Gaussian-Stick Breaking Process

Jay Yoon Lee
Computer Science Department
jaylee@cs.cmu.edu

Manzil Zaheer
Machine Learning Department
manzil@cmu.edu

1 Introduction

Dirichlet process [1, 2] forms an important part of Bayesian nonparametrics. In the seminal paper by Sethuraman, a constructive definition of such priors where given through the stick breaking process [3, 4]. Although traditional dirichlet process favors models whose complexity grows with the dataset size as it is a nonparametric prior, yet it only allows the number of clusters (tables) to grow as $\alpha \log(N)$ as $N \to \infty$ almost surely. This imposes restrictions and thus cannot capture behaviours with larger number of clusters, e.g. the power law behaviour arising in many situations.

To overcome this, many literature in Bayesian nonparametrics focused on models for collections of distributions based on extensions of the Dirichlet process and other stick-breaking priors. These extensions are generated by playing around with the generative sequence in the stick-breaking construction [5, 6, 7, 8, 9, 10, 11].

We suggest a generative model for the stick breaking process which should be able to handle arbitrary tail behaviours or capture behaviours of varying degree of growth of number of cluster with data size.

2 Notations and Preliminary Results

Definition 1 *Falling factorial*

$$(x)_{n\uparrow a} = x(x+a)\cdots(x+(n-1)a) = \prod_{i=0}^{n-1}(x+ia) \tag{1}$$

In particular recognize that for $x = 1$ and $a = 1$, we have $(1)_{n\uparrow 1} = n!$.

Definition 2 *(Unsigned) Stirling Number of first kind $s(n,k)$:* They represent the number of permutations of an n-set with precisely k cycles. Next we prove some basic properties of these numbers. Conventionally we define $s(0,0) = 1$, $s(n,0) = 0$ and $s(n,k) = 0$, if $n < k$.

Lemma 1 *Stirling numbers of first kind satisfy the following recurrence relation:*

$$s(n,k) = s(n-1,k-1) + (n-1)s(n-1,k) \tag{2}$$

Proof: Consider forming a new permutation with n objects from a permutation of $n-1$ objects by adding a distinguished object. There are exactly two ways in which this can be accomplished.

1. First, we could form a singleton cycle, leaving the extra object fixed. This increases the number of cycles by 1 and so accounts for the $s(n-1,k-1)$ term in the recurrence.
2. Second, we could insert the object into one of the existing cycles. Consider an arbitrary permutation of $n-1$ objects with k cycles. To form the new permutation, we insert the new object before any of the $n-1$ objects already present. This explains the $(n-1)s(n-1,k)$ term of the recurrence.

These two cases include all of the possibilities, so the recurrence relation follows with the given initial conditions.

Next we have the following relationship:

Lemma 2 *Stirling numbers of first kind satisfy the following:*

$$(x)_{n\uparrow 1} = \sum_{k=0}^{n} s(n,k)x^k \tag{3}$$

Proof: We proceed by induction on n. For $n = 0$ and $n = 1$ we have that:

$$(x)_{0\uparrow 1} = s(0,0) = 1 \quad \text{and} \quad (x)_{1\uparrow 1} = s(1,0) + xs(1,1) = x \tag{4}$$

Now assume the claim is true for $n = l$, i.e.:

$$(x)_{l\uparrow 1} = \sum_{k=0}^{l} s(l,k)x^k \tag{5}$$

Then the inductive step for $n = l + 1$ can be shown as:

$$\begin{aligned}
(x)_{l+1\uparrow 1} &= (x+l)(x)_{l\uparrow 1} \\
&= x(x)_{l\uparrow 1} + l(x)_{l\uparrow 1} \\
&= x\sum_{k=0}^{l} s(l,k)x^k + l\sum_{k=0}^{l} s(l,k)x^k \\
&= \sum_{k=1}^{l+1} s(l,k-1)x^k + l\sum_{k=0}^{l} s(l,k)x^k \\
&= \sum_{k=0}^{l+1} \left[s(l,k-1) + ls(l,k)\right] x^k \\
&= \sum_{k=0}^{l+1} s(l+1,k)x^k
\end{aligned} \tag{6}$$

Thus by principle of mathematical induction, we complete the proof for (3).

Definition 3 *Generalized Stirling number:* Generalizing the striling number of first kind, we define Generalized Stirling number $s_{p,q}(n,k)$ as the coefficients satisfying:

$$(x)_{n\uparrow 1} = \sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} \tag{7}$$

Lemma 3 *Generalized Stirling number satisfy the following recurrence relation:*

$$s_\theta(n,k) = s_\theta(n-1,k-1) + (n-1-k\theta)s_\theta(n-1,k) \tag{8}$$

Proof: Starting with the definition of generalized Stirling number (7) as:

$$\begin{aligned}
(x)_{n\uparrow 1} &= (x+n-1)(x)_{n-1\uparrow 1} \\
\sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} &= (x+n-1)\sum_{k=0}^{n-1} s_\theta(n-1,k)(x)_{k\uparrow\theta} \\
\sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} &= \sum_{k=0}^{n-1} s_\theta(n-1,k)\left((x+k\theta)(x)_{k\uparrow\theta} + (n-1-k\theta)(x)_{k\uparrow\theta}\right) \\
\sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} &= \sum_{k=0}^{n-1} s_\theta(n-1,k)(x)_{k+1\uparrow\theta} + \sum_{k=0}^{n-1} s_\theta(n-1,k)(n-1-k\theta)(x)_{k\uparrow\theta} \\
\sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} &= \sum_{k=1}^{n} s_\theta(n-1,k-1)(x)_{k\uparrow\theta} + \sum_{k=0}^{n-1} s_\theta(n-1,k)(n-1-k\theta)(x)_{k\uparrow\theta}
\end{aligned} \tag{9}$$

$$\sum_{k=0}^{n} s_\theta(n,k)(x)_{k\uparrow\theta} = \sum_{k=0}^{n} \left(s_\theta(n-1,k-1) + s_\theta(n-1,k)(n-1-k\theta)\right)(x)_{k\uparrow\theta}$$

From which the recurrence relationship follows by comparing coefficients of $(x)_{k\uparrow\theta}$.

3 Review and Motivation

Clustering is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. Clustering is the task of grouping a set of objects S in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). In other words clustering aims to find a partition $\mathcal{Q}$ of the set S respecting the similarity considerations. Formally a partition $\mathcal{Q}$ of set S is a disjoint family of non-empty subsets of S whose union is S. Now in a Bayesian setting for cluster analysis we would need to introduce prior over $\mathcal{Q}$ and cluster parameter and compute posterior over both. Towards this end, we have to work with random partitions.

A basic method for sequential construction of consistent and exchangeable random partitions is through Chinese restaurant process. By exchangeablity we mean that for the random partitions generated relabelling $\{1,...,n\}$ does not change the distribution of the partition, and it is consistent in the sense that the distribution of the partition of $n1$ obtained by removing the element n from the random partition at time n is the same as the law of the random partition at time $n1$. These properties are crucial to typical clustering applications. (Sometimes these properties are inapplicable, e.g. exchangibility must be given up for timestamped data and then we can use distance dependent CRP [12].)

The Chinese restaurant process is a discrete-time stochastic process indexed by positive-integer n is a partition $\mathcal{Q}_n$ of the set $\{1,...,n\}$ whose probability distribution is determined as follows. Let $\{P_k\}_{k=0}^{\infty}$ be an arbitrary sequence of random variables with $P_i \geq 0$ and $\sum_k P_i \leq 1$. Given the entire sequence, at time $n = 1$, the trivial partition $\{\{1\}\}$ is obtained with probability 1. Or in terms of the Chinese restaurant metaphor the first customer be seated at the first table. Then for $n \geq 1$, given the partition $\mathcal{Q}_n$, regarded as a placement of the first n customers at tables of the Chinese restaurant, with K_n occupied tables, at time $n+1$ the customer would be placed at:

- placed at occupied table k with probability P_k
- placed at new table with probability $1 - \sum_{k=1}^{K_n} P_k$

For the tradition Dirichlet Process [1, 2] we have $P_k = \frac{n_k}{n+\alpha}$, where n_k is the number of customers at the k-th table and in case of Pitman-Yor Process we have $P_k = \frac{n_k-\theta}{n+\alpha}$.

Now to study the growth of number of tables/clusters with data, we need to look at behaviour K_n. It is easy to observe from the construction that, the sequence $\{Kn\}_{n=1}^{\infty}$ forms a Markov chain, starting at $K_1 = 1$, with increments in $\{0,1\}$, and (possibly inhomogeneous) transition probabilities.

3.1 Dirichlet Process

The transition probabilities are:

$$\begin{aligned} P(K_n = k | K_{n-1} = k) &= \frac{n-1}{\alpha+n-1} \\ P(K_n = k | K_{n-1} = k-1) &= \frac{\alpha}{\alpha+n-1} \end{aligned} \tag{10}$$

So we get the recurrence relation:

$$P(K_n = k) = \frac{\alpha}{\alpha+n-1} P(K_{n-1} = k-1) + \frac{n-1}{\alpha+n-1} P(K_{n-1} = k) \tag{11}$$

One can easily check that following is the solution to the recurrence relation:

$$P(K_n = k) = \frac{\alpha^k}{(\alpha)_{n\uparrow1}} s(n,k) \tag{12}$$

The expected value of number of cluster/tables goes as:

$$\mathbb{E}[K_n] = \sum_{k=0}^{n} kP(K_n = k) = \frac{1}{(\alpha)_{n\uparrow 1}} \sum_{k=0}^{n} s(n,k)k\alpha^k \tag{13}$$

Differentiating (3) with respect to x yields:

$$(x)_{n\uparrow 1} \sum_{i=1}^{n} \frac{1}{x+i-1} = \sum_{k=0}^{n} s(n,k)kx^{k-1} \tag{14}$$

which can be used for $x = \alpha$ to give us:

$$\mathbb{E}[K_n] = \sum_{i=1}^{n} \frac{\alpha}{\alpha+i-1} \tag{15}$$

Further for asymptotic behaviour, we can directly see it is $\alpha \log n$, however we can show a stronger result that:

$$\frac{K_n}{\alpha \log n} \xrightarrow{a.s.} 1 \tag{16}$$

by simply defining indicator Bernoulli random variables W_n representing the event of a new table created and then applying strong law of large numbers.

3.2 Pitman-Yor Process

Assume $\theta \neq 0$. The transition probabilities are:

$$\begin{aligned} P(K_n = k | K_{n-1} = k) &= \frac{n-1-k\theta}{\alpha+n-1} \\ P(K_n = k | K_{n-1} = k-1) &= \frac{\alpha+k\theta}{\alpha+n-1} \end{aligned} \tag{17}$$

So we get the recurrence relation:

$$P(K_n = k) = \frac{\alpha+k\theta}{\alpha+n-1} P(K_{n-1} = k-1) + \frac{n-1-k\theta}{\alpha+n-1} P(K_{n-1} = k) \tag{18}$$

One can easily check that following is the solution to the recurrence relation:

$$P(K_n = k) = \frac{(\alpha)_{k\uparrow\theta}}{(\alpha)_{n\uparrow 1}} s_\theta(n,k) \tag{19}$$

The expected value of number of cluster/tables goes as:

$$\begin{aligned} \mathbb{E}[K_n] &= \sum_{k=0}^{n} kP(K_n = k) \\ &= \frac{1}{(\alpha)_{n\uparrow 1}} \sum_{k=0}^{n} k(\alpha)_{k\uparrow\theta} s_\theta(n,k) \\ &= \frac{1}{\theta(\alpha)_{n\uparrow 1}} \sum_{k=0}^{n} (\alpha+k\theta)(\alpha)_{k\uparrow\theta} s_\theta(n,k) - \alpha(\alpha)_{k\uparrow\theta} s_\theta(n,k) \\ &= \frac{1}{\theta(\alpha)_{n\uparrow 1}} \sum_{k=0}^{n} (\alpha)_{k+1\uparrow\theta} s_\theta(n,k) - \alpha(\alpha)_{k\uparrow\theta} s_\theta(n,k) \\ &= \frac{\alpha}{\theta(\alpha)_{n\uparrow 1}} \left(\sum_{k=0}^{n} (\alpha+\theta)_{k\uparrow\theta} s_\theta(n,k) - \sum_{k=0}^{n} (\alpha)_{k\uparrow\theta} s_\theta(n,k) \right) \\ &= \frac{\alpha}{\theta(\alpha)_{n\uparrow 1}} \left((\alpha+\theta)_{n\uparrow 1} - (\alpha)_{k\uparrow 1} \right) \\ &= \frac{\alpha}{\theta} \left(\frac{(\alpha+\theta)_{n\uparrow 1}}{(\alpha)_{n\uparrow 1}} - 1 \right) \end{aligned} \tag{20}$$

where we used the (7) and again be reminded that $\theta \neq 0$. The asymptotic behaviour can be easily seen as n^θ.

4 General Stick Breaking

An equivalent way to construct the random partitions is through stick breaking processing, i.e. directly using the limiting frequencies of partition size counts. The connection between Chinese Restaurant process and stick breaking process is a simple (and quite beautiful) application of Polya's urn theorem.

In the Chinese Restaurant process, consider the table #1 without loss of generality. Assign an indicator variable $W_i(1)$ of belonging to table 1 to every customer i. Then polya's urn theorem says that $\sum_i W_i^{(1)}/n \overset{a.s.}{\to} U_i$ where U_i is a beta distributed random variable. Now looking at table #2, it is easy to see by the same Polya urn argument that the asymptotic fraction of customers among those that are not in table #1, is again beta distributed, say U_2. Now the overall fraction would be $(1-U_1)U_2$. Arguing by induction as above, one obtains the stick breaking process.

The appeal for stick breaking process lies in the fact that often in practical applications we have some domain knowledge in these limiting frequencies and we want our models to posses them. One potential application we are looking forward is the topic modelling of WWW. In WWW we have on one hand the web graph and on the other hand document associated with each webpage. Now, for a given "sane" webpage it is expected that topic distribution for the links and content would be similar. A natural way to enforce similarity in both the topic distribution, is to make share some latent parameters through the stick breaking process dictating link and content topics. (Stick breaking process for random graphs have also been studied, e.g. see [**?**], which is nothing but a different view to look at random partitions)

Formally, the stick-breaking priors are almost surely discrete random probability measures $\mathcal{P}$ that can be represented generally as:

$$\mathcal{P}(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}(\cdot) \tag{21}$$

where δ_{θ_k} denotes a discrete measure concentration at θ_k. In (21), the π_k are random variables (called random weights) chosen to be independent of θ_k such that $0 \leq \pi_k \leq 1$ and $\sum_k \pi_k = 1$ almost surely. The $\theta_k \sim H$ are drawn independently and identically from a continuous base measure. The method of construction for the random weights is what sets stick-breaking priors apart from general random measures. The general stick breaking construction goes as follows:

1. Draw $V_j \sim \mathsf{Beta}(a_j, b_j)$
2. Set $\pi_k = V_k \prod_{j<k}(1-V_j)$

where $a_j, b_j > 0$. Informally, this construction can be thought of as a stick- breaking procedure, where at each stage we independently and randomly, break what is left of a stick of unit length and assign the length of this break to the current π_k value. By suitably choosing a_j, b_j many different measures have been proposed already:

1. the Ferguson Dirichlet process [1, 2] with $a_j = 1$ and $b_j = \alpha$
2. the Pitman-Yor Process (also known as two-parameter Poisson Dirichlet process) [5] with $a_j = 1-\theta$ and $b_j = \alpha + j\theta$
3. the Kernel Stick Breaking Process [6] with a data (position) dependent sequence a_j, b_j and $\pi_k = V_k \mathcal{K}(x, \Gamma_k) \prod_{j<k}(1 - V_j \mathcal{K}(x, \Gamma_j))$
4. the Beta Two-Parameter Processes and finite dimensional Dirichlet priors [7] with finite sequences a_i, b_j

We also target to engineer the sequence a_j, b_j in a novel manner, details of which follows.

However one is not totally free in selecting any sequence a_j, b_j, because finally the stick breaking process should be distribution, i.e. the random weights must sum upto 1. Towards this end Ishwaran gave a necessary and sufficient condition which V_k must satisfy [4]

Lemma 4 *For the random weights in the Gaussian-Stick Breaking process,*

$$\sum_{k=1}^{\infty} \pi_k \overset{a.s.}{=} 1 \qquad \Leftrightarrow \qquad \sum_{k=1}^{\infty} \mathbb{E}(\log(1-V_k)) = -\infty \tag{22}$$

Proof: To establish eq.(22), first consider the following equation where $V_N = 1$ will make the probability sum up to 1.

$$1 - \sum_{k=1}^{N-1} \pi_k = (1 - V_1) \cdots (1 - V_{N-1}) \tag{23}$$

Now, take the limit of eq.(23) as $N \to \infty$ and take logs, similar to HW3, to see that:

$$\sum_{k=1}^{\infty} \pi_k \overset{a.s.}{=} 1 \qquad \Leftrightarrow \qquad \sum_{k=1}^{\infty} (\log(1 - V_k)) \overset{a.s.}{=} -\infty$$

The expression on the right-hand side is a sum of independent random variables and, therefore, by the Kolmogorov three series theorem, equals $-\infty$ almost surely iff $\sum_{k=1}^{\infty} \mathbb{E}(\log(1 - V_k)) = -\infty$. Alternatively, by (23),

$$\prod_{N=1}^{\infty} \frac{\mathbb{E}\left(\sum_{k=N+1}^{\infty} p_k\right)}{\mathbb{E}\left(\sum_{k=1}^{\infty} p_k\right)} = \prod_{N=1}^{\infty} \mathbb{E}(1 - V_N) = \prod_{N=1}^{\infty} \frac{b_N}{a_N + b_N}.$$

If $\sum_{N=1}^{\infty} \log(1 + a_N/b_N) = +\infty$, then the right-hand side equals zero and we must have that $\mathbb{E}\left(\sum_{k=1}^{\infty} p_k\right) \to 1$. However, because $\sum_{k=1}^{N} p_k$ is positive and increasing, it follows that $\sum_{k=1}^{\infty} p_k \overset{a.s.}{=} 1$.

The above lemma trivially holds for traditional Dirichlet Process. In case of Pitman-Yor Process, we verify by directly showing $\sum_{N=1}^{\infty} \log(1 + a_N/b_N) = +\infty$ using $\log(1 + x) \geq x - x^2/2$and then rest follows as in the proof of lemma 4.

$$\begin{aligned} \sum_{N=1}^{\infty} \log\left(1 + \frac{a_N}{b_N}\right) &= \sum_{N=1}^{\infty} \log\left(1 + \frac{1-\theta}{\alpha + N\theta}\right) \\ &\geq \underbrace{\sum_{N=1}^{\infty} \frac{1-\theta}{\alpha + N\theta}}_{\longrightarrow\infty} - \underbrace{\sum_{N=1}^{\infty} \left(\frac{1-\theta}{\alpha + N\theta}\right)^2}_{\text{something finite}} \end{aligned} \tag{24}$$

For the Kernel Stick Breaking Process, by design we have $\log(1 - V_k \mathcal{K}(x, \Gamma_k)) < 0$, so the expectation is strictly negative and this condition is satisfied.

5 Gaussian-Stick Breaking (GSB) model

In this section, we introduce the Gaussian-Stick Breaking (GSB) process. The main idea behind Gaussian-Stick is that we use hyperparameters to generate parameters of beta distributions while general sticks fix the parameters to constant or to some sequence a_j, b_j, i.e. $\mathsf{Beta}(a_j, b_j)$. To exploit well established properties of exponential family in designing applications, we have chosen gaussian random variable ξ_i and exponential function to generate the parameter α_i which we will use in generating the sticks. We formally define the GSB as:

- $\theta \sim \mathsf{Gaussian\text{-}Stick}(\rho, \mu = 0, \sigma^2 = 1)$
  1. $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$
  2. $\alpha_i = \rho_i e^{\xi_i}$
  3. $\beta_i \sim \mathsf{Beta}(1, \alpha_i)$
  4. $\theta_k = \beta_k \prod_{j<k} (1 - \beta_j)$

With the provided definition, in the next section, we will examine the properties of the tail properties of the GSB and also prove that GSB is a valid process by showing the sticks θ_k sums up to 1 as sequence goes to infinity.

6 Properties of Gaussian-Stick Breaking Process

The lack of control in the distribution was the biggest motivation in constructing new stick breaking process. Therefore in this section, we first examine the tail behaviours of GSB: behaviour of moments and required design constraints in order to have finite moments. And to show that distribution generated from GSB is a valid probability function, we prove that the sticks θ_k sums up to 1.

6.1 Tail Properties

Starting from the first moment, here we will generalize the expression for the kth moment.

- **1st order:** $\mathbb{E}\left[\sum \rho_i e^{\xi_i}\right] = \sum_i \rho_i \int e^{\xi_i} e^{-\frac{\xi_i^2}{2}} d\xi_i = \sum_i \rho_i e^{\frac{1}{2}}$
- **2nd order:** $\mathbb{E}\left[\left(\sum_i \rho_i e^{\xi_i}\right)^2\right] = e||\rho||_1^2 + (e^2 - e)||\rho||_2^2$

  $\vdots$
- **kth order:** $\mathbb{E}\left[\left(\sum_i \rho_i e^{\xi_i}\right)^2\right]$ becomes a combination of $||\rho||_1^k, ||\rho||_2^k, \dots, ||\rho||_k^k||$.

Analyzing the resutls, we can conclude the kth order moment is a combination of $||\rho||_1^k, ||\rho||_2^k, \dots, ||\rho||_k^k||$. Since $||\rho||_1 \geq ||\rho||_k$ for any $k \geq 1$, the only constraint we need for the existence of the moments are $||\rho||_1 < \infty$. Therefore we conclude that given $||\rho||_1 < \infty$, the moments exist and we can control the moments by controlloing ρ_is.

6.2 Proof on the validity of Gaussian-Stick Brekaing

We prove that $\sum_{k=1}^{\infty} \theta_k \overset{a.s.}{=} 1$ by proving that $\sum_{k=1}^{\infty} \mathbb{E}\left[\log(1-\beta_k)\right] \overset{a.s.}{=} -\infty$ using the Lemma 4 we mentioned in the previous section [4].

To show the right hand side of eq.(22), we first obtain an expression for $\mathbb{E}\left[\log(1-\beta_k)\right]$.

6.2.1 Evaluation of $\mathbb{E}\left[\log(1-\beta_k)\right]$

$$
\begin{aligned}
\mathbb{E}\left[\log(1-\beta_k)\right] &= \mathbb{E}\left[\log(\gamma_k)\right] && \text{where } \gamma_k \sim \mathsf{Beta}(\alpha_k, 1) \\
&= \frac{1}{C}\int_0^1 \log\gamma_k \ \gamma_k^{\alpha_k - 1} d\gamma_k && \text{where } C \text{ is normalizer for } \mathsf{Beta}(\alpha_k, 1) \\
&= \frac{1}{C\alpha_k}\left[\left[\log\gamma_k \ \gamma_k^{\alpha_k}\right]_0^1 - \int_0^1 \frac{1}{\gamma_k}\gamma_k^{\alpha_k} d\gamma_k\right] \\
&= -\frac{1}{\alpha_k}
\end{aligned}
$$

With the above derivation of $\mathbb{E}\left[\log(1-\beta_k)\right] = -\frac{1}{\alpha_k}$ gives us result that $\sum_{k=1}^{\infty} \mathbb{E}\left[\log(1-\beta_k)\right] = -\sum_{k=1}^{\infty} \frac{1}{\alpha_k}$. Therefore to show that $\sum_{k=1}^{\infty} \mathbb{E}\left[\log(1-\beta_k)\right] \overset{a.s.}{=} -\infty$, we will show that $\sum_{k=1}^{\infty} \frac{1}{\alpha_k} \overset{a.s.}{=} \infty$.

Lemma 5 *For series $\rho_k \geq 0$ such that $||\rho||_1 < \infty$, $\lim_{l\to\infty} \sum_{k=1}^{l} \frac{1}{\rho_k} = \infty$.*

Proof: We will prove the result using 2nd Borel-Canteli Lemma, which states that $\sum_{j=1}^{\infty} P(V_j > \epsilon) = \infty \longrightarrow P(V_j > \epsilon \text{ infinitely often}) = 1$. Using this fact, we will prove the following:

$$\sum_{j=1}^{\infty} P\left(\frac{1}{\alpha_j} > \epsilon\right) = \infty \Rightarrow P\left(\frac{1}{\alpha_j} > \epsilon \text{ infinitely often}\right) = 1 \Rightarrow P\left(\sum_{j=1}^{\infty} \frac{1}{\alpha_j} = \infty\right) = 1 \quad \text{almost surely}$$

Now, inspecting $P(\frac{1}{\alpha_k} > \epsilon)$ for $0 < \epsilon < 1$ will let us know the sufficient condition for 2nd Borel-Canteli Lemma holds.

$$\begin{aligned} P\left(\frac{1}{\alpha_j} > \epsilon\right) &= P\left(\frac{1}{\rho_j} e^{\xi_j} > \epsilon\right) \\ &= (\xi_j < -\log(\rho_j \epsilon)) \end{aligned}$$

Looking at the last term of the above equation, as $\rho_j \to 0$, $P(\xi_j < -\log(\rho_j \epsilon)) \to 1$ and therefore $\sum_{j=1}^{\infty} P(\frac{1}{\alpha_j} > \epsilon) = \infty$, the sufficient condition for 2nd Borel-Canteli lemma, holds.

7 Conclusion and Future Work

We study the growth of cluster size and corresponding stick-breaking process for Chinese Restaurant type processes with the aim to eventually design stick-breaking process which can handle arbitrary cluster size growth. To achieve this we have come up with a novel stick breaking process and proved its validity. Future work remains to find exact design methodology for the sequence ρ to achieve desired cluster growth.

References

[1] T. S. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, vol. 1, no. 2, pp. pp. 209–230, 1973.

[2] C. E. Antoniak, "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, vol. 2, no. 6, pp. 1152–1174, 1974.

[3] J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica*, vol. 4, pp. 639–650, 1994.

[4] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 161–173, 2001.

[5] J. Pitman and M. Yor, "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator," *The Annals of Probability*, vol. 25, no. 2, pp. 855–900, April 1997.

[6] D. B. Dunson and J.-H. Park, "Kernel stick-breaking processes," *Biometrika*, vol. 95, no. 2, pp. 307–323, 2008.

[7] H. Ishwaran and M. Zarepour, "Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models," *Biometrika*, vol. 87, no. 2, pp. 371–390, 2000.

[8] S. N. MacEachern, "Dependent Dirichlet Processes," Dept. of Statistics, Ohio State University, Tech. Rep., 2000.

[9] A. E. Gelfand, A. Kottas, and S. N. MacEachern, "Bayesian nonparametric spatial modeling with dirichlet process mixing," *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 1021–1035, 2005.

[10] J. E. Griffin and M. F. J. Steel, "Order-based dependent dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 473, pp. 179–194, 2006.

[11] J. A. Duan, M. Guindani, and A. E. Gelfand, "Generalized spatial dirichlet process models," *Biometrika*, vol. 94, no. 4, pp. 809–825, 2007.

[12] D. M. Blei and P. I. Frazier, "Distance dependent chinese restaurant processes," *J. Mach. Learn. Res.*, vol. 12, pp. 2461–2488, 2011.

Appendix: Contribution of Work

7.1 Jay Yoon Lee

- Development of Gaussian Stick Breaking
- GSB formulation
- GSB Properties
- Lemma 4 proof

7.2 Manzil Zaheer

- Development of Gaussian Stick Breaking
- Introduction, Notations & Conclusion
- DP/Pitman-Yor/CRP
- Traditional Stick-Breaking