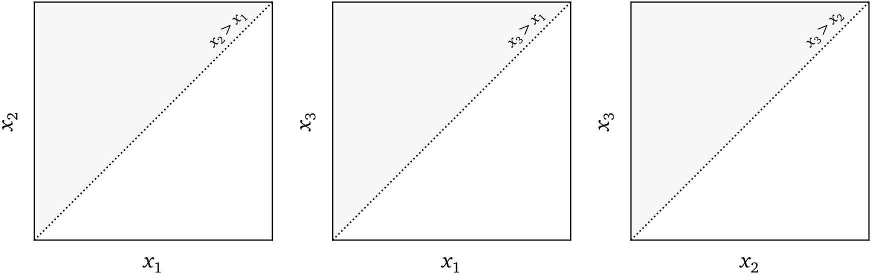
- Supported by the *ordered region*

$$\mathscr{R}_K(x_0) = \{\boldsymbol{x}|x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_K\}$$



Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

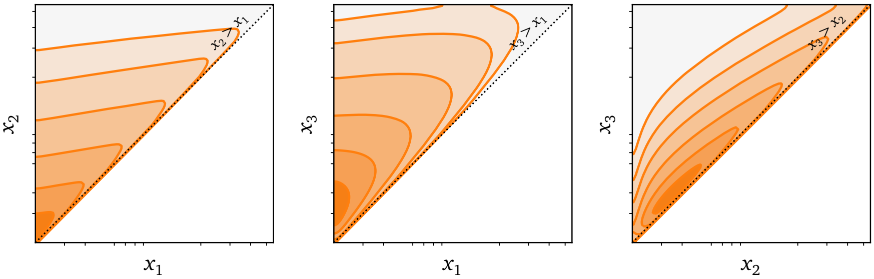
$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



Support $\mathscr{R}_3(x_0) = \{\boldsymbol{x}|x_0 \leq x_1 \leq x_2 \leq x_3\}$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

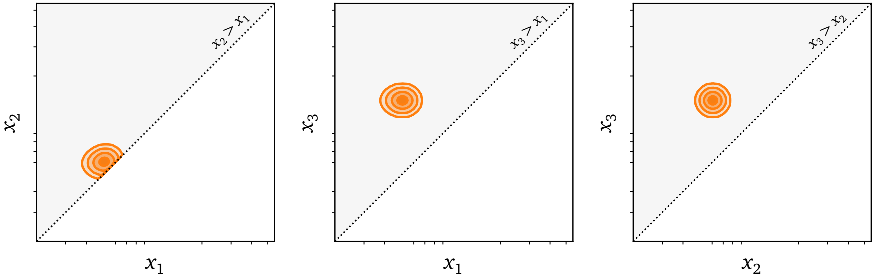
$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \text{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



Pairwise prior distribution $\pi_3(\boldsymbol{x}|\boldsymbol{\lambda})$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

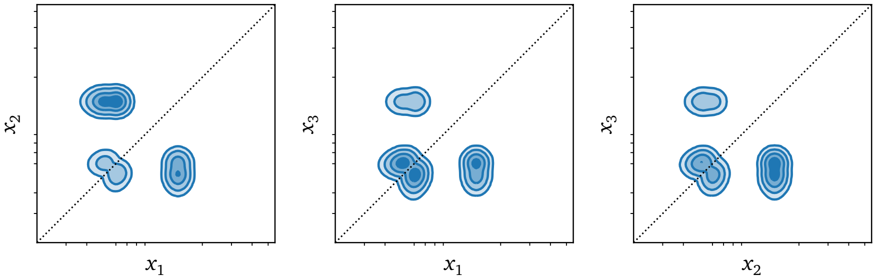
$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathrm{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



Pairwise posterior distribution $P_3(\boldsymbol{x}|\boldsymbol{\lambda}) \propto \mathscr{L}(\boldsymbol{x})\pi_3(\boldsymbol{x}|\boldsymbol{\lambda})$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

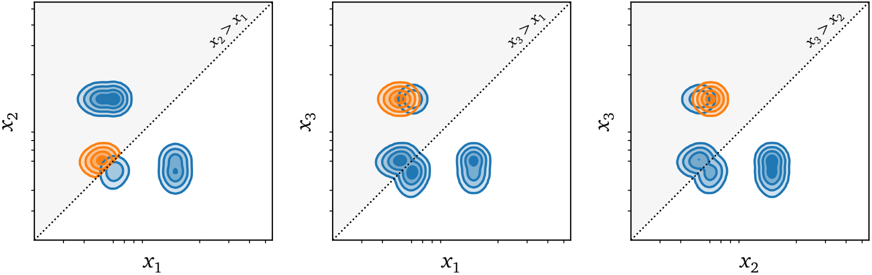
$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \text{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



Pairwise posterior distribution $P_1(\boldsymbol{x}|a,b) \propto \mathscr{L}(\boldsymbol{x})\pi_1(\boldsymbol{x}|a,b)$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

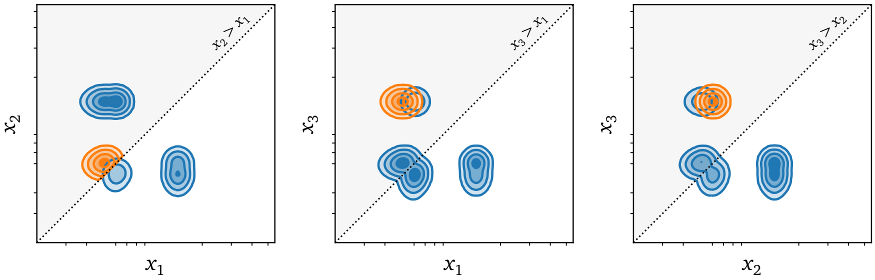
$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1},\lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



$P_1(\boldsymbol{x}|a,b)$ vs. $P_3(\boldsymbol{x}|\boldsymbol{\lambda})$

Third candidate: $\pi_3(\boldsymbol{x}|\cdot)$

$$\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \text{Pareto}(x_k|x_{k-1}, \lambda_k) \qquad \boldsymbol{x} \in \mathscr{R}_K(x_0)$$



$P_1(\boldsymbol{x}|a, b)$ vs. $P_3(\boldsymbol{x}|\boldsymbol{\lambda})$

~~Label switching problem~~ ✔ ~~Multiplet problem~~ ✔

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

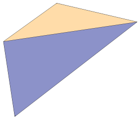
Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Ansatz: Jeffreys prior $m(x_k) \propto 1/x_k$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Ansatz: Jeffreys prior $m(x_k) \propto 1/x_k$

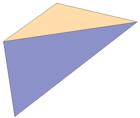
$\mathscr{R}_3(x_0)$



- $m(\boldsymbol{x}) = \begin{cases} \prod_{k=1}^{K} \frac{1}{x_k} & \boldsymbol{x} \in \mathscr{R}_K(x_0) \\ 0 & \text{else} \end{cases}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Ansatz: Jeffreys prior $m(x_k) \propto 1/x_k$



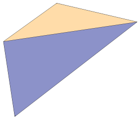
- $m(\boldsymbol{x}) = \begin{cases} \prod_{k=1}^{K} \frac{1}{x_k} & \boldsymbol{x} \in \mathscr{R}_K(x_0) \\ 0 & \text{else} \end{cases}$

- $\boldsymbol{x} \to \boldsymbol{u}$: $\begin{cases} u_1 = \log(x_1/x_0) \\ u_2 = \log(x_2/x_1) \\ \quad \dots \\ u_K = \log(x_K/x_{K-1}) \end{cases}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Ansatz: Jeffreys prior $m(x_k) \propto 1/x_k$

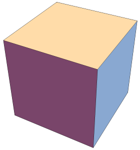
$\mathscr{R}_3(x_0)$



- $m(\boldsymbol{x}) = \begin{cases} \prod_{k=1}^{K} \frac{1}{x_k} & \boldsymbol{x} \in \mathscr{R}_K(x_0) \\ 0 & \text{else} \end{cases}$

- $\boldsymbol{x} \to \boldsymbol{u}$: $\begin{cases} u_1 = \log(x_1/x_0) \\ u_2 = \log(x_2/x_1) \\ \quad \dots \\ u_K = \log(x_K/x_{K-1}) \end{cases}$

$\mathscr{R}_3(x_0)$ in $\boldsymbol{u}$ space



- $m(\boldsymbol{u}) = m(\boldsymbol{x}(\boldsymbol{u}))\left|\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{u}}\right| = \begin{cases} 1 & \boldsymbol{u} \geq 0 \\ 0 & \text{else} \end{cases}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Normalize $m(\boldsymbol{u}) = \begin{cases} 1 & \boldsymbol{u} \geq 0 \\ 0 & \text{else} \end{cases}$ by constraining first moments $\langle \boldsymbol{u} \rangle := \overline{\boldsymbol{u}}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Normalize $m(\boldsymbol{u}) = \begin{cases} 1 & \boldsymbol{u} \geq 0 \\ 0 & \text{else} \end{cases}$ by constraining first moments $\langle \boldsymbol{u} \rangle := \overline{\boldsymbol{u}}$

- Minimize

$$D_{\mathrm{KL}}(\pi_3|m) = \int \mathrm{d}^K\boldsymbol{u}\, \pi_3(\boldsymbol{u}) \log \frac{\pi_3(\boldsymbol{u})}{m(\boldsymbol{u})}$$

  subject to

$$\langle \boldsymbol{u} \rangle \equiv \int \mathrm{d}^K\boldsymbol{u}\, \boldsymbol{u} \pi_3(\boldsymbol{u}) := \overline{\boldsymbol{u}}$$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Normalize $m(\boldsymbol{u}) = \begin{cases} 1 & \boldsymbol{u} \geq 0 \\ 0 & \text{else} \end{cases}$ by constraining first moments $\langle \boldsymbol{u} \rangle := \overline{\boldsymbol{u}}$

- Minimize

$$D_{\mathrm{KL}}(\pi_3|m) = \int \mathrm{d}^K\boldsymbol{u}\, \pi_3(\boldsymbol{u}) \log \frac{\pi_3(\boldsymbol{u})}{m(\boldsymbol{u})}$$

  subject to

$$\langle \boldsymbol{u} \rangle \equiv \int \mathrm{d}^K\boldsymbol{u}\, \boldsymbol{u}\pi_3(\boldsymbol{u}) := \overline{\boldsymbol{u}}$$

- Solution:

$$\pi_3(\boldsymbol{u}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Exp}(u_k|\lambda_k) \qquad (\boldsymbol{u} \geq 0)$$

  where the rates $\lambda_k = 1/\overline{u_k}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Normalize $m(\boldsymbol{u}) = \begin{cases} 1 & \boldsymbol{u} \geq 0 \\ 0 & \text{else} \end{cases}$ by constraining first moments $\langle \boldsymbol{u} \rangle := \overline{\boldsymbol{u}}$

- Minimize

$$D_{\mathrm{KL}}(\pi_3|m) = \int \mathrm{d}^K\boldsymbol{u}\, \pi_3(\boldsymbol{u}) \log \frac{\pi_3(\boldsymbol{u})}{m(\boldsymbol{u})}$$

  subject to

$$\langle \boldsymbol{u} \rangle \equiv \int \mathrm{d}^K\boldsymbol{u}\, \boldsymbol{u}\pi_3(\boldsymbol{u}) := \overline{\boldsymbol{u}}$$

- Solution:

$$\pi_3(\boldsymbol{u}|\boldsymbol{\lambda}) = \prod_{k=1}^{K} \mathsf{Exp}(u_k|\lambda_k) \qquad (\boldsymbol{u} \geq 0)$$

  where the rates $\lambda_k = 1/\overline{u_k}$

- Equivalent to Jaynes' *principle of maximum entropy* with $m(\boldsymbol{u})$ serving as the invariant measure [Jaynes 1968]

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

- $\boldsymbol{u} \to \boldsymbol{x}$:
$$\begin{cases} x_1 = x_0 \exp\{u_1\} \\ x_2 = x_0 \exp\{u_1 + u_2\} \\ \quad \cdots \\ x_K = x_0 \exp\{u_1 + u_2 + \cdots + u_K\} \end{cases}$$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

- $\boldsymbol{u} \to \boldsymbol{x}$: $$\begin{cases} x_1 = x_0 \exp\{u_1\} \\ x_2 = x_0 \exp\{u_1 + u_2\} \\ \quad \cdots \\ x_K = x_0 \exp\{u_1 + u_2 + \cdots + u_K\} \end{cases}$$
- $\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \pi_3(\boldsymbol{u}(\boldsymbol{x})|\boldsymbol{\lambda}) \left| \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\boldsymbol{x}} \right| = \prod_{k=1}^{K} \mathsf{Pareto}(x_k | x_{k-1}, \lambda_k)$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

- $\boldsymbol{u} \to \boldsymbol{x}$: $$\begin{cases} x_1 = x_0 \exp\{u_1\} \\ x_2 = x_0 \exp\{u_1 + u_2\} \\ \quad \cdots \\ x_K = x_0 \exp\{u_1 + u_2 + \cdots + u_K\} \end{cases}$$
- $\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \pi_3(\boldsymbol{u}(\boldsymbol{x})|\boldsymbol{\lambda}) \left| \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\boldsymbol{x}} \right| = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k)$
- How to set **hyperparameters** $\lambda_k^{-1} = \overline{u_k} = \overline{\log x_k / x_{k-1}}$?

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

- $\boldsymbol{u} \to \boldsymbol{x}$: $$\begin{cases} x_1 = x_0 \exp\{u_1\} \\ x_2 = x_0 \exp\{u_1 + u_2\} \\ \quad \cdots \\ x_K = x_0 \exp\{u_1 + u_2 + \cdots + u_K\} \end{cases}$$
- $\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \pi_3(\boldsymbol{u}(\boldsymbol{x})|\boldsymbol{\lambda}) \left| \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\boldsymbol{x}} \right| = \prod_{k=1}^{K} \mathsf{Pareto}(x_k | x_{k-1}, \lambda_k)$
- How to set **hyperparameters** $\lambda_k^{-1} = \overline{u_k} = \overline{\log x_k / x_{k-1}}$?
  - Use identity: $\langle x_k \rangle \equiv \int \mathrm{d}^K \boldsymbol{x}\, x_k \pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \frac{\lambda_k}{\lambda_k - 1} \langle x_{k-1} \rangle$

Deriving $\pi_3(\boldsymbol{x}|\cdot)$

Transform $\pi_3(\boldsymbol{u}|\boldsymbol{\lambda})$ to $\boldsymbol{x}$ space and re-express $\boldsymbol{\lambda}$

- $\boldsymbol{u} \to \boldsymbol{x}$: $$\begin{cases} x_1 = x_0 \exp\{u_1\} \\ x_2 = x_0 \exp\{u_1 + u_2\} \\ \quad \cdots \\ x_K = x_0 \exp\{u_1 + u_2 + \cdots + u_K\} \end{cases}$$
- $\pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \pi_3(\boldsymbol{u}(\boldsymbol{x})|\boldsymbol{\lambda}) \left| \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\boldsymbol{x}} \right| = \prod_{k=1}^{K} \mathsf{Pareto}(x_k|x_{k-1}, \lambda_k)$
- How to set **hyperparameters** $\lambda_k^{-1} = \overline{u_k} = \overline{\log x_k / x_{k-1}}$?
  - Use identity: $\langle x_k \rangle \equiv \int \mathrm{d}^K\boldsymbol{x}\, x_k \pi_3(\boldsymbol{x}|\boldsymbol{\lambda}) = \frac{\lambda_k}{\lambda_k - 1} \langle x_{k-1} \rangle$
  - Thus:

$$\boxed{\lambda_k = \frac{\overline{x_k}}{\overline{x_k} - \overline{x_{k-1}}}}$$

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

- **Convenient parametrization**: $\boxed{\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}}) \quad \text{where} \quad \overline{\boldsymbol{x_0}} \equiv (x_0, \overline{x_1}, \overline{x_2}, \cdots, \overline{x_K})}$

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

- **Convenient parametrization**: $\boxed{\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}}) \quad \text{where} \quad \overline{\boldsymbol{x_0}} \equiv (x_0, \overline{x_1}, \overline{x_2}, \cdots, \overline{x_K})}$

- **Inference insensitive** to values of the hyperparameters $\overline{\boldsymbol{x_0}}$
  - Maximum entropy: weak inductive bias
  - Heavy tails
  - Terrible sample statistics $\frac{1}{n}\sum_i x_k^{(i)} \to \overline{x_k}$

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

- **Convenient parametrization**: $\boxed{\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}}) \quad \text{where} \quad \overline{\boldsymbol{x_0}} \equiv (x_0, \overline{x_1}, \overline{x_2}, \cdots, \overline{x_K})}$

- **Inference insensitive** to values of the hyperparameters $\overline{\boldsymbol{x_0}}$
  - Maximum entropy: weak inductive bias
  - Heavy tails
  - Terrible sample statistics $\frac{1}{n}\sum_i x_k^{(i)} \to \overline{x_k}$

- **Sampling** trivial: $\boldsymbol{u} \to \boldsymbol{x}$

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

- **Convenient parametrization**: $\boxed{\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}}) \quad \text{where} \quad \overline{\boldsymbol{x_0}} \equiv (x_0, \overline{x_1}, \overline{x_2}, \cdots, \overline{x_K})}$

- **Inference insensitive** to values of the hyperparameters $\overline{\boldsymbol{x_0}}$
  - Maximum entropy: weak inductive bias
  - Heavy tails
  - Terrible sample statistics $\frac{1}{n}\sum_i x_k^{(i)} \to \overline{x_k}$

- **Sampling** trivial: $\boldsymbol{u} \to \boldsymbol{x}$

- **"Consistent"**
  - Marginalizing out higher frequencies $\equiv$ having set up $\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}})$ without knowledge of those frequencies

Further properties of $\pi_3(\boldsymbol{x}|\cdot)$

- **Convenient parametrization**: $\boxed{\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}}) \quad \text{where} \quad \overline{\boldsymbol{x_0}} \equiv (x_0, \overline{x_1}, \overline{x_2}, \cdots, \overline{x_K})}$

- **Inference insensitive** to values of the hyperparameters $\overline{\boldsymbol{x_0}}$
  - Maximum entropy: weak inductive bias
  - Heavy tails
  - Terrible sample statistics $\frac{1}{n}\sum_i x_k^{(i)} \to \overline{x_k}$

- **Sampling** trivial: $\boldsymbol{u} \to \boldsymbol{x}$

- **"Consistent"**
  - Marginalizing out higher frequencies $\equiv$ having set up $\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}})$ without knowledge of those frequencies

- **Scale invariant**: $\pi_3(c\boldsymbol{x}|\overline{\boldsymbol{x_0}}) = f(c)\pi_3(\boldsymbol{x}|\overline{\boldsymbol{x_0}})$ [Newman 2005]

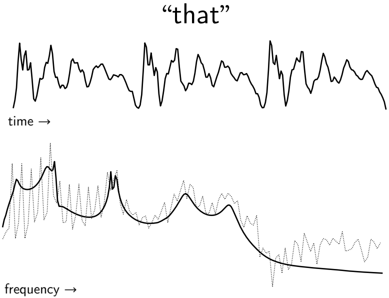
Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.

Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.

- Measure resonance frequencies of the human vocal tract
- Five representative vowel sounds taken from the CMU ARCTIC database [Kominek and Black 2004]
- $D \in \{\mathsf{shore}, \mathsf{that}, \mathsf{you}, \mathsf{little}, \mathsf{until}\}$
- Compare $Z_i(K)$ and $H_i(K)$ for each D and $i \in \{1, 2, 3\}$



Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.

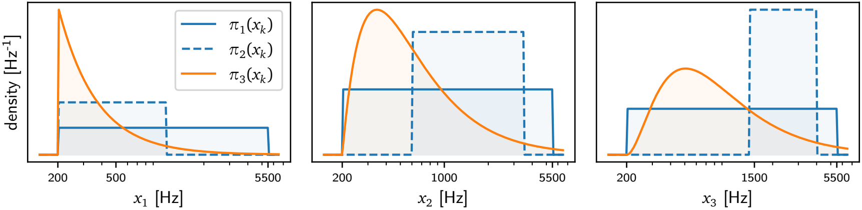
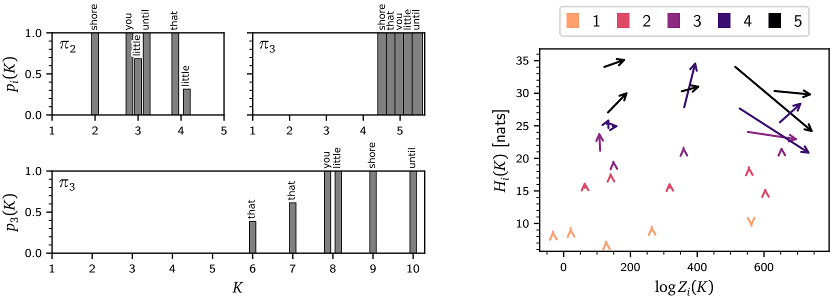


Figure: Comparison of π_1, π_2 and π_3 in terms of the marginal priors $\pi_i(x_k|\cdot)$ for the case $K := 3$. The marginal $\pi_i(x_k|\cdot)$ is obtained by integrating out the two other frequencies; for example, $\pi_i(x_1|\cdot) = \iint \mathrm{d}x_2\,\mathrm{d}x_3\,\pi_i(\boldsymbol{x}|\cdot)$. The pdfs are shown on a common log scale and are scaled by the appropriate Jacobian determinant $|\,\mathrm{d}x_k/\mathrm{d}\log x_k\,| = x_k$.

$\pi_1(\boldsymbol{x}|\cdot)$ **is excluded for** $K \geq 4$

Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.



(a)

(b)

Figure: **(a)** Model selection in Experiment I (top row) and Experiment II (bottom row). **(b)** In Experiment I, π_2 and π_3 are compared in terms of evidence $[\log Z_i(K)]$ and uninformativeness $[H_i(K)]$ for each (D, K). The arrows point from π_2 to π_3 and are color-coded by the value of K. For small values of K, the arrow lengths are too small to be visible on this scale.

Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.

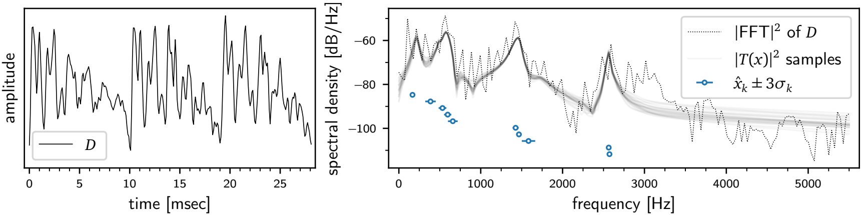


Figure: The VTR problem for the case ($D :=$ until, $K := 10$). Left panel: The data D, i.e., the quasi-periodic steady-state part consisting of 3 highly correlated pitch periods. Right panel: Inferred VTR frequency estimates $\{\hat{x}_k\}_{k=1}^{K}$ for $K := 10$ at 3 sigma. They describe the power spectral density of the vocal tract transfer function $|T(x)|^2$, represented here by 25 posterior samples and compared to the Fast Fourier Transform (FFT) of D. All $\hat{x}_k$ are well resolved and most have error bars too small to be seen on this scale.

Application

Compare the $\pi_i(\boldsymbol{x}|\cdot)$ candidates on a simple inference task.

Conclusions:

1. π_1 can't be used for $K \geq 4$
2. π_2 dominated by π_3 in terms of **evidence** $Z_i(K)$
3. π_2 about as uninformative as π_3 in terms of the **information** $H_i(K)$
4. π_3 can push K further

Summary

Summary

- The prior facilitates model selection problems in which the number K of resonance frequencies is unknown by enabling the use of more robust evidence-based methods, even in the presence of multiplets of arbitrary order.
    1. Solves label switching problem
    2. Solves multiplet problem

Summary

- The prior facilitates model selection problems in which the number K of resonance frequencies is unknown by enabling the use of more robust evidence-based methods, even in the presence of multiplets of arbitrary order.
  1. Solves label switching problem
  2. Solves multiplet problem

- The prior
  1. is in the exponential family,
  2. encodes a weakly inductive bias,
  3. provides a reasonable density everywhere,
  4. is easily parametrizable,
  5. is easy to sample from.

  That's enough! Meant to be overwhelmed.

Summary

- The prior facilitates model selection problems in which the number K of resonance frequencies is unknown by enabling the use of more robust evidence-based methods, even in the presence of multiplets of arbitrary order.
    1. Solves label switching problem
    2. Solves multiplet problem

- The prior
    1. is in the exponential family,
    2. encodes a weakly inductive bias,
    3. provides a reasonable density everywhere,
    4. is easily parametrizable,
    5. is easy to sample from.

  That's enough! Meant to be overwhelmed.

- The prior is valid for any collection of scale variables which are intrinsically ordered.
    – Does it apply to modeling spectra directly?

References I

Celeux, Gilles, Kaniav Kamary, Gertraud Malsiner-Walli, Jean-Michel Marin, and Christian P. Robert (2018). "Computational Solutions for Bayesian Inference in Mixture Models". In: *arXiv:1812.07240 [stat]*.

Jaynes (1968). "Prior Probabilities". In: *IEEE Transactions on Systems Science and Cybernetics* 4.3, pp. 227–241.

Kominek, John and Alan W Black (2004). "The CMU Arctic Speech Databases". In: *Fifth ISCA Workshop on Speech Synthesis*.

Littenberg, Tyson B. and Neil J. Cornish (2015). "Bayesian Inference for Spectral Estimation of Gravitational Wave Detector Noise". In: *Physical Review D* 91.8, p. 084034.

Newman, M. E. J. (2005). "Power Laws, Pareto Distributions and Zipf's Law". In: *Contemporary Physics* 46.5, pp. 323–351.

Wilson, Andrew Gordon+ (2014). "Bayesian Inference for NMR Spectroscopy with Applications to Chemical Quantification". In: *arXiv:1402.3580 [stat]*.

Xu, K., G. Marrelec, S. Bernard, and Q. Grimal (2019). "Lorentzian-Model-Based Bayesian Analysis for Automated Estimation of Attenuated Resonance Spectrum". In: *IEEE Transactions on Signal Processing* 67.1, pp. 4–16.