Construction and properties of Bayesian nonparametric regression models

Isaac Newton Institute, Cambridge

Models beyond the Dirichlet process

Antonio Lijoi

University of Pavia

6 August 2007

Outline

- Discrete nonparametric priors
  - Neutral to the right priors and models for the cumulative hazard
  - Normalized random measures with independent increments
  - Poisson–Kingman models
  - Species sampling models & stick–breaking priors
- Priors for density estimation
  - Hierarchical mixture models
  - Random Bernstein polynomials
  - Pólya trees

Sequence of exchangeable observations

- Suitable mathematical framework for inference alternative to the classical approach with “fixed and unknown probability distribution P”

Sequence of exchangeable observations

- Suitable mathematical framework for inference alternative to the classical approach with "fixed and unknown probability distribution P"
- Observations are recorded under homogeneous physical conditions

Sequence of exchangeable observations

- Suitable mathematical framework for inference alternative to the classical approach with "fixed and unknown probability distribution P"
- Observations are recorded under homogeneous physical conditions
- Inferences are not affected by the order with which observations are obtained

Sequence of exchangeable observations

- Suitable mathematical framework for inference alternative to the classical approach with "fixed and unknown probability distribution P"
- Observations are recorded under homogeneous physical conditions
- Inferences are not affected by the order with which observations are obtained

For any $n \geq 1$,

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\pi(1)}, \dots, X_{\pi(n)})$$

for any permutation π of the vector $(1, \dots, n)$

$X^{(\infty)} = (X_n)_{n \geq 1}$	sequence of $\mathbb{X}$–valued random variables
$\mathscr{P}(\mathbb{X})$	space of probability measures on $(\mathbb{X}, \mathscr{X})$

$X^{(\infty)} = (X_n)_{n \geq 1}$	sequence of $\mathbb{X}$–valued random variables
$\mathscr{P}(\mathbb{X})$	space of probability measures on $(\mathbb{X}, \mathscr{X})$

de Finetti's representation theorem

The sequence $X^{(\infty)}$ is exchangeable if and only if

$$\mathbb{P}\left[X^{(\infty)} \in A\right] = \int_{\mathscr{P}(\mathbb{X})} p^{\infty}(A)\, Q(\mathrm{d}p) \qquad \forall A \in \mathscr{X}^{\infty}$$

where p^{∞} is the infinite product measure $p \times p \times \cdots$ on $(\mathbb{X}^{\infty}, \mathscr{X}^{\infty})$

Conditionally on a random probability measure $\tilde{p}$, the observations $X_1, \ldots, X_n$ are iid with common distribution $\tilde{p}$, *i.e.*

$$\Pr\left[X_1 \in A_1, \ldots, X_n \in A_n \,\middle|\, \tilde{p}\right] = \prod_{i=1}^{n} \tilde{p}(A_i)$$

Conditionally on a random probability measure $\tilde{p}$, the observations $X_1, \ldots, X_n$ are iid with common distribution $\tilde{p}$, *i.e.*

$$\Pr\left[X_1 \in A_1, \ldots, X_n \in A_n \,\middle|\, \tilde{p}\right] = \prod_{i=1}^{n} \tilde{p}(A_i)$$

Q is the probability distribution of $\tilde{p}$ and it is also known as the *de Finetti measure*

Conditionally on a random probability measure $\tilde{p}$, the observations $X_1, \ldots, X_n$ are iid with common distribution $\tilde{p}$, *i.e.*

$$\Pr\left[X_1 \in A_1, \ldots, X_n \in A_n \,\middle|\, \tilde{p}\right] = \prod_{i=1}^{n} \tilde{p}(A_i)$$

Q is the probability distribution of $\tilde{p}$ and it is also known as the *de Finetti measure*

$$Q = \text{prior distribution}$$

Conditionally on a random probability measure $\tilde{p}$, the observations $X_1, \ldots, X_n$ are iid with common distribution $\tilde{p}$, *i.e.*

$$\Pr\left[X_1 \in A_1, \ldots, X_n \in A_n \,\middle|\, \tilde{p}\right] = \prod_{i=1}^{n} \tilde{p}(A_i)$$

Q is the probability distribution of $\tilde{p}$ and it is also known as the *de Finetti measure*

$$Q = \text{prior distribution}$$

Moreover

$$\hat{p}_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \;\Rightarrow\; \tilde{p} \qquad \text{a.s.--}\mathbb{P}$$

and $\Rightarrow$ denotes "weak convergence"

Dirichlet process

Common specification of Q

- Dirichlet process
- Mixture of the Dirichlet process

Dirichlet process

Common specification of Q

- Dirichlet process
- Mixture of the Dirichlet process

Reasons for the use of the Dirichlet process

Dirichlet process

Common specification of Q

- Dirichlet process
- Mixture of the Dirichlet process

Reasons for the use of the Dirichlet process

- Conjugacy

Dirichlet process

Common specification of Q

- Dirichlet process
- Mixture of the Dirichlet process

Reasons for the use of the Dirichlet process

- Conjugacy
- Simple system of predictive distributions which gives rise to simulation algorithms useful, *e.g.*, in mixture models

Other specifications of Q

Other specifications of Q

- Suitable transformations of *completely random measures*

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes
- Mixtures of discrete random probability measures $\tilde{p}$

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes
- Mixtures of discrete random probability measures $\tilde{p}$
- Partitioning of the sample space: neutral random probabilities, Pólya trees

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes
- Mixtures of discrete random probability measures $\tilde{p}$
- Partitioning of the sample space: neutral random probabilities, Pólya trees
- Linear convex combinations of beta densities with random weights: random Bernstein polynomials

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes
- Mixtures of discrete random probability measures $\tilde{p}$
- Partitioning of the sample space: neutral random probabilities, Pólya trees
- Linear convex combinations of beta densities with random weights: random Bernstein polynomials
- Transformations of Gaussian processes

Other specifications of Q

- Suitable transformations of *completely random measures*
- Series representation: species sampling models, stick–breaking priors, dependent processes
- Mixtures of discrete random probability measures $\tilde{p}$
- Partitioning of the sample space: neutral random probabilities, Pólya trees
- Linear convex combinations of beta densities with random weights: random Bernstein polynomials
- Transformations of Gaussian processes
- ...

Completely random measures, Kingman (1967, 1993)

Completely random measures, Kingman (1967, 1993)

Let $\tilde{\mu}$ be a a function on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in the space of all measures on $\mathbb{X}$ such that

- for any $A_1, \ldots, A_n$ in $\mathscr{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent

Completely random measures, Kingman (1967, 1993)

Let $\tilde{\mu}$ be a a function on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in the space of all measures on $\mathbb{X}$ such that

- for any $A_1, \ldots, A_n$ in $\mathscr{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent

The random measure $\tilde{\mu}$ is a *completely random measure*

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \, \delta_{x_i}$$

where

Completely random measures, Kingman (1967, 1993)

Let $\tilde{\mu}$ be a a function on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in the space of all measures on $\mathbb{X}$ such that

- for any $A_1, \ldots, A_n$ in $\mathscr{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent

The random measure $\tilde{\mu}$ is a *completely random measure*

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \, \delta_{x_i}$$

where

- the fixed jump points $x_1, \ldots, x_M$ are in $\mathbb{X}$

Completely random measures, Kingman (1967, 1993)

Let $\tilde{\mu}$ be a a function on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in the space of all measures on $\mathbb{X}$ such that

- for any $A_1, \ldots, A_n$ in $\mathscr{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent

The random measure $\tilde{\mu}$ is a *completely random measure*

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \, \delta_{x_i}$$

where

- the fixed jump points $x_1, \ldots, x_M$ are in $\mathbb{X}$
- the (non–negative) random jumps $V_1, \ldots, V_M$ are mutually independent and they are independent from $\tilde{\mu}_c$

Completely random measures, Kingman (1967, 1993)

Let $\tilde{\mu}$ be a a function on $(\Omega, \mathscr{F}, \mathbb{P})$ with values in the space of all measures on $\mathbb{X}$ such that

- for any $A_1, \ldots, A_n$ in $\mathscr{X}$ such that $A_i \cap A_j = \varnothing$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent

The random measure $\tilde{\mu}$ is a *completely random measure*

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \, \delta_{x_i}$$

where

- the fixed jump points $x_1, \ldots, x_M$ are in $\mathbb{X}$
- the (non–negative) random jumps $V_1, \ldots, V_M$ are mutually independent and they are independent from $\tilde{\mu}_c$
- the random measure $\tilde{\mu}_c$ has a Lévy measure ν on $\mathbb{X} \times \mathbb{R}^+$

The Lévy–Khintchine representation yields

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} f(x)\,\tilde{\mu}_c(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{X}\times\mathbb{R}^+}\left[1-\mathrm{e}^{-s\,f(x)}\right]\,\nu(\mathrm{d}x,\mathrm{d}s)\right\}$$

for any measurable $f : \mathbb{X} \to \mathbb{R}$

The Lévy–Khintchine representation yields

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} f(x)\,\tilde{\mu}_c(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{X}\times\mathbb{R}^+}\left[1-\mathrm{e}^{-s\,f(x)}\right]\,\nu(\mathrm{d}x,\mathrm{d}s)\right\}$$

for any measurable $f:\mathbb{X}\to\mathbb{R}$

The probability distribution of $\tilde{\mu}_c$ is fully characterized by ν

The Lévy–Khintchine representation yields

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} f(x)\,\tilde{\mu}_c(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{X}\times\mathbb{R}^+}\left[1-\mathrm{e}^{-s\,f(x)}\right]\,\nu(\mathrm{d}x,\mathrm{d}s)\right\}$$

for any measurable $f : \mathbb{X} \to \mathbb{R}$

The probability distribution of $\tilde{\mu}_c$ is fully characterized by ν

Suppose there exist

- measure η on $\mathbb{X}$
- measure ρ_x on $\mathbb{R}^+$, for any $x \in \mathbb{X}$

such that

$$\nu(\mathrm{d}x,\mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

The Lévy–Khintchine representation yields

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} f(x)\,\tilde{\mu}_c(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{X}\times\mathbb{R}^+}\left[1-\mathrm{e}^{-s\,f(x)}\right]\,\nu(\mathrm{d}x,\mathrm{d}s)\right\}$$

for any measurable $f : \mathbb{X} \to \mathbb{R}$

The probability distribution of $\tilde{\mu}_c$ is fully characterized by ν

Suppose there exist

- measure η on $\mathbb{X}$
- measure ρ_x on $\mathbb{R}^+$, for any $x \in \mathbb{X}$

such that

$$\nu(\mathrm{d}x,\mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

If $\rho_x = \rho$, then $\tilde{\mu}_c$ is said *homogeneous*

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Let $Y_t = \tilde{\mu}((0, t])$ for any $t \geq 0$. Then

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Let $Y_t = \tilde{\mu}((0, t])$ for any $t \geq 0$. Then

$Y_{t_1} = \tilde{\mu}((0, t_1]),$

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Let $Y_t = \tilde{\mu}((0, t])$ for any $t \geq 0$. Then

$$Y_{t_1} = \tilde{\mu}((0, t_1]), \quad Y_{t_2} - Y_{t_1} = \tilde{\mu}((t_1, t_2]), \quad \ldots \quad ,$$

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Let $Y_t = \tilde{\mu}((0, t])$ for any $t \geq 0$. Then

$$Y_{t_1} = \tilde{\mu}((0, t_1]), \quad Y_{t_2} - Y_{t_1} = \tilde{\mu}((t_1, t_2]), \quad \ldots \quad , Y_{t_k} - Y_{t_{k-1}} = \tilde{\mu}((t_{k-1}, t_k])$$

Neutral to the right priors

Let $\mathbb{X} = \mathbb{R}^+$

Definition (Doksum, 1974)

A random distribution function F on $\mathbb{R}^+$ is neutral to the right (NTR) if, for any $0 < t_1 < t_2 < \cdots < t_k < \infty$ and $k \geq 1$, the random variables

$$F(t_1), \quad \frac{F(t_2) - F(t_1)}{1 - F(t_1)}, \quad \ldots \quad , \frac{F(t_k) - F(t_{k-1})}{1 - F(t_{k-1})}$$

are independent

Let $Y_t = \tilde{\mu}((0, t])$ for any $t \geq 0$. Then

$$Y_{t_1} = \tilde{\mu}((0, t_1]), \quad Y_{t_2} - Y_{t_1} = \tilde{\mu}((t_1, t_2]), \quad \ldots \quad , Y_{t_k} - Y_{t_{k-1}} = \tilde{\mu}((t_{k-1}, t_k])$$

and $Y = \{Y_t : \ t \geq 0\}$ is an independent increments process

Connection with independent increments processes (Doksum, 1974)

F is NTR if and only if there exists a process with independent increments $Y = \{Y_t : \ t \geq 0\}$ such that $\mathbb{P}\left[\lim_{t\to\infty} Y_t = \infty\right] = 1$, $\mathbb{P}\left[Y_0 = 0\right] = 1$ and

$$F(t) = 1 - \mathrm{e}^{-Y_t} \qquad \forall t \geq 0$$

Connection with independent increments processes (Doksum, 1974)

F is NTR if and only if there exists a process with independent increments $Y = \{Y_t : t \geq 0\}$ such that $\mathbb{P}[\lim_{t\to\infty} Y_t = \infty] = 1$, $\mathbb{P}[Y_0 = 0] = 1$ and

$$F(t) = 1 - \mathrm{e}^{-Y_t} \qquad \forall t \geq 0$$

Implications of the above representation

Connection with independent increments processes (Doksum, 1974)

F is NTR if and only if there exists a process with independent increments $Y = \{Y_t : t \geq 0\}$ such that $\mathbb{P}\left[\lim_{t\to\infty} Y_t = \infty\right] = 1$, $\mathbb{P}\left[Y_0 = 0\right] = 1$ and

$$F(t) = 1 - \mathrm{e}^{-Y_t} \qquad \forall t \geq 0$$

Implications of the above representation

- F is a discrete distribution function with probability 1

Connection with independent increments processes (Doksum, 1974)

F is NTR if and only if there exists a process with independent increments $Y = \{Y_t : \ t \geq 0\}$ such that $\mathbb{P}\left[\lim_{t\to\infty} Y_t = \infty\right] = 1$, $\mathbb{P}\left[Y_0 = 0\right] = 1$ and

$$F(t) = 1 - \mathrm{e}^{-Y_t} \qquad \forall t \geq 0$$

Implications of the above representation

- F is a discrete distribution function with probability 1
- Prior guess at the shape of F

$$\mathbb{E}[F(t)] = 1 - \mathbb{E}\left[\mathrm{e}^{-Y_t}\right] =$$

Connection with independent increments processes (Doksum, 1974)

F is NTR if and only if there exists a process with independent increments $Y = \{Y_t : \ t \geq 0\}$ such that $\mathbb{P}[\lim_{t\to\infty} Y_t = \infty] = 1$, $\mathbb{P}[Y_0 = 0] = 1$ and

$$F(t) = 1 - \mathrm{e}^{-Y_t} \qquad \forall t \geq 0$$

Implications of the above representation

- F is a discrete distribution function with probability 1
- Prior guess at the shape of F

$$\mathbb{E}[F(t)] = 1 - \mathbb{E}\left[\mathrm{e}^{-Y_t}\right] = 1 - \mathrm{e}^{-\int_{(0,t]\times\mathbb{R}^+}\left[1-\mathrm{e}^{-s}\right]\eta(\mathrm{d}x)\,\rho_x(\mathrm{d}s)}$$

where we suppose $\tilde{\mu}$ has no fixed points of discontinuity

In the sequel we write that

$$F \sim \mathrm{NTR}(Y)$$

to mean that $F(t) = 1 - \mathrm{e}^{-Y_t}$. We will also suppose that the prior process Y does not have fixed points of discontinuity

In the sequel we write that

$$F \sim \mathrm{NTR}(Y)$$

to mean that $F(t) = 1 - \mathrm{e}^{-Y_t}$. We will also suppose that the prior process Y does not have fixed points of discontinuity

NTR priors are conjugate, Doksum (1974)

$$F \sim \ \mathrm{NTR}(Y) \quad \Longrightarrow \quad F \,|\, \text{ Data } \sim \ \mathrm{NTR}(Y^*)$$

where $Y^* = \{Y_t^* : \ t \geq 0\}$ is an independent increments process with fixed points of discontinuity

In the sequel we write that

$$F \sim \mathrm{NTR}(Y)$$

to mean that $F(t) = 1 - \mathrm{e}^{-Y_t}$. We will also suppose that the prior process Y does not have fixed points of discontinuity

NTR priors are conjugate, Doksum (1974)

$$F \sim \ \mathrm{NTR}(Y) \quad \Longrightarrow \quad F \,|\; \mathrm{Data} \ \sim \ \mathrm{NTR}(Y^*)$$

where $Y^* = \{Y_t^* :\ t \geq 0\}$ is an independent increments process with fixed points of discontinuity

Description of Y^* (Ferguson, 1974)

If $X_1^*, \ldots, X_k^*$ are the distinct observations among the X_i's,

$$Y_t^* = Y_t^c + \sum_{\{i\colon X_i^* \leq t\}} J_i$$

Updating of Y_t

Updating of Y_t

- $Y^c = \{Y_t^c : \; t \geq 0\}$ with updated Lévy measure

$$\nu^*((0,t], \mathrm{d}s) = \int_{(0,t]} \mathrm{e}^{-\bar{N}(x)\, s}\, \eta(\mathrm{d}x)\, \rho_x(\mathrm{d}s)$$

where

$$\bar{N}(x) = \text{ “at risk process” } = \text{ “nr of obs } \geq x\text{”}$$

Updating of Y_t

- $Y^c = \{Y_t^c :\ t \geq 0\}$ with updated Lévy measure

$$\nu^*((0,t], \mathrm{d}s) = \int_{(0,t]} \mathrm{e}^{-\bar{N}(x)\, s}\, \eta(\mathrm{d}x)\, \rho_x(\mathrm{d}s)$$

  where

$$\bar{N}(x) = \text{ “at risk process” } = \text{ “nr of obs } \geq x\text{”}$$

- The random jumps $J_1, \ldots, J_k$ are mutually independent and independent from Y^c

Updating of Y_t

- $Y^c = \{Y_t^c :\ t \geq 0\}$ with updated Lévy measure

$$\nu^*((0,t], \mathrm{d}s) = \int_{(0,t]} \mathrm{e}^{-\bar{N}(x)\, s}\, \eta(\mathrm{d}x)\, \rho_x(\mathrm{d}s)$$

  where

$$\bar{N}(x) = \text{ “at risk process” } = \text{ “nr of obs } \geq x\text{”}$$

- The random jumps $J_1, \ldots, J_k$ are mutually independent and independent from Y^c
- Probability distribution of each jump J_i in terms of $\nu(\mathrm{d}x, \mathrm{d}s)$ and $\bar{N}$

Updating of Y_t

- $Y^c = \{Y_t^c : \ t \geq 0\}$ with updated Lévy measure

$$\nu^*((0,t], \mathrm{d}s) = \int_{(0,t]} \mathrm{e}^{-\bar{N}(x)\, s}\, \eta(\mathrm{d}x)\, \rho_x(\mathrm{d}s)$$

where

$$\bar{N}(x) = \text{“at risk process”} \ = \ \text{“nr of obs } \geq x\text{”}$$

- The random jumps $J_1, \ldots, J_k$ are mutually independent and independent from Y^c
- Probability distribution of each jump J_i in terms of $\nu(\mathrm{d}x, \mathrm{d}s)$ and $\bar{N}$

$$\rho_{x_i^*} = \rho \quad \Longrightarrow \quad \text{distribution of } J_i \text{ does not depend on } X_i^*$$

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

- $c_1, \ldots, c_n$ are n censoring times

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

- $c_1, \ldots, c_n$ are n censoring times
- To each X_i associate

$$\Delta_i = \mathbb{I}_{(0,c_i)}(X_i) = \begin{cases} 1 & \text{if } X_i \text{ is actually observed} \\ 0 & \text{if } X_i \text{ is censored} \end{cases}$$

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

- $c_1, \ldots, c_n$ are n censoring times
- To each X_i associate

$$\Delta_i = \mathbb{I}_{(0,c_i)}(X_i) = \begin{cases} 1 & \text{if } X_i \text{ is actually observed} \\ 0 & \text{if } X_i \text{ is censored} \end{cases}$$

- $T_i = \min\{X_i, c_i\}$

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

- $c_1, \ldots, c_n$ are n censoring times
- To each X_i associate

$$\Delta_i = \mathbb{I}_{(0,c_i)}(X_i) = \begin{cases} 1 & \text{if } X_i \text{ is actually observed} \\ 0 & \text{if } X_i \text{ is censored} \end{cases}$$

- $T_i = \min\{X_i, c_i\}$
- Data $= (T_1, \Delta_1), \ldots, (T_n, \Delta_n)$

Typical application: **survival analysis**

$$\text{Survival function } = S(t) = \mathrm{e}^{-Y_t} \qquad t \geq 0$$

and the X_i's represent survival times or times to an event

Right censored data

- $c_1, \ldots, c_n$ are n censoring times
- To each X_i associate

$$\Delta_i = \mathbb{I}_{(0,c_i)}(X_i) = \begin{cases} 1 & \text{if } X_i \text{ is actually observed} \\ 0 & \text{if } X_i \text{ is censored} \end{cases}$$

- $T_i = \min\{X_i, c_i\}$
- Data $= (T_1, \Delta_1), \ldots, (T_n, \Delta_n)$

Ferguson & Phadia (1979): NTR are conjugate also with respect to right–censored data

$$F \sim \text{NTR}(Y) \quad \Longrightarrow \quad F \mid \text{ Data } \sim \text{ NTR}(Y^*)$$

Example: The Dirichlet process

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

There exists an independent increments process $Y = \{Y_t : t \geq 0\}$ such that $F(t) = 1 - \mathrm{e}^{-Y_t}$

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

There exists an independent increments process $Y = \{Y_t : t \geq 0\}$ such that $F(t) = 1 - \mathrm{e}^{-Y_t}$

Lévy measure of Y?

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

There exists an independent increments process $Y = \{Y_t : \; t \geq 0\}$ such that $F(t) = 1 - \mathrm{e}^{-Y_t}$

Lévy measure of Y?

- $F_\eta(t) = \eta((0, t])$
- $a = \eta(\mathbb{R}^+)$

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

There exists an independent increments process $Y = \{Y_t : t \geq 0\}$ such that $F(t) = 1 - \mathrm{e}^{-Y_t}$

Lévy measure of Y?

- $F_\eta(t) = \eta((0, t])$
- $a = \eta(\mathbb{R}^+)$

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\mathbf{e}^{-(a - F_\eta(x))s}}{(1 - \mathbf{e}^{-s})}\, \eta(\mathrm{d}x) = \rho_x(\mathbf{d}s)\, \eta(\mathrm{d}x)$$

Example: The Dirichlet process

If F is a Dirichlet process with parameter measure η, then F is NTR

There exists an independent increments process $Y = \{Y_t : t \geq 0\}$ such that $F(t) = 1 - \mathrm{e}^{-Y_t}$

Lévy measure of Y?

- $F_\eta(t) = \eta((0, t])$
- $a = \eta(\mathbb{R}^+)$

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\mathbf{e}^{-(a-F_\eta(x))s}}{(1-\mathbf{e}^{-s})}\, \eta(\mathrm{d}x) = \rho_x(\mathbf{ds})\, \eta(\mathrm{d}x)$$

Susarla and van Ryzin (1976)

- $F|\text{Data} \neq$ Dirichlet
- Kaplan Meier estimator as a limit of the Bayes' estimate $\mathbb{E}[S(t)|\text{data}]$, with $\eta(\mathbb{R}^+) \to 0$

Example: the beta–Stacy process (Walker & Muliere, 1997)

The Dirichlet process is not conjugate with respect to right censoring

Example: the beta–Stacy process (Walker & Muliere, 1997)

The Dirichlet process is not conjugate with respect to right censoring

- m is some measure on $\mathbb{R}^+$ with $m(\mathbb{R}^+) = \bar{m}$ and with distribution function M, *i.e.* $M(x) = m((0, x])$ for any $x \geq 0$
- $x_1, x_2, \ldots$ are the (possible) points of discontinuities of M
- M_c is the continuous part of M

Example: the beta–Stacy process (Walker & Muliere, 1997)

The Dirichlet process is not conjugate with respect to right censoring

- m is some measure on $\mathbb{R}^+$ with $m(\mathbb{R}^+) = \bar{m}$ and with distribution function M, *i.e.* $M(x) = m((0, x])$ for any $x \geq 0$
- $x_1, x_2, \ldots$ are the (possible) points of discontinuities of M
- M_c is the continuous part of M

Definition

F a *beta–Stacy* process with parameters c and M

Example: the beta–Stacy process (Walker & Muliere, 1997)

The Dirichlet process is not conjugate with respect to right censoring

- m is some measure on $\mathbb{R}^+$ with $m(\mathbb{R}^+) = \bar{m}$ and with distribution function M, *i.e.* $M(x) = m((0,x])$ for any $x \geq 0$
- $x_1, x_2, \ldots$ are the (possible) points of discontinuities of M
- M_c is the continuous part of M

Definition

F a *beta–Stacy* process with parameters c and M if $F \sim$NTR(Y) and Y is an independent increments process with Lévy measure

$$\nu((0,t], ds) = \frac{ds}{1 - e^{-s}} \int_{(0,t]} e^{-s\,c(x)\,[\bar{m} - M(x)]}\, c(x)\, dM_c(x)$$

Example: the beta–Stacy process (Walker & Muliere, 1997)

The Dirichlet process is not conjugate with respect to right censoring

- m is some measure on $\mathbb{R}^+$ with $m(\mathbb{R}^+) = \bar{m}$ and with distribution function M, *i.e.* $M(x) = m((0, x])$ for any $x \geq 0$
- $x_1, x_2, \ldots$ are the (possible) points of discontinuities of M
- M_c is the continuous part of M

Definition

F a *beta–Stacy* process with parameters c and M if $F \sim$NTR(Y) and Y is an independent increments process with Lévy measure

$$\boldsymbol{\nu((0,t], ds) = \frac{ds}{1 - e^{-s}} \int_{(0,t]} e^{-s\, c(x)\, [\bar{m} - M(x)]}\, c(x)\, dM_c(x)}$$

In symbols $F \sim \mathscr{S}(c, M)$

The Dirichlet process arises as a special case if

- M is continuous
- $c(x) \equiv c > 0$ for any $x \geq 0$

The Dirichlet process arises as a special case if

- M is continuous
- $c(x) \equiv c > 0$ for any $x \geq 0$

Log–beta process

The Dirichlet process arises as a special case if

- M is continuous
- $c(x) \equiv c > 0$ for any $x \geq 0$

Log–beta process

Suppose

- η is a measure on $\mathbb{R}^+$ and set $F_\eta(x) = \eta((0, x])$
- $\{x_1, x_2, \ldots\}$ is the set of possible discontinuities of F_η
- F_η^c is the continuous part of F_η

The Dirichlet process arises as a special case if

- M is continuous
- $c(x) \equiv c > 0$ for any $x \geq 0$

Log–beta process

Suppose

- η is a measure on $\mathbb{R}^+$ and set $F_\eta(x) = \eta((0,x])$
- $\{x_1, x_2, \ldots\}$ is the set of possible discontinuities of F_η
- F_η^c is the continuous part of F_η

An independent increments process $Y = \{Y_t : \ t \geq 0\}$ with Lévy measure

$$\boldsymbol{\nu((0,t],ds) = \frac{ds}{1-e^{-s}} \int_{(0,t]} e^{-s\,[\beta(x)+\eta(\{x\})]}\, dF_\eta^c(x)}$$

is a log-beta process

Suitable choices of the measure η and of the function β of a log–beta process $Y = \{Y_t : t \geq 0\}$ lead to a beta–Stacy process $F = 1 - \mathrm{e}^{-Y}$ with parameters

$$M(x) = 1 - \mathrm{e}^{-\int_{(0,x]} \eta(\mathrm{d}v)/\beta(v)} \qquad c(x) = \frac{\beta(x)}{1 - M(x)}$$

Suitable choices of the measure η and of the function β of a log–beta process $Y = \{Y_t :\ t \geq 0\}$ lead to a beta–Stacy process $F = 1 - \mathrm{e}^{-Y}$ with parameters

$$M(x) = 1 - \mathrm{e}^{-\int_{(0,x]} \eta(\mathrm{d}v)/\beta(v)} \qquad c(x) = \frac{\beta(x)}{1 - M(x)}$$

Conjugacy

Let $F \sim$NTR(Y) and Y be a log–beta process with parameters η and β. Then

$$F \,|\ \text{Data} \ \sim \ \text{NTR}(Y^*)$$

Suitable choices of the measure η and of the function β of a log–beta process $Y = \{Y_t : t \geq 0\}$ lead to a beta–Stacy process $F = 1 - \mathrm{e}^{-Y}$ with parameters

$$M(x) = 1 - \mathrm{e}^{-\int_{(0,x]} \eta(\mathrm{d}v)/\beta(v)} \qquad c(x) = \frac{\beta(x)}{1 - M(x)}$$

Conjugacy

Let $F \sim$NTR(Y) and Y be a log–beta process with parameters η and β. Then

$$F \mid \text{ Data } \sim \text{ NTR}(Y^*)$$

where $Y^* = \{Y_t^* : t \geq 0\}$ is a log–beta process with parameters η^* and β^* such that

$$\mathbf{F}_{\eta^*}(\mathbf{x}) = \mathbf{F}_{\eta}(\mathbf{x}) + \sum_{\{\mathbf{i}:\, \boldsymbol{\Delta}_\mathbf{i}=1\}} \mathbb{I}(\mathbf{T}_\mathbf{i} \leq \mathbf{x})$$

$$\boldsymbol{\beta}^*(\mathbf{x}) = \boldsymbol{\beta}(\mathbf{x}) + \bar{\mathbf{N}}(\mathbf{x}) - \sum_{\{\mathbf{i}:\, \boldsymbol{\Delta}_\mathbf{i}=1\}} \mathbb{I}(\mathbf{T}_\mathbf{i} = \mathbf{x})$$

As a corollary one has conjugacy of the beta–Stacy process

As a corollary one has conjugacy of the beta–Stacy process

Let $F \sim \mathscr{S}(c, M)$. Then

$$F \mid \text{ Data } \sim \mathscr{S}(c^*, M^*)$$

for suitable M^* and c^*. For details see Walker & Muliere (1997)

As a corollary one has conjugacy of the beta–Stacy process

Let $F \sim \mathscr{S}(c, M)$. Then

$$F \,|\, \text{ Data } \sim \mathscr{S}(c^*, M^*)$$

for suitable M^* and c^*. For details see Walker & Muliere (1997)

Bayes estimate of F, under squared loss, is

$$\hat{F}(t) = \mathbb{E}[F(t) \,|\, \text{data}] = M^*(t)$$

and, as $c(\,\cdot\,) \to 0$, $\hat{F}$ converges to the Kaplan–Meier estimator of F

Some remarks

Some remarks

- Walker & Muliere (1997) and Walker & Damien (1998): prior specification with NTR processes

Some remarks

- Walker & Muliere (1997) and Walker & Damien (1998): prior specification with NTR processes
- The posterior characterizations are typically useful if one is interested in evaluating $\mathbb{E}[F(t)|\text{data}]$.

Some remarks

- Walker & Muliere (1997) and Walker & Damien (1998): prior specification with NTR processes
- The posterior characterizations are typically useful if one is interested in evaluating $\mathbb{E}[F(t)|\text{data}]$.
- More difficult is a full Bayesian analysis, apart from some cases.

Some remarks

- Walker & Muliere (1997) and Walker & Damien (1998): prior specification with NTR processes
- The posterior characterizations are typically useful if one is interested in evaluating $\mathbb{E}[F(t)|\text{data}]$.
- More difficult is a full Bayesian analysis, apart from some cases.
- Computational algorithms: Ferguson & Klass (1972), Damien, Laud & Smith (1995), Walker & Damien (1998, 2000), Wolpert & Ickstadt (1998)

Some remarks

- Walker & Muliere (1997) and Walker & Damien (1998): prior specification with NTR processes
- The posterior characterizations are typically useful if one is interested in evaluating $\mathbb{E}[F(t)|\text{data}]$.
- More difficult is a full Bayesian analysis, apart from some cases.
- Computational algorithms: Ferguson & Klass (1972), Damien, Laud & Smith (1995), Walker & Damien (1998, 2000), Wolpert & Ickstadt (1998)
- Connections with some recent results on combinatorics. See Gnedin & Pitman (2005)

Models for cumulative hazards

As an alternative approach to inferential problems in survival analysis, instead of specifying a prior for F one could focus on the hazard rate

$$\frac{\mathrm{d}F(x)}{1 - F(x-)} = \text{Pr}[\text{death at } (x, x + \mathrm{d}x] \mid \text{ survival time } \geq x]$$

Models for cumulative hazards

As an alternative approach to inferential problems in survival analysis, instead of specifying a prior for F one could focus on the hazard rate

$$\frac{\mathrm{d}F(x)}{1 - F(x-)} = \Pr[\text{death at } (x, x + \mathrm{d}x] \mid \text{ survival time } \geq x]$$

or on the cumulative hazard

$$A_x(F) = \int_0^x \frac{\mathrm{d}F(v)}{1 - F(v-)}$$

Models for cumulative hazards

As an alternative approach to inferential problems in survival analysis, instead of specifying a prior for F one could focus on the hazard rate

$$\frac{\mathrm{d}F(x)}{1-F(x-)} = \Pr[\text{death at } (x, x+\mathrm{d}x] \mid \text{ survival time } \geq x]$$

or on the cumulative hazard

$$A_x(F) = \int_0^x \frac{\mathrm{d}F(v)}{1-F(v-)}$$

Note that from the above relation one deduces

$$F(t) = 1 - \prod_{[0,t]} \{1 - \mathrm{d}A_x(F)\}$$

where $\prod_{[0,t]}$ is the product integral on $[0, t]$

The beta process (Hjort, 1990)

A suitable prior for A should be concentrated on those cumulative hazard functions for which

$$A_x - A_{x-} \leq 1 \qquad \forall x \geq 0$$

The beta process (Hjort, 1990)

A suitable prior for A should be concentrated on those cumulative hazard functions for which

$$A_x - A_{x-} \leq 1 \qquad \forall x \geq 0$$

Let

- A_0 be the baseline cumulative hazard with discontinuities at $x_1, x_2, \ldots$
- $\mathrm{d}A_{0,x_j}$ be the height of the jump of A_0 at x_j
- A_0^c continuous part of A_0
- $A_c = \{A_t^c : \ t \geq 0\}$ independent increments process with no fixed points of discontinuity and with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \mathbf{c(x)\, s^{-1}\, (1 - s)^{c(x)-1}\, ds}\, \mathrm{d}A_{0,c}(x) = \boldsymbol{\rho_x(ds)}\, \eta(\mathrm{d}x)$$

  for any $x \geq 0$ and $0 < s < 1$ and for some function $c(\,\cdot\,)$
- $S_1, S_2, \ldots$ are mutually independent, independent from A_c and

$$S_j \sim \mathrm{Beta}(c(x_j)\mathrm{d}A_{0,x_j},\ c(x_j)[1 - \mathrm{d}A_{0,x_j}])$$

Definition

The process A is a beta process with parameters c and A_0 if

$$A_x = A_x^c + \sum_{\{j:\, x_j \leq x\}} S_j$$

Definition

The process A is a beta process with parameters c and A_0 if

$$A_x = A_x^c + \sum_{\{j:\, x_j \leq x\}} S_j$$

In symbols $A \sim \mathrm{Beta}(c, A_0)$

Definition

The process A is a beta process with parameters c and A_0 if

$$A_x = A_x^c + \sum_{\{j:\, x_j \leq x\}} S_j$$

In symbols $A \sim \text{Beta}(c, A_0)$

Connection with NTR priors?

Definition

The process A is a beta process with parameters c and A_0 if

$$A_x = A_x^c + \sum_{\{j:\, x_j \le x\}} S_j$$

In symbols $A \sim \text{Beta}(c, A_0)$

Connection with NTR priors?

F is $\text{NTR}(Y) \iff A(F) = \{A_x(F) :\ x \geq 0\}$ has independent increments

Definition

The process A is a beta process with parameters c and A_0 if

$$A_x = A_x^c + \sum_{\{j:\, x_j \le x\}} S_j$$

In symbols $A \sim \mathrm{Beta}(c, A_0)$

Connection with NTR priors?

F is $\mathrm{NTR}(Y) \iff A(F) = \{A_x(F) : \ x \geq 0\}$ has independent increments

See Hjort (1990) and Dey, Erickson & Ramamoorthi (2003)

Conjugacy of the beta process (Hjort, 1990)

Let $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ be a sample of (possibly right censored) survival times.

Conjugacy of the beta process (Hjort, 1990)

Let $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ be a sample of (possibly right censored) survival times. Set

$$\bar{L}(x) = \sum_{\{i:\, \Delta_i = 1\}} \mathbb{I}_{(0,x]}(T_i) = \sharp\{ \text{ exact observations } \leq x\}$$

and recall that $\bar{N}(x) = \sum_{i=1}^{n} \mathbb{I}_{[x,\infty)}(T_i)$.

Conjugacy of the beta process (Hjort, 1990)

Let $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ be a sample of (possibly right censored) survival times. Set

$$\bar{L}(x) = \sum_{\{i\colon \Delta_i = 1\}} \mathbb{I}_{(0,x]}(T_i) = \sharp\{\text{ exact observations } \le x\}$$

and recall that $\bar{N}(x) = \sum_{i=1}^{n} \mathbb{I}_{[x,\infty)}(T_i)$. Hence

$$\mathrm{d}\bar{L}(x) = \sharp\{\text{ exact observations at } x\}$$

Conjugacy of the beta process (Hjort, 1990)

Let $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ be a sample of (possibly right censored) survival times. Set

$$\bar{L}(x) = \sum_{\{i:\, \Delta_i = 1\}} \mathbb{I}_{(0,x]}(T_i) = \sharp\{\text{ exact observations } \leq x\}$$

and recall that $\bar{N}(x) = \sum_{i=1}^{n} \mathbb{I}_{[x,\infty)}(T_i)$. Hence

$$\mathrm{d}\bar{L}(x) = \sharp\{\text{ exact observations at } x\}$$

Posterior distribution

Let $A \sim \text{Beta}(c, A_0)$. Then

$$A|\,\text{Data} \sim \text{Beta}\left(c + \bar{N}, \int \frac{\mathrm{d}\bar{L} + c\,\mathrm{d}A_0}{c + \bar{N}}\right)$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{ \text{ discontinuities of } A_0 \} \cup \{ \text{ exact observations } \}$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{ \text{ discontinuities of } A_0 \} \cup \{ \text{ exact observations } \}$$

then

$$A_x \,|\, \text{data } = A_{c,x}^* + \sum_{\{i:\, x_i^* \leq x,\, \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* : \, x \geq 0\}$ is an independent increments process with no fixed points of discontinuity

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{\text{ discontinuities of } A_0 \} \cup \{\text{ exact observations }\}$$

then

$$A_x \,|\, \text{data} \; = A_{c,x}^* + \sum_{\{i:\, x_i^* \leq x,\; \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* : \; x \geq 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \left[c(x) + \bar{N}(x)\right] \; s^{-1} \, (1-s)^{c(x)+\bar{N}(x)-1} \; \frac{c(x)\, \mathrm{d}A_{0,x}^c}{c(x) + \bar{N}(x)}$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{ \text{ discontinuities of } A_0 \} \cup \{ \text{ exact observations } \}$$

then

$$A_x \,|\, \text{data} \;= A_{c,x}^* + \sum_{\{i:\, x_i^* \leq x,\ \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* : \; x \geq 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \left[c(x) + \bar{N}(x)\right] \; s^{-1} \, (1-s)^{c(x)+\bar{N}(x)-1} \, \frac{c(x)\,\mathrm{d}A_{0,x}^c}{c(x) + \bar{N}(x)}$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{\text{ discontinuities of } A_0 \} \cup \{\text{ exact observations }\}$$

then

$$A_x \,|\, \text{data} \; = A_{c,x}^* + \sum_{\{i:\, x_i^* \le x,\; \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* \,:\, x \ge 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \left[c(x) + \bar{N}(x)\right] \, s^{-1} \, (1-s)^{c(x)+\bar{N}(x)-1} \, \frac{c(x) \, \mathrm{d}A_{0,x}^c}{c(x) + \bar{N}(x)}$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{\text{ discontinuities of } A_0 \} \cup \{ \text{ exact observations } \}$$

then

$$A_x \,|\, \text{data} \; = A_{c,x}^* + \sum_{\{i:\, x_i^* \leq x,\, \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* : \; x \geq 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \left[c(x) + \bar{N}(x)\right] \; s^{-1} (1-s)^{c(x)+\bar{N}(x)-1} \; \frac{c(x)\, \mathrm{d}A_{0,x}^c}{c(x) + \bar{N}(x)}$$

and, if $\mathrm{d}A_{0,x}^* = [c(x)\mathrm{d}A_{0,x} + \mathrm{d}\bar{L}(x)]/[c(x) + \bar{N}(x)]$, then

$$\mathrm{d}A_{x_i^*} \; \sim \; \textsc{Beta}\left([c(x_i^*) + \bar{N}(x_i^*)]\mathrm{d}A_{0,x_i^*}^* \; ; \; [c(x_i^*) + \bar{N}(x_i^*)] \left\{ 1 - \mathrm{d}A_{0,x_i^*}^* \right\} \right)$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{\text{ discontinuities of } A_0 \} \cup \{\text{ exact observations }\}$$

then

$$A_x \,|\, \text{data} \;= A_{c,x}^* + \sum_{\{i:\, x_i^* \le x,\, \Delta_i = 1\}} \mathrm{d}A_{x_i}^*$$

where $A_c^* = \{A_{c,x}^* : \; x \geq 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \left[c(x) + \bar{N}(x)\right] \, s^{-1} \, (1-s)^{c(x)+\bar{N}(x)-1} \, \frac{c(x)\, \mathrm{d}A_{0,x}^c}{c(x) + \bar{N}(x)}$$

and, if $\mathrm{d}A_{0,x}^* = [c(x)\mathrm{d}A_{0,x} + \mathrm{d}\bar{L}(x)]/[c(x) + \bar{N}(x)]$, then

$$\mathrm{d}A_{x_i^*} \;\sim\; \textsc{Beta}\left([c(x_i^*) + \bar{N}(x_i^*)]\mathrm{d}A_{0,x_i^*}^* \,;\, [c(x_i^*) + \bar{N}(x_i^*)] \left\{1 - \mathrm{d}A_{0,x_i^*}^*\right\}\right)$$

The previous statement implies that, if

$$\{x_1^*, x_2^*, \ldots\} = \{\text{ discontinuities of } A_0 \} \cup \{ \text{ exact observations } \}$$

then

$$A_x \,|\, \text{data} \; = A^*_{c,x} + \sum_{\{i:\, x_i^* \le x,\, \Delta_i = 1\}} \mathrm{d}A^*_{x_i}$$

where $A_c^* = \{A^*_{c,x} : \; x \geq 0\}$ is an independent increments process with no fixed points of discontinuity and Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \; \left[c(x) + \bar{N}(x)\right] \; s^{-1} \, (1-s)^{c(x)+\bar{N}(x)-1} \, \frac{c(x)\, \mathrm{d}A^c_{0,x}}{c(x) + \bar{N}(x)}$$

and, if $\mathrm{d}A^*_{0,x} = [c(x)\mathrm{d}A_{0,x} + \mathrm{d}\bar{L}(x)]/[c(x) + \bar{N}(x)]$, then

$$\mathrm{d}A_{x_i^*} \;\sim\; \textsc{Beta}\left([c(x_i^*) + \bar{N}(x_i^*)] \mathrm{d}A^*_{0,x_i^*} \; ; \; [c(x_i^*) + \bar{N}(x_i^*)] \left\{ 1 - \mathrm{d}A^*_{0,x_i^*} \right\} \right)$$

The Bayes estimators of A and F, under squared loss, are

$$\hat{A}_x = \int_{(0,x)} \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}}$$

and

$$\hat{F}(t) = 1 - \prod_{[0,t]} \left\{ 1 - \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}} \right\}$$

The Bayes estimators of A and F, under squared loss, are

$$\hat{A}_x = \int_{(0,x)} \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}}$$

and

$$\hat{F}(t) = 1 - \prod_{[0,t]} \left\{ 1 - \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}} \right\}$$

As $c \to 0$, they tend to the Nelson–Aalen and Kaplan–Meier estimator, respectively

The Bayes estimators of A and F, under squared loss, are

$$\hat{A}_x = \int_{(0,x)} \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}}$$

and

$$\hat{F}(t) = 1 - \prod_{[0,t]} \left\{ 1 - \frac{c\,\mathrm{d}A_0 + \mathrm{d}\bar{L}}{c + \bar{N}} \right\}$$

As $c \to 0$, they tend to the Nelson–Aalen and Kaplan–Meier estimator, respectively

Extensions of Hjort's work

- Discrete model with dependent hazard rates: Gamerman (1991), Gray (1994), Arias & Gasbarra (1994), Nieto–Barajas & Walker (2002)
- Inference on counting processes: Kim (1999)
- Spatial neutral to the right: James (2006)
- Connections with the Indian buffet process: Griffiths & Ghahramani (2006), Thibaux & Jordan (2007) and Teh, Görür & Ghahramani (2007). See Yee Whye's talk

Models for hazard rates

Instead of defining a prior for A one can set a prior for the hazard rate

$$\alpha(s) = \frac{F'(s)}{1 - F(s)}$$

If $\tilde{\mu}$ is a completely random measure on a complete and separable metric space $\mathbb{Y}$ and $k(\,\cdot\,|\,\cdot\,)$ is some kernel such that

$$\alpha(s) = \int_{\mathbb{Y}} k(s|y)\,\tilde{\mu}(\mathrm{d}y)$$

and $\lim_{T\to\infty} \int_0^T \alpha(s)\,\mathrm{d}s = \infty$

Models for hazard rates

Instead of defining a prior for A one can set a prior for the hazard rate

$$\alpha(s) = \frac{F'(s)}{1 - F(s)}$$

If $\tilde{\mu}$ is a completely random measure on a complete and separable metric space $\mathbb{Y}$ and $k(\cdot\,|\,\cdot)$ is some kernel such that

$$\alpha(s) = \int_{\mathbb{Y}} k(s|y)\, \tilde{\mu}(\mathrm{d}y)$$

and $\lim_{T\to\infty} \int_0^T \alpha(s)\, \mathrm{d}s = \infty$
Multiplicative intensity model: data from counting processes

$$\text{Intensity rate at } s \;=\; Z_{i,s}\, \alpha(s)$$

Models for hazard rates

Instead of defining a prior for A one can set a prior for the hazard rate

$$\alpha(s) = \frac{F'(s)}{1 - F(s)}$$

If $\tilde{\mu}$ is a completely random measure on a complete and separable metric space $\mathbb{Y}$ and $k(\cdot\,|\,\cdot)$ is some kernel such that

$$\alpha(s) = \int_{\mathbb{Y}} k(s|y)\, \tilde{\mu}(\mathrm{d}y)$$

and $\lim_{T\to\infty} \int_0^T \alpha(s)\,\mathrm{d}s = \infty$

Multiplicative intensity model: data from counting processes

$$\text{Intensity rate at } s \;=\; Z_{i,s}\,\alpha(s)$$

The factor $Z_{i,s}$ characterizes the censoring mechanism

$$Z_{i,s} = \mathbb{I}(T_i \geq s) \qquad \text{right censored data}$$

$$Z_{i,s} = \mathbb{I}(a_i < T_i \leq b_i) \qquad \text{left and right censored data}$$

Models for hazard rates

Instead of defining a prior for A one can set a prior for the hazard rate

$$\alpha(s) = \frac{F'(s)}{1 - F(s)}$$

If $\tilde{\mu}$ is a completely random measure on a complete and separable metric space $\mathbb{Y}$ and $k(\cdot\,|\,\cdot)$ is some kernel such that

$$\alpha(s) = \int_{\mathbb{Y}} k(s|y)\,\tilde{\mu}(\mathrm{d}y)$$

and $\lim_{T\to\infty} \int_0^T \alpha(s)\,\mathrm{d}s = \infty$

Multiplicative intensity model: data from counting processes

$$\text{Intensity rate at } s \;=\; Z_{i,s}\,\alpha(s)$$

The factor $Z_{i,s}$ characterizes the censoring mechanism

$$Z_{i,s} = \mathbb{I}(T_i \geq s) \qquad \text{right censored data}$$

$$Z_{i,s} = \mathbb{I}(a_i < T_i \leq b_i) \qquad \text{left and right censored data}$$

Different choices of $\tilde{\mu}$, $Z_{i,s}$ and of the kernel k yield different models used in the literature. Dykstra & Laud (1981) and Lo & Weng (1989).

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\, \eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\, \eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations
- $X_{m+1}, \ldots, X_n$ partially observed data

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations
- $X_{m+1}, \ldots, X_n$ partially observed data
- $Y_1, \ldots, Y_m$ latent random variables, corresponding to the exact observations, with $k \leq m$ distinct values $Y_1^*, \ldots, Y_k^*$

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\, \eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations
- $X_{m+1}, \ldots, X_n$ partially observed data
- $Y_1, \ldots, Y_m$ latent random variables, corresponding to the exact observations, with $k \leq m$ distinct values $Y_1^*, \ldots, Y_k^*$

$$\tilde{\mu} \,|\, (\text{ data, latent }) \;\stackrel{d}{=} \tilde{\mu}_n + \sum_{i=1}^{k} J_i\, \delta_{Y_i^*}$$

where

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\, \eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations
- $X_{m+1}, \ldots, X_n$ partially observed data
- $Y_1, \ldots, Y_m$ latent random variables, corresponding to the exact observations, with $k \leq m$ distinct values $Y_1^*, \ldots, Y_k^*$

$$\tilde{\mu} \,|\, (\text{ data, latent }) \;\overset{d}{=} \tilde{\mu}_n + \sum_{i=1}^{k} J_i\, \delta_{Y_i^*}$$

where

- $\tilde{\mu}_n$ completely random measure with no fixed points of discontinuity: Lévy measure easy to determine

Multiplicative intensity models governed by any completely random measure $\tilde{\mu}$ with Lévy measure

$$\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

- $X_1, \ldots, X_m$ exact observations
- $X_{m+1}, \ldots, X_n$ partially observed data
- $Y_1, \ldots, Y_m$ latent random variables, corresponding to the exact observations, with $k \leq m$ distinct values $Y_1^*, \ldots, Y_k^*$

$$\tilde{\mu} \,|\, (\text{ data, latent }) \;\stackrel{d}{=} \tilde{\mu}_n + \sum_{i=1}^{k} J_i\, \delta_{Y_i^*}$$

where

- $\tilde{\mu}_n$ completely random measure with no fixed points of discontinuity: Lévy measure easy to determine
- the jumps J_i's are independent, independent from $\tilde{\mu}_n$ and their densities depends on $\rho_{Y_i^*}$

This general result was obtained in James (2002, 2005).

This general result was obtained in James (2002, 2005).

Once the Lévy measure is specified, the posterior distribution of $\tilde{\mu}$, given the data and the latent variables, is completely known

This general result was obtained in James (2002, 2005).

Once the Lévy measure is specified, the posterior distribution of $\tilde{\mu}$, given the data and the latent variables, is completely known

Posterior distribution of $\tilde{\mu}$ given $X_1, \ldots, X_n$: integrate with respect to the distribution of $Y_1, \ldots, Y_m$ given $X_1, \ldots, X_n$

This general result was obtained in James (2002, 2005).

Once the Lévy measure is specified, the posterior distribution of $\tilde{\mu}$, given the data and the latent variables, is completely known

Posterior distribution of $\tilde{\mu}$ given $X_1, \ldots, X_n$: integrate with respect to the distribution of $Y_1, \ldots, Y_m$ given $X_1, \ldots, X_n$

Analytical difficulties: one needs to average with respect to all possible partitions of the set of integers $\{1, \ldots, m\}$

This general result was obtained in James (2002, 2005).

Once the Lévy measure is specified, the posterior distribution of $\tilde{\mu}$, given the data and the latent variables, is completely known

Posterior distribution of $\tilde{\mu}$ given $X_1, \ldots, X_n$: integrate with respect to the distribution of $Y_1, \ldots, Y_m$ given $X_1, \ldots, X_n$

Analytical difficulties: one needs to average with respect to all possible partitions of the set of integers $\{1, \ldots, m\}$

Full Bayesian analysis requires implementation of simulation scheme: Nieto–Barajas & Walker (2004), Ishwaran & James (2004) and James (2005)

Normalized random measures with independent increments

Normalized random measures with independent increments

- $\tilde{\mu}$ completely random measure on $\mathbb{X}$ with no fixed points of discontinuity

Normalized random measures with independent increments

- $\tilde{\mu}$ completely random measure on $\mathbb{X}$ with no fixed points of discontinuity
- $\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$ Lévy measure on $\mathbb{X} \times \mathbb{R}^+$

Normalized random measures with independent increments

- $\tilde{\mu}$ completely random measure on $\mathbb{X}$ with no fixed points of discontinuity
- $\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$ Lévy measure on $\mathbb{X} \times \mathbb{R}^+$

Suppose ν satisfies conditions such that

$$\mathbb{P}[\tilde{\mu}(\mathbb{X}) \notin \{0, \infty\}] = 1$$

Normalized random measures with independent increments

- $\tilde{\mu}$ completely random measure on $\mathbb{X}$ with no fixed points of discontinuity
- $\nu(\mathrm{d}x, \mathrm{d}s) = \rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$ Lévy measure on $\mathbb{X} \times \mathbb{R}^+$

Suppose ν satisfies conditions such that

$$\mathbb{P}[\tilde{\mu}(\mathbb{X}) \notin \{0, \infty\}] = 1$$

Definition

The random probability measures

$$\tilde{p} = \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

is a *normalized random measure with independent increments* (NRMI)

Use of NRMIs

- Ferguson (1973): Dirichlet process as a normalized gamma process
- Kingman (1975): normalized α–stable process for application to optimal storage
- Regazzini, L. & Prünster (2003): distributions of linear functionals of non–homogeneous NRMIs on $\mathbb{R}$
- James (2002): non–homogeneous NRMI on more general spaces
- James, L. & Prünster (2005): posterior analysis of non–homogeneous NRMIs
- Griffin (2007): use of NRMIs for constructing time dependent random probability measures

Examples

Let $\alpha \in (0, 1)$ and $\tau > 0$

Examples

Let $\alpha \in (0,1)$ and $\tau > 0$

(1) Let $\tilde{\mu}_\alpha$ be a completely random measure with

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{d}s\, \eta(\mathrm{d}x)$$

Then $\tilde{p} = \tilde{\mu}_\alpha / \tilde{\mu}_\alpha(\mathbb{X})$ is a *normalized α–stable process* with baseline measure η. See Kingman (1975)

Let $\alpha \in (0,1)$ and $\tau > 0$

(1) Let $\tilde{\mu}_\alpha$ be a completely random measure with

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{d}s\, \eta(\mathrm{d}x)$$

Then $\tilde{p} = \tilde{\mu}_\alpha / \tilde{\mu}_\alpha(\mathbb{X})$ is a *normalized* α*–stable process* with baseline measure η. See Kingman (1975)

(2) Let $\tilde{\mu}_{\alpha,\tau}$ be a completely random measure with

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{e}^{-\tau s} \mathrm{d}s\, \eta(\mathrm{d}x)$$

Brix (1999).

Examples

Let $\alpha \in (0,1)$ and $\tau > 0$

(1) Let $\tilde{\mu}_\alpha$ be a completely random measure with

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{d}s\, \eta(\mathrm{d}x)$$

Then $\tilde{p} = \tilde{\mu}_\alpha / \tilde{\mu}_\alpha(\mathbb{X})$ is a *normalized α–stable process* with baseline measure η. See Kingman (1975)

(2) Let $\tilde{\mu}_{\alpha,\tau}$ be a completely random measure with

$$\nu(\mathrm{d}x, \mathrm{d}s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{e}^{-\tau s} \mathrm{d}s\, \eta(\mathrm{d}x)$$

Brix (1999). The random probability measure $\tilde{p} = \tilde{\mu}_{\alpha,\tau} / \tilde{\mu}_{\alpha,\tau}(\mathbb{X})$ is a *generalized gamma process prior*. See Pitman (2003), James (2002) and L., Mena & Prünster (2007)

Inference with NRMIs

Suppose

$$\begin{aligned} X_i \,|\, \tilde{p} &\stackrel{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \ldots, n \\ \tilde{p} &= \tilde{\mu}/\tilde{\mu}(\mathbb{X}) \end{aligned}$$

What about

$$\tilde{p} \,|\, X_1, \ldots, X_n$$

Inference with NRMIs

Suppose

$$\begin{aligned} X_i \mid \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \ldots, n \\ \tilde{p} &= \tilde{\mu}/\tilde{\mu}(\mathbb{X}) \end{aligned}$$

What about

$$\tilde{p} \mid X_1, \ldots, X_n$$

- U_0 is a positive random variable with $U_0 \sim q_0$

Inference with NRMIs

Suppose

$$\begin{aligned} X_i \mid \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \ldots, n \\ \tilde{p} &= \tilde{\mu}/\tilde{\mu}(\mathbb{X}) \end{aligned}$$

What about

$$\tilde{p} \mid X_1, \ldots, X_n$$

- U_0 is a positive random variable with $U_0 \sim q_0$
- For any $n \geq 1$ U_n is a positive random variable such that

$$U_n | (X_1, \ldots X_n) \sim q_{n_1, \ldots, n_k}$$

Inference with NRMIs

Suppose

$$
\begin{aligned}
X_i \mid \tilde{p} &\stackrel{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \ldots, n \\
\tilde{p} &= \tilde{\mu}/\tilde{\mu}(\mathbb{X})
\end{aligned}
$$

What about

$$\tilde{p} \mid X_1, \ldots, X_n$$

- U_0 is a positive random variable with $U_0 \sim q_0$
- For any $n \geq 1$ U_n is a positive random variable such that

$$U_n | (X_1, \ldots X_n) \sim q_{n_1, \ldots, n_k}$$

- The density functions q_0 and $q_{n_1, \ldots, n_k}$ expressed in terms of the Lévy measure $\rho_x(\mathrm{d}s)\eta(\mathrm{d}x)$

Posterior distribution of a NRMI

Posterior distribution of a NRMI

Data $X_1, \ldots, X_n$ consist of k distinct values $X_1^*, \ldots, X_k^*$. Then

$$\tilde{\mu} \,|\, (U_n,\ \text{data}) \stackrel{d}{=} \tilde{\mu}_{U_n} + \sum_{i=1}^{k} J_i^{(U_n)} \, \delta_{X_i^*}$$

Posterior distribution of a NRMI

Data $X_1, \ldots, X_n$ consist of k distinct values $X_1^*, \ldots, X_k^*$. Then

$$\tilde{\mu} \,|\, (U_n,\ \text{data}) \stackrel{d}{=} \tilde{\mu}_{U_n} + \sum_{i=1}^{k} J_i^{(U_n)} \, \delta_{X_i^*}$$

where

(i) $\tilde{\mu}_u$ is a completely random measure with no fixed jumps and with Lévy measure

$$\nu^{(u)}(\mathrm{d}x, \mathrm{d}s) = \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s) \, \eta(\mathrm{d}x)$$

(ii) jump $J_i^{(u)}$ has density $f_i(s)\mathrm{d}s \propto s^{n_i} \mathrm{e}^{-us} \rho_{X_i^*}(\mathrm{d}s)$

(iii) $\tilde{\mu}_u$ and $J_i^{(u)}$ $(i = 1, \ldots, k)$ are independent

Posterior distribution of a NRMI

Data $X_1, \ldots, X_n$ consist of k distinct values $X_1^*, \ldots, X_k^*$. Then

$$\tilde{\mu} \,|\, (U_n,\ \text{data}) \stackrel{d}{=} \tilde{\mu}_{U_n} + \sum_{i=1}^{k} J_i^{(U_n)} \, \delta_{X_i^*}$$

where

(i) $\tilde{\mu}_u$ is a completely random measure with no fixed jumps and with Lévy measure

$$\nu^{(u)}(\mathrm{d}x, \mathrm{d}s) = \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s) \, \eta(\mathrm{d}x)$$

(ii) jump $J_i^{(u)}$ has density $f_i(s)\mathrm{d}s \propto s^{n_i} \mathrm{e}^{-us} \rho_{X_i^*}(\mathrm{d}s)$

(iii) $\tilde{\mu}_u$ and $J_i^{(u)}$ $(i = 1, \ldots, k)$ are independent

James, L. & Prünster (2005)

Posterior distribution of $\tilde{p}$

From the previous result one deduces

$$\tilde{p} \,|\, (U_n,\ \text{Data}) \quad \stackrel{d}{=} \quad \text{NRMI}$$

Posterior distribution of $\tilde{p}$

From the previous result one deduces

$$\tilde{p} \,|\, (U_n,\ \text{Data}) \quad \stackrel{d}{=} \quad \text{NRMI}$$

Predictive distribution

Set $P_0 = \eta/\eta(\mathbb{X})$. Then

$$\Pr[X_{n+1} \in \mathrm{d}x_{n+1} | X_1, \ldots, X_n] = w^{(n)}\, P_0(\mathrm{d}x_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_j^{(n)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where

$$w^{(n)} = \frac{1}{n} \int_0^{+\infty} u\, \tau_1(u|x_{n+1})\, q_{n_1,\ldots,n_k}(u)\, \mathrm{d}u$$

$$w_j^{(n)} = \int_0^{+\infty} u\, \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, q_{n_1,\ldots,n_k}(u)\, \mathrm{d}u$$

and $\tau_m(u|x) = \int_0^\infty s^m\, \mathrm{e}^{-us}\, \rho_x(\mathrm{d}s)$

Sampling conditional on U_n

Sampling conditional on U_n

Predictive

$$\Pr[X_{i+1} \in \mathrm{d}x_{i+1} | X_1, \ldots, X_i, U_i = u] = m(\mathrm{d}x_{i+1} | x_1, \ldots, x_i, u)$$

$$\propto \kappa_1(u)\, \tau_1(u|x_{i+1})\, P_0(\mathrm{d}x_{i+1}) + \sum_{j=1}^{k} \frac{\tau_{n_{j,i}+1}(u|X^*_{j,i})}{\tau_{n_{j,i}}(u|X^*_{j,i})}\, \delta_{X^*_{j,i}}(\mathrm{d}x_{i+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \eta(\mathrm{d}x)$

Sampling conditional on U_n

Predictive

$$\Pr[X_{i+1} \in \mathrm{d}x_{i+1} | X_1, \ldots, X_i, U_i = u] = m(\mathrm{d}x_{i+1} | x_1, \ldots, x_i, u)$$

$$\propto \kappa_1(u)\, \tau_1(u|x_{i+1})\, P_0(\mathrm{d}x_{i+1}) + \sum_{j=1}^{k} \frac{\tau_{n_{j,i}+1}(u|X^*_{j,i})}{\tau_{n_{j,i}}(u|X^*_{j,i})}\, \delta_{X^*_{j,i}}(\mathrm{d}x_{i+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \eta(\mathrm{d}x)$

Moreover

$$\Pr[X_1 \in \mathrm{d}x_1 \mid U_1 = u_1] = m(\mathrm{d}x_1 | u) \propto \tau_1(u|x_1)\, \eta(\mathrm{d}x_1)$$

(i) Generalization of Pólya urn scheme: one can generate a sample $X_1, \ldots, X_n$ from $\tilde{p}$

- Generate U_0 from q_0
- Generate X_1 from $m(\mathrm{d}x|U_0)$
- Generate U_1 from q_1
- Generate X_2 from $m(\mathrm{d}x_2|X_1, U_1)$
- ...

See James, L. & Prünster (2005)

(i) Generalization of Pólya urn scheme: one can generate a sample $X_1, \ldots, X_n$ from $\tilde{p}$

- Generate U_0 from q_0
- Generate X_1 from $m(\mathrm{d}x|U_0)$
- Generate U_1 from q_1
- Generate X_2 from $m(\mathrm{d}x_2|X_1, U_1)$
- ...

See James, L. & Prünster (2005)

(ii) Sample from the posterior random measure, given U_n and the data,

- Sample U_n from $q_{n_1,\ldots,n_k}$
- Sample $J_i^{(U_n)}$ from $f_i(s)\mathrm{d}s \propto s^{n_i}\,\mathrm{e}^{-U_n s}\,\rho_{X_i^*}(\mathrm{d}s)$
- Simulate $\tilde{\mu}_{U_n}$ via the Ferguson & Klass algorithm

See Luis' talk.

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)
Let (α, θ) be such that $\alpha \in (0, 1)$ and $\theta > -\alpha$

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)
Let (α, θ) be such that $\alpha \in (0, 1)$ and $\theta > -\alpha$

- $\mathbb{P}_{\boldsymbol{\alpha}}$ probability distribution of the α–stable random measure $\tilde{\mu}_\alpha$ with parameter measure $\eta(\,\cdot\,)$

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)
Let (α, θ) be such that $\alpha \in (0, 1)$ and $\theta > -\alpha$

- $\mathbb{P}_{\alpha}$ probability distribution of the α–stable random measure $\tilde{\mu}_{\alpha}$ with parameter measure $\eta(\,\cdot\,)$
- $\mathbb{P}_{(\alpha,\theta)} \ll \mathbb{P}_{\alpha}$ and such that

$$\frac{\mathbb{P}_{(\alpha,\theta)}(\mathrm{d}\mu)}{\mathbb{P}_{\alpha}(\mathrm{d}\mu)} = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)}\,\mu(\mathbb{X})^{-\theta}$$

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)
Let (α, θ) be such that $\alpha \in (0, 1)$ and $\theta > -\alpha$

- $\mathbb{P}_{\boldsymbol{\alpha}}$ probability distribution of the α–stable random measure $\tilde{\mu}_\alpha$ with parameter measure $\eta(\,\cdot\,)$
- $\mathbb{P}_{\boldsymbol{(\alpha,\theta)}} \ll \mathbb{P}_\alpha$ and such that

$$\frac{\mathbb{P}_{(\alpha,\theta)}(\mathrm{d}\mu)}{\mathbb{P}_\alpha(\mathrm{d}\mu)} = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)}\,\mu(\mathbb{X})^{-\theta}$$

- $\boldsymbol{\tilde{\mu}_{(\alpha,\theta)}}$ random measure with p.d. $\mathbb{P}_{(\alpha,\theta)}$: it does not have independent increments

Remark

An important random probability measure does not fit into this framework: two parameter Poisson–Dirichlet process. See, *e.g.*, Pitman & Yor (1997)
Let (α, θ) be such that $\alpha \in (0,1)$ and $\theta > -\alpha$

- $\mathbb{P}_{\alpha}$ probability distribution of the α–stable random measure $\tilde{\mu}_{\alpha}$ with parameter measure $\eta(\,\cdot\,)$
- $\mathbb{P}_{(\alpha,\theta)} \ll \mathbb{P}_{\alpha}$ and such that

$$\frac{\mathbb{P}_{(\alpha,\theta)}(\mathrm{d}\mu)}{\mathbb{P}_{\alpha}(\mathrm{d}\mu)} = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)}\,\mu(\mathbb{X})^{-\theta}$$

- $\tilde{\mu}_{(\alpha,\theta)}$ random measure with p.d. $\mathbb{P}_{(\alpha,\theta)}$: it does not have independent increments

The random probability measure

$$\tilde{\mathrm{p}}_{(\alpha,\theta)} = \frac{\tilde{\mu}_{(\alpha,\theta)}}{\tilde{\mu}_{(\alpha,\theta)}(\mathbb{X})}$$

is a two parameter Poisson–Dirichlet process

Poisson–Kingman models

Family of discrete random probability measures introduced by Pitman (2003)

Poisson–Kingman models

Family of discrete random probability measures introduced by Pitman (2003)

Let $J_1 \geq J_2 \geq \ldots \geq 0$ be the ranked points of a Poisson process $\tilde{\Pi}$ on $\mathbb{R}^+$ with Lévy measure ρ

Poisson–Kingman models

Family of discrete random probability measures introduced by Pitman (2003)

Let $J_1 \geq J_2 \geq \ldots \geq 0$ be the ranked points of a Poisson process $\tilde{\Pi}$ on $\mathbb{R}^+$ with Lévy measure ρ

- For any $I \subset \mathbb{R}^+$

$$\tilde{\Pi}(I) = \text{card}(\{J_1, J_2, \ldots, \} \cap I)$$

Poisson–Kingman models

Family of discrete random probability measures introduced by Pitman (2003)
Let $J_1 \geq J_2 \geq \ldots \geq 0$ be the ranked points of a Poisson process $\tilde{\Pi}$ on $\mathbb{R}^+$ with Lévy measure ρ

- For any $I \subset \mathbb{R}^+$

$$\tilde{\Pi}(I) = \text{card}(\{J_1, J_2, \ldots,\} \cap I)$$

- For any I_1 and I_2 such that $I_1 \cap I_2 = \varnothing$

$$\tilde{\Pi}(I_1) \perp \tilde{\Pi}(I_2)$$

Poisson–Kingman models

Family of discrete random probability measures introduced by Pitman (2003)
Let $J_1 \geq J_2 \geq \ldots \geq 0$ be the ranked points of a Poisson process $\tilde{\Pi}$ on $\mathbb{R}^+$ with Lévy measure ρ

- For any $I \subset \mathbb{R}^+$

$$\tilde{\Pi}(I) = \text{card}(\{J_1, J_2, \ldots, \} \cap I)$$

- For any I_1 and I_2 such that $I_1 \cap I_2 = \varnothing$

$$\tilde{\Pi}(I_1) \perp \tilde{\Pi}(I_2)$$

- For any I such that $\rho(I) < \infty$

$$\mathbb{P}[\tilde{\Pi}(I) = k] = \frac{\mathrm{e}^{-\rho(I)}\, \rho(I)^k}{k!} \qquad k = 0, 1, \ldots$$

It is assumed that the Lévy measure ρ satisfies suitable conditions that guarantee

$$\mathbb{P}\left[\sum_{i\geq 1} J_i < \infty\right] = 1$$

and we set

$$T = \sum_{i=1}^{\infty} J_i$$

It is assumed that the Lévy measure ρ satisfies suitable conditions that guarantee

$$\mathbb{P}\left[\sum_{i\geq 1} J_i < \infty\right] = 1$$

and we set

$$T = \sum_{i=1}^{\infty} J_i$$

It is also assumed that the distribution of T has density function f_T

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

and denote by $S^* = \{(p_1, p_2, \dots) : \ p_1 \geq p_2 \geq \ \cdots \ \geq 0, \ \sum_{i \geq 1} p_i = 1\}$ the set of all sequences of non–negative real numbers that sum up to 1.

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

and denote by $S^* = \{(p_1, p_2, \ldots) : \; p_1 \geq p_2 \geq \cdots \geq 0, \; \sum_{i \geq 1} p_i = 1\}$ the set of all sequences of non–negative real numbers that sum up to 1.

Definition

Let

$$p(\rho|t) \; = \; \text{conditional distribution of } (P_i)_{i\geq 1} \text{ on } S^*, \text{ given } T = t$$

$$\gamma \; = \; \text{probability distribution on } \mathbb{R}^+$$

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

and denote by $S^* = \{(p_1, p_2, \dots) : \; p_1 \geq p_2 \geq \cdots \geq 0, \; \sum_{i \geq 1} p_i = 1\}$ the set of all sequences of non–negative real numbers that sum up to 1.

Definition

Let

$$p(\rho|t) = \text{ conditional distribution of } (P_i)_{i\geq 1} \text{ on } S^*, \text{ given } T = t$$

$$\gamma = \text{ probability distribution on } \mathbb{R}^+$$

The distribution

$$\int_{\mathbb{R}^+} p(\rho|t)\, \gamma(\mathrm{d}t)$$

on S^* is said the *Poisson–Kingman distribution with Lévy density ρ and mixing distribution γ*.

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

and denote by $S^* = \{(p_1, p_2, \ldots) : \; p_1 \geq p_2 \geq \cdots \geq 0, \; \sum_{i \geq 1} p_i = 1\}$ the set of all sequences of non–negative real numbers that sum up to 1.

Definition

Let

$$p(\rho|t) = \text{ conditional distribution of } (P_i)_{i\geq 1} \text{ on } S^*, \text{ given } T = t$$

$$\gamma = \text{ probability distribution on } \mathbb{R}^+$$

The distribution

$$\int_{\mathbb{R}^+} p(\rho|t)\, \gamma(\mathrm{d}t)$$

on S^* is said the *Poisson–Kingman distribution with Lévy density* ρ *and mixing distribution* γ. It is denoted by $\mathrm{PK}(\rho, \gamma)$

Set

$$P_1 = \frac{J_1}{T} \geq P_2 = \frac{J_2}{T} \geq P_3 = \frac{J_3}{T} \geq \quad \cdots$$

and denote by $S^* = \{(p_1, p_2, \dots) : \; p_1 \geq p_2 \geq \cdots \geq 0, \; \sum_{i \geq 1} p_i = 1\}$ the set of all sequences of non–negative real numbers that sum up to 1.

Definition

Let

$$p(\rho|t) \; = \; \text{conditional distribution of } (P_i)_{i\geq 1} \text{ on } S^*, \text{ given } T = t$$

$$\gamma \; = \; \text{probability distribution on } \mathbb{R}^+$$

The distribution

$$\int_{\mathbb{R}^+} p(\rho|t)\, \gamma(\mathrm{d}t)$$

on S^* is said the *Poisson–Kingman distribution with Lévy density* ρ *and mixing distribution* γ. It is denoted by $\mathrm{PK}(\rho, \gamma)$

$\mathrm{PK}(\rho, \gamma)$ distribution = discrete nonparametric prior

Example: the two parameter Poisson–Dirichlet process

Let

Example: the two parameter Poisson–Dirichlet process

Let

- $\alpha \in (0, 1)$ and $\theta > -\alpha$

Example: the two parameter Poisson–Dirichlet process

Let

- $\alpha \in (0, 1)$ and $\theta > -\alpha$
- $\rho_\alpha(s) = s^{-1-\alpha}\mathbb{I}(s > 0)$ (Lévy measure of an α–stable process)

Example: the two parameter Poisson–Dirichlet process

Let

- $\alpha \in (0,1)$ and $\theta > -\alpha$
- $\rho_\alpha(s) = s^{-1-\alpha}\mathbb{I}(s > 0)$ (Lévy measure of an α–stable process)
- mixing measure

$$\gamma_{\alpha,\theta}(\mathrm{d}t) = \frac{\alpha\Gamma(\theta)}{\Gamma(\frac{\theta}{\alpha})}\, t^{-\theta}\, f_\alpha(t)\mathrm{d}t$$

where f_α is the α–stable density of T

Example: the two parameter Poisson–Dirichlet process

Let

- $\alpha \in (0,1)$ and $\theta > -\alpha$
- $\rho_\alpha(s) = s^{-1-\alpha}\mathbb{I}(s > 0)$ (Lévy measure of an α–stable process)
- mixing measure

$$\gamma_{\alpha,\theta}(\mathrm{d}t) = \frac{\alpha\Gamma(\theta)}{\Gamma(\frac{\theta}{\alpha})}\, t^{-\theta}\, f_\alpha(t)\mathrm{d}t$$

  where f_α is the α–stable density of T

The *Poisson–Dirichlet process with parameters* (α, θ) is the random probability measure with $\mathrm{PK}(\rho_\alpha, \gamma_{\alpha,\theta})$ distribution on S^*

Example: the two parameter Poisson–Dirichlet process

Let

- $\alpha \in (0,1)$ and $\theta > -\alpha$
- $\rho_\alpha(s) = s^{-1-\alpha}\mathbb{I}(s>0)$ (Lévy measure of an α–stable process)
- mixing measure

$$\gamma_{\alpha,\theta}(\mathrm{d}t) = \frac{\alpha\Gamma(\theta)}{\Gamma(\frac{\theta}{\alpha})}\, t^{-\theta}\, f_\alpha(t)\mathrm{d}t$$

  where f_α is the α–stable density of T

The *Poisson–Dirichlet process with parameters* (α,θ) is the random probability measure with $\mathrm{PK}(\rho_\alpha, \gamma_{\alpha,\theta})$ distribution on S^*

If $\alpha \to 0$, the $\mathrm{PD}(\alpha,\theta)$ process reduces to the one–parameter $\mathrm{PD}(\theta)$ process introduced by Kingman (1975)

Since a PK(ρ, γ) distribution selects (almost surely) discrete probability distributions, if

$$\begin{aligned} X_i \,|\, \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \\ \tilde{p} &\sim \text{PK}(\rho, \gamma) \end{aligned}$$

Since a PK(ρ, γ) distribution selects (almost surely) discrete probability distributions, if

$$\begin{aligned} X_i \mid \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \\ \tilde{p} &\sim \text{PK}(\rho, \gamma) \end{aligned}$$

then

$$\mathbb{P}[X_i = X_j] > 0 \qquad \forall i \neq j$$

and the n data $X_1, \ldots, X_n$ can be partitioned into

- $K_n = k$ classes such that two observations belong to the same class if and only if they coincide
- the K_n classes have frequencies $N_{1,n}, \ldots, N_{K_n,n}$ such that $\sum_{i=1}^{K_n} N_{i,n} = n$

For the PD(α, θ) process one has

$$\Pr(\text{data in } k \text{ classes with frequencies } n_1, \ldots, n_k)$$

$$= \Pr[K_n = k, N_{1,n} = n_1, \ldots, N_{k,n} = n_k]$$

$$= \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta+1)_{n-1}} \prod_{j=1}^{k} (1-\alpha)_{n_j - 1}$$

where $(a)_n = \Gamma(a+n)/\Gamma(a) = a(a+1)\cdots(a+n-1)$

For the PD(α, θ) process one has

$$\text{Pr(data in } k \text{ classes with frequencies } n_1, \ldots, n_k)$$

$$= \Pr[K_n = k, N_{1,n} = n_1, \ldots, N_{k,n} = n_k]$$

$$= \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta + 1)_{n-1}} \prod_{j=1}^{k} (1 - \alpha)_{n_j - 1}$$

where $(a)_n = \Gamma(a + n)/\Gamma(a) = a(a + 1) \cdots (a + n - 1)$

- The previous distribution is also termed *exchangeable partition probability function* (EPPF)

For the PD(α, θ) process one has

$$\begin{aligned}
&\Pr(\text{data in } k \text{ classes with frequencies } n_1, \ldots, n_k) \\
&\qquad = \Pr[K_n = k, N_{1,n} = n_1, \ldots, N_{k,n} = n_k] \\
&\qquad = \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta+1)_{n-1}} \prod_{j=1}^{k} (1-\alpha)_{n_j - 1}
\end{aligned}$$

where $(a)_n = \Gamma(a+n)/\Gamma(a) = a(a+1)\cdots(a+n-1)$

- The previous distribution is also termed *exchangeable partition probability function* (EPPF)
- As $\alpha \to 0$ one has

$$\frac{\theta^k}{(\theta)_n} \prod_{j=1}^{k} \Gamma(n_j)$$

Antoniak (1974)

If $X_1, \ldots, X_n$ take on k distinct values $X_1^*, \ldots, X_k^*$ and n_j of them are equal to X_j^*

If $X_1, \ldots, X_n$ take on k distinct values $X_1^*, \ldots, X_k^*$ and n_j of them are equal to X_j^*

$$\Pr[X_{n+1} = \text{ new} \mid X_1, \ldots, X_n] = \frac{\theta + k\alpha}{\theta + n}$$

If $X_1, \ldots, X_n$ take on k distinct values $X_1^*, \ldots, X_k^*$ and n_j of them are equal to X_j^*

$$\Pr[X_{n+1} = \text{ new} \mid X_1, \ldots, X_n] = \frac{\theta + k\alpha}{\theta + n}$$

and

$$\Pr[X_{n+1} = X_j^* \mid X_1, \ldots, X_n] = \frac{n_j - \alpha}{\theta + n}$$

Species sampling models (Pitman, 1996)

Suppose

Species sampling models (Pitman, 1996)

Suppose

- $(\tilde{p}_i)_{i \geq 1}$ is a sequence of non–negative random weights with $\sum_{i \geq 1} \tilde{p}_i \leq 1$

Species sampling models (Pitman, 1996)

Suppose

- $(\tilde{p}_i)_{i \geq 1}$ is a sequence of non–negative random weights with $\sum_{i \geq 1} \tilde{p}_i \leq 1$
- $(\hat{X}_i)_{i \geq 1}$ iid, independent from $(\tilde{p}_i)_{i \geq 1}$, with non–atomic distribution P_0

Species sampling models (Pitman, 1996)

Suppose

- $(\tilde{p}_i)_{i\geq 1}$ is a sequence of non–negative random weights with $\sum_{i\geq 1} \tilde{p}_i \leq 1$
- $(\hat{X}_i)_{i\geq 1}$ iid, independent from $(\tilde{p}_i)_{i\geq 1}$, with non–atomic distribution P_0

The random probability measure

$$\tilde{p} = \sum_{i\geq 1} \tilde{p}_i \, \delta_{\hat{X}_i} + (1 - \sum_{i\geq 1} \tilde{p}_i) \, P_0$$

is a *species sampling model*

Species sampling models (Pitman, 1996)

Suppose

- $(\tilde{p}_i)_{i\geq 1}$ is a sequence of non–negative random weights with $\sum_{i\geq 1} \tilde{p}_i \leq 1$
- $(\hat{X}_i)_{i\geq 1}$ iid, independent from $(\tilde{p}_i)_{i\geq 1}$, with non–atomic distribution P_0

The random probability measure

$$\tilde{p} = \sum_{i\geq 1} \tilde{p}_i \, \delta_{\hat{X}_i} + (1 - \sum_{i\geq 1} \tilde{p}_i) \, P_0$$

is a *species sampling model*

A sequence $(X_i)_{i\geq 1}$ such that

$$\begin{aligned} X_i \mid \tilde{p} \quad &\sim \quad \tilde{p} \qquad\qquad i = 1, 2, \ldots \\ \tilde{p} \quad &= \quad \text{species sampling model} \end{aligned}$$

is a *species sampling sequence*

We suppose $\sum_{i \geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

We suppose $\sum_{i\geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

We suppose $\sum_{i\geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

- $\tilde{p}_i$ proportion of the i–th species in the population

We suppose $\sum_{i \geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

- $\tilde{p}_i$ proportion of the i–th species in the population
- $\hat{X}_i$ label assigned to species i

We suppose $\sum_{i \geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

- $\tilde{p}_i$ proportion of the i–th species in the population
- $\hat{X}_i$ label assigned to species i

Sampling rule (Pitman, 1996)

Let $(\xi_i)_{i \geq 1}$ be i.i.d. from P_0.

We suppose $\sum_{i \geq 1} \tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

- $\tilde{p}_i$ proportion of the i–th species in the population
- $\hat{X}_i$ label assigned to species i

Sampling rule (Pitman, 1996)

Let $(\xi_i)_{i \geq 1}$ be i.i.d. from P_0. Then $(X_i)_{i \geq 1}$ is a species sampling sequence if and only if there exists a collection of weights

$$\{p_{j,n}(n_1, \ldots, n_k) : \, 1 \leq j \leq k, \, 1 \leq k \leq n, \, n \geq 1\}$$

such that

$$X_1 = \xi_1$$

$$X_{i+1} \,|\, (X_1, \ldots, X_i) = \begin{cases} \xi_{i+1} & \text{with prob } p_{k_i+1,i}(n_1, \ldots, n_{k_i}, 1) \\ X_j^* & \text{with prob } p_{k_i,i}(n_1, \ldots, n_j + 1, \ldots, n_{k_i}) \end{cases}$$

We suppose $\sum_{i\geq 1}\tilde{p}_i = 1$ almost surely, that is the model is *proper*

Interpretation of the weights

- $\tilde{p}_i$ proportion of the i–th species in the population
- $\hat{X}_i$ label assigned to species i

Sampling rule (Pitman, 1996)

Let $(\xi_i)_{i\geq 1}$ be i.i.d. from P_0. Then $(X_i)_{i\geq 1}$ is a species sampling sequence if and only if there exists a collection of weights

$$\{p_{j,n}(n_1,\ldots,n_k):\ 1\leq j\leq k,\ 1\leq k\leq n,\ n\geq 1\}$$

such that

$$X_1 = \xi_1$$

$$X_{i+1}\,|\,(X_1,\ldots,X_i) = \begin{cases} \xi_{i+1} & \text{with prob } p_{k_i+1,i}(n_1,\ldots,n_{k_i},1) \\ X_j^* & \text{with prob } p_{k_i,i}(n_1,\ldots,n_j+1,\ldots,n_{k_i}) \end{cases}$$

and k_i is the number of distinct $X_1^*,\ldots,X_{k_i}^*$ among $X_1,\ldots,X_i$

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

$(V_i)_{i\geq 1}$ sequence of independent random variables taking values in $[0, 1]$

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

$(V_i)_{i\geq 1}$ sequence of independent random variables taking values in $[0,1]$

$$\tilde{p}_1 = V_1 \qquad \tilde{p}_i = V_i \prod_{j=1}^{i-1}(1-V_j) \quad i \geq 2$$

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

$(V_i)_{i \geq 1}$ sequence of independent random variables taking values in $[0, 1]$

$$\tilde{p}_1 = V_1 \qquad \tilde{p}_i = V_i \prod_{j=1}^{i-1} (1 - V_j) \quad i \geq 2$$

These random weights define a proper species sampling model if an only if

$$\sum_{i \geq 1} \mathbb{E}[\log(1 - V_i)] = -\infty$$

See Ishwaran & James (2001)

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

$(V_i)_{i\geq 1}$ sequence of independent random variables taking values in $[0,1]$

$$\tilde{p}_1 = V_1 \qquad \tilde{p}_i = V_i \prod_{j=1}^{i-1}(1-V_j) \quad i \geq 2$$

These random weights define a proper species sampling model if an only if

$$\sum_{i\geq 1} \mathbb{E}[\log(1-V_i)] = -\infty$$

See Ishwaran & James (2001)

$$V_i \sim \text{beta}(1,\theta) \qquad \Longrightarrow \qquad \tilde{p} \sim \text{Dir}(\theta\, P_0)$$

Stick–breaking construction

Issue: specification of the random weights $\tilde{p}_i$

$(V_i)_{i\geq 1}$ sequence of independent random variables taking values in $[0,1]$

$$\tilde{p}_1 = V_1 \qquad \tilde{p}_i = V_i \prod_{j=1}^{i-1}(1 - V_j) \quad i \geq 2$$

These random weights define a proper species sampling model if an only if

$$\sum_{i\geq 1} \mathbb{E}[\log(1 - V_i)] = -\infty$$

See Ishwaran & James (2001)

$$V_i \sim \mathrm{beta}(1,\theta) \qquad \Longrightarrow \qquad \tilde{p} \sim \mathrm{Dir}(\theta\, P_0)$$

$$V_i \sim \mathrm{beta}(\theta + i\alpha, 1-\alpha) \qquad \Longrightarrow \qquad \tilde{p} \sim \mathrm{PD}(\alpha,\theta)$$

Generating a sample from PD(α, θ)

Generating a sample from PD(α, θ)

(1) Generate an i.i.d. sample $\xi_1, \ldots, \xi_n$ from P_0

Generating a sample from PD(α, θ)

(1) Generate an i.i.d. sample $\xi_1, \ldots, \xi_n$ from P_0

(2) $X_1 = \xi_1$

Generating a sample from PD(α, θ)

(1) Generate an i.i.d. sample $\xi_1, \ldots, \xi_n$ from P_0

(2) $X_1 = \xi_1$

(3) For $2 \leq i \leq n$

$$X_i \,|\, (X_1, \ldots, X_{i-1}) = \begin{cases} \xi_i & \text{prob } (\theta + \alpha k_{i-1})/(\theta + i - 1) \\ X^*_{j,i-1} & \text{prob } (n_{j,i-1} - \alpha)/(\theta + i - 1) \end{cases}$$

where

Generating a sample from PD(α, θ)

(1) Generate an i.i.d. sample $\xi_1, \ldots, \xi_n$ from P_0

(2) $X_1 = \xi_1$

(3) For $2 \leq i \leq n$

$$X_i \,|\, (X_1, \ldots, X_{i-1}) = \begin{cases} \xi_i & \text{prob } (\theta + \alpha k_{i-1})/(\theta + i - 1) \\ X^*_{j,i-1} & \text{prob } (n_{j,i-1} - \alpha)/(\theta + i - 1) \end{cases}$$

where

$$X^*_{j,i-1} = j\text{–th distinct species in } X_1, \ldots, X_{i-1}$$

$$n_{j,i-1} = \text{ number of } X_1, \ldots, X_{i-1} \text{ equal to } X^*_{j,i-1}$$

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$.

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

Posterior distribution of a $\mathrm{PD}(\alpha, \theta)$ model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \mathrm{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

- they take on k distinct values $X_1^*, \ldots, X_k^*$

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

- they take on k distinct values $X_1^*, \ldots, X_k^*$
- $n_j = \text{card}\{X_i : X_i = X_j^*\}$

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

- they take on k distinct values $X_1^*, \ldots, X_k^*$
- $n_j = \text{card}\{X_i : X_i = X_j^*\}$

Then

$$\tilde{p} \mid (X_1, \ldots, X_n) \overset{d}{=} (1 - \sum_{i=1}^{k} p_i^*)\, \tilde{p}^{(k)} + \sum_{j=1}^{k} p_j^*\, \delta_{X_j^*}$$

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \stackrel{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

- they take on k distinct values $X_1^*, \ldots, X_k^*$
- $n_j = \text{card}\{X_i : X_i = X_j^*\}$

Then

$$\tilde{p} \mid (X_1, \ldots, X_n) \stackrel{d}{=} (1 - \sum_{i=1}^{k} p_i^*)\, \tilde{p}^{(k)} + \sum_{j=1}^{k} p_j^*\, \delta_{X_j^*}$$

where $\tilde{p}^{(k)} = \text{PD}(\alpha, \theta + k\alpha)$ with $\mathbb{E}[\tilde{p}^{(k)}(A)] = P_0(A)$

Posterior distribution of a PD(α, θ) model (Pitman, 1996)

Suppose

$$X_i \mid \tilde{p} \overset{\text{iid}}{\sim} \tilde{p}$$

$$\tilde{p} = \text{PD}(\alpha, \theta)$$

with $\mathbb{E}[\tilde{p}(A)] = P_0(A)$. Let the observed data $X_1, \ldots, X_n$ be such that

- they take on k distinct values $X_1^*, \ldots, X_k^*$
- $n_j = \text{card}\{X_i : X_i = X_j^*\}$

Then

$$\tilde{p} \mid (X_1, \ldots, X_n) \overset{d}{=} (1 - \sum_{i=1}^{k} p_i^*)\, \tilde{p}^{(k)} + \sum_{j=1}^{k} p_j^*\, \delta_{X_j^*}$$

where $\tilde{p}^{(k)} = \text{PD}(\alpha, \theta + k\alpha)$ with $\mathbb{E}[\tilde{p}^{(k)}(A)] = P_0(A)$ and

$$(p_1^*, \ldots, p_k^*) \sim \text{Dir}(n_1 - \alpha, \ldots, n_k - \alpha, \theta + k\alpha)$$

Other descriptions of the posterior

Other descriptions of the posterior

- Ishwaran & James (2003): species sampling models with stick–breaking construction

Other descriptions of the posterior

- Ishwaran & James (2003): species sampling models with stick–breaking construction
- James, L. & Prünster (2005): species sampling models with weights obtained by normalizing jumps of a completely random measure

Density estimation: nonparametric mixture models

Continuous data

Density estimation: nonparametric mixture models

Continuous data

- $Y_1, \ldots, Y_n$ taking values in $\mathbb{Y}$ (observables)

Density estimation: nonparametric mixture models

Continuous data

- $Y_1, \ldots, Y_n$ taking values in $\mathbb{Y}$ (observables)
- $X_1, \ldots, X_n$ taking values in $\mathbb{X}$ (latent)

Density estimation: nonparametric mixture models

Continuous data

- $Y_1, \ldots, Y_n$ taking values in $\mathbb{Y}$ (observables)
- $X_1, \ldots, X_n$ taking values in $\mathbb{X}$ (latent)

Nonparametric hierarchical model

$$
\begin{aligned}
Y_i | X_i &\overset{\text{ind}}{\sim} f(Y_i | X_i) \qquad i = 1, \ldots, n \\
X_i | \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \ldots, n \\
\tilde{p} &= \text{discrete random probability measure}
\end{aligned}
$$

Density estimation: nonparametric mixture models

Continuous data

- $Y_1, \dots, Y_n$ taking values in $\mathbb{Y}$ (observables)
- $X_1, \dots, X_n$ taking values in $\mathbb{X}$ (latent)

<u>Nonparametric hierarchical model</u>

$$\begin{aligned} Y_i | X_i &\overset{\text{ind}}{\sim} f(Y_i | X_i) \qquad i = 1, \dots, n \\ X_i | \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \qquad i = 1, \dots, n \\ \tilde{p} &= \text{discrete random probability measure} \end{aligned}$$

Equivalently

$$\begin{aligned} Y_i | \tilde{f} &\overset{\text{iid}}{\sim} \tilde{f} \qquad i = 1, \dots, n \\ \tilde{f}(\,\cdot\,) &= \int_{\mathbb{X}} f(\,\cdot\,|x)\, \tilde{p}(\mathrm{d}x) \end{aligned}$$

Clustering of the observations

Clustering of the observations

$\tilde{p}$ discrete $\quad \Longrightarrow \quad$ ties among the latent X_i

Clustering of the observations

$$\tilde{p} \text{ discrete} \qquad \Longrightarrow \qquad \text{ties among the latent } X_i$$

Recall that $X_1^*, \ldots, X_k^*$ denote the k distinct values among the $X_1, \ldots, X_n$ and

$$I_j = \{i : \, X_i = X_j^*\} \qquad j = 1, \ldots, k$$

Clustering of the observations

$$\tilde{p} \text{ discrete} \qquad \Longrightarrow \qquad \text{ties among the latent } X_i$$

Recall that $X_1^*, \dots, X_k^*$ denote the k distinct values among the $X_1, \dots, X_n$ and

$$I_j = \{i : \, X_i = X_j^*\} \qquad j = 1, \dots, k$$

The I_j's describe a grouping of $X_1, \dots, X_n$ such that

$$i, l \in I_j \qquad \Leftrightarrow \qquad X_i = X_l$$

Clustering of the observations

$$\tilde{p} \text{ discrete} \qquad \Longrightarrow \qquad \text{ties among the latent } X_i$$

Recall that $X_1^*, \ldots, X_k^*$ denote the k distinct values among the $X_1, \ldots, X_n$ and

$$I_j = \{i : \; X_i = X_j^*\} \qquad j = 1, \ldots, k$$

The I_j's describe a grouping of $X_1, \ldots, X_n$ such that

$$i, l \in I_j \qquad \Leftrightarrow \qquad X_i = X_l$$

Number of distinct X_i's

$\Downarrow$

Number K_n of clusters for the observations Y_i

- Evaluate $\mathbb{E}[\tilde{f} \mid Y_1, \ldots, Y_n]$
- Posterior distribution of the number of clusters K_n, given $Y_1, \ldots, Y_n$

- Evaluate $\mathbb{E}[\tilde{f} \mid Y_1, \ldots, Y_n]$
- Posterior distribution of the number of clusters K_n, given $Y_1, \ldots, Y_n$

Let

$\mathscr{P}_{n,k}$ set of all partitions $\mathbf{p}$ of $\{1, \ldots, n\}$ into k sets

$\mathbf{Y}^{(n)}$ vector $(Y_1, \ldots, Y_n)$

and

$$\mathscr{L}(\mathrm{d}x_1^*, \ldots, \mathrm{d}x_k^* \mid \mathbf{Y}^{(n)}, \mathbf{p}) = \text{ posterior distribution of the latent}$$

A posterior estimate of $f(y)$

$$\mathbb{E}\left[\tilde{f}(y) \mid \mathbf{Y}^{(n)}\right] = \sum_{k=1}^{n} \sum_{\mathbf{p} \in \mathscr{P}_{n,k}} \int_{\mathbb{X}^k} \left\{ \int_{\mathbb{X}} f(y|x)\, \mathbb{E}\left[\tilde{P}(\mathrm{d}x) | x_1^*, \ldots, x_k^*\right] \right\}$$

$$\times \; \mathscr{L}(\mathrm{d}x_1^*, \ldots, \mathrm{d}x_k^* \mid \mathbf{Y}^{(n)}, \mathbf{p})$$

A posterior estimate of $f(y)$

$$\mathbb{E}\left[\tilde{f}(y) \mid \mathbf{Y}^{(n)}\right] = \sum_{k=1}^{n} \sum_{\mathbf{p}\in\mathscr{P}_{n,k}} \int_{\mathbb{X}^k} \left\{ \int_{\mathbb{X}} f(y|x)\, \mathbb{E}\left[\tilde{P}(\mathrm{d}x)|x_1^*,\ldots,x_k^*\right] \right\} \times \mathscr{L}(\mathrm{d}x_1^*,\ldots,\mathrm{d}x_k^* \mid \mathbf{Y}^{(n)},\mathbf{p})$$

$\mathbb{E}\left[\tilde{P}(\mathrm{d}x)|x_1^*,\ldots,x_k^*\right]$ predictive distribution

$\sum_{\mathbf{p}\in\mathscr{P}_{\mathbf{n},\mathbf{k}}}$ hard to evaluate

A posterior estimate of $f(y)$

$$\mathbb{E}\left[\tilde{f}(y) \,|\, \mathbf{Y}^{(n)}\right] = \sum_{k=1}^{n} \sum_{\mathbf{p}\in\mathscr{P}_{n,k}} \int_{\mathbb{X}^k} \left\{ \int_{\mathbb{X}} f(y|x)\, \mathbb{E}\left[\tilde{P}(\mathrm{d}x)|x_1^*, \ldots, x_k^*\right] \right\} \times \mathscr{L}(\mathrm{d}x_1^*, \ldots, \mathrm{d}x_k^* \,|\, \mathbf{Y}^{(n)}, \mathbf{p})$$

$\mathbb{E}\left[\tilde{P}(\mathrm{d}x)|x_1^*, \ldots, x_k^*\right]$ predictive distribution

$\sum_{\mathbf{p}\in\mathscr{P}_{\mathbf{n},\mathbf{k}}}$ hard to evaluate

The posterior distribution of K_n cannot be evaluated exactly, as well

MCMC algorithm

Suppose

$$\tilde{p} = \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

where $\tilde{\mu}$ has intensity $\nu(\mathrm{d}x, \mathrm{d}s) = \rho(\mathrm{d}s)\,\eta(\mathrm{d}x)$.

MCMC algorithm

Suppose

$$\tilde{p} = \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

where $\tilde{\mu}$ has intensity $\nu(\mathrm{d}x, \mathrm{d}s) = \rho(\mathrm{d}s)\,\eta(\mathrm{d}x)$.

In order $X_1, \ldots, X_n$ from the posterior $\mathscr{L}(\,\cdot\,|\mathbf{Y}^{(n)})$ one makes use of

$$\Pr\left[X_i \in \mathrm{d}x_i \,\middle|\, \mathbf{X}_{-i}^{(n)}, \mathbf{Y}^{(n)}\right] = q_{i,0}^*\, P_0(\mathrm{d}x_i)\, f(Y_i|x_i) + \sum_{j=1}^{k_{i,n-1}} q_{i,j}^*\, \delta_{X_{i,j}^*}(\mathrm{d}x_i)$$

MCMC algorithm

Suppose

$$\tilde{p} = \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})}$$

where $\tilde{\mu}$ has intensity $\nu(\mathrm{d}x, \mathrm{d}s) = \rho(\mathrm{d}s)\,\eta(\mathrm{d}x)$.

In order $X_1, \ldots, X_n$ from the posterior $\mathscr{L}(\,\cdot\,|\mathbf{Y}^{(n)})$ one makes use of

$$\Pr\left[X_i \in \mathrm{d}x_i \middle| \mathbf{X}_{-i}^{(n)}, \mathbf{Y}^{(n)}\right] = q_{i,0}^*\, P_0(\mathrm{d}x_i)\, f(Y_i|x_i) + \sum_{j=1}^{k_{i,n-1}} q_{i,j}^*\, \delta_{X_{i,j}^*}(\mathrm{d}x_i)$$

where

- $\mathbf{X}_{-i}^{(n)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$
- $k_{i,n-1}$ = number of distinct observations, $X_{i,j}^*$, in $\mathbf{X}_{-i}^{(n)}$

Weights

$$q^*_{i,0} \propto w_i^{(n-1)} \int_{\mathbb{X}} f(Y_i|x)\, P_0(\mathrm{d}x) \qquad q^*_{i,j} \propto w_{i,j}^{(n-1)}\, f(Y_i|X^*_{i,j})$$

such that $\sum_{j=0}^{k_{i,n-1}} q^*_{i,j} = 1$. Moreover,

$$w_{i,0}^{(n-1)} = \frac{1}{n-1} \int_0^\infty u\, \tau_1(u)\, q_{n_1,\dots,n_{k_{i,n-1}}}(u)\, \mathrm{d}u$$

$$w_{i,j}^{(n-1)} = \int_0^\infty u\, \frac{\tau_{n_j+1}(u)}{\tau_{n_j}(u)}\, q_{n_1,\dots,n_{k_{i,n-1}}}(u)\, \mathrm{d}u$$

Weights

$$q^*_{i,0} \propto w_i^{(n-1)} \int_{\mathbb{X}} f(Y_i|x)\, P_0(\mathrm{d}x) \qquad q^*_{i,j} \propto w_{i,j}^{(n-1)}\, f(Y_i|X^*_{i,j})$$

such that $\sum_{j=0}^{k_{i,n-1}} q^*_{i,j} = 1$. Moreover,

$$w_{i,0}^{(n-1)} = \frac{1}{n-1} \int_0^\infty u\, \tau_1(u)\, q_{n_1,\dots,n_{k_{i,n-1}}}(u)\, \mathrm{d}u$$

$$w_{i,j}^{(n-1)} = \int_0^\infty u\, \frac{\tau_{n_j+1}(u)}{\tau_{n_j}(u)}\, q_{n_1,\dots,n_{k_{i,n-1}}}(u)\, \mathrm{d}u$$

Non–homogeneous case: the weights depend on the $X^*_{i,j}$ as well

(1) Generate starting values $X_{i,0}$ from P_0, $(i = 1, \ldots, n)$

(1) Generate starting values $X_{i,0}$ from P_0, $(i = 1, \ldots, n)$

(2) At iteration $t \geq 1$ generate $\mathbf{X}_t^{(n)} = (X_{1,t}, \ldots, X_{n,t})$ as follows

$$
\begin{aligned}
X_{1,t} &\sim \Pr\left[X_1 \,\middle|\, X_{2,t-1}, \ldots, X_{n,t-1}, \mathbf{Y}^{(n)}\right] \\
X_{2,t} &\sim \Pr\left[X_2 \,\middle|\, X_{1,t}, X_{3,t-1}, \ldots, X_{n,t-1}, \mathbf{Y}^{(n)}\right] \\
\vdots \quad & \quad \vdots \\
X_{n,t} &\sim \Pr\left[X_n \,\middle|\, X_{1,t}, \ldots, X_{n-1,t}, \mathbf{Y}^{(n)}\right]
\end{aligned}
$$

(1) Generate starting values $X_{i,0}$ from P_0, $(i = 1, \ldots, n)$

(2) At iteration $t \geq 1$ generate $\mathbf{X}_t^{(n)} = (X_{1,t}, \ldots, X_{n,t})$ as follows

$$X_{1,t} \sim \Pr\left[X_1 \,\middle|\, X_{2,t-1}, \ldots, X_{n,t-1}, \mathbf{Y}^{(n)}\right]$$

$$X_{2,t} \sim \Pr\left[X_2 \,\middle|\, X_{1,t}, X_{3,t-1}, \ldots, X_{n,t-1}, \mathbf{Y}^{(n)}\right]$$

$$\vdots \qquad \vdots$$

$$X_{n,t} \sim \Pr\left[X_n \middle| X_{1,t}, \ldots, X_{n-1,t}, \mathbf{Y}^{(n)}\right]$$

(3) Go to (2)

Output of the algorithm is used to evaluate

Output of the algorithm is used to evaluate

(I) A posterior estimate of f

$$\hat{f}(y) = \frac{1}{N} \sum_{t=1}^{N} \int_{\mathbb{X}} f(y|x)\, \mathbb{E}\left[\tilde{P}(\mathrm{d}x) \,|\, X^*_{1,t}, \dots, X^*_{k_t,t}\right]$$

Output of the algorithm is used to evaluate

(I) A posterior estimate of f

$$\hat{f}(y) = \frac{1}{N} \sum_{t=1}^{N} \int_{\mathbb{X}} f(y|x)\, \mathbb{E}\left[\tilde{P}(\mathrm{d}x) \,|\, X^*_{1,t}, \ldots, X^*_{k_t,t}\right]$$

(II) Posterior of the number of clusters K_n

$$\mathbb{P}\left[K_n = k \,|\, \mathbf{Y}^{(n)}\right] \approx \frac{1}{N} \sum_{t=1}^{N} \mathbb{I}\{k^{(t)} = k\}$$

where $k^{(t)}$ is the number of distinct values among $X_{1,t}, \ldots, X_{n,t}$ generated at iteration t.

MCMC algorithm inspired by the work of Escobar (1988,1994) and Escobar & West (1995) for the MDP.

MCMC algorithm inspired by the work of Escobar (1988,1994) and Escobar & West (1995) for the MDP.

- Additional acceleration step: slow mixing if

$$q^*_{i,j} \gg q^*_{i,0}$$

Iteration t with $k^{(t)}$ distinct components and, then, sample $X^*_{1,t}, \ldots, X^*_{k^{(t)},t}$

$$\pi(x_1, \ldots, x_{k^{(t)}} \,|\, \mathbf{Y}, k^{(t)}) \propto \prod_{j=1}^{k^{(t)}} P_0(\mathrm{d}x_j) \prod_{i \in I_j} f(y_i | x_j)$$

See MacEachern (1994,1998) and Ishwaran & James (2001).

MCMC algorithm inspired by the work of Escobar (1988,1994) and Escobar & West (1995) for the MDP.

- Additional acceleration step: slow mixing if

$$q^*_{i,j} \gg q^*_{i,0}$$

  Iteration t with $k^{(t)}$ distinct components and, then, sample $X^*_{1,t}, \ldots, X^*_{k^{(t)},t}$

$$\pi(x_1, \ldots, x_{k^{(t)}} \,|\, \mathbf{Y}, k^{(t)}) \propto \prod_{j=1}^{k^{(t)}} P_0(\mathrm{d}x_j) \prod_{i \in I_j} f(y_i|x_j)$$

  See MacEachern (1994,1998) and Ishwaran & James (2001).

- Non–conjugate models: $\int_{\mathbb{X}} f(Y_i|x)\, P_0(\mathrm{d}x)$ and, then, $q^*_{i,0}$ are difficult to evaluate. See MacEachern & Müller (1998) and Neal (2000).

MCMC algorithm inspired by the work of Escobar (1988,1994) and Escobar & West (1995) for the MDP.

- Additional acceleration step: slow mixing if

$$q^*_{i,j} \gg q^*_{i,0}$$

  Iteration t with $k^{(t)}$ distinct components and, then, sample $X^*_{1,t}, \ldots, X^*_{k^{(t)},t}$

$$\pi(x_1, \ldots, x_{k^{(t)}} \,|\, \mathbf{Y}, k^{(t)}) \propto \prod_{j=1}^{k^{(t)}} P_0(\mathrm{d}x_j) \prod_{i \in I_j} f(y_i|x_j)$$

  See MacEachern (1994,1998) and Ishwaran & James (2001).

- Non–conjugate models: $\int_{\mathbb{X}} f(Y_i|x)\, P_0(\mathrm{d}x)$ and, then, $q^*_{i,0}$ are difficult to evaluate. See MacEachern & Müller (1998) and Neal (2000).
- Algorithm for stick–breaking priors: Ishwaran & James (2001,2003), Papaspiliopoulos & Roberts (2004) and Walker (2006)

Important: knowledge of

$$\Pr[X_{n+1} \in \cdot \mid X_1, \ldots, X_n]$$

Important: knowledge of

$$\Pr[X_{n+1} \in \cdot \mid X_1, \ldots, X_n]$$

Mixture of the Dirichlet process: $\tilde{p}$ a Dirichlet process with baseline measure $\theta\, P_0$

$$\tilde{f}(y) = \int_{\mathbb{X}} f(y|x)\, \tilde{p}(\mathrm{d}x)$$

See Lo (1984).

Important: knowledge of

$$\Pr[X_{n+1} \in \cdot \,|\, X_1, \ldots, X_n]$$

Mixture of the Dirichlet process: $\tilde{p}$ a Dirichlet process with baseline measure $\theta\, P_0$

$$\tilde{f}(y) = \int_{\mathbb{X}} f(y|x)\, \tilde{p}(\mathrm{d}x)$$

See Lo (1984).

Mixture of more general discrete priors: one can take $\tilde{p}$ not to be the Dirichlet process. For example

$$\tilde{p} = \text{ normalized random measure}$$

$$\tilde{p} = \text{ stick–breaking prior}$$

Why using a mixing measure different from the Dirichlet process?

Why using a mixing measure different from the Dirichlet process? There are discrete random probabilities that give rise to completely different clustering structures

Why using a mixing measure different from the Dirichlet process? There are discrete random probabilities that give rise to completely different clustering structures

- If $\tilde{p}$ is the Dirichlet process with baseline measure $\theta\, P_0$, then

$$\lim_{n\to\infty} \frac{K_n}{\log n} \to \theta$$

Why using a mixing measure different from the Dirichlet process? There are discrete random probabilities that give rise to completely different clustering structures

- If $\tilde{p}$ is the Dirichlet process with baseline measure $\theta\, P_0$, then

$$\lim_{n\to\infty} \frac{K_n}{\log n} \to \theta$$

- If $\tilde{p}$ is a PD(α, θ) or a generalized gamma process with parameters (α, τ) then

$$\lim_{n\to\infty} \frac{K_n}{n^\alpha} = \text{ positive random variable}$$

Why using a mixing measure different from the Dirichlet process? There are discrete random probabilities that give rise to completely different clustering structures

- If $\tilde{p}$ is the Dirichlet process with baseline measure θP_0, then

$$\lim_{n \to \infty} \frac{K_n}{\log n} \to \theta$$

- If $\tilde{p}$ is a PD(α, θ) or a generalized gamma process with parameters (α, τ) then

$$\lim_{n \to \infty} \frac{K_n}{n^{\alpha}} = \text{ positive random variable}$$

- The value of α in a PD(α, θ) or in a generalized gamma process strongly influences the clustering structure

Why using a mixing measure different from the Dirichlet process? There are discrete random probabilities that give rise to completely different clustering structures

- If $\tilde{p}$ is the Dirichlet process with baseline measure $\theta\, P_0$, then

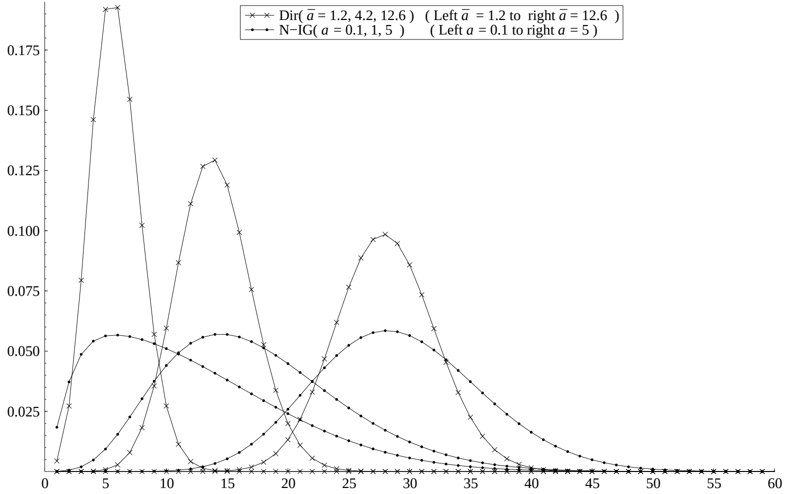
$$\lim_{n\to\infty} \frac{K_n}{\log n} \to \theta$$

- If $\tilde{p}$ is a PD(α, θ) or a generalized gamma process with parameters (α, τ) then

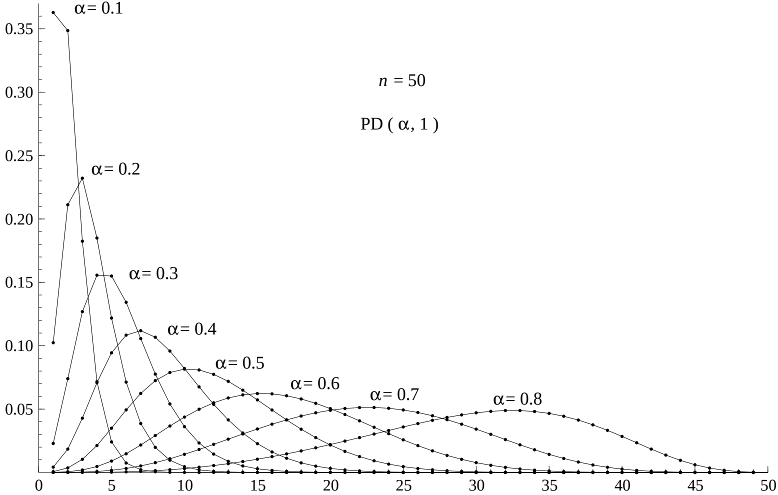
$$\lim_{n\to\infty} \frac{K_n}{n^{\alpha}} = \text{ positive random variable}$$

- The value of α in a PD(α, θ) or in a generalized gamma process strongly influences the clustering structure

For mixtures with more general mixing random probabilities see Ishwaran & James (2001, 2003), L., Mena & Prünster (2005, 2007) and, for an application to AFT models, see Argiento, Guglielmi & Pievatolo (2007).



Dir($\bar{a}$ = 1.2, 4.2, 12.6)   (Left $\bar{a}$ = 1.2 to right $\bar{a}$ = 12.6)
N−IG(a = 0.1, 1, 5)   (Left a = 0.1 to right a = 5)
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0
5
10
15
20
25
30
35
40
45
50
55
60



α= 0.1
α= 0.2
α= 0.3
α= 0.4
α= 0.5
α= 0.6
α= 0.7
α= 0.8
n = 50
PD (α, 1)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0
5
10
15
20
25
30
35
40
45
50

Density estimation: random Bernstein polynomial (Petrone, 1999)

Estimation of a density function on a bounded interval $[a, b]$

Density estimation: random Bernstein polynomial (Petrone, 1999)

Estimation of a density function on a bounded interval $[a, b]$

Take $[a, b] = [0, 1]$ and define

$$B(x; k, F) = \sum_{j=1}^{k} F(j/k) \binom{k}{j} x^j (1 - x)^{k-j} \qquad x \in [0, 1]$$

Density estimation: random Bernstein polynomial (Petrone, 1999)

Estimation of a density function on a bounded interval $[a, b]$

Take $[a, b] = [0, 1]$ and define

$$B(x; k, F) = \sum_{j=1}^{k} F(j/k) \binom{k}{j} x^j (1-x)^{k-j} \qquad x \in [0, 1]$$

Case where

- $F : [0, 1] \to \mathbb{R}$ is a cumulative distribution function with density f
- $w_{j,k} = F(j/k) - F((j-1)/k)$
- $\mathbf{w}_k = (w_{1,k}, \ldots, w_{k,k})$

Density estimation: random Bernstein polynomial (Petrone, 1999)

Estimation of a density function on a bounded interval $[a, b]$

Take $[a, b] = [0, 1]$ and define

$$B(x; k, F) = \sum_{j=1}^{k} F(j/k) \binom{k}{j} x^j (1-x)^{k-j} \qquad x \in [0, 1]$$

Case where

- $F : [0, 1] \rightarrow \mathbb{R}$ is a cumulative distribution function with density f
- $w_{j,k} = F(j/k) - F((j-1)/k)$
- $\mathbf{w}_k = (w_{1,k}, \ldots, w_{k,k})$

$$\tilde{f}_k(x; \mathbf{w}_k) = \sum_{j=1}^{k} w_{j,k} \text{Beta}(x; j, k-j+1)$$

$$\lim_{k\to\infty} \tilde{f}_k(x;\mathbf{w}_k) = f(x) \qquad \forall x \in [0,1]$$

$$\lim_{k \to \infty} \tilde{f}_k(x; \mathbf{w}_k) = f(x) \qquad \forall x \in [0,1]$$

Prior on $(K, \mathbf{W}_k)$

$$(W_{1,K}, \ldots, W_{K,K}) | K \sim \mathrm{Dir}(\alpha_{1,K}, \ldots, \alpha_{K,K})$$
$$K \sim \pi$$

$$\lim_{k\to\infty} \tilde{f}_k(x; \mathbf{w}_k) = f(x) \qquad \forall x \in [0,1]$$

Prior on $(K, \mathbf{W}_k)$

$$(W_{1,K}, \ldots, W_{K,K})|K \sim \mathrm{Dir}(\alpha_{1,K}, \ldots, \alpha_{K,K})$$
$$K \sim \pi$$

Posterior of $(K, \mathbf{W}_K)$

$$\pi(k, \mathbf{w}_k|X_1, \ldots, X_n) \propto \pi(k)\pi(\mathbf{w}_k|k) \prod_{i=1}^{n} \tilde{f}_k(x_i; \mathbf{w}_k)$$

and density estimate

$$\hat{f}(x) = \sum_{k\geq 1} \tilde{f}_K(x; \mathbb{E}[\mathbf{W}_K|X_1, \ldots, X_n, K])\, \pi(K|X_1, \ldots, X_n)$$

$$\lim_{k\to\infty} \tilde{f}_k(x;\mathbf{w}_k) = f(x) \qquad \forall x \in [0,1]$$

Prior on $(K, \mathbf{W}_k)$

$$(W_{1,K},\ldots,W_{K,K})|K \sim \mathrm{Dir}(\alpha_{1,K},\ldots,\alpha_{K,K})$$
$$K \sim \pi$$

Posterior of $(K, \mathbf{W}_K)$

$$\pi(k,\mathbf{w}_k|X_1,\ldots,X_n) \propto \pi(k)\pi(\mathbf{w}_k|k)\prod_{i=1}^{n}\tilde{f}_k(x_i;\mathbf{w}_k)$$

and density estimate

$$\hat{f}(x) = \sum_{k\geq 1} \tilde{f}_K(x;\mathbb{E}[\mathbf{W}_K|X_1,\ldots,X_n,K])\,\pi(K|X_1,\ldots,X_n)$$

mixture of random Bernstein polynomials

MCMC for random Bernstein polynomials

Introduce auxiliary random variables Y_i such that

$$
\begin{aligned}
X_i | Y_i, \tilde{F}, k &\overset{\text{ind}}{\sim} \text{Beta}(j, k + j - 1) \qquad \text{if } Y_i \in \left(\frac{j-1}{k}, \frac{j}{k}\right] \\
Y_i | \tilde{F}, k &\overset{\text{iid}}{\sim} \tilde{F} \\
\tilde{F} | k &\sim \text{Dir}(\theta\, F_0) \\
k &\sim \pi
\end{aligned}
$$

MCMC for random Bernstein polynomials

Introduce auxiliary random variables Y_i such that

$$
\begin{aligned}
X_i | Y_i, \tilde{F}, k &\overset{\text{ind}}{\sim} \text{Beta}(j, k+j-1) \qquad \text{if } Y_i \in \left(\frac{j-1}{k}, \frac{j}{k}\right] \\
Y_i | \tilde{F}, k &\overset{\text{iid}}{\sim} \tilde{F} \\
\tilde{F} | k &\sim \text{Dir}(\theta F_0) \\
k &\sim \pi
\end{aligned}
$$

Algorithm

- Generate $(k^{(t)}, \mathbf{y}^{(t)})$ from $\pi(K, Y_1, \ldots, Y_n | X_1, \ldots, X_n)$
- Generate $\mathbf{w}_k^{(t)}$ from $\pi(\mathbf{W}_{k^{(t)}} | k^{(t)}, \mathbf{y}^{(t)}, X_1, \ldots, X_n)$ which is

$$\text{Dir}(\alpha^*_{1,k}, \ldots, \alpha^*_{k,k})$$

In order to sample $(k^{(t)}, \mathbf{y}^{(t)})$ one makes use of the full conditionals

- $\pi(K|\mathbf{Y}, \mathbf{X})$ is proportional to

$$\pi(k) \prod_{i=1}^{n} \text{Beta}(x_i; \zeta_k(Y_i),\ k - \zeta_k(Y_i) + 1)$$

where $\zeta_k(Y_i) = j$ if $Y_i \in (\frac{j-1}{k}, \frac{j}{k}]$

In order to sample $(k^{(t)}, \mathbf{y}^{(t)})$ one makes use of the full conditionals

- $\pi(K|\mathbf{Y}, \mathbf{X})$ is proportional to

$$\pi(k) \prod_{i=1}^{n} \text{Beta}(x_i; \zeta_k(Y_i),\ k - \zeta_k(Y_i) + 1)$$

  where $\zeta_k(Y_i) = j$ if $Y_i \in (\frac{j-1}{k}, \frac{j}{k}]$

- $\pi(Y_i|k, \mathbf{Y}_{-i}, \mathbf{X})$ is proportional to

$$q(x_i, k) f_0(Y_i) \text{Beta}(x_i; \zeta_k(Y_i),\ k - \zeta_k(Y_i) + 1)\, \mathrm{d}Y_i + \sum_{j \neq i} q^*_{x_j,k}\, \delta_{Y_j}(\mathrm{d}Y_i)$$

  where

$$q(x_i, k) \propto \theta\, \text{Beta}(x_i; k, F_0) \qquad q^*_{x_j,k} \propto \text{Beta}(x_j; \zeta_k(Y_j),\ k - \zeta_k(Y_j) + 1)$$

  are such that $q(x_i, k) + \sum_{j \neq i} q^*_{x_j,k} = 1$

Density estimation: Pólya trees

Consider a sequence of nested partitions of $\mathbb{X} \subset \mathbb{R}$

- $\Pi_1 = \{B_0, B_1\}$
- $\Pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}$ where

$$B_{00} \cup B_{01} = B_0 \qquad B_{10} \cup B_{11} = B_1$$

- ....

Density estimation: Pólya trees

Consider a sequence of nested partitions of $\mathbb{X} \subset \mathbb{R}$

- $\Pi_1 = \{B_0, B_1\}$
- $\Pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}$ where

$$B_{00} \cup B_{01} = B_0 \qquad B_{10} \cup B_{11} = B_1$$

- ....

If $E^k = \{0,1\}^k$, then $\Pi_k = \{B_\varepsilon : \varepsilon \in E^k\}$.

Density estimation: Pólya trees

Consider a sequence of nested partitions of $\mathbb{X} \subset \mathbb{R}$

- $\Pi_1 = \{B_0, B_1\}$
- $\Pi_2 = \{B_{00}, B_{01}, B_{10}, B_{11}\}$ where

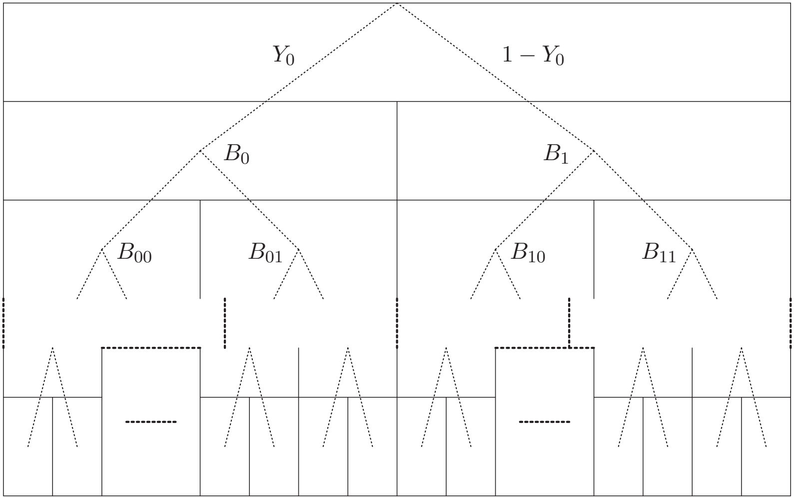
$$B_{00} \cup B_{01} = B_0 \qquad B_{10} \cup B_{11} = B_1$$

- ....

If $E^k = \{0,1\}^k$, then $\Pi_k = \{B_{\varepsilon} : \varepsilon \in E^k\}$.

If $\varepsilon \in E^k$, then

$$B_{\varepsilon 0} \cup B_{\varepsilon 1} = B_{\varepsilon}$$



Y_0
$1 - Y_0$
B_0
B_1
B_{00}
B_{01}
B_{10}
B_{11}

Definition

Let $E^* = \cup_{k\geq 1} E^k$ and $\mathscr{A} = \{\alpha_{\boldsymbol{\varepsilon}} : \boldsymbol{\varepsilon} \in E^*\}$ is a set of non negative real numbers.

Definition

Let $E^* = \cup_{k \geq 1} E^k$ and $\mathscr{A} = \{\alpha_{\varepsilon} : \varepsilon \in E^*\}$ is a set of non negative real numbers. A random probability measure $\tilde{p}$ is a Pólya tree process with respect to $\Pi = \{\Pi_k : k \geq 1\}$ and $\mathscr{A}$ if

- $\{\tilde{p}(B_{\varepsilon 0} | B_{\varepsilon}) : \varepsilon \in E^*\}$ is a collection of independent random variables
- $\tilde{p}(B_{\varepsilon 0} | B_{\varepsilon}) \sim \text{Beta}(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$

In symbols, $\tilde{\mathbf{p}} \sim \mathbf{PT}(\mathscr{A}, \mathbf{\Pi})$

Definition

Let $E^* = \cup_{k\geq 1} E^k$ and $\mathscr{A} = \{\alpha_{\varepsilon} : \varepsilon \in E^*\}$ is a set of non negative real numbers. A random probability measure $\tilde{p}$ is a Pólya tree process with respect to $\Pi = \{\Pi_k : k \geq 1\}$ and $\mathscr{A}$ if

- $\{\tilde{p}(B_{\varepsilon 0}|B_{\varepsilon}) : \varepsilon \in E^*\}$ is a collection of independent random variables
- $\tilde{p}(B_{\varepsilon 0}|B_{\varepsilon}) \sim \text{Beta}(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$

In symbols, $\mathbf{\tilde{p} \sim PT(\mathscr{A}, \Pi)}$

See Ferguson (1974), Mauldin, Sudderth & Williams (1992), Lavine (1992, 1994)

Definition

Let $E^* = \cup_{k\geq 1} E^k$ and $\mathscr{A} = \{\alpha_{\varepsilon} : \varepsilon \in E^*\}$ is a set of non negative real numbers. A random probability measure $\tilde{p}$ is a Pólya tree process with respect to $\Pi = \{\Pi_k : k \geq 1\}$ and $\mathscr{A}$ if

- $\{\tilde{p}(B_{\varepsilon 0}|B_{\varepsilon}) : \varepsilon \in E^*\}$ is a collection of independent random variables
- $\tilde{p}(B_{\varepsilon 0}|B_{\varepsilon}) \sim \text{Beta}(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$

In symbols, $\tilde{\mathbf{p}} \sim \mathbf{PT}(\mathscr{A}, \boldsymbol{\Pi})$

See Ferguson (1974), Mauldin, Sudderth & Williams (1992), Lavine (1992, 1994)

If

$$\alpha_{\varepsilon} = \alpha_{\varepsilon 0} + \alpha_{\varepsilon 1}$$

then $\tilde{p}$ is the Dirichlet process

Let $Y_{\boldsymbol{\varepsilon}} = P(B_{\boldsymbol{\varepsilon}0} | B_{\boldsymbol{\varepsilon}})$. For any $\boldsymbol{\varepsilon} \in E^k$

$$\tilde{p}(B_{\boldsymbol{\varepsilon}}) = \left\{ \prod_{j=1,\, \varepsilon_j = 0}^{k} Y_{\varepsilon_1 \cdots \varepsilon_{j-1}} \right\} \left\{ \prod_{j=1,\, \varepsilon_j = 1}^{k} (1 - Y_{\varepsilon_1 \cdots \varepsilon_{j-1}}) \right\}$$

Let $Y_{\boldsymbol{\varepsilon}} = P(B_{\boldsymbol{\varepsilon}0}|B_{\boldsymbol{\varepsilon}})$. For any $\boldsymbol{\varepsilon} \in E^k$

$$\tilde{p}(B_{\boldsymbol{\varepsilon}}) = \left\{ \prod_{j=1,\, \varepsilon_j=0}^{k} Y_{\varepsilon_1 \cdots \varepsilon_{j-1}} \right\} \left\{ \prod_{j=1,\, \varepsilon_j=1}^{k} (1 - Y_{\varepsilon_1 \cdots \varepsilon_{j-1}}) \right\}$$

Absolute continuity

Suppose $\mathbb{X} = [0, 1]$ and Π is a sequence of partitions of $[0, 1]$ into dyadic intervals: when $\boldsymbol{\varepsilon} \in E^k$ a set $B_{\boldsymbol{\varepsilon}} \in \Pi_k$ is of the form

$$\left(\sum_{j=1}^{k} \varepsilon_j \, 2^{-j}, \; \sum_{j=1}^{k} \varepsilon_j \, 2^{-j} + 2^{-k} \right]$$

If $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ and $\alpha_{\varepsilon_1 \cdots \varepsilon_k} = k^2$ then $\tilde{p}$ is (almost surely) a probability distribution absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$

Marginal distribution: Using independence of the beta random variables one has

$$\mathbb{P}[X_1 \in B_{\boldsymbol{\varepsilon}}] = \mathbb{E}[\tilde{p}(B_{\boldsymbol{\varepsilon}})] = \prod_{i=1}^{k} \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_i}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 1}}$$

Marginal distribution: Using independence of the beta random variables one has

$$\mathbb{P}[X_1 \in B_{\boldsymbol{\varepsilon}}] = \mathbb{E}[\tilde{p}(B_{\boldsymbol{\varepsilon}})] = \prod_{i=1}^{k} \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_i}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 1}}$$

Posterior distribution: If

$$\begin{aligned} X_i \,|\, \tilde{p} \quad &\sim \quad \tilde{p} \qquad\qquad i = 1, \ldots, n \\ \tilde{p} \quad &\sim \quad \mathrm{PT}(\mathscr{A}, \Pi) \end{aligned}$$

then

$$\tilde{p} \,|\, (X_1, \ldots, X_n) \sim \mathrm{PT}(\mathscr{A}^*, \Pi)$$

where $\mathscr{A}^* = \{\alpha_{\boldsymbol{\varepsilon}}(\mathbf{X}) : \boldsymbol{\varepsilon} \in E^*\}$ and

$$\alpha_{\boldsymbol{\varepsilon}}(\mathbf{X}) = \alpha_{\boldsymbol{\varepsilon}} + \sum_{i=1}^{n} \mathbb{I}_{B_{\boldsymbol{\varepsilon}}}(X_i)$$

Using the representations of the marginal and of the posterior distribution one obtains

The predictive distribution: For any $\boldsymbol{\varepsilon} \in E^k$

$$\Pr[X_{n+1} \in B_{\boldsymbol{\varepsilon}} | X_1, \ldots, X_n] = \frac{\alpha_{\varepsilon_1} + n_{\varepsilon_1}}{\alpha_0 + \alpha_1 + n} \times$$

$$\times \; \frac{\alpha_{\varepsilon_1 \varepsilon_2} + n_{\varepsilon_1 \varepsilon_2}}{\alpha_{\varepsilon_1 0} + \alpha_{\varepsilon_1 1} + n_{\varepsilon_1}} \cdots \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_k} + n_{\varepsilon_1 \cdots \varepsilon_{k-1}}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{k-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{k-1} 1} + n_{\varepsilon_1 \cdots \varepsilon_k}}$$

where $n_{\varepsilon_1 \cdots \varepsilon_i} =$ number of X_j's in $B_{\varepsilon_1 \cdots \varepsilon_i}$ $(j = 1, \ldots, n)$

Centering of $\tilde{p}$

Suppose one wants to specify $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ such that $\mathbb{E}[\tilde{p}] = P_0$

Centering of $\tilde{p}$

Suppose one wants to specify $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ such that $\mathbb{E}[\tilde{p}] = P_0$

- Set $F_0(x) = P_0(-\infty, x]$ for any x

Centering of $\tilde{p}$

Suppose one wants to specify $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ such that $\mathbb{E}[\tilde{p}] = P_0$

- Set $F_0(x) = P_0(-\infty, x]$ for any x
- For any k and $\boldsymbol{\varepsilon} \in E^k$

$$B_{\boldsymbol{\varepsilon}} = \left(F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i\, 2^{-i}\Big),\; F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i\, 2^{-i} + 2^{-k}\Big) \right]$$

Centering of $\tilde{p}$

Suppose one wants to specify $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ such that $\mathbb{E}[\tilde{p}] = P_0$

- Set $F_0(x) = P_0(-\infty, x]$ for any x
- For any k and $\boldsymbol{\varepsilon} \in E^k$

$$B_{\boldsymbol{\varepsilon}} = \left(F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i \, 2^{-i}\Big),\ F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i \, 2^{-i} + 2^{-k}\Big) \right]$$

- Take $\alpha_0 = \alpha_1$ and $\alpha_{\boldsymbol{\varepsilon}0} = \alpha_{\boldsymbol{\varepsilon}1}$ for any $\boldsymbol{\varepsilon} \in E^*$

Centering of $\tilde{p}$

Suppose one wants to specify $\tilde{p} \sim \mathrm{PT}(\mathscr{A}, \Pi)$ such that $\mathbb{E}[\tilde{p}] = P_0$

- Set $F_0(x) = P_0(-\infty, x]$ for any x
- For any k and $\boldsymbol{\varepsilon} \in E^k$

$$B_{\boldsymbol{\varepsilon}} = \left(F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i \, 2^{-i}\Big),\ F_0^{-1}\Big(\sum_{i=1}^{k} \varepsilon_i \, 2^{-i} + 2^{-k}\Big) \right]$$

- Take $\alpha_0 = \alpha_1$ and $\alpha_{\boldsymbol{\varepsilon}0} = \alpha_{\boldsymbol{\varepsilon}1}$ for any $\boldsymbol{\varepsilon} \in E^*$

Then, for any $\boldsymbol{\varepsilon} \in E^k$ and $k \geq 1$ and

$$\mathbb{E}\left[\tilde{p}(B_{\boldsymbol{\varepsilon}})\right] = \prod_{i=1}^{k} \frac{\alpha_{\varepsilon_1 \cdots \varepsilon_i}}{\alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 0} + \alpha_{\varepsilon_1 \cdots \varepsilon_{i-1} 1}} = 2^{-k} = P_0(B_{\boldsymbol{\varepsilon}})$$

- Drawback: posterior inferences depend on the partition Π
  Lavine (1992): mixture of Pólya trees

$$\begin{aligned}\tilde{p} \,|\, \theta, \xi &\sim \mathrm{PT}(\mathscr{A}^{\theta}, \Pi^{\xi}) \\ \theta &\sim \pi_1(\theta) \\ \xi &\sim \pi_2(\theta)\end{aligned}$$

  See also Hanson & Johnson (2002)