Asymptotic analysis and dependent construction of Bayesian nonparameric models

by

Junxi Zhang

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

Bayesian nonparametric models have gained increasing attention due to their flexibility in modelling natural and social phenomena and have been widely applied in machine learning, biology, social science and so on. Unlike traditional Bayesian parametric models, Bayesian nonparametric models place priors on an *infinite* dimensional space and allow the model itself to be determined by data. To understand and apply Bayesian nonparametric models, the properties, especially the asymptotic analysis, of the priors and posteriors of Bayesian nonparametric models should be studied. In this dissertation, various asymptotic problems for Bayesian nonparametric priors and posteriors are studied, and two dependent Bayesian nonparametric priors are constructed. This thesis includes three main parts corresponding to three papers. In the first part, we obtain the strong law of large numbers, Glivenko-Cantelli theorem, central limit theorem, functional central limit theorem for various Bayesian nonparametric priors which include the stick-breaking process with general iid stick-breaking weights, the two-parameter Poisson-Dirichlet process, the normalized inverse Gaussian process, the normalized generalized gamma process, and the generalized Dirichlet process. For the stick-breaking process with general iid stick-breaking weights, two general conditions are formulated such that the asymptotic theorems hold. In the second part, we present the posterior consistency analysis for normalized random measures with independent increments (NRMIs) through the corresponding Lévy intensities, which can be used to characterize the completely random measures in the construction of NRMIs. An assumption based on the Lévy intensities for analysing the posterior consistency of NRMIs is introduced and verified with multiple examples. Furthermore, we derive the Bernstein-von Mises theorem for the normalized generalized gamma process, based on which, credible intervals are constructed with some discussions and numerical illustration. In the third part, we construct two classes of dependent Bayesian nonparametric models through the normalization of completely random measures driven by Cox processes. We provide multiple distribution theories for the two constructions including

moments, probabilistic characterizations of the induced random partition structures by the hierarchical models, distributions of the random partition numbers.

Preface

The content in this thesis is based on the joint works with my supervisor Prof. Yaozhong Hu. These works include two published papers and two preprints.

Chapter 1 of this thesis includes the joint work with Prof. Yaozhong Hu that has been published as "Dirichlet process and Bayesian nonparametric models (in Chinese)" in *SCIENTIA SINICA Mathematica*, 2021, 51 (11), 1895–1932. This paper contains a supplementary material of 37 pages that is available online.

Chapter 2 of this thesis is a joint work with Prof. Yaozhong Hu that has been published as "Functional central limit theorems for stick-breaking priors" in *Bayesian Analysis*, 2022, 17 (4), 1101–1120.

Chapter 3 of this thesis is a joint work with Prof. Yaozhong Hu and the work is a complete preprint entitled "Large sample asymptotic analysis for normalized random measures with independent increments".

Chapter 4 of this thesis is a joint work with Prof. Yaozhong Hu and the work is a complete preprint entitled "Normalized random measures with independent increments driven by Cox process".

Junxi Zhang (PhD Candidate)

Acknowledgements

My graduate study has been an exciting and memorable experience in my life. I would like to take the opportunity to thank everyone who accompanied me and helped me during my PhD.

My utmost gratitude goes to my supervisor, Professor Yaozhong Hu, for both his extraordinary guidance in my research, continuous support and kind help in my daily life. Prof. Hu brought me to the word of research, he patiently taught me how to preform research from the beginning of my graduate program. We started our Bayesian nonparametric research from reading Ferguson's paper (Ferguson, 1973), and Prof. Hu patiently leaded me to understand the paper line by line. He taught me how to formulate problems, how to think problems, how to write clear proofs, how to write quality papers, and more. I benefit greatly from his experience, and without his help, nothing in this thesis would come true.

I would like to express my sincere gratitude to my examination committee members Professor Linglong Kong, Professor Bei Jiang, Professor Michael Y. Li, Professor Zhongwei Shen and Professor Shui Feng from McMaster University for their service and help. I want to express my sincere thanks to Prof. Linglong Kong, Prof. Bei Jiang and Prof. Shui Feng for their enthusiastic help and helpful discussion of my future research. I gratefully thank all the professors who taught me and all the supporting staff in the department of Mathematical and Statistical Sciences for their great assistance.

I also own many thanks to my friend Dr. Xiong Wang, who is a talented mathematician, for his accompanying and selfless help during my PhD. I want to thank Dr. Panqiu Xia for his accompanying and kind help during the time in the University of Kansas. Many thanks to all my friends, including but not limited to Jian Song, Le Chen, Fei Lu, Hongjuan Zhou, Jiayu Zheng, Yiying Cheng, Yanru Su, Jinjin Zhang, Peng Hu, Yuejuan Xi, Min Li, Chuying Huang, Shuo liu, Wenyi Pei, Yeli Niu, Weiwei Qi, Lei Ding, Meichen

Liu, Jiaqi Li, Hao Zhou, Qun Shi, Chunxiao Zhu, Shuhui Liu, Chao Mi, Haiyi Shi, Wen Liu, Shuyu Zhang, and many others.

Last but not least, my special thanks and forever love goes to my parents and my sister for their unconditional support and love.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Summary

Probabilistic models are widely used throughout statistics and machine learning to model distributions for observed data. People observe a sequence of data $\mathbf{X} = \{X_1, \cdots, X_n\}$, which is assumed to be generated from a certain unknown probabilistic model M. Parametric models assume the true probability distribution is determined and parametrized by fixed and finite many parameters $\theta \in \Theta$, where Θ is a finite dimensional parameter space. For example, a Gaussian model is determined by its mean and variance, which are the model parameters. Usually, the model parameters for M are estimated by using the observations to make the model suitable for the observed data. The estimation of the model parameters can be obtained by two popular ways. One way is to assume that θ is a deterministic value and its value is obtained, foe example, by maximizing the likelihood function $f(\mathbf{X}|M) = f(\mathbf{X}|\theta)$, the maximizer $\hat{\theta}_n$ is called the maximum likelihood estimator (MLE) of θ. The other way is to assume that the parameter θ is random and to assign a prior distribution $\pi(\theta)$ for the parameter. Then, the estimation of θ can be found by maximizing the posteriori (MAP) via point estimation or by using Bayesian inference. The MAP estimation is not flexible, since the estimation is mostly determined by the optimization information of the posterior distribution. Thus, Bayesian inference is usually preferable as it produces the parameter estimation by using the entire information of the posterior distribution. In Bayesian inference, the estimator of θ is the expectation of the posterior distribution that is formulated by the celebrating Bayesian rule:

$$f(\theta|\mathbf{X}) = \frac{f(\theta, \mathbf{X})}{f(\mathbf{X})} = \frac{f(\theta)f(\mathbf{X}|\theta)}{\int_\Theta f(\theta)f(\mathbf{X}|\theta)d\theta}.$$

In traditional Bayesian inference, although the model M is random, it is still assumed to be parametrized by a fixed and finite number of parameters. That is to say, the parameter space Θ is finite dimensional, and the model is known as *Bayesian parametric model*. However, the methods that are mentioned above can suffer from over-fitting and under-fitting problems when there is a misfit between the model complexity (usually expressed by the dimension of the parameter space) and the data size. As a result, model selection and determining the suitable model complexity become important problems in these methods. Fortunately, *Bayesian nonparametric models* allow the model M to be parametrized by infinite number of parameters, i.e., Θ is infinite dimensional, and the model complexity could grow (determined by the data) with the sample size. Thus, both the under-fitting and over-fitting issues are mitigated. In this thesis, Bayesian nonparametric setting is considered.

To understand Bayesian nonparametric models in detail, we let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space, let $\mathbb{X}$ be a complete, separable metric space whose σ-algebra is denoted by $\mathcal{X}$ and let $(\mathbb{M}_\mathbb{X}, \mathcal{M}_\mathbb{X})$ be the space of all probability measures on $\mathbb{X}$. We first recall the definition of random measure on $(\mathbb{X}, \mathcal{X})$.

Definition 1.0.1. *A random measure is a mapping P from $\Omega \times \mathcal{X}$ to $\mathbb{R}_+$ (we denote this random measure by $P = (P(\omega, A),\, \omega \in \Omega\,, A \in \mathcal{X})$) such that*

(i) when $\omega \in \Omega$ is fixed, $P(\omega, \cdot)$ is a measure on $(\mathbb{X}, \mathcal{X})$;

(ii) when $A \in \mathcal{X}$ is fixed, $P(\cdot, A)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

A sample $\mathbf{X} = (X_1, \cdots, X_n)$ that takes values in $\mathbb{X}^n$ is drawn iid from a random probability measure (see 2.7.1 for details of random probability measures) P conditional on P. A Bayesian approach in this setting is to place a prior distribution Q, on $(\mathbb{M}_\mathbb{X}, \mathcal{M}_\mathbb{X})$, for the random probability measure P. That is to say,

$$X_1, \cdots, X_n | P \overset{iid}{\sim} P; \qquad P \sim Q. \tag{1.0.1}$$

Thus, one fundamental question is to determine the prior distribution Q. Addressing this question, (Ferguson, 1973) suggests that the prior distribution should have large enough support and the posterior distribution should be manageable analytically. A large amount of priors along this line have been proposed, among which, the most popular ones are the Dirichlet process (Ferguson, 1973), the two-parameter Poisson-Dirichlet process (also known as Pitman-Yor process (Pitman and Yor, 1997)), the σ-stable process (Kingman, 1975), the normalized inverse Gaussian process (Lijoi et al., 2005b), the normalized generalized gamma process (Lijoi et al., 2003, 2007), and generalized Dirichlet process (Lijoi et al., 2005a). For the applications of these Bayesian nonparametric models, including mixture models and hierarchical models, we refer interested readers to (Müller and Quintana, 2004; Lijoi et al., 2010; Zhang and Hu, 2021; Ghosal and Van der Vaart, 2017) and the references therein.

In this thesis, we mainly consider two important subclasses of Bayesian nonparametric priors: *stick-breaking process* and *normalized random measures with independent increments (NRMIs)*. All the previously mentioned processes are included in stick-breaking process and, all except the two-parameter Poisson-Dirichlet process are included in NRMIs. The topics covered in this thesis are as follows: (i) the asymptotic behaviour of the stick-breaking process when its concentration parameter (more details can be found in 2.1) $a \to \infty$; (ii) the posterior consistency analysis of NRMIs and the Bernstein-von Mises theorem for the normalized generalized gamma process when the sample size $n \to \infty$; (iii) the constructions of two classes of dependent NRMIs by using Cox process and the corresponding hierarchical structures. The results of the first two topics provide the theoretical support for constructing Bayesian credible intervals, simplifying Bayesian statistics, approximating these processes in the limiting sense. The last topic provides two flexible dependent Bayesian nonparametric models that can be used for partially exchangeable data.

Before going into the details of these topics, we briefly sketch the two subclasses of Bayesian nonparametric models.

1.1 Stick-breaking process

Let H be a nonatomic probability measure on $(\mathbb{X}, \mathcal{X})$ (i.e., $H(\{x\}) = 0$ for any $x \in \mathbb{X}$).

Definition 1.1.1. *A random measure $P = (P(\omega, A), \omega \in \Omega, A \in \mathcal{X})$ is said to be a stick-breaking process with the base measure H, if it has the following representation:*

$$\begin{cases} P = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}, & \textit{where} \\ w_1 = v_1, \quad w_i = v_i \prod_{j=1}^{i-1} (1 - v_j) & \textit{for} \quad i = 2, 3, \cdots, \end{cases}$$

where $\theta_i, i = 1, 2, \cdots$ are iid random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(\mathbb{X}, \mathcal{X})$ such that for each i, the law of θ_i is H; δ_{θ_i} denotes the Dirac measure on $(\mathbb{X}, \mathcal{X})$, and $v_i, i = 1, 2, \cdots$ are random variables with values in $[0, 1]$, independent of $\{\theta_i\}$, which are called the stick-breaking weights.

An illustration of the stick-breaking process is given in figure 1.1. The law of $(w_1, w_2, \cdots)$ is called the GEM distribution (details can be found in e.g., (Feng, 2010; Ewens, 2004)), named for the contributions of Griffiths (Griffiths, 1980), Engen (Engen, 1978) and McCloskey (McCloskey, 1965). One breakthrough in this topic is made by (Sethuraman, 1994) who shows that the Dirichlet process admits the stick-breaking representation, where the stick-breaking weights are iid Beta random variables, i.e., $v_i \overset{iid}{\sim} \text{Beta}(1, a)$ (throughout this thesis the notation $\text{Beta}(\alpha, \beta)$ denotes the Beta distribution whose density is $g(x; \alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \ 0 < x < 1$). (Perman et al., 1992) obtains the stick-breaking representation for the two-parameter Poisson-Dirichlet process and shows the stick-breaking weights $v_i \overset{ind}{\sim} \text{Beta}(1 - b, a + ib)$ with $b > 0$, $a > -b$. For the stick-breaking representations of other special stick-breaking processes, we include the details

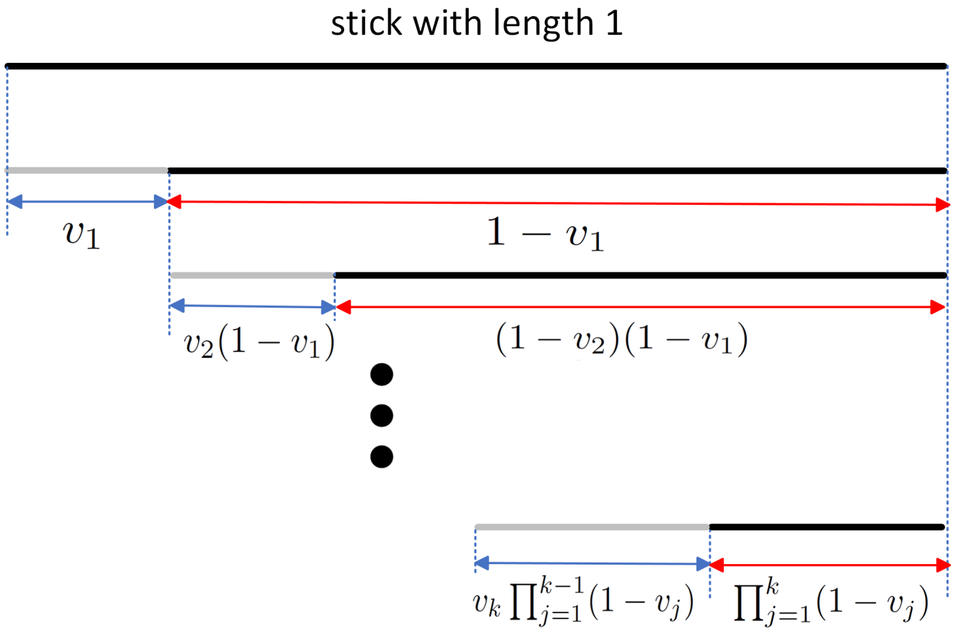


Figure 1.1: Illustration of the stick-breaking process.

in 2.7.1. Based on the definition of the stick-breaking process, it shows benefits in the computational aspect as the infinite summation can be truncated with a tolerable error (Ghosal and Van der Vaart, 2017). However, it is worth pointing out that the infinite summation is not analytical friendly.

1.2 Normalized random measures with independent increments

We start by recalling the definitions of completely random measures (see e.g., (Kingman, 1967, 1993) and references therein for more details) and Poisson random measure, which play important roles in the construction of NRMIs.

Definition 1.2.1. *Let μ be a finite random measure on $(\mathbb{X}, \mathcal{X})$. We call μ a completely random measure (CRM) if the random variables $\mu(A_1), \cdots, \mu(A_d)$ are mutually independent, for any pairwise disjoint sets $A_1, \cdots, A_d$, where $d \geq 2$ is a finite integer.*

Definition 1.2.2. *Let $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$ and denote its Borel σ-algebra by $\mathcal{S}$. A Poisson random* *measure* $\tilde{N}$ *on* $\mathbb{S}$ *with finite intensity measure* $\nu(ds, dx)$ *is a random measure from* $\Omega \times \mathbb{S}$ *to* $\mathbb{R}_+$ *satisfying*

(i) $\tilde{N}(A) \sim Poisson(\nu(A))$ *for any* A *in* $\mathcal{S}$*;*

(ii) for any pairwise disjoint sets $A_1, \cdots, A_m$ *in* $\mathcal{S}$*, the random variables* $\tilde{N}(A_1), \cdots, \tilde{N}(A_m)$ *are mutually independent.*

The Poisson intensity measure ν satisfies the condition (see (Daley and Vere-Jones, 2008) for details of Poisson random measures) that

$$\int_0^\infty \int_{\mathbb{X}} \min(s, 1)\nu(ds, dx) < \infty \,.$$

Let $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$ be the space of bounded finite measures on $(\mathbb{X}, \mathcal{X})$ endowed with the topology of weak convergence and let $\tilde{\mu}$ be the random measure defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$ defined as follows,

$$\tilde{\mu}(A) := \int_0^\infty \int_A s\tilde{N}(ds, dx), \quad \forall A \in \mathcal{X} \,. \tag{1.2.1}$$

It is trivial to verify that $\tilde{\mu}$ is a completely random measure. It is also well-known that for any $B \in \mathcal{X}$, $\tilde{\mu}(B)$ is discrete and is uniquely characterized by its Laplace transform as follows:

$$\mathbb{E}\left[e^{-\lambda \tilde{\mu}(B)}\right] = \exp\left\{-\int_0^\infty \int_B \left[1 - e^{-\lambda s}\right] \nu(ds, dx)\right\} \,. \tag{1.2.2}$$

The measure ν is called the *Lévy intensity* of $\tilde{\mu}$ and we denote the Laplace exponent by

$$\psi_B(\lambda) = \int_0^\infty \int_B \left[1 - e^{-\lambda s}\right] \nu(ds, dx) \,. \tag{1.2.3}$$

From the Laplace transform in (1.2.2), we shall study the completely random measure $\tilde{\mu}$ by its Lévy intensity ν, which usually takes the following forms in the literature:

(a) $\nu(ds, dx) = \rho(ds)\alpha(dx)$, where $\rho : \mathcal{B}(\mathbb{R}^+) \to \mathbb{R}^+$ is some measure on $\mathbb{R}^+$ and α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ so that $\alpha(\mathbb{X}) = a < \infty$. The corresponding $\tilde{\mu}$ is called *homogeneous* completely random measure.

(b) $\nu(ds, dx) = \rho(ds|x)\alpha(dx)$, where ρ is defined on $\mathcal{B}(\mathbb{R}^+) \times \mathbb{X}$ such that for any $x \in \mathbb{X}$, $\rho(\cdot|x)$ is a σ-finite measure on $\mathcal{B}(\mathbb{R}^+)$ and for any $A \in \mathcal{X}$, $\rho(A|x)$ is $\mathcal{B}(\mathbb{R}^+)$ measurable. The corresponding $\tilde{\mu}$ is called *non-homogeneous* completely random measure.

It is obvious that case (a) is a special case of case (b). Usually, we assume that α is a finite measure so we may write $\alpha(dx) = aH(dx)$ for some probability measure H and some constant $a = \alpha(\mathbb{X}) \in (0, \infty)$.

To construct NRMIs, the completely random measure will be normalized, and thus one needs the total mass $\tilde{\mu}(\mathbb{X})$ to be finite and positive almost surely. This happens under the condition that $\rho(\mathbb{R}^+) = \infty$ in homogeneous case and that $\rho(\mathbb{R}^+|x) = \infty$ in non-homogeneous case (Regazzini et al., 2002). Under the above conditions, an NRMI P on $(\mathbb{X}, \mathcal{X})$ is a random probability measure defined by

$$P(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}. \tag{1.2.4}$$

Based on the construction of NRMIs, its posterior distribution can be obtained analytically (see (James et al., 2009) for the details of the posterior analysis).

1.3 Summary of this thesis

This thesis is a collection of joint works with my supervisor. It consists of the following four papers.

1. (Zhang and Hu, 2021) Dirichlet process and Bayesian nonparametric models (in chinese), *with Yaozhong Hu, SCIENTIA SINICA Mathematica (2021) 51 (11), 1895-1932;*

2. (Hu and Zhang, 2022) Functional central limit theorems for stick-breaking priors, *with Yaozhong Hu, Bayesian Analysis (2022) 17 (4), 1101-1120;* Along with the supplementary material of this paper;

3. Large sample asymptotic analysis for normalized random measures with independent increments, *joint work with Yaozhong Hu, Preprint.*

4. Normalized random measures with independent increments driven by Cox process, *joint work with Yaozhong Hu, Preprint.*

In chapter 2, we study the asymptotic behaviours for various Bayesian nonparametric priors when their concentration parameters $a \to \infty$. We obtain the strong law of large numbers, Glivenko-Cantelli theorem, central limit theorem, functional central limit theorem for the *stick-breaking process with general iid stick-breaking weights*, the Dirichlet process, the two-parameter Poisson-Dirichlet process, the normalized inverse Gaussian process, the normalized generalized gamma process, and the generalized Dirichlet process. The stick-breaking process with general iid stick-breaking weights is introduced as the general stick-breaking process when the stick-breaking weight v_i's are iid, we deduce two general conditions on the stick-breaking weights such that the central limit theorem and functional central limit theorem hold for this general stick-breaking process. These asymptotic results seem true intuitively since a is an important parameter to send the variances of these probability random measures to 0. However, people haven't found way to show them, since the finite dimensional distributions of these processes are either hard to obtain or are complicated to use even when they are available. We derive the asymptotic theorems for the above mentioned processes except the generalized Dirichlet process by the method of moments (see 2.1) and their stick-breaking representations. Thus, the key to this work is to find the moment results (see 2.3) by using the stick-breaking representations of these processes. And the main results are proved by the moment results and some combinatorial analysis. In the case of the generalized Dirichlet process, as its marginal density is available, we derive the asymptotic results by showing the convergence of its marginal density to the normal density, providing an alternative way of achieving the asymptotic theorems. The numerical illustration shows that the convergences are fast.

In chapter 3, we study the posterior asymptotic behaviours for NRMIs when the sample $\mathbf{X} \overset{iid}{\sim} P_0$ and when the sample size $n \to \infty$. Here, P_0 is assumed to be the true distribution of $\mathbf{X}$. We first provide the posterior consistency analysis for NRMIs through the corresponding Lévy intensities, which can be used to characterize the completely random measures in the construction of NRMIs. An assumption based on the Lévy intensities for analysing the posterior consistency of NRMIs is formulated. To show the applicability of the proposed assumption, we verify it with multiple known processes including the normalized generalized gamma process (NGGP), the generalized Dirichlet process, the extended gamma process. Our results show that the posterior consistency holds for the Dirichlet process, the generalized Dirichlet process, and the extended gamma process for any P_0 when the Lévy intensity is gamma type. However, for the general NRMIs, the posterior consistency holds when P_0 is discrete or when P_0 is continuous with $\bar{C}_1 = 0$ (see e.g., assumption 3.3.2). The posterior consistency results suggest that one should avoid using NRMIs when the true distribution is continuous and $\bar{C}_1 \neq 0$. Furthermore, we derive the Bernstein-von Mises theorem for the NGGP, which is a flexible sub-subclass of NRMIs. From the posterior consistency result, the NGGP is posterior consistent when P_0 is discrete or when the model parameter $\sigma = 0$. However, the NGGP is reduced to the Dirichlet process with the latter case ($\sigma = 0$), and thus only the former case is studied. It is interesting to note that there exists a bias term in the Bernstein-von Mises results that is closely related to the number of atoms of P_0 when P_0 is discrete. Therefore, a bias correction is necessary when constructing credible intervals by using the Bernstein-von Mises theorem. We illustrate how the bias correction affects the coverage of the true value by the credible intervals when P_0 is discrete with different types of atoms by using numerical experiments. We also discuss the affect of the estimators for the model parameters of the NGGP under the Bernstein-von Mises convergences.

In chapter 4, we construct two forms of dependent normalized random measures with independent increments through the normalization of completely random measures that is constructed through Cox process. Bayesian nonparameteric models are popular not only for their flexibility, but also for their property of generating exchangeable samples. Hierarchical Bayesian nonparametric models extend the exchangeable assumption of the sample to partially exchangeable, thus they are widely used for two important properties: (i) they naturally represent multiple heterogeneous sub-populations; (ii) there is a tie across all sub-populations to represent the dependence across sub-populations. Instead of using the Poisson random measure to construct NRMIs in the hierarchical NRMIs models, we suggest a more flexible approach using Cox random measures to construct NRMIs. We derive two forms along this direction, one is conditionally independent NRMIs driven by Cox process, and the other is conditionally dependent NRMIs driven by Cox process. In our constructions, two vectors of dependent NRMIs $(P_1, \cdots, P_d)$ are generated such that the dependence between P_i and P_j with $i \neq j$ in each construction is determined by both the hierarchical structure and multiple user-controlled tuning parameters. Based on the two constructions, we first derive the moment results that include the variance and covariance between P_i and P_j. Secondly, we provide probabilistic characterizations of the induced partially exchangeable random partition structures in the hierarchical models, including the partially exchangeable partition probability functions, distribution of the number of partition sets. To give a clear interpretation, the random partition structure of the hierarchical models is explained by *local special Chinese restaurant franchise* as introduced in 4.5.

Chapter 2

Functional central limit theorems for stick-breaking priors

2.1 Introduction

Ever since the work of (Ferguson, 1973) the Dirichlet process has become a critical tool in Bayesian nonparametric statistics and has found applications in various areas, including machine learning, biological science, social science and so on. One of the important features of the Dirichlet process is that when the prior is a Dirichlet process its posterior is also a Dirichlet process (see e.g., (Ferguson, 1973)). This makes the complex computation in the Bayesian nonparametric analysis possible and enables the Dirichlet process to become a backbone of the Bayesian nonparametric statistics.

To widen the applicability of the Bayesian nonparametric statistics, researchers have tried to extend the concept of Dirichlet process. One of these efforts is the introduction of the stick-breaking process. The first breakthrough along this path is due to (Sethuraman, 1994) who shows that the Dirichlet process admits the stick-breaking representation (see (2.2.1)-(2.2.2) in the next section), where the stick-breaking weights are independent and identically distributed (iid) random variables satisfying the Beta distribution $\text{Beta}(1, a)$. Within this stick-breaking representation, we can extend the class of Dirichlet processes to many other priors by assuming that the stick-breaking weights are iid with other distributions; satisfy some other kinds of dependence; or satisfy some specific (joint) distributions. Among various such extensions, let us mention the following works which we shall deal with in this chapter. (Perman et al., 1992) obtain a general formulae for sized-biased sampling from a Poisson point process where the size of a point is defined by an arbitrary strictly positive function. From this formulae, they identify the stick-breaking representation of the two-parameter Poisson-Dirichlet process, which admits a stick-breaking process with the stick-breaking weights $v_i \overset{ind}{\sim} \text{Beta}(1-b, a+ib)$, where $b > 0$, $a > -b$ and $i = 1, 2, \cdots$. (Favaro et al., 2012) introduce the normalized inverse Gaussian process through its stick-breaking representation by identifying the explicit finite dimensional joint density functions of its stick-breaking weights. (Favaro et al., 2016) present the stick-breaking representation of homogeneous normalized random measures with independent increments (hNRMIs) (see e.g., (Regazzini et al., 2003) for more details of NRMIs), which include the normalized generalized gamma process and the generalized Dirichlet process, two widely used priors in Bayesian nonparametric statistics.

Strong law of large numbers, central limit theorem and functional central limit theorem have always been ones of the central topics in statistics and in probability theory. Without exception the asymptotic behaviors of the Dirichlet process and other Bayesian nonparametric priors play important roles in the Bayesian nonparametric analysis, for example in the construction of asymptotic Bayesian confidence intervals, regression analysis and functional estimations. Compared to the vast literature in the field of parametric statistics relevant to these issues the achievements in the field of Bayesian nonparametrics are quite limited. However, let us mention the following works pioneered this work. (Sethuraman and Tiwari, 1982) discuss the weak convergences of the Dirichlet measure P when its parameter measure (i.e the measure aH in this chapter) approaches to a non-zero measure or a zero measure respectively. (Lo, 1983) studies the central limit theorem of the posterior distribution of Dirichlet process which is analogous to our central limit theorem for the Dirichlet process. Based on this result, (Lo, 1987) obtains the asymptotic confidence bounds and establishes the asymptotic validity of the Bayesian bootstrap method. The above mentioned Lo's results are extended to the mixtures of Dirichlet process by (Brunner and Lo, 1996). (James, 2008) reveals the consistency behavior (the posterior distribution converges to the true distribution weakly) and the functional central limit theorem for the posterior distribution of the two-parameter Poisson-Dirichlet process (with fixed a and when the sample size goes to infinity). The consistency of the posterior is discussed by (Ho Jang et al., 2010) when the priors are the two-parameter Poisson-Dirichlet prior and the species sampling prior. Furthermore, (De Blasi et al., 2013) investigate the consistency of the Gibbs-type priors. (Kim and Lee, 2004) show that the Bernstein-von Mises theorem holds in survival models for the Dirichlet process, Beta process and Gamma process. (Dawson and Feng, 2006) establish the large deviation principle for the Poisson-Dirichlet distribution and give the explicit rate functions when the parameter a (which represents the mutation rate in the context of population genetics) approaches infinity. (Labadi and Zarepour, 2013) present the functional central limit theorem for the normalized inverse Gaussian process on $D(\mathbb{R})$ when its parameter a is large by using its finite dimensional joint density. (Labadi and Abdelrazeq, 2016) obtain the functional central limit theorem for the Dirichlet process by using the finite dimensional densities and for the Beta process on $D(\mathbb{R})$ by using the characteristic function.

From the above mentioned works we see that there are only very limited results on the asymptotics of the stick-breaking processes. Relevant to the asymptotics as $a \to \infty$, there have been established the central limit theorem and functional central limit theorem only for two processes: the Dirichlet process and the normalized inverse Gaussian process. The reason for the above limitation is that the most commonly used technique appeals to the explicit forms of the finite dimensional densities of the process itself. This method is effective only when the finite dimensional distributions have explicit forms and are possible to handle. It cannot be applied to study other processes when the explicit forms for the finite dimensional marginal densities of the process itself are unavailable or they are too complex to analyze even though they are available.

This chapter is to introduce the method of moments into this study and to provide a systematic study of the asymptotics as $a \to \infty$ for various stick-breaking processes depending on a parameter $a > 0$. Let us emphasize that the method of moments in this chapter refers to the fact that if the distribution of the random variable X is determined by its moments, and the random variables $\{X_i\}_{i=1}^n$ have all moments, and if $\lim_{n\to\infty} \mathbb{E}\left[X_n^r\right] = \mathbb{E}\left[X^r\right]$ for $r = 1, 2, \cdots$, then $X_n \xrightarrow{d} X$ (see e.g., (Billingsley, 1995, Theorem 30.2)). We are mainly concerned with three types of the asymptotics (strong law of large numbers, central limit theorem, and functional central limit theorem) for a number of processes, which include the stick-breaking process with general stick-breaking weights, the classical Dirichlet process $\mathrm{DP}(a, H)$ (see (Ferguson, 1973)), the two-parameter Poisson-Dirichlet process $\mathrm{PDP}(a, b, H)$ (also known as Pitman-Yor process (Pitman and Yor, 1997)), the normalized inverse Gaussian process $\text{N-IG}(a, H)$ (Lijoi et al., 2005b), the normalized generalized gamma process $\mathrm{NGGP}(\sigma, a, H)$ (see (Brix, 1999; Lijoi et al., 2007, 2003)), and the generalized Dirichlet process $\mathrm{GDP}(a, r, H)$ (see (Lijoi et al., 2005a)).

All of the mentioned processes depend on a parameter a which is usually called the *concentration parameter*. It is of the same order as the inverse of the variance of the process (see Remark 2.3.6 for more precise meaning). It has also some more specific meanings for various processes. For example, if $\{X_i\}_{i=1}^n$ is a sample from the Dirichlet process $\mathrm{DP}(a, H)$, then it is known that the posterior mean is $\mathbb{E}\left[P(\cdot)|X_1, \cdots, X_n\right] = \frac{a}{a+n} H(\cdot) + \frac{n}{a+n} \frac{\sum_{i=1}^n X_i}{n}$, which means that a plays the key role of the weight of the prior.

For the generalized Dirichlet process since the finite dimensional marginal distributions of the process itself are available we shall use them to obtain the asymptotics directly although the computations are very technical. Let us point out that this process also admits a stick-breaking representation. However, it seems to us that it is more complex to use the method of moments than to use the finite dimensional marginal distributions of the process itself.

Let us stress the following points of the work about the well-known Bayesian nonparametric priors.

(1) (for Dirichlet process) Both the finite dimensional distributions of the stick-breaking weights and the process itself are explicit and are easy to handle. Prior to this work the central limit theorem and the functional central limit theorem have been

established for this process by using the finite dimensional distribution of the process itself.

For the Dirichlet process the stick-breaking weights $\{v_i\}$ are iid and follow the Beta distribution Beta$(1, \alpha)$. We introduce the concept of stick-breaking process with general stick-breaking weights, where we still require the stick-breaking weights $\{v_i\}$ to be iid but the law μ they follow can be arbitrary. In this case there is no way to obtain the explicit form of the joint distributions of the process itself. We use the method of moments to establish the central limit theorem and the functional central limit theorem for this process. For example, $v_i \sim$ Beta(ρ_a, a), where ρ_a is a function of a such that $\rho_a/a \to 0$ as $a \to \infty$. In this case the joint distributions of the process itself is unavailable except in the case $\rho_a = 1$, i.e., in the case of the Dirichlet process.

(ii) (for the normalized inverse Gaussian process and for the generalized Dirichlet process) Both the finite dimensional distributions of the stick-breaking weights and that of the process itself are explicit. Prior to this work the central limit theorem and the functional central limit theorem have been established only for the normalized inverse Gaussian process by using the finite dimensional distributions of the process itself. We shall also use the finite dimensional distributions of the process itself to obtain the central limit theorem and the functional central limit theorem for the generalized Dirichlet process. We shall use the method of moments to re-derive the central limit theorem and the functional central limit theorem for the normalized inverse Gaussian process, providing an alternative tool for this process.

(iii) (for the two-parameter Poisson-Dirichlet process and the normalized generalized gamma process) The finite dimensional distributions of the stick-breaking weights are known but the finite dimensional distributions of the process itself are not available. We use the method of moments to obtain the central limit theorem and the functional central limit theorem for these processes.

Now we explain the organization of this chapter. In Section 2.2, we recall the general stick-breaking process and introduce the stick-breaking process with general stick-breaking weights (SPG(μ, H)). In Section 2.3, we present the moment results for various stick-breaking processes, including SPG(μ, H), PDP(a, b, H), N-IG(a, H), and NGGP(σ, a, H), GDP(a, r, H) separately since the computations are different for different processes. In Section 2.4, we state the strong law of large numbers, central limit theorem, and functional central limit theorem. The stick-breaking process with general stick-breaking weights are new and we allow the stick-breaking weights to be some very general iid random variables defined on $(0, 1)$. With different choices of the stick-breaking weights we can obtain various known stick-breaking processes. Because of this generality of the stick-breaking weights we state one theorem on the central limit theorem and functional central limit theorem for this type of processes. We state a similar theorem for all other processes (PDP(a, b, H), N-IG(a, H), NGGP(σ, a, H), GDP(a, r, H)). The details of the proofs will be provided in a supplementary file where we also include some definitions and some well-known propositions of the mentioned processes to provide the necessary background. Interested readers are referred to (Zhang and Hu, 2021) and references therein for a recent survey of some of these processes and their applications.

Finally, let us emphasize that all the processes we dealt with in this chapter are actually "*random probability measures*". However, we follow the convention in the literature to continue to call them "*processes*".

2.2 Preliminary Notations

2.2.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $(\mathbb{X}, \mathcal{X})$ be a measurable Polish space, namely, $\mathbb{X}$ is a separable complete metric space and $\mathcal{X}$ is the Borel σ-algebra of $\mathbb{X}$. Let H be a nonatomic probability measure on $(\mathbb{X}, \mathcal{X})$ (i.e., $H(\{x\}) = 0$ for any $x \in \mathbb{X}$). Now we give the definition of the stick-breaking process (more appropriately a stick-breaking random probability measure).

Definition 2.2.1. *A random measure $P = (P(\omega, A), \omega \in \Omega, A \in \mathcal{X})$ is said to be a stick-breaking process with the base measure H, if it has the following representation:*

$$\begin{cases} P = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}, \qquad \text{where} & (2.2.1) \\ w_1 = v_1, \quad w_i = v_i \prod_{j=1}^{i-1} (1 - v_j) \quad \text{for} \quad i = 2, 3, \cdots, & (2.2.2) \end{cases}$$

where $\theta_i, i = 1, 2, \cdots$ are iid random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(\mathbb{X}, \mathcal{X})$ such that for each i, the law of θ_i is H; δ_{θ_i} denotes the Dirac measure on $(\mathbb{X}, \mathcal{X})$, and $v_i, i = 1, 2, \cdots$ are random variables with values in $[0, 1]$, independent of $\{\theta_i\}$, which are called the stick-breaking weights.

Since we assume that $\{\theta_i\}$ are iid and follow the distribution H, if H is given and fixed, then the random probability measure P depends only on the choice of $\{v_i\}$.

Remark 2.2.2. *To make sure that P is well-defined (namely, (2.2.1) is convergent), one needs to impose the condition that $\sum_{i=1}^{\infty} w_i = 1$ almost surely, which is equivalent to the condition that $\sum_{i=1}^{\infty} \log \mathbb{E}\left[(1 - v_i)\right] = -\infty$ (e.g., (Ghosal and Van der Vaart, 2017, Lemma 3.4)).*

Remark 2.2.3. *Throughout the entire chapter, we shall assume that a is a positive real number and H is a nonatomic measure on $(\mathbb{X}, \mathcal{X})$ unless otherwise specified.*

For potential applications in practice we introduce the concept of *stick-breaking process with general stick-breaking weights.*

Definition 2.2.4. *P is called the stick-breaking process with general stick-breaking weights, denoted by $P \sim \mathrm{SPG}(\mu, H)$, if the stick-breaking weights $\{v_1, v_2, \cdots\}$ in (2.2.1)-(2.2.2) are iid and follow a general distribution μ.*

Remark 2.2.5. *The law μ on $(0, 1)$ can be of continuous or discrete types, or the mixture. An interesting special example is the quasi Bernoulli stick-breaking process (Zeng and Duan, 2020), where the $v_i \sim g(x) = pf(x) + \frac{1-p}{\varepsilon} f(x/\varepsilon)$ for the Bernoulli density $f(x) \sim$ $\mathrm{Beta}(1, a)$ and for some $p \in (0, 1), \varepsilon > 0$.*

Based on the expectation and variance of P, we introduce the following quantities that are investigated in the main theorems:

$$D_a(\cdot) = \frac{P(\cdot) - \mathbb{E}[P(\cdot)]}{\sqrt{\mathrm{Var}[P(\cdot)]}} = \frac{P(\cdot) - H(\cdot)}{\sqrt{H(A)(1-H(A))\mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\right]}}, \tag{2.2.3}$$

where the last identity follows from (7.12)-(7.13) (in the supplementary material). Up to a constant we may just consider the following quantity for notational simplicity:

$$Q_{H,a}(\cdot) = \frac{P(\cdot) - \mathbb{E}[P(\cdot)]}{\sqrt{\mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\right]}}. \tag{2.2.4}$$

2.3 Moment results

We use the method of moments to show the announced asymptotics. This requires to have some nice estimates of the moments of the random probability measure P, which in turn requires some nice bounds for the moments of $\{w_i\}_{i=1}^{\infty}$. Thus, in this section we present the asymptotic behaviors of the joint moments of w_i's for various processes introduced in the introduction. These results will play the key roles in the proofs of our main theorems. On the other hand, they also have their own interest.

In the following proposition and throughout the chapter we use the notation $p_{m:n} := \sum_{i=m}^{n} p_i$ for $m \leq n$, and let the sequence $\{w_i\}_{i=1}^{\infty}$ be defined as in (2.2.2).

Proposition 2.3.1. *Let $P \sim \mathrm{SPG}(\mu_a, H)$, i.e., the law of the i.i.d stick-breaking weights v_i is μ_a, where $a > 0$ is a certain parameter. We assume that v_i is not identically 0. If $\lim_{a\to\infty} \frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]} = 0$ for all $n \in \mathbb{Z}_+$ (set of nonnegative integers), then for any nonnegative integers m, n,*

$$\begin{cases} \mathbb{E}\left[v_i^n(1-v_i)^m\right] = \mathbb{E}[v_1^n] + o\left(\mathbb{E}[v_1^n]\right), & (2.3.1) \\ \displaystyle\sum_{j=0}^{\infty} \left(\mathbb{E}\left[(1-v_i)^m\right]\right)^j = \frac{1}{m\mathbb{E}[v_1]} + o\left(\frac{1}{m\mathbb{E}[v_1]}\right). & (2.3.2) \end{cases}$$

Furthermore, for any positive integers $p_1, \cdots, p_k$*, we have*

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right] = \frac{\mathbb{E}[v_1^{p_1}]\cdots\mathbb{E}[v_1^{p_k}]}{p_{1:k}p_{2:k}\cdots p_{k:k}\left(\mathbb{E}[v_1]\right)^k} + o\left(\frac{\mathbb{E}[v_1^{p_1}]\cdots\mathbb{E}[v_1^{p_k}]}{\left(\mathbb{E}[v_1]\right)^k}\right). \tag{2.3.3}$$

In particular, when $p_j = 2$ *for all* $j \in \{1, \cdots, k\}$ *(hence* $p_{1:k} = 2k$*), the asymptotics* (2.3.3) *becomes*

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{2} w_{i_2}^{2}\cdots w_{i_k}^{2}\right] = \frac{1}{2^k k!}\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^k + o\left(\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^k\right). \tag{2.3.4}$$

Proposition 2.3.2. *Let* $P \sim \mathrm{PDP}(a, b, H)$*. Namely, let the stick-breaking weights* $v_1, v_2, \cdots$ *be given by* (7.2) *(in the supplementary material). Then, for any positive integers* $p_1, \cdots, p_k$*, we have the following identity.*

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right] = \frac{1}{(a+kb)(a+1)_{(p_{1:k}-1)}}\prod_{i=1}^{k}\frac{(1-b)_{p_i}(a+bi)}{p_{i:k}-(k-i+1)b}. \tag{2.3.5}$$

In particular, when $p_j = 2$ *for all* $j \in \{1, \cdots, k\}$*, the above expectation becomes*

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{2} w_{i_2}^{2}\cdots w_{i_k}^{2}\right] = \frac{(1-b)^k(a+b)\cdots(a+b(k-1))}{k!(a+1)\cdots(a+2k-1)}. \tag{2.3.6}$$

Proposition 2.3.3. *Let* $P \sim \text{N-IG}(a, H)$*. Namely, let the stick-breaking weights* $\{v_i\}_{i=1}^{\infty}$ *be given by* (7.3)-(7.4) *(in the supplementary material). Then, for any positive integers* $p, p_1, \cdots, p_k$*, we have*

$$\mathbb{E}\left[\sum_{n=1}^{\infty} w_n^p\right] = O\left(\frac{1}{a^{p-1}}\right) \quad \text{as } a\to\infty, \tag{2.3.7}$$

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right] = O\left(\frac{1}{a^{p_{1:k}-k}}\right) \quad \text{as } a\to\infty. \tag{2.3.8}$$

Furthermore, when $p = p_1 = \cdots = p_k = 2$, we have

$$\mathbb{E}\left[\sum_{n=1}^{\infty} w_n^2\right] = \frac{1}{a} + o\left(\frac{1}{a}\right) \quad \text{as } n \to \infty\,, \tag{2.3.9}$$

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^2 w_{i_2}^2 \cdots w_{i_k}^2\right] = \frac{1}{k!a^k} + o\left(\frac{1}{a^k}\right)\,. \tag{2.3.10}$$

[namely, the leading coefficient in (2.3.7) *is* 1 *and the leading coefficient in* (2.3.8) *is* $\frac{1}{k!}$*.]*

Proposition 2.3.4. *Let $P \sim \mathrm{NGGP}(\sigma, a, H)$. Namely, let the distribution of the stick-breaking weights $\{v_1, v_2, \cdots\}$ be given by* (7.5)-(7.6) *(in the supplementary material). Then, for any positive integers $p_1, \cdots, p_k$, we have*

$$\mathbb{E}\left[\sum_{n=1}^{\infty} w_n^p\right] = O\left(\frac{1}{a^{p-1}}\right) \quad \text{as } a \to \infty, \tag{2.3.11}$$

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \cdots w_{i_k}^{p_k}\right] = O\left(\frac{1}{a^{p_{1:k}-k}}\right) \quad \text{as } a \to \infty. \tag{2.3.12}$$

Furthermore, when $p = p_1 = \cdots = p_k = 2$ and when $\sigma = \frac{1}{m}$ for some arbitrarily fixed integer $m \ge 2$, we have

$$\mathbb{E}\left[\sum_{n=1}^{\infty} w_n^2\right] = \frac{1}{a} + o\left(\frac{1}{a}\right) \quad \text{as } a \to \infty, \tag{2.3.13}$$

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^2 w_{i_2}^2 \cdots w_{i_k}^2\right] = \frac{1}{k!a^k} + o\left(\frac{1}{a^k}\right) \quad \text{as } a \to \infty. \tag{2.3.14}$$

Proposition 2.3.5. *Let $P \sim \mathrm{GDP}(a, r, H)$ and let $p_1, \cdots, p_k$ be positive integers. Then, as $a \to \infty$,*

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \cdots w_{i_k}^{p_k}\right] = O\left(\frac{1}{a^{p_{1:k}-k}}\right)\,. \tag{2.3.15}$$

In particular, when $p_j = 2$ for all $j \in \{1, \cdots, k\}$, the above expectation becomes

$$\mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^2 w_{i_2}^2 \cdots w_{i_k}^2\right] = \frac{\left[\frac{\sum_{k=1}^{r}\left(\frac{1}{k}\right)^2}{\left(\sum_{j=1}^{r}\frac{1}{j}\right)^2}\right]^k}{k!a^k} + o\left(\frac{1}{a^k}\right). \tag{2.3.16}$$

Remark 2.3.6. *As for* $\mathrm{SPG}(\mu_a, H)$*, a is a parameter such that v_i converges in distribution to 1 as $a \to \infty$. And we will give more details in Remark 2.4.13 on page 13 later on. For the specified processes in Propositions 2.3.2-2.3.5, the parameter a is the prior precision or the concentration parameter as we mentioned in the introduction. We can also see that the parameter a is the same order as* $\frac{1}{\mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\right]}$.

Remark 2.3.7. *The special cases when $p_1 = \cdots = p_k = 2$ in Propositions 2.3.1-2.3.5 are particularly important, since the corresponding terms in Theorem 2.4.4 will not converge to zero and we need to use them to identify the limits. Other terms will converge to* 0*. This is because otherwise some p_i will be greater than* 2 *and then there will be fewer factors in the product for the same value $p_1 + \cdots + p_k$ (see the proof of Theorem 2.4.4, Cases 1 and 2).*

Remark 2.3.8. *The quantity $p(n_1, \cdots, n_k) = \sum_{i_1,\cdots,i_k} \mathbb{E}\left[w_{i_1}^{n_1} w_{i_2}^{n_2} \cdots w_{i_k}^{n_k}\right]$ bears the same form of the exchangeable partition probability function (EPPF) in the random partition theory (see e.g (Pitman, 1996), (Pitman, 2003)), where the w_i is replaced by the so-called size biased permutation from a random partition. In the study of Poisson-Kingman model, the order statistics $w_1^*, w_2^*, \cdots$ of $w_1, w_2 \cdots$ are given by $w_i^* = \frac{J_i}{J_1+J_2+\cdots}$, where $J_1, J_2, \cdots$ are the ranked points of a Poisson process with Lévy density ρ (see (Pitman, 2003, Definition 3)). When v_i's are iid* $\mathrm{Beta}(1, a)$*, w_i^* is Poisson-Dirichlet distribution (see (Pitman, 1996, Theorem 5)). In general case it seems hard to find the distribution of w_i^* from v_i's. However, it remains interesting to apply our method of moments to study the asymptotics for the Poisson-Kingman model.*

2.4 Main results

2.4.1 Strong law of large numbers

The strong law of large numbers and the Glivenko-Cantelli theorem play undoubtedly important roles in statistics. In this subsection we state the strong law of large numbers and the Glivenko-Cantelli theorem for various processes introduced in the introduction. But before we state our theorem, we need an additional condition on the stick-breaking weights v_i in the case of $\mathrm{SPG}(\mu_a, H)$.

Assumption 2.4.1. *Let the iid stick-breaking weights* $\{v_i\}$ *satisfy*

$$\mathbb{E}[v_i^p] = \frac{C_p}{a^{k_p}} + o\left(\frac{1}{a^{k_p}}\right) \quad \text{as } a \to \infty\,, \tag{2.4.1}$$

for any $p \in \mathbb{N}$*, where* k_p *is a positive sequence satisfying* $jk_i \geq ik_j$ *for* $i \geq j$ *and* C_p *is a sequence of finite constants, independent of* a*.*

Theorem 2.4.2. *Let* P *be one of the stick-breaking process with general stick-breaking weights* $\mathrm{SPG}(\mu_a, H)$ *satisfying Assumption 2.4.1, the two-parameter Poisson-Dirichlet process* $\mathrm{PDP}(a, b, H)$*, the normalized inverse Gaussian process* $\text{N-IG}(a, H)$*, the normalized generalized gamma process*
$\mathrm{NGGP}(\sigma, a, H)$*, and the generalized Dirichlet process* $\mathrm{GDP}(a, r, H)$*. Assume that* $a = N^\tau$ *for some arbitrarily fixed* $\tau > 0$*. Then, as* $N \to \infty$*,*

$$P(A) \stackrel{a.s.}{\to} H(A) \tag{2.4.2}$$

for any measurable set $A \in \mathcal{X}$*.*

Once we have the strong law of large numbers for P, we can deduce the Glivenko-Cantelli theorem for P (see e.g., Theorem 20.6 in (Patrick, 1995) for a general discussion).

Theorem 2.4.3. *Let* $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Let* P *be one of the stick-breaking process with general stick-breaking weights* $\mathrm{SPG}(\mu_a, H)$ *satisfying Assumption 2.4.1, the two-*

parameter Poisson-Dirichlet process $\mathrm{PDP}(a,b,H)$*, the normalized inverse Gaussian process* $\text{N-IG}(a,H)$*, the normalized generalized gamma process* $\mathrm{NGGP}(\sigma,a,H)$*, and the generalized Dirichlet process* $\mathrm{GDP}(a,r,H)$*. Assume that* $a=N^\tau$ *for some arbitrarily fixed* $\tau>0$*. Then, as* $N\to\infty$*,*

$$\sup_{x\in\mathbb{R}} |P\left((-\infty,x]\right)-H\left((-\infty,x]\right)| \xrightarrow{a.s.} 0\,.$$

2.4.2 Central limit theorems and functional central limit theorems

In this subsection, we state the central limit theorems corresponding to the strong law of large numbers of the form (2.4.2).

We shall state the central limit theorems and functional central limit theorems for various processes as the following three theorems. The first one is for the stick-breaking process with general stick-breaking weights defined by Definition 2.2.4. We will assume mild convergence conditions on the stick-breaking weights.

Theorem 2.4.4. *Let* $P\sim\mathrm{SPG}(\mu_a,H)$*, where the stick-breaking weights* $v_1,v_2,\cdots$ *(whose distributions) depending on a parameter* $a>0$ *(we omit the explicit dependence on* a *of the* v_i*'s). Let* D_a *and* $Q_{H,a}$ *be defined by* (2.2.3) *and (2.2.4) respectively. Assume that the stick-breaking weights* $v_1,v_2,\cdots$ *satisfy the following two conditions.*

(i) For all $n\in\mathbb{Z}^+$*, we have*

$$\lim_{a\to\infty}\frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]}=0\,. \tag{2.4.3}$$

(ii) For any multi-index $(p_1,\cdots,p_k)$ *such that* $p_i\geq 2$ *and* $\frac{p_{1:k}}{2}>k$*, where* $p_{1:k}=\sum_{i=1}^k p_i$*, we have*

$$\lim_{a\to\infty}\frac{\left(\mathbb{E}[v_1]\right)^{\frac{p_{1:k}}{2}-k}\prod_{i=1}^k\mathbb{E}[v_1^{p_i}]}{\left(\mathbb{E}[v_1^2]\right)^{\frac{p_{1:k}}{2}}}=0. \tag{2.4.4}$$

Then we have the following results.

(i) *(Central limit theorem) Let $A_1, A_2, \cdots, A_n$ be any disjoint measurable subsets of $\mathbb{X}$. Then, as $a \to \infty$,*

$$(D_a(A_1), D_a(A_2), \cdots, D_a(A_n)) \xrightarrow{d} (X_1, X_2, \cdots, X_n)\,, \tag{2.4.5}$$

where $(X_1, X_2, \cdots, X_n) \sim N(0, \Sigma)$ and $\Sigma = (\sigma_{ij})_{1\le i,j\le n}$ is given by

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i = j\,, \\ -\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}} & \text{if } i \neq j\,. \end{cases} \tag{2.4.6}$$

(ii) *(Functional central limit theorem) Let $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be the d-dimensional Euclidean space with the Borel σ-algebra. Then*

$$Q_{H,a} \overset{\text{weakly}}{\to} \quad B_H^o \qquad \text{in} \qquad D(\mathbb{R}^d) \tag{2.4.7}$$

with respect to the Skorohod topology.

Remark 2.4.5. *For central limit theorem we use D_a because each component converges to a standard Gaussian. For functional central limit theorem we use $Q_{H,a}$ since it converges to a Brownian bridge with parameter H. We can presumably use D_a (or $Q_{H,a}$) in both (2.4.6) and (2.4.7) with a scaling.*

The conditions (i) and (ii) in Theorem 2.4.4 are implied by many other conditions. One of them is given below.

Remark 2.4.6. *Assumption 2.4.1 implies the conditions (i) and (ii) in Theorem 2.4.4.*

Proof. It is obviously that $\{k_p\}$ is an increasing sequence, and thus the condition (i) of Theorem 2.4.4 (i.e (2.4.3)) holds.

For any nonnegative integer m, let $\mathfrak{N}$ be a certain collection of integers $j's$ such that $\sum_{j\in\mathfrak{N}} j = m$. The condition (ii) in Theorem 2.4.4 is equivalent to the following statement:

If $j \geq 2$ and $|\mathfrak{N}| < \frac{\mathfrak{m}}{2}$, then

$$\frac{m(k_2 - k_1)}{2} < \sum_{j \in \mathfrak{N}} k_j - |\mathfrak{N}| k_1. \tag{2.4.8}$$

Thus, to prove (2.4.4) it is sufficient to show $\frac{mk_2}{2} < \sum_{j \in \mathfrak{N}} k_j$. This is a simple consequence of $jk_i \geq ik_j$ for $i \geq j$. In fact, taking $i = 2$, we have for all $j \geq 2$, $2k_j \geq jk_2$ holds and thus we have $\sum_{j \in \mathfrak{N}} 2k_j \geq \sum_{j \in \mathfrak{N}} jk_2$, which implies $\frac{mk_2}{2} < \sum_{j \in \mathfrak{N}} k_j$. Hence we have (2.4.8). □

The conditions (i) and (ii) in Theorem 2.4.4 are satisfied by many interesting processes including the Dirichlet process. We give three examples to illustrate the applicability of our above theorem.

Corollary 2.4.7. *Theorem 2.4.4 holds true when* $P \sim \mathrm{DP}(a, H)$.

Proof. It is sufficient to verify the condition (2.4.1) in Assumption 2.4.1. Since $v_i \overset{iid}{\sim}$ Beta$(1, a)$, we have for any positive integer p,

$$\mathbb{E}[v_i^p] = \frac{\Gamma(a+1)\Gamma(p+1)}{\Gamma(1)\Gamma(a+p+1)} = \frac{p!}{(a+1)\cdots(a+p)} = \frac{p!}{a^p} + o\left(\frac{1}{a^p}\right).$$

Hence, $k_p = p$ and $C_p = p!$. Obviously, for $i \geq j$, $jk_i \geq ik_j$ always holds true. □

Remark 2.4.8. *Since the posterior of the Dirichlet process is still a Dirichlet process, the above result can be applied to the posterior process in the Bayesian nonparametric models when the prior is the Dirichlet process for the following situations: (i) with large sample size and finite parameter a; (ii) with large parameter a and finite sample size, (iii) with parameter a and sample size both large.*

The assumption of the Beta$(1, a)$-distribution in Corollary 2.4.7 can be replaced by a general Beta(ρ_a, a), where $\rho_a / a \to 0$. In fact, in this case, we have

$$\mathbb{E}\left[v_1^n\right] = \left(\frac{\rho_a}{a}\right)^n + o\left(\left(\frac{\rho_a}{a}\right)^n\right).$$

It is easy to verify that the conditions (2.4.3)-(2.4.4) in Theorem 2.4.4 are satisfied. Thus we have

Corollary 2.4.9. *Theorem 2.4.4 holds true when* $P \sim \mathrm{SPG}(\mu_a, H)$*, where* $v_i \overset{iid}{\sim} \mathrm{Beta}(\rho_a, a)$ *with* $\lim_{a\to\infty} \frac{\rho_a}{a} = 0$.

Remark 2.4.10. *It is not clear yet what is the finite dimensional distribution of stick-breaking process* P *if the corresponding stick-breaking weights* $v_i \overset{iid}{\sim} \mathrm{Beta}(\rho_a, a)$.

The next corollary is about the asymptotic behaviour of the prior P, when the corresponding stick-breaking weights v_i follow a linear combination of Beta distributions, whose precise meaning is give below.

Definition 2.4.11. *Let* s *be any positive integer and let* $\{r_1, \cdots, r_s\}$ *and* $\{t_1, \cdots, t_s\}$ *be two sets of positive real numbers such that* $\sum_{\ell=1}^{s} t_\ell = 1$*. Let* $u_{1,1}, \cdots, u_{1,s}, u_{2,1}, \cdots, u_{2,s}, \cdots$ *be independent and let* $u_{i,\ell} \sim \mathrm{Beta}(1, a^{r_\ell})$ *,* $i = 1, 2, \cdots, \ell = 1, \cdots, s$*. Then the random variables*

$$v_i = \sum_{\ell=1}^{s} t_\ell u_{i,\ell}\,, \quad i = 1, 2, \cdots, \tag{2.4.9}$$

are called linear combinations of Beta random variables.

Corollary 2.4.12. *Theorem 2.4.4 holds true when* P *is the stick-breaking process as defined in Definition 2.2.1, where the weights* v_i *are the linear combinations of Beta random variables defined by* (2.4.9).

Proof. By the independence of $\{u_{i,\ell}\}_{\ell=1}^{s}$, we can compute the p-th moment of v_i as follows.

$$\mathbb{E}[v_i^p] = \mathbb{E}\left[\left(\sum_{\ell=1}^{s} t_\ell u_{i,\ell}\right)^p\right] = \sum_{\substack{q_1,\cdots,q_s\in\mathbb{Z}_+\\ q_1+\cdots+q_s=p}} \binom{p}{q_1, \cdots, q_s} \prod_{\ell=1}^{s} \mathbb{E}\left[(t_\ell u_{i,\ell})^{q_\ell}\right]$$

$$= \sum_{\substack{q_1,\cdots,q_s\in\mathbb{Z}_+\\ q_1+\cdots+q_s=p}} \binom{p}{q_1, \cdots, q_s} \prod_{\ell=1}^{s} t_\ell^{q_\ell} \left(\frac{q_\ell!}{a^{q_\ell r_\ell}} + o\left(\frac{1}{a^{q_\ell r_\ell}}\right)\right) = \frac{t^p p!}{a^{pr}} + o\left(\frac{1}{a^{pr}}\right),$$

where $r = \min(r_1, \cdots, r_s)$. Taking $k_p = pr$ and $C_p = t^p p!$ in Assumption 2.4.1 we see the condition $i \geq j$, $jk_i \geq ik_j$ is always verified. □

Remark 2.4.13. *Let us return to Corollary 2.4.7. This is a typical case and we take a close look of the density* $f_a(x) = a(1-x)^{a-1}$, $0 \le x \le 1$, *of the Beta distribution* Beta$(1, a)$.

For any continuous function $g : \mathbb{R} \to \mathbb{R}$, *it is easy to verify that*

$$\int_{\mathbb{R}} [g(x) - g(1)]\, f_a(x)dx = \int_0^1 [g(x) - g(1)]\, f_a(x)dx \to 0 \quad \text{as } a \to \infty\,.$$

This means that $\int_{\mathbb{R}} g(x)f_a(x)dx \to g(1)$. *In other word,* f_a *converges to the Dirac delta function* $\delta(x-1)$. *This observation hints that when the distribution* f_a *of* v_i*'s converges to the Dirac delta function* $\delta(x-1)$, *or the random variable* v_i *converges in distribution to* 1 *(as* $a \to \infty$*) we should have the convergence of the random process* $Q_{H,a}$. *But we still need to impose some more technical conditions. We give a further illustration by the following corollary.*

Corollary 2.4.14. *Let the stick-breaking process* P *be defined as in Definition 2.2.1, where the corresponding* v_i *follows the following distribution:*

$$f_a(x) = \begin{cases} a(1-g(a)) & \text{if } 0 < x \le 1/a; \\ \frac{ag(a)}{a-1} & \text{if } 1/a < x \le 1, \end{cases}$$

where $g(a) = e^{-a^\epsilon}, a > 1$, *for a certain arbitrarily fixed* $\epsilon > 0$. *Then, as* $a \to \infty$, *the conditions* (2.4.3) *and* (2.4.4) *of Theorem 2.4.4 hold for this density* f_a. *Thus the statements* (2.4.5) *and* (2.4.7) *of Theorem 2.4.4 hold true.*

Proof. Before we proceed to the proof. Let us note the obvious fact that f_a converges to the Dirac delta distribution $\delta(x-1)$.

For any $n > 0$, we see $\lim\limits_{a\to\infty} a^n g(a) = 0$. A trivial calculation implies that for any positive integer p,

$$\mathbb{E}[v_i^p] = \frac{(1/a)^p + g(a)\sum_{i=0}^p (1/a)^i}{p+1} = \frac{1}{(p+1)a^p} + o(b^{-p})\,.$$

An application of Assumption 2.4.1 with $k_p = p$ yields the desired statement. □

When the stick-breaking weights are iid, Theorem 2.4.4 that we obtained for the stick-breaking random measure P covers very general situation and the conditions (2.4.3)-(2.4.4) are minimal and are easy to verify. But when the stick-breaking weights are not iid the situation becomes much more sophisticated like in other statistical situations. We shall consider some well-known processes introduced in the introduction. For these processes the explicit forms of the joint finite dimensional distributions of the stick-breaking weights, although complicated, are given (in the supplementary material). We can state similar results as those in Theorem 2.4.4 in one theorem for all these processes.

Theorem 2.4.15. *Let P be one of the Poisson-Dirichlet process* $\mathrm{PDP}(a, b, H)$*, the normalized inverse Gaussian process* $\text{N-IG}(a, H)$*, the normalized generalized gamma process* $\mathrm{NGGP}(\sigma, a, H)$*, and the generalized Dirichlet process* $\mathrm{GDP}(a, r, H)$*. Then, we have the following results.*

(i) As $a \to \infty$,

$$(D_a(A_1), D_a(A_2), \cdots, D_a(A_n)) \xrightarrow{d} (X_1, X_2, \cdots, X_n)\,, \tag{2.4.10}$$

where $(X_1, X_2, \cdots, X_n) \sim N(0, \Sigma)$ and $\Sigma = (\sigma_{ij})_{1\le i,j\le n}$ is given by

$$\sigma_{ij} = \begin{cases} 1 & \text{if } i = j \\ -\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}} & \text{if } i \neq j\,. \end{cases} \tag{2.4.11}$$

(ii) Let $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be the d-dimensional Euclidean space with the Borel σ-algebra. Then

$$Q_{H,a} \overset{\text{weakly}}{\to} B_H^o \quad \text{in} \quad D(\mathbb{R}^d) \tag{2.4.12}$$

with respect to the Skorohod topology.

Corollary 2.4.16. *Theorem 2.4.15 holds true when the random measure P is the Beta process (denoted by $P \sim$* $\mathrm{BP}(a, \gamma H)$*), whose stick-breaking representation is given in*

Definition 7.8 (in the supplementary material). Our method of moments still works and in fact, due to the independence of the weights $w_{i,j}$ in (7.9) (in the supplementary material) the computation is much simpler.

As long as the central limit theorem of P is obtained, it is trivial to use the delta-method to show the similar theorem for the nonlinear functional of this process. Using Theorem 3.9.4 in (van der Vaart and Wellner, 1996), we can state the following theorem.

Theorem 2.4.17. *Let P be one of* $\text{N-IG}(a, H)$, $\text{PDP}(a, b, H)$, $\text{NGGP}(\sigma, a, H)$, $\text{GDP}(a, r, H)$ *or* $\text{SPG}(\mu_a, H)$ *satisfying* (2.4.3)-(2.4.4) *of Theorem 2.4.4. Let $\mathbb{D}$ be the metric space of all probability measures on $(\mathbb{X}, \mathcal{X})$ with the total variation distance. Let $\phi : \mathbb{D} \to \mathbb{R}^d$ be a continuous functional which is Hadamard differentiable on $\mathbb{D}$. Then, as $a \to \infty$, we have*

$$\frac{1}{\sqrt{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}} \left(\phi\left(P(\cdot)\right) - \phi\left(H(\cdot)\right)\right) \overset{weakly}{\to} \phi'_{H(\cdot)}\left(B_H^o\right).$$

Remark 2.4.18. *When P is Dirichlet process or the normalized inverse Gaussian process, the above conclusion have been known (e.g., (Labadi and Zarepour, 2013; Labadi and Abdelrazeq, 2016)).*

One application of the above theorem is the limiting distribution of the quantile process of P.

Example 2.4.19. *Suppose* $(\mathbb{X}, \mathcal{X}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *and suppose that P is one of* $\text{N-IG}(a, H)$, $\text{PDP}(a, b, H)$, $\text{NGGP}(\sigma, a, H)$, $\text{GDP}(a, r, H)$ *or* $\text{SPG}(\mu_a, H)$ *satisfying* (2.4.3)-(2.4.4) *of Theorem 2.4.4. Let H be absolutely continuous with positive derivative h. By Lemma 3.9.23 of (van der Vaart and Wellner, 1996), we have*

$$\frac{1}{\sqrt{\mathbb{E}[\sum_{i=1}^{\infty} w_i^2]}} \left(P^{-1}(\cdot) - H^{-1}(\cdot)\right) \overset{weakly}{\to} -\frac{B^o(\cdot)}{h\left(H^{-1}(\cdot)\right)} = G(\cdot), \tag{2.4.13}$$

where $H^{-1}(s) = \inf\{t : H(t) \geq s\}$. The limiting process G is a Gaussian process with zero-mean and with covariance function

$$\text{Cov}\left(G\left((0, s]\right), G\left((0, t]\right)\right) = \frac{s \wedge t - st}{h\left(H^{-1}\left((0, s]\right)\right) h\left(H^{-1}\left((0, t]\right)\right)}$$

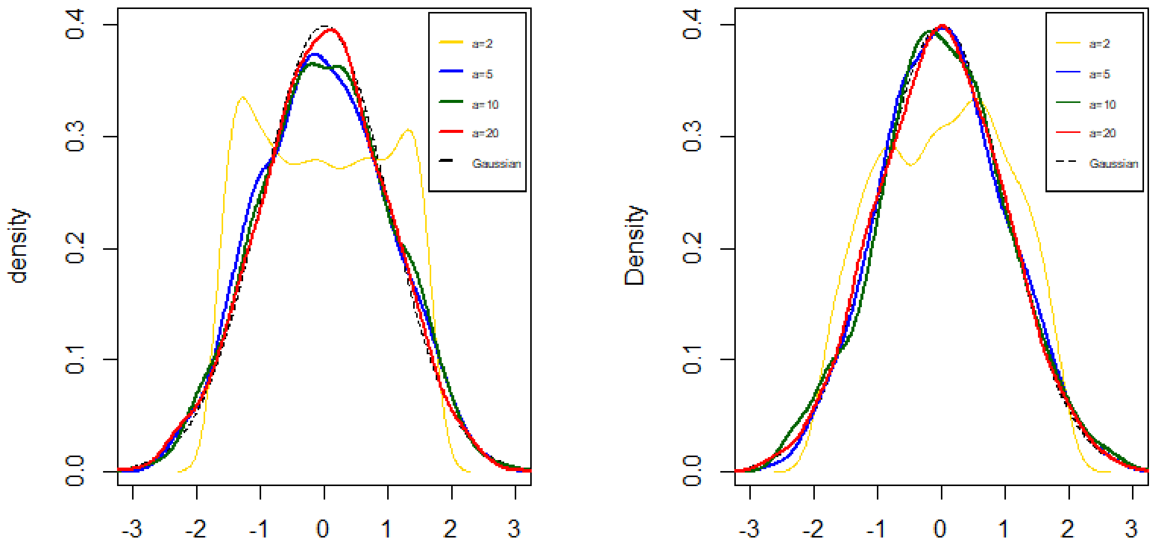
for $s, t \in \mathbb{R}$,

2.5 Numerical Illustration

Theorem 2.4.15 states that $(D_a(A_1), D_a(A_2), \cdots, D_a(A_n))$ converges to a joint normal distribution as $a \to \infty$. In this section we shall perform some numerical simulations to illustrate this convergence. To be specific the processes we choose to simulate are Poisson-Dirichlet process PDP(a, b, H) and the stick-breaking process with general stick-breaking weights constructed in Corollary 2.4.14. For PDP(a, b, H) we consider the cases the parameter $b = 0.2$ and $b = 0.5$ and for the process constructed in Corollary 2.4.14 we consider the cases $\epsilon = 1$ and $\epsilon = 5$.

For both of these two processes, the base measure H is assumed to be uniform distribution on $\mathbb{X} = (0, 1)$ and we take $n = 3$ and fix the partition of $\mathbb{X}$ as $A_1 = (0, 0.3]$, $A_2 = (0.3, 0.7]$, $A_3 = (0.7, 1)$. In our simulations we truncate the infinite series (2.2.1) to 5000 terms and we simulate 2000 samples of $(D_a(A_1), D_a(A_2), D_a(A_3))$. Since it is rather messy to visualize the joint densities of $(D_a(A_1), D_a(A_2), D_a(A_3))$, we plot the histograms of the linear combination $1.6 \times D_a(A_1) + 1.4 \times D_a(A_2) + 0.5 \times D_a(A_3)$ (other linear combinations will produce similar results with different variances). The histograms for PDP(a, b, H) with $b = 0.2$ and $b = 0.5$ and with $a = 2, 5, 10, 20$ are plotted in Figure 2.1 and the histograms for the stick-breaking process with general stick-breaking weights constructed in Corollary 2.4.14 with $\epsilon = 1$ and $\epsilon = 5$ and with $a = 2, 5, 10, 20$ are plotted in Figure 2.2. Graphs corresponding to different parameters (different b or different ϵ) are plotted in different figures but those corresponding to different values of a are plotted in the same figure with different colored curves so that one can observe the convergence to mean zero normal curves easily.

It is easy to observe that the convergence to the normal shape is very fast as a is getting larger.



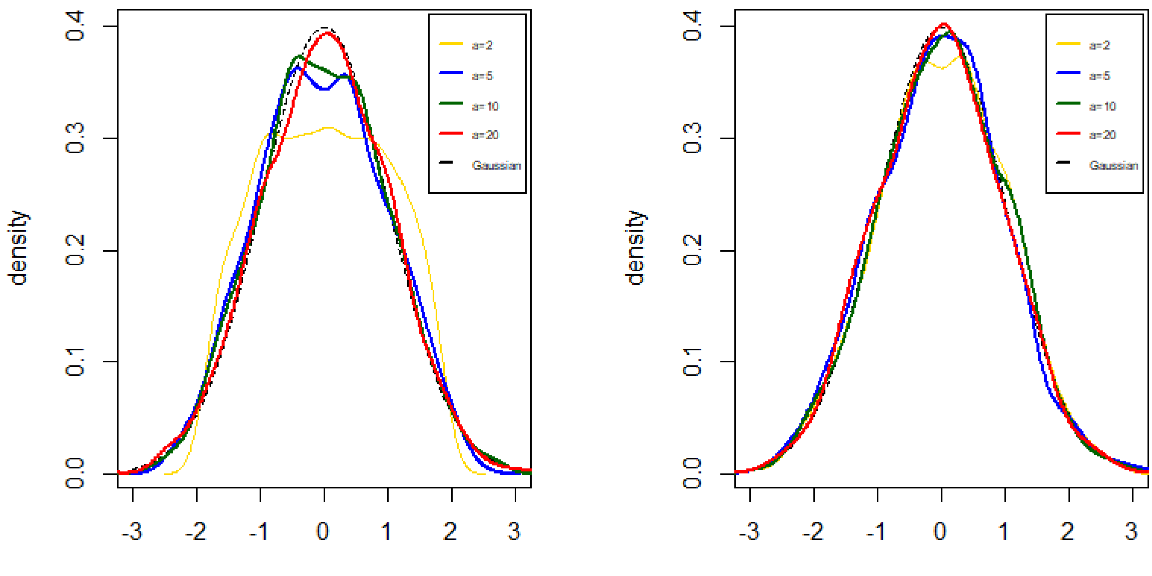
(a) PDP(a, b, H) with $b = 0.2$.

(b) PDP(a, b, H) with $b = 0.5$.

Figure 2.1: Convergence of D_a with respect to PDP(a, b, H) for $a = 2, 5, 10, 20$.

2.6 Concluding remarks

The method of moments used in this chapter could be applied to the study of asymptotics for some Bayesian nonparametric posterior processes in the following situations: (i) when the parameter a is finite and the sample size is large; (ii) when the parameter a is large and the sample size is finite; (iii) when the parameter a and the sample size are both large. As is well-known it is usually very hard to obtain the explicit form of the posterior distribution (even in the parametric cases) and even when the posterior distribution is obtained sometimes it is still very hard to use it to compute the needed statistics. A particularly interesting example is the posterior distribution of a homogeneous normalized random measure with independent increments (hNRMI) obtained by (James et al., 2009) and (Favaro et al., 2016, Proposition 4). The hNRMI is a large class of priors, which contains the normalized generalized gamma processes (Definition 7.6 in the supplementary material) and the generalized Dirichlet processes (Definition 7.7 in the supplementary



(a) Parameter $\epsilon = 1$.

(b) Parameter $\epsilon = 5$.

Figure 2.2: Convergence of D_a with respect to the constructed process in Corollary 2.4.14 for $a = 2, 5, 10, 20$.

material) mentioned in this chapter as special cases. Assume that P^a is an hNRMI depending on a parameter $a > 0$ and some other parameters studied in (Favaro et al., 2016), where the parameter a is the same as the one in our work when the hNRMI becomes a normalized generalized gamma process or a generalized Dirichlet process. If $\{X_i\}_{i=1}^n$ is a sample from the hNRMI P^a, in the sense that the sequence of exchangeable observations $\{X_i\}_{i=1}^n$ are defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{X}$ in such a way that, given P^a, $X_1, \cdots, X_n$ are i.i.d with distribution P^a, then the posterior distribution of P^a can be computed with the help of a latent variable U_n as follows ((Favaro et al., 2016, Proposition 4))

$$P^a|U_n, X_1, \ldots, X_n \;\sim\; \hat{P}_n^a := \varphi_{0,U_n}^a \tilde{P}_{U_n}^a + \sum_{j=1}^k \varphi_{j,U_n}^a \delta_{X_j^*},$$

where given $U_n = u$, $\tilde{P}_u^a$ is an hNRMI admitting a stick-breaking representation and $\{X_j^*\}_{j=1}^k$ are the distinct values of $\{X_i\}_{i=1}^n$. To compute a Bayesian statistic, we need to compute some functional of the posterior probability measure $\hat{P}_n^a$. For example, to find the quantile t_q such that $\hat{P}_n^a((-\infty, t_q)) \leq q \leq \hat{P}_n^a((-\infty, t_q])$ for some given $q \in (0,1)$ or to compute $\int_{\mathbb{X}^d} f(x_1, \cdots, x_d)\hat{P}_n^a(dx_1)\cdots\hat{P}_n^a(dx_d)$ and so on (e.g., (Ferguson, 1973), (Zhang and Hu, 2021)), which is usually complicated due to the complexity of $\tilde{P}_{U_n}^a$. However, when a is sufficiently large the probability measure $\tilde{P}_{U_n}^a$ is approximately a normal distribution, then we can use the normal distribution to approximate $\tilde{P}_{U_n}^a$ in the computation of these Bayesian statistics. Let us also point out that in some situations the normal approximation is sufficiently good for reasonable size a (from the figures in this chapter, we see that when $a = 5$, the graphs are already close to normal distributions).

2.7 Appendix

In this section, we present the necessary definitions and propositions for the processes (random measures) considered in this chapter and give the proofs for the propositions and theorems in the main body.

2.7.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $(\mathbb{X}, \mathcal{X})$ be a measurable Polish space, namely, $\mathbb{X}$ is a separable complete metric space and $\mathcal{X}$ is the Borel σ-algebra of $\mathbb{X}$. Let H be a nonatomic probability measure on $(\mathbb{X}, \mathcal{X})$ (i.e., $H(\{x\}) = 0$ for any $x \in \mathbb{X}$). A random measure is a mapping P from $\Omega \times \mathcal{X}$ to $\mathbb{R}_+$ (we denote this random measure by $P = (P(\omega, A), \, \omega \in \Omega \, , A \in \mathcal{X}))$ such that

(i) when $\omega \in \Omega$ is fixed, $P(\omega, \cdot)$ is a measure on $(\mathbb{X}, \mathcal{X})$;

(ii) when $A \in \mathcal{X}$ is fixed, $P(\cdot, A)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

In the following definitions we shall always assume that P is a random probability measure, which are given by their stick-breaking representation. Different assumptions on the stick-breaking weights give rise to different processes. The first one is the classical Dirichlet process.

Definition 2.7.1. *Let $a > 0$ and let H be a nonatomic measure on $(\mathbb{X}, \mathcal{X})$. A random probability measure P is called the Dirichlet process with parameter (a, H), denoted by $P \sim \mathrm{DP}(a, H)$, if it has the representation* (2.1)-(2.2)*, where $v_i \overset{iid}{\sim} \mathrm{Beta}(1, a)$.*

In fact, this process is defined differently and the above definition is a result of (Sethuraman, 1994). To state the original definition of the Dirichlet process as the following proposition, we need to recall the concept of the Dirichlet distribution. Throughout this section we use the following notation to denote the standard simplex in $\mathbb{R}^n$:

$$\mathbb{S}_n = \left\{ (s_1, \cdots, s_n) \in \mathbb{R}^n : \quad s_i \geq 0, \quad \sum_{i=1}^{n} s_i = 1 \right\} . \tag{2.7.1}$$

In case of no ambiguity we also write $\mathbb{S} = \mathbb{S}_n$. A random vector $(X_1, \cdots, X_n) \in \mathbb{S}$ follows the Dirichlet distribution with parameters $(\alpha_1, \cdots, \alpha_n) \in [0, \infty)^n$, denoted by $(X_1, \cdots, X_n) \sim \mathrm{Dir}(\alpha_1, \cdots, \alpha_n)$, if the joint probability density function of $(X_1, \cdots, X_n)$

is given by

$$f(x_1, \cdots, x_n) = \frac{\Gamma(|\alpha|)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n x_i^{\alpha_i - 1} \mathbb{1}_{\mathbb{S}}(x_1, \cdots, x_n),$$

where $|\alpha| = \sum_{i=1}^n \alpha_i$, $\Gamma(a) = \int_0^\infty x^{a-1} dx$ $(a > 0)$, is the gamma function, and $\mathbb{1}_{\mathbb{S}}$ is the indicator function of the simplex $\mathbb{S}$. With this notion of Dirichlet distribution we can write the following proposition.

Proposition 2.7.2. *A random probability measure P is the Dirichlet process with parameter (a, H) if for any measurable partition $(A_1, \cdots, A_n)$ of $\mathbb{X}$ (i.e., $A_1, \cdots, A_n \in \mathcal{X}$, $A_1 \cup \cdots \cup A_d = \mathbb{X}$ and $A_i \cap A_j = \emptyset$ for $1 \le i < j \le n$), the random vector $(P(A_1), \cdots, P(A_n))$ follows the Dirichlet distribution with parameters $(aH(A_1), \cdots, aH(A_n))$.*

Proof. We refer to (Sethuraman, 1994) or (Zhang and Hu, 2021) for the proof of the equivalence between Definition 2.7.1 and Proposition 2.7.2. □

Definition 2.7.3 (Pitman and Yor (1997))**.** *Let $b \in (0, 1)$ and let $-b < a < \infty$. A random probability measure P is called the two-parameter Poisson-Dirichlet process or the Pitman-Yor process, denoted by* $\mathrm{PDP}(a, b, H)$*, if the stick-breaking weights satisfy the following:*

$$\begin{cases} v_1, v_2, \cdots \ \textit{are independent}, \\ v_i \sim \mathrm{Beta}(1 - b, a + ib), \quad i = 1, 2, \cdots \end{cases} \tag{2.7.2}$$

Definition 2.7.4 (Favaro et al. (2012))**.** *A random probability measure P is called the normalized inverse Gaussian process with parameters a and H, denoted by $P \sim$* N-IG(a, H)*, if the joint distributions of the stick-breaking weights $\{v_1, v_2, \cdots\}$ are given through the* *following conditional probability densities recursively:*

$$\begin{cases} f_{v_1}(x) = \dfrac{a^{\frac{1}{2}} x^{-\frac{1}{2}} (1-x)^{-1}}{(2\pi)^{\frac{1}{2}} K_{-\frac{1}{2}}(a)} K_{-1}\left(\dfrac{a}{\sqrt{1-x}}\right), & (2.7.3) \\ f_{v_n|v_1,\cdots,v_{n-1}}(x) = \dfrac{a^{\frac{1}{2}} \prod_{i=1}^{n-1}(1-v_i)^{-\frac{1}{4}} x^{-\frac{1}{2}} (1-x)^{-\frac{5}{4}+\frac{n}{4}}}{(2\pi)^{\frac{1}{2}} K_{-\frac{n}{2}}\left(\frac{a}{\sqrt{\prod_{i=1}^{n-1}(1-v_i)}}\right)} \\ \qquad \times K_{-\frac{1}{2}-\frac{n}{2}}\left(\dfrac{a}{\sqrt{(1-x)\prod_{i=1}^{n-1}(1-v_i)}}\right), \\ \qquad n = 2, 3, \cdots & (2.7.4) \end{cases}$$

where $a > 0$ and K_μ is the modified Bessel function of the third type (see e.g., (Gradshteyn and Ryzhik, 2014)).

Similar to what we did for the Dirichlet process, we present the original definition of the normalized inverse Gaussian process as a proposition.

Proposition 2.7.5. *A random probability measure P is the normalized inverse Gaussian process with parameter (a, H) if for any measurable partition $(A_1, \cdots, A_n)$ of $\mathbb{X}$, the random vector $(P(A_1), \cdots, P(A_n))$ follows the normalized inverse Gaussian distribution with parameters $(aH(A_1), \cdots, aH(A_n))$ given by the following form:*

$$f(x_1, \cdots, x_n) = \frac{e^a a^n \prod_{i=1}^n H(A_i)}{2^{\frac{n}{2}-1} \pi^{\frac{n}{2}}} \times K_{-\frac{n}{2}}\left(\sqrt{\sum_{i=1}^n \frac{(aH(A_i))^2}{x_i}}\right)$$
$$\times \left(\sum_{i=1}^n \frac{(aH(A_i))^2}{x_i}\right)^{-\frac{n}{4}} \times \prod_{i=1}^n x_i^{-\frac{3}{2}} \times \mathbb{1}_{\mathbb{S}}(x_1, \cdots, x_n),$$

where $\mathbb{S}$ is the simplex defined by (2.7.1).

Proof. We refer to (Favaro et al., 2012) for the proof of the equivalence between Definition 2.7.4 and Proposition 2.7.5. □

Definition 2.7.6 (Favaro et al. (2016))**.** *P is called the normalized generalized gamma process with parameters $\sigma \in (0, 1)$, $a > 0$ and H, denoted by $P \sim \text{NGGP}(\sigma, a, H)$, if the* *finite dimensional joint distributions of the stick-breaking weights* $\{v_1, v_2, \cdots\}$ *are given by the following conditional distributions:*

$$\begin{cases} f_{v_1}(x) = \dfrac{x^{-\sigma}(1-x)^{\sigma-1}e^a}{\Gamma(\sigma)\Gamma(1-\sigma)} \displaystyle\sum_{j=0}^{\infty} \frac{(1-\sigma)_j}{j!} \frac{a^{\frac{j}{\sigma}}}{(1-x)^j} \Gamma\left(1-\frac{j}{\sigma}; \frac{a}{(1-x)^\sigma}\right), & (2.7.5) \\ f_{v_n|v_1,\cdots,v_{n-1}}(x) = \dfrac{\sigma\Gamma((n-1)\sigma)x^{-\sigma}(1-x)^{n\sigma-1}}{\Gamma(1-\sigma)\Gamma(n\sigma)} \\ \qquad \dfrac{\sum_{j=0}^{\infty} \frac{(1-n\sigma)_j}{j!} \frac{a^{\frac{j}{\sigma}}}{(1-x)^j \prod_{i=1}^{n-1}(1-v_i)^j} \Gamma\left(n-\frac{j}{\sigma}; \frac{a}{(1-x)^\sigma \prod_{i=1}^{n-1}(1-v_i)^\sigma}\right)}{\sum_{j=0}^{\infty} \frac{(1-(n-1)\sigma)_j}{j!} \frac{a^{\frac{j}{\sigma}}}{\prod_{i=1}^{n-1}(1-v_i)^j} \Gamma\left(n-1-\frac{j}{\sigma}; \frac{a}{\prod_{i=1}^{n-1}(1-v_i)^\sigma}\right)}, \\ \qquad n = 2, 3, \cdots, & (2.7.6) \end{cases}$$

where $\Gamma(c, x) = \int_x^\infty u^{c-1}e^{-u}du$ *is the upper incomplete gamma function.*

Definition 2.7.7. *We call a random probability measure* P *on* $(\Omega, \mathcal{F})$ *the generalized Dirichlet process with parameters* $a > 0$, $r \in \mathbb{N}^+$ *and* H, *denoted by* $P \sim \mathrm{GDP}(a, r, H)$, *if for any measurable partition* $(A_1, \cdots, A_n)$ *of* $\mathbb{X}$, *the joint density of* $(P(A_1), \cdots, P(A_n))$ *is given by*

$$f(x_1, \cdots, x_n) = \frac{(r!)^a}{\prod_{i=1}^n \Gamma(ra_i)} \int_0^\infty t^{ra-1}e^{-rt} \left[\prod_{j=1}^n \Phi_2^{(r-1)}\left(a_j \mathbf{I}_{r-1}; ra_j; tx_j\mathbf{J}_{r-1}\right)\right] dt$$
$$\times \prod_{i=1}^n x_i^{ra_i-1} \times \mathbb{1}_{\mathbb{S}}(x_1, \cdots, x_n), \qquad (2.7.7)$$

where $a_i = aH(A_i)$; $\mathbf{I}_{r-1} = (1, \cdots, 1)^T$, $\mathbf{J}_{r-1} = (1, \cdots, r-1)$ *are* $r-1$ *dimensional vectors and* $\Phi_2^N(\mathbf{b}; c; \mathbf{x})$ *is the confluent form of the fourth Lauricella hypergeometric function (see e.g., (Exton, 1976)), and* $\mathbb{S}$ *is the simplex defined by* (2.7.1).

It is trivial to verify that the Dirichlet process is a special case of the generalized Dirichlet process with parameter $r = 1$. Although the expression (2.7.7) looks very sophisticated, its mean, variance, and predictive distribution have been computed (see e.g., (Lijoi et al., 2005b)). This process also admits a stick-breaking representation (e.g., (Favaro et al., 2016)). However, the corresponding stick-breaking representation is more complicated to use for our study of the limiting theorems. So, we rather use this sophisticated finite dimensional distribution than the more sophisticated stick-breaking representation, which we omit.

For the Beta process, the stick-breaking representations are given in (Paisley et al., 2010) and (Teh et al., 2007). We use the former as our definition below.

Definition 2.7.8. *A random measure P is called the Beta process with parameters $a > 0$, $\gamma > 0$, H, denoted by $P \sim \mathrm{BP}(a, \gamma H)$, if it has the following representation:*

$$\begin{cases} P = \sum_{i=1}^{\infty} \sum_{j=1}^{\kappa_i} w_{i,j} \delta_{\theta_{i,j}}, \qquad \textit{where} & (2.7.8) \\ w_{1,j} = v_{1,j}^{(1)}, \quad w_{i,j} = v_{i,j}^{(i)} \prod_{l=1}^{i-1} (1 - v_{i,j}^{(l)}) \quad \textit{for} \quad i = 2, 3, \cdots, \textit{and} \quad j = 1, 2, \cdots, & (2.7.9) \end{cases}$$

where all variables are iid and $\kappa_i \overset{iid}{\sim} \mathrm{Poisson}(\gamma)$, $v_{i,j}^{(l)} \overset{iid}{\sim} \mathrm{Beta}(1, a)$, $\theta_{i,j} \overset{iid}{\sim} H$ are mutually independent.

As we are presenting the functional central limit theorem of P, we need to recall the definition of the Brownian bridge process of parameter H (see e.g., (Kim and Bickel, 2003) for more details).

Definition 2.7.9. *Let H be a measure on $(\mathbb{X}, \mathcal{X})$ and let $B_H^o = (B_H^o(\omega, A), \omega \in \Omega, A \in \mathcal{X})$ be a stochastic process (random measure) with parameter $A \in \mathcal{X}$. It is called the Brownian bridge with parameter H if the following two conditions are satisfied.*

(i) B_H^o is Gaussian. Namely, for any elements $A_1, \cdots, A_n \in \mathcal{X}$, $B_H^o(A_1), \cdots, B_H^o(A_n)$ are jointly centered Gaussian random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

(ii) For any $A_1, A_2 \in \mathcal{X}$, the covariance of $B_H^o(A_1)$ and $B_H^o(A_2)$ is given by

$$\mathbb{E}\ [B_H^o(A_1) B_H^o(A_2)] = H(A_1 \cap A_2) - H(A_1) H(A_2)\,. \qquad (2.7.10)$$

To state the functional central limit theorem we also need the space $D(\mathbb{R}^d)$ introduced in Section 3 of (Bickel and Wichura, 1971). The characteristics of the elements (functions) in $D(\mathbb{R}^d)$ are given by their continuity properties described as follows. For $1 \le p \le d$, let R_p be one of the relations $<$ or $\ge$ and for $t = (t_1, \cdots, t_d) \in \mathbb{R}^d$ let $\mathcal{Q}_{R_1,\cdots,R_d}$ be the quadrant

$$\mathcal{Q}_{R_1,\cdots,R_d} := \left\{(s_1, \cdots, s_d) \in \mathbb{R}^d : \, s_p R_p t_p, \, 1 \le p \le d\right\}.$$

Then, $x \in D(\mathbb{R}^d)$ if and only if (see e.g., (Straf, 1972)) for each $t \in \mathbb{R}^d$, the following two conditions hold: (i) $x_{\mathcal{Q}} = \lim_{s \to t, \, s \in \mathcal{Q}} x(s)$ exists for each of the 2^d quadrants $\mathcal{Q} = \mathcal{Q}_{R_1,\cdots,R_d}(t)$ (namely, for all the combinations that $R_1 =$ "$<$", or "$\ge$", $\cdots$, $R_d =$ "$<$" or "$\ge$"), and (ii) $x(t) = x_{\mathcal{Q}_{\ge,\cdots,\ge}}$. In other words, $D(\mathbb{R}^d)$ is the space of functions that are "continuous from above with limits from below", which are similar to the space of the càdlàg (French word abbreviation for "right continuous with left limits") functions in one variable (i.e., $d = 1$). The metric on $D(\mathbb{R}^d)$ is introduced as follows. Let $\Lambda = \{\lambda : \mathbb{R}^d \to \mathbb{R}^d : \, \lambda(t_1, \cdots, t_d) = (\lambda_1(t_1), \cdots, \lambda_d(t_d))\}$, where each $\lambda_p : \mathbb{R} \to \mathbb{R}$ is continuous, strictly increasing and has limits at both infinities. Denote the Skorohod distance between $x, y \in D(\mathbb{R}^d)$ by

$$d(x, y) = \inf\{\min(\| x - y\lambda \|, \| \lambda \|) : \, \lambda \in \Lambda\},$$

where $\| x - y\lambda \| = \sum_{n=1}^{\infty} \sup_{|t| \le n} |x(t) - y(\lambda(t))|$ and $\| \lambda \| = \sum_{n=1}^{\infty} \sup_{|t| \le n} |\lambda(t) - t|$.

Having introduced the metric space $D(\mathbb{R}^d)$ we can now explain the concept of weak convergence of a random measure on this space with respect to its Skorohod topology (the topology on $D(\mathbb{R}^d)$ induced by the Skorohod distance $d(x, y)$). Let $\mathbb{Q}_a : \Omega \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be a family of random probability measures depending on a parameter $a > 0$ and let $\mathbb{B} : \Omega \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ be another random probability measure. Define

$$\mathbb{Q}_a(t_1, \cdots, t_d) = \mathbb{Q}_a((-\infty, t_1] \times \cdots \times (-\infty, t_d]), \quad (t_1, \cdots, t_d) \in \mathbb{R}^d.$$

Definition 2.7.10. *We say $\mathbb{Q}_a$ converges to $\mathbb{B}$ weakly on $D(\mathbb{R}^d)$ with respect to the Skorohod topology, denoted by $\mathbb{Q}_a \overset{weakly}{\to} \mathbb{B}$ in $D(\mathbb{R}^d)$, if for any bounded continuous (continuous* *with respect to Skorohod topology) functional $f : D(\mathbb{R}^d) \to \mathbb{R}$ we have*

$$\lim_{a\to\infty} \mathbb{E}\left[f(\mathbb{Q}_a(\cdot,\cdots,\cdot))\right] = \mathbb{E}\left[f(\mathbb{B}(\cdot,\cdots,\cdot))\right]. \tag{2.7.11}$$

Before the proofs of our results in the main body of the chapter, we would like to give the mean and variance of the stick-breaking random measure P as defined in (2.1)-(2.2). For any $A \in \mathcal{X}$,

$$\begin{aligned}
\mathbb{E}\,(P(A)) &= \mathbb{E}\left[\sum_{i=1}^{\infty} w_i \delta_{\theta_i}(A)\right] = \sum_{i=1}^{\infty} \mathbb{E}\,(w_i)\mathbb{E}\,\left[\mathbb{1}_A(\theta_i)\right] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\,(w_i) H(A) = \mathbb{E}\left[\sum_{i=1}^{\infty} w_i\right] H(A) = H(A),
\end{aligned} \tag{2.7.12}$$

since $\sum_{i=1}^{\infty} w_i = 1$ a.s. and

$$\begin{aligned}
\mathrm{Var}\left[P(A)\right] &= \mathbb{E}\left[(P(A)-H(A))^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{\infty} w_i\left(\delta_{\theta_i}(A)-H(A)\right)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\left(\delta_{\theta_i}(A)-H(A)\right)^2\right] \\
&\quad + 2\mathbb{E}\left[\sum_{1\le i<j<\infty} w_i w_j \left(\delta_{\theta_i}(A)-H(A)\right)\left(\delta_{\theta_j}(A)-H(A)\right)\right] \\
&= \mathbb{E}\left[\left(\delta_{\theta_i}(A)-H(A)\right)^2\right]\mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\right] \\
&= H(A)(1-H(A))\mathbb{E}\left[\sum_{i=1}^{\infty} w_i^2\right].
\end{aligned} \tag{2.7.13}$$

2.7.2 Proof of Proposition 2.3.1

Proof of Proposition 2.3.1. Using the binomial expansion and using the fact that $v_i \in [0,1]$ we have

$$\mathbb{E}\left[v_i^n(1-v_i)^m\right] = \sum_{k=0}^{m} \binom{m}{k}(-1)^k \mathbb{E}[v_i^{n+k}]$$

$$= \mathbb{E}[v_i^n] - m\mathbb{E}[v_i^{n+1}] + \cdots + (-1)^m \mathbb{E}[v_i^{m+n}]$$

$$= \mathbb{E}[v_1^n] + O\left(\mathbb{E}[v_1^{n+1}]\right) = \mathbb{E}[v_1^n] + o\left(\mathbb{E}[v_1^n]\right) , \tag{2.7.14}$$

where the last equality follows from the assumption $\lim_{a\to\infty} \frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]} = 0$ for all $n \in \mathbb{Z}_+$. Since $v_i \in [0,1]$ and since we assume that v_0 is not identically zero, we have $\mathbb{E}[(1-v_i)^m] \in [0,1)$ and

$$\begin{aligned}
\sum_{j=0}^{\infty} \left(\mathbb{E}\left[(1-v_i)^m\right]\right)^j &= \frac{1}{1-\mathbb{E}\left[(1-v_i)^m\right]} \\
&= \frac{1}{1-\sum_{j=0}^{m}\binom{m}{j}(-1)^j\mathbb{E}[v_i^j]} \\
&= \frac{1}{m\mathbb{E}[v_1] + \sum_{j=2}^{m}\binom{m}{j}(-1)^j\mathbb{E}[v_i^j]} \\
&= \frac{1}{m\mathbb{E}[v_1]} + o\left(\frac{1}{m\mathbb{E}[v_1]}\right) ,
\end{aligned} \tag{2.7.15}$$

where the last equality also follows from the assumption $\lim_{a\to\infty} \frac{\mathbb{E}[v_1^{n+1}]}{\mathbb{E}[v_1^n]} = 0$ for all $n \in \mathbb{Z}_+$. This proves (3.1)-(3.2).

Now we use (3.1)-(3.2) to show (3.3). Denote

$$\mathcal{I} = \mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \cdots w_{i_k}^{p_k}\right] .$$

By the construction of the stick-breaking sequence $\{w_i\}_{i=1}^{\infty}$, we may rewrite $\mathcal{I}$ as

$$\mathcal{I} = \sum_{1\le i_1<i_2<\cdots<i_k<\infty} \mathbb{E}\left[v_{i_1}^{p_1} \prod_{\ell_1=1}^{i_1-1}(1-v_{\ell_1})^{p_1} \cdots v_{i_m}^{p_m} \prod_{\ell_m=1}^{i_m-1}(1-v_{\ell_m})^{p_m} \cdots v_{i_k}^{p_k} \prod_{\ell_k=1}^{i_k-1}(1-v_{\ell_k})^{p_k}\right] .$$

Since $1 \le i_1 < i_2 < \cdots < i_k < \infty$, we can rearrange $\mathcal{I}$ by putting v's with the same index together to obtain

$$\mathcal{I} = \sum_{1\le i_1<i_2<\cdots<i_k<\infty} \mathbb{E}\Bigg[v_{i_1}^{p_1}(1-v_{i_1})^{p_{2:k}} \prod_{\ell_1=1}^{i_1-1}(1-v_{\ell_1})^{p_{1:k}} v_{i_2}^{p_2}(1-v_{i_2})^{p_{3:k}}$$

$$\prod_{\ell_2=i_1+1}^{i_2-1}(1-v_{\ell_2})^{p_{2:k}}\cdots v_{i_m}^{p_m}(1-v_{i_m})^{p_{m+1:k}} \prod_{\ell_m=i_{m-1}+1}^{i_m-1}(1-v_{\ell_m})^{p_{m:k}} \tag{2.7.16}$$

$$\cdots v_{i_{k-1}}^{p_{k-1}}(1-v_{i_{k-1}})^{p_k} \prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1}(1-v_{\ell_{k-1}})^{p_{k-1:k}} v_{i_k}^{p_k} \prod_{\ell_k=i_{k-1}+1}^{i_k-1}(1-v_{\ell_k})^{p_k}\Bigg].$$

From the independence of $\{v_1, v_2, \cdots\}$ it follows

$$\mathcal{I} = \sum_{i_1=1}^{\infty}\mathbb{E}\left[v_{i_1}^{p_1}(1-v_{i_1})^{p_{2:k}}\right]\prod_{\ell_1=1}^{i_1-1}\mathbb{E}\left[(1-v_{\ell_1})^{p_{1:k}}\right]\sum_{i_2=i_1+1}^{\infty}\mathbb{E}\left[v_{i_2}^{p_2}(1-v_{i_2})^{p_{3:k}}\right]$$

$$\prod_{\ell_2=i_1+1}^{i_2-1}\mathbb{E}\left[(1-v_{\ell_2})^{p_{2:k}}\right]\cdots\sum_{i_m=i_{m-1}+1}^{\infty}\mathbb{E}\left[v_{i_m}^{p_m}(1-v_{i_m})^{p_{m+1:k}}\right]$$

$$\prod_{\ell_m=i_{m-1}+1}^{i_m-1}\mathbb{E}\left[(1-v_{\ell_m})^{p_{m:k}}\right]\cdots\sum_{i_{k-1}=i_{k-2}+1}^{\infty}\mathbb{E}\left[v_{i_{k-1}}^{p_{k-1}}\left(1-v_{i_{k-1}}\right)^{p_k}\right]$$

$$\prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1}\mathbb{E}\left[\left(1-v_{\ell_{k-1}}\right)^{p_{k-1:k}}\right]\sum_{i_k=i_{k-1}+1}^{\infty}\mathbb{E}\left[v_{i_k}^{p_k}\right]\prod_{\ell_k=i_{k-1}+1}^{i_k-1}\mathbb{E}\left[(1-v_{\ell_k})^{p_k}\right].$$

Denoting the general factor in the above expression by

$$\mathcal{S}_m = \sum_{i_m=i_{m-1}+1}^{\infty}\mathbb{E}\left[v_{i_m}^{p_m}(1-v_{i_m})^{p_{m+1:k}}\right]\prod_{\ell_m=i_{m-1}+1}^{i_m-1}\mathbb{E}\left[(1-v_{\ell_m})^{p_{m:k}}\right],$$

for $m \in \{1, 2, \cdots, k\}$, we can write

$$\mathcal{I} = \mathcal{S}_1\mathcal{S}_2\cdots\mathcal{S}_k\,. \tag{2.7.17}$$

From the fact that $\{v_1, v_2, \cdots\}$ are identically distributed and by (3.1)-(3.2) we have for

$m = 1, \cdots, k,$

$$
\begin{aligned}
\mathcal{S}_m &= \mathbb{E}\left[v_1^{p_m}(1-v_1)^{p_{m+1:k}}\right] \sum_{i_m=i_{m-1}+1}^{\infty} \left(\mathbb{E}\left[(1-v_1)^{p_{m:k}}\right]\right)^{i_m-i_{m-1}-1} \\
&= \left(\mathbb{E}\left[v_1^{p_m}\right] + o\left(\mathbb{E}\left[v_1^{p_m}\right]\right)\right)\left(\frac{1}{p_{m:k}\mathbb{E}[v_1]} + o\left(\frac{1}{\mathbb{E}[v_1]}\right)\right) \\
&= \frac{\mathbb{E}[v_1^{p_m}]}{p_{m:k}\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^{p_m}]}{p_{m:k}\mathbb{E}[v_1]}\right).
\end{aligned}
$$

Substituting this estimate into (2.7.17), we see

$$
\begin{aligned}
\mathcal{I} &= \prod_{m=1}^{k}\left(\frac{\mathbb{E}[v_1^{p_m}]}{p_{m:k}\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^{p_m}]}{p_{m:k}\mathbb{E}[v_1]}\right)\right) \\
&= \frac{\mathbb{E}[v_1^{p_1}]\cdots\mathbb{E}[v_1^{p_m}]}{p_{1:k}p_{2:k}\cdots p_{k:k}(\mathbb{E}[v_1])^k} + o\left(\frac{\mathbb{E}[v_1^{p_1}]\cdots\mathbb{E}[v_1^{p_m}]}{(\mathbb{E}[v_1])^k}\right).
\end{aligned}
$$

This prove (3.3). If we take $p_j = 2$ for all $j \in \{1, \cdots, k\}$, then $p_{1:k} = 2k$. The identity (3.4) is hence a straightforward consequence of (3.3). □

2.7.3 Proof of Proposition 2.3.2

Since $v_1, v_2, \cdots$ are no longer identically distributed, the results established in the previous proof cannot be applied and we need some new computations. We shall still use the general method of moments. To this end, we need to recall some results about the hypergeometric functions and we refer to (Aomoto et al., 2011) for further reading.

Definition 2.7.11. *The hypergeometric function* ${}_2F_1(a, b, c; x)$ *(of parameters* $a, b, c \in \mathbb{C}$*, the complex plane) is defined by the series*

$$
{}_2F_1(a, b, c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} x^n
$$

for $|x| < 1$*, where* $(q)_n$ *is the Pochhammer symbol defined by*

$$
(q)_n = \begin{cases} 0 & \textit{for} \quad n = 0 \\ q(q+1)\dots(q+n-1) & \textit{for} \quad n > 0\,. \end{cases}
$$

This function is defined for $|x| < 1$ and may be extended to $x = 1$ and/or $x = -1$ by continuation.

We need the following result obtained by Gauss: when $\mathrm{Re}(c - a - b) > 0$ (real part of $c - a - b$), the hypergeometric function can be extended to $x = 1$ and its value at this point is given by

$$ {}_2F_1(a, b, c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)}. \tag{2.7.18} $$

We introduce a variant of the hypergeometric function that will be needed in the following calculations.

Definition 2.7.12. *For any $b < 2$, $n \in \mathbb{N}^+$, $a > 0$, $m > 0$, $c > 0$, define the increasing coefficient hypergeometric function ${}_2Q_1\left((a, b), c, m, n; x\right)$ by the series*

$$ {}_2Q_1\left((a, b), c, m, n; x\right) = \sum_{k=0}^{\infty} \prod_{\ell=1}^{n-1} (a + b(k + \ell)) \frac{\left(\frac{a}{b} + 1\right)_k (c)_k}{\left(\frac{a+m}{b} + 1\right)_k k!} x^k $$

for $|x| < 1$ and we may extend the definition to $x = 1$ and/or $x = -1$ by continuation. In the above product we use the convention that $\prod_{\ell=1}^{0} c_\ell = 1$.

The next proposition describes a Gauss type result for the increasing coefficient hypergeometric function.

Proposition 2.7.13. *Let $b < 2$, $n \in \mathbb{N}^+$, $a > 0$, $m > 0$, $c > 0$. Then, the increasing coefficient hypergeometric function can be extended to $x = 1$ and its value at this point is given by*

$$ {}_2Q_1\left((a, b), c, m, n; 1\right) = \prod_{\ell=1}^{n-1} (a + b\ell) \frac{a + m}{m - nb}. $$

Proof. By (2.7.18) we have

$$ {}_2Q_1\left((a, b), c, m, n; 1\right) = \prod_{\ell=1}^{n-1} (a + b\ell) \sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{a + (n + i - 1)b}{a + ib + m} $$

$$ = \prod_{\ell=1}^{n-1} (a + b\ell) \ {}_2F_1\left(\frac{a}{b} + n, 1, \frac{a + m}{b} + 1; 1\right) = \prod_{\ell=1}^{n-1} (a + b\ell) \frac{a + m}{m - nb}, \tag{2.7.19} $$

proving the proposition. □

Now we are in the position of proving Proposition 3.2.

Proof of Proposition 3.2 Denote

$$\mathcal{S}_m = \sum_{i_m=i_{m-1}+1}^{\infty} \mathbb{E}\left[v_{i_m}^{p_m}(1-v_{i_m})^{p_{m+1:k}}\right] \prod_{\ell_m=i_{m-1}+1}^{i_m-1} \mathbb{E}\left[(1-v_{\ell_m})^{p_{m:k}}\right],$$

for $m \in \{1, 2, \cdots, k\}$. Then we can write

$$\mathcal{I} := \mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \cdots w_{i_k}^{p_k}\right] = \mathcal{S}_1\mathcal{S}_2\cdots\mathcal{S}_k\,.$$

We shall compute $\mathcal{I}$ by computing $\mathcal{S}_m\mathcal{S}_{m+1}\cdots\mathcal{S}_k$ recursively on $m = k, k-1, \cdots, 2, 1$. First, by using (2.7.18) we have

$$\begin{aligned}\mathcal{S}_k &= \sum_{i_k=i_{k-1}+1}^{\infty} \frac{(1-b)_{p_k}}{(1+a+b(i_k-1))_{p_k}} \prod_{\ell_k=i_{k-1}+1}^{i_k-1} \frac{(a+b\ell_k)_{p_k}}{(1+a+b(\ell_k-1))_{p_k}}\\ &= \frac{(1-b)_{p_k}}{(1+a+bi_{k-1})_{p_k}}\ {}_2F_1\left(\frac{a}{b}+i_{k-1}+1, 1, \frac{a+p_k}{b}+i_{k-1}+1; 1\right)\\ &= \frac{(1-b)_{p_k}}{(p_k-b)(1+a+bi_{k-1})_{p_k-1}}.\end{aligned} \tag{2.7.20}$$

Now we want to compute $\mathcal{S}_m\mathcal{S}_{m+1}\cdots\mathcal{S}_k$ assuming that we have already computed $\mathcal{S}_{m+1}\cdots\mathcal{S}_k$. To make thing clear we will explain how to compute $\mathcal{S}_1\cdots\mathcal{S}_k$ from the expression of $\mathcal{S}_2\cdots\mathcal{S}_k$. General case is similar. We assume

$$\mathcal{S}_2\cdots\mathcal{S}_k = \frac{(a+b(i_1+1))\cdots(a+b(i_1+k-2))}{(1+a+bi_1)_{p_{2:k}-1}} \prod_{i=2}^{k} \frac{(1-b)_{p_i}}{p_{i:k}-(k-i+1)b}.$$

Then, by Proposition 2.7.13

$$
\begin{aligned}
\mathcal{I} = \mathcal{S}_1\mathcal{S}_2\cdots\mathcal{S}_k &= (1-b)_{p_1}\prod_{i=2}^{k}\frac{(1-b)_{p_i}}{p_{i:k}-(k-i+1)b} \\
&\quad \sum_{i=1}^{\infty}\frac{(a+bi_1)\cdots(a+b(i_1+k-2))}{(1+a+b(i_1-1))_{p_{1:k}}}\prod_{\ell_1=1}^{i_1-1}\frac{(a+b\ell_1)_{p_{1:k}}}{(1+a+b(\ell_1-1))_{p_{1:k}}} \\
&= \frac{(1-b)_{p_1}}{(a+1)_{p_{1:k}}}\prod_{i=2}^{k}\frac{(1-b)_{p_i}}{p_{i:k}-(k-i+1)b}\;{}_2Q_1\left((a,b),1,p_{1:k},k;1\right) \\
&= \frac{(1-b)_{p_1}}{(a+1)_{p_{1:k}}}\prod_{i=2}^{k}\frac{(1-b)_{p_i}}{p_{i:k}-(k-i+1)b}\prod_{\ell=1}^{k-1}(a+b\ell)\frac{a+p_{1:k}}{p_{1:k}-kb} \\
&= \frac{(a+b)\cdots(a+b(k-1))}{(1+a)_{p_{1:k}-1}}\prod_{i=1}^{k}\frac{(1-b)_{p_i}}{p_{i:k}-(k-i+1)b} \\
&= \frac{1}{(a+kb)(a+1)_{(p_{1:k}-1)}}\prod_{i=1}^{k}\frac{(1-b)_{p_i}(a+bi)}{p_{i:k}-(k-i+1)b}.
\end{aligned}
$$

This proves (3.5).

When $p_j = 2$ for all $j \in \{1,\cdots,n\}$, we have easily

$$
\mathcal{I} = \mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_n<\infty} w_{i_1}^2 w_{i_2}^2\cdots w_{i_n}^2\right] = \frac{(1-b)^n(a+b)\cdots(a+b(n-1))}{n!(a+1)\cdots(a+2n-1)}
$$

proving (3.6). □

2.7.4 Proof of Proposition 2.3.3

Proof of Proposition 3.3. By the stick-breaking representation of N-IG(a,H) and the formula (3.471.9) in (Gradshteyn and Ryzhik, 2014), we find that the joint distribution of $\{v_i\}_{i=1}^{n}$ can be written as

$$
f(v_1,\cdots,v_n) = \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}}\prod_{i=1}^{n} v_i^{-\frac{1}{2}}(1-v_i)^{-\frac{n+3-i}{2}}\int_0^{\infty} t^{-\frac{n+3}{2}} e^{-\frac{t}{2}-\frac{a^2}{2t\prod_{i=1}^{n}(1-v_i)}}\,dt. \qquad (2.7.21)
$$

Thus, by using Fubini's theorem, we have

$$\begin{aligned}\mathcal{I} =& \mathbb{E}\left[\sum_{n=1}^{\infty} w_n^p\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} v_n^p \prod_{i=1}^{n-1}(1-v_i)^p\right] \\ =& \sum_{n=1}^{\infty}\int_0^{\infty}\int_0^1\cdots\int_0^1 \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}}\left(\prod_{i=1}^{n-1} v_i^{-\frac{1}{2}}(1-v_i)^{p-\frac{n+3-i}{2}}\right) v_n^{p-\frac{1}{2}}(1-v_n)^{-\frac{3}{2}} \\ & \qquad t^{-\frac{n+3}{2}} e^{-\frac{t}{2}-\frac{a^2}{2t\prod_{i=1}^{n}(1-v_i)}}\, dv_1\cdots dv_n dt\,. \end{aligned} \tag{2.7.22}$$

To evaluate the above multiple integral, we shall use the formula (3.471.2) in (Gradshteyn and Ryzhik, 2014):

$$\int_0^1 (1-v)^{\eta-1} v^{\mu-1} e^{-\frac{\beta}{1-v}}\, dv = \Gamma(\mu)\beta^{\frac{\eta-1}{2}} e^{-\frac{\beta}{2}} W_{\frac{1-2\mu-\eta}{2},\frac{\eta}{2}}(\beta), \tag{2.7.23}$$

where W is the Whittaker function. For large β, by the formula (9.227) in (Gradshteyn and Ryzhik, 2014), we have

$$\int_0^1 (1-v)^{\eta-1} v^{\mu-1} e^{-\frac{\beta}{1-v}}\, dv = \Gamma(\mu)\beta^{-\mu} e^{-\beta}\left(1+o\left(\frac{1}{\beta}\right)\right), \tag{2.7.24}$$

as $\beta\to\infty$. In particular, when $\mu=\frac{1}{2}$, we have

$$\int_0^1 (1-v)^{\eta-1} v^{-\frac{1}{2}} e^{-\frac{\beta}{1-v}}\, dv = \Gamma\left(\frac{1}{2}\right)\beta^{-\frac{1}{2}} e^{-\beta}\left(1+o\left(\frac{1}{\beta}\right)\right). \tag{2.7.25}$$

Denote $\beta_i = \frac{a^2}{2t(1-v_n)\prod_{\ell=1}^{i}(1-v_\ell)}$ for $i\in\{1,\cdots,n-1\}$. We rewrite (2.7.22) as

$$\begin{aligned}\mathcal{I} =& \sum_{n=1}^{\infty}\int_0^{\infty}\int_0^1\cdots\int_0^1 \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}} t^{-\frac{n+3}{2}} e^{-\frac{t}{2}}\left(\prod_{i=1}^{n-2} v_i^{-\frac{1}{2}}(1-v_i)^{p-\frac{n+3-i}{2}}\right) \\ & v_n^{p-\frac{1}{2}}(1-v_n)^{-\frac{3}{2}}\int_0^1 v_{n-1}^{-\frac{1}{2}}(1-v_{n-1})^{p-\frac{4}{2}} e^{-\frac{\beta_{n-2}}{(1-v_{n-1})}}\, dv_{n-1}dv_1\cdots dv_{n-2}dv_n dt. \end{aligned}$$

Integrating with respect to v_{n-1} by applying (2.7.25) yields

$$\mathcal{I} = \sum_{n=1}^{\infty}\int_0^{\infty}\int_0^1\cdots\int_0^1 \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}}\Gamma\left(\frac{1}{2}\right)\left(\frac{a^2}{2t(1-v_n)}\right)^{-\frac{1}{2}}$$

$$\left(1+o\left(\frac{1}{a}\right)\right) t^{-\frac{n+3}{2}} e^{-\frac{t}{2}} (1-v_n)^{-\frac{3}{2}} v_n^{p-\frac{1}{2}} \left(\prod_{i=1}^{n-2} v_i^{-\frac{1}{2}} (1-v_i)^{p-\frac{n+2-i}{2}}\right)$$
$$\int_0^1 v_{n-2}^{-\frac{1}{2}} (1-v_{n-2})^{p-\frac{4}{2}} e^{-\frac{\beta_{n-3}}{(1-v_{n-2})}} dv_1 \cdots dv_{n-2} dv_n dt.$$

We repeatedly apply the above procedure to integrate $v_{n-2}, v_{n-3}, \cdots, v_1$, each time using (2.7.25). After these computations we obtain

$$\begin{aligned}
\mathcal{I} &= \sum_{n=1}^{\infty} \int_0^{\infty} \int_0^1 \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}} \left(\Gamma\left(\frac{1}{2}\right)\left(\frac{a^2}{2t(1-v_n)}\right)^{-\frac{1}{2}}\right)^{n-1} \left(1+o\left(\frac{1}{a}\right)\right) \\
&\qquad \times t^{-\frac{n+3}{2}} e^{-\frac{t}{2}} v_n^{p-\frac{1}{2}} (1-v_n)^{-\frac{3}{2}} dv_n dt \\
&= \sum_{n=1}^{\infty} \int_0^{\infty} \int_0^1 \frac{e^a a^2}{2\pi} v^{p-\frac{1}{2}} (1-v)^{\frac{n-4}{2}} e^{-\frac{a^2}{2t(1-v)}} t^{-2} e^{-\frac{t}{2}} \left(1+o\left(\frac{1}{a}\right)\right) dv dt. \qquad (2.7.26)
\end{aligned}$$

Again by using Fubini's theorem and by the fact that $v \in (0,1)$, we take the sum to obtain

$$\begin{aligned}
\mathcal{I} &= \int_0^{\infty} \int_0^1 \frac{e^a a^2}{2\pi} v^{p-\frac{1}{2}} \left(\sum_{n=1}^{\infty} (1-v)^{\frac{n-4}{2}}\right) e^{-\frac{a^2}{2t(1-v)}} t^{-2} e^{-\frac{t}{2}} \left(1+o\left(\frac{1}{a}\right)\right) dv dt \\
&= \int_0^{\infty} \int_0^1 \frac{e^a a^2}{2\pi} v^{p-\frac{3}{2}} \left(1+(1-v)^{\frac{1}{2}}\right) e^{-\frac{a^2}{2t(1-v)}} t^{-2} e^{-\frac{t}{2}} \left(1+o\left(\frac{1}{a}\right)\right) dv dt. \qquad (2.7.27)
\end{aligned}$$

Then, the result (3.7) is obtained by first applying (2.7.24) when we integrate the integral with respect to v, and then by applying the formula (3.471.9) in (Gradshteyn and Ryzhik, 2014) when we integrate t. Here, we also use the approximation that for fixed ν and for large a, $K_\nu(a) = \sqrt{\frac{\pi}{2}} a^{-1/2} e^{-a} \left(1+o\left(\frac{1}{a}\right)\right)$.

When $p = 2$, we want to show that the leading coefficient in (3.7) is 1. This needs some more delicate computations. First, we have

$$\begin{aligned}
\mathcal{I} = \sum_{n=1}^{\infty} \int_0^{\infty} \int_0^1 \cdots \int_0^1 \frac{e^a a^{n+1}}{(2\pi)^{\frac{n+1}{2}}} \left(\prod_{i=1}^{n-1} v_i^{-\frac{1}{2}} (1-v_i)^{-\frac{n-1-i}{2}}\right) v_n^{\frac{3}{2}} (1-v_n)^{-\frac{3}{2}} \\
\times t^{-\frac{n+3}{2}} e^{-\frac{t}{2} - \frac{a^2}{2t \prod_{i=1}^{n}(1-v_i)}} dv_1 \cdots dv_n dt.
\end{aligned}$$

Notice the fact that the power of v_{n-4} in the above integrand is $-\frac{3}{2}$ and to compute the integral with respect to v_{n-4} we can use the following nice integral identity:

$$\int_0^1 (1-v)^{-\frac{3}{2}} v^{-\frac{1}{2}} e^{-\frac{\beta}{1-v}} dv = \Gamma(\frac{1}{2})\beta^{-\frac{1}{2}} e^{-\beta}. \tag{2.7.28}$$

After this integration with respect to v_{n-4}, we obtain an expression for v_{n-5} which also has this form and we then integrate v_{n-5} and so on. This procedure can continue until integrating v_1. Hence, we compute the integrals for v_{n-4}, and then for v_{n-5}, $\cdots$ and then for v_1 recursively to obtain

$$\begin{aligned}\mathcal{I} = \sum_{n=1}^{\infty} \int_0^{\infty} \int_0^1 \int_0^1 \int_0^1 \int_0^1 & \frac{e^a a^5}{(2\pi)^{\frac{5}{2}}} (v_{n-3} v_{n-2} v_{n-1})^{-\frac{1}{2}} v_n^{\frac{3}{2}} \\ & \times (1-v_{n-3})^{\frac{n-6}{2}} (1-v_{n-2})^{\frac{n-5}{2}} (1-v_{n-1})^{\frac{n-4}{2}} (1-v_n)^{\frac{n-7}{2}} \\ & \times e^{-\frac{a^2}{2t(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}} t^{-\frac{7}{2}} e^{-\frac{t}{2}} dv_{n-3} dv_{n-2} dv_{n-1} dv_n dt. \end{aligned} \tag{2.7.29}$$

By Fubini's theorem and by the fact that $v_i \in (0,1)$, we can take the sum first [There is no need to sum up the index n in $v_n, v_{n-1}, v_{n-2}, v_{n-3}$ since we can call them by other notations. But for consistency we still keep these notations.]

$$\begin{aligned}\mathcal{I} = \int_0^{\infty} \int_0^1 \int_0^1 \int_0^1 \int_0^1 & \frac{e^a a^5}{(2\pi)^{\frac{5}{2}}} (v_{n-3} v_{n-2} v_{n-1})^{-\frac{1}{2}} v_n^{\frac{3}{2}} (1-v_{n-3})^{-\frac{5}{2}} (1-v_{n-2})^{-\frac{4}{2}} \\ & \times (1-v_{n-1})^{-\frac{3}{2}} (1-v_n)^{-\frac{6}{2}} \frac{1+\sqrt{(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}}{1-(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)} \\ & \times e^{-\frac{a^2}{2t(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}} t^{-\frac{7}{2}} e^{-\frac{t}{2}} dv_{n-3} dv_{n-2} dv_{n-1} dv_n dt. \end{aligned} \tag{2.7.30}$$

Now we integrate t by using the formula (3.471.9) in (Gradshteyn and Ryzhik, 2014) to obtain

$$\begin{aligned}\mathcal{I} = \int_0^1 \int_0^1 \int_0^1 \int_0^1 & \frac{2e^a a^{\frac{5}{2}}}{(2\pi)^{\frac{5}{2}}} (v_{n-3} v_{n-2} v_{n-1})^{-\frac{1}{2}} v_n^{\frac{3}{2}} (1-v_{n-3})^{-\frac{5}{4}} (1-v_{n-2})^{-\frac{3}{4}} \\ & (1-v_{n-1})^{-\frac{1}{4}} (1-v_n)^{-\frac{7}{4}} \frac{1+\sqrt{(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}}{1-(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)} \end{aligned}$$

$$K_{-\frac{5}{2}}\left(\frac{a}{\sqrt{(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}}\right) dv_{n-3}dv_{n-2}dv_{n-1}dv_n\,.$$

Approximating the above modified Bessel function $K_{-\frac{5}{2}}$ of the third type by the formula (8.451.6) in Gradshteyn and Ryzhik (2014) we have

$$\begin{aligned}\mathcal{I} = &\int_0^1\int_0^1\int_0^1\int_0^1 \frac{e^a a^2}{(2\pi)^2}(v_{n-3}v_{n-2}v_{n-1})^{-\frac{1}{2}} v_n^{\frac{3}{2}}(1-v_{n-3})^{-\frac{2}{2}}(1-v_{n-2})^{-\frac{1}{2}}\\ &\times (1-v_{n-1})^{-\frac{0}{2}}(1-v_n)^{-\frac{3}{2}}\frac{1+\sqrt{(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}}{1-(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}\\ &\times e^{-\frac{a^2}{\sqrt{(1-v_{n-3})(1-v_{n-2})(1-v_{n-1})(1-v_n)}}}\left(1+o\left(\frac{1}{a}\right)\right) dv_{n-3}dv_{n-2}dv_{n-1}dv_n.\end{aligned}$$

To evaluate the above integral we make the following variable substitutions.

$$\begin{cases} v_n = 1-\frac{a^2}{(y_0+a)^2}\,; \\ v_{n-1} = 1-\frac{(y_0+a)^2}{(y_0+y_1+a)^2}\,; \\ v_{n-2} = 1-\frac{(y_0+y_1+a)^2}{(y_0+y_1+y_2+a)^2}\,; \\ v_{n-3} = 1-\frac{(y_0+y_1+y_2+a)^2}{(y_0+y_1+y_2+y_3+a)^2}\,. \end{cases}$$

The integral $\mathcal{I}$ can then be written as

$$\begin{aligned}\mathcal{I} = &\int_0^1\int_0^1\int_0^1\int_0^1 \frac{e^a a^2}{(2\pi)^2}\left(1-\frac{a^2}{(y_0+a)^2}\right)^{\frac{3}{2}}\left(1-\frac{(y_0+a)^2}{(y_0+y_1+a)^2}\right)^{-\frac{1}{2}}\\ &\times\left(1-\frac{(y_0+y_1+a)^2}{(y_0+y_1+y_2+a)^2}\right)^{-\frac{1}{2}}\left(1-\frac{(y_0+y_1+y_2+a)^2}{(y_0+y_1+y_2+y_3+a)^2}\right)^{-\frac{1}{2}}\\ &\times\left(\frac{(y_0+y_1+y_2+a)^2}{(y_0+y_1+y_2+y_3+a)^2}\right)^{-\frac{2}{2}}\left(\frac{(y_0+y_1+a)^2}{(y_0+y_1+y_2+a)^2}\right)^{-\frac{1}{2}}\\ &\times\left(\frac{(y_0+a)^2}{(y_0+y_1+a)^2}\right)^{-\frac{0}{2}}\left(\frac{a^2}{(y_0+a)^2}\right)^{-\frac{3}{2}}\frac{1+\frac{a}{(y_0+y_1+y_2+y_3+a)}}{1-\frac{a^2}{(y_0+y_1+y_2+y_3+a)^2}}\\ &\times\frac{2^4a^2}{(y_0+a)(y_0+y_1+a)(y_0+y_1+y_2+a)(y_0+y_1+y_2+y_3+a)^3}\\ &e^{-y_0-y_1-y_2-y_3-a}\left(1+o\left(\frac{1}{a}\right)\right)dy_0dy_1dy_2dy_3. \end{aligned} \tag{2.7.31}$$

When a is large, we have the following asymptotics:

$$\begin{aligned}
&1-\frac{a^2}{(y_0+a)^2}=\frac{2ay_0+y_0^2}{a^2+2ay_0+y_0^2}=\frac{2y_0}{a}+o\left(\frac{1}{a}\right) \ ;\\
&1-\frac{(y_0+a)^2}{(y_0+y_1+a)^2}=\frac{2y_1}{a}+o\left(\frac{1}{a}\right) \ ;\\
&1-\frac{(y_0+y_1+a)^2}{(y_0+y_1+y_2+a)^2}=\frac{2y_2}{a}+o\left(\frac{1}{a}\right) \ ;\\
&1-\frac{(y_0+y_1+y_2+a)^2}{(y_0+y_1+y_2+y_3+a)^2}=\frac{2y_3}{a}+o\left(\frac{1}{a}\right) \ ;\\
&\frac{a^2}{(y_0+a)^2}=\frac{(y_0+a)^2}{(y_0+y_1+a)^2}=\frac{(y_0+y_1+a)^2}{(y_0+y_1+y_2+a)^2} \, ;\\
&\qquad\qquad=\frac{(y_0+y_1+y_2+a)^2}{(y_0+y_1+y_2+y_3+a)^2}=1+o\left(\frac{1}{a}\right) \ ;\\
&1+\frac{a}{(y_0+y_1+y_2+y_3+a)}=2+o\left(\frac{1}{a}\right) \ ;\\
&1-\frac{a^2}{(y_0+y_1+y_2+y_3+a)^2}=\frac{2(y_0+y_1+y_2+y_3)}{a}+o\left(\frac{1}{a}\right) \ ;\\
&\frac{2^4a^2}{(y_0+a)(y_0+y_1+a)(y_0+y_1+y_2+a)(y_0+y_1+y_2+y_3+a)^3}=\frac{2^4}{a^4}+o\left(\frac{1}{a^4}\right).
\end{aligned}$$

Substituting the above asymptotics into (2.7.31), we see that when a is large the integral $\mathcal{I}$ is

$$\begin{aligned}
\mathcal{I}=&\int_0^1\int_0^1\int_0^1\int_0^1 \frac{e^a a^2}{(2\pi)^2}\left(\frac{2y_0}{a}\right)^{\frac{3}{2}}\left(\frac{2y_1}{a}\right)^{-\frac{1}{2}}\left(\frac{2y_2}{a}\right)^{-\frac{1}{2}}\left(\frac{2y_3}{a}\right)^{-\frac{1}{2}}\\
&\qquad\times\frac{a}{(y_0+y_1+y_2+y_3)}\frac{2^4}{a^4}e^{-y_0-y_1-y_2-y_3-a}\left(1+o\left(\frac{1}{a}\right)\right)dy_0dy_1dy_2dy_3\\
=&\int_0^1\int_0^1\int_0^1\int_0^1 \frac{4}{a\pi^2}\frac{y_0^{\frac{3}{2}}y_1^{-\frac{1}{2}}y_2^{-\frac{1}{2}}y_3^{-\frac{1}{2}}}{y_0+y_1+y_2+y_3}\\
&\qquad\times e^{-y_0-y_1-y_2-y_3}\left(1+o\left(\frac{1}{a}\right)\right)dy_0dy_1dy_2dy_3\,. \qquad (2.7.32)
\end{aligned}$$

The above integral can be further evaluated by making the following variable substitutions:

$$\begin{cases} y_0 = t_0\,; \\ y_0 + y_1 = t_1\,; \\ y_0 + y_1 + y_2 = t_2\,; \\ y_0 + y_1 + y_2 + y_3 = t_3\,. \end{cases}$$

With these substitutions, (2.7.32) can be written as

$$\mathcal{I} = \int_0^{\infty}\int_0^{t_3}\int_0^{t_2}\int_0^{t_1} \frac{4}{a\pi^2}\frac{t_0^{\frac{3}{2}}(t_1-t_0)^{-\frac{1}{2}}(t_2-t_1)^{-\frac{1}{2}}(t_3-t_2)^{-\frac{1}{2}}}{t_3} \times e^{-t_3}\left(1+o\left(\frac{1}{a}\right)\right) dt_0dt_1dt_2dt_3.$$

Now we integrate t_0 by letting $t_0 = ut_1$,

$$\begin{aligned} \mathcal{I} &= \int_0^{\infty}\int_0^{t_3}\int_0^{t_2}\int_0^{1} \frac{4}{a\pi^2}\frac{t_1(ut_1)^{\frac{3}{2}}(t_1-ut_1)^{-\frac{1}{2}}(t_2-t_1)^{-\frac{1}{2}}(t_3-t_2)^{-\frac{1}{2}}}{t_3} \\ &\qquad \times e^{-t_3}\left(1+o\left(\frac{1}{a}\right)\right) dudt_1dt_2dt_3 \\ &= \int_0^{\infty}\int_0^{t_3}\int_0^{t_2} \frac{4}{a\pi^2}\frac{\Gamma(\frac{5}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{6}{2})}\frac{t_1^2(t_2-t_1)^{-\frac{1}{2}}(t_3-t_2)^{-\frac{1}{2}}}{t_3} \\ &\qquad \times e^{-t_3}\left(1+o\left(\frac{1}{a}\right)\right) dt_1dt_2dt_3. \end{aligned}$$

Similarly, we integrate t_1, t_2, t_3 one by one in this order to obtain

$$\begin{aligned} \mathcal{I} &= \frac{4}{a\pi^2}\frac{\Gamma(\frac{5}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{6}{2})}\frac{\Gamma(\frac{6}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{7}{2})}\frac{\Gamma(\frac{7}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{8}{2})}\Gamma(\frac{6}{2})\left(1+o\left(\frac{1}{a}\right)\right) \\ &= \frac{1}{a} + o\left(\frac{1}{a}\right), \end{aligned} \tag{2.7.33}$$

completing the proof of (3.9).

To prove the results (3.8) and (3.10), we denote

$$\mathcal{L} = \mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2} \cdots w_{i_k}^{p_k}\right].$$

We have

$$\begin{aligned}\mathcal{L} = \sum_{1\le i_1<i_2<\cdots<i_k<\infty} \mathbb{E}\Bigg[& v_{i_1}^{p_1}(1-v_{i_1})^{p_{2:k}} \prod_{\ell_1=1}^{i_1-1}(1-v_{\ell_1})^{p_{1:k}} v_{i_2}^{p_2}(1-v_{i_2})^{p_{3:k}} \\ &\times \prod_{\ell_2=i_1+1}^{i_2-1}(1-v_{\ell_2})^{p_{2:k}} \cdots v_{i_m}^{p_m}(1-v_{i_m})^{p_{m+1:k}} \prod_{\ell_m=i_{m-1}+1}^{i_m-1}(1-v_{\ell_m})^{p_{m:k}} \\ &\cdots v_{i_{k-1}}^{p_{k-1}}(1-v_{i_{k-1}})^{p_k} \prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1}(1-v_{\ell_{k-1}})^{p_{k-1:k}} v_{i_k}^{p_k} \prod_{\ell_k=i_{k-1}+1}^{i_k-1}(1-v_{\ell_k})^{p_k}\Bigg].\end{aligned} \tag{2.7.34}$$

Using the explicit form of the joint density of $v_1, \cdots, v_{i_k}$, we have

$$\begin{aligned}\mathcal{L} = \sum_{1\le i_1<i_2<\cdots<i_k<\infty} & \int_0^\infty \int_0^1 \cdots \int_0^1 \frac{e^a a^{i_k+1}}{(2\pi)^{\frac{i_k+1}{2}}} t^{-\frac{i_k+3}{2}} e^{-\frac{t}{2}} e^{-\frac{a^2}{2t\prod_{j=1}^{i_k}(1-v_j)}} \\ &\times \prod_{\ell_1=1}^{i_1-1} v_{\ell_1}^{-\frac{1}{2}}(1-v_{\ell_1})^{p_{1:k}-\frac{i_k+3-\ell_1}{2}} v_{i_1}^{p_1-\frac{1}{2}}(1-v_{i_1})^{p_{2:k}-\frac{i_k+3-i_1}{2}} \\ &\cdots \prod_{\ell_m=i_{m-1}+1}^{i_m-1} v_{\ell_m}^{-\frac{1}{2}}(1-v_{\ell_m})^{p_{m:k}-\frac{i_k+3-\ell_m}{2}} v_{i_m}^{p_m-\frac{1}{2}}(1-v_{i_m})^{p_{m+1:k}-\frac{i_k+3-i_m}{2}} \\ &\cdots \prod_{\ell_k=i_{k-1}+1}^{i_k-1} v_{\ell_k}^{-\frac{1}{2}}(1-v_{\ell_k})^{p_k-\frac{i_k+3-\ell_k}{2}} v_{i_k}^{p_k-\frac{1}{2}}(1-v_{i_k})^{-\frac{3}{2}} dv_1\cdots dv_{i_k} dt.\end{aligned} \tag{2.7.35}$$

Notice that the integrals of $v_{i_{k-1}+1}, v_{i_{k-1}+2}, \cdots v_{i_k}$ with the sum of i_k from $i_{k-1}+1$ to ∞ is the same form as (2.7.22). Thus, by the computation (2.7.26), we have

$$\begin{aligned}\mathcal{L} = \sum_{1\le i_1<i_2<\cdots<i_k<\infty} & \int_0^\infty \int_0^1 \cdots \int_0^1 \frac{e^a a^{i_k+1}}{(2\pi)^{\frac{i_k+1}{2}}} t^{-\frac{i_k+3}{2}} e^{-\frac{t}{2}} e^{-\frac{a^2}{2t(1-v_{i_k})\prod_{j=1}^{i_{k-1}}(1-v_j)}} \\ & \prod_{\ell_1=1}^{i_1-1} v_{\ell_1}^{-\frac{1}{2}}(1-v_{\ell_1})^{p_{1:k}-\frac{i_k+3-\ell_1}{2}} v_{i_1}^{p_1-\frac{1}{2}}(1-v_{i_1})^{p_{2:k}-\frac{i_k+3-i_1}{2}} \cdots\end{aligned}$$

$$\prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1} v_{\ell_{k-1}}^{-\frac{1}{2}}(1-v_{\ell_{k-1}})^{p_k-\frac{i_k+3-\ell_{k-1}}{2}} v_{i_{k-1}}^{p_{k-1}-\frac{1}{2}}(1-v_{i_{k-1}})^{-\frac{3}{2}}$$

$$\left(\Gamma(\frac{1}{2})\left(\frac{a^2}{2t(1-v_{i_k})\prod_{j=1}^{i_{k-1}}(1-v_j)}\right)^{-\frac{1}{2}}\right)^{i_k-i_{k-1}-1}$$

$$\times\left(1+o\left(\frac{1}{a}\right)\right) dv_1\cdots dv_{i_{k-1}} dv_{v_{i_k}} dt \qquad (2.7.36)$$

$$= \sum_{1\le i_1<i_2<\cdots<i_{k-1}<\infty} \sum_{i_k=i_{k-1}+1}^{\infty} \int_0^\infty \int_0^1 \cdots \int_0^1 \frac{e^a a^{i_{k-1}+2}}{(2\pi)^{\frac{i_{k-1}+2}{2}}} t^{-\frac{i_{k-1}+4}{2}} e^{-\frac{t}{2}}$$

$$\prod_{\ell_1=1}^{i_1-1} v_{\ell_1}^{-\frac{1}{2}}(1-v_{\ell_1})^{p_{1:k}-\frac{i_{k-1}+4-\ell_1}{2}} v_{i_1}^{p_1-\frac{1}{2}}(1-v_{i_1})^{p_{2:k}-\frac{i_{k-1}+4-i_1}{2}}\cdots$$

$$\prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1} v_{\ell_{k-1}}^{-\frac{1}{2}}(1-v_{\ell_{k-1}})^{p_k-\frac{i_{k-1}+4-\ell_{k-1}}{2}} v_{i_{k-1}}^{p_{k-1}-\frac{1}{2}}(1-v_{i_{k-1}})^{-\frac{4}{2}}$$

$$v_{i_k}^{p_k}(1-v_{i_k})^{\frac{i_k-i_{k-1}-4}{2}} e^{-\frac{a^2}{2t(1-v_{i_k})\prod_{j=1}^{i_{k-1}}(1-v_j)}}\left(1+o\left(\frac{1}{a}\right)\right) dv_{i_k} dv_1\cdots dv_{i_{k-1}} dt. \qquad (2.7.37)$$

By a similar calculation to that of (2.7.27),

$$\mathcal{L} = \sum_{1\le i_1<i_2<\cdots<i_{k-1}<\infty} \int_0^\infty \int_0^1 \cdots \int_0^1 O\left(\frac{1}{a}\right) \frac{e^a a^{i_{k-1}+1}}{(2\pi)^{\frac{i_{k-1}+1}{2}}} t^{-\frac{i_{k-1}+3}{2}} e^{-\frac{t}{2}}$$

$$\times \prod_{\ell_1=1}^{i_1-1} v_{\ell_1}^{-\frac{1}{2}}(1-v_{\ell_1})^{p_{1:k}-\frac{i_{k-1}+3-\ell_1}{2}} v_{i_1}^{p_1-\frac{1}{2}}(1-v_{i_1})^{p_{2:k}-\frac{i_{k-1}+3-i_1}{2}}$$

$$\cdots \prod_{\ell_{k-1}=i_{k-2}+1}^{i_{k-1}-1} v_{\ell_{k-1}}^{-\frac{1}{2}}(1-v_{\ell_{k-1}})^{p_k-\frac{i_{k-1}+3-\ell_{k-1}}{2}} v_{i_{k-1}}^{p_{k-1}-\frac{1}{2}}(1-v_{i_{k-1}})^{-\frac{3}{2}}$$

$$\times e^{-\frac{a^2}{2t\prod_{j=1}^{i_{k-1}}(1-v_j)}} dv_{v_{i_k}} dv_1\cdots dv_{i_{k-1}} dt. \qquad (2.7.38)$$

We can perform the analogous computations for i_{k-1}, i_{k-2}, $\cdots$, i_1 in this order repeatedly to obtain (3.8).

When $p_1 = \cdots = p_k = 2$, similar computations to that in the proof of (3.9) can be carried out to obtain (3.10). □

2.7.5 Proof of Proposition 2.3.4

Proof of Proposition 3.4. By the identities $\Gamma(c,x) = e^{-x}x^c \int_0^\infty e^{-xu}(1+u)^{c-1}du$ and $\sum_{j=0}^\infty \frac{(n)_j}{j!}x^j = (1-x)^n$, we can rewrite the joint density of stick-breaking weights $v_1, \cdots, v_n$ as

$$f(v_1,\cdots,v_n) = \frac{a^n\sigma^{n-1}}{[\Gamma(1-\sigma)]^n\Gamma(n\sigma)}\prod_{i=1}^n v_i^{-\sigma}(1-v_i)^{-(n-i)\sigma-1}e^{-\frac{a}{\prod_{i=1}^n(1-v_i)^\sigma}}$$
$$\times\int_0^\infty (1-(1+t)^{-\frac{1}{\sigma}})^{n\sigma-1}(1+t)^{n-1}e^{-\frac{at}{\prod_{i=1}^n(1-v_i)^\sigma}}dt. \tag{2.7.39}$$

We make the substitution $t = \frac{\prod_{i=1}^n(1-v_i)^\sigma s}{a}$ in the above integral. Then, when a is large, namely, when t is small, we have

$$(1+t)^{-\frac{1}{\sigma}} \asymp 1 - \frac{t}{\sigma} = 1 - \frac{\prod_{i=1}^n(1-v_i)^\sigma}{\sigma a}s\,,$$

where and throughout this chapter we use $\mu \asymp \nu$ to represent $\lim \frac{\mu}{\nu} = 1$.

The integral in (2.7.39) is then approximated by

$$\int_0^\infty (1-(1+t)^{-\frac{1}{\sigma}})^{n\sigma-1}(1+t)^{n-1}e^{-\frac{at}{\prod_{i=1}^n(1-v_i)^\sigma}}dt \asymp \frac{\prod_{i=1}^n(1-v_i)^{n\sigma^2}\Gamma(n\sigma)}{\sigma^{n\sigma-1}a^{n\sigma}}. \tag{2.7.40}$$

Thus, for large a, the joint density of $v_1, \cdots, v_n$ has the following asymptotics:

$$f(v_1,\cdots,v_n) \asymp \frac{(a\sigma)^{n-n\sigma}}{[\Gamma(1-\sigma)]^n}\prod_{i=1}^n v_i^{-\sigma}(1-v_i)^{n\sigma^2-(n-i)\sigma-1}e^{-\frac{a}{\prod_{i=1}^n(1-v_i)^\sigma}}. \tag{2.7.41}$$

Now the identities (3.11) and (3.12) follows from the same arguments as that in the proof of Proposition 3.3 and from the use of the following identity, which holds true for any $q \in \mathbb{R}$:

$$\int_0^1 x^{-\sigma}(1-x)^q e^{-\frac{a}{(1-x)^\sigma}}dx$$
$$= \frac{e^{-a}}{a\sigma}\int_0^\infty (1-(\frac{a}{a+s})^{\frac{1}{\sigma}})^{-\sigma}(\frac{a}{a+s})^{\frac{q+1}{\sigma}+1}e^{-s}ds$$

$$= e^{-a}(a\sigma)^{\sigma-1}(1+o(\frac{1}{a}))\int_0^1 s^{-\sigma}e^{-s}ds$$
$$= e^{-a}(a\sigma)^{\sigma-1}\Gamma(1-\sigma)\left(1+o\left(\frac{1}{a}\right)\right), \tag{2.7.42}$$

where the last equality follows from the substitution $s = \frac{a}{(1-x)^{-\sigma}} - a$ and the following asymptotics:

$$\left(\frac{a}{a+s}\right)^{\frac{1}{\sigma}} = \left(1 - \frac{s}{a+s}\right)^{\frac{1}{\sigma}} = 1 - \frac{s}{\sigma(a+s)} + o\left(\frac{1}{a}\right).$$

To obtain the exact asymptotics in the case when $\sigma = \frac{1}{m}$ and $p = p_1 = \cdots = p_k = 2$, we first prove (3.13) using the same argument as that in the proof of (3.9). The only differences are as follows. First, we integrate v_{n-m}, v_{n-m-1}, $\cdots$, v_1 recursively in this order by using (2.7.42). After these integrations, it remains to integrate the variables $v_{n-m+1}, \cdots, v_n$. This multiple integral is now evaluated simultaneously by using the substitution

$$v_{n-i} = 1 - \left(\frac{a + y_0 + \cdots + y_{i-1}}{a + y_0 + \cdots + y_i}\right)^{\frac{1}{\sigma}}, \qquad i = 0, \cdots, m-1.$$

The identity (3.14) follows from (3.13) by the same argument as that in the proof of (3.10). □

2.7.6 Proof of Proposition 2.3.5

Proof of Proposition 3.5. When $P \sim \mathrm{GDP}(a, r, H)$, we will prove the weak convergence of D_a in the part (i) of Theorem 4.15. Combining with the fact that $\mathrm{GDP}(a, r, H)$ also admits the general stick-breaking representation as in (2.1) and (2.2), we can obtain the desired results by using the same argument as that in the proof of (4.5) in Theorem 4.4. □

2.7.7 Proof of Theorem 2.4.2

Proof.

$$
\begin{aligned}
\mathbb{P}\left(|P(A)-H(A)|>\epsilon\right) &\leq \frac{\mathbb{E}\ \left[|P(A)-H(A)|^m\right]}{\epsilon^m} \\
&= \frac{\mathbb{E}\ \left[|\sum_{i=1}^{\infty} w_i \left(\delta_{\theta_i}(A)-H(A)\right)|^m\right]}{\epsilon^m} \\
&= \sum \frac{c(p_1,\cdots,p_k)}{\epsilon^m} \prod_{i=1}^{k} \mathbb{E}\ \left[|\delta_{\theta_i}(A)-H(A)|^{p_i}\right] \\
&\qquad \times \mathbb{E}\left[\sum_{1\leq i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right], \qquad (2.7.43)
\end{aligned}
$$

where the first sum is taken over all combinations of nonnegative integers $\{p_1,\cdots,p_k\}$ such that $k\in\{1,\cdots,m\}$ and $\sum_{i=1}^{k} p_i = m$ and where $c(p_1,\cdots,p_k)=\binom{m}{p_1,\cdots,p_k}=\frac{m!}{p_1!\cdots p_k!}$ are the corresponding combinatorial coefficients. For any combination of $\{p_1,\cdots,p_k\}$, if there exists $i\in\{1,\cdots,k\}$ such that $p_i=1$, then the product in (2.7.43) will be 0 due to the fact that

$$
\mathbb{E}\left[\delta_{\theta_i}(A_i)-H(A_i)\right]=\mathbb{E}\left[\mathbb{1}_{A_i}(\theta_i)\right]-H(A_i)=\int_{A_i} dH - H(A_i) = 0\,.
$$

That is to say, $p_i \geq 2$ for all i and thus $k\leq \frac{m}{2}$.

First, assume P is one of $\mathrm{DP}(a,H)$, $\mathrm{PDP}(a,b,H)$, $\mathrm{N\text{-}IG}(a,H)$, $\mathrm{NGGP}(\sigma,a,H)$, $\mathrm{GDP}(a,r,H)$. We choose $m=\lfloor\frac{4}{\tau}\rfloor$, where $\lfloor x\rfloor$ is the smallest integer that is greater than or equal to x. Then, from Propositions 3.2-3.5, we have

$$
\sum \mathbb{E}\left[\sum_{1\leq i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right] \asymp \frac{1}{a^{m-k}} \leq \frac{1}{a^{\frac{m}{2}}} = \frac{1}{N^{\frac{m\tau}{2}}} \leq \frac{1}{N^2}\,.
$$

If $p\sim \mathrm{SPG}(\mu_a,H)$, then we can choose m such that for all $1\leq k\leq m/2$,

$$
\sum_{1\leq i\leq k, p_1+\cdots+p_k=m} q_{p_i} - kq_1 \geq \frac{m(q_2-q_1)}{2} \geq \frac{2}{\tau}\,.
$$

Then, from Proposition 3.1,

$$\sum \mathbb{E}\left[\sum_{1\le i_1<i_2<\cdots<i_k<\infty} w_{i_1}^{p_1} w_{i_2}^{p_2}\cdots w_{i_k}^{p_k}\right] \asymp \frac{1}{a^{\sum_{i=1}^k q_{p_i}-kq_1}} \le \frac{1}{N^2}\,.$$

Since the series $\sum_{N=1}^{\infty}\frac{1}{N^2}$ converges, it follows

$$\sum_{N=1}^{\infty}\mathbb{P}\left(|P(A)-H(A)|>\epsilon\right)<\infty \tag{2.7.44}$$

for any of the processes presented in the theorem. This implies (4.2) by the Borel-Cantelli lemma. □

2.7.8 Proof of Theorem 2.4.4

Before we proceed to the proof of Theorem 4.4, we need a preparatory result about the joint moments of multivariate normal distribution. To state this result, we introduce the following notations. Let n be a positive integer and let $\vec{p}=(p_{ij}, 1\le i<j\le n)$ be a multi-index. Denote

$$|\vec{p}| = \sum_{1\le i<j\le n} p_{ij} \tag{2.7.45}$$

and denote

$$|\vec{p}|_m = \begin{cases} \sum_{j>1} p_{1j} & \text{when } m=1,\\ \sum_{j>m} p_{mj} + \sum_{i<m} p_{im} & \text{when } m=2,\cdots,n-1,\\ \sum_{i<n} p_{in} & \text{when } m=n\,. \end{cases} \tag{2.7.46}$$

The following proposition is about the joint moments of Gaussian random variables. Similar or more general results may be found in literature under the terminology of "Feynman diagram" (e.g., (Hu, 2017, Theorem 5.7) and references therein). But we could not find the exact result we need. So, we give the following proposition.

Proposition 2.7.14. *Let the random vector $(X_1, X_2, \cdots, X_n)$ follow a multivariate nor-*

mal distribution $N_n(0,\Sigma)$*, where* $\Sigma = (\sigma_{ij} = \mathbb{E}\,(X_iX_j))_{1\le i,j\le n}$ *and* $\sigma_{ii} = \sigma_i^2$*. For any nonnegative integers* $r_1,\cdots,r_n$*, the joint* $(r_1,\cdots,r_n)$ *moments of* $(X_1,X_2,\cdots,X_n)$ *is given by the following formulas.*

(i) When $\sum_{\ell=1}^n r_\ell$ *is an odd integer, we have*

$$\mathbb{E}\left[X_1^{r_1}\cdots X_n^{r_n}\right] = 0\,. \tag{2.7.47}$$

(ii) When $\sum_{\ell=1}^n r_\ell$ *is an even integer, we have (we use the convention that* $0^0 = 1$*)*

$$\begin{aligned}&\mathbb{E}\left[X_1^{r_1}\cdots X_n^{r_n}\right]\\ &=\sum_{\vec{p}\in\mathcal{C}}\frac{r_1!\cdots r_n!\ \prod_{m=1}^n \sigma_m^{r_m-|\vec{p}|_m}\ \prod_{1\le i<j\le n}\sigma_{ij}^{p_{ij}}}{2^{\frac{|r|}{2}-|\vec{p}|}\ \prod_{m=1}^n((r_m-|\vec{p}|_m)/2)!\ \prod_{1\le i<j\le n}p_{ij}!}\,,\end{aligned} \tag{2.7.48}$$

where $|r| = r_1+\cdots+r_n$ *and*

$$\begin{aligned}\mathcal{C} = \Big\{&\vec{p} = \{p_{ij},1\le i<j\le n\}\,;\quad 0\le p_{ij}\le r_i\wedge r_j\,,\\ &r_m-|\vec{p}|_m\ \ m=1,\cdots,n,\ \ \textit{are all even}\Big\}\,.\end{aligned} \tag{2.7.49}$$

Proof. We shall use the moment generating function to prove it. On one hand, we see

$$\mathbb{E}\left[e^{\sum_{\ell=1}^n t_\ell X_\ell}\right] = \sum_{r_1,\cdots,r_n=0}^{\infty}\frac{t_1^{r_1}\cdots t_n^{r_n}}{r_1!\cdots r_n!}\mathbb{E}\left[X_1^{r_1}\cdots X_n^{r_n}\right]\,. \tag{2.7.50}$$

On the other hand, by the moment generating function formula for multivariate normal variables, we have

$$\begin{aligned}\mathbb{E}\left[e^{\sum_{\ell=1}^n t_\ell X_\ell}\right] &= e^{\frac12\mathbb{E}\left[(\sum_{\ell=1}^n t_\ell X_\ell)^2\right]}\\ &=\exp\left\{\frac12\sum_{\ell=1}^n\mathbb{E}\left[(t_\ell X_\ell)^2\right]+\sum_{1\le i<j\le n}\mathbb{E}\left[t_it_jX_iX_j\right]\right\}\\ &=\prod_{\ell=1}^n\exp\left\{\frac{t_\ell^2\sigma_\ell^2}{2}\right\}\prod_{1\le i<j\le n}\exp\left\{t_it_j\sigma_{ij}\right\}\end{aligned} \tag{2.7.51}$$

$$= \sum_{\substack{0 \le p_l, p_{ij} < \infty \\ 1 \le \ell \le n \\ 1 \le i < j \le n}} \prod_{\ell=1}^{n} \frac{(t_\ell \sigma_\ell)^{2p_\ell}}{2^{p_\ell} p_\ell!} \prod_{1 \le i < j \le n} \frac{(t_i t_j \sigma_{ij})^{p_{ij}}}{p_{ij}!}$$

$$= \sum_{\substack{0 \le p_l, p_{ij} < \infty \\ 1 \le \ell \le n \\ 1 \le i < j \le n}} \frac{t_1^{2p_1 + |\vec{p}|_1} \cdots t_n^{2p_n + |\vec{p}|_m}}{2^{p_1 + \cdots + p_n}} \prod_{\ell=1}^{n} \frac{\sigma_\ell^{2p_\ell}}{p_\ell!} \prod_{1 \le i < j \le n} \frac{\sigma_{ij}^{p_{ij}}}{p_{ij}!} . \tag{2.7.52}$$

Comparing the coefficients of $t_1^{r_1} \cdots t_n^{r_n}$ of the two representations (2.7.50) and (2.7.52) we obtain

$$\mathbb{E}\left[X_1^{r_1} \cdots X_n^{r_n}\right] = \frac{r_1! \cdots r_n!}{2^{p_1 + \cdots + p_n}} \prod_{\ell=1}^{n} \frac{\sigma_\ell^{2p_\ell}}{p_\ell!} \prod_{1 \le i < j \le n} \frac{\sigma_{ij}^{p_{ij}}}{p_{ij}!} , \tag{2.7.53}$$

where p_ℓ, p_{ij}, $1 \le \ell \le n, 1 \le i < j \le n$ satisfies the relation $r_\ell = 2p_\ell + |\vec{p}|_\ell$ for $\ell = 1, \cdots, n$, which implies that $\sum_{\ell=1}^{n} r_\ell = 2(\sum_{\ell=1}^{n} p_\ell + \sum_{1 \le i,j \le n} p_{ij})$ is an even number. This also proves that $\mathbb{E}\left[X_1^{r_1} \cdots X_n^{r_n}\right] = 0$ when $\sum_{\ell=1}^{n} r_\ell$ is an odd integer. This proves part (i) of the proposition. When $\sum_{\ell=1}^{n} r_\ell$ is even, by the fact that $\{p_1, \cdots, p_n\}$ are nonnegative integers and by the relationship $p_\ell = \frac{r_\ell - |\vec{p}|_\ell}{2}$ for $\ell = 1, \cdots, n$, one see the summation in (2.7.53) is over the set $\mathcal{C}$ defined by (2.7.49). □

Now, we are in the position to prove Theorem 4.4.

Proof of Theorem 4.4. We first prove the part (i) of this theorem.

For any $A \in \mathcal{X}$, the variance of $P(A)$ is given by (2.7.13). Using Proposition 3.1, we have

$$\mathrm{Var}(P(A)) = H(A)\,(1 - H(A)) \left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]} \right) \right).$$

From the definition of $D_a(A)$ we have

$$D_a(A) = \left[H(A)\,(1 - H(A)) \left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]} \right) \right) \right]^{-\frac{1}{2}} (P(A) - H(A)) .$$

By the Cramér-Wold theorem (e.g., (Billingsley, 1995, Theorem 29.4)), to show (4.5) it is sufficient to show that for any $(t_1, \cdots, t_n) \in \mathbb{R}^n$

$$\sum_{i=1}^n t_i D_a(A_i) \xrightarrow{d} \sum_{i=1}^n t_i X_i,$$

where and throughout the remaining part of the chapter $(X_1, \cdots, X_n)$ are jointly Gaussian with mean zero and covariance given by (4.6). For any positive integer n and a nonnegative integer sequence $\{r_i\}_{i=1}^n$, consider the joint moments of $D_a(A_1), D_a(A_2), \cdots, D_a(A_n)$:

$$\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^n D_a^{r_i}(A_i)\right] \\
&= \prod_{i=1}^n \left[H(A_i)(1-H(A_i))\left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)\right)\right]^{-\frac{r_i}{2}} \mathbb{E}\left[\prod_{i=1}^n (P(A_i) - H(A_i))^{r_i}\right] \\
&= \prod_{i=1}^n \left[H(A_i)(1-H(A_i))\left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)\right)\right]^{-\frac{r_i}{2}} \times \\
&\qquad \mathbb{E}\left[\prod_{i=1}^n \left(\sum_{j_i=1}^\infty w_{j_i}\left(\delta_{\theta_{j_i}}(A_i) - H(A_i)\right)\right)^{r_i}\right]. \qquad (2.7.54)
\end{aligned}$$

We expand the product of the infinite sums inside the above last expectation as

$$\begin{aligned}
&\mathbb{E}\left[\prod_{i=1}^n \left(\sum_{j_i=1}^\infty w_{j_i}\left(\delta_{\theta_{j_i}}(A_i) - H(A_i)\right)\right)^{r_i}\right] \\
&\qquad = \sum C(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}) I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}),
\end{aligned}$$

where the sum is over all the nonnegative integers $\{q, s_1, \cdots, s_n; s_{j,i} : j \in \{1, \cdots, q\}; i \in \{1, \cdots, n\}\}$ satisfying

(i) $s_j = \sum_{i=1}^n s_{j,i} \geq 1$ for all $j \in \{1, \cdots, q\}$

(ii) $r_i = \sum_{j=1}^q s_{j,i} \geq 1$ for all $i \in \{1, \cdots, n\}$

(iii) $1 \leq q \leq \sum_{i=1}^n r_i$;

(iv) $\sum_{j=1}^q s_j = \sum_{i=1}^n r_i$

(v) $C(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ are some constants (that are found later on) depending on $q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}$ and

$$
\begin{aligned}
& I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}) \\
& \qquad := \mathbb{E}\Bigg[\sum_{1 \le e_1 < \cdots < e_q < \infty} \prod_{j=1}^{q} w_{e_j}^{s_j} \Big(\delta_{\theta_{e_j}}(A_1) - H(A_1)\Big)^{s_{j,1}} \\
& \qquad\qquad \cdots (\delta_{\theta_{e_j}}(A_n) - H(A_n))^{s_{j,n}} \Bigg] . \qquad (2.7.55)
\end{aligned}
$$

With these notations, we can write (2.7.54) as

$$
\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^{n} D_a^{r_i}(A_i)\right] = \sum \prod_{i=1}^{n} & \left[H(A_i)(1 - H(A_i)) \left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)\right)\right]^{-\frac{r_i}{2}} \\
& C(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}) I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n}) . \qquad (2.7.56)
\end{aligned}
$$

We will divide the discussion of the limit as $a \to \infty$ of the terms in (2.7.56) into three cases according to the indice $\{q; s_1, \cdots, s_n; s_{j,i} : j \in \{1, \cdots, q\}; i \in \{1, \cdots, n\}\}$ appeared in (2.7.56) satisfying (i)-(iv).

Case 1: *There exists at least one $j \in \{1, \cdots, q\}$ such that $s_j = 1$ or there exists at least one pair (j, k) such that $s_{j,k} = 1$.*

From the fact that

$$
\mathbb{E}\left[\delta_{\theta_i}(A_i) - H(A_i)\right] = \mathbb{E}\left[\mathbb{1}_{A_i}(\theta_i)\right] - H(A_i) = \int_{A_i} dH - H(A_i) = 0 ,
$$

we see that in this case the corresponding terms in the sum of $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ are identically equal to 0.

Case 2 *All $s_j \ge 2$, $\sum_{i=1}^{n} r_i$ is odd or $\sum_{i=1}^{n} r_i$ is even but $\frac{\sum_{i=1}^{n} r_i}{2} > q$.*

We substitute (2.7.55) into (2.7.56) and we consider the expectations of v_1's. By Proposition 3.1, when excluding the terms discussed in *Case 1* the remaining terms corresponding to this case have the following asymptotics

$$O\left(\left[\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right]^{-\frac{\sum_{i=1}^n r_i}{2}} \frac{\prod_{j=1}^q \mathbb{E}[v_1^{s_j}]}{(\mathbb{E}[v_1])^q}\right) = O\left(\frac{(\mathbb{E}[v_1])^{\frac{\sum_{j=1}^q s_j}{2}-q} \prod_{j=1}^q \mathbb{E}[v_1^{s_j}]}{(\mathbb{E}[v_1^2])^{\frac{\sum_{j=1}^q s_j}{2}}}\right). \qquad (2.7.57)$$

From the assumption (4.4), it follows that when $\sum_{i=1}^n r_i$ is even and when $q < \frac{\sum_{i=1}^n r_i}{2}$, the expectation of the terms in the sum of $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ will converge to 0 as $a \to \infty$.

Similarly, when $\sum_{i=1}^n r_i$ is odd, since q is an integer and $s_j \geq 2$ for all j, we always have $q < \frac{\sum_{i=1}^n r_i}{2}$. Therefore, the expectation of the corresponding terms satisfying the condition that $\sum_{i=1}^n r_i$ is odd in the sum of $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ will always converge to 0 as $a \to \infty$.

Case 3 *All* $s_j \geq 2$, $\sum_{i=1}^n r_i$ *is even and* $\frac{\sum_{i=1}^n r_i}{2} = q$.

The only terms that may not converge to zero are the terms that are not covered in *Case 1* and *Case 2*. This means that the only terms that have nontrivial limits are the terms satisfying the conditions that

$$\sum_{i=1}^n r_i \text{ is even and } \quad q = \frac{\sum_{i=1}^n r_i}{2}.$$

With the condition (iv), we have $\sum_{j=1}^q s_j = \sum_{i=1}^n r_i = 2q$. But $s_j \geq 2$ for all j. Thus it is easy to see this is possible only when

$$s_1 = \cdots = s_q = 2 \quad \text{and} \quad s_{j,i} \in \{0, 1, 2\}$$

for all $j \in \{1, \cdots, q\}$ and for all $i \in \{1, \cdots, n\}$. We shall discuss this nontrivial case in 3 steps.

Step 1: The form of $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$.

For $\ell \in \{1, \cdots, \frac{\sum_{i=1}^n r_i}{2}\}$ the factors in $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ are either of the form $w_{e_\ell}^2(\delta_{\theta_{e_\ell}}(A_i) - H(A_i))(\delta_{\theta_{e_\ell}}(A_j) - H(A_j))$ for $1 \leq i < j \leq n$ (we call this factor the (ij)-*mixed term*, and if there is no ambiguity to omit the pre-index (ij), we call this kind of factor the *mixed term*) or of the form $w_{e_\ell}^2(\delta_{\theta_{e_\ell}}(A_d) - H(A_d))^2$ (which we call the *power two term of* A_d) for $d \in \{1, \cdots, n\}$.

Step 2: Computation of $C(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$ **.**

For each $\ell \in \{1, \cdots, \frac{\sum_{i=1}^n r_i}{2}\}$, let

$$p_{ij} := \# \{(i,j); 1 \le i < j \le n\,,\ s_{\ell,i} = s_{\ell,j} = 1\}\ .$$

Namely, for each pair of i, j such that $1 \le i < j \le n$, p_{ij} is the number of (ij)-mixed terms in the product of $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$. Notice that, in order to obtain a (ij)-mixed term, we need to multiply the form $w_{e_\ell}(\delta_{\theta_{e_\ell}}(A_i) - H(A_i))$ (we call this form the *power* 1 *term of* A_i and there are r_i power 1 terms of A_i) and the form $w_{e_\ell}(\delta_{\theta_{e_\ell}}(A_j) - H(A_j))$ (we call this form the *power* 1 *term of* A_j and there are r_j power 1 terms of A_i). Moreover, there are $p_{ij}!$ ways to get as many as p_{ij} (ij)-mixed terms. Therefore, there are totally $\prod_{1 \le i < j \le n} p_{ij}!$ mixed terms in $I(q; s_1, \cdots, s_n, s_{1,1}, \cdots, s_{q,n})$.

Now for each $d \in \{1, \cdots, n\}$, after picking up p_{dj} power 1 terms of A_d, there will be $\frac{r_d - \sum_{j>d} p_{dj} - \sum_{i<d} p_{id}}{2}$ power 2 terms of A_d left for us to pick up. Thus, for each $d \in \{1, \cdots, n\}$, there are r_d power 1 terms of A_d, in which $\sum_{j>d}^n p_{dj} + \sum_{i<d} p_{id}$ of them will be used to construct the mixed terms and $r_d - \sum_{j>d}^n p_{dj} - \sum_{i<d} p_{id}$ of them will be used to construct the power 2 terms of A_d. They have to satisfy the conditions (denoted by $\mathcal{C}$)

$$\begin{cases} r_1 - \sum_{j>1} p_{1j} \ \text{ is even,} \\ r_\ell - \sum_{j>\ell} p_{\ell j} - \sum_{i<\ell} p_{i\ell} \ \text{ is even,} \quad \text{for } \ell \in \{2, \cdots, n-1\}, \\ r_n - \sum_{i<n} p_{in} \ \text{ is even.} \end{cases}$$

Thus, for each $d \in \{1, \cdots, n\}$, we can construct the mixed terms and power 2 terms of A_d as follows. Since there are $\sum_{j>d}^n p_{dj} + \sum_{i<d} p_{id}$ mixed terms, we choose $(p_{1d}, \cdots, p_{(d-1)d}, p_{d(d+1)}, \cdots, p_{dn})$ (some of them could be 0) out of r_d and then combine the remaining $r_d - \sum_{j>d}^n p_{dj} - \sum_{i<d} p_{id}$ power 1 terms of A_d as the power 2 terms of A_d.

Thus, the number of terms from the above steps is

$$\frac{\binom{r_d}{p_{1d},\cdots,p_{(d-1)d},p_{d(d+1)},\cdots,p_{dn}}\binom{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}\binom{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}-2}{2}\cdots\binom{2}{2}}{\left(\frac{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}\right)!}$$

without ordering. After the above steps, all the mixed terms and power 2 terms of A_d can be ordered in $\frac{\sum_{l=1}^n r_l}{2}!$ ways. Noticing that $p_{d_1 d_2}\in\{0,1,\cdots,r_{d_1}\wedge r_{d_2}\}$, the coefficient $C(q;s_1,\cdots,s_n,s_{1,1},\cdots,s_{q,n})$ is then

$$\begin{aligned}
&C(q;s_1,\cdots,s_n,s_{1,1},\cdots,s_{q,n})\\
&=\sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j}\prod_{d=1}^n\frac{\binom{r_d}{p_{1d},\cdots,p_{(d-1)d},p_{d(d+1)},\cdots,p_{dn}}\binom{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}\cdots\binom{2}{2}}{\left(\frac{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}\right)!}\\
&\qquad\left(\prod_{1\le i<j\le n}p_{ij}!\right)\left(\frac{\sum_{\ell=1}^n r_\ell}{2}\right)!\\
&=\sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j}\prod_{d=1}^n\frac{\binom{r_d}{p_{1d},\cdots,p_{(d-1)d},p_{d(d+1)},\cdots,p_{dn}}\left(r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}\right)!}{\left(\frac{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}\right)!2^{\frac{r_d-\sum_{j>d}^n p_{dj}-\sum_{i<d}p_{id}}{2}}}\\
&\qquad\left(\prod_{1\le i,j\le n}p_{ij}!\right)\left(\frac{\sum_{\ell=1}^n r_\ell}{2}!\right)\\
&=\sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j}\frac{r_1!\cdots r_n!}{\left(\frac{r_1-\sum_{j>1}p_{1j}}{2}\right)!\cdots\left(\frac{r_m-\sum_{j>m}p_{mj}-\sum_{i<m}p_{im}}{2}\right)!\cdots\left(\frac{r_n-\sum_{i<n}p_{in}}{2}\right)!}\\
&\qquad\frac{\left(\frac{\sum_{\ell=1}^n r_\ell}{2}\right)!}{2^{\frac{1}{2}\sum_{\ell=1}^n r_\ell-\sum_{1\le i,j\le n}p_{ij}}\left(\prod_{1\le i<j\le n}p_{ij}!\right)}.
\end{aligned}\tag{2.7.58}$$

Step 3: Computation of $\mathbb{E}\left[\prod_{i=1}^n D_a^{r_i}(A_i)\right]$.

By Proposition 3.1, $I(q;s_1,\cdots,s_n,s_{1,1},\cdots,s_{q,n})$ can be rewritten as

$$\begin{aligned}
&I(q;s_1,\cdots,s_n,s_{1,1},\cdots,s_{q,n})=\\
&\mathbb{E}\left[\sum_{1\le e_1<\cdots<e_{\frac{\sum_{i=1}^n r_i}{2}}<\infty}\prod_{j=1}^{\frac{\sum_{i=1}^n r_i}{2}}w_{e_j}^2\right]\prod_{d=1}^n\left(\mathbb{E}\left[(\delta_\theta(A_d)-H(A_d))^2\right]\right)^{\frac{r_d-\sum_{j>d}p_{dj}-\sum_{i<d}p_{id}}{2}}
\end{aligned}$$

$$\prod_{1\le i<j\le n} \left(\mathbb{E}\left[(\delta_\theta(A_i)-H(A_i))\left(\delta_\theta(A_j)-H(A_j)\right)\right]\right)^{p_{ij}}$$

$$= \left(\frac{1}{2^{\frac{\sum_{i=1}^n r_i}{2}}\left(\frac{\sum_{i=1}^n r_i}{2}\right)!}\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^{\frac{\sum_{i=1}^n r_i}{2}} + o\left(\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^{\frac{\sum_{i=1}^n r_i}{2}}\right)\right)$$

$$\prod_{d=1}^n (H(A_d)(1-H(A_d)))^{\frac{r_d-\sum_{j>d}p_{dj}-\sum_{i<d}p_{id}}{2}} \prod_{1\le i<j\le n} (-H(A_i)H(A_j))^{p_{ij}} . \tag{2.7.59}$$

From (2.7.56) and (2.7.58), (2.7.59), we can compute $\mathbb{E}\left[\prod_{i=1}^n D_a^{r_i}(A_i)\right]$ as follows.

$$\mathbb{E}[\prod_{i=1}^n D_a^{r_i}(A_i)] = \prod_{i=1}^n \left[H(A_i)(1-H(A_i))\left(\frac{\mathbb{E}[v_1^2]}{2\mathbb{E}[v_1]} + o\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)\right)\right]^{-\frac{r_i}{2}}$$

$$\left(\frac{1}{2^{\frac{\sum_{i=1}^n r_i}{2}}\left(\frac{\sum_{i=1}^n r_i}{2}\right)!}\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^{\frac{\sum_{i=1}^n r_i}{2}} + o\left(\left(\frac{\mathbb{E}[v_1^2]}{\mathbb{E}[v_1]}\right)^{\frac{\sum_{i=1}^n r_i}{2}}\right)\right)$$

$$\sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j} \frac{r_1!\cdots r_n!}{\left(\frac{r_1-\sum_{j>1}p_{1j}}{2}\right)!\cdots\left(\frac{r_m-\sum_{j>m}p_{mj}-\sum_{i<m}p_{im}}{2}\right)!\cdots\left(\frac{r_n-\sum_{i<n}p_{in}}{2}\right)!}$$

$$\frac{\left(\frac{\sum_{i=1}^n r_i}{2}\right)!}{2^{\frac{1}{2}\sum_{\ell=1}^n r_\ell-\sum_{1\le i<j\le n}p_{ij}}\left(\prod_{1\le i<j\le n}p_{ij}!\right)}$$

$$\prod_{d=1}^n (H(A_d)(1-H(A_d)))^{\frac{r_d-\sum_{j>d}p_{dj}-\sum_{i<d}p_{id}}{2}} \prod_{1\le i<j\le n} (-H(A_i)H(A_j))^{p_{ij}}$$

$$= \sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j} \frac{r_1!\cdots r_n!}{\left(\frac{r_1-\sum_{j>1}p_{1j}}{2}\right)!\cdots\left(\frac{r_m-\sum_{j>m}p_{mj}-\sum_{i<m}p_{im}}{2}\right)!\cdots\left(\frac{r_n-\sum_{i<n}p_{in}}{2}\right)!}$$

$$\frac{1}{2^{\frac{1}{2}\sum_{\ell=1}^n r_\ell-\sum_{1\le i<j\le n}p_{ij}}\left(\prod_{1\le i<j\le n}p_{ij}!\right)}$$

$$\prod_{1\le i<j\le n}\left(-\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}}\right)^{p_{ij}} + o(1)$$

$$\overset{a\to\infty}{\to} \sum_{\substack{p_{ij}=0,1\le i<j\le n\\(i,j)\in\mathcal{C}}}^{r_i\wedge r_j} \frac{r_1!\cdots r_n!}{\left(\frac{r_1-\sum_{j>1}p_{1j}}{2}\right)!\cdots\left(\frac{r_m-\sum_{j>m}p_{mj}-\sum_{i<m}p_{im}}{2}\right)!\cdots\left(\frac{r_n-\sum_{i<n}p_{in}}{2}\right)!}$$

$$\frac{1}{2^{\frac{1}{2}\sum_{\ell=1}^n r_\ell-\sum_{1\le i<j\le n}p_{ij}}\left(\prod_{1\le i<j\le n}p_{ij}!\right)}$$

$$\prod_{1\le i<j\le n}\left(-\sqrt{\frac{H(A_i)H(A_j)}{(1-H(A_i))(1-H(A_j))}}\right)^{p_{ij}}, \tag{2.7.60}$$

which is equal to $\mathbb{E}\left[\prod_{i=1}^n X_i^{r_i}\right]$, where $X_1,\cdots,X_n$ are the multi-normal distribution defined in Proposition 2.7.14. Now, by multinomial expansion we see that for any positive integer k and for any $(t_1,\cdots,t_n)\in\mathbb{R}^d$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^n t_i D_a(A_i)\right]^k &= \sum_{k_1+\cdots+k_n=k}\binom{k}{k_1!\cdots k_n!}\mathbb{E}\left[\prod_{i=1}^n t_i^{k_i} D_a^{k_i}(A_i)\right]\\
&\overset{a\to\infty}{\to} \sum_{k_1+\cdots+k_n=k}\binom{k}{k_1!\cdots k_n!}\mathbb{E}\left[\prod_{i=1}^n t_i^{k_i} X_i^{k_i}\right]\\
&= \mathbb{E}\left[\sum_{i=1}^n t_i X_i\right]^k.
\end{aligned}$$

By the method of moments (see e.g. Billingsley (1995, Theorem 30.2)) it follows that

$$\sum_{i=1}^n t_i D_a(A_i) \overset{d}{\to} \sum_{i=1}^n t_i X_i \quad \text{as } a\to\infty.$$

Part (i) of Theorem 4.4 follows then from the Cramér-Wold theorem (e.g., (Billingsley, 1995, Theorem 29.4)).

Now, we prove the part (ii) of this theorem by proving the weak convergence of finite dimensional distributions and by verifying a tightness condition. The finite dimensional weak convergence of $Q_{H,a}$ can be shown directly by part (i), i.e., for any finite measurable sets $A_1,\cdots,A_n$ in $\mathcal{X}^d$, we have

$$(Q_{H,a}(A_1),\cdots,Q_{H,a}(A_n)) \overset{d}{\to} (B_H^o(A_1),\cdots,B_H^o(A_n)).$$

By Theorem 2 of (Bickel and Wichura, 1971), to show (4.7) we only need to check the tightness condition, i.e, inequality (3) of (Bickel and Wichura, 1971), with $\gamma_1=\gamma_2=2$, $\beta_1=\beta_2=1$ and $\mu=2H$. Obviously, μ is finite and nonatomic. For every pair of Borel sets A and B in $\mathcal{B}(\mathbb{R}^d)$, by the proof of part (i) of this theorem and the Isserlis' theorem

(Isserlis, 1918) we have

$$\begin{aligned}
&\mathbb{E}[|Q_{H,a}(A)|^2|Q_{H,a}(B)|^2] \\
&\qquad = [H(A)(1-H(A))][H(B)(1-H(B))]\mathbb{E}[D_a^2(A)D_a^2(B)] \\
&\qquad = [H(A)(1-H(A))][H(B)(1-H(B))]\left(1+2\frac{H(A)H(B)}{(1-H(A))(1-H(B))}+o(1)\right) \\
&\qquad = 3H(A)^2H(B)^2 - H(A)^2H(B) - H(A)H(B)^2 + H(A)H(B) + o(1) \\
&\qquad \leq \mu(A)\mu(B).
\end{aligned}$$

The last inequality is due to the fact that $H(\cdot) \in (0,1)$ and thus $H(\cdot)^2 \leq H(\cdot)$. Therefore, the tightness condition on $D(\mathbb{R}^d)$ is verified. □

2.7.9 Proof of Theorem 2.4.15

Proof of Theorem 4.15. Once we have Proposition 3.2-3.4, the proofs of part (i) of this theorem for the various processes except the generalized Dirichlet process follow from a similar argument to that in the proof of part (i) of Theorem 4.4. So, we shall omit the details.

When $P \sim \text{GDP}(a,r,H)$, We need the following result about the variance of P from (Lijoi et al., 2005a):

$$\text{Var}\left[P(A)\right] = H(A)\left(1-H(A)\right)\mathcal{I}_{a,r}\,,$$

where $\mathcal{I}_{a,r}$ is given by

$$\begin{aligned}
\mathcal{I}_{a,r} &= a(r!)^a \sum_{k=1}^{r}\int_0^\infty \frac{x}{(k+x)^2\prod_{j=1}^{r}(j+x)^a}dx \\
&= \frac{a(r!)^a\Gamma(ra)}{r^{ra}\Gamma(ra+2)}\sum_{j=1}^{r} F_D^{(r-1)}\left(ra, \mathbf{a}_\mathbf{k}^*; ra+2; \frac{1}{r}\mathbf{J}_{\mathbf{r-1}}\right). \qquad (2.7.61)
\end{aligned}$$

Here $\mathbf{a}_k^* = (a,\cdots,a+2,\cdots,a)^T$ is a $r-1$ dimensional vector where the k-th element is $a+2$ and all other elements are equal to a; $\mathbf{J}_{r-1} = (1,\cdots,r-1)^T$; and $F_D^{(r-1)}$ is the fourth

Lauricella multiple hypergeometric function (see e.g., (Exton, 1976) for more details).

Letting $x = \frac{t}{a}$ we have

$$\begin{aligned}\mathcal{I}_{a,r} &= a\sum_{k=1}^{r}\frac{1}{k^2}\int_0^{\infty}\frac{x}{\left(1+\frac{x}{k}\right)^2\prod_{j=1}^{r}\left(1+\frac{x}{j}\right)^a}dx\\ &= a\sum_{k=1}^{r}\frac{1}{ak^2}\int_0^{\infty}\frac{t}{a\left(1+\frac{t}{ak}\right)^2\prod_{j=1}^{r}\left(1+\frac{t}{aj}\right)^a}dt\,.\end{aligned}$$

When a is large, we can approximate $\mathcal{I}_{a,r}$ by

$$\begin{aligned}\mathcal{I}_{a,r} &= \frac{1}{a}\sum_{k=1}^{r}\frac{1}{k^2}\int_0^{\infty}te^{-(\sum_{j=1}^{r}\frac{1}{j})t}dt + o\left(\frac{1}{a}\right)\\ &= \frac{\sum_{k=1}^{r}\left(\frac{1}{k}\right)^2}{\left(\sum_{j=1}^{r}\frac{1}{j}\right)^2 a} + o\left(\frac{1}{a}\right)\,. \qquad (2.7.62)\end{aligned}$$

Denote

$$c = \frac{\left(\sum_{j=1}^{r}\frac{1}{j}\right)^2}{\sum_{k=1}^{r}\left(\frac{1}{k}\right)^2}\,. \qquad (2.7.63)$$

Then

$$D_a(\cdot) \approx \frac{1}{\sqrt{\frac{H(\cdot)(1-H(\cdot))}{ca}}}\left(P(\cdot)-H(\cdot)\right)\,, \qquad (2.7.64)$$

where $\approx$ means that the two sides converge to the same distribution as $a \to \infty$. We shall prove the result for $n = 3$ and the case for general n can be handled in a similar way.

By the integral representation of the confluent form of the fourth Lauricella hypergeometric function (see e.g., formula (1.4.3.9) in (Exton, 1976)), the joint probability density of $P(A_1), P(A_2)$ admits the following form:

$$\begin{aligned}\rho(x_1,x_2) =& \frac{\Gamma(ra)}{\Gamma(raH_1)\Gamma(raH_2)\Gamma(raH_3)}\frac{(r!)}{r^{ra}\Gamma(ra)}\frac{\Gamma(raH_1)\Gamma(raH_2)\Gamma(raH_3)}{[\Gamma(aH_1)\Gamma(aH_2)\Gamma(aH_3)]^r}\\ &\times x_1^{2aH_1-1}x_2^{2aH_2-1}(1-x_1-x_2)^{2aH_3-1}\\ &\times\int_0^{\infty}\int_{0\le u_1^{(1)}+\cdots+u_{r-1}^{(1)}\le 1}\int_{0\le u_1^{(2)}+\cdots+u_{r-1}^{(2)}\le 1}\int_{0\le u_1^{(3)}+\cdots+u_{r-1}^{(3)}\le 1}\xi^{ra-1}\end{aligned}$$

$$\times \exp\left\{ -\xi + \xi\left(\frac{\sum_{k=1}^{r-1} ku_k^{(1)}x_1 + \sum_{k=1}^{r-1} ku_k^{(2)}x_2 + \sum_{k=1}^{r-1} ku_k^{(3)}(1-x_1-x_2)}{r}\right)\right\}$$

$$\times \prod_{i=1}^{3}\left[u_1^i \cdots u_{r-1}^{(i)}\left(1-u_1^i-\cdots-u_{r-1}^{(i)}\right)\right]^{aH_i-1} du_1^{(1)}\cdots du_{r-1}^{(1)}$$

$$du_1^{(2)}\cdots du_{r-1}^{(2)}du_1^{(3)}\cdots du_{r-1}^{(3)}d\xi, \tag{2.7.65}$$

where $H_i = H(A_i)$ for $i = 1, 2, 3$. Using the expression of (2.7.64) and the above density form (2.7.65), we can obtain the probability density function of $D(A_1), D(A_2)$ as follows:

$$f(y_1, y_2) = \mathcal{J}_1 \times f_1(y_1, y_2)\,, \tag{2.7.66}$$

where

$$\mathcal{J}_1 = \frac{\sqrt{H_1(1-H_1)H_2(1-H_2)}\Gamma(ra)}{ca\Gamma(raH_1)\Gamma(raH_2)\Gamma(raH_3)}\left(\sqrt{\frac{H_1(1-H_1)}{ca}}y_1 + H_1\right)^{raH_1-1}$$
$$\times\left(\sqrt{\frac{H_2(1-H_2)}{ca}}y_2 + H_2\right)^{raH_2-1}\left(H_3 - \frac{\sqrt{H_1(1-H_1)}y_1 + \sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}\right)^{raH_3-1} \tag{2.7.67}$$

and

$$f_1(y_1, y_2) = \int_0^\infty \int\cdots\int_{0\le u_1^{(1)}+\cdots+u_{r-1}^{(1)}\le 1}\int\cdots\int_{0\le u_1^{(2)}+\cdots+u_{r-1}^{(2)}\le 1}$$
$$\int\cdots\int_{0\le u_1^{(3)}+\cdots+u_{r-1}^{(3)}\le 1}\prod_{i=1}^{3}\left[u_1^i\cdots u_{r-1}^{(i)}(1-u_1^i-\cdots-u_{r-1}^{(i)})\right]^{aH_i-1}\xi^{ra-1}$$
$$\times\exp\left\{-\xi+\frac{\xi}{r}\left[\sum_{k=1}^{r-1}ku_k^{(1)}\left(\frac{\sqrt{H_1(1-H_1)}y_1}{\sqrt{ca}}+H_1\right)\right.\right.$$
$$+\sum_{k=1}^{r-1}ku_k^{(2)}\left(\frac{\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}+H_2\right)$$
$$\left.\left.+\sum_{k=1}^{r-1}ku_k^{(3)}\left(H_3-\frac{\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}\right)\right]\right\}$$

$$du_1^{(1)}\cdots du_{r-1}^{(1)}du_1^{(2)}\cdots du_{r-1}^{(2)}du_1^{(3)}\cdots du_{r-1}^{(3)}d\xi\,. \tag{2.7.68}$$

To determine if the density of $f(y_1, y_2)$ has a limit or not and if yes, to find the limiting density, we shall find the limits of $\mathcal{J}_1$ and $f_1(y_1, y_2)$ separately.

Step 1: Limit of $\mathcal{J}_1$.

By Stirling's formula $\Gamma(z) = \sqrt{2\pi}z^{z-\frac{1}{2}}e^{-z}\left(1+o\left(\frac{1}{z}\right)\right)$ when z is large, we have the following approximation:

$$\begin{aligned}
\mathcal{J}_1 =&\frac{ra}{ca2\pi}\sqrt{\frac{(1-H_1)(1-H_2)}{H_3}}\left(\sqrt{\frac{(1-H_1)}{caH_1}}y_1+1\right)^{raH_1-1}\left(\sqrt{\frac{(1-H_2)}{caH_2}}y_2+1\right)^{raH_2-1}\\
&\times\left(1-\frac{\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2}{H_3\sqrt{ca}}\right)^{raH_3-1}\left(1+o\left(\frac{1}{a}\right)\right),\\
=&\frac{r}{c2\pi}\sqrt{\frac{(1-H_1)(1-H_2)}{H_3}}\left(1+o\left(\frac{1}{a}\right)\right)\exp\left\{ra\left[H_1\log\left(\sqrt{\frac{(1-H_1)}{caH_1}}y_1+1\right)\right.\right.\\
&+H_2\log\left(\sqrt{\frac{(1-H_2)}{caH_2}}y_2+1\right)\\
&\left.\left.+H_3\log\left(1-\frac{\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2}{H_3\sqrt{ca}}\right)\right]\right\}.
\end{aligned}$$

An application of $\log(1+\frac{1}{\sqrt{z}}) = \frac{1}{\sqrt{z}} - \frac{1}{2z} + o\left(\frac{1}{z}\right)$ for large z yields

$$\begin{aligned}
\mathcal{J}_1 =&\frac{r}{c}\frac{\sqrt{(1-H_1)(1-H_2)}}{2\pi\sqrt{H_3}}\\
&\times\exp\left\{-\frac{r}{c}\frac{y_1^2+y_2^2+2y_1y_2\sqrt{\frac{H_1H_2}{(1-H_1)(1-H_2)}}}{2\frac{H_3}{(1-H_1)(1-H_2)}}\right\}\left(1+o\left(\frac{1}{a}\right)\right).
\end{aligned} \tag{2.7.69}$$

Step 2: The limit of $f_1(y_1, y_2)$. This is much more complicated. We first obtain the leading terms of f_1 as $a\to\infty$.

To make the presentation clear, denote the integrating variables by

$$\mathbf{z} = \left(u_1^{(1)},\cdots,u_{r-1}^{(1)},u_1^{(2)},\cdots,u_{r-1}^{(2)},u_1^{(3)},\cdots,u_{r-1}^{(3)},\xi\right)^T$$

and denote

$$
\begin{aligned}
&g\left(u_1^{(1)},\cdots,u_{r-1}^{(1)},u_1^{(2)},\cdots,u_{r-1}^{(2)},u_1^{(3)},\cdots,du_{r-1}^{(3)},\xi\right)\\
&\quad:=\sum_{i=1}^{3}(aH_i-1)\log\left[u_1^{(i)}\cdots u_{r-1}^{(i)}(1-u_1^{(i)}-\cdots-u_{r-1}^{(i)})\right]+(ra-1)\log(\xi)\\
&\quad-\xi+\frac{\xi}{r}\Bigg[\sum_{k=1}^{r-1}ku_k^{(1)}\left(\frac{\sqrt{H_1(1-H_1)}y_1}{\sqrt{ca}}+H_1\right)+\sum_{k=1}^{r-1}ku_k^{(2)}\left(\frac{\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}+H_2\right)\\
&\qquad+\sum_{k=1}^{r-1}ku_k^{(3)}\left(H_3-\frac{\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}\right)\Bigg]. \qquad (2.7.70)
\end{aligned}
$$

This function g attains its maximum at its critical point

$$
\mathbf{z_0}=\left(u_{1,0}^{(1)},\cdots,u_{r-1,0}^{(1)},u_{1,0}^{(2)},\cdots,u_{r-1,0}^{(2)},u_{1,0}^{(3)},\cdots,u_{r-1,0}^{(3)},\xi_0\right)^T, \qquad (2.7.71)
$$

where

$$
\begin{cases}
u_{k,0}^{(i)}=\dfrac{1}{(r-k)\left(\sum_{j=1}^{r}\frac{1}{j}\right)}, & k=1,\cdots,r-1,\ i=1,2,3;\\
\xi_0=(ra-1)\left(\displaystyle\sum_{j=1}^{r}\frac{1}{j}\right).
\end{cases}
$$

By an elementary calculation we have

$$
g'(\mathbf{z_0})=\begin{pmatrix}
\left[\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(1-\frac{H_1}{r}\right)+\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(a-\frac{1}{r}\right)\frac{\sqrt{H_1(1-H_1)}y_1}{\sqrt{ca}}\right]\mathbf{J}_{r-1}\\
\left[\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(1-\frac{H_2}{r}\right)+\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(a-\frac{1}{r}\right)\frac{\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}\right]\mathbf{J}_{r-1}\\
\left[\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(1-\frac{H_1}{r}\right)-\left(\sum_{j=1}^{r}\frac{1}{j}\right)\left(a-\frac{1}{r}\right)\frac{\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2}{\sqrt{ca}}\right]\mathbf{J}_{r-1}\\
0
\end{pmatrix},
$$

where $\mathbf{J}_{r-1} = (1, \cdots, r-1)^T$ and

$$g''(\mathbf{z_0}) = \begin{pmatrix} M_1 & & & N_1 \\ & M_2 & & N_2 \\ & & M_3 & N_3 \\ N_1^T & N_2^T & N_3^T & -\frac{1}{(ra-1)\left(\sum_{k=1}^{r-1}\frac{1}{k}\right)^2} \end{pmatrix},$$

where the empty entries should be filled with a $(r-1) \times (r-1)$ dimensional zero matrix and where for $i = 1, 2, 3$, M_i is a $(r-1) \times (r-1)$ matrix given by

$$M_i = -(aH_i - 1)\left(\sum_{k=1}^{r-1}\frac{1}{k}\right)^2 \begin{pmatrix} [(r-1)^2 + r^2] & r^2 & \cdots & r^2 \\ r^2 & [(r-2)^2 + r^2] & \cdots & r^2 \\ \vdots & \vdots & \ddots & \vdots \\ r^2 & r^2 & \cdots & [(1)^2 + r^2] \end{pmatrix},$$

and N_i is a $(r-1)$ column vector given by

$$N_i = \begin{pmatrix} \frac{1\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}} + H_i\right)}{r} \\ \frac{2\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}} + H_i\right)}{r} \\ \vdots \\ \frac{(r-1)\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}} + H_i\right)}{r} \end{pmatrix}.$$

With these notations we have when a is large

$$\begin{aligned} f_1(y_1, y_2) =& \frac{(r!)}{r^{ra}\Gamma(ra)} \frac{\Gamma(raH_1)\Gamma(raH_2)\Gamma(raH_3)}{[\Gamma(aH_1)\Gamma(aH_2)\Gamma(aH_3)]^r} \int_0^\infty \int \cdots \int_{0 \le u_1^{(1)} + \cdots + u_{r-1}^{(1)} \le 1} \\ & \int \cdots \int_{0 \le u_1^{(2)} + \cdots + u_{r-1}^{(2)} \le 1} \int \cdots \int_{0 \le u_1^{(3)} + \cdots + u_{r-1}^{(3)} \le 1} e^{g(\mathbf{z})} d\mathbf{z} \\ =& \frac{(r!)(aH_1)^{\frac{r-1}{2}}(aH_2)^{\frac{r-1}{2}}(aH_3)^{\frac{r-1}{2}}}{r^{\frac{3}{2}}(\sqrt{2\pi})^{3r-2}e^{-ra}(ra)^{ra-\frac{1}{2}}} \int_0^\infty \int \cdots \int_{0 \le u_1^{(1)} + \cdots + u_{r-1}^{(1)} \le 1} \\ & \int \cdots \int_{0 \le u_1^{(2)} + \cdots + u_{r-1}^{(2)} \le 1} \int \cdots \int_{0 \le u_1^{(3)} + \cdots + u_{r-1}^{(3)} \le 1} \exp\Big\{ g(\mathbf{z_0}) \end{aligned}$$

$$+ g'(\mathbf{z_0})^T(\mathbf{z}-\mathbf{z_0}) + \frac{1}{2}(\mathbf{z}-\mathbf{z_0})^T g''(\mathbf{z_0})(\mathbf{z}-\mathbf{z_0})\Big\}\left(1+o\left(\frac{1}{a}\right)\right)d\mathbf{z}\,. \quad (2.7.72)$$

Step 3: Evaluation of the leading term of $f_1(y_1, y_2)$.

In order to evaluate the integral of (2.7.72), we use the change of variables

$$u_k^{(i)} - u_{k,0}^{(i)} = \frac{t_k^{(i)}}{\sqrt{aH_i}\sqrt{(r-k)^2 + r^2\left(\sum_{k=1}^{r-1}\frac{1}{k}\right)}},$$

for $k = 1, \cdots, r-1$, $i = 1, 2, 3$, and

$$\xi - \xi_0 = \sqrt{ra-1}\left(\sum_{k=1}^{r-1}\frac{1}{k}\right)s\,.$$

Thus, we have

$$\begin{aligned}
f_1(y_1,y_2) =& \frac{(r!)(aH_1)^{\frac{r-1}{2}}(aH_2)^{\frac{r-1}{2}}(aH_3)^{\frac{r-1}{2}}}{r^{\frac{3}{2}}(\sqrt{2\pi})^{3r-2}e^{-ra}(ra)^{ra-\frac{1}{2}}}\sqrt{ra-1}\left(\sum_{k=1}^{r}\frac{1}{k}\right)\\
&\times\prod_{i=1}^{3}\prod_{k=1}^{r-1}\frac{1}{\sqrt{aH_i}\sqrt{(r-k)^2+r^2\left(\sum_{k=1}^{r}\frac{1}{k}\right)}}\\
&\times\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{g(\mathbf{z_0})+\mathbf{b}^T\mathbf{t}-\frac{1}{2}\mathbf{t}^TA\mathbf{t}\right\}\left(1+o\left(\frac{1}{a}\right)\right)d\mathbf{t}\\
=&\frac{(r!)^3\left(\sum_{k=1}^{r}\frac{1}{k}\right)^3}{r^{\frac{3}{2}}\prod_{k=1}^{r-1}\left(\sqrt{(r-k)^2+r^2}\right)^3}(\det A)^{-\frac{1}{2}}\\
&\qquad\times\exp\left\{\frac{1}{2}\mathbf{b}^TA^{-1}\mathbf{b}\right\}\left(1+o\left(\frac{1}{a}\right)\right),
\end{aligned} \quad (2.7.73)$$

where $\mathbf{t} = \left(t_1^{(1)}, \cdots, t_{r-1}^{(2)}, t_1^{(2)}, \cdots, t_{r-1}^{(2)}, t_1^{(3)}, \cdots, t_{r-1}^{(3)}, s\right)^T$, and where a direct calculation from $g'(\mathbf{z_0})$ gives

$$\mathbf{b} = \begin{pmatrix} -\frac{\sqrt{r}\sqrt{1-H_1}y_1}{\sqrt{c}}B \\ -\frac{\sqrt{r}\sqrt{1-H_2}y_2}{\sqrt{c}}B \\ -\frac{\sqrt{r}\left(\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2\right)}{\sqrt{c}\sqrt{H_3}}B \\ 0 \end{pmatrix}. \quad (2.7.74)$$

The matrix A in (2.7.73) can be found directly from $g''(\mathbf{z_0})$ and will be given below when we study it. The above last identity (2.7.73) follows from the fact that

$$e^{g(\mathbf{z_0})} = \frac{\left[(ra-1)\left(\sum_{k=1}^{r}\frac{1}{k}\right)\right]^{ra-1}}{\left[r!\left(\sum_{k=1}^{r}\frac{1}{k}\right)^{r}\right]^{a-3}} e^{-(ra-1)}$$

and the multivariate Gaussian integral formula. By a simple algebra we can write

$$A = \begin{pmatrix} A_0 & & & B_1 \\ & A_0 & & B_2 \\ & & A_0 & B_3 \\ B_1^T & B_2^T & B_3^T & 1 \end{pmatrix}, \tag{2.7.75}$$

where A_0 is a $(r-1)\times(r-1)$ matrix whose entries are

$$[A_0]_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{r^2}{\sqrt{(r-i)^2+r^2}\sqrt{(r-j)^2+r^2}} & \text{if } i \neq j\,. \end{cases}$$

and for $i = 1, 2, 3$

$$B_i = -\begin{pmatrix} \frac{1\sqrt{ra-1}\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}}+H_i\right)}{r\sqrt{aH_i}\sqrt{(r-1)^2+r^2}} \\ \frac{2\sqrt{ra-1}\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}}+H_i\right)}{r\sqrt{aH_i}\sqrt{(r-2)^2+r^2}} \\ \vdots \\ \frac{(r-1)\sqrt{ra-1}\left(\frac{\sqrt{H_i(1-H_i)}y_i}{\sqrt{ca}}+H_i\right)}{r\sqrt{aH_i}\sqrt{(1)^2+r^2}} \end{pmatrix} = -\begin{pmatrix} \frac{1\sqrt{H_1}}{\sqrt{r}\sqrt{(r-1)^2+r^2}}\left(1+o\left(\frac{1}{\sqrt{a}}\right)\right) \\ \frac{2\sqrt{H_1}}{\sqrt{r}\sqrt{(r-2)^2+r^2}}\left(1+o\left(\frac{1}{\sqrt{a}}\right)\right) \\ \vdots \\ \frac{(r-1)\sqrt{H_1}}{\sqrt{r}\sqrt{(1)^2+r^2}}\left(1+o\left(\frac{1}{\sqrt{a}}\right)\right) \end{pmatrix}.$$

Step 4: The inverse and the determinant of A

We need to find A^{-1}. First we find A_0^{-1}. From the expression of A_0 we can write

$$A_0 = D + \mathbf{v}\mathbf{v}^T = D^{\frac{1}{2}}\left(I + D^{-\frac{1}{2}}\mathbf{v}\mathbf{v}^T D^{-\frac{1}{2}}\right)D^{\frac{1}{2}},$$

where

$$\mathbf{v} = \begin{pmatrix} \frac{r}{\sqrt{(r-1)^2+r^2}} \\ \frac{r}{\sqrt{(r-2)^2+r^2}} \\ \vdots \\ \frac{r}{\sqrt{(1)^2+r^2}} \end{pmatrix}$$

and

$$D = \begin{pmatrix} \frac{(r-1)^2}{(r-1)^2+r^2} & 0 & \cdots & 0 \\ 0 & \frac{(r-2)^2}{(r-2)^2+r^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1^2}{1^2+r^2} \end{pmatrix}.$$

Hence, the determinant of A_0 is given by

$$\begin{aligned} \det(A_0) &= \det\left(D^{\frac{1}{2}}\right)\det\left(I + D^{-\frac{1}{2}}\mathbf{v}\mathbf{v}^T D^{-\frac{1}{2}}\right)\det\left(D^{\frac{1}{2}}\right) \\ &= \det\left(D^{\frac{1}{2}}\right)\left(1 + \left(D^{-\frac{1}{2}}\mathbf{v}\right)^T\left(D^{-\frac{1}{2}}\mathbf{v}\right)\right)\det\left(D^{\frac{1}{2}}\right) \\ &= \left(\frac{(r-1)!}{\prod_{k=1}^{r-1}\sqrt{(r-k)^2+r^2}}\right)^2\left(1 + \sum_{k=1}^{r-1}\frac{r^2}{(r-k)^2}\right) \\ &= \frac{\sum_{k=1}^{r}(\frac{r!}{k})^2}{\prod_{k=1}^{r-1}(\sqrt{(r-k)^2+r^2})^2}. \end{aligned}$$

Now, by the Sherman-Morrison formula we have

$$A_0^{-1} = (D + ff^T)^{-1} = D^{-1} - \frac{1}{1 + f^T D^{-1} f} D^{-1} f f^T D^{-1}.$$

As a result we obtain

$$[A_0^{-1}]_{ij} = \begin{cases} -\frac{1}{\sum_{k=1}^{r}(\frac{1}{k})^2}\frac{\sqrt{(r-i)^2+r^2}\sqrt{(r-j)^2+r^2}}{(r-i)^2(r-j)^2} & \text{if } i \neq j \\ \frac{[\sum_{k=1}^{r}(\frac{1}{k})^2 - \frac{1}{(r-i)^2}]}{\sum_{k=1}^{r}(\frac{1}{k})^2}\frac{[(r-i)^2+r^2]}{(r-i)^2} & \text{if } i = j. \end{cases}$$

The determinant of A can be computed as follows.

$$\begin{aligned}\det(A) &= (\det A_0)^3 \left(1 - B_1^T A_0^{-1} B_1 - B_2^T A_0^{-1} B_2 - B_3^T A_0^{-1} B_3\right) \\ &= (\det A_0)^3 \left(1 - B^T A_0^{-1} B\right) ,\end{aligned} \tag{2.7.76}$$

where

$$B = -\begin{pmatrix} \frac{1}{\sqrt{r}\sqrt{(r-1)^2+r^2}} \left(1 + o\left(\frac{1}{\sqrt{a}}\right)\right) \\ \frac{2}{\sqrt{r}\sqrt{(r-2)^2+r^2}} \left(1 + o\left(\frac{1}{\sqrt{a}}\right)\right) \\ \vdots \\ \frac{(r-1)}{\sqrt{r}\sqrt{(1)^2+r^2}} \left(1 + o\left(\frac{1}{\sqrt{a}}\right)\right) \end{pmatrix}.$$

From the relation (2.7.76) of expressing $\det(A)$ by $\det(A_0)$ and B we have

$$\det(A) = (\det A_0)^3 \left(1 - B^T A_0^{-1} B\right) = \frac{\left(\sum_{k=1}^r \frac{r!}{k}\right)^2 \left(\sum_{k=1}^r \left(\frac{r!}{k}\right)^2\right)^2}{r \prod_{k=1}^{r-1} \left(\sqrt{(r-k)^2 + r^2}\right)^6}. \tag{2.7.77}$$

Moreover, by the block Gaussian elimination method, we find A^{-1} as

$$\begin{pmatrix} A_0^{-1}\left(I_0 + \frac{B_1 B_1^T A_0^{-1}}{m}\right) & \frac{A_0^{-1} B_1 B_2^T A_0^{-1}}{m} & \frac{A_0^{-1} B_1 B_3^T A_0^{-1}}{m} & -\frac{A_0^{-1} B_1}{m} \\ \frac{A_0^{-1} B_2 B_1^T A_0^{-1}}{m} & A_0^{-1}\left(I_0 + \frac{B_2 B_2^T A_0^{-1}}{m}\right) & \frac{A_0^{-1} B_2 B_3^T A_0^{-1}}{m} & -\frac{A_0^{-1} B_2}{m} \\ \frac{A_0^{-1} B_3 B_1^T A_0^{-1}}{m} & \frac{A_0^{-1} B_3 B_2^T A_0^{-1}}{m} & A_0^{-1}\left(I_0 + \frac{B_3 B_3^T A_0^{-1}}{m}\right) & -\frac{A_0^{-1} B_3}{m} \\ -\frac{B_1^T A_0^{-1}}{m} & -\frac{B_2^T A_0^{-1}}{m} & -\frac{B_3^T A_0^{-1}}{m} & \frac{1}{m} \end{pmatrix}, \tag{2.7.78}$$

where

$$m = 1 - B_1^T A_0^{-1} B_1 - B_2^T A_0^{-1} B_2 - B_3^T A_0^{-1} B_3$$

and I_0 is the $(r-1) \times (r-1)$-dimensional identity matrix.

Step 5: the limit of $f_1(y_1, y_2)$.

Combining (2.7.78) and (2.7.74), one finds

$$\mathbf{b}^T A^{-1} \mathbf{b} = \frac{r}{c} B^T A_0^{-1} B \Bigg((1 - H_1) y_1^2 + (1 - H_2) y_2^2$$

$$+\frac{\left(\sqrt{H_1(1-H_1)}y_1+\sqrt{H_2(1-H_2)}y_2\right)^2}{H_3}\Bigg)$$

$$=(\frac{r}{c}-1)\frac{y_1^2+y_2^2+2y_1y_2\sqrt{\frac{H_1H_2}{(1-H_1)(1-H_2)}}}{\frac{H_3}{(1-H_1)(1-H_2)}}\left(1+o\left(\frac{1}{a}\right)\right). \tag{2.7.79}$$

Substituting the expression (2.7.79) and the formula (2.7.77) for the determinant into (2.7.73) yields

$$f_1(y_1,y_2)=\frac{c}{r}\exp\left\{\left(\frac{r}{c}-1\right)\frac{y_1^2+y_2^2+2y_1y_2\sqrt{\frac{H_1H_2}{(1-H_1)(1-H_2)}}}{2\frac{H_3}{(1-H_1)(1-H_2)}}\right\}\left(1+o\left(\frac{1}{a}\right)\right). \tag{2.7.80}$$

Combining (2.7.80) with the asymptotic (2.7.69) of $\mathcal{J}_1$ and the relationship $f(y_1,y_2)=\mathcal{J}_1\times f_1(y_1,y_2)$, we see $f(y_1,y_2)$ converges to the desired Gaussian density, which completes the proof of part (i) of the theorem when $P\sim\text{GDP}(a,r,H)$.

The proof of part (ii) of this theorem follows from the same argument as that in the proof of part (ii) of Theorem 4.4. □

Chapter 3

Large sample asymptotic analysis for normalized random measures with independent increments

3.1 Introduction

Bayesian nonparametrics has been undergone major study due to its various applications to many areas, such as biology, economics, machine learning and so on. As a lavish class of Bayesian nonparametric priors, normalized random measures with independent increments (NRMIs), introduced by (Regazzini et al., 2003), include the famous Dirichlet process (Ferguson, 1973), the σ-stable NRMIs (Kingman, 1975), the normalized inverse Gaussian process (Lijoi et al., 2005b), the normalized generalized gamma process (Lijoi et al., 2003, 2007), and generalized Dirichlet process (Lijoi et al., 2005a). We refer to (Müller and Quintana, 2004; Lijoi et al., 2010; Zhang and Hu, 2021) as reviews of these processes with their properties and applications.

In Bayesian nonparametric statistics, samples are drawn from a random probability measure that is equipped with a prior distribution. To be more precise, let $(\Omega, \mathcal{F}, \mathbb{P})$ be any probability space, let $\mathbb{X}$ be a complete, separable metric space whose σ-algebra is denoted by $\mathcal{X}$ and let $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ be the space of all probability measures on $\mathbb{X}$. A sample

$\mathbf{X} = (X_1, \cdots, X_n)$ that take values in $\mathbb{X}^n$ is drawn iid from a random probability measure P conditional on P, which follows a prior distribution Q on $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$. That is to say,

$$X_1, \cdots, X_n | P \overset{iid}{\sim} P; \qquad P \sim Q. \tag{3.1.1}$$

Two natural questions in literature are raised as follows.

(i) A frequentist analysis of the Bayesian consistency (Freedman and Diaconis, 1983): by assuming the "true" distribution of $\mathbf{X}$ is P_0, we are interested in whether the posterior law, that is the conditional law of $P|\mathbf{X}$, denoted by Q_n, converges to δ_{P_0}, the Dirac mass at the "true" distribution, as $n \to \infty$.

(ii) What is the limiting distribution of centered and rescaled $P|\mathbf{X}$? In particular, is there a Bernstein-von Mises like theorem and central limit theorem for P? If so, what is the limiting process of $\sqrt{n}(P|\mathbf{X} - \mathbb{E}[P|\mathbf{X}])$?

These two questions are always very important in statistics, as the posterior consistency can guarantee the model behaves "good" when the sample size is sufficiently large, and the limiting distribution of the posterior process is the key to construct confident intervals and conduct hypothesis tests.

Many inspiring works corresponding to the above questions have been done. Referred to question (i), (James, 2008) obtains the posterior consistency and weak convergence for the two-parameter Poisson-Dirichlet process, which is not, but closely related to an NRMI (Pitman and Yor, 1997; Perman et al., 1992; Ghosal and Van der Vaart, 2017). The posterior consistency of the species sampling priors (Pitman, 1996; Aldous et al., 1985) and the Gibbs-type priors (Gnedin and Pitman, 2006) are discussed in (Ho Jang et al., 2010) and (De Blasi et al., 2013). It is worthy to point out that there are overlaps among the species sampling priors, the Gibbs-type priors and the homogeneous NRMIs. Whereas, non-homogeneous NRMIs are totally independent from the species sampling priors and the Gibbs-type priors. As for question (ii), the Bernstein-von Mises results have been established for the Dirichlet process (Lo, 1983, 1986; Ray and van der Vaart, 2021; Hu and Zhang, 2022) and for the two-parameter Poisson-Dirichlet process (James, 2008; Franssen and van der Vaart, 2022). Along the same line, we would like to answer the above two mentioned questions when P is an NRMI.

Since NRMIs are constructed by the normalization of completely random measures (Kingman, 1967, 1993) associated with their Lévy intensities (see e.g., section 3.2), it is quite natural to study their properties based on the corresponding Lévy intensities. In this work, we discuss the posterior consistency of non-homogeneous NRMIs (including the homogeneous case as a particular case) and provide a simple condition to guarantee the posterior consistency of non-homogeneous NRMIs. As a result, when P_0 is continuous, the posterior consistency doesn't hold for NRMIs generally, and when P_0 is discrete, the posterior consistency holds as long as our proposed condition is satisfied.

Furthermore, we obtain the Bernstein-von Mises theorem for the normalized generalized gamma process (NGGP), which is a flexible class of Bayesian nonparametric priors includes the Dirichlet process, the normalized inverse-Gaussian process and the $\sigma-$stable process. Through the posterior consistency analysis, the NGGP is posterior consistent when the true distribution P_0 is discrete or when the parameter $\sigma \to 0$. The case that $\sigma \to 0$ would reduce the NGGP to the Dirichlet process. Thus, we should emphasis the case when the true distribution P_0 is discrete. However, there should be a bias correction when we use the Bernstein-von Mises theorem for the NGGP when P_0 is discrete. Thus, in order to construct the "correct" confidence intervals that cover the true parameter value, we suggest to de-bias the bias term. The comparison of confidence intervals with bias correction and without bias correction is illustrated in the numerical computation. In the application, the model parameters of NGGP are chosen by some estimators, we show that the Bayesian estimator or maximum likelihood estimators of the model parameters of the NGGP won't affect the convergences in the Bernstein-von Mises results.

The outline of this chapter is as follows. In Section 3.2, we recall the construction of the NRMIs and their posterior distributions. In Section 3.3, we discuss the posterior consistency of the homogeneous NRMIs and present a simple assumption on the corresponding Lévy intensities to guarantee the posterior consistency of the homogeneous

NRMIs. Examples are given to see the applicability of the assumption and the posterior consistency results for some well-known Bayesian nonparametric priors. In Section 3.4, we derive the Bernstein-von Mises theorem for the NGGP and provide the analysis of the bias correction with an numerical illustration. Finally, in Section 3.5, we provides a discussion of our results and some ideas that can be studied in the future. In order to ease the flow of the ideas, we delay the proofs in the Supplementary Materials 3.6.

3.2 Normalized random measures with independent increments

3.2.1 Constructions of NRMIs

We start by recalling the notions of completely random measures (see e.g., (Kingman, 1967, 1993) and references therein for more details), which play important role in the construction of NRMIs.

Let $\mathbb{B}_{\mathbb{X}}$ be the space of bounded finite measures on $(\mathbb{X}, \mathcal{X})$ endowed with a suitable topology so that the associated Borel $\sigma-$algebra $\mathcal{B}_{\mathbb{X}}$ can be introduced (Daley and Vere-Jones, 2008).

Definition 3.2.1. *Let μ be a measurable function defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$. We call μ is a completely random measure (CRM) if the random variables $\mu(A_1), \cdots, \mu(A_d)$ are mutually independent, for any pairwise disjoint sets $A_1, \cdots, A_d$, where $d \geq 2$ is a finite integer.*

See e.g., (Regazzini et al., 2003; Lijoi et al., 2010) for a more detailed discussion of constructing Bayesian nonparametric priors by using completely random measures.

As CRMs can be defined via Poisson random measure, we shall first recall this concept. Denote $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$ and denote its Borel σ-algebra by $\mathcal{S}$. A Poisson random measure $\tilde{N}$ on $\mathbb{S}$ with finite intensity measure $\nu(ds, dx)$ is a random measure from $\Omega \times \mathbb{S}$ to $\mathbb{R}_+$ satisfying

(i) $\tilde{N}(A) \sim \text{Poisson}(\nu(A))$ for any A in $\mathcal{S}$;

(ii) for any pairwise disjoint sets $A_1, \cdots, A_m$ in $\mathcal{S}$, the random variables $\tilde{N}(A_1), \cdots, \tilde{N}(A_m)$ are mutually independent.

The Poisson intensity measure ν satisfies the condition (see (Daley and Vere-Jones, 2008) for details of Poisson random measures) that

$$\int_0^\infty \int_{\mathbb{X}} \min(s, 1)\nu(ds, dx) < \infty\,.$$

Let $\tilde{\mu}$ be the random measure on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$ defined as follows,

$$\tilde{\mu}(A) := \int_0^\infty \int_A s\tilde{N}(ds, dx), \quad \forall A \in \mathcal{X}\,. \tag{3.2.1}$$

It is trivial to verify that $\tilde{\mu}$ is a completely random measure. It is also well-known that for any $B \in \mathcal{X}$, $\tilde{\mu}(B)$ is discrete and is uniquely characterized by its Laplace transform as follows:

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(B)}\right] = \exp\left\{-\int_0^\infty \int_B \left[1 - e^{-\lambda s}\right] \nu(ds, dx)\right\}\,. \tag{3.2.2}$$

The measure ν is called the *Lévy intensity* of $\tilde{\mu}$ and we denote the Laplace exponent by

$$\psi_B(\lambda) = \int_0^\infty \int_B \left[1 - e^{-\lambda s}\right] \nu(ds, dx)\,. \tag{3.2.3}$$

From the Laplace transform in eq. (4.2.2), we shall study the completely random measure $\tilde{\mu}$ by its Lévy intensity ν, which usually takes the following forms in the literature:

(a) $\nu(ds, dx) = \rho(ds)\alpha(dx)$, where $\rho : \mathcal{B}(\mathbb{R}^+) \to \mathbb{R}^+$ is some measure on $\mathbb{R}^+$ and α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ so that $\alpha(\mathbb{X}) = a < \infty$. The corresponding $\tilde{\mu}$ is called *homogeneous* completely random measure.

(b) $\nu(ds, dx) = \rho(ds|x)\alpha(dx)$, where ρ is defined on $\mathcal{B}(\mathbb{R}^+) \times \mathbb{X}$ such that for any $x \in \mathbb{X}$, $\rho(\cdot|x)$ is a σ-finite measure on $\mathcal{B}(\mathbb{R}^+)$ and for any $A \in \mathcal{X}$, $\rho(A|x)$ is $\mathcal{B}(\mathbb{R}^+)$

measurable. The corresponding $\tilde{\mu}$ is called *non-homogeneous* completely random measure.

It is obvious that the case (a) is a special case of case (b). Usually, we assume that α is a finite measure so we may write $\alpha(dx) = aH(dx)$ for some probability measure H and some constant $a = \alpha(\mathbb{X}) \in (0, \infty)$.

To construct NRMIs, the completely random measure will be normalized, and thus one needs the total mass $\tilde{\mu}(\mathbb{X})$ to be finite and positive almost surely. This happens under the condition that $\rho(\mathbb{R}^+) = \infty$ in homogeneous case and that $\rho(\mathbb{R}^+|x) = \infty$ in non-homogeneous case (Regazzini et al., 2002). Under the above conditions, an NRMI P on $(\mathbb{X}, \mathcal{X})$ is a random probability measure defined by

$$P(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}\,. \tag{3.2.4}$$

P is discrete due to the discreteness of $\tilde{\mu}$. For notional simplicity, we denote $T = \tilde{\mu}(\mathbb{X})$ and let $f_T(t)$ be the density of T throughout this chapter.

3.2.2 Posterior of NRMIs

We will recall the posterior analysis (James et al., 2009) of NRMIs, which is a key topic in Bayesian nonparametric analysis. Let P be an NRMI on $\mathbb{X}$. A sample of size n from P as in eq. (3.1.1) is an exchangeable sequence of random variables $\mathbf{X} = (X_i)_{i=1}^n$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathbb{X}^n$, such that given P, $(X_i)_{i\geq 1}$ are iid with distribution P, i.e.,

$$\mathbb{P}[X_1 \in A_1, \cdots, X_n \in A_n | P] = \prod_{i=1}^{n} P(A_i)\,. \tag{3.2.5}$$

Let $\mathbf{Y} = (Y_j)_{j=1}^{n(\pi)}$ be the distinct observations of the sample $\mathbf{X}$ and let $n(\pi)$ be the number of unique values of $\mathbf{X}$. This means, $\pi = (i_1, \cdots, i_{n_1}, \cdots, i_{n_{\pi(n)-1}}, \cdots, i_{n_{n(\pi)}})$ is the partition of $\{1, \cdots, n\}$ of size $n(\pi)$. The number of the jth set of the partition is n_j, so that

$\sum_{j=1}^{n(\pi)} n_j = n$, and $Y_1 := X_{i_1} = \cdots = X_{i_{n_1}}, \cdots, Y_{n(\pi)} := X_{n_{\pi(n)-1}+1} = \cdots = X_{n_{\pi(n)}}$. Let

$$\tau_k(u, Y) = \int_0^\infty s^k e^{-us} \rho(ds|Y) \quad \text{for any positive integer } k \text{ and } Y \in \mathbb{X}. \tag{3.2.6}$$

With these notations, the posterior distribution of P conditional on the observations of the sample $X_1, \cdots, X_n$ is given by the following theorem.

Theorem 3.2.2 (James et al. (2009)). *Let P be an NRMI with intensity $\nu(ds, dx) = \rho(ds|x)\alpha(dx)$. The posterior distribution of P, given a latent random variable U_n, is an NRMI that coincides in distribution with the random measure*

$$\kappa_n \frac{\tilde{\mu}_{(U_n)}}{T_{(U_n)}} + (1 - \kappa_n) \sum_{j=1}^{n(\pi)} \frac{J_j \delta_{Y_j}}{\sum_{j=1}^{n(\pi)} J_j}, \tag{3.2.7}$$

where

(i) The random variable U_n has density

$$f_{U_n}(u) = \frac{u^{n-1}}{\Gamma(n)} \int_0^\infty t^n e^{-ut} f_T(t) dt. \tag{3.2.8}$$

(ii) Given U_n, $\tilde{\mu}_{(U_n)}$ is the conditional completely random measure of $\tilde{\mu}$ with the Lévy intensity $\nu_{(U_n)} = e^{-U_n s}\rho(ds|x)\alpha(dx)$.

(iii) $\{J_1, \cdots, J_{n(\pi)}\}$ are random variables depending on U_n and Y_j and having density

$$f_{J_j}(s|U_n = u, \boldsymbol{X}) = \frac{s^{n_j} e^{-us} \rho(s|Y_j)}{\int_0^\infty s^{n_j} e^{-us} \rho(ds|Y_j)}. \tag{3.2.9}$$

(iv) The random elements $\tilde{\mu}_{(U_n)}$ and J_j, $j \in \{1, \cdots, n(\pi)\}$ are independent.

(v) $T_{(U_n)} = \tilde{\mu}_{(U_n)}(\mathbb{X})$ and $\kappa_n = \frac{T_{(U_n)}}{T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j}$.

(vi) The conditional density of U_n given $\boldsymbol{X}$ is given by

$$f_{U_n|\boldsymbol{X}}(u|\boldsymbol{X}) \propto u^{n-1} e^{-\psi(u)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(u, Y_j). \tag{3.2.10}$$

The above theorem shows that, given the latent variable U_n, the posterior of P is a weighted sum of another NRMI $\frac{\tilde{\mu}_{(U_n)}}{T_{(U_n)}}$ and the normalization of Delta measure of distinct observations Y_j, δ_{Y_j}, multiplied by its corresponding jumps J_j. This gives a rather complete description of the posterior distribution of NRMIs. More details of the posterior analysis of $\tilde{\mu}$ and P is discussed in (James et al., 2009).

3.3 Posterior consistency analysis for the NRMIs

In this section, we aim at discussing the posterior consistency for NRMIs as pointed out in question (i) in the introduction. Assume that $\mathbf{X} = \{X_1, \cdots, X_n\}$ is a sample from the "true" distribution P_0 in $\mathbb{M}_{\mathbb{X}}$. Namely, $\mathbf{X} = \{X_1, \cdots, X_n\}$ is iid P_0- distributed. Let Q_n denote the probability law of the posterior random probability measure $P|\mathbf{X}$. The posterior distribution is said to be weakly consistent if Q_n concentrates on the weak neighbourhood of P_0 almost surely. More precisely, for any weak neighbourhood $O_\epsilon \in \mathcal{M}_{\mathbb{X}}$ of P_0 with arbitrary radius $\epsilon > 0$,

$$Q_n(O_\epsilon) \to 1 \qquad a.s. - P_0^n\,,$$

as $n \to \infty$. And P_0^∞ is the infinite product measure on $\mathbb{X}^\infty$.

Before presenting the main result, we shall give the following lemma, which provides the moments of the posterior P. And the lemma plays an important role in the proof of the main theorem. By recalling ψ_A in eq. (4.2.3), we denote

$$V_{\alpha(A)}^{(k)}(y) = (-1)^k e^{\psi_A(y)} \frac{d^k}{dy^k} e^{-\psi_A(y)}\,, \tag{3.3.1}$$

for any $A \in \mathcal{X}$.

Lemma 3.3.1. *Let $\mathbf{X} = (X_i)_{i=1}^n$ be a random sample from a normalized random measure with independent increments P. The moments and the mixed moments of the posterior moments of P given $\mathbf{X}$ are given as follows (we use the notation of Theorem 3.2.2).*

(i) For any $A \in \mathcal{X}$ and $m \in \mathbb{N}$, the posterior moments of P are given by

$$\mathbb{E}[(P(A))^m|\mathbf{X}] = \frac{\Gamma(n)}{\Gamma(m+n)} \sum_{0 \le l_1+\cdots+l_{n(\pi)} \le m}^{m} \binom{m}{l_1, \cdots, l_{n(\pi)}} \int_0^\infty u^m f_{U_n|\mathbf{X}}(u|\mathbf{X}) \\ V_{\alpha(A)}^{(m-(l_1+\cdots+l_{n(\pi)}))}(u) \left(\prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A) \right) du \, . \tag{3.3.2}$$

(ii) For any family of pairwise disjoint subsets $\{A_1, \cdots, A_q\}$ of $\mathcal{X}$ and any integers $\{m_1, \cdots, m_q\}$, we have

$$\mathbb{E}\left[P(A_1)^{m_1} \cdots P(A_q)^{m_q}|\mathbf{X}\right] = \frac{\Gamma(n)}{\Gamma(m+n)} \int_0^\infty u^m f_{U_n|\mathbf{X}}(u|\mathbf{X}) \\ \prod_{i=1}^{q+1} \left\{ \sum_{0 \le l_1+\cdots+l_{\#(\lambda_i)} \le m_i}^{m_i} \binom{m_i}{l_1, \cdots, l_{\#(\lambda_i)}} V_{\alpha(A_i)}^{(m_i-(l_1+\cdots+l_{\#(\lambda_i)}))}(u) \left(\prod_{j \in \lambda_i} \frac{\tau_{n_j+l_j}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \right) \right\} du \, , \tag{3.3.3}$$

where $m = \sum_{i=1}^q m_i$, $A_{q+1} = (\cup_{i=1}^q A_i)^c$, $m_{q+1} = 0$, $\lambda_i = \{j : Y_j \in A_i\}$ is the set of the index of Y_j's that are in A_i, and $\#(\lambda_i)$ is the number of components in λ_i.

The above lemma provides the posterior moments of NRMIs without including the latent random variable U_n. Such results can be reduced to the moments of NRMIs by letting the sample size $n = 0$. The proof of lemma 3.3.1 is inspired by the idea in (James et al., 2006) and the details are given in the Supplementary Materials 3.6. To apply the above lemma, one needs to deal with the term $V_{\alpha(A)}^{(k)}(y)$ defined by (3.3.1). We give the following recursion formula

$$V_{\alpha(A)}^{(k)}(y) = \sum_{i=0}^{k-1} \binom{k-1}{i} \xi_{k-i}(y) V_{\alpha(A)}^{(i)}(y) \, ,$$

where $\xi_i(y) = \int_A \tau_i(y, x) \alpha(dx)$.

To answer question (i) that is mentioned in the introduction, we shall study the weak consistency for more general NRMIs. To do so, we need the following assumption.

Assumption 3.3.2. *Let $\tau_k(u,x)$ be defined by (3.2.6) and let $\rho(s|x)$ be a function such that $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)}$ is nondecreasing in u and bounded from above by $k - C_k(x)$ uniformly for all $k \in \mathbb{Z}^+$ and $x \in \mathbb{X}$, where $\{C_k(x)\}$ is a sequence of functions from $\mathbb{X}$ to $[0,1)$. Namely, there is an increasing positive function $\phi(u)$ with $\lim_{u\to\infty} \phi(u) = 1$ such that*

$$\sup_{k\in\mathbb{Z}^+, x\in\mathbb{X}} \frac{u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)}}{k - C_k(x)} \leq \phi(u) , \quad \forall\ u \in \mathbb{R}_+ .$$

Theorem 3.3.3. *Let P be an NRMI with Lévy intensity $\nu(ds, dx) = \rho(s|x)ds\alpha(dx)$, where $\rho_x(s)$ satisfies Assumption 3.3.2. Then, we have the following results.*

1. *If P_0 is continuous, then the posterior of P converges weakly to a point mass at $\bar{C}_1 H(\cdot) + (1 - \bar{C}_1)P_0(\cdot)$ a.s.$-P_0^\infty$, where $\bar{C}_1$ is the population mean of $\{C_1(X_i)\}_{i=1}^\infty$, that is to say $\bar{C}_1 = \lim_{n\to\infty} \frac{\sum_{i=1}^n C_1(X_i)}{n}$.*

2. *If P_0 is discrete with $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$, then P is weakly consistent, i.e., the posterior of P converges weakly to a point mass at $P_0(\cdot)$ a.s.$-P_0^\infty$.*

Although the assumption 3.3.2 looks complicate, it is quite easy to check as long as $\rho(s|x)$ is given. For instance, the intensities $\rho(s|x)$ for almost all popular NRMIs are gamma type, and we shall check assumption 3.3.2 for these NRMIs in example 3.3.8, example 3.3.9 and example 3.3.10 to show how the assumption 3.3.2 works for these processes. This allows more applicability of Theorem 3.3.3.

As a comparison between Theorem 3.3.3 and the results in (Ho Jang et al., 2010) for the species sampling priors and (De Blasi et al., 2013) for the Gibbs-type priors, Theorem 3.3.3 considers the consistency results for the non-homogeneous NRMIs, which is a more general class of Bayesian nonparametric priors than both the species sampling priors and the Gibbs-type priors. On the other hand, the conditions in (Ho Jang et al., 2010; De Blasi et al., 2013) are not trivial to verify for homogeneous NRMIs, even though the predictive distribution of homogeneous NRMIs is given (Pitman, 2003; James et al., 2009).

In Theorem 3.3.3, we require $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ as a condition to guarantee the posterior consistency result when P_0 is discrete. This condition is true almost surely by the following remark.

Remark 3.3.4. *When P_0 is discrete, $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$, almost surely. When P_0 is continuous, $\lim_{n\to\infty} \frac{n(\pi)}{n} = 1$, almost surely.*

Proof. Note that P_0 is the true distribution of $\mathbf{X}$, i.e., $\mathbf{X} \overset{iid}{\sim} P_0$. Recall that $\pi(n)$ is the number of distinct observations of $\mathbf{X}$. Let $\mathbb{P}_n(\cdot) = \frac{\sum_{i=1}^n \delta_{X_i}(\cdot)}{n}$ be the empirical probability measure.

If P_0 is discrete, we denote the collection of atoms of P_0 is $\mathbb{D}$, then $\mathbb{D} = \{x_1, x_2, \cdots\}$. For any $k \in \mathbb{Z}^+$, we have $\pi(n) \leq k + n\mathbb{P}_n(\{x_{k+1}, x_{k+2}, \cdots\})$. Thus,

$$\begin{aligned}
\lim_{n\to\infty} \frac{\pi(n)}{n} &= \lim_{k\to\infty} \lim_{n\to\infty} \frac{\pi(n)}{n} \\
&\leq \lim_{k\to\infty} \lim_{n\to\infty} \frac{k}{n} + \mathbb{P}_n(\{x_{k+1}, x_{k+2}, \cdots\}) \\
&= \lim_{k\to\infty} P_0(\{x_{k+1}, x_{k+2}, \cdots\}) = 0\,,
\end{aligned}$$

almost surely, where we use the Borel–Cantelli lemma when taking the limit of $n \to \infty$.

If P_0 is continuous, we have $\pi(n) = n\mathbb{P}_n(\mathbb{X})$ and thus $\lim_{n\to\infty} \frac{\pi(n)}{n} = \lim_{n\to\infty} \mathbb{P}_n(\mathbb{X}) = P_0(\mathbb{X}) = 1$, almost surely. □

By the identity that $\frac{d}{du}\tau_k(u,x) = \frac{d}{du}\int_0^\infty s^k e^{-us}\rho(s|x)ds = -\tau_{k+1}(u,x)$, we can have the following assumption that is equivalent to assumption 3.3.2.

Assumption 3.3.5. *$\rho_x(s)$ is a function such that $u\frac{d}{du}\ln(\tau_k(u,x))$ is nonincreasing in u and bounded from below by $C_k(x) - k$ for all $k \in \mathbb{Z}^+$ and $x \in \mathbb{X}$.*

Remark 3.3.6. *Theorem 3.3.3 can be extended to more general constructions of NRMIs. For example, (James, 2002) introduced the h-biased random measures $\tilde{\mu}$ by $\int_{\mathbb{Y}\times\mathbb{X}} g(s)\tilde{N}(ds, dx)$, where $g : \mathbb{Y} \to \mathbb{R}^+$ is an integrable function on any complete and separable metric space $\mathbb{Y}$.*

One interesting quantity to be considered is $n(\pi)$, the number of distinct observations of the sample $\{X_i\}_{i=1}^n$. In Bayesian nonparametric mixture models, $n(\pi)$ is the number of clusters in the sample observations and thus is studied in a number of works that are concerning the clustering and so on. Among the literatures let us mention that the distribution of $n(\pi)$ is obtained in (Korwar and Hollander, 1973) for the Dirichlet process; in (Antoniak, 1974) for the mixture of Dirichlet process; in (Pitman, 2003) for the two-parameter Poisson-Dirichlet process. For the general NRMIs we have by a result of (James et al., 2009):

Proposition 3.3.7. *For any positive integer n, the distribution of $n(\pi)$ is*

$$\mathbb{P}(n(\pi)=k)=\int_0^\infty \frac{nu^n-1}{k!}e^{-\int_{\mathbb{X}}\int_0^\infty(1-e^{-ys})\rho(ds|x)\alpha(dx)}\sum_{(n_1,\cdots,n_k)}\prod_{j=1}^k\frac{\int_{\mathbb{X}}\tau_{n_j}(u,x)\alpha(dx)}{n_j!}du, \tag{3.3.4}$$

where $k=1,\cdots,n$, and the summation is over all vectors of positive integers $(n_1,\cdots,n_k)$ such that $\sum_{j=1}^k n_j=n$.

As we mentioned above, the assumption 3.3.2 is in fact quite easy to verify. We provide in the following examples to see the applicability of Theorem 3.3.3.

Example 3.3.8. *The normalized generalized gamma process* NGGP(a,σ,θ,H) *(Lijoi et al., 2003, 2007) is an NRMI with the following homogeneous Lévy intensity*

$$\nu(ds,dx)=\frac{1}{\Gamma(1-\sigma)}s^{-1-\sigma}e^{-\theta s}ds\alpha(dx)\,, \tag{3.3.5}$$

where the parameters $\sigma\in(0,1)$ and $\theta>0$. It is easy to see that the Laplace transform for $\tilde{\mu}(A)$ is

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(A)}\right]=\exp\left\{-\frac{\alpha(A)}{\sigma}[(\lambda+\theta)^\sigma-\theta^\sigma]\right\}.$$

When $\theta\to 0$, this NRMI yields the homogeneous σ-stable NRMI introduced by (Kingman, 1975). Letting $\sigma\to 0$, this NRMI becomes the Dirichlet process (Ferguson, 1973). If we let $\sigma=\theta=\frac{1}{2}$ then this NRMI becomes the normalized inverse-Gaussian process (Lijoi et al., 2005b).

It is easy to check that for any nonnegative integer k,

$$\tau_k(u,x) = \tau_k(u) = \frac{1}{\Gamma(1-\sigma)} \int_0^\infty s^{k-\sigma-1} e^{-(u+\theta)s} ds = \frac{\Gamma(k-\sigma)}{\Gamma(1-\sigma)(u+\theta)^{k-\sigma}} .$$

It is obvious that $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k-\sigma}{u+\theta}$ is increasing in u with the upper bound $k-\sigma$. Thus, the assumption 3.3.2 is verified and Theorem 3.3.3 yields that the normalized generalized gamma process is posterior consistent when $\sigma \to 0$ (i.e., the Dirichlet process), or when P_0 is discrete.

Example 3.3.9. *The generalized Dirichlet process* $\mathrm{GDP}(a,\gamma,H)$ *(Lijoi et al., 2005a) is an NRMI with the following homogeneous Lévy intensity*

$$\nu(ds,dx) = \sum_{j=1}^{\gamma} \frac{e^{-js}}{s} ds\alpha(dx) , \tag{3.3.6}$$

where γ is a positive integer. The corresponding Laplace transform of $\tilde{\mu}(A)$ is

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(A)}\right] = \left(\frac{(\gamma!)}{(\lambda+1)_\gamma}\right)^{\alpha(A)} ,$$

where for $c > 0$, $c_k = \frac{\Gamma(c+k)}{\Gamma(c)}$ is the ascending factorial of c for any positive integer k. When $\gamma = 1$, the generalized Dirichlet process is reduced to the Dirichlet process.

We see easily that for any nonnegative integer k,

$$\tau_k(u,x) = \tau_k(u) = \sum_{j=1}^{\gamma} \frac{k}{(u+j)^k} .$$

This means $\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = k\frac{\sum_{j=1}^{\gamma}(u+j)^{-k-1}}{\sum_{j=1}^{\gamma}(u+j)^{-k}} \in (\frac{k}{u+\gamma}, \frac{k}{u+1})$, which implies $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k}{u+c(\gamma)}$ with some constant $c(\gamma) \in (1,\gamma)$. Thus, $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)}$ is increasing in u with the upper bound k. Theorem 3.3.3 can then be used to conclude that the generalized Dirichlet process is posterior consistent.

Example 3.3.10. *As a non-homogeneous example, we consider the extended gamma*

NRMI whose non-homogeneous Lévy intensity is given by

$$\nu(ds, dx) = \frac{e^{-\beta(x)s}}{s} ds\alpha(dx)\,, \tag{3.3.7}$$

where $\beta(x) : \mathbb{X} \to \mathbb{R}^+$ is an integrable function (with respect to $\alpha(dx)$). Such NRMI is constructed by the normalization of the extended gamma process on $\mathbb{R}$ introduced by (Dykstra and Laud, 1981). More generally, (Lo, 1982) studied the extended Gamma process, called weighted Gamma process on abstract spaces.

By a trivial computation, for any nonnegative integer k, $\tau_k(u, x) = \frac{\Gamma(k)}{(u+\beta(x))^k}$ and thus $u\frac{\tau_{k+1}(u,x)}{\tau_k(u,x)} = u\frac{k}{u+\beta(x)}$ and the assumption 3.3.2 is satisfied. Theorem 3.3.3 implies that the extended gamma NRMI is posterior consistent when $\beta(x)$ is integrable with respect to $\alpha(dx)$.

Our theorem can also be applied to more general NRMIs which haven't been investigated in previous works. For example, we may naturally consider the following *generalized extended gamma NRMI* by letting the Lévy intensity be as follows:

$$\nu(ds, dx) = \sum_{i=1}^{r} \frac{e^{-\beta_i(x)s}}{s} ds\alpha(dx),$$

where $r \in \mathbb{Z}^+$ and $\beta_i(x) : \mathbb{X} \to \mathbb{R}^+$ are integrable functions (with respect to $\alpha(dx)$). A similar argument to that of example 3.3.9 and example 3.3.10 implies that the generalized extended gamma NRMI is posterior consistent when $\beta_i(x)$ are finite for all $i \in \{1, \cdots, r\}$.

Relying on the results in this section, we have answered the question (i) addressed in the introduction. The posterior consistency of NRMIs when P_0 is continuous doesn't hold generally, as the posterior distribution of NRMIs is inconsistent when $\bar{C}_1 \neq 0$ or $H \neq P_0(\mathbb{P}_n)$. However, it is rare to choose H to be the "true" distribution P_0 and it is not possible to let $H = \mathbb{P}_n$ before a sample is observed. Thus, the assumption $\bar{C}_1 = 0$ should be made to guarantee the posterior consistency for the NRMIs when P_0 is continuous. And, whenever $\rho_x(ds)$ is gamma type, $\bar{C}_1 = 0$ would reduce the corresponding P to the Dirichlet process or the generalized Dirichlet process.

3.4 Bernstein-von Mises theorem for the generalized normalized gamma process

The Bernstein-von Mises theorem links Bayesian inference with frequentist inference. Similarly to the Bernstein-von Mises theorem (Vaart, 1998) in Bayesian parametric framework, one can derive the Bernstein-von Mises theorem in Bayesian nonparametric framework. There has been some works in the literature. One example is the Bernstein-von Mises theorem for the empirical process $\mathbb{P}_n = \frac{\sum_{i=1}^n \delta_{X_i}}{n}$ (van der Vaart and Wellner, 1996; Vaart, 1998). With the fact that the maximum likelihood estimator of P_0 in the Bayesian nonparametric sense is $\mathbb{P}_n = \frac{\sum_{i=1}^n \delta_{X_i}}{n}$, one can conclude the limit law of $\sqrt{n}(\mathbb{P}_n - P_0)$ is normal distribution. Based on a similar idea, we would consider the limit law of the posterior distribution of $\sqrt{n}(P - \mathbb{P}_n)$ given an iid sample $\mathbf{X}$ from P_0. To explain the Bernstein-von Mises theorem in the Bayesian nonparametric case, we temporarily let $P \in \mathbb{M}_{\mathbb{X}}$ be any random probability measure and define the functional as follows:

$$Pf = \int f dP, \qquad P_0 f = \int f dP_0, \qquad \mathbb{P}_n f = \int f d\mathbb{P}_n = \frac{\sum_{i=1}^n f(X_i)}{n},$$

where $f : \mathbb{X} \to \mathbb{R}$ is any measurable functions.

Let $\mathbb{F}$ be the collection of functions f, the Bernstein-von Mises theorem in the Bayesian nonparametric case considers the distribution of $\{\sqrt{n}(Pf - \mathbb{P}_n f)|\mathbf{X} : f \in \mathbb{F}\}$ and $\{\sqrt{n}(\mathbb{P}_n f - P_0 f) : f \in \mathbb{F}\}$. It is worth to point out that there have been many works for the weak convergence of stochastic processes indexed by elements of Banach space of functions, we refer the statisticians to (van der Vaart and Wellner, 1996; Vaart, 1998) for further reading. When the function collection $\mathbb{F}$ is finite, both $\{\sqrt{n}(Pf - \mathbb{P}_n f)|\mathbf{X} : f \in \mathbb{F}\}$ and $\{\sqrt{n}(\mathbb{P}_n f - P_0 f) : f \in \mathbb{F}\}$ are random vectors in Euclidean space. Otherwise, it is convenient to consider the $\mathbb{F}$ to be P_0-Donsker. Here we recall that $\mathbb{F}$ is P_0-Donsker if the sequence $\sqrt{n}(\mathbb{P}_n f - P_0 f)$ converges to $\mathbb{B}^o_{P_0}$ in distribution in the metric space $l^\infty(\mathbb{F})$ of bounded functions $g : \mathbb{F} \to \mathbb{R}$, equipped with the uniform norm $||g||_{\mathbb{F}} = \sup_{f\in\mathbb{F}} |g(f)|$. And $\mathbb{B}^o_{P_0}$ is a Brownian bridge with parameter P_0 or P_0- Brownian bridge, so that $\mathbb{E}[\mathbb{B}^o_{P_0} f] = 0$

and $\mathbb{E}[\mathbb{B}^o_{P_0} f_1 \mathbb{B}^o_{P_0} f_2] = P_0(f_1 f_2) - P_0 f_1 P_0 f_2$. An notable result is that a finite set $\mathbb{F}$ is P_0-Donsker if and only if $P_0 f^2 < \infty$ for every $f \in \mathbb{F}$. For the infinite P_0-Donsker classes, one can find details and examples in (van der Vaart and Wellner, 1996).

In order to define the weak convergence of $\sqrt{n}(P - \mathbb{P}_n)$ conditional on $\mathbf{X}$ to $\mathbb{B}^o_{P_0}$, we can use the conditional weak convergence in the bounded Lipschitz metric (van der Vaart and Wellner, 1996) as follows:

$$\sup_{h \in \mathrm{BL}_1} \left| \mathbb{E}\left\{ h(\sqrt{n}(P - \mathbb{P}_n) | \mathbf{X} \right\} - \mathbb{E}[h(\mathbb{B}^o_{P_0})] \right| \to 0\,, \tag{3.4.1}$$

as $n \to \infty$. The expectation in (3.4.1) is taken for the random probability measure P, and thus the left side of (3.4.1) is a function of $\mathbf{X}$. The convergence in (3.4.1) refers to the iid sample $\mathbf{X}$ from P_0 and can be in probability or almost surely. The supreme is taken over the set BL_1 of all functions $h : l^\infty(\mathbb{F}) \to [0, 1]$ such that $|h(f_1) - h(f_2)| \le ||f_1 - f_2||_{\mathbb{F}}$, for all $f_1, f_2 \in l^\infty(\mathbb{F})$. We denote the above convergence as

$$\sqrt{n}(P - \mathbb{P}_n) | \mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}\,.$$

Under the convergence criteria we explained above, we will present the Bernstein-von Mises theorem when $P \sim \mathrm{NGGP}(a, \sigma, \theta, H)$. For simplicity of interpretation, let $\tilde{\mathbb{P}}_n = \frac{\sum_{i=1}^{n(\pi)} \delta_{Y_i}}{n(\pi)}$.

Theorem 3.4.1. *Let $\mathbf{X}$ be a sample as defined in (3.1.1) with $P \sim \mathrm{NGGP}(a, \sigma, \theta, H)$. Let $\mathbb{F}$ be the finite collection of functions such that $P_0 f^2 < \infty$ and $H f^2 < \infty$ for any $f \in \mathbb{F}$. We have the following convergences almost surely under P_0^∞.*

(i) If P_0 is discrete,

$$\sqrt{n}\left(P - \left\{ \mathbb{P}_n + \frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n) \right\} \right) | \mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}\,, \tag{3.4.2}$$

$$\sqrt{n}\left(P - \mathbb{E}[P|\mathbf{X}] \right) | \mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}\,. \tag{3.4.3}$$

(ii) If P_0 is continuous,

$$\sqrt{n}\left(P - \{(1-\sigma)\mathbb{P}_n + \sigma H\}\right)|\mathbf{X} \\ \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}^o_{P_0} + \sqrt{\sigma(1-\sigma)}\mathbb{B}^o_H + \sqrt{\sigma}Z(P_0 - H)\,, \qquad (3.4.4)$$

$$\sqrt{n}(P - \mathbb{E}[P|\mathbf{X}])|\mathbf{X} \\ \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}^o_{P_0} + \sqrt{\sigma(1-\sigma)}\mathbb{B}^o_H + \sqrt{\sigma}Z(P_0 - H)\,. \qquad (3.4.5)$$

Here $\mathbb{B}^o_{P_0}$, $\mathbb{B}^o_H$ are independent Brownian bridges, independent of the standard normal random variable Z. Moreover, the convergences hold in probability in $l^\infty(\mathbb{F})$, for every P_0-Donsker class of functions $\mathbb{F}$ for which the NGGP(a, σ, θ, H) *process satisfies the central limit theorem in $l^\infty(\mathbb{F})$. If in addition that $P_0||f - P_0 f||^2_{\mathbb{F}} < \infty$, then the convergences is also $P_0^\infty-$almost surely.*

We refer to Theorem 2.11.1 and 2.11.9 in (van der Vaart and Wellner, 1996) for more details of the discussion for $\mathbb{F}$ such that the convergence holds in $l^\infty(\mathbb{F})$.

When P_0 is continuous, there is a "bias" term $\sigma(H - \mathbb{P}_n)$ in the convergence in (3.4.4). And the term vanishes only when $\sigma = 0$, under which P becomes the Dirichlet process, or when $H = \mathbb{P}_n$ ($H = P_0$), which is unrealistic. Moreover, the σ equals the $\bar{C}_1$ in Theorem 3.3.3. Thus, it suggests that one is not expected to use NGGP for continuous P_0.

On the other hand, it is interesting to see that there is a "bias" term $\frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n)$ in the convergence in (3.4.2) when P_0 is discrete to make the limiting process is $\mathbb{B}^o_{P_0}$. We can not drop this "bias" term directly, although $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ a.s.. The term can be dropped as long as $\lim_{n\to\infty} \frac{n(\pi)}{\sqrt{n}} = 0$, in the sense that the number of atoms $\{x_j\}$ in P_0 should decrease fast enough when $n \to \infty$. For a formal condition of P_0 to make $\lim_{n\to\infty} \frac{n(\pi)}{\sqrt{n}} = 0$, we have the following Corollary.

Corollary 3.4.2. *Under the conditions in Theorem 3.4.1, when P_0 is discrete, we have the following results.*

(i) If $P_0(\{x_j\}) \leq \frac{C}{j^\alpha}$, for some positive constant C and $\alpha > 2$ and $\mathbb{F}$ is the class of uniformly bounded functions, then $\sqrt{n}(P_{U_n} - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ in probability in $l^\infty(\mathbb{F})$.

(ii) *If the function* $h(t) := \#\{x : P_0(\{x\}) \geq \frac{1}{t}\}$ *is regularly varying at* ∞ *of exponent* η *with* $\eta < \frac{1}{2}$ *and* $\mathbb{F}$ *is the class of uniformly bounded functions, then* $\sqrt{n}(P_{U_n} - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ *a.s. in* $l^\infty(\mathbb{F})$.

(iii) *If* $\mathbb{F}$ *is a class of functions* f *such that* $f(\{x_j\}) \asymp j^\beta$ *for some* $\beta > 0$ *and* $P_0(\{x_j\}) \leq \frac{C}{j^\alpha}$, *for some positive constant* C *and* $\alpha > 2 + 2\beta$, *then* $\sqrt{n}(P_{U_n} - \mathbb{P}_n)|\mathbf{X} \rightsquigarrow \mathbb{B}^o_{P_0}$ *in probability in* $l^\infty(\mathbb{F})$.

The proof of the above Corollary follows directly from the Corollary 2 in (Franssen and van der Vaart, 2022). And we recall that if h is regularly varying at ∞ with exponent $\eta \in (0, 1)$, then for any $t > 0$, we have $\lim_{n\to\infty} \frac{h(nt)}{h(n)} = t^\eta$. Moreover, for such regularly varying function h, we have $\frac{n(\pi)}{h(n)} \to \Gamma(1-\eta)$ a.s., and $h(n)$ is n^η up to a slowly varying factor. We refer the appendix in (Haan and Ferreira, 2006) and (Bingham et al., 1987) for more details of the regularly varying function.

As the application of the Bernstein-von Mises results in Theorem 3.4.1, we may construct the confidence intervals for $P_0 f$ when $n \to \infty$. The choices of f determine the parameters $P_0 f$, for which the credible intervals are constructed. For example, if $f(x) = x$, the credible interval is for the mean. Since the posterior consistency does not hold for the case when P_0 is continuous, the credible intervals for $P_0 f$ is not correct in this case, thus we shall only give the credible interval for Pf when P_0 is discrete.

Corollary 3.4.3. *If* P_0 *is discrete, under the conditions in Theorem 3.4.1, we have the probability of* $P_0 f \in \left(L_{n,\alpha} f - \frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f), L_{n,\beta} f - \frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f)\right)$ *is* $\beta - \alpha$ *for any* f *such that* $P_0 f^2 < \infty$ *and* $Hf^2 < \infty$. *Here* $L_{n,\alpha}$ *is the* $\alpha-$*quantile of the posterior distribution of* $Pf|\mathbf{X}$ *and* $\beta > \alpha$.

One direct interpretation of the above Corollary is one may want $\frac{n(\pi)}{n} \to 0$ to make the "bias" term vanish and therefore the credible interval for $P_0 f$ becomes a regular form $(L_{n,\alpha} f, L_{n,\beta} f))$. Otherwise, the correction $\frac{\sigma n(\pi)}{n}(Hf - \tilde{\mathbb{P}}_n f)$ is necessary as a bias correction to the credible interval. We provide the numerical illustration that corresponding to this scenario in Section 3.4.1.

However, P_0 is of course unknown in the real application and we shall consider Theorem 3.4.1 without the information from P_0. One important parameter that one needs to pay attention especially is σ, and it is easy to see from both Theorem 3.3.3 and Theorem 3.4.1 that if $\sigma \to 0$, P is posterior consistent and the Bernstein-von Mises results hold without the bias terms for any P_0. But this corresponds to the case that P becomes the Dirichlet process. Thus, one should at least expect the parameter σ to be small. Usually, the model parameters are chosen by the empirical Bayesian method, and people can estimate the model parameters by using the maximum likelihood estimators conditional on the observations $\mathbf{X}$. A well known conclusion (Pitman, 2003, 2006) in Bayesian nonparametric framework is the observation $\mathbf{X}$ from NRMIs induces a random partition structure for $\{1, \cdots, n\}$ as we introduced in Section 3.2.2. The random partition structure is characterized by the exchangeable partition probability function (EPPF) (Pitman, 2003), which also plays the rule as the likelihood function of σ as explained in e.g., (Favaro and Naulet, 2021; Ghosal and Van der Vaart, 2017; Franssen and van der Vaart, 2022). And the EPPF for the NGGP is given as

$$\Pi_\sigma(n_1, \cdots, n_{n(\pi)}) = \frac{\prod_{j=1}^{n(\pi)} (1-\sigma)_{(n_j-1)}}{\Gamma(n)} \int_0^\infty u^{n-1}(u+\theta)^{n(\pi)\sigma - n} e^{\frac{a}{\sigma}((u+\theta)^\sigma - \theta^\sigma} du \,,$$

where $(1-\sigma)_{(n_j-1)} = \frac{\Gamma(n_j-\sigma)}{\Gamma(1-\sigma)}$. From Theorem 1 in (Favaro and Naulet, 2021), the maximum likelihood estimator $\hat{\sigma}_n$ exists uniquely. Furthermore, the results in Theorem 2 in (Favaro and Naulet, 2021) implies that $\hat{\sigma}_n \to \sigma_0$ in probability with a rate $\sqrt{\log(n)} n^{-\frac{\sigma_0}{2}}$, when P_0 is discrete with atoms x satisfying $h(t) = \#\{P_0(\{x\}) \geq \frac{1}{t}\}$ is a regularly varying function of exponent $\sigma_0 \in [0, 1)$.

Theorem 3.4.4. *Under the assumptions in Theorem 3.4.1, we have the following results.*

(i) If $\hat{\sigma}_n$ is an estimator based on $\mathbf{X}$ that converges to σ_0 in probability, then the convergences in Theorem 3.4.1 hold in probability by replacing σ_n by $\hat{\sigma}_n$ and replacing σ by σ_0. In particular, this is true for the maximum likelihood estimator $\hat{\sigma}_n$, if P_0 is discrete with atoms x satisfying the condition that $h(t) = \#\{P_0(\{x\}) \geq \frac{1}{t}\}$ is a

regularly varying function of exponent $\sigma_0 \in [0,1)$.

(ii) If $\sigma \sim L_\sigma$, where L_σ is the prior probability law on $[0,1]$ that plays the law of σ, then the convergences in Theorem 3.4.1 hold by replacing σ_n by σ on the left hand side, and replacing σ by σ_0 on the limiting processes.

The proof of the above theorem follows the same constructions as the proof in section 4.2 of (Franssen and van der Vaart, 2022). For the posterior consistency of $\hat{\sigma}_n$, we refer to the details with proofs in section 4.3 of (Franssen and van der Vaart, 2022). The maximum likelihood estimator is not quite interesting since $\hat{\sigma}_n \to \sigma_0$ with $\sigma_0 = 1$ when P_0 is continuous, and $\sigma_0 \neq 0$ when P_0 is discrete (Favaro and Naulet, 2021).

Besides the parameter σ, the parameters a and θ don't appear in the asymptotic results in Theorem 3.3.3 and Theorem 3.4.1, and thus estimators of a and θ based on prior distributions or maximum likelihood method won't affect the convergences when $a << \sqrt{n}$ and $\theta << n^\sigma$. And the cases when $\hat{a}_n$ and $\hat{\theta}_n$ converge to ∞ as $n \to \infty$ are beyond the scope of this work and can be considered in the future works.

3.4.1 Numerical illustration

We present the credible intervals for $P_0 f$ when P_0 is discrete with different behaviours of the number of atoms. To be more precise, let $P_0 f = P_0([2,\infty])$ for $P_0 = P_1, P_2, P_3, P_4$, where we describe P_1, P_2, P_3, P_4 as follows. Let the probability distributions of P_1, P_2, P_3, P_4 be on $\mathbb{Z}^+$ are as follows.

$$P_1(X=1) = 0.2, P_1(X=2) = 0.2, P_1(X=3) = 0.2, P_1(X=4) = 0.3, P_1(X=5) = 0.1\,,$$
$$P_2(X=k) \propto k^{-3}\,, \qquad P_3(X=k) \propto k^{-2}\,, \qquad P_4(X=k) \propto k^{-\frac{3}{2}}\,.$$

Obviously, $n(\pi) = 5$ for P_1. From the result (see e.g., Example 4) in (Karlin, 1967), we have the regularly varying functions $h(t)$ corresponding to P_2, P_3, P_4 are proportional to $t^{\frac{1}{3}}, t^{\frac{1}{2}}, t^{\frac{2}{3}}$ respectively. And when $n \to \infty$, the distinct numbers $n(\pi)$ of P_2, P_3, P_4 are proportional to $n^{\frac{1}{3}}, n^{\frac{1}{2}}, n^{\frac{2}{3}}$, respectively, from Theorem 1 in (Karlin, 1967). Thus, the

n	10	100	1000	10000	100000
P_1	0.791	0.952	0.961	0.967	0.986
P_2	0.695	0.857	0.928	0.917	0.931
P_3	0.712	0.785	0.811	0.727	0.754
P_4	0.601	0.292	0.078	0.000	0.000

Table 3.1: Proportion of coverage of the true value for the 95% credible interval without "bias" correction.

n	10	100	1000	10000	100000
P_1	0.977	0.989	0.991	0.995	0.997
P_2	0.914	0.938	0.951	0.933	0.941
P_3	0.863	0.931	0.962	0.960	0.978
P_4	0.901	0.955	0.969	0.966	0.956

Table 3.2: Proportion of coverage of the true value for the 95% credible interval with "bias" correction.

"bias" term for P_1, P_2, P_3, P_4 goes to $0, 0$, some constant, ∞, respectively.

For the NGGP, we let $P \sim \text{NGGP}(1, \sigma = 0.5, 1, H)$, where H is standard normal distribution. We simulate P through its stick-breaking representation with the generating algorithm in (Favaro et al., 2016). To make sure the simulation of $P = \sum_{i=1}^{\infty} w_i \delta_{X_i}$ is accurate, we truncate the infinite sum at some N such that the weight of the tail $\sum_{i=N}^{\infty} w_i < \frac{1}{\sqrt{n}}$, where n is the sample size. We simulate 10000 replications of the sample $\mathbf{X}$ from P_1, P_2, P_3, P_4 with the sample size $n = 10, 100, 1000, 10000, 100000$ respectively. For the sample from P_1, we construct one 95% credible interval for each sample for $P_1([2, \infty))$ with the "bias" correction as in Corollary 3.4.3 and compute the proportion that the true value $P_1([2, \infty))$ belongs to the intervals of 10000 replications. And we also compute the same proportion without the "bias" correction. The results of P_1, P_2, P_3, P_4 are given in tables 3.1 and 3.2.

Since the "bias" terms for P_1 and P_2 vanish as $n \to \infty$, the proportions of the coverage of the true value are large for both with and without "bias" correction. And the 95% credible intervals for $P_3 f$ and $P_4 f$ are not performing good without "bias" correction.

As for the normality convergence, we draw the marginal density plots in figure 3.1

for $P_1([2,\infty))$ given the sample $\mathbf{X}$ with size $n = 10, 100, 1000, 10000, 100000$ respectively. Both plots are generated from 1000000 replicates, the true mean of $P_1([2,\infty))$ is 0.8 The marginal density for $P_1([2,\infty))$ is skewed when $n = 10, 100$, and symmetric when $n = 1000$ and larger.

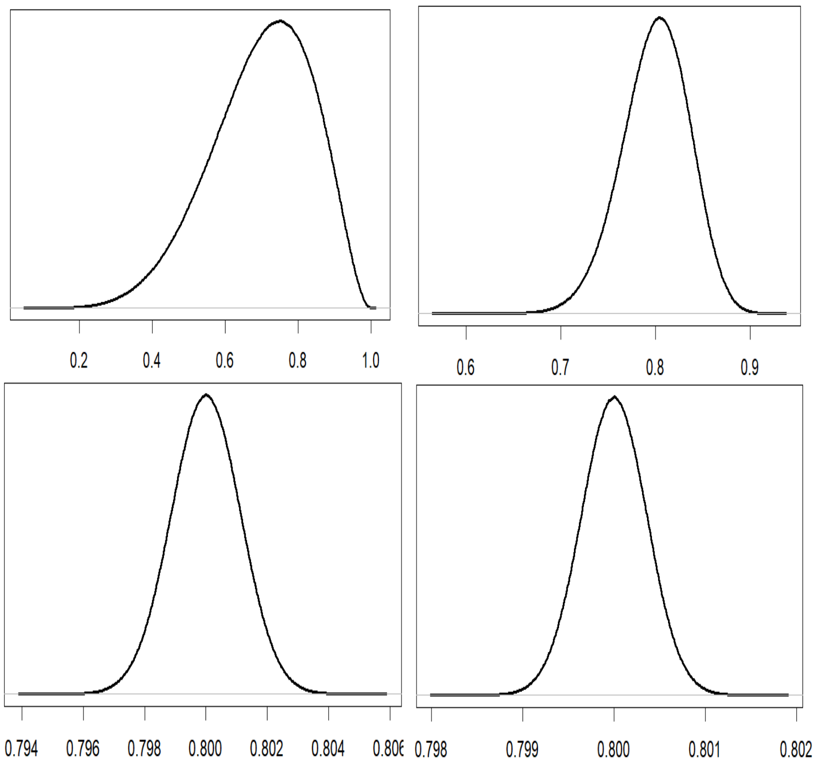


Figure 3.1: The marginal densities for $P_1([2,\infty))$ with sample size $n = 10, 100, 1000, 10000, 100000$ follow the order from top left to bottom right.

3.5 Discussion

To the best of our knowledge, the Lévy intensities of the well-studied NRMIs up-to-date are given in the form of the gamma density: $s^{-\sigma-1}e^{-\beta s}$. It turns out that with the shape parameter $\sigma = 0$, the posterior consistency is always guaranteed for any "true" prior distribution P_0. Otherwise, the posterior consistency only holds for discrete prior P_0 but not for diffusive prior P_0. Such phenomenon does naturally make sense due to the discreteness of NRMIs (the completely random measures (Kingman, 1975)). As explained in the Bayesian literature, if P_0 is diffusive and the prior guess for the sample distribution

$\alpha \neq P_0$, then the prior guess will always contribute to the posterior, no matter how large is the sample size. In such sense, the Bayesian nonparametric models never behave "better" than the empirical models asymptomatically. However, this doesn't mean the NRMIs are not useful. On the one hand, we are not able to know the "true" distribution of a given sample with any size n, also the sample size n will never be ∞, a prior guess of the random probability measure based on experience could make the model suitable. On the other hand, the NRMIs behave great for the data from discrete distributions. Furthermore, the mixture and hierarchical Bayesian nonparametric models based on NRMIs are showing great success in the applications and consistency behaviours (Lijoi et al., 2005). And the class of NRMIs is much larger than we expected, so that more study is necessary to develop more flexible subclasses of NRMIs or more general NRMIs like classes that are satisfying the consistency property. The results in this work provides a guideline of choosing the proper intensity $\rho_x(s)$, for example, the generalized Dirchlet process and the generalized extended gamma NRMI are good choice in the Bayesian nonparametric applications and they both show some flexibility. Besides, we may let $\sigma \to 0$ by assigning a randomness on σ, or one may construct α to depend on ρ_x to deduct $\bar{C}_1$.

Due to the complexity of the posterior of the NRMIs, it is not easy to present a Bernstein-von Mises like result to give the limiting process of posterior of general NRMIs. The result for the normalized generalized gamma process, along with the works in (Lo, 1983, 1986; Ray and van der Vaart, 2021; Hu and Zhang, 2022; James, 2008; Franssen and van der Vaart, 2022), shed some light in discovering the Bernstein-von Mises theorem for general NRMIs.

3.6 Appendix

In this section, we prove Lemma 3.3.1, Theorem 3.3.3 and Theorem 3.4.1.

Proof of Lemma 3.3.1

Let $\mathcal{I} = \mathbb{E}[(P(A)|\mathbf{X})^m]$. Then, by Theorem 3.2.2, $\mathcal{I}$ can be computed as follows.

$$\begin{aligned}
\mathcal{I} &= \int_0^\infty \mathbb{E}[(P(A)|U_n = u, \mathbf{X})^m] f_{U_n|\mathbf{X}}(u|\mathbf{X}) du \\
&= \int_0^\infty \mathbb{E}\left[\left(\frac{\tilde{\mu}_{(U_n)}(A)}{T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j} + \sum_{j=1}^{n(\pi)} \frac{J_j \delta_{Y_j}(A)}{T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j}\right)^m\right] f_{U_n|\mathbf{X}}(u|\mathbf{X}) du \\
&= \sum_{k=0}^m \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} \mathbb{E}\Bigg[e^{-y(T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j)} \tilde{\mu}_{(U_n)}(A)^{m-k} \\
&\qquad \times \left(\sum_{j=1}^{n(\pi)} J_j \delta_{Y_j}(A)\right)^k \Bigg] f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du \,.
\end{aligned} \tag{3.6.1}$$

Noticing that $T_{(U_n)} = \tilde{\mu}_{(U_n)}(A) + \tilde{\mu}_{(U_n)}(A^c)$, where $\tilde{\mu}_{(U_n)}(A)$ and $\tilde{\mu}_{(U_n)}(A^c)$ are independent, we can rewrite the expectation in (3.6.1) as

$$\begin{aligned}
\mathcal{I} =& \sum_{k=0}^m \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} \mathbb{E}\left[e^{-y\tilde{\mu}_{(U_n)}(A)} \tilde{\mu}_{(U_n)}(A)^{m-k}\right] \mathbb{E}\left[e^{-y\tilde{\mu}_{(U_n)}(A^c)}\right] \\
&\qquad \mathbb{E}\left[e^{-y(\sum_{j=1}^{n(\pi)} J_j)} \left(\sum_{j=1}^{n(\pi)} J_j \delta_{Y_j}(A)\right)^k\right] f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du \\
=& \sum_{k=0}^m \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} (-1)^{m-k} \mathbb{E}\left[\frac{d^{m-k}}{dy^{m-k}} e^{-y\tilde{\mu}_{(U_n)}(A)}\right] \mathbb{E}\left[e^{-y\tilde{\mu}_{(U_n)}(A^c)}\right] \\
&\qquad \left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \mathbb{E}\left[e^{-y(\sum_{j=1}^{n(\pi)} J_j)} \prod_{j=1}^{n(\pi)} J_j^{l_j} \delta_{Y_j}(A)\right]\right) f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du \\
=& \sum_{k=0}^m \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} (-1)^{m-k} \mathbb{E}[\frac{d^{m-k}}{dy^{m-k}} e^{-y\tilde{\mu}_{(U_n)}(A)}] \mathbb{E}[e^{-y\tilde{\mu}_{(U_n)}(A^c)}] \\
&\qquad \left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \mathbb{E}[\prod_{j=1}^{n(\pi)} (-1)^{l_j} \frac{d^{l_j}}{dy^{l_j}} e^{-yJ_j} \delta_{Y_j}(A)]\right) f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du \,,
\end{aligned}$$

where the sum in front of $\binom{k}{l_1, \cdots, l_{n(\pi)}}$ is over all the vector $(l_1, \cdots, l_{n(\pi)})$ such that $\sum_{j=1}^{n(\pi)} l_j = k$. Taking the derivatives inside the expectation and using the Laplace transform of

$\tilde{\mu}_{(U_n)}(A)$, we have

$$\mathcal{I} = \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} (-1)^{m-k} \frac{d^{m-k}}{dy^{m-k}} \mathbb{E}[e^{-y\tilde{\mu}_{(U_n)}(A)}] \mathbb{E}[e^{-y\tilde{\mu}_{(U_n)}(A^c)}]$$
$$\left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(u+y, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A) \right) f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du$$
$$= \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} V_{\alpha(A)}^{(m-k)}(u, y) e^{-\psi_{\mathbb{X}}(u,y)}$$
$$\left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(u+y, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A) \right) f_{U_n|\mathbf{X}}(u|\mathbf{X}) dy du \,, \tag{3.6.2}$$

where $\psi_{\mathbb{X}}(u, y) = \int_{\mathbb{X}} \int_0^\infty (1 - e^{-ys}) e^{-us} \rho(ds|x) \alpha(dx)$. By the fact that

$$f_{U_n|\mathbf{X}}(u|\mathbf{X}) \propto u^{n-1} e^{-\psi_{\mathbb{X}}(u)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(u, Y_j)$$

and $e^{-\psi_{\mathbb{X}}(u)} e^{-\psi_{\mathbb{X}}(u,y)} = e^{-\psi_{\mathbb{X}}(u+y)}$, we further simplify (3.6.2) to

$$\mathcal{I} = \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} u^{n-1} V_{\alpha(A)}^{(m-k)}(u+y) e^{-\psi_{\mathbb{X}}(u+y)}$$
$$\left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \tau_{n_j+l_j}(u+y, Y_j) \delta_{Y_j}(A) \right) dy du$$
$$= \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^\infty y^{m-1} u^{n-1} V_{\alpha(A)}^{(m-k)}(u+y) e^{-\psi_{\mathbb{X}}(u+y)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(u+y, Y_j)$$
$$\left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(u+y, Y_j)}{\tau_{n_j}(u+y, Y_j)} \delta_{Y_j}(A) \right) dy du \,.$$

The change of variable $(w, z) = (u + y, u)$ yields

$$\mathcal{I} = \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \int_0^\infty \int_0^w (w-z)^{m-1} z^{n-1} V_{\alpha(A)}^{(m-k)}(w) e^{-\psi_{\mathbb{X}}(w)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(w, Y_j)$$

$$\left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(w, Y_j)}{\tau_{n_j}(w, Y_j)} \delta_{Y_j}(A)\right) dzdw\,.$$

Using

$$\int_0^w (w-z)^{m-1} z^{n-1} dz = w^{m+n-1} \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}\,,$$

we obtain

$$\begin{aligned}
\mathcal{I} =& \sum_{k=0}^{m} \binom{m}{k} \frac{1}{\Gamma(m)} \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \int_0^\infty w^{n+m-1} V_{\alpha(A)}^{(m-k)}(w) e^{-\psi_{\mathbb{X}}(w)} \prod_{j=1}^{n(\pi)} \tau_{n_j}(w, Y_j) \\
& \left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(w, Y_j)}{\tau_{n_j}(w, Y_j)} \delta_{Y_j}(A)\right) dw \\
=& \frac{\Gamma(n)}{\Gamma(m+n)} \sum_{k=0}^{m} \binom{m}{k} \int_0^\infty w^m f_{U_n}(w) V_{\alpha(A)}^{(m-k)}(w) \\
& \left(\sum \binom{k}{l_1, \cdots, l_{n(\pi)}} \prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(w, Y_j)}{\tau_{n_j}(w, Y_j)} \delta_{Y_j}(A)\right) dw \\
=& \frac{\Gamma(n)}{\Gamma(m+n)} \sum_{0 \le l_1+\cdots+l_{n(\pi)} \le m}^{m} \binom{m}{l_1, \cdots, l_{n(\pi)}} \int_0^\infty w^m f_{U_n}(w) V_{\alpha(A)}^{(m-(l_1+\cdots+l_{n(\pi)}))}(w) \\
& \left(\prod_{j=1}^{n(\pi)} \frac{\tau_{n_j+l_j}(w, Y_j)}{\tau_{n_j}(w, Y_j)} \delta_{Y_j}(A)\right) dw\,.
\end{aligned}$$

This is (3.3.2).

For any family of pairwise disjoint sets $\{A_1, \cdots, A_q\}$ in $\mathcal{X}$ and for any positive integers $\{m_1, \cdots, m_q\}$ we denote $A_{q+1} = (\cup_{i=1}^q A_i)^c$, $m_{q+1} = 0$, and $m = \sum_{i=1}^q m_i$. For any sample $\{X_i\}_{i=1}^n$ from P, let $\{Y_j\}_{j=1}^{n(\pi)}$ be the distinct values of $\{X_i\}_{i=1}^n$. Let $\lambda_i = \{j : Y_j \in A_i\}$ be the set of the index of Y_j's that in A_i and we denote by $\max(\lambda_i)$ the maximal value in λ_i. We can compute the following moments easily.

$$\begin{aligned}
\mathcal{L} :=& \mathbb{E}\left[P(A_1)^{m_1} \cdots P(A_q)^{m_q} | \mathbf{X}\right] \qquad (3.6.3) \\
=& \int_0^\infty \mathbb{E}\left[P(A_1)^{m_1} \cdots P(A_q)^{m_q} | U_n = u, \mathbf{X}\right] f_{U_n|\mathbf{X}}(u|\mathbf{X}) du
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{\Gamma(m)}\int_0^\infty\int_0^\infty y^{m-1}\mathbb{E}\left[e^{-y(T_u+\sum_{j=1}^{n(\pi)}J_j)}\prod_{i=1}^{q+1}\left(\mu_u(A_i)+\sum_{j=1}^{n(\pi)}J_j\delta_{Y_j}(A_i)\right)^{m_i}\right]f_{U_n|\mathbf{X}}(u|\mathbf{X})dydu\\
=&\frac{1}{\Gamma(m)}\int_0^\infty\int_0^\infty y^{m-1}\prod_{i=1}^{q+1}\left\{\sum_{k=0}^{m_i}\mathbb{E}\left[e^{-y\mu_u(A_i)}\mu_u(A_i)^{m_i-k}\right]\right.\\
&\left.\mathbb{E}\left[e^{-y(\sum_{j\in\lambda_i}J_j)}\left(\sum_{j\in\lambda_i}J_j\right)^k\right]\right\}f_{U_n|\mathbf{X}}(u|\mathbf{X})dydu\,.
\end{aligned}
$$

A similar computation as that for $\mathcal{I}$ yields

$$
\begin{aligned}
\mathcal{L}=&\frac{\Gamma(n)}{\Gamma(m+n)}\int_0^\infty w^{n+m-1}e^{-\psi_{\mathbb{X}}(w)}\prod_{j=1}^{n(\pi)}\tau_{n_j}(w,Y_j)\prod_{i=1}^{q+1}\left\{\sum_{0\le l_1+\cdots+l_{\max(\lambda_i)}\le m_i}\binom{m_i}{l_1,\cdots,l_{max(\lambda_i)}}\right.\\
&\left.V_{\alpha(A_i)}^{(m_i-(l_1+\cdots+l_{max(\lambda_i)}))}(w)\left(\prod_{j\in\lambda_i}\frac{\tau_{n_j+l_j}(w,Y_j)}{\tau_{n_j}(w,Y_j)}\right)\right\}dw\\
=&\frac{\Gamma(n)}{\Gamma(m+n)}\int_0^\infty w^m f_{U_n|\mathbf{X}}(w|\mathbf{X})\prod_{i=1}^{q+1}\left\{\sum_{0\le l_1+\cdots+l_{\max(\lambda_i)}\le m_i}\binom{m_i}{l_1,\cdots,l_{max(\lambda_i)}}\right.\\
&\left.V_{\alpha(A_i)}^{(m_i-(l_1+\cdots+l_{max(\lambda_i)}))}(w)\left(\prod_{j\in\lambda_i}\frac{\tau_{n_j+l_j}(w,Y_j)}{\tau_{n_j}(w,Y_j)}\right)\right\}dw\,.
\end{aligned}
$$

This is part (ii) of the theorem. Then the proof of lemma 3.3.1 is completed.

Proof of Theorem 3.3.3

We need the following lemma to prove Theorem 3.3.3.

Lemma 3.6.1. *Under the assumption 3.3.2, we have for any $y\in\mathbb{X}$ and $k\in\mathbb{Z}^+$,*

$$
\lim_{n\to\infty}\int_0^\infty\tau_k(u,y)f_{U_n|\mathbf{X}}(u|\mathbf{X})du=k-C_k(y)\,. \tag{3.6.4}
$$

Proof. Let $g_n(u)$ be a constant multiple of the density of $f_{U_n|\mathbf{X}}(u|\mathbf{X})$ given by (3.2.10). Namely,

$$
g_n(u)=u^{n-1}e^{-\psi(u)}\prod_{j=1}^{n(\pi)}\tau_{n_j}(u,Y_j)
$$

$$=u^{n-1}e^{-a\int_{\mathbb{X}}\int_0^\infty(1-e^{-us})\rho(ds|x)H(dx)}\prod_{j=1}^{n(\pi)}\int_0^\infty s^{n_j}e^{-us}\rho(ds|Y_j) \tag{3.6.5}$$

$$f_{U_n|\mathbf{X}}(u|\mathbf{X}) =\frac{g_n(u)}{\int_0^\infty g_n(u)du}\,. \tag{3.6.6}$$

The derivative of $g_n(u)$ is computed as follows,

$$g_n'(u) = u^{n-2}e^{-\psi_{\mathbb{X}}(u)}\prod_{j=1}^{n(\pi)}\tau_{n_j}(u,Y_j)\left[n-1-\left(u\int_{\mathbb{X}}\tau_1(u,y)\alpha(dy)+\sum_{j=1}^{n(\pi)}u\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\right)\right].$$

Let $h_n(u) = u\int_{\mathbb{X}}\tau_1(u,y)\alpha(dy)$, then $h_n'(u) = \int_{\mathbb{X}}(\tau_1(u,y)-u\tau_2(u,y))\,\alpha(dy)$. By the assumption 3.3.2, $u\frac{\tau_2(u,y)}{\tau_1(u,y)} \le 1$. This means $h_n'(u)\ge 0$ and then $h_n(u)$ is nondecreasing in u. Similarly, from the assumption 3.3.2, it follows that $u\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}$ is also nondecreasing in u for all n_j. Thus, we have

$$\tilde{g}_n(u) := u\int_{\mathbb{X}}\tau_1(u,y)\alpha(dy)+\sum_{j=1}^{n(\pi)}u\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}$$

is nondecreasing in u. Since $g_n(u)$ is a continuously differentiable function such that $\int_0^\infty g_n(u)du<\infty$, it is then bounded and attain its maximum point at some point $u^2_{n,n(\pi)}$ satisfying $g_n'(u^2_{n,n(\pi)})=0$ or $\tilde{g}_n(u^2_{n,n(\pi)})=n-1$. Note that $\tilde{g}_n$ is also a continuous function and is then bounded on bounded interval. We claim that $u^2_{n,n(\pi)}\to\infty$ as $n\to\infty$. In fact, by assumption 3.3.2, $u\frac{\tau_{k+1}(u,y)}{\tau_k(u,y)}\le\phi(u)(k-C_k(y))$, $\forall k\in\mathbb{Z}^+$ and $y\in\mathbb{X}$, for some function $\phi(u)\in(0,1)$ which is nondecreasing in u and $\lim_{u\to\infty}\phi(u)=1$. Assume that $u^2_{n,n(\pi)}<\infty$ as $n\to\infty$. Then, $\phi(u^2_{n,n(\pi)})=\alpha<1$, which implies $\sum_{j=1}^{n(\pi)}u\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}<\alpha\left(n-\sum_{j=1}^{n(\pi)}C_j(Y_j)\right)$. Therefore,

$$n-1=\tilde{g}_n(u^2_{n,n(\pi)})<u^2_{n,n(\pi)}\int_{\mathbb{X}}\tau_1(u^2_{n,n(\pi)},y)\alpha(dy)+\alpha\left(n-\sum_{j=1}^{n(\pi)}C_j(Y_j)\right),$$

which implies

$$n(1-\alpha)+\alpha\sum_{j=1}^{n(\pi)}C_j(Y_j)-1<u_{n,n(\pi)}^2\int_{\mathbb{X}}\tau_1(u_{n,n(\pi)}^2,y)\alpha(dy)<\infty\,,$$

which is a contradiction.

Denote

$$\tilde{\tau}_k(u,y)=u\frac{\tau_{k+1}(u,y)}{\tau_k(u,y)}\,.$$

And let $u_{n,n(\pi)}$ be the positive square root of $u_{n,n(\pi)}^2$, thus $u_{n,n(\pi)}\to\infty$ as $n\to\infty$. Then, we have the following inequalities,

$$\begin{aligned}
k-C_k(y) &\geq \int_0^\infty u\frac{\tau_{k+1}(u,y)}{\tau_k(u,y)}f_{U_n|\mathbf{X}}(u|\mathbf{X})du=\frac{\int_0^\infty \tilde{\tau}_k(u,y)g_n(u)du}{\int_0^\infty g_n(u)du}\\
&\geq \frac{\int_{u_{n,n(\pi)}}^\infty \tilde{\tau}_k(u,y)g_n(u)du}{\int_0^\infty g_n(u)du}\geq \tilde{\tau}_k(u_{n,n(\pi)},y)\frac{\int_{u_{n,n(\pi)}}^\infty g_n(u)du}{\int_0^\infty g_n(u)du}\\
&=\tilde{\tau}_k(u_{n,n(\pi)},y)\left(1+\frac{\int_0^{u_{n,n(\pi)}} g_n(u)du}{\int_{u_{n,n(\pi)}}^\infty g_n(u)du}\right)^{-1}\\
&\geq \tilde{\tau}_k(u_{n,n(\pi)},y)\left(1+\frac{\int_0^{u_{n,n(\pi)}} g_n(u)du}{\int_{u_{n,n(\pi)}}^{u_{n,n(\pi)}^2} g_n(u)du}\right)^{-1}\\
&\geq \tilde{\tau}_k(u_{n,n(\pi)},y)\left(1+\frac{\int_0^{u_{n,n(\pi)}} g_n(u_{n,n(\pi)})du}{\int_{u_{n,n(\pi)}}^{u_{n,n(\pi)}^2} g_n(u_{n,n(\pi)})du}\right)^{-1}\\
&=\tilde{\tau}_k(u_{n,n(\pi)},y)\left(1+\frac{u_{n,n(\pi)}g_n(u_{n,n(\pi)})}{(u_{n,n(\pi)}^2-u_{n,n(\pi)})g_n(u_{n,n(\pi)})}\right)^{-1}\\
&=(u_{n,n(\pi)}-1)\frac{\tau_{k+1}(u_{n,n(\pi)},y)}{\tau_k(u_{n,n(\pi)},y)}\overset{n\to\infty}{\to}k-C_k(y)\,.
\end{aligned}\tag{3.6.7}$$

The last limit in eq. (3.6.7) is due to the following form

$$\lim_{n\to\infty}(u_{n,n(\pi)}-1)\frac{\tau_{k+1}(u_{n,n(\pi)},y)}{\tau_k(u_{n,n(\pi)},y)}=\lim_{n\to\infty}\frac{(u_{n,n(\pi)}-1)}{u_{n,n(\pi)}}\tilde{\tau}_k(u_{n,n(\pi)},y)\,,$$

and $\lim_{n\to\infty}\frac{(u_{n,n(\pi)}-1)}{u_{n,n(\pi)}}=1$, $\lim_{n\to\infty}\tilde{\tau}_k(u_{n,n(\pi)},y)=\lim_{u\to\infty}\tilde{\tau}_k(u,y)=k-C_k(y)$ by assumption 3.3.2. This completes the proof of the lemma. □

Now we are ready to give the proof of Theorem 3.3.3. To emphasise the finiteness of α, we use the notation that $\alpha = aH$, where $a = \alpha(\mathbb{X})$ is finite and H is some probability measure.

We would follow the similar idea as that in (Freedman and Diaconis, 1983) to define a class of semi-norms on $\mathbb{M}_{\mathbb{X}}$ such that convergence under such norms implies weak convergence. Let $\mathcal{A} = \{A_i\}_{i=1}^{\infty}$ be a measurable partition of $\mathbb{X}$. The semi-norm between two probability measures P_1 and P_2 in $\mathbb{M}_{\mathbb{X}}$ with respect to the partition $\mathcal{A}$ is defined by

$$|P_1 - P_2|_{\mathcal{A}} = \sqrt{\sum_{i=1}^{\infty}[P_1(A_i) - P_2(A_i)]^2}\,. \tag{3.6.8}$$

In order to show the posterior distribution of NRMI concentrates around its posterior mean, we have the following lemma.

Lemma 3.6.2. *For any given measurable partition $\mathcal{A}$,*

$$\mathbb{E}\left[|P - \mathbb{E}[P|\mathbf{X}]|_{\mathcal{A}}^2|\mathbf{X}\right] = \sum_{i=1}^{\infty}\mathrm{Var}[P(A_i)|\mathbf{X}] \to 0\,, \tag{3.6.9}$$

a.s.-P_0^{∞} as $n \to \infty$.

Proof. To prove this claim, we shall evaluate the first and second posterior moments of P for any $A \in \mathcal{X}$. For the first moment we have

$$\begin{aligned}\mathbb{E}[P(A)|\mathbf{X}] &= \frac{1}{n}\int_0^{\infty} u f_{U_n}(u) V_{\alpha(A)}^{(1)}(u)du + \frac{1}{n}\sum_{j=1}^{n(\pi)}\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A)du \\ &= \frac{a}{n}\int_0^{\infty} u f_{U_n}(u)\int_A \tau_1(u,x)H(dx)du + \frac{1}{n}\sum_{j=1}^{n(\pi)}\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A)du\,.\end{aligned}$$

For the second moment we have

$$\begin{aligned}\mathbb{E}[P(A)^2|\mathbf{X}] =& \frac{a}{n(n+1)}\int_0^{\infty} u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_A \tau_2(u,x)H(dx)du \\ &+ \frac{a^2}{n(n+1)}\int_0^{\infty} u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_A \tau_1(u,x)H(dx)\right)^2 du\end{aligned}$$

$$+ 2\frac{a}{n(n+1)}\sum_{j=1}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A)\int_A \tau_1(u,x)H(dx)du$$

$$+ \frac{1}{n(n+1)}\sum_{j=1}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A)du$$

$$+ 2\frac{1}{n(n+1)}\sum_{j\neq k}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_k+1}(u,Y_k)}{\tau_{n_k}(u,Y_k)}\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_i}(A)\delta_{Y_j}(A)du\,.$$

Then, we can write

$$\sum_{i=1}^\infty \mathrm{Var}[P(A_i)|\mathbf{X}] = \sum_{i=1}^\infty \left(\mathbb{E}[P(A)^2|\mathbf{X}] - \mathbb{E}[P(A)|\mathbf{X}]^2\right) = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 + \mathcal{J}_4\,,$$

where the terms $\mathcal{J}_1$, $\mathcal{J}_2$, $\mathcal{J}_3$, $\mathcal{J}_4$ are defined as follows.

$$\begin{aligned}\mathcal{J}_1 =& \frac{a}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{\mathbb{X}}\tau_2(u,x)H(dx)du \\ &+ \frac{a^2}{n(n+1)}\sum_{i=1}^\infty\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_{A_i}\tau_1(u,x)H(dx)\right)^2 du \\ &- \frac{a^2}{n^2}\sum_{i=1}^\infty\left(\int_0^\infty u f_{U_n}(u)\int_{A_i}\tau_1(u,x)H(dx)du\right)^2\,; \end{aligned} \qquad (3.6.10)$$

$$\begin{aligned}\mathcal{J}_2 =& 2\frac{a}{n(n+1)}\sum_{i=1}^\infty\sum_{j=1}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A_i) \\ &\times\int_{A_i}\tau_1(u,x)H(dx)du - 2\frac{a}{n^2}\sum_{i=1}^\infty\sum_{j=1}^{n(\pi)}\int_0^\infty u f_{U_n}(u)\int_{A_i}\tau_1(u,x)H(dx)du \\ &\times\int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A_i)du\,; \end{aligned} \qquad (3.6.11)$$

$$\begin{aligned}\mathcal{J}_3 =& \frac{1}{n(n+1)}\sum_{i=1}^\infty\sum_{j=1}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A_i)du \\ &- \frac{1}{n^2}\sum_{i=1}^\infty\sum_{j=1}^{n(\pi)}\left(\int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A_i)du\right)^2 \end{aligned} \qquad (3.6.12)$$

$$= \frac{1}{n(n+1)} \sum_{j=1}^{n(\pi)} \int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+2}(u, Y_j)}{\tau_{n_j}(u, Y_j)} du$$

$$- \frac{1}{n^2} \sum_{j=1}^{n(\pi)} \left(\int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} du \right)^2 ; \tag{3.6.13}$$

and

$$\begin{aligned} \mathcal{J}_4 =& 2\frac{1}{n(n+1)} \sum_{i=1}^{\infty} \sum_{j \neq k}^{n(\pi)} \int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_k+1}(u, Y_k)}{\tau_{n_k}(u, Y_k)} \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_k}(A_i) \delta_{Y_j}(A_i) du \\ & - 2\frac{1}{n^2} \sum_{i=1}^{\infty} \sum_{j \neq k}^{n(\pi)} \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_k+1}(u, Y_k)}{\tau_{n_k}(u, Y_k)} \delta_{Y_k}(A_i) du \\ & \times \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A_i) du \,. \end{aligned} \tag{3.6.14}$$

We will first consider the terms $\mathcal{J}_2$, $\mathcal{J}_3$, $\mathcal{J}_4$ and then $\mathcal{J}_1$. But before dealing with them, we need some prior preparations. By the identity $\mathbb{E}[P(\mathbb{X})|\mathbf{X}] = 1$ we have

$$\frac{a}{n} \int_0^\infty u f_{U_n}(u) \int_{\mathbb{X}} \tau_1(u, x) H(dx) du + \frac{1}{n} \sum_{j=1}^{n(\pi)} \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} du = 1 \,.$$

By Lemma 3.6.1, we have the approximation

$$\frac{a}{n} \int_0^\infty u f_{U_n|\mathbf{X}}(u) \int_{\mathbb{X}} \tau_1(u, x) H(dx) du \overset{n}{\approx} \frac{\sum_{j=1}^{n(\pi)} C_{n_j}(Y_j)}{n} \tag{3.6.15}$$

as n is large. On the other hand, let $u_{n,n(\pi)}$ be the maximal point of $g_n(u)$ as in Lemma 3.6.1. Under the assumption 3.3.2, we know that $u\tau_1(u, x)$ is nondecreasing in u for all x. We have

$$\begin{aligned} a & \int_0^{u_{n,n(\pi)}} u f_{U_n|\mathbf{X}}(u) \int_{\mathbb{X}} \tau_1(u, x) H(dx) du \\ &= a\frac{\int_0^{u_{n,n(\pi)}} u g_n(u) \int_{\mathbb{X}} \tau_1(u, x) H(dx) du}{\int_0^\infty g_n(u) du} \\ &\leq a u_{n,n(\pi)} \int_{\mathbb{X}} \tau_1(u_{n,n(\pi)}, x) H(dx) \frac{\int_0^{u_{n,n(\pi)}} g_n(u) du}{\int_0^\infty g_n(u) du} \end{aligned} \tag{3.6.16}$$

$$= au_{n,n(\pi)} \int_{\mathbb{X}} \tau_1(u_{n,n(\pi)}, x)H(dx) \left(1 + \frac{\int_{u_{n,n(\pi)}}^{\infty} g_n(u)du}{\int_0^{u_{n,n(\pi)}} g_n(u)du}\right)^{-1}$$

$$\le au_{n,n(\pi)} \int_{\mathbb{X}} \tau_1(u_{n,n(\pi)}, x)H(dx) \left(1 + \frac{\int_{u_{n,n(\pi)}}^{u^2_{n,n(\pi)}} g_n(u_{n,n(\pi)})du}{\int_0^{u_{n,n(\pi)}} g_n(u_{n,n(\pi)})du}\right)^{-1}$$

$$= au_{n,n(\pi)} \int_{\mathbb{X}} \tau_1(u_{n,n(\pi)}, x)H(dx) \left(1 + \frac{u_{n,n(\pi)}(u_{n,n(\pi)} - 1)g_n(u_{n,n(\pi)})}{u_{n,n(\pi)}g_n(u_{n,n(\pi)})}\right)^{-1}$$

$$= a \int_{\mathbb{X}} \tau_1(u_{n,n(\pi)}, x)H(dx)\,, \tag{3.6.17}$$

which goes to 0 as $n \to \infty$ by the Monotone convergence theorem, since $\tau_1(u, x)$ is decreasing to 0 in u for all x. Combining the above computation with the approximation (3.6.15), we have

$$\lim_{n\to\infty} \frac{a}{\sum_{j=1}^{n(\pi)} C_{n_j}(Y_j)} \int_{u_{n,n(\pi)}}^{\infty} u f_{U_n|\mathbf{X}}(u) \int_{\mathbb{X}} \tau_1(u, x)H(dx)du = 1\,. \tag{3.6.18}$$

Step 1: Evaluation of $\mathcal{J}_2$.

Notice first that for any A_i and Y_j, by the assumption 3.3.2, we will have

$$\begin{aligned} I_1 &:= \int_0^{\infty} u^2 \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A_i) \int_{A_i} \tau_1(u, x)H(dx) f_{U_n|\mathbf{X}}(u|\mathbf{X})du \\ &\le \left(n_j - C_{n_j}(Y_j)\right) \delta_{Y_j}(A_i) \int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \int_{A_i} \tau_1(u, x)H(dx)du\,. \end{aligned} \tag{3.6.19}$$

On the other hand,

$$\begin{aligned} I_1 &\ge \int_{u_{n,n(\pi)}}^{\infty} u^2 \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A_i) \int_{A_i} \tau_1(u, x)H(dx) f_{U_n|\mathbf{X}}(u|\mathbf{X})du \\ &\ge u_{n,n(\pi)} \frac{\tau_{n_j+1}(u_{n,n(\pi)}, Y_j)}{\tau_{n_j}(u_{n,n(\pi)}, Y_j)} \delta_{Y_j}(A_i) \int_{u_{n,n(\pi)}}^{\infty} u f_{U_n|\mathbf{X}}(u) \int_{\mathbb{X}} \tau_1(u, x)H(dx)du\,. \end{aligned} \tag{3.6.20}$$

By the above inequalities (3.6.19), (3.6.20) and the approximation (3.6.15), (3.6.18), we can see as n becomes large

$$\int_0^{\infty} u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u, Y_j)}{\tau_{n_j}(u, Y_j)} \delta_{Y_j}(A_i) \int_{A_i} \tau_1(u, x)H(dx)du$$

$$\overset{n}{\sim} \int_0^\infty u f_{U_n}(u) \int_{A_i} \tau_1(u,x) H(dx) du \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)} \delta_{Y_j}(A_i) du \,.$$

Thus, for large n, we have

$$\begin{aligned}
\mathcal{J}_2 \overset{n}{\sim} & 2\left(\frac{a}{n(n+1)} - \frac{a}{n^2}\right) \sum_{i=1}^{\infty} \sum_{j=1}^{n(\pi)} \left(n_j - C_{n_j}(Y_j)\right) \delta_{Y_j}(A_i) \\
& \times \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \int_{A_i} \tau_1(u,x) H(dx) du \\
\overset{n}{\sim} & 2\left(\frac{a}{n(n+1)} - \frac{a}{n^2}\right) \sum_{j=1}^{n(\pi)} \left(n_j - C_{n_j}(Y_j)\right) \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \int_{\mathbb{X}} \tau_1(u,x) H(dx) du \,.
\end{aligned}$$

This combined with (3.6.15) yields

$$\mathcal{J}_2 \overset{n}{\sim} -2 \frac{\left(n - \sum_{j=1}^{n(\pi)} C_{n_j}(Y_j)\right) \left(\sum_{j=1}^{n(\pi)} C_{n_j}(Y_j)\right)}{n^2(n+1)} \,, \tag{3.6.21}$$

which has order $O(\frac{1}{n})$.

Step 2: Evaluation of $\mathcal{J}_3$.

For $\mathcal{J}_3$, notice that under the assumption 3.3.2, we have

$$u^2 \frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j}(u,Y_j)} = u \frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j+1}(u,Y_j)} \times u \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}$$

is nondecreasing in u and is bounded by $(n_j + 1 - C_{n_j+1}(Y_j))(n_j - C_{n_j}(Y_j))$. Using a similar approach as that in Lemma 3.6.1, we have as n is large,

$$\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j}(u,Y_j)} du \overset{n}{\sim} (n_j + 1 - C_{n_j+1}(Y_j))(n_j - C_{n_j}(Y_j)) \,. \tag{3.6.22}$$

Combining it with Lemma 3.6.1, we have as n becomes large

$$\begin{aligned}
\mathcal{J}_3 \overset{n}{\sim} & \frac{1}{n(n+1)} \sum_{j=1}^{n(\pi)} (n_j + 1 - C_{n_j+1}(Y_j))(n_j - C_{n_j}(Y_j)) - \frac{1}{n^2} \sum_{j=1}^{n(\pi)} (n_j - C_{n_j}(Y_j))^2 \\
= & \frac{1}{n^2(n+1)} \sum_{j=1}^{n(\pi)} (n_j - C_{n_j}(Y_j)) \left(n + (n+1) C_{n_j} - n_j - n C_{n_j+1}\right)
\end{aligned}$$

$$\leq 2\frac{n - \left(\sum_{j=1}^{n(\pi)} C_{n_j}(Y_j)\right)}{n(n+1)}, \tag{3.6.23}$$

which has order at most $O(\frac{1}{n})$.

Step 3: Evaluation of $\mathcal{J}_4$.

For $\mathcal{J}_4$, we have that under the assumption 3.3.2, $u^2 \frac{\tau_{n_k+1}(u,Y_k)}{\tau_{n_k}(u,Y_k)} \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}$ is nondecreasing in u and is bounded by $(n_k - C_{n_k}(Y_k))(n_j - C_{n_j}(Y_j))$. Using a similar argument to that in Lemma 3.6.1 leads to

$$\begin{aligned}
&\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_k+1}(u,Y_k)}{\tau_{n_k}(u,Y_k)} \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)} du \\
&\qquad \overset{n}{\sim} \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_k+1}(u,Y_k)}{\tau_{n_k}(u,Y_k)} du \int_0^\infty u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)} du \\
&\qquad \overset{n}{\sim} (n_k - C_{n_k}(Y_k))(n_j - C_{n_j}(Y_j)) \,.
\end{aligned}$$

Thus

$$\begin{aligned}
\mathcal{J}_4 &\overset{n}{\sim} 2\left(\frac{1}{n(n+1)} - \frac{1}{n^2}\right) \sum_{i=1}^{\infty} \sum_{j\neq k}^{n(\pi)} (n_k - C_{n_k}(Y_k))(n_j - C_{n_j}(Y_j)) \delta_{Y_k}(A_i) \delta_{Y_j}(A_i) \\
&\overset{n}{\sim} -\frac{2\sum_{j\neq k}^{n(\pi)} (n_k - C_{n_k}(Y_k))(n_j - C_{n_j}(Y_j))}{n^2(n+1)},
\end{aligned}$$

which has an order at most $O(\frac{1}{n})$.

Step 4: Evaluation of $\mathcal{J}_1$.

Finally, we deal with the term $\mathcal{J}_1$. Notice that $\mathbb{E}[P(\mathbb{X})^2|\mathbf{X}] = 1$. Using the computation we obtained for $\mathcal{J}_2$, $\mathcal{J}_3$, $\mathcal{J}_4$, we have

$$\begin{aligned}
1 = \mathbb{E}[P(\mathbb{X})^2|\mathbf{X}] =& \frac{a}{n(n+1)} \int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \int_{\mathbb{X}} \tau_2(u,x) H(dx) du \\
&+ \frac{a^2}{n(n+1)} \int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \left(\int_{\mathbb{X}} \tau_1(u,x) H(dx)\right)^2 du \\
&+ 2\frac{a}{n(n+1)} \sum_{j=1}^{n(\pi)} \int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)} \int_{\mathbb{X}} \tau_1(u,x) H(dx) du
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n(n+1)}\sum_{j=1}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+2}(u,Y_j)}{\tau_{n_j}(u,Y_j)}du\\
&+2\frac{1}{n(n+1)}\sum_{j\neq k}^{n(\pi)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_k+1}(u,Y_k)}{\tau_{n_k}(u,Y_k)}\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}du\\
\overset{n}{\sim}&\frac{a}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{\mathbb{X}}\tau_2(u,x)H(dx)du\\
&+\frac{a^2}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_{\mathbb{X}}\tau_1(u,x)H(dx)\right)^2du\\
&+2\frac{\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)\left(n-\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)\right)}{n(n+1)}\\
&+\frac{\sum_{j=1}^{n(\pi)}(n_j+1-C_{n_j+1}(Y_j)-\frac{n_j-C_{n_j}}{n})(n_j-C_{n_j}(Y_j))}{n(n+1)}\\
&+2\frac{\sum_{j\neq k}^{n(\pi)}(n_k-C_{n_k}(Y_k))(n_j-C_{n_j}(Y_j))}{n(n+1)}.
\end{aligned}
$$

This implies

$$
\begin{aligned}
&\frac{a}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{\mathbb{X}}\tau_2(u,x)H(dx)du\\
&+\frac{a^2}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_{\mathbb{X}}\tau_1(u,x)H(dx)\right)^2du\\
&\overset{n}{\sim}\frac{n+\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)^2-\sum_{j=1}^{n(\pi)}(n_j-C_{n_j}(Y_j))(1+C_{n_j}-C_{n_j+1}-\frac{n_j-C_{n_j}}{n})}{n(n+1)}.
\end{aligned}
\tag{3.6.24}
$$

Combining the approximations (3.6.15) and (3.6.24), we have

$$
\begin{aligned}
\mathcal{J}_1=&\frac{a}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{\mathbb{X}}\tau_2(u,x)H(dx)du\\
&+\frac{a^2}{n(n+1)}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_{\mathbb{X}}\tau_1(u,x)H(dx)\right)^2du\\
&-\frac{a^2}{n^2}\left(\int_0^\infty u f_{U_n}(u)\int_{\mathbb{X}}\tau_1(u,x)H(dx)du\right)^2\\
&+2\sum_{i\neq l}^\infty\int_0^\infty u f_{U_n}(u)\int_{A_i}\tau_1(u,x)H(dx)du\int_0^\infty u f_{U_n}(u)\int_{A_l}\tau_1(u,x)H(dx)du\\
&-2\sum_{i\neq l}^\infty\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{A_i}\tau_1(u,x)H(dx)\int_{A_l}\tau_1(u,x)H(dx)du
\end{aligned}
$$

$$
\begin{aligned}
&\overset{n}{\sim}\frac{n+\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)^2-\sum_{j=1}^{n(\pi)}(n_j-C_{n_j}(Y_j))(1+C_{n_j}-C_{n_j+1}-\frac{n_j-C_{n_j}}{n})}{n(n+1)}\\
&-\frac{\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)^2}{n^2}\\
&+2\frac{a^2}{n^2}\sum_{i\neq l}^{\infty}\int_0^\infty uf_{U_n}(u)\int_{A_i}\tau_1(u,x)H(dx)du\int_0^\infty uf_{U_n}(u)\int_{A_l}\tau_1(u,x)H(dx)du\\
&-2\frac{a^2}{n(n+1)}\sum_{i\neq l}^{\infty}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{A_i}\tau_1(u,x)H(dx)\int_{A_l}\tau_1(u,x)H(dx)du\,. \quad (3.6.25)
\end{aligned}
$$

We now treat the above last two summation terms. First, we have

$$
\begin{aligned}
&2\sum_{i\neq l}^{\infty}\int_0^\infty uf_{U_n}(u)\int_{A_i}\tau_1(u,x)H(dx)du\int_0^\infty uf_{U_n}(u)\int_{A_l}\tau_1(u,x)H(dx)du\\
&\overset{n}{\sim}\left(\int_0^\infty uf_{U_n}(u)\int_{\mathbb{X}}\tau_1(u,x)H(dx)du\right)^2
\end{aligned}
$$

and

$$
\begin{aligned}
&2\sum_{i\neq l}^{\infty}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\int_{A_i}\tau_1(u,x)H(dx)\int_{A_l}\tau_1(u,x)H(dx)du\\
&\overset{n}{\sim}\int_0^\infty u^2 f_{U_n|\mathbf{X}}(u|\mathbf{X})\left(\int_{\mathbb{X}}\tau_1(u,x)H(dx)\right)^2du\,.
\end{aligned}
$$

Thus,

$$
\mathcal{J}_1\overset{n}{\sim}\frac{n^2-n\sum_{j=1}^{n(\pi)}(n_j-C_{n_j}(Y_j))(1+C_{n_j}-C_{n_j+1})-\left(\sum_{j=1}^{n(\pi)}C_{n_j}(Y_j)\right)^2}{n^2(n+1)}\,.
$$

It is easy to have that

$$
\sum_{j=1}^{n(\pi)}(n_j-C_{n_j}(Y_j))(1+C_{n_j}-C_{n_j+1})\leq 3n\sum_{j=1}^{n(\pi)}(n_j-C_{n_j}(Y_j))\leq 3n^2\,.
$$

Thus $\mathcal{J}_1$ has an order $O(\frac{1}{n})$.

Summarizing the above four steps for evaluating $\mathcal{J}_1$, $\mathcal{J}_2$, $\mathcal{J}_3$, $\mathcal{J}_4$, we have

$$\sum_{i=1}^{\infty} \mathrm{Var}[P(A_i)|\mathbf{X}] \overset{n}{\sim} O(\frac{1}{n}) \to 0 \qquad \text{as } n \to \infty \,.$$

□

Now, we can give the completion of theorem 3.3.3.

Proof. By lemma 3.6.2, the distribution of $P(\cdot|\mathbf{X})$ converges weakly to the point mass at the distribution of $\lim_{n\to\infty} \mathbb{E}[P(dx)|\mathbf{X}]$.

If the "true" distribution P_0 of $\mathbf{X}$ is continuous, the posterior expectation has the following form for any $A \in \mathcal{X}$.

$$\begin{aligned}\mathbb{E}[P(A)|\mathbf{X}] =& \frac{a}{n}\int_0^{\infty} u f_{U_n}(u) \int_A \tau_1(u,x)H(dx)du \\ &+ \frac{1}{n}\sum_{j=1}^{n}\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_2(u,X_j)}{\tau_1(u,X_j)}\delta_{X_j}(A)du \,. \end{aligned} \tag{3.6.26}$$

As $n \to \infty$, by (3.6.7), the weight $\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_2(u,X_j)}{\tau_1(u,X_j)}du \overset{n}{\sim} 1 - C_1(X_j)$, thus the second part of the summation in (3.6.26) has the form $\sum_{j=1}^{n} \frac{1-C_1(X_j)}{n}\delta_{X_j}(A)$ that converges uniformly over Glivenko-Cantelli classes to $(1-\bar{C}_1)P_0$. Since the sum of the weights of $H(dx)$ and $\delta_{X_j}(dx)$ is equal to 1, we have $\lim_{n\to\infty}\mathbb{E}[P(\cdot)|\mathbf{X}] = \bar{C}_1 H(\cdot) + (1-\bar{C}_1)P_0(\cdot)$.

If the "true" distribution P_0 of $\mathbf{X}$ is discrete with $\lim_{n\to\infty}\frac{n(\pi)}{n} = 0$, the posterior expectation has the following form:

$$\begin{aligned}\mathbb{E}[P(A)|\mathbf{X}] =& \frac{a}{n}\int_0^{\infty} u f_{U_n}(u) \int_A \tau_1(u,x)H(dx)du \\ &+ \frac{1}{n}\sum_{j=1}^{n(\pi)}\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X}) \frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)}\delta_{Y_j}(A)du \,. \end{aligned} \tag{3.6.27}$$

As $n \to \infty$, by (3.6.7), $\int_0^{\infty} u f_{U_n|\mathbf{X}}(u|\mathbf{X})\frac{\tau_{n_j+1}(u,Y_j)}{\tau_{n_j}(u,Y_j)} \overset{n}{\sim} n_j - C_{n_j}(Y_j)$. Hence, the second part of the summation in (3.6.27) has the form $\sum_{j=1}^{n(\pi)} \frac{n_j - C_{n_j}(Y_j)}{n} \delta_{Y_j}(A)$. Notice that

$$\sum_{j=1}^{n(\pi)} \frac{n_j - C_{n_j}(Y_j)}{n} \delta_{Y_j}(A) = \sum_{i=1}^{n} \frac{1}{n} \delta_{X_i}(A) - \sum_{j=1}^{n(\pi)} \frac{C_{n_j}(Y_j)}{n} \delta_{Y_j}(A) \,,$$

where the term $\sum_{j=1}^{n(\pi)} \frac{C_{n_j}(Y_j)}{n} \leq \sum_{j=1}^{n(\pi)} \frac{1}{n} = \frac{n(\pi)}{n}$ converges to 0. Thus the weight of $H(dx)$ is 0 and we have $\lim_{n\to\infty} \mathbb{E}[P(\cdot)|\mathbf{X}] = P_0(\cdot)$. This completes the proof of Theorem 3.3.3. $\square$

Proof of Theorem 3.4.1

As the preparation of the proof of theorem 3.4.1, we shall present the posterior process of NGGP(a, σ, θ, H) followed by theorem 3.2.2.

Lemma 3.6.3. *If $P \sim \mathrm{NGGP}(a, \sigma, \theta, H)$, conditionally on $\mathbf{X}$ and a latent random variable U_n, P coincides in distribution with the random probability measure*

$$\kappa_n P_{U_n} + (1 - \kappa_n) \sum_{j=1}^{n(\pi)} D_{n,j} \delta_{Y_j} \,, \tag{3.6.28}$$

where

(i) The random variable U_n has density

$$f_{U_n}(u) \propto \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^{\sigma}} \,. \tag{3.6.29}$$

(ii) Given $U_n = u$, $P_{U_n} \sim \mathrm{NGGP}(a, \sigma, \theta + u, H)$.

(iii) $D_n := (D_{n,1}, \cdots, D_{n,n(\pi)}) \sim \mathrm{Dir}(n(\pi); n_1 - \sigma, \cdots n_{n(\pi)})$ is independent of κ_n and P_{U_n}.

(iv) The random elements P_{U_n} and J_j, $j \in \{1, \cdots, n(\pi)\}$ are independent.

(v) $T_{(U_n)} = \tilde{\mu}_{(U_n)}(\mathbb{X})$ and $\kappa_n = \frac{T_{(U_n)}}{T_{(U_n)} + \sum_{j=1}^{n(\pi)} J_j}$.

Proof. The lemma is an immediate result from theorem 3.2.2 and the NGGP intensity given in example 3.3.8. To verify (iii), we let $D_{n,j} := \frac{J_j}{\sum_{j=1}^{n(\pi)} J_j}$. Since $\{J_1, \cdots, J_{n(\pi)}\}$ are independent $G(n_j - \sigma, u + \theta)$ random variables with density

$$f_{J_j}(t|U_n = u) = \frac{(u+\theta)^{n_j - \sigma}}{\Gamma(n_j - \sigma)} t^{n_j - \sigma - 1} e^{-(u+\theta)t},$$

By the Proposition G.2 in (Ghosal and Van der Vaart, 2017), we have

$$D_n := (n(\pi); D_{n,1}, \cdots, D_{n,n(\pi)}) \sim \text{Dir}(n_1 - \sigma, \cdots n_{n(\pi)}),$$

which is totally independent of U_n, thus independent of κ_n and P_{U_n}. To understand the independence, we can use the relationship between Dirichlet distribution and the gamma distribution form the Proposition G.2 in (Ghosal and Van der Vaart, 2017) and let $D_{n,j} = \frac{\gamma_j}{\sum_{j=1}^{n(\pi)} \gamma_j}$, where $\gamma_j \sim G(n_j - \sigma, 1)$. □

The convergences 3.4.2 and 3.4.3 are equivalent in theorem 3.4.1, and also the convergences 3.4.4 and 3.4.5 are equivalent. These equivalences can be shown by the following lemma. To make the results lavish, we will assume $\{\sigma_i\}_{i=1}^n$ be a sequence such that $\lim_{n\to\infty} \sigma_n = \sigma \in [0, 1)$ in the following proofs. It is worth to point that, we always assume that $\sigma_i < 1$ and $\sigma < 1$ to make sure all quantities in this work are well-defined. To be more precise, this assumption would make the forms $\int_0^\infty s^{n_j - \sigma_i - 1} e^{-(u+\theta)s} ds < \infty$ and $\int_0^\infty s^{n_j - \sigma - 1} e^{-(u+\theta)s} ds < \infty$ for any integer $n_j \geq 1$.

Lemma 3.6.4. *For any P_0, we have*

(i) If P_0 is discrete,

$$\lim_{n\to\infty} \mathbb{E}[P|\mathbf{X}] = \lim_{n\to\infty} \mathbb{P}_n + \frac{\sigma_n n(\pi)}{n}(H - \tilde{\mathbb{P}}_n) = P_0 . \tag{3.6.30}$$

(ii) If P_0 is continuous,

$$\lim_{n\to\infty} \mathbb{E}[P|\mathbf{X}] = \lim_{n\to\infty} (1 - \sigma_n)\mathbb{P}_n + \sigma_n H = (1 - \sigma_n)P_0 + \sigma_n H . \tag{3.6.31}$$

Proof. Since the convergence of σ_n to σ won't affect the proof, and σ_n is well-defined as discussed previously, we may fix σ_n and use σ for the sake of notational simplicity in the proof.

By applying the NGGP intensity 3.3.5 to lemma 3.3.1, we have

$$\mathbb{E}[P|\mathbf{X}] = \frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)duH + \frac{1}{n}\sum_{j=1}^{n(\pi)}\int_0^\infty \frac{(n_j-\sigma)u}{u+\theta} f_{U_n}(u)du\delta_{Y_j}\,. \tag{3.6.32}$$

To evaluate $\lim_{n\to\infty}\mathbb{E}[P|\mathbf{X}]$, we need to find the limits of

$$\frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)du\,, \tag{3.6.33}$$

$$\frac{1}{n}\int_0^\infty \frac{u}{u+\theta} f_{U_n}(u)du\,. \tag{3.6.34}$$

We will find the limit of 3.6.33 and then 3.6.34. For 3.6.33 by the density of U_n, we have

$$\frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)du = \frac{\frac{a}{n}\int_0^\infty \frac{u^n}{(u+\theta)^{n+1-(n(\pi)+1)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}du}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}du}\,. \tag{3.6.35}$$

By the similar arguments in lemma 3.6.1, we use the Laplace method to find the limit of the nominator and denominator of 3.6.35. Let

$$g_1(u) = \frac{u^n}{(u+\theta)^{n+1-(n(\pi)+1)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}\,, \qquad g_2(u) = \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}\,.$$

Thus,

$$g_1'(u) = \{n(u+\theta) + ((n(\pi)+1)\sigma - (n+1))u - au(u+\theta)^\sigma\}\frac{u^{n-1}}{(u+\theta)^{n+2-(n(\pi)+1)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}\,,$$

$$g_2'(u) = \{(n-1)(u+\theta) + (n(\pi)\sigma - n)u - au(u+\theta)^\sigma\}\frac{u^{n-2}}{(u+\theta)^{n+1-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma}$$

As $n\to$, by the similar arguments in lemma 3.6.1, $g_1(u)$ and $g_2(u)$ attain their maximums at $u_{1,n}$, $u_{2,n}$ that are both infinity large. Thus, $u_{1,n} \approx \left\{\frac{(n(\pi)+1)\sigma-1}{a}\right\}^{\frac{1}{\sigma}} - \theta$, and $u_{2,n} \approx$

$\left\{\frac{n(\pi)\sigma-1}{a}\right\}^{\frac{1}{\sigma}} - \theta$. Therefore, followed by eq. (3.6.35),

$$\begin{aligned}
\lim_{n\to\infty} \frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)du &= \lim_{n\to\infty} \frac{a}{n}\frac{g_1(u_{1,n})}{g_2(u_{2,n})} \\
&= \lim_{n\to\infty} \frac{a}{n} \frac{\left\{\left(\frac{(n(\pi)+1)\sigma-1}{a}\right)^{\frac{1}{\sigma}} - \theta\right\}^n \left(\frac{(n(\pi)+1)\sigma-1}{a}\right)^{n(\pi)+1-\frac{n+1}{\sigma}} e^{-(n(\pi)+1)-1/\sigma}}{\left\{\left(\frac{n(\pi)\sigma-1}{a}\right)^{\frac{1}{\sigma}} - \theta\right\}^{n-1} \left(\frac{n(\pi)\sigma-1}{a}\right)^{n(\pi)-\frac{n}{\sigma}} e^{-n(\pi)-1/\sigma}} \\
&= \lim_{n\to\infty} \frac{((n(\pi)+1)\sigma-1)^{(n(\pi)+1)-\frac{1}{\sigma}}}{n\,(n(\pi)\sigma-1)^{n(\pi)-\frac{1}{\sigma}}} \frac{e^{-\theta a^{1/\sigma}\left(\frac{n}{(n(\pi)+1)\sigma-1}\right)^{1/\sigma} n^{1-1/\sigma}}}{e^{-\theta a^{1/\sigma}\left(\frac{n}{n(\pi)\sigma-1}\right)^{1/\sigma}(n-1)n^{-1/\sigma}}} e^{-1}\,. \qquad (3.6.36)
\end{aligned}$$

Recall remark 3.3.4, when P_0 is discrete, $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$,almost surely. The limit in 3.6.36 becomes

$$\begin{aligned}
&\lim_{n\to\infty} \frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)du \\
&= \lim_{n\to\infty} \frac{1}{n}\frac{((n(\pi)+1)\sigma-1)^{(n(\pi)+1)-\frac{1}{\sigma}}}{(n(\pi)\sigma-1)^{n(\pi)-\frac{1}{\sigma}}} e^{-\theta a^{1/\sigma}\left\{\left(\frac{1}{(n(\pi)+1)\sigma-1}\right)^{1/\sigma} n - \left(\frac{1}{n(\pi)\sigma-1}\right)^{1/\sigma}(n-1)\right\}} e^{-1} \\
&= 0\,,
\end{aligned}$$

where the exponential part in the last equation converges to 0 by the fact that $\left(\frac{1}{(n(\pi)+1)\sigma-1}\right)^{1/\sigma} - \left(\frac{1}{n(\pi)\sigma-1}\right)^{1/\sigma} > 0$ for $\sigma \in (0,1)$.

When P_0 is continuous, $\lim_{n\to\infty} \frac{n(\pi)}{n} = 1$,almost surely. The limit in 3.6.36 becomes

$$\begin{aligned}
&\lim_{n\to\infty} \frac{a}{n}\int_0^\infty \frac{u}{(u+\theta)^{1-\sigma}} f_{U_n}(u)du \\
&= \lim_{n\to\infty} \frac{1}{n}\frac{((n+1)\sigma-1)^{(n+1)-\frac{1}{\sigma}}}{(n\sigma-1)^{n-\frac{1}{\sigma}}} e^{-\theta a^{1/\sigma}\left\{\left(\frac{1}{(n)+1)\sigma-1}\right)^{1/\sigma} n - \left(\frac{1}{n\sigma-1}\right)^{1/\sigma}(n-1)\right\}} e^{-1} \\
&= \lim_{n\to\infty} \frac{((n+1)\sigma-1)}{n}\frac{((n+1)\sigma-1)^{-\frac{1}{\sigma}}}{(n\sigma-1)^{-\frac{1}{\sigma}}}\left(1+\frac{\frac{n\sigma}{n\sigma-1}}{n}\right)^n e^{-\theta a^{1/\sigma}\left\{\left(\frac{1}{(n)+1)\sigma-1}\right)^{1/\sigma} n - \left(\frac{1}{n\sigma-1}\right)^{1/\sigma}(n-1)\right\}} e^{-1} \\
&= \sigma\,,
\end{aligned}$$

where we emphasis that $\frac{1}{\sigma} > 1$ when dealing with the convergence of the exponential part.

By using the same arguments above for finding the limit of 3.6.33, we can find the limit of 3.6.34. We omit the details of the computation and can obtain the following results.

When P_0 is discrete,

$$\lim_{n\to\infty} \frac{1}{n} \int_0^\infty \frac{u}{u+\theta} f_{U_n}(u)du = 1\,.$$

Thus,

$$\lim_{n\to\infty} \mathbb{E}[P|\mathbf{X}] = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n(\pi)} (n_j - \sigma)\delta_{Y_j} = \lim_{n\to\infty} \mathbb{P}_n + \frac{\sigma n(\pi)}{n}(H - \tilde{\mathbb{P}}_n) = P_0\,,$$

where the last equation is due to $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ and the Borel–Cantelli lemma . That is to say, the result in 3.6.30 is completed by combining the limit of 3.6.33 and 3.6.34 when P_0 is discrete.

When P_0 is continuous,

$$\lim_{n\to\infty} \frac{1}{n} \int_0^\infty \frac{u}{u+\theta} f_{U_n}(u)du = 1 - \sigma\,.$$

Thus, combining the limit of 3.6.33 and 3.6.34, we have

$$\lim_{n\to\infty} \mathbb{E}[P|\mathbf{X}] = \lim_{n\to\infty} \sigma H + \frac{1}{n} \sum_{j=1}^{n} (1-\sigma)\delta_{X_j} = \lim_{n\to\infty} \sigma H + (1-\sigma)\mathbb{P}_n = \sigma H + (1-\sigma)P_0\,.$$

Thus the proof of the result in 3.6.31 is completed.

□

With the lemma 3.6.4, it is sufficient to proof theorem 3.4.1 by only showing the convergences 3.4.2 and 3.4.4. The following lemma plays an important role in the proof of theorem 3.4.1. Here, we recall that an *envelop function* of $\mathbb{F}$ is a measurable function $f_e : \mathbb{X} \to \mathbb{R}$ such that $|f| < f_e$, for any $f \in \mathbb{F}$.

Lemma 3.6.5. *Let $\mathbb{F}$ be a finite set of $H-$square integrable functions. Assume that $n(\pi) \to \infty$ as $n \to \infty$, which includes the case when P_0 is continuous so that $n(\pi) = n$*

and the case when P_0 is discrete but $n(\pi)$ converges to ∞ with a lower rate than n do. Then

$$\sqrt{\sigma_n n(\pi)}(P_{U_n} - H)|\mathbf{X} \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}_H^o \,, \qquad a.s., \tag{3.6.37}$$

in $\mathbb{R}^{\mathbb{F}}$. The convergence holds a.s. in $l^\infty(\mathbb{F})$ with an envelop function f_e such that $H(f_e^2) < \infty$, and thus the central limit theorem holds for $P_{U_n}|\mathbf{X}$ in $l^\infty(\mathbb{F})$.

Proof. The proof relies on the stick-breaking representation of P_{U_n} in (Favaro et al., 2016) and the functional central limit theorem of NGGP in (Hu and Zhang, 2022). And similarly as discussed in the proof of last lemma, we use σ instead of σ_n to make the interpretation easy to read.

By section 4.2 in (Favaro et al., 2016), P_{U_n} admits a stick-breaking representation with dependent stick-breaking weights $\{v_i\}_{i=1}^\infty$, and the joint distribution of $\{v_i\}_{i=1}^\infty$ are given (Hu and Zhang, 2022) as

$$\begin{aligned} f(v_1, \cdots, v_k) = & \frac{\beta_n^k \sigma^{k-1}}{[\Gamma(1-\sigma)]^k \Gamma(k\sigma)} \prod_{i=1}^k v_i^{-\sigma}(1-v_i)^{-(k-i)\sigma-1} e^{-\frac{\beta_n}{\prod_{i=1}^k (1-v_i)^\sigma}} \\ & \times \int_0^\infty (1-(1+t)^{-\frac{1}{\sigma}})^{k\sigma-1}(1+t)^{k-1} e^{-\frac{\beta_n t}{\prod_{i=1}^k (1-v_i)^\sigma}} dt \,, \end{aligned} \tag{3.6.38}$$

where $\beta_n = \frac{a(u+\theta)^\sigma}{\sigma}$. We will follow the same idea as in the proof of Proposition 3.4 and the theorem 4.4 in (Hu and Zhang, 2022). To obtain the similar result as the Proposition 3.4 in (Hu and Zhang, 2022), we will consider the asymptotic result of the following quantity as $n \to \infty$.

$$\begin{aligned} & \mathbb{E}\left\{\sum_{k=1}^\infty w_k^p|\mathbf{X}\right\} = \mathbb{E}\left\{\mathbb{E}\left\{\sum_{k=1}^\infty w_k^p|U_n = u, \mathbf{X}\right\}\right\} = \mathbb{E}\left\{\mathbb{E}\left\{\sum_{k=1}^\infty v_k^p \prod_{l=1}^{k-1}(1-v_l)^p|U_n = u, \mathbf{X}\right\}\right\} \\ & = \int_0^\infty \frac{\beta_n^k \sigma^{k-1}}{[\Gamma(1-\sigma)]^k \Gamma(k\sigma)} \prod_{i=1}^k v_i^{-\sigma}(1-v_i)^{-(n-i)\sigma-1} e^{-\frac{\beta_n}{\prod_{i=1}^k (1-v_i)^\sigma}} \\ & \qquad \times \int_0^\infty (1-(1+t)^{-\frac{1}{\sigma}})^{k\sigma-1}(1+t)^{k-1} e^{-\frac{\beta_n t}{\prod_{i=1}^k (1-v_i)^\sigma}} dt f_{U_n}(u) du \,, \end{aligned} \tag{3.6.39}$$

where p is any positive integer. To evaluate 3.6.39 as $n \to \infty$, we shall have a further analysis of the integral with respect to u, which is the only term that relates to n. Consider the following integral for any $b > 0$, and any positive integer k.

$$\int_0^\infty \beta_n^k e^{b\beta_n} f_{U_n}(u)du = \frac{\left(\frac{a}{\sigma}\right)^k \int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-(n(\pi)+k)\sigma}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma} du}$$
$$= \frac{\left(\frac{a}{\sigma}\right)^k \left(\int_0^M \frac{u^{n-1}}{(u+\theta)^{n-(n(\pi)+k)\sigma}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du + \int_M^\infty \frac{u^{n-1}}{(u+\theta)^{n-(n(\pi)+k)\sigma}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du\right)}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma} du}, \tag{3.6.40}$$

for any $M > 0$. For any n and any M, we have

$$\begin{aligned}
&\int_0^M \frac{u^{n-1}}{(u+\theta)^{n-(n(\pi)+k)\sigma}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du \\
&= \int_0^M \left(\frac{u}{(u+\theta)^{1-\sigma}}\right)^{n-1} \frac{1}{(u+\theta)^{(n-n(\pi)-k-1)\sigma+1}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du \\
&\leq \left(\frac{M}{(M+\theta)^{1-\sigma}}\right)^{n-1} \frac{(M+\theta)^{(n(\pi)+k+1-n)\sigma}}{(n(\pi)+k+1-n)\sigma} e^{-\frac{(b+1)a}{\sigma}(\theta)^\sigma},
\end{aligned}$$

which goes to 0 as $n \to \infty$ due to the fact that either $\lim_{n\to\infty} \frac{n(\pi)}{n} = 0$ or $n(\pi) = n$. The last inequality holds because $\left(\frac{u}{(u+\theta)^{1-\sigma}}\right)^{n-1}$ is nondecreasing in u for any $\sigma \in [0, 1)$. Thus when $n \to \infty$, we have

$$\int_0^\infty \beta_n^k e^{b\beta_n} f_{U_n}(u)du = \frac{\left(\frac{a}{\sigma}\right)^k \int_M^\infty \frac{u^{n-1}}{(u+\theta)^{n-(n(\pi)+k)\sigma}} e^{-\frac{(b+1)a}{\sigma}(u+\theta)^\sigma} du}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma}} e^{-\frac{a}{\sigma}(u+\theta)^\sigma} du}.$$

This would imply

$$\begin{aligned}
\mathbb{E}\left\{\sum_{k=1}^\infty w_k^p | \mathbf{X}\right\} &= \int_{M_n}^\infty \frac{\beta_n^k \sigma^{k-1}}{[\Gamma(1-\sigma)]^k \Gamma(k\sigma)} \prod_{i=1}^k v_i^{-\sigma} (1-v_i)^{-(n-i)\sigma-1} e^{-\frac{\beta_n}{\prod_{i=1}^k (1-v_i)^\sigma}} \\
&\times \int_0^\infty (1-(1+t)^{-\frac{1}{\sigma}})^{k\sigma-1} (1+t)^{k-1} e^{-\frac{\beta_n t}{\prod_{i=1}^k (1-v_i)^\sigma}} dt f_{U_n}(u) du,
\end{aligned} \tag{3.6.41}$$

in which we choose $M = M_n$ that goes to ∞ as $n \to \infty$. In this case, when $n \to \infty$, $\beta_n \to \infty$ as well and we are safe to use the results in Proposition 3.4 in (Hu and Zhang,

2022) to obtain that when $n \to \infty$ (thus $n(\pi) \to \infty$)

$$\mathbb{E}\left[\sum_{n=1}^{\infty} w_n^2\right] = \int_{M_n}^{\infty} \left(\frac{1-\sigma}{a(u+\theta)^\sigma}\right) f_{U_n}(u)du = \frac{1-\sigma}{n(\pi)\sigma} + o\left(\frac{1}{n(\pi)}\right),$$

where the last equation can be computed by the same argument as in eq. (3.6.35) and the computation afterwards. The result of 3.6.37 follows immediately by applying the theorem 4.4 in (Hu and Zhang, 2022). □

By the above lemma and its proof, it is interesting to see that when $n \to \infty$, we can have $P_{n(\pi)} \overset{d}{=} P_{U_n}$, where $P_{n(\pi)} \sim \text{NGGP}(n(\pi), \sigma, \theta, H)$ for any $n(\pi) \to \infty$. Thus, we can replace P_{U_n} by $P_{n(\pi)}$ in the proof of theorem 3.4.1, the benefit of such replacement is $P_{n(\pi)}$ is independent of κ_n when $n \to \infty$.

The next lemma provides the convergence of κ_n.

Lemma 3.6.6. *(i) If P_0 is discrete, when $n \to \infty$,*

$$\sqrt{n}(\kappa_n - \frac{\sigma_n n(\pi)}{n}) \rightsquigarrow 0 \qquad a.s.$$

(ii) If P_0 is continuous, when $n \to \infty$, $\kappa_n \to \sigma$ in probability.

Proof. We shall compute the moments of $\kappa_n = \frac{T_{(U_n)}}{T_{(U_n)}+\sum_{j=1}^{n(\pi)} J_j}$ by the same method that we use in the proof of lemma 3.3.1. To make it clear, we present the details for $\mathbb{E}[\kappa_n]$ as follows.

$$\begin{aligned}
\mathbb{E}[\kappa_n] &= \mathbb{E}\left\{\mathbb{E}[\kappa_n|U_n]\right\} = \mathbb{E}\left\{\mathbb{E}\left\{\int_0^\infty e^{-(T_{(U_n)}+\sum_{j=1}^{n(\pi)} J_j)y} T_{(U_n)} dy|U_n\right\}\right\} \\
&= \mathbb{E}\left\{\int_0^\infty \left(-\frac{d}{dy}\mathbb{E}\left\{e^{-(T_{(U_n)}y|U_n}\right\}\right)\prod_{j=1}^{n(\pi)}\mathbb{E}\left\{e^{-yJ_j}|U_n\right\}dy\right\} \\
&= \mathbb{E}\left\{\int_0^\infty a(y+U_n+\theta)^{\sigma_n-1} e^{-\frac{a}{\sigma_n}((y+U_n+\theta)^{\sigma_n}-(U_n+\theta)^{\sigma_n})}\left(\frac{U_n+\theta}{y+U_n+\theta}\right)^{n-n(\pi)\sigma_n} dy\right\},
\end{aligned}$$

where the last equation is a direct use of the Laplace transform of T_{U_n} and the distribution of J_j in lemma 3.6.3. Solving the expectation with respect to U_n, we have

$$\mathbb{E}[\kappa_n] = \frac{\int_0^\infty \int_0^\infty a \frac{u^{n-1}}{(y+u+\theta)^{(n+1)-(n(\pi)+1)\sigma_n}} e^{-\frac{a}{\sigma_n}(y+u+\theta)^{\sigma_n}} dydu}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma_n}} e^{-\frac{a}{\sigma_n}(u+\theta)^{\sigma_n}} du} .$$

By the substitution $v = u + y$ and $z = u$ on the nominator of the above form.

$$\mathbb{E}[\kappa_n] = \frac{a}{n} \frac{\int_0^\infty \frac{v^n}{(v+\theta)^{(n+1)-(n(\pi)+1)\sigma_n}} e^{-\frac{a}{\sigma_n}(v+\theta)^{\sigma_n}} dv}{\int_0^\infty \frac{u^{n-1}}{(u+\theta)^{n-n(\pi)\sigma_n}} e^{-\frac{a}{\sigma_n}(u+\theta)^{\sigma_n}} du} ,$$

which implies $\mathbb{E}[\kappa_n] = \frac{n(\pi)\sigma_n}{n}$ by the analysis of 3.6.33. And we have $\mathbb{E}[\kappa_n] \to 0$ if P_0 is discrete and $\mathbb{E}[\kappa_n] \to \sigma$ if P_0 is continuous when $n \to \infty$.

Similarly, we can obtain the second moment of κ_n in the same way as $n \to \infty$.

$$\mathbb{E}[\kappa_n^2] = \frac{\sigma_n^2 n(\pi)^2}{n^2} + \frac{(n(\pi)+1)(1-\sigma_n)\sigma_n}{n(n+1)} ,$$

followed by which, we have $\mathrm{Var}[\kappa_n] = \frac{(n(\pi)+1)(1-\sigma_n)\sigma_n}{n(n+1)}$. And $\lim_{n\to\infty} \mathrm{Var}[\kappa_n] = 0$ for both continuous and discrete P_0. In particular, if P_0 is discrete, $\mathrm{Var}[\sqrt{n}\kappa_n] \to 0$ as well when $n \to \infty$. This complete the proof of the lemma. □

Now, theorem 3.4.1 can be proved by using the previous lemmas. And we give the details as follows.

Proof. Proof of theorem 3.4.1

We proof theorem 3.4.1 in two parts corresponding to when P_0 is discrete and when P_0 is continuous. We denote $R_n = \sum_{j=1}^{n(\pi)} \frac{J_j}{\sum_{j=1}^{n(\pi)}} \delta_{Y_j}$. Then, $R_n = \sum_{j=1}^{n(\pi)} D_{n,j}\delta_{Y_j}$.

(i) When P_0 is discrete.

It is convenient to decompose $\sqrt{n}\left(P - \mathbb{P}_n - \frac{\sigma_n n(\pi)}{n}(H - \tilde{\mathbb{P}}_n)\right)|\mathbf{X}$ as

$$\sqrt{n}\left(\kappa_n - \frac{\sigma_n n(\pi)}{n}\right)(P_{U_n} - R_n) + \sqrt{n}\left(\sqrt{\sigma_n n(\pi)}(P_{U_n} - H)\sqrt{\frac{\sigma_n n(\pi)}{n}}\right)$$

$$+\sqrt{n}\left(R_n\left(1-\frac{\sigma_n n(\pi)}{n}\right)-\left(\mathbb{P}_n-\frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n\right)\right)|\mathbf{X}\,. \tag{3.6.42}$$

The first term in decomposition 3.6.42 converges to 0 by using lemma 3.6.6 and the fact that $P_{U_n}-R_n$ is uniformly bounded. The second term in decomposition 3.6.42 converges to 0 by using lemma 3.6.5 and the fact that $\frac{\sigma_n n(\pi)}{n}\to 0$ a.s.. And the convergence for the first two terms in decomposition 3.6.42 holds for both $n(\pi)$ is finite and goes to ∞ when $n\to\infty$.

The convergence of the last term in decomposition 3.6.42 relaying on the gamma representation of $D_{n,j}$ in R_n. For each $j\in\{1,\cdots,n(\pi)\}$, we rewrite

$$D_{n,j}=\frac{\gamma_{j,0}+\sum_{i=1}^{n_j-1}\gamma_{j,i}}{\sum_{j=1}^{n(\pi)}\left(\gamma_{j,0}+\sum_{i=1}^{n_j-1}\gamma_{j,i}\right)}\,,$$

where the independent random variables $\gamma_{j,0}\sim G(1-\sigma_n,1)$ and $\gamma_{j,i}\sim G(1,1)$ for all j and all i. That is to say, there are n $\gamma_{j,i}$'s (i can take 0) for $j\in\{1,\cdots,n(\pi)\}$, during which, there are $n(\pi)$ independent $G(1-\sigma_n,1)$ random variables and $n-n(\pi)$ independent $G(1,1)$ random variables. Relabel all these n gamma random variables as $\{\tilde{\gamma}_{n,l}\}_{l=1}^{n}$ (the order doesn't matter). Then

$$R_n=\sum_{j=1}^{n(\pi)}D_{n,j}\delta_{Y_j}=\frac{n^{-1}\sum_{l=1}^{n}\tilde{\gamma}_{n,l}\delta_{X_l}}{n^{-1}\sum_{l=1}^{n}\tilde{\gamma}_{n,l}}\,. \tag{3.6.43}$$

To make the interpretation clear, we denote $R_nf=\frac{\bar{R}_nf}{\bar{R}_n1}$, where $\bar{R}_nf=\frac{\sum_{l=1}^{n}\tilde{\gamma}_{n,l}f(X_l)}{n}$ and $\bar{R}_n1=\frac{\sum_{l=1}^{n}\tilde{\gamma}_{n,l}}{n}$. Thus,

$$\begin{aligned}&\sqrt{n}\left(R_n\left(1-\frac{\sigma_n n(\pi)}{n}\right)-\left(\mathbb{P}_n-\frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n\right)\right)f\\&=-R_nf\sqrt{n}\left(\bar{R}_n1-\left(1-\frac{\sigma_n n(\pi)}{n}\right)\right)+\sqrt{n}\left(\bar{R}_nf-\left(\mathbb{P}_nf-\frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_nf\right)\right)\,.\end{aligned} \tag{3.6.44}$$

It is clear that $\mathbb{P}_nf+\frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_nf\to P_0f$ outer almost surely, by the Borel-Cantelli lemma and the fact that $\mathbb{F}$ is a finite set such that $P_0(f^2)<\infty$, and thus $\mathbb{F}$ is a P_0-Donsker class. By the distributions and independence of $\{\tilde{\gamma}_{n,l}\}_{l=1}^{n}$, we have

$$\mathbb{E}\left\{\bar{R}_n 1\right\} = \mathbb{E}\left\{\frac{\sum_{l=1}^{n}\tilde{\gamma}_{n,l}}{n}\right\} = 1 - \frac{\sigma_n n(\pi)}{n} \to 1 \qquad \text{a.s.}\,,$$
$$\mathrm{Var}\left\{\bar{R}_n 1\right\} = \mathrm{Var}\left\{\frac{\sum_{l=1}^{n}\tilde{\gamma}_{n,l}}{n}\right\} = \frac{1}{n}\sum_{l=1}^{n}\mathrm{Var}\left\{\tilde{\gamma}_{n,l}\right\} = \frac{1}{n} - \frac{\sigma_n n(\pi)}{n} \to 0 \qquad \text{a.s.}\,.$$

Thus, we have the convergence $\sqrt{n}\left(\bar{R}_n 1 - \left(1 - \frac{\sigma_n n(\pi)}{n}\right)\right) \rightsquigarrow 0$. By noting that R_n is uniformly bounded, we obtain $-R_n f\sqrt{n}\left(\bar{R}_n 1 - \left(1 - \frac{\sigma_n n(\pi)}{n}\right)\right) \rightsquigarrow 0$. To find the convergence of $\sqrt{n}\left(\bar{R}_n f - \left(\mathbb{P}_n f - \frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n f\right)\right)$, we follow the similar way and check the Linderberg-Feller condition as follows.

$$\mathbb{E}\left\{\bar{R}_n f\right\} = \frac{1}{n}\sum_{l=1}^{n}\mathbb{E}\left\{\tilde{\gamma}_{n,l}\right\} f(X_l) = \mathbb{P}_n f - \frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n f\,,$$
$$\mathrm{Var}\left\{\bar{R}_n f\right\} = \frac{1}{n^2}\sum_{l=1}^{n}\mathrm{Var}\left\{\tilde{\gamma}_{n,l}\right\} f^2(X_l) = \frac{1}{n}\left(\mathbb{P}_n f^2 - \frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n f^2\right)\,,$$
$$\frac{1}{n}\sum_{l=1}^{n}\mathbb{E}\left\{\tilde{\gamma}_{n,l}^2 f^2(X_l)\mathbb{1}_{|\tilde{\gamma}_{n,l} f(X_l)|>\epsilon\sqrt{n}}\right\}$$
$$\leq \max\left(\mathbb{E}\left\{\gamma_{j,0}^2 f^2(X_l)\mathbb{1}_{|\gamma_{j,0}|\max_{1\leq l\leq n}|f(X_l)|>\epsilon\sqrt{n}}\right\}, \mathbb{E}\left\{\gamma_{j,i}^2 f^2(X_l)\mathbb{1}_{|\gamma_{j,i}|\max_{1\leq l\leq n}|f(X_l)|>\epsilon\sqrt{n}}\right\}\right)\mathbb{P}_n f^2\,,$$

where the last inequality is a verification of Linderberg-Feller condition, and the right hand side converges to 0 for every sequence $\mathbf{X}$, since $P_0 f^2 < \infty$ and $\max_{1\leq l\leq n}|f(X_l)/\sqrt{n} \to 0$. By the Cramér-Wold device and the linearity of f, we have

$$\sqrt{n}\left(\bar{R}_n f - \left(\mathbb{P}_n f - \frac{\sigma_n n(\pi)}{n}\tilde{\mathbb{P}}_n f\right)\right) \rightsquigarrow \mathbb{B}_{P_0}^{o} f$$

for any $f \in \mathbb{F}$.

To show the convergence in $l^\infty(\mathbb{F})$ for any P_0-Donsker class, we shall prove the asymptotic tightness, see e.g., Theorem 1.5.4 in (van der Vaart and Wellner, 1996). The multipliers of the multiplier process $\frac{1}{\sqrt{n}}\sum_{l=1}^{n}(\tilde{\gamma}_{n,l} - \mathbb{E}\left\{\tilde{\gamma}_{n,l}\right\})f(X_l)$ are independent with 0 means. Thus, the multiplier central limit theorem in Theorem 2.9.7 (van der Vaart and Wellner, 1996) can be applied once we have the following inequality for any collection $\mathbb{H}$

of functions.

$$\mathbb{E}_{\tilde{\gamma}}\left\|\sum_{l=1}^{n}(\tilde{\gamma}_{n,l}-\mathbb{E}\left\{\tilde{\gamma}_{n,l}\right\})f(X_l)\right\|_{\mathbb{H}}^{*}\leq \mathbb{E}_{\tilde{\gamma},\tilde{\gamma}'}\left\|\sum_{l=1}^{n}(\tilde{\gamma}_{n,l}-\mathbb{E}\left\{\tilde{\gamma}_{n,l}\right\}+\tilde{\gamma}'_{n,l}-\mathbb{E}\left\{\tilde{\gamma}'_{n,l}\right\})f(X_l)\right\|_{\mathbb{H}}^{*},$$

by Jensen's inequality, for any random variable $\tilde{\gamma}'_{n,l}$ independent of $\tilde{\gamma}_{n,l}$. It is safe to choose all $\tilde{\gamma}'_{n,l} \overset{iid}{\sim} G(1,1)$ and $\tilde{\gamma}_{n,l} \overset{iid}{\sim} G(1,1)$. Then, the multiplier central limit theorem that is given as Theorem 2.9.7 (see also 2.9.6,2.9.9,3.6.13) in (van der Vaart and Wellner, 1996), the asymptotic tightness follows immediately. (We apply the inequality with $\mathbb{H}$ to be the set of $f_1 - f_2$ for any $f_1, f_2 \in \mathbb{F}$, with $L_2(P_0)$ norm of $f_1 - P_0 f_1 - (f_2 - P_0 f_2)$ smaller than δ.)

This complete the proof of the theorem when P_0 is discrete.

(ii) When P_0 is continuous.

In this case, $n(\pi) = n$. We can decompose $\sqrt{n}\left(P - \{(1-\sigma)\mathbb{P}_n + \sigma H\}\right)|\mathbf{X}$ as

$$\sqrt{n}\left(P_{U_n} - H\right)\kappa_n + \sqrt{n}(1-\kappa_n)(R_n - \mathbb{P}_n) + \sqrt{n}(\kappa_n - \sigma_n)(H - \mathbb{P}_n)\,. \tag{3.6.45}$$

For the convergence of the first term in 3.6.45, we first use the discussion below the proof of lemma 3.6.5 to use $P_n \overset{d}{=} P_{U_n}$ when $n \to \infty$, where $P_n \sim \mathrm{NGGP}(n, \sigma, \theta, H)$. Thus, we can consider the convergence of $\sqrt{n}\left(P_n - H\right)\kappa_n$ instead of $\sqrt{n}\left(P_{U_n} - H\right)\kappa_n$, the benefit of the former form is P_n and κ_n are independent. Thus, by lemma 3.6.6, $\kappa_n \to \sigma$ in probability. By using the result in lemma 3.6.5, we have $\sqrt{n}\left(P_n - H\right) \rightsquigarrow \sqrt{\frac{1-\sigma}{\sigma}}\mathbb{B}_H^o$ a.s.. Thus, we have $\sqrt{n}\left(P_{U_n} - H\right)\kappa_n \rightsquigarrow \sqrt{\sigma(1-\sigma)}$.

For the second term in 3.6.45, $R_n = \sum_{j=1}^{n} D_{n,j}\delta_{X_j}$, where $D_{n,j} = \frac{\gamma_j}{\sum_{j=1}^{n}\gamma_j}$ with $\gamma \overset{iid}{\sim} G(1-\sigma_n, 1)$. A direct application of the result of Theorem 2.1 in (Præstgaard and Wellner, 1993) implies $\sqrt{n}(R_n - \mathbb{P}_n) \rightsquigarrow \frac{1}{\sqrt{1-\sigma}}\mathbb{B}_{P_0}^o$ a.s., in $l^\infty(\mathbb{F})$ if there is a P_0-square-integrable envelope function for $\mathbb{F}$. Furthermore, the convergence in probability is a direct application of Theorem 2.9.7 in (van der Vaart and Wellner, 1996). By noting the fact that $(1-\kappa_n) \to 1-\sigma$ in probability, we have $\sqrt{n}(1-\kappa_n)(R_n - \mathbb{P}_n) \rightsquigarrow \sqrt{1-\sigma}\mathbb{B}_{P_0}^o$ a.s. in $l^\infty(\mathbb{F})$.

For the last term in 3.6.45, we follow the same argument as that in lemma 3.6.6 and will have $\text{Var}[\sqrt{n}\kappa_n] = (1-\sigma_n)\sigma_n \to (1-\sigma)\sigma$, thus $\sqrt{n}(\kappa_n - \sigma_n) \rightsquigarrow \sqrt{\sigma(1-\sigma)}Z$. Furthermore, by the Borel-Cantelli lemma, $\mathbb{P}_n \to P_0$ a.s., and thus $\sqrt{n}(\kappa_n - \sigma_n)(H - \mathbb{P}_n) \rightsquigarrow \sqrt{\sigma(1-\sigma)}Z(H - P_0)$.

The result in theorem 3.4.1 when P_0 is continuous follows by combining the convergences of the three terms in 3.6.45. □

Chapter 4

Normalized random measures with independent increments driven by Cox process

4.1 Introduction

The random partition structure induced by Bayesian nonparametric priors has been shown very useful in the statistical inference problems related to clustering, density estimation, and prediction. Various works have been devoted to study the probabilistic theory (Kingman, 1982; Pitman, 1996; Gnedin and Pitman, 2006; Pitman, 2006) and statistical aspects (De Blasi et al., 2013; James, 2005; James et al., 2009) of the random partition structure of exchangeable observations (e.g. (Kallenberg, 2005) and the references therein for more details). Based on the celebrating work in (MacEachern, 1999, 2000), dependent Bayesian nonparametric models have gained particular attention due to their flexibilities and the nonexchangeable assumption (Camerlenghi et al., 2019; Quintana et al., 2022). One type of the well-studied dependent Bayesian nonparametric models is the hierarchical Bayesian nonparametric models (see e.g., (Teh et al., 2006; Teh and Jordan, 2010; Gasthaus and Teh, 2010; NGUYEN, 2016; Zhang and Hu, 2021; Camerlenghi et al., 2019)). In hierarchical models, the analytic forms of the random partition structure are complex due to the hierarchical levels. To the best of our knowledge, the distribution theory and posterior characterizations are known only for the hierarchical Dirichlet process (Teh et al., 2006), the hierarchical Pitman-Yor process (Camerlenghi et al., 2019) and the hierarchical normalized random measures with independent increments (NRMIs) (Camerlenghi et al., 2019). However, hierarchical structures are "auto" dependent structure, where the dependence are controlled by the concentration parameters of each hierarchy. Since the concentration parameters are finite and represent the faithful of the prior, we prefer to give another "tuning" parameter, which can be used to control the dependence by users.

In this Chapter, we present two flexible constructions of hierarchical NRMIs. We will construct a vector of dependent random probability measures through two suitable transformations in the construction of NRMIs. On the one hand, the new constructions allow the observations to be partially exchangeable. On the other hand, each component of the dependent vector of random probability measures is assigned a tuning parameter as the control of component-wise dependence by users. Multiple distributional quantities include the moments, distribution of the induced random partition structures, distribution of the number of partition numbers are obtained. Furthermore, we allow each component of the vector of dependent random probability measure itself to follow different distributions.

1.1 Partial exchangeability As a more general framework, partial exchangeability (see e.g., (Teh et al., 2004; Camerlenghi et al., 2019) and the references therein) extends the definition of exchangeability in a natural way. A random sequence is partially exchangeable if its distribution is invariant under all finite permutations within subgroups. Partial exchangeability is also a natural behaviour in many scenarios, when a population is decomposed into multiple sub-populations, each of which is exchangeable, and thus the population is overall partially exchangeable. Formally, let $(\mathbb{X}, \mathcal{X})$ be a Polish space and $\mathbb{M}_{\mathbb{X}}$ be the space of all probability measures on $\mathbb{X}$ with the corresponding Borel $\sigma-$algebra $\mathcal{M}_{\mathbb{X}}$. Consider a sequence of partially exchangeable random variables $\mathbf{X} = \{\mathbf{X}^{(N_i)} = (X_{i,k}) : k = 1, \cdots, N_i; i = 1, \cdots, d\}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that are taking values in $(\mathbb{X}, \mathcal{X})$. That is to say $\mathbf{X}$ contains d groups of exchangeable groups $\mathbf{X}^{(N_i)}$ of random variables. By the celebrating de Finetti's representation theorem, for any $A_1, \cdots, A_d \in \mathcal{X}$, we have

$$\mathbb{P}(\mathbf{X}^{(N_1)} \in A_1, \cdots, \mathbf{X}^{(N_d)} \in A_d) = \int_{\mathbb{M}_{\mathbb{X}}^d} \prod_{i=1}^d P_i^{N_i}(A_i) Q_d(dP_1, \cdots, dP_d), \tag{4.1.1}$$

where $P_i^{N_i} = \times_{k=1}^{N_i} P_i$ represents the N_i-ford product measure on $\mathbb{X}^{N_i}$ for each $i \in \{1, \cdots, d\}$, Q_d is a probability measure on $(\mathbb{M}_{\mathbb{X}}^d, \mathcal{M}_{\mathbb{X}}^d)$ serving as the prior distribution of $(P_1, \cdots, P_d)$. Following (Ferguson, 1973), we call $\mathbf{X}$ a sample of size $N = \sum_{i=1}^d N_i$ from $(P_1, \cdots, P_d)$, i.e., the sample $\mathbf{X}^{(N_i)} \overset{i.i.d}{\sim} P_i$ conditionally on P_i.

In the literature of Bayesian nonparametric framework, Q_d is constructed so that the random probability measures $(P_1, \cdots, P_d)$ are discrete with probability 1. Therefore, there are positive probabilities that the random variables $X_{i,k} = X_{j,l}$, that is to say, the observations within group or between groups could be equal each other. Such feature induces a natural way to cluster the observations $\mathbf{X}$, and thus induces a random partition structure of $(1, \cdots, N)$ under Q_d. The random partition structure is identified by the exchangeable partition probability function (EPPF) in the exchangeable setting (e.g. (Pitman, 1996, 2003, 2006)). The EPPF provides a probability function of random partition of $(1, \cdots, N)$ and thus is very important in the study of clustering, sampling schemes, prediction rules. In the partially exchangeable setting, one can identify the random partition structure by using partially exchangeable partition probability function (pEPPF) as follows. Let K be the number of distinct observations of $\mathbf{X}$, and $(X_1^*, \cdots, X_K^*)$ be the K distinct observations. For each $i \in \{1, \cdots, d\}$, let $\mathbf{n}_i = (n_{i,1}, \cdots, n_{i,K})$ be the vector of frequency of $(X_1^*, \cdots, X_K^*)$ appears in the ith group of observations accordingly, namely, $n_{i,k} = \#\{X_{i,j} : X_{i,j} = X_k^*; j = 1, \cdots, N_i\}$. Thus, $n_{i,j}$ could be 0, $\sum_{i=1}^d n_{i,j} \geq 1$ for any $j \in \{1, \cdots, K\}$ and $\sum_{j=1}^K n_{i,j} = N_i$ for each $i \in \{1, \cdots, d\}$. The pEPPF is defined as

$$\Pi_K^{(N)}(\mathbf{n}_1, \cdots, \mathbf{n}_d) = \mathbb{E}\left[\int_{\mathbb{X}^K} \prod_{j=1}^K \prod_{i=1}^d P_i^{n_{i,j}}(dx_j^*)\right]. \tag{4.1.2}$$

1.2 Outline The outline of this chapter is as follows. In Section 4.2, we recall the construction of NRMIs through Poisson random measures. In Section 4.3, we introduce the two constructions of dependent vector of random probability measures based on the construction of NRMIs. In Section 4.4, we obtain the moments of the two constructions that can be a direct view of the dependence structures of our models. In Section 4.5, we derive the probability distribution of the random partition induced by the two models. In Section 4.6, we discuss the distribution of partition number. In order to ease the flow of the ideas, we delay the proofs in the Appendix 4.8.

4.2 Normalized random measures with independent increments

4.2.1 Constructions of NRMIs

We start by recalling the notions of completely random measures (see e.g., (Kingman, 1967, 1993) and references therein for more details), which play important role in the construction of NRMIs.

Let $\mathbb{B}_{\mathbb{X}}$ be the space of bounded finite measures on $(\mathbb{X}, \mathcal{X})$ endowed with a suitable topology so that the associated Borel $\sigma-$algebra $\mathcal{B}_{\mathbb{X}}$ can be introduced (Daley and Vere-Jones, 2008).

Definition 4.2.1. *Let μ be a measurable function on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{B}_{\mathbb{X}}, \mathcal{B}_{\mathbb{X}})$. We call μ is a completely random measure (CRM) if the random variables $\mu(A_1), \cdots, \mu(A_d)$ are mutually independent, for any pairwise disjoint sets $A_1, \cdots, A_d$ in $\mathcal{X}$, where $d \geq 2$ is a finite integer.*

We refer to (Regazzini et al., 2003; Lijoi et al., 2010; James et al., 2009; Camerlenghi et al., 2019) for more detailed discussions of constructing NRMIs by using completely random measures.

In this work, we construct NRMIs driven by Cox random measures (Cox, 1955) in the similar way as the construction of NRMIs. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathbb{X}$ be a complete, separable metric space whose Borel σ-algebra is denoted by $\mathcal{X}$.

Denote $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$ and denote its Borel σ-algebra by $\mathcal{S}$. Let $\xi(dx, ds)$ be a random measure defined on $(\mathbb{S}, \mathcal{S})$. A Cox random measure $\tilde{N}$ on $\mathbb{S}$ is a random measure from $\Omega \times \mathbb{S}$ to $\mathbb{R}_+$ such that $\tilde{N}$ is a Poisson random measure with intensity ν conditioning on $\xi(dx, ds) = \nu(ds, dx)$. That is to say,

(i) $\tilde{N}(A)|\xi = \nu \sim \text{Poisson}(\nu(A))$ for any A in $\mathcal{S}$;

(ii) for any pairwise disjoint sets $A_1, \cdots, A_m$ in $\mathcal{S}$, the random variables $\tilde{N}(A_1), \cdots, \tilde{N}(A_m)$ are mutually independent conditionally on ξ.

The random intensity measure ξ satisfies the condition (see (Daley and Vere-Jones, 2008) for details of Poisson random measures) that

$$\int_0^\infty \int_{\mathbb{X}} \min(s, 1)\xi(ds, dx) < \infty\,,$$

almost surely. Let $\tilde{\mu}$ be the random measure defined on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes values in $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ defined as follows,

$$\tilde{\mu}(A) := \int_0^\infty \int_A s\tilde{N}(ds, dx), \quad \forall A \in \mathcal{X}\,. \tag{4.2.1}$$

It is trivial to verify that $\tilde{\mu}$ is a completely random measure conditioning on ξ. It is also well-known that for any $B \in \mathcal{X}$, $\tilde{\mu}(B)$ is discrete and is uniquely characterized by its Laplace transform as follows:

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(B)}|\xi = \nu\right] = \exp\left\{-\int_0^\infty \int_B \left[1 - e^{-\lambda s}\right]\nu(ds, dx)\right\}. \tag{4.2.2}$$

The measure ν is called the *Lévy intensity* of $\tilde{\mu}$ and we denote the Laplace exponent by

$$\psi_B(\lambda) = \int_0^\infty \int_B \left[1 - e^{-\lambda s}\right]\nu(ds, dx)\,. \tag{4.2.3}$$

From the Laplace transform in eq. (4.2.2), we aware that the completely random measure $\tilde{\mu}$ is characterized completely by its Lévy intensity ν conditional on $\xi = \nu$. To make the interpretation easy to read, the measure ν that ξ can take is usually represented in the following two cases.

(a) $\nu(ds, dx) = \rho(ds)\alpha(dx)$, where $\rho : \mathcal{B}(\mathbb{R}^+) \to \mathbb{R}^+$ is some measure on $\mathbb{R}^+$ and α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ so that $\alpha(\mathbb{X}) = a < \infty$. Conditional on $\xi = \nu$, the corresponding $\tilde{\mu}$ is called *homogeneous* completely random measure.

(b) $\nu(ds, dx) = \rho(ds|x)\alpha(dx)$, where ρ is defined on $\mathcal{B}(\mathbb{R}^+) \times \mathbb{X}$ such that for any $x \in \mathbb{X}$, $\rho(\cdot|x)$ is a σ-finite measure on $\mathcal{B}(\mathbb{R}^+)$ and for any $A \in \mathcal{X}$, $\rho(A|x)$ is $\mathcal{B}(\mathbb{R}^+)$ measurable. Conditional on $\xi = \nu$, the corresponding $\tilde{\mu}$ is called *non-homogeneous* completely random measure.

It is obvious that the case (a) is a special case of case (b).

To construct NRMIs driven by Cox process, the completely random measure will be normalized, and thus one needs the total mass $\tilde{\mu}(\mathbb{X})$ to be finite and positive almost surely. This happens under the condition that $\rho(\mathbb{R}^+) = \infty$ in homogeneous case and that $\rho(\mathbb{R}^+|x) = \infty$ in non-homogeneous case (See e.g., (Regazzini et al., 2002) for a proof). Under the above conditions, we call the random probability measure P on $(\mathbb{X}, \mathcal{X})$ an NRMI driven by Cox process, denoted as $P \sim \text{CoxNRMI}(\xi)$, if P is defined by

$$P(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}. \tag{4.2.4}$$

P is discrete due to the discreteness of $\tilde{\mu}$.

In this work, we only focus on the homogeneous case and denote the homogeneous NRMIs driven by Cox process by $P \sim \text{hCoxNRMI}(\alpha, \rho)$. Usually, the intensity ρ is used to identify the distributional structure of P. For example, the Dirichlet process is a CoxNRMI with $\xi = \nu$ with probability 1 and $\nu(ds, dx) = \alpha(dx)\frac{e^{-s}}{s}ds$; the $\sigma-$stable process is a CoxNRMI with $\xi = \nu$ with probability 1 and $\nu(ds, dx) = \alpha(dx)\frac{\sigma}{\Gamma(1-\sigma)s^{1+\sigma}}ds$; the normalized generalized gamma process is a CoxNRMI with $\xi = \nu$ with probability 1 and $\nu(ds, dx) = \alpha(dx)\frac{1}{\Gamma(1-\sigma)}s^{-1-\sigma}e^{-\theta s}ds$, where the parameters $\sigma \in (0, 1)$ and $\theta > 0$. To keep the distributional structure of P, we shall assign the randomness of ν only on

α. That is to say, $\xi = \alpha(dx)\rho(ds)$, where $\rho(ds)$ is a given fixed intensity and $\alpha(dx)$ is a random measure on $(\mathbb{X}, \mathcal{X})$.

It is worthy to point out that the hierarchical NRMIs discussed in (Camerlenghi et al., 2019) are homogeneous CoxNRMIs in the case that $\alpha = aH(dx)$, where a is a positive finite constant number and $H(dx)$ is an NRMI.

4.3 Models: NRMIs driven by Cox process

With the construction in Section 4.2, we present two models of dependent random probability measures $(P_1, \cdots, P_d)$. The first model follows the idea similar to the regular hierarchical NRMIs, and the dependence between P_i and $(P_j)_{j\neq i}$ are controlled by a tuning parameter. The second model involves two dependent relationships, one is between P_i and P_j, another is due to the hierarchical structure across the vector. We use the notation $H_0 \sim \text{NRMI}(aH, \rho)$ to denote the random probability measure H_0, which is an NRMI with Lévy intensity $aH(dx)\rho(ds)$.

4.3.1 Conditionally independent hCoxNRMIs

Model 4.3.1. *Let $\{z_i\}_{i=1}^d$ be a sequence of random variables that take values in $[0, 1]$. For any $i \in \{1, \cdots, d\}$, let $\tilde{\mu}_i$ be a CRM with intensity $\nu_i(dx, ds) = z_i a_i H_i(dx)\rho_i(ds)$ and let $\tilde{\mu}_{i,0}$ be a CRM with intensity $\nu_{i,0}(dx, ds) = (1 - z_i)a_0 H_0(dx)\rho_i(ds)$, where $H_0 \sim$ NRMI(aH, ρ), for some nonatomic probability measure H, and a, a_0, $\{a_i\}_{i=1}^d$ are positive numbers. For each $i \in \{1, \cdots, d\}$, define*

$$\mu_i = \tilde{\mu}_i + \tilde{\mu}_{i,0} \,. \tag{4.3.1}$$

Then the normalization of μ_i

$$P_i = \frac{\mu_i}{\mu_i(\mathbb{X})} \,, \tag{4.3.2}$$

is a sequence of dependent random probability measures. We call such sequences of random probability measures ***conditionally independent hCoxNRMIs****.*

A trivial result from the above construction 4.3.1 is that μ_i are conditionally independent with Laplace transform

$$\mathbb{E}\left[e^{-\sum_{i=1}^{d}\lambda_i\mu_i(A)}\right] \\ = \exp\left\{-\sum_{i=1}^{d}\left(z_i a_i H_i(A) + (1-z_i)a_0 H_0(A)\right)\phi_i(\lambda_i)\right\}, \tag{4.3.3}$$

conditional on H_0 and $\{z_i\}_{i=1}^d$, for any $A \in \mathcal{X}$.

Remark 4.3.2. *The random sequence $\{z_i\}_{i=1}^d$ is the control of the dependence. When $z_i = 0$, model 4.3.1 reduces to the general hierarchical NRMIs as introduced in Camerlenghi et al. (2019), and there is an "auto" dependence structure only due to the hierarchical structure. When $z_i = 1$, the hierarchical structure in model 4.3.1 is gone, and P_i's are totally independent, but may not be identical, since ρ_i's may not be the same. Thus, z_i is a tuning parameter that control how "heavy" the hierarchical dependence is on P_i. A smaller z_i implies a "heavier" dependence of P_i induced by the hierarchical dependence.*

We can trivially extend the model 4.3.1 as follows.

Remark 4.3.3. *Let $D_{q\times q}$ be a $q \times q$ matrix with all entries taking values in $\{0,1\}$. For each $i \in \{1,\cdots,d\}$, let $\{z_{i,j}\}_{j=1}^q$ be a q-dimensional standard simplex sequence, i.e. $(z_{i,1},\cdots,z_{i,q}) \in \mathbb{S}_q := \{x \in \mathbb{R}^q : \sum_{i=1}^q x_i = 1, x_i \geq 0 \text{ for } i = 1,\cdots,q\}$. Let q_1, q_2 be two integers such that $q_1 + q_2 = q$. Let $\tilde{\mu}_{i,j}$ and $\tilde{\mu}_{i,0,l}$ be defined similarly to $\tilde{\mu}_i$ and $\tilde{\mu}_{i,0}$ in model 4.3.1 with the random weight $z_{i,j}$ for $j \in \{1,\cdots,q_1\}$ and $z_{i,l}$ for $l \in \{q_1+1,\cdots,q\}$ correspondingly.*

Define

$$\mu = D_{q\times q}(\tilde{\mu}_{i,1},\cdots,\tilde{\mu}_{i,q_1},\tilde{\mu}_{i,0,q_1+1},\cdots,\tilde{\mu}_{i,0,q})^T$$

as the random vector of CRM $(\mu_1, \cdots, \mu_q)^T$*. And* $P_i = \frac{\mu_i}{\mu_i(\mathbb{X})}$ *presents the general case of model 4.3.1.*

We will continue to consider the following example of model 4.3.1 in the next sections.

Example 4.3.4. *If* $\rho_i(ds) = \frac{e^{-s}}{s} ds$ *for all* $i \in \{1, \cdots, d\}$ *and* $\rho(ds) = \frac{\sigma}{\Gamma(1-\sigma)s^{1+\sigma}} ds$*, for some* $\sigma \in (0, 1)$*, then* P_i*'s are independent Dirichlet processes conditionally on* H_0*, and* H_0 *is a* σ*-stable NRMIs.*

4.3.2 Conditionally dependent hCoxNRMIs

Model 4.3.5. *Let* $\{z_i\}_{i=1}^d$ *be a sequence of random variables that take values in* $[0, 1]$ *and* a*,* a_0*,* $\{a_i\}_{i=1}^d$ *are positive numbers. For any* $i \in \{1, \cdots, d\}$*, let* $\tilde{\mu}_i$ *be a CRM with intensity* $\nu_i(dx, ds) = z_i a_i H_i(dx)\rho_i(ds)$*. Let* $\tilde{\mu}_0$ *be a CRM with intensity* $\nu_0(dx, ds) = z_0 a_0 H_0(dx)\rho_0(ds)$*, where* $H_0 \sim$ *NRMI*(aH, ρ)*, for some nonatomic probability measure* H *and* $z_0 = d - \sum_{i=1}^d z_i$*. For each* $i \in \{1, \cdots, d\}$*, define*

$$\mu_i = \tilde{\mu}_i + \tilde{\mu}_0 \,. \tag{4.3.4}$$

Then the normalization of μ_i

$$P_i = \frac{\mu_i}{\mu_i(\mathbb{X})} \,, \tag{4.3.5}$$

is a sequence of dependent random probability measures. We call such sequences of random probability measures ***conditionally dependent hCoxNRMIs****.*

The dependence of P_i is given by the dependence structure of μ_i. It is trivial to see the joint Laplace transform of μ_i as follows.

$$\begin{aligned} &\mathbb{E}\left[e^{-\sum_{i=1}^d \lambda_i \mu_i(A)}\right] \\ &\quad = \exp\left\{-\sum_{i=1}^d z_i a_i H_i(A)\phi_i(\lambda_i) - z_0 a_0 H_0(A)\phi_0\left(\sum_{i=1}^d \lambda_i\right)\right\} \,, \end{aligned} \tag{4.3.6}$$

conditional on H_0 and $\{z_i\}_{i=1}^d$, for any $A \in \mathcal{X}$.

Remark 4.3.6. *The construction in model 4.3.5 presents a complicate dependent structure. When $z_i = 0$ for all i, P_i's are identical to each other, which shows a completely dependence of P_i's. When $z_i = 1$ for all i, P_i's are independent. Thus, the independence of P_i is no longer controlled only by z_i but $\sum_{i=1}^d z_i$. Furthermore, when H_0 is deterministic, $(P_1, \cdots, P_d)$ is still a vector of dependent probability measures due to the common component induced by $\tilde{\mu}_0$.*

The model 4.3.5 can be extended by the similar manner as in remark 4.3.3. We will consider the following example in the next sections.

Example 4.3.7. *We can take $\rho_i(ds) = \frac{e^{-s}}{s}ds$ for all $i \in \{1, \cdots, d\}$, $\rho_0(ds) = \frac{e^{-s}}{s}ds$ and $\rho(ds) = \frac{\sigma}{\Gamma(1-\sigma)s^{1+\sigma}}ds$, for some $\sigma \in (0, 1)$.*

The two models we introduced in model 4.3.1 and model 4.3.5 are both constructed under the hierarchical Bayesian nonparametric framework. However, the linear random intensity presents a more flexible dependent structure than general hierarchical Bayesian model.

Remark 4.3.8. *The distribution of H_0 in model 4.3.1 and model 4.3.5 is not necessary an NRMI. A flexible form is $H_0 \sim \mathbb{Q}_H$, where $\mathbb{Q}_H$ represents the distribution of random probability measure H_0 with mean measure H. One example is the two-parameter Poisson-Dirichlet process.*

4.4 Dependent results: Moments

In this section, we will obtain the moment results of the dependent hCoxNRMIs introduced in model 4.3.1 and model 4.3.5. The variance and covariance results would present a clear view of the dependent structures.

To make the notations simple, we introduce the following symbols. For any positive integer m, k and $i \in \{0, 1, \cdots, d\}$, let

$$\begin{aligned}
\tau_{i,m}(u) &= \int_0^\infty s^m e^{-us} \rho_i(ds), \qquad \phi_i(u) = \int_0^\infty (1 - e^{-us}) \rho_i(ds)\,, \\
\mathcal{I}_{i,m}^{(k)} &= \int_0^\infty u e^{-a_i \phi_i(u)} \tau_{i,m}^k(u) du\,, \\
\tau_m(u) &= \int_0^\infty s^m e^{-us} \rho(ds), \qquad \phi(u) = \int_0^\infty (1 - e^{-us}) \rho(ds)\,, \\
\mathcal{I}_m^{(k)} &= \int_0^\infty u e^{-a\phi(u)} \tau_m^k(u) du\,.
\end{aligned}$$

4.4.1 Moment results of model 4.3.1

Proposition 4.4.1. *Let $\{P_i\}_{i=1}^d$ be the CoxNRMIs on $(\mathbb{X}, \mathcal{X})$ defined as in model 4.3.1. For any $A, B \in \mathcal{X}$, we have*

$$\begin{aligned}
&\mathbb{E}[P_i(A)] = \frac{z_i a_i H_i(A) + (1 - z_i) a_0 H(A)}{z_i a_i + (1 - z_i) a_0}\,, \\
&\mathrm{Var}[P_i(A)] = \frac{(1 - z_i)^2 a_0^2 a H(A) H(A^c) \mathcal{G}_2}{(z_i a_i + (1 - z_i) a_0)^2} + \\
&\frac{[(z_i a_i H_i(A) + (1 - z_i) a_0 H(A)) (z_i a_i H_i(A^c) + (1 - z_i) a_0 H(A^c)) - (1 - z_i)^2 a_0^2 a H(A) H(A^c) \mathcal{G}_2]}{z_i a_i + (1 - z_i) a_0} \mathcal{G}_{i,2}\,, \\
&\mathrm{Cov}[P_i(A), P_j(B)] = \frac{(1 - z_i)(1 - z_j) a_0^2 a [H(A \cap B) - H(A) H(B)] \mathcal{G}_2}{(z_i a_i + (1 - z_i) a_0) (z_j a_j + (1 - z_j) a_0)}\,,
\end{aligned} \tag{4.4.1}$$

where

$$\begin{aligned}
\mathcal{G}_2 &= \int_0^\infty u \tau_2(u) e^{-a\phi(u)} du\,, \\
\mathcal{G}_{i,2} &= \int_0^\infty u \tau_{i,2}(u) e^{-(z_i a_i + (1 - z_i) a_0) \phi_i(u)} du\,.
\end{aligned}$$

The moment results of example 4.3.4 can be calculated by using $\mathcal{G}_2 = \frac{1-\sigma}{a}$ and $\mathcal{G}_{i,2} = \frac{1}{(z_i a_i + (1 - z_i) a_0)(z_i a_i + (1 - z_i) a_0 + 1)}$. And thus, we have

$$\begin{aligned}
\mathbb{E}[P_i(A)] &= \frac{z_i a_i H_i(A) + (1 - z_i) a_0 H(A)}{z_i a_i + (1 - z_i) a_0}\,, \\
\mathrm{Var}[P_i(A)] &= \frac{(1 - \sigma)(1 - z_i)^2 a_0^2 H(A) H(A^c)}{(z_i a_i + (1 - z_i) a_0) (z_i a_i + (1 - z_i) a_0 + 1)} +
\end{aligned}$$

$$\frac{\left[(z_ia_iH_i(A)+(1-z_i)a_0H(A))\left(z_ia_iH_i(A^c)+(1-z_i)a_0H(A^c)\right)\right]}{(z_ia_i+(1-z_i)a_0)^2\left(z_ia_i+(1-z_i)a_0+1\right)},$$

$$\text{Cov}[P_i(A),P_j(B)]=\frac{(1-\sigma)(1-z_i)(1-z_j)a_0^2[H(A\cap B)-H(A)H(B)]}{(z_ia_i+(1-z_i)a_0)\left(z_ja_j+(1-z_j)a_0\right)}. \tag{4.4.2}$$

Interestingly, the variance and the mutual covariance of the dependent random probability measures $\{P_i\}_{i=1}^d$ are not affected by the concentration parameter a of H_0.

4.4.2 Moment results for model 4.3.5

Proposition 4.4.2. *Let $\{P_i\}_{i=1}^d$ be the CoxNRMIs on $(\mathbb{X},\mathcal{X})$ defined as in model 4.3.5. For any $A,B\in\mathcal{X}$, we have*

$$\begin{aligned}\mathbb{E}[P_i(A)]&=H(A)z_0a_0\mathcal{I}_{i,1}^{(0,1)}+H_i(A)z_ia_i\mathcal{L}_{i,1}^{(0,1)}\\&=H_i(A)+(H(A)-H_i(A))z_0a_0\mathcal{I}_{i,1}^{(0,1)}=H(A)+(H_i(A)-H(A))z_ia_i\mathcal{I}_{j,1}^{(0,1)},\end{aligned}\tag{4.4.3}$$

$$\begin{aligned}\text{Var}[P_i(A)]&=H_i(A)^2z_i^2a_i^2\left(\mathcal{L}_{i,1}^{(1,2)}-\left(\mathcal{L}_{i,1}^{(0,1)}\right)^2\right)+H(A)^2z_0^2a_0^2\left(\mathcal{L}_{0,1}^{(1,2)}-\left(\mathcal{I}_{i,1}^{(0,1)}\right)^2\right)\\&\quad+2H_i(A)H(A)z_iz_0a_ia_0\left(\mathcal{J}_{i,0}-\mathcal{L}_{i,1}^{(0,1)}\mathcal{I}_{i,1}^{(0,1)}\right)+H_i(A)z_ia_i\mathcal{L}_{i,2}^{(1,1)}\\&\quad+H(A)z_0a_0\mathcal{I}_{i,2}^{(1,1)}+az_0^2a_0^2H(A)H(A^c)\mathcal{I}_{i,1}^{(1,2)}\mathcal{G}_2,\end{aligned}\tag{4.4.4}$$

$$\begin{aligned}&\text{Cov}[P_i(A),P_j(B)]=z_iz_ja_ia_jH_i(A)H_j(B)\left(\mathcal{K}_{i,j}-\mathcal{L}_{i,1}^{(0,1)}\mathcal{L}_{j,1}^{(0,1)}\right)\\&+z_iz_0a_ia_0H_i(A)H(B)\left(\mathcal{K}_{i,0}-\mathcal{L}_{i,1}^{(0,1)}\mathcal{I}_{j,1}^{(0,1)}\right)+z_jz_0a_ja_0H(A)H_j(B)\left(\mathcal{K}_{j,0}-\mathcal{L}_{j,1}^{(0,1)}\mathcal{I}_{i,1}^{(0,1)}\right)\\&+z_0^2a_0^2H(A)H(B)\left(\mathcal{K}_{0,0}-\mathcal{I}_{i,1}^{(0,1)}\mathcal{I}_{j,1}^{(0,1)}\right)+z_0^2a_0^2a(H(A\cap B)-H(A)H(B))\mathcal{G}_2\mathcal{K}_{0,0}\\&+z_0a_0H(A\cap B)\mathcal{H}_{0,2},\end{aligned}\tag{4.4.5}$$

where for any $i,j\in\{1,\cdots,d\}$, $k,n\in\mathbb{Z}$,

$$\mathcal{I}_{i,m}^{(n,k)}=\int_0^\infty u^n\tau_{0,m}^k(u)e^{-z_0a_0\phi_0(u)-z_ia_i\phi_i(u)}du,$$

$$\mathcal{L}_{i,m}^{(n,k)}=\int_0^\infty u^n\tau_{i,m}^k(u)e^{-z_0a_0\phi_0(u)-z_ia_i\phi_i(u)}du,$$

$$\mathcal{J}_{i,0}=\int_0^\infty u\tau_{i,1}(u)\tau_{0,1}(u)e^{-z_0a_0\phi_0(u)-z_ia_i\phi_i(u)}du,$$

$$\mathcal{K}_{i,j} = \int_0^\infty \int_0^\infty \tau_{i,1}(u_1)\tau_{j,1}(u_2)e^{-z_0a_0\phi_0(u_1+u_2)-z_ia_i\phi_i(u_1)-z_ja_j\phi_j(u_2)}du_1du_2$$
$$\mathcal{K}_{i,0} = \int_0^\infty \int_0^\infty \tau_{i,1}(u_1)\tau_{0,1}(u_1+u_2)e^{-z_0a_0\phi_0(u_1+u_2)-z_ia_i\phi_i(u_1)-z_ja_j\phi_j(u_2)}du_1du_2$$
$$\mathcal{K}_{j,0} = \int_0^\infty \int_0^\infty \tau_{j,1}(u_2)\tau_{0,1}(u_1+u_2)e^{-z_0a_0\phi_0(u_1+u_2)-z_ia_i\phi_i(u_1)-z_ja_j\phi_j(u_2)}du_1du_2$$
$$\mathcal{K}_{0,0} = \int_0^\infty \int_0^\infty \tau_{0,1}(u_1+u_2)\tau_{0,1}(u_1+u_2)e^{-z_0a_0\phi_0(u_1+u_2)-z_ia_i\phi_i(u_1)-z_ja_j\phi_j(u_2)}du_1du_2$$
$$\mathcal{H}_{0,2} = \int_0^\infty \int_0^\infty \tau_{0,2}(u_1+u_2)e^{-z_0a_0\phi_0(u_1+u_2)-z_ia_i\phi_i(u_1)-z_ja_j\phi_j(u_2)}du_1du_2\,.$$

Note that $z_0a_0\mathcal{L}_{0,i}^{(0,1)} + z_ia_i\mathcal{L}_{i,0}^{(0,1)} = 1$ for any $i \in \{1, \cdots, d\}$.

The moment results of example 4.3.7 can be evaluated by some trivial algebra and we have

$$\begin{aligned}
&\mathbb{E}[P_i(A)] = \frac{H(A)z_0a_0 + H_i(A)z_ia_i}{a_0z_0 + a_iz_i}\,,\\
&\mathrm{Var}[P_i(A)] = \frac{(1-\sigma)z_0^2a_0^2H(A)H(A^c)}{(z_ia_i + z_0a_0)\,(z_ia_i + z_0a_0 + 1)} +\\
&\frac{[(z_ia_iH_i(A) + z_0a_0H(A))\,(z_ia_iH_i(A^c) + z_0a_0H(A^c))]}{(z_ia_i + z_0a_0)^2\,(z_ia_i + z_0a_0 + 1)}\,,\\
&\mathrm{Cov}[P_i(A), P_j(B)]\\
&= z_iz_ja_ia_jH_i(A)H_j(B)\left(\frac{{}_3F_2(1, a_0z_0, 1; a_0z_0 + a_iz_i + 1, a_0z_0 + a_jz_j + 1); 1) - 1}{(a_0z_0 + a_iz_i)(a_0z_0 + a_jz_j)}\right)\\
&+ z_iz_0a_ia_0H_i(A)H(B)\Big(\frac{{}_3F_2(1, a_0z_0 + 1, 1; a_0z_0 + a_iz_i + 2, a_0z_0 + a_jz_j + 1); 1)}{(a_0z_0 + a_iz_i + 1)(a_0z_0 + a_jz_j)}\\
&\qquad - \frac{1}{(a_0z_0 + a_iz_i)(a_0z_0 + a_jz_j)}\Big)\\
&+ z_jz_0a_ja_0H(A)H_j(B)\Big(\frac{{}_3F_2(1, a_0z_0 + 1, 1; a_0z_0 + a_jz_j + 2, a_0z_0 + a_iz_i + 1); 1)}{(a_0z_0 + a_jz_j + 1)(a_0z_0 + a_iz_i)}\\
&\qquad - \frac{1}{(a_0z_0 + a_iz_i)(a_0z_0 + a_jz_j)}\Big)\\
&+ z_0^2a_0^2H(A)H(B)\Big(\frac{{}_3F_2(1, a_0z_0 + 2, 1; a_0z_0 + a_jz_j + 2, a_0z_0 + a_iz_i + 2); 1)}{(a_0z_0 + a_jz_j + 1)(a_0z_0 + a_iz_i + 1)}\\
&\qquad - \frac{1}{(a_0z_0 + a_iz_i)(a_0z_0 + a_jz_j)}\Big)\\
&+ \big((1-\sigma)z_0^2a_0^2(H(A\cap B) - H(A)H(B)) + z_0a_0H(A\cap B)\big) \times\\
&\quad \frac{{}_3F_2(1, a_0z_0 + 2, 1; a_0z_0 + a_jz_j + 2, a_0z_0 + a_iz_i + 2); 1)}{(a_0z_0 + a_jz_j + 1)(a_0z_0 + a_iz_i + 1)}\,. \qquad (4.4.6)
\end{aligned}$$

The form ${}_3F_2(a, b, c; \alpha, \beta; x)$ is the generalized hypergeometric function that is defined by

$$ {}_3F_2(a, b, c; \alpha, \beta; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n (c)_n}{(\alpha)_n (\beta)_n n!} x^n \,, $$

where $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$ for any $a > 0$ and $n \in \mathbb{Z}^+$. The above hypergeometric function converges when $|x| < 1$, or $x = 1$ under the condition $Rz(\alpha + \beta - a - b - c) > 0$, with $Rz()$ denoting the real part of a complex number.

4.5 Random partition structure

Consider the partially exchangeable random sequence $\mathbf{X} = \{\mathbf{X}^{(N_i)} : i = 1, \cdots, d\}$ that is sampled from $(P_1, \cdots, P_d)$ given by either model 4.3.1 or model 4.3.5. Based on the discreteness of NRMIs, there is a positive probability (when $z_i \neq 1$) of $X_{i,k} = X_{j,l}$, no matter that $X_{i,k}$ and $X_{j,l}$ are in the same group or in different group. As we discussed in the introduction, a random partition structure is thus induced, since $X_{i,k}$ and $X_{j,l}$ will be in the same partition subset of $\mathbb{X}$ as long as $X_{i,k} = X_{j,l}$. The induced random partition distribution is identified by pEPPF $\Pi_k^{(N)}(\mathbf{n}_1, \cdots, \mathbf{n}_d)$ as defined in eq. (4.1.2). We will derive the pEPPFs corresponding to model 4.3.1 and model 4.3.5 in this section.

In order to have a detailed view of the pEPPFs induced by model 4.3.1 and model 4.3.5, we would introduce the **Local special Chinese restaurant franchise**, which is similar to the well-known *Chinese restaurant franchise* for the hierarchical Dirichlet process ((Teh et al., 2006)) and the hierarchical NRMIs ((Camerlenghi et al., 2019)). Assume that there is a Chinese restaurant franchise consisting of d restaurants located at d different locations. Each restaurant has infinite number of tables, a shared menu (same for all the d restaurants) that includes infinite number of shared common white dishes (generated by H), and a local special menu (different in different restaurant) that includes infinite number of red dishes (generated by H_i). Although we use red dishes to represent the local special dishes for all d restaurants, we assumes the red dishes between different restaurants are totally different. The first customer of each restaurant i will choose a table and order a dish, which will be shared with all the customers who afterwards join the same table. Thus, the second customer of the restaurant i will either choose a new table and order a new dish, or joint the first table to share the dish ordered by the first customer. According to the scheme, the same white dish can be served at different tables within the same restaurant and across different restaurants, whereas the red dish in restaurant i can only be served at different tables within the restaurant i. The preference of ordering white dishes or red dishes in a restaurant i is controlled by a historical rating score z_i, namely the customers in the restaurant i will prefer to order red dishes if z_i is close to 1 and prefer to order white dishes if z_i is close to 0. We denote the dish served to the jth customer in the restaurant i by $X_{i,j}$ and denote the frequency count of the number of customers in restaurant i who eat the dish $X_{i,j}$ as $n_{i,j}$ for $i \in \{1, \cdots, d\}$ and $j \in \{1, \cdots, N_i\}$. Let $r_{i,j}$ be the number of tables in restaurant i with the jth dish being served and then let $q_{i,j,l}$ be the number of customers in restaurant i at table l eating the jth dish. Hence, there are following relationships for the above quantities.

(1) $\sum_{l=1}^{r_{i,j}} q_{i,j,l} = n_{i,j}$ and $l \in \{1, \cdots, r_{i,j}\}$;

(2) $\sum_{i=1}^{d} r_{i,j} = |r_{\bullet,j}|$ is the number of tables with dish j across all the d restaurants;

(3) $\sum_{j=1}^{K} r_{i,j} = |r_{i,\bullet}|$ is the number of tables occupied in restaurant i, where K is the number of distinct dishes served in all the d restaurants to all the N customers.

The local special Chinese restaurant franchise actually describes a hierarchical partition structure: the N customers are clustered to $|\mathbf{r}| = \sum_{i=1}^{d} \sum_{j=1}^{K} r_{i,j}$ tables; then the $|\mathbf{r}|$ tables are clustered to K subgroups, which are identified by the K distinct served dishes. It is worthy to point out that the subgroups corresponding to red dishes only includes the costumers within the restaurant, but not across the restaurants.

4.5.1 pEPPF for model 4.3.1

Theorem 4.5.1. *Suppose the partially exchangeable random sequence* $\mathbf{X}$ *is a sample of size* N *from* $(P_1, \cdots, P_d)$ *that is constructed in model 4.3.1. Assuming the following* *condition*

Assumption A *Let* $\frac{\prod_{i=1}^{d} H_i(dx_j^*)^{m_{i,j}}}{H(dx_j^*)} = C(m_{1,j}, \cdots, m_{d,j})$*, where* $C(m_{1,j}, \cdots, m_{d,j})$ *is finite (could be* 0*) for any* $j \in \{1, \cdots, K\}$*,* $m_{i,j} \in \mathbb{Z}^+$*.*

Then

$$\begin{aligned}
&\Pi_K^{(N)}(\mathbf{n}_1, \cdots, \mathbf{n}_d) \\
&= \sum_{\mathbf{r}} \sum_{\mathbf{q}} \left[\left(\prod_{i=1}^{d} (z_i a_i)^{|r_{i,\bullet}|} \right) \left(\prod_{j=1}^{K} C(r_{1,j}, \cdots, r_{d,j}) \right) + \right. \\
&\quad \left. \Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|, \cdots, |r_{\bullet,K}|) a_0^{|\mathbf{r}|} \left(\prod_{i=1}^{d} (1 - z_i)^{|r_{i,\bullet}|} \right) \right] \\
&\quad \times \prod_{i=1}^{d} \prod_{j=1}^{K} \frac{1}{r_{i,j}!} \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{i,j}}} \Psi_i^{(N_i)}(\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,K}),
\end{aligned} \tag{4.5.1}$$

where the leading sum are taken of all vectors of $\mathbf{r}$ *and* $\mathbf{q}$ *such that for any* $i \in \{1, \cdots, d\}$ *and* $j \in \{1, \cdots, K\}$*:*

(1) $\mathbf{r} = (\mathbf{r}_{1,\bullet}, \cdots, \mathbf{r}_{d,\bullet})$ *and* $\mathbf{r}_{i,\bullet} = (r_{i,1}, \cdots, r_{i,K}) \in \times_{j=1}^{K} \{1, \cdots, n_{i,j}\}$*;*

(2) $\mathbf{q} = (\mathbf{q}_{1,1}, \cdots, \mathbf{q}_{1,K}, \cdots, \mathbf{q}_{d,1}, \cdots, \mathbf{q}_{d,K})$ *and* $\mathbf{q}_{i,j} = (q_{i,j,1}, \cdots, q_{i,j,r_{i,j}})$ *are vectors of positive integers such that* $\sum_{l=1}^{r_{i,j}} q_{i,j,l} = n_{i,j}$*;*

(3) $\Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|, \cdots, |r_{\bullet,K}|) = \frac{a^K}{\Gamma(|\mathbf{r}|)} \int_0^\infty u^{|\mathbf{r}|-1} e^{-a\phi(u)} \prod_{j=1}^{K} \tau_{|r_{\bullet,j}|}(u) du$

(4) $\Psi_i^{(N_i)}(\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,K}) = \frac{1}{\Gamma(N_i)} \int_0^\infty u^{N_i-1} e^{-(z_i a_i + (1-z_i)a_0)\phi_i(u)} \prod_{j=1}^{K} \prod_{t=1}^{r_{i,j}} \tau_{i,q_{i,j,t}}(u) du.$

It is worthy to point out that the partially exchangeable random partition structure is a multiplication of two hierarchies:

(i) The exchangeable random partition structure of each of the d groups, this is identified by the EPPF $\Psi_i^{(N_i)}(\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,k})$ of each group $i \in \{1, \cdots, d\}$. Correspondingly to the local special Chinese restaurant franchise metaphor, this is the partition probability function of separating N_i costumers in restaurant i by K distinct dishes.

(ii) The exchangeable random partition structure of the whole d group after the partition in (i), this is identified by the EPPF $\Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|, \cdots, |r_{\bullet,K}|)$, which corresponds to the randomness of H_0; and $\prod_{j=1}^{K} C(r_{1,j}, \cdots, r_{d,j})$, which corresponds to the unique deterministic part H_i of group H_i. As to the local special Chinese restaurant franchise metaphor, this acts the partition across d groups based on the K distinct dishes.

We also present the pEPPF for the case in example 4.3.4 as follows.

Example 4.5.2 (Continued of example 4.3.4)**.** *If $(P_1, \cdots, P_d)$ is constructed as in example 4.3.4, then the pEPPF can be computed under the Assumption A as*

$$
\begin{aligned}
&\Pi_K^{(N)}(\mathbf{n}_1, \cdots, \mathbf{n}_d) \\
&= \sum_{\mathbf{r}} \Bigg[\left(\prod_{i=1}^{d} (z_i a_i)^{|r_{i,\bullet}|} \right) \left(\prod_{j=1}^{K} C(r_{1,j}, \cdots, r_{d,j}) \right) + \\
&\quad \frac{\sigma^{K-1}\Gamma(K)}{\Gamma(|\mathbf{r}|)} \left(\prod_{j=1}^{K} (1-\sigma)_{|r_{\bullet,j}|-1} \right) a_0^{|\mathbf{r}|} \left(\prod_{i=1}^{d} (1-z_i)^{|r_{i,\bullet}|} \right) \Bigg] \\
&\quad \times \left(\prod_{i=1}^{d} \frac{1}{(z_i a_i + (1-z_i) a_0)_{N_i}} \right) \left(\prod_{j=1}^{K} (|r_{\bullet,j}|-1)! \prod_{i=1}^{d} \mathcal{S}(n_{i,j}, r_{i,j}) \right),
\end{aligned}
$$

where $\mathcal{S}(n,m)$ denotes the unsigned Stirling number of the first kind.

4.5.2 pEPPF for model 4.3.5

Theorem 4.5.3. *Suppose the partially exchangeable random sequence $\mathbf{X}$ is a sample of size N from $(P_1, \cdots, P_d)$ that is constructed in model 4.3.5. Under the assumption A in theorem 4.5.1, we have*

$$
\begin{aligned}
&\Pi_K^{(N)}(\mathbf{n}_1, \cdots, \mathbf{n}_d) \\
&= \sum_{\mathbf{r}_{\max}} \sum_{\mathbf{q}} \Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^{d} \frac{u_i^{N_i-1}}{\Gamma(N_i)} \right) e^{-\sum_{i=1}^{d} z_i a_i \phi_i(u_i) - z_0 a_0 \phi_0\left(\sum_{i=1}^{d} u_i\right)} \times \\
&\quad \left\{ \prod_{j=1}^{K} \frac{1}{r_{\max,j}!} \prod_{i=1}^{d} \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}}} \right\} \times
\end{aligned}
$$

$$\Bigg[\Psi^{(r_{\max})}(r_{\max,1},\cdots,r_{\max,K})\,(z_0a_0)^{r_{\max}}\prod_{j=1}^{K}\prod_{t_j=1}^{r_{\max,j}}\tau_{0,\sum_{i=1}^{d}q_{i,j,t_j}}(\sum_{i=1}^{d}u_i)+$$
$$\prod_{j=1}^{K}C(r_{1,j},\cdots,r_{d,j})\left(\prod_{t_j=1}^{r_{\max,j}}\mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}}\right)\prod_{i=1}^{d}(z_ia_iH_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|}\prod_{t_j=1}^{r_{\max,j}}\tau_{i,q_{i,j,t_j}}(u_i)\Bigg]d\mathbf{u}\Bigg], \tag{4.5.2}$$

where for any $i\in\{1,\cdots,d\}$, $j\in\{1,\cdots,K\}$ *the sum of* $\mathbf{r}_{\max}$ *and* $\mathbf{q}$ *are taking over the vector of* $\mathbf{r}_{\max}$ *and* $\mathbf{q}$, *and*

(1) $\mathbf{r}_{\max}=(r_{\max,1},\cdots,r_{\max,K})$ *with* $r_{\max,j}\in\{1,\cdots,n_{\bullet,j}\}$ *and* $r_{\max}=\sum_{j=1}^{K}r_{\max,j}$;

(2) $\mathbf{q}=(\mathbf{q}_{1,1},\cdots,\mathbf{q}_{1,K},\cdots,\mathbf{q}_{d,1},\cdots,\mathbf{q}_{d,K})$ *and* $\mathbf{q}_{i,j}=(q_{i,j,1},\cdots,q_{i,j,r_{\max,j}})$ *are vectors of nonnegative integers such that* $\sum_{t_j=1}^{r_{\max,j}}q_{i,j,t_j}=n_{i,j}$;

(3) We define $\tau_{i,0}(u_i)=1$ *temporarily in this theorem for the sake of notational simplicity;*

(4) $|\overrightarrow{\mathbf{q}_{j,t_j}}|$ *is the length of the vector* $(q_{1,j,t_j},\cdots,q_{d,j,t_j})$ *and* $|\overrightarrow{\mathbf{q}_{i,j}}|$ *is the length of the vector* $(q_{i,j,1},\cdots,q_{i,j,r_{\max,j}})$.

The pEPPF corresponding to the case in example 4.3.7 is illustrated as follows.

Example 4.5.4 (Continued of example 4.3.7). *If* $(P_1,\cdots,P_d)$ *is constructed as in example 4.3.7, then the pEPPF can be computed under the Assumption A as*

$$\Pi_K^{(N)}(\mathbf{n}_1,\cdots,\mathbf{n}_d)$$
$$=\sum_{\mathbf{r}_{\max}}\sum_{\mathbf{q}}\Bigg\{\frac{\sigma^K(z_0a_0)^{r_{\max}}}{\Gamma(r_{\max})\left(\prod_{i=1}^{d}\Gamma(N_i)\right)}\Phi^{(d)}(N_1,\cdots,N_d;a_1z_1,\cdots,a_dz_d;N+a_0z_0)\times$$
$$\left\{\prod_{j=1}^{K}\frac{1}{r_{\max,j}!}\prod_{i=1}^{d}\binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{\max,j}}}\right\}\prod_{j=1}^{K}\left\{(1-\sigma)_{(r_{\max,j}-1)}\prod_{t_j=1}^{r_{\max,j}}\Gamma(\sum_{i=1}^{d}q_{i,j,t_j})\right\}+$$
$$\left(\prod_{i=1}^{d}\frac{(a_iz_i)^{|\overrightarrow{\mathbf{q}_{i,\bullet}}|}}{\Gamma(N_i)(a_iz_i)_{N_i}}\right)\Phi^{(d)}(N_1,\cdots,N_d;a_1z_1+N_1,\cdots,a_dz_d+N_d;a_0z_0)\times$$
$$\prod_{j=1}^{K}\left\{\frac{C(r_{i,j},\cdots,r_{d,j})}{r_{\max,j}!}\left(\prod_{t_j=1}^{r_{\max,j}}\mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}}\right)\left(\prod_{i=1}^{d}\binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{\max,j}}}\prod_{t_j=1}^{r_{\max,j}}\Gamma(q_{i,j,t_j})\right)\right\}\Bigg\},$$

where

$$\Phi^{(d)}(\alpha_1, \cdots, \alpha_d; \beta_d, \cdots, \beta_d; \gamma)$$
$$= \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{d} \left(u_i^{\alpha_i - 1}(1+u_i)^{-\beta_i}\right)(1+u_1+\cdots+u_d)^{-\gamma} du_1 \cdots du_d \,.$$

From Theorem4.5.1 and Theorem 4.5.3, we see that the random partition structure induced by model 4.3.1 and model 4.3.5 are both two-level structure partition models described by the local special Chinese restaurant franchise. The two-level partition structure of model 4.3.1 can be separated level by level, in the sense that the first level partition of each d groups of observations are not affected by the second overall level partition. Whereas the two level partition structures of model 4.3.5 are affected each other.

4.6 Cluster numbers K_N

It is natural to consider the partition structures induced by model 4.3.1 and model 4.3.5 after the pEPPFs is the distributions of the number of random partitions K_N. According to the local special Chinese restaurant franchise, K_N represents the number of distinct dishes served to the N customers in the d restaurants. To derive the distribution of K_N and to make the presentation easy to follow, we introduce a collection of sequences $\{(Y_{i,j})_{j=1}^{N_i} : i = 1, \cdots, d\}$, where $Y_{i,j}|\tilde{P}_i \overset{iid}{\sim} \tilde{P}_i$ and $\tilde{P}_i$ is defined similarly to P_i as in model 4.3.1 and model 4.3.5 but with non-random H_0. That is to say, we only consider the first level partition structure of model 4.3.1 and model 4.3.5. For $i = 1, \cdots, d$, let $K_{i,N_i} = K_{i,N_i}^{(1)} + K_{i,N_i}^{(2)}$ be the number of distinct observations in $\mathbf{Y}^{N_i} = (Y_{i,1}, \cdots, Y_{i,N_i})$, where $K_{i,N_i}^{(1)}$ is the number of tables occupied in restaurant i and served common white dishes, $K_{i,N_i}^{(2)}$ is the number of tables occupied in restaurant i and served special red dishes. Let $K_{0,t}$ be the number of distinct observations for t exchangeable observations from $H_0 \sim \text{NRMI}(a, H, \rho)$, then $K_{0,t}$ represent the number of distinct common white dishes dishes served on the t occupied tables.

Due to the different constructions of model 4.3.1 and model 4.3.5, $(K_{1,N_1}, \cdots, K_{d,N_d})$

are independent for the former case and dependent for the later case.

Theorem 4.6.1. *Suppose the partially exchangeable random sequence* $\mathbf{X}$ *is a sample of size* N *from* $(P_1, \cdots, P_d)$ *that is constructed in either model 4.3.1 or model 4.3.5. Then, for any* $k \in \{1, \cdots, N\}$*, we have*

$$\mathbb{P}(K_N = k) = \sum_{\eta=0}^{k} \sum_{t=k-\eta}^{N} \mathbb{P}(K_{0,t} = k - \eta) \mathbb{P}\left(\sum_{i=1}^{d} K_{i,N_i}^{(1)} = t, \sum_{i=1}^{d} K_{i,N_i}^{(2)} = \eta\right).$$

For model 4.3.1, we further have

$$\mathbb{P}(K_N = k) = \sum_{\eta=0}^{k} \sum_{t=k-\eta}^{N} \mathbb{P}(K_{0,t} = k - \eta) \sum_{\substack{(t_1,\cdots,t_d)\in\Delta_{d,t} \\ (\eta_1,\cdots,\eta_d)\in\Delta_{d,\eta}}} \prod_{i=1}^{d} \mathbb{P}\left(K_{i,N_i}^{(1)} = t_i, K_{i,N_i}^{(2)} = \eta_i\right),$$

where $\Delta_{d,n} = \{(r_1, \cdots, r_d) : r_i \geq 0, \sum_{i=1}^{d} r_i = n\}$.

The distribution of $K_{0,t}$ and $(K_{i,N_i}^{(1)}, K_{i,N_i}^{(2)})$ can be derived from Theorem 4.5.1 and Theorem 4.5.3. For model 4.3.1, we have

$$\mathbb{P}(K_{0,t} = k_0) = \frac{1}{k_0!} \sum_{(r_1,\cdots,r_{k_0})\in\Lambda_{k_0,t}} \binom{t}{r_1 \cdots r_{k_0}} \Psi^{(t)}(r_1, \cdots, r_{k_0})$$

for any $k_0 \in \{1, \cdots, t\}$, where $\Lambda_{d,n} = \{(r_1, \cdots, r_d) : r_i \geq 1, \sum_{i=1}^{d} r_i = n\}$. And for any (t_i, η_i) such that $t_i + \eta_i \in \{1, \cdots, N_i\}$, we have

$$\begin{aligned}
&\mathbb{P}\left(K_{i,N_i}^{(1)} = t_i, K_{i,N_i}^{(2)} = \eta_i\right) \\
&= \frac{z_i a_i + (1 - z_i) a_0}{(t_i + \eta_i)!} \sum_{(r_1,\cdots,r_{t_i},\cdots,r_{t_i+\eta_i})\in\Lambda_{t_i+\eta_i,N_i}} \binom{N_i}{r_1 \cdots r_{t_i} \cdots r_{t_i+\eta_i}} \Psi_i^{(N_i)}(r_1, \cdots, r_{t_i}, \cdots, r_{t_i+\eta_i}).
\end{aligned}$$

For model 4.3.5, we have

$$\mathbb{P}(K_{0,t} = k_0) = \frac{1}{k_0!} \sum_{(r_1,\cdots,r_{k_0})\in\Lambda_{k_0,t}} \binom{t}{r_1 \cdots r_{k_0}} \Psi^{(t)}(r_1, \cdots, r_{k_0})$$

for any $k_0 \in \{1, \cdots, t\}$. And for $(t_i, \eta_i)_{i=1}^d$ such that $t_i + \eta_i \in \{1, \cdots, N_i\}$ for all i, we have

$$\begin{aligned}
&\mathbb{P}\left(K_{1,N_1}^{(1)} = t_1, K_{1,N_1}^{(2)} = \eta_1, \cdots, K_{d,N_d}^{(1)} = t_d, K_{d,N_d}^{(2)} = \eta_d\right) \\
&= \sum_{\substack{(r_{1,1},\cdots,r_{1,t_1+\eta_1})\in\Lambda_{t_1+\eta_1,N_1} \\ \cdots \\ (r_{d,1},\cdots,r_{d,t_d+\eta_d})\in\Lambda_{t_d+\eta_d,N_d}}} \left(\prod_{i=1}^d \frac{1}{(t_i+\eta_i)!}\binom{N_i}{r_{i,1}\cdots r_{i,t_i+\eta_i}}\right) \times \\
&\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^d \frac{u_i^{N_i}}{\Gamma(N_i)}\right) e^{-\sum_{i=1}^d z_i a_i \phi_i(u_i) - z_0 a_0 \phi_0\left(\sum_{i=1}^d u_i\right)} \times \\
&\left\{\prod_{i=1}^d z_i a_i \prod_{j_i=1}^{t_i+\eta_i} \tau_{i,r_{i,j_i}}(u_i) + (z_0 a_0)^d \prod_{i=1}^d \tau_{0,N_i}\left(\sum_{i=1}^d u_i\right)\right\} d\mathbf{u}\,.
\end{aligned}$$

Example 4.6.2 (Continuation of example 4.3.4). *If $(P_1, \cdots, P_d)$ is constructed as in example 4.3.4, the distribution of K_N can be found as*

$$\begin{aligned}
\mathbb{P}(K_N = k) = \sum_{\eta=0}^k \sum_{t=k}^N \frac{t\sigma^{k-\eta-1}}{k-\eta} &\left(\sum_{(r_1,\cdots,r_{k-\eta})\in\Lambda_{k-\eta,t}} \prod_{j=1}^{k-\eta} \frac{(1-\sigma)_{r_j-1}}{r_j!}\right) \times \\
&\sum_{\substack{(t_1,\cdots,t_d)\in\Delta_{d,t} \\ (\eta_1,\cdots,\eta_d)\in\Delta_{d,\eta}}} \prod_{i=1}^d \left(\frac{N_i!\left(\sum_{(r_{i,1},\cdots,r_{i,t_i+\eta_i})\in\Lambda_{t_i+\eta_i,N_i}} \prod_{j_i=1}^{t_i+\eta_i} r_j^{-1}\right)}{(t_i+\eta_i)!(a_i z_i + (1-z_i)a_0 + 1)_{N_i-1}}\right).
\end{aligned}$$

Example 4.6.3 (Continuation of example 4.3.7). *If $(P_1, \cdots, P_d)$ is constructed as in example 4.3.7, the distribution of K_N can be found as*

$$\begin{aligned}
&\mathbb{P}(K_N = k) = \sum_{\eta=0}^k \sum_{t=k}^N \frac{t\sigma^{k-\eta-1}}{k-\eta} \left(\sum_{(r_1,\cdots,r_{k-\eta})\in\Lambda_{k-\eta,t}} \prod_{j=1}^{k-\eta} \frac{(1-\sigma)_{r_j-1}}{r_j!}\right) \times \\
&\sum_{\substack{(t_1,\cdots,t_d)\in\Delta_{d,t} \\ (\eta_1,\cdots,\eta_d)\in\Delta_{d,\eta}}} \sum_{\substack{(r_{1,1},\cdots,r_{1,t_1+\eta_1})\in\Lambda_{t_1+\eta_1,N_1} \\ \cdots \\ (r_{d,1},\cdots,r_{d,t_d+\eta_d})\in\Lambda_{t_d+\eta_d,N_d}}} \left(\prod_{i=1}^d \frac{N_i z_i a_i \Phi^{(d)}(N_1,\cdots,N_d; a_1 z_1 + N_1, \cdots, a_d z_d + N_d; a_0 z_0)}{(t_i+\eta_i)! \prod_{j_i=1}^{t_i+\eta_i} r_{i,j_i}}\right. \\
&\qquad \left. + \prod_{i=1}^d \frac{a_0 z_0 N_i! \Phi^{(d)}(N_1,\cdots,N_d; a_1 z_1, \cdots, a_d z_d; N + a_0 z_0)}{(t_i+\eta_i)! \prod_{j_i=1}^{t_i+\eta_i} r_{i,j_i}!}\right).
\end{aligned}$$

4.7 Discussion

Although the proposed constructions provide flexible tools in modelling partially exchangeable observations, they show some complications in deriving the theoretical results. There are further works that should be done for our constructions, for example, the asymptotic behaviour of K_N when N is large, and the posterior analysis for the two constructions.

4.8 Appendix

Proof of Theorem 4.4.1

Proof.

$$\begin{aligned}
\mathbb{E}[P_i(A)|H_0] &= \mathbb{E}\left[\frac{\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A)}{\tilde{\mu}_i(\mathbb{X})+\tilde{\mu}_{i,0}(\mathbb{X})}|H_0\right] \\
&= \mathbb{E}\left[\int_0^\infty e^{-u(\tilde{\mu}_i(\mathbb{X})+\tilde{\mu}_{i,0}(\mathbb{X}))}(\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A))du|H_0\right] \\
&= \int_0^\infty -\frac{d}{du}\mathbb{E}\left[e^{-u(\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A))}\right]\mathbb{E}\left[e^{-u(\tilde{\mu}_i(A^c)+\tilde{\mu}_{i,0}(A^c))}\right]du \\
&= (z_i a_i H_i(A)+(1-z_i)a_0 H_0(A))\int_0^\infty e^{-(z_i a_i+(1-z_i)a_0)\phi_i(u)}\frac{d\phi(u)}{du}du \\
&= \frac{z_i a_i H_i(A)+(1-z_i)a_0 H_0(A)}{z_i a_i+(1-z_i)a_0}.
\end{aligned}$$

Then, $\mathbb{E}[P_i(A)] = \mathbb{E}[\mathbb{E}[P_i(A)|H_0]]$ is followed by the fact $\mathbb{E}[H_0(A)] = H(A)$.

We can further calculate the second moments as follows.

$$\begin{aligned}
\mathbb{E}[P_i(A)^2|H_0] &= \mathbb{E}\left[\left(\frac{\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A)}{\tilde{\mu}_i(\mathbb{X})+\tilde{\mu}_{i,0}(\mathbb{X})}\right)^2|H_0\right] \\
&= \mathbb{E}\left[\int_0^\infty ue^{-u(\tilde{\mu}_i(\mathbb{X})+\tilde{\mu}_{i,0}(\mathbb{X}))}(\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A))^2du|H_0\right] \\
&= \int_0^\infty u\frac{d^2}{du^2}\mathbb{E}\left[e^{-u(\tilde{\mu}_i(A)+\tilde{\mu}_{i,0}(A))}\right]\mathbb{E}\left[e^{-u(\tilde{\mu}_i(A^c)+\tilde{\mu}_{i,0}(A^c))}\right]du \\
&= \int_0^\infty ue^{-(z_i a_i+(1-z_i)a_0)\phi_i(u)}\left[(z_i a_i H_i(A)+(1-z_i)a_0 H_0(A))^2\tau_{i,1}(u)^2\right.
\end{aligned}$$

$$+ (z_i a_i H_i(A) + (1-z_i)a_0 H_0(A))\tau_{i,2}(u)\big] du\,.$$

By the fact that $\mathbb{E}[P_i(\mathbb{X})^2|H_0] = 1$, we have

$$\int_0^\infty u e^{-(z_i a_i + (1-z_i)a_0)\phi_i(u)} \tau_{i,1}(u)^2 du$$
$$= \frac{1}{(z_i a_i + (1-z_i)a_0)^2} - \frac{\mathcal{G}_{i,2}}{z_i a_i + (1-z_i)a_0}\,.$$

Combining with the fact that $\mathbb{E}[H_0(A)^2] = H(A)^2 + aH(A)H(A^c)\mathcal{G}_2$ (James et al. (2006), Proposition 1), we can find $\mathrm{Var}[P_i(A)]$ by $\mathbb{E}[\mathbb{E}[P_i(A)^2|H_0]] - \mathbb{E}[P(A)]^2$.

Since $P_i(A)$ and $P_j(B)$ are conditionally independent, we have

$$\begin{aligned}\mathbb{E}[P_i(A)P_j(B)|H_0] &= \mathbb{E}[P_i(A)|H_0]\mathbb{E}[P_j(B)|H_0]\\ &= \left(\frac{z_i a_i H_i(A) + (1-z_i)a_0 H_0(A)}{z_i a_i + (1-z_i)a_0}\right)\left(\frac{z_j a_j H_j(B) + (1-z_j)a_0 H_0(B)}{z_j a_j + (1-z_j)a_0}\right)\,.\end{aligned}$$

Noting that $\mathbb{E}[H_0(A)H_0(B)] = (H(A\cap B) - H(A)H(B))\mathcal{G}_2 + H(A)H(B)$, we can further calculate

$$\begin{aligned}\mathbb{E}[P_i(A)P_j(B)] &= \mathbb{E}[\mathbb{E}[P_i(A)P_j(B)|H_0]]\\ &= \left(\frac{z_i a_i H_i(A) + (1-z_i)a_0 H(A)}{z_i a_i + (1-z_i)a_0}\right)\left(\frac{z_j a_j H_j(B) + (1-z_j)a_0 H(B)}{z_j a_j + (1-z_j)a_0}\right)\\ &\quad + \frac{a_0^2 a(1-z_i)(1-z_j)(H(A\cap B) - H(A)H(B))\mathcal{G}_2}{(z_i a_i + (1-z_i)a_0)(z_j a_j + (1-z_j)a_0)}\,.\end{aligned}$$

The covariance result then follows. ☐

Proof of Theorem 4.4.2

Proof. The results of $\mathbb{E}[P_i(A)]$ and $\mathbb{E}[P_i(A)^2]$ can be found by the similar steps to those in the proof 4.8.

For the computation of Cov$[P_i(A), P_j(B)]$, it is sufficient to compute

$$\begin{aligned}
\mathbb{E}[P_i(A)P_j(B)|H_0] &= \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\mu_i(A)\mu_j(B)|H_0\right] du_i du_j \\
&= \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}(\tilde{\mu}_i(A)+\tilde{\mu}_0(A))(\tilde{\mu}_j(B)+\tilde{\mu}_0(B))|H_0\right] du_i du_j \\
&= \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_i(A)\tilde{\mu}_j(B)|H_0\right] du_i du_j + \\
&\quad \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_i(A)\tilde{\mu}_0(B)|H_0\right] du_i du_j + \\
&\quad \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_0(A)\tilde{\mu}_j(B)|H_0\right] du_i du_j + \\
&\quad \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_0(A)\tilde{\mu}_0(B)|H_0\right] du_i du_j \,,
\end{aligned}$$

where the first three forms in the last equation can be computed in the similar way as follows.

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_i(A)\tilde{\mu}_j(B)|H_0\right] du_i du_j \\
&= \int_0^\infty \int_0^\infty \left(-\frac{d}{du_i}\mathbb{E}\left[e^{-u_i\tilde{\mu}_i(A)}|H_0\right]\right)\left(-\frac{d}{du_j}\mathbb{E}\left[e^{-u_j\tilde{\mu}_j(B)}|H_0\right]\right) \\
&\qquad \mathbb{E}\left[e^{-u_i\tilde{\mu}_i(A^c)-u_j\tilde{\mu}_j(B^c)-(u_i+u_j)\tilde{\mu}_0(\mathbb{X})}|H_0\right] du_i du_j \\
&= z_i z_j a_i a_j H_i(A) H_j(B) \mathcal{K}_{i,j} \,.
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\mu_i(\mathbb{X})-u_j\mu_j(\mathbb{X})}\tilde{\mu}_0(A)\tilde{\mu}_0(B)|H_0\right] du_i du_j \\
&= \int_0^\infty \int_0^\infty \mathbb{E}\Big[e^{-u_i\tilde{\mu}_i(\mathbb{X})-u_j\tilde{\mu}_j(\mathbb{X})-(u_i+u_j)\tilde{\mu}_0(\mathbb{X})} \\
&\qquad (\tilde{\mu}_0(A\cap B)+\tilde{\mu}_0(A\cap B^c))(\tilde{\mu}_0(B\cap A)+\tilde{\mu}_0(B\cap A^c))|H_0\Big] du_i du_j \\
&= \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\tilde{\mu}_i(\mathbb{X})-u_j\tilde{\mu}_j(\mathbb{X})-(u_i+u_j)\tilde{\mu}_0((A\cap B)^c)}|H_0\right] \frac{d^2}{d(u_i+u_j)^2}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(A\cap B)}|H_0\right] du_i du_j \\
&\quad + \int_0^\infty \int_0^\infty \mathbb{E}\left[e^{-u_i\tilde{\mu}_i(\mathbb{X})-u_j\tilde{\mu}_j(\mathbb{X})-(u_i+u_j)\tilde{\mu}_0(((A\cap B)\cup(B\cap A^c))^c)}|H_0\right] \\
&\qquad \frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(A\cap B)}|H_0\right] \frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(B\cap A^c)}|H_0\right] du_i du_j
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^\infty\int_0^\infty \mathbb{E}\left[e^{-u_i\tilde{\mu}_i(\mathbb{X})-u_j\tilde{\mu}_j(\mathbb{X})-(u_i+u_j)\tilde{\mu}_0(((A\cap B)\cup(B^c\cap A))^c)}|H_0\right] \\
&\quad \frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(A\cap B)}|H_0\right]\frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(B^c\cap A)}|H_0\right]du_idu_j \\
&+\int_0^\infty\int_0^\infty \mathbb{E}\left[e^{-u_i\tilde{\mu}_i(\mathbb{X})-u_j\tilde{\mu}_j(\mathbb{X})-(u_i+u_j)\tilde{\mu}_0(((A\cap B^c)\cup(B\cap A^c))^c)}|H_0\right] \\
&\quad \frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(A\cap B^c)}|H_0\right]\frac{-d}{d(u_i+u_j)}\mathbb{E}\left[e^{-(u_i+u_j)\tilde{\mu}_0(B\cap A^c)}|H_0\right]du_idu_j \\
&= z_0^2a_0^2H_0(A)H_0(B)\mathcal{K}_{0,0}+z_0a_0H_0(A\cap B)\mathcal{H}_{0,2}\,.
\end{aligned}
$$

By combining the above calculations, we have

$$
\begin{aligned}
\mathbb{E}[P_i(A)P_j(B)|H_0] = {} & z_iz_ja_ia_jH_i(A)H_j(B)\mathcal{K}_{i,j}+z_iz_0a_ia_0H_i(A)H_0(B)\mathcal{K}_{i,0}+ \\
& z_0z_ja_0a_jH_0(A)H_j(B)\mathcal{K}_{i,j}+z_0^2a_0^2H_0(A)H_0(B)\mathcal{K}_{0,0}+z_0a_0H_0(A\cap B)\mathcal{H}_{0,2}\,.
\end{aligned}
$$

The result of $\mathrm{Cov}(P_i(A),P_j(B))$ then follows by the fact that $\mathbb{E}[H_0(A)H_0(B)] = H(A)H(B)+(H(A\cap B)-H(A)H(B))\mathcal{G}_2$. □

Proof of Theorem 4.5.1

Proof. From eq. (4.1.2), it follows

$$
\Pi_K^{(N)}(\mathbf{n}_1,\cdots,\mathbf{n}_d) = \int_{\mathbb{X}^K}\mathbb{E}\left[\prod_{i=1}^d\prod_{j=1}^K P_i(dx_j^*)^{n_{i,j}}\right].
$$

We shall evaluate the form $\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*),\cdots,B_\epsilon(x_K^*))$ as follows.

$$
\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*),\cdots,B_\epsilon(x_K^*)) = \mathbb{E}\left[\prod_{i=1}^d\prod_{j=1}^k P_i(B_\epsilon(x_j^*))^{n_{i,j}}\right],
$$

where $B_\epsilon(x_j^*)$ is a ball of radius ϵ and center x_j^* such that $\epsilon>0$ is small enough to make $B_\epsilon(x_j^*)\cap B_\epsilon(x_i^*)=\emptyset$ for $i\neq j$.

Note that

$$
\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*),\cdots,B_\epsilon(x_K^*))\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{d}\prod_{j=1}^{K}P_i(B_\epsilon(x_j^*))^{n_{i,j}}|H_0\right]\right]\\
&\quad=\mathbb{E}\left[\prod_{i=1}^{d}\mathbb{E}\left[\prod_{j=1}^{K}\left(\frac{\mu_i(B_\epsilon(x_j^*))}{\mu_i(\mathbb{X})}\right)^{n_{i,j}}|H_0\right]\right]\\
&\quad=\mathbb{E}\left[\prod_{i=1}^{d}\int_0^\infty\frac{u^{N_i-1}}{\Gamma(N_i)}\mathbb{E}\left[e^{-u\mu_i(\mathbb{X})}\prod_{j=1}^{K}\mu_i(B_\epsilon(x_j^*))^{n_{i,j}}|H_0\right]du\right]\\
&\quad=\mathbb{E}\left[\prod_{i=1}^{d}\int_0^\infty\frac{u^{N_i-1}}{\Gamma(N_i)}\mathbb{E}\left[e^{-u\mu_i(B^*)}|H_0\right]\prod_{j=1}^{K}(-1)^{n_{i,j}}\frac{d^{n_{i,j}}}{du^{n_{i,j}}}\mathbb{E}\left[e^{-u\mu_i(B_\epsilon(x_j^*))}|H_0\right]du\right],
\end{aligned}
\tag{4.8.1}
$$

where $B^*=\mathbb{X}\backslash\left(\cup_{j=1}^{K}B_\epsilon(x_j^*)\right)$ and the last equation is due to the conditional independence of $\mu_i(B_\epsilon(x_j^*))$ for different j. By the famous Faà di Bruno formula (see e.g., Hardy (2006)), we can calculate the form

$$
\begin{aligned}
&(-1)^{n_{i,j}}\frac{d^{n_{i,j}}}{du^{n_{i,j}}}\mathbb{E}\left[e^{-u\mu_i(B_\epsilon(x_j^*))}|H_0\right]\\
&=(-1)^{n_{i,j}}\frac{d^{n_{i,j}}}{du^{n_{i,j}}}e^{-[z_ia_iH_i(B_\epsilon(x_j^*))+(1-z_i)a_0H_0(B_\epsilon(x_j^*))]\phi_i(u)}\\
&=e^{-[z_ia_iH_i(B_\epsilon(x_j^*))+(1-z_i)a_0H_0(B_\epsilon(x_j^*))]\phi_i(u)}\times\\
&\sum_{r_{i,j}=1}^{n_{i,j}}\left[z_ia_iH_i(B_\epsilon(x_j^*))+(1-z_i)a_0H_0(B_\epsilon(x_j^*))\right]^{r_{i,j}}\frac{1}{r_{i,j}!}\sum_{\mathbf{q}_{i,j}}\binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{i,j}}}\prod_{t=1}^{r_{i,j}}\tau_{i,q_{i,j,t}}(u),
\end{aligned}
\tag{4.8.2}
$$

where the sum of $\mathbf{q}_{i,j}$ is taken all over the vector of positive integers $(q_{i,j,1},\cdots,q_{i,j,r_{i,j}})$ such that $\sum_{t=1}^{r_{i,j}}q_{i,j,t}=n_{i,j}$.

Applying eq. (4.8.2) to eq. (4.8.1), we have

$$\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*),\cdots,B_\epsilon(x_K^*))$$

$$= \mathbb{E}\Bigg[\prod_{i=1}^{d}\int_0^\infty \frac{u^{N_i-1}}{\Gamma(N_i)} e^{-(z_i a_i + (1-z_i)a_0)\phi_i(u)} \prod_{j=1}^{K}\Big(\sum_{r_{i,j}=1}^{n_{i,j}} \left[z_i a_i H_i(B_\epsilon(x_j^*)) + (1-z_i)a_0 H_0(B_\epsilon(x_j^*))\right]^{r_{i,j}}$$

$$\times \frac{1}{r_{i,j}!}\sum_{\mathbf{q}_{i,j}} \binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{i,j}}} \prod_{t=1}^{r_{i,j}} \tau_{i,q_{i,j,t}}(u)\Big)du\Bigg]. \tag{4.8.3}$$

Since the expectation is taken with respect to H_0, and $H_0 \sim \text{NRMI}(aH,\rho)$, for any sequence of positive integers $\mathbf{m} = (m_1,\cdots,m_K)$, we have the following identity.

$$\begin{aligned}
&\mathbb{E}\left[\prod_{j=1}^{K} H_0(B_\epsilon(x_j^*))^{m_j}\right] \\
&\quad= \prod_{j=1}^{K} H(B_\epsilon(x_j^*)) \frac{a^K}{\Gamma(|\mathbf{m}|)} \int_0^\infty u^{|\mathbf{m}|-1} e^{-a\phi(u)} \prod_{j=1}^{K}\tau_{m_j}(u)du \\
&\quad= \Psi^{(|\mathbf{m}|)}(m_1,\cdots,m_K) \times \prod_{j=1}^{K} H(B_\epsilon(x_j^*)).
\end{aligned} \tag{4.8.4}$$

Thus, we can rearrange eq. (4.8.3) as

$$\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*),\cdots,B_\epsilon(x_K^*)) \\
&\quad= \sum_{\mathbf{r}}\left(\mathbb{E}\left[\prod_{i=1}^{d}\prod_{j=1}^{K}\left(z_i a_i H_i(B_\epsilon(x_j^*)) + (1-z_i)a_0 H_0(B_\epsilon(x_j^*))\right)^{r_{i,j}}\right]\right)\times \\
&\quad\prod_{i=1}^{d}\int_0^\infty \frac{u^{N_i-1}}{\Gamma(N_i)} e^{-(z_i a_i+(1-z_i)a_0)\phi_i(u)} \prod_{j=1}^{k}\frac{1}{r_{i,j}!}\binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{i,j}}}\prod_{t=1}^{r_{i,j}}\tau_{i,q_{i,j,t}}(u)du \\
&\quad= \sum_{\mathbf{r}}\Bigg\{\prod_{i=1}^{d}\prod_{j=1}^{K}\left(z_i a_i H_i(B_\epsilon(x_j^*))\right)^{r_{i,j}} + \\
&\quad\Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|,\cdots,|r_{\bullet,K}|)a_0^{|\mathbf{r}|}\left(\prod_{i=1}^{d}(1-z_i)^{|r_{i,\bullet}|}\right)\left(\prod_{j=1}^{K}H(B_\epsilon(x_j^*))\right) + \mathcal{R}^*\Bigg\}\times \\
&\quad\prod_{i=1}^{d}\int_0^\infty \frac{u^{N_i-1}}{\Gamma(N_i)} e^{-(z_i a_i+(1-z_i)a_0)\phi_i(u)} \prod_{j=1}^{k}\frac{1}{r_{i,j}!}\binom{n_{i,j}}{q_{i,j,1},\cdots,q_{i,j,r_{i,j}}}\prod_{t=1}^{r_{i,j}}\tau_{i,q_{i,j,t}}(u)du,
\end{aligned} \tag{4.8.5}$$

where the last equation is computed by expanding

$$\mathbb{E}\left[\prod_{i=1}^{d}\prod_{j=1}^{K}\left(z_i a_i H_i(B_\epsilon(x_j^*)) + (1-z_i)a_0 H_0(B_\epsilon(x_j^*))\right)^{r_{i,j}}\right]$$

and applying eq. (4.8.4). The notation $\mathcal{R}^* = \mathcal{R}(B_\epsilon(x_1^*), \cdots, B_\epsilon(x_K^*))$ is the sum of all the cross terms with the form

$$\begin{aligned}
&\mathbb{E}\left\{\prod_{j=1}^{K} C_{i,j,0} H_i(B_\epsilon(x_j^*))^{\alpha_{i,j}} H_0(B_\epsilon(x_j^*))^{\beta_{i,j}}\right\} \\
&= \Psi^{|\beta_i|}(\beta_{i,1}, \cdots, \beta_{i,K}) \times \prod_{j=1}^{K} C_{i,j,0} H_i(B_\epsilon(x_j^*))^{\alpha_{i,j}} H_0(B_\epsilon(x_j^*)) \, ,
\end{aligned}$$

where $\alpha_{i,j}$, $\beta_{i,j}$ are some positive integers, and $C_{i,j,0}$ is the coefficient corresponding to the form.

Rearranging the form in eq. (4.8.5), we obtain

$$\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*), \cdots, B_\epsilon(x_K^*)) \\
&= \sum_{\mathbf{r}}\sum_{\mathbf{q}}\left\{\prod_{i=1}^{d}\prod_{j=1}^{K}\left(z_i a_i H_i(B_\epsilon(x_j^*))\right)^{r_{i,j}} + \right. \\
&\qquad \left. \Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|, \cdots, |r_{\bullet,K}|) a_0^{|\mathbf{r}|}\left(\prod_{i=1}^{d}(1-z_i)^{|r_{i,\bullet}|}\right)\left(\prod_{j=1}^{K} H(B_\epsilon(x_j^*))\right) + \mathcal{R}^*\right\} \times \\
&\qquad \prod_{i=1}^{d}\prod_{j=1}^{k}\frac{1}{r_{i,j}!}\binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{i,j}}}\Psi_i^{(N_i)}(\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,k}) \, .
\end{aligned}$$

Letting $\epsilon \to 0$, one has

$$\begin{aligned}
&\int_{\mathbb{X}^K} \mathcal{R}^* = 0 \, , \\
&\int_{\mathbb{X}^K} \prod_{i=1}^{d}\prod_{j=1}^{K}\left(z_i a_i H_i(dx_j^*)\right)^{r_{i,j}} = \left(\prod_{i=1}^{d}(z_i a_i)^{|r_{i,\bullet}|}\right)\left(\prod_{j=1}^{K} H(dx_j^*) C(r_{1,j}, \cdots, r_{d,j})\right) .
\end{aligned}$$

Thus, we have

$$
\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(dx_1^*, \cdots, dx_K^*) \\
&= \sum_{\mathbf{r}} \sum_{\mathbf{q}} \Bigg[\left(\prod_{i=1}^{d} (z_i a_i)^{|r_{i,\bullet}|} \right) \left(\prod_{j=1}^{K} H(dx_j^*) C(r_{1,j}, \cdots, r_{d,j}) \right) + \\
&\quad \Psi^{(|\mathbf{r}|)}(|r_{\bullet,1}|, \cdots, |r_{\bullet,K}|) a_0^{|\mathbf{r}|} \left(\prod_{i=1}^{d} (1 - z_i)^{|r_{i,\bullet}|} \right) \left(\prod_{j=1}^{K} H(B_\epsilon(x_j^*)) \right) \Bigg] \\
&\quad \times \prod_{i=1}^{d} \prod_{j=1}^{K} \frac{1}{r_{i,j}!} \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{i,j}}} \Psi_i^{(N_i)}(\mathbf{q}_{i,1}, \cdots, \mathbf{q}_{i,K}).
\end{aligned}
$$

The result then follows. □

Proof of Theorem 4.5.3

Proof. The proof follows the similar steps in Section 4.8. For the sake of notational simplicity, we denote $B_j = B_\epsilon(x_j^*)$, which is defined in Section 4.8. And let $\mathbf{u} = (u_1, \cdots, u_d)$. By noticing that $\{\mu_i(B)\}_{i=1}^d$ are no longer independent conditional on H_0 and $\{z_i\}_{i=1}^d$ for any $B \in \mathcal{X}$, we can have the following decomposition similar to that in eq. (4.8.1).

$$
\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(B_\epsilon(x_1^*), \cdots, B_\epsilon(x_K^*)) \\
&= \mathbb{E}\left[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^{d} \frac{u_i^{N_i - 1}}{\Gamma(N_i)} \right) \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(\mathbb{X})} \prod_{i=1}^{d} \prod_{j=1}^{K} \mu_i(B_j)^{n_{i,j}} | H_0 \right] d\mathbf{u} \right] \\
&= \mathbb{E}\Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^{d} \frac{u_i^{N_i - 1}}{\Gamma(N_i)} \right) \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(B^*)} | H_0 \right] \times \\
&\quad \prod_{j=1}^{K} \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(B_j)} \prod_{i=1}^{d} \mu_i(B_j)^{n_{i,j}} | H_0 \right] d\mathbf{u} \Bigg] \\
&= \mathbb{E}\Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^{d} \frac{u_i^{N_i - 1}}{\Gamma(N_i)} \right) \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(B^*)} | H_0 \right] \times \\
&\quad \prod_{j=1}^{K} (-1)^{n_{\bullet,j}} \frac{d^{n_{\bullet,j}}}{\prod_{i=1}^{d} du_i^{n_{i,j}}} \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(B_j)} | H_0 \right] d\mathbf{u} \Bigg]. \qquad (4.8.6)
\end{aligned}
$$

The joint Laplace transform of $\{\mu_i\}_{i=1}^d$ and the multivariate Faa di Bruno formula (Constantine and Savits, 1996) lead to

$$
\begin{aligned}
&(-1)^{n_{\bullet,j}} \frac{d^{n_{\bullet,j}}}{\prod_{i=1}^d du_i^{n_{i,j}}} \mathbb{E}\left[e^{-\sum_{i=1}^d u_i \mu_i(B_j)} | H_0 \right] \\
&= (-1)^{n_{\bullet,j}} \frac{d^{n_{\bullet,j}}}{\prod_{i=1}^d du_i^{n_{i,j}}} e^{-\sum_{i=1}^d z_i a_i H_i(B_j)\phi_i(u_i) - z_0 a_0 H_0(B_j)\phi_0(\sum_{i=1}^d u_i)} \\
&= e^{-\eta(B_j, \mathbf{u})} \sum_{r_{max,j}=1}^{n_{\bullet,j}} \sum_{\mathbf{q}_{\bullet,j}} \left\{ \frac{1}{r_{\max,j}!} \prod_{i=1}^d \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}}} \right\} \times \\
&\quad \frac{d^{\sum_{i=1}^d q_{i,j,1}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,1}}} \cdots \frac{d^{\sum_{i=1}^d q_{i,j,r_{\max,j}}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,r_{\max,j}}}}, \qquad (4.8.7)
\end{aligned}
$$

where the sum of $\mathbf{q}_{\bullet,j}$ is taking over all vectors $\mathbf{q}_{\bullet,j}$ such that

(1) $\mathbf{q}_{\bullet,j} = (q_{1,j,1}, \cdots, q_{1,j,r_{\max,j}}, \cdots, q_{d,j,1}, \cdots, q_{d,j,r_{\max,j}})$ is a vector of nonnegative integers such that $\sum_{t_j=1}^{r_{\max,j}} q_{i,j,t_j} = n_{i,j}$. Note that $\sum_{i=1}^d q_{i,j,t_j} \geq 1$ for any j and $t_j \in \{1, \cdots, r_{\max,j}\}$.

The product of the partial derivatives in eq. (4.8.7) can be computed in the following two cases.

Case I: For any $j \in \{1, \cdots, K\}$, if there exists at least one $t_j \in \{1, \cdots, r_{\max,j}\}$ such that $\sum_{i=1}^d \mathbb{1}_{\{r_{i,j,t_j}>0\}} > 1$, then

$$
\begin{aligned}
&\frac{d^{\sum_{i=1}^d q_{i,j,1}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,1}}} \cdots \frac{d^{\sum_{i=1}^d q_{i,j,r_{\max,j}}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,r_{\max,j}}}} \\
&= (z_0 a_0 H_0(B_j))^{r_{\max,j}} \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}}(\sum_{i=1}^d u_i) + \mathcal{R}_I^*(B_j),
\end{aligned}
$$

where $\mathcal{R}_I^*(B_j)$ is a linear combination of the form $H_i(B_j)^{k_i} H_0(B_j)^{k_0}$ for $k_i > 0$, $k_0 > 0$.

Case II: All the cases except the Case I.

$$
\begin{aligned}
&\frac{d^{\sum_{i=1}^d q_{i,j,1}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,1}}} \cdots \frac{d^{\sum_{i=1}^d q_{i,j,r_{\max,j}}} \eta(B_j, \mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,r_{\max,j}}}} \\
&= (z_0 a_0 H_0(B_j))^{r_{\max,j}} \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}}(\sum_{i=1}^d u_i) + \prod_{i=1}^d (z_i a_i H_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|} \prod_{t_j=1}^{r_{\max,j}} \tau_{i,q_{i,j,t_j}}(u_i) +
\end{aligned}
$$

$$\mathcal{R}^*_{II}(B_j)\,,$$

where $\mathcal{R}^*_{II}(B_j)$ has the similar form to that of $\mathcal{R}^*_I(B_j)$ and $|\overrightarrow{\mathbf{q}_{i,j}}|$ is the length of the vector $(q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}})$.

Overall, by the analysis of Case I and Case II, we can have a general form of the partial derivatives in eq. (4.8.7).

$$\frac{d^{\sum_{i=1}^d q_{i,j,1}}\eta(B_j,\mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,1}}} \cdots \frac{d^{\sum_{i=1}^d q_{i,j,r_{\max,j}}}\eta(B_j,\mathbf{u})}{\prod_{i=1}^d du_i^{q_{i,j,r_{\max,j}}}}$$
$$= (z_0 a_0 H_0(B_j))^{r_{\max,j}} \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}}(\sum_{i=1}^d u_i) +$$
$$\left(\prod_{t_j=1}^{r_{\max,j}} \mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}}\right) \prod_{i=1}^d (z_i a_i H_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|} \prod_{t_j=1}^{r_{\max,j}} \tau_{i,q_{i,j,t_j}}(u_i) + \mathcal{R}^*(B_j)\,,$$

where $|\overrightarrow{\mathbf{q}_{j,t_j}}|$ is the length of the vector $(q_{1,j,t_j}, \cdots, r_{d,j,t_j})$. And $\mathcal{R}^*(B_j)$ is either $\mathcal{R}^*_I(B_j)$ or $\mathcal{R}^*_{II}(B_j)$.

Combining the above form and eq. (4.8.7), we have

$$\mathcal{I}_{\mathbf{n}}(B_1, \cdots, B_K)$$
$$= \mathbb{E}\Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^d \frac{u_i^{N_i-1}}{\Gamma(N_i)}\right) e^{-\sum_{i=1}^d z_i a_i \phi_i(u_i) - z_0 a_0 \phi_0\left(\sum_{i=1}^d u_i\right)} \times$$
$$\prod_{j=1}^K \sum_{\mathbf{r}_{\bullet,j}} \sum_{\mathbf{q}_{\bullet,j}} \left\{\frac{1}{r_{\max,j}!} \prod_{i=1}^d \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}}}\right\} \times$$
$$\Bigg[(z_0 a_0 H_0(B_j))^{r_{\max,j}} \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}}(\sum_{i=1}^d u_i) +$$
$$\left(\prod_{t_j=1}^{r_{\max,j}} \mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}}\right) \prod_{i=1}^d (z_i a_i H_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|} \prod_{t_j=1}^{r_{\max,j}} \tau_{i,q_{i,j,t_j}}(u_i) + \mathcal{R}^*(B_j)\Bigg] d\mathbf{u}\Bigg]. \quad (4.8.8)$$

By a rearrangement, we have

$$\mathcal{I}_{\mathbf{n}}(B_1, \cdots, B_K)$$

$$
\begin{aligned}
&= \sum_{\mathbf{r}} \sum_{\mathbf{q}} \Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^d \frac{u_i^{N_i-1}}{\Gamma(N_i)} \right) e^{-\sum_{i=1}^d z_i a_i \phi_i(u_i) - z_0 a_0 \phi_0 \left(\sum_{i=1}^d u_i\right)} \times \\
&\quad \left\{ \prod_{j=1}^K \frac{1}{r_{\max,j}!} \prod_{i=1}^d \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}}} \right\} \times \\
&\quad \Bigg[\mathbb{E} \left\{ \prod_{j=1}^K H_0(B_j)^{r_{\max,j}} \right\} (z_0 a_0)^{r_{\max}} \prod_{j=1}^K \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}} (\sum_{i=1}^d u_i) + \\
&\quad \prod_{j=1}^K \left(\prod_{t_j=1}^{r_{\max,j}} \mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}} \right) \prod_{i=1}^d (z_i a_i H_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|} \prod_{t_j=1}^{r_{\max,j}} \tau_{i,q_{i,j,t_j}}(u_i) + \mathcal{R}^*(B_1, \cdots, B_K) \Bigg] d\mathbf{u} \Bigg],
\end{aligned} \tag{4.8.9}
$$

where $\mathcal{R}^*(B_1, \cdots, B_K)$ is a linear combination of $\prod_{i=1}^d \prod_{j=1}^K H_i(B_j)^{k_{i,j}}$ $H_0(B_j)_{k_{0,j}}$ such that there exists at least one group of (i, j) satisfying $k_{i,j} > 0$ and $k_{0,j} > 0$.

The same analysis as that of eq. (4.8.4) gives when $\epsilon \to 0$,

$$
\begin{aligned}
&\mathcal{I}_{\mathbf{n}}(dx_1^*, \cdots, dx_K^*) \\
&\quad = \sum_{\mathbf{r}} \sum_{\mathbf{q}} \Bigg[\int_0^\infty \cdots \int_0^\infty \left(\prod_{i=1}^d \frac{u_i^{N_i-1}}{\Gamma(N_i)} \right) e^{-\sum_{i=1}^d z_i a_i \phi_i(u_i) - z_0 a_0 \phi_0 \left(\sum_{i=1}^d u_i\right)} \times \\
&\quad \left\{ \prod_{j=1}^K \frac{1}{r_{\max,j}!} \prod_{i=1}^d \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,r_{\max,j}}} \right\} \left(\prod_{j=1}^K H(dx_j^*) \right) \times \\
&\quad \Bigg[\Psi^{(r_{\max})}(r_{\max,1}, \cdots, r_{\max,K}) \, (z_0 a_0)^{r_{\max}} \prod_{j=1}^K \prod_{t_j=1}^{r_{\max,j}} \tau_{0,\sum_{i=1}^d q_{i,j,t_j}} (\sum_{i=1}^d u_i) + \\
&\quad \prod_{j=1}^K C(r_{1,j}, \cdots, r_{d,j}) \left(\prod_{t_j=1}^{r_{\max,j}} \mathbb{1}_{\{|\overrightarrow{\mathbf{q}_{j,t_j}}|=1\}} \right) \prod_{i=1}^d (z_i a_i H_i(B_j))^{|\overrightarrow{\mathbf{q}_{i,j}}|} \prod_{t_j=1}^{r_{\max,j}} \tau_{i,q_{i,j,t_j}}(u_i) \Bigg] d\mathbf{u} \Bigg].
\end{aligned} \tag{4.8.10}
$$

The result follows easily. □

Proof of Theorem 4.6.1

Proof. The proof is similar to that of Theorem 5 in (Camerlenghi et al., 2019). In our constructions, we have

$$\mathbb{P}\left(K_N = k\right) = \mathbb{P}\left(\cup\{K^{(1)}_{1,N_1} = t_1, \cdots, K^{(1)}_{d,N_d} = t_d, K^{(2)}_{1,N_1} = \eta_1, \cdots, K^{(2)}_{d,N_d} = \eta_d, K_{0,t} = k_0\}\right),$$

where the union is taking over all combinations of nonnegative integers $\{t_1, \cdots, t_d, \eta_1, \cdots, \eta_d\}$ and positive integers $\{k_0, t\}$ such that

(i) $\sum_{i=1}^{d} t_i = t$, $k_0 + \sum_{i=1}^{d} \eta_i = k$;

(ii) $t_i + \eta_i \in \{1, \cdots, N_i\}$ for all $i \in \{1, \cdots, d\}$.

The result in this theorem follows immediately. As for the case in model 4.3.1, the result is implied by the independence of $(K^{(1)}_{i,N_i}, K^{(2)}_{i,N_i})$ and $(K^{(1)}_{l,N_l}, K^{(2)}_{l,N_l})$ for $i \neq l$. □

Bibliography

Aldous, D. J., I. A. Ibragimov, J. Jacod, and D. J. Aldous (1985). *Exchangeability and related topics*. Springer.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 1152–1174.

Aomoto, K., M. Kita, T. Kohno, and K. Iohara (2011). *Theory of hypergeometric functions*. Tokyo: Springer.

Bickel, P. J. and M. J. Wichura (1971, oct). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics 42*(5), 1656–1670.

Billingsley, P. (1995). *Probability and measure* (Third ed.). Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.

Bingham, N. H., C. M. Goldie, and J. L. Teugels (1987). *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press.

Brix, A. (1999, dec). Generalized Gamma measures and shot-noise Cox processes. *Advances in Applied Probability 31*(04), 929–953.

Brunner, L. J. and A. Y. Lo (1996). Limiting posterior distributions under mixture of conjugate priors. *Statistica Sinica 6*(1), 187–197.

Camerlenghi, F., A. Lijoi, P. Orbanz, and I. Prünster (2019). Distribution theory for hierarchical processes. *The Annals of Statistics 47*(1), 67–92.

Constantine, G. M. and T. H. Savits (1996). A multivariate Faa di Bruno formula with applications. *Transactions of the American Mathematical Society 348*, 503–520.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B (Methodological) 17*(2), 129–157.

Daley, D. and D. Vere-Jones (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and structure* (Second ed.). Probability and its Applications. Springer, New York.

Dawson, D. A. and S. Feng (2006, may). Asymptotic behavior of the Poisson–Dirichlet distribution for large mutation rate. *The Annals of Applied Probability 16*(2), 562–582.

De Blasi, P., S. Favaro, A. Lijoi, R. H. Mena, I. Prünster, and M. Ruggiero (2013). Are gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE transactions on pattern analysis and machine intelligence 37*(2), 212–229.

De Blasi, P., A. Lijoi, and I. Pruenster (2013). An asymptotic analysis of a class of discrete nonparametric priors. *Statistica Sinica 23*(3), 1299–1322.

Dykstra, R. and P. Laud (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics 9*(2), 356–367.

Engen, S. (1978). *Abundance Models with Emphasis on Biological Communities and Species Diversity*. Chapman-Hall, London.

Ewens, W. J. (2004). *Mathematical population genetics: theoretical introduction*, Volume 27. Springer.

Exton, H. (1976). *Multiple hypergeometric functions and applications.* Ellis Horwood: Chichester.

Favaro, S., A. Lijoi, C. Nava, B. Nipoti, I. Prünster, and Y. W. Teh (2016, sep). On the stick-breaking representation for homogeneous NRMIs. *Bayesian Analysis 11*(3), 697–724.

Favaro, S., A. Lijoi, and I. Prünster (2012, jun). On the stick-breaking representation of normalized inverse Gaussian priors. *Biometrika 99*(3), 663–674.

Favaro, S. and Z. Naulet (2021). Near-optimal estimation of the unseen under regularly varying tail populations. *arXiv preprint arXiv:2104.03251*.

Feng, S. (2010). *The Poisson-Dirichlet distribution and related topics: models and asymptotic behaviors*. Springer Science & Business Media.

Ferguson, T. S. (1973, mar). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*(2), 209–230.

Franssen, S. and A. van der Vaart (2022). Bernstein-von Mises theorem for the Pitman-Yor process of nonnegative type. *Electronic Journal of Statistics 16*(2), 5779–5811.

Freedman, D. and P. Diaconis (1983). On inconsistent Bayes estimates in the discrete case. *The Annals of Statistics*, 1109–1118.

Gasthaus, J. and Y. Teh (2010). Improvements to the sequence memoizer. *Advances in Neural Information Processing Systems 23*.

Ghosal, S. and A. Van der Vaart (2017). *Fundamentals of nonparametric Bayesian inference* (First ed.), Volume 44. Cambridge: Cambridge University Press.

Gnedin, A. and J. Pitman (2006). Exchangeable Gibbs partitions and stirling triangles. *Journal of Mathematical sciences 138*, 5674–5685.

Gradshteyn, I. S. and I. M. Ryzhik (2014). *Table of integrals, series, and products* (eight ed.). New York: Academic press.

Griffiths, R. C. (1980). Unpublished notes.

Haan, L. and A. Ferreira (2006). *Extreme value theory: an introduction*, Volume 3. Springer.

Hardy, M. (2006). Combinatorics of partial derivatives. *the electronic journal of combinatorics 13*(R1), 1.

Ho Jang, G., J. Lee, and S. Lee (2010). Posterior consistency of species sampling priors. *Statistica Sinica 20*(2), 581–593.

Hu, Y. (2017). *Analysis on Gaussian spaces.* World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ.

Hu, Y. and J. Zhang (2022). Functional central limit theorems for stick-breaking priors. *Bayesian Analysis 17*(4), 1101–1120.

Isserlis, L. (1918, nov). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika 12*(1-2), 134–139.

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. *arXiv preprint math/0205093*.

James, L. F. (2005). Bayesian poisson process partition calculus with an application to bayesian lévy moving averages. *The Annals of Statistics 33*(4), 1771–1799.

James, L. F. (2008). Large sample asymptotics for the two-parameter Poisson–Dirichlet process. In *In Clarke, B. and Ghosal, S. (eds.), Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh,*, Volume 3, Cambridge, MA, pp. 187–199. MIT Press.

James, L. F., A. Lijoi, and I. Prünster (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scandinavian Journal of Statistics 33*(1), 105–120.

James, L. F., A. Lijoi, and I. Prünster (2009, jul). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics 36*(1), 76–97.

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles*, Volume 9. Springer.

Karlin, S. (1967). Central limit theorems for certain infinite urn schemes. *Journal of Mathematics and Mechanics 17*(4), 373–401.

Kim, N. and P. J. Bickel (2003). The limit distribution of a test statistic for bivariate normality. *Statistica Sinica 13*, 327–349.

Kim, Y. and J. Lee (2004, aug). A Bernstein–von Mises theorem in the nonparametric right-censoring model. *The Annals of Statistics 32*(4), 1492–1512.

Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics 21*(1), 59–78.

Kingman, J. F. (1975). Random discrete distributions. *Journal of the Royal Statistical Society: Series B (Methodological) 37*(1), 1–15.

Kingman, J. F. C. (1982). The coalescent. *Stochastic processes and their applications 13*(3), 235–248.

Kingman, J. F. C. (1993). *Poisson processes*, Volume 3 of *Oxford Studies in Probability.* Clarendon Press, Oxford University Press, New York.

Korwar, R. M. and M. Hollander (1973, aug). Contributions to the theory of Dirichlet processes. *The Annals of Probability 1*(4).

Labadi, L. A. and I. Abdelrazeq (2016, aug). On functional central limit theorems of Bayesian nonparametric priors. *Statistical Methods & Applications 26*(2), 215–229.

Labadi, L. A. and M. Zarepour (2013, sep). On asymptotic properties and almost sure approximation of the normalized inverse-Gaussian process. *Bayesian Analysis 8*(3), 553–568.

Lijoi, A., R. H. Mena, and I. Prünster (2005a, dec). Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Statistical Inference for Stochastic Processes 8*(3), 283–309.

Lijoi, A., R. H. Mena, and I. Prünster (2005b, dec). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association 100*(472), 1278–1291.

Lijoi, A., R. H. Mena, and I. Prünster (2007, sep). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(4), 715–740.

Lijoi, A., I. Prünster, et al. (2003). On a normalized random measure with independent increments relevant to Bayesian nonparametric inference. In *13th European Young Statisticians Meeting*, pp. 123–134. Staempfli.

Lijoi, A., I. Prünster, et al. (2010). Models beyond the Dirichlet process. *Bayesian nonparametrics 28*(80), 342.

Lijoi, A., I. Prünster, and S. G. Walker (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *Journal of the American Statistical Association 100*(472), 1292–1296.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 59*(1), 55–66.

Lo, A. Y. (1983). Weak convergence for Dirichlet processes. *Sankhyā: The Indian Journal of Statistics, Series A*, 105–111.

Lo, A. Y. (1986). A remark on the limiting posterior distribution of the multiparameter Dirichlet process. *Sankhyā: The Indian Journal of Statistics, Series A*, 247–249.

Lo, A. Y. (1987, mar). A large sample study of the Bayesian bootstrap. *The Annals of Statistics 15*(1), 360–375.

MacEachern, S. N. (1999). Dependent nonparametric processes. Alexandria, VA. Amereican Statistical Association.

MacEachern, S. N. (2000). Dependent dirichlet processes. Columbus, OH. Technical Report, Department of Statistics, Ohio State University.

McCloskey, J. (1965). *A Model for the Distribution of Individuals by Species in an Environment, unpublished Ph. D.* Ph. D. thesis, thesis, Michigan State University.

Müller, P. and F. A. Quintana (2004). Nonparametric Bayesian data analysis. *Statistical Science 19*(1), 95–110.

NGUYEN, X. (2016). Borrowing strengh in hierarchical bayes: Posterior concentration of the dirichlet base measure. *Bernoulli*, 1535–1571.

Paisley, J. W., A. K. Zaas, C. W. Woods, G. S. Ginsburg, and L. Carin (2010). A stick-breaking construction of the Beta process. In *In International Conference on Machine Learning*, Haifa, Israel.

Patrick, B. (1995). *Probability and measure* (Third ed.). New York: John Wiley & Sons.

Perman, M., J. Pitman, and M. Yor (1992, mar). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields 92*(1), 21–39.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory, Lecture Notes-Monograph Series*, Volume 30, Hayward, CA, pp. 245–267.

Pitman, J. (2003). Poisson-Kingman partitions. In *Statistics and Science: a Festschrift for Terry Speed, Lecture Notes-Monograph Series*, Volume 40, Beachwood, OH, pp. 1–34.

Pitman, J. (2006). *Combinatorial stochastic processes: Ecole d'eté de probabilités de saint-flour xxxii-2002*. Springer.

Pitman, J. and M. Yor (1997, apr). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability 25*(2), 855–900.

Præstgaard, J. and J. A. Wellner (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, 2053–2086.

Quintana, F. A., P. Müller, A. Jara, and S. N. MacEachern (2022). The dependent dirichlet process and related models. *Statistical Science 37*(1), 24–41.

Ray, K. and A. van der Vaart (2021). On the Bernstein-von Mises theorem for the Dirichlet process. *Electronic Journal of Statistics 15*, 2224–2246.

Regazzini, E., A. Guglielmi, G. Di Nunno, et al. (2002). Theory and numerical analysis for exact distributions of functionals of a Dirichlet process. *The Annals of Statistics 30*(5), 1376–1411.

Regazzini, E., A. Lijoi, and I. Prünster (2003, apr). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics 31*(2), 560–585.

Sethuraman, J. (1994, may). A constructive definition of Dirichlet priors. *Statistica sinica*, 639–650.

Sethuraman, J. and R. C. Tiwari (1982). Convergence of Dirichlet measures and the interpretation of their parameter. In *Statistical decision theory and related topics III*, pp. 305–315. Elsevier.

Straf, M. L. (1972). Weak convergence of stochastic processes with several parameters. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2, pp. 187–221.

Teh, Y., M. Jordan, M. Beal, and D. Blei (2004). Sharing clusters among related groups: Hierarchical dirichlet processes. *Advances in neural information processing systems 17.*

Teh, Y. W., D. Grür, and Z. Ghahramani (2007). Stick-breaking construction for the Indian buffet process. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics, 11*, San Juan, Puerto Rico.

Teh, Y. W. and M. I. Jordan (2010). Hierarchical bayesian nonparametric models with applications. *Bayesian nonparametrics 1*, 158–207.

Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association 101*(476), 1566–1581.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* New York: Springer.

Zeng, C. and L. L. Duan (2020). Quasi-Bernoulli stick-breaking: Infinite mixture with cluster consistency. *arXiv preprint arXiv:2008.09938*.

Zhang, J. and Y. Hu (2021). Dirichlet process and Bayesian nonparametric models (in chinese). *SCIENTIA SINICA Mathematica 51*(11), 1895–1932.