∘ Dirichlet Process:

Un Dirichlet process è un processo che mira a generare una distribuzione di probabilità discreta (anche se il supporto è definito come continuo). È caratterizzato da due parametri:

- G_0 – misura base
  misura di base delle categorie, funziona come una sorta di media. Se $\alpha \uparrow$, allora $G \to G_0$
- α – parametro di concentrazione (precisione)

$$\begin{cases} \text{alto} \to \text{distribuzione generata sarà concentrata sulla media } G_0 \\ \text{basso} \to \text{opposto} \end{cases} \qquad \alpha \approx \frac{1}{Var(G)}$$

$\Rightarrow G \sim DP(\alpha, G_0)$ è la variabile casuale definita da questo processo. g = sample da G è una distribuzione di probabilità discreta

Supponendo che G_0 sia definita su Ω, $G \sim DP(\alpha, G_0)$ è t.c. $\forall \{A_1 \dots A_K\}$ partizione di Ω,

il vettore $(G(B_1), \dots, G(B_K)) \sim \text{Dirichlet}(\alpha \cdot G_0(B_1), \dots, \alpha \cdot G_0(B_K))$: $\vec{p}_0 = (G_0(B_i))_{i=1 \dots K}$

OSS: $\sum_{i=1}^{K} G_0(B_i) = 1 \rightsquigarrow \sum_{i=1}^{K} \alpha \cdot G_0(B_i) = \alpha \rightsquigarrow \alpha \approx n$ numero osservazioni

α = fattore di scala per $\vec{p}_0$: α alto → è come se avessi fatto più osservazioni nell'estrazione delle palline ⇝ più certezza sulle mie stime!

> **Esempio Dirichlet:**
> Sacco con palline rossi, verdi e blu [K=3]
> Voglio modellizzare la distribuzione discreta
> $\vec{p} = (p_{rosso}, p_{verde}, p_{blu})$
> Pesco $n=10$ palline dal sacco e ottengo: $r=2$, $v=3$, $b=5$
> Quindi posso usare come distribuzione di partenza $\vec{p}_0 = \left(\frac{2}{10}, \frac{3}{10}, \frac{5}{10}\right)$ e modellare l'incertezza rimasta con distribuzioni di probabilità per $p_{rosso}, p_{verde}, p_{blu}$ ⇝ per $\vec{p}$!
> ⇝ Posso dire che $\vec{p} \sim \text{Dirichlet}(\vec{p}_0)$

Sample from a Dirichlet Process:

(1) Chinese Restaurant

$C = [\ ]$ class of categories
At each iteration, sample $x \sim G_0$

$\forall c_K \in C: \ \mathbb{P}(x \in c_K) = \frac{|c_K|}{\alpha + \sum_J |c_J|}$ Assign x to an already existent category

$\mathbb{P}(x \in c_{New}) = \frac{\alpha}{\alpha + \sum_J |c_J|}$ Create a new category c_{new} and assign x to it

$C.\text{append}(c_{new})$

Stop when we reach K categories.

(2) Stick-Breaking Process:

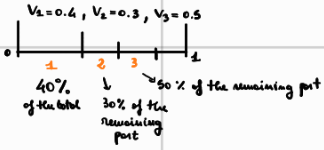
(1) Sample K locations from G_0: $\{m_1 \dots m_K\}$

(2) $\forall$ class c_K generate $V_K \sim \text{Beta}(1, \alpha) \in [0,1]$

(3) compute the associated probabilities:

$$\pi_1 = V_1 \cdot 1 \qquad (V_1\text{ \% of the total})$$
$$\pi_2 = V_2 \cdot (1 - V_1) \qquad (V_2\text{ \% of the remaining part})$$
$$\vdots$$
$$\pi_K = V_K \cdot \prod_{j=1}^{K-1} (1 - V_j)$$

Es: $K=3$
$V_1 = 0.4$, $V_2 = 0.3$, $V_3 = 0.5$



(4) $G = \sum_{J=1}^{K} \pi_J \cdot \delta_{m_J} \sim DP(\alpha, G_0)$

Properties:

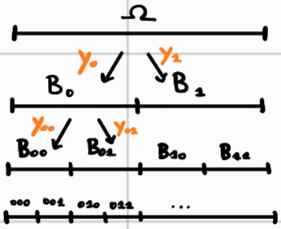
∘ **Posterior:** If $G \sim DP(\alpha, G_0)$ and $y_1 \dots y_n \overset{iid}{\sim}$, $y_i \in \{c_1 \dots c_K\}$, then $G | \vec{y} \sim DP\left(\alpha + n, \frac{\alpha}{\alpha+n} \cdot G_0 + \sum_{k=1}^{K} \frac{n_K}{\alpha+n} \cdot \delta_{c_K}\right)$

n_K: # i observed this category in $\vec{y}$; c_K: category

∘ **Prediction:** $(y_{n+1} | y_1 \dots y_n) = \frac{n_K}{\alpha + n} + \frac{\alpha}{\alpha+n} \cdot G_0(y_{n+1})$

∘ Polya Trees

The Polya Tree aims to generate a probability distribution over Ω, by dividing the sample space in 2^m parts (m = # steps) which will be indexed as $B_{\varepsilon_1 \dots \varepsilon_m}$, where $\varepsilon_i \in \{0,1\}$ and indicates the "path" of the splitting



define also:

$$y_{\varepsilon_1 \dots \varepsilon_J} = \mathbb{P}(B_{\varepsilon_1 \dots \varepsilon_J} \mid B_{\varepsilon_1 \dots \varepsilon_{J-1}}), \ \mathbb{P} \text{ is called } G$$

as a consequence, using Bayes' theorem: $G(B_{\varepsilon_1 \dots \varepsilon_m}) = \underbrace{\prod_{J=1}^{m}}_{\text{1 possible path}} y_{\varepsilon_1 \dots \varepsilon_J}$

Distribution of $y_{\varepsilon_1 \dots \varepsilon_J}$: Observe first of all that $y_{\varepsilon_1 \dots \varepsilon_{J-1} 0} + y_{\varepsilon_1 \dots \varepsilon_{J-1} 1} = 1$, so we have to define the distribution for only one of the two: we choose a Beta, st we will be able to easily compute the distribution of $1 - y_{\varepsilon_1 \dots \varepsilon_{J-1} 1}$, which will be still a Beta

$$y_{\varepsilon_1 \dots \varepsilon_{J-1} 0} \sim \text{Beta}(\alpha_{\varepsilon_1 \dots \varepsilon_{J-1} 0}, \alpha_{\varepsilon_1 \dots \varepsilon_{J-1} 1})$$
$$y_{\varepsilon_1 \dots \varepsilon_{J-1} 1} \sim \text{Beta}(\alpha_{\varepsilon_1 \dots \varepsilon_{J-1} 1}, \alpha_{\varepsilon_1 \dots \varepsilon_{J-1} 0})$$

$\overset{\text{Def}}{\Longrightarrow}$ $G \sim \text{PolyaTree}(\Pi, \mathcal{A})$

Π: defines the partitions B of the space

$\mathcal{A}$: defines the parameters of the y's: $\alpha_{\varepsilon_1 \dots J-1, 0/1}$

- **Prior Centering:**

We have two ways of centering G around G_0: we can either change Π (how we partition Ω) or $\mathcal{A}$ (the probabilities assigned to the partitions)

(1) Centering by Π: $PT(G_0, \mathcal{A})$

Given the sequence $\varepsilon_1 \dots \varepsilon_J$ that defines the path, call it $\vec{\varepsilon}$. $\vec{\varepsilon} \in \{0,1\}^J$ and can be interpreted as a binary coding of an integer. Call $N = N(\vec{\varepsilon})$ that integer. es: $\vec{\varepsilon} = \{\underset{1}{0}, \underset{2}{0}, \underset{4}{1}, \underset{8}{1}\}$
$\rightsquigarrow N = 4 + 8 = 12$

Then, by putting:

$$B_{\vec{\varepsilon}} = \Big[\underbrace{G_0^{-1}\Big(\frac{N(\vec{\varepsilon})}{2^m}\Big)}, \underbrace{G_0^{-1}\Big(\frac{N(\vec{\varepsilon})+1}{2^m}\Big)}\Big] \quad \text{and} \quad \alpha_{\vec{\varepsilon} \dots 0} = \alpha_{\vec{\varepsilon} \dots 1} \ \forall \ y_{\vec{\varepsilon}}$$

quantiles

(so, probability density of $y_{\vec{\varepsilon}}$ centered on 50%)

we obtain that $\mathbb{E}[G(B_{\vec{\varepsilon}})] = G_0(B_{\vec{\varepsilon}}) \ \forall \vec{\varepsilon}$, so G is centered on G_0

This is the most common strategy, cause we can put all the $\alpha_{\vec{\varepsilon}}$ with the same value:

$$\alpha_{\vec{\varepsilon}} = \alpha_{\varepsilon_1 \dots \varepsilon_m} \overset{\forall \vec{\varepsilon} \in \mathbb{R}^m, \forall m:}{=} \begin{cases} m^2 \\ c \cdot m^2 \\ 8^m \end{cases}$$

In general, we want the "freedom" of the model to increase quickly with m = # steps
["freedom" = possibilità di allontanarsi da G_0 prior iniziale]

$\alpha_{\vec{\varepsilon}} = c \cdot m^2$

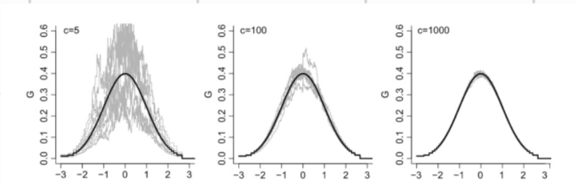


Fig. 3.1 Plots of $n = 10$ density g_i samples from $G_i \sim \text{PT}(G^o, \mathcal{A}^c)$, $i = 1, \dots, n$, centered on a standard normal distribution G^o, for $c = 5$, 100 and 1000. In all cases, $E(g_i)$ is overlaid over a plot of the realizations of g_i

(2) Centering $\mathcal{A}$: $PT(\Pi, G_0)$

$\forall$ split $B_{\vec{\varepsilon}}$ into $B_{\vec{\varepsilon}\_0}$ and $B_{\vec{\varepsilon}\_1}$, by putting $\begin{cases} \alpha_{\vec{\varepsilon}\_0} = c \cdot G_0(B_{\vec{\varepsilon}\_0}) \\ \alpha_{\vec{\varepsilon}\_1} = c \cdot G_0(B_{\vec{\varepsilon}\_1}) \end{cases}$ we obtain that G is centered on G_0

- **Posterior Distribution:**

Suppose $G \sim PT(\Pi, \mathcal{A})$, and we observe the data $y_1 \dots y_n \in \Omega$. Then, the posterior distribution of G is still a Polya Tree with the same partitions Π:

$$G \mid \vec{y} \sim PT(\Pi, \mathcal{A}^*), \text{ where } \mathcal{A}^*: \ \alpha^*_{\vec{\varepsilon}} = \alpha_{\vec{\varepsilon}} + n_{\vec{\varepsilon}}, \text{ with } n_{\vec{\varepsilon}} = \#\{y_i \in B_{\vec{\varepsilon}}\}$$

- **Marginal distribution of $\vec{y} \sim G$**: mega formula