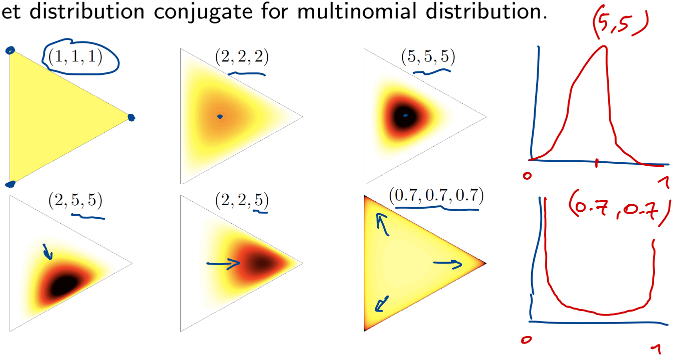
Dirichlet distribution

A *Dirichlet distribution* on a simplex Δ_K is a probability distribution with parameters $\alpha_i > 0$ and a density function

$$f(x_1, \dots, x_K; \alpha_1, \dots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}.$$

It is common to refer to Dirichlet distribution as $\mathrm{Dir}(\alpha_1, \dots, \alpha_k)$.

Remark Dirichlet distribution conjugate for multinomial distribution.



[Image by Y.W. Teh]

Dirichlet process

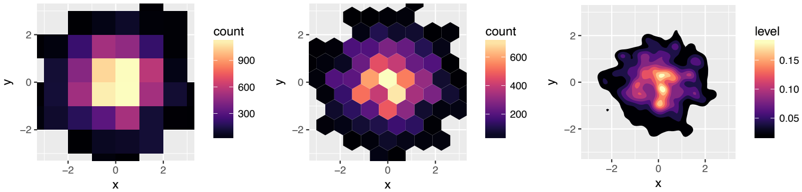
A central Bayesian nonparametric prior (Ferguson, 1973) A. S 1

Definition (Dirichlet process)

A Dirichlet process on the space $\mathcal{Y}$ is a random process P such that there exist α (precision parameter) and G_0 (base/centering distribution) such that for any finite partition $\{A_1, \dots, A_d\}$ of $\mathcal{Y}$, the random vector $(P(A_1), \dots, P(A_d))$ is Dirichlet distributed

$$(P(A_1), \dots, P(A_d)) \sim \mathrm{Dir}(\alpha G_0(A_1), \dots, \alpha G_0(A_d))$$

Notation: $P \sim DP(\alpha, G_0)$



Dirichlet process

A central Bayesian nonparametric prior (Ferguson, 1973)

Definition (Dirichlet process)

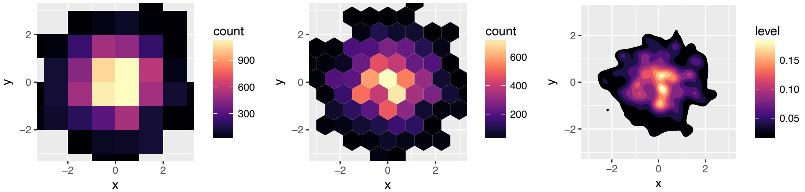
std.

A Dirichlet process on the space $\mathcal{Y}$ is a random process P such that there exist α (precision parameter) and G_0 (base/centering distribution) such that for any finite partition $\{A_1, \ldots, A_d\}$ of $\mathcal{Y}$, the random vector $(P(A_1), \ldots, P(A_d))$ is Dirichlet distributed

$\alpha > 0$

$$(P(A_1), \ldots, P(A_d)) \sim \mathrm{Dir}(\alpha G_0(A_1), \ldots, \alpha G_0(A_d))$$

Notation: $P \sim DP(\alpha, G_0)$



Dirichlet process

A central Bayesian nonparametric prior (Ferguson, 1973)

Definition (Dirichlet process)

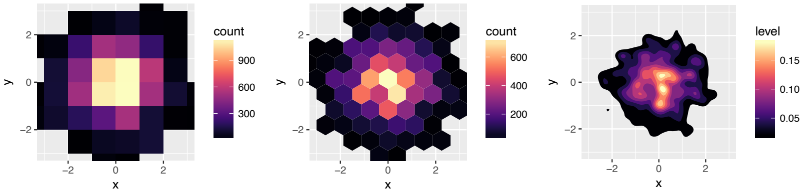
A Dirichlet process on the space $\mathcal{Y}$ is a random process P such that there exist α (precision parameter) and G_0 (base/centering distribution) such that for any finite partition $\{A_1, \ldots, A_d\}$ of $\mathcal{Y}$, the random vector $(P(A_1), \ldots, P(A_d))$ is Dirichlet distributed

$$\boxed{(P(A_1), \ldots, P(A_d)) \sim \text{Dir}(\alpha G_0(A_1), \ldots, \alpha G_0(A_d))}$$

Notation: $P \sim DP(\alpha, G_0)$

$\mathcal{Y} = \mathbb{R}^2$, $G_0 = N_2(0, I_2)$

α



Moments of Dirichlet process I

$X \sim Be(a,b)$

$\mathbb{E}X = \frac{a}{a+b}$

$Var\, X = \frac{ab}{(a+b)^2(a+b+1)}$

$$\mathbb{E}[P(A)] = \frac{\alpha P_0(A)}{\alpha P_0(A) + \alpha(1-P_0(A))} = P_0(A)$$

$$Var[P(A)] = \frac{\alpha P_0(A)\cdot\alpha(1-P_0(A))}{\alpha^2(\alpha+1)} = \frac{P_0(A)(1-P_0(A))}{\alpha+1}$$

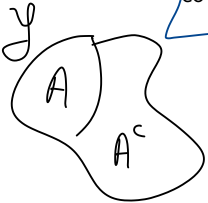
Proposition

Let $P \sim DP(\alpha, P_0)$ then for every measurable sets A, B we have

$$\mathbb{E}(P(A)) = P_0(A), \quad \longleftrightarrow \quad \mathbb{E}[P] = P_0 \tag{1}$$

$$Var(P(A)) = \frac{P_0(A)(1-P_0(A))}{1+\alpha}, \quad \longleftrightarrow \quad \alpha : \text{"concentra"} \tag{2}$$

$$\text{cov}(P(A), P(B)) = \frac{P_0(A\cap B) - P_0(A)P_0(B)}{1+\alpha}. \tag{3}$$



(A, A^c) : $(P(A), P(A^c)) \sim Dir(\alpha P_0(A), \alpha \underbrace{P_0(A^c)}_{1-P_0(A)})$

$P \sim DP(\alpha, P_0)$

$P(A) \sim Be(\alpha P_0(A), \alpha(1-P_0(A)))$

Moments of Dirichlet process II

Proof

We will make use of $p(A) \sim \text{Beta}(\alpha P_0(A), \alpha(1 - P_0(A)))$. From this we obtain

$$\mathbb{E}(p(A)) = \frac{\alpha P_0(A)}{\alpha(P_0(A) + 1 - P_0(A))} = P_0(A)$$

and

$$\text{Var}(p(A)) = \frac{\alpha^2 P_0(A)(1 - P_0(A))}{\alpha^2(\alpha + 1)}.$$

We derive the covariance term in two cases, firstly taking into consideration the one with $A \cap B = \emptyset$. In that case any space Ω may be decomposed into three sets:

$$\Omega = \{A, B, (A \cup B)^c\}.$$

Using de Morgan's law the last can be written as $(A \cup B)^c = A^c \cap B^c =: C$. Therefore we may write a joint probability vector

$$\big(p(A), p(B), p(A^c \cap B^c)\big) \sim \text{Dir}\big(\alpha P_0(A), \alpha P_0(B), \alpha P_0(C)\big)$$

Moments of Dirichlet process III

and hence $\text{cov}(p(A), p(B)) = -P_0(A)P_0(B)/(1+\alpha)$. In the more general case one may decompose

$$A = (A \cap B) \cup (A \cap B^c)$$
$$B = (B \cap A) \cup (B \cap A^c),$$

so that

$$\text{cov}(P(A), P(B)) = \text{cov}(P(A \cap B) + P(A \cap B^c), P(B \cap A) + P(B \cap A^c))$$

and so forth using the linearity of covariance.

Marginalizing out the DP

$\mathbb{E}\{P(A)\} = P_0(A)$

Property 1 can be written equivalently as

$$\underbrace{\mathbb{E}(P(A)) = P_0(A)} = \underbrace{\int P(A)dDP(P).} \quad (4)$$

A Dirichlet process model can be constructed as two level sampling:

prior

model

$$\begin{cases} P \sim DP(\alpha, P_0) \\ X|P \sim P, \end{cases}$$

P : proba measures

(X, P)

i.e. we sample probability measure P from the Dirichlet process and then given P we sample random variables X_i.

Marginalizing out P, we obtain the marginal distribution of X:

$$\boxed{X \sim P_0}$$

Posterior distribution I

Let $(X_1, \dots, X_n) =: X_{1:n}$ be sampled from the hierarchical model

$$\boxed{\begin{cases} P \sim DP(\alpha, P_0) & \text{prior} \\ X_{1:n} | P \overset{i.i.d.}{\sim} P, \end{cases}} \tag{5}$$

This model is usually used as a building block in a larger hierarchical model, e.g. mixture models, graphs etc.

Theorem (Ferguson [1973]) Conjugacy

The posterior of P as presented in (5) is

$$P|X_{1:n} \sim DP(\alpha P_0 + \sum_{i=1}^{n} \delta_{X_i}). \tag{6}$$

The predictive distribution of a next observation is given by

$$\boxed{\mathbb{P}(X_{n+1}|X_{1:n}) = \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{X_i}.} \tag{7}$$

Posterior distribution II

The predictive (7) is also called *Polya Urn schema* or *Blackwell-MacQueen Urn Schema*.

Posterior distribution III

Proof

Property (6) can be obtained by remarking that the posterior distribution of $P(A_1), \dots, P(A_k)$ depends on the observations only via their cell counts (it comes from *tail–free* property). Denote $N_j = \#\{1 \le i \le n : x_i \in A_j\}$, i.e. the number of observations in each partition of X. Then we have

$$\underbrace{(P(A_1), \dots, P(A_k))\big|X_{1:n}} \overset{d}{=} \underbrace{(P(A_1), \dots, P(A_k))\big|N_{1:k}}.$$

Lets use shorthand notation: $\alpha = (\alpha_1, \dots, \alpha_k) = (P(A_1), \dots, P(A_k))$ and $N = (N_1, \dots, N_k)$. Then

$$\begin{cases} N|P \sim \overset{\text{Multinoulli}}{\mathrm{Multinom}_k}(P(A_1), \dots, P(A_k)) & \text{Model} \\ \alpha = (P(A_1), \dots, P(A_k)) \sim \mathrm{Dir}_k(\alpha P_0(A_1), \dots, \alpha P_0(A_k)) & \text{Prior} \end{cases}$$

and hence we obtain the prior of the form

$$p(\alpha) \propto \alpha_1^{\alpha P_0(A_1)-1} \dots \alpha_k^{\alpha P_0(A_k)-1},$$

while sampling model is

$$p(N|\alpha) \propto \alpha_1^{N_1} \dots \alpha_k^{N_k}.$$

Bayes' thm.

Posterior distribution IV

Notation: $\boxed{\times}$ $P \sim DP(\underset{\text{mass}}{\alpha}, \underset{\text{prob. m.}}{P_0}) \overset{[\text{or}]}{=} DP(\underset{\text{meas. } G}{\alpha P_0})$

$\alpha = G(\mathcal{Y})$

$P_0 = \frac{G}{G(\mathcal{Y})} = \frac{G}{\alpha}$

$\boxed{\times}$ Emp. proba. meas. of $X_{1:n}$: $\overline{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

This results in the posterior of form

$$p(\alpha|N) \propto \alpha_1^{\alpha P_0(A_1)+N_1-1} \ldots \alpha_k^{\alpha P_0(A_k)+N_k-1} = \mathrm{Dir}_k(\alpha P_0(A_1)+N_1, \ldots, \alpha P_0(A_k)+N_k).$$

Property (7) is a result of taking the expected value of (6).

$P | X_{1:n} \sim DP(\alpha_n, P_n) = DP(\alpha_n P_n) = DP\left(\alpha P_0 + \sum_{i=1}^{n} \delta_{X_i}\right)$

$\alpha_n = \sum_i \left(\alpha P_0(A_i) + N_i\right) = \alpha + n$ $\qquad = DP(\alpha P_0 + n \overline{P}_n)$

$P_n = \frac{\alpha}{\alpha+n} P_0 + \frac{1}{\alpha+n}\sum_{i=1}^{n} \delta_{X_i} = \underbrace{\frac{\alpha}{\alpha+n}}\, P_0 + \underbrace{\frac{n}{\alpha+n}}\, \overline{P}_n$

Combinatorial properties: Number of distinct values I

Assume that the base measure P_0 is non-atomic. Then with probability 1:

$$\boxed{X_i \notin \{X_1, \dots, X_{i-1}\} \Leftrightarrow X_i \sim P_0.}$$

Let $D_i = \mathbb{I}(X_i \text{ is a new value})$ and lets denote $\boxed{K_n = \sum_{i=1}^n D_i}$ a number of distinct values $X_1, \dots, X_n$ with distribution $\mathcal{L}(K_n)$.

$$\mathbb{P}(X_i | X_{1:i-1}) = \frac{\alpha}{\alpha+i-1} P_0 + \frac{1}{\alpha+i-1} \sum_{1}^{i-1} \delta_{X_j}$$

Proposition

Random variables D_i are distributed ~~i.i.d.~~ $\perp\!\!\!\perp$ *with respect to Bernoulli$(\alpha/(\alpha+i-1))$. Therefore for fixed α and for $n \to \infty$ we have:*

- i) $\mathbb{E}K_n \sim \alpha \log n \sim Var(K_n)$:

$$\mathbb{E}K_n = \sum \mathbb{E}D_i = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1} \sim \alpha \log n$$

$$Var\, K_n = \sum Var\, D_i = \sum \frac{\alpha(i-1)}{(\alpha+i-1)^2} \sim \alpha \log n$$

- ii) $K_n/\log(n) \xrightarrow{a.s.} \alpha$
- iii) $(K_n - \mathbb{E}K_n)/sd(K_n) \to N(0,1)$
- iv) $d_{TV}(\mathcal{L}(K_n), Poisson(\mathbb{E}K_n)) = o(1/\log(n))$ *where*

$$d_{TV}(P, Q) = sup|P(A) - Q(A)|$$

over measurable partition A

Combinatorial properties: Number of distinct values II

Proof

i) $\mathbb{E}K_n = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1}$ and $\mathrm{Var}(K_n) = \sum_{i=1}^{n} \frac{\alpha(i-1)}{(\alpha+i-1)^2}$.

ii) Since D_i's are $\mathbb{I}$ one may use Kolmogorov law of strong numbers and

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}(D_i)}{(\log i)^2} = \sum_{i=1}^{\infty} \frac{\alpha(i-1)}{(\alpha+i-1)^2(\log i)^2} < \infty \quad \times$$

by e.g. the fact that $\sum_i (1/i(\log i)^2)$ converges.

iii) By Lindeberg central limit theorem. α

iv) This is implied from Chein–Stein approximation.

Combinatorial properties: Number of distinct values III

Theorem Lindeberg

Suppose X_i are i.i.d. such that $\mathbb{E}X_i = \mu_i$ and $VarX_i = \sigma_i^2 < \infty$. Define $Y_i = X_i - \mu_i$, $T_n = \sum_{i=1}^n Y_i$, $s_n^2 = Var(T_n) = \sum_{i=1}^n \sigma_i^2$. Then provided that

$$\forall \epsilon > 0 \quad \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E}\big(Y_i^2 \mathbb{I}(|Y_i| > \epsilon s_n)\big) \xrightarrow{n\to\infty} 0$$

we have $T_n/s_n \xrightarrow{d} N(0,1)$.

Combinatorial properties: Distribution of distinct values I

We have now the limits of K_n and we know its approximate distribution $\mathcal{L}(K_n)$. The **exact distribution of K_n** is:

PROPOSITION

If P_0 is non-atomic then : $\forall k \in \{1, \dots, n\}$

$$\underbrace{\mathbb{P}(K_n = k)}_{\text{p.m.f.}} = \mathfrak{C}_n(k) n! \alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)}, \tag{8}$$

where

$$\mathfrak{C}_n(k) = \frac{1}{n!} \sum_{S \in \mathfrak{J}_n(k)} \prod_{j \in S} j \tag{9}$$

and $\mathfrak{J}_n(k) = \{S \subset \{1, \dots, n-1\},\ |S| = n - k\}$.

Recall the definition of the **Gamma function** $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$.

Combinatorial properties: Distribution of distinct values II

Let us consider when we may deal with events $K_n = k$: we have two cases

$$\begin{cases} K_{n-1} = k-1 \text{ and } X_n \text{ is a new value} \\ K_{n-1} = k \text{ and } X_n \text{ is not a new value.} \end{cases}$$

This results in

$$p_n(k,\alpha) := \mathbb{P}(k_n = k|\alpha) = \frac{\alpha}{\alpha+n-1} p_{n-1}(k-1,\alpha) + \frac{n-1}{\alpha+n-1} p_{n-1}(k,\alpha). \tag{10}$$

Now let us remark that $\mathfrak{C}_n(k) = p_n(k,\alpha=1)$. Therefore

$$\mathfrak{C}_n(k) = \frac{1}{n}\mathfrak{C}_{n-1}(k-1) + \frac{n-1}{n}\mathfrak{C}_{n-1}(k). \tag{11}$$

By induction over n: first we check case $n=1$:

$$p_1(1,\alpha) = \mathfrak{C}_1(1)\frac{\alpha}{\alpha} = \mathfrak{C}_1(1). \tag{12}$$

Combinatorial properties: Distribution of distinct values III

To check case $n > 1$ we use (8) and then (10):

$$
\begin{aligned}
p_n(k,\alpha) &= \frac{\alpha}{\alpha+n-1} p_{n-1}(k-1,\alpha) + \frac{n-1}{\alpha+n-1} p_{n-1}(k,\alpha) \\
&= \frac{\alpha}{\alpha+n-1} \mathfrak{C}_{n-1}(k-1)(n-1)!\alpha^{k-1} \frac{\Gamma(\alpha)}{\Gamma(\alpha+n-1)} + \\
&+ \frac{n-1}{\alpha+n-1} \mathfrak{C}_{n-1}(k)(n-1)!\alpha^{k} \frac{\Gamma(\alpha)}{\Gamma(\alpha+n-1)} \\
&= \frac{\alpha^k}{\alpha+n-1}(n-1)! \frac{\Gamma(\alpha)}{\Gamma(\alpha+n-1)} n \left(\frac{1}{n} \mathfrak{C}_{n-1}(k-1) + \frac{n-1}{n} \mathfrak{C}_{n-1}(k) \right) \\
&= \mathfrak{C}_n(k) n! \alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)},
\end{aligned}
$$

which proves property (8).

Combinatorial properties: Distribution of distinct values IV

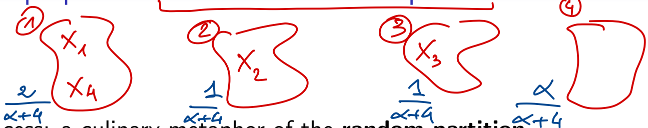
To prove (9) let use define a polynomial $A_n(s)$ as $A_n(s) = \sum_{i=1}^{\infty} \mathfrak{C}_n(k)s^k$. Then using (11) polynomial $A_n(s)$ can be written as

$$\begin{aligned} A_n(s) &= \sum_{k=1}^{\infty} \left(\frac{1}{n}\mathfrak{C}_{n-1}(k-1) + \frac{n-1}{n}\mathfrak{C}_{n-1}(k) \right) s_k \\ &= \frac{1}{n}(sA_{n-1}(s) + (n-1)A_{n-1}(s)) = \frac{s+n-1}{n}A_{n-1}(s) \\ &= \ldots = A_1(s)\prod_{j=2}^{n} \frac{s+j-1}{j} = \frac{s(s+1)\cdot \ldots \cdot (s+n-1)}{n!}. \end{aligned}$$

Last equality implies from the fact that $\mathfrak{C}_1(k) = 1\delta_{k1}$ and hence $A_1(s) = s$. Checking terms after the expansion finishes the proof of (9).

Combinatorial properties: Chinese Restaurant process I

$$\left\{\begin{array}{l} P \sim DP(\alpha, P_0) \\ X_1, \dots, X_n \mid P \overset{iid}{\sim} P \end{array}\right. :$$



Chinese restaurant process: a culinary metaphor of the **random partition induced by the DP**. Customers join a populated table with probability $n_j/(\alpha+n)$, where n_j denotes the number of clients already sitting around the table or sit at new table with probability $\alpha/(\alpha+n)$.

Proposition

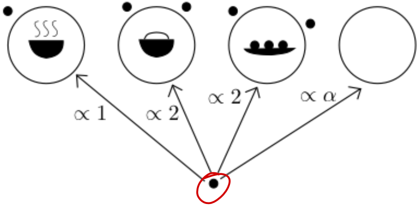
A random sample $X_{1:n}$ from a DP with precision parameter α induces a partition of $\{1,\dots,n\}$ into k sets of sizes $n_1,\dots,n_k$ with probability

$$\boxed{p(n_1,\dots,n_k) = p(\{n_1,\dots,n_k\}) = \alpha^k \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \prod_{j=1}^{k} \Gamma(n_j).}$$

$(n_j - 1)!$

⚠ indep. on P_0, only α

Combinatorial properties: Chinese Restaurant process II



Combinatorial properties: Chinese Restaurant process III

X_1^*: x_1, x_4 — $n_1=2$ X_2^*: x_2 — $n_2=1$ X_3^*: x_3

Proof

We will use the Polya urn schema slightly changed by using $n_1, \dots, n_k$

$$\boxed{\mathbb{P}(X_{n+1}|X_{1:n}) = \frac{\alpha}{\alpha+n}P_0 + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j \delta_{X_j^*}.}$$

$$(\alpha)_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} = \alpha(\alpha+1)\cdots(\alpha+n-1)$$

By exchangeability, the distribution of $\{n_1, \dots, n_k\}$ does not depend on the order of the observations. Let's compute $p(n_1, \dots, p_k)$ as the probability of one draw where the first table consists of first n_1 observations etc.

To proceed, let us use Polya urn scheme: we denote $\bar{n}_j = \sum_{i=1}^{j} n_i$ and hence $\bar{n}_k = n$, the total number of observations. We can observe the following pattern: first ball open new table, following $n_j - 1$ ones fill in that table and so forth. That quantity can be rewritten as

$\frac{\alpha}{\alpha}$

$$\left[1\cdot\frac{1}{\alpha+1}\frac{2}{\alpha+2}\cdots\frac{n_1-1}{\alpha+n_1-1}\right]\cdot\left[\frac{\alpha}{\alpha+n_1}\frac{1}{\alpha+n_1+1}\cdots\frac{n_2-1}{\alpha+n_1+n_2-1}\right]\cdots\left[\frac{\alpha}{\alpha+n_1+\dots+n_{k-1}}\cdots\frac{1}{\cdots}\cdots\frac{n_k-1}{\alpha+n-1}\right]$$

$$\frac{\alpha^k}{\alpha(\alpha+1)\dots(\alpha+n-1)}\prod_{j=1}^{k}(n_j-1)!,$$

$$\frac{\alpha^k}{\alpha\dots(\alpha+n-1)}\prod_{j=1}^{k}(n_j-1)! = \frac{\alpha^k}{(\alpha)_n}\prod_{1}^{k}\Gamma(n_j)$$

Combinatorial properties: Chinese Restaurant process IV

where one can rewrite both terms using Gamma function $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$: the first term can be written as

$$\frac{\alpha^k}{\alpha(\alpha+1)\dots(\alpha+n-1)} = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)},$$

while the second one as $(n_j - 1)! = \Gamma(n_j)$.
One should remark that for ordered partitions we have

$$\bar{p}(n_1,\dots,n_k) = \frac{p(n_1,\dots,n_k)}{k!}.$$

Combinatorial properties: Ewens sampling formula I

Ewens sampling formula (ESF), presented originally by Ewens [1972], is the distribution of multiplicities $m = (m_1, \ldots, m_n)$, m_ℓ is the number of groups of size ℓ.
Also known as allelic partitions in population genetics, when there is no selective difference between types: null hypothesis in non Darwinian theory.

PROPOSITION (Ewens [1972]; Antoniak [1974])

Random variables $X_1, \ldots, X_n$ generated from a DP has multiplicity class $(m_1, \ldots, m_n)$ with probability

$$p(m_1, \ldots, m_n) = \frac{\alpha^k}{\alpha_{(n)}} \frac{n!}{\prod_{\ell=1}^n \ell^{m_\ell} m_\ell!}.$$

Notation $n_{(k)} := n(n-1) \cdot \ldots \cdot (n-k+1)$.

Combinatorial properties: Ewens sampling formula II

Proof

Two steps: 1) Compute probability of particular sequence of $X_1, \dots, X_n$ in given class $(m_1, \dots, m_n)$, note that all such sequences are equally likely and 2) multiply obtained quantity by the number of such sequences.

1) Consider a sequence $X_1, \dots, X_n$ such that $X_1, \dots, X_{m_1}$ occur each only once, then the next m_2 occur only twice and so on. This sequence has probability which may be obtained by the Polya Urn scheme in the same fashion as CRP:

$$\frac{\alpha^{m_1}(\alpha \cdot 1)^{m_2} \dots \left(\alpha \cdot 1 \cdot \dots \cdot (n-1)\right)^{m_n}}{\alpha_{(n)}} = \frac{\alpha^k}{\alpha_{(n)}} \prod_{\ell=1}^{n} ((\ell-1)!)^{m_\ell}.$$

2) Number of sequences $X_1, \dots, X_n$ with frequencies $(m_1, \dots, m_n)$ is a number of ways of putting n distinct objects into bins, so called multinomial coefficient. Since ordering of the m_ℓ bins of frequency ℓ is irrelevant, divide by $m_\ell!$:

$$\frac{1}{\prod_{l=1}^{n}(m_\ell)!} \binom{n}{1 \times \# m_1, 2 \times \# m_2, \dots, n \times \# m_n} = \frac{n!}{\prod_{\ell=1}^{n} m_\ell! (\ell!)^{m_\ell}}$$

To finish one needs to multiply results obtained in 1) and 2).

Stick-breaking representation

The DP has almost surely discrete realizations (Sethuraman, 1994)

$$P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}$$

- locations $\theta_j \overset{\text{iid}}{\sim} G_0$
- weights $\pi_j = \tilde{\pi}_j \prod_{l<j}(1-\tilde{\pi}_l)$ with $\tilde{\pi}_j \overset{\text{iid}}{\sim} Beta(1, \alpha)$,

Stick-breaking representation

The DP has almost surely discrete realizations (Sethuraman, 1994)

$$P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}$$

- locations $\theta_j \overset{\text{iid}}{\sim} G_0$
- weights $\pi_j = \tilde{\pi}_j \prod_{l<j}(1 - \tilde{\pi}_l)$ with $\tilde{\pi}_j \overset{\text{iid}}{\sim} Beta(1, \alpha)$,

Stick-breaking representation

The DP has almost surely discrete realizations (Sethuraman, 1994)

$$P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}$$

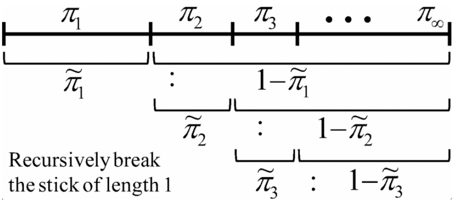
- locations $\theta_j \stackrel{\text{iid}}{\sim} G_0$
- weights $\pi_j = \tilde{\pi}_j \prod_{l<j}(1 - \tilde{\pi}_l)$ with $\tilde{\pi}_j \stackrel{\text{iid}}{\sim} Beta(1, \alpha)$,

Stick-breaking representation

The DP has almost surely discrete realizations (Sethuraman, 1994)

$$P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}$$

- locations $\theta_j \overset{\text{iid}}{\sim} G_0$
- weights $\pi_j = \tilde{\pi}_j \prod_{l<j}(1 - \tilde{\pi}_l)$ with $\tilde{\pi}_j \overset{\text{iid}}{\sim} Beta(1, \alpha)$,

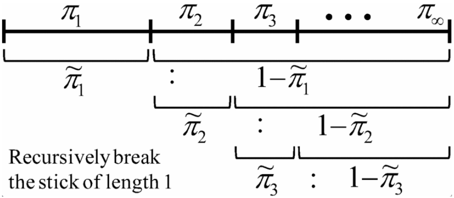


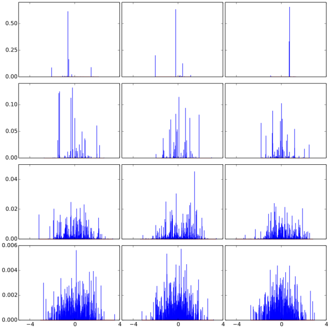
Stick-breaking representation

The DP has almost surely discrete realizations (Sethuraman, 1994)

$$P = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}$$

- locations $\theta_j \overset{\text{iid}}{\sim} G_0$
- weights $\pi_j = \tilde{\pi}_j \prod_{l<j}(1 - \tilde{\pi}_l)$ with $\tilde{\pi}_j \overset{\text{iid}}{\sim} Beta(1, \alpha)$,





Stick-breaking representation I

Theorem (Sethuraman [1994])

If $V_1, V_2, \ldots \overset{i.i.d.}{\sim} Be(1, \alpha)$ *and* $\phi_1, \phi_2, \ldots \overset{i.i.d.}{\sim} P_0$ *are i.i.d. variables, then define* $p_1 = V_1$ *and*

$$p_j = V_j \prod_{1 \leq l \leq j} (1 - V_l)$$

then

$$P = \sum_{i=1}^{\infty} p_i \delta_{\phi_i} \sim DP(\alpha, P_0).$$

Stick-breaking representation II

Lemma

For independent $\phi \sim P_0$ and $V \sim Be(1, \alpha)$ the DP is the only solution of the distributional equation

$$P \stackrel{\mathrm{d}}{=} V\delta_\phi + (1 - V)P, \tag{13}$$

where $P \sim DP(\alpha, P_0)$.

Stick-breaking representation III

Proof

1) The weights $(p_1, p_2, \ldots)$ need to form a probability vector. The leftover mass at stage j is

$$1 - \left(\sum_{i=1}^{j} p_i\right) = \prod_{i=1}^{j}(1 - V_i) =: R_j.$$

One may notice that R_j is decreasing and for every j we have $R_j \in [0, 1]$, hence we obtain almost sure convergence which is equivalent with convergence in mean. Therefore

$$\mathbb{E}R_j = \mathbb{E}\prod_j(1 - V_j) = \prod_j \mathbb{E}(1 - V_j) = \left(\frac{\alpha}{\alpha + 1}\right)^j \to 0.$$

So $(p_1, \ldots)$ is a probability vector almost surely and P is a probability measure almost surely.

Stick-breaking representation IV

2) Now one may write

$$P = p_1 \delta_{\phi_1} + \sum_{j=2}^{\infty} p_j \delta_{\phi_j} = V_1 \delta_{\phi_1} + (1 - V_1) \sum_{j=1}^{\infty} \tilde{p}_j \delta_{\tilde{\phi}_j},$$

where $\tilde{p}_j = \frac{p_{j+1}}{1-V_1} = V_{j+1} \prod_{l=2}^{j} (1 - V_l)$ and $\tilde{\phi}_j = \phi_{j+1}$, then $(\tilde{p}_j)$ and $(\tilde{\phi}_j)$ satisfy the same distributional definitions as (p_j) and (ϕ_j), hence $\tilde{P} \overset{\mathrm{d}}{=} P$ and so P is solution of the Lemma equation (13) whose only solution is the DP.

DP as a normalized Gamma process I

The DP can be obtained by normalizing a Gamma process. It is a generic way to obtain independently distributed probability measures from almost surely finite random measures. Let us investigate for the case $\mathcal{Y} = \mathbb{R}$.

Definition

Gamma process on $\mathbb{R}_+$ is a process $(S(u) : u \geq 0)$ with independent increments satisfying

$$\forall u_1 : 0 \leq u_1 \leq u_2 : \quad S(u_2) - S(u_1) \overset{\perp}{\sim} Ga(u_2 - u_1, 1).$$

This ensures that the process has non-decreasing right continuous sample path $u \mapsto S(u)$.

Theorem

For every $\alpha > 0$ and for every cumulative distribution function G, a random cumulative distribution function such that

$$F(t) = \frac{S(\alpha G(t))}{S(\alpha)}$$

is the distribution of a $DP(\alpha, G)$.

DP as a normalized Gamma process II

Proof

For any set of t_i satisfying $-\infty = t_0 < t_1 < \ldots < t_k = \infty$ we have

$$S(\alpha G(t_i)) - S(\alpha G(t_{i-1})) \sim Ga\big(\alpha G(t_i) - \alpha G(t_{i-1}), 1\big).$$

Use property that if $Y_i \overset{ind}{\sim} Ga(\alpha_i, 1)$ then $(Y_1, \ldots, Y_n)/\sum_i Y_i \sim \mathrm{Dir}_n(\alpha_1, \ldots, \alpha_n)$ to obtain

$$\big(F(t_1)-F(t_0), \ldots, F(t_k)-F(t_{k-1})\big) \sim \mathrm{Dir}_k\big(\alpha G(t_1)-\alpha G(t_0), \ldots, \alpha G(t_k)-\alpha G(t_{k-1})\big).$$

Hence the definition of DP holds for every partition in intervals. These form a measure determining class, so that the definition holds for every partition in general.

Definition via the Polya Urn Scheme

A Polya sequence with parameter αP_0 is a sequence of random variables $X_1, \ldots, X_n$ whose joint distribution satisfies

$$X_1 \sim P_0,\ X_{n+1}|X_1, \ldots, X_n \sim \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{X_i}. \quad (14)$$

Theorem

If $X_1, X_2, \ldots$ is a Polya sequence then exists random probability measure P such that $X_i|P \overset{i.i.d.}{\sim} P$ and $P \sim DP(\alpha, P_0)$.

Proof

We can consider Polya sequence as an outcome of Polya urn, we see that it is exchangeable. By de Finetti theorem exists such probability measure P such that $X_i|P \overset{i.i.d.}{\sim} P$. So far we have proved existence of the DP and know that DP generates a Polya sequence. Since the RPM given by de Finetti's theorem is unique this proves that $P \sim DP(\alpha, P_0)$.

Definition via the Polya Urn Scheme

A Polya sequence with parameter αP_0 is a sequence of random variables $X_1, \ldots, X_n$ whose joint distribution satisfies

$$X_1 \sim P_0, \; X_{n+1} | X_1, \ldots, X_n \sim \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{X_i}. \tag{14}$$

Theorem

If $X_1, X_2, \ldots$ is a Polya sequence then exists random probability measure P such that $X_i | P \overset{i.i.d.}{\sim} P$ and $P \sim DP(\alpha, P_0)$.

Proof

We can consider Polya sequence as an outcome of Polya urn, we see that it is exchangeable. By de Finetti theorem exists such probability measure P such that $X_i | P \overset{i.i.d.}{\sim} P$. So far we have proved existence of the DP and know that DP generates a Polya sequence. Since the RPM given by de Finetti's theorem is unique this proves that $P \sim DP(\alpha, P_0)$.

Definition via the Polya Urn Scheme

A Polya sequence with parameter αP_0 is a sequence of random variables $X_1, \ldots, X_n$ whose joint distribution satisfies

$$X_1 \sim P_0,\ X_{n+1}|X_1, \ldots, X_n \sim \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n} \delta_{X_i}. \tag{14}$$

Theorem

If $X_1, X_2, \ldots$ is a Polya sequence then exists random probability measure P such that $X_i|P \overset{i.i.d.}{\sim} P$ and $P \sim DP(\alpha, P_0)$.

Proof

We can consider Polya sequence as an outcome of Polya urn, we see that it is exchangeable. By de Finetti theorem exists such probability measure P such that $X_i|P \overset{i.i.d.}{\sim} P$. So far we have proved existence of the DP and know that DP generates a Polya sequence. Since the RPM given by de Finetti's theorem is unique this proves that $P \sim DP(\alpha, P_0)$.