An Introduction to Bayesian Nonparametric Inference.

Dario Spanò

MASDOC seminar - Statistical Frontiers, 2014

6 February, 2014
Motivation

Statistical models.

- Data $\mathbf{x} = (x_1, \ldots, x_n) \in E^n$ (say) modelled as a realization of $E$-valued random variables $X_1, \ldots, X_n$ whose joint distribution $P$ is unknown.
Motivation

Statistical models.

- Data $\mathbf{x} = (x_1, \ldots, x_n) \in E^n$ (say) modelled as a realization of $E$-valued random variables $X_1, \ldots, X_n$ whose joint distribution $P$ is unknown.

- Many possible assumptions on dependence structure:
  - $(X_i) \overset{iid}{\sim} P_\theta : \theta \in \Theta$;
  - $X_i = \phi X_{i-1} + \epsilon_i, \ i = 1, 2, \ldots$;
  - $dX_t = a(X_t) + b(X_t)dW_t, \ t \geq 0$;
  - $X_i = \alpha + \beta Y_i + \epsilon_i, \ i = 1, 2, \ldots$;

  etc...
Repeated sampling.

- Frequentist approach. Independent replications are assumed:

\[
\{X_i\}_{i=1}^n \overset{iid}{\sim} P_\theta, \ \theta \in \Theta.
\]

If parameter \( \theta \) known, \( P_\theta \) fully available.
Repeated sampling.

- Frequentist approach. Independent replications are assumed:

\[
\{X_i\}_{i=1}^n \overset{iid}{\sim} P_\theta, \quad \theta \in \Theta.
\]

If parameter \( \theta \) known, \( P_\theta \) fully available.

- Glivenko-Cantelli. As \( n \to \infty \),

\[
\sup_{A \in B(E)} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \in A) - P_\theta(X \in A) \right| \to 0, \quad \text{a.s.} P_\theta
\]
Repeated sampling.

- Frequentist approach. Independent replications are assumed:
  \[ \{X_i\}_{i=1}^n \overset{iid}{\sim} P_\theta, \; \theta \in \Theta. \]

  If parameter \( \theta \) known, \( P_\theta \) fully available.

  ▶ Glivenko-Cantelli. As \( n \to \infty \),
  \[
  \sup_{A \in \mathcal{B}(E)} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \in A) - P_\theta(X \in A) \right| \to 0, \quad a.s. P_\theta
  \]

  ▶ If \( T_n = T(X_1, \ldots, X_n) \) is an estimator for \( \theta \) and a sufficient statistic, then
  \[ T_n \to \theta, \quad a.s. P_\theta, \forall \theta \in \Theta \]
Classical Parametric vs non-parametric

Glivenko-Cantelli is independent of the dimension of $\Theta$. We can even take

$$\Theta = \mathcal{M}_1(E) := \{\text{all probability measures on } E\}.$$ 

Note that, under iid $(P)$ assumption $(P \in \mathcal{M}_1(E)$ unknown),

$$T_n := \frac{1}{n} \sum_{1}^{n} \delta_{X_n}$$

is a sufficient statistic for $P$ (1:1 function of order statistics), or for $P_\theta$ in the parametric case.
Predictive probabilities.

Thus, under the iid assumption, \( T_n = \frac{1}{n} \sum_{1}^{n} \delta_{X_n} \) is a good estimator for \( P \). However, suppose

\[
(X_1, X_2, \ldots) \overset{iid}{\sim} \text{ber}(\theta) : \theta \in (0, 1)
\]

Suppose data are \((1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1)\). How do you predict, based on such data, the value for \( X_{n+1} \)?

- \( X_{n+1} = 1 \) with probability \( T_n\{1\} \) (proportion of of ones in the past);
- It should be, by iid property, \( \mathbb{P}[X_{n+1} = 1] = \mathbb{P}[X_1 = 1] \) and \( X_{n+1} \) independent of data.
Predictive probabilities.

Thus, under the iid assumption, \( T_n = \frac{1}{n} \sum_{1}^{n} \delta X_n \) is a good estimator for \( P \). However, suppose

\[
(X_1, X_2, \ldots) \overset{iid}{\sim} ber(\theta) : \theta \in (0, 1)
\]

Suppose data are \((1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1)\). How do you predict, based on such data, the value for \( X_{n+1} \)?

- \( X_{n+1} = 1 \) with probability \( T_n \{ 1 \} \) (proportion of of ones in the past);
- It should be, by iid property, \( \mathbb{P}[X_{n+1} = 1] = \mathbb{P}[X_1 = 1] \) and \( X_{n+1} \) independent of data.
- Difficulties in assessing the predictive distribution of the first outcome \( X_1 \).
Johnson’s postulate

“The probability with which any value appears at the next sample step should be proportional to the number of times it has appeared in the past”

Figure: Zabell (1982)
Example: two-color Pólya urn.

Let $\alpha > 0$ and $\beta > 0$ represent the number of balls colored, respectively, in white and black, contained in an urn.

Rule: at each sample stage $n \geq 1$: pick out a ball, note its color $X_n$, and return it into the urn along with an additional ball of the same color.
Example: two-color Pólya urn.

Let $\alpha > 0$ and $\beta > 0$ represent the number of balls colored, respectively, in white and black, contained in an urn.

Rule: at each sample stage $n \geq 1$: pick out a ball, note its color $X_n$, and return it into the urn along with an additional ball of the same color.

- $\mathbb{P}[X_1 = 1] = \frac{\alpha}{\alpha + \beta}$;

- For $n \geq 1$, 
  
  $$
  \mathbb{P}[X_{n+1} = 1 \mid X_1 = x_1, \ldots, X_n = x_n] = \frac{n}{\alpha + \beta + n} t_n + \frac{\alpha + \beta}{\alpha + \beta + n} \left( \frac{\alpha}{\alpha + \beta} \right),
  $$

  where $t_n = \frac{1}{n} \sum_{1}^{n} \delta_{x_i}$.

  Johnson’s postulate satisfied!
Beta prior, Binomial likelihood

After $n$ observations,

$$
\mathbb{P} [X_1 = x_1, \ldots, X_n = x_n] = \frac{\alpha (nt_n) \beta (1-t_n)}{(\alpha + \beta)_n} \\
= \int_0^1 p^{nt_n} (1-p)^{n(1-t_n)} \pi_{\alpha,\beta} (dp).
$$

Here, $a(x) = \Gamma(a + x)/\Gamma(a)$ (generalized incr. factorials) and

$$
\pi_{\alpha,\beta} (dp) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} p^{\alpha-1} (1 - p)^{\beta - 1} \mathbb{I}(p \in (0, 1)) dp.
$$
Beta prior, Binomial likelihood

After $n$ observations,

$$\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \frac{\alpha(n t_n) \beta n(1-t_n)}{(\alpha + \beta)(n)}$$

$$= \int_0^1 p^{nt_n} (1-p)^{n(1-t_n)} \pi_{\alpha, \beta}(dp).$$

Here, $a(x) = \Gamma(a + x)/\Gamma(a)$ (generalized incr. factorials) and

$$\pi_{\alpha, \beta}(dp) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \mathbb{1}(p \in (0, 1)) dp.$$

- $(T_n : n \geq 1)$ u.i. martingale. Then $P := \lim_{n \to \infty} T_n$ almost surely exists and its law is $\pi_{\alpha, \beta}$. 
Repeated sampling

- Bayesian approach. Conditionally independent replications are assumed:

\[ \{X_i\}_{i=1}^n \mid \theta \overset{iid}{\sim} P_\theta, \quad \theta \sim \pi \quad \text{(prior)}. \]
Repeated sampling

- Bayesian approach. Conditionally independent replications are assumed:

\[ \{X_i\}_{i=1}^n \mid \theta \sim P_\theta, \quad \theta \sim \pi \quad \text{(prior)}. \]

Thus

\[ \mathbb{P} [X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\Theta} \prod_{i=1}^{n} P_\theta(A_i) \pi(d\theta) \]
Repeated sampling

- Bayesian approach. Conditionally independent replications are assumed:
  \[
  \{X_i\}_{i=1}^n \mid \theta \overset{iid}{\sim} P_\theta, \quad \theta \sim \pi \quad (\text{prior}).
  \]

  Thus
  \[
  \mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\Theta} \prod_{i=1}^n P_\theta(A_i) \pi(d\theta)
  \]

- Posterior:
  \[
  \pi^*(\theta \in B \mid X = x) \propto \left[ \prod_{i=1}^n P_\theta(dx_i) \right] \times \pi(d\theta)
  \]
  posterior $\propto$ likelihood $\times$ prior
Existence.

- Typical criticism: How can you justify the presence (existence) of a random parameter? Prior taken out of the blue.
Existence.

- Typical criticism: How can you justify the presence (existence) of a random parameter? Prior taken out of the blue.
- A proper justification is available at least for replications whose distribution does not depend on the order in which data are collected: *exchangeable data*. 
Existence.

- Typical criticism: How can you justify the presence (existence) of a random parameter? Prior taken out of the blue.
- A proper justification is available at least for replications whose distribution does not depend on the order in which data are collected: exchangeable data.
- General assumptions:
  (a) Experimental observation can be repeated indefinitely in the same condition;
  (b) The order of appearance of experimental outcomes does not affect their joint distribution.
de Finetti’s representation theorem

**Definition**

A sequence of r.v.’s \( \{ X_i : i \geq 1 \} \) is (infinitely) exchangeable if, for every \( n \),

\[
(X_1, \ldots, X_n) \overset{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)})
\]

for any permutation \( \sigma \) of \( \{1, \ldots, n\} \).
de Finetti’s representation theorem

**Definition**

A sequence of r.v.’s \( \{X_i : i \geq 1\} \) is (infinitely) exchangeable if, for every \( n \),

\[
(X_1, \ldots, X_n) \overset{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)})
\]

for any permutation \( \sigma \) of \( \{1, \ldots, n\} \).

**Theorem (B. de Finetti 1931)**

A seq. of binary r.v.s \( \{X_i : i \geq 1\} \) is infinitely exchangeable iff there exists a distribution \( \pi \) on \( (0, 1) \) such that

\[
\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \int_0^1 p^{nt_n}(1-p)^{n(1-t_n)} \pi(dp)
\]

where \( t_n := n^{-1} \sum_{i=1}^n x_i \).
de Finetti’s representation theorem

**Definition**

A sequence of r.v.’s \( \{X_i : i \geq 1\} \) is (infinitely) exchangeable if, for every \( n \),

\[
(X_1, \ldots, X_n) \overset{d}{=} (X_{\sigma(1)}, \ldots, X_{\sigma(n)})
\]

for any permutation \( \sigma \) of \( \{1, \ldots, n\} \).

**Theorem (B. de Finetti 1931)**

A seq. of binary r.v.s \( \{X_i : i \geq 1\} \) is infinitely exchangeable iff there exists a distribution \( \pi \) on \((0, 1)\) such that

\[
\mathbb{P}[X_1 = x_1, \ldots, X_n = x_n] = \int_0^1 p^{nt_n} (1 - p)^{n(1-t_n)} \pi(dp)
\]

where \( t_n := n^{-1} \sum_{i=1}^n x_i \).

The distribution \( \pi \) (prior) is uniquely characterized as the distribution of the (a.s.) random limit \( P := \lim_{n \to \infty} T_n \).
Observables.

Bayesian notions of priors and posterior are thus justified by assumptions (a) and (b) on observables (i.e. not out of the blue):

- $(X_i) \mid P \overset{iid}{\sim} \text{ber}(P)$ a.s. (*conditional likelihood*)
- $P \sim \pi$ (*prior*)
- 1:1 correspondence between the prior $\pi$ on $P$ via a system of prediction rules

\[
\mathbb{P} [X_1 = 1];
\]

\[
\mathbb{P} [X_{n+1} = 1 \mid X_1 = x_1, \ldots, X_n = x_n], \quad n \geq 1
\]

(Ionescu-Tulcea).
de Finetti’s Representation Theorem in general dimensions.

**Theorem (Hewitt and Savage, 1955)**

Let $E$ be a Polish space endowed with its Borel $\sigma$-field $\mathcal{E}$. A sequence of r.v.’s $\{X_i : i \geq 1\} \in E^\infty$ is (infinitely) exchangeable if and only if there exists a distribution $\pi$ on $\mathcal{M}_1(E)$ such that

$$
\mathbb{P}[X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\mathcal{M}_1(E)} \left[ \prod_{i=1}^{n} P(A_i) \right] \pi(dP)
$$
de Finetti’s Representation Theorem in general dimensions.

**Theorem (Hewitt and Savage, 1955)**

Let $E$ be a Polish space endowed with its Borel $\sigma$-field $\mathcal{E}$. A sequence of r.v.’s $\{X_i : i \geq 1\} \in E^\infty$ is (infinitely) exchangeable if and only if there exists a distribution $\pi$ on $\mathcal{M}_1(E)$ such that

$$
P[X_1 \in A_1, \ldots, X_n \in A_n] = \int_{\mathcal{M}_1(E)} \left[ \prod_{i=1}^{n} P(A_i) \right] \pi(dP)
$$

- The distribution $\pi$ (prior) is uniquely characterized as the distribution of the (a.s.) random limit $P := \lim_{n \to \infty} T_n$
- $P^\otimes n := \prod_{i=1}^{n} P(dx_i)$ (likelihood) is a version of the conditional probability of $(X_1, \ldots, X_n)$, given $P$. 
de Finetti’s Representation Theorem in general dimensions.

**Theorem (Hewitt and Savage, 1955)**

Let $E$ be a Polish space endowed with its Borel $\sigma$-field $\mathcal{E}$. A sequence of r.v.’s $\{X_i : i \geq 1\} \in E^\infty$ is (infinitely) exchangeable if and only if there exists a distribution $\pi$ on $\mathcal{M}_1(E)$ such that

$$
P \left[ X_1 \in A_1, \ldots, X_n \in A_n \right] = \int_{\mathcal{M}_1(E)} \left[ \prod_{i=1}^n P(A_i) \right] \pi(dP)$$

- The distribution $\pi$ (prior) is uniquely characterized as the distribution of the (a.s.) random limit $P := \lim_{n \to \infty} T_n$
- $P \otimes^n := \prod_{i=1}^n P(dx_i)$ (likelihood) is a version of the conditional probability of $(X_1, \ldots, X_n)$, given $P$.
- In other words,

$$(X_i) \mid P \overset{iid}{\sim} P; \quad P \sim \pi.$$
de Finetti and Bayes via predictive laws.

- Bijection between prior $\pi$ and predictive distributions
  $\mathbb{P} [X_{n+1} \mid X_1, \ldots, X_n], n \geq 1$.
- Distribution of $X_1$ is the prior mean:
  
  $P_0(A) := \mathbb{P} (X_1 \in A) = \mathbb{E}_\pi [P(A)]$
de Finetti and Bayes via predictive laws.

- Bijection between prior $\pi$ and predictive distributions
  \[ \mathbb{P} [X_{n+1} \mid X_1, \ldots, X_n], \ n \geq 1. \]
- Distribution of $X_1$ is the prior mean:
  \[ P_0(A) := \mathbb{P} (X_1 \in A) = \mathbb{E}_\pi [P(A)] \]
- Distribution of $X_{n+1}$ given $X_1, \ldots, X_n$ is posterior mean:
  \[
  \mathbb{P} (X_{n+1} \in A_{n+1} \mid X_1 \in A_1, \ldots, X_n \in A_n) = \frac{\mathbb{E}_\pi \left[ \prod_{i=1}^{n+1} P(A_i) \right]}{E_\pi \left[ \prod_{i=1}^{n} P(A_i) \right]} = E_{\pi_{A(n)}} [P(A)],
  \]
de Finetti and Bayes via predictive laws.

- Bijection between prior $\pi$ and predictive distributions
  $$P [X_{n+1} \mid X_1, \ldots, X_n], n \geq 1.$$
- Distribution of $X_1$ is the prior mean:
  $$P_0(A) := P (X_1 \in A) = \mathbb{E}_\pi [P(A)]$$

- Distribution of $X_{n+1}$ given $X_1, \ldots, X_n$ is posterior mean:
  $$P (X_{n+1} \in A_{n+1} \mid X_1 \in A_1, \ldots, X_n \in A_n) = \frac{\mathbb{E}_\pi \left[ \prod_{i=1}^{n+1} P(A_i) \right]}{\mathbb{E}_\pi \left[ \prod_{i=1}^{n} P(A_i) \right]} = E_{\pi_A^{(n)}} [P(A)],$$

where (Bayes’ Theorem):

$$\pi_A^{(n)} (dP) = \frac{\left[ \prod_{i=1}^{n+1} P(A_i) \right] \pi (dP)}{\mathbb{E}_\pi \left[ \prod_{i=1}^{n} P(A_i) \right]} = \frac{L(data \mid P) \times prior}{marg. \ likelihood}$$

is the posterior distribution of $P$ given data in $A^{(n)} := A_1 \times \cdots \times A_n$. 
Predictive distributions and posterior laws.

- If \((X_n : n \geq 1)\) infinitely exchangeable, then \(T_n(X^{(n)}) \Rightarrow \pi\).
Predictive distributions and posterior laws.

- If \((X_n : n \geq 1)\) infinitely exchangeable, then \(T_n(X^{(n)}) \Rightarrow \pi\).
- If \((X_n : n \geq 1)\) infinitely exchangeable, then also

\[
(X_{n+m} : m \geq 1 \mid X^{(n)} = x^{(n)})
\]

is infinitely exchangeable.
Predictive distributions and posterior laws.

- If \( (X_n : n \geq 1) \) infinitely exchangeable, then \( T_n(X^{(n)}) \Rightarrow \pi \).
- If \( (X_n : n \geq 1) \) infinitely exchangeable, then also
  \[
  (X_{n+m} : m \geq 1 \mid X^{(n)} = x^{(n)})
  \]
  is infinitely exchangeable.
- Thus, given \( X^{(n)} = x^{(n)} \) the conditional law of
  \[
  \lim_{m \to \infty} T_m(X^{(n:m)})
  \]
  where \( X^{(n:m)} := (X_{n+j} : j = 1, \ldots, m) \), is precisely \( \pi_{x^{(n)}}(dP) \), the posterior probability of \( P \) given data.
Predictive distributions and posterior laws.

- If \((X_n : n \geq 1)\) infitly exchble, then \(T_n(X^{(n)}) \Rightarrow \pi\).
- If \((X_n : n \geq 1)\) infitly exchble, then also

\[
(X_{n+m} : m \geq 1 | X^{(n)} = x^{(n)})
\]

is infitly exchble.
- Thus, given \(X^{(n)} = x^{(n)}\) the conditional law of

\[
\lim_{m \to \infty} T_m(X^{(n:m)})
\]

where \(X^{(n:m)} := (X_{n+j} : j = 1, \ldots, m)\), is precisely \(\pi_{x^{(n)}}(dP)\), the posterior probability of \(P\) given data.
- One-to-one correspondence between \(\pi_{x^{(n)}}(dP)\) and

\[
\mathbb{P} \left( X_{m+1}^{n:\infty} | X_1^{n:\infty}, \ldots, X_m^{n:\infty}; x_1, \ldots, x_n \right).
\]
Clustering and predictive distributions.

Assume an infinite population is partitioned into $d$ distinct species or clustering classes. Unknown proportions $(P_j : j = 1, \ldots, d)$ of species $j \in \{1, \ldots, d\}$, respectively.
Clustering and predictive distributions.

Assume an infinite population is partitioned into \( d \) distinct species or clustering classes. Unknown proportions \((P_j : j = 1, \ldots, d)\) of species \( j \in \{1, \ldots, d\} \), respectively.

Set \( X_i = j \) iff the individual \( X_i \) is of species \( j \).
Cluster and predictive distributions.

Assume an infinite population is partitioned into \( d \) distinct species or clustering classes. Unknown proportions \((P_j : j = 1, \ldots, d)\) of species \( j \in \{1, \ldots, d\} \), respectively.

Set \( X_i = j \) iff the individual \( X_i \) is of species \( j \).

Pólya sequences naturally extend to \( d \)-color urn schemes. Exchangeability + Johnson’s postulate respected. Generating prior on

\[
\Delta_{(d-1)} := \left\{ (x_1, \ldots, x_d) \in [0, 1]^d : \sum_j x_j = 1 \right\}
\]
Pólya urn in higher dimensions.

Same procedure with $d$ colors. Let $\alpha_i = n.$ balls of color $i$ ($i = 1, \ldots, d$) and $\theta = \sum_{j=1}^{d} \alpha_i$.

$$\mathbb{P}[X_1 = 1] = \frac{\alpha_i}{\theta};$$

$$\mathbb{P}[X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n] = \frac{n}{\theta + n} t_n(i) + \frac{\theta}{\theta + n} \left( \frac{\alpha_i}{\theta} \right),$$
Pólya urn in higher dimensions.

Same procedure with $d$ colors. Let $\alpha_i = n$ balls of color $i$ ($i = 1, \ldots, d$) and $\theta = \sum_1^d \alpha_i$.

\[ P[X_1 = 1] = \frac{\alpha_i}{\theta}; \]

\[ P[X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n] = \frac{n}{\theta + n} t_n(i) + \frac{\theta}{\theta + n} \left( \frac{\alpha_i}{\theta} \right), \]

$\theta$ = how much importance I give to my prior opinions.
Pólya urn in higher dimensions.

Same procedure with $d$ colors. Let $\alpha_i = n$. balls of color $i$ ($i = 1, \ldots, d$) and $\theta = \sum_1^d \alpha_i$.

\[
P[X_1 = 1] = \frac{\alpha_i}{\theta};
\]

\[
P[X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n] = \frac{n}{\theta + n} t_n(i) + \frac{\theta}{\theta + n} \left(\frac{\alpha_i}{\theta}\right),
\]

- $\theta = \text{how much importance I give to my prior opinions.}$
- $\text{As } n \to \infty, \ T_n \Rightarrow \pi_{\alpha_1, \ldots, \alpha_d}$

\[
\pi_{\alpha_1, \ldots, \alpha_d}(dx) \propto \left[ \prod_{j=1}^{d-1} x_j^{\alpha_j - 1} \right] (1 - |x|)^{\alpha_d - 1} \mathbb{1}(x \in \Delta_{d-1}) \, dx_1 \cdots dx_{d-1}.
\]
Dirichlet distribution.

- Let $\Gamma_j \sim \text{Ga}(\alpha_j, \beta)$ and $(\Gamma_j : j = 1, \ldots, d)$ independent $(\alpha_i > 0, \beta > 0)$.
- Then $\Gamma := \sum_{j=1}^{d} \Gamma_j \sim \text{Ga}(\theta, \beta)$, where $\theta = \sum_j \alpha_j$.
- Moreover, the vector $\mathbf{P} = (P_1, \ldots, P_d)$ defined by
  \[ P_i := \frac{\Gamma_j}{\sum_{j=1}^{d} \Gamma_j} : j = 1, \ldots, d \]
  is independent of $\Gamma$ and has distribution $\pi_{\alpha_1, \ldots, \alpha_d}$. 
Dirichlet distribution properties.

- $\mathbb{E}[P_j] = \alpha_j/\theta$ for every $j = 1, \ldots, d$ (cf distribution of $X_1$ in polya urn);
- $\text{Var} P_j = \frac{1}{\theta+1} \frac{\alpha_j}{\theta} (1 - \frac{\alpha_j}{\theta^2})$;
- Joint moments:
  
  $$
  \mathbb{E} \left[ \frac{n!}{\prod_{i=1}^d n_j!} \prod_{i=1}^d P_j^{n_j} \right] = \frac{n!}{\prod_{i=1}^d n_j!} \frac{\prod_{i=1}^d (\alpha_j)(n_j)}{(\theta)(n)} =: DM_{\alpha_1,\ldots,\alpha_d}(n_1, \ldots, n_d)
  $$

  $\rightarrow$ **Dirichlet-Multinomial distribution**: distribution of sample cluster partition from Pólya urn !!!!

  $$
  DM_{\alpha_1,\ldots,\alpha_d}(n_1, \ldots, n_d) = \mathbb{P} \left( \frac{n_j}{n} : j = 1, \ldots, d \right)
  $$
Dirichlet distribution properties.

- $\mathbb{E}[P_j] = \frac{\alpha_j}{\theta}$ for every $j = 1, \ldots, d$ (cf distribution of $X_1$ in Polya urn);
- $\text{Var} P_j = \frac{1}{\theta + 1} \frac{\alpha_j}{\theta} (1 - \frac{\alpha_j}{\theta^2})$;
- Joint moments:

$$
\mathbb{E} \left[ \frac{n!}{\prod_{i=1}^d n_j!} \prod_{i=1}^d P_j^{n_j} \right] = \frac{n!}{\prod_{i=1}^d n_j!} \frac{\prod_{i=1}^d (\alpha_j)(n_j)}{(\theta)(n)} =: DM_{\alpha_1, \ldots, \alpha_d}(n_1, \ldots, n_d)
$$

$\rightarrow$ Dirichlet-Multinomial distribution: distribution of sample cluster partition from Polya urn !!!!

$$
DM_{\alpha_1, \ldots, \alpha_d}(n_1, \ldots, n_d) = \mathbb{P} \left( T_n\{j\} = \frac{n_j}{n} : j = 1, \ldots, d \right)
$$

- Posterior distribution, given $nT_n = (n_1, \ldots, n_d)$:

$$
\pi_{\alpha_1 + n_1, \ldots, \alpha_d + n_d}(dp) = \frac{\frac{n!}{\prod_{i=1}^d n_j!} \prod_{i=1}^d P_j^{n_j}}{DM_{\alpha_1, \ldots, \alpha_d}(n_1, \ldots, n_d)} \pi_{\alpha_1, \ldots, \alpha_d}(dp)
$$
Prior on partitions

We have seen that Pólya urns generate priors $\pi$ on the parameter $p$ of conditionally iid samples:

$$(X_1, \ldots) \mid p \sim p,\quad (N_1, \ldots, N_d) \mid p \sim \text{Multinomial}(|N|; p).$$

and of the induced conditionally multinomial cluster partitions

This is true for every infinitely exchangeable sequence $(X_1, X_2, \ldots)$ in $E = \{1, \ldots, d\}$, i.e. for any choice of prior !!!!
Prior on partitions

We have seen that Pólya urns generate priors $\pi$ on the parameter $p$ of conditionally iid samples:

$$(X_1, \ldots) \mid p \sim p,$$

and of the induced conditionally multinomial cluster partitions

$$(N_1, \ldots, N_d) \mid p \sim \text{Multinomial}(|N|; p).$$

- This is true for every infinitely exchangeable sequence $(X_1, X_2, \ldots)$ in $E = \{1, \ldots, d\}$, i.e. for any choice of prior !!!!

Specificities of Dirichlet:
  - Johnson’s postulate preserved
  - Conjugate wrt model (i.e. posterior is again Dirichlet);
  - Absolute continuity on the simplex
Prior on partitions

We have seen that Pólya urns generate priors $\pi$ on the parameter $p$ of conditionally iid samples:

$$(X_1, \ldots) \mid p \overset{iid}{\sim} p,$$

and of the induced conditionally multinomial cluster partitions

$$(N_1, \ldots, N_d) \mid p \sim \text{Multinomial}(|N|; p).$$

- This is true for every infinitely exchangeable sequence $(X_1, X_2, \ldots)$ in $E = \{1, \ldots, d\}$, i.e. for any choice of prior !!!!

- Specificities of Dirichlet:
  - Johnson’s postulate preserved
  - Conjugate wrt model (i.e. posterior is again Dirichlet);
  - Absolute continuity on the simplex

- What about $d \to \infty$ ?
“Predicting the unpredictable (Zabell (1992))”.

Suppose we sample sequentially items and classify them according to their species. Suppose, however, we do not know in advance $d$, i.e., how many species there are in the population. We want to retain exchangeability, and Johnson’s postulate.

- $E =$ continuous space of colors (diagonal-measurable)
- $\tilde{E} = \{1, 2, \ldots\}$ space of labels (labelling is immaterial)
Suppose we sample sequentially items and classify them according to their species. Suppose, however, we do not know in advance \(d\) i.e. how many species there are in the population. We want to retain exchangeability, and Johnson’s postulate.

- \(E\) = continuous space of colors (diagonal-measurable)
- \(\tilde{E} = \{1, 2, \ldots\}\) space of labels (labelling is immaterial)
- Map
  \[
  (x_1, x_2, \ldots) \in E^\infty \mapsto (y_1, y_2, \ldots) \in \tilde{E}^\infty
  \]
  via \(y_1 = 1\) and
  \[
  y_n = \inf\{j \leq n : X_n = X_j\}.
  \]
- Define partition \(\pi = \Pi(x)\) of \(\mathbb{N}\) induced by \(x \in E^\infty\) as the one where each integer \(i\) is in the \(j-th\) block of \(\pi\) if and only if \(y_i = j\) (order of appearance of blocks).
“Predicting the unpredictable (Zabell (1992))”.

Suppose we sample sequentially items and classify them according to their species. Suppose, however, we do not know in advance $d$ i.e. how many species there are in the population. We want to retain exchangeability, and Johnson’s postulate.

- $E$ = continuous space of colors (diagonal-measurable)
- $\tilde{E} = \{1, 2, \ldots\}$ space of labels (labelling is immaterial)
- Map
  $$(x_1, x_2, \ldots) \in E^\infty \mapsto (y_1, y_2, \ldots) \in \tilde{E}^\infty$$
  via $y_1 = 1$ and
  $$y_n = \inf\{j \leq n : X_n = X_j\}.$$
- Define partition $\pi = \Pi(x)$ of $\mathbb{N}$ induced by $x \in E^\infty$ as the one where each integer $i$ is in the $j – th$ block of $\pi$ if and only if $y_i = j$ (order of appearance of blocks).
Partition structure.

By construction,

\[ \Pi_n(y) = \Pi_n(x) \]

Example:

\[ x^{(n)} = ABBACABBBABCD \mapsto y^{(n)} = 12213211234 \]

\[ \Pi_n = \{(1, 4, 6, 9), (2, 3, 7, 8, 10), (5, 11), (12)\} \]

\[ N_n = (4, 5, 2, 1) \]
Measures and partitions.

- consistency of \((x^{(n)} = (x_1, \ldots, x_n) : n \geq 1)\) implies consistency of the induced partition
  \[\pi_n = (\pi_1, \ldots, \pi_{k_n}), \quad n \geq 1\]
  of \(\{1, \ldots, n\}\), where \(k_n = \max\{y_1, \ldots, y_n\}\).
- Let \(n_n = (n_1, \ldots, n_{n,k_n})\) be the block sizes of \(\pi_n\). For every \(x \in E^\infty\), both limits are well defined:
  \[p(\cdot)(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(\cdot) \in \mathcal{M}_1(E),\]
  \[p(x) = \lim_{n \to \infty} \frac{n_n}{n} \in \Delta_\infty.\]
Exchangeable random partitions.

- 1:1 correspondence between $\pi$ on $\mathcal{M}_1(E)$ and exchangeable law on $E^\infty$
- Let $X \in E^\infty$ be exchangeable with de Finetti measure $\pi$. What is the induced measure on the induced partition $p(X) \in \Delta_\infty$?
- Interest in: number of non-zero frequencies, skewness/diversity, gaps, etc..
- Interest in: predicting waiting times to next $k^*$ new distinct species given past observations, etc,
Blackwell and MacQueen/Hoppe urns.

Construct an infinitely many colors Polya sequence. Fix $\theta > 0$ and $P_0 \in \mathcal{M}_1(E)$ non-atomic, where $E =$ continuous space of colors. Set:

$$P(X_1 \in A) = P_0(A);$$

$$P(X_{n+1} \in A \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\theta}{\theta + n} P_0(A) + \frac{n}{n + \theta} T_n(A),$$

$(A \in \mathcal{B}(E))$. 
Blackwell and MacQueen/Hoppe urns.

Construct an infinitely many colors Polya sequence. Fix $\theta > 0$ and $P_0 \in \mathcal{M}_1(E)$ non-atomic, where $E= \text{continuous space of colors}$. Set:

$$
P( X_1 \in A ) = P_0(A);$$

$$
P(X_{n+1} \in A \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\theta}{\theta + n} P_0(A) + \frac{n}{n + \theta} T_n(A),$$

$(A \in \mathcal{B}(E))$.

Then

$$
P( Y_1 = 1 ) = 1;$$

$$
P(Y_{n+1} = j \mid Y_2 = i_2, \ldots, Y_n = i_n) = \frac{\theta}{\theta + n} \mathbb{I}(j = k_n + 1) + \frac{n}{n + \theta} \frac{n_j}{n} \mathbb{I}(j = k_n),$$

$(j = 1, \ldots, k_n + 1)$. 
Exchangeable partition function.

Let $X_1, \ldots, X_n$ be the first $n$ observations from a Pólya-BMQ sequence $(\theta, P_0)$, and $\Pi_n$ the induced partition on $[n] = \{1, \ldots, n\}$. Then

$$P(X_1 \in dx_1, \ldots, X_n \in dx_n) = \left[ \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)! \right] \prod_{j=1}^{k} P_0(dx^*_k).$$

$$P(\Pi_n = (\pi_{n,1}, \ldots, \pi_{n,k})) = \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)!$$
Exchangeable partition function.

Let $X_1, \ldots, X_n$ be the first $n$ observations from a Pólya-BMQ sequence $(\theta, P_0)$, and $\Pi_n$ the induced partition on $[n] = \{1, \ldots, n\}$. Then

\[
P(X_1 \in dx_1, \ldots, X_n \in dx_n) = \left[ \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)! \right] \prod_{j=1}^{k} P_0(dx^*_k).
\]

\[
P(\Pi_n = (\pi_{n,1}, \ldots, \pi_{n,k})) = \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)!
\]

\[
P(K_n = k) = \frac{\theta^k}{\theta(n)} |S(n, k)|
\]

where $|S(n, k)| = \text{unsigned Stirling numbers}$. 


Exchangeable partition function.

Let $X_1, \ldots, X_n$ be the first $n$ observations from a Pólya-BMQ sequence $(\theta, P_0)$, and $\Pi_n$ the induced partition on $[n] = \{1, \ldots, n\}$. Then

$$
\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) = \left[ \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)! \right] \prod_{j=1}^{k} P_0(dx_j^*).
$$

$$
\mathbb{P}(\Pi_n = (\pi_{n,1}, \ldots, \pi_{n,k})) = \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)!.
$$

$$
\mathbb{P}(K_n = k) = \frac{\theta^k}{\theta(n)} |S(n, k)|
$$

where $|S(n, k)| =$ unsigned Stirling numbers.

Distribution of $\Pi_n$ given $K_n$ does not depend on $\theta$. $X^{(n)}$ given $\Pi_n$ depends only on $P_0$. Frequencies and labels are independent.
Dirichlet Process.

Let $X$ be a BMQ sequence $(\theta, P_0)$. Then

(i) the empirical distribution function $T_n = T_n(X)$ converges a.s. to a random measure $F$ on $E$ whose finite-dimensional distributions are, for every $d \in \mathbb{N}$ and every partition $A_1, \ldots, A_d$ of $E$,

$$(F(A_1), \ldots, F(A_d)) \sim \pi_{\alpha_1, \ldots, \alpha_d}$$

where $\alpha_i := \theta P_0(A_i), \ i = 1, \ldots, d$.

(Proof: binning urn scheme reduces to $d$-color Pólya urns).
Dirichlet Process.

Let $X$ be a BMQ sequence $(\theta, P_0)$. Then

(i) the empirical distribution function $T_n = T_n(X)$ converges a.s. to a random measure $F$ on $E$ whose finite-dimensional distributions are, for every $d \in \mathbb{N}$ and every partition $A_1, \ldots, A_d$ of $E$,

$$(F(A_1), \ldots, F(A_d)) \sim \pi_{\alpha_1, \ldots, \alpha_d}$$

where $\alpha_i := \theta P_0(A_i), \ i = 1, \ldots, d$.

(Proof: binning urn scheme reduces to $d$-color Pólya urns).

**Definition**

The RPM $F$ above is called a Dirichlet process with parameter $(\theta, P_0)$. 

Normalized (time-changed) subordinators.

Let $E = \mathbb{R}$. Define with $(S_t : t \geq 0)$ a subordinator with Lévy measure $\nu(\mathbb{R}_+) = +\infty$ and, for a finite measure $\alpha$ on $\mathbb{R}$ s.t. $\alpha(\mathbb{R}) = \theta$, define $A(x) = \alpha(-\infty, x]$. If

$$\nu(dx) = \frac{e^{-x}}{x} dx$$

Then the random probability measure

$$F((-\infty, x]) := \frac{S_{A(x)}}{S_\theta}$$

is a Dirichlet Process with parameter $(\theta, P_0)$, where $P_0(\cdot) = \alpha(\cdot)/\theta$. (Ferguson 1973).
Dirichlet Process.

Let $\mathbf{X}$ be a BMQ sequence $(\theta, P_0)$. Then

(ii) The number of distinct values (blocks) $K_n$ scales as $K_n \sim \theta \log n$. 

F is therefore almost surely atomic. In particular,

$F = \sum_{j=1}^{\infty} P_j \delta_{X^*_j}$

where $X^*_j \overset{iid}{\sim} P_0$ independent of $(P_j)_{j \geq 1}$ defined by

$B_j := P_j / \left(1 - j - 1 \sum_{i=1}^{j} P_i \right) \overset{iid}{\sim} \text{beta}(1, \theta)$.
Dirichlet Process.

Let $\mathbf{X}$ be a BMQ sequence $(\theta, P_0)$. Then

(ii) The number of distinct values (blocks) $K_n$ scales as $K_n \sim \theta \log n$.

(iii) Every distinct value $X_j^*$, appearing in $\mathbf{X}$, occurs infinitely often in $\mathbf{X}$ almost surely. $F$ is therefore almost surely atomic. In particular,

$$F = \sum_{j=1}^{\infty} P_j \delta_{X_j^*}$$

where

- $(X_j^*) \overset{iid}{\sim} P_0$ independent of $(P_j)$
- $P_j = \lim_n N_{n,j} / n$ where $N_{n,j} = \sum_{i=1}^{n} \mathbb{I}(Y_i = j)$
- $(B_j)_{j \geq 1}$ defined by

$$B_j := \frac{P_j}{1 - \sum_{i=1}^{j-1} P_i}$$

is $\overset{iid}{\sim} \text{beta}(1, \theta)$. 
Exchangeable frequencies.

The function

\[ p_{\theta}(n) = \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)! \]

is called the exchangeable partition probability function (EPPF) of both X and Y, and of the common induced partition \( \Pi = ((\Pi_n)_{n \geq 1}) \).
Exchangeable frequencies.

The function

\[ p_\theta(n_n) = \frac{\theta^k}{\theta(n)} \prod_{j=1}^{k} (n_j - 1)! \]

is called the exchangeable partition probability function (EPPF) of both \(X\) and \(Y\), and of the common induced partition \(\Pi = ((\Pi_n)_{n \geq 1})\).

**Note:** The eppf is, for every \(n\) a probability on the space of partitions of \([n]\). However it is a symmetric function of frequencies \(n_1, \ldots, n_k\). Vector of ranked frequencies is sufficient statistic for \(p_\theta\).
Predictive distributions again.

Rewriting urn scheme:

\[
\begin{align*}
\mathbb{P}(X_1 \in A) &= P_0(A); \\
\mathbb{P}(X_{n+1} \in A \mid X_1 = x_1, \ldots, X_n = x_n) &= \frac{p_\theta(n_n + e_{k+1})}{p_\theta(n_n)} P_0(A) \\
&+ \sum_{j=1}^{k} \frac{p_\theta(n_n + e_j)}{p_\theta(n_n)} \delta_{x_j^*}(A),
\end{align*}
\]

\((A \in \mathcal{B}(E)).\)
EPPF and frequencies in order of appearance

Distribution of frequencies and EPPF: For every $k$

$$p_{\theta}(n_1, \ldots, n_k) = \mathbb{E} \left[ \prod_{j=1}^{k} P_j^{n_j-1} (1 - \sum_{i=1}^{j-1} P_i) \right]$$
EPPF and frequencies in order of appearance

Distribution of frequencies and EPPF: For every \( k \)

\[
p_\theta(n_1, \ldots, n_k) = \mathbb{E} \left[ \prod_{j=1}^{k} P_j^{n_j-1} (1 - \sum_{i=1}^{j-1} P_i) \right]
\]

Interpretation: given limit frequencies in order of appearance, all sequences generating partition with frequencies \((n_1, \ldots, n_k)\) have the same probability as

\[
P \left( Y^{(n)} = (1, \ldots, 1, 2, \ldots, 2, \ldots, k, \ldots, k) \mid (P_j) \right) = \prod_{j=1}^{k} P_j^{n_j-1} (1 - \sum_{i=1}^{j-1} P_i).
\]
EPPF and frequencies in order of appearance

Distribution of frequencies and EPPF: For every $k$

$$p_\theta(n_1, \ldots, n_k) = \mathbb{E} \left[ \prod_{j=1}^{k} P_j^{n_j - 1}(1 - \sum_{i=1}^{j-1} P_i) \right]$$

**Interpretation:** given limit frequencies in order of appearance, all sequences generating partition with frequencies $(n_1, \ldots, n_k)$ have the same probability as

$$\mathbb{P} \left( Y^{(n)} = (1, \ldots, 1, 2, \ldots, 2, \ldots, k, \ldots, k) \mid (P_j) \right) = \prod_{j=1}^{k} P_j^{n_j - 1}(1 - \sum_{i=1}^{j-1} P_i).$$

**Exercise:** Derive BMQ EPPF from $P_j/(1 - \sum_{1}^{j-1} P_i) \overset{iid}{\sim} \text{beta}(1, \theta)$. 
General exchangeable random partitions.

A random partition $\Pi_n = (\Pi_{n,1}, \ldots, \Pi_{n,K_n})$ of $n$ objects is a random equivalence relation among the $n$ objects. $i \sim_j$ if and only if $i$ and $j$ are in the same equivalence class (block).

Convention: order the blocks by their least elements: $\Pi_{n,1} \ni 1$ and so on.

A sequence $((\Pi_n)_{n \geq 1})$ of random partitions is consistent if the law of $\Pi_n$ is the law of $\Pi_{n+m}$ restricted to its first $n$ elements, for every $m$.

The projective limit $\Pi$ of $(\Pi_n)$ is a partition of $\mathbb{N}$.

**Definition**

A partition $\Pi$ of $\mathbb{N}$ is infinitely exchangeable if it is the projective limit of a consistent partition structure $((\Pi_n)_{n \geq 1})$ s.t., for every $n$, the law of $\Pi_n$ is invariant under permutation of $\{1, \ldots, n\}$.
Representation for exchangeable partitions.

Let \( \Pi = ((\Pi_n)_{n\geq 1}) \) be an exchangeable partition structure, with blocks ranked by their least elements. Let \( P_j = \lim_{n \to \infty} \frac{|\Pi_{n,j}|}{n} , j \geq 1 \).

**Theorem (Pitman (1995))**

- There exist (random) \( P_j = \lim_{n \to \infty} \frac{|\Pi_{n,j}|}{n} , j \geq 1 \).
- For every \( n, k \leq n \),

\[
P(\Pi_n = (\pi_{n,1}, \ldots, \pi_{n,k})) = q(|\pi_{n,1}|, \ldots, |\pi_{n,k}|)
\]

for a function \( q \) symmetric in its arguments, determined by

\[
q(|\pi_{n,1}|, \ldots, |\pi_{n,k}|) = \mathbb{E} \left[ \prod_{j=1}^{k} P_j^{nj-1} (1 - \sum_{i=1}^{j-1} P_i) \right].
\]
Representation for exchangeable partitions.

Theorem (Pitman (1995))

- $\Pi$ is uniquely characterized by the predictive distributions: for every $\pi_n$ s.t. $(|\pi_{n,1}|, \ldots, |\pi_{n,k}|) = n_n$,
  
  (i) For $j = 1, \ldots, k$,

  $$
  \mathbb{P}(\{n+1\} \text{ starts a new block } | \Pi_n = \pi_n) = \frac{q(n_n + e_{k+1})}{q(n_n)}; 
  $$

  (ii)

  $$
  \mathbb{P}(\{n+1\} \text{ joins } j\text{-th block of } \pi_n | \Pi_n = \pi_n) = \frac{q(n_n + e_j)}{q(n_n)}. 
  $$
Generalizing Polya sequences: species sampling.

Fix $P_0 \in \mathcal{M}_1(E)$ diffuse. Choose an exchangeable random partition with EPPF $q$. Sample sequentially $(\Pi_n)$ via associated predictive rule.

(i) Set $X_1 = X_1^* \sim P_0$;

(ii) After any sample stage $n$, if $k$ blocks formed:

(ii).a) if \{n + 1\} starts a new block, sample $X_n \sim P_0$ and set $X_k^* = X_n$;

(ii).b) If \{n + 1\} joins $j$-th old block, set $X_{n+1} = X_j^*$.

Theorem

The predictive distribution of the sequence $X_1, X_2, \ldots$ converges in total variation to a RPM

\[
F \overset{d}{=} \sum_j P_j \delta_{X^*_j} + \left(1 - \sum_1^{\infty} P_j\right) P_0
\]

where $(X^*_j) \overset{iid}{\sim} P_0$ independent of $(P_j)$, and $(P_j) \overset{1:1}{\sim} q$
2 Parameter Poisson-Dirichlet models.

Choose for suitable $\alpha, \theta$,

- Probability that $\{n+1\}$ starts a new block (when $k$ existing blocks):
  \[
  \frac{\theta + \alpha k}{\theta + n}
  \]

- Probability $\{n+1\}$ joins $j$-th old block ($j \leq k$):
  \[
  \frac{n_j - \alpha k}{\theta + n}.
  \]
2 Parameter Poisson-Dirichlet models.

Choose for suitable $\alpha, \theta$,

- Probability that $\{n + 1\}$ starts a new block (when $k$ existing blocks):
  \[
  \frac{\theta + \alpha k}{\theta + n}
  \]

- Probability $\{n + 1\}$ joins $j$-th old block ($j \leq k$):
  \[
  \frac{n_j - \alpha k}{\theta + n}.
  \]

- EPPF:
  \[
  q(n_1, \ldots, n_k) = \frac{\prod_1^k(\theta + \alpha (j - 1))}{\theta(n)} \prod_{j=1}^k(1 - \alpha)n_j - 1.
  \]
As $n \to \infty$,

\[ \frac{K_n}{n^\alpha} \to M \]

where $M$ = generalized Mittag Leffler r.v.

Almost surely $\sum_j P_j = 1$ i.e. limit measure purely atomic, and

\[ B_j := \frac{P_j}{1 - \sum_{i=1}^{j-1} P_i}, j \geq 1 \]

is an independent sequence where $B_j \sim \text{beta}(1 - \alpha, \theta + \alpha j)$. 
Gibbs partition models.

In common with Poisson-Dirichlet \((\alpha, \theta)\): \(K_n\) sufficient to predict \(K_{n+1}\).

\[
\mathbb{P}(K_{n+1} = K_n + 1 \mid K_n = k) = \frac{V_{n+1,k+1}}{V_{n,k}}
\]

for a sequence \((V_{n,k})\) such that \(V_{1,1} = 1\) and

\[
V_{n,k} = (n - \alpha k)V_{n+1,k} + V_{n+1,k+1}
\]

for some \(\alpha\)

- EPPF:

\[
q(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^{k} (1 - \alpha)^{n_j-1}
\]

- Again, \(K_n \sim n^\alpha S\) for a random variable \(S\).
- Extreme points: conditional on \(S = s\), the distribution of \(\Pi_n\) is a known family induced by (inverse) stable density - Poisson-Kingman partitions.
- Interpretation: put a prior on \(S\), assume P-K likelihood given \(S = s\) as parameter.
Example: Generalized Gamma priors.

\[ V_{n,k} = \frac{\alpha^{k-1}e^\beta}{\Gamma(n)} \sum_{i=0}^{n-1} \left( \binom{n-1}{i} (-1)^i \beta^{1/\alpha} \Gamma \left( k - \frac{i}{\alpha}; \beta \right) \right) . \]
Example: Generalized Gamma priors.

\[ V_{n,k} = \frac{\alpha^{k-1} e^{\beta}}{\Gamma(n)} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i \beta^{1/\alpha} \Gamma \left( k - \frac{i}{\alpha}; \beta \right). \]

The corresponding random measure is also a normalized subordinator with Lévy measure

\[ \nu(dx) = \frac{e^{-\tau x} x^{-(1+\alpha)}}{\Gamma(1 - \alpha)} \, dx \]

with \( \beta = \alpha^{-1} \tau^\alpha. \)