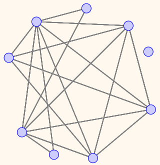
Projective Limit Techniques in Bayesian Nonparametrics

Peter Orbanz

University of Cambridge

Question

What is a reasonable nonparametric Bayesian model for graph-valued data?

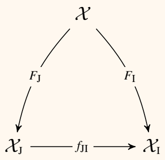


Bayesian Model on $\mathcal{X}$

Random measure $\Pi : \mathcal{T} \to \mathbf{M}(\mathcal{X})$ on probability space $(\mathcal{T}, \mathcal{A}, \mathbb{P})$.

- Prior = distribution $Q = \Pi(\mathbb{P})$
- Posterior = $Q[\Pi \in \ . \,|X_1, \ldots, X_n]$

Conditional probabilities $P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]$ for all $I \in \Gamma$.



Call family $\big\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]\big\rangle_{\Gamma}$

- a conditional promeasure if

$$P_{\mathrm{J}}[X_{\mathrm{J}} \in f_{\mathrm{JI}}^{-1}\,.\,|\Theta_{\mathrm{J}}] =_{\text{a.e.}} P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}] \qquad \text{if } I \preceq J$$

- tight on $\mathcal{X}$ if
  $\forall \varepsilon > 0\ \exists$ compact $K \in \mathcal{X}$: $P_{\mathrm{I}}[\mathcal{X}_{\mathrm{I}} \smallsetminus F_{\mathrm{I}}K|\theta_{\mathrm{I}}] \leq 1 - \varepsilon$

Theorem (O. 2010)

There is a conditional probabiliy $P[X \in \,.\,|\Theta]$ on $\mathcal{X}$ with

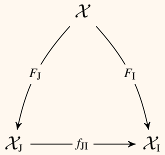
$$F_{\mathrm{I}}P[X \in \,.\,|\Theta] =_{\text{a.e.}} P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]$$

if and only if $\big\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]\big\rangle_{\Gamma}$ is a conditional promeasure and tight on $\mathcal{X}$.

Intuition

$$\big\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]\big\rangle_{\Gamma} \quad = \quad \text{proxy for } P[X \in \,.\,|\Theta]$$

Conditional probabilities $P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}]$ for all $I \in \Gamma$.



Call family $\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}]\rangle_{\Gamma}$

- a conditional promeasure if

$$P_{\mathrm{J}}[X_{\mathrm{J}} \in f_{\mathrm{JI}}^{-1}\,.\,|\Theta_{\mathrm{J}}] =_{\text{a.e.}} P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}] \qquad \text{if } I \preceq J$$

- tight on $\mathcal{X}$ if it satisfies concentration condition

Theorem (O. 2010)

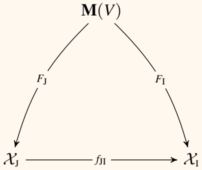
There is a conditional probabiliy $P[X \in \,.\, |\Theta]$ on $\mathcal{X}$ with

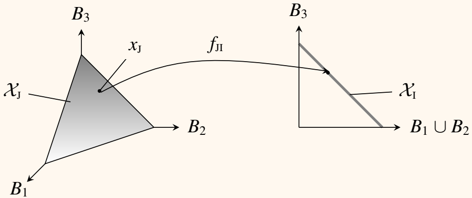
$$F_{\mathrm{I}}P[X \in \,.\, |\Theta] =_{\text{a.e.}} P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}]$$

if and only if $\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}]\rangle_{\Gamma}$ is a conditional promeasure and tight on $\mathcal{X}$.

Intuition

$$\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\, |\Theta_{\mathrm{I}}]\rangle_{\Gamma} \quad = \quad \text{proxy for } P[X \in \,.\, |\Theta]$$



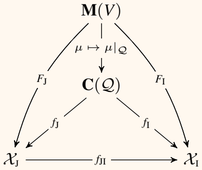


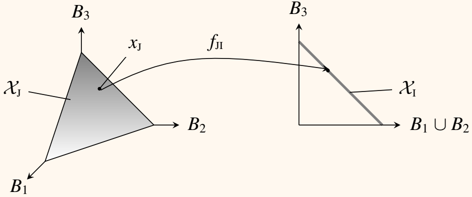
Theorem (O. 2011)

There is a unique $P[X \in \,.\,|\Theta]$ on $\mathbf{M}(V)$ with marginals $P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]$ if:

1. $\left\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]\right\rangle_{\Gamma}$ conditional promeasure.
2. For all $\theta \in \mathcal{T}$ exists $\mu_\theta \in \mathbf{M}(V)$ such that

$$\mathbb{E}_{Q_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|F_{\mathrm{I}}\theta]}[X_{\mathrm{I}}] = F_{\mathrm{I}}\mu_\theta \qquad \text{for all } I \in \Gamma \;.$$





Theorem (O. 2011)

There is a unique $P[X \in \,.\,|\Theta]$ on $\mathbf{M}(V)$ with marginals $P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]$ if:

1. $\left\langle P_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|\Theta_{\mathrm{I}}]\right\rangle_{\Gamma}$ conditional promeasure.
2. For all $\theta \in \mathcal{T}$ exists $\mu_\theta \in \mathbf{M}(V)$ such that

$$\mathbb{E}_{Q_{\mathrm{I}}[X_{\mathrm{I}} \in \,.\,|F_{\mathrm{I}}\theta]}[X_{\mathrm{I}}] = F_{\mathrm{I}}\mu_\theta \qquad \text{for all } I \in \Gamma\,.$$

Bayesian model

Likelihoods $P_{\text{I}}[X_{\text{I}} \in \,.\, |\Theta_{\text{I}}]$ and priors $Q_{\text{I}}[\Theta_{\text{I}} \in \,.\, |Y_{\text{I}}]$ for each I.

Theorem (O. 2010)

$$\begin{array}{ccc} Q & \xrightarrow{\{X = x\}} & Q[\Theta \in \,.\, |X] \\ \downarrow \uparrow \overleftarrow{\text{lim}} & & \downarrow \uparrow \overleftarrow{\text{lim}} \\ Q_{\text{I}} & \xrightarrow{\{X_{\text{I}} = F_{\text{I}} x\}} & Q_{\text{I}}[\Theta_{\text{I}} \in \,.\, |X_{\text{I}}] \end{array}$$

1. Weak convergence on $\mathcal{T}$ $\Leftrightarrow$ weak convergence on all $\mathcal{T}_{\text{I}}$
2. If there is a sufficient statistic S_{I} for each $\mathbf{p}_{\text{I}}$ and $\langle S_{\text{I}} \rangle_{\Gamma}$ projective: $S = \underleftarrow{\lim} \langle S_{\text{I}} \rangle_{\Gamma}$ is sufficient for $\mathbf{p}$
3. Each $\mathbf{p}_{\text{I}}$ exponential family, $\mathbf{q}_{\text{I}}$ its canonical conjugate prior: Nonparametric model conjugate with posterior update

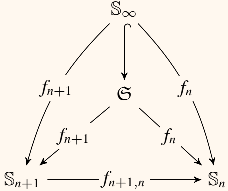
$$(\lambda, \gamma) \quad \mapsto \quad \Big(\lambda + n, \gamma + \sum_{i=1}^{n} S(X^{(i)})\Big) \; .$$

Virtual permutations (Kerov, Olshanski, Vershik, 2004)

$$\mathfrak{S} = \varprojlim \mathbb{S}_n = \text{virtual permutations}$$

Elements of $\mathfrak{S}$:

$$\pi = \sigma_{k_1}(1)\sigma_{k_2}(2)\sigma_{k_3}(3)\cdots$$



Nonparametric Bayesian model ("Cayley process")

$$P_n[\pi_n|\Theta=\theta] = \frac{1}{Z(\theta)}\exp\Big(-\sum_{j=1}^{n}\theta_j W_j(\pi)\Big)$$

Nonparametric model: $P_\infty[\pi|\Theta] = \varprojlim \big\langle P_n[\pi_n|\Theta]\big\rangle_\Gamma$ + conjugate prior

- Concentrates on $\mathbb{S}_\infty \subset \mathfrak{S}$ if $\mathbb{E}[W(\pi)] \in \ell_1$.
- For concentration on $\mathbb{S}_\infty$, sequence $\theta = (\theta^{(j)})_j$ has to diverge.

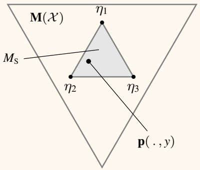
Problem

Given a model $P[\,.\,|Y] = \varprojlim \langle P_{\mathrm{I}}[\,.\,|Y_{\mathrm{I}}]\rangle_{\Gamma}$, what is the "natural" parameter space?

Assumptions

Sufficient statistic S_{I} for each I: $\quad P_{\mathrm{I}}[X_{\mathrm{I}} \in\ .\,|S_{\mathrm{I}}, Y] = P_{\mathrm{I}}[X_{\mathrm{I}} \in\ .\,|S_{\mathrm{I}}]$
Let M_{S} = maximal set of measures for which $S = \varprojlim S_{\mathrm{I}}$ is sufficient.

Theorem (O. 2011)



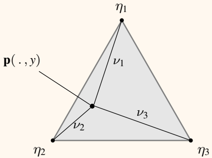
1. M_{S} is a convex.
2. Extreme points = measures $\eta \in M_{\mathrm{S}}$ with

$$\eta\{S = s_\eta\} = 1 \qquad \text{for some } s_\eta\ .$$

   Define: $\quad \mathcal{T}$ = space of all such s_η

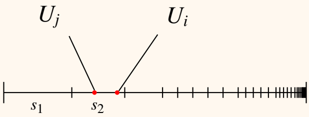
3. There is a unique conditional $\mathbf{v} : \mathcal{Y} \to \mathbf{M}(\mathcal{T})$ with

$$P[\,.\,|Y = y] =_{\text{a.e.}} \int_{\mathcal{T}} P[\,.\,|S = \theta]\mathbf{v}(d\theta, y)\ .$$



Ordered paint-box distribution ρ

S = set of infinite partitions $s = (s_1 \geq s_2 \geq \dots)$ of $[0, 1]$



Random virtual permutation $\pi \sim \rho(\,.\,, s)$:

1. Sample i.i.d. uniform $U_1, U_2, \dots$
2. i, j on same cycle in $\pi \;\Leftrightarrow\; i, j$ in same interval
3. i preceeds j on cycle $\Leftrightarrow U_i < U_j$

Corollary

Let Λ be the group of conjugations of $\mathbb{S}_\infty$. For any measure $P \in \mathbf{M}(\mathfrak{S})$ satisfying $\lambda_\sigma P = P$ for all $\lambda_\sigma \in \Lambda$, there is a unique ν such that

$$P = \int_{\mathbf{S}} \rho(\,.\,, s)\nu(ds)\;.$$

Projective Limit Representations: Summary

Required steps

Domain:

1. Set up $\mathcal{X}_{\mathrm{I}}$, $\mathcal{X}$, F_{I}, f_{JI}.
2. Get the topological structure right. All spaces must be Polish.

Model:

3. Projectivity
4. Concentration (tightness)

Results

- Nonparametric Bayesian model exists and unique
- Several properties follow from properties of marginals
- Exponential family marginals: Conjugacy generic
- Ergodic decomposition helps to identify $\mathcal{T}$ and symmetry properties