Bayesian inference as iterated random functions (IRF)

with applications to sequential inference in graphical models

- $\mathcal{P}_d := \mathcal{P}(\{0,1\}^d) :=$ class of probability measures on $\{0,1\}^d$. (A subset of $\mathbb{R}_+^{2^d}$.)
- View sequential Bayesian inference as an iteration on random elements of $\mathcal{P}_d$.
- An instance of the general iteration

$$Q_n = q_{\theta_n}(T(Q_{n-1})), \quad n \geq 1$$

Q_n	Probability measure in $\mathcal{P}_d$.
θ_n	Random element of $\mathbb{R}_+^{2^d}$. (Models likelihood)
T	Nonrandom (possibly nonlinear) map on $\mathcal{P}_d$.
q_θ	Bayesian update operator on $\mathcal{P}_d$:

$$q_\theta(x) := \frac{x \odot \theta}{\langle x, \theta \rangle}, \quad x \in \mathcal{P}_d$$

$x \odot \theta$	Pointwise multiplication.
$\langle x, \theta \rangle$	Euclidean inner product.

Bayesian inference as iterated random functions (IRF)

with applications to sequential inference in graphical models

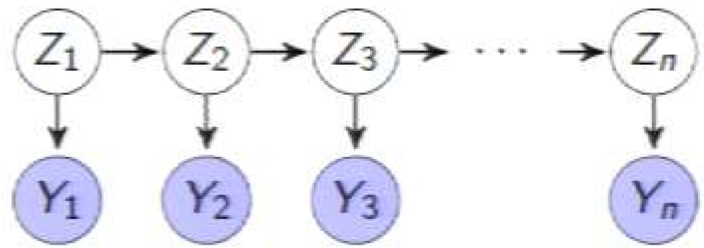
- $\mathcal{P}_d := \mathcal{P}(\{0,1\}^d) :=$ class of probability measures on $\{0,1\}^d$. (A subset of $\mathbb{R}_+^{2^d}$.)
- View sequential Bayesian inference as an iteration on random elements of $\mathcal{P}_d$.
- An instance of the general iteration

$$Q_n = q_{\theta_n}(T(Q_{n-1})), \quad n \geq 1$$

- Certain approximate inference algorithms are obtained by changing T.
- Original motivating example:
  Inference in a graphical model for *distributed change point detection*.
- General theory for convergence of $\{Q_n\}$ to an extreme point $\boldsymbol{e}$ of $\mathcal{P}_d$.
- Demonstrate the theory in concrete setting of
  approximate and exact inference for latent variable graphical model.

Bayesian inference as iterated random functions (IRF)

Simple example: Hidden Markov Model



- For exact inference,

$$\overbrace{P(Z_n | Y_n, Y_{n-1}, \ldots, Y_1)}^{Q_n} \overset{Z_n}{\propto} P(Z_n, Y_n | Y_{n-1}, \ldots, Y_1)$$

$$= \underbrace{P(Y_n | Z_n)}_{\theta_n} \underbrace{P(Z_n | Y_{n-1}, \ldots, Y_1)}_{R_n}$$

so that $Q_n = q_{\theta_n}(R_n)$, and

$$P(Z_n | Y_{n-1}, \ldots, Y_1) = \sum_{Z_{n-1}} P(Z_n, Z_{n-1} | Y_{n-1}, \ldots, Y_1)$$

$$= \sum_{Z_{n-1}} P(Z_n | Z_{n-1}) \underbrace{P(Z_{n-1} | Y_{n-1}, \ldots, Y_1)}_{Q_{n-1}}$$

so that $R_n = T_n(Q_{n-1})$. Linear operator via matrix multiplication with $[P(Z_n | Z_{n-1})]$.

- General theory for convergence of $\{Q_n\}$ to an extreme point $\boldsymbol{e}$ of $\mathcal{P}_d$.
- Classical approach to IRF, via contraction in mean, fails.
- We rely on semi-group property of $\{q_\theta\}$. ($\circ$ is function composition)

$$q_{\theta \odot \theta'} = q_\theta \circ q_{\theta'}, \quad \boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}_+^m.$$

Theorem

For all $n \geq 1$, and $\varepsilon > 0$, with probability at least $1 - e^{-Cn\varepsilon^2/\sigma_*^2}$, we have

$$\|Q_n - \mathbf{e}\|_1 \leq 2 \frac{1 - Q_0^{\mathbf{e}}}{Q_0^{\mathbf{e}}} \left(L e^{-I_* + \varepsilon} \right)^n$$

where $Q_0^{\mathbf{e}} := \langle Q_0, \mathbf{e} \rangle$.

Let $\boldsymbol{\theta}_n^* := \|\boldsymbol{\theta}_n - \langle \boldsymbol{\theta}_n, \mathbf{e} \rangle \mathbf{e}\|_\infty$. Assumptions:

$\{\log \boldsymbol{\theta}_n^*\}$	i.i.d. sub-Gaussian with mean $\leq -I_* < 0$ and norm $\leq \sigma_* \in (0, \infty)$,
T	Lipschitz with $L := \mathrm{Lip}(T) \leq 1$,
$\boldsymbol{e}$	a fixed point of T.