UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS
Y DE LA ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

Random Partitions Models

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRA EN CIENCIAS

PRESENTA:
MARÍA FERNANDA GIL LEYVA VILLA

DIRECTOR DE LA TESIS:
RAMSÉS HUMBERTO MENA CHÁVEZ
I.I.M.A.S.

MÉXICO, D.F., AGOSTO 2016

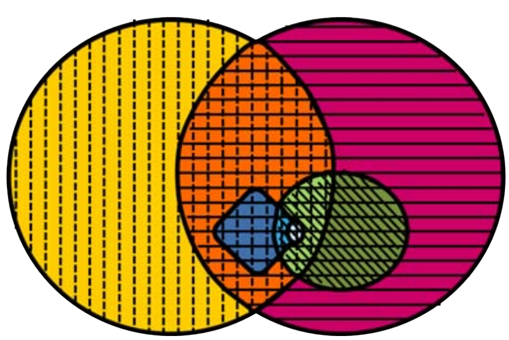
Random partitions models

Author: María Fernanda Gil Leyva Villa
Advisor: PhD. Ramsés Humberto Mena Chávez

Professional examination committee members:

1. PhD. Ramsés Humberto Mena Chávez
2. PhD. Arno Siri-Jégousse
3. PhD. Gerónimo Uribe Bravo
4. PhD. Asael Fabian Martínez Martínez
5. PhD. Sergio Iván López Ortega

A thesis presented for the degree of
Master in Mathematical Science



Contents

Acknowledgements

In a general scenario I want to thank all the beings that somehow collaborated in then realization of this thesis, including the snails that motivated the drawings and the guys that harvested, recollected, processed and finally prepared the coffee I drank while writing the thesis. As you can imagine, that is big group of beings, so let us just leave it at that.

In a particular scenario, first of all I want to thank my parents who have always supported me, and loved me the way I am. Without them I would not be standing where I am (obviously). I want to thank them for their kindness, willingness, courage, serenity and intelligence, they have always been and always will be the absolutely most standing persons in my life. Thank you mom and dad for your guidance and extreme patience (specially towards me). Thank you for offering your smile as a priceless prize I earn even without deserving it.

Second of all I want to, particularly, thank the rest of my family, my grandmother, my aunts, uncles, cousins, and my second mother. You might not be con aware of it, but the fact is that you all are a huge motive for my to keep going forward and making an effort. It is of great importance to me to make you proud, but more importantly to make you (even though it is occasionally) happy.

Not fourth of all, I want to thank the most amazing group of people I have ever met, my friends. I assume they are the smartest, funniest, most multifaceted and capable group of persons, I will ever get to know. I am not sure how big their accumulated I.Q. is, I am not even sure of how they measure such number, but I am definitely sure that I strongly believe that their accumulated I.Q. must be bigger than the accumulated I.Q. of a set including a white shark, a belluga, an elephant and an ant. Anyway, thank you for keeping me smiling, I really admire each and every one of you.

Fourth of all, I want to thank Ramses, my thesis advisor, for sharing his knowledge with me and supporting me in the endeavor of achieving my goals. I also want to say thank you to all the members of my professional examination committee for investing your time in reading my, not that short, thesis. Thank you for all your advices and comments that improved this project.

Almost finally, I want to thank (even though they did not helped at all in the elaboration of the thesis) to my three dogs and to all the pretty girls walking out there for being pretty. Also thank you to the guys I train parkour with (not for being pretty, but for jumping pretty).

Finally I want to thank to Darwin and/or Newton who might have helped.

Agradecimientos

En un escenario general quiero agradecer a todos los seres que de alguna manera colaboraron en la realización de esta tesis, incluyendo a los caracoles que motivaron los dibujos y a las personas que cosecharon, recolectaron, procesaron y finalmente prepararon el café que bebí mientras escribía la tesis. Como podrán imaginarse este conjunto de seres es bastante grande, así que dejémoslo aquí.

En un escenario particular, primero que nada quiero agradecer a mis padres, quienes siempre me han apoyado y amado de la manera que soy, sin ellos no estaría parada en donde estoy (obviamente). Quiero agradecerles por toda su bondad, fortaleza, serenidad e inteligencia, siempre han sido y simpre serán las personas más enteras y completas en mi vida. Gracias mamá y papá for su ejemplo y extrema paciencia (especialmente hacia mí). Gracias por siempre ofrecerme su sonrisa como motivo (incluso en las ocasiones en las que no lo merezco).

En segundo lugar quiero, partícularmente, agradecer al resto de mi familia. A mi Tita, a mis tías, tíos, primos y a mi segunda mamá. Tal vez no sean consientes de ello, pero la verdad es que todos ustedes son un impulso gigante en mí, una razón por la cual seguir esforzándome. Es muy importante para mí hacerlos sentir orgullosos, pero sobretodo felices (así sea ocasionalmente).

No en cuarto lugar, quiero agradecer al grupo de personas más increíble que conozco, mis amigos. Supongo que son el grupo de personas más brillantes, divertidas, multifacéticas y capacez que conoceré. No estoy segura de cual es su I.Q. acumulado, de hecho no estoy ni segura de como miden esa cifra, pero si estoy segura de que creo fuertemente que su I.Q. acumulado debe de ser mayor que el I.Q. acumulado del conjunto que incluye a un tiburón blanco, una belluga, un elefante y una hormiga. De cualquier forma, gracias por mantenerme sonriendo, realmente los admiro y quiero mucho a todos y cada uno de ustedes.

En cuarto lugar, quiero agradecer a Ramses, mi asesor de tesis, por compartir su conocimiento conmigo, por apoyarme en la labor de alcanzar mis metas. También quiero agradecer a todos mis sinodales for invertir su tiempo en leer mi, no tan corta, tesis. Gracias por sus consejos y comentarios que ayudaron a mejorar este trabajo.

Casi finalmente, quiero agradecer (a pesar de que no ayudaron para nada en la elaboración de la tesis) a mis tres perras, Gluck, Leeloo y Gamma, y también a todas las niñas bonitas caminando en el mundo, por ser bonitas. Adicionalmente quiero agradecer a mis compañeros de parkour (no por ser bonitos, pero si por saltar bonito).

Finalmente quiero agradecer a Darwin y/o a Newton, quienes tal vez ayudaron.

Introduction

Random partitions occur in mathematics and physics in a wide variety of contexts. Often, it happens that a certain quantity of interest can be expressed as a sum over partitions, commonly one can recognize in such a sum a discrete version of some random matrix integral. Random partitions are also closely related to genetics and the study of populations. For instance, a partition can record a state of a certain growth or branching process. In genetics the main stochastic model, called the *Kingman's Coalescent*, is a time reversed continuous Markov process $\{X_t\}_{t\geq 0}$ that takes values over the set of partitions of $[n]$ (where n is the number of individuals in the population of study) that is, for every $t \geq 0$, X_t, is a random partition of $[n]$. Some more examples of the uses of random partitions in physics, geometry, ergodic theory and other areas can be found on [15] in the bibliography. As random partitions occur in many mathematical contexts, the study of these random elements could be quite extensive, this is why we will restrict the analysis.

In the thesis we will analyse random partitions, conceiving them as one of the key concepts in Bayesian non-parametric statistics. For instance, given a finite sample, one can approximate its joint density (or probability) function through a mixture of parametric distributions, here the random partitions play a main role, as the number of components in the mixture is typically regarded as the random number of blocks in a certain random partition, this matter is deeply studied in the work of Fabian Martínez [14] in the bibliography. Another situation, where the random partitions play a main role, is that of estimating the number of distinct species in some population. As the species determine an unknown partition of the population, the number of species can be regarded as the number of blocks in a random partition, the work of Favaro, Lijoi, Mena and Prunster [3] as well as the work of Lijoi, Mena and Prunster [13] specialize in this topic. Like these two, there are many other problems that appear in a Bayesian non-parametric context, in which random partitions arise naturally. Hence, the understanding of the theoretical properties of random partitions becomes essential to the endeavor of providing a precise statistical analysis.

Introducción

Las particiones aleatorias ocurren en matemáticas y física en muchos contextos. Comúnmente sucede que cierta cantidad de interés puede ser expresada como una suma sobre particiones, y en ocasiones es posible reconocer en dicha suma una versión discreta de alguna matriz integral aleatoria. Los modelos de particiones aleatorias también se encuentran fuertemente ligados con áreas como la genética y estudios de poblaciones. Por ejemplo, una partición puede representar el estado de cierto proceso de ramificaión o crecimiento de alguna población. En genética el principal modelo estocástico, llamado *Coalescente de Kingman*, (usado para rastrear genealogias de alguna población) es una cadena de Markov reversa en el tiempo, a tiempo continuo, $\{X_t\}_{t\geq 0}$, que toma valores en el conjunto de particiones de $[n]$ (donde n es el número de individuos en la población objetivo). Esto es, para toda $t \geq 0$, X_t, es una partición de $[n]$. Otros ejemplos acerca del uso de particiones aleatorias en física, geometría, teoría ergódica y otras áreas, pueden ser encontrados en [15] de la bibliografía. Como las particiones aleatorias ocurren en muchos contextos, el estudio acerca de estos objetos aleatorios puede resultar muy extenso, por esta razón, en la tesis se restringirá un poco el análisis.

En este documento se estudiará a las particiones aleatorias considerándolas como un concepto clave de la estadística Bayesiana no paramétrica. Por ejemplo, dada una muestra finita, es posible aproximar su función de densidad (o de masa de probabilidad) a través de una mezcla de distribuciones paramétricas, en este contexto las particiones aleatorias son una pieza esencial, pues el número de componentes de la mezcla es típicamente considerado como el número de bloques de una partición aleatoria, en el trabajo de Fabian Martínez [14] se analiza a detalle esta situación. Otro problema en el que las particiones aleatorias juegan un papel principal, es aquél de estimar el número de especies en alguna población. Como las especies determinan una partición desconocida de la población, el número de especies puede ser considerado como el número de bloques de alguna partición aleatoria, el trabajo de Favaro, Lijoi, Mena y Prunster [3] así como el trabajo de Lijoi, Mena y Prunster [13] se especializan en este tema. Como estos, existen muchos otros problemas que aparecen en un contexto Bayesiano no paramétrico, en los que las particiones aleatorias surgen de manera natural. De esta forma, el entendimiento de las propiedades teóricas de los modelos de partitionces aleatorias resulta esencial para la labor the realizar un análisis estadístico preciso.

Organization of the thesis

In the thesis we will analyse some of the most important results and properties about random partitions, clarifying related concepts and linking them in order to provide a firm understanding of the matter. As mentioned in the introduction, throughout the thesis, we will regard the random partitions models as a key concept in the context of Bayesian non-parametric analysis, for this reason, we will emphasize the relation between random partitions, exchangeable random variables and random probability measures.

The aim of first two chapters of the document is to introduce the basic concepts about random partitions. In Chapter 1 we give the preliminary concepts about the combinatorial structures that will be used throughout the thesis, such as a partition of a set with a finite number of elements and the composition of an integer n, this chapter is based mainly in the work of Jim Pitman [18] and María Luisa Pérez Segui [16]. Chapter 2 is divided in two parts, the first one is dedicated to introduce the notion of a probability distribution over the set of partitions of a finite set, here we define the notion of exchangeability in the partitions context as well as the notion of consistency of a family of partitions. In the second section of this chapter we introduce the idea of the generation of random partitions by the consecutive sampling from a exchangeable sequence random variables $\{X_i\}_{i=1}^{\infty}$. The work developed in Chapter 2 is based on the work of Fabián Martínez [14] and Jim Pitman [18].

Chapter 3 based on the work of Kingman [11] and chapter 4, based on the work Hjort, Holmes, Múller and Walker [6] (chapter 3.3) and James, Lijoi and Prunster, [9], are respectively dedicated to the study of completely random measures and normalized independent random measure and how to generate an exchangeable partition by sampling from a sequence of exchangeable random variables $\{X_i\}_{i=1}^{\infty}$ with random distribution given by the normalized independent random measure $\tilde{p}$. Here we also study the subordinators as a particular case of a completely random measure.

The work developed on Chapter 5 is based on the work of Pitman [18] and Gnedin [4], this chapter is mainly dedicated to study the converse relation between sequences of exchangeable random variables and families of exchangeable partitions, that is to say, we see that every consistent family of exchangeable partitions can be thought as if it was generated by sampling from a sequence of exchangeable random variables, unlike the approach undertaken in previous chapters where we study that a sequence of exchangeable random variables generates a consistent family of exchangeable partitions. Also, in this chapter the connection between random partitions and the normalization of completely random measures (particularly subordinators) is made more explicit.

The following Section 6 provides a sequential construction of random partitions by defining a prediction rule, that is the probability that a certain partition π of $\{1, ..., n, n+1\}$ takes place given that it will be obtained by adding the element $\{n+1\}$ to an already fixed partition π^* of $\{1, 2, ..., n\}$. In this chapter we will also study how to generate a random partition by continuously breaking a stick of original size 1. The developing on this chapter was also based on the work of Pitman [18].

Chapter 7, which is based on the work of Pitman [18], Perman, Pitman and Yor [17], and Hollander and Korwar [7], provides the characterization and study of two very important processes called the *Dirichlet process* and *Poisson-Dirichlet process*, this characterization is mainly made through its parameters. Chapter 7 also provides a branching construction of these models.

The Chapter 8 is the last one of the thesis, here we will introduce a family of partitions called Gibbs partions and prove the fact that by adding a few restrictions to this family of partitions, the restricted family becomes that of the Poisson-Dirichlet model or the Coupon's collector model. To the end of the chapter, we also talk a little about models beyond the families discussed so far. The main material used to construct this chapter was the work of Kerov [10].

By the end of the thesis we discuss, in an additional chapter, the final remarks about what has been discussed so far, summarizing the most relevant results and making a more explicit connection between them.

Chapter 1

Preliminaries. Combinatorial structures

As the main subject that we will be studying throughout the thesis is random partitions (over a finite set). Let us start by stating the definition of a partition, and some other important concepts that arise in the study of them.

Definition 1.1 (Partition)**.** *Let A be a finite set such that $|A| = n$ and $1 \leq k \leq n$. A partition of A into k blocks is an unordered collection of sets $\{A_1, A_2, ..., A_k\}$ such that.*

1. $A_i \subseteq A$ *for all* $i = 1, 2, ..., k$.
2. $A_i \neq \emptyset$ *for all* $i = 1, 2, ..., k$.
3. $A_i \cap A_j = \emptyset$ *for all* $i \neq j$, $i, j \in \{1, 2, ..., k\}$.
4. $\bigcup_{i=1}^{k} A_i = A$.

If we consider $[n] := \{1, 2, ..., n\}$ *and* $1 \leq k \leq n$, *let* $\mathcal{P}_{[n]}^{k}$ *denote the set of all partitions of* $[n]$ *into exactly* k *blocks and define* $\mathcal{P}_{[n]} := \bigcup_{k=1}^{n} \mathcal{P}_{[n]}^{k}$ *the set of all possible partitions of* $[n]$.

Example 1.1. *For the set* $[3] = \{1, 2, 3\}$ *we have that*

- $\mathcal{P}_{[3]}^{1} = \Big\{\{\{1, 2, 3\}\}\Big\}$.
- $\mathcal{P}_{[3]}^{2} = \Big\{\{\{1\}, \{2, 3\}\}, \{\{2\}, \{1, 3\}\}, \{\{3\}, \{1, 2\}\}\Big\}$.
- $\mathcal{P}_{[3]}^{3} = \Big\{\{\{1\}, \{2\}, \{3\}\}\Big\}$

- $\mathcal{P}_{[3]} = \Big\{ \{\{1,2,3\}\}, \{\{1\},\{2,3\}\}, \{\{2\},\{1,3\}\}, \{\{3\},\{1,2\}\}, \{\{1\},\{2\},\{3\}\} \Big\}$.

Definition 1.2. *Let $\pi^* \in \mathcal{P}_{[n]}$, let k be the number of blocks of π^*, $A_1, ..., A_k$ its blocks and $\pi \in \mathcal{P}_{[n+1]}$. We say that π can be obtained from π^* if one of the following conditions holds (renaming the blocks of π if it is necessary):*

i) π has exactly $k+1$ blocks $B_1 = A_1, B_2 = A_2, ..., B_k = A_k, B_{k+1} = \{n+1\}$.

ii) π has exactly k blocks $B_1, B_2, ..., B_k$ such that there exists $j \in \{1,2,...,k\}$ such that $B_j = A_j \cup \{n+1\}$ and $B_i = A_i$ for all $i \neq j$.

Also, define $\mathcal{P}_{n+1}(\pi^*)$ as the set of the partitions of $[n+1]$ that can be obtained from π^*

Remark: Equivalently, we say that $\pi \in \mathcal{P}_{[n+1]}$ can be obtained from $\pi^* \in \mathcal{P}_{[n]}$ if the restriction of π to $[n]$ equals π^*. That is, one can recover π^* through the deletion of the element $n+1$ from π.

Example 1.2. *Consider $\pi^* = \big\{\{1\},\{2,3\}\big\} \in \mathcal{P}_{[3]}$ then*

$$\mathcal{P}_{n+1}(\pi^*) = \Big\{ \{\{1\},\{2,3\},\{4\}\}, \{\{1,4\},\{2,3\}\}, \{\{1\},\{2,3,4\}\} \Big\}$$

Definition 1.3 (Composition)**.** *Let n be a positive integer. A composition of n into $1 \leq k \leq n$ parts is a sequence $\{n_1, n_2, ...n_k\}$ of positive integers such that $\sum_{i=1}^{k} n_k = n$. Let $\mathcal{C}_n^k$ denote the set of all the compositions of n into k parts and define $\mathcal{C}_n := \bigcup_{k=1}^{n} \mathcal{C}_n^k$ the set of all the compositions of n.*

Note the the sequence $\{|A_1|, |A_2|, ..., |A_k|\}$ of sizes of the blocks of a partition $\{A_1, A_2, ..., A_k\}$ of $[n]$ defines a composition of n.

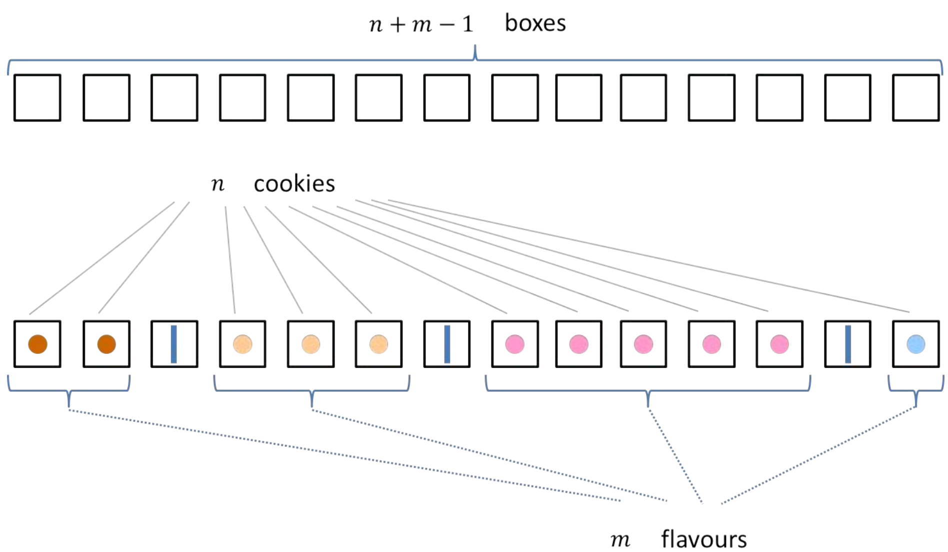
Proposition 1.1. *Let n be a positive integer, then*

i) $\left|\mathcal{C}_n^k\right| = \binom{n-1}{k-1}$ for every $1 \leq k \leq n$.

ii) $|\mathcal{C}_n| = 2^{n-1}$

In order to understand the proof this theorem let us consider the following small problem. Suppose there is a child that heading towards a cookies store, the kid wants to buy n cookies for his mother. When he arrives to the store, the seller tells him that currently there are only m flavours available, let us call them $f_1, f_2, ..., f_m$. How many different ways does the child have to buy his mother cookies?

Note that each collection of n cookies can be represented by $n+m-1$ boxes in which $m-1$ separators have been placed, the boxes before the first separator will represent cookies of flavour f_1, the boxes between the first and second separator will represent cookies of flavour f_2, and so on, as illustrated in the following image.



Thus, to answer the question it suffices to count the number of distinct ways we can place the $m-1$ separators in the n boxes, which is evidently

$$\binom{n+m-1}{m-1}$$

Note that if in two consecutive boxes there is a separator placed, then this means that there will be a flavour that does not appear in the collection of cookies. For instance, if the first and second separators are placed in consecutive boxes, then in the corresponding collection there will be no cookie with flavour f_2.

Proposition 1.2. *Given $k \in \mathbb{N}$ and $n \in \mathbb{N}$ such that $k \leq n$, the number of sequences $(n_1, ..., n_k)$that satisfy $\sum_{j=1}^{k} n_j = n$ subject to the constraints $n_1 \geq a_1$, $n_2 \geq a_2$,...,$n_k \geq a_k$, where $a_1, ..., a_k$ are given integers, is*

$$\binom{n-(a_1+a_2+\cdots+a_k)+k-1}{k-1}$$

Proof:

For $j = 1, 2, ..., k$ define $b_j := n_j - a_j$. Observe that this means $b_j \geq 0$ for all $j = 1, ..., k$, and the condition $\sum_{j=1}^{k} n_j = n$ becomes $\sum_{j=1}^{k} b_j = n - (a_1 + a_2 + \cdots + a_k)$. This new problem is the equivalent to the cookie problem with $n - (a_1 + a_2 + \cdots + a_k)$ cookies to be bought of $m = k$ distinct flavours. Thus the number of sequences $(n_1, ..., n_k)$ that satisfy $\sum_{j=1}^{k} n_j = n$ subject to the constraints $n_1 \geq a_1$, $n_2 \geq a_2$,...,$n_k \geq a_k$ is

$$\binom{n-(a_1+a_2+\cdots+a_k)+k-1}{k-1}$$

$\square$

With this in mind the proof of the Proposition 1.1 becomes trivial.

Proof of Proposition 1.1:

Note that $\left|\mathcal{C}_n^k\right|$ is the number of sequences $(n_1, ..., n_k)$ that satisfy $\sum_{j=1}^k n_j = n$ subject to the constraints $n_1 \geq 1$, $n_2 \geq 1$,...,$n_k \geq 1$, proposition 1.2

$$\left|\mathcal{C}_n^k\right| = \binom{(n-k)+k-1}{k-1} = \binom{n-1}{k-1}$$

this proves i). The second part of the proposition follows from i)

$$|\mathcal{C}_n| = \sum_{k=1}^{n} \left|\mathcal{C}_n^k\right| = \sum_{k=1}^{n} \binom{n-1}{k-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} = (1+1)^{n-1} = 2^{n-1}$$

$\square$

The cardinality of the sets $\mathcal{P}_{[n]}$ and $\mathcal{P}_{[n]}^k$ will be analysed towards the end of this chapter.

Remark. Let $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ such that $|A_i| = n_i$ for every $1 \leq i \leq k$. Let $m_j := \sum_{i=1}^k \mathbf{1}_{(n_i=j)}$ for $1 \leq j \leq n$, that is, m_j counts the number of blocks of π that have exactly j elements. Then the following equalities hold.

$$n = \sum_{j=1}^{k} j m_j, \qquad k = \sum_{j=1}^{n} m_j$$

Before moving on, let us introduce the following notation, same that we will use throughout the whole work.

Factorial powers. For $n = 0, 1, 2, ...$ and arbitrary real numbers x and α, let $(x)_{n\uparrow\alpha}$ denote *the nth factorial power with increment* α, that is

$$(x)_{n\uparrow\alpha} := x(x+\alpha)\cdots(x+(n-1)\alpha) = \prod_{k=0}^{n-1}(x+k\alpha)$$

Note that for $\alpha \neq 0$ the following equality holds.

$$(x)_{n\uparrow\alpha} = \alpha^n \left(\frac{x}{\alpha}\right)_{n\uparrow 1}$$

Similarly let $(x)_{n\downarrow\alpha}$ denote *the nth factorial power with decrement* α, i.e.

$$(x)_{n\downarrow\alpha} := (x)_{n\uparrow -\alpha} = \prod_{k=0}^{n-1}(x-k\alpha).$$

In the case $\alpha = 1$ we are also going to use the notation $(x)_{n\uparrow} := (x)_{n\uparrow 1}$ and $(x)_{n\downarrow} := (x)_{n\downarrow 1}$.

Some useful properties related to the factorial powers defined above, are stated in the following

1) For x a positive integer and $n \leq x$,

$$(x)_{n\downarrow} = x(x-1)\cdots(x-(n-1)),$$

is the number of permutations of length n of a set with x elements.

2) Using the recursion for the gamma function $\Gamma(k+1) = k\Gamma(k)$, we obtain that

$$\begin{aligned}
(x)_{n\uparrow} &= x(x+1)\cdots(x+n-1) \\
&= \frac{x\Gamma(x)(x+1)\cdots(x+n-1)}{\Gamma(x)} \\
&= \frac{(x+1)\Gamma(x+1)(x+2)\cdots(x+n-1)}{\Gamma(x)} \\
&\;\vdots \\
&= \frac{\Gamma(x+n)}{\Gamma(x)}
\end{aligned}$$

3) Using 2) we obtain

$$\begin{aligned}
\lim_{x\to\infty} \frac{\Gamma(x+n)}{\Gamma(x)x^n} &= \lim_{x\to\infty} \frac{x(x+1)\cdots(x+n-1)}{x^n} \\
&= \lim_{x\to\infty} \left[\frac{x}{x}\right]\left[\frac{x+1}{x}\right]\left[\frac{x+n-1}{x}\right] \\
&= 1
\end{aligned}$$

Note that from 3) the following is evident

$$\frac{\Gamma(x+n)}{\Gamma(x)} \sim x^n \text{ as } x \to \infty. \tag{1.1}$$

This last equation, known as *Stirling's approximation for the gamma function* will be very helpful later.

Power series. Let $[x^n]\, f(x)$ be the notation for the coefficient of x^n in $f(x)$. That is, if for example

$$f(x) = \sum_{k=1}^{m} c_k x^k$$

and $1 \leq n \leq m$ then $[x^n]\, f(x) = c_n$. Also note that, e.g.

$$\left[\frac{x^n}{n!}\right] f(x) = n!\, [x^n]\, f(x).$$

1.1 Composite structures

Let $v := (v_1, v_2, ...)$ and $w := (w_1, w_2, ...)$ be two sequences of non-negative integers. Let V be a combinatorial structure, so for each finite set F_n with $|F_n| = n$ there is some construction of a set $V(F_n)$ of V-structures on F_n, such that the number of V-structures on a set of n elements is $|V(F_n)| = v_n$. For instance if v is such that $v_j = j^2$, or $v_j = (j-1)!$, then V might be $F_n \times F_n$ or the cyclic permutations from F_n to F_n. Let W be another combinatorial structure, such that the number of W-structures on a set of j elements is w_j.

Definition 1.4 (Composite structures)**.** *Let $(V \circ W)(F_n)$ denote the composite structure on F_n defined as the set of all ways to partition F_n into blocks $\{A_1, A_2, ..., A_k\}$ for some $1 \le k \le n$, assign this collection of blocks a V-structure, and assign each block A_i a W-structure.*

Evidently, for each set F_n with n elements, the number of such composite structures is

$$|(V \circ W)(F_n)| = B_n(v, w) := \sum_{k=1}^{n} v_k B_{n,k}(w)$$

where

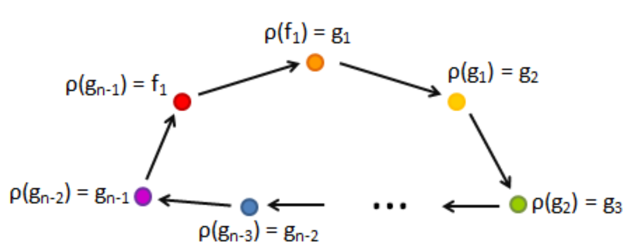
$$B_{n,k}(w) := \sum_{\{A_1,...,A_k\} \in \mathcal{P}^k_{[n]}} \prod_{i=1}^{k} w_{|A_i|}$$

Before providing an example, let us recall the definition of *cyclic permutation*

Definition 1.5 (Cyclic permutation)**.** *Let F_n be a finite set with n elements $f_1, f_2, ..., f_n$, let $\tau : F_n \to F_n$ be a bijection. We say that τ is a cyclic permutation of F_n if*

$$\begin{aligned} \tau(f_1) &= g_1 \in F_n \setminus \{f_1\} \\ \tau(g_i) &= g_{i+1} \in F_n \setminus \{f_1, g_1, ..., g_i\} \quad \textit{for } i = 2, ..., n-2 \\ \tau(g_{n-1}) &= g_n = f_1 \end{aligned}$$

The name of the cyclic permutation comes from the fact that if each element of F_n is assigned to a vertex in a directed graph, in which an arrow goes from f_i to f_j iff $\tau(f_i) = f_j$, then the directed graph forms a cycle as shown in the next figure.



Example 1.3. *Let V be a cyclic permutation and W a permutation (not necessarily cyclic), then clearly $v = (v_1, v_2, ...)$ and $w = (w_1, w_2, ...)$ satisfy $v_n = |V(F_n)| = (n-1)!$ and $w_n = |W(F_n)| = n!$ for each $n \geq 1$, and where F_n is a set with exactly n elements.*

Let us consider the set $[3]$, as seen before the set of the partitions of $[3]$ is

$$\begin{aligned}\mathcal{P}_{[3]} &= \Big\{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\Big\} \\ &= \Big\{\big\{\{1\},\{2\},\{3\}\big\}, \big\{\{1\},\{2,3\}\big\}, \big\{\{2\},\{1,3\}\big\}, \big\{\{3\},\{1,2\}\big\}, \big\{\{1,2,3\}\big\}\Big\}\end{aligned}$$

Let us consider $\pi_1 = \big\{\{1\},\{2\},\{3\}\big\}$ note that this partition has three blocks $A_{1,1} = \{1\}, A_{1,2} = \{2\}$ and $A_{1,3} = \{3\}$. Let $F_3 = \{A_{1,1}, A_{1,2}, A_{1,3}\}$ then the number of cyclic permutations from F_3 to F_3 is $v_3 = (3-1)!$. Now, each block $A_{1,i}$ of π_1 consists in one unique element so for each block there is only one permutation from the block to itself, so there is also only one way ($1 = 1 * 1 * 1 = w_{|A_{1,1}|} * w_{|A_{1,2}|} * w_{|A_{1,3}|}$) to assign each block a permutation. So over all the number of ways to assign π_1 a V-structure and each block of π_1 a W-structure is

$$(3-1)!\prod_{i=1}^{3} w_{|A_{1,i}|} = 2 * 1 * 1 * 1 = 2$$

Now, $\pi_2 = \big\{\{1\},\{2,3\}\big\}$ has 2 blocks $A_{2,1} = \{1\}$ and $A_{2,2} = \{2,3\}$. There is clearly only one cyclic permutation from $\{A_{2,1}, A_{2,2}\}$ to itself ($v_2 = (2-1)! = 1$), for $A_{2,1}$ there is one permutation from $A_{2,1}$ to $A_{2,1}$, and for $A_{2,2}$ which consists in 2 elements there are $w_2 = 2!$ possible ways to assign this block a permutation. Over all, the number of distinct ways to assign π_2 a V-structure and each block of π_2 a W-structure is

$$(2-1)!\prod_{i=1}^{2} w_{|A_{2,i}|} = 1 * 1 * 2 = 2$$

The same happens to π_3 and π_4.

Let us consider the last case. π_5 only has one block $A_{5,1} = \{1,2,3\}$ so the number of cyclic permutations that can be assigned to $\{A_{5,1}\}$ is only 1. The only one block, $A_{5,1}$, has 3 elements, so there are $w_3 = 3!$ different permutations that can be assigned to it. This way $1 * 3! = 6$ is the number of ways to assign π_5 a V-structure and each block of π_5 a W-structure.

From this we can finally conclude that the number of composite structures is

$$\begin{aligned}|(V \circ W)([3])| = B_3(v,w) &= \sum_{k=1}^{3} v_k B_{3,k}(w)\\ &= \sum_{k=1}^{3} v_k \sum_{\{A_1,\dots,A_k\}\in\mathcal{P}^k_{[n]}} \prod_{i=1}^{k} w_{|A_i|}\\ &= (3-1)![1*1*1] + (2-1)![(3)[1*2]] + (1-1)![3!]\\ &= 2+2+2+2+6 = 14.\end{aligned}$$

Note that $B_{n,k}(w)$ is a polynomial in variables $w_1, w_2, \dots$ known as the (n,k)*th partial Bell polynomial*. Let $n \in \mathbb{N}$, $1 \le k \le n$, and $m = (m_1, m_2, \dots, m_n)$ be a sequence of non-negative integers such that

$$\sum_{j=1}^{k} j m_j = n, \text{ and } \sum_{j=1}^{n} m_j = k$$

also let

$$\mathcal{P}_m := \{\pi \in \mathcal{P}^k_{[n]} : \pi \text{ has } m_j \text{ blocks that have exactly } j \text{ elements}\}$$

for the sequence w, consider the coefficient of $\prod_{j=1}^{n} w_j^{m_j}$ in $B_{n,k}(w)$, note that

$$B_{n,k}(w) = \sum_{\{A_1,\dots,A_k\}\in\mathcal{P}^k_{[n]}} \prod_{i=1}^{k} w_i = \sum_{m} \sum_{\pi\in\mathcal{P}_m} \prod_{j=1}^{n} {w_j}^{m_j}$$

so that this coefficient indicates $|\mathcal{P}_m|$. That is to say

$$\left[\prod_{j=1}^{n} {w_j}^{m_j}\right] B_{n,k}(w) = \frac{n!}{\prod_{j=1}^{n} (j!)^{m_j} m_j!}$$

Some Bell partial polynomials are indicated in the next table for $1 \le n \le 5$ and $1 \le k \le n$.

n	$B_{n,1}(w)$	$B_{n,2}(w)$	$B_{n,3}(w)$	$B_{n,4}(w)$	$B_{n,5}(w)$
1	w_1				
2	w_2	w_1^2			
3	w_3	$3w_1w_2$	w_1^3		
4	w_4	$4w_1w_3 + 3w_2^2$	$6w_1^2w_2$	w_1^4	
5	w_5	$5w_1w_4 + 10w_2w_3$	$10w_1^2w_3 + 15w_1w_2^2$	$10w_1^3w_2$	w_1^5

Stirling numbers as we will see below are obtained of as evaluations of $B_{n,k}(w)$ for particular sequences $w = (w_1, w_1, \dots)$.

Useful alternative expressions for $B_{n,k}(w)$ and $B_n(v,w)$ can be given as follows. For each partition π of $[n]$ into k blocks, there are $k!$ different ordered partitions of $[n]$ into k blocks (in terms of the order in which the blocks appear) that correspond to π. Corresponding to each composition $(n_1, ..., n_k)$ of n with k parts, there are

$$\binom{n}{n_1, ..., n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

different ordered partitions, meaning that there are

$$\frac{n!}{k!} \frac{1}{n_1! n_2! \cdots n_k!}$$

partitions (not ordered) corresponding to such composition. So the definition of $B_{n,k}(w)$ as a sum of products over $\mathcal{P}_{[n]}^k$ implies

$$B_{n,k}(w) = \frac{n!}{k!} \sum_{(n_1,...,n_k) \in \mathcal{C}_n^k} \prod_{i=1}^{k} \frac{w_{n_i}}{n_i!}. \tag{1.2}$$

With this in mind, it is natural to introduce the *exponential generating functions* associated with the sequences $w = (w_1, w_2, ...)$ and $v = (v_1, v_2, ...)$:

$$v(\theta) := \sum_{k=1}^{\infty} v_k \frac{\theta^k}{k!}, \text{ and } w(\xi) := \sum_{j=1}^{\infty} w_j \frac{\xi^j}{j!}$$

Then it is easy to see that (1.2) reads

$$B_{n,k}(w) == \left[\frac{\xi^n}{n!}\right] \frac{w(\xi)^k}{k!} \tag{1.3}$$

and also recalling that $B_n(v,w) = \sum_{k=1}^{n} v_k B_{n,k}(w)$,

$$B_n(v,w) = \left[\frac{\xi^n}{n!}\right] v(w(\xi)) \tag{1.4}$$

known as the *compositional formula*.

Remark. Note that (1.3) is a particular case of (1.4), when $v_k = \mathbf{1}_{\{j=k\}}$.

1.2 Stirling numbers and generalized factorial coefficients

Let $\mathbf{1} = (1, 1, ...)$ and define

$$S_{n,k} := B_{n,k}(\mathbf{1}) = \sum_{\{A_1,...,A_k\} \in \mathcal{P}_{[n]}^k} \prod_{i=1}^{k} 1 = \left|\mathcal{P}_{[n]}^k\right|.$$

The numbers $S_{n,k}$ are known as the *Stirling numbers of second kind* and as stated above, they indicate the number of ways to partition the set $[n]$ into k blocks. Also, let us define

$$B_n = \sum_{k=1}^{n} S_{n,k} = \sum_{k=1}^{n} \left|\mathcal{P}_{[n]}^{k}\right| = \left|\mathcal{P}_{[n]}\right|$$

The numbers B_n are known as the *Bell number* and they tell us the number of different ways to partition a set of n distinct elements. By making the corresponding evaluation of the Bell polynomials the next table can be obtained, it shows these numbers for $1 \leq k \leq n \leq 5$

n	$S_{n,1}$	$S_{n,2}$	$S_{n,3}$	$S_{n,4}$	$S_{n,5}$	B_n
1	1					1
2	1	1				2
3	1	3	1			5
4	1	7	6	1		15
5	1	15	25	10	1	52

Observe that the Stirling numbers of second kind are such that

$$x^n = \sum_{k=1}^{n} S_{n,k} (x)_{k\downarrow}.$$

One way to corroborate this is realizing that both sides of the equality are polynomials of degree n. Thus it suffices to check that their value as functions coincide in at least n numbers; in fact, we are going to justify the values are the same in every natural number. Given $a \in \mathbb{N}$, substituting $x = a$ in the identity, both sides count the number of functions of $[n]$ into $[a]$, where the right side does this labour varying the size of the image.

Another numbers closely related to the Bell polynomials are the *Stirling numbers of the first kind* and the *Generalized Stirling numbers.* Let $w = (w_1, w_2, ..)$ where $w_i = (i-1)!$ and let

$$c_{n,k} := B_{n,k}(w) = \sum_{\{A_1,\ldots,A_k\}\in\mathcal{P}_{[n]}^{k}} \prod_{i=1}^{k} (i-1)! = |\{ \text{ permutations of } [n] \text{ with } k \text{ cycles}\}|.$$

Since $(n-1)!$ is the number of cyclic permutations of $[n]$, the last equality corresponds to the representation of a permutation of $[n]$ as the product of cyclic permutations acting on the blocks of some $\pi \in \mathcal{P}_{[n]}^{k}$. The $c_{n,k}$ numbers are known as the *unsigned Stirling numbers of the first kind.*

If now we let $w = (w_1, w_2, ...)$ to be such that $w_i = (-1)^{i-1}(i-1)!$ and let

$$s_{n,k} := B_{n,k}(w)$$

we obtain the so-called *Stirling numbers of the first kind.* Some properties that these numbers hold are stated below

- $s_{n,k} = (-1)^{n-k} c_{n,k}$
- $(x)_{n\uparrow} = \sum_{k=1}^{n} c_{n,k} x^k$
- $(x)_{n\downarrow} = \sum_{k=1}^{n} s_{n,k} x^k$

Generalized Stirling numbers. Let α and β be arbitrary real numbers define the numbers $S_{n,k}^{\alpha,\beta}$ as the connection coefficients of

$$(x)_{n\downarrow\alpha} = \sum_{k=1}^{n} S_{n,k}^{\alpha,\beta} (x)_{k\downarrow\beta}$$

The numbers $S_{n,k}^{\alpha,\beta}$ are known as the *Generalized Stirling numbers*. The name of these numbers comes from the fact that $S_{n,k}^{0,1} = S_{n,k}$ and $S_{n,k}^{1,0} = s_{n,k}$. The generalized Stirling numbers, can also be regarded as

$$S_{n,k}^{\alpha,\beta} = B_{n,k}(w)$$

with $w = (w_1, w_2, ...)$ such that $w_i = (\beta - \alpha)_{i-1\downarrow\alpha}$. The justification of this last assertions as well as other properties about these numbers may be found, for instance, in the work of Toscano [20], Riordan [19], Charalambides and Singh [1], Hsu and Shiue [8].

Generalized factorial coefficients. Let $n \geq 0$ and $k = 1, 2, ..., n$, we define the *generalized factorial coefficient*, $\mathcal{C}_{n,k}(\sigma)$, and the *non-central generalized factorial coefficients* $\mathcal{C}_{n,k}(\sigma, \gamma)$ as the connection coefficients of

$$(\sigma x)_{n\uparrow} = \sum_{k=0}^{n} \mathcal{C}_{n,k}(\sigma)(x)_{k\uparrow}$$

and

$$(\sigma x - \gamma)_{n\uparrow} = \sum_{k=0}^{n} \mathcal{C}_{n,k}(\sigma, \gamma)(x)_{k\uparrow}$$

respectively and together with the provision $\mathcal{C}_{0,0}(\sigma) = 1$ and $\mathcal{C}_{n,0}(\sigma) = 0$ for all $n \geq 1$.

In the work of Charalambides and Singh [1] and Charalambides [2] there are provided some helpful representations of the above defined numbers:

$$\mathcal{C}_{n,k}(\sigma) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\sigma)_{n\uparrow}$$

as well as

$$\mathcal{C}_{n,k}(\sigma, \gamma) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\sigma - \gamma)_{n\uparrow}$$

Moreover, it is possible to relate the non-central and central generalized factorial coefficients by the following equation

$$\mathcal{C}_{n,k}(\sigma,\gamma) = \sum_{s=k}^{n} \binom{n}{s} \mathcal{C}_{s,k}(\sigma)(-\gamma)_{n-s\uparrow}$$

Another concerning property of the central and non-central generalized factorial coefficients is that

$$\lim_{\sigma\to 0} \frac{\mathcal{C}_{s,k}(\sigma)}{\sigma^k} = c_{n,k}$$

where $c_{n,k}$ is a Stirling numbre of the first kind. Thus

$$\lim_{\sigma\to 0} \frac{\mathcal{C}_{s,k}(\sigma,\gamma)}{\sigma^k} = \sum_{s=k}^{n} \binom{n}{s} c_{n,k}(-\gamma)_{n-s\uparrow}$$

Chapter 2

Random partitions. Basic concepts

In the past chapter we study concepts that will aid us in the analysis of random partitions. Up to this point everything has been deterministic, so now let us start treating the partitions as random objects. We start with the definition of a random partition.

Definition 2.1 (Random Partition)**.** *A random partition Π_n is a random variable that takes values in the set of all the partitions of $[n]$, $\mathcal{P}_{[n]}$.*

Definition 2.2. *We say that $\{\Pi_n\}_{n\geq 1}$ is a family of random partitions if for each $n \geq 1$ Π_n is a $\mathcal{P}_{[n]}$-valued random variable.*

Note that in the definition of a random partition there are other random variables implicitly defined, one of them is K_n which refers to the number of blocks in the random partition Π_n. It is clear that

$$(\Pi_n = \pi) \cap (K_n = k) = \begin{cases} (\Pi_n = \pi) & \text{if the number of blocks of } \pi \text{ is exactly } k \leq n \\ \emptyset & \text{otherwise.} \end{cases}$$

Hence, if $\mathbb{P}[\Pi_n = \pi]$ is known for each $\pi \in \mathcal{P}_{[n]}$ then we can obtain the marginal distribution of K_n as follows

$$\mathbb{P}[K_n = k] = \sum_{\pi \in \mathcal{P}_{[n]}^k} \mathbb{P}[\Pi_n = \pi, K_n = k] = \sum_{\pi \in \mathcal{P}_{[n]}^k} \mathbb{P}[\Pi_n = \pi].$$

Other random variables implicitly defined in a random partition are $N_{n,1}, ...N_{n,K_n}$ which model the number of elements in each of the blocks of the random partition. It is clear that these random variables must satisfy $N_{n,1} + ... + N_{n,K_n} = n$, in other words $\{N_{n,1}, ...N_{n,K_n}\}$ forms a random composition of n. As done before, note that

$$(\Pi_n = \pi) \cap (K_n = k) \cap \left[\bigcap_{i=1}^{k} (N_{n,i} = n_i)\right] = (\Pi_n = \pi)$$

if $\pi = \{A_1, ..., A_k\}$ where $|A_i| = n_i$, $i = 1, ..., k$, and

$$(\Pi_n = \pi) \cap (K_n = k) \cap \left[\bigcap_{i=1}^{k} (N_{n,i} = n_i)\right] = \emptyset$$

otherwise.

In the following sections two important properties of random partitions are studied, the *projectivity* and the *exchangeability* properties.

2.1 Projective distributions and consistent families of random partitions

The projectivity property corresponds to Kolmogorov's consistency theorem, and is stated in the following definition.

Definition 2.3 (Consistency in distribution)**.** *Let $\{\Pi_n\}_{n\geq 1}$ be a family of random partitions and $f_n(\cdot) := \mathbb{P}[\Pi_n = \cdot]$ be the probability distribution of the random partition Π_n taking values on $\mathcal{P}_{[n]}$. Then the sequence of distributions $\{f_j\}_{j\geq 1}$ is projective if*

$$f_n(\pi^*) = \sum_{\pi \in \mathcal{P}_{n+1}(\pi^*)} f_{n+1}(\pi) \tag{2.1}$$

If $\{f_j\}_{j\geq 1}$ is projective we say that the family of random partitions $\{\Pi_n\}_{n\geq 1}$ is consistent in distribution.

Recall that $\mathcal{P}_{n+1}(\pi^*)$ is the set of all partitions π of $[n+1]$ that can be obtained from π^*, see Definition 1.2.

Definition 2.4 (Consistency)**.** *Let $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ be a family of random partitions such that for every $n \in \mathbb{N}$, Π_n is a partition of $[n]$ we say that the family of random partitions is consistent if for $m < n$, the restriction of Π_n to $[m]$ is Π_m almost surely.*

For instance, if for some family of consistent random partitions it is known that

$$\Pi_6 = \big\{\{1,3,6\},\{2,4\},\{5\}\big\}$$

then conditionally on the event above, we get that

$$\begin{aligned}
\Pi_5 &= \big\{\{1,3\},\{2,4\},\{5\}\big\} \\
\Pi_4 &= \big\{\{1,3\},\{2,4\}\big\} \\
\Pi_3 &= \big\{\{1,3\},\{2\}\big\} \\
\Pi_2 &= \big\{\{1\},\{2\}\big\} \\
\Pi_1 &= \big\{\{1\}\big\}
\end{aligned}$$

almost surely.

The sequence Π_∞, as in the above definition, can be regarded as a random element of the set $\mathcal{P}_{\mathbb{N}}$ of partitions of $\mathbb{N}$, such that Π_n is its restriction to $[n]$.

It is easy to see that every consistent family of random partitions $\{\Pi_n\}_{n\geq 1}$ is also consistent in distribution as

$$\begin{aligned}\mathbb{P}[\Pi_n = \pi^*] &= \mathbb{P}[\Pi_n = \pi^*, \Pi_{n+1} \in \mathcal{P}_{n+1}(\pi^*)] \\ &= \sum_{\pi\in\mathcal{P}_{n+1}(\pi^*)} \mathbb{P}[\Pi_n = \pi^*, \Pi_{n+1} = \pi] \\ &= \sum_{\pi\in\mathcal{P}_{n+1}(\pi^*)} \mathbb{P}[\Pi_{n+1} = \pi]\end{aligned}$$

The converse is clearly false. For instance consider any consistent family of random partitions $\{\Pi_n\}_{n\geq 1}$ and let $\{\widehat{\Pi}_n\}_{n\geq 1}$ be a family of *independent* random partitions such that for every $n \geq 1$ and for every $\pi \in \mathcal{P}_{[n]}$

$$\mathbb{P}[\Pi_n = \pi] = \mathbb{P}[\widehat{\Pi}_n = \pi].$$

As $\{\Pi_n\}_{n\geq 1}$ is consistent, it is consistent in distribution, which means that $\{\widehat{\Pi}_n\}_{n\geq 1}$ is also consistent in distribution. But $\{\widehat{\Pi}_n\}_{n\geq 1}$ is a family of independent random objects, so for every $n \geq 1$, $\pi \in \mathcal{P}_{[n]}$ and $\pi^* \in \mathcal{P}_{[n+1]}$ we get

$$\mathbb{P}[\widehat{\Pi}_n = \pi | \widehat{\Pi}_n = \pi^*] = \mathbb{P}[\widehat{\Pi}_n = \pi]$$

which is not necessarily equal to 1. Thus $\{\widehat{\Pi}_n\}_{n\geq 1}$ is not necessarily consistent.

2.2 Exchangeable random partitions

Let $\rho : [n] \to [n]$ be a permutation of $[n]$, $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$, $n_i = |A_i|$ the size of the ith element of π for every $i = 1, ..., k$ and $a_{i,1}, ..., a_{i,n_i}$ the elements in A_i. Define $\rho^* : \mathcal{P}^k_{[n]} \to \mathcal{P}^k_{[n]}$ as follows $\rho^*(\{A_1, ..., A_k\}) = \{A^*_1, ..., A^*_k\}$ where for every $i = 1, ..., k$ $A^*_i = \{\rho(a_{i,1}), ..., \rho(a_{i,n_i})\}$, it is clear that $|A^*_i| = n_i$. Continuing with this notation let us state the following definition

Definition 2.5. *A random partition Π_n is exchangeable if, for any partition $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ and for every permutation ρ of $[n]$,*

$$\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}] = \mathbb{P}[\Pi_n = \rho^*(\{A_1, ..., A_k\})]$$

Equivalently, if $n_i = |A_i|$, $i = 1, ..., k$

$$\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}] =: p_n(n_1, ..., n_k) \tag{2.2}$$

for some symmetric function p_n of its arguments. Furthermore, the function p_n is called an exchangeable partition probability function (EPPF).

Remark: Note that the equivalence mentioned in the above definition follows from the fact $\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}]$ does not depend in which element belongs to which block, hence, at most it depends on the number of blocks, k, and the number of elements in each block, n_i, $i = 1, 2, ..., k$. The symmetry part comes from the fact that $\{A_1, ..., A_k\} = \{A_{\hat{\rho}(1)}, ..., A_{\hat{\rho}(k)}\}$ for every permutation $\hat{\rho} : [k] \to [k]$.

When working with families of exchangeable random partitions it is easy to see that the projectivity property (2.1) is satisfied if the *addition rule* holds for the EPPF:

$$p_n(n_1, ...n_k) = p_{n+1}(n_1, ..., n_k, 1) + \sum_{j=1}^{k} p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k) \tag{2.3}$$

It might be of interest to study the relation between families of random partitions which are consistent in distribution and families of exchangeable random partitions. With this in mind, the next couple of examples show that not all the families of exchangeable random partitions posses the projectivity property and not every family of consistent (in distribution) random partitions is exchangeable.

Example 2.1. *Let $\{\Pi_n\}_{n\geq 1}$ be a family of random partitions such that Π_n is a random partition uniformly distributed on $\mathcal{P}_{[n]}$, i.e. for every $\pi \in \mathcal{P}_{[n]}$ $\mathbb{P}[\Pi_n = \pi] = \frac{1}{B_n}$ where $B_n = |\mathcal{P}_{[n]}|$ is the nth Bell number. Clearly for every $n \geq 1$ Π_n is exchangeable as*

$$\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}] = p_n(n_1, ..., n_k) = \frac{1}{B_n}$$

To see that the family is not projective it suffices to check that the addition rule is not satisfied. Consider $n = 2$, $k = 2$ and $(n_1, n_2) = (1, 1)$, then

$$p_2(1, 1) = \frac{1}{B_2} = \frac{1}{2},$$

on the other side

$$p_3(1, 1, 1) = p_3(1, 2) = p_3(2, 1) = \frac{1}{B_3} = \frac{1}{5}$$

so

$$p_2(1, 1) = \frac{1}{2} \neq \frac{3}{5} = p_3(1, 1, 1) + 2p_3(1, 2).$$

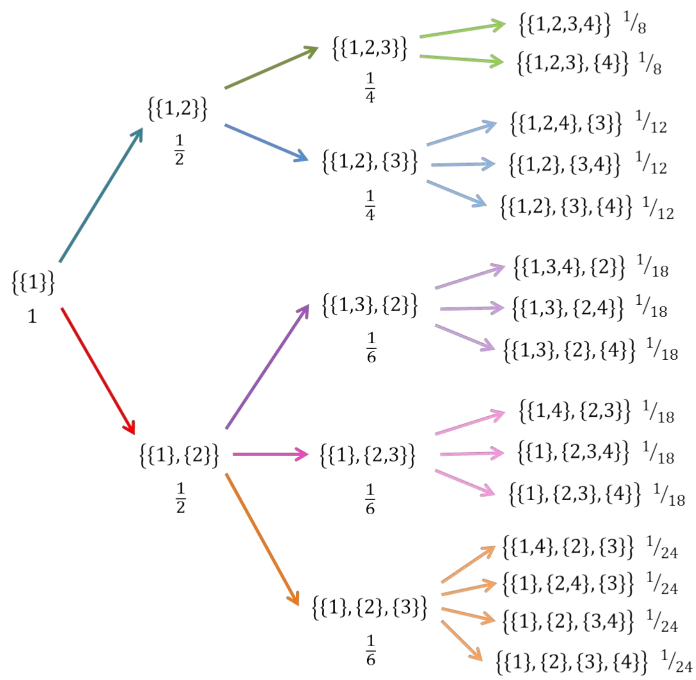
This proves that the addition rule does not hold for this family, and so it is not consistent in distribution.

Example 2.2. *Let $\{\Pi_n\}_{n\geq 1}$ be a family of random partitions such that for every $n \geq 1$, for every $\pi^* \in \mathcal{P}_{[n]}$ and for every $\pi \in P_{n+1}(\pi^*)$ we have that*

$$f_{n+1}(\pi) = \mathbb{P}[\Pi_{n+1} = \pi] = \frac{f_n(\pi^*)}{|\mathcal{P}_{n+1}(\pi^*)|} = \frac{\mathbb{P}[\Pi_n = \pi^*]}{|\mathcal{P}_{n+1}(\pi^*)|}$$

and clearly $f_1(\{\{1\}\}) = 1$. *Note that these recursive probability functions are well defined as for every* $\pi \in \mathcal{P}_{[n+1]}$ *there exist one unique* $\pi^* \in \mathcal{P}_{[n]}$ *such that* $\pi \in \mathcal{P}_{n+1}(\pi^*)$.

The next image ilustrates this for $n = 1, 2, 3, 4$.



Every vertex of the tree is a random partition, if there is an arrow from π^* *to* π *then this means that* $\pi \in \mathcal{P}_{n+1}(\pi^*)$. *The number that is either next or below the partition* π *refers to* $f_n(\pi)$.

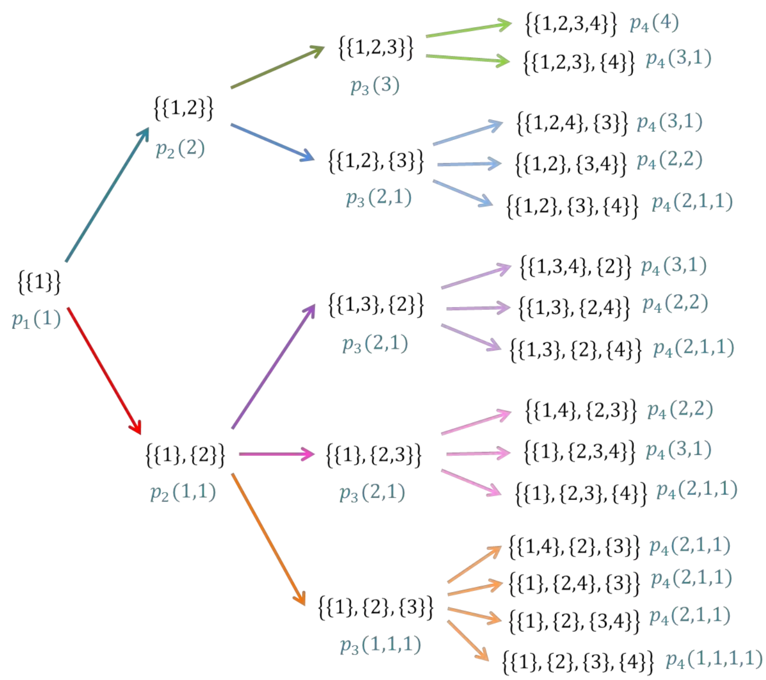
It is very easy to see that this family of random partitions is consistent as for every $\pi^* \in \mathcal{P}_{[n]}$

$$\sum_{\pi \in \mathcal{P}_{n+1}(\pi^*)} f_{n+1}(\pi) = \sum_{\pi \in \mathcal{P}_{n+1}(\pi^*)} \frac{f_n(\pi^*)}{|\mathcal{P}_{n+1}(\pi^*)|} = \frac{f_n(\pi^*)}{|\mathcal{P}_{n+1}(\pi^*)|} |\mathcal{P}_{n+1}(\pi^*)| = f_n(\pi^*)$$

It is also easy to see that this family is not exchangeable. Consider $\{\{1,2\},\{3\}\}$ *and* $\{\{1,3\},\{2\}\}$, *the number of blocks on both partition is two, also both partitions have one block with two elements and other one with only one element. If the family was exchangeable, in particular this two partitions would have the* *same probability of occurring, instead*

$$\mathbb{P}[\Pi_3 = \{\{1,2\},\{3\}\}] = \frac{1}{4} \neq \frac{1}{6} = \mathbb{P}[\Pi_3 = \{\{1,3\},\{2\}\}].$$

In general a consistent (in distribution) family of exchangeable random partitions would look as shown in the next figure (for $n = 1, 2, 3, 4$).



Every vertex of the tree is a random partition, if there is an arrow from π^* to π then this means that $\pi \in \mathcal{P}_{n+1}(\pi^*)$. The number that is either next or below the partition π refers to $f_n(\pi)$ where $\{f_n\}_{n\geq 1}$ is projective.

Remark: If the family of random partitions illustrated above was consistent (not only consistent in distribution) we could assing a probability to each arrow as follows. If an arrow goes from π^* to π then the probability assigned to this arrow would be the one of going from π^* to π that is

$$\mathbb{P}[\Pi_{n+1} = \pi | \Pi_n = \pi^*] = \frac{\Pi_{n+1} = \pi}{\Pi_n = \pi^*} = \frac{f_{n+1}(\pi)}{f_n(\pi^*)}$$

When working with a consistent family of exchangeable random partitions $\{\Pi_n\}_{n\geq 1}$ it is easy to derive a *prediction rule*. Let $\pi = \{A_1, ...A_k\} \in \mathcal{P}_{[n]}$, with $n_j = |A_j|$ and let

$$\pi^{(j)} = \{A_1, ..., A_{j-1}, A_j \cup \{n+1\}, A_{j+1}, ..., A_k\} \quad j = 1, 2, ..., k$$
$$\pi^{(0)} = \{A_1, ..., A_k, \{n+1\}\}$$

then as π is the only partition on $\mathcal{P}_{[n]}$ such that $\pi^{(i)}$ can be obtained from π for $i = 0, 1, ..., k$,

$$\begin{aligned}\mathbb{P}[\Pi_{n+1} = \pi^{(0)} | \Pi_n = \pi] &= \frac{\mathbb{P}[\Pi_{n+1} = \pi^{(0)}, \Pi_n = \pi]}{\mathbb{P}[\Pi_n = \pi]} \\ &= \frac{\mathbb{P}[\Pi_{n+1} = \pi^{(0)}]}{\mathbb{P}[\Pi_n = \pi]} \\ &= \frac{p_{n+1}(n_1, ..., n_k, 1)}{p_n(n_1, ...n_k)}\end{aligned}$$

and analogously

$$\begin{aligned}\mathbb{P}[\Pi_{n+1} = \pi^{(j)} | \Pi_n = \pi] &= \frac{\mathbb{P}[\Pi_{n+1} = \pi^{(j)}, \Pi_n = \pi]}{\mathbb{P}[\Pi_n = \pi]} \\ &= \frac{\mathbb{P}[\Pi_{n+1} = \pi^{(j)}]}{\mathbb{P}[\Pi_n = \pi]} \\ &= \frac{p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k)}{p_n(n_1, ...n_k)}\end{aligned}$$

for $j = 1, 2, ..., k$.

The consistent families of exchangeable random partitions will be the main topic along the document.

2.3 Exchangeable random variables and exchangeable random partitions

Definition 2.6 (Exchangeable Random Variables). *We say that $\{X_i\}_{i=1}^n$ is a finite sequence of exchangeable random variables iff*

$$(X_1, X_2, ..., X_n) \stackrel{d}{=} (X_{\rho(1)}, X_{\rho(2)}, ..., X_{\rho(n)})$$

where $\rho : [n] \to [n]$ is any permutation of $[n]$. Analogously, we say that $\{X_i\}_{i=1}^\infty$ is an (infinite) sequence of random variables iff for all $A \subset \{1, 2, ..., n\}$, with $|A| < \infty$ we have that $\{X_i\}_{i \in A}$ is a finite sequence of exchangeable random variables.

Theorem 2.1 (de Finetti's representation Theorem). *Let $\{X_i\}_{i=1}^\infty$ be a sequence of exchangeable random variables taking values on a Polish space $\mathbb{X}$ endowed with its Borel σ-field $\mathcal{X}$, also let $\mathcal{P}_\mathbb{X}$ be the space of all probability measures over $\mathbb{X}$, then*

1. *For every $n \geq 1$, and $A_1, A_2, ..., A_n \in \mathcal{X}$*

$$\mathbb{P}[X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n] = \int_{\mathcal{P}_\mathbb{X}} \prod_{i=1}^n \tilde{p}[A_i] Q(d\tilde{p})$$

where $\tilde{p}$ is known as the directing random probability measure and its distribution, Q, is known as the de Finetti's measure.

2. *For every $A \in \mathcal{X}$, and as n goes to infinity, the empiric distribution satisfies*

$$F_n(A) := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(A) \to \tilde{p}(A)\, a.s. \tag{2.4}$$

where $\tilde{p} \sim Q$.

Remark: Note that the converse is also true, in the sense that if $\{X_i\}_{i\geq 1}$ is a sequence of random variable such that 1. of the last theorem holds, then $\{X_i\}_{i\geq 1}$ is a sequence of exchangeable random variables.

Another way of looking at the first part of Bruno de Finetti's representation theorem is that there exists a probability measure Q over $\mathcal{P}_{\mathbb{X}}$ such that, $X_1, X_2, ...$ are conditionally independent random variables given the random probability measure $\tilde{p}$ where Q is the distribution of $\tilde{p}$, i.e.

$$X_i|\tilde{p} \overset{i.i.d.}{\sim} \tilde{p}, \quad i = 1, 2, ... \tag{2.5}$$

$$\tilde{p} \sim Q. \tag{2.6}$$

Now, imagine that somehow we have managed to sample $X_1, ..., X_n$, it can be of interest to make a prediction for the value of X_{n+1}, we can do this as follows,

$$\mathbb{P}[X_{n+1} \in \bullet|X_1, ..., X_n] = \int_{\mathcal{P}_{\mathbb{X}}} \tilde{p}(\bullet)Q(d\tilde{p}|X_1, ..., X_n) = \mathbb{E}_{Q(*)|X_1,...,X_n}[\tilde{p}(\bullet)] \tag{2.7}$$

From the last equation it can be stated that the (one-step ahead) predictive distribution coincides with the posterior expected value of $\tilde{p}$.

For now assume $\tilde{p}$ is discrete a.s. then it is known that $\tilde{p}$ can be represented as

$$\tilde{p} = \sum_{j=1}^{\infty}\tilde{p}_j\delta_{Z_j}$$

where $\{Z_j\}_{j\geq 1}$ are $\mathbb{X}$-valued random variables from a non-atomic distribution P_0 (so $\mathbb{P}[Z_j = Z_i] = 0$ for $i \neq j$) and $\{\tilde{p}_j\}_{j\geq 1}$ are non-negative random variables such that $\sum \tilde{p}_j = 1$ almost surely, and δ_Z stands for the unit point mass at Z

$$\delta_Z(\bullet) = \begin{cases} 1 \text{ if } Z \in \bullet \\ 0 \text{ otherwise.} \end{cases}$$

That is, if X is an $\mathbb{X}$-valued random variable such that $X|\tilde{p} \sim \tilde{p}$ then

$$X = \begin{cases} Z_1 & \text{with probability } \tilde{p}_1 \\ Z_2 & \text{with probability } \tilde{p}_2 \\ & \vdots \\ Z_k & \text{with probability } \tilde{p}_k \\ & \vdots \end{cases}$$

Now, the discrete nature of Q implies that any sample $X_1, ..., X_n$, such that (2.5) holds, will feature ties with positive probability, i.e. $\mathbb{P}[X_i = X_j] > 0$, generating $K_n = k \leq n$ different values $X_1^*, ..., X_k^*$ with frequencies $N_{n,1} = n_1, ..., N_{n,k} = n_k$ such that $\sum_{i=1}^k n_i = n$. So, given the discreteness of Q, $\tilde{p}$ induces a partition Π_n of $[n]$ in such way that i and j belong to the same block of Π_n iff $X_i = X_j = X_r^*$ for some $r \in \{1, 2, ..., k\}$. As $X_1, ..., X_n$ are exchangeable then

$$\mathbb{P}[X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = \mathbb{P}[X_{\rho(1)} = x_1, X_{\rho(2)} = x_2, ..., X_{\rho(n)} = x_n]$$

for any permutation of $[n]$, so the distribution of Π_n will only depend on the number of clusters K_n and the frequencies of the blocks $N_{n,1}, ..., N_{n,K_n}$, this is, if $\pi = \{A_1, ..., A_k\}$ is a partition of $[n]$ and $\tilde{\pi} = \{\tilde{A}_1, ..., \tilde{A}_k\}$ is another partition such that $|A_j| = |\tilde{A}_j|$ for every $1 \leq j \leq k$ then $\mathbb{P}[\Pi_n = \pi] = \mathbb{P}[\Pi_n = \tilde{\pi}]$ (in other words the distribution of Π_n does not depend on which elements belong to each block, it only depends on the number of blocks and the number of elements in each one). This way, we have that

$$\mathbb{P}[\Pi_n = \pi] = \mathbb{P}[K_n = k, N_{n,1} = n_1, ..., N_{n,k} = n_k] = p_n(n_1, ..., n_k)$$

for some symmetric function $p_n(n_1, ..., n_k)$ of $n_1, ...n_k$. So it is clear that exchangeable random variables not only induce a random partition, but the random partition they induce turns out to be exchangeable and its EPPF is given by

$$p_n(n_1, ..., n_k) = \int_{\mathbb{X}^k} \mathbb{E}_Q[\tilde{p}^{n_1}(dx_1)...\tilde{p}^{n_k}(dx_k)]$$

Given the sample $X_1, ..., X_n$, if we sample one more value X_{n+1} then the updated sample $X_1, ..., X_n, X_{n+1}$ will generate a random partition Π_{n+1} of $[n+1]$ as explained before. Thus, the sampling procedure generates a family $\{\Pi_n\}_{n\geq 1}$ of exchangeable random partitions. One might wonder if this family is consistent, the answer is yes and it is very easy and intuitive to verify this affirmation. Note that given the n-sized sample $X_1, ..., X_n$, the next observed value X_{n+1} will necessarily be such that $X_{n+1} \in \{X_1^*, ..., X_k^*\}$ or X_{n+1} takes a value not yet observed in the original n-sized sample.

In the first case there must exist $j \in [k]$ such that $X_{n+1} = X_j^*$ so the frequencies will now be

$$\{N_{n+1,1}, ..., N_{n+1,j-1}, N_{n+1,j}, N_{n+1,j+1}, ..., N_{n+1,k}\}$$
$$= \{N_{n,1}, ..., N_{n,j-1}, N_{n,j} + 1, N_{j+1}, ..., N_k\}$$

and the number of clusters in the $n+1$-sized sample will remain the same $K_{n+1} = K_n$. On the second case the number of clusters will now be $K_{n+1} = K_n + 1$ and the frequencies

$$\{N_{n+1,1}, ..., N_{n+1,k}, N_{n+1,k+1}\} = \{N_{n,1}, ..., N_{n,k}, 1\}.$$

From this it is clear that the addition rule holds for the corresponding EPPF:

$$p_n(n_1, ...n_k) = p_{n+1}(n_1, ..., n_k, 1) + \sum_{j=1}^{k} p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k).$$

So summarising, given a sequence of exchangeable random variables $\{X_i\}_{i=1}^{\infty}$ such that (2.5) holds where $\tilde{p}$ is a.s. discrete we can construct a consistent family of exchangeable random partitions $\{\Pi_n\}_{n\geq 1}$ by sampling in such way that if $X^{(n)} = \{X_1, ..., X_n\}$ is the n-sized sample then the $n+1$-sized sample will be such that $X^{(n+1)} = X^{(n)} \cup \{X_{n+1}\}$.

Chapter 3

Completely random measures

Let us start this chapter with the definition of a completely random measure.

Definition 3.1 (Completely random measures)**.** *Let $\mathbb{X}$ be a polish space with corresponding σ-field $\mathcal{X}$. Let $\mathbb{M}$ be the space of measures on $(\mathbb{X}, \mathcal{X})$ with corresponding σ-field $\mathcal{M}$. Let Φ be a random measurable mapping taking values in $(\mathbb{M}, \mathcal{M})$ such that for every $A_1, ..., A_n \in \mathcal{X}$ with $A_i \cap A_j = \emptyset$ whenever $i \neq j$, the random variables $\Phi(A_1), ..., \Phi(A_n)$ are independent. Then Φ is called a completely random measure (CRM).*

Definition 3.2 (The Poisson Process)**.** *Let Ξ be a random countable subset of $\mathbb{X}$. Define the random variable*

$$N(A) = |\Xi \cap A| \quad \forall A \in \mathcal{X}$$

We say that Ξ is a Poisson Process on $\mathbb{X}$ if

1. *For any disjoint sets $A_1, ..., A_n \in \mathcal{X}$ the random variables $N(A_1), ..., N(A_n)$ are independent, and*
2. *$N(A) \sim Poisson(\mu(A))$ where μ is a measure on $(\mathbb{X}, \mathcal{X})$, μ is called the mean measure.*

It is easy to see that

i) $N(\emptyset) = |\Xi \cap \emptyset| = |\emptyset| = 0$, and

ii) If $\{A_i\}_{i=1}^{\infty}$ is a collection of measurable disjoint subsets of $\mathbb{X}$, then

$$
\begin{aligned}
N\left(\bigcup_{i=1}^{\infty} A_i\right) &= \left|\Xi \cap \bigcup_{i=1}^{\infty} A_i\right| \\
&= \left|\bigcup_{i=1}^{\infty} \Xi \cap A_i\right| \\
&= \sum_{i=1}^{\infty} |\Xi \cap A_i| \\
&= \sum_{i=1}^{\infty} N(A_i)
\end{aligned}
$$

so N is in fact a random measure, from the definition of Poisson Process we can see N is not only a random measure, but it is also a CRM

Theorem 3.1 (Campbell's theorem)**.** *Let Ξ be a Poisson process on* $(\mathbb{X}, \mathcal{X})$ *with mean measure μ, let $f : \mathbb{X} \to \mathbb{R}$ be a measurable function, and*

$$\Upsilon = \sum_{X \in \Xi} f(X)$$

if we also assume that

$$\int_{\mathbb{X}} min\{|f(x)|, 1\}\mu(dx) < \infty.$$

then for $t \in \mathbb{R}$

$$\mathbb{E}[e^{t\Upsilon}] = \exp\left\{-\int_{\mathbb{X}} \left(1 - e^{tf(x)}\right)\mu(dx)\right\}$$

Proof:

Recall that if $Z \sim Poisson(\lambda)$ then its moment generator function is given by:

$$M_z(t) = \mathbb{E}\left[e^{tZ}\right] = \exp\{(e^t - 1)\lambda\} \quad \forall t \in \mathbb{R}.$$

Let $N(A) = |\Xi \cap A|$ as in the definition of a Poisson Process. First, we consider the case where f is a non-negative simple function, that is

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

where $A_1, ..., A_n \in \mathcal{X}$ and $a_1, ..., a_n \in \mathbb{R}^+$, we can assume without loss of generality that this is the canonical form of f, that is $A_1, ..., A_n$ are disjoint. Considering the fact that $A_1, ..., A_n$ are disjoint and so $N(A_1), ..., N(A_n)$ are independent random variables such that $N(A_i) \sim Poisson(\mu(A_i))$, we have that

$$
\begin{aligned}
\mathbb{E}\left[e^{t\Upsilon}\right] &= \mathbb{E}\left[\exp\left\{t\sum_{X\in\Xi}\sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(X)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{t\sum_{i=1}^{n} a_i \sum_{X\in\Xi} \mathbf{1}_{A_i}(X)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{t\sum_{i=1}^{n} a_i N(A_i)\right\}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[e^{ta_i N(A_i)}\right] \\
&= \prod_{i=1}^{n} \exp\left\{(e^{ta_i}-1)\mu(A_i)\right\} \\
&= \exp\left\{-\sum_{i=1}^{n}\int_{\mathbb{X}} (1-e^{ta_i})\mathbf{1}_{A_i}(x)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{t\sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf(x)}\right)\mu(dx)\right\}
\end{aligned}
$$

Now assume f is a non-negative function. Then there is a sequence $\{f_n\}_{n=1}^{\infty}$ of simple functions such that $0 \leq f_1 \leq f_2 \leq ...$ and $\lim_{n\to\infty} f_n = f$. Define

$$\Upsilon_n = \sum_{X\in\Xi} f_n(X)$$

By monotone convergence lemma (when $t > 0$) or lebesgue dominated convergence (when $t < 0$)

$$
\begin{aligned}
\mathbb{E}\left[e^{t\Upsilon}\right] &= \mathbb{E}\left[e^{t\lim_{n\to\infty}\Upsilon_n}\right] \\
&= \lim_{n\to\infty}\mathbb{E}\left[e^{t\Upsilon_n}\right] \\
&= \lim_{n\to\infty}\exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf_n(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\lim_{n\to\infty}\int_{\mathbb{X}}\left(1-e^{tf_n(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{t\lim_{n\to\infty}f_n(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf(x)}\right)\mu(dx)\right\}
\end{aligned}
$$

Note that if $\int_{\mathbb{X}} min\{|f(x)|,1\}\mu(dx) < \infty$ does not hold then for $t < 0$ the last integral diverges so that $\mathbb{E}[e^{t\Upsilon}] = 0$ showing that $\Upsilon = \infty$ with probability 1.

Finally assume that f is any measurable function consider

$$f^+(\bullet) = max\{f(\bullet), 0\}, \quad f^- = -min\{f(\bullet), 0\}.$$

Then f^+ and f^- are non-negative functions such that $f = f^+ - f^-$, Let

$$\Upsilon^+ = \sum_{X\in\Xi} f^+(X) = \sum_{X\in\Xi^+} f^+(X),$$

and

$$\Upsilon^- = \sum_{X\in\Xi} f^-(X) = \sum_{X\in\Xi^-} f^-(X),$$

where $\Xi^+ = \{X \in \Xi : f(X) > 0\}$ and $\Xi^- = \{X \in \Xi : f(X) < 0\}$. As Ξ^+ and

Ξ^- are disjoint then Υ^+ and Υ^- are independent, and so

$$\begin{aligned}
\mathbb{E}\left[e^{t\Upsilon}\right] &= \mathbb{E}\left[e^{t\Upsilon^+ - t\Upsilon^-}\right] \\
&= \mathbb{E}\left[e^{t\Upsilon^+}\right]\mathbb{E}\left[e^{-t\Upsilon^-}\right] \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf^+(x)}\right)\mu(dx)\right\}\exp\left\{-\int_{\mathbb{X}}\left(1-e^{-tf^-(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf^+(x)}\right)\mu(dx)-\int_{\mathbb{X}}\left(1-e^{-tf^-(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\{x:f(x)>0\}}\left(1-e^{tf^+(x)}\right)\mu(dx)-\int_{\{x:f(x)<0\}}\left(1-e^{-tf^-(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\{x:f(x)>0\}}\left(1-e^{tf(x)}\right)\mu(dx)-\int_{\{x:f(x)<0\}}\left(1-e^{tf(x)}\right)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}}\left(1-e^{tf(x)}\right)\mu(dx)\right\}
\end{aligned}$$

this finishes the proof of the theorem.

If we set $f \geq 0$ and $t = -1$ in Campbell's theorem we get the so-called *characteristic funtional*

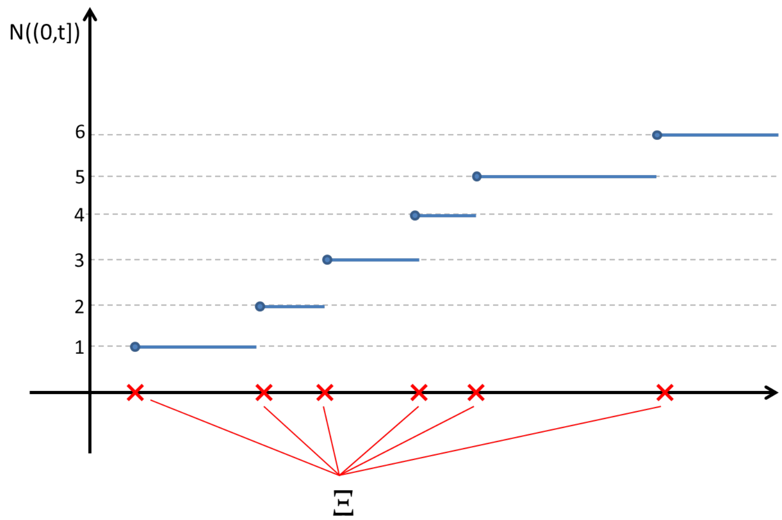
$$\mathbb{E}\left[e^{-\Upsilon}\right] = \exp\left\{-\int_{\mathbb{X}}\left(1-e^{-f(x)}\right)\mu(dx)\right\} \tag{3.1}$$

where

$$\Upsilon = \sum_{X\in\Xi} f(X)$$

The characteristic functional, as its name suggests, characterizes the Poisson process.

The next figure illustrates one possible realization of the particular case when $\mathbb{X} = \mathbb{R}$ and $\mathcal{X} = \mathscr{B}(\mathbb{R})$ and for $A \in \mathscr{B}(\mathbb{R})$ the mean measure μ is $\alpha\tilde{\mu}(A)$ where $\tilde{\mu}$ is the lebesgue measure on $\mathbb{R}$ and α is a positive constant, that is the following figure is a realization of a simple Poisson Process on $\mathbb{R}$ with parameter α.



Note that the random measure N over $(\mathbb{X}, \mathcal{X})$ in Definition 3.2 could be written in the form

$$N(\bullet) = |\bullet \cap \Xi| = \sum_{X \in \Xi} \delta_X(\bullet).$$

So, if each random location $X \in \Xi$ had assigned a mass of 1 then $N(\bullet)$ can be regarded as the sum of the masses assigned to each random location in $|\bullet \cap \Xi|$.

Generalizing this concept, let Ξ be a Poisson process on $\mathbb{X}$ with mean measure μ, and assume that to each location $X \in \Xi$ we assign a random mass m_X taking values in $\mathbb{R}^+$ such that the random mass m_X may (or may not) depend on its corresponding location X, but is independent of any other $Y \in \Xi$. Define $X^* = (m_X, X)$ then X^* is clearly a random variable taking values in the product space $\mathbb{X}^* := \mathbb{R}^+ \times \mathbb{X}$. Thus

$$\Xi^* := \{(m_X, X) : X \in \Xi\} \tag{3.2}$$

is a countable random subset of $\mathbb{X}^*$. Next we will try to justify that, as a matter of fact, Ξ^* is Poisson process on $\mathbb{X}^*$.

Formally, let $\rho(\bullet, x)$ be a probability distribution on $\mathbb{R}^+$ depending on $x \in \mathbb{X}$ in such way that for $B \in \mathscr{B}(\mathbb{R}^+)$, $\rho(B, \bullet)$ is a measurable function on $\mathbb{X}$. We say that the random countable subset Ξ^* of $\mathbb{R}^+ \times \mathbb{X}$, defined as in (3.2) is a marking of Ξ if its projection onto $\mathbb{X}$ is Ξ and the conditional distribution of Ξ^* given Ξ makes the sequence $\{m_X\}_{X \in \Xi}$ independent with respective distributions $\rho(\bullet, X)$.

Theorem 3.2 (Marking theorem)**.** *The random countable subset Ξ^* is a Poisson process on $\mathbb{R}^+ \times \mathbb{X}$ with mean measure*

$$\mu(A^*) = \int\int_{(m,x) \in A^*} \rho(dm, x)\mu(dx)$$

Proof:

For any measurable function on $\mathbb{R}^+ \times \mathbb{X}$ define

$$\Upsilon^* = \sum_{X \in \Xi} f(m_X, X)$$

Given Ξ, Υ^* is the sum of independente random variables, thus

$$\begin{aligned}
\mathbb{E}\left[e^{-\Upsilon^*}|\Xi\right] &= \prod_{X\in\Xi} \mathbb{E}\left[e^{-f(m_X,X)}\right] \\
&= \prod_{X\in\Xi} \int_{\mathbb{R}^+} e^{-f(m,X)} \rho(dm, X)
\end{aligned}$$

so that

$$\begin{aligned}
\mathbb{E}[e^{-\Upsilon^*}] &= \mathbb{E}\left[\mathbb{E}\left[e^{-\Upsilon^*}|\Xi\right]\right] \\
&= \mathbb{E}\left[\prod_{X\in\Xi} \int_{\mathbb{R}^+} e^{-f(m,X)} \rho(dm, X)\right] \\
&= \mathbb{E}\left[\exp\left\{-\sum_{X\in\Xi} -\log \int_{\mathbb{R}^+} e^{-f(m,X)} \rho(dm, X)\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{-\sum_{X\in\Xi} f^*(X)\right\}\right]
\end{aligned}$$

where

$$f^*(x) = -\log \int_{\mathbb{R}^+} e^{-f(m,x)} \rho(dm, x)$$

thus, applying (3.1) to the Poisson process we get

$$\begin{aligned}
\mathbb{E}\left[e^{-\Upsilon^*}\right] &= \exp\left\{-\int_{\mathbb{X}} \left(1 - e^{-f^*(x)}\right) \mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}} \left(1 - \int_{\mathbb{R}^+} e^{-f(m,x)\rho(dm,x)}\right) \mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{X}} \int_{\mathbb{R}^+} \left(1 - e^{-f(m,x)}\right) \rho(dm, x)\mu(dx)\right\} \\
&= \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{X}} \left(1 - e^{-f(m,x)}\right) d\mu^*\right\}.
\end{aligned}$$

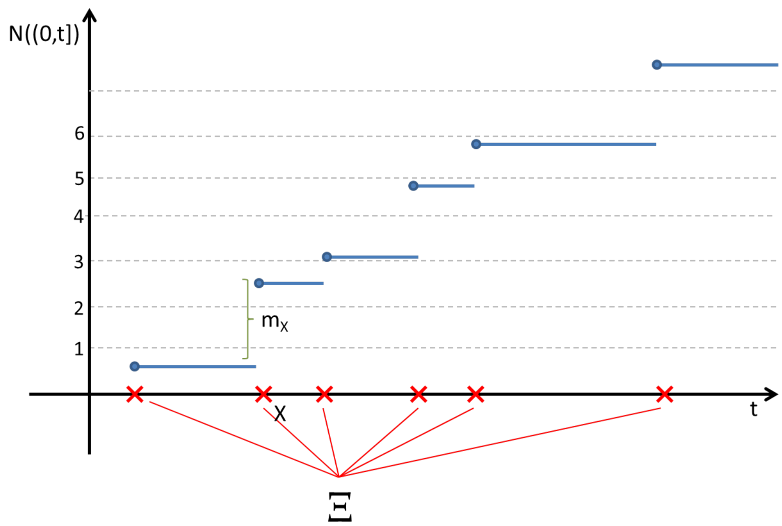
$\square$

If we, particularly, set $\{m_X\}_{X\in\Xi}$ to be a sequence of i.i.d. random variables such that m_X is also independent of its corresponding location X, and define

$$N(A) := \sum_{X\in\Xi} m_X \quad \forall A \in \mathcal{X}$$

then N is a CRM on $(\mathbb{X}, \mathcal{X})$.

The next figure illustrates one possible realization when $\mathbb{X} = \mathbb{R}$ and $\mathcal{X} = \mathscr{B}(\mathbb{R})$, and for $A \in \mathscr{B}(\mathbb{R})$ the mean measure of the Poisson Process Ξ is $\mu = \alpha\tilde{\mu}(A)$ where $\tilde{\mu}$ is the lebesgue measure on $\mathbb{R}$ and α a positive constant, that is the following figure is a realization of a Compound Poisson Process on $\mathbb{R}$.



In the next segment we are going to discuss CRMs more generally, and we are going to try and justify that as a matter of fact any complete random measure can be seen as a generalization of the Poisson Process.

Proposition 3.1. *Consider a completely random measure, Φ, as in Definition 3.1, let $A \in \mathcal{X}$ and for any $t > 0$ define*

$$\lambda_t(A) = -\log \mathbb{E}\left[e^{-t\Phi(A)}\right].$$

Then

1. *Φ and λ_t are mutually absolutely continuous with respect to each other.*
2. *Φ and λ_t are either finite or infinite together.*
3. *$\lambda_t : \mathcal{X} \to [0, \infty)$ is a measure on $(\mathbb{X}, \mathcal{X})$.*

Proof:

Clearly $0 \leq \lambda_t(A) \leq \infty$.

To prove the first part of the proposition it suffices to check that

$$\Phi(A) = 0\, a.s. \quad \Longleftrightarrow \quad e^{-t\Phi(A)} = 1 \quad \Longleftrightarrow \quad -\log \mathbb{E}\left[e^{-t\Phi(A)}\right] = \lambda_t(A) = 0,$$

as Φ is a measure then $\Phi(\emptyset) = 0$ a.s. the first part of the proposition implies $\lambda_t(\emptyset) = 0$.

From the definition of λ_t it is clear that

$$\lambda_t(A) = \infty \Longleftrightarrow \Phi(A) = \infty\, a.s.$$

The last thing we need to check to complete the proof is that λ_t is countably additive. So let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{X}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, it is clear that, as they are disjoint $\{\Phi(A_n)\}_{n=1}^{\infty}$ is a collection of independent random variables, so

$$\begin{aligned}
\lambda_t\left(\bigcup_{n=1}^{\infty} A_n\right) &= -\log \mathbb{E}\left[e^{-t\Phi\left(\bigcup_{n=1}^{\infty} A_n\right)}\right] \\
&= -\log \mathbb{E}\left[e^{-t\sum_{n=1}^{\infty} \Phi(A_n)}\right] \\
&= -\log \mathbb{E}\left[\prod_{n=1}^{\infty} e^{-t\Phi(A_n)}\right] \\
&= -\log \prod_{n=1}^{\infty} \mathbb{E}\left[e^{-t\Phi(A_n)}\right] \\
&= -\sum_{n=1}^{\infty} \log \mathbb{E}\left[e^{-t\Phi(A_n)}\right] \\
&= \sum_{n=1}^{\infty} \lambda_t(A_n)
\end{aligned}$$

This ends the proof and we can conclude that λ_t is a measure on $(\mathbb{X}, \mathcal{X})$ such that λ_t and Φ are mutually absolutely continuous and also Φ and λ_t are either finite or infinite together.

$\square$

Consider Φ a CRM and λ_t as in the last proposition. We now impose a further restriction, that the measure λ_t (for some $t > 0$ and then for all) be σ-finite. That is, there exists a partition $\{S_j\}_{j=1}^{\infty}$ of $\mathbb{X}$ such that $\lambda_t(S_j) < \infty$.

As for every $A \in \mathcal{X}$, $\lambda_t(A) = \infty$ if and only $\mathbb{P}[\Phi(A) = \infty] = 1$, then λ_t is σ-finite if and only if there exists a partition $\{S_j\}_{j=1}^{\infty}$ of $\mathbb{X}$ such that for every

$j \geq 1$ $\mathbb{P}[\Phi(S_j) < \infty] > 0$ which is not the same as saying that Φ is σ-finite a.s, and we shall say that Φ is Σ-finite when this holds.

From now on we are going to assume that Φ is Σ-finite. Define

$$\mathscr{A} = \{x \in \mathbb{X} : \lambda_t(\{x\}) > 0\},$$

that is, the set of atoms of λ_t, in fact as λ_t is a measure then $\mathscr{A}$ is at most countable. Note that as λ_t and Φ are mutually and absolutely continuous then

$$x \in \mathscr{A} \iff \mathbb{P}[\Phi(\{x\}) > 0] > 0,$$

therefore the set $\mathscr{A}$ is called the set of fixed atoms of Φ. Let us define $\phi(x) := \Phi(\{x\})$ then $\{\phi(x)\}_{x\in\mathscr{A}}$ is a countable collection of independent random variables.

If we write

$$\Phi_f(A) = \Phi(A \cap \mathscr{A}), \quad \Phi_c(A) = \Phi(A \cap (\mathbb{X} \setminus \mathscr{A}))$$

it is clear that both Φ_f and Φ_c are both completely random measures, they are independent of each other and $\Phi = \Phi_f + \Phi_c$. Furthermore

$$\Phi_f(A) = \sum_{x\in\mathscr{A}} \phi(x)\delta_x(A)$$

Let us focus on the non-fixed atoms component Φ_c of Φ. As the fixed atoms of λ_t can be removed by simply surgery, there is no loss of generality in assuming assume $\Phi = \Phi_c$ this is equivalent to say that λ_t is a non-atomic measure so $\mathscr{A} = \emptyset$.

Before continuing let us us state the following definition

Definition 3.3 (Infinitely divisible random variable)**.** *Let X be a random variable, we say that it is infinitely divisible if for every positive integer n*

$$X \stackrel{d}{=} \sum_{i=1}^{n} X_{n,i}$$

where $\{X_{n,i}\}_{i=1}^n$ is a collection of i.i.d. random variables.

Now, let $A \in \mathcal{X}$ such that $l = \lambda_1(A) < \infty$. A special case of Lyapunov's theorem ensures that for every $n \in \mathbb{N}$ there exists a partition $\{A_{n,j}\}_{j=1}^n$ of A such that

$$\lambda_1(A_{n,j}) = \frac{l}{n}$$

thus

$$\mathbb{E}\left[e^{-\phi(A_{n,j})}\right] = e^{-\lambda_1(A_{n,j})} = e^{-\frac{l}{n}},$$

and by Markov's inequality we have that for every $c > 0$

$$\mathbb{P}[\Phi(A_{n,j}) \geq c] \leq \frac{1 - e^{-\frac{l}{n}}}{1 - e^{-c}}$$

which tends to zero uniformly on j when n goes to ∞.

As $\Phi(A) = \sum_{j=1}^{n} \Phi(A_{n,j})$ where $\{\Phi(A_{n,j})\}_{j=1}^{n}$ are i.i.d. random variables, then we conclude that $\Phi(A)$ is an infinitely divisible random variable. The Lévy-Khintchine represention for positive infinitely divisible random variables specialises to the form

$$\mathbb{E}\left[e^{-t\phi(A)}\right] = \exp\left\{-\beta(A)t - \int_{\mathbb{R}^+} \left(1 - e^{-ts}\right) \nu(ds, A)\right\} \tag{3.3}$$

for $t > 0$ where $\beta \geq 0$ and $\nu(\bullet, A)$ is a measure on $\mathbb{R}^+$ which makes the integral converge. Therefore

$$\lambda_t(A) = \beta(A)t + \int_{\mathbb{R}^+} \left(1 - e^{ts}\right) \nu(ds, A) \tag{3.4}$$

By this representation, we see that β and ν characterize Φ completely through λ_t, where β is a measure on $(\mathbb{X}, \mathcal{X})$, and $\nu(\bullet, \bullet)$ is a two argument function such that

- For every fixed $A \in \mathcal{X}$ $\nu(\bullet, A)$ is a measure on $(\mathbb{R}^+, \mathscr{B}(\mathbb{R}^+))$, and,
- For every $B \in \mathscr{B}(\mathbb{R}^+)$ $\nu(B, \bullet)$ is a measure on $(\mathbb{X}, \mathcal{X})$.

As λ_t is non-atomic, both β and ν must be non-atomic on $\mathbb{X}$, and ν must be such that $\int_{\mathbb{R}^+} (1 - e^{-ts})\, \nu(ds, A) < \infty$ for every $A \in \mathcal{X}$.

If we compare equation (3.3) to Campbell's theorem we can intuit that there is in fact a very close relationship between CRMs and Poisson Process, in other to prove this close connection let us take a look at the situation from another perspective. Let us forget for a moment about β.

Suppose first that the measure

$$\mu(A) := \nu(\mathbb{R}^+, A)$$

is σ-finite. It is clear that for $z > 0$ the measure

$$\mu_z(A) := \nu((0, z], A)$$

is absolutely continuous with respect to μ, thus, it has a Radon-Nikodym derivate F:

$$\mu_z(A) = \int_{\mathbb{A}} F(z, x)\mu(dx)$$

Recall that the last equation determines $F(z, \bullet)$ almost everywhere (for each z) A version could be taken for each rational $z \in \mathbb{R}^+$ and then extended to $z \in \mathbb{R}^+$ by right continuity, so that for each $x \in \mathbb{X}$ fixed, $F(z,x)$ is the distribution function of a random variable taking values in $\mathbb{R}^+$.

Let

$$\mathbb{X}^* = \mathbb{R}^+ \times \mathbb{X},$$

we can now define a measure in this product space by

$$\mu^*(A^*) = \int\int_{A*} dF(z,x)\mu(dx) \quad \forall A^* \subseteq \mathbb{X}^*$$

such that $A^* = B \times A$ for some $B \in \mathscr{B}(\mathbb{R}^+)$, $A \in \mathcal{X}$. This becomes

$$\mu^*(B \times A) = \int_A \int_B dF(z,x)\mu(dx) = \int_B d\nu((0,z], A) = \nu(B, A)$$

Now in general $\mu(A) = \nu(\mathbb{R}^+, A)$ will not be σ-finite, but ν must satisfy

$$\int \left(1 - e^{-z}\right) \nu(dz, S_j) < \infty$$

so that $\nu((\epsilon, \infty), S_j) < \infty$ for every $\epsilon > 0$, where $\{S_j\}_{j=1}^{\infty}$ is a partition of $\mathbb{X}$. Thus, if we define for each $k \in \{0, 1, 2, ...\}$ the σ-finite measure

$$\mu^{(k)}(A) := \nu\left(\left[\frac{1}{k+1}, \frac{1}{k}\right), A\right)$$

we can write

$$\mu = \sum_{k=0}^{\infty} \mu^{(k)}$$

Applying the argument to each $\mu^{(k)}$, just as we did in the case that we assumed μ to be σ-finite, and summing over k we get that we can construct a measure μ^* on the product space $\mathbb{R}^+ \times \mathbb{X}$ such that for every $B \in \mathscr{B}(\mathbb{R}^+)$ and for every $A \in \mathcal{X}$

$$\mu^*(B \times A) = \nu(B, A).$$

Now let Ξ^* be a Poisson process on $\mathbb{X}^*$ with mean measure μ^*, and define the random measure on $\mathbb{X}$ given by

$$\Psi(A) = \sum \{z : (z,x) \in \Xi^*, x \in A\}$$

Then Ψ is a purely atomic measure on $\mathbb{X}$ whose atoms correspond to the location in Ξ in such way that if $(z,x) \in \Xi^*$ then Ψ assigns a mass of z to the location $x \in \mathbb{X}$. It follows from the definition of a Poisson Process that if $A_1, ..., A_n \in \mathcal{X}$ are disjoint sets, then $\Psi(A_1), ..., \Psi(A_n)$ are independent random variables, so Ψ

is a completely random measure. The distribution of Ψ can be calculated by Campell's Theorem, for $t > 0$, $A \in \mathbb{X}$.

$$\begin{aligned}\mathbb{E}\left[e^{-t\Psi(A)}\right] &= \exp\left\{-\int_A \int_{\mathbb{R}^+} \left(1 - e^{-tz}\right) \mu^*(dzdx)\right\} \\ &= \exp\left\{-\int_{\mathbb{R}^+} \left(1 - e^{-tz}\right) \nu(dz, A)\right\}\end{aligned}$$

Comparing this last equation with Equation (3.3) we obtain $\Phi = \beta + \Psi$. Therefore we have justified the next theorem

Theorem 3.3. *Let Φ be a Σ-finite completely random measure in $\mathbb{X}$. Then Φ can be decomposed as follows*

$$\Phi = \Phi_f + \Psi + \beta$$

where

i) Φ_f is a measure with random masses at fixed locations, $\mathscr{A} = \{x_1, x_2, ...\} \subseteq \mathbb{X}$

$$\Phi_f = \sum_{i=1}^{\infty} V_i \delta_{x_i},$$

non-negative random jumps $\{V_i\}_{i=1}^{\infty}$ independent of each other and independent of $\Psi + \beta$

ii) Ψ is a measure that can be written as

$$\Psi = \sum_{i=1}^{\infty} Z_i \delta_{X_i},$$

where $\{Z_i\}_{i \geq 1}$ are positive random jumps and $\{X_i\}_{i \geq 1}$ are $\mathbb{X}$-valued random locations, such that $\Xi = \{X_i\}_{i \geq 1}$ is a Poisson Process on $\mathbb{X}$ i.e. Ψ is a measure with random masses at random locations.

iii) β is simply a deterministic non-atomic measure.

For our purposes assume that $\Phi = \Psi$, i.e. we will be working with CRMs such that the Lévy-Khintchine representation theorem specializes to the form

$$\mathbb{E}\left[e^{-t\Phi(A)}\right] = \exp\left\{-\int_{\mathbb{R}^+} \left(1 - e^{-ts}\right) \nu(ds, A)\right\}$$

The measure ν is known as the Lévy intensity. Note that ν contains all the information of the random locations $\{X_i\}_{i \geq 1}$ and the random jumps $\{Z_i\}_{i \geq 1}$. There are two cases which are of interest to us.

1. $\nu(ds, dx) = \rho(ds, x)\mu(dx) = \rho_x(ds)\mu(dx)$. In this case we have that the random jumps depend on its corresponding random location and given the set of random locations Ξ (which is Poisson Process on $\mathbb{X}$ with mean μ) the jumps are independent of each other (this can be seen following the construction previously done).

2. $\nu(ds, dx) = \rho(ds)\mu(dx)$. In this second case we have that the random jumps are independent of each other and they are independent of the set of random locations Ξ which forms a Poisson Process on $\mathbb{X}$ with mean measure μ.

In first case we say that ν in *non-homogeneous*, whereas for the second case we say that the Lévy intensity ν is *homogeneous*.

Subordinators

We now specialize the theory to the case where $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A completely random measure on Φ $\mathbb{R}$, such that $\Phi(A) < \infty$ almost surely whenever A is bounded, defines a random process as follows: Let $\phi : \mathbb{R} \to \mathbb{R}$ be defined by

$$\phi(s) = \begin{cases} \Phi((0, s]) & \text{if } s \geq 0 \\ -\Phi([s, 0)) & \text{if } s < 0 \end{cases}$$

Note that ϕ is right-continuous and determines Φ uniquely. The independence of Φ on disjoint sets implies that ϕ has independent increments, in the sense that for $s_1 < s_2 < ... < s_n$ the increments

$$\phi(s_{k+1}) - \phi(s_k) = \Phi((s_k, s_{k+1}]) \quad k \in \{1, ..., n-1\}$$

are independent positive random variables. So the theory of completely random measures on $\mathbb{R}$ is conterminous with the theory of random processes with independent and positive increments. In particular, from the theory developed above, one can say that such process is the sum of three independent process:

- One is a deterministic increasing function,
- the second one is a random increasing function that jumps only at certain fixed discontinuities, and
- the third and last component is derived from a Poisson Process Ξ^* on the half plane $S^* = \{(x, z) : z > 0\}$ by the formula

$$\phi(s) = \begin{cases} \sum\{z : (x, z) \in \Xi^*, 0 < x \leq s\} & \text{if } s \geq 0 \\ -\sum\{z : (x, z) \in \Xi^*, s \leq x < 0\} & \text{if } s < 0 \end{cases}$$

An extremely relevant case is that in which for every $s < r$ the distribution of $\phi(r) - \phi(s)$ depends on s and r only through its difference, that is $\phi(r) - \phi(s) \stackrel{d}{=} \phi(r - s)$. Such process is called a *subordinator*. In other words a *subordinator* is a Lévy process[1] with non-decreasing trajectories.

[1] A Lévy process $\{X_s\}_{s \in \mathbb{R}}$ is a stochastic process such that $X_0 = 0$ a.s., its has independent and stationary increments and its trajectories are right-continuous and their left limit exists

A subordinator cannot have fixed discontinuities, this is because the discontinuities set is countable and therefore cannot be invariant under translations so any process having fixed discontinuities does not have stationary increments. Similarly, for the stationarity property to hold, the deterministic component can only be a constant multiple of s. Hence a subordinator has in it $t \in \mathbb{R}$, the representation

$$\phi(s) = \begin{cases} \beta s + \sum\{z : (x, z) \in \Xi^*, 0 < x \leq s\} & \text{if } s \geq 0 \\ \beta s - \sum\{z : (x, z) \in \Xi^*, s \leq x < 0\} & \text{if } s < 0 \end{cases}$$

where β is a constant and Ξ^* is a Poisson Process on the half plane which now must be invariant under translations parallel to the x-axis. Hence its mean measure μ^* must be of the form

$$\mu^*(dz, dx) = \rho(dz)dx$$

where ρ is a measure on $(0, \infty)$ such that

$$\int (1 - e^{-z})\rho(dz) < \infty.$$

So for $s < r$, equation (3.3) reduces to the form

$$\mathbb{E}\left[e^{t[\phi(r)-\phi(s)]}\right] = \exp\left\{-(r-s)\left[t\beta + \int_{\mathbb{R}^+} \left(1 - e^{-tz}\right)\rho(dz)\right]\right\}$$

Chapter 4

Normalized independent random measures

Consider a CRM Φ as in Definition 3.1 such that $\Phi(\mathbb{X}) < \infty$ a.s. Define for every $A \in \mathcal{A}$

$$\tilde{p}(A) := \frac{\Phi(A)}{\Phi(\mathbb{X})}.$$

Then $\tilde{p}$ is a random probability measure called *Normalized independent random measure* (NRMI for short).

By Theorem 3.3 any Σ-finite CRM Φ can be decomposed as $\Phi = \Psi + \beta + \Phi_f$, if we assume that the deterministic measure $\beta \equiv 0$ and the fixed atoms component $\Phi_f \equiv 0$, then can be written ass follows:

$$\Phi = \sum_{i=1}^{\infty} J_i \delta_{\xi_i},$$

where $\{J_i\}_{i \geq 1}$ are positive random jumps and $\{\xi_i\}_{i \geq 1}$ are $\mathbb{X}$-valued random locations, such that $\Xi = \{\xi_i\}_{i \geq 1}$ is a Poisson Process on $\mathbb{X}$

By the Lévy Khintchine representation, Φ is characterized by its Lévy intensity ν, which holds

$$\mathbb{E}\left[e^{-t\Phi(A)}\right] = \exp\left\{-\int_{\mathbb{R}^+} \left(1 - e^{-ts}\right) \nu(ds, A)\right\},$$

where ν contains all the information of the random jumps and locations. Let us assume that ν can be factorized as

$$\nu(ds, dx) = \rho_x(ds)\mu(dx)$$

where μ is a non-atomic measure, and $\rho_x(\bullet)$ may or may not depend on x, depending if we lay on the case where the random jumps are dependent of their corresponding random location or not.

If the CRM Φ holds all the previous assumptions then $\tilde{p}$ can be written in the form

$$\tilde{p}(A) = \sum_{i=1}^{\infty} \tilde{p}_i \delta_{\xi_i}(A)$$

where $\{\tilde{p}_i\}_{i\geq 1}$ is a sequence of non-negative random variables such that $\sum_{i=1}^{\infty} \tilde{p}_i = 1$ a.s. and given the random locations $\{\xi_i\}_{i\geq 1}$,the masses $\{\tilde{p}_i\}_{i\geq 1}$ are independent.

Before moving on, it is important to highlight a couple of things about a NRMIs

- As Φ is characterized by its Lévy intensity ν, then $\tilde{p}$ is also characterized by ν.
- $\tilde{p}$ is not necessarily a CRM itself. It is clear that $\tilde{p}$ is a random measure, but if there exist $A \in \mathcal{X}$ such that $\mu(A)$ and $\mu(B)$ are not a constant almost surely, (where $B = \mathbb{X} \setminus A$), then

  $$\tilde{p}(A) = \frac{\mu(A)}{\mu(\mathbb{X})} \text{ and } \tilde{p}(B) = \frac{\mu(B)}{\mu(\mathbb{X})}$$

  must sum up to 1, this means that the value of $\tilde{p}(B)$ is determined given $\tilde{p}(A)$. Thus we have found a couple of disjoint sets A and B such that $\tilde{p}(A)$ is not independent of $\tilde{p}(B)$.

Now we recall the discussion in Chapter 2.3, let $\{X_i\}_{i=1}^{\infty}$ be a sequence of $\mathbb{X}$-valued random variables. By Bruno de Finetti's representation theorem, $\{X_i\}_{i=1}^{\infty}$ are exchangeable iff there exists a probability measure Q over the space of all probability measures on $(\mathbb{X}, \mathcal{X})$ such that

$$\begin{aligned} X_i|\tilde{p} &\overset{i.i.d.}{\sim} \tilde{p} \quad i = 1, 2, ... \\ \tilde{p} &\sim Q \end{aligned}$$

Now, let $\{X_i\}_{i=1}^{\infty}$ be a sequence of $\mathbb{X}$-valued random variables such that the last couple of equations holds where $\tilde{p}$ is a NRMI, then it is clear that they are exchangeable and that the distribution of $\tilde{p}$, Q, is characterized by the corresponding Lévy intensity $\nu(ds, dx) = \rho_x(ds)\mu(dx)$. It has also been stated that $\tilde{p}$ is discrete a.s. (as a consequence of the construction we have made of an NRMI) so we stand on the case we discussed in Chapter 2.3.

We have already seen that every finite sample of size n, $X_1, ...X_n$ will feature ties with positive probability, in fact if $X \sim \tilde{p} = \sum_{i=1}^{\infty} \tilde{p}_i \delta_{\xi_i}$, then

$$\mathbb{P}[X = \xi_i] = \tilde{p}_i \quad i = 1, 2, ...$$

so every n-sized sample $X_1, ..., X_n$ generates a random partition Π_n of $[n]$ as follows. If $X_1^*, ..., X_k^*$ are the distinct values observed in the sample the induced partition Π_n will have k blocks and for every $i, j \in [n]$ i will belong to the same block as j if and only if $X_i = X_j = X_r^*$ for some $r \in \{1, 2, ..., k\}$. We have also discussed that as a consequence of the exchangeability of sample Π_n turns out to be an exchangeable partition as for every $\pi = \{B_1, ..., B_k\} \in \mathcal{P}_{[n]}$.

$$\mathbb{P}[\Pi_n = \pi] = \mathbb{P}[K_n = k, N_{n,1} = n_1, ..., N_{n,k} = n_k] = p_n(n_1, ..., n_k)$$

for some symmetric function p_n of its arguments, where K_n is the random number of blocks of Π_n, and $N_{n,j} = |B_j|$ for every block B_j of Π_n.

Also, if we consider $\{X^{(n)}\}_{n\geq 1}$ where $X^{(n)} = \{X_1, ..., X_n\}$ is an n-sized sample, such that for every $n \geq 1$, $X^{(n+1)} = \{X^{(n)}\} \cup \{X_{n+1}\}$ (this is, to obtain $X^{(n+1)}$ we take the already obtained sample, $X^{(n)}$, of size n, and sample one more value) then the family of induced exchangeable random partitions $\{\Pi_n\}_{n\geq 1}$ holds the addition rule, as for every $n \geq 1$

$$p_n(n_1, ...n_k) = p_{n+1}(n_1, ..., n_k, 1) + \sum_{j=1}^{k} p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k). \tag{4.1}$$

To see this, it suffices to realize that given $X^{(n)}$, X_{n+1} will necessarily be equal to an already observed value ($X_{n+1} \in \{X_1^*, ..., X_k^*\}$) or will be an unobserved value ($X_{n+1} \notin \{X_1^*, ..., X_k^*\}$).

Summarizing, given a Lévy intensity ν associated to a NRMI, we can construct a consistent family of exchangeable random partitions $\{\Pi_n\}_{n\geq 1}$ by taking samples from $\{X_i\}_{i=1}^{\infty}$ where

$$\begin{aligned} X_i|\tilde{p} &\overset{i.i.d.}{\sim} \tilde{p} \quad i = 1, 2, ... \\ \tilde{p} &\sim Q \end{aligned}$$

and Q is characterized by ν. The explicit computation of the EPPF it is not at all trivial, in the next couple of sections we will discuss this problem.

4.1 The gamma process

In this section we will discuss one particular case in which the computation of the EPPF is done quite easily. First let us recall the definition of the Dirichlet distribution.

Definition 4.1. *Let $\tilde{p} = (\tilde{p}_1, ..., \tilde{p}_n)$ be a random probability vector (i.e. $\tilde{p}_i \geq 0$ for every $i \in \{1, ..., n\}$ and $\sum_{i=1}^n \tilde{p}_i = 1$) whose density (relative to the $(n-1)$-dimensional lebesgue measure) is given by*

$$f(p_1, ..., p_n) = \frac{\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n p_i^{\alpha_i - 1} \tag{4.2}$$

where $\alpha_i > 0$ for every $i = 1, ..., n$, then we say that $\tilde{p}$ has a Dirichlet distribution with parameters $(\alpha_1, ..., \alpha_n)$ ($\tilde{p} \sim Dir(\alpha_1, ..., \alpha_n)$).

Direct manipulation of (4.2) can be complicated. A better way to handle this is the following. Let $Y_1, ...Y_n$ be independent positive independent random variables such that $Y_i \sim Gamma(\alpha_i, 1)$ i.e. Y_i has a density function given by

$$f_{Y_i}(y) = \frac{y^{\alpha_i - 1} e^{-y}}{\Gamma(\alpha_i)} \mathbf{1}_{\{y \in (0,\infty)\}}.$$

Define $Y = \sum_{i=1}^n Y_i$ and $\tilde{p}_i = \frac{Y_i}{Y}$, $i = 1, ..., n$, it can easily be seen that

$$(\tilde{p}_1, ..., \tilde{p}_n) \sim Dir(\alpha_1, ..., \alpha_n)$$

The proof is immediate considering the change of variables

$$(Y_1, ..., Y_n) \longmapsto \left(\sum_{i=1}^n Y_i, \frac{Y_1}{\sum_{i=1}^n Y_i}, ..., \frac{Y_{n-1}}{\sum_{i=1}^n Y_i} \right) = (Y, \tilde{p}_1, ..., \tilde{p}_{n-1})$$

It is widely known that $Y \sim Gamma(\alpha, 1)$ where $\alpha = \sum_{i=1}^n \alpha_i$, thus it is an infinitely divisible random variable with Lévy-Khintchine representation given by

$$\mathbb{E}\left[e^{-tY}\right] = \frac{1}{(1+t)^\alpha} = \exp\left\{ -\alpha \int_0^\infty (1 - e^{-ts}) \frac{e^{-s}}{s} ds \right\}$$

Corresponding to the last Lévy-Khintchine representation there is a subordinator called *Moran gamma process*, this process is defined exactly as in Chapter 3 with

$$\beta = 0, \qquad \nu(dz) = \frac{e^{-z}}{z} dz$$

so for $s < r$

$$\mathbb{E}\left[e^{-t(\phi(r)-\phi(s))}\right] = \exp\left\{ -(r-s) \int_0^\infty (1 - e^{-tz}) \frac{e^{-z}}{z} dz \right\} = \frac{1}{(1+t)^{(r-s)}}$$

hence $\phi(r) - \phi(s) \sim Gamma((r-s), 1)$. Note that

$$\int_0^\infty \frac{e^{-z}}{z} dz = \infty$$

this means that the jumps are dense everywhere.

Let $\alpha_1, \alpha_2, ...\alpha_n > 0$, and define

$$s_0 = 0, \qquad s_k = \sum_{i=1}^k \alpha_i \quad k \in \{1, ..., n\}$$

Then $Y_k = \phi(s_k) - \phi(s_{k-1}) \sim Gamma(\alpha_k, 1)$ and $Y_1, ..., Y_k$ are independent. Since

$$Y = \sum_{k=1}^{n} Y_k = \phi(s_n)$$

we see that

$$\tilde{p}_k = \frac{\phi(s_k) - \phi(s_{k-1})}{\phi(s_n)}$$

defines a n-dimesional random probability vector such that

$$(\tilde{p}_1, ..., \tilde{p}_n) \sim Dir(\alpha_1, ..., \alpha_n).$$

This generalizes to quite arbitrary spaces in a very natural form. As we have seen, the Gamma process corresponds to a completely random measure on $\mathbb{R}$ having the property that for every $A \in \mathscr{B}(\mathbb{R})$ $\Phi(A) \sim Gamma(\alpha, 1)$ where α is the lebesgue measure of A. Now let μ we a non-atomic finite measure over $\mathbb{X}$, and define the measure over $\mathbb{X} \times \mathbb{R}^+$

$$\nu(B, A) = \int_B \frac{e^{-s}}{s} ds \mu(A)$$

Thus the Poisson Process Ξ^* on $\mathbb{X}^* = \mathbb{X} \times \mathbb{R}^+$ has mean measure which is the product on μ and $\nu(ds) = \frac{e^{-s}}{s} ds$, and

$$\Phi(A) = \sum \{s : (s, x) \in \Xi^*, x \in A\}$$

defines a CMR on $\mathbb{X}$ with

$$\mathbb{E}\left[e^{-t\Phi(A)}\right] = \exp\left\{-\int_0^\infty \mu(A)(1 - e^{-ts})\frac{e^{-s}}{s} ds\right\} = \frac{1}{(1+t)^{\mu(A)}}$$

that is $\Phi(A) \sim Gamma(\mu(A), 1)$.

As μ is finite then $\Phi(\mathbb{X}) < \infty$ a.s. and we can define the NRMI

$$\tilde{p}(A) := \frac{\Phi(A)}{\Phi(\mathbb{X})} \quad A \in \mathcal{X}$$

and it is clear that for any finite partition $\{A_i\}_{i=1}^n$ of $\mathbb{X}$

$$(\tilde{p}(A_1), ..., \tilde{p}(A_n)) \sim Dir(\mu(A_1), ..., \mu(A_n))$$

This particular NRMI is called a *Dirichlet Process* with correponding measure μ and we write $\tilde{p} \sim \mathcal{D}_\mu$.

Let $\{X_i\}_{i=1}^n$ be a sequence of exchangeable random variables such that

$$\begin{aligned} X_i | \tilde{p} &\overset{i.i.d.}{\sim} \tilde{p} \\ \tilde{p} &\sim \mathcal{D}_\mu \end{aligned}$$

Note that for every $A \in \mathcal{X}$ $\mathbb{P}[X \in A|\tilde{p}] = \tilde{p}(A)$, and

$$P_0(A) := \mathbb{P}[X \in A] = \int_{\mathcal{P}_{\mathbb{X}}} \tilde{p}(A)\mathcal{D}_\mu(d\tilde{p}) = \mathbb{E}_{\mathcal{D}_\mu}[\tilde{p}(A)] = \frac{\mu(A)}{\theta}$$

where $\theta := \mu(\mathbb{X})$. It is important to highlight that as μ is non-atomic then P_0 must be a non-atomic probability measure.

Let $\{A_1, ..., A_m\}$ be a partition of $\mathbb{X}$ so that $\widehat{p} := (\tilde{p}(A_1), ..., \tilde{p}(A_m)) \sim Dir(\alpha_1, ..., \alpha_m)$ where $\alpha_i = \mu(A_i)$ for $i = 1, ..., m$. Assume we have managed to obtain a sample $X^{(n)} = X_1, ..., X_n$. We are now interested in the posterior distribution of $\widehat{p}$ given $X^{(n)}$. As $\{A_i\}_{i=1}^m$ is a partition of $\mathbb{X}$ then for every random variable X taking values in $\mathbb{X}$ there exists one unique $k \in \{1, ..., m\}$ such that $X \in A_k$. So we can define the random variable Z_k taking values in $\{1, ..., m\}$ as follows

$$Z_k = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \in A_k\}}$$

that is Z_k counts the number of X_is such that $X_i \in A_k$. It is clear that $Z = (Z_1, ..., Z_m)$ and $X^{(n)}$ provide the same information about $\widehat{p} = (\tilde{p}(A_1), ..., \tilde{p}(A_m))$, i.e.

$$\mathbb{P}[\widehat{p} = (p_1, ..., p_m)|X^{(n)}] = \mathbb{P}[\widehat{p} = (p_1, ..., p_m)|Z].$$

It is also easy to see that $Z|(\tilde{p}(A_1), ..., \tilde{p}(A_m)) \sim Multinomial(\tilde{p}(A_1), ..., \tilde{p}(A_m), n)$, that is

$$\mathbb{P}[Z = (z_1, ..., z_m)|\widehat{p}] = \frac{n!}{z_1! \cdots z_m!}\tilde{p}(A_1)^{z_1} \cdots \tilde{p}(A_m)^{z_m}.$$

Hence we can compute

$$\begin{aligned}
&\mathbb{P}[\widehat{p} = (p_1, ..., p_m)|Z = (z_1, ..., z_m)] \\
&\qquad \propto \mathbb{P}[Z = (z_1, ..., z_m)|\widehat{p} = (p_1, ..., p_m)]\mathbb{P}[\widehat{p} = (p_1, ..., p_m)] \\
&\qquad = \left(\frac{n!}{z_1! \cdots z_m!} p_1^{z_1} \cdots p_m^{z_m}\right)\left(\frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m p_i^{\alpha_i - 1}\right) \\
&\qquad \propto \left(\prod_{i=1}^m p_i^{z_i}\right)\left(\prod_{i=1}^m p_i^{\alpha_i - 1}\right) \\
&\qquad = \prod_{i=1}^m p_i^{\alpha_i + z_i - 1} \\
&\qquad \propto \frac{\Gamma\left(\sum_{i=1}^m \alpha_i + z_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i + z_i)} \prod_{i=1}^m p_i^{\alpha_i + z_i - 1}
\end{aligned}$$

as $z_i = \sum_{j=1}^n \mathbf{1}_{\{x_j \in A_i\}} = \sum_{j=1}^n \delta_{x_j}(A_i)$ then we can conclude that the posterior distribution of $\widehat{p} = (\tilde{p}(A_1), ..., \tilde{p}(A_m))$ given $X^{(n)}$ is

$$(\tilde{p}(A_1), ..., \tilde{p}(A_m))|X^{(n)} \sim Dir\left(\alpha_1 + \sum_{j=1}^n \delta_{X_j}(A_1), ..., \alpha_m + \sum_{j=1}^n \delta_{X_j}(A_m)\right)$$

As this happens for every partition of $\mathbb{X}$ then we have that $\tilde{p}|X^{(n)}$ is also a Dirichlet process with mean measure

$$\mu + \sum_{i=1}^{n} \delta_{X_i},$$

hence by (2.7), for every $A \in \mathcal{X}$

$$\begin{aligned}
\mathbb{P}[X_{n+1} \in A|X^{(n)}] &= \mathbb{E}_{\mathcal{D}_{\mu+\sum_{i=1}^{n}\delta_{X_i}}}[\tilde{p}(A)] \\
&= \frac{\mu(A) + \sum_{i=1}^{n} \delta_{X_i}(A)}{\mu(\mathbb{X}) + \sum_{i=1}^{n} \delta_{X_i}(\mathbb{X})} \\
&= \frac{\theta P_0(A) + \sum_{i=1}^{n} \delta_{X_i}(A)}{\theta + n}
\end{aligned}$$

In other words given an n-sized sample which exhibits k distinct values $X_1^*, ..., X_k^*$ with corresponding frequencies $n_1, ..., n_k$

$$X_{n+1} = \begin{cases} X_j & \text{with probability } \frac{n_j}{\theta+n},\ j = 1, 2, ..., k \\ X & \text{with probability } \frac{\theta}{\theta+n} \end{cases}$$

where X is a random variable independent of $X_1, ..., X_n$ such that $X \sim P_0$. As P_0 is non-atomic, $\mathbb{P}[X = X_j^*] = 0$ for every $j = 1, 2, ..., k$.

From this and the fact that the addition rule holds (4.1) for every $n \geq 1$,

$$\begin{aligned}
\frac{\theta}{\theta + n} &= \mathbb{P}[X_{n+1} =\notin \{X_1^*, ..., X_k^*\}|X^{(n)}] = \frac{p_{n+1}(n_1, ..., n_k, 1)}{p_n(n_1, ...n_k)} \\
\frac{n_j}{\theta + n} &= \mathbb{P}[X_{n+1} = X_j^*|X^{(n)}] = \frac{p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k)}{p_n(n_1, ...n_k)} \quad j = 1, ..., k,
\end{aligned}$$

thus

$$\begin{aligned}
p_{n+1}(n_1, ..., n_k, 1) &= \frac{\theta}{\theta + n} p_n(n_1, ...n_k) \\
p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k) &= \frac{n_j}{\theta + n} p_n(n_1, ...n_k) \quad j = 1, ..., k
\end{aligned}$$

from these equations, together with $p_1(1) = 1$, one can recursively prove that the EPPF of the corresponding consistent family of exchangeable partitions takes the form

$$p_n(n_1, ...n_k) = \frac{\theta^k}{(\theta)_{n\uparrow 1}} \prod_{j=1}^{k} (n_j - 1)!$$

where $(\theta)_{n\uparrow 1} = (\theta)(\theta + 1) \cdots (\theta + n - 1)$.

4.2 The general case

In the Dirichlet Process scenario the feature that made everything tractable, is the fact that if $\tilde{p}$ is a Dirichlet Process then the posterior distribution of $\tilde{p}$ given the sample turns out to be a Dirchlet process as well, these characteristic is known as conjugacy. In the general case where the Lévy intensity ν takes the form $\nu(ds, dx) = \rho_x(ds)\mu(dx)$, the calculations are not at all simple. However, it has been possible to show that, conditional on a specific latent variable, the posterior distribution of an NRMI coincides with the posterior distribution of another NRMI having a rescaled intensity and fixed points of discontinuity. This can be seen as conditional conjugacy.

Now again, let $\{X_i\}_{i=1}^{\infty}$ be a sequence of exchangeable random variables such that $X_i \overset{i.i.d.}{\sim} \tilde{p}$ where $\tilde{p}$ is a NRMI. Since NRMIs are a.s. discrete data can display ties, let $X^{(n)}$ be a sample of size n and $X_1^*, ..., X_k^*$ be the k distinct observations with corresponding frequencies $n_1, ..., n_k$ presented in the n-sized sample. Let us introduce a key latent variable, let U_n be a positive random variable whose density function conditional on the sample $X^{(n)}$ satisfies

$$q_{X^{(n)}}(u) \propto u^{n-1} e^{-\Theta(u)} \prod_{j=1}^{j} \tau_{n_j}(u|X_j^*)$$

where $\Theta(u)$ denotes the Laplace exponent of Φ, that is

$$\Theta(u) = \mathbb{E}[e^{-u\Phi(\mathbb{X})}] = \int_{\mathbb{X}} \int_{\mathbb{R}^+} (1 - e^{-us})\rho_x(ds)\mu(dx)$$

and for every $m \geq 1$

$$\tau_m(u|x) := \int_{\mathbb{R}^+} s^m e^{-us} \rho_x(ds)$$

The posterior distribution of Φ and of $\tilde{p}$ given the sample $X^{(n)}$ does not have tractable form, but the distribution of Φ and of $\tilde{p}$ given the sample $X^{(n)}$ and U_n does, as is stated in the following result.

Theorem 4.1. *If $\tilde{p}$ is an NRMI obtained from normalizing the CRM Φ with Lévy intensity $\nu(ds, dx) = \rho_x(ds)\mu(dx)$ then*

i) The posterior distribution of Φ given $X^{(n)}$ and U_n is

$$\Phi|(X^{(n)}, U_n) \overset{d}{=} \Phi^{(U_n)} + \sum_{j=1}^{k} J_j^{(U_n)} \delta_{X_j^*}$$

ii) The posterior distribution of $\tilde{p}$ given $X^{(n)}$ and U_n is

$$\tilde{p}|(X^{(n)}, U_n) \overset{d}{=} w \frac{\Phi^{(U_n)}}{\Phi^{(U_n)}(\mathbb{X})} + (1 - w) \frac{\sum_{j=1}^{k} J_j^{(U_n)} \delta_{X_j^*}}{\sum_{j=1}^{k} J_j^{(U_n)}}$$

where

- $\Phi^{(U_n)}$ *is a CRM with Lévy intensity* $\nu^{(U_n)}(ds, dx) = e^{-U_n s}\rho_x(ds)\mu(dx)$.
- *The non-negative jumps* $\{J_j^{(U_n)}\}_{j=1}^k$ *are mutually independent and are also independent from* $\Phi^{(U_n)}$ *and* $J_j^{(U_n)}$ *has density function* $f_j(s) \propto s^{n_i} e^{-U_n s}\rho_{X_j^*}(ds)$.
- $w = \dfrac{\Phi^{(U_n)}(\mathbb{X})}{\Phi^{(U_n)}(\mathbb{X}) + \sum_{j=1}^k J_j^{(U_n)}}$.

As the predictive distribution of X_{n+1} coincides with the posterior expected value of $\tilde{p}$, then

$$\begin{aligned}\mathbb{P}[X_{n+1} \in \bullet | X^{(n)}] &= \int_0^\infty \mathbb{P}[X_{n+1} \in \bullet | X^{(n)}, U_n = u](c * q_{X^{(n)}}(u))du \\ &= \int_0^\infty \int_{\mathcal{P}_\mathbb{X}} \tilde{p}(\bullet)Q(d\tilde{p}|X^{(n)}, U_n = u)q_{X^{(n)}}(u)du\end{aligned}$$

Thus the last theorem leads to the following proposition

Proposition 4.1. *Let* $\tilde{p}$ *be an NRMI with Lévy intensity* $\nu(ds, dx) = \rho_x(ds)\mu(dx)$, *then the predictive distribution of* X_{n+1} *given* $X^{(n)}$ *coincides with*

$$\mathbb{P}[X_{n+1} \in dx | X^{(n)}] = w^{(n)} P_0(dx) + \frac{1}{n}\sum_{i=1}^k w_i^{(n)} \delta_{X_i^*}(dx)$$

where $P_0 = \frac{\mu}{\mu(\mathbb{X})}$, *and for* $i \in \{1, 2, ..., k\}$

$$w^{(n)} = \frac{1}{n}\int_{\mathbb{R}^+} u\tau_1(u|x)q_{X^{(n)}}(u)du, \quad w_i^{(n)} = \int_{\mathbb{R}^+} u\frac{\tau_{n_i+1}(u|X_i^*)}{\tau_{n_i}(u|X_i^*)}q_{X^{(n)}}(u)du$$

Note that the predictive distribution has a quite and intuitive form as it consist of a linear combination of P_0 and a weighted version of the empirical distribution function of the original sample. This prediction rule gives us a sampling scheme and so a way to construct a consistent family of exchangeable random partitions as done earlier. As it turns out there is an expression to the EPPF in the general case, as is stated bellow.

Proposition 4.2. *Let* $\tilde{p}$ *be an NRMI. Then the corresponding EPPF is given by*

$$p_n(n_1, ...n_k) = \frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}e^{-\Theta(u)}\left[\prod_{i=1}^k \kappa_{n_i}(u)\right]$$

where for every $i = 1, 2, ..., k$

$$\kappa_{n_i}(u) = \int_\mathbb{X} \tau_{n_i}(u|x)\mu(dx)$$

For a proof of the above theorems, review [9] and [6] in the bibliography.

4.3 Other examples of NRMIs

We have already studied an NRMI, the Dirichlet process, in this segment we will name some other important NRMIs in the literature.

The σ-stable NRMI

Let $\sigma \in (0,1)$. Consider a CRM $\Phi = \tilde{\mu}_\sigma$ with corresponding Lévy intensity

$$\nu(ds,dx) = \frac{\sigma}{\Gamma(1-\sigma)s^{1+\sigma}} ds\mu(dx)$$

Then, $\tilde{\mu}_\sigma$ is called a σ-stable process with parameter σ on $\mathbb{X}$. The Laplace transform takes the form

$$\mathbb{E}\left[e^{-t\Phi(A)}\right] = \exp\left\{-\int_0^\infty \frac{\sigma(1-e^{-ts})}{\Gamma(1-\sigma)s^{1+\sigma}} ds\mu(A)\right\} = e^{-t^\sigma \mu(A)}$$

The normalized generalized gamma process

A third example of a NRMI is based on the Generalized Gamma process with Lévy intensity

$$\nu(ds,dx) = \frac{e^{-\tau s}s^{-(1+\sigma)}}{\Gamma(1-\sigma)} ds\mu(dx)$$

for $\tau \geq 0$ and $\sigma \in (0,1)$. Note that when $\tau = 0$ the NRMI we obtain is that of the σ-stable NRMI, also when $\tau = 1$ and $\sigma \to 0$ the case of the Dirichlet process arises. In the case of $\sigma = \frac{1}{2}$ we obtain the so-called Normalized Inverse Gaussian process to which the Lévy intensity reduces to the form

$$\nu(ds,dx) = \frac{\exp\left\{-\frac{\tau}{2}\right\}}{\sqrt{s^3\pi}} ds\mu(dx)$$

4.4 Poisson-Kingman models

Consider once again a CRM Φ such that its Lévy intensity can be written as $\nu(ds,dx) = \rho(ds)\mu(dx)$ with $\rho(\mathbb{R}^+) = \infty$ and μ is a non-atomic measure, that is

$$\Phi = \sum_{i=1}^{\infty} J_i \delta_{\xi_i},$$

where $\{J_i\}_{i\geq 1}$ are positive random jumps and $\{\xi_i\}_{i\geq 1}$ are $\mathbb{X}$-valued random locations, such that $\Xi = \{\xi_i\}_{i\geq 1}$ is a Poisson Process on $\mathbb{X}$ and the jumps are independent of the locations.

Denote $J_1^\downarrow \geq J_2^\downarrow \geq \cdots$ the ranked jumps of the CRM, set

$$T := \Phi(\mathbb{X}) = \sum_{i=1}^{\infty} J_i \delta_{\xi_i}(\mathbb{X}) = \sum_{i=1}^{\infty} J_i = \sum_{i=1}^{\infty} J_i^\downarrow$$

and assume that the probability distribution of the total mass T is absolutely continuous with respect to the lebesgue measure on $\mathbb{R}^+$. Now, let us define

$$\tilde{p}_i^{\downarrow} := \frac{J_i^{\downarrow}}{T} \quad i \geq 1$$

Let $S^* = \{(p_1, p_2, ...) : p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i\geq 1} p_i = 1\}$ the set of all the sequences of ordered of non-negative numbers that sum up to 1.

Definition 4.2. *Let $P_{\rho,t}$ denote the conditional distribution of the sequence $\{\tilde{p}_{(i)}\}_{i\geq 1}$ of ranked normalized jumps generated by a CRM with Lévy intensity $\nu(ds, dx) = \rho(ds)\mu(dx)$, given $T = t$. Let γ be a probability distribution on $\mathbb{R}^+$. The distribution*

$$\int_{\mathbb{R}^+} P_{\rho,t}\gamma(dt)$$

on S^* is termed a Poisson-Kingman distribution with Lévy intensity ρ and mixing intensity γ. It is denoted by $PK(\rho, \gamma)$.

Define

$$\tilde{p} = \sum_{i\geq 1} \tilde{p}_i^{\downarrow}\delta_{\xi_i}$$

where $(\tilde{p}_1^{\downarrow}, \tilde{p}_2^{\downarrow}, ...)$ follow a $PK(\rho, \gamma)$ distribution, then $\tilde{p}$ is termed a $PK(\rho, \gamma)$ probability measure.

If γ coincides with the probability distribution of T, then we write $PK(\rho)$ instead of $P(\rho, \gamma)$ in the previous definition. Also, given the construction we have made it is easy to corroborate that, as the random jumps and the random locations of a homogeneous CRM are independent (and we have not yet established a particular order in the locations of the underlying Poisson Process on $\mathbb{X}$), then

$$\tilde{p} = \sum_{i\geq 1} \tilde{p}_i^{\downarrow}\delta_{\xi_i} = \sum_{i\geq 1} \frac{J_i^{\downarrow}}{T}\delta_{\xi_i} \stackrel{d}{=} \sum_{i\geq 1} \frac{J_i}{T}\delta_{\xi_i} = \frac{\Phi}{\Phi(\mathbb{X})}$$

so the $PK(\rho)$ process are equivalent to NRMIs generated by the same homogeneous CRM.

One very important subclass of the Poisson-Kingman models is the *Poisson Dirichlet* process also known as the *Pitman-Yor* process, this can be constructed by considering the NRMI to be a Dirichlet process or a stable process and normalizing its jumps. We will study this process with detail further on.

Chapter 5

Exchangeable partitions and partitions generated by sampling without replacement

In the past chapters we studied how to generate a consistent family of exchangeable partitions by sampling from a sequence of exchangeable random variables, now we see that any consistent family of exchangeable partitions can be thought as if it was generated this way.

Let $\{\Pi_n\}_{n\geq 1}$ be a consistent family of exchangeable random partitions. We have seen in Chapter 2 that together with the random object Π_n there are implicitly defined the random variables K_n which corresponds to the number of blocks in Π_n, and $N_{n,1}, ..., N_{n,K_n}$ which refer to the sizes of the blocks of Π_n. It is clear that the sequence of positive random variables $\mathbf{N} = (N_{n,1}, ..., N_{n,K_n})$ defines a random composition of n, that is, $\mathbf{N}$ is a random object taking values on $\mathcal{C}_n$.

As one can intuit from the EPPF, in an exchangeable random partition, it is of primary interest the study of the block sizes. Next we show three different ways to encode these block sizes as a random composition on $[n]$ and show how the distribution of these encodings are determined by the EPPF p.

Remark: From now on we will denote the EPPF by p instead of p_n in order to simplify the notation. Note as the sizes of the blocks of any partition of $[n]$ sum up to n, then by summing over all the arguments in p on can determine to which n does p refers to.

Exchangeable random order

Let $(N^{ex}_{n,1}, ..., N^{ex}_{n,K_n})$ be the random composition of n induced by an exchangeable partition $\Pi_n = \{A_1, ..., A_{K_n}\}$ by uniformly permuting the blocks of Π_n, that is let τ be a uniformly distributed random permutation of $\{1, ..., K_n\}$ then for every $j \in \{1, ..., K_n\}$ $N^{ex}_{n,j} = |A_{\tau(j)}|$. Then every $(n_1, ..., n_k) \in \mathcal{C}_n$

$$\begin{aligned}\mathbb{P}[(N^{ex}_{n,1}, ..., N^{ex}_{n,K_n}) = (n_1, ..., n_k)] &= \binom{n}{n_1, ..., n_k} \frac{1}{k!} p(n_1, ..., n_k) \\ &= \left[\frac{n!}{n_1! n_2! \cdots n_k!}\right] \frac{1}{k!} p(n_1, ..., n_k)\end{aligned}$$

To see this recall that $p(n_1, ..., n_k) = p_n(n_1, ..., n_k)$ is the probability is that Π_n equals any particular partition with block sizes $(n_1, ..., n_k)$ in some order. As there are $k!$ distinct ways to order the blocks of the partition, diving by $k!$ gives the probability of obtaining a particular partition of $[n]$ with a particular order of the blocks after randomizing the order of blocks, the multinomial coefficient is the number of such ordered partitions consistent with $(n_1, ..., n_k)$.

Decreasing order

Let $\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$ denote the random composition of n induced by Π_n such that $N^{\downarrow}_{n,i}$ is the size ith largest block, that is

$$n \geq N^{\downarrow}_{n,1} \geq N^{\downarrow}_{n,2} \geq \cdots \geq N^{\downarrow}_{n,K_n} \geq 1$$

Note that for every $(n_1, ..., n_k) \in \mathcal{C}_n$ such that $n_1 \geq n_2 \geq ... \geq n_k$ there are exactly $\dfrac{n!}{\prod_{i=1}^n (i!)^{m_i} m_i!}$ different partitions that induce that particular ordered composition,

where

$$m_i = |\{j : n_j = i, 1 \leq j \leq k\}| = \sum_{j=1}^k \mathbf{1}_{\{n_j = i\}}.$$

For instance, $(5, 3, 2, 1, 1) \in \mathcal{C}_{[12]}$ could be induced by

$$\{\{1, 2, 3, 4, 5\}, \{6, 7, 8\}, \{9, 10\}, \{11\}, \{12\}\}$$

or it could be induced by

$$\{\{1, 3, 7, 8, 5\}, \{2, 6, 9\}, \{4, 12\}, \{10\}, \{11\}\}$$

etcetera.

So for $(n_1, ..., n_k) \in \mathcal{C}_n$ such that $n_1 \geq n_2 \geq ... \geq n_k$

$$\mathbb{P}\left[\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right) = (n_1, ..., n_k)\right] = \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!} p(n_1, ..., n_k). \tag{5.1}$$

Now define, the random variables $M_1, ..., M_n$ where M_j denotes the number of blocks in the random partition that contain exactly j elements. Then M_j can be written in terms of the induced random composition of n and the number of blocks K_n as follows

$$M_j = \sum_{j=1}^{K_n} \mathbf{1}_{\{N_{n,j}=i\}} \quad \text{for } j \in \{1, ..., n\}$$

It is easy to corroborate that given $(M_1, ..., M_n)$ one can determine the induced ordered random composition $\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$ induced by Π_n (without knowing Π_n). Due to this, for every collection of non-negative integers $(m_1, ..., m_n)$ such that $\sum_{j=1}^{n} j m_j = n$ and $\sum_{j=1}^{n} m_j = k$

$$\mathbb{P}\left[(M_1, ..., M_n) = (m_1, ..., m_n)\right] = \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!} p(n_1, ..., n_k) \tag{5.2}$$

where clearly $m_i = \sum_{j=1}^{k} \mathbf{1}_{\{n_j=i\}}$

Size-biased order of the least element

Let $(\tilde{N}_{n,1}, ..., \tilde{N}_{n,K_n})$ denote the random composition of n induced by Π_n with blocks in order of appearance. That is, let A_1 be the block of Π_n that contains the element 1, let A_2 be the block of the random partition such that $min([n] \setminus A_1) \in A_2$, inductively for $j \in \{3, ..., K_n\}$ let A_j be the block of Π_n such that $min([n] \setminus \bigcup_{i=1}^{j-1} A_i) \in A_j$, then $\tilde{N}_{n,j} = |A_j|$ for $j \in \{1, ..., K_n\}$.

For instance, if $\Pi_7 = \big\{\{1,3,6\}, \{5\}, \{4\}, \{2,7\}\big\}$ then $A_1 = \{1,3,6\}$, $A_2 = \{2,7\}$, $A_3 = \{4\}$, $A_4 = \{5\}$, and $(\tilde{N}_{7,1}, \tilde{N}_{7,2}, \tilde{N}_{7,3}, \tilde{N}_{7,4}) = (3,2,1,1)$.

Then for all $(n_1, ..., n_k) \in \mathcal{C}_n$

$$\begin{aligned}&\mathbb{P}[(\tilde{N}_{n,1}, ..., \tilde{N}_{n,K_n}) = (n_1, ..., n_k)] \\ &\quad = \frac{n!}{(n_k)(n_k + n_{k-1}) \cdots (n_k + ... + n_1) \prod_{i=1}^{k}(n_i - 1)!} p(n_1, ..., n_k)\end{aligned}$$

where

$$\frac{n!}{(n_k)(n_k + n_{k-1}) \cdots (n_k + ... + n_1) \prod_{i=1}^{k}(n_i - 1)!}$$

is the the number of partitions of $[n]$ with the pre-described block sizes in order of appearance. Note that $(\tilde{N}_{n,1}, ..., \tilde{N}_{n,K_n})$ is a size-biased random permutation of $\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$, meaning that given $\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$, the blocks appear in the random order in which they would be discovered in a process of simple random sampling without replacement.

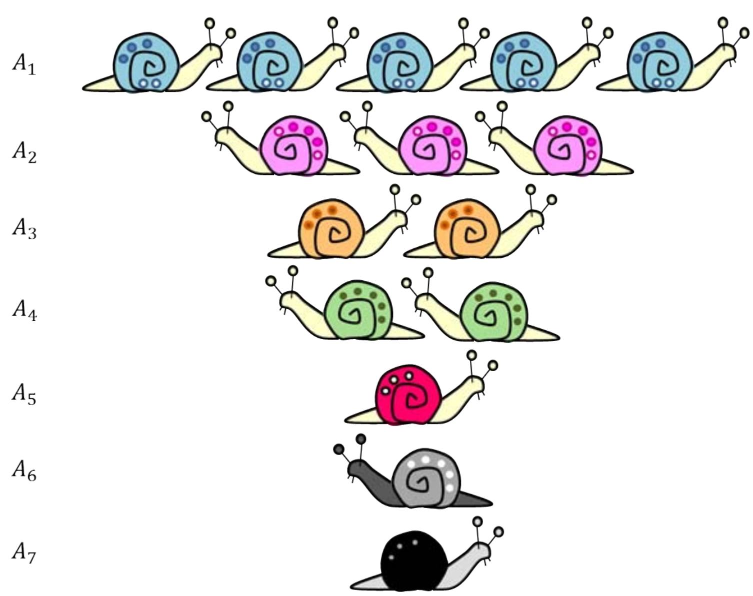
5.1 Partitions generated by sampling without replacement

Finite Case

Let $(x_1, ..., x_n)$ be a finite sequence of numbers. Define $\Pi(x_1, ..., x_n) \in \mathcal{P}_{[n]}$ as the partition generated by $(x_1, ..., x_n)$ that is, the blocks of $\Pi(x_1, ..., x_n)$ are the equivalence classes for the equivalence relation $i \sim j$ if and only if $x_i = x_j$. Now, if $(X_1, ..., X_n)$ is a random vector, it is clear that $\Pi(X_1, ..., X_n)$ is a random partition. Moreover, as we have mentioned before, if $X_1, ..., X_n$ are exchangeable then $\Pi(X_1, ..., X_n)$ is an exchangeable random partition.

First consider the next example

Example 5.1. *Imagine, we have* 15 *snails of* 7 *different types* $A_1, ..., A_7$*. There are* 5 *snails of type* A_1*,* 3 *snails of type* A_2*,* 2 *snails of type* A_3*,* 2 *snails of type* A_4 *and for the types* A_5, A_6 *and* A_7 *we have one snail of each type.*



Suppose that we start sampling snails without replacement from this same group of snails. Let X_i *denote the type of the* i*th snail that is sampled. And define* $\Pi(X_1, ..., X_{15}) \in \mathcal{P}_{[15]}$ *as the partition generated by the equivalence relation*

$i \sim j$ iff X_i and X_j are of the same type. We are interested in the distribution of $\Pi(X_1, ..., X_{15})$.

Let $\mathbf{N}^{\downarrow} = (n_1, ..., n_7) = (5, 3, 2, 2, 1, 1, 1)$ be the vector (in decreasing order) that specifies the number of snails of each type. Now let

$$\pi = \big\{\{1,3\}, \{2,7,11,12,14\}, \{4,5,15\}, \{6\}, \{8\}, \{9,13\}\{10\}\big\}$$

Let us calculate $\mathbb{P}[\Pi(X_1, ..., X_{15}) = \pi]$. In order for $\Pi(X_1, ..., X_{15}) = \pi$ it is necessary that $X_2, X_7, X_{11}, X_{12}, X_{14} \in A_1$, $X_4, X_5, X_{15} \in A_2$, $X_1, X_3 \in A_{\rho(3)}$, $X_9, X_{13} \in A_{\rho(4)}$, $X_6 \in A_{\tau(5)}$, $X_8 \in A_{\tau(6)}$ and $X_{10} \in A_{\tau(7)}$ for some permutation ρ of $\{3, 4\}$ and some permutation τ of $\{5, 6, 7\}$, and where $X_i \in A_j$ means that the ith snail to be sampled is of the jth type. Note that for ρ and τ fixed

$$\mathbb{P}[X_1 \in A_{\rho(3)}] = \frac{2}{15},$$

so

$$\mathbb{P}[X_2 \in A_1, X_1 \in A_{\rho(3)}] = \frac{2}{15}\frac{7}{14}$$

hence

$$\mathbb{P}[X_3 \in A_{\rho(3)}, X_2 \in A_1, X_1 \in A_{\rho(3)}] = \frac{2}{15}\frac{7}{14}\frac{1}{13}$$

inductively

$$\mathbb{P}[E_{\rho,\tau}] = \frac{5!3!2!2!1!1!1}{15!} = \frac{1}{\binom{15}{5,3,2,2,1,1,1}}$$

where $E_{\rho,\tau}$ is the event

$$\{(X_2, X_7, X_{11}, X_{12}, X_{14} \in A_1), (X_4, X_5, X_{15} \in A_2), (X_1, X_3 \in A_{\rho(3)}),$$
$$(X_9, X_{13} \in A_{\rho(4)}), (X_6 \in A_{\tau(5)}), (X_8 \in A_{\tau(6)}), (X_{10} \in A_{\tau(7)})\}$$

As there are $2!$ permutations of $\{3, 4\}$ and $3!$ different permutations of $\{5, 6, 7\}$ then

$$\begin{aligned}
&\mathbb{P}[\Pi(X_1, ..., X_{15}) = \pi] \\
&= \sum_{\rho,\tau} \mathbb{P}[E_{\rho,\tau}] \\
&= \sum_{\rho,\tau} \frac{5!3!2!2!1!1!1!}{15!} \\
&= 2!3!\left[\frac{5!3!2!2!1!1!1}{15!}\right]
\end{aligned}$$

It is easy to see that the random variables in this scenario are exchangeable, and so $\Pi(X_1, ..., X_{15})$ is also exchangeable, so basically for every partition $\pi \in \mathcal{P}_{[15]}$ with 7 blocks of sizes $5, 3, 2, 2, 1, 1, 1$

$$\mathbb{P}[\Pi(X_1, ..., X_{15}) = \pi] = 2!3! \left[\frac{5!3!2!2!1!1!1!}{15!}\right]$$

Generalizing this reasoning, imagine that now we have n snails of k distinct types $A_1, ..., A_k$, with n_j snails of type A_j. Suppose we start sampling from this population without replacement, let X_i denote the type of the ith sampled snail, and set once again $\Pi(X_1, ..., X_n)$ the partition generated by $(X_1, ..., X_n)$ as before in this example. Then it is clear that for every $\pi \in \mathcal{P}_{[n]}$ having k blocks of sizes $n_1, ..., n_k$

$$\begin{aligned}\mathbb{P}[\Pi(X_1, ..., X_n) = \pi] &= \prod_{i=1}^{n} m_i! \left[\frac{n_1! \cdots n_k!}{n!}\right] \\ &= \frac{\prod_{i=1}^{n} (i!)^{m_i} m_i!}{n!}\end{aligned}$$

where $m_i = |\{j : n_j = i\}|$ for $i \in \{1, ..., n\}$.

Let Π_n be an exchangeable partition taking values in $\mathcal{P}_{[n]}$, with EPPF given by p. Let $\mathbf{N}^{\downarrow} = \left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$ the random composition of n (arranged in decreasing order) incuded by Π_n. Then we have for every $(n_1, ..., n_k) \in \mathcal{C}_n$ such that $n_1 \geq n_2 \geq ... \geq n_k$.

$$\mathbb{P}\left[\left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right) = (n_1, ..., n_k)\right] = \frac{n!}{\prod_{i=1}^{n} (i!)^{m_i} m_i!} p(n_1, ..., n_k).$$

where once again $m_i = |\{j : n_j = i\}|$.

Note that given Π_n the value of $\mathbf{N}^{\downarrow}$ is determined, but given $\mathbf{N}^{\downarrow}$ there could be more than one partition on $[n]$ that induced that particular ordered composition, it is of interest to compute the conditional distribution of Π_n given $\mathbf{N}^{\downarrow}$.

If $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ is a partition with k blocks such that $|A_j| = n_{\sigma(j)}$ for some permutation σ of $\{1, ...k\}$, then

$$\mathbb{P}[\Pi_n = \pi, \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] = \mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k)$$

And so

$$\begin{aligned}\mathbb{P}[\Pi_n = \pi | \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] &= \frac{\mathbb{P}[\Pi_n = \pi, \mathbf{N}^{\downarrow} = (n_1, ..., n_k)]}{\mathbb{P}[\mathbf{N}^{\downarrow} = (n_1, ..., n_k)]} \\ &= \frac{\prod_{i=1}^{n} (i!)^{m_i} m_i!}{n!}\end{aligned} \tag{5.3}$$

We now generalize the situation given in Example 5.1. Consider a set of n elements divided into K_n groups where there are $N_{n,j}$ elements in the jth group, let $\mathbf{N}^{\downarrow} = \left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n}\right)$ be the sizes of the groups arranged in decreasing order. Imagine we start sampling without replacement, let $X_1, ..., X_n$ be the obtained sample where X_i corresponds to the ith draw and set $\Pi(X_1, ..., X_n)$ the partition of $[n]$ induced by $(X_1, ..., X_n)$, where $i \sim j$ iff X_i and X_j belong to the same group. In the last example we analysed the particular case when K_n and $\mathbf{N}^{\downarrow}$ were known. Our next aim is to treat the situation when these quantities are unknown, in particular when $\mathbf{N}^{\downarrow}$ is the ordered random composition induced by an exchangeable partition Π_n.

We have already seen in Example 5.1, that given $\mathbf{N}^{\downarrow} = (n_1, ..., n_k)$, for every partition $\pi \in \mathcal{P}_{[n]}$ with k blocks of sizes $n_1, ...n_k$

$$\begin{aligned} \mathbb{P}[\Pi(X_1, ..., X_n) = \pi | \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] &= \prod_{i=1}^{n} m_i! \left[\frac{n_1! \cdots n_k!}{n!} \right] \\ &= \frac{\prod_{i=1}^{n} (i!)^{m_i} m_i!}{n!} \end{aligned} \tag{5.4}$$

where $m_i = |\{j : n_j = i\}|$ for $i \in \{1, ..., n\}$.

Comparing equations (5.4) and (5.3) we get

$$\begin{aligned} &\mathbb{P}[\Pi(X_1, ..., X_n) = \pi, \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \\ &\qquad = \mathbb{P}[\Pi(X_1, ..., X_n) = \pi | \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \mathbb{P}[\mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \\ &\qquad = \mathbb{P}[\Pi_n = \pi | \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \mathbb{P}[\mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \\ &\qquad = \mathbb{P}[\Pi_n = \pi, \mathbf{N}^{\downarrow} = (n_1, ..., n_k)] \\ &\qquad = \mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k) \end{aligned}$$

and therefore we have proved the next proposition.

Proposition 5.1. *Let Π_n be an exchangeable partition of $[n]$, and*

$$\mathbf{N}^{\downarrow} = \left(N^{\downarrow}_{n,1}, ..., N^{\downarrow}_{n,K_n} \right)$$

be the corresponding composition of n defined by decreasingly arranging the block sizes of Π_n. Then the joint law of Π_n and of $\mathbf{N}^{\downarrow}$ is that of $\Pi(X_1, ..., X_n)$ and $\mathbf{N}^{\downarrow}$, where conditionally given $\mathbf{N}^{\downarrow}$ the sequence $(X_1, ..., X_n)$ is defined by simple sampling without replacement from a list $x_1, ..., x_n$ with $N^{\downarrow}_j$ values equal to j for each $1 \leq j \leq K_n$.

Infinite case

Let $\Pi_\infty = \{\Pi_n\}_{n \geq 1}$ be a consistent family of exchangeable random partitions, that is, for every $n \in \mathbb{N}$ Π_n is an exchangeable partition of $[n]$, and for $m < n$, Π_m is the restriction of Π_n to $[m]$. As we have mentioned is Chapter 2.1, Π_∞ can be regarded as an exchangeable random element of the set $\mathcal{P}_{\mathbb{N}}$ of partitions of $\mathbb{N}$, such that Π_n is its restriction to $[n]$.

Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of exchangeable real random variables. In Chapter 2.3 we mentioned de Finetti's representation theorem for exchangeable random variables that took values in a Polish space $\mathbb{X}$ endowed with its corresponding σ-field $\mathcal{X}$, to our aim it suffices to restrict our study to $\mathbb{X} = \mathbb{R}$ and $\mathcal{X} = \mathscr{B}(\mathbb{R})$. According to Bruno de Finetti's theorem $\{X_i\}_{i=1}^{\infty}$ is obtained by sampling without replacement from some random probability distribution F. That is conditionally given F $X_1, X_2, ...$ are independent and identically distributed random variables. To be more explicit if

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}((-\infty, x]) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}$$

then

$$F(x) = \lim_{n\to\infty} F_n(x) \text{ uniformly in } x \text{ almost surely}$$

Let $\Pi(X_i, i \geq 1)$ be the exchangeable partition on $\mathbb{N}$ generated by $(X_1, X_2, ...)$ meaning that its restriction to $[n]$ is $\Pi(X_1, ..., X_n)$. Let $(P_j^{\downarrow}, j \geq 1)$ be the magnitudes of the atoms of F arranged in decreasing order, that is $P_j^{\downarrow}$ is the magnitude of the jth largest atom of F. Note that $1 - \sum_{j\geq 1} P_j^{\downarrow}$ is the magnitude of the continuous component of F, which might be strictly positive and that almost surely each i such that X_i is not an atom of F contributes to a singleton component $\{i\}$ to $\Pi(X_i, i \geq 1)$.

According to the next theorem every infinite exchangeable random partition Π_∞ has the same distribution as one generated this way. In the proposition 5.1 it was stated that every finite exchangeable random partition can be generated by a process of sampling without replacement from some population.

Theorem 5.1 (Kingman's representation)**.** *Let $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ be an exchangeable random partition of $\mathbb{N}$, and let K_n be the number of blocks of Π_n and $(N_{n,j}^{\downarrow}, j \geq 1)$ be the decreasing rearrangement of the block sizes of its blocks sizes, where $N_{n,j}^{\downarrow} = 0$ for every $j > K_n$. Then $\dfrac{N_{n,j}^{\downarrow}}{n}$ has an almost sure limit $P_j^{\downarrow}$ as $n \to \infty$ for each $j \geq 1$. Moreover the conditional distribution of Π_∞ given $(P_j^{\downarrow}, j \geq 1)$ is that of $\Pi(X_1, X_2, ...)$ where $X_1, X_2, ...$ are obtained by sampling without replacement from some distribution F with ranked atoms $(P_j^{\downarrow}, j \geq 1)$.*

Proof:

Suppose without loss of generality that on the same probability space where Π_∞ is defined, there exist a sequence $\{U_i\}_{i\geq 1}$ of i.i.d. uniform $[0,1]$ random variables. Let $X_n = U_i$ if n falls into the ith block of Π_∞ in order of appearance, hence, as Π_∞ is exchangeable, $X_1, X_2, ...$ are exchangeable. Now, considering the restriction Π_n of Π_∞ to $[n]$, let us rename the uniform random variables in the following way: If $(N_{n,j}^{\downarrow}, j \geq 1)$ is the decreasing rearrangement of the block sizes of Π_n and $(A^{\downarrow}_{n,j}, 1 \leq j \leq K_n)$ is a rearrangement of the blocks in such way that $|A^{\downarrow}_{n,j}| = N^{\downarrow}_{n,j}$. Then

$$\hat{U}_{n,j} := U_i \text{ iff there exist } \exists k \in A^{\downarrow}_{n,j} \text{ such that } X_k = U_i.$$

Hence we could consider that Π_∞ is generated by random sampling from F which is uniform almost sure limit (as $n \to \infty$) of

$$F_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq u\}} = \sum_{j \geq 1} \frac{N^{\downarrow}_{n,j}}{n} \mathbf{1}_{\{\hat{U}_{n,j} \leq u\}}$$

By almost sure uniformity of convergence of F_n to F we get that $\frac{N^{\downarrow}_{n,j}}{n}$ must converge almost surely to $P^{\downarrow}_j$ where $P^{\downarrow}_j$ is the magnitude of the jth largest atom of F.

□

In order to make more tractable the proof of the last theorem, we illustrate in the next example the procedure described above.

Example 5.2. *Consider for instance, that for some realization, the restriction of* Π_∞ *to* $[10]$ *is*

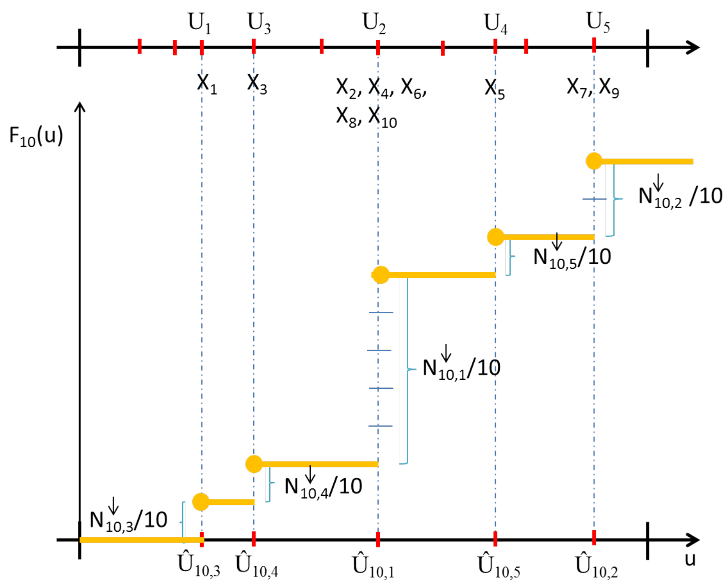
$$\Pi_{10} = \big\{\{1\}, \{2, 4, 6, 8, 10\}, \{3\}, \{5\}, \{7, 9\}\big\}$$

Then $X_1 = U_1$, $X_2, X_4, X_6, X_8, X_{10} = U_2$, $X_3 = U_3$, $X_5 = U_4$ *and* $X_7, X_9 = U_5$. *Rearranging the block sizes of* Π_{10} *in decreasing order we get*

$$\begin{aligned}
(N^{\downarrow}_{10,j}, j \geq 1) &= (5, 2, 1, 1, 1, 0, 0, ...) \\
A^{\downarrow}_{10,1} &= \{2, 4, 6, 8, 10\} \\
A^{\downarrow}_{10,2} &= \{7, 9\} \\
A^{\downarrow}_{10,3} &= \{1\} \\
A^{\downarrow}_{10,4} &= \{3\} \\
A^{\downarrow}_{10,5} &= \{5\}
\end{aligned}$$

and therefore $\hat{U}_{10,1} = U_2$, $\hat{U}_{10,2} = U_5$, $\hat{U}_{10,3} = U_1$, $\hat{U}_{10,4} = U_3$, *and* $\hat{U}_{10,5} = U_4$.

The next figure illustrates the empirical distribution function of $X_1, ..., X_5$ *for this particular realization*



Last theorem sets up a bijection between probability distributions of infinite exchangeable random partitions, as specified by an infinite EPPF, and probability distributions of $(P_j^{\downarrow}, j \geq 1)$ on the set

$$\mathcal{P}_{[0,1]}^{\downarrow} := \left\{ (p_1, p_2, ...) : p_1 \geq p_2 \geq ... \geq 0, \sum_{i=1}^{\infty} p_i \leq 1 \right\} \tag{5.5}$$

of ranked sub-probability distributions of $\mathbb{N}$. This bijection is known in literature as *Kingman's correspondence*. In order to make explicit this correspondence it suffices to note the following:

a) Given a probability distribution of $(P_j^{\downarrow}, j \geq 1)$ on the set $\mathcal{P}_{[0,1]}^{\downarrow}$, we can build a random distribution function F where the sizes of its atoms of in non-decreasing order distribute as $(P_j^{\downarrow}, j \geq 1)$ and the magnitude of its continuous component is $1 - \sum_{j \geq 1} P_j^{\downarrow}$. Now, let $\{X_i\}_{i \geq 1}$ be a sequence of random variables such that

$$X_i | F \overset{i.i.d.}{\sim} F$$

so that $\{X_i\}_{i \geq 1}$ is a sequence of exchangeable random variables. By consecutive sampling, we can generate an exchangeable partition of $\mathbb{N}$, $\Pi(X_1, X_2, ...)$, with a certain (infinite) EPPF, p. As the distribution of the random partition generated is not affected by the locations of the atoms of F or the shape of the continuous component, but only the sizes of the atoms, we have established that to each probability distribution on the set $\mathcal{P}_{[0,1]}^{\downarrow}$ corresponds one unique probability distribution of infinite exchangeable random partitions.

b) Conversely, consider an (infinite) EPPF p and let Π_∞ be some exchangeable partition of $\mathbb{N}$, with such EPPF. Kingman's representation theorem establishes (following the notation of the statement of this theorem) that the almost sure limits,

$$\lim_{n\to\infty} \frac{N_{n,j}^{\downarrow}}{n} = P_j^{\downarrow},$$

exist and that the limiting random variables $(P_j^{\downarrow}, j \geq 1)$ are such that

$$P_1^{\downarrow} \geq P_2^{\downarrow} \geq \cdots$$

and $\sum_{j\geq 1} P_j^{\downarrow} \leq 1$ almost surely. So that the distribution of the limiting random variables $(P_j^{\downarrow}, j \geq 1)$ is some distribution on the set $\mathcal{P}_{[0,1]}^{\downarrow}$. Thus, to each distribution of infinite exchangeable random partitions, corresponds one unique distribution on the set $\mathcal{P}_{[0,1]}^{\downarrow}$.

Example 5.3. *Let F be a random distribution function that is continuous a.s., that is if $(P_j^{\downarrow}, j \geq 1)$ are its ranked atoms then $P_j^{\downarrow} = 0$ a.s. for every $j \geq 1$. Let $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ be the partition of $\mathbb{N}$ derived by sampling from F, that is $\Pi_n = \Pi(X_1, ..., X_n)$. As F is continuous almost surely $\mathbb{P}[X_i = X_j] = 0$ for every $i \neq j$ and so each X_i contributes to a singleton in Π_∞, so*

$$\mathbb{P}[\Pi_n = \big\{\{1\}, \{2\}, ..., \{n\}\big\}] = 1, \quad n \geq 1$$

and the EPPF of Π_∞ takes the form

$$p(n_1, ..., n_k) = \prod_{j=1}^{k} \mathbf{1}_{\{n_i=1\}}.$$

Therefore the distribution over $\mathcal{P}_{[0,1]}^{\downarrow}$ that corresponds to the EPPF defined by the last equation is $P_1^{\downarrow} = P_2^{\downarrow} = ... = 0$ a.s.

Remark: In terms of the distributions of $(P_j^{\downarrow}, j \geq 1)$ on the set $\mathcal{P}_{[0,1]}^{\downarrow}$ there two important scenarios.

- *Proper case:* This scenario happens when $\sum_{j\geq 1} P_j^{\downarrow} = 1$ almost surely.
- *Improper case:* This scenario takes place when

$$\mathbb{P}\left[\sum_{j\geq 1} P_j^{\downarrow} < 1\right] > 0.$$

Note that any random distribution function, F, with sizes of its atoms $(P_j^{\downarrow}, j \geq 1)$ that lay in the proper cases, must be atomic almost surely (this is the case of the NRMIs defined in Chapter (4) when the deterministic measure of the CRM

$\beta \equiv 0$, and $\mathbb{X} = \mathbb{R}$.)

Returning to Kingman's correspondence, it can be made more explicit as follows. If p is the EPPF of an exchangeable random partition of $\mathbb{N}$, then by Kingman's representation it can be thought as the random partition generated by sampling $X_1, X_2, \ldots$ without replacement from a distribution F with ranked frequencies $(P_j^{\downarrow}, j \geq 1)$ where $P_j^{\downarrow}$ is the almost sure limit of $\frac{N_{n,j}^{\downarrow}}{n}$. Assume we stand in the proper case, so $\sum_{i=1}^{\infty} P_i^{\downarrow} = 1$ a.s. Thus $p(n_1, \ldots, n_k)$ (where $\sum_{i=1}^{k} n_i = n$) is the probability that a sample of size n exhibits k different values with frequencies $(n_1, \ldots, n_k)$, as the random variables are i.i.d given F then

$$p(n_1, \ldots, n_k) = \sum_{(j_1, \ldots, j_k)} \mathbb{E}\left[\prod_{i=1}^{k} (P_{j_i}^{\downarrow})^{n_i}\right]$$

where $(j_1, \ldots, j_k)$ ranges over all ordered k-tuples of distinct positive integers. Note that for any rearrangement $(P_j, j \geq 1)$ of $(P_j^{\downarrow}, j \geq 1)$ we also have

$$p(n_1, \ldots, n_k) = \sum_{(j_1, \ldots, j_k)} \mathbb{E}\left[\prod_{i=1}^{k} (P_{j_i})^{n_i}\right] \tag{5.6}$$

as the right side of the formula is an expectation of a function of $(P_1, P_2, \ldots)$, same that is invariant under finite or infinite permutation of its arguments.

Let us consider more examples illustrating Kingman's correspondence:

Example 5.4 (Sampling from exchangeable frequencies)**.** *Let $(P_1^{\downarrow}, \ldots, P_m^{\downarrow})$ be a sequence of random ranked frequencies such that $\sum_{i=1}^{m} P_i^{\downarrow} = 1$ a.s. for some $m < \infty$. Let $\Pi_\infty = \{\Pi_n\}_{n \geq 1}$ be the partition of $\mathbb{N}$ obtained by sampling from a distribution F such that the ranked sizes of its atoms are $(P_1^{\downarrow}, \ldots, P_m^{\downarrow})$. Note that as $\sum_{i=1}^{m} P_i^{\downarrow} = 1$ a.s. then for every $n \geq 1$ $K_n \leq m$ a.s. where K_n denotes the number of blocks of Π_n as usual. By equation (5.6) and the argument above*

$$p(n_1, \ldots, n_k) = \begin{cases} 0 & \text{if } k > m \\ \sum_{(j_1, \ldots, j_k)} \mathbb{E}\left[\prod_{i=1}^{k} (P_{j_i}^{\downarrow})^{n_i}\right] & \text{if } k \leq m \end{cases}$$

Note that if $k \leq m$ then, we can simplify the form of the EPPF as follows, let $(P_1, \ldots, P_m)$ be the exchangeable sequence obtained by putting this random frequencies in exchangeable random order, then

$$\begin{aligned} p(n_1, \ldots, n_k) &= \binom{m}{k} k! \mathbb{E}\left[\prod_{i=1}^{k} (P_i)^{n_i}\right] \\ &= \frac{m!}{(m-k)!} \mathbb{E}\left[\prod_{i=1}^{k} (P_i)^{n_i}\right] \\ &= (m)_{k \downarrow 1} \mathbb{E}\left[\prod_{i=1}^{k} (P_i)^{n_i}\right] \end{aligned}$$

where $(r)_{j\downarrow c} = (r)(r-c)(r-2c) * \cdots * (r-(j-1)c)$

Example 5.5 (Coupon's Collecting)**.** *Let* $(P_1^{\downarrow}, ..., P_m^{\downarrow})$ *be a sequence of random ranked frequencies such that* $P_i^{\downarrow} = \frac{1}{m}$ *a.s. for* $1 \leq i \leq m$ *and some* $m < \infty$*. Note that this is a particular case of the last example. Any rearrangement* $(P_1, ..., P_m)$ *of the ranked sequence is an exchangeable sequence, hence the corresponding EPPF is*

$$p(n_1, ..., n_k) = \begin{cases} 0 & \text{if } k > m \\ \frac{(m)_{k\downarrow 1}}{m^n} & \text{if } k \leq m, \quad n := \sum_{i=1}^k n_i \end{cases}$$

Example 5.6 (Sampling from exchangeable Dirichlet frequencies)**.** *Consider once again the scenario of Example 5.4, if* $(P_1, ..., P_m)$ *has a symmetric Dirichlet distribution, that is the density of* $(P_1, ..., P_m)$ *is given by equation (4.2) with parameters* $\alpha_1 = ... = \alpha_m = \alpha$ *then*

$$p(n_1, ..., n_k) = \begin{cases} 0 & \text{if } k > m \\ (m)_{k\downarrow 1} \mathbb{E}\left[\prod_{i=1}^k (P_i)^{n_i}\right] & \text{if } k \leq m \end{cases}$$

it is easy to compute the above expectation in this scenario, let $n_i = 0$ *for every* $i > k$ *then*

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^k (P_i)^{n_i}\right] &= \mathbb{E}\left[\prod_{i=1}^m (P_i)^{n_i}\right] \\
&= \int \cdots \int \prod_{i=1}^m (p_i)^{n_i} \left[\frac{\Gamma(m\alpha)}{\Gamma(\alpha)^m} \prod_{i=1}^m p_i^{\alpha-1}\right] dp_1 \cdots dp_{m-1} \\
&= \frac{\prod_{i=1}^m (\alpha)_{n_i\uparrow 1}}{(m\alpha)_{n\uparrow 1}} \int \cdots \int \frac{\Gamma(m\alpha)(m\alpha)_{n\uparrow 1}}{\prod_{i=1}^m \Gamma(\alpha)(\alpha)_{n_i\uparrow 1}} \prod_{i=1}^m p_i^{\alpha+n_i-1} dp_1 \cdots dp_{m-1} \\
&= \frac{\prod_{i=1}^m (\alpha)_{n_i\uparrow 1}}{(m\alpha)_{n\uparrow 1}} \int \cdots \int \frac{\Gamma(m\alpha+n)}{\prod_{i=1}^m \Gamma(\alpha+n_i)} \prod_{i=1}^m p_i^{\alpha+n_i-1} dp_1 \cdots dp_{m-1} \\
&= \frac{\prod_{i=1}^m (\alpha)_{n_i\uparrow 1}}{(m\alpha)_{n\uparrow 1}} \\
&= \frac{\prod_{i=1}^k (\alpha)_{n_i\uparrow 1}}{(m\alpha)_{n\uparrow 1}}
\end{aligned}$$

where $n := \sum_{i=1}^m n_i$ *and* $(\alpha)_{0\uparrow 1} := 1$*, therefore*

$$p(n_1, ..., n_k) = \begin{cases} 0 & \text{if } k > m \\ (m)_{k\downarrow 1} \frac{\prod_{i=1}^k (\alpha)_{n_i\uparrow 1}}{(m\alpha)_{n\uparrow 1}} & \text{if } k \leq m \end{cases}$$

5.2 Structural distributions

Let us start with the following defintion.

Definition 5.1. *Let $(P_1, P_2, ...)$ be a random discrete probability (i.e. $P_j \geq 0$ a.s. for every $j \geq 1$ and $\sum_j P_j = 1$ a.s.). Its size-biased permutation is the same sequence in a random order $(P_{\sigma(1)}, P_{\sigma(2)}, ...)$ where σ is a permutation of $[n]$ such that*

$$\mathbb{P}[\sigma(1) = j] = P_j$$

and for $k \geq 2$ and distinct numbers $j_1, ..., j_k$

$$\mathbb{P}[\sigma(k) = j_k | \sigma(1) = j_1, ..., \sigma(k-1) = j_{k-1}] = \frac{P_{j_k}}{1 - \sum_{i=1}^{k-1} P_{j_i}}.$$

Note that, in the last definition a number j with bigger corresponding probability P_j tends to appear earlier in the permutation, hence the name size-biased.

Let $\{P_j\}_{j\geq 1}$ be a random discrete probability with size-biased permutation $\{\tilde{P}_j\}_{j\geq 1}$. In particular

$$\tilde{P}_1 = P_{\sigma(1)} \quad \text{ where } \mathbb{P}[\sigma(1) = j] = P_j \quad j \in \{1, 2, ...\}$$

The random variable $\tilde{P}_1$ is called size-biased pick from $\{P_j\}_{j\geq 1}$. Let $\tilde{\nu}$ be the distribution of $\tilde{P}_1$ on $(0, 1]$, this distribution is known in the literature as *structural distribution* associated to the random discrete probability distribution $\{P_j\}_{j\geq 1}$.

Let Π_∞ be a random partition derived by sampling from $\{P_j\}_{j\geq 1}$, then the size-biased permutation $\{\tilde{P}_j\}_{j\geq 1}$ can be constructed as the sequence of class frequencies of Π_∞ in order of appearance, it is very easy to see this.

Consider an urn with a numerable number of coloured balls, the number of distinct colours is unknown as well as the proportion of balls with certain colour. Say we number the distinct colours and let P_i be the proportion of balls that have the ith colour. Then we sample without replacement balls from this urn, and we are going to rename the colours in other of appearance in the following way. Let Π_∞ be the infinite partition derived from this procedure, let $\sigma(1)$ be the first sampled colour, $\sigma(2)$ the second sampled coloured and so on, for instance, if the first ball we extract is red and in the original naming red was assigned the number 4 then $\sigma(1) = 4$. Let $\{\tilde{P}_j = P_{\sigma(j)}\}_{j\geq 1}$ be the proportions of the balls having certain colour according to the new naming. Then it is clear that

$$\mathbb{P}[\sigma(1) = j] = \mathbb{P}[\text{ The first ball dropped has colour } j] = P_j \quad j \in \{1, 2, ...\}$$

and also for $i \geq 2$

$$\mathbb{P}[\sigma(k) = j_k | \sigma(1) = j_1, ..., \sigma(k-1) = j_{k-1}] = \frac{P_{j_k}}{1 - \sum_{i=1}^{k-1} P_{j_i}}.$$

Example 5.7. *In this example we are going to assume that the number of colours is infinite and the proportion of balls having each colour is known. Assume we have an urn with an infinite number of balls such that $\frac{1}{2}$ of the balls are blue, $\frac{1}{4}$ are red, $\frac{1}{8}$ are purple, $\frac{1}{16}$ are orange and the rest of the balls ($\frac{1}{16}$ of them) have a unique distinct colour. Say we number the colours in the following way, blue $= 1$, purple $= 2$, red $= 3$, orange $= 4$, yellow $= 5$, green $= 6$, and so on. Imagine we start sampling without replacement from this urn. Let X_i be the colour of the ball obtained in the ith extraction and Π_∞ the partition of $\mathbb{N}$ derived from this procedure. Say that the first 20 extractions are*

$$X_1, X_3, ..., X_9, X_{19} = 1, \quad X_2, X_{11} = 4, \quad X_{10}, X_{13}, X_{14}, X_{17} = 3,$$

$$X_{12}, X_{15}, X_{18} = 2, \quad X_{16} = 7, \quad X_{20} = 19.$$

Then

$$\Pi_{20} = \big\{\{1, 3, ..., 9, 19\}, \{2, 11\}, \{10, 13, 14, 17\}, \{12, 15, 18\}, \{16\}, \{20\}\big\}$$

with blocks in order of appearance. Now we rename the colours according to the order in wich they appear, let $\sigma(i)$ be the ith colour to appear, then we have

$$\sigma(1) = 1, \quad \sigma(2) = 4, \quad \sigma(3) = 3, \quad \sigma(4) = 2, \quad \sigma(5) = 7, \quad \sigma(6) = 19$$

let $\tilde{P}_j = P_{\sigma(j)}$ the proportion of the balls having the jth colour to appear. Then for this particular case it is clear that the structural distribution is

$$\mathbb{P}[\tilde{P}_1 = P_i] = \frac{1}{2}\mathbf{1}_{\{i=1\}} + \frac{1}{8}\mathbf{1}_{\{i=2\}} + \frac{1}{4}\mathbf{1}_{\{i=3\}} + \frac{1}{16}\mathbf{1}_{\{i=4\}} + \frac{1}{16}\mathbf{1}_{\{i\notin\{1,2,3,4\}\}}$$

Note that $\tilde{P}_1$ is the frequency of the class of Π_∞ containing 1. The structural distribution, as we are about to see, encodes much information about the entire sequence of frequencies. Let g be an arbitrary non-negative measurable function, then

$$\int_0^1 g(p)\tilde{\nu}(dp) = \mathbb{E}[g(\tilde{P}_1)] = \mathbb{E}[\mathbb{E}[g(\tilde{P}_1)|(P_1, P_2, ...)]] = \mathbb{E}\left[\sum_i P_i g(P_i)\right]$$

taking $g(p) = \frac{f(p)}{p}$ for an arbitrary non-negative measurable function f we get

$$\mathbb{E}\left[\sum_i f(P_i)\right] = \mathbb{E}\left[\sum_i P_i \frac{f(P_i)}{P_i}\right] = \mathbb{E}\left[\frac{f(\tilde{P}_1)}{\tilde{P}_1}\right] = \int_0^1 \frac{f(p)}{p}\tilde{\nu}(dp)$$

Due to the above equation, particular choices of f contribute to the description of some particular random partition models. See [18] in the bibliography for further discussion about Structural distributions.

5.3 The ordered paintbox

Here we provide a construction that might help us understand the connection between exchangeable partitions, sequences of exchangeable random variables and completely random measures, particularly subordinators. This construction is just a more explicit way of looking at Kingsman's correspondence.

Let $\mathcal{R}$ be a random closed subset of $[0,1]$. The open complement set $\mathcal{R}^c := [0,1] \setminus \mathcal{R}$ has a canonical representation as a disjoint union of countably many open interval components, which we are going to call the *gaps* of $\mathcal{R}$. That is $\mathcal{R}^c = \bigcup_{j=1}^{\infty} R_j$ where R_j is an open interval of $[0,1]$, and $R_i \cap R_j = \emptyset$ for every $i \neq j$. If this representation turns out to be the union of m open intervals, just set $R_j = \emptyset$ for every $j > m$.

Imagine we have a countable number of uncoloured balls, and we decide to color them according to the next procedure. Let $U_1, U_2, ...$ be Uniform independent random variables which are also independent of $\mathcal{R}$ and assume that to each gap of $\mathcal{R}$ we assign a different colour. Now if U_i falls into R_j we are going to paint the ith ball with the color j (previously assigned to R_j), if on the other hand U_i falls into $\mathcal{R}$ we are going to paint the ith ball with a unique color different from the colors assigned to the gaps and also different from any other previously used color. Note that the colors of the balls generate an exchangeable partition Π_∞ of $\mathbb{N}$ by the equivalence relation $i \sim k$ iff the ith ball and the kth ball to be painted have the same colour. It is clear that if U_i falls into $\mathcal{R}$ the ith ball will contribute to a singleton in Π_∞.

Let P_j be the length of R_j so that $\mathbb{P}[U_i \in R_j] = P_j$, and note that the P_j's can be thought as the sizes of the atoms of some random distribution function F. So that if we sample $X_1, X_2, ...$ from F and consider $\Pi(X_1, X_2, ...)$, then $\Pi(X_1, X_2, ...)$ has the same distribution as Π_∞ generated by the coloured balls given the sequence $\{P_i\}_{i \geq 1}$. Realize once again that the distribution of the exchangeable partition only depends on the sizes of the atoms and not in the form of the continuous part of F. Let $P_1^\downarrow, P_2^\downarrow, ...$ be the ranked lengths of the gaps. It is easy to see that each random closed subset $\mathcal{R}$ of $[0,1]$ defines a distribution over $\mathcal{P}_{[0,1]}^\downarrow$ (as in (5.5)) by the previous construction. Hence, the ordered paintbox is provides another way of understanding Kingman's correspondence.

If $\mathcal{R}$ has a lebesgue measure zero a.s. then we fall into the *proper case*, that is $\sum_{i=1}^{\infty} P_i^\downarrow = 1$ a.s. On the other hand, if there is a positive probability that the lebesgue measure of $\mathcal{R}$ is greater than zero, then we stand in the *improper case*. Usually $\mathcal{R}$ is referred as the paintbox.

The ordered paintbox and subordinators

Recall that a subordinator $\{\phi(t)\}_{t \in \mathbb{R}}$ is a stochastic process such that

- $\phi(0) = 0$ a.s.

- It has independent and stationary increments.

- Its trajectories are non-decreasing, right-continuous and their left limit exists.

We have seen in Chapter 3 that we can write

$$\phi(t) = \begin{cases} \beta t + \sum\{z : (x,z) \in \Xi^*, 0 < x \leq t\} & \text{if } t \geq 0 \\ \beta t - \sum\{z : (x,z) \in \Xi^*, t \leq x < 0\} & \text{if } t < 0 \end{cases}$$

where β is a non-negative constant and Ξ^* a Poisson Process on the half plane $S^* = \{(x,z) : z > 0\}$.

Let us concentrate in the restriction of ϕ to $\mathbb{R}^+$ that is, in the process $\{\phi(t)\}_{t\geq 0}$. Now, let $\varsigma := \inf\{t : \phi(t) = \infty\}$ (where will consider $\inf(\emptyset) = \infty$) and

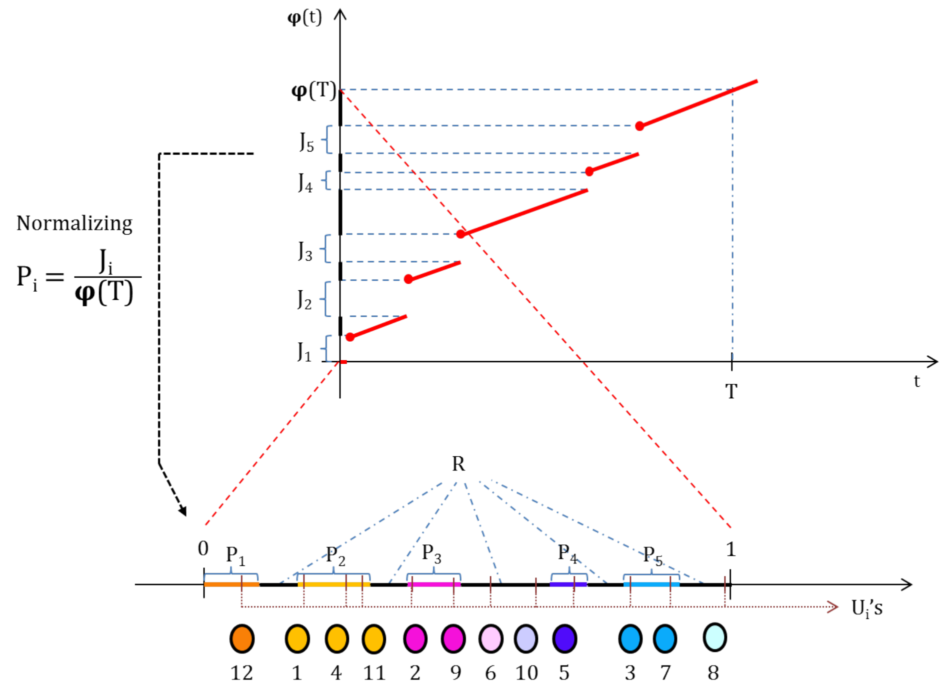
$$\mathscr{R} = \overline{\{\phi(t) : 0 \leq t < \varsigma\}}$$

where $\overline{A}$ denotes the closure of the set A. $\mathscr{R}$ above is called the *range* of ϕ, it is clear by definition that $\mathscr{R}$ is a closed subset of $[0,\infty)$. Let τ be a stopping time and consider the set $\mathscr{R}_\tau = [0,\phi(\tau)] \cap \mathscr{R}$, then this set is a closed subset of $[0,\phi(\tau)]$. Let $\mathscr{R}_\tau^c = [0,\phi(\tau)] \setminus \mathscr{R}_\tau$ be the open component of $\mathscr{R}_\tau$ with canonical representation

$$\mathscr{R}_\tau^c = \bigcup_{j=1}^{\infty} \mathscr{R}_j$$

where $\mathscr{R}_j$ is open interval and $\mathscr{R}_j \cap \mathscr{R}_i = \emptyset$ whenever $i \neq j$. Let J_i be the lebesgue measure of $\mathscr{R}_i$, then it is clear that $\{J_i\}_{i=1}^\infty$ are the jumps of the subordinator up to time τ. Now, consider the transformation $G : [0,\phi(\tau)] \to [0,1]$ given by $G(t) = \frac{t}{\phi(\tau)}$, that is, we normalize the set $[0,\phi(\tau)]$, then $\mathcal{R} = G(\mathscr{R}_\tau)$ is a random closed subset of $[0,1]$, hence $\mathcal{R}$ could be thought as a paintbox. Define, once again, $\mathcal{R}^c = [0,1] \setminus \mathcal{R}$ and let $\bigcup_{j=1}^\infty R_j$ be its canonical representation. It is clear, then, that (by possibly renaming the open intervals) $R_j = G(\mathscr{R}_j)$ hence the lebesgue measure of R_j will be given by $P_j = \frac{J_j}{\phi(T)}$. Now given the random closed subset $\mathcal{R}$ of $[0,1]$ we can generate a random exchangeable partition, by the procedure described above.

This construction is illustrated in the image bellow.



It is easily seen that if $\beta = 0$, for every stopping time τ, $\mathscr{R}_\tau$ will have a lebesgue measure zero a.s. thus the derived paintbox $\mathcal{R}$ will also have lebesgue measure zero a.s. Therefore we stand in the *proper* scenario, i.e. $\sum_{i=1}^{\infty} P_i = 1$ a.s.

Two very important and illustrative subordinators are the *Moran Gamma* subordinator and the *stable* subordinator which we introduced in past chapters and will continue to discuss on the following.

Chapter 6

Sequential construction of random partitions

The main aim of this chapter is to provide a sequential construction of random partitions, we start with a simple model and later on we will generalize it.

6.1 The chinese restaurant process

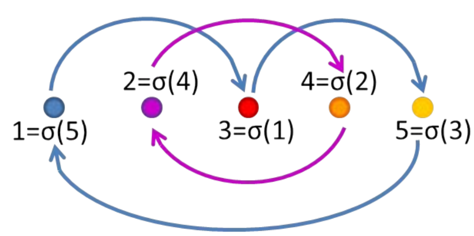
Definition 6.1 (Consistent random permutations)**.** *Consider a sequence of random permutations $\{\sigma_n\}_{n\geq 1}$ such that*

a) σ_n is a uniformly distributed random permutation of $[n]$, for every $n \geq 1$

b) If σ_n is written as a product of cycles, then σ_{n-1} is derived by the deletion of the element n from its cycle.

Example 6.1. *Consider the permutation σ_5 of the set $[5]$ as follows*

$$\sigma(1) = 3, \quad \sigma(2) = 4, \quad \sigma(3) = 5, \sigma(4) = 2, \sigma(5) = 1$$



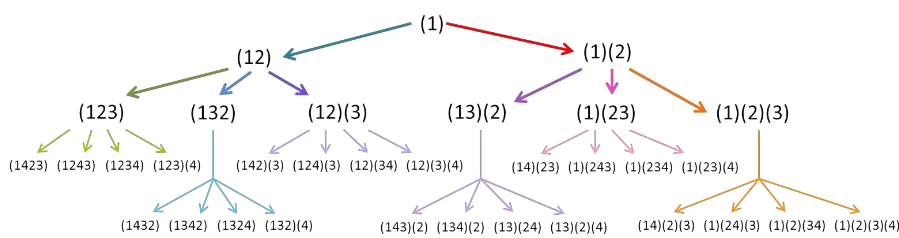
As shown in the picture σ has two cycles $1 \to 3 \to 5$ and $2 \to 4$, so using standard cycle notation for permutations, writing σ_5 as a product of its cycles *we obtain* $\sigma_5 = (135)(24)$. *Let* $\{\sigma_n\}_{n\geq 1}$ *be a family of random consistent permutations, then given* σ_5 *equal to the one described above, in order for the family to be consistent, it is necessary by* (b) *that*

$$\sigma_4 = (13)(24), \quad \sigma_3 = (13)(2), \quad \sigma_2 = (1)(2), \quad \sigma_1 = (1)$$

so σ_4 *is obtained by deleting the element* 5 *from* σ_5*,* σ_3 *can be derived by the deletion of the element* 4 *in* σ_4*, and so on. By* (a) *it would also be necessary that given* σ_5

$$\sigma_6 = \begin{cases} (1635)(24) & \textit{with probability } \frac{1}{6} \\ (1365)(24) & \textit{with probability } \frac{1}{6} \\ (1356)(24) & \textit{with probability } \frac{1}{6} \\ (135)(264) & \textit{with probability } \frac{1}{6} \\ (135)(246) & \textit{with probability } \frac{1}{6} \\ (1356)(24)(6) & \textit{with probability } \frac{1}{6} \end{cases}$$

In general a consistent random permutation evolves in the next way. Consider the next tree where each vertex in the graph stands for a permutation written as the product of its cycles



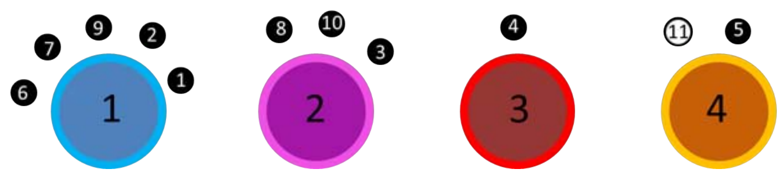
Note that $\sigma_1 = (1)$ a.s. Now for $n \geq 2$ to obtain σ_n given σ_{n-1} we choose uniformly between one of the arrows that have its origin at σ_{n-1} and set σ_n equal to the permutation of $[n]$ that is at the end of the chosen arrow.

Hence, it is easily seen that conditions (a) and (b) in the definition of a consistent random permutation determine one unique distribution for the sequence $\{\sigma_n\}_{n\geq 1}$, as follows.

Consider an initially empty restaurant with an unlimited number of tables where each table has capacity for an unlimited number of customers. Numbered customers arrive one by one to the restaurant. The first customer to arrive seats at table 1, the second customer chooses to sit either to the right of the first customer with probability $\frac{1}{2}$ or at table 2 with probability $\frac{1}{2}$. The third customer decides, uniformly, to seat either to the right of customer 1, to the right of customer 2 or at an unoccupied table. Inductively, given that are already n customers seated at k different tables, the $n+1$th customer chooses with equal probability to sit at one of the following $n+1$ places: to the right of customer

j for some $1 \leq j \leq n$ or at a table $k+1$, same that has not yet been occupied. Define $\sigma_n : [n] \to [n]$ as follows, if after exactly n customers have entered, the ith and the jth are seated in the same table and the customer number j is seated to the immediate right of the customer number i then $\sigma_n(i) = j$, on the other side if the customer number i is seated at a table alone then we set $\sigma_n(i) = i$. Let Π_n the partition of $[n]$ determined by the cycles of σ_n i.e. if i and j belong to the same cycle of σ_n (that is the ith and the jth customer are seated at the same table) then i and j belong to the same block of Π_n.

Example 6.2. *Imagine that after the arrival of* 11 *customers their configuration at the chinese restaurant looks as shown in the next picture*



Then

$$\begin{aligned}
\sigma_1 &= (1)\\
\sigma_2 &= (12)\\
\sigma_3 &= (12)(3)\\
\sigma_4 &= (12)(3)(4)\\
&\vdots\\
\sigma_{10} &= (1\,2\,9\,7\,6)(3\,8\,10)(4)(5)\\
\sigma_{11} &= (1\,2\,9\,7\,6)(3\,8\,10)(4)(5\,11)
\end{aligned}$$

and

$$\begin{aligned}
\Pi_1 &= \big\{\{1\}\big\}\\
\Pi_2 &= \big\{\{1,2\}\big\}\\
\Pi_3 &= \big\{\{1,2\},\{3\}\big\}\\
\Pi_4 &= \big\{\{1,2\},\{3\}\{4\}\big\}\\
&\vdots\\
\Pi_{10} &= \big\{\{1,2,6,7,9\},\{3,10,8\}\{4\}\{5\}\big\}\\
\Pi_{11} &= \big\{\{1,2,6,7,9\},\{3,10,8\}\{4\}\{5,11\}\big\}
\end{aligned}$$

Many asymptotic properties of consistent random permutations can be read from this construction, for instance if K_n denotes the number of occupied tables after the nth customer arrived then

$$K_n = |\{\text{cycles of } \sigma_n\}| = |\{\text{blocks of } \Pi_n\}| = \sum_{i=1}^{n} \mathbf{1}_{A_i}$$

where A_i is the event that the ith customer seats at a new table. As $\mathbb{P}[A_i] = \frac{1}{i}$, $\mathbf{1}_{A_i} \sim Bernouilli\left(\frac{1}{i}\right)$, and so

$$\frac{K_n}{\log(n)} \to 1 \text{ almost surely}$$

as will be later proved in a more general scenario.

Π_n is an exchangeable random partition and is consistent as n varies, the exchangeability follows from the fact that σ_n is uniformly distributed over the set of permutations of $[n]$, and the consistency of Π_n as n varies follows directly from the consistency of σ_n as n varies. Thus $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ is an exchangeable partition of $\mathbb{N}$. Let $X_n = \mathbf{1}_{B_n}$ where B_n is the event that the $(n+1)$th person sits at table 1 and $S_n := \sum_{i=1}^n X_i$. Then for $n \geq 0$

$$\mathbb{P}[X_{n+1} = 1|X_1, ..., X_n] = \frac{|\{\text{persons sitting at table 1}\}|}{n+2} = \frac{1+S_n}{n+2},$$

this equation is the same as the equation (A.1), in the appendix section, for $a = b = m = 1$.

If we consider a urn containing one white ball and one black ball, at each time n we draw a ball form the urn and return it together with an additional ball having the same color and let Y_i be the indicator function of the event that the ith drawn ball is white, then $X_1, X_2, ...$ follow the same distributional law of $Y_1, Y_2, ...$ Thus $\frac{S_n}{n} \to \rho$ a.s. as $n \to \infty$ where $\rho \sim Beta(1,1)$, i.e. ρ is uniformly distributed of $[0,1]$.

Now $S_n + 1$ denotes the size of the cycle of σ_{n+1} containing the element 1 and therefore the size of the block of Π_{n+1} containing the element 1. Thus the frequency of the class of Π_∞ containing the element 1, $\tilde{P}_1$ is $\lim_{n\to\infty} \frac{S_n+1}{n+1} = \rho$, so we have proved that the structural distribution associated to Π_∞, same that we discussed in Chapter 5.2, is a uniform distribution over $[0,1]$.

Also note that equation (A.3), for $a = b = m = 1$ reduces to the form

$$\mathbb{P}[S_n = k] = \binom{n}{k}\frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1} \quad k \in \{0, 1, ..., n\}$$

so S_n distributes uniformly on the set $\{0, 1, ..., n\}$, hence $S_n + 1$ distributes uniformly over $[n]$

The limit frequencies

Let $(N_{n,1}, ...N_{n,K_n})$ be the frequencies of the blocks of Π_n in order of the least element. In terms of the restaurant construction $N_{n,k}$ denotes the number of customers sitting at the kth table after n customers arrived. From above $N_{n,1} = S_n + 1$ distributes uniformly on the set $[n]$. Similarly given $N_{n,1} = n_1 < n$, $N_{n,2}$

distributes uniformly on the set $[n - n_1]$; given $N_{n,1} = n_1$, $N_{n,2} = n_2$ with $n_1 + n_2 < n$, $N_{n,3}$ distributes uniformly in $[n - (n_1 + n_2)]$, and so on. It is quite obvious the asymptotic behaviour of the relative sizes frequencies of the blocks of Π_n, $(N_{n,1}/n, ...N_{n,K_n}/n)$, which are in size-biased random other, converge in distribution (as $n \to \infty$) to the so called *stick breaking* sequence

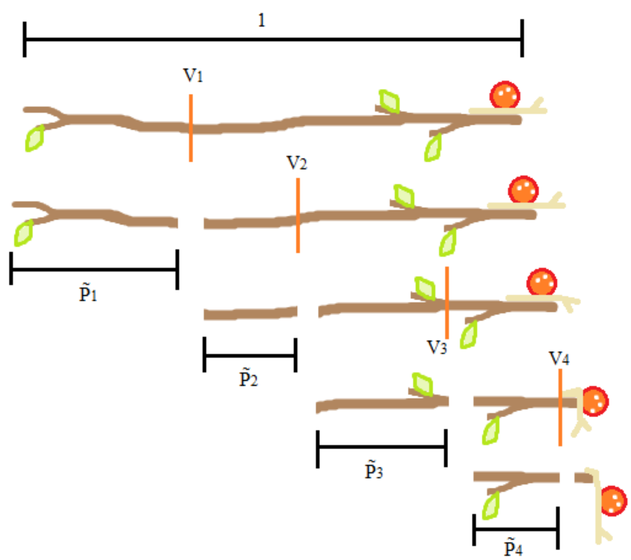
$$(\tilde{P}_1, \tilde{P}_2, \tilde{P}_3, ...) = (U_1, \overline{U_1}U_2, \overline{U}_1\overline{U}_2U_3, ...)$$

where $U_1, U_2, ...$ are independent $Uniform(0, 1)$ random variables and $\overline{U}_i := 1 - U_i$ for every $i \geq 1$.

The name *stick breaking* comes from the fact that this distribution can be obtained as follows: Consider a stick of size 1

1. Break the stick at uniformly distributed (over the points of the stick) random point V_1 and name the bit on the left $\tilde{P}_1$.
2. Let V_2 be a random variable that distributes uniformly over the points of the right-side bit of the stick, break this remaining bit at V_2 and once again name the left bit $\tilde{P}_2$,
3. Take the remaining piece of the stick and break it a uniformly distributed random point, name the left-side bit $\tilde{P}_3$.

And so on.



It is clear that in this case the frequencies are proper, since it is obvious from the stick-breaking construction

$$\sum_{j=1}^{\infty} \tilde{P}_j = 1$$

Now considering $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ the partition of $\mathbb{N}$ derived from the chinese restaurant by a simple combinatorial argument it is easy to realize that the EPPF of π_∞ is given by

$$p(n_1, ..., n_k) = \frac{1}{n!}\prod_{i=1}^{k}(n_i - 1)!$$

see Theorem 6.2 for a proof of this statement.

By the form of the EPPF and Kingman's correspondence, one can realize that we are taking about the partition generated by the Dirichlet process (with parameter $\theta = 1$) which we constructed in Chapter 4.1 by Normalizing the Gamma Process. Thus, putting together the work developed in this chapter with the theory established in Chapter 4.1 we can see that the continuous uniform stick-breaking sequence $(\tilde{P}_1, \tilde{P}_2, ...)$ has the same distribution as a size-biased permutation of the jumps of the Dirichlet process with exchangeable increments

$$\left(\frac{\phi(u)}{\phi(1)}, 0 \leq u \leq 1\right)$$

where $\{\phi(u)\}_u$ is the Moran Gamma process.

Generalization

Our aim now is to generalize the chinese restaurant construction in such way that the partition of $\mathbb{N}$ associated to the construction coincides in distribution with an already specified exchangeable random partition of $\mathbb{N}$.

Recall that the distribution of a exchangeable partition $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ of $\mathbb{N}$ is specified by means of its EPPF p, where $p(n_1, ..., n_k)$ is a symmetric function of its arguments and if $n = \sum_{j=1}^{n} n_j$ then $p(n_1, ..., n_k) = \mathbb{P}[\Pi_n = \pi]$ where π is any partition of $[n]$ having k blocks with corresponding sizes $n_1, ..., n_k$. Also, Π_∞ is a partition of $\mathbb{N}$ then the consistency property must hold, which reduces to the addition rule

$$p(n_1, ..., n_k) = p(n_1, ..., n_k, 1) + \sum_{j=1}^{k} p(n_1, ...n_{j-1}, n_j + 1, n_{j+1}..., n_k)$$

hence

$$1 = \frac{p(n_1, ..., n_k, 1)}{p(n_1, ..., n_k)} + \sum_{j=1}^{k}\frac{p(n_1, ...n_{j-1}, n_j + 1, n_{j+1}..., n_k)}{p(n_1, ..., n_k)} \tag{6.1}$$

Now let us get back to the chinese restaurant construction. Assume that now the seating plan for the customers is as follows. The first customer to arrive is seated at table 1. The second customer to arrive is seated at table 1, next to the first customer, with probability $\frac{p(2)}{p(1)} = p(2)$ or at a new table with probability $\frac{p(1,1)}{p(1)} = p(1,1)$, as a consequence of the addition rule it is clear that $p(1,1) = 1 - p(2)$ so the above is well defined. Now given that $n \geq 1$ customers have arrived and there are k occupied tables where n_j is the number of persons sitting at the jth table $(1 \leq j \leq k)$ then, the next person to arrive will be seated

- at the occupied table j with probability

$$\frac{p(n_1, ...n_{j-1}, n_j + 1, n_{j+1}..., n_k)}{p(n_1, ..., n_k)}, \quad 1 \leq j \leq k,$$

- or at a new table with probability

$$\frac{p(n_1, ..., n_k, 1)}{p(n_1, ..., n_k)}$$

by equation (6.1), this seating plan is well defined.

Note that the partition Π_∞ generated by this seating plan (that is if after arrival of the n customer, the ith person to arrive and the jth person to arrive are sitting in the same table then i and j will belong to the same block of Π_n) will be exchangeable by construction and its EPPF will be given by p which is a symmetric function of its arguments.

We call the *prediction rule* (in terms of the restaurant construction) to

$$\mathbb{P}[\text{the } n+1 \text{ costumer sits at table } j | \text{ the configuration after the } n\text{th arrival}]$$

for $j \in \{1, ..., k+1\}$ where k is the number of already occupied tables.

In general any sequential seating plan corresponding to a prediction rule can be used to construct a partition $\Pi_\infty = \{\Pi_n\}_{n \geq 1}$ of $\mathbb{N}$, but nevertheless most seating plans will fail to produce a partition Π_∞ that is exchangeable. Let us experiment with simple plans to see which ones do generate exchangeable partitions.

According to Kingman's representation Theorem 5.1 in order for Π_∞ to be exchangeable, a necessary condition is that the proportion of customers seating at table j has an almost sure limit P_j as the number of customers in the restaurant tends to ∞, that is $\frac{N_{n,j}}{n} \to P_j$ as $n \to \infty$ where $N_{n,j}$ is the number of customers sitting at table j after the arrival of the nth customer. More formally, this is the limit frequency of the jth block of Π_∞ when the blocks are ordered according to the least element. This suggests the next seating plan.

Example 6.3. *Let* $(P_1, P_2, ...)$ *be a deterministic sequence such that* $P_j \geq 0$ *for every* $j \geq 1$ *and* $\sum_{j=1}^{\infty} P_j = 1$. *Consider the next seating plan. As usual the first customer to arrive will seat at table* 1. *For* $n \geq 1$, *if after the* n*th* *customer arrived there are k occupied tables with n_j persons sitting at table j for $1 \le j \le k$, then the customer number $n+1$ will sit at the occupied table j with probability P_j and at table $k+1$ with probability $1 - \sum_{j=1}^{k} P_j$. Let Π_∞ be the partition of $\mathbb{N}$ derived by this construction and Π_n its restriction to $[n]$ as done before. It is easy to see that*

$$\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}] = \prod_{j=1}^{k} P_j^{n_j - 1} \prod_{j=1}^{k-1} \left(1 - \sum_{i=1}^{j} P_i\right) =: p(n_1, ..., n_k)$$

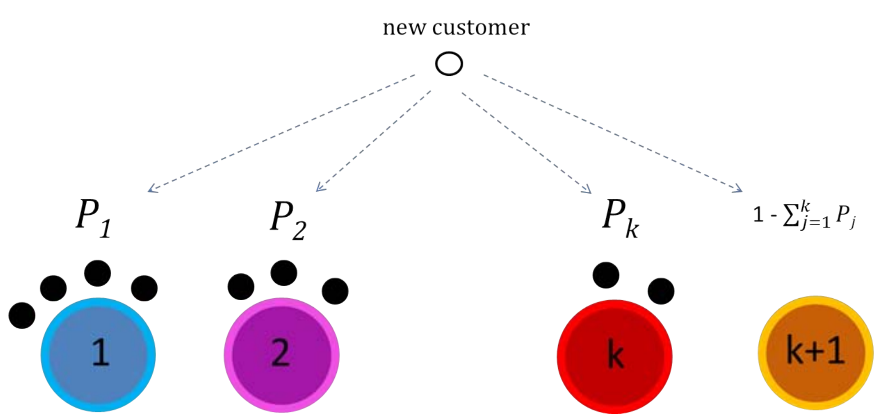
where $n_i = |A_i|$ and the blocks are ordered according to the least element ordering. Note that in order for the partition to be exchangeable it is necessary for p to be a symmetric function of its arguments, which holds only if $P_j = P_i$ for every $j, i \ge 1$, but as $\sum_{j=1}^{\infty} P_j = 1$ and $P_j \ge 0$ for every $j \ge 1$ we can conclude that there are no deterministic sequences $(P_1, P_2, ...)$ that produce exchangeable partitions according to the chinese restaurant construction.

Random Seating Plan for a Partially Exchangeable Partition

Let $(P_1, P_2, ...)$ be an arbitrary random sequence such that $P_j \ge 0$ for all $j \in \mathbb{N}$ and $\sum_{j=1}^{\infty} P_j \le 1$. Given the entire sequence $(P_1, P_2, ...)$, the chinese restaurant manager decides to use the following seating plan: The first customer to arrive will be seated at table 1. For $n \ge 1$, after the arrival of the nth customer, the customer number $n+1$ will seat

- at the occupied table j with probability P_j, for $1 \le j \le k$
- at the non-occupied table $k+1$ with probability $1 - \sum_{j=1}^{k} P_j$

where k is the number of already occupied tables after the nth customer arrived.



By construction and the law of large numbers, for each j the limiting frequency of the proportion of customers sitting at table j exists and equals P_j.

Moreover, if Π_∞ is the partition of $\mathbb{N}$ derived by this seating plan and Π_n its restriction to $[n]$, then by conditioning on the entire sequence and Kingman's correspondence we get that

$$\mathbb{P}[\Pi_n = \{A_1, ..., A_k\}] = \mathbb{E}\left[\prod_{j=1}^{k} P_j^{n_j-1} \prod_{j=1}^{k-1}\left(1 - \sum_{i=1}^{j} P_i\right)\right] =: p(n_1, ..., n_k)$$

where $n_j = |A_j|$ and the blocks $A_1, ..., A_k$ are in order of the least element. Note that p may or may not be a symmetric function of its arguments, in the case when it is, is clear that the derived partition Π_∞ of $\mathbb{N}$ will be exchangeable. Whether p is symmetric or not, such partition is called *partially exchangeable.* These conditions lead to the following variation of Kingman's representation.

Theorem 6.1. *Let $(P_1, P_2, ...)$ be an arbitrary sequence of random variables such that $P_j \geq 0$ for every $j \geq 1$ and $\sum_{j=1}^{\infty} P_j \leq 1$ and define*

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

i) There exists an exchangeable partition Π_∞ of $\mathbb{N}$ whose block frequencies on order of appearance $(\tilde{P}_1, \tilde{P}_2, ...)$ distribute like $(P_1, P_2, ...)$ iff $p(n_1, ..., n_k)$ is a symmetric function of $n_1, ..., n_k$ for all k.

ii) If Π_∞ is such random partition of $\mathbb{N}$ with block frequencies $(\tilde{P}_1, \tilde{P}_2, ...)$ then its EPPF is given by $p(n_1, ..., n_k)$ defined above for $P_i = \tilde{P}_i$ and the conditional law of Π_∞ given its blocks frequencies in order of appearance is governed by the random seating plan described before.

6.1.1 The two-parameter model

Taking into account the chinese restaurant construction, we now suggest the next seating plan with two parameters. Let $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ be the exchangeable partition of $\mathbb{N}$ derived by the seating plan we are about to describe.

The (σ, θ) seating plan

Let (σ, θ) such that one of the following holds

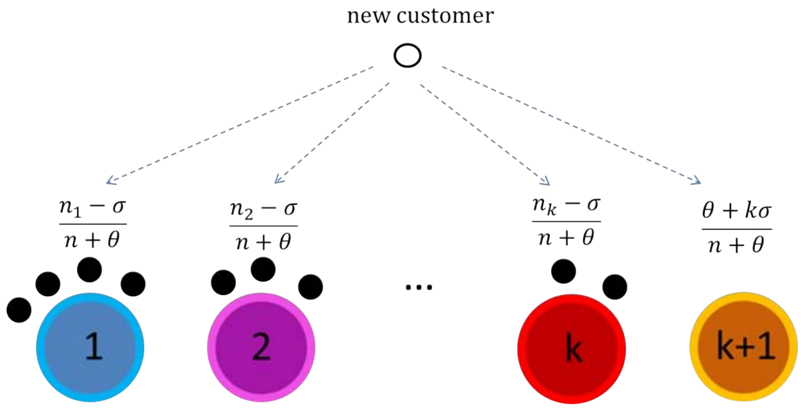
i) $0 < \sigma < 1$ and $\theta > -\sigma$

ii) $\sigma = -r < 0$ and $\theta = mr$ for some $m = 1, 2, ...$

As before, the first customer to arrive at the restaurant sits at table 1 so $\Pi_1 = \{1\}$. Imagine that at stage n there are k occupied tables such that at the ith table there are n_i persons sitting at it (clearly $\sum_{i=1}^{k} n_i = n$), the next costumer to arrive will

- sit at table i with probability $\frac{n_i - \sigma}{n+\theta}$
- sit at a new table with probability $\frac{\theta + k\sigma}{n+\theta}$

as shown in the next figure



Note that given the restrictions we imposed on the parameters (σ, θ), this seating plan is well defined.

Now, let $X_n = \mathbf{1}_{B_n}$ where B_n is the event that the $(n+1)$th customer sits at table 1 and let $(N_{n,1}, ..., N_{n,K_n})$ be the frequencies of the blocks of Π_n in order of appearance. It is clear that if we define $S_n := \sum_{i=1}^{n} X_n$ then $N_{n,1} = S_{n-1} + 1$. Also note that

$$\begin{aligned}\mathbb{P}[X_{n+1} = 1 | X_1, ..., X_n] &= \frac{(N_{n+1,1}) - \sigma}{n+1+\theta} \\ &= \frac{S_n + 1 - \sigma}{n+1+\theta} \\ &= \frac{(1-\sigma) + S_n}{(1-\sigma) + (\theta+\sigma) + n}\end{aligned}$$

Hence $X_1, X_2, ...$ follow the same distributional law of $Y_1, Y_2, ...$ where, standing in the context of a Pólya Eggenberger urn, Y_i is the indicator function of the event that the ith ball drawn is white, starting from an urn containg $1-\sigma$ white balls and $\theta + \sigma$ black balls (see A.1). Thus $\frac{S_n}{n} \to \rho$ a.s. and as $n \to \infty$ where $\rho \sim Beta(1-\sigma, \theta+\sigma)$. In order words, as $n \to \infty$

$$\frac{N_{n,1}}{n} \to W_1 \quad \text{a.s. where } W_1 \sim Beta(1-\sigma, \theta+\sigma)$$

Note that given $P_1 = \lim_{n\to\infty} \frac{N_{n,1}}{n}$, that is the frequency of the block of Π_∞ containing the element 1, the proportion of customers that will not be sitting at table 1 is $1 - P_1$, now among them we are interested in the proportion of those that are sitting in table 2. To this aim assume all the customers that will end up sitting at table 1 have made a reservation, so now if a person arrives to the restaurant and has a reservation (at table 1) he (or she) will just pass through, on the other hand if that person does not have a reservation he (or she) will be placed in some other table than table 1. In order for this to be consistent with the last construction the first customer, without reservation, to arrive will be seated at table 2. At stage n (i.e. after n customers without reservation arrived) if there are n_2 customers sitting at table 2 the $(n+1)$ customer will be sited at table 2 with probability $\frac{n_2-\sigma}{n+\theta+\sigma}$ or at another table with probability $1-\frac{n_2-\sigma}{n+\theta+\sigma}$. Define C_n the event that the $(n+1)$th customer without reservation sits at table 2, $Z_n = \mathbf{1}_{C_n}$ and $T_n = \sum_{i=1}^{n} Z_n$. Then $T_n + 1$ stands for the number of customers without reservation that are sitting at table 2 after $n+1$ customers (not having a reservation) arrived. Analogously as before

$$\mathbb{P}[Z_{n+1} = 1 | Z_1, ..., Z_n] = \frac{T_n + 1 - \sigma}{n + 1 + \theta + \sigma} = \frac{(1-\sigma) + T_n}{(1-\sigma) + (\theta + 2\sigma) + n}$$

So the proportion of customers without reservation that will end up sitting at table 2 is

$$\lim_{n\to\infty} \frac{T_n + 1}{n+1} = \lim_{n\to\infty} \frac{T_n}{n} = W_2 \text{ almos surely}$$

where $W_2 \sim Beta(1-\sigma, \theta + 2\sigma)$.

Therefore we can conclude that given P_1 (the frequency of the first block of Π_∞ in order of appearance) the frequency $P_2 = \lim_{n\to\infty} \frac{N_{n,2}}{n}$ of the second block will be $(1-P_1)W_2$ where $W_2 \sim Beta(1-\sigma, \theta+2\sigma)$. Recall that we have already seen that $P_1 = W_1 \sim Beta(1-\sigma, \theta+\sigma)$, thus it can be stated that

$$(P_1, P_2) = (W_1, (1-W_1)W_2)$$

where W_1, W_2 are independent random variables such that $W_1 \sim Beta(1-\sigma, \theta+\sigma)$ and $W_2 \sim Beta(1-\sigma, \theta+2\sigma)$.

Inductively, and under similar arguments we obtain that if P_i is the frequency of the ith block of Π_∞ in order of appearance then

$$(P_1, P_2, P_3, ...) = (W_1, \overline{W}_1 W_2, \overline{W}_1 \overline{W}_2 W_3, ...)$$

where $\overline{W}_i = 1 - W_i$ and $\{W_i\}_{i\geq 1}$ are independent random variables such that $W_i \sim Beta(1-\sigma, \theta+i\sigma)$. Hence, we have proved the second part of the following theorem.

Theorem 6.2. *For each pair of parameters (σ, θ) such that one of the following holds*

- $0 < \sigma < 1$ *and* $\theta > -\sigma$
- $\sigma = -r < 0$ *and* $\theta = mr$ *for some* $m = 1, 2, ...$

the chinese restaurant construction with the (σ,θ) seating plan generates an exchangeable partition Π_∞ of $\mathbb{N}$. The corresponding EPPF is

$$p(n_1,...,n_k)=\frac{(\theta+\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}} \tag{6.2}$$

where $n=\sum_{i=1}^{k}n_i$ and

$$(x)_{n\uparrow\alpha}=\prod_{i=0}^{n-1}(x+i\alpha).$$

The corresponding limit frequencies of classes, in size-biased order of the least element, can be represented as

$$(\tilde{P}_1,\tilde{P}_2,\tilde{P}_3,...)=(W_1,\overline{W}_1W_2,\overline{W}_1\overline{W}_2W_3,...) \tag{6.3}$$

where $\{W_i\}_{i\geq 1}$ are independent random variables such that $W_i\sim Beta(1-\sigma,\theta+i\sigma)$ and $\overline{W}_i:=1-W_i$.

Proof:

Let $\pi=\{A_1,...,A_k\}$ be a partition of $[n]$, let $\pi_j=\{A_1,...,A_j\cup\{n+1\},...A_k\}$ for $1\leq j\leq k$ and $\pi_{k+1}=\{A_1,...,A_k,\{n+1\}\}$. From the construction it is clear that Π_n is consistent as n varies, so

$$\mathbb{P}[\Pi_n=\pi]=\sum_{j=1}^{k+1}\mathbb{P}[\Pi_{n+1}=\pi_j]$$

Note that given Π_{n+1}, Π_n can be obtained by the deletion of the element $n+1$, so $(\Pi_n=\pi)\cap(\Pi_{n+1}=\pi_j)=(\Pi_{n+1}=\pi_j)$. Also $\Pi_n=\pi$ means that after the nth costumer arrived there are $n_i=|A_i|$ sitting at table i for $i\in[k]$ where the blocks of π, $A_1,...,A_k$, are in order of the least element. Thus

$$\frac{\mathbb{P}[\Pi_{n+1}=\pi_j]}{\mathbb{P}[\Pi_n=\pi]}=\mathbb{P}[\Pi_{n+1}=\pi_j|\Pi_n=\pi]=\frac{n_j-\sigma}{n+\theta},\quad j=1,...k$$

and

$$\frac{\mathbb{P}[\Pi_{n+1}=\pi_{k+1}]}{\mathbb{P}[\Pi_n=\pi]}=\mathbb{P}[\Pi_{n+1}=\pi_{k+1}|\Pi_n=\pi]=\frac{\theta+k\sigma}{n+\theta}$$

Hence

$$\mathbb{P}[\Pi_{n+1}=\pi_j]=\frac{n_j-\sigma}{n+\theta}\mathbb{P}[\Pi_n=\pi],\quad j\in[k]$$
$$\mathbb{P}[\Pi_{n+1}=\pi_{k+1}]=\frac{\theta+k\sigma}{n+\theta}\mathbb{P}[\Pi_n=\pi]$$

from these equations, together with $\mathbb{P}[\Pi_1=\{\{1\}\}]=1$ one can inductively realize that

$$\mathbb{P}[\Pi_n=\pi]=\frac{(\theta+\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}}=:p(n_1,...,n_k)$$

where p is a symmetric function of its arguments, so Π_∞ is an exchangeable random partition of $\mathbb{N}$ and its EPPF is given by the last formula.

The second part of the theorem has already been proved in the previous discussion.

Definition 6.2 (GEM and PD distributions)**.** *Let (σ, θ) such that one of the following conditions holds*

- $0 < \sigma < 1$ *and* $\theta > -\sigma$
- $\sigma = -r < 0$ *and* $\theta = mr$ *for some* $m = 1, 2, ...$

The partition of $\mathbb{N}$ *derived by the chinese reatuarant (σ, θ)-seating plan is named a (σ, θ)-partition*

i) We call the distribution of the size biased frequencies $\{\tilde{P}_j\}_{j\geq 1}$ defined by (6.3) the Griffiths-Engen-McCloskey distribution *with parameters (σ, θ) abbreviated $GEM(\sigma, \theta)$.*

ii) We call the corresponding distribution on

$$\mathcal{P}^{\downarrow}_{[0,1]} = \left\{ (p_1, p_2, ...) : p_1 \geq p_2 \geq ... \geq 0, \sum_{i=1}^{\infty} p_i \leq 1 \right\}$$

of the ranked frequencies of a (σ, θ)-partition (obtained by ranking $\{\tilde{P}_j\}_{j\geq 1}$ with $GEM(\sigma, \theta)$ distribution) the Poisson-Dirichlet distribution *with parameters (σ, θ) abbreviated $PD(\sigma, \theta)$.*

Chapter 7

Characterization of the two-parameter model

Let us study and characterize the model according to its parameters.

$\sigma = -r < 0$ and $\theta = mr$ for some $m = 1, 2, ...$

Going back to the chinese restaurant construction recall that after k tables are occupied, the next costumer will seat at a new table with probability $\frac{\theta+k\sigma}{n+\theta} = \frac{r(m-k)}{n+mr}$. Note that for $k < m$ the probability is positive, so there is a chance for a non-occupied table to become occupied, when $k = m$ the probability last becomes zero so there is no chance for a new customer to seat at a non-occupied table, thus obviously the scenario $k > m$ does not has a chance to take place. In terms of the partition generated $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ it is clear that for each $n \geq 1$, Π_n will have at most m blocks, that is $K_n \leq m$ a.s. By the last argument and the last theorem we know that

$$p(n_1, ..., n_k) = \begin{cases} 0 & \text{if } k > m \\ \frac{(r(m-1))_{k-1\uparrow -r} \prod_{i=1}^{k} (r+1)_{n_i-1\uparrow 1}}{(mr+1)_{n-1\uparrow 1}} & \text{if } k \leq m \end{cases}$$

where $n := \sum_{i=1}^{k} n_i$.

Note that $(x+1)_{c-1\uparrow 1} = \frac{(x)c\uparrow 1}{x}$ for $c \geq 1$, so if $k \leq m$

$$p(n_1, ..., n_k) = [(r(m-1))_{k-1\uparrow -r}] \frac{(mr)}{r^k} \frac{\prod_{i=1}^{k} (r)_{n_i\uparrow 1}}{(mr)_{n\uparrow 1}}$$

also

$$\begin{aligned}
[(r(m-1))_{k-1\uparrow -r}]\,\frac{(mr)}{r^k} &= \frac{r(m)*r(m-1)*r(m-2)*\cdots*r(m-(k-1))}{r^k} \\
&= (m)_{k\downarrow 1}
\end{aligned}$$

so we can conclude that

$$p(n_1,...,n_k) = \begin{cases} 0 & \text{if } k > m \\ (m)_{k\downarrow 1}\frac{\prod_{i=1}^{k}(r)_{n_i\uparrow 1}}{(mr)_{n\uparrow 1}} & \text{if } k \leq m \end{cases}$$

Comparing this EPPF with the one obtained in Example 5.6 with $\alpha = r$, and by Kingman's correspondence it is clear that Π_∞ is distributed as if by sampling from a symmetric Dirichlet distribution with m parameters equal to r.

Note that the corresponding limit frequencies of classes, in size-biased order of the least element, can be represented as

$$(\tilde{P}_1, \tilde{P}_2, ..., \tilde{P}_{m-1}, \tilde{P}_m) = \left(W_1, \overline{W}_1 W_2, ..., \prod_{i=1}^{m-2}(1-\overline{W_i})W_{m-1}, \prod_{i=1}^{m-1}(1-\overline{W_i})\right)$$

where $\{W_i\}_{i=1}^{m-1}$ are independent random variables such that $W_i \sim Beta(1 + r, r(m-i))$ and $\overline{W}_i := 1 - W_i$.

$\sigma = 0$ and $\theta > 0$

Note that for this case the EPPF reduces to the form

$$p(n_1,...,n_k) = \frac{\theta^k}{(\theta)_{n-1\uparrow 1}}\prod_{i=1}^{k}(n_i - 1)!$$

by the form of the EPPF and Kingman's correspondence, we get that Π_∞ is distributed as if sampling from a Dirichlet process with parameter θ. We first discussed this process in Example **??**, and we also constructed it in Chapter 4.1 by normalizing the Moran-Gamma process.

We have seen that limiting classes frequencies of the blocks in size-biased order of the least element can be seen as

$$(\tilde{P}_1, \tilde{P}_2, \tilde{P}_3 ...,) = \left(W_1, \overline{W}_1 W_2, \overline{W}_1\overline{W}_2 W_3, ...\right)$$

where $\{W_i\}_{i\geq 1}$ are independent random variables such that $W_i \sim Beta(1,\theta)$ and $\overline{W}_i := 1 - W_i$. Also recall that the particular case when $\sigma = 0$ and $\theta = 1$

corresponds to Π_∞ generated by the cycles of a consistent sequence of uniform random permutations, this scenario was analysed with some detail in the first chinese restaurant construction.

In the *ordered paintbox scenario* (see Chapter 5.3), if we choose the subordinator to be a Moran-Gamma process and choose the stopping $\tau = \theta$ so the length of the Lebesgue measure of the interval $[0, \tau]$ equals θ, then it is clear that the partition generated by the ordered paintbox construction will be identically distributed as a partition governed by a $(0, \theta)$ model. We have already proved what has been mentioned in these paragraphs using the form of the EPPF, nevertheless an alternative proofe of the relationship between the $(0, \theta)$-partition and the Moran-Gamma process will be provided in the following.

$0 < \sigma < 1$ and $\theta > -\alpha$

In order to complete the analysis of this scenario, we are first going to concentrate in the case where $\theta = 0$, and then introduce the parameter θ by making a slight modification of the construction we are about to make.

$0 < \sigma < 1$ and $\theta = 0$

As we are about to see, this case turns out to be related with the σ-stable subordinator in the similar, yet not identical, way that the $(0, \theta)$ model was related to the Moran-Gamma process. The main different is the meaning of the parameter (either σ or θ) in the construction. In the case of the $(0, \theta)$ model, θ is the stopping time that we choose in the *ordered paintbox* construction, while in the case $(\sigma, 0)$, σ is the parameter of the stable process.

In order to justify what has been said in the last paragraph, let us concentrate for now in a more general scenario.

Let $\{\phi(t)\}_{t \geq 0}$ be the restriction to $\mathbb{R}^+$ of a subordinator with $\beta = 0$ (following the notation of Chapter 3). Let $(J_1, J_2, ...)$ be the jumps of the process with sum $T := \sum_{i=1}^{\infty} J_i$, define $P_i := \frac{J_i}{T}$ and let $(\tilde{P}_1, \tilde{P}_2, ...)$ be a size-biased permutation of $(P_1, P_2, ...)$ and $(\tilde{J}_1, \tilde{J}_2, ...)$ the corresponding reordering of the jumps. Set

$$
\begin{aligned}
U_1 &= 1 - \tilde{P}_1, \\
U_2 &= \frac{1 - \tilde{P}_1 - \tilde{P}_2}{1 - \tilde{P}_1} \\
U_3 &= \frac{1 - \tilde{P}_1 - \tilde{P}_2 - \tilde{P}_3}{1 - \tilde{P}_1 - \tilde{P}_2} \\
&\vdots
\end{aligned}
$$

and so on, so that

$$\tilde{P}_i = (1 - U_i) \prod_{j=1}^{i-1} U_j$$

or equivalently

$$\tilde{P}_i = (W_i) \prod_{j=1}^{i-1} (1 - W_j)$$

where we have the relation $W_i = 1 - U_i$ for every $i \geq 1$.

Let us set $T_0 = T$, and for every $n \geq 1$ let

$$T_n := T - \sum_{i=1}^{n} \tilde{J}_i$$

Note that

$$\frac{T_i}{T_{i-1}} = \frac{T - \sum_{j=1}^{i} \tilde{J}_j}{T - \sum_{j=1}^{i-1} \tilde{J}_j} = \frac{1 - \sum_{j=1}^{i} \tilde{P}_j}{1 - \sum_{j=1}^{i-1} \tilde{P}_j} = U_i$$

It is clear from Chapter 3 that the jumps $(J_1, J_2, ...)$ must be i.i.d. random variables, this means that T is a positive and infinitely divisible random variable and hence has its Lévy-Khintchine representation takes the form

$$\mathbb{E}\left[e^{-yT}\right] = \exp\left\{-\int_0^\infty (1 - e^{-yx})\Delta(dx)\right\}, \quad y > 0. \tag{7.1}$$

Blanket assumptions on Δ, to ensure $0 < T < \infty$ a.s., are

$$\Delta((0,\infty)) = \infty, \qquad \Delta((1,\infty)) < \infty, \quad \text{and} \quad \int_0^1 x\Delta(dx) < \infty$$

Also assume that Δ has a density ρ, that is

$$\Delta(dx) = \rho(x)dx, \quad x > 0.$$

Then T has a probability density

$$\mathbb{P}[T \in dt] = f(t)dt, \qquad t > 0,$$

for some f determined by equation (7.1). In this context, we have the next theorem.

Theorem 7.1. *The sequence $T, T_1, T_2, ...$ is a Markov chain with stationary transition probabilities*

$$\mathbb{P}[T_1 \in dt_1 | T = t] = \frac{\Theta(t - t_1)}{t} \frac{f(t_1)}{f(t)} dt_1, \tag{7.2}$$

$$\text{where } \Theta(x) = x\rho(x). \tag{7.3}$$

Consequently, the joint density of (T, T_1, T_2) *is*

$$f(t, t_1, t_2) = \frac{\Theta(t - t_1)}{t} \frac{\Theta(t_1 - t_2)}{t_1} f(t_2); \tag{7.4}$$

the joint density of $(T, U_1, U_2) = \left(T, \frac{T_1}{T}, \frac{T_2}{T_1}\right)$ *is*

$$g(t, u_1, u_2) = \Theta(\overline{u}_1 t)\Theta(u_1 \overline{u}_2 t) f(u_1 u_2 t), \tag{7.5}$$

where $\overline{u} = 1 - u$*; the joint density of* (T_2, U_1, U_2) *is*

$$h(t_2, u_1, u_2) = \frac{1}{u_1 u_2} \Theta\left(\frac{\overline{u}_1 t_2}{u_1 u_2}\right) \Theta\left(\frac{\overline{u}_2 t_2}{u_2}\right) f(t_2); \tag{7.6}$$

and for every $n \geq 1$ *there is a similar product formula for the* $n+1$ *dimensional joint densities of* $(T, T_1, ...T_n)$*,* $(T, U_1, ..., U_n)$ *and* $(T_n, U_1, ..., U_n)$*.*

The proof of this theorem is based on the following lemma

Lemma 7.1. *Let* $J^{\downarrow *} = (J_1^{\downarrow *}, J_2^{\downarrow *}, ...)$ *be the ranked sequence obtained from the deletion of the point* J_1 *from the ranked sequence*

$$J^{\downarrow} = (J_1^{\downarrow}, J_2^{\downarrow}, ...).$$

Then for $x > 0$*,* $t_1 > 0$ *and measurable subsets* B *of the sequence space,*

$$\mathbb{P}[\tilde{J}_1 \in dx, T_1 \in dt_1, J^{\downarrow *} \in B] = (\rho(x)dx)\mathbb{P}[T \in dt_1, J^{\downarrow} \in B]\frac{x}{x + t_1} \tag{7.7}$$

Proof:

Consider the event $(\tilde{J}_1 \in dx, T_1 \in dt_1, J^{\downarrow *} \in B)$ as the following chain of events

1) Some point $J_i^{\downarrow} \in dx$. Note that

$$\mathbb{P}[J_i^{\downarrow} \in dx] = \rho(x)dx$$

this can be seen by the Lévy-Khintchine representation of T and the fact that $T = \sum_{i=1}^{\infty} J_i$.

2) Given 1), the sequence $J^{\downarrow *}(x)$ obtained by the deletion of point in dx is such that $\sum_j J^{\downarrow *}(x)_j \in dt_1$ and $J^{\downarrow *}(x) \in B$.

Note that if we condition the Poisson process defined by $J^{\downarrow}$ to have a point in dx, then the remaining points $J^{\downarrow *}(x)$ are distributed just like the original points meaning that the conditional probability of 2) given 1) is $\mathbb{P}[T \in dt_1, J^{\downarrow} \in B]$.

3) Given 1) and 2) the value $J_i^{\downarrow} \in dx$ is chosen by size-biased sampling from $J^{\downarrow}$. Note that $\mathbb{P}[\tilde{J}_1 = J_i^{\downarrow} | J_i^{\downarrow} = x] = \frac{x}{x+t_1}$.

The product of the probabilities of 1), 2) and 3) is the right side of equation (7.7).

□

Proof of Theorem 7.1:

First of all, observe that summing over all the measurable subsets B in both sides of the equation (7.7) we get that

$$\mathbb{P}[\tilde{J}_1 \in dx, T_1 \in dt_1] = (\rho(x)dx)\mathbb{P}[T \in dt_1]\frac{x}{x+t_1}, \tag{7.8}$$

by the last lemma and last equation and since $T = \tilde{J}_1 + T_1$

$$\begin{aligned}
\frac{\mathbb{P}[J^{\downarrow *} \in B, T \in dt, T_1 \in dt_1]}{\mathbb{P}[T \in dt, T_1 \in dt_1]} &= \frac{\mathbb{P}[J^{\downarrow *} \in B, \tilde{J}_1 \in dx, T_1 \in dt_1]}{\mathbb{P}[\tilde{J}_1 \in dx, T_1 \in dt_1]} \\
&= \frac{(\rho(x)dx)\mathbb{P}[T \in dt_1, J^{\downarrow} \in B]\frac{x}{x+t_1}}{\rho(x)dx)\mathbb{P}[T \in dt_1]\frac{x}{x+t_1}} \\
&= \frac{\mathbb{P}[T \in dt_1, J^{\downarrow} \in B]}{\mathbb{P}[T \in dt_1]}
\end{aligned}$$

where $x = t - t_1$. Hence the conditional distribution of $J^{\downarrow *}$ given T and $T_1 = t_1$ is identical to the conditional distribution of $J^{\downarrow}$ given $T = t_1$. Also given the sequence $J^{\downarrow *}$ the sequence $T_2, T_3, ...$ is obtained by the same procedure of size-biased deletion that was used to obtain $T_1, T_2, ...$ from $J^{\downarrow}$. This means that the conditional distribution of $T_1, T_2, ...$ given $T = t_1$ is identical to the conditional distribution of $T_2, T_3, ...$ given T and $T_1 = t_1$. This proves that $T, T_1, T_2, ..$ is a stationary Markov Chain.

Now, by equation (7.8) and the fact the the event $(\tilde{J}_1 \in dx, T_1 \in dt_1)$ is the same as the event $(T \in dt, T_1 \in dt_1)$ with $t = x + t_1$, we obtain

$$\mathbb{P}[T \in dt, T_1 \in dt_1] = (\rho(t-t_1)d(t-t_1))f(t_1)dt_1\frac{t-t_1}{t}$$

(recalling the assumption that T has a probability density f), thus

$$\mathbb{P}[T_1 \in dt_1 | T = t] = \frac{(t-t_1)\rho(t-t_1)}{t}\frac{f(t_1)}{f(t)}dt_1$$

this proves the equation (7.2).

As the Markov chain is stationary, it is clear that

$$\mathbb{P}[T_2 \in dt_2 | T_1 = t_1] = \frac{(t_1 - t_2)\rho(t_1 - t_2)}{t_1} \frac{f(t_2)}{f(t_1)} dt_2$$

it is immediate then, that the joint density of (T, T_1, T_2) is

$$\begin{aligned} f(t, t_1, t_2) &= \left[\frac{(t_1 - t_2)\rho(t_1 - t_2)}{t_1} \frac{f(t_2)}{f(t_1)} \right] \left[\frac{(t - t_1)\rho(t - t_1)}{t} \frac{f(t_1)}{f(t)} \right] f(t) \\ &= \frac{\Theta(t_1 - t_2)}{t_1} \frac{\Theta(t - t_1)}{t} f(t_2) \end{aligned}$$

hence we have proved equation (7.4).

The joint density of (T, U_1, U_2) is very easy to calculate given the fact that we already now the joint density of (T, T_1, T_2) and by making the elementary changes of variables

$$(T, T_1, T_2) \to \left(T, \frac{T_1}{T}, \frac{T_2}{T_1} \right).$$

with this, equation (7.5) is proved.

Finally, the joint density of (T_2, U_1, U_2) can be obtained by making the change of variables

$$(T, U_1, U_2) \to \left(\frac{T_2}{U_1 U_2}, U_1, U_2 \right)$$

and using equation (7.5) One can inductively compute the joint densities of $(T, T_1, ...T_n)$, $(T, U_1, ..., U_n)$ and $(T_n, U_1, ..., U_n)$ by repeating over and over the previous arguments.

□

Corolary 7.1. *In the setting of Theorem 7.1, with* $U_1 = \frac{T_1}{T}$*:*

i) The random variables U_1 *and* T *are independent if and only if* T *has the Gamma distribution with density*

$$f(t) = \frac{1}{\Gamma(\theta)} \lambda^\theta t^{\theta - 1} e^{-\lambda t} \tag{7.9}$$

for some $\theta > 0$, $\lambda > 0$. *Equivalently*

$$\Theta(x) = \theta e^{-\lambda x}.$$

Then $T, U_1, ..., U_n$ *are independent and* $U_n \sim Beta(\theta, 1)$ *for all* $n \geq 1$.

ii) The random variables U_1 *and* T_1 *are independent if and only if* T *has the stable distribution with Laplace transform*

$$\mathbb{E}[e^{-yT}] = \int_0^\infty e^{-yt} f(t) dt = \exp\{-cy^\sigma\} \tag{7.10}$$

for some $\sigma > 0$ and $c > 0$. Equivalently,

$$\Theta(x) = Kx^{-\sigma}, \tag{7.11}$$

where $K = \frac{c\sigma}{\Gamma(1-\sigma)}$. Then $T_n, U_1, ..., U_n$ are independent and $U_n \sim Beta(n\sigma, 1-\sigma)$ for all $n \geq 1$

Proof of i):

First assume that T has gamma distribution with parameters (θ, λ) for some $\theta > 0$, $\lambda > 0$. Substituting the formulae for f and Θ into the equation (7.5) we get

$$\begin{aligned}
g(t, u_1, u_2) &= \theta e^{-\overline{u}_1 t\lambda} \theta e^{-u_1 \overline{u}_2 t\lambda} \frac{1}{\Gamma(\theta)} \lambda^\theta (u_1 u_2 t)^{\theta-1} e^{-u_1 u_2 t\lambda} \\
&= \left[\frac{1}{\Gamma(\theta)} \lambda^\theta t^{\theta-1} e^{-((1-u_1)+(u_1(1-u_2))+u_1 u_2)t\lambda}\right] \left[\theta u_1^{\theta-1}\right] \left[\theta u_2^{\theta-1}\right] \\
&= \left[\frac{1}{\Gamma(\theta)} \lambda^\theta t^{\theta-1} e^{-t\lambda}\right] \left[\theta u_1^{\theta-1}\right] \left[\theta u_2^{\theta-1}\right]
\end{aligned}$$

meaning that T, U_1 and U_2 are independent. The joint density of $(T, U_1, ..., U_n)$ can be factorized in a similar way:

$$g(t, u_1, ..., u_n) = \left[\frac{1}{\Gamma(\theta)} \lambda^\theta t^{\theta-1} e^{-t\lambda}\right] \prod_{i=1}^{n} \left[\theta u_i^{\theta-1}\right] \tag{7.12}$$

Conversely, if T and U_1 are independent then from the equation (7.4) be must have

$$g(t, u_1) = \Theta(\overline{u}_1 t) f(u_1 t) = \alpha(u_1) f(t)$$

where α is the density probability function of U_1. This means that the joint density of (T, U_1, U_2) must have the form:

$$\begin{aligned}
g(t, u_1, u_2) &= \Theta(\overline{u}_1 t)[\Theta(u_1 \overline{u}_2 t) f(u_1 u_2 t)] \\
&= \Theta(\overline{u}_1 t)[\alpha(\overline{u}_2) f(u_1 t)] \\
&= \alpha(\overline{u}_2)[\alpha(\overline{u}_2) f(u_1 t)] \\
&= \alpha(\overline{u}_2) \alpha(\overline{u}_1) f(t).
\end{aligned}$$

By repeating again and again this argument one gets that the joint density of $(T, U_1, ..., U_n)$ has the form:

$$g(t, u_1, ...u_n) = f(t) \prod_{i=1}^{n} \alpha(\overline{u}_i)$$

meaning that $U_1, ..., U_n$ are i.i.d. and also independent from T. In order to prove that T has a gamma distribution we are going to make use of the following result of Lukacs (1955).

Result 7.1. *If X and Y are two random variables, independent, positive, non-constant and such that $X+Y$ is independent of $\frac{X}{X+Y}$, then X and Y have gamma distribution with common scale parameter.*

Let $\{\varphi(s)\}_{s\geq 0}$ be a subordinator with $\varphi(1)=T$ and such that the jumps are dense everywhere. Define $X=\varphi(1/2)$ and $Y=T-\varphi(1/2)$ and let I_n be the indicator function of the event that the nth jump (in size-biased order) of size $\tilde{J}_n=T_{n-1}-T_n$ occurs before time 1/2. Note that I_n is independent of the size of the corresponding jump, then $I_1, I_2, ...$ are independent $Bernoulli(1/2)$ variables and are also independent from $T, T_1, T_2, ...$ Now we can rewrite X in the following way

$$
\begin{aligned}
X=\varphi(1/2) &= \sum_{n\geq 1} I_n(T_{n-1}-T_n) \\
&= \sum_{n\geq 1} I_n\left[T\frac{T_1}{T}\frac{T_2}{T_1}\cdots\frac{T_{n-1}}{T_{n-2}}\frac{(T_{n-1}-T_n)}{T_{n-1}}\right] \\
&= \sum_{n\geq 1} I_n\left[TU_1U_2\cdots U_{n-1}\overline{U}_n\right]
\end{aligned}
$$

so that

$$
\frac{X}{X+Y}=\frac{X}{T}=\sum_{n\geq 1} I_n\left[U_1U_2\cdots U_{n-1}\overline{U}_n\right]
$$

is a function of $(I_1, I_2, ...)$ and of $(U_1, U_2, ..)$, as both sequences are independent of T, then $\frac{X}{X+Y}$ is independent of T. The fact that X and Y are independent, positive and non-constant follow directly from the definition of a subordinator. Thus by the result of Lukacs we mentioned earlier, X and Y have are independent random variables that have a gamma distribution with common scale parameter, i.e $X\sim Gamma(\theta_1,\lambda)$ and $Y\sim Gamma(\theta_2,\lambda)$ therefore $T=X+Y\sim Gamma(\theta_1+\theta_2,\lambda)$.

Finally, from the density of $(T, U_1, ..., U_n)$ exhibited in the formula (7.12) it is immediate that $U_1, ..., U_n$ are independent and such that $U_n\sim Beta(\theta,1)$ for all $n\geq 1$.

□

Proof of ii):

For the stable case, inserting the formulae for Θ into the equation (7.6) we obtain

$$
\begin{aligned}
h(t_2,u_1,u_2) &= \frac{1}{u_1u_2}K\left(\frac{\overline{u}_1t_2}{u_1u_2}\right)^{-\sigma}K\left(\frac{\overline{u}_2t_2}{u_2}\right)^{-\sigma}f(t_2) \\
&= K^2[u_1^{\sigma-1}\overline{u}_1^{-\sigma}][u_2^{2\sigma-1}\overline{u}_2^{-\sigma}]t_2^{-2\sigma}f(t_2)
\end{aligned}
$$

hence T_2, U_1 and U_2 are independent. In a similar way, the joint density of $(T_n, U_1, ..., U_n)$ factors:

$$h(t_n, u_1, ..., u_n) = K^n t_n^{-n\sigma} f(t_n) \prod_{i=1}^{n} [u_i^{i\sigma-1} \overline{u}_i^{-\sigma}] \tag{7.13}$$

Thus $T_n, U_1, ..., U_n$ are independent and $U_i \sim Beta(i\sigma, 1-\sigma)$.

Conversely, assuming T_1 and U_1 are independent, by equation (7.6), the joint density of (T_1, U_1) is such that

$$h(t_1, u_1) = \frac{1}{u_1} \Theta\left(\frac{\overline{u}_1 t_1}{u_1}\right) f(t_1) = \alpha(u_1)\beta(t_1)$$

for some density functions α and β. This forces Θ to be of the form $\Theta(tx) = \gamma(t)\delta(x)$ for some positive measurable functions γ and δ. Thus, there must exist some constants K and σ such that $\Theta(x) = Kx^{-\sigma}$. Finally by the constraints on the Lévy measure of a stable subordinator, $0 < \sigma < 1$.

□

Remarks on the Corollary 7.1:

1. Note that along the proof of the part i) of the last corollary, it has been implicitly established that

$$\rho(x) = \frac{\theta e^{-\lambda x}}{x}$$

this means that the original subordinator ϕ is a Moran-Gamma subordinator. Hence the part i) of the corollary is just an alternative proof that the Moran-Gamma process is related to the $(0, \theta)$-partition due to the Kingman's correspondence.

2. The part ii) of the corollary establishes that by normalizing the jumps of a σ-stable subordinator and the reordering them by size-biased sampling we obtain that the ith normalized jump (under the new ordering) will be such that

$$\tilde{P}_i \stackrel{d}{=} W_i \prod_{j=1}^{i-1} (1 - W_j)$$

where $W_1, W_2, ..$ are independent random variables such that $W_i \sim Beta(1-\sigma, i\sigma)$. The comparison of this with the stick-breaking construction of the $(\sigma, 0)$ model and taking into account Kingman's correspondence, we get that the partition generated by the procedure of the *oredered paintbox* taking as a starting point the stable subordinator and a $(\sigma, 0)$-partition are identically distributed.

$0 < \sigma < 1$ and $\theta > -\sigma$

In the set up of the last discussion, let us concentrate in the stable case. From the formulae (7.13) it is easily seen that for every $n \geq 0$

$$\mathbb{P}[T_n \in dt] = K_n^{-1} t^{-n\sigma} f(t) dt$$

Now, the constant of integration K_n can be obtain in the following way, by last equation

$$1 = \int_0^\infty K_n^{-1} t^{-n\sigma} f(t) dt = K_n^{-1} \mathbb{E}\left[T^{-n\sigma}\right]$$

As we will later prove,

$$K_n = \mathbb{E}\left[T^{-n\sigma}\right] = \frac{\Gamma(n+1)}{\Gamma(n\sigma+1)} c^{-n} \tag{7.14}$$

and this formulae is valid, not only for every positive integer n, but also for every real number $r > -1$. For such r let μ_r denote the distribution $(0, \infty)$ whose density at t is $K_r^{-1} t^{-r\sigma} f(t)$.

Let us consider the Markov chain $T = T_0, T_1, T_2, ...$ as described before $T = T_0$ has density f and the transition probabilities are given by (7.2) in the set up of the stable case: Let $\hat{T}_0, \hat{T}_1, \hat{T}_2, ...$ be another Markov chain such that this new process has probability transition as the original chain $T_0, T_1, T_2, ..$, but now we assign $\hat{T}_0$ an arbitrary initial distribution μ. Also let $U_i = \frac{\hat{T}_i}{\hat{T}_{i-1}}$.

We claim

a) If we assign $\hat{T}_0$ has the law μ_r, then for all $n = 1, 2, ...$ the law of $\hat{T}_n$ is μ_{r+n}, the law of U_n is $Beta((r+n)\sigma, 1-\sigma)$ and $\hat{T}_n, U_1, ..., U_n$ are independent.

b) The μ_r laws are the only laws for $\hat{T}_0$ that make $\hat{T}_1$ and U_1 independent.

Note that if $r = n$ for some positive integer n, then the law of $\hat{T}_0$ is the exact same law of T_n in the original chain. As the transition probabilities of both chains are exactly the same, this simply means that we are shifting everything n steps ahead. By corollary 7.1 the the part a) of the last assertion is proved in the case r is a positive integer. Now we see this is also true for every $r > -1$.

If $\hat{T}_0$ has initial density μ with density in t given by $f_0(t)f(t)$ then, by Theorem 7.1 the new chain will be such that

- The joint density of $(\hat{T}_0, \hat{T}_1)$ has the form

$$f(t, t_1) = \frac{\Theta(t - t_1)}{t} \frac{f(t_1)}{f(t)} f_0(t) f(t) = \frac{\Theta(t - t_1)}{t} f_0(t) f(t_1)$$

- The joint density of $(\hat{T}_0, U_1) = \left(\hat{T}, \frac{\hat{T}_1}{\hat{T}_0}\right)$ is

$$g(t, u_1) = \Theta(\overline{u}_1 t) f_0(t) f(u_1 t)$$

- The joint density of $(\hat{T}_1, U_1) = (TU_1, U_1)$ is given by

$$h(t_1, u_1) = \frac{1}{u_1}\Theta\left(\frac{\overline{u}_1 t_1}{u_1}\right) f_0\left(\frac{t_1}{u_1}\right) f(t_1)$$

If the initial distribution is $\mu = \mu_r$ then $f_0(t)f(t) = K_r^{-1}t^{-r\sigma}f(t)$, inserting this and the formulae for Θ, the joint distribution of $(\hat{T}_1, U_1)$ becomes

$$h(t_1, u_1) = KK_r^{-1}u_1^{(r+1)\sigma-1}\overline{u}_1^{-\sigma}t_1^{-(r+1)\sigma}f(t_1)$$

Now, the part a) of the assertion is clear to hold for every real number r, and by using induction, provided the moment K_r is finite, which is seen bellow to be so iff $r > -1$. The part b of the assertion follows by analogous arguments of the ones given in the proof of ii) of corollary 7.1. Note that the right side of the last equation defines a density iff

$$1 = \left[KK_r^{-1}\frac{\Gamma((r+1)\sigma)\Gamma(1-\sigma)}{\Gamma(r\sigma+1)}\right]K_r + 1$$

thus, substituting the formulae for K we obtain

$$K_{r+1} = \frac{\Gamma(r\sigma+1)K_r}{c\sigma\Gamma((r+1)\sigma)} \qquad r > -1 \tag{7.15}$$

The equation (7.14) can by easily proved now by induction over n with $K_0 = 1$, and using the last equation and the recursion for the gamma function $\Gamma(s+1) = s\Gamma(s)$. To show that this formulae is also valid for any real number $r > 0$ we use the fact that for every random variable T with Laplace transform $\alpha(\lambda) = \mathbb{E}\left[e^{-\lambda T}\right]$, and any $p > 0$

$$\mathbb{E}\left[T^{-p}\right] = \frac{1}{\Gamma(p)}\int_0^\infty \lambda^{p-1}\alpha(\lambda)d\lambda.$$

In our set up we get

$$\begin{aligned}\mathbb{E}\left[T^{-r\sigma}\right] &= \frac{1}{\Gamma(r\sigma)}\int_0^\infty \lambda^{r\sigma-1}\exp\left\{-c\lambda^\sigma\right\}d\lambda \\ &= \frac{1}{\sigma\Gamma(r\sigma)}\int_0^\infty \lambda^{\sigma(r-1)}\exp\left\{-c\lambda^\sigma\right\}\sigma\lambda^{\sigma-1}d\lambda\end{aligned}$$

so making the change of variable $y = \lambda^\sigma$ and using the identity

$$\frac{\Gamma(p)}{t^p} = \int_0^\infty y^{p-1}e^{-cy}dy$$

we obtain

$$\begin{aligned}\mathbb{E}\left[T^{-r\sigma}\right] &= \frac{1}{\sigma\Gamma(r\sigma)}\int_0^\infty y^{r-1}e^{-cy}dy \\ &= \frac{r\Gamma(r)}{r\sigma\Gamma(r\sigma)}c^{-r} \\ &= \frac{\Gamma(r+1)}{\Gamma(r\sigma+1)}c^{-r}\end{aligned}$$

The case $-1 < r < 0$ follows from the case $0 < r < 1$ and the recursion (7.15).

Finally, set $\theta := r\sigma$ (note that $-1 < r$ iff $\theta > -\sigma$) so $U_n \sim Beta(\theta + n\sigma, 1 - \sigma)$. If we consider once again

$$\tilde{P}_i = (1 - U_i) \prod_{j=1}^{i} U_j$$

we arrive to the stick-breaking construction of the (σ, θ) model.

A Branching Process Construction of the Two-parameter Model

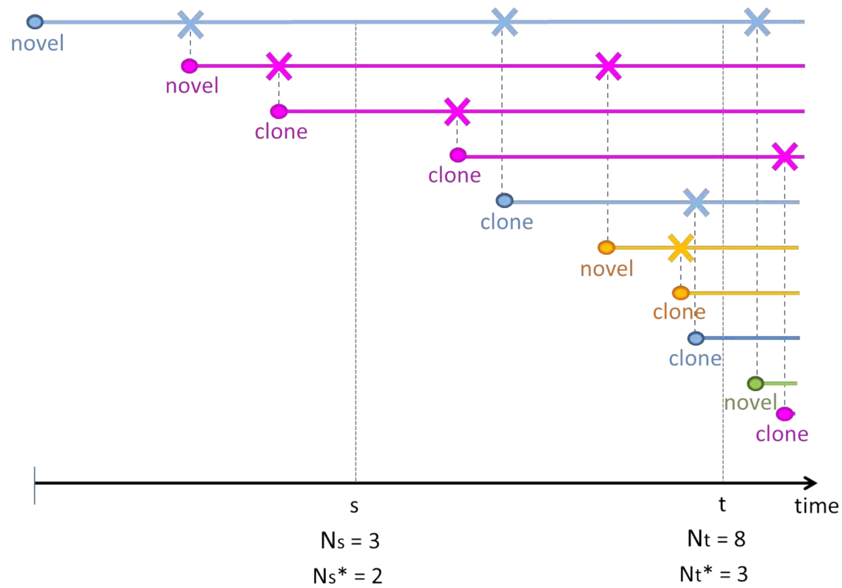
Let $0 < \sigma < 1$, and consider a population of individuals of two types, *novel* and *clone*. The population behaves as follows

- All the individuals, have an infinite lifetime and a fixed colour assigned.
- A novel individual will produce offspring according to a simple Poisson process of rate 1, such that each descendent will be novel with probability σ or clone with probability $1 - \sigma$. That is, a novel individual produces novel individuals according to a Poisson process of rate σ and independently produces clone individuals according to a Poisson process of rate $1 - \sigma$. Each time the novel being produces a clone, the colour of this new member will be exactly the same as his predecessor, if on the other hand, it produces a novel member, the descendant will be assigned a new colour different from all the colours that already appear at the population
- A clone individual will reproduce according to a Poisson process of rate 1, and its descendant can only be clone, so a clone individual is unable to produce novel offspring. Also, all the descendants of a clone member will have the exact same colour as their predecessor.
- At time $t = 0$ there is exactly one novel individual.

Now let

$$\begin{aligned} N_t &:= \text{ number of individuals in the population at time } t. \\ N_t^* &:= \text{ number of novel individuals at time } t. \end{aligned}$$

Note that N_t^* also indicates the number of different colours of the individuals ate time t. Now, it is clear that $N_0 = N_0^* = 1$ and for every $t \geq 0$, $1 \leq N_t^* \leq N_t$. The process $\{N_t^*\}_{t \geq 0}$ is a *Yule process* of equivalently a *pure birth process* with transition rate σk from state k to state $k + 1$. Similarly $\{N_t\}_{t \geq 0}$ is a Yule process with rate 1. The following image illustrates one possible realization of this process.



The colouring scheme of the individuals defines a random partition of $\mathbb{N}$ as will we explained now. Assume that at time t, $N_t = n$, consider the partition Π_n of $[n]$ determined by the process as follows. Let $1 \leq i < j \leq n$, then i and j are in the same block of Π_n iff the ith and the jth individual to be born have the same colour, it is clear that the number of blocks in Π_n will be $K_n = N_t^*$.

Assume that $K_n = k$, and $\Pi_n = \{A_1, ..., A_k\}$ with $|A_i| = n_i$, when a new individual is born, then conditioned on the event $(K_n = k, \Pi_n = \{A_1, ..., A_k\})$ we are interest on the probabilities that this new member is

a) novel,

b) clone and with colour j born from a novel individual,

c) clone and with colour j born from a clone individual,

d) clone and with colour j.

In order to compute this probabilities we first track the predecessor, recall that each individual in the population produces offspring independently according to a Poisson Process of rate 1. Hence if there are n members in the population the probability that the newborn comes from one particular individual is $\frac{1}{n}$ (this follows from standard properties of the Yule process). Now, if its predecessor is clone it has been stated that the new member will be clone with the same colour of its parent, if unlike, its predecessor is novel then this new individual will be novel with probability σ and clone with $1 - \sigma$ conditioned on the event that its predecessor is fixed and novel. With this in mind it is very easy to compute the above probabilities.

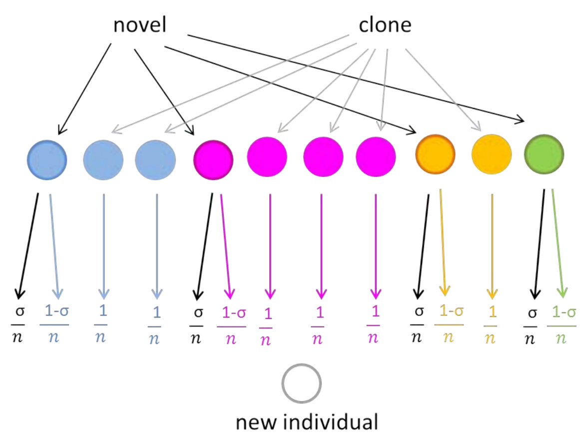
To compute probability of a) it suffices to realize that the probability that this new individual is born novel from one fixed novel member is $\frac{\sigma}{n}$, as there are already k novel members, then the probability of a) is $\frac{k\sigma}{n}$.

In the case of b) there is exactly one novel individual with colour j hence the probability of b) is $\frac{1-\sigma}{n}$.

Now, there are $n_j - 1$ clone members in the population with colour j, thus the probability of c) becomes $\frac{n_j-1}{n}$.

At last d) is the sum of the probabilities of c) and d), so the probability that the newborn is clone and with colour j is $\frac{n_j-\sigma}{n}$.

The following image illustrates what has been stated above. In the image, each novel individual has two arrows below it, indicating that if it is the predecessor of the new being, there are two possibilities: that the new born is novel (black arrow) and that the newborn is clone (coloured arrow). In contrast, each clone individual has one arrow below it. The number below the arrows indicates the probability that the newborn comes from that history.



In terms of the generated partition, it is clear that the following holds

$$\mathbb{P}\left[\Pi_{n+1} = \{A_1, ..., A_k, \{n+1\}\}\right] = \frac{k\sigma}{n}$$
$$\mathbb{P}\left[\Pi_{n+1} = \{A_1, ..., A_j \cup \{n+1\}, ..., A_k\}\right] = \frac{n_j - \sigma}{n}$$

comparing this with the chinese construction of the (σ, θ)-partition, it is evident that the partition generated by this branching construction will be a $(\sigma, 0)$-partition.

In the following we are going to make a variation of the previous construction attempting to generate a (σ, θ)-partition. In the last set up, let us make a modification in the way the first (novel) individual produces offspring, instead of σ, let $\sigma + \theta$ be the rate at which the first individual produces a novel member, the rate at which it produces clone individuals remains $(1 - \sigma)$. The rest of the population behaves as before, and also the colouring scheme, as well as the way to generate partitions remains.

Assume that given that at a certain time t the number of individuals is $N_t = n$, and the partition of $[n]$ generated is $\Pi_n = \{A_1, ..., A_k\}$ with $|A_i| = n_i$. Imagine an individual is born into the population, once again we are interested in the probabilities that the newborn is clone and with a certain colour or novel. In this case we have that the first member produces novel offspring with rate $\sigma + \theta$ and clone descendants with rate $1 - \sigma$, so overall it produces offspring with rate $1 + \theta$ instead of 1. The other $k - 1$ novel members produce novel offspring with rate σ and clone offspring at rate $1 - \sigma$, as for the rest of the population, they produce clone offspring at rate 1 each. Hence the newborn will be

a) clone from the first individual with probability $\frac{1-\sigma}{n+\theta}$.

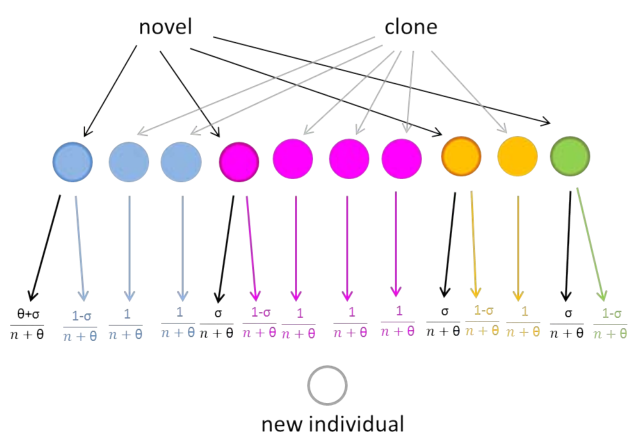
b) novel from the clone individual with probability $\frac{\theta+\sigma}{n+\theta}$.

c) clone from a novel individual, other that the first, with probability $\frac{1-\sigma}{n+\theta}$.

d) novel from a novel individual, other that the first, with probability $\frac{\sigma}{n+\theta}$.

e) clone from a clone individual with probability $\frac{1}{n+\theta}$.

The image below illustrates this.



With this in mind it is clear, that summing over the necessary probabilities,

$$\mathbb{P}\left[\Pi_{n+1} = \{A_1, ..., A_k, \{n+1\}\}\right] = \frac{\theta + k\sigma}{n+\theta}.$$
$$\mathbb{P}\left[\Pi_{n+1} = \{A_1, ..., A_j \cup n+1, ..., A_k\}\right] = \frac{n_j - \sigma}{n+\theta}.$$

which coincides with the chinese restaurant construction for the (σ, θ)-model.

7.1 Asymptotic behaviour

One very important variable implicitly defined in a random partition Π_n is the number of clusters K_n. In the following we are going to analyse the way K_n behaves as n tends to infinity in the context of the two-parameter model. As we will see the comportment of K_n depends on whether the parameter σ is positive, zero or negative.

Recall that K_n can be written as

$$K_n = \sum_{i=1}^{n} X_i$$

where $X_i = \mathbf{1}_{A_i}$ and A_i is the event that i falls into a new cluster distinct than the ones containing $1, 2, ...i-1$.

Case $\sigma = -r < 0$ and $\theta = mr$ for some $m = 1, 2, ...$ In this case, as we have previously discussed for n sufficiently large $K_n = m$ almost surely.

Case $\sigma = 0$ and $\theta > 0$ Recall that in this scenario $\{X_i\}_{i=1}^{n}$ are independent Bernoulli distributed random variables, where $X_i \sim Bernoulli\left(\frac{\theta}{i-1+\theta}\right)$. One easy way to see this is to think in the chinese restaurant construction, in that context A_i is the event that the ith customer seats at a new table, same that occurs with probability $\frac{\theta}{i-1+\theta}$ and does not depend on the arrangement of the first $i-1$ customers.

Theorem 7.2. *Let Π_∞ be a $(0, \theta)$-partition of $\mathbb{N}$. Let Π_n be its restriction to $[n]$ and K_n the number of blocks of Π_n, then*

$$\lim_{n\to\infty} \frac{K_n}{\log(n)} = \theta \text{ almost surely and for every } \theta > 0$$

In order to prove the last theorem we will to make use of the following lemma

Lemma 7.2. *Let $X_1, X_2, ...$ be independent random variables such that $\mathbb{E}[X_i^2] < \infty$ and $\sum_{i=1}^{\infty} \frac{1}{(b_i)^2} Var(X_i) < \infty$ for some sequence of non-negative numbers $b_1, b_2, ...$ such that $b_i \to \infty$ as $i \to \infty$. Then as $n \to \infty$*

$$\frac{S_n - \mathbb{E}[S_n]}{b_n} \to 0 \text{ almost surely}$$

where $S_n = \sum_{i=1}^{n} X_i$.

Proof of Theorem 7.2:
It suffices to show that

i) $\sum_{i=2}^{n} \frac{1}{\log(i)^2} Var(X_i)$ is bounded, and

ii) $\frac{\mathbb{E}[K_n]}{\log(n)} \to \theta$

Since $X_i \sim Bernoulli\left(\frac{\theta}{\theta+i-1}\right)$

$$\begin{aligned}
\sum_{i=2}^{n} \frac{1}{\log(i)^2} Var(X_i) &= \sum_{i=1}^{\infty} \frac{1}{\log(i)^2} \left(\frac{\theta}{\theta+i-1}\right)\left(1 - \frac{\theta}{\theta+i-1}\right) \\
&= \theta \sum_{i=2}^{n} (i-1) \frac{1}{[(\theta+i-1)\log(i)]^2} \\
&< \theta \sum_{i=2}^{n} \frac{1}{(i-1)\log(i)^2} \\
&< \theta \left(\frac{1}{\log(2)^2} + \sum_{i=2}^{n-1} \frac{1}{i \log(i)^2}\right)
\end{aligned}$$

and the term on the right side of the last equation is bounded, since

$$\lim_{n\to\infty} \sum_{i=2}^{n-1} \frac{1}{i\log(i)^2} < \infty$$

thus $\sum_{i=1}^{\infty} \frac{1}{(\log(i))^2} Var(X_i) < \infty$ and this proves *i)*. On the other side

$$\begin{aligned}
\frac{\mathbb{E}[K_n]}{\log(n)} &= \frac{\theta}{\log(n)} \sum_{i=1}^{n} \frac{1}{\theta+i-1} \\
&= \frac{1}{\log(n)} + \frac{\theta}{\log(n)} \sum_{i=2}^{n} \frac{1}{\theta+i-1} \\
&= \frac{1}{\log(n)} + \theta + a_n\theta
\end{aligned}$$

where

$$a_n := \frac{1}{\log(n)} \left[\sum_{i=2}^{n} \frac{1}{\theta+i-1} - \log(n)\right]$$

In order to prove that $\dfrac{\mathbb{E}[K_n]}{\log(n)} \to \theta$ it is enough to show that $a_n \to 0$ as n tends to ∞.

As $\theta > 0$

$$\begin{aligned} a_n &< \frac{1}{\log(n)} \left[\sum_{i=2}^{n} \frac{1}{i-1} - \log(n) \right] \\ &= \frac{1}{\log(n)} \left[\sum_{i=1}^{n} \frac{1}{i} - \log(n) \right] - \frac{1}{n \log(n)} =: b_n \end{aligned}$$

Now

$$b_n \log(n) = \left[\sum_{i=1}^{n} \frac{1}{i} - \log(n) \right] - \frac{1}{n} \to \gamma$$

where γ is the Euler's constant, Hence $b_n \to 0$.

Futhermore, let $m = \lfloor \theta \rfloor$, then $\theta - 1 \leq m$, so

$$a_n \geq \frac{1}{\log(n)} \left[\sum_{i=2}^{n} \frac{1}{m+i} - \log(n) \right] =: c_n$$

rewriting c_n, we obtain

$$\begin{aligned} c_n &= \frac{1}{\log(n)} \left[\sum_{i=1}^{n+m} \frac{1}{i} - \sum_{i=1}^{m+1} \frac{1}{i} \right] - 1 + \frac{\log(n+m)}{\log(n)} - \frac{\log(n+m)}{\log(n)} \\ &= \frac{1}{\log(n)} \left[\sum_{i=1}^{n+m} \frac{1}{i} - \log(n+m) \right] - \frac{1}{\log(n)} \sum_{i=1}^{m+1} \frac{1}{i} + \frac{\log(n+m)}{\log(n)} - 1. \end{aligned}$$

It can be easily seen that the right side of the last equation goes to 0 as n tends to ∞. Thus we have shown

$$0 = \lim_{n\to\infty} c_n \leq \lim_{n\to\infty} a_n < \lim_{n\to\infty} b_n = 0$$

this proves *ii)*. By the Lemma 7.2 and *i)*

$$\lim_{n\to\infty} \frac{K_n}{\log(n)} - \frac{\mathbb{E}[K_n]}{\log(n)} = 0 \text{ almost surely}$$

using *ii)* we finally obtain

$$\lim_{n\to\infty} \frac{K_n}{\log(n)} = \theta \text{ almost surely}$$

□

Note that this theorem gives us a way to estimate the parameter θ. For instance, imagine we have a partition Π_∞ of $\mathbb{N}$ which we now that is governed by a $(0, \theta)$ model, where the parameter θ is unknown. Let Π_n be its restriction to $[n]$, if we manage to obtain an observation of $\Pi_n = \pi_n$ for n large enough, then we can obtain an estimator for θ by dividing the the number of clusters in the observed partition π_n by $\log(n)$. That is if $k_n = |\{\text{blocks of } \pi_n\}|$ our estimator of the unknown parameter would be

$$\hat{\theta} = \frac{k_n}{\log(n)}.$$

Case $0 < \sigma < 1$ **and** $\theta > -\sigma$ In this case K_n is the sum of dependent indicators. For instance, consider

$$K_3 = X_1 + X_2 + X_3.$$

In the chinese construction context, the event $X_i = 1$ is the same event that the *ith* customer seats at an occupied table. Now, $X_1 = 1$ a.s., if the second customer seats at a new table ($X_2 = 1$) then $K_2 = 2$ and given this event $X_3 \sim \textit{Bernoulli}\left(\frac{\theta+2\sigma}{2+\theta}\right)$ that is

$$\mathbb{P}[X_3 = 1 | X_1 = 1, X_2 = 1] = \frac{\theta + 2\sigma}{2 + \theta}$$

on the other hand, if the second customer seats at the same table of the first customer ($X_2 = 0$) then $K_2 = 2$ and conditionally on this event $X_3 \sim \textit{Bernoulli}\left(\frac{\theta+\sigma}{2+\theta}\right)$, i.e.

$$\mathbb{P}[X_3 = 1 | X_1 = 1, X_2 = 0] = \frac{\theta + \sigma}{2 + \theta}.$$

This dependence between the X_is makes things slightly more complicated. In order to obtain a right normalization for a limit law let us study the behaviour of the stochastic process $\{K_n\}_{n \geq 1}$.

Proposition 7.1. *Let $\mathbb{P}_{(\sigma,\theta)}$ govern Π_∞ as a (σ, θ) partition. Let K_n be the number of blocks of Π_n.*

i) $\{K_n\}_{n \geq 1}$ is a Markov chain with initial state $K_1 = 1$ with increments $\{0, 1\}$, and inhomogeneous transition probabilities

$$\mathbb{P}_{(\sigma,\theta)}[K_{n+1} = k + 1 | K_n = k] = \frac{\theta + k\sigma}{n + \theta}$$
$$\mathbb{P}_{(\sigma,\theta)}[K_{n+1} = k | K_n = k] = \frac{n - k\sigma}{n + \theta}$$

ii) The distribution on K_n is given by

$$\mathbb{P}_{(\sigma,\theta)}[K_n = k] = \frac{(\theta + \sigma)_{k-1\uparrow\sigma}}{(\theta + 1)_{n-1\uparrow}} S_{n,k}^{-1,-\sigma}$$

where $S^{-1,-\sigma}_{n,k}$ is a generalized Stirling number as defined towards the end of Chapter 1.2.

iii) The expectation of K_n in the case $\sigma \neq 0$ is

$$\mathbb{E}_{(\sigma,\theta)}\left[K_n\right] = \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{\theta}{\sigma} \tag{7.16}$$

Proof:

As $\Pi_1 = \{\{1\}\}$ a.s. then clearly $K_1 = 1$ a.s. Now, to prove the rest of i) one may consider the chinese restaurant construction, in this context the event $K_{n+1} = k+1$ given $K_n = k$ is read as the $(n+1)$th costumer to arrive is seated at a new table, given that there are already k tables occupied by the first n costumers, thus evidently

$$\mathbb{P}_{(\sigma,\theta)}[K_{n+1} = k+1 | K_n = k] = \frac{\theta + k\sigma}{n+\theta}.$$

In the same context the event $K_{n+1} = k$ given $K_n = k$, can be read as the $(n+1)$th costumer seats at an occupied table, given that the n first costumers are currently occupying a total of k tables. To compute this probability let n_j be the number of costumers seating at table j after the n costumer arrived. Note that the probability of this event can be regarded as the sum over $j \in [k]$ of the probability that the $(n+1)$th costumer seats at table j (taking into account that there k occupied tables and n_j persons seating at table j), thus

$$\mathbb{P}_{(\sigma,\theta)}[K_{n+1} = k | K_n = k] = \sum_{j=1}^{k} \frac{n_j - \sigma}{n+\theta} = \frac{n-k\sigma}{n+\theta}.$$

This proves i). In order to prove ii) recall that

$$\mathbb{P}_{(\sigma,\theta)}[\Pi_n = \pi] = \frac{(\theta+\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}}$$

where π is a partition of $[n]$ having k blocks of sizes $n_1, ..., n_k$, so

$$\begin{aligned}
\mathbb{P}_{(\sigma,\theta)}[K_n = k] &= \sum_{\pi\in\mathcal{P}^k_{[n]}} \mathbb{P}_{(\sigma,\theta)}[\Pi_n = \pi] \\
&= \sum_{\pi\in\mathcal{P}^k_{[n]}} \frac{(\theta+\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}} \\
&= \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow 1}} \sum_{\pi\in\mathcal{P}^k_{[n]}} \prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1} \\
&= \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow 1}} B_{n,k}(w) = \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k}
\end{aligned}$$

with $w = (w_1, w_2, ...)$ such that $w_i = (1-\sigma)_{i-1\uparrow}$ and $S^{-1,-\sigma}_{n,k}$ is a generalized Stirling number, see the Chapter 1 for details, in particular the Sections 1.1 and 1.2 about composite structures and Stirling numbers.

It remains the computing of the expectation of K_n. From ii)

$$\begin{aligned}
\mathbb{E}_{(\sigma,\theta)}\left[K_n\right] &= \sum_{k=1}^{n} k\mathbb{P}_{(\sigma,\theta)}[K_n = k] \\
&= \sum_{k=1}^{n} k\frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k} \\
&= \sum_{k=1}^{n} \frac{k\sigma(\theta+\sigma)(\theta+2\sigma)\cdots(\theta+(k-1)\sigma)}{\sigma(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k} \\
&= \sum_{k=1}^{n} \frac{(\theta+k\sigma)(\theta+\sigma)(\theta+2\sigma)\cdots(\theta+(k-1)\sigma)}{\sigma(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k} \\
&\quad - \sum_{k=1}^{n} \frac{(\theta)(\theta+\sigma)(\theta+2\sigma)\cdots(\theta+(k-1)\sigma)}{\sigma(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k} \\
&= \sum_{k=1}^{n} \frac{(\theta+\sigma)_{k\uparrow\sigma}}{\sigma(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k} - \sum_{k=1}^{n} \frac{(\theta)_{k\uparrow\sigma}}{\sigma(\theta+1)_{n-1\uparrow}} S^{-1,-\sigma}_{n,k}
\end{aligned}$$

Using the identity for the generalized Stirling numbers

$$(x)_{n\downarrow\alpha} = \sum_{k=1}^{n} S^{\alpha,\beta}_{n,k}(x)_{k\downarrow\beta}$$

with $\alpha = -1$ and $\beta = -\sigma$, i.e.

$$(x)_{n\uparrow} = \sum_{k=1}^{n} S^{-1,-\sigma}_{n,k}(x)_{k\uparrow\sigma}$$

we obtain

$$\begin{aligned}
\mathbb{E}_{(\sigma,\theta)}\left[K_n\right] &= \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{(\theta)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} \\
&= \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{(\theta)(\theta+1)\cdots(\theta+n-1)}{\sigma(\theta+1)(\theta+2)\cdots(\theta+n-1)} \\
&= \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{\theta}{\sigma}
\end{aligned}$$

□

Now, using the equation (7.16), the relationship between the ascending factorial and the gamma function, and Stirling's approximation for the gamma function we get

$$
\begin{aligned}
\mathbb{E}_{(\sigma,\theta)}\left[K_n\right] &= \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{\theta}{\sigma} \\
&= \frac{\frac{\Gamma(\theta+\sigma+n)}{\Gamma(\theta+\sigma)}}{\sigma\frac{\Gamma(\theta+n)}{\Gamma(\theta+1)}} - \frac{\theta}{\sigma} \\
&= \left[\frac{\Gamma(\theta+1)}{\sigma\Gamma(\theta+\sigma)}\right]\left[\frac{\Gamma(\theta+\sigma+n)}{\Gamma(\theta+n)}\right] - \frac{\theta}{\sigma} \\
&\sim \frac{\Gamma(\theta+1)}{\sigma\Gamma(\theta+\sigma)} n^{\sigma} (\text{as } n \to \infty)
\end{aligned}
$$

this indicates the right normalization for a limit law.

Theorem 7.3. *For $0 < \sigma < 1$, $\theta > -\sigma$, under $\mathbb{P}_{(\sigma,\theta)}$, as $n \to \infty$*

$$\frac{K_n}{n^{\sigma}} \to S_{\sigma} \text{ almost surely}$$

and in the pth mean (for $p > 0$), for a positive random variable S_{σ} , with continuous density

$$\frac{d}{ds}\mathbb{P}_{(\sigma,\theta)}[S_{\sigma} \in ds] = g_{(\sigma,\theta)}(s) := \frac{\Gamma(\theta+1)}{\Gamma(\frac{\theta}{\sigma}+1)} s^{\frac{\theta}{\sigma}} g_{\sigma}(s) \quad (s > 0)$$

where $g_{\sigma} = g_{(\sigma,0)}$ is the Mittag-Leffler density *of the $\mathbb{P}_{(\sigma,0)}$ distribution of S_{σ}, whose pth moment is*

$$\frac{\Gamma(p+1)}{\Gamma(p\sigma+1)}.$$

that is

$$g_{\sigma}(s) = \frac{f_{\sigma}(s^{-\frac{1}{\sigma}})}{\sigma s^{1+\frac{1}{\sigma}}}$$

where f_{σ} denotes the probability density function of a positive σ-stable random variable.

Before proving the last theorem let us prove the following proposition.

Proposition 7.2. *For $p > 0$, let $[k]_p = \frac{\Gamma(k+p)}{\Gamma(k)}$ so that $[k]_p = (k)_{p\uparrow}$ for $p = 1, 2, \dots$. For $0 < \alpha < 1$, and real $p > 0$,*

$$\mathbb{E}_{(\sigma,0)}\left[[K_n]_p\right] = \frac{\Gamma(p)[p\sigma]_n}{\sigma\Gamma(n)}. \tag{7.17}$$

Proof:

$$
\begin{aligned}
\mathbb{E}_{(\sigma,0)}\left[[K_n]_p\right] &= \sum_{k=1}^{n}[k]_p \mathbb{P}_{(\sigma,0)}[K_n = k] \\
&= \sum_{k=1}^{n} \frac{\Gamma(k+p)}{\Gamma(k)} \frac{(\sigma)_{k-1\uparrow\sigma}}{(1)_{n-1\uparrow}} S_{n,k}^{-1,-\sigma} \\
&= \sum_{k=1}^{n} \frac{\Gamma(k+p)}{\Gamma(k)} \frac{(k-1)!\sigma^{k-1}}{(n-1)!} S_{n,k}^{-1,-\sigma} \\
&= \sum_{k=1}^{n} \frac{\Gamma(k+p)}{\Gamma(k)} \frac{\Gamma(k)\sigma^{k}}{\sigma\Gamma(n)} S_{n,k}^{-1,-\sigma} \\
&= \frac{\Gamma(p)}{\sigma\Gamma(n)} \sum_{k=1}^{n} \frac{\Gamma(k+p)}{\Gamma(p)} \sigma^k S_{n,k}^{-1,-\sigma} \\
&= \frac{\Gamma(p)}{\sigma\Gamma(n)} \sum_{k=1}^{n} (p)_{k\uparrow} \sigma^k S_{n,k}^{-1,-\sigma} \\
&= \frac{\Gamma(p)}{\sigma\Gamma(n)} \sum_{k=1}^{n} (p\sigma)_{k\uparrow\sigma} S_{n,k}^{-1,-\sigma} \\
&= \frac{\Gamma(p)}{\sigma\Gamma(n)} (p\sigma)_{n\uparrow} \\
&= \frac{\Gamma(p)[p\sigma]_n}{\sigma\Gamma(n)}
\end{aligned}
$$

Proof of Theorem 7.3:

Let $0 < \sigma < 1$. Let $\mathscr{F}_n$ be the field of events generated by Π_n. Recall

$$
\mathbb{P}_{(\sigma,\theta)}[\Pi_n = \pi] = \frac{(\theta+\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}}
$$

so that

$$
\mathbb{P}_{(\sigma,0)}[\Pi_n = \pi] = \frac{(\sigma)_{k-1\uparrow\sigma}\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(1)_{n-1\uparrow 1}}
$$

which gives the likelihood ratio

$$
M_{\sigma,\theta,n} := \left.\frac{d\mathbb{P}_{(\sigma,\theta)}}{d\mathbb{P}_{(\sigma,0)}}\right|_{\mathscr{F}_n} = \frac{\frac{(\theta+\sigma)_{K_n-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow 1}}}{\frac{(\sigma)_{K_n-1\uparrow\sigma}}{(1)_{n-1\uparrow 1}}} = \frac{\frac{(\theta+\sigma)_{K_n-1\uparrow\sigma}}{(\sigma)_{K_n-1\uparrow\sigma}}}{\frac{(\theta+1)_{n-1\uparrow 1}}{(1)_{n-1\uparrow 1}}} = \frac{f_{\sigma,\theta}(K_n)}{f_{1,\theta}(n)}
$$

where, for $\theta > -\sigma$

$$
f_{\sigma,\theta}(k) := \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\sigma)_{k-1\uparrow\sigma}}
$$

Thus for $\theta > -\sigma$,

$\{M_{\sigma,\theta,n}\}_{n=1,2,\ldots}$ is a positive $\mathbb{P}_{(\sigma,0)}$-martingale respect to $\{\mathscr{F}_n\}_{n=1,2,\ldots}$

(see Chapter **??** for a proof). By the martingale convergence theorem, $M_{\sigma,\theta,n}$ has a limit $M_{\sigma,\theta}$ almost surely ($\mathbb{P}_{(\sigma,0)}$) as $n \to \infty$. Now, note that

$$\begin{aligned} f_{\sigma,\theta}(k) &= \frac{(\theta+\sigma)(\theta+2\sigma)\cdots(\theta+(k-1)\sigma)}{(\sigma)(2\sigma)\cdots((k-1)\sigma)} \\ &= \frac{\left(\frac{\theta}{\sigma}+1\right)\left(\frac{\theta}{\sigma}+2\right)\cdots\left(\frac{\theta}{\sigma}+(k-1)\right)}{\Gamma(k)} \\ &= \frac{\left(\frac{\theta}{\sigma}+1\right)_{k-1\uparrow}}{\Gamma(k)} \\ &= \frac{\Gamma\left(\frac{\theta}{\sigma}+k\right)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(k)} \sim \frac{k^{\frac{\theta}{\sigma}}}{\Gamma\left(\frac{\theta}{\sigma}+1\right)} \text{ (as } k \to \infty) \end{aligned}$$

Theorem 6.2 tells us the Π_∞ has infinitely many blocks with strictly positive frequencies, so that $K_n \to \infty$ almost surely as $n \to \infty$, hence

$$M_{\sigma,\theta,n} = \frac{f_{\sigma,\theta}(K_n)}{f_{1,\theta}(n)} = \left(\frac{\Gamma\left(\frac{\theta}{\sigma}+K_n\right)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(K_n)}\right)\left(\frac{\Gamma(\theta+1)\Gamma(n)}{\Gamma(\theta+n)}\right) \sim \frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)}\left(\frac{K_n}{n^\sigma}\right)^{\frac{\theta}{\sigma}} \tag{7.18}$$

Moreover, the ratio of the two sides in last equation is bounded away from 0 and ∞. Using equation (7.17) we get that the martingale $\{M_{\sigma,\theta,n}\}_{n\geq 1}$ is bounded in $L^p(\mathbb{P}_{(\sigma,0)})$ and hence convergent in $L^p(\mathbb{P}_{(\sigma,0)})$ to $M_{(\sigma,\theta)}$ for every $p > 1$. Also, as

$$\begin{aligned} \mathbb{E}_{(\sigma,0)}[M_{\sigma,\theta,n}] &= \frac{\Gamma(n)\Gamma(\theta+1)\mathbb{E}_{(\sigma,0)}\left[[K_n]_{\frac{\theta}{\sigma}}\right]}{\Gamma(\theta+n)\Gamma\left(\frac{\theta}{\sigma}+1\right)} \\ &= \frac{\Gamma(n)\Gamma(\theta+1)\Gamma\left(\frac{\theta}{\sigma}\right)[\theta]_n}{\Gamma(\theta+n)\Gamma\left(\frac{\theta}{\sigma}+1\right)\sigma\Gamma(n)} \\ &= \frac{\Gamma(\theta+1)\Gamma\left(\frac{\theta}{\sigma}\right)}{\sigma\Gamma\left(\frac{\theta}{\sigma}+1\right)\Gamma(\theta)} = \frac{(\theta)_{1\uparrow}}{\sigma\left(\frac{\theta}{\sigma}\right)_{1\uparrow}} = \frac{\theta}{\sigma\left(\frac{\theta}{\sigma}\right)} = 1 \end{aligned}$$

then

$$\mathbb{E}_{(\sigma,0)}\left[M_{\sigma,\theta}\right] = 1 \tag{7.19}$$

Furthermore, using (7.18)

$$\frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)}\left(\frac{K_n}{n^\sigma}\right)^{\frac{\theta}{\sigma}} \to M_{\sigma,\theta} = \frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)}(S_\sigma)^{\frac{\theta}{\sigma}} \tag{7.20}$$

almost surely ($\mathbb{P}_{(\sigma,0)}$) and in $L^p(\mathbb{P}_{(\sigma,0)})$, where $S_\sigma := \frac{M_{\sigma,\sigma}}{\Gamma(\sigma+1)}$. Now equations (7.19) and (7.20) yield the moments of the $\mathbb{P}_{(\sigma,0)}$ distribution of S_σ, since these moments are those of the Mittag-Leffler distributions, the conclusions of the theorem in the case $\theta = 0$ are evident. The corresponding results for the case $\theta > -\sigma$ follow the results for $\theta = 0$ due to the below corollary of the above martingale argument.

□

Corolary 7.2. *Let $\mathbb{P}_{(\sigma,\theta)}$ denote the distribution on $\mathcal{P}_{\mathbb{N}}$ of a (σ,θ)-partition $\Pi_\infty := \{\Pi_n\}_{n\geq 1}$. For each $0<\sigma<1$ and $\theta>-\sigma$, the laws $\mathbb{P}_{(\sigma,\theta)}$ and $\mathbb{P}_{(\sigma,0)}$ are mutually absolutely continuous with density*

$$\frac{d\mathbb{P}_{(\sigma,\theta)}}{d\mathbb{P}_{(\sigma,0)}} = \frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)}(S_\sigma)^{\frac{\theta}{\sigma}}$$

where S_σ is the almost sure limit of $\frac{K_n}{n^\sigma}$ under $\mathbb{P}_{(\sigma,\theta)}$ for every $\theta>-\sigma$, and K_n denotes the number of blocks of Π_n.

The limit random variable

$$S_\sigma := \lim_{n\to\infty} \frac{K_n}{n^\sigma}$$

plays a key role in the further analysis of the asymptotic properties of a (σ,θ)-partition Π_∞, see [18] for a insight on this analysis. This leads us to the following definition

Definition 7.1. *We say that Π_∞, an exchangeable partition on $\mathbb{N}$ has σ-diversity S_σ, if the limit*

$$S_\sigma := \lim_{n\to\infty} \frac{K_n}{n^\sigma}$$

exists and is positive an finite almost surely.

Chapter 8

Gibbs partitions, the two-parameter model and Gibbs-type partitions

8.1 Gibbs partitions

Let us introduce a wide family of exchangeable random partitions called *Gibbs partitions*. As in Section 1.1 let $(V \circ W)([n])$ denote the set of all composite $V \circ W$ structures built over $[n]$ for some species of combinatorial structures V and W. Recall that v_j and w_j denote the number of $V-$structures and $W-$structures respectively on a set with j elements, that is $v_j = |V([j])|$ and $w_j = |W([j])|$. Also recall that

$$|(V \circ W)([n])| = B_n(v,w) := \sum_{k=1}^{n} v_k B_{n,k}(w)$$

where

$$B_{n,k}(w) := \sum_{\{A_1,...,A_k\} \in \mathcal{P}_{[n]}^k} \prod_{i=1}^{k} w_{|A_i|}$$

Let us choose uniformly one composite structure from $(V \circ W)([n])$ and let Π_n be the random partition over $[n]$ generated by the blocks of this random composite structure. Then for every particular partition $\pi = \{B_1, ..., B_k\}$ of $[n]$, the number $\mathbb{P}[\Pi_n = \pi]$ will only depend on the numbers of blocks k of π, on $|B_i|$ for $i = 1, 2, ..., k$, and on the sequences of positive numbers $v = (v_1, v_2, ...)$ and $w = (w_1, w_2, ...)$. Thus

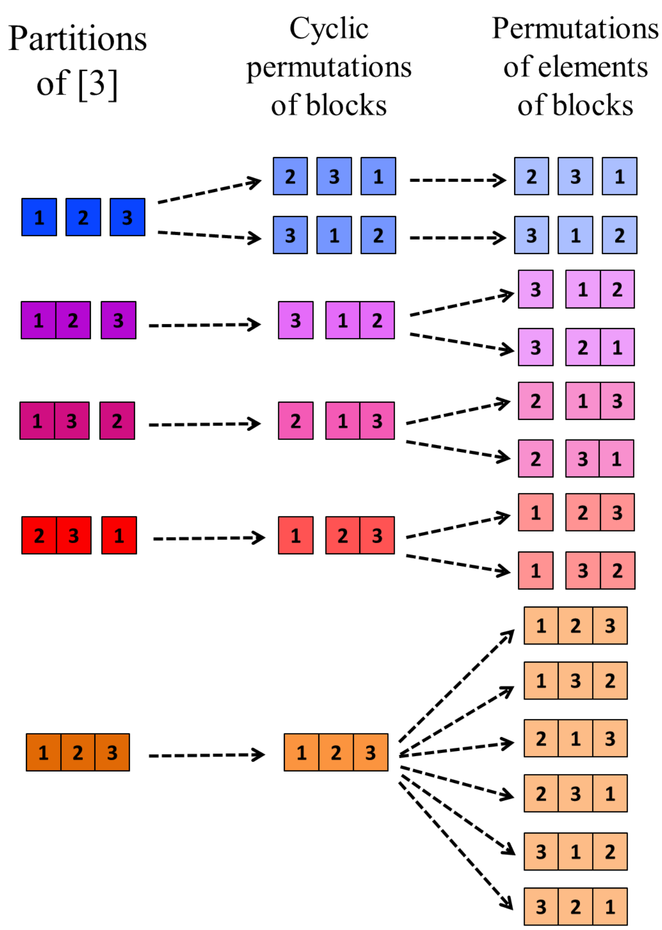
$$\mathbb{P}[\Pi_n = \pi] = p(|B_1|, |B_2|, ..., |B_k|)$$

where for every composition $\{n_1, ..., n_k\}$ of n

$$p(n_1, n_2, ..., n_k) = \frac{v_k \prod_{i=1}^{k} w_{n_i}}{B_n(v, w)}$$

Example 8.1 (Example 1.3 continue)**.** *Recall Example 1.3 where we considered the set $(V \circ W)([3])$ where V is a cyclic permutation and W is a permutation (not necessarily cyclic) so for every $j \in \mathbb{N}$ $v_j = V([j]) = (j-1)!$ and $w_j = W([j]) = j!$.*

Each element in $(V \circ W)([3])$ can be represented by a path (starting in the first column and ending in the third one) on the next figure



Thus, choosing uniformly an element of $(V \circ W)([3])$ is equivalent to uniformly choosing an element of the third column of the last figure. In order to locate the random partition corresponding to this random composite structure *we just have to travel backwards on the path to see from which element of the first column did the path came from.*

We have seen that $|(V \circ W)([3])| = B_3(v,w) = 14$ *so each composite structure will have a probability of* $\frac{1}{14}$ *of occurring. From the figure it is clear that there are* 2 *composite structures that lead to* $\{\{1\},\{2\},\{3\}\}$, 2 *leading to* $\{\{1\},\{2,3\}\}$, 2 *more that correspond to* $\{\{2\},\{1,3\}\}$, 2 *corresponding to* $\{\{3\},\{1,2\}\}$ *and finally* 6 *composite structures that generate* $\{\{1,2,3\}\}$. *Thus*

$$\mathbb{P}[\Pi_3 = \{\{1\},\{2\},\{3\}\}] = \frac{v_3 * (w_1 * w_1 * w_1)}{B_3(v,w)} = \frac{2}{14} = \frac{1}{7}$$
$$\mathbb{P}[\Pi_3 = \{\{1\},\{2,3\}\}] = \frac{v_2 * (w_1 * w_2)}{B_3(v,w)} = \frac{2}{14} = \frac{1}{7}$$
$$\mathbb{P}[\Pi_3 = \{\{2\},\{1,3\}\}] = \frac{v_2 * (w_1 * w_2)}{B_3(v,w)} = \frac{2}{14} = \frac{1}{7}$$
$$\mathbb{P}[\Pi_3 = \{\{3\},\{1,2\}\}] = \frac{v_2 * (w_1 * w_2)}{B_3(v,w)} = \frac{2}{14} = \frac{1}{7}$$
$$\mathbb{P}[\Pi_3 = \{\{1,2,3\}\}] = \frac{v_1 * (w_3)}{B_3(v,w)} = \frac{6}{14} = \frac{3}{7}$$

This leads us to the following definition.

Definition 8.1. *Let* $v = (v_1, v_2, ...)$ *and* $w = (w_1, w_2, ...)$ *be two sequences of non-negative numbers. Let* Π_n *be a random partition taking values in* $\mathcal{P}_{[n]}$. *We call* Π_n *a* $Gibbs_{[n]}(v,w)$ *partition if for every* $\pi \in \mathcal{P}_{[n]}$ *we have that*

$$\mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k) = \frac{v_k \prod_{i=1}^{k} w_{n_i}}{B_n(v,w)}$$

where k *is the number of blocks in* π *and* n_i *is the number of elements in the* i*th block of* π.

Note that the number as $\prod_{i=1}^{k} w_{n_i}$ is a symmetric function of $(n_1, ..., n_k)$ then $p_n(n_1, n_2, ..., n_k)$ is also a symmetric function of $(n_1, ..., n_k)$ so, if Π_n is a $Gibbs_{[n]}(v,w)$ partition, then it turns out to be exchangeable and its EPPF is given by

$$p(n_1, ..., n_k) = v_{n,k} \prod_{i=1}^{k} w_{n_i}$$

where $v_{n,k} := \frac{v_k}{B_n(v,w)}$.

One might wonder if, given the sequences of non-negative numbers v and w, the family $\{\Pi_n\}_{n \geq 1}$, where Π_n is a $Gibbs_{[n]}(v,w)$ partition, is consistent. The answer is no, consider for instance, $v = (1,1,...)$ and $w = (1,1,...)$ then

$$p(n_1, ..., n_k) = \frac{1}{B_n}$$

where B_n is the Bell number. That is Π_n is uniformly distributed over $\mathcal{P}_{[n]}$ for every $n \geq 1$, in Example 2.1 we have already seen that this family is not consistent.

Physical interpretation

Suppose that we have n labelled particles $q_1, ..., q_n$ and that they are partitioned into clusters in such way that every particle belongs to one unique cluster, then clearly this collection of clusters is represented by a partition of $[n]$ in such way that q_i and q_j belong to the same cluster iff i and j belong to the same block of the partition of $[n]$. Furthermore, assume that, for a sequence of non-negative numbers $w = (w_1, w_2, ...)$, each cluster with j elements can be in one of w_j different internal states. Let the *configuration* of the system be the partition of the set of $[n]$ together with the assignment of a internal state to each cluster. Then for each partition $\pi \in \mathcal{P}_{[n]}$ with k blocks, where the jth block has a size of n_j, there are $\prod_{i=1}^{k} w_{n_i}$ different configurations of the system corresponding to π.

Let $v = (v_1, v_2, ...)$ such that $v_j = \mathbf{1}_{\{j=k\}}$. Let Π_n be a $Gibbs_{[n]}(v, w)$ partition, then for any $\pi \in \mathcal{P}_{[n]}^k$ with block sizes $n_1, ..., n_k$

$$\mathbb{P}[\Pi_n = \pi] = \frac{\prod_{i=1}^{k} w_{n_i}}{B_{n,k}(w)}$$

and for $\pi \in \mathcal{P}_{[n]} \setminus \mathcal{P}_{[n]}^k$

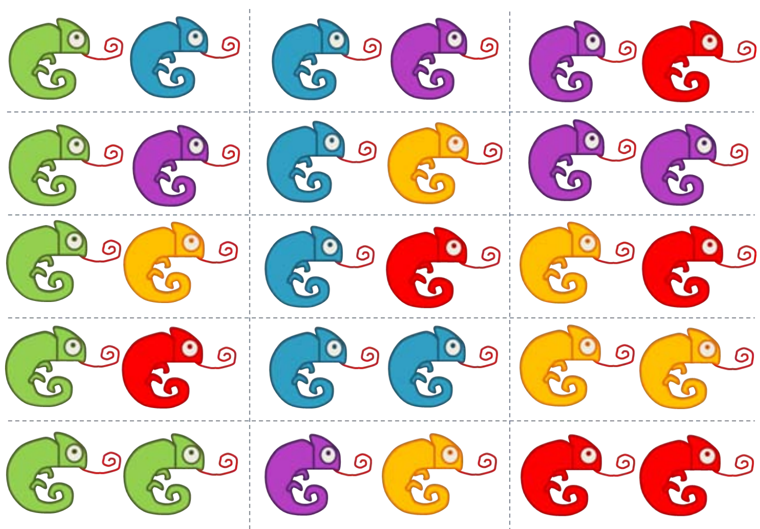
$$\mathbb{P}[\Pi_n = \pi] = 0$$

This distribution, restricted to $\mathcal{P}_{[n]}^k$ is known in the physics literature as a *Microcanonical Gibbs state*, and we will denote this distribution over $\mathcal{P}_{[n]}^k$ by $Gibbs_{[n]}^{(k)}(w)$.

Example 8.2. *Imagine we have n chameleons such that each chameleon can paint itself in one of m colours, and we also have $1 \leq k \leq n$ very big cages, the chameleons are to be placed into the k cages in such way that every cage has at least one chameleon. In this scenario each of the colouring schemes of the chameleons within each cage is what we called before the internal states. For instance a cage having 1 chameleon can be in 1 of m distinct internal states, determined by the colour of the chameleon in it. If $m = 5$ all the possible internal states of a cage having 1 chameleon are illustrated below.*

5 distinct possible internal states of cages having 1 chameleon

Again if $m = 5$, for a cage having 2 chameleons the 15 possible internal states are shown in the picture below, note that if the internal state is blue − green it does not matter which of the 2 chameleons is blue and which one is green, as long as one of them is blue and the other one green.



Similarly for a cage having n_i chameleons such that each one can paint itself into m distinct colours, the number possible internal states of such case is

$$w_{n_i} := \binom{n_i + m - 1}{m - 1}.$$

If the chameleons are placed into the cages one to one, this generates a partition $\Pi_{n,k}$ of $[n]$ into exactly k blocks determined by the following equivalence relation $i \sim j$ iff the ith chameleon and the jth chameleon are placed into the same cage. The partition of $[n]$ together with the internal state of each of the k clusters is the configuration of the system. If we assume that each of these configurations have the same probability of occurring then clearly $\Pi_{n,k}$ is a Gibbs$_{[n]}^{(k)}(w)$ partition with $w = (w_1, w_2, ...)$ where

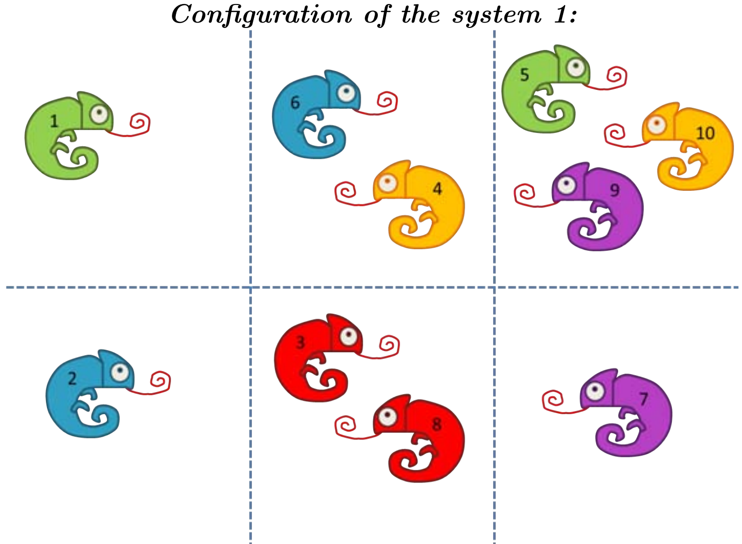
$$w_i = \binom{i + m - 1}{m - 1}$$

That is, if $\pi \in \mathcal{P}^k_{[n]}$ is such that π has k blocks with frequencies $n_1, ..., n_k$, then

$$\mathbb{P}[\Pi_{n,k} = \pi] = \frac{\prod_{i=1}^{k} w_{n_i}}{B_{n,k}(w)}$$

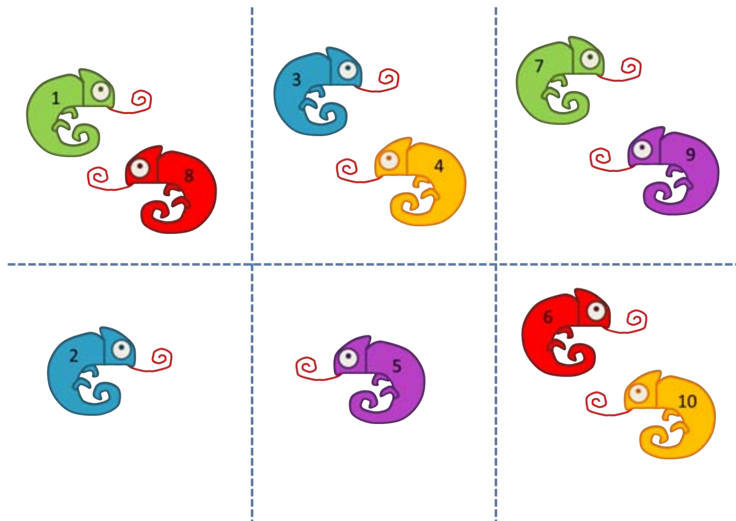
as there are exactly $\prod_{i=1}^{k} w_{n_i}$ configurations of the system corresponding to π. If $n = 10$, $m = 5$ and $k = 6$, 3 possible configurations of the system together with the partition associated to them are shown bellow.

Configuration of the system 1:



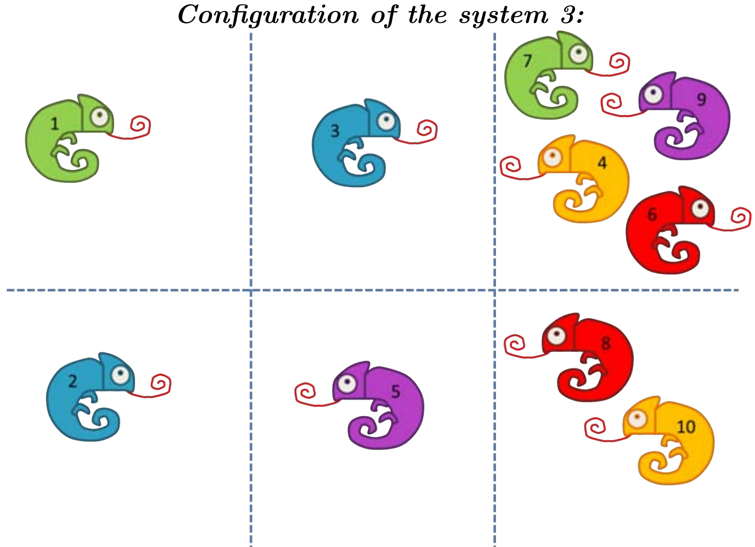
Corresponding partition of [10]: {{1},{2},{3,8},{4,6},{5,9,10},{7}}

Configuration of the system 2:



Corresponding partition of [10]: $\{\{1,8\},\{2\},\{3,4\},\{5\},\{6,10\},\{7,9\}\}$

Configuration of the system 3:



Corresponding partition of [10]: $\{\{1\},\{2\},\{3\},\{5\},\{4,6,7,9\},\{8,10\}\}$

A general weight sequence $v = (v_1, v_2 ...)$ randomizes k to allow any probabilistic mixture of these n microcanonical Gibbs states. That is if $\Pi_{n,k}$ is a

$Gibbs^{(k)}_{[n]}(w)$ partition, and Π_n is a $Gibbs_{[n]}(v,w)$ partition then

$$\Pi_n = \begin{cases} \Pi_{n,1} & \text{with probability } p_1 \\ \Pi_{n,2} & \text{with probability } p_2 \\ & \vdots \\ \Pi_{n,n} & \text{with probability } p_n \end{cases}$$

where for every $k \in \{1, ..., n\}$

$$p_k = \sum_{\{A_1,\ldots,A_k\}\in\mathcal{P}^k_{[n]}} \frac{v_k \prod_{i=1}^k w_{|A_i|}}{\sum_{k=1}^n v_k \sum_{\{A_1,\ldots,A_k\}\in\mathcal{P}^k_{[n]}} \prod_{i=1}^k w_{|A_i|}} = \frac{v_k B_{n,k}(w)}{\sum_{k=1}^n v_k B_{n,k}(w)}$$

clearly the $p_k \geq 0$ and $\sum_{k=1}^n p_k = 1$.

In other words, if K_n is the random number of blocks then

$$\Pi_n = \sum_{k=1}^n \Pi_{n,k} \mathbf{1}_{\{K_n=k\}}$$

and

$$\mathbb{P}[K_n = k] = p_k$$

Thus for w and n fixed, the set of all $Gibbs_{[n]}(v,w)$ distributions on partitions of $[n]$, as v varies, is a $(n-1)$-dimensional simplex whose set of extreme points is the colletion $\left\{Gibbs^{(k)}_{[n]}(w)\right\}_{k=1}^n$ of microcanonical Gibbs states.

Example 8.3. *Let us consider the same setting as in Example 8.2, with the difference that now, the number of available cages K_n is random (recall that n is the number of chameleons to be placed in the K_n cages). Once the number of cages $K_n = k$ is fixed, it all works out as in Example 8.2. Let $p_1, ..., p_n$ be such that $0 \leq p_i \leq 1$ for all $i \in \{1, 2, ..., n\}$ and $\sum_{i=1}^n p_i = 1$. Also assume that for $k = 1, 2, ..., n$.*

$$\mathbb{P}[K_n = k] = p_k$$

For instance if we only have 3 chameleons, such that each one can paint itself in either color blue or green, then

1. *If $K_3 = 3$, this means that there are 3 boxes available, considering the above mentioned restriction that every box has got to contain at least 1 chameleon then the generated partition of $[3]$ has got to be $\Pi_{3,3} = \{\{1\},\{2\},\{3\}\}$. Also note that each block containing one element can be in 1 of 2 internal states, determined by whether the chameleon is blue or green so overall the number of configurations of the system corresponding to $\{\{1\},\{2\},\{3\}\}$ is $2^3 = 8$.*

2. *If $K_3 = 2$ then we have 2 cages to be filled with three chameleons, this means that in one cage there will be 1 chameleon and in the other one there will be 2 of them, so that the generated partition of [3] has got to be one of the following 3*

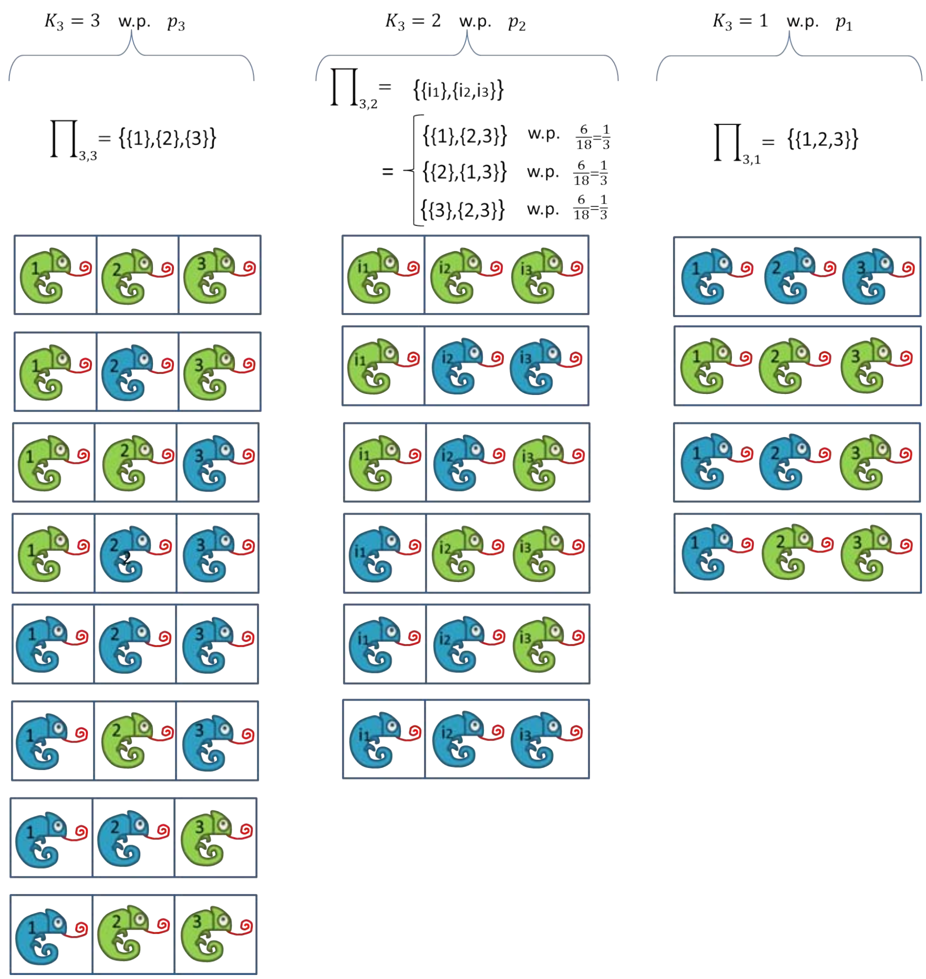
$$\{\{1\},\{2,3\}\}, \quad \{\{2\},\{1,3\}\}, \ or \ \{\{3\},\{1,2\}\}$$

As we mentioned before a cluster having 1 chameleon can be in 1 of 2 internal states, as for a cluster having 2 chameleons can be in 1 of $\binom{2+1}{1} = 3$ internal states so for a fixed partition $\{\{i_1\},\{i_2,i_3\}\}$ there are $2(3) = 6$ internal states corresponding to it. Overall there will be 18 configurations of the system given $K_3 = 2$. If we choose uniformly between one of this configurations and then set $\Pi_{3,2}$ to be the partition corresponding to such configuration then it is clear that

$$\Pi_{3,2} = \begin{cases} \{\{1\},\{2,3\}\} & \textit{with probability } \frac{6}{18} = \frac{1}{3} \\ \{\{2\},\{1,3\}\} & \textit{with probability } \frac{6}{18} = \frac{1}{3} \\ \{\{3\},\{1,2\}\} & \textit{with probability } \frac{6}{18} = \frac{1}{3} \end{cases}$$

3. *If $K_3 = 1$, this means that there is only 1 cage available, so that all the chameleons are going to be placed in it, clearly the generated partition has got to be $\Pi_{3,1} = \{\{1,2,3\}\}$. Observe that a cluster having 3 chameleons can be in 1 of $\binom{3+1}{1} = 4$ internal states, determined by the number of chameleons that are blue and green.*

This is illustrated in the following image:



Now let Π_3 be the random partition of $[n]$ generated by the procedure of the 3 chameleons placed in the K_3 available boxes, that is

$$\Pi_3 = \Pi_{3,1}\mathbf{1}_{\{K_3=1\}} + \Pi_{3,2}\mathbf{1}_{\{K_3=2\}} + \Pi_{3,3}\mathbf{1}_{\{K_3=3\}}$$

and where

$$\mathbb{P}[K_3 = k] = p_k \quad k = 1, 2, 3.$$

Observe that Π_3 is a $Gibbs_{[3]}(v, w)$ partition iff there exist $v = (v_1, v_2, ...)$ such that the following equations hold for every $k = 1, 2, 3$.

$$p_k = \frac{v_k B_{3,k}(w)}{\sum_{k=1}^{3} v_k B_{3,k}(w)}$$

where $w = (w_1, w_2, ...)$ is such that $w_i = \binom{i+1}{1} = i + 1$, and v_k does not depend on $n = 3$.

Before leaving this example behind, realize one last thing, if instead of fixing p_1, p_2, p_3 *we let* $v = (v_1, v_2, ...) = (1, 1, ...)$ *and then compute* p_k *through*

$$p_k = \frac{v_k B_{3,k}(w)}{\sum_{k=1}^{3} v_k B_{3,k}(w)} = \frac{B_{3,k}(w)}{\sum_{k=1}^{3} B_{3,k}(w)}$$

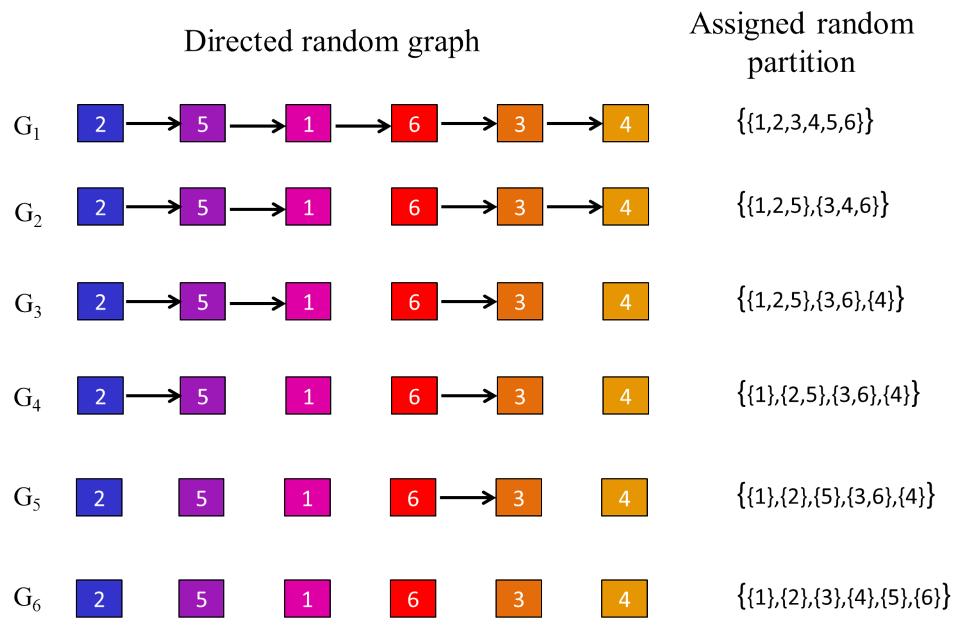
then Π_3 *is equally distributed as the partition* Π *generated by uniformly choosing one of the* $8{+}3(6){+}4 = 30$ *configurations of the system exhibited in the last figure and then setting* Π *to be the partition corresponding to the chosen configuration.*

Example 8.4 (Cutting a rooted random segment)**.** *Suppose that the internal state of a cluster* C *of size* j *is one of* $w_j = j!$ *linear orderings of* C*. Where if, for instance,* $C = \{a, b, c\}$ *then the set of possible linear orderings of* C *is*

$$\{[a \to b \to c], [a \to c \to b], [b \to a \to c], [b \to c \to a], [c \to a \to b], [c \to b \to a]\}$$

We can identify each ordering of a cluster with a directed graph in which there is an arrow from a *to* b *iff* a *is immediate predecessor of* b *in the linear ordering. We call such a graph a* rooted segment*. Then* $B_{n,k}(w)$ *is the number of directed graphs labelled by* $[n]$ *with* k *such segments as its connected components. Note that assigned to each of this graphs there is a partition* π *of* $[n]$ *such that* i *and* j *are in the same block* π *iff they belong to the same connected component of the graph.*

Now let G_1 *be a uniformly distributed random rooted segment labelled by* $[n]$*, let* $\Pi_{n,1}$ *be the partition assigned to* G_1 *(clearly* $\Pi_{n,1} = \{\{1, 2, ..., n\}\}$ *a.s.). For* $1 < k < n$*, given the directed graph* G_{k-1} *with* $k - 1$ *connected components let* G_k *be obtained by the deletion of an arrow in* G_{k-1} *uniformly chosen and let* $\Pi_{n,k}$ *be the partition assigned to it. The next figure illustrates one possible realization.*



The random sequence $\{\Pi_{n,k}\}_{k=1}^{n}$ *is a* fragmenting sequence, *such that marginally* $\Pi_{n,k}$ *has a microcanonical Gibbs distribution on* $\mathcal{P}_{[n]}^{k}$ *governed by the sequence* $w = (w_1, w_2, ...)$ *where* $w_j = j!$. *So for every* $k \in [n]$ *and any* $\pi \in \mathcal{P}_{[n]}^{k}$ *with block sizes* $n_1, ..., n_k$

$$\mathbb{P}[\Pi_{n,k} = \pi] = \frac{\prod_{i=1}^{k} n_i!}{B_{n,k}(w)}$$

for this particular case we have that

$$B_{n,k}(w) = \binom{n-1}{k-1}\frac{n!}{k!}$$

thus

$$\mathbb{P}[\Pi_{n,k} = \pi] = \frac{(k-1)!(n-k)!k!}{n!(n-1)!}\prod_{i=1}^{n} n_i!$$

The time reversed sequence $\{\Pi_{n,n-k}\}_{k=0}^{n-1}$ *is such that conditionally given the partition* $\Pi_{n,k}$ *with* k *components,* $\Pi_{n,k-1}$ *is equally likely to be one of the* $\binom{k}{2}$ *partitions of* $[n]$ *(with* $k-1$ *components) obtained by merging two blocks of* $\Pi_{n,k}$. *This sequence is the underlying jump chain of a very famous time-continuous process taking values in* $\mathcal{P}_{[n]}$ *called the* Kingman's n-coalescent, *the importance of this last process is due to its relationship with genetics.*

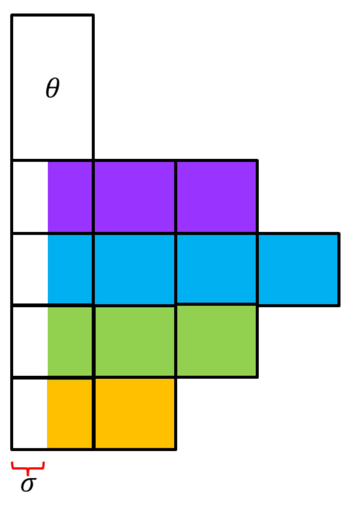
Our aim in the remaining part of this chapter is to prove the following theorem as well as introduce a family of exchangeable partitions, that generalizes the family of Gibbs partitions, called Gibbs-type priors.

Theorem 8.1. *Assume that an exchangeable partition Π_∞ of $\mathbb{N}$ has block frequencies $\tilde{P}_j$ (in order of least elements) such that $0 < \tilde{P}_1 < \infty$ almost surely and the restriction Π_n of Π_∞ to $[n]$ is a Gibbs$_{[n]}(v, w)$ partition for some non-negative sequences $v = (v_1, v_2, ...)$ and $w = (w_1, w_2, ...)$. Then the distribution of Π_∞ is either determined by a (σ, θ) model for some (σ, θ), or that of the coupon collectors partition, for some $m = 2, 3, ...$*

8.2 The coloured squares construction and colour spectrum

Let us construct the two-parameter model as follows. Let $0 < \sigma < 1$ and $\theta > -\sigma$. Imagine we have table with an infinite size and we are going to accommodate squares over it in the following way. We start by putting a square with a total area of 1 that is partially white and partially painted by a color, blue for instance, such that the portion of the square that is white is exactly σ. Also, if $\theta \geq 0$ at this first stage we are going to put over the table a rectangle of area θ that is totally white. If on the contrary, $\theta < 0$, we are going to remove an appropriate part of the initial square so that the total white area equals $\sigma + \theta$.

At each stage of the process we are going to pick uniformly a point on the surface of the squares. If the point we picked is coloured, then we are going to add a square of area 1 totally painted with the same colour of the touched point. If, on the other hand, the point we picked is white, then we are going to add a square of area 1 such that it is partially white and partially a new colour different from the ones that already appear in the table, the white portion of this square is also going to be exactly σ. The next figure illustrates this.



At stage n this generates a partition Π_n of $[n]$. Note n is the number of painted squares (including the ones that are partially painted), that is n equals the number of squares at the table disregarding the one that is totally white. Let us define the equivalence relationship for $i, j \in [n]$ by $i \sim j$ iff the ith and the jth squares to be put over the table are painted in the same colour. Let K_n be number of different colours over the table except white, that is K_n is the number of blocks in the partition defined above, and let N_{n_j} be the number of squares painted in colour j. Note that if at stage n the partition generated is $\Pi_n = \pi = \{A_1, A_2, ...A_k\}$ with $|A_j| = n_j$ then

$$\mathbb{P}[\Pi_{n+1} = \pi_j | \Pi_n = \pi] = \frac{n_j - \sigma}{n + \theta}$$

where $\pi_j = \{A_1, ..., A_j \cup \{n+1\}, ..., A_k\}$ for $j = 1, 2, ..., k$,

$$\mathbb{P}[\Pi_{n+1} = \pi_{k+1} | \Pi_n = \pi] = \frac{\theta + k\sigma}{n + \theta}$$

where $\pi_{k+1} = \{A_1, ..., A_k, \{n+1\}\}$. Hence this is just another construction of the (σ, θ)-partition for the case $0 < \sigma < 1$, $\theta > -\sigma$. Thus the EPPF takes the form

$$\mathbb{P}[\Pi_n = \pi] = \frac{(\theta + \sigma)_{k-1\uparrow\sigma} \prod_{i=1}^{k} (1 - \sigma)_{n_i - 1\uparrow 1}}{(\theta + 1)_{n-1\uparrow 1}}$$

Let us encode this in a slightly different way, let $M_{n,i} := \sum_{j=1}^{K_n} \mathbf{1}_{\{N_{n,j} = i\}}$ for $i \in [n]$, that is $M_{n,i}$ is the number of colours represented by exactly j squares in the table. Note that

$$n = \sum_{i=1}^{n} i M_{n,i}, \text{ and } K_n = \sum_{i=1}^{n} M_{n,i}$$

Now define $\Lambda_n := (M_{n,1}, M_{n,2}, ...)$ where $M_{n,i} = 0$ for $i = n+1, n+2, ...$ and for $N = 1, 2, ...$ let

$$\mathcal{Y}_N := \left\{ (m_1, m_2, ...) : m_i \in \{0, 1, ..., n\} \forall i, \quad \sum_{i \geq 1} i m_i = N \right\}$$

Clearly Λ_n is a random variable that takes values in $\mathcal{Y}_n$, we will call Λ_n the *colour spectrum* generated by the square construction at stage n.

Assume that at stage n, $\Lambda_n = \lambda = (m_1, m_2, ...) \in \mathcal{Y}_N$ is the color spectrum generated, and that $k = \sum_{i \geq 1} m_i$. We are interested in computing the probabilities $p(\lambda, \lambda^*) := \mathbb{P}[\Lambda_{n+1} = \lambda^* | \Lambda_n = \lambda]$ for $\lambda \in \mathcal{Y}_{n+1}$. Let $\lambda^{(i)} = (m_1, ..., m_{i-1}, m_i - 1, m_{i+1} + 1, m_{i+2}, ...) \in \mathcal{Y}_{n+1}$ for every $i \geq 1$ such that $m_i > 0$ also let $\lambda^{(0)} = (m_1 + 1, m_2, ...) \in \mathcal{Y}_{n+1}$. Note that

$$p(\lambda, \lambda^{(i)}) = \frac{(i - \sigma) m_i}{\theta + n}, \text{ for } i \geq 2$$

is the probability that at stage $n+1$ a new coloured square is added to i previously existing squares of this colour, and

$$p(\lambda, \lambda^{(0)}) = \frac{\theta + k\sigma}{\theta + n}$$

is the probability of adding a square painted in a distinct colour, where k represents the numbers of different colours in the table at stage n. Also note that for $\lambda^* \in \mathcal{Y}_{n+1}$ such that $\lambda^* \neq \lambda^{(i)}$ for every $i \geq 0$, we have that $p(\lambda, \lambda^*) = 0$.

Now we want to compute $P_n(\lambda) := \mathbb{P}[\Lambda_n = \lambda]$ for certain $\lambda \in \mathcal{Y}_n$, so by what has been stated in Chapter 5, particularly equation (5.1), and inserting the formula for the EPPF we obtain

$$\mathbb{P}[\Lambda_n = \lambda] = \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!} \frac{(\theta+\sigma)_{k-1\uparrow\sigma} \prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1}}{(\theta+1)_{n-1\uparrow 1}}$$

where $\lambda = (m_1, m_2, ...)$, such that $m_i = \sum_{j=1}^{k} \mathbf{1}_{\{n_j = i\}}$. Note that

$$\begin{aligned}
\prod_{i=1}^{k}(1-\sigma)_{n_i-1\uparrow 1} &= \prod_{i=1}^{n}\left((1-\sigma)_{i-1\uparrow 1}\right)^{m_i} \\
&= \prod_{i=1}^{n}\left(\prod_{j=1}^{i-1}(j-\sigma)\right)^{m_i} \\
&= \prod_{i=1}^{n}\left(\prod_{j=1}^{i-1} j\left(1-\frac{\sigma}{j}\right)\right)^{m_i} \\
&= \prod_{i=1}^{n}((i-1)!)^{m_i}\left(\prod_{j=1}^{i-1}\left(1-\frac{\sigma}{j}\right)\right)^{m_i}
\end{aligned}$$

Thus we conclude

$$\mathbb{P}[\Lambda_n = \lambda] = \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow 1}} \frac{n!}{\prod_{i=1}^{n}(i)^{m_i} m_i!} \prod_{i=1}^{n}\left(\prod_{j=1}^{i-1}\left(1-\frac{\sigma}{j}\right)\right)^{m_i} =: P_n^{(\sigma,\theta)}(\lambda) \tag{8.1}$$

we call the equation above the *Ewens-Pitman formula* and the family of distributions $\{P_n^{(\sigma,\theta)}\}_{n\geq 1}$ over the set $\mathcal{Y}_n$, $n \geq 1$ the *Ewens-Pitman distributions.*

Remark: Even though we made the construction assuming $0 < \sigma < 1$ and $\theta > -\sigma$, the Ewens-Pitman distributions are also well defined for the cases $\sigma = 0$ and $\theta > 0$ and $\sigma < 0$ and $\theta = -\sigma r$ for some $r = 1, 2, ...$

Now, let us think in a backwards process. Imagine we have at the table a colour spectrum $\Lambda_n \in \mathcal{Y}_n$ and now instead of putting a square in the table we are going to choose uniformly a square among the painted ones and take it of the table so that we obtain the colour spectrum $\Lambda_{n-1} \in \mathcal{Y}_{n-1}$. Assume that $\Lambda_n = \lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$. For i such that $m_i > 0$, define $\lambda^{(-i)} = (r_1, r_2, ...) \in \mathcal{Y}_{n-1}$ where $r_i = m_i - 1$, $r_{i-1} = m_{i-1} + 1$, and $r_j = m_j$ for every $j \neq i, i-1$. If we set $q(\lambda, \lambda^*) := \mathbb{P}[\Lambda_{n-1} = \lambda^* | \Lambda_n = \lambda]$, then clearly

$$q(\lambda, \lambda^*) = \begin{cases} \frac{im_i}{n} & \text{if } \lambda^* = \lambda^{(-i)} \text{ for some } i \text{ as above} \\ 0 & \text{if } \lambda^* \neq \lambda^{(-i)} \text{ for every } i \text{ as above} \end{cases}$$

The numbers $q(\lambda, \lambda^*)$, are called the *cotransition probabilities* and one can use them to define, starting from a probability distribution P_n over $\mathcal{Y}_n$, a new distribution P_{n-1} over $\mathcal{Y}_{n-1}$ by

$$P_{n-1}(\lambda^*) = \sum_{\lambda \in \mathcal{Y}_n} P_n(\lambda) q(\lambda, \lambda^*)$$

This leads us to the next definition

Definition 8.2. *A family of distributions $\{P_n\}_{n \geq 1}$, such that P_n is a distribution on the set $\mathcal{Y}_n$, is called* coherent *if for all $n \geq 1$ and for every $\lambda^* \in \mathcal{Y}_{n-1}$ the following identity holds*

$$P_{n-1}(\lambda^*) = \sum_{\lambda \in \mathcal{Y}_n} P_n(\lambda) q(\lambda, \lambda^*) \tag{8.2}$$

Now we see that for $0 < \sigma < 1$ and $\theta > -\sigma$ the distribution $P_n^{(\sigma,\theta)}$ obtained by the forward coloured squares construction is coherent. By inserting the corresponding formulae it is direct that the following equation holds for every $n \geq 2$:

$$P_{n-1}^{(\sigma,\theta)}(\lambda^*) p(\lambda^*, \lambda) = P_n^{(\sigma,\theta)}(\lambda) \frac{jm_j}{n}$$

where $\lambda = (m_1, m_2, ...)$ and $\lambda^* = (r_1, r_2, ...)$ are such that $r_j = m_j - 1$, $r_{j-1} = m_{j-1} + 1$ and $r_i = m_i$ for every $i \neq j, j-1$. It follows that

$$\sum_{\lambda \in \mathcal{Y}_n} \frac{jm_j}{n} P_n^{(\sigma,\theta)}(\lambda) = \sum_{\lambda \in \mathcal{Y}_n} P_{n-1}^{(\sigma,\theta)}(\lambda^*) p(\lambda^*, \lambda) = P_{n-1}^{(\sigma,\theta)}(\lambda^*).$$

Hence the family $\{P_n^{(\sigma,\theta)}\}_{n \geq 1}$ is coherent.

Before moving on, let us take a look at a few degenerate limiting cases of Ewens-Pitman formula

i) If $\sigma + \theta = 0$ then only partitions with a single component have non-zero probabilities, i.e. for $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$

$$P_n^{(\sigma,\theta)}(\lambda) = \begin{cases} 1 & \text{if } m_n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

ii) If $m = \frac{-\theta}{\sigma} = 1, 2, ...$ and $\sigma \to -\infty$, $\theta \to \infty$ then the Ewens-Pitman distributions converge to the limiting distribution

$$P_n(\lambda) = \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i}(m_i!)} \frac{(m)_{k\downarrow 1}}{m^n} \tag{8.3}$$

where $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$ such that $\sum_{i=1}^{n} m_i = k$.

iii) If $\sigma \to 1$ while $\theta > c > -1$ is bounded below by a constant c that does not depend on σ, then only components of length 1 have positive probabilities, that is

$$P_n^{(\sigma,\theta)}(\lambda) = \begin{cases} 1 & \text{if } m_1 = n \\ 0 & \text{otherwise.} \end{cases}$$

where $\lambda = (m_1, m_2, ...)$. The same happens if we let $\theta \to \infty$ and $\sigma \geq 0$.

iv) If $\sigma \to 1$, $\theta \to -1$ in such way that the following limits exist

$$p = \lim_{\sigma \to 1, \theta \to -1} \frac{1-\sigma}{\theta+1} \qquad \text{and} \qquad q = \lim_{\sigma \to 1, \theta \to -1} \frac{\theta+\sigma}{\theta+1},$$

then for $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$ the Ewens-Pitman distribution converge to the distribution

$$P_n(\lambda) = \begin{cases} p & \text{if } m_n = 1, \\ q & \text{if } m_1 = n, \\ 0 & \text{otherwise.} \end{cases}$$

8.3 The Kolchin's model

We start with two positive integer valued random variables K and N with distributions

$$\mathbb{P}[K = j] = \gamma_j; \quad j = 1, 2, ... \tag{8.4}$$

$$\mathbb{P}[N = j] = \alpha_j; \quad j = 1, 2, ... \tag{8.5}$$

Consider the random vector $(N_1, N_2, ..., N_K)$ such that K distributes as (8.4) and given K the components of the vector are independent and identically distributed with common distribution (8.5). Let $N^* := \sum_{j=1}^{K} N_j$ and $M_i := \sum_{j=1}^{k} \mathbf{1}_{\{N_j = i\}}$ so that M_i denotes the number of components of the random vector that equal to i. It is clear that the following identities hold

$$N^* = \sum_{i=1}^{N^*} iM_i, \text{ and } K = \sum_{i=1}^{N^*} M_i$$

Hence $\Lambda = (M_1, M_2, ...)$ is a random variable taking values in $\mathcal{Y}_{N^*}$. It is very direct to compute

$$\mathbb{P}[\Lambda = \lambda] = \gamma_k \frac{k!}{m_1! m_2! \cdots} \alpha_1^{m_1} \alpha_2^{m_2} \cdots$$

We are interested in the conditional probabilities $\mathbb{P}[\Lambda = \lambda | N^* = n]$, for a fixed value of $N^* = n$, and according to the last equation

$$P_n(\lambda) := \mathbb{P}[\Lambda = \lambda | N^* = n] = \frac{\gamma_k}{C_n} \frac{k!}{\prod_{i \geq 1} m_i!} \prod_{j \geq 1} \alpha_j^{m_j} \tag{8.6}$$

where the normalizing constant is $C_n = \mathbb{P}[N_1 + \cdots + N_K = n] = \sum_{\Lambda \in \mathcal{Y}_n} \mathbb{P}[\Lambda = \lambda]$.

Definition 8.3. *We say that the distributions (8.6) are* generated via Kolchin's model. *The positive integer distributions (8.4) and (8.5) are called the parameters of the model.*

8.4 Quasi-binomial distributions

Let us introduce a class of integer distributions similar to the binomial distributions. As we will see bellow if we choose from this class the parameters in the Kolchin model, this leads us to the Ewens-Pitman formula

Definition 8.4. *We call a positive integer distribution* $(p_1(a, x), p_2(a, x), ...)$ quasi-binomial*, if the ratios*

$$\hat{p}_k = \frac{(k+1)p_{k+1}}{p_k}; \quad k = 1, 2, ... \tag{8.7}$$

form an arithmetic series with parameters a, x:

$$\hat{p}_k = a + xk; k = 1, 2, ... \tag{8.8}$$

Lemma 8.1. *Every quasi-binomial distribution can be obtained by one of the following formulae*

$$p_k(a, x) = \frac{1}{\left((1-x)^{-a/x} - 1\right)} \frac{(a)_{k \uparrow x}}{k!}; \quad k = 1, 2, ... \tag{8.9}$$

where $0 < x < 1$, $a > -x$ or $x = 1$, $-1 < a < 0$;

$$p_k(a, x) = \frac{1}{(e^a - 1)} \frac{a^k}{k!} \quad k = 1, 2, ... \tag{8.10}$$

where $x = 0$ and $a > 0$; or

$$p_k(-mx, x) = \frac{(-x)^k}{((1-x)^m - 1)} \binom{m}{k}; k = 1, 2, ... \tag{8.11}$$

where $x < 0$ and $-\frac{a}{x} = m = 1, 2, ...$

Remark: Note that formulae (8.9) is meaningless for $a = 0$, in this case we replace it with

$$p_k(0, x) = \frac{x^k}{-k \log(1-x)}; k = 1, 2, ...$$

with the parameter $0 < x < 1$.

Sketch of the Proof:

For every quasi-binomial distribution $p_k(a, x)$ the following holds

$$k!\frac{p_k}{p_1} = \hat{p}_1\hat{p}_2 \cdots \hat{p}_{k-1} = (a + x)_{k-1\uparrow x}$$

so that

$$p_k = p_1 \frac{(a + x)_{k-1\uparrow x}}{k!}$$

Using the condition $\sum_k p_k(a, x) = 1$ we can write uniquely the probabilities $p_k(a, x)$ as functions of the parameters a, x. The series $\sum_k p_k(a, x)$ converges only whether $|x| < 1$, or the series is finite, that is is $a + mx = 0$ for some $m = 1, 2, ...$ It is not difficult to see that the numbers $p_k(a, x)$ are positive only in the cases indicated in the lemma. As a matter of fact, we have that

$$0 \leq \frac{2p_2}{p_1} = \hat{p}_1 = a + x;$$

and since the probabilities $p_k(a, x)$ decrease to zero as $k \to \infty$,

$$1 \geq \lim_{k\to\infty} \frac{(k + 1)p_{k+1}}{p_k} = \lim_{k\to\infty} \frac{\hat{p}_k}{k} = \lim_{k\to\infty} \frac{a + kx}{k} = x$$

If $x < 0$, necessarily $-\frac{a}{x}$ is an integer (if it was not, then there exists an integer m such that $m - 1 < -\frac{a}{x} < m$, implying $\hat{p}_m = a + mx < 0$ thus $p_{m+1} < 0$ which can not be because p_m is a probability).

$\square$

In the following we are going to use the generating functions

$$\varphi^{(a,x)}(z) = \sum_{k=1}^{\infty} z^k p_k(a, x)$$

of quasi-binomial distributions. As a consequence of Lemma 8.1 and the remark below it, these functions are given by

$$\varphi^{(a,x)}(z) = \frac{(1 - xz)^{-a/x} - 1}{(1 - x)^{-a/x} - 1} \tag{8.12}$$

if $0 < x < 1$ and $a > -x$, $a \neq 0$;

$$\varphi^{(0,x)}(z) = \frac{\log(1 - xz)}{\log(1 - x)} \tag{8.13}$$

if $0 < x < 1$ and $a = 0$;

$$\varphi^{(a,x)}(z) = \frac{e^{az} - 1}{e^a - 1} \tag{8.14}$$

if $x = 0$ and $a > 0$; finally

$$\varphi^{(-mx,x)}(z) = \frac{(1 - xz)^m - 1}{(1 - x)^m - 1} \tag{8.15}$$

if $x < 0$ and $a = -mx$ for some $m = 1, 2, ...$

Quasi-binomial distributions and Ewens-Pitman formula

Lemma 8.2. *Let K and N in the Kolchin model such that they have quasi binomial distributions, that is $\alpha_k = p_k(a, x)$ and $\gamma_k = p_k(b, y)$ for some admissible a, x, b, y, also assume that $y = -\frac{a}{\alpha_1}$, then the distribution of N^* as in Chapter 8.3 is quasi-binomial and written as*

$$C_n = \mathbb{P}[N^* = n] = p_n\left(-\frac{ab}{y}, x\right), \quad n = 1, 2, ... \tag{8.16}$$

Proof:

Recall that $N^* = \sum_{i=1}^{K} N_i$ where the random variables $(N_1, N_2, ..., N_K)$ are independent and identically distributed given K, with common distribution as N. It is widely known that the generating function of N^* can be written in terms of the generating functions of K and N as follows:

$$\varphi_{N^*}(z) = \varphi_K(\varphi_N(z)).$$

Now we consider the cases indicated in lemma 8.1 separately.

i) If $0 < x < 1$ and $a > -x$, $a \neq 0$, then $y = -\frac{a}{\alpha_1} = 1 - (1 - x)^{-a/x}$ so that $1 - y = (1 - x)^{-a/x}$, hence the generating functions for K and N have the form

$$\varphi_K(z) = \frac{(1 - yz)^{-b/y} - 1}{(1 - x)^{-b/y} - 1}, \quad \varphi_N(z) = \frac{(1 - xz)^{-a/x} - 1}{(1 - x)^{-a/x} - 1}$$

Since

$$1 - y\varphi_N(z) = 1 - \left[(1 - (1 - x)^{-a/x})\frac{(1 - xz)^{-a/x} - 1}{(1 - x)^{-a/x} - 1}\right] = (1 - xz)^{-a/x}$$

then

$$\varphi_{N^*} = \varphi_K(\varphi_N(z)) = \frac{(1 - xz)^{ab/xy} - 1}{(1 - x)^{ab/xy} - 1},$$

corresponding to the quasi-binomial distribution

$$C_n = p_n\left(-\frac{ab}{y}, x\right).$$

If $a = 0$, we obtain $\alpha_1 = -\frac{x}{\log(1-x)}$, so $y = -\frac{a}{\alpha_1} = 0$ and the generating functions of K and N are

$$\varphi_K(z) = \frac{e^{bz}-1}{e^b-1}, \quad \varphi_N(z) = \frac{\log(1-xz)}{\log(1-x)}$$

Thus

$$\varphi_{N^*} = \varphi_K(\varphi_N(z)) = \frac{(1-xz)^{b/\log(1-x)}-1}{(1-x)^{b/\log(1-x)}-1},$$

so that N^* has a quasi binomial distribution with

$$C_n = p_n\left(-\frac{bx}{\log(1-x)}, x\right).$$

ii) Now if $x = 0$ and $a > 0$, then $y = -\frac{a}{\alpha_1} = 1 - e^{\sigma} < 0$ and $(1-y) = e^a$ so that $b = -my$ for some $m = 1, 2, ...$ and the generating functions for K and N take the form

$$\varphi_K(z) = \frac{(1-yz)^m-1}{(1-y)^m-1}, \quad \varphi_N(z) = \frac{e^{az}-1}{e^a-1}$$

Since

$$1 - y\varphi_N(z) = 1 - \left[1-(1-e^a)\frac{e^{az}-1}{e^a-1}\right] = e^{az}$$

we obtain

$$\varphi_{N^*} = \varphi_K(\varphi_N(z)) = \frac{e^{amz}-1}{e^{am}-1}$$

Hence N^* follows a quasi-binomial distribution where

$$C_n = p_n(am, 0)$$

iii) At last, if $x < 0$ and $a = -nx$ for some $n = 1, 2, ...$, this implies $y = -\frac{a}{\alpha_1} = 1-(1-x)^n < 0$, $1-y = (1-x)^n$ and $b = -my$ for some $m = 1, 2, ...$ Thus the generating functions of K and N are

$$\varphi_N(z) = \frac{(1-xz)^n-1}{(1-x)^n-1}, \quad \varphi_K(z) = \frac{(1-yz)^m-1}{(1-y)^m-1}$$

Note that

$$1 - y\varphi_N(z) = 1 - \left[1-(1-x)^n\frac{(1-xz)^n-1}{(1-x)^n-1}\right] = (1-xz)^n$$

so that

$$\varphi_{N^*} = \varphi_K(\varphi_N(z)) = \frac{(1-xz)^{nm}-1}{(1-x)^{nm}-1}$$

Concluding that N^* has the quasi-binomial distribution determined by

$$C_n = p_n(-kmx, x).$$

$\square$

The next proposition establishes a connection between the quasi-binomial distributions and Ewens-Pitman formula.

Proposition 8.1. *Let P_n denote the conditional distribution (given $N^* = n$) on the set $\mathcal{Y}_n$ obtained via Kolchin's model, i.e. equation (8.6) with the parameter distribution*

$$\alpha_k = p_k(a, x), \quad \gamma_k = p_k(b, y); \qquad k = 1, 2, ...$$

and assume $y = -\frac{a}{\alpha_1}$. Then the distribution P_n is given by Ewens-Pitman formula with parameters

$$\sigma = -\frac{a}{x}, \quad \theta = -\frac{ab}{xy}.$$

Remark: Note that we have the following particular cases.

1) If $0 < x < 1$, $-x < a < 0$ and $y = 1-(1-x)^{-a/x}$, $b > -y$; then $0 < \sigma < 1$ and $\theta > -\sigma$.

2) If $0 < x < 1$, $a = 0$ and $y = 0$, $b > 0$; then $\sigma = 0$ and $\theta = -\frac{b}{\log(1-x)}$.

3) If $0 < x < 1$, $a > 0$ and $y = 1-(1-x)^{-a/x}$, $b = -my$ for some $m = 1, 2, ...$; then $\sigma < 0$ and $\theta = -r\sigma$ for some $r = 1, 2, ...$

4) If $x = 0$, $a > 0$ and $y = e^a - 1$, $b = -my$ for some $m = 1, 2, ...$; then the distributions P_n coincide with the limiting case (8.3).

Proof:

1) First let us consider the case $0 < x < 1$ and $-x < a < 0$, so that $y = 1 - (1-x)^{-a/x} = -\frac{a}{\alpha_1} > 0$, $b > -y$. Note that this also means $0 < \sigma < 1$ and $\theta > -\sigma$. Since $1 - y = (1-x)^{-a/x}$, the normalizing constant in equation (8.6) equals

$$C_n = p_n\left(-\frac{ab}{y}, x\right) = \frac{1}{n!\left((1-y)^{-\frac{b}{y}} - 1\right)} \prod_{j=0}^{n-1} \left(-\frac{ab}{y} + jx\right)$$

Thus for $\lambda \in \mathcal{Y}_n$

$$\begin{aligned}
P_n(\lambda) &= \frac{p_k(b,y)}{p_n\left(-\frac{ab}{y},x\right)} \frac{k!}{m_1!m_2!\cdots} \prod_{i=1}^{n} (p_i(a,x))^{m_i} \\
&= \frac{n!\left((1-y)^{-b/y}-1\right)}{\left(-\frac{ab}{y}\right)_{n\uparrow x}} \frac{(b)_{k\uparrow y}}{k!\left((1-y)^{-b/y}-1\right)} \frac{k!}{m_1!m_2!\cdots} \times \\
&\quad \times \prod_{i=1}^{n} \left(\frac{(a)_{i\uparrow x}}{i!\left((1-y)^{-b/y}-1\right)} \right)^{m_i} \\
&= \frac{(b)_{k\uparrow y}}{\left(-\frac{ab}{y}\right)_{n\uparrow x}} \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i} m_i!} \prod_{i=1}^{n} \left(\frac{(a)_{i\uparrow x}}{\left((1-y)^{-b/y}-1\right)} \right)^{m_i} \\
&= \frac{\left(-\frac{a}{xy}\right)^k (b)_{k\uparrow y}}{\left(\frac{1}{x}\right)^n \left(-\frac{ab}{y}\right)_{n\uparrow x}} \frac{n!}{\prod_{i=1}^{n}(i)^{m_i} m_i!} \prod_{i=1}^{n} \left(\prod_{j=1}^{i-1} \left(1+\frac{a}{jx}\right) \right)^{m_i} \\
&= \frac{\left(-\frac{ab}{xy}-\frac{a}{x}\right)_{k-1\uparrow -\frac{a}{x}}}{\left(-\frac{ab}{xy}+1\right)_{n-1\uparrow 1}} \frac{n!}{\prod_{i=1}^{n}(i)^{m_i} m_i!} \prod_{i=1}^{n} \left(\prod_{j=1}^{i-1} \left(1+\frac{a}{jx}\right) \right)^{m_i} \\
&= \frac{(\theta+\sigma)_{k-1\uparrow\sigma}}{(\theta+1)_{n-1\uparrow 1}} \frac{n!}{\prod_{i=1}^{n}(i)^{m_i} m_i!} \prod_{i=1}^{n} \left(\prod_{j=1}^{i-1} \left(1+\frac{\sigma}{j}\right) \right)^{m_i}
\end{aligned}$$

where $\sigma = -\frac{a}{x}$ and $\theta = -\frac{ab}{xy}$.

2) If $0 < x < 1$ and $a = 0$, then $y = -\frac{a}{\alpha_1} = 0$ and $b > 0$. Using the notation

$$\theta = -\frac{b}{\log(1-x)}$$

we have $(1-x)^{-\theta} = e^b$ and

$$C_n = p_n\left(-\frac{bx}{\log(1-x)}, x\right) = \frac{\prod_{j=0}^{n-1}(x\theta + jx)}{n!(e^b - 1)}$$

Thus, by equation (8.6)

$$
\begin{aligned}
P_n(\lambda) &= \frac{p_k(b,0)}{p_n\left(x\theta,x\right)}\frac{k!}{m_1!m_2!\cdots}\prod_{i=1}^{n}(p_i(0,x))^{m_i}\\
&= \frac{n!(e^b-1)}{(\theta x)_{n\uparrow x}}\frac{b^k}{k!(e^b-1)}\frac{k!}{\prod_{i=1}^{n}m_i!}\prod_{i=1}^{n}\left(\frac{x^i}{i}\frac{\theta}{b}\right)^{m_i}\\
&= \frac{1}{(\theta x)_{n\uparrow x}}\frac{n!b^k}{\prod_{i=1}^{n}(i)^{m_i}m_i!}\frac{x^n\theta^k}{b^k}\\
&= \frac{\theta^k}{\left(\frac{1}{x}\right)^n(\theta x)_{n\uparrow x}}\frac{n!}{\prod_{i=1}^{n}(i)^{m_i}m_i!}\\
&= \frac{\theta^k}{(\theta+1)_{n-1\uparrow 1}}\frac{n!}{\prod_{i=1}^{n}(i)^{m_i}m_i!}
\end{aligned}
$$

which coincides with the Ewens-Pitman formula for the scenario $\sigma = 0$.

3) If $0 < x < 1$ and $a > 0$, then $y = 1-(1-x)^{-a/x} < 0$, thus we can only choose b such that $-\frac{b}{y} = m$ for some $m = 1, 2, ...$ the and the rest of the proof is similar (1).

4) Assume that $x = 0$ and $a > 0$, thus $y = 1 - e^a$ and once again the choice of the parameter has to hold the identity $-\frac{b}{y} = m$ for some $m = 1, 2, ...$ Hence

$$C_n = p_n(am,0) = \frac{(am)^n}{n!\left(e^{am}-1\right)}$$

and formula (8.6) becomes

$$
\begin{aligned}
P_n(\lambda) &= \frac{p_k(-my,y)}{p_n\left(am,0\right)}\frac{k!}{m_1!m_2!\cdots}\prod_{i=1}^{n}(p_i(a,0))^{m_i}\\
&= \frac{n!\left(e^{am}-1\right)}{(am)^n}\frac{(-y)^k}{(1-y)^m-1}\binom{m}{k}\frac{k!}{m_1!m_2!\cdots}\prod_{i=1}^{n}\left(\frac{a^i}{i!\,(-y)}\right)^{m_i}\\
&= \frac{(e^{am}-1)}{a^n}\frac{(-y)^k}{(e^a)^m-1}\frac{n!}{\prod_{i=1}^{n}(i!)^{m_i}m_i!}\frac{(m)_{k\downarrow 1}}{m^n}\frac{a^n}{(-y)^k}\\
&= \frac{n!}{\prod_{i=1}^{n}(i!)^{m_i}m_i!}\frac{(m)_{k\downarrow 1}}{m^n}
\end{aligned}
$$

which is identical to equation (8.3)

□

Remark: The choice $x < 0$, $-\frac{a}{x} = k = 1, 2, ...$ of the parameters leads $y = 1-(1-x)^k < 0$ and $-\frac{b}{y} = m = 1, 2, ...$ hence $C_n = p_n(-kmx, x)$ and the variable N^* cannot exceed km. This case is of no interest to us.

Theorem 8.2. *Consider the family of distributions $\{P_n\}_{n\geq 1}$ on the set of partitions $\mathcal{Y}_n$, $n = 1, 2, ...$ obtained via Kolchin's model, and assume that the family $\{P_n\}_{n\geq 1}$ is coherent. Then the distributions $P_n = P_n(\sigma, \theta)$ are given by the Ewens-Pitman formula for either $0 \leq \sigma < 1$ and $\theta > -\sigma$ or $\sigma < 0$ and $\theta = -\sigma r$ for some $r = 1, 2, ...$. Also a limiting case can occur, so that*

$$P_n(\lambda) = \begin{cases} 1 & \text{if } m_1 = n \\ 0 & \text{otherwise.} \end{cases}$$

for $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$.

Proof:

Let $\lambda = (m_1, m_2, ..) \in \mathcal{Y}_n$ such that $\sum_{i=1}^{n} m_i = k$, and for $i \geq 1$ such that $m_i > 0$ let

$$\lambda^{(i)} = (r_1^{(i)}, r_2^{(i)}, ...) = (m_1, ...m_{i-1}, m_i - 1, m_{i+1} + 1, m_{i+2}, ...) \in \mathcal{Y}_{n+1}$$

also define

$$\lambda^{(0)} = (r_1^{(0)}, r_2^{(0)}, r_3^{(0)} ...) = (m_1 + 1, m_2, m_3, ...) \in \mathcal{Y}_{n+1}$$

Note that for $i \geq 1$, $\sum_{j=1}^{n+1} r_j^{(i)} = k$ and $\sum_{j=1}^{n+1} r_j^{(0)} = k + 1$; that is, $\lambda^{(i)}$ has the same number of parts as λ and $\lambda^{(0)}$ has one more part. Also note that for a fixed $\lambda \in \mathcal{Y}_n$, $q(\lambda^*, \lambda) > 0$ iff $\lambda^* = \lambda^{(i)}$ for some $i \geq 0$ as defined above. Thus the condition of coherency (8.2) can be written as

$$P_{n+1}(\lambda^{(0)})\frac{m_1 + 1}{n + 1} + \sum_{i\geq 1} \frac{(i + 1)(m_{i+1} + 1)}{n + 1} P_{n+1}(\lambda^{(i)}) = P_n(\lambda) \tag{8.17}$$

for all $\lambda \in \mathcal{Y}_n$.

Lemma 8.3. *The condition of coherency for the distributions, i.e. equation (8.2), obtained via Kolchin's model is equivalent to the identity*

$$\frac{(k + 1)\gamma_{k+1}}{\gamma_k}\alpha_1 + \sum_{i=1}^{n} m_i \frac{(i + 1)\alpha_{i+1}}{\alpha_i} = \frac{(n + 1)c_{n+1}}{c_n} \tag{8.18}$$

where $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$ is any partition such that $\sum_{i=1}^{n} m_i = k$ and $\sum_{i=1}^{n} i m_i = n$.

Proof:

By equation (8.6) we get that

$$P_{n+1}(\lambda^{(0)}) = \frac{\gamma_{k+1}}{C_{n+1}} \frac{(k + 1)!}{(m_1 + 1)\prod_{i\geq 1} m_i!} \alpha_1 \prod_{j\geq 1} \alpha_j^{m_j},$$

$$P_{n+1}(\lambda^{(i)}) = \frac{\gamma_k}{C_{n+1}} \frac{(k)!m_i}{(m_{i+1}+1)\prod_{i\geq 1} m_i!} \frac{\alpha_{i+1}}{\alpha_i} \prod_{j\geq 1} \alpha_j^{m_j} \qquad \text{for } i \geq 2$$

and

$$P_n(\lambda) = \frac{\gamma_k}{C_n} \frac{k!}{\prod_{i\geq 1} m_i!} \prod_{j\geq 1} \alpha_j^{m_j},$$

hence, we have that

$$(m_1+1)P_{n+1}(\lambda^{(0)}) = \frac{c_n}{c_{n+1}} \frac{(k+1)\gamma_{k+1}}{\gamma_k} \alpha_1 P_n(\lambda)$$

and

$$(i+1)(m_{i+1}+1)P_{n+1}(\lambda^{(i)}) = \frac{c_n}{c_{n+1}} \frac{m_j(j+1)\sigma_{j+1}}{\gamma_j} P_n(\lambda).$$

substituting into equation (8.17) we obtain

$$(n+1)P_n(\lambda) = \frac{c_n}{c_{n+1}} \frac{(k+1)\gamma_{k+1}}{\gamma_k} \alpha_1 P_n(\lambda) + \sum_{i\geq 1} \frac{c_n}{c_{n+1}} \frac{m_j(j+1)\sigma_{j+1}}{\gamma_j} P_n(\lambda)$$

from where, cancelling the common factor $P_n(\lambda)$, the conclusion is evident.

Lemma 8.4. *Assume that the equation (8.18) holds for every $\lambda \in \mathcal{Y}_n$. Then the numbers*

$$\hat{\alpha}_i := \frac{(i+1)\alpha_{i+1}}{\alpha_i}, \quad i = 1, 2, ...$$

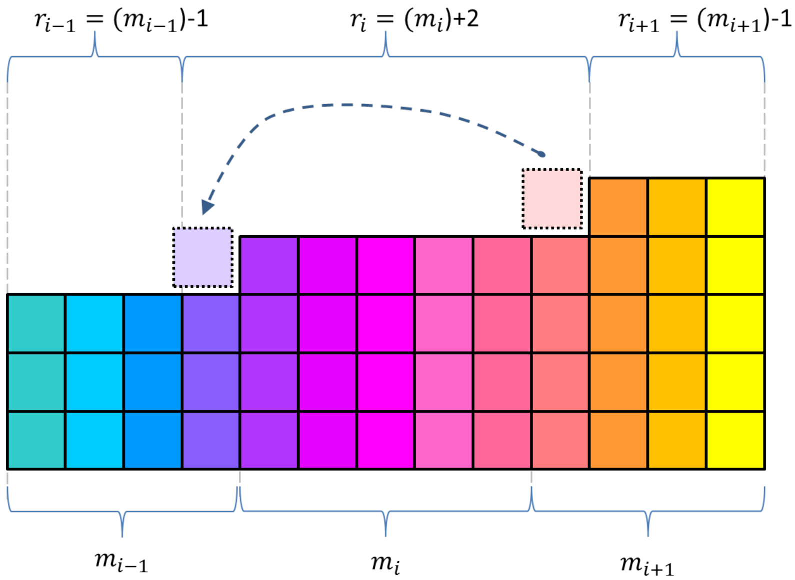
form an arithmetic sequence

$$\hat{\alpha}_i = xi + a, \quad i = 1, 2...$$

for appropriate values of x and a.

Sketch of the Proof:

Assume that there exists $2 \leq i \leq n-1$ such that the numbers m_{i-1} and m_{i+1} are non-zero. Let $\widehat{\lambda} = (r_1, r_2, ...) \in \mathcal{Y}_n$ such that $r_j = m_j$ for every $j \notin \{i-1, i, i+1\}$ and $r_{i-1} = m_{i-1} - 1$, $r_i = m_i + 2$, $r_{i+1} = m_{i+1} - 1$. Note that the transition from λ to $\widehat{\lambda}$ simply means that an element is taken from a set with $i+1$ elements and attached to a set with $i-1$ elements, so that now there are $m_{i+1} - 1$ sets with $i+1$ elements, $m_{i-1} - 1$ sets with $i-1$ elements and $m_i + 2$ sets with exactly i elements as illustrated in the next figure.



Realize that $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} m_i = k$. Obtaining, for $\widehat{\lambda}$ an analogue equation to (8.18) ans subtracting this new equation to (8.18) we get the following identity

$$\hat{\alpha}_{i+1} - \hat{\alpha}_i = \hat{\alpha}_i - \hat{\alpha}_{i-1}$$

By possibly making n grow, changing the choice of $\lambda \in \mathcal{Y}_n$ so that for some $2 \leq j \leq n-1$, $j \neq i$ it holds that m_{j-1} and m_{j+1} are positive, and repeating the argument above we obtain that the last identity holds for every $i \geq 2$. This means that $(\hat{\alpha}_1, \hat{\alpha}_2, ...)$ actually form an arithmetic sequence and the conclusion follows.

□

Lemma 8.5. *It follows from equation (8.18) that the ratios*

$$\hat{\gamma}_k = \frac{(k+1)\gamma_{k+1}}{k}, \quad k = 1, 2, ...$$

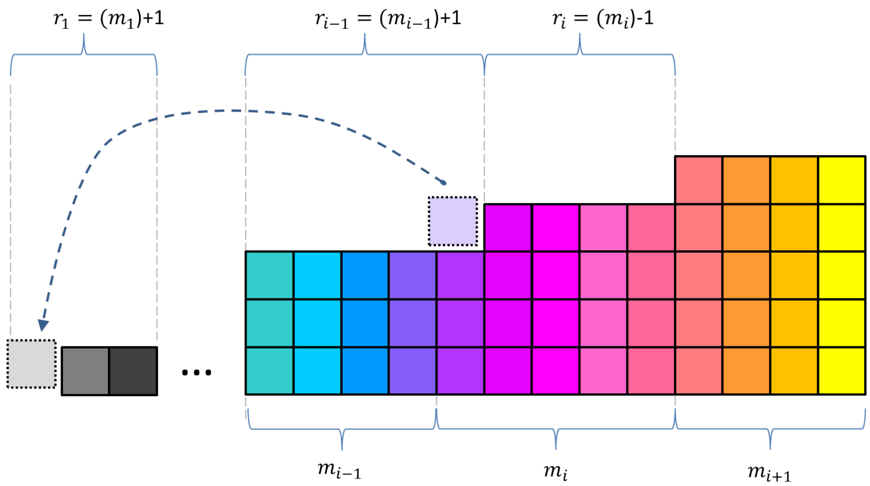
form an arithmetic sequence

$$\hat{\gamma}_k = yk + b, \quad k = 1, 2, ...$$

where the coefficients of the last lemma and this one satisfy $a + y\alpha_1 = 0$.

Sketch of the Proof:

Let $2 \leq i \leq n-1$ such that $m_i > 0$. Define $\widehat{\lambda} = (r_1, r_2, ...) \in \mathcal{Y}_n$ such that $r_j = m_j$ for every $j \notin \{1, i-1, i\}$ and $r_1 = m_1 + 1$, $r_{i-1} = m_{i-1} + 1$, $r_i = m_i - 1$. Going from λ to $\widehat{\lambda}$ can be thought as taking an element from a set with i elements and putting it in a set apart, so that $\widehat{\lambda}$ has, in terms of λ, m_1+1 sets with one unique elements, $m_{i-1}+1$ sets with $i-1$ elements and $m_i - 1$ sets with exactly i elements the following picture illustrates the transition.



Note that in this case $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} m_i + 1 = k+1$. Taking this into account and comparing the identities (8.18) for λ and $\widehat{\lambda}$ we obtain

$$(\hat{\gamma}_{k+1} - \hat{\gamma}_k)\alpha_1 = \hat{\alpha}_i - \hat{\alpha}_{i-1} - \alpha_1 = -a$$

by the last lemma and last identity, one obtains

$$\gamma k = -k\frac{a}{\alpha_1} + b$$

for some b, and the lemma is proved.

The proof of Theorem 8.2 follows from the last three results and proposition 8.1.

□

With this in mind, the proof of Theorem 8.1 becomes easy, it suffices to check that the following two statements are true

i) Every family $\{\Pi_n\}_{n\geq 1}$ of exchangeable and consistent random partitions generates a family $\{\Lambda_n\}_{n\geq 1}$ of coherent random variables where Λ_n takes values over $\mathcal{Y}_n$

ii) The block sizes of a $Gibbs_{[n]}(v,w)$ can be realised by Kolchin's model.

Let us concentrate on i). Recall that every exchangeable and consistent family of random partitions $\{\Pi_n\}_{n\geq 1}$ satisfies that for every $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$

$$\mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k)$$

where $n_j = |A_j|$ and p is a symmetric function of its arguments. Also the function p satisfies the addition rule

$$p(n_1, ...n_k) = p(n_1, ..., n_k, 1) + \sum_{j=1}^{k} p(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k)$$

in order for the family to be consistent. From this last equation one can derive the *prediction rule*, that is, if we let $\pi^{(0)} = \{A_1, ..., A_k, \{n+1\}\}$ and $\pi^{(j)} = \{A_1, ..., A_{j-1}, A_j \cup \{n+1\}, A_{j+1}, ...A_k\}$ then

$$\mathbb{P}[\Pi_{n+1} = \pi^{(0)} | \Pi_n = \pi] = \frac{p(n_1, ..., n_k, 1)}{p(n_1, ...n_k)}, \text{ and}$$
$$\mathbb{P}[\Pi_{n+1} = \pi^{(j)} | \Pi_n = \pi] = \frac{p(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k)}{p(n_1, ...n_k)} \quad j = 1, 2, ..., k.$$

As we have seen before every partition $\pi = \{A_1, ...A_k\} \in \mathcal{P}_{[n]}$ defines a spectrum $\lambda = (m_1, m_2, ...) \in \mathcal{Y}_n$ by

$$m_i = \sum_{j=1}^{k} \mathbf{1}_{\{n_j = i\}} \quad i = 1, 2, ...$$

where $n_j = |A_j|$. Let $\{\Lambda_n\}_{n \geq 1}$ be the family of random spectrum generated by $\{\Pi_n\}_{n \geq 1}$ in such way, by equation (5.2)

$$\mathbb{P}[\Lambda_n = \lambda] = \frac{n!}{\prod_{i=1}^{n} (i!)^{m_i} m_i!} p(n_1, ..., n_k).$$

Also note that, for $\lambda^* \in \mathcal{Y}_n$ and $\lambda \in \mathcal{Y}_n$, we can write the transition probabilities

$$p(\lambda, \lambda^*) = \mathbb{P}[\Lambda_{n+1} = \lambda^* | \Lambda_n = \lambda]$$

in terms of the EPPF as follows. Assume $\Pi_n = \pi = \{A_1, ..., A_k\}$ with $|A_j| = n_j$ and $\Lambda_n = \lambda = (m_1, m_2, ...)$, let $\lambda^{(0)} = (m_1+1, m_2, ...)$ and $\lambda^{(i)} = (m_1, ..., m_{i-1}, m_i - 1, m_{i+1} + 1, m_{i+2}, ...)$ for every i such that $m_i \neq 0$. Note that the transition $\lambda \to \lambda^{(0)}$ takes place when the element $\{n+1\}$ is added to π in a new block apart from $A_1, ..., A_k$. Thus

$$p(\lambda, \lambda^{(0)}) = \mathbb{P}[\Lambda_{n+1} = \lambda^{(0)} | \Lambda_n = \lambda] = \frac{p(n_1, ..., n_k, 1)}{p(n_1, ...n_k)}$$

Also, transition $\lambda \to \lambda^{(i)}$ takes place when the element $\{n+1\}$ is added to one of the blocks of π containing i elements, as there are exactly m_i such blocks in π, then

$$p(\lambda, \lambda^{(i)}) = \mathbb{P}[\Lambda_{n+1} = \lambda^{(i)} | \Lambda_n = \lambda] = m_i \frac{p(n_1, ..., n_{r-1}, n_r + 1, n_{r+1}, ..., n_k)}{p(n_1, ...n_k)}$$

where $n_r = i$.

Any order $\lambda^* \in \mathcal{Y}_{n+1}$ different from the ones defined above cannot be obtained from λ following the explained procedure thus $p(\lambda, \lambda^*) = 0$ for such spectrum. Furthermore, the cotransition probabilities as defined in Chapter 8.2, are such that

$$q(\lambda^*, \lambda) = \begin{cases} \frac{m_1+1}{n+1} & \text{if } \lambda^* = \lambda^{(0)} \\ \frac{(i+1)(m_{i+1}+1)}{n+1} & \text{if } \lambda^* = \lambda^{(i)} \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$\begin{aligned}
\mathbb{P}[\Lambda_{n+1} = \lambda^{(0)}] q(\lambda^{(0)}, \lambda) &= \left[\frac{(n+1)!}{(m_1+1)\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, ..., n_k, 1) \right] \left[\frac{m_1+1}{n+1} \right] \\
&= \frac{n!}{\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, ..., n_k, 1) \\
&= \left[\frac{n!}{\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, ..., n_k) \right] \left[\frac{p(n_1, ..., n_k, 1)}{p(n_1, ...n_k)} \right] \\
&= \mathbb{P}[\Lambda_n = \lambda] p(\lambda, \lambda^{(0)}), \mathbb{P}[\Lambda_{n+1} = \lambda^*] q(\lambda^*, \lambda)
\end{aligned}$$

and for every i such that $m_i \neq 0$

$$\begin{aligned}
&\mathbb{P}[\Lambda_{n+1} = \lambda^{(i)}] q(\lambda^{(i)}, \lambda) \\
&= \left[\frac{i! m_i}{(i+1)!(m_{i+1}+1)} \frac{(n+1)!}{\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, .., n_{r-1}, n_r+1, n_{r+1}, ...n_k) \right] \times \\
&\qquad\qquad \times \left[\frac{(i+1)(m_{i+1}+1)}{n+1} \right] \\
&= \frac{n! m_i}{\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, .., n_{r-1}, n_r+1, n_{r+1}, ...n_k) \\
&= \left[\frac{n!}{\prod_{j=1}^{n}(j!)^{m_j} m_j!} p(n_1, ..., n_k) \right] \left[m_i \frac{p(n_1, .., n_{r-1}, n_r+1, n_{r+1}, ...n_k)}{p(n_1, ...n_k)} \right] \\
&= \mathbb{P}[\Lambda_n = \lambda] p(\lambda, \lambda^{(i)})
\end{aligned}$$

So, for every $\lambda^* \in \mathcal{Y}_{n+1}$ the following equation holds

$$\mathbb{P}[\Lambda_{n+1} = \lambda^*] q(\lambda^*, \lambda) = \mathbb{P}[\Lambda_n = \lambda] p(\lambda, \lambda^*)$$

and summing over every $\lambda^* \in \mathcal{Y}_{n+1}$ we finally get

$$\mathbb{P}[\Lambda_n = \lambda] = \sum_{\lambda^* \in \mathcal{Y}_{n+1}} \mathbb{P}[\Lambda_{n+1} = \lambda^*] q(\lambda^*, \lambda)$$

so that the family $\{\Lambda_n\}_{n\geq 1}$ derived from $\{\Pi_n\}_{n\geq 1}$ is coherent and i) holds.

Now let us focus on ii). Let Π_n be a $Gibbs_{[n]}(v,w)$ partition for some sequences of non-negative numbers $v = (v_1, v_2, ...)$ and $w = (w_1, w_2, ...)$. Let K_n be the number of blocks in Π_n and $(N_{n,1}, ..., N_{n,K_n})$ be the sizes of the blocks, then for every $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$

$$\mathbb{P}[\Pi_n = \pi] = \frac{v_k \prod_{i=1}^k w_{n_i}}{B_n(v,w)}$$

where $n_i = |A_i|$. Again let $\Lambda_n \in \mathcal{Y}_n$ be the spectrum determined by Π_n so that

$$\begin{aligned}
\mathbb{P}[\Lambda_n = \lambda] &= \frac{n!}{\prod_{i=1}^n (i!)^{m_i} m_i!} \frac{v_k \prod_{i=1}^k (w_i)^{m_i}}{B_n(v,w)} \\
&= \frac{v_k}{B_n(v,w)} \frac{n!}{\prod_{i=1}^n m_i!} \prod_{i=1}^k \left(\frac{w_i}{i!}\right)^{m_i} \\
&= \frac{v_k}{B_n(v,w)} \frac{n!}{\prod_{i=1}^n m_i!} \prod_{i=1}^k (w_i^*)^{m_i}
\end{aligned}$$

where $m_i = \sum_{j=1}^k \mathbf{1}_{\{n_j = i\}}$ and $w_i^* = \frac{w_i}{i!}$. Comparing this equation with equation (8.6) one can realise that the number of blocks K_n and the block sizes $(N_{n,1}, ...N_{n,K_n})$ behave as the variables in the random sum

$$\sum_{i=1}^K N_i = N^*$$

in Kolchin's models conditionally given K and $N^* = n$ where the probabilities

$$\begin{aligned}
\gamma_k &= \mathbb{P}[K = k] \quad k = 1, 2, .. \\
\alpha_j &= \mathbb{P}[N = j] \quad j = 1, 2, ...
\end{aligned}$$

are proportional to the weights $v_1, v_2, ..., v_n$ and $w_1^*, w_2^*, ...w_n^*$ respectively, given $N^* = n$. Thus, we have corroborated ii).

Now every exchangeable partition Π_∞ of $\mathbb{N}$ defines a consistent family of exchangeable random partitions $\{\Pi_n\}_{n\geq 1}$ where Π_n is the restriction of Π_∞ to $[n]$, so from i) and ii) the proof of Theorem 8.1 follows.

8.5 Gibbs-type partitions

Let us consider an exchangeable partition $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ of $\mathbb{N}$ (where Π_n is the restriction of Π_∞ to $[n]$) such that for every $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$, the

EPPF takes the form

$$\mathbb{P}[\Pi_n = \pi] = v_{n,k} \prod_{i=1}^{k} w_{n_i}$$

where $n_i = |A_i|$, it is clear that when

$$v_{n,k} = \frac{v_k}{B_n(v, w)}$$

for some sequences of non-negative number $v = (v_1, v_2, ...)$ and $w = (w_1, w_2, ...)$, then Π_n is a $Gibbs_{[n]}(v, w)$ partition, hence this family of partitions generalizes the one introduced in Section 8.1.

In this general scenario, it can be seen that in order for the EPPF to be defined correctly the weight sequence w has got to be such that

$$w_i = (1 - \sigma)_{i\uparrow}$$

for some $\sigma \leq 1$, and furthermore, the numbers $\{v_{n,k}\}_{n\geq 1, 1\leq k\leq n}$ have to satisfy the forward recursive equation

$$v_{n,k} = (n - \sigma k)v_{n+1,k} + v_{n+1,k+1},$$

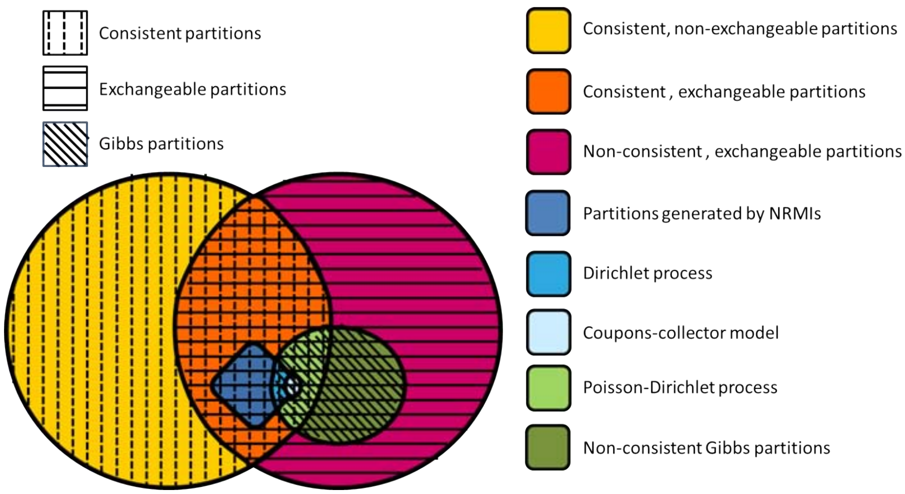
the proof of this can be seen, for instance, in [18] or [5]. Such partitions are known as *Gibbs-type* partitions.

Chapter 9

Final remarks

In the first chapters of the thesis we introduced the concepts of random partitions and families of random partitions. Two properties of some of these families turned out to be of interest to us: the projectivity or consistency and the exchangeability. When an infinite family of partitions is consistent, it can be regarded as a partition of $\mathbb{N}$. As we showed not all of the consistent families are exchangeable, and not all of the exchangeable families are consistent. There are, though, some families which satisfy both properties, the further analysis was mainly made over this set of families of partitions

Another wide family that we introduced is that of Gibbs partitions. As we have seen every Gibbs partition is exchangeable, but the are some families of Gibbs partitions that are not consistent. When a Gibbs partition is consistent, it is forced to be either that of the Coupon's collector partition or distributed according to the Poisson-Dirichlet model, characterized by two parameters (α, θ). We also showed that when the parameter $\alpha = 0$, then the model we are talking about is the Dirichlet process, which was probably the more widely discussed model throughout the thesis. The following image illustrates this classification of families of partitions.



Throughout the thesis we also discussed some ways to generate exchangeable and consistent families of random partitions.

First approach

The first approach we analysed was obtained through the normalization of a completely random measure Φ over some space $\mathbb{X}$ with corresponding σ-field $\mathcal{X}$. The resulting normalized measure $\tilde{p}(A) := \frac{\Phi(A)}{\Phi(\mathbb{X})}$, for every $A \in \mathcal{X}$, is clearly a random probability measure. So if we take a sequence of random variables $\{X_i\}_{i=1}^{\infty}$ such that given $\tilde{p}$ they are i.i.d. with $X_i \sim \tilde{p}$ and define the equivalence relation over $\mathbb{N}$, $i \sim j$ iff $X_i = X_j$, then this generates a random exchangeable partition of $\mathbb{N}$, $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$, where Π_n is the restriction of Π_∞ to $[n]$.

In the particular case when $\mathbb{X}$ equals $\mathbb{R}$ or is a measurable subset of $\mathbb{R}$, and if we add to the completely random measure the condition of having stationary increments, then it turns out to be a subordinator. Note that if we have a subordinator $\{\phi(t)\}_{t\in\mathbb{R}}$, take some stopping time T with respect to it such that $0 < \phi(T) < \infty$ almost surely. and set $F(t) := \frac{\phi(t)}{\phi(T)}$ for every $t \in [0, T]$ then F is a probability distribution function over $[0, T]$ with discontinuities at some random locations $\{\xi_i\}_{i\geq 1}$ (which are the locations where the subordinator jumps) and where the sizes of the atoms of F will be the corresponding normalized sizes of the jumps $\{J_i\}_{i\geq 1}$. If the subordinator is such that its deterministic measure is zero, $\beta = 0$, then clearly F will be discrete a.s., on the other side if $\beta > 0$ then F will have a continuous part almost surely. Once again if we start sampling without replacement from F, that is we let $\{X_i\}_{i=1}^{\infty}$ to be a sequence of i.i.d. random variables given F, we can generate an exchangeable partition $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ of $\mathbb{N}$ by the procedure described above. Note that in terms of the distribution of the random partition, the locations $\{\xi_i\}_{i\geq 1}$ turn out to be irrelevant, it only matters the sizes of the jumps and whether F has a continuous part or not. Also note that if F has a continuous part and some X_i falls into such part the i will contribute to a singleton a.s. to the partition.

In the case when we took the subordinator to be a Gamma process, we were able to compute the distribution of the random partition $\{\Pi_n\}_{n\geq 1}$ for every $n \geq 1$

$$\mathbb{P}[\Pi_n = \pi] \text{ for } \pi \in \mathcal{P}_{[n]}$$

as well as the prediction rule

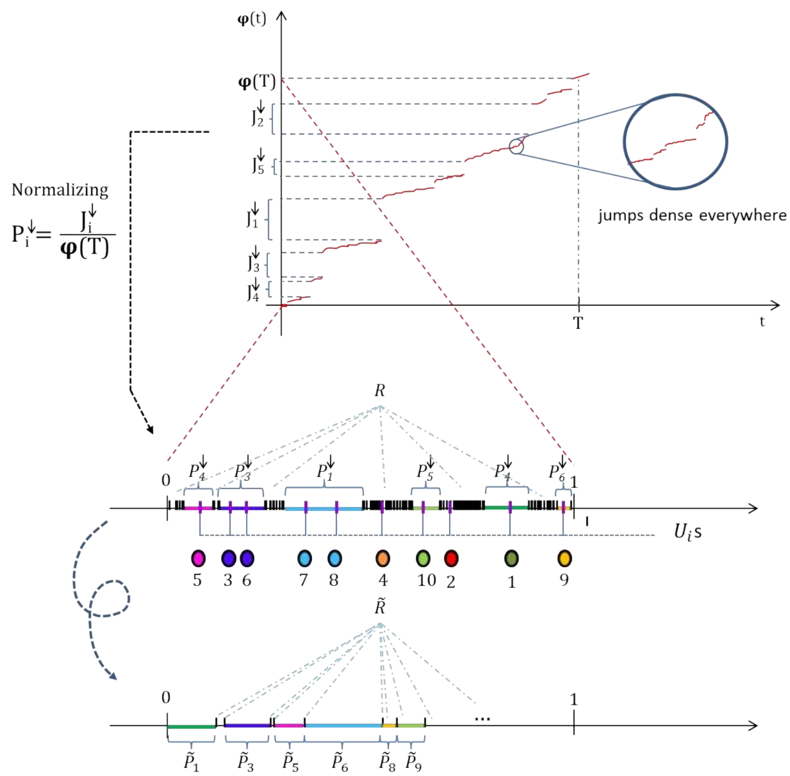
$$\mathbb{P}[\Pi_{n+1} = \pi^*|\Pi_n = \pi] \text{ for } \pi \in \mathcal{P}_{[n]}, \pi^* \in \mathcal{P}_{[n+1]}$$

In general it is not easy to compute directly this quantities, it can be done, though by introducing a latent random variable.

Second approach

The second approach, discussed in Chapter 5.3 was that of the ordered paintbox. We consider a random closed subset R of $(0, 1)$ and $\{U_i\}_{i\geq 1}$ a sequence of independent and Uniform$(0, 1)$ random variables. Then we assigned a distinct colour to each of the open gaps of R and painted numbered balls according to where the corresponding uniform random variable fell, that is if U_i and U_j fell into the same open gap of R then the ball number i and the ball number j will be painted in the same colour (the colour assigned to the open gap where the uniform random variables fell), if some U_k falls into R the kth ball will be painted in some unique colour, distinct to every other. These considerations generate a partition $\Pi_\infty = \{\Pi_n\}_{n\geq 1}$ of $\mathbb{N}$ by setting the equivalence relation $i \sim j$ iff the balls number i and j are painted in the same colour. We can recognize here two scenarios that influence greatly on the distribution of the random partition generated, the first one in when R has lebesgue measure 0 a.s., and the second one is when the lebesgue measure of R is bigger than 0 a.s. Also if R has a lebesgue measure of 0 and a finite number of open gaps a.s. the partition Π_∞ of $\mathbb{N}$ will have a finite number of blocks that will be at most the number of open gaps of R.

A ordered paintbox R can be constructed through the normalization of a subordinator as explained with detail in the Chapter 5.3 and illustrated bellow



Note that if the random probability distribution function F defined through the normalization of a subordinator has continuous part almost surely (that is the deterministic measure of the is $\beta > 0$) then the generated paintbox R will have lebesgue measure bigger than 0 a.s., on the other side, if F is discrete almost surely ($\beta = 0$) then R will have lebesgue measure of 0 almost surely. Also if the subordinator is such that its jumps are dense everywhere and its deterministic measure equals zero, then R will have infinitely many gaps and its lebesgue measure will be zero (this is the case when we take the subordinator to be the Gamma process or the σ-stable process). Note that in the latter scenario we could redefine R by permuting its open gaps according to a size-biased permutation of them, let us denote by $\tilde{R}$ this new ordered paintbox, if Π_∞ is a random partition of $\mathbb{N}$ generated through the original paintbox and $\tilde{\Pi}_\infty$ is a partition of $\mathbb{N}$ generated through the second one, then clearly Π_∞ and $\tilde{\Pi}_\infty$ will have the same distribution. The second ordered paintbox, though, might be regarded as if constructed by some stick breaking sequence.

Third approach

The third approach we used to generate an exchangeable random partition of $\mathbb{N}$ was that of the Chinese restaurant construction. Customers arrive to a restaurant one by one, and they are seated at the numbered tables in the following way: The first customer to arrive will be seated at table 1. After the arrival of the nth customer, if there are k occupied tables with frequencies $n_1, ..., n_k$ the $(n+1)$th customer will be seated at table j with probability p_{n,k,n_j}, $j = 1, ..., k$ or at a new table with probability $1 - \sum_{j=1}^{k} p_{n,k,n_j}$. The numbers p_{n,k,n_j} might or might not depend on n, k and n_j but they must satisfy that $0 \leq p_{n,k,n_j} \leq 1$ and $\sum_{j=1}^{k} p_{n,k,n_j} \leq 1$ for every $n \geq 1$, $1 \leq k \leq n$ and $n_1, ..., n_k$ such that $1 \leq n_j \leq n$ for $j = 1, ..., k$ and $\sum_{j=1}^{k} n_j = n$, the numbers p_{n,k,n_j} might also depend on some extra parameters.

At stage n (after the arrival of the nth customer) this generates a partition of $[n]$ by setting the equivalence relation $i \sim j$ iff the ith customer and the jth customer are seated at the same table. Notice that this is equivalent to setting a prediction rule such that for $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ with $|A_j| = n_j$, $j = 1, 2, ..., k$ and $\pi^* \in \mathcal{P}_{[n+1]}$

$$\mathbb{P}[\Pi_{n+1} = \pi^* | \Pi_n = \pi] = \begin{cases} p_{n,k,n_j} & \text{if } \pi^* = \pi^{(j)}, j = 1, ..., k \\ 1 - \sum_{j=1}^{k} p_{n,k,n_j} & \text{if } \pi^* = \pi^{(0)} \\ 0 & \text{otherwise} \end{cases}$$

where

$$\pi^{(j)} = \{A_1, ..., A_{j-1}, A_j \cup \{n+1\}, A_{j+1}, ..., A_k\} \in \mathcal{P}_{[n+1]}$$
$$\pi^{(0)} = \{A_1, ...A_k, \{n+1\}\} \in \mathcal{P}_{[n+1]}$$

In general the family of partitions, $\{\Pi_n\}_{n \geq 1}$, generated through this procedure will be consistent,though necessarily exchangeable, (see Example 2.2). The scenarios that we analysed in the thesis (the one of the consistent random permutation and the one of the two-parameter model) are particular cases of setting the prediction rule such that the generated random partition of $\mathbb{N}$ is also exchangeable.

Notice that in order for the partition of $\mathbb{N}$ to be exchangeable, the prediction rule has to take the following form: For every $n \geq 1$ and $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ with $n_i = |A_i|$, for every $i = 1, 2, ..., k$

$$\mathbb{P}[\Pi_{n+1} = \pi^{(j)} | \Pi_n = \pi] = \frac{p_{n+1}(n_1, ..., n_{j-1}, n_j + 1, n_{j+1}, ..., n_k)}{p_n(n_1, ...n_k)}$$
$$\mathbb{P}[\Pi_{n+1} = \pi^{(0)} | \Pi_n = \pi] = \frac{p_{n+1}(n_1, ..., n_k, 1)}{p_n(n_1, ...n_k)}$$
$$\mathbb{P}[\Pi_{n+1} = \pi^* | \Pi_n = \pi] = 0$$

for every $\pi^* \in \mathcal{P}_{[n+1]}$, $\pi^* \neq \pi^{(j)}$, $j = 0, 1, ..., k$. where

$$\pi^{(j)} = \{A_1, ..., A_{j-1}, A_j \cup \{n+1\}, A_{j+1}, ..., A_k\} \in \mathcal{P}_{[n+1]}$$
$$\pi^{(0)} = \{A_1, ...A_k, \{n+1\}\} \in \mathcal{P}_{[n+1]}$$

and the numbers $p_l(l_1, ..., l_m)$ refer to the probability of the event $\Pi_l = \tilde{\pi}$ where $\tilde{\pi} \in \mathcal{P}_{[l]}$ is any partition of $[l]$ into m blocks with frequencies $l_1, ..., l_m$. We also ask p_l to be a symmetric function of its arguments for every $l \geq 1$.

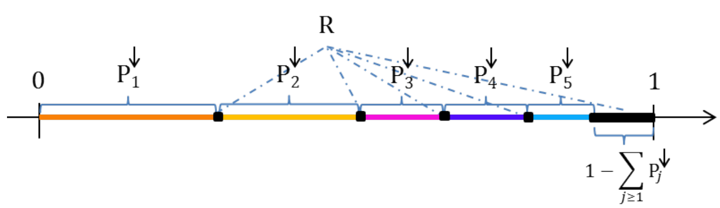
If the generated partition of $\mathbb{N}$ is exchangeable. At stage n let K_n be the number of occupied tables and let $\mathbf{N} = (N_{n,j}, j \leq 1)$ be the random sequence such that $N_{n,j}$ stands for number of customers seating at table j for $j \leq 1$. Let $\mathbf{N}^{\downarrow} = (N^{\downarrow}_{n,j}, j \leq 1)$ be the rearrangement in decreasing order of $\mathbf{N}$. Clearly $\sum_{j=1}^{K_n} N^{\downarrow}_{n,j} = n$ and $N^{\downarrow}_{n,j} \geq 1$ for $j = 1, ..., K_n$ and $N^{\downarrow}_{n,j} = 0$ for $j > K_n$. By Kingman's representation theorem the following almost sure limits exist

$$\lim_{n \to \infty} \frac{N^{\downarrow}_{n,j}}{n} \to P^{\downarrow}_j$$

and hence

$$R := \left(\left[\sum_{j=1}^{\infty} P^{\downarrow}_j, 1 \right] \right) \bigcup \left(\bigcup_{i=1}^{\infty} \left\{ \sum_{j=1}^{i} P^{\downarrow}_j \right\} \right)$$

is a random closed subset of $[0, 1]$ so that it might be regarded as an ordered paintbox. As illustrated below

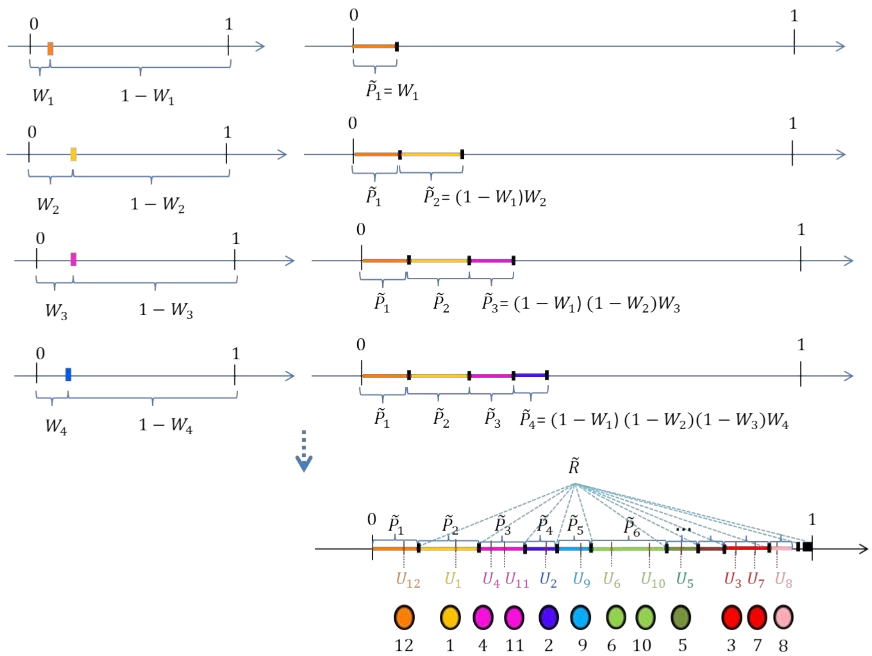


Fourth approach

The fourth and last approach we reviewed was the stick breaking procedure. That is we let $\{W_i\}_{i \geq 1}$ be a sequence of random variables such that W_i takes values on $[0, 1]$ and define

$$\tilde{P}_j = W_j \prod_{i=1}^{j-1} (1 - W_i)$$

as illustrated below



once again

$$\tilde{R} = \bigcup_{i=1}^{\infty} \left\{ \sum_{j=1}^{i} \tilde{P}_j \right\} \setminus \{0, 1\} \tag{9.1}$$

defines a random closed subset of $(0, 1)$ so that it might be regarded as an ordered paintbox. The generation of a random exchangeable partition of $\mathbb{N}$ follows as in the second approach, that is assigning colours to the open gaps of $\tilde{R}$, throwing independent uniform random variables into $[0, 1]$ and colouring numbered balls according to where did the corresponding uniform random variable fell.

In the thesis we did not analysed what happens in this general scenario, we only analysed the case where $\{W_i\}_{i\geq 1}$ is a sequence of independent random variables such that $W_i \sim Beta(1-\sigma, \theta + i\sigma)$ with (σ, θ) that hold one of the following

i) $0 \leq \sigma \leq 1$ and $\theta > -\sigma$

ii) $\sigma = -r < 0$ and $\theta = mr$ for some $m = 1, 2, ...$

A result of Pitman in [18] states that if the $\{W_i\}_{i\geq 1}$ are independent random variables such that $0 < W_1 < 1$ a.s. then the distribution of the random partition of $\mathbb{N}$ generated as above is either determined by a (σ, θ) model, or that of the Coupon's collector partition.

Dirichlet process and Poisson-Dirichlet process

In the thesis the above mentioned approaches were particularly analysed for the Dirichlet process and the Poisson-Dirichlet process cases.

The Dirichlet process with parameter $\theta > 0$, which is a particular case of the Poisson-Dirichlet process ($\sigma = 0$) can be constructed by normalizing a Gamma process Φ on any space $(\mathbb{X}, \mathcal{X})$ such that $\Phi(\mathbb{X}) \sim Gamma(\theta, 1)$. This is equivalent as taking the Gamma subordinator on $\mathbb{R}^+$, choosing the stopping time $T = \theta$ and generate the ordered paintbox through the procedure explained in Chapter 5.3, this order paintbox can also be constructed by the stick breaking sequence

$$\tilde{P}_j = W_j \prod_{i=1}^{j-1}(1 - W_i)$$

where $\{W_i\}_{i\geq 1}$ are independent random variables such that $W_i \sim Beta(1, \theta)$ and defining the closed subset of $(0, 1)$ as in equation (9.1). We also constructed the Dirichlet process by means of the chinese restaurant construction by setting $p_{n,k,n_j} = \frac{n_i}{n+\theta}$ (following the notation p_{n,k,n_j} defined above in this section). For the Dirchlet process one has that for every $n \geq 1$ and $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ with $n_i = |A_i|$

$$\mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k) = \frac{\theta^k}{(\theta)_{n-1\uparrow 1}} \prod_{i=1}^{k}(n_i - 1)!$$

The Poisson Dirchlet process with parameters (σ, θ) can also be constructed in many ways, one possible way is through the chinese restaurant procedure by setting $p_{n,k,n_j} = \frac{n_i-\sigma}{n+\theta}$, another one is through the stick breaking sequence

$$\tilde{P}_j = W_j \prod_{i=1}^{j-1}(1 - W_i)$$

where $\{W_i\}_{i\geq 1}$ are independent random variables such that $W_i \sim Beta(1 - \sigma, \theta + i\sigma)$. The third way to construct it in the case $\theta = 0$ is by means of the normalization of a σ-stable process on $\mathbb{R}$, choosing any stopping time greater than 0 with respect to it and constructing the paintbox as in 5.3, as for the case $\theta \neq 0$, in Chapter 7 we showed how to obtain a Poisson-Dirichlet process with parameters (σ, θ) with $\theta \neq 0$ by slightly modifying the procedure of generating the paintbox also based on a σ-stable subordinator. In the case of this two parameter model we also reviewed, in Chapter 7, how to generate it by means of a branching process construction. For the Poisson Dirichlet process the EPPF takes the form

$$\mathbb{P}[\Pi_n = \pi] = p(n_1, ..., n_k) = \frac{(\theta + \sigma)_{k-1\uparrow\sigma} \prod_{i=1}^{k}(1 - \sigma)_{n_i-1\uparrow 1}}{(\theta + 1)_{n-1\uparrow 1}}$$

for every $n \geq 1$ and $\pi = \{A_1, ..., A_k\} \in \mathcal{P}_{[n]}$ with $n_i = |A_i|$.

For these models we also analysed the asymptotic behaviour of the number of blocks in Π_n, K_n, as $n \to \infty$. For the case of the Dirichlet process with parameter $\theta > 0$ we obtained that

$$\lim_{n\to\infty} \frac{K_n}{\log(n)} = \theta \text{ almost surely}$$

as for the Poisson Dirichlet process with parameter (σ, θ) with $0 < \sigma < 1$, $\theta > -\sigma$ we showed that

$$\frac{K_n}{n^\sigma} \to S_\sigma \text{ almost surely}$$

and in the pth mean (for $p > 0$), for a strickly positive random variable S_σ , with continuous density

$$\frac{d}{ds}\mathbb{P}_{(\sigma,\theta)}[S_\sigma \in ds] = g_{(\sigma,\theta)}(s) := \frac{\Gamma(\theta+1)}{\Gamma(\frac{\theta}{\sigma}+1)} s^{\frac{\theta}{\sigma}} g_\sigma(s) \quad (s > 0)$$

where

$$g_\sigma(s) = \frac{f_\sigma(s^{-\frac{1}{\sigma}})}{\sigma s^{1+\frac{1}{\sigma}}}$$

and f_σ denotes the probability density function of a positive σ-stable random variable.

Appendix A

A.1 Pólya Eggenberger urn

Imagine we have an urn containing a white balls and b black balls. At each time n a ball is drawn from the urn and returned with m additional balls having the same color. Let X_i for $i = 1, 2, ...$ be the indicator function of the event that the ith drawn ball was white. Then $S_n := \sum_{i=1}^{n} X_i$ denotes the number of white balls that have been drawn to time n and $W_n := a + m\sum_{i=1}^{n} X_i$ denotes the number of white balls at the urn after the nth draw. Note that For $n \geq 1$

$$\mathbb{P}[X_{n+1} = 1|X_1, X_2, ..., X_n] = \frac{|\{\text{white balls in the urn}\}|}{|\{\text{balls in the urn}\}|} = \frac{a + mS_n}{a + b + mn} = \frac{W_n}{a + b + mn} \tag{A.1}$$

It is easy to see that

$$\mathbb{P}[X_1 = x_1, ..., X_n = x_n] = \frac{(a)_{x\uparrow m}(b)_{n-x\uparrow m}}{(a + b)_{n\uparrow m}} \tag{A.2}$$

where $x = \sum_{i=1}^{n} x_i$ and $(r)_{j\uparrow c} = r(r + c)(r + 2c) * \cdots * (r + (j - 1)c)$.

The probability $\mathbb{P}[X_1 = x_1, ..., X_n = x_n]$ only depends on $x_1, ..., x_n$ be means of a symmetric function of them, hence $X_1, X_2, ...$ are exchangeable random variables.

Now we are interested in the distribution of $\frac{S_n}{n}$ as n goes to ∞, this quantity clearly denotes the proportion of white balls drawn on the long run. Note that

$$\lim_{n\to\infty} \frac{S_n}{m} = \lim_{n\to\infty} \frac{a + mS_n}{a + b + mn} = \lim_{n\to\infty} \frac{W_n}{a + b + mn}$$

We are going to prove that $\frac{W_n}{a+b+nm}$ has an almost sure limit, by proving it is a non-negative martingale respect to $\mathcal{F}_n = \sigma(X_1, ..., X_n)$ (where $\sigma(X_1, X_2, ..., X_n)$ stands for the generated σ-algebra by $X_1, ..., X_n$)

i) As $0 \leq S_n \leq a + b + nm$ then $0 \leq \mathbb{E}\left[\left|\frac{W_n}{a+b+nm}\right|\right] \leq 1$

ii) $\frac{W_n}{a+b+nm}$ is a continuous function of $X_1, ..., X_n$ and hence a measurable function.

iii)

$$\begin{aligned}
\mathbb{E}\left[\frac{W_{n+1}}{a+b+(n+1)m}\middle|\mathcal{F}_n\right] &= \frac{1}{a+b+(n+1)m}\left\{\mathbb{E}[W_n + mX_{n+1}|\mathcal{F}_n]\right\} \\
&= \frac{1}{a+b+(n+1)m}\left\{m\mathbb{E}[X_{n+1}|\mathcal{F}_n] + \mathbb{E}[W_n|\mathcal{F}_n]\right\} \\
&= \frac{1}{a+b+(n+1)m}\left\{m\frac{W_n}{a+b+mn} + W_n\right\} \\
&= \frac{1}{a+b+(n+1)m}\left\{\frac{mW_n + (a+b+nm)W_n}{a+b+mn}\right\} \\
&= \frac{1}{a+b+(n+1)m}\left\{W_n\frac{a+b+(n+1)m}{a+b+mn}\right\} \\
&= \frac{W_n}{a+b+nm}
\end{aligned}$$

With this we have proved that $\frac{W_n}{a+b+nm}$ is a non-negative martingale and hence it has an almost sure limit ρ. Thus

$$\lim_{n\to\infty}\frac{S_n}{n} = \lim_{n\to\infty}\frac{W_n}{a+b+nm} = \rho \text{ almost surely}$$

Using the fact that the random variables are exchangeable and (A.2) we get

$$\mathbb{P}[S_n = k] = \binom{n}{k}\frac{(a)_{k\uparrow m}(b)_{n-k\uparrow m}}{(a+b)_{n\uparrow m}}, \tag{A.3}$$

manipulating $\mathbb{E}[e^{it\frac{S_n}{n}}]$ and using lebesgue dominated convergence it can be proven that $\rho \sim Beta\left(\frac{a}{m}, \frac{b}{m}\right)$.

Hence

$$\frac{S_n}{n} \to \rho \text{ a.s.} \qquad \text{where } \rho \sim Beta\left(\frac{a}{m}, \frac{b}{m}\right)$$

A.2 Proof of de Finetti's representation theorem

Theorem A.1 (de Finetti's representation Theorem)**.** *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of exchangeable random variables taking values on a Polish space $\mathbb{X}$ endowed with its Borel σ-field $\mathcal{X}$, also let $\mathcal{P}_{\mathbb{X}}$ be the space of all probability measures over $\mathbb{X}$, then*

1. For every $A_1, A_2, ..., A_n \in \mathcal{X}$

$$\mathbb{P}[X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n] = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} \tilde{p}[A_i] Q(d\tilde{p})$$

where $\tilde{p}$ is known as the directing random probability measure and its distribution, Q, is known as the de Finetti's measure.

2. For every $A \in \mathcal{X}$, and as n goes to infinity, the empiric distribution satisfies

$$F_n(A) := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(A) \to \tilde{p}(A) \, a.s. \tag{A.4}$$

where $\tilde{p} \sim Q$.

For simplicity we are going to demonstrate the theorem when $\mathbb{X} = \mathbb{R}$ and $\mathcal{X} = \mathscr{B}(\mathbb{R})$. For $n \geq 1$ let us define the sub-σ-fields

$$\begin{aligned} \overline{\mathscr{F}}_n &:= \sigma(X_{n+1}, X_{n+2}, ...) \\ \mathscr{G}_n &:= \sigma(f(X_1, ..., X_n) : f \text{ is a symmetric function}) \\ \mathscr{H}_n &:= \sigma(\mathscr{G}_n \cup \overline{\mathscr{F}}_n) \end{aligned}$$

In order to complete the proof, we will divide the sequentially prove the following statements:

a) For every $n \geq 1$ $\mathscr{H}_{n+1} \subseteq \mathscr{H}_n$.

b) For every bounded function $\gamma : \mathbb{R} \to \mathbb{R}$ we have that as $n \to \infty$

$$\frac{1}{n} \sum_{j=1}^{n} \gamma(X_j) \to \mathbb{E}[\gamma(X_1) | \mathscr{H}] = \mathbb{E}[\gamma(X_j) | \mathscr{H}]$$

almost surely.

c) For every bounded function $\gamma : \mathbb{R}^k \to \mathbb{R}$ (where $1 \leq k \leq n$) we have that as $n \to \infty$

$$\frac{1}{n^k} \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} \gamma(X_{j_1}, ..., X_{j_k}) \to \mathbb{E}[\gamma(X_1, ..., X_k) | \mathscr{H}]$$

almost surely.

d) For the particular choice of $\gamma = \prod_{i=1}^{k} \gamma_i$ where $\gamma_i : \mathbb{R} \to \mathbb{R}$ is a bounded function for each i, and using b) and c) we obtain

$$\mathbb{E}\left[\prod_{i=1}^{k} \gamma_i(X_i) \middle| \mathscr{H} \right] = \prod_{i=1}^{k} \mathbb{E}[\gamma_i(X_i) | \mathscr{H}]$$

e) For the particular choice $\gamma(x) = \mathbf{1}_{x \in A}$ for some $A \in \mathcal{B}(\mathbb{R})$, b) becomes

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{X_j \in A} \to \mathbb{E}[\mathbf{1}_{X_1 \in A}|\mathscr{H}] = \mathbb{P}[X_1 \in A|\mathscr{H}] =: \tilde{p}(A)$$

which proves 2. of the de Finetti's representation theorem, where Q y the probability distribution of $\tilde{p}$ defined above.

f) Taking $k = n$, and $\gamma_i(x) = \mathbf{1}_{x \in A_i}$ for some $A_1, ...A_n \in \mathcal{B}(\mathbb{R})$, d) reduces to the form

$$\mathbb{P}[X_1 \in A_1, ..., X_n \in A_n|\mathscr{H}] = \prod_{i=1}^{n} \mathbb{P}[X_i \in A_i|\mathscr{H}] = \prod_{i=1}^{n} \tilde{p}(A_i)$$

g) Finally we will see that f) implies 1. of the de Finetti's representation theorem.

First of all let us see that for every $n \geq 1$ $\mathscr{H}_{n+1} \subseteq \mathscr{H}_n$. Note that for every symmetric function $f : \mathbb{R}^{n+1} \to \mathbb{R}$ we can write

$$f(X_1, ..., X_n, X_{n+1}) = h(g(X_1, ..., X_n), X_{n+1})$$

for some functions $h : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ where g is a symmetric function. Thus

$$\mathscr{G}_{n+1} \subseteq \mathscr{G}_n \cup \sigma(X_{n+1}) \subseteq \mathscr{G}_n \cup \overline{\mathscr{F}}_n,$$

and considering the generated σ-fields we obtain

$$\mathscr{G}_{n+1} = \sigma(\mathscr{G}_{n+1}) \subseteq \sigma(\mathscr{G}_n \cup \overline{\mathscr{F}}_n) = \mathscr{H}_n.$$

Also

$$\overline{\mathscr{F}}_{n+1} \subseteq \overline{\mathscr{F}}_n \subseteq \mathscr{G}_n \cup \overline{\mathscr{F}}_n$$

so generating the corresponding σ-fields we get

$$\overline{\mathscr{F}}_{n+1} = \sigma(\overline{\mathscr{F}}_{n+1}) \subseteq \sigma(\mathscr{G}_n \cup \overline{\mathscr{F}}_n) = \mathscr{H}_n.$$

Finally, as $\mathscr{G}_{n+1} \cup \overline{\mathscr{F}}_{n+1} \subseteq \mathscr{H}_n$,

$$\mathscr{H}_{n+1} = \sigma(\mathscr{G}_{n+1} \cup \overline{\mathscr{F}}_{n+1}) \subseteq \sigma(\mathscr{H}_n) = \mathscr{H}_n.$$

Exchangeabilty of the random variables implies that for any random variable Z of the form $Z = g(f(X_1, ..., X_n), Z_{n+1}, Z_{n+2}, ...)$, where f is a symmetric function, we have that

$$(X_1, Z) \stackrel{d}{=} (X_j, Z), \quad 1 \leq j \leq n.$$

and the equation above is valid for every $\mathscr{H}_n$-measurable random variable Z. So for a bounded function γ we obtain

$$\mathbb{E}[\gamma(X_1)|\mathscr{H}_n] = \mathbb{E}[\gamma(X_j)|\mathscr{H}_n] \quad 1 \leq j \leq n. \tag{A.5}$$

As $\frac{1}{n}\sum_{j=1}^{n} X_j$ is a symmetric function of $X_1, ..., X_n$ and hence $\mathscr{H}_n$-measurable,

$$\begin{aligned}\mathbb{E}[\gamma(X_1)|\mathscr{H}_n] &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\gamma(X_1)|\mathscr{H}_n] \\ &= \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}[\gamma(X_j)|\mathscr{H}_n] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\gamma(X_j)\middle|\mathscr{H}_n\right] \\ &= \frac{1}{n}\sum_{j=1}^{n}\gamma(X_j)\end{aligned}$$

That is

$$\left\{\frac{1}{n}\sum_{j=1}^{n}\gamma(X_j)\right\}_{n\geq 1}$$

is a closed and reversed martingale, and by the reversed martingale convergence theorem and equation (A.5), as $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n}\gamma(X_j) \to \mathbb{E}[\gamma(X_1)|\mathscr{H}] = \mathbb{E}[\gamma(X_j)|\mathscr{H}] \quad 1 \leq j \leq n \tag{A.6}$$

almost surely, where $\mathscr{H} = \bigcap_{n\geq 1}\mathscr{H}_n$. Now let us define for every $n \geq 1$ and $1 \leq k < n$

$$\begin{aligned}A_{n,k} &:= \{(j_1, ..., j_k) : j_i \in [n], i \in [k]\} \\ B_{n,k} &:= \{(j_1, ..., j_k) \in A_{n,k} : j_i \neq j_l, \forall i \neq l\} \\ C_{n,k} &:= A_{n,k} \setminus B_{n,k}\end{aligned}$$

As the random variables are exchangeable, then we also have that for every $\mathscr{H}_n$-measurable function, Z,

$$(X_1, ..., X_k, Z) \stackrel{d}{=} (X_{j_1}, ..., X_{j_k}, Z), \quad (j_1, ..., j_k) \in B_{n,k}$$

Thus for every bounded function $\gamma : \mathbb{R}^k \to \mathbb{R}$ and $(j_1, ..., j_k) \in B_{n,k}$

$$\mathbb{E}[\gamma(X_1, ..., X_k)|\mathscr{H}_n] = \mathbb{E}[\gamma(X_{j_1}, ..., X_{j_k})|\mathscr{H}_n].$$

As $\frac{1}{|B_{n,k}|}\sum_{(j_1,...j_k)\in B_{n,k}}\gamma(X_{j_1}, ..., X_{j_k})$ is a symmetric function of $X_1, ..., X_n$, and hence $\mathscr{H}_n$-measurable

$$\begin{aligned}
\mathbb{E}[\gamma(X_1,...,X_k)|\mathscr{H}_n] &= \frac{1}{|B_{n,k}|}\sum_{(j_1,...j_k)\in B_{n,k}} \mathbb{E}[\gamma(X_{j_1},...,X_{j_k})|\mathscr{H}_n] \\
&= \mathbb{E}\left[\frac{1}{|B_{n,k}|}\sum_{(j_1,...j_k)\in B_{n,k}} \gamma(X_{j_1},...,X_{j_k})\middle|\mathscr{H}_n\right] \\
&= \frac{1}{|B_{n,k}|}\sum_{(j_1,...j_k)\in B_{n,k}} \gamma(X_{j_1},...,X_{j_k}) \\
&= \frac{1}{|B_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} \gamma(X_{j_1},...,X_{j_k})\mathbf{1}_{\{(j_1,...,j_k)\in B_{n,k}\}}
\end{aligned}$$

So by the reversed martingale convergence theorem, as $n\to\infty$

$$\frac{1}{|B_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} \gamma(X_{j_1},...,X_{j_k})\mathbf{1}_{\{(j_1,...,j_k)\in B_{n,k}\}} \to \mathbb{E}[\gamma(X_1,...,X_k)|\mathscr{H}]. \tag{A.7}$$

almost surely. Now, note that for k fixed

$$\lim_{n\to\infty}\frac{|B_{n,k}|}{|A_{n,k}|} = \lim_{n\to\infty}\frac{k!\binom{n}{k}}{n^k} = \lim_{n\to\infty}\frac{n(n-1)\cdots(n-(k-1))}{n^k} = 1 \tag{A.8}$$

and

$$\lim_{n\to\infty}\frac{|C_{n,k}|}{|A_{n,k}|} = \lim_{n\to\infty}\frac{|A_{n,k}|-|B_{n,k}|}{|A_{n,k}|} = \lim_{n\to\infty} 1-\frac{|B_{n,k}|}{|A_{n,k}|} = 0 \tag{A.9}$$

this last couple of equations tell us that for k fixed and as n grows, the proportion vectors $(j_1,...,j_k)\in C_{n,k}$ (relative to the number of vectors in $A_{n,k}$) becomes irrelevant. As γ is bounded, there exist $c>0$ such that for every $(j_1,...,j_k)\in C_{n,k}$, $|\gamma(X_{j_1},...,X_{j_k})|\le c$. So

$$\begin{aligned}
0 &\le \left|\frac{1}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} \gamma(X_{j_1},...,X_{j_k})\mathbf{1}_{\{(j_1,...,j_k)\in C_{n,k}\}}\right| \\
&\le \frac{1}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} |\gamma(X_{j_1},...,X_{j_k})|\mathbf{1}_{\{(j_1,...,j_k)\in C_{n,k}\}} \\
&\le \frac{c}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} \mathbf{1}_{\{(j_1,...,j_k)\in C_{n,k}\}} \\
&= \frac{c|C_{n,k}|}{|A_{n,k}|}
\end{aligned}$$

hence, making n grow and by equation (A.9)

$$\frac{1}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n} \gamma(X_{j_1},...,X_{j_k})\mathbf{1}_{\{(j_1,...,j_k)\in C_{n,k}\}} \to 0. \tag{A.10}$$

Also

$$\begin{aligned}
&\frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\\
&\quad=\frac{1}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\mathbf{1}_{\{(j_1,...,j_k)\in A_{n,k}\}}\\
&\quad=\frac{1}{|A_{n,k}|}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\left(\mathbf{1}_{\{(j_1,...,j_k)\in B_{n,k}\}}+\mathbf{1}_{\{(j_1,...,j_k)\in C_{n,k}\}}\right)
\end{aligned}$$

so putting together this last equation with equations (A.7), (A.10) and (A.8) we obtain

$$\frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\to\mathbb{E}[\gamma(X_1,...,X_k)|\mathscr{H}]. \tag{A.11}$$

If we consider $\gamma_j:\mathbb{R}\to\mathbb{R}$ bounded functions and $\gamma:\mathbb{R}^k\to\mathbb{R}$ defined by $\gamma(x_1,...,x_k)=\prod_{j=1}^{k}\gamma_j(x_j)$, as γ is bounded, by equation (A.11)

$$\frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\to\mathbb{E}[\gamma(X_1,...,X_k)|\mathscr{H}]=\mathbb{E}\left[\prod_{i=1}^{k}\gamma_i(X_i)\middle|\mathscr{H}\right] \tag{A.12}$$

On the other hand

$$\frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})=\prod_{i=1}^{k}\left(\frac{1}{n}\sum_{j_i=1}^{n}\gamma_i(X_{j_i})\right)$$

so equation (A.6) implies

$$\frac{1}{n^k}\sum_{j_1=1}^{n}\cdots\sum_{j_k=1}^{n}\gamma(X_{j_1},...,X_{j_k})\to\prod_{i=1}^{k}\mathbb{E}[\gamma_i(X_i)|\mathscr{H}], \tag{A.13}$$

thus we can conclude that almost surely

$$\mathbb{E}\left[\prod_{i=1}^{k}\gamma_i(X_i)\middle|\mathscr{H}\right]=\prod_{i=1}^{k}\mathbb{E}[\gamma_i(X_i)|\mathscr{H}] \tag{A.14}$$

Now let $A\in\mathscr{B}(\mathbb{R})$ and let $\gamma(x)=\mathbf{1}_{\{X_1\in A\}}$, clearly γ is bounded so equation (A.6) reduces to the form

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\in A\}}\to\mathbb{E}[\mathbf{1}_{\{X_1\in A\}}|\mathscr{H}]=\mathbb{P}[X_1\in A|\mathscr{H}]. \tag{A.15}$$

Let us define $\tilde{p}(A) := \mathbb{P}[X_1 \in A|\mathscr{H}]$ for every $A \in \mathscr{B}(\mathbb{R})$. Note that for $A \in \mathscr{B}(\mathbb{R})$ fixed, $p(A)$ is a $\mathscr{H}$-measurable random variable and for $\omega \in \Omega$ fixed $p(\bullet)(\omega)$ is a probability measure over $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

By equation (A.5) we have that

$$p(A) = \mathbb{E}[X_j \in A|\mathscr{H}] \quad 1 \leq j$$

almost surely. Now, let $A_1, ..., A_n \in \mathscr{B}(\mathbb{R})$, let us consider $\gamma_j(x) = \mathbf{1}_{\{x \in A_j\}}$ for $1 \leq j \leq n$, obviously γ_j is bounded for every j, hence by equation (A.14) we obtain that

$$\begin{aligned}
\mathbb{P}[X_1 \in A_1, ..., X_n \in A_n|\mathscr{H}] &= \mathbb{E}[\mathbf{1}_{\{X_1 \in A_1, ..., X_n \in A_n\}}|\mathscr{H}] \\
&= \mathbb{E}\left[\prod_{j=1}^{n} \mathbf{1}_{\{X_j \in A_j\}} \middle| \mathscr{H}\right] \\
&= \prod_{j=1}^{n} \mathbb{E}[\mathbf{1}_{\{X_j \in A_j\}}|\mathscr{H}] \\
&= \prod_{j=1}^{n} \mathbb{P}[X_j \in A_j|\mathscr{H}] \\
&= \prod_{j=1}^{n} \tilde{p}(A_j)
\end{aligned}$$

almost surely. If we define the event $E = \{X_1 \in A_1, ..., X_n \in A_n\}$ this reads as

$$\mathbb{P}[E|\mathscr{H}] = \mathbb{E}[\mathbf{1}_E|\mathscr{H}] = \prod_{j=1}^{n} \tilde{p}(A_j) \tag{A.16}$$

almost surely. Let

$$\mathscr{T} = \sigma(\tilde{p}) := \sigma(\tilde{p}(A) : A \in \mathscr{B}(\mathbb{R})),$$

As mentioned above, $p(A)$ is an $\mathscr{H}$-measurable random variable for every $A \in \mathscr{B}(\mathbb{R})$, hence it is obvious that $\mathscr{T} \subseteq \mathscr{H}$ and by the *tower property* of conditional expectation and equation (A.16)

$$\begin{aligned}
\mathbb{P}[E|\tilde{p}] &= \mathbb{E}[\mathbf{1}_E|\tilde{p}] \\
&= \mathbb{E}[\mathbf{1}_E|\mathscr{T}] \\
&= \mathbb{E}[\mathbb{E}[\mathbf{1}_E|\mathscr{H}]|\mathscr{T}] \\
&= \mathbb{E}\left[\prod_{j=1}^{n} \tilde{p}(A_j) \middle| \mathscr{T}\right] \\
&= \prod_{j=1}^{n} \tilde{p}(A_j)
\end{aligned}$$

Thus for every $A_1, ..., A_n \in \mathscr{B}(\mathbb{R})$

$$\mathbb{P}[X_1 \in A_1, ..., X_n \in A_n | \mathscr{H}] = \prod_{j=1}^{n} \tilde{p}(A_j) = \mathbb{P}[X_1 \in A_1, ..., X_n \in A_n | \tilde{p}] \quad \text{(A.17)}$$

Finally let Q be the distribution of $\tilde{p}$ over the space of all probability measure over $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then

$$\mathbb{P}[X_1 \in A_1, ..., X_n \in A_n] = \int_{\mathcal{P}_{\mathbb{R}}} \mathbb{P}[X_1 \in A_1, ..., X_n \in A_n | \tilde{p}] Q(d\tilde{p})$$

and by equation (A.17)

$$\mathbb{P}[X_1 \in A_1, ..., X_n \in A_n] = \int_{\mathcal{P}_{\mathbb{R}}} \prod_{j=1}^{n} \tilde{p}(A_j) Q(d\tilde{p})$$

which finishes the proof of de Finetti's representation theorem. $\square$

A.3 The Radon-Nikodym's martingale

Let $(\Omega, \mathscr{F})$ a measurable space. Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $(\Omega, \mathscr{F})$ such that $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{Q}$. Let $\{\mathscr{F}_n\}_{n \geq 1}$ be a family of sub σ-fields of $\mathscr{F}$ such that for ever $n \geq 1$ $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$. Then

$$\left\{ \left. \frac{d\mathbb{P}}{d\mathbb{Q}} \right|_{\mathscr{F}_n} \right\}_{n \geq 1}$$

is a positive $\mathbb{Q}$-martingale with respect to $\{\mathscr{F}_n\}_{n \geq 1}$.

1. The fact that it is positive follows from the fact that $\mathbb{P}$ and $\mathbb{Q}$ are both probability measures and hence are non-negative measures.

2. By definition $\left. \frac{d\mathbb{P}}{d\mathbb{Q}} \right|_{\mathscr{F}_n}$ is $\mathscr{F}_n$-measurable.

3. Note that also by definition, for every $n \geq 1$, for every $A \in \mathscr{F}_n$

$$\mathbb{E}_{\mathbb{Q}} \left[\mathbf{1}_A \left. \frac{d\mathbb{P}}{d\mathbb{Q}} \right|_{\mathscr{F}_n} \right] = \mathbb{E}_{\mathbb{Q}} \left[\mathbf{1}_A \frac{d\mathbb{P}}{d\mathbb{Q}} \right] = \int_A \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{Q} = \int_A d\mathbb{P} = \mathbb{P}[A]$$

4. As a consequence of 3. we get that, as $\Omega \in \mathscr{F}_n$ for every $n \geq 1$ then

$$\mathbb{E}_{\mathbb{Q}} \left[\left. \frac{d\mathbb{P}}{d\mathbb{Q}} \right|_{\mathscr{F}_n} \right] = \mathbb{E}_{\mathbb{Q}} \left[\mathbf{1}_\Omega \left. \frac{d\mathbb{P}}{d\mathbb{Q}} \right|_{\mathscr{F}_n} \right] = \mathbb{P}[\Omega] = 1 < \infty.$$

5. Also as a consequence of 3. we obtain that for every $n \geq 1$ and $A \in \mathscr{F}_n \subseteq \mathscr{F}_{n+1}$

$$\mathbb{E}_{\mathbb{Q}}\left[\mathbf{1}_A \frac{d\mathbb{P}}{d\mathbb{Q}}\bigg|_{\mathscr{F}_n}\right] = \mathbb{P}[A] = \mathbb{E}_{\mathbb{Q}}\left[\mathbf{1}_A \frac{d\mathbb{P}}{d\mathbb{Q}}\bigg|_{\mathscr{F}_{n+1}}\right]$$

As $\frac{d\mathbb{P}}{d\mathbb{Q}}\big|_{\mathscr{F}_n}$ is a $\mathscr{F}_n$-measurable. Then by definition of conditional expectation

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\bigg|_{\mathscr{F}_{n+1}}\right)\bigg|\mathscr{F}_n\right] = \frac{d\mathbb{P}}{d\mathbb{Q}}\bigg|_{\mathscr{F}_n}$$

The above considerations prove the statement.

Bibliography

[1] CHARALAMBIDES, C. A. and SINGH, J. (1988) *A Review of the Stirling Numbers, their Generalizations and Statistical Applications*. Communications in Statistics-Theory and Methods, **17**.

[2] CHARALAMBIDES, C. A. (2005). *Combinatorial Methods in Discrete Distributions*. Hoboken, NJ: Wiley.

[3] FAVARO, S., LIJOI, A., MENA, R., PRÜNSTER, I. (2009) *Bayesian Nonparametric Inference for Species Variety with a Two Parameter Poisson-Dirichlet Process Prior*. Collegio Carlo Alberto **123**.

[4] GNEDIN, A. V. (2010) *Regeneration in Random Combinatorial Structures.* Probability Surveys, Vol. **7**.

[5] GNEDIN, A. V. and PITMAN, J. (2006) *Exchangeable Gibbs Partitions and Stirling Triangles.* Journal of Mathematical Science, Vol. **138**, No. **3**

[6] HJORT, N. L., HOLMES, C., MÜLLER, P. and WALKER, S. G. (2010) *Bayesian Nonparametrics.* Cambridge Series in Statistical and Probabilistic Mathematics.

[7] HOLLANDER, M. and KORWAR, R. M. (1973) *Contributions to the theory of Dirichlet Processes.* The Annals of Probability, Vol. **1**, No. **4**.

[8] HSU, L. C. and SHIUE, P. J.-S. (1998) *A Unified Approach to Generalized Stirling Numbers.* Advances in Applied Mathematics, **20**.

[9] JAMES, L. F., LIJOI, A. and PRÜNSTER, I. (2009) *Posterior Analysis for Normalized Random Measures with Independent Increments.* Scandinavian Journal of Statistics, Vol. **36**, Issue **1**.

[10] KEROV, S. V. (2005) *Coherent Random Allocations, and the Ewens Pitman Sampling Formula.* Zap. Nauchn. Sem. POMI, Vol. **325**.

[11] KINGMAN, J.F.C. (1993) *Poisson Processes.* Oxford Studies in Probability **3**.

[12] LIJOI, A., MENA, R. and PRÜNSTER, I. (2007) *Bayesian Nonparametric Estimation of the Probability of Discovering New Species.* Biometrika **94**.

[13] LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2008) *Investigating Nonparametric Priors with Gibbs Structure*. Statistica Sinica **18**.

[14] MARTÍNEZ, A. F. (2015) *Usages Of Random Combinatorial Structures in Statistics; a Bayesian Nonparametric Approach*. UNAM.

[15] OKOUNKOV, A. (2003) *The Uses of Random Partitions*. arXiv:math-ph/0309015

[16] PÉREZ SEGUI, M. L. (2010) *Combinatoria Avanzada*. Cuadernos de Olimpiadas de Matemáticas.

[17] PERMAN, M., PITMAN, J. and YOR, M. (1991) *Size-biased Sampling of Poisson Point Processes and Excursions.*

[18] PITMAN, J. (2006) *Combinatorial Stochastic Processes*. Lecture Notes in Mathematics **1875**.

[19] RIORDAN, J. (1958) *An Introduction to Combinatorial Analysis*. Wiley, New York.

[20] TOSCANO, L. (1939) *Numeri Di Stirling Generalizzati Operatori Differenziali e Dolinomi Ipergeometrici*. Commentationes Pontificia Academica Scientarum **3**.