Estadística bayesiana
y aplicaciones en ciencia de datos

Ramsés H. Mena
IIMAS-UNAM, México

2024

1. Statistical induction
2. Dirichlet process: the canonical BNP prior
3. BNP mixtures

1. Statistical induction

2. Dirichlet process: the canonical BNP prior

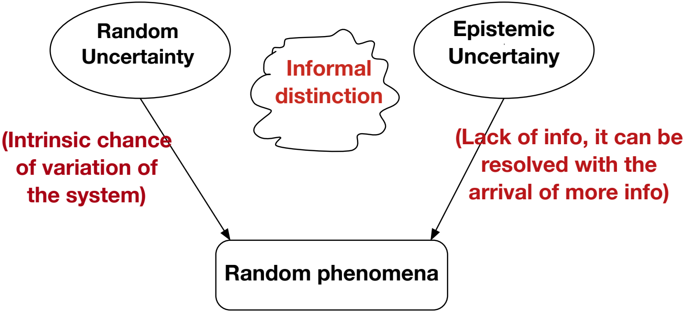
3. BNP mixtures

Random phenomena drive many aspects of this world

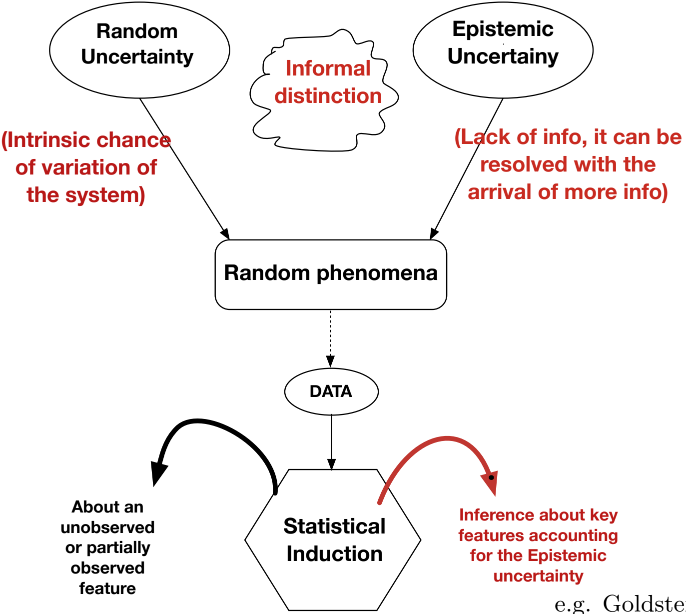
The Bayesian approach to statistical induction

Random phenomena

The Bayesian approach to statistical induction



The Bayesian approach to statistical induction



e.g. Goldstein (2013)

The basic probabilistic setup

- $(\Omega, \mathcal{A}, \mathbb{P})$: ***Probability space***
  - Ω –***sample space.*** Set of all possible outcomes
  - $\mathcal{A}$ –σ***-field.*** Collection of subsets of Ω with all events of interest
  - $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ –***Probability measure.*** Mathematically coherent measure to quantify all ***events*** $A \in \mathcal{A}$
- Features of interest can be translated into "numeric" quantities via
  - $(\mathbb{X}, \mathcal{X})$-valued functions, $X : \Omega \mapsto \mathbb{X}$. ***random variables*** (r.v.'s)
- Given a r.v. X, the set function defined by

$$\mathsf{P}_X(B) = \mathbb{P}(X^{-1}(B)), \qquad \text{for all } B \in \mathcal{X} \tag{1}$$

is termed the ***distribution or law*** of the random variable X.

  - When $\mathbb{X} = \mathbb{R}$ and $B = (-\infty, x]$ we write

$$\mathsf{F}_X(x) = \mathsf{P}_X((-\infty, x]) = \mathbb{P}(X \le x) \qquad \to \text{the } \textbf{\textit{(cdf)}} \text{ of } X$$

The basic probabilistic setup

- $(\Omega, \mathcal{A}, \mathbb{P})$: ***Probability space***
  - ▷ Ω –***sample space.*** Set of all possible outcomes
  - ▷ $\mathcal{A}$ –σ***-field.*** Collection of subsets of Ω with all events of interest
  - ▷ $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ –***Probability measure.*** Mathematically coherent measure to quantify all ***events*** $A \in \mathcal{A}$
- Features of interest can be translated into "numeric" quantities via
  - ▷ $(\mathbb{X}, \mathcal{X})$-valued functions, $X : \Omega \mapsto \mathbb{X}$. ***random variables*** (r.v.'s)
- Given a r.v. X, the set function defined by

$$\mathsf{P}_X(B) = \mathbb{P}(X^{-1}(B)), \qquad \text{for all } B \in \mathcal{X} \tag{1}$$

is termed the ***distribution or law*** of the random variable X.

- ▷ When $\mathbb{X} = \mathbb{R}$ and $B = (-\infty, x]$ we write

$$\mathsf{F}_X(x) = \mathsf{P}_X((-\infty, x]) = \mathbb{P}(X \leq x) \qquad \rightarrow \text{the } \textbf{\textit{(cdf)}} \text{ of } X$$

The basic probabilistic setup

- $(\Omega, \mathcal{A}, \mathbb{P})$: ***Probability space***
  - ▷ Ω –***sample space.*** Set of all possible outcomes
  - ▷ $\mathcal{A}$ –σ***-field.*** Collection of subsets of Ω with all events of interest
  - ▷ $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ –***Probability measure.*** Mathematically coherent measure to quantify all ***events*** $A \in \mathcal{A}$
- Features of interest can be translated into "numeric" quantities via
  - ▷ $(\mathbb{X}, \mathcal{X})$-valued functions, $X : \Omega \mapsto \mathbb{X}$. ***random variables*** (r.v.'s)
- Given a r.v. X, the set function defined by

$$\mathsf{P}_X(B) = \mathbb{P}(X^{-1}(B)), \qquad \text{for all } B \in \mathcal{X} \tag{1}$$

is termed the ***distribution or law*** of the random variable X.

- ▷ When $\mathbb{X} = \mathbb{R}$ and $B = (-\infty, x]$ we write

$$\mathsf{F}_X(x) = \mathsf{P}_X((-\infty, x]) = \mathbb{P}(X \leq x) \quad \rightarrow \text{the } \textbf{\textit{(cdf)}} \text{ of } X$$

Ex. toss a coin

$\Omega = \{\text{head}, \text{tail}\} = \{\omega_1, \omega_2\} = \{0, 1\}$

$\mathcal{A} = \{\Omega, \{0\}, \{1\}, \emptyset\}$

Let X the r.v. that assigns 1 if the outcome is tail and 0 otherwise, i.e. $\mathsf{P}_X(\{1\}) = \mathbb{P}(X(\omega_1) = 1)$ with $\mathbb{X} = \{0, 1\}$

- For such quantity, we might assign a value $\theta \in [0, 1]$, i.e.

$$\mathsf{P}_X(\{1\}) = \theta$$

$\Rightarrow$ Uncertainty about X is transferred to the ***parameter*** of interest θ.

How can we improve our knowledge about θ in the presence of observations from the random phenomena?

The basic setup

- Availability of more info about a random phenomenon

  ⇒ better uncertainty quantification

  ⇒ better statistical induction

- Realizations of a given phenomenon encoded via r.v.'s $\{X_i\}_{i\in\mathcal{I}}$
  - ▷ Logical/physical independence $\not\Rightarrow$ stochastic independence

    so $\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} \in B)$ not always a good idea!
  - ▷ Statistical learning requires stochastic dependence !

Under physical independence all we can assume is certain stochastic symmetry among $\{X_i\}_{i\in\mathcal{I}}$

The basic setup

- Availability of more info about a random phenomenon

  ⇒ better uncertainty quantification

  ⇒ better statistical induction

- Realizations of a given phenomenon encoded via r.v.'s $\{X_i\}_{i\in\mathcal{I}}$
  - ▷ Logical/physical independence $\not\Rightarrow$ stochastic independence

    so $\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} \in B)$ not always a good idea!

  - ▷ Statistical learning requires stochastic dependence !

Under physical independence all we can assume is certain stochastic symmetry among $\{X_i\}_{i\in\mathcal{I}}$

The basic setup

- Availability of more info about a random phenomenon

$\Rightarrow$ better uncertainty quantification

$\Rightarrow$ better statistical induction

- Realizations of a given phenomenon encoded via r.v.'s $\{X_i\}_{i \in \mathcal{I}}$
  - ▷ Logical/physical independence $\not\Rightarrow$ stochastic independence

    so $\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) = \mathbb{P}(X_{n+1} \in B)$ not always a good idea!

  - ▷ Statistical learning requires stochastic dependence !

Under physical independence all we can assume is certain stochastic symmetry among $\{X_i\}_{i \in \mathcal{I}}$

The basic setup

- Symmetry/Stability principles in the law modelling $\{X_i\}$'s are fundamental for statistical induction

  ▷ e.g. the past and future have similar behaviour

- Major symmetries used in statistics

  ▷ IID r.v.'s: physical & stochastic independence (rare in real apps!)

  ▷ Exchangeability: physical indep. + sampling order invariace!

  ▷ Stationarity: Uncertainty is not "time" invariant

  "...practitioners seem to prefer the language of populations: theoreticians, that of exchangeability"

  Lindley and Novick, 1981.

The basic setup

- Symmetry/Stability principles in the law modelling $\{X_i\}$'s are fundamental for statistical induction

  ▷ e.g. the past and future have similar behaviour

- Major symmetries used in statistics

  ▷ IID r.v.'s: physical & stochastic independence (rare in real apps!)

  ▷ Exchangeability: physical indep. + sampling order invariace!

  ▷ Stationarity: Uncertainty is not "time" invariant

"... practitioners seem to prefer the language of populations: theoreticians, that of exchangeability"

Lindley and Novick, 1981.

The basic setup

- Symmetry/Stability principles in the law modelling $\{X_i\}$'s are fundamental for statistical induction

    ▷ e.g. the past and future have similar behaviour

- Major symmetries used in statistics

  ▷ IID r.v.'s: physical & stochastic independence (rare in real apps!)

  ▷ Exchangeability: physical indep. + sampling order invariace!

  ▷ Stationarity: Uncertainty is not "time" invariant

  "... practitioners seem to prefer the language of populations: theoreticians, that of exchangeability"

  Lindley and Novick, 1981.

Exchangeable sequences

A finite sequence of r.v.'s, $\{X_n\}_{i=1}^{n}$, is said to be ***finite exchangeable*** if, for any permutation π of $(1, \ldots, n)$

$$(X_1, \ldots, X_n) \overset{\mathrm{d}}{=} (X_{\pi(1)}, \ldots, X_{\pi(n)})$$

An infinite sequence $\{X_n\}_{i=1}^{\infty}$ is said to be ***exchangeable*** if every subcollection is exchangeable.

- ≈ Distributional invariance under sampling order
- ▷ What can we say about the law of an exchangeable sequence

▷ B. de Finetti's representation characterises exchangeable sequences

de Finetti's representation Theorem: $\mathbb{X} = \{0, 1\}$ case

- B. de Finetti 1931: A seq. of binary r.v.'s $\{X_i\}_{i=1}^{\infty}$, *e.g.* with values in $\mathbb{X} = \{0, 1\}$, is exchangeable iff there exists a dist. q on $[0, 1]$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n} \mathsf{q}(\mathrm{d}\theta)$$

where $s_n := \sum_{i=1}^{n} x_i$.

- $\mathsf{q}(\cdot)$ is the distribution of $\lim_{n\to\infty} \frac{s_n}{n}$
- Conditional independence

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i \mid \theta) = \theta^{s_n}(1-\theta)^{n-s_n}$$

- This decomposition in "conditionally independent sample" given a random parameter "θ" justifies the Bayesian approach

de Finetti's representation Theorem: $\mathbb{X} = \{0, 1\}$ case

- B. de Finetti 1931: A seq. of binary r.v.'s $\{X_i\}_{i=1}^{\infty}$, *e.g.* with values in $\mathbb{X} = \{0, 1\}$, is exchangeable iff there exists a dist. q on $[0, 1]$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n}\mathsf{q}(\mathrm{d}\theta)$$

where $s_n := \sum_{i=1}^{n} x_i$.

  - $\mathsf{q}(\cdot)$ is the distribution of $\lim_{n\to\infty} \frac{s_n}{n}$
  - Conditional independence

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i \mid \theta) = \theta^{s_n}(1-\theta)^{n-s_n}$$

  - This decomposition in "conditionally independent sample" given a random parameter "θ" justifies the Bayesian approach

de Finetti's representation Theorem: $\mathbb{X} = \{0, 1\}$ case

- B. de Finetti 1931: A seq. of binary r.v.'s $\{X_i\}_{i=1}^{\infty}$, *e.g.* with values in $\mathbb{X} = \{0, 1\}$, is exchangeable iff there exists a dist. q on $[0, 1]$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n}\mathsf{q}(\mathrm{d}\theta)$$

where $s_n := \sum_{i=1}^{n} x_i$.

- $\mathsf{q}(\cdot)$ is the distribution of $\lim_{n\to\infty} \frac{s_n}{n}$
- Conditional independence

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i \mid \theta) = \theta^{s_n}(1-\theta)^{n-s_n}$$

- This decomposition in "conditionally independent sample" given a random parameter "θ" justifies the Bayesian approach

de Finetti's representation Theorem: $\mathbb{X} = \{0, 1\}$ case

- B. de Finetti 1931: A seq. of binary r.v.'s $\{X_i\}_{i=1}^{\infty}$, *e.g.* with values in $\mathbb{X} = \{0, 1\}$, is exchangeable iff there exists a dist. q on $[0, 1]$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n}\mathsf{q}(\mathrm{d}\theta)$$

where $s_n := \sum_{i=1}^{n} x_i$.

- $\mathsf{q}(\cdot)$ is the distribution of $\lim_{n\to\infty} \frac{s_n}{n}$
- Conditional independence

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n \mid \theta) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i \mid \theta) = \theta^{s_n}(1-\theta)^{n-s_n}$$

- This decomposition in "conditionally independent sample" given a random parameter "θ" justifies the Bayesian approach

Example: $\{X_i\}_{i=1}^{\infty}$ be Bernoulli r.v.'s

Two different Bernoulli exchangeable laws by two different persons

$$\mathbb{P}(x_1,\ldots,x_n) = \frac{12}{s_n+2}\frac{1}{\binom{n+4}{s_n+2}} \quad \text{and} \quad \mathbb{P}(x_1,\ldots,x_n) = \frac{1}{[n+1]\binom{n}{s_n}},$$

- These persons believe that $\mathbb{P}(X_1 = 1) = 0.4$ & $\mathbb{P}(X_1 = 1) = 0.5$ resp.
- Both believe that $\Theta := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$ exists & $\mathbb{P}(X_1 = 1 \mid \Theta = \theta) = \theta$
- Since $\mathbb{P}(X_1 = 1) = \mathbb{E}(\Theta)$ they must have different values for $\mathbb{E}(\Theta)$
- Assume we observe the result of $n = 20$ given by 14 "1s" and 6 "0s".

$$\mathbb{P}[X_{21} \mid x_1,\ldots,x_{20}] = 0.64 \quad \text{and} \quad \mathbb{P}[X_{21} \mid x_1,\ldots,x_{20}] = 0.68.$$

• Regardless of the prior mean on Θ, they should modify their opinion about the prop. of 1's!

• Consequence due to exchangeability, regardless of frequencies being interpreted as probabilities.

Schervish, 1995.

In Bayesian terms

If X is a r.v. on $\mathbb{X} = \{1, 2\}$ with prob. p_1 and $p_2 = 1 - p_1$ assigned to each element of $\mathbb{X}$. That is $\{X \mid p_1, p_2\} \sim \text{Bernoulli}(p_1, p_2)$

- Each value of $\mathbf{p} = (p_1, p_2)$ defines probability measure on $\mathbb{X}$
- $q = \mathsf{Be}(\alpha_1, \alpha_2)$ defines a probability measure on

$\mathcal{P}_{\mathbb{X}} :=$ {Space of prob. measures on $\mathbb{X}$}$=\{(p_1, p_2); p_i \geq 0 \text{ y } p_1 + p_2 = 1\}$

$$\Rightarrow \quad p_1 \mid X^{(n)} \sim \mathsf{Be}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(\{2\})\right)$$

▷ If $\mathbb{X} = \{1, 2, \ldots, K\}$ then $X \mid (p_1, p_2, \ldots, p_k) \sim \prod_{i=1}^{k} p_i^{\delta_X(\{i\})}$

$$\mathcal{P}_{\mathbb{X}} \quad = \{(p_1, \ldots, p_k); p_i \geq 0 \text{ y } p_1 + \cdots + p_k = 1\}$$

$$\Rightarrow (p_1, p_2, \ldots, p_k) \mid X^{(n)} \sim \text{Dirichlet}(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(\{k\}))$$

One parameter per value in the support $\mathbb{X}$!

In Bayesian terms

If X is a r.v. on $\mathbb{X} = \{1, 2\}$ with prob. p_1 and $p_2 = 1 - p_1$ assigned to each element of $\mathbb{X}$. That is $\{X \mid p_1, p_2\} \sim \text{Bernoulli}(p_1, p_2)$

- Each value of $\mathbf{p} = (p_1, p_2)$ defines probability measure on $\mathbb{X}$
- $q = \mathsf{Be}(\alpha_1, \alpha_2)$ defines a probability measure on

$\mathcal{P}_{\mathbb{X}} :=$ {Space of prob. measures on $\mathbb{X}$}$=\{(p_1, p_2); p_i \geq 0 \text{ y } p_1 + p_2 = 1\}$

$$\Rightarrow \quad p_1 \mid X^{(n)} \sim \mathsf{Be}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(\{2\})\right)$$

▷ If $\mathbb{X} = \{1, 2, \ldots, K\}$ then $X \mid (p_1, p_2, \ldots, p_k) \sim \prod_{i=1}^{k} p_i^{\delta_X(\{i\})}$

$$\mathcal{P}_{\mathbb{X}} \quad = \{(p_1, \ldots, p_k); p_i \geq 0 \text{ y } p_1 + \cdots + p_k = 1\}$$

$$\Rightarrow (p_1, p_2, \ldots, p_k) \mid X^{(n)} \sim \text{Dirichlet}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(\{k\})\right)$$

One parameter per value in the support $\mathbb{X}$!

In Bayesian terms

If X is a r.v. on $\mathbb{X} = \{1, 2\}$ with prob. p_1 and $p_2 = 1 - p_1$ assigned to each element of $\mathbb{X}$. That is $\{X \mid p_1, p_2\} \sim \text{Bernoulli}(p_1, p_2)$

- Each value of $\mathbf{p} = (p_1, p_2)$ defines probability measure on $\mathbb{X}$
- $q = \mathsf{Be}(\alpha_1, \alpha_2)$ defines a probability measure on

$\mathcal{P}_{\mathbb{X}} :=$ {Space of prob. measures on $\mathbb{X}$}$=\{(p_1, p_2); p_i \geq 0 \text{ y } p_1 + p_2 = 1\}$

$$\Rightarrow \quad p_1 \mid X^{(n)} \sim \mathsf{Be}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(\{2\})\right)$$

▷ If $\mathbb{X} = \{1, 2, \ldots, K\}$ then $X \mid (p_1, p_2, \ldots, p_k) \sim \prod_{i=1}^{k} p_i^{\delta_X(\{i\})}$

$$\mathcal{P}_{\mathbb{X}} \quad = \{(p_1, \ldots, p_k); p_i \geq 0 \text{ y } p_1 + \cdots + p_k = 1\}$$

$$\Rightarrow (p_1, p_2, \ldots, p_k) \mid X^{(n)} \sim \text{Dirichlet}(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(\{k\}))$$

One parameter per value in the support $\mathbb{X}$!

In Bayesian terms

If X is a r.v. on $\mathbb{X} = \{1,2\}$ with prob. p_1 and $p_2 = 1 - p_1$ assigned to each element of $\mathbb{X}$. That is $\{X \mid p_1, p_2\} \sim \text{Bernoulli}(p_1, p_2)$

- Each value of $\mathbf{p} = (p_1, p_2)$ defines probability measure on $\mathbb{X}$
- $q = \mathsf{Be}(\alpha_1, \alpha_2)$ defines a probability measure on

$\mathcal{P}_{\mathbb{X}} := \{\text{Space of prob. measures on } \mathbb{X}\} = \{(p_1, p_2); p_i \geq 0 \text{ y } p_1 + p_2 = 1\}$

$$\Rightarrow \quad p_1 \mid X^{(n)} \sim \mathsf{Be}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(\{2\})\right)$$

▷ If $\mathbb{X} = \{1, 2, \ldots, K\}$ then $X \mid (p_1, p_2, \ldots, p_k) \sim \prod_{i=1}^{k} p_i^{\delta_X(\{i\})}$

$$\mathcal{P}_{\mathbb{X}} \quad = \{(p_1, \ldots, p_k); p_i \geq 0 \text{ y } p_1 + \cdots + p_k = 1\}$$

$$\Rightarrow (p_1, p_2, \ldots, p_k) \mid X^{(n)} \sim \text{Dirichlet}(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(\{k\}))$$

One parameter per value in the support $\mathbb{X}$!

In Bayesian terms

If X is a r.v. on $\mathbb{X} = \{1, 2\}$ with prob. p_1 and $p_2 = 1 - p_1$ assigned to each element of $\mathbb{X}$. That is $\{X \mid p_1, p_2\} \sim \text{Bernoulli}(p_1, p_2)$

- Each value of $\mathbf{p} = (p_1, p_2)$ defines probability measure on $\mathbb{X}$
- $q = \mathsf{Be}(\alpha_1, \alpha_2)$ defines a probability measure on

$\mathcal{P}_{\mathbb{X}} :=$ {Space of prob. measures on $\mathbb{X}$}$=\{(p_1, p_2); p_i \geq 0 \text{ y } p_1 + p_2 = 1\}$

$$\Rightarrow \quad p_1 \mid X^{(n)} \sim \mathsf{Be}\left(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \alpha_2 + \sum_{i=1}^{n} \delta_{X_i}(\{2\})\right)$$

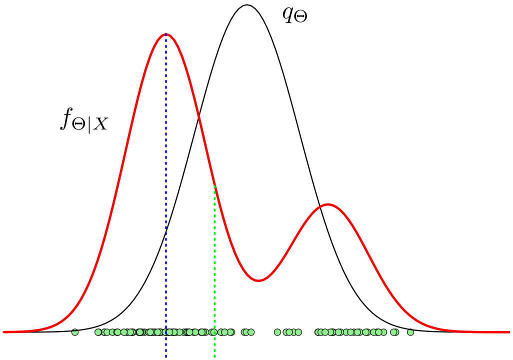
▷ If $\mathbb{X} = \{1, 2, \ldots, K\}$ then $X \mid (p_1, p_2, \ldots, p_k) \sim \prod_{i=1}^{k} p_i^{\delta_X(\{i\})}$

$$\mathcal{P}_{\mathbb{X}} \quad = \{(p_1, \ldots, p_k); p_i \geq 0 \text{ y } p_1 + \cdots + p_k = 1\}$$

$$\Rightarrow (p_1, p_2, \ldots, p_k) \mid X^{(n)} \sim \text{Dirichlet}(\alpha_1 + \sum_{i=1}^{n} \delta_{X_i}(\{1\}), \ldots, \alpha_k + \sum_{i=1}^{n} \delta_{X_i}(\{k\}))$$

One parameter per value in the support $\mathbb{X}$!

Prior to posterior effect



Exchangeable sequences: general $\mathbb{X}$

- Let $\mathcal{P}_{\mathbb{X}}$ be the space of all probability measures on $(\mathbb{X}, \mathcal{X})$

A seq. $\{X_i\}_{i=1}^{\infty}$ is exchangeable iff there exists Q on $\mathcal{P}_{\mathbb{X}}$ such that

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \mathsf{P}(A_i)\, \mathsf{Q}(d\mathsf{P}), \quad \forall n \geq 1 \text{ and } A_i \in \mathcal{X}$$

Alternatively: $X_i \mid \mathsf{P} \overset{\text{iid}}{\sim} \mathsf{P}$ and $\mathsf{P} \sim \mathsf{Q}$ (conditionally iid) .

Hewitt and Savage 1955

- If $P_n(A) := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(A)$ denotes the empirical dist. hence, Q is the dist. of the RPM P, where $\mathbb{P}[P_n \Rightarrow \mathsf{P}] = 1 \quad (\mathsf{P} \sim \mathsf{Q})$
- Q is unique
- "The unknown", P, that allows us to disaggregate the elements of $X^{(\infty)}$ as a conditional iid sample, is random.

Consequences of de Finetti's representation

- There is a clear bijection between the law of $\{X_i\}_{i=1}^{\infty}$ and Q
  - Pick $\mathsf{Q} \Rightarrow$ we have a law for $\{X_i\}_{i=1}^{\infty}$
  - Pick a law for $\{X_i\}_{i=1}^{\infty} \Rightarrow$ there exist a unique Q

    i.e. $\mathsf{P}[X_{n+1} \in A \mid X^{(n)}] = \mathbb{E}_{\mathsf{Q}_{X^{(n)}}}[\mathsf{P}(A)]$ characterizes Q with $\mathsf{Q}_{X^{(n)}}(B) := \mathbb{P}(\mathsf{P} \in B \mid X^{(n)})$,
- However, any $\mathsf{P} \in \mathcal{P}_{\mathbb{X}}$ can be seen as the limit of P_n!
- Bayesian interpretation:
  Q takes the interpretation of prior distributions on P

Probabilistically speaking the Bayesian approach is equivalent to the exchangeability assumption of the $\{X_i\}_{i=1}^{\infty}$

Consequences of de Finetti's representation

- There is a clear bijection between the law of $\{X_i\}_{i=1}^{\infty}$ and Q
  - Pick Q $\Rightarrow$ we have a law for $\{X_i\}_{i=1}^{\infty}$
  - Pick a law for $\{X_i\}_{i=1}^{\infty}$ $\Rightarrow$ there exist a unique Q

    i.e. $\mathsf{P}[X_{n+1} \in A \mid X^{(n)}] = \mathbb{E}_{\mathsf{Q}_{X^{(n)}}}[\mathsf{P}(A)]$ characterizes Q with $\mathsf{Q}_{X^{(n)}}(B) := \mathbb{P}(\mathsf{P} \in B \mid X^{(n)})$,
- However, any $\mathsf{P} \in \mathcal{P}_{\mathbb{X}}$ can be seen as the limit of P_n!
- Bayesian interpretation:
  Q takes the interpretation of prior distributions on P

Probabilistically speaking the Bayesian approach is equivalent to the exchangeability assumption of the $\{X_i\}_{i=1}^{\infty}$

Consequences of de Finetti's representation

- There is a clear bijection between the law of $\{X_i\}_{i=1}^{\infty}$ and Q
  - Pick Q $\Rightarrow$ we have a law for $\{X_i\}_{i=1}^{\infty}$
  - Pick a law for $\{X_i\}_{i=1}^{\infty}$ $\Rightarrow$ there exist a unique Q

    i.e. $\mathsf{P}[X_{n+1} \in A \mid X^{(n)}] = \mathbb{E}_{\mathsf{Q}_{X^{(n)}}}[\mathsf{P}(A)]$ characterizes Q with $\mathsf{Q}_{X^{(n)}}(B) := \mathbb{P}(\mathsf{P} \in B \mid X^{(n)})$,
- However, any $\mathsf{P} \in \mathcal{P}_{\mathbb{X}}$ can be seen as the limit of P_n!
- Bayesian interpretation:
  Q takes the interpretation of prior distributions on P

Probabilistically speaking the Bayesian approach is equivalent to the exchangeability assumption of the $\{X_i\}_{i=1}^{\infty}$

Consequences of de Finetti's representation

- There is a clear bijection between the law of $\{X_i\}_{i=1}^{\infty}$ and Q
  - Pick Q $\Rightarrow$ we have a law for $\{X_i\}_{i=1}^{\infty}$
  - Pick a law for $\{X_i\}_{i=1}^{\infty}$ $\Rightarrow$ there exist a unique Q

    i.e. $\mathsf{P}[X_{n+1} \in A \mid X^{(n)}] = \mathbb{E}_{\mathsf{Q}_{X^{(n)}}}[\mathsf{P}(A)]$ characterizes Q with $\mathsf{Q}_{X^{(n)}}(B) := \mathbb{P}(\mathsf{P} \in B \mid X^{(n)})$,
- However, any $\mathsf{P} \in \mathcal{P}_{\mathbb{X}}$ can be seen as the limit of P_n!
- Bayesian interpretation:
  Q takes the interpretation of prior distributions on P

Probabilistically speaking the Bayesian approach is equivalent to the exchangeability assumption of the $\{X_i\}_{i=1}^{\infty}$

de Finetti and the Bayesian approach

The law of the exchangeable r.v's (and thus Q) is characterized by the conditional probabilities (or predictive distributions)

$$\mathbb{P}\left[X_{n+1} \in A_{n+1} \mid X_1 \in A_1, \ldots, X_n \in A_n\right] = \frac{\mathsf{E}_{\mathsf{Q}}\left[\prod_{i=1}^{n+1} \mathsf{P}(A_i)\right]}{\mathsf{E}_{\mathsf{Q}}\left[\prod_{i=1}^{n} \mathsf{P}(A_i)\right]} = \mathsf{E}_{\mathsf{Q}_{X^{(n)}}}\left[\mathsf{P}(A_{n+1})\right]$$

for all $n > 1$, with $P_0 := \mathbb{P}[X_1 \in A_1] = \mathsf{E}_{\mathsf{Q}}[\mathsf{P}(A_1)]$ and where

$$\mathsf{Q}_{X^{(n)}}(\mathrm{d}\mathsf{P}) = \frac{\prod_{i=1}^{n} \mathsf{P}(A_i)\, \mathsf{Q}(\mathrm{d}\mathsf{P})}{\mathsf{E}_{\mathsf{Q}}\left[\prod_{i=1}^{n} \mathsf{P}(A_i)\right]}, \quad \text{(dominated case)}$$

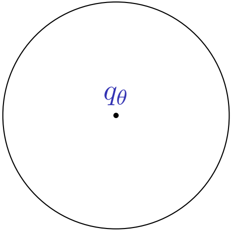
the posterior distribution of P given $X^{(n)} := (X_1, \ldots, X_n)$

Exchangeability: statistical learning for physically independent observations

Random phenomena encoded in $\mathbb{X}$-valued $\{X_i\}_{i=1}^{\infty}$ exchangeable sequence driven by $\mathsf{P} \sim \mathsf{Q}$

- $\mathsf{Q}(\cdot) = \delta_{q_\theta}(\cdot) \Rightarrow X_i$'s are iid

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \mathsf{P}(A_i)\, \delta_{q_\theta}(d\mathsf{P}) = \prod_{i=1}^{n} q_\theta(A_i)$$



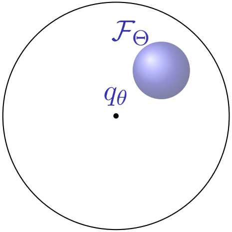
$\mathcal{P}_{\mathbb{X}}$: Space of all distributions on $\mathbb{X}$

Exchangeability: statistical learning for physically independent observations

Random phenomena encoded in $\mathbb{X}$-valued $\{X_i\}_{i=1}^{\infty}$ exchangeable sequence driven by $\mathsf{P} \sim \mathsf{Q}$

- $\mathsf{Q}(\mathcal{F}_\Theta) = 1 \Rightarrow$ Parametric family

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{F}_\Theta} \prod_{i=1}^{n} \underbrace{F_\theta(A_i)}_{\text{Random uncertainty via param. model}} \overbrace{\pi_\theta(d\theta)}^{\text{Epistemic uncertainty}}$$



$\mathcal{P}_{\mathbb{X}}$: Space of all distributions on $\mathbb{X}$

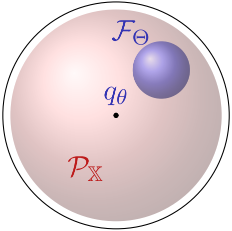
Exchangeability: statistical learning for physically independent observations

Random phenomena encoded in $\mathbb{X}$-valued $\{X_i\}_{i=1}^{\infty}$ exchangeable sequence driven by $\mathsf{P} \sim \mathsf{Q}$

- $\mathsf{Q}(\mathsf{P} : \mathsf{d}(\mathsf{P}, \eta) < \varepsilon) > 0, \ \forall\, \eta \in \mathcal{P}_{\mathbb{X}}$ y $\varepsilon > 0 \Rightarrow$ BNP

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \underbrace{\mathsf{P}(A_i)\, \mathsf{Q}(d\mathsf{P})}$$

Random and epistemic uncertainties in one stroke!



$\mathcal{P}_{\mathbb{X}}$: Space of all distributions on $\mathbb{X}$
…or other infinite dimensional sub-spaces of interest, $\mathcal{P}_{\mathbb{X}}^{d}$, $\mathcal{P}_{\mathbb{X}}^{c}$, etc.

Bayesian nonparametrics

What happens if $\mathbb{X}$ is of an infinite nature?

- We could $\mathcal{P}_{\mathbb{X}}\big|_{\mathcal{F}_\Theta}$, but doesn't resolve the "random uncertainty"
- We want models Q giving positive prob. to all elements of $\mathcal{P}_{\mathbb{X}}$, or at least some infinite subset, e.g. set of densities, cdf's, etc.
- de Finetti's representation Th. for general $\mathbb{X}$ gives an answer...
- Remember: $\{X_i\}_{i=n}^{\infty}$ exchangeable is driven by $\mathrm{P} \sim \mathrm{Q}$

How to construct suitable models for Q (nonparametric priors!)?

Bayesian nonparametrics

What happens if $\mathbb{X}$ is of an infinite nature?

- We could $\mathcal{P}_{\mathbb{X}}\big|_{\mathcal{F}_{\Theta}}$, but doesn't resolve the "random uncertainty"
- We want models Q giving positive prob. to all elements of $\mathcal{P}_{\mathbb{X}}$, or at least some infinite subset, e.g. set of densities, cdf's, etc.
- de Finetti's representation Th. for general $\mathbb{X}$ gives an answer...
- Remember: $\{X_i\}_{i=n}^{\infty}$ exchangeable is driven by $\mathsf{P} \sim \mathsf{Q}$

How to construct suitable models for Q (nonparametric priors!)?

The Dirichlet distribution

Let $Z_i \stackrel{\text{iid}}{\sim} \mathsf{Ga}(\alpha_i, 1)$, $i = 1, \ldots, m$ and $\mathbf{W} := (W_1, \ldots, W_m)$ with

$$W_i = \frac{Z_i}{\sum_{i=1}^m Z_i}, \quad i = 1, \ldots, m \qquad \Rightarrow \mathbf{W} \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_m)$$

and is independent of $Z := \sum_{i=1}^m Z_i \sim \mathsf{Ga}(\sum_{i=1}^m \alpha_i, 1)$ with density

$$f(\mathbf{w}) = \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^{m-1} w_i^{\alpha_i - 1} \left(1 - \sum_{i=1}^{n-1} w_i\right)^{\alpha_m - 1} \mathbb{I}_{\Delta_{m-1}}(\mathsf{w}),$$

where $\Delta_{m-1} := \left\{(w_1, \ldots, w_{m-1}) : w_i \geq 0, \sum_{i=1}^{m-1} w_i \leq 1\right\}$

Properties of Dirichlet distribution

Moments

Let $\alpha := \sum_{i=1}^{m} \alpha_i$ and $p_i := \alpha_i/\alpha$ hence

- $\mathsf{E}[w_i] = p_i$
- $\mathsf{Var}[w_i] = \frac{p_i(1-p_i)}{\alpha+1}$
- $\mathsf{Corr}[w_i, w_j] = -\frac{p_i p_j}{\alpha+1}$

Addition property

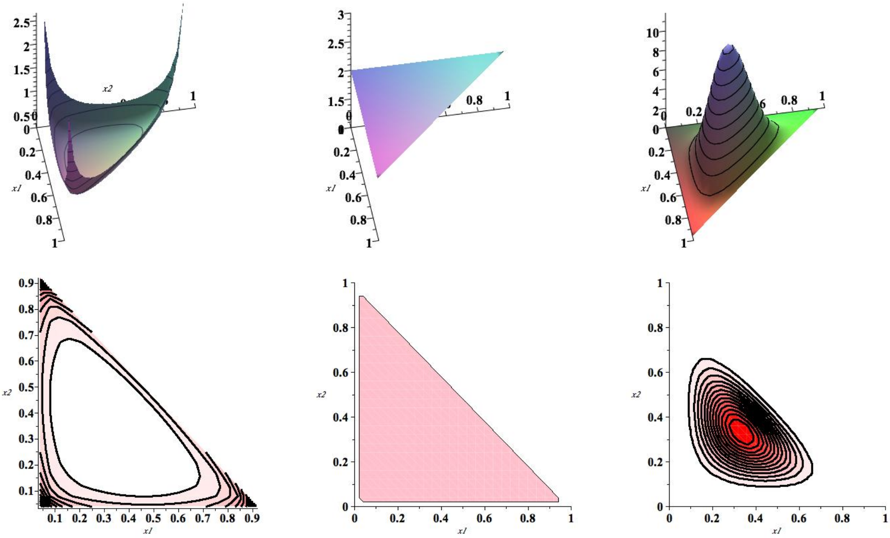
If W $\sim$ Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_m)$ then

i) For any partition $A_1, \ldots, A_k$ of $\{1, \ldots, n\}$, the vector

$$\left(\sum_{i \in A_1} w_i, \sum_{i \in A_2} w_i, \ldots, \sum_{i \in A_k} w_i\right) \sim \text{Dirichlet}(\alpha'_1, \ldots, \alpha'_k)$$

where $\alpha'_i := \sum_{j \in A_i} \alpha_j$

$\alpha_1 = \alpha_2 = \alpha_3 = 0.2$ | $\alpha_1 = \alpha_2 = \alpha_3 = 1$ | $\alpha_1 = \alpha_2 = \alpha_3 = 5$



Ferguson 1973: The canonical example

1. Via infinite dimensional distributions with pre-scribed fdds

Let $\alpha > 0$ a non-atomic finite measure on a Polish space $(\mathbb{X}, \mathcal{X})$. A $\mathcal{P}_{\mathbb{X}}$-valued RPM, P, is said to have a Dirichlet process $(\mathcal{D}_\alpha)$ distribution, if for all measurable partition $(B_1, \ldots, B_k)$ de $\mathbb{X}$

$$(\mathsf{P}(B_1), \ldots, \mathsf{P}(B_k)) \sim \mathrm{Dir}(\alpha(B_1), \ldots, \alpha(B_k))$$

- Ferguson 73' proved that the Dirichlet dist. is projective and therefore Daniel-Kolmogorov's existence theorem ensures the existence of $\mathcal{D}_\alpha$. Namely, a stochastic process indexed on $\mathcal{X}$.

The Dirichlet process $\mathcal{D}_\alpha$: The canonical example

Extending the finite-dim properties to the infinite-dim object it can be seen that if $X_i \stackrel{\text{iid}}{\sim} \mathsf{P}$ and $\mathsf{P} \sim \mathcal{D}_\alpha$ then

- $P_0(B) := \mathsf{E}_{\mathcal{D}_\alpha}[\mathsf{P}] = \frac{\alpha(B)}{\theta}$ for $B \in \mathcal{X}$ and where $\theta := \alpha(\mathbb{X})$
- $\mathsf{Var}_{\mathcal{D}_\alpha}[\mathsf{P}(B)] = \frac{P_0(B)(1-P_0(B))}{\theta+1}$
- $\mathsf{Cov}(\mathsf{P}(B_1), \mathsf{P}(B_2)) = \frac{P_0(B_1 \cap B_2) - P_0(B_1)P_0(B_2)}{\theta+1}$

If $X_i \mid \mathsf{P} \stackrel{\text{iid}}{\sim} \mathsf{P}$ y $\mathsf{P} \sim \mathcal{D}_{\theta P_0}$, then $X_i \sim P_0, \; \forall \; i = 1, 2, \ldots$

$$\mathsf{P} \mid X_1, \ldots, X_n \sim \mathcal{D}_{\theta P_0 + n P_n} \quad \text{(conjugacy)}$$

$$\mathsf{E}[\mathsf{P} \mid X_1, \ldots, X_n] = \frac{\theta}{\theta+n} P_0 + \frac{n}{\theta+n} \sum_{i=1}^{n} \frac{\delta_{X_i}}{n}, \quad \text{(Bayes estimator)}$$

Ferguson (1973)

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

2. Specification of Q via predictive distributions.

- Q can be characterized by its predictive dist. (Bayes estimator)

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \mathsf{E}\left[\mathsf{P}(A) \mid X_1, \ldots, X_n\right] = \frac{\alpha_n(A)}{\alpha_n(\mathbb{X})}$$

with $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^n \delta_{X_i}(A)$. In other terms

$$\mathbb{P}[X_{n+1} \in \cdot \mid X^{(n)}] = \underbrace{\frac{\theta}{\theta + n}}_{\mathbb{P}[X_{n+1} = \text{“new”} \mid X^{(n)}]} \overbrace{P_0(\cdot)}^{\text{Prior guess}} + \underbrace{\frac{n}{\theta + n}}_{\mathbb{P}[X_{n+1} = \text{“old”} \mid X^{(n)}]} \overbrace{\sum_{i=1}^n \frac{\delta_{X_i}}{n}(\cdot)}^{\text{empirical measure}},$$

- Q is a DP iff the predictive is a linear combination of P_0 and the empirical measure

Regazzini (1978); Lo (1991)

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

2. Specification of Q via predictive distributions.
   - Q can be characterized by its predictive dist. (Bayes estimator)

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \mathsf{E}\left[\mathsf{P}(A) \mid X_1, \ldots, X_n\right] = \frac{\alpha_n(A)}{\alpha_n(\mathbb{X})}$$

with $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^n \delta_{X_i}(A)$. In other terms

$$\mathbb{P}[X_{n+1} \in \cdot \mid X^{(n)}] = \underbrace{\frac{\theta}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{“new”} \mid X^{(n)}]} \overbrace{P_0(\cdot)}^{\text{Prior guess}} + \underbrace{\frac{n}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{“old”} \mid X^{(n)}]} \overbrace{\sum_{i=1}^n \frac{\delta_{X_i}}{n}(\cdot)}^{\text{empirical measure}},$$

- Q is a DP iff the predictive is a linear combination of P_0 and the empirical measure

Regazzini (1978); Lo (1991)

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

2. Specification of Q via predictive distributions.

- Q can be characterized by its predictive dist. (Bayes estimator)

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \mathsf{E}\left[\mathsf{P}(A) \mid X_1, \ldots, X_n\right] = \frac{\alpha_n(A)}{\alpha_n(\mathbb{X})}$$

with $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^n \delta_{X_i}(A)$. In other terms

$$\mathbb{P}[X_{n+1} \in \cdot \mid X^{(n)}] = \underbrace{\frac{\theta}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{“new”} \mid X^{(n)}]} \overbrace{P_0(\cdot)}^{\text{Prior guess}} + \underbrace{\frac{n}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{“old”} \mid X^{(n)}]} \overbrace{\sum_{i=1}^n \frac{\delta_{X_i}}{n}(\cdot)}^{\text{empirical measure}},$$

- Q is a DP iff the predictive is a linear combination of P_0 and the empirical measure

Regazzini (1978); Lo (1991)

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

2. Specification of Q via predictive distributions.
   - Q can be characterized by its predictive dist. (Bayes estimator)

$$\mathbb{P}(X_{n+1} \in A \mid X_1, \ldots, X_n) = \mathsf{E}\left[\mathsf{P}(A) \mid X_1, \ldots, X_n\right] = \frac{\alpha_n(A)}{\alpha_n(\mathbb{X})}$$

with $\alpha_n(\cdot) = \alpha(\cdot) + \sum_{i=1}^n \delta_{X_i}(A)$. In other terms

$$\mathbb{P}[X_{n+1} \in \cdot \mid X^{(n)}] = \underbrace{\frac{\theta}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{"new"} \mid X^{(n)}]} \overbrace{P_0(\cdot)}^{\text{Prior guess}} + \underbrace{\frac{n}{\theta+n}}_{\mathbb{P}[X_{n+1} = \text{"old"} \mid X^{(n)}]} \overbrace{\sum_{i=1}^n \frac{\delta_{X_i}}{n}(\cdot)}^{\text{empirical measure}},$$

- Q is a DP iff the predictive is a linear combination of P_0 and the empirical measure

Regazzini (1978); Lo (1991)

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

Blackwell and MacQueen 73' observed that when $n \to \infty$

$$\frac{\alpha_n()}{\alpha_n(\mathbb{X})} \overset{\text{a.s.}}{\to} \mathsf{P}, \qquad \text{with} \qquad \mathsf{P} \sim \mathcal{D}_\alpha$$

→ Very appealing for MCMC implementations

→ A direct consequence is that

$$\mathbb{P}(X_i = X_j) = \frac{1}{\theta + 1} > 0, \qquad i \neq j$$

Blackwell 73' proved that

- $\mathcal{D}_\alpha(\mathrm{P} : \mathrm{P} \text{ is discrete }) = 1$

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

Blackwell and MacQueen 73' observed that when $n \to \infty$

$$\frac{\alpha_n()}{\alpha_n(\mathbb{X})} \xrightarrow{\text{a.s.}} \mathsf{P}, \qquad \text{with} \qquad \mathsf{P} \sim \mathcal{D}_\alpha$$

→ Very appealing for MCMC implementations

→ A direct consequence is that

$$\mathbb{P}(X_i = X_j) = \frac{1}{\theta + 1} > 0, \qquad i \neq j$$

Blackwell 73' proved that

- $\mathcal{D}_\alpha(\mathsf{P} : \mathsf{P} \text{ is discrete }) = 1$

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

Blackwell and MacQueen 73' observed that when $n \to \infty$

$$\frac{\alpha_n()}{\alpha_n(\mathbb{X})} \stackrel{\text{a.s.}}{\to} \mathsf{P}, \qquad \text{with} \qquad \mathsf{P} \sim \mathcal{D}_\alpha$$

→ Very appealing for MCMC implementations

→ A direct consequence is that

$$\mathbb{P}(X_i = X_j) = \frac{1}{\theta + 1} > 0, \qquad i \neq j$$

Blackwell 73' proved that

- $\mathcal{D}_\alpha(\mathsf{P} : \mathsf{P} \text{ is discrete }) = 1$

The Dirichlet process $\mathcal{D}_\alpha$: Pólya urn representation

Blackwell and MacQueen 73' observed that when $n \to \infty$

$$\frac{\alpha_n()}{\alpha_n(\mathbb{X})} \overset{\text{a.s.}}{\to} \mathsf{P}, \qquad \text{with} \qquad \mathsf{P} \sim \mathcal{D}_\alpha$$

→ Very appealing for MCMC implementations

→ A direct consequence is that

$$\mathbb{P}(X_i = X_j) = \frac{1}{\theta + 1} > 0, \qquad i \neq j$$

Blackwell 73' proved that

- $\mathcal{D}_\alpha(\mathsf{P} : \mathsf{P} \text{ is discrete }) = 1$

References. . .

- Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, **2**, 1152–1174.
- Anzarut, M. and Mena, R.H. (2016). A Harris process to model stochastic volatility. Submitted manuscript.
- Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-Gaussian Ornstein?Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B* **63**, 167–241.
- Banfield, J.D. and Raftery, A.E. (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics*, **7**, 1-34.
- Bissiri, P. and Ongaro, A. (2014) On the topological support of species sampling priors. *Electron. J. Statist.*, **8**, 861–882.
- Blackwell, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1**, 356–358.
- Blackwell, D., and MacQueen, J. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355.
- Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics.* **69**, 542–547.
- Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. in Appl. Probab.*, **31**, 929–953.

References. . .

- Bühlmann, H. (1960). Austauschbare stochastische Variablen und ihre Grenzwertsätze *PhD thesis*, University of California, Berkeley.
- *****De Blasi, P., Favaro, S., Lijoi, A., Mena, R., Prünster, I. and Ruggiero, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process?** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37, 212–229.**
- de Finetti, B. (1931). Le funzioni caratteristiche di legge istantanea dotate di valori eccezionali. *Atti Reale Accademia Nazionale dei Lincei, Serie, VI Rend.* **14**, 259–265.
- Escobar, M.D. (1988). Estimating the means of several normal populations by nonparametric estimation of the distribution of the means. *Unpublished Ph.D. dissertation, Department of Statistics, Yale University*
- Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *J. Am. Stat. Assoc.*, **89**, *268?-277*
- *Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. J. Amer. Stat. Assoc.,* **90**, *577–588.*
- *Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. Theor. Popul. Biol.,* ***3***, *87-112.*

References. . .

- Feng, S. (2010). *The Poisson-Dirichlet Distribution and Related Topics: Models and Asymptotic Behaviors.* Springer.
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.
- Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.
- Fuentes-García, R., Mena, R. H. and Walker, S. G. (2009). A nonparametric dependent process for Bayesian regression. *Statistics and Probability Letters*. **79**, 1112–1119.
- Fuentes-García, R., Mena, R. H. and Walker, S. G. (2010). A new Bayesian nonparametric mixture model. *Communications in Statistics-Simulation and Computation*. **39**, 669–682.
- Fuentes-García, R., Mena, R. H. and Walker, S. G. (2010). A probability for classification based on the mixture of Dirichlet process model. *Journal of Classification*. **27**, 389–403.
- * **Goldstein, M. (2013). Observables and models: exchaengeability and the inductive argument. In** *Bayesian Theory and Applications.* **Damien, P., Dellaportas, P., Polson, N. G. and Stephen, D.A. Eds. Oxford University Press.**
- Gnedin, A. and Pitman, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *J. Math. Sci. (N.Y.)* **138**, 5674-85.

References. . .

- Gutierrez Inostroza, L., Mena, R.H., Ruggiero, M. (2016). time dependent Bayesian nonparametric model for air quality analysis. *Computational Statistics and Data Analysis.* 95, 161-175.
- Gutierrez Inostroza, L., Mena, R.H., and Ruggiero, M. (2016). On GEM diffusive mixtures. In *JSM Proceedings 2016, Section on Nonparametric Statistics.* Alexandria, VA: American Statistical Association.
- Hewitt, E. and Savage, L. J. (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, **80**, 470–501.
- * **Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.*, 96, 161–173.**
- James, L.F., Lijoi, A., and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scandinavian Journal of Statistics*, **33**, 105–120
- * **James, L.F., Lijoi, A., and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments.** *Scandinavian Journal of Statistics*, **36, 76–97**
- Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability.* **33**, 664–77.
- Kallenberg, O. (1973). Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrsch. verw. Geb.* **27**, 23–36.

References. . .

- Kallenberg, O. (1975). Infinitely divisible processes with interchangeable increments and random measures under convolution. *Z. Wahrsch. verw. Geb.* **32**, 309–321.
- Kallenberg, O. (1990). Exchangeable random measures in the plane. *J. Theor. Probab.* **3**, 81–136.
- Kalli, M. and Griffin, J.E. and Walker, S.G. (2011). Slice sampling mixture models. *Statistics and Computing.* **21**, 93–105.
- ***Kingman, J.F.C. (1975). Random discrete distributions.** *J. Roy. Statist. Soc. Ser. B*, **37, 1–22.**
- Kingman, J.F.C. (1978). The representation of partition structures. *Journal of the London Mathematical Society*, **18**, 374–380.
- Kingman, J.F.C. (1993). *Poisson processes* Oxford University Press.
- Lenk, P. J. (1988). The logistic normal distribution for Bayesian nonparametric, predictive densities. *J. Amer. Statist. Asoc.*, **83**, 509–516.
- Lijoi, A., Mena, R.H. and Prünster, I. (2005). Bayesian nonparametric estimation of the probability of discovering new species. *JASA*, **100**, 1278–1291.
- * **Lijoi, A., Mena, R.H.** and **Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models.** *J. R. Statist. Soc. B*, **69, 715–740.**
- Lijoi, A. and Prünster, I. (2010). Models Beyond the Dirichlet process. In *Bayesian nonparametrics.* (Eds. Hjort, N., Holmes, C., Müller, P. and Walker, S.G.). Cambridge Univ. Press.

References. . .

- Lijoi, A., Prünster, I. and Walker, S.G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.*, **18**, 1519-1547.
- Lindley, D. V. and Novick, M. R. (1981). The role of exchangeability in inference. *Annals of Statistics.*, **9**, 45–58.
- Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I Density estimates. *Annals of Statistics*, **12**, 351–357.
- MacEachern, S.N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Commun. Statist. Simulation Comp.*, **23**, 727–741.
- MacEachern, S.N. (1998). Computational methods for mixture of Dirichlet process models. In *Practical nonparametric and semiparametric Bayesian statistics* (eds D. Dey, P. Müller and D. Sinha). New York: Springer, 23–43.
- MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria: American Statistical Association, 50-55.
- * **Mena, R.H. (2013). Geometric Weight Priors and their Applications in Bayesian Nonparametrics. In *Bayesian Theory and Applications*. Damien, P., Dellaportas, P., Polson, N. G. and Stephen, D.A. Eds. Oxford University Press.**
- Mena, R. H. and Walker, S. G. (2005). Stationary autoregressive models via a Bayesian nonparametric approach. *Journal of Time Series Analysis*, 26, 789–805.

References. . .

- Mena, R. H. and Walker, S. G. (2007). On the stationary version of the generalized hyperbolic ARCH model. *Annals of the Institute of Statistical Mathematics*. 59, 325–348.
- Mena, R. H. and Walker, S. G. (2007). Stationary Mixture Transition Distribution (MTD) models via predictive distributions. *Journal of Statistical Planning and Inference*. 137, 3103–3112.
- Mena, R. H. and Walker, S. G. (2009). Construction of Markov processes in continuous time. *Metron*. 67, 303-323.
- ***Mena, R.H., Ruggiero, M. and Walker, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *Journal of Statistical Planning and Inference*, 141, 3217-3230.**
- Mena, R.H., Ruggiero, M. (2016). Dynamic density estimation with diffusive Dirichlet mixtures. *Bernoulli*. 22, 901-926.
- Mena, R.H. and Walker, S. G. (2017). Bayesian mixtures of Feller processes.
- Müller, P. (2017). Nonparametric Bayesian Mixure Models. In *Handbook of Mixtures*. Frühwirth-Schnatter, S., Robert, C. and Celeux, G. (eds.), CRC-Press

References...

- Müller, P. Xu, Y. and Jara, A. (2016). A Short Tutorial on Bayesian Nonparametrics, *Journal of Statistical Research*, **48-50**, 1–19.
- Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102**, 145–158.
- Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in honor of David Blackwell* (Eds. Ferguson, T.S., *et al.*). Lecture Notes, Monograph Series, **30**, 245–267. Institute of Mathematical Statistics, Hayward.
- Quintana, F. A. and Iglesias, P. L. (2003). Bayesian clustering and product partition models. *J. R. Statist. Soc. B*, **65**, 557?-574.
- Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.*, **31**, 560–585.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–650.
- Walker, S.G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics*, **36**, 45–54.