University of Pavia

DEPARTMENT OF MATHEMATICS "FELICE CASORATI"
Joint Ph.D. program in Mathematics Milano–Bicocca, Pavia, INdAM

Ph.D. Thesis

Novel asymptotical results for Ewens–Pitman partitions, with statistical applications

Candidate:
Claudia Contardi

Supervisor:
Prof. Emanuele Dolera

Academic Year 2024/25

To my family, old and new.

Contents

Preface

This (very brief) chapter is not intended as a scientific introduction to the work, as much as an account of the research projects I have carried out during the three years of my Ph.D. and the respective stages of development they have reached.

The main topic of my doctorate research has been the study of asymptotic properties of the Ewens–Pitman random partitions in the "large θ" regime, and the application of such properties to species sampling problems in Bayesian statistics. Overall, this is the research project that has reached the more advanced state, leading to the drafting of three articles, one of which published, which constitute the basis for all material in this thesis. The work has been carried out in collaboration with professors Emanuele Dolera and Stefano Favaro, and with professor Bernard Bercu during my research visit at Bordeaux University.

A secondary project, which I have begun while visiting professor Giovanni Peccati at the University of Luxembourg, aims at establishing conditions for a quantitative central limit theorem for the second chaos of the Dirichlet–Ferguson process. While the results obtained so far are promising, and collaboration is ongoing, this project has not yet reached a stage suitable for publication.

The last project, started in collaboration with professors Emanuele Dolera and Stefano Favaro, regards the possibility of extending the "Bayes Empirical Bayes" framework of Deely and Lindley to the nonparametric setting, establishing Bayesian consistency with posterior contraction rates and quantifying the efficiency of a fully Bayes version of Robbins' Empirical Bayes model. Similarly to the previous topic, while we have been able to prove some interesting results, the project is not yet fully mature.

I wish to express my outmost gratitude to professors Emanuele Dolera and Stefano Favaro for their constant support and guidance during my Ph.D., and to professors Giovanni Peccati and Bernard Bercu for welcoming me in their departments and spending time to initiate and discuss engaging new projects with me. Finally, I wish to thank professors Sandra Fortini and Shui Feng, whose reviews and comments have significantly improved this work.

Abstract

This thesis investigates the large-sample behavior of the Ewens–Pitman model for random partitions under a non-standard asymptotic regime in which the concentration parameter θ grows linearly with the sample size.

The Ewens-Pitman model for random partitions first appeared in [79] as a two-parameter generalization of the Ewens model in population genetics [33]. The model is indexed by two parameters $\alpha \in [0, 1)$ and $\theta > -\alpha$, with the Ewens model corresponding to the case $\alpha = 0$. The large sample asymptotic properties of the partition for fixed parameters are well understood, with the number of blocks and the numbers of blocks of fixed numerosity exhibiting different scalings depending on the value of α. More recently, motivated by its significance in population genetics in the case $\alpha = 0$, different asymptotical regimes have been considered, in which the parameter θ scales with the sample size. We focus in particular on the so-called "large θ" regime, where $\theta = \lambda n$ for some $\lambda > 0$. While asymptotic in this regime is reasonably well understood in the case $\alpha = 0$, much less is known for the two–parameter generalization. This thesis contributes to filling this gap by providing novel asymptotical results, in the form of laws of large numbers and central limit theorems, in the general case $\alpha \in [0, 1)$.

More specifically, after a general introduction to the model and a review of existing literature, a strong law of large numbers and a Gaussian central limit theorem are established for the number K_n of blocks in the partition when $\theta = \lambda n$ for all $\alpha \in [0, 1)$, extending the analogous result established in [92] for $\alpha = 0$. These theorems show that K_n scales linearly in n for all $\alpha \in [0, 1)$ and exhibits deterministic limits, a behaviour notably different than the one observed in the standard asymptotic regime. Our proof relies on different arguments depending on whether $\alpha = 0$ or $\alpha \in (0, 1)$. In the Ewens case ($\alpha = 0$), K_n admits a representation as a sum of independent but non-identically distributed Bernoulli random variables, which also allows to refine the central limit theorem to a Berry–Esseen theorem. Instead, for $\alpha \in (0, 1)$, the analysis exploits a compound Poisson construction of K_n, leading to prove a LLN, a CLT and a Berry-Esseen theorem for the number of blocks of the negative-Binomial compound Poisson random partition, which are of independent interest.

An alternative proof is developed in the second part of the thesis, where martingale techniques are employed to re-derive the LLN and the CLT for K_n. A generalization of this approach further enables to prove a strong LLN and a joint CLT for the $d-$dimensional vectors $(K_{1,n}, ..., K_{d,n})$ containing the number of blocks of fixed numerosity r, for $r \in \{1, ...d\}$, thereby providing a finer description of the partition structure in the large θ regime.

The final part of the thesis addresses statistical implications of these asymptotic results within a Bayesian nonparametric (BNP) framework. Focusing on the classical unseen-

species problem first addressed in [47], a Gaussian central limit theorem is derived for the posterior distribution of the number of unseen species under a Pitman–Yor prior. This result is then used to construct Gaussian asymptotic credible intervals for the BNP estimator of the number of new unseen species. Our method improves upon competitors in two key aspects: firstly, it enables the full parameterization of the Pitman-Yor prior, including the Dirichlet prior; secondly, it is fully analytical, enhancing computational efficiency. We validate the proposed method on synthetic and real data, demonstrating that it also improves the empirical performance of competitors in terms of coverage of the exact intervals.

Finally, possible generalizations of the results obtained in the thesis are discussed, such as the possibility to consider further non–standard asymptotic regimes, or to extend the asymptotic analysis to other variables of interest beyond K_n and $K_{r,n}$. Overall, the thesis advances the theoretical understanding of Ewens–Pitman random partitions under non-standard asymptotic scalings and illustrates how such probabilistic insights can be applied to statistical problems, especially in the field of Bayesian nonparametrics.

Chapter 1

Introduction

1.1 Organization of the thesis

This thesis presents some novel results for the large sample behaviour of the Ewens–Pitman random partition under a non–standard asymptotical scaling of the parameter θ, referred to as the "large θ" regime. It is organized as follows:

- Chapter 1 is dedicated to a general introduction to the Ewens–Pitman model for random partitions, with a focus on the asymptotic properties of some quantities of interest. Further, it contains a review of existing literature regarding asymptotics in the regime $\theta = \lambda n$, which is the focus of the original work carried out in the rest of the thesis.

- Chapter 2 establishes a law of large numbers (LLN) and a central limit theorem (CLT) for K_n, the number of blocks in the Ewens–Pitma random partition, in the regime $\theta = \lambda n$. Depending on whether $\alpha = 0$ or $\alpha \in (0, 1)$, our results rely on different arguments. For $\alpha = 0$ they rely on the representation of K_n as a sum of independent, but not identically distributed, Bernoulli random variables, leading to a refinement of the CLT in terms of a Berry-Esseen theorem. Instead, for $\alpha \in (0, 1)$, they rely on a compound Poisson construction of K_n. Technical details for this chapter are deferred to appendices A, B and C.

  All material in this chapter is based on work [21].

- Chapter 3 contains an alternative proof of the CLT of chapter 2 based on a martingale argument. An extension of the same argument allows to prove a joint CLT for the $d-$dimensional vector $(K_{1,n}, ..., K_{d,n})$ of the number of blocks with given numerosity r, for $r \in \{1, ..., d\}$. Technical details are deferred to appendix D

  All material in this chapter is based on work [8].

- Chapter 4 presents an analogous of the central limit theorem of chapter 2 for the posterior distribution of K_n, and shows an application of such theorem to the problem of uncertainty quantification in the Bayesian nonparametric (BNP) approach to the so–called "unseen species problem". For the sake of readability, this chapter is self–contained, including an introduction to the statistical problem at hand and to the BNP approach to its solution. Technical details are deferred to appendices E, F and G.

  All material in this chapter is based on work [22], which has been submitted for publication.

- Chapter 5 contains a very brief discussion of the results of the previous chapters and presents some directions for future work.

1.2 The Ewens–Pitman model for random partitions

The Ewens-Pitman model for random partitions first appeared in [79] as a two-parameter generalization of the Ewens model in population genetics [33]; see [23] and references therein. It now plays a critical role in a variety of research areas, e.g., population genetics, Bayesian statistics, combinatorics, machine learning and statistical physics. See [80, Chapter 3] for an overview of the Ewens-Pitman model and generalizations thereof.

1.2.1 Exchangeable random partitions

This section is meant to give a brief and discursive overview of the theory of random partitions, introducing the main properties of interest. For a detailed and rigorous theory, see [64] and [80, Chapter 2].

Let $[n] = \{1, ..., n\}$. We call a *random partition* any random variable Π_n taking values in the set of partitions of $[n]$. Further, we say that Π_n is *exchangeable* if or every $\sigma \in S_n$,

$$P[\Pi_n = \pi_n] = P[\Pi_n = \sigma(\pi_n)],$$

where $\sigma(\pi_n)$ is the partition obtained by permuting the members of clusters in π_n by σ. This is tantamount to saying that $P[\Pi_n = \pi_n]$ only depends on the number k of blocks and the sizes $(n_1, ..., n_k)$ of the blocks in π_n; in particular

$$P[\Pi_n = \pi_n] = f_n(n_1, ..., n_k) \tag{1.2.1}$$

with f_n a symmetric function of its arguments, defined on $\{(x_1, ..., x_k) \in \mathbb{N}^k : x_i \geq 1 \text{ for all } i, \sum_{j=1}^k n_j = n\}$.

Now consider a sequence of random partitions $\Pi_1, \Pi_2, ...$ such that Π_i is a random partition of $[i]$. We call it *projective* or *extendable* if for all $n \geq m$ and all π_m partitions of $[m]$,

$$P[I_{[m]}\Pi_n = \pi_m] = P[\Pi_{[m]} = \pi_m]$$

where $I_{[m]}\Pi_n$ denotes the partition of $[m]$ obtained by removing all the members of clusters in Π_n which are in $[n] \setminus [m]$. In particular, this means that if for all i we let

$$P[\Pi_i = \pi_i] = f_i(n_1, ..., n_k)$$

the f_i must satisfy the following compatibility relation:

$$f_n(n_1, ..., n_k) = f_{n+1}(n_1, ..., n_k, 1) + \sum_{j=1}^{k} f_{n+1}(n_1, ..., n_j + 1, ..., n_k) \tag{1.2.2}$$

We call any sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ satisfying (1.2.1) and (1.2.2) an *exchangeable partition probability function* (EPPF). Note that the compatibility formula 1.2.2 entails that a projective sequence of partitions can be constructed iteratively in the following way: suppose that after n iterations the partition has k clusters of sizes $n_1, ..., n_k$. At step $n + 1$, a new member is added into a new cluster with probability proportional to

$f_{n+1}(n_1, ..., n_k, 1)$, while it is added to the $j-$th cluster with probability proportional to $f_{n+1}(n_1, ..., n_j + 1, ..., n_k)$.

Finally, we say that a random partition is *generated by (sampling from) a random distribution* μ if given a sequence $(X_i)_{i\geq 1}$ such that

$$X_i \,|\, \mu \overset{i.i.d.}{\sim} \mu$$

we construct a partition of $[n]$ by the following equivalence relation: given $i, j \in [n]$,

$$i \sim j \iff X_i = X_j.$$

Clearly, since the X_i thus defined are exchangeable, the resulting partition is also exchangeable. A celebrated theorem of [64] states that every projective sequence of random partitions can be generated by sampling from a random distribution.

1.2.2 The sampling formula

For $n \in \mathbb{N}$, let Π_n be an (exchangeable) random partition of $[n]$ into $K_n \in \{1, \ldots, n\}$ blocks of sizes (or frequencies) $\mathbf{N}_n = (N_{1,n}, \ldots, N_{K_n,n}) \in \mathbb{N}^{K_n}$ such that $n = \sum_{1\leq i\leq K_n} N_{i,n}$. For $\alpha \in [0, 1)$ and $\theta > -\alpha$, the Ewens-Pitman model assigns to Π_n the probability

$$P[K_n = k, \mathbf{N}_n = (n_1, \ldots, n_k)] = \frac{1}{k!}\binom{n}{n_1, \ldots, n_k}\frac{(\theta)_{(k,\alpha)\uparrow}}{(\theta)_{n\uparrow}}\prod_{i=1}^{n}(1-\alpha)_{(n_i-1)\uparrow}, \tag{1.2.3}$$

where $(x)_{(n,a)\uparrow}$ denotes the rising factorial of x of order n and increment a, i.e. $(x)_{(n,a)\uparrow} := \prod_{0\leq i\leq n-1}(x + ia)$ and $(x)_{n\uparrow} := (x)_{(n,1)\uparrow}$.

It is easy to see that the function on the right-end side of (1.2.3) is an EPPF. The distribution (1.2.3) admits a sequential construction in terms of the Chinese restaurant process [79; 96] and a Poisson process construction by random sampling the two-parameter Poisson-Dirichlet distribution [77; 81]; see also [27; 28] for a construction through the negative-Binomial compound Poisson model for random partitions [17]. For $\alpha = 0$ the Ewens-Pitman model reduces to the Ewens model, arising by random sampling the Poisson-Dirichlet distribution [63]. We briefly introduce the sequential construction and the Poisson process construction in sections 1.2.3 and 1.2.4, respectively.

1.2.3 Sequential construction: the Chinese Restaurant Process

The sequential construction of the Ewens–Pitman model was introduced in the seminal work of [79, Proposition 9]. For $\alpha \in [0, 1)$ and $\theta > -\alpha$, it is possible to show that the following recursive construction yields an exchangeable random partition of the set $[n]$.

Conditionally on the number of blocks $K_n = k$ and on the partition subsets (the "tables", in the restaurant metaphor) $\{T_1, \ldots, T_k\}$ with corresponding sizes $(n_1, \ldots, n_k)$, the partition of $[n+1]$ is obtained by extending that of $[n]$ in such a way that the element $n + 1$ is assigned to an existing subset T_i, for $1 \leq i \leq k$, with probability

$$\frac{n_i - \alpha}{n + \theta},$$

or initiates a new subset with probability

$$\frac{\alpha k+\theta}{n+\theta}.$$

Since $n=n_1+\cdots+n_k$, we clearly have

$$\frac{1}{n+\theta}\sum_{i=1}^{k}(n_i-\alpha)+\frac{\alpha k+\theta}{n+\theta}=\frac{n-\alpha k+\alpha k+\theta}{n+\theta}=1.$$

From now on, for $K_n\in\{1,\ldots,n\}$, let

$$\mathbf{N}_n=(N_{1,n},\ldots,N_{K_n,n})$$

denote the sizes of the partition subsets $\{T_1,\ldots,T_{K_n}\}$. As shown in [79, Proposition 9], the above construction yields the joint distribution (1.2.3) of the random vector $(K_n,\mathbf{N}_n)$.

For $r=1,...,n$, we introduce the notation $K_{r,n}$ to denote the number of tables with given numerosity r. Note that the above construction immediately yields a recursive structure not only for K_n, but also for the $K_{r,n}$. In fact, letting

$$\mathbf{K}_n=(K_n,K_{1,n},...,K_{n,n})$$

and

$$\mathcal{F}_n=\sigma\left(\mathbf{K}_1,...,\mathbf{K}_n\right)$$

there hold

$$P(K_{n+1}=K_n+1\mid\mathcal{F}_n)=\frac{\alpha K_n+\theta}{n+\theta}.$$

and for $r\geq 1$,

$$P\left[K_{r,n+1}=K_{r,n}+k\,|\,\mathcal{F}_n\right]=\begin{cases}p_{r,n} & \text{if } k=1\\ q_{r,n} & \text{if } k=-1\\ 1-p_{r,n}-q_{r,n} & \text{if } k=0\end{cases}\tag{1.2.4}$$

with

$$p_{r,n}=\begin{cases}\frac{\alpha K_n+\theta}{\theta+n} & \text{if } r=1\\ \frac{(r-1-\alpha)K_{r-1,n}}{\theta+n} & \text{if } r\geq 2\end{cases}$$

and

$$q_{r,h-1}^{\{n\}}=\frac{(r-\alpha)K_{r,n}}{\theta+n}$$

1.2.4 Construction via sampling from the Pitman–Yor prior

A simple and intuitive definition of $P\sim\mathrm{PYP}(\alpha,\theta)$ follows from its stick-breaking construction [77]. Specifically, let: i) $(V_i)_{i\geq 1}$ be independent random variables, with each V_i following a Beta distribution with parameter $(1-\alpha,\theta+i\alpha)$; ii) $(S_j)_{j\geq 1}$ be random variables following a non-atomic distribution ν on $\mathbb{S}$ and independent of each other as well as of the V_i's. If $P_1=V_1$ and $P_j=V_j\prod_{1\leq i\leq j-1}(1-V_i)$ for $j\geq 1$, so that $P_j\in(0,1)$ for any $j\geq 1$ and $\sum_{j\geq 1}P_j=1$ almost surely, then $P=\sum_{j\geq 1}P_j\delta_{S_j}\sim\mathrm{PYP}(\alpha,\theta)$, with the Dirichlet prior corresponding to $\alpha=0$.

Remark 1.2.1. *If $(P_{(j)})_{j\geq 1}$ are the decreasingly ordered stick-breaking random probabilities P_j's of $P \sim PYP(\alpha,\theta)$, then, for $\alpha \in (0,1)$, as $j \to +\infty$ the $P_{(j)}$'s follow a power-law distribution of exponent $c = \alpha^{-1}$ [81]. The parameter $\alpha \in (0,1)$ controls the power-law tail of P through the small $P_{(j)}$'s: the larger α, the heavier the tail of P. As $\alpha \to 0$, the Dirichlet prior features geometric tails [80, Chapter 4].*

We consider $n \geq 1$ observations with values in the space of species' labels or symbols $\mathbb{S}$, modeled as random samples $\mathbf{X}_n = (X_1, \ldots, X_n)$ such that

$$\begin{aligned} X_1, \ldots, X_n \,|\, P &\overset{\text{iid}}{\sim} P, \\ P &\sim \text{PYP}(\alpha,\theta), \end{aligned} \tag{1.2.5}$$

where $\text{PYP}(\alpha,\theta)$ is the Pitman-Yor prior indexed by $\alpha \in [0,1)$ and $\theta > -\alpha$. For short, we say that $\mathbf{X}_n$ is a random sample from $\text{PYP}(\alpha,\theta)$. Due to the discreteness of the Pitman-Yor prior, the random sample $\mathbf{X}_n$ from $P \sim \text{PYP}(\alpha,\theta)$ induces a random partition of $[n] = \{1,\ldots,n\}$ into $K_n = j \leq n$ blocks, labelled by $\{S_1^*, \ldots, S_{K_n}^*\}$, with frequencies $(N_{1,n}, \ldots, N_{K_n,n}) = (n_1, \ldots, n_k)$ such that the n_i's are positive and $\sum_{1\leq i\leq j} n_i = n$. The distribution of the random partition is determined by the predictive distribution, or generative scheme, of the Pitman-Yor prior [79, Proposition 9], i.e.,

$$\Pr[X_1 \in \cdot] = \nu(\cdot),$$

and, for $n \geq 1$,

$$\Pr[X_{n+1} \in \cdot \,|\, \mathbf{X}_n] = \frac{\theta + j\alpha}{\theta + n}\nu(\cdot) + \frac{1}{\theta + n}\sum_{i=1}^{j}(n_i - \alpha)\delta_{S_i^*}(\cdot). \tag{1.2.6}$$

The expression in (1.2.6) allows to recover the conditional distribution (1.2.4)of the random partition of $[n+1]$ obtained after sampling one additional point X_{n+1}, given the previous partition of $[n]$.

1.2.5 Large n asymptotics for K_n and $K_{r,n}$

Under the Ewens-Pitman model (1.2.3), there have been several works on the large n asymptotic behaviour of K_n, showing different behaviours depending on whether $\alpha = 0$ or $\alpha \in (0,1)$. Denote the almost sure and weak convergence by $\xrightarrow{a.s.}$ and $\xrightarrow{w}$, respectively. For $\alpha = 0$ and $\theta > 0$, [65, Theorem 2.3] and Lindeberg-Feller central limit theorem show that as $n \to +\infty$

$$\frac{K_n}{\log n} \xrightarrow{a.s.} \theta \tag{1.2.7}$$

and

$$\sqrt{\log n}\left(\frac{K_n}{\log n} - \theta\right) \xrightarrow{w} \sqrt{\theta}\, N(0,1), \tag{1.2.8}$$

where $N(0,1)$ denotes the standard Gaussian random variable. Instead, for $\alpha \in (0,1)$ and $\theta > -\alpha$, [80, Theorem 3.8] and [9, Theorem 2.1 and Theorem 2.3] show that as $n \to +\infty$

$$\frac{K_n}{n^\alpha} \xrightarrow{a.s.} S_{\alpha,\theta} \tag{1.2.9}$$

and

$$\sqrt{n^\alpha}\left(\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right) \xrightarrow{w} \sqrt{\tilde{S}_{\alpha,\theta}}\, N(0,1), \tag{1.2.10}$$

where $S_{\alpha,\theta}$ and $\tilde{S}_{\alpha,\theta}$ are scaled Mittag-Leffler random variables [97], sharing the same distribution, and $\tilde{S}_{\alpha,\theta}$ is independent of $N(0,1)$. See [80, Chapter 3] for details on (1.2.7) and (1.2.9).

For what regards the $K_{r,n}$'s, [2, Theorem 1] entails that for $\alpha = 0$

$$K_{r,n} \xrightarrow[n\to+\infty]{w} Z_r \tag{1.2.11}$$

and

$$(K_{1,n}, K_{2,n}, \ldots) \xrightarrow[n\to+\infty]{w} (Z_1, Z_2, \ldots), \tag{1.2.12}$$

where the Z_r's are independent Poisson random variables with $\mathbb{E}[Z_r] = \theta/r$, for all $r \geq 1$. See [3] for generalizations and refinements of these asymptotic results. For $\alpha \in (0,1)$, it follows from [80, Lemma 3.11] that

$$\lim_{n\to+\infty} \frac{K_{r,n}}{n^\alpha} = p_\alpha(r) S_{\alpha,\theta} \qquad \text{a.s.}, \tag{1.2.13}$$

where

$$p_\alpha(r) = \frac{\alpha(1-\alpha)^{(r-1)}}{r!}.$$

Furthermore, [9] shows that

$$\sqrt{n^\alpha}\left(\frac{K_{r,n}}{n^\alpha} - \frac{A_{r,n}}{b_{r,n}\, n^\alpha}\right) \xrightarrow{w} \sqrt{p_\alpha(r) S'_{\alpha,\theta}}\, N(0,1),$$

where

$$b_{r,n} = \prod_{k=r}^{n-1} \frac{k+\theta}{k+\theta-r+\alpha},$$

$$A_{r,n} = \sum_{k=1}^{n-1} b_{r,k+1} \frac{(r-1-\alpha)}{k+\theta} \cdot K_{r-1,k}$$

and $S'_{\alpha,\theta}$ denotes a scaled Mittag-Leffler random variable independent of $N(0,1)$.

Remark 1.2.2. *Beyond the almost-sure and Gaussian fluctuations displayed in (1.2.7)-(1.2.5), K_n and $K_{r,n}$ have been investigated with respect to large and moderate deviations [45; 36; 37; 10; 75] and laws of iterated logarithm [9]. Further, non-asymptotic results for K_n have been established in terms of Berry-Esseen theorems [27] and concentration inequalities [76; 10].*

1.3 The "large θ" regime

In population genetics, the Poisson-Dirichlet distribution (i.e. $\alpha = 0$) describes the distribution of gene frequencies in a large neutral population at a locus, with the parameter $\theta > 0$ taking on the interpretation of the scaled population mutation rate [44, Chapter 2]. This has motivated the study of the so–called "large θ" behaviour of the random partition in the Ewens model, with some extensions to the case $\alpha \in (0,1)$ (see section 1.3.2 for a brief review). In particular, [40] first considered the regime

$$\theta = \lambda n$$

with $\lambda > 0$, in the study of the large n asymptotic behaviour of the number K_n of blocks in the Ewens model. Following this work, in most of this thesis we refer to such scaling as the the "large θ" regime.

1.3.1 Sequential construction: arrays of CRPs

It is possible to adapt the sequential construction of section 1.2.3 to the "large θ" setting. However, this requires a careful specification of the objects at hand, to avoid confusion with the model in the classical setting. In particular, the sequential construction in this case is used to build finite arrays of partitions, defined as follows.

For $n \in \mathbb{N}$ and $h \in \{1, ..., n\}$, let

$$\mathbf{K}_h^{\{n\}} = \left(K_h^{\{n\}}, K_{1,h}^{\{n\}}, ..., K_{h,h}^{\{n\}}\right)$$

be a variable encoding a random partition of $[h] = \{1, ..., h\}$, with $K_h^{\{n\}}$ the number of blocks and $K_{r,h}^{\{n\}}$ the number of blocks of numerosity r. Letting

$$\mathcal{F}_{h-1}^{\{n\}} = \sigma\left(\mathbf{K}_1^{\{n\}}, ..., \mathbf{K}_{h-1}^{\{n\}}\right),$$

for $h \in \{2, ..., n\}$, the following sequential relation holds:

$$P\left[K_h^{\{n\}} = K_{h-1}^{\{n\}} + 1 \mid \mathcal{F}_{h-1}^{\{n\}}\right] = \frac{\alpha K_n + \theta}{n + \theta}.$$

and for $r \geq 1$,

$$P\left[K_{r,h}^{\{n\}} = K_{r,h-1}^{\{n\}} + k \,|\, \mathcal{F}_{h-1}^{\{n\}}\right] = \begin{cases} p_{h-1}^{\{n\}} & \text{if } k = 1 \\ q_{h-1}^{\{n\}} & \text{if } k = -1 \\ 1 - p_{h-1}^{\{n\}} - q_{h-1}^{\{n\}} & \text{if } k = 0 \end{cases} \tag{1.3.1}$$

with

$$p_{r,h-1}^{\{n\}} = \begin{cases} \frac{\alpha K_{h-1}^{\{n\}} + \lambda n}{\lambda n + h - 1} & \text{if } r = 1 \\ \frac{(r-1-\alpha) K_{r-1,h-1}^{\{n\}}}{\lambda n + h - 1} & \text{if } r \geq 2 \end{cases}$$

and

$$q_{r,h-1}^{\{n\}} = \frac{(r - \alpha) K_{r,h-1}^{\{n\}}}{\lambda n + h - 1}.$$

Clearly, (1.3.1) is recovered by substituting $\theta = \lambda n$ in (1.2.4). It should be noted, however, that in this regime the recursive relation only holds for $h \leq n$, and the array $\left(\mathbf{K}_1^{\{n\}}, ..., \mathbf{K}_n^{\{n\}}\right)$ is not further extendable to an exchangeable random partition of $[n+1]$. Furthermore, two partitions $\mathbf{K}_i^{\{n\}}$ and $\mathbf{K}_j^{\{m\}}$ with $m \neq n$ do not satisfy compatibility relations of any kind. In other words, we are considering triangular arrays of the form

$$\begin{array}{ll} (\theta = \lambda) & \mathbf{K}_1^{\{1\}} \\ (\theta = 2\lambda) & \mathbf{K}_1^{\{2\}}\, \mathbf{K}_2^{\{2\}} \\ \vdots & \qquad\qquad \ddots \\ (\theta = n\lambda) & \mathbf{K}_1^{\{n\}}\, \mathbf{K}_2^{\{n\}} \ldots \mathbf{K}_{n-1}^{\{n\}}\, \mathbf{K}_n^{\{n\}} \\ \vdots & \qquad\qquad\qquad\qquad \ddots \end{array}$$

Equilvalently, for every $n \in \mathbb{N}$ we can see the partition of $[n]$ as the random partition generated by sampling from a finite vector of exchangeable random variables $X_1^{\{n\}}, ..., X_n^{\{n\}}$ such that

$$X_i^{\{n\}} \,|\, \mathrm{PYP}(\alpha, \lambda n) \overset{i.i.d.}{\sim} \mathrm{PYP}(\alpha, \lambda n).$$

Once again, because the prior $\mathrm{PYP}(\alpha, \lambda n)$ depends on n, the sequence $(X_1^{\{n\}}, ..., X_n^{\{n\}})$ is not extendable to an infinite exchangeable sequence, and for $n \neq m$ it satisfies no compatibility relation with the sequence $(X_1^{\{m\}}, ..., X_m^{\{m\}})$.

Our results will be concerned with the asymptotic behaviour of the variables $\mathbf{K}_n^{\{n\}}$. Because the notation introduced in this section is heavy, whenever there is no ambiguity in the meaning of the symbols we will drop the superscript $\{n\}$ and denote

$$\mathbf{K}_n := \mathbf{K}_n^{\{n\}}$$

In particular, we will return to the full notation only in chapter 3, where we will make explicit use of the partition structure 1.3.1.

1.3.2 Large n asymptotics for $K_n^{\{n\}}$ and $K_{r,n}^{\{n\}}$

There exists a rich literature on the large θ asymptotic behaviour of the two-parameter Poisson-Dirichlet distribution, assuming $\alpha \in [0, 1)$ and $\theta > 0$.

The genetic interpretation of θ has motivated the study of the large θ asymptotic behaviour of the Poisson-Dirichlet distribution, as well as of statistics thereof, providing Gaussian fluctuations and large (and moderate) deviations [93; 53; 60; 25; 40; 42]. As already mentioned, [40] first considered the regime $\theta = \lambda n$, with $\lambda > 0$, in the study of the large n asymptotic behaviour of the number $K_n^{\{n\}}$ of blocks in the Ewens model, providing a large deviation principle for $K_n^{\{n\}}$. Subsequently, [91] derived both a LLN and a CLT for $K_n^{\{n\}}$ under the same regime. The large θ asymptotic behaviour of $K_{r,n}^{\{n\}}$ has been explored in [7], establishing a microclustering property, namely the size of the largest cluster grows sub-linearly with the sample size, while the number of clusters grows linearly, by means of a weak LLN.

Some of the large θ asymptotic results for the Poisson-Dirichlet distribution have been extended to two-parameter Poisson-Dirichlet distribution (i.e. $\alpha \in (0, 1)$), in terms of both Gaussian fluctuations and large deviations [41; 44]. By contrast, our work is, to the best of our knowledge, the first to explore the large θ asymptotic behaviour of $K_n^{\{n\}}$ in the Ewens-Pitman model, establishing a LLN and a CLT. The large θ asymptotic behaviour of $K_{r,n}^{\{n\}}$ has been explored in [7] along with the case $\alpha = 0$, establishing an analogous weak LLN.

Chapter 2

Laws of large numbers and central limit theorem for the number of blocks in the Ewens–Pitman partition in the "large θ" regime

Under the Ewens-Pitman model (1.2.3), we study the large n asymptotic behaviour of $K_n^{\{n\}}$ when the parameter θ is allowed to depend linearly on $n \in \mathbb{N}$, namely $\theta = \lambda n$ with $\lambda > 0$. This non-standard asymptotic regime first appeared in [40] for the special case of the Ewens model, i.e. $\alpha = 0$. In particular, for $\alpha = 0$ and $\theta = \lambda n$, both a law of large numbers (LLN) and a central limit theorem (CLT) were established in [91, Proposition 2 and Theorem 2]. In the present chapter, we extend the LLN and the CLT to the more general case of the Ewens-Pitman model, i.e. $\alpha \in (0, 1)$.

2.1 Main result

Denoting by $\xrightarrow{p}$ the convergence in probability and by $\xrightarrow{w}$ weak convergence, the next theorem states our main result. For completeness of exposition, we state the theorem for $\alpha \in [0, 1)$, thereby including the original results of [91] for $\alpha = 0$.

Theorem 2.1.1. *For $n \in \mathbb{N}$, let $K_n^{\{n\}}$ be the number of partition blocks under the Ewens-Pitman model with parameter $\alpha \in [0, 1)$ and $\theta = \lambda n$, with $\lambda > 0$. If*

$$\mathfrak{m}_{\alpha,\lambda} := \begin{cases} \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{\alpha} - 1\right] & \text{for } \alpha \in (0,1) \\ \lambda \log\left(1+\frac{1}{\lambda}\right) & \text{for } \alpha = 0 \end{cases}$$

and

$$\mathfrak{s}^2_{\alpha,\lambda} := \begin{cases} \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{2\alpha}\left(1-\frac{\alpha}{1+\lambda}\right) - \left(1+\frac{1}{\lambda}\right)^{\alpha}\right] & \text{for } \alpha \in (0,1) \\ \lambda \log\left(1+\frac{1}{\lambda}\right) - \frac{\lambda}{\lambda+1} & \text{for } \alpha = 0, \end{cases}$$

then, as $n \to +\infty$ there hold:

i)

$$\mathbb{E}\left[K_n^{\{n\}}\right] = n\mathfrak{m}_{\alpha,\lambda} + O(1) \tag{2.1.1}$$

and

$$\mathrm{Var}\left(K_n^{\{n\}}\right) = n\mathfrak{s}_{\alpha,\lambda}^2 + O(1); \tag{2.1.2}$$

ii)

$$\frac{K_n^{\{n\}}}{n} \xrightarrow{p} \mathfrak{m}_{\alpha,\lambda}; \tag{2.1.3}$$

iii)

$$\frac{K_n^{\{n\}} - n\mathfrak{m}_{\alpha,\lambda}}{\sqrt{n\mathfrak{s}_{\alpha,\lambda}^2}} \xrightarrow{w} N(0,1). \tag{2.1.4}$$

Further, for any $\lambda > 0$ *there hold that* $\mathfrak{m}_{0,\lambda} = \lim_{\alpha\to 0^+} \mathfrak{m}_{\alpha,\lambda}$ *and* $\mathfrak{s}_{0,\lambda}^2 = \lim_{\alpha\to 0^+} \mathfrak{s}_{\alpha,\lambda}^2$.

Theorem 2.1.1 provides analogues of the almost-sure fluctuations (1.2.7) and (1.2.9), as well as of the Gaussian fluctuations (1.2.8) and (1.2.5), under the non-standard asymptotic regime $\theta = \lambda n$, with $\lambda > 0$. By comparing the LLN (2.1.3) with the almost-sure fluctuation results (1.2.7) and (1.2.9), we can assess how the linear dependence of θ on n, i.e. $\theta = \lambda n$, affects the large n asymptotic behaviour of $K_n^{\{n\}}$ in terms of both scaling and limiting behaviour. Specifically, while the almost-sure fluctuations scale as $\log n$ for $\alpha = 0$ and as n^α for $\alpha \in (0,1)$, the LLN (2.1.3) exhibits the "usual" linear scaling in n for all $\alpha \in [0,1)$. Moreover, in the almost-sure fluctuation results (1.2.7) and (1.2.9), the limiting behaviour is non-random for $\alpha = 0$ and random for $\alpha \in (0,1)$, which in turn determines the non-random and random centerings in (1.2.8) and (1.2.5), respectively. In contrast, the LLN (2.1.3) yields a non-random limit for all $\alpha \in [0,1)$, leading to a non-random centering in the CLT (2.1.4).

The LLN (2.1.3) relies on combining Chebyshev's inequality with the asymptotic expansions (2.1.1) and (2.1.2). These expansions, in turn, are obtained from the distribution of $K_n^{\{n\}}$, which follows by marginalizing (1.2.3) with $\theta = \lambda n$, for $\lambda > 0$ [80, Equation 3.11]. For $\alpha \in [0,1)$, a stronger version of the LLN (2.1.3) will be also established, showing that as $n \to +\infty$

$$\frac{K_n^{\{n\}}}{n} \xrightarrow{a.s.} \mathfrak{m}_{\alpha,\lambda}. \tag{2.1.5}$$

While the LLN (2.1.3) relies on the sole (marginal) distribution of $K_n^{\{n\}}$, the strong LLN (2.1.5) requires to considering the finite-dimensional laws of the sequence of random variables $\left\{K_n^{\{n\}}\right\}_{n\geq 1}$. That is, for each $n \geq 1$, the strong LLN requires to look at $K_n^{\{n\}}$ as the number of blocks of the random partition of $\{1,\dots,n\}$ induced by random sampling the two-parameter Poisson-Dirichlet distribution with $\alpha \in [0,1)$ and $\theta = \lambda n$, for $\lambda > 0$ [63; 77; 81].

The CLT (2.1.4) relies on different techniques depending on whether $\alpha = 0$ or $\alpha \in (0,1)$. For $\alpha = 0$, we provide an alternative proof of the CLT (2.1.4), distinct from that given in [91, Theorem 2]. In particular, by representing $K_n^{\{n\}}$ as a sum of independent, but not identically distributed, Bernoulli random variables, our proof also yields a Berry-Esseen theorem for $K_n^{\{n\}}$ with respect to the Kolmogorov metric $\|\cdot\|_\infty$. More precisely, if F_n and Φ denote the cumulative distribution functions of the distributions of $n^{-1/2}\mathfrak{s}_{0,\lambda}^{-1}(K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda})$ and $N(0,1)$, respectively, then we show that there exist a constant $C > 0$ and $\bar{n} \in \mathbb{N}$ such that

$$\|F_n - \Phi\|_\infty \leq \frac{C\,\log(n)}{n^{1/8}} \tag{2.1.6}$$

holds for every $n \geq \bar{n}$. For $\alpha \in (0,1)$ the proof of the CLT (2.1.4) relies on the compound Poisson construction of $K_n^{\{n\}}$ [27; 28]. This approach naturally leads to a LLN, a CLT and a Berry-Esseen theorem for the number of blocks of the negative-Binomial compound Poisson random partition, results of independent interest [17, Example 3.1 and Example 3.2].

2.1.1 Organization of the chapter

The proof of Theorem 2.1.1 has a rather neat structure, but calls for many technically complex auxiliary results. Therefore, for the sake of readability, this chapter is structured as follows. Section 2.2 contains the proof of Theorem 2.1.1 for $\alpha \in (0,1)$, with technical results deferred to Appendix A. Section 4.3 contains an alternative proof of the CLT (2.1.4) for $\alpha = 0$, and the proof of the Berry-Esseen inequality (2.1.6), with technical results deferred to Appendix B. Section 4.4 contains the proof of the strong LLN (2.1.5), with technical results deferred to Appendix C.

2.2 Proof of Theorem 2.1.1 for $\alpha \in (0,1)$

The proof of Theorem 2.1.1 is structured as follows. In Section 2.2.1 we prove the asymptotic expansions (2.1.1) and (2.1.2). In Section 2.2.2 we prove the LLN (2.1.3). In Section 2.2.3 we prove the CLT (2.1.4). Technical lemmas and propositions are deferred to Appendix A.

2.2.1 Proof of the asymptotic expansions (2.1.1) and (2.1.2)

Denote by $(x)_{\downarrow n}$ the falling (or descending) factorial of x of order n, i.e. $(x)_{\downarrow n} := \prod_{i=0}^{n-1}(x-i)$. Based on the distribution of K_n [80, Equation 3.11], a direct calculation leads to

$$\mathbb{E}\left[(K_n)_{\downarrow j}\right] = \left[\frac{\theta}{\alpha}\right]_{(j)} \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} \frac{[\theta + i\alpha]_{(n)}}{[\theta]_{(n)}} \qquad j \in \mathbb{N}$$

where $(x)_{\downarrow j} := \prod_{i=0}^{j-1}(x-i)$ denotes the falling factorial of x. Accordingly, by simple algebra

$$\mathbb{E}\left[K_n\right] = \mathbb{E}\left[(K_n)_{\downarrow 1}\right] = \left(\frac{\theta}{\alpha}\right)\left[-1 + \frac{\Gamma(\theta+n+\alpha)}{\Gamma(\theta+\alpha)}\frac{\Gamma(\theta+n)}{\Gamma(\theta)}\right] \tag{2.2.1}$$

and

$$\begin{aligned}
&\operatorname{Var}\left(K_n\right) \qquad (2.2.2)\\
&= \mathbb{E}\left[(K_n)_{\downarrow 2}\right] + \mathbb{E}\left[K_n\right] - \left(\mathbb{E}\left[K_n\right]\right)^2\\
&= \frac{\theta}{\alpha}\left(\frac{\theta}{\alpha}+1\right)\left[1 - 2\frac{\Gamma(\theta+n+\alpha)}{\Gamma(\theta+n)}\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)} + \frac{\Gamma(\theta+n+2\alpha)}{\Gamma(\theta+n)}\frac{\Gamma(\theta)}{\Gamma(\theta+2\alpha)}\right]\\
&\quad + \frac{\theta}{\alpha}\left[-1 + \frac{\Gamma(\theta+n+\alpha)}{\Gamma(\theta+n)}\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)}\right]\\
&\quad - \left(\frac{\theta}{\alpha}\right)^2\left[1 + \left(\frac{\Gamma(\theta+n+\alpha)}{\Gamma(\theta+n)}\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)}\right)^2 - 2\frac{\Gamma(\theta+n+\alpha)}{\Gamma(\theta+n)}\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)}\right].
\end{aligned}$$

See the proof of [9, Theorem 2.1] for a derivation of (2.2.1) and (2.2.2) based on the sequential construction of the Ewens-Pitman model. If $\theta = \lambda n$, then (2.2.1) and (2.2.2)

become

$$\mathbb{E}\left[K_n^{\{n\}}\right] = \frac{\lambda n}{\alpha}\left[-1+\frac{\Gamma\left(\alpha+(\lambda+1)n\right)\ \Gamma(\lambda n)}{\Gamma(\alpha+\lambda n)\ \Gamma\left((\lambda+1)n\right)}\right] \tag{2.2.3}$$

and

$$\begin{aligned}
&\mathrm{Var}\left(K_n^{\{n\}}\right) \qquad (2.2.4)\\
&= \frac{\lambda n}{\alpha}\left(\frac{\lambda n}{\alpha}+1\right)\left[1-2\frac{\Gamma(n(\lambda+1)+\alpha)}{\Gamma(n(\lambda+1))}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+\alpha)}+\frac{\Gamma(n(\lambda+1)+2\alpha)}{\Gamma(n(\lambda+1))}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+2\alpha)}\right]\\
&\quad+\frac{\lambda n}{\alpha}\left[-1+\frac{\Gamma(n(\lambda+1)+\alpha)}{\Gamma(n(\lambda+1))}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+\alpha)}\right]\\
&\quad-\left(\frac{\lambda n}{\alpha}\right)^2\left[1+\left(\frac{\Gamma(n(\lambda+1)+\alpha)}{\Gamma(n(\lambda+1))}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+\alpha)}\right)^2-2\frac{\Gamma(n(\lambda+1)+\alpha)}{\Gamma(n(\lambda+1))}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+\alpha)}\right],
\end{aligned}$$

respectively. By applying to (2.2.3) the asymptotic expansion for the Gamma function [89, Equation 1],

$$\mathbb{E}\left[K_n^{\{n\}}\right] = n\left\{\frac{\lambda}{\alpha}\left[-1+\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}+O\left(\frac{1}{n}\right)\right]\right\} = n\mathfrak{m}_{\alpha,\lambda}+O(1), \tag{2.2.5}$$

with

$$\mathfrak{m}_{\alpha,\lambda} = \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{\alpha}-1\right]. \tag{2.2.6}$$

Similarly, by applying to (2.2.4) the asymptotic expansion for the Gamma function [89, Equation 1],

$$\begin{aligned}
&\mathrm{Var}\left(K_n^{\{n\}}\right) \qquad (2.2.7)\\
&= \left[\left(\frac{\lambda n}{\alpha}\right)^2+\frac{\lambda n}{\alpha}\right]\left[1-2\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}+\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\frac{\alpha(\alpha-1)}{\lambda(\lambda+1)n}+\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\right.\\
&\qquad\left.-\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\frac{\alpha(2\alpha-1)}{\lambda(\lambda+1)n}+O\left(\frac{1}{n^2}\right)\right]\\
&\quad+\left(\frac{\lambda n}{\alpha}\right)\left[-1+\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}+O\left(\frac{1}{n}\right)\right]\\
&\quad-\left(\frac{\lambda n}{\alpha}\right)^2\left[1+\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}-\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\frac{\alpha(\alpha-1)}{\lambda(\lambda+1)n}-2\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\right.\\
&\qquad\left.+\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\frac{\alpha(\alpha-1)}{\lambda(\lambda+1)n}+O\left(\frac{1}{n^2}\right)\right]\\
&= \left(\frac{\lambda n}{\alpha}\right)\cdot\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\left[-1+\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\cdot\left(1-\frac{\alpha}{\lambda+1}\right)\right]+O(1)\\
&= n\mathfrak{s}_{\alpha,\lambda}^2+O(1),
\end{aligned}$$

with

$$\mathfrak{s}_{\alpha,\lambda}^2 = \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{2\alpha}\left(1-\frac{\alpha}{1+\lambda}\right)-\left(1+\frac{1}{\lambda}\right)^{\alpha}\right]. \tag{2.2.8}$$

The proof of the asymptotic expansions (2.1.1) and (2.1.2) is completed in view of (2.2.5)-(2.2.6) and (2.2.7)-(2.2.8).

2.2.2 Proof of the LLN (2.1.3)

Based on the asymptotic expansions (2.1.1), it is useful to consider the following identity:

$$\frac{K_n^{\{n\}} - n\mathfrak{m}_{\alpha,\lambda}}{n} = \frac{K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]}{n} + \frac{\mathbb{E}\left[K_n^{\{n\}}\right] - n\mathfrak{m}_{\alpha,\lambda}}{n}. \tag{2.2.9}$$

The second term in the right-end side is a deterministic term, and it converges to 0 in view of (2.1.1). Therefore, fix $\varepsilon > 0$ and combine Chebyshev's inequality with (2.1.2) to obtain

$$P\left[\left|\frac{K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]}{n}\right| > \varepsilon\right] = P\left[\left|K_n^{\{n\}} - \mathbb{E}[K_n^{\{n\}}]\right| > n\varepsilon\right] \leq \frac{\mathrm{Var}\left(K_n^{\{n\}}\right)}{n^2\varepsilon^2} = O\left(\frac{1}{n}\right)$$

as $n \to +\infty$. Accordingly, the proof of the LLN (2.1.3) is completed in view of the identity (2.2.9).

2.2.3 Proof of the CLT (2.1.4)

To prove the CLT (2.1.4), we recall the compound Poisson construction of K_n [27, Lemma 1 and Proposition 1]. For $\alpha \in (0,1)$ and $\theta > 0$ let $S_{\alpha,\theta}$ be the scaled Mittag-Leffler random variable in (1.2.9). More precisely, $S_{\alpha,\theta}$ is a positive random variable with density function

$$f_{S_{\alpha,\theta}}(s) = \frac{\Gamma(\theta)}{\Gamma\left(\frac{\theta}{\alpha}\right)} s^{\frac{\theta-1}{\alpha}-1} f_\alpha(s^{-1/\alpha}) \qquad s > 0,$$

where f_α denotes the positive α-stable density function [82]. We refer to [97, Sections 2.2 and 2.4] for details on f_α, and to [80, Chapter 0] for details on $S_{\alpha,\theta}$. In particular, there hold

$$\mathbb{E}\left[S_{\alpha,\theta}\right] = \frac{\theta}{\alpha}\frac{\Gamma(\theta)}{\Gamma(\alpha+\theta)} \tag{2.2.10}$$

and

$$\mathrm{Var}\left[S_{\alpha,\theta}\right] = \frac{\theta}{\alpha}\left(\frac{\theta}{\alpha}+1\right)\frac{\Gamma(\theta)}{\Gamma(\theta+2\alpha)} - \left(\frac{\theta}{\alpha}\right)^2\left(\frac{\Gamma(\theta)}{\Gamma(\theta+\alpha)}\right)^2. \tag{2.2.11}$$

Now, for $\alpha \in (0,1)$, $z > 0$ and $n \in \mathbb{N}$, let us introduce a random variable $R(\alpha,n,z)$ with values in the set $\{1,\ldots,n\}$, whose distribution has probability mass function given by

$$P[R(\alpha,n,z) = k] = \frac{\mathscr{C}(n,k;\alpha)z^k}{\sum_{j=1}^n \mathscr{C}(n,j;\alpha)z^j} \qquad k \in \{1,\ldots,n\}, \tag{2.2.12}$$

where

$$\mathscr{C}(n,k;\alpha) = \frac{1}{k!}\sum_{i=0}^k (-1)^i \binom{k}{i} [-i\alpha]_{(n)} \geq 0$$

is the generalized factorial coefficient, with the proviso that $\mathscr{C}(0,0;\alpha) := 1$ and $\mathscr{C}(n,0;\alpha) := 0$ for $n \geq 1$ [16, Chapter 2]. Specifically, (2.2.12) is the distribution of the number of blocks in a random partition of $\{1,\ldots,n\}$ under the negative-binomial compound Poisson model [17, Example 3.2].

Let $G_{a,b}$ be a Gamma random variable with shape parameter $a > 0$ and scale parameter $b > 0$, namely the distribution of $G_{a,b}$ has density function given by: $f_{a,b}(x) =$

$[b^a/\Gamma(a)]x^{a-1}e^{-bx}$, for $x > 0$. For $\alpha \in (0,1)$ and $\theta > 0$, [27, Proposition 1] shows that for any $n \in \mathbb{N}$

$$K_n \stackrel{d}{=} R\left(\alpha, n, Z_{\theta,n}\right), \tag{2.2.13}$$

and

$$Z_{\theta,n} := S_{\alpha,\theta} G^{\alpha}_{\theta+n,1}$$

such that $S_{\alpha,\theta}$ and $G_{\theta+n,1}$ are independent random variables, and independent of $R(\alpha, n, z)$, for any $z > 0$. Here and throughout the paper, $\stackrel{d}{=}$ denotes identity in distribution. See also [28] for details on the distributional identity (2.2.13), as well as for related results.

The proof of the CLT (2.1.4) relies on (2.2.13). We consider (2.2.13) in the regime $\theta = \lambda n$. That is, for any $n \in \mathbb{N}$

$$K_n^{\{n\}} \stackrel{d}{=} R\left(\alpha, n, Z_n\right), \tag{2.2.14}$$

where

$$Z_n := Z_{\lambda n, n} = S_{\alpha,\lambda n} G^{\alpha}_{(\lambda+1)n,1} \tag{2.2.15}$$

such that $S_{\alpha,\lambda n}$ and $G_{(\lambda+1)n,1}$ are independent random variables, and independent of $R(\alpha, n, z)$, for any $z > 0$. To simplify the notation, for any $n \in \mathbb{N}$ and $z > 0$ it is useful to set

$$R_n(z) := R(\alpha, n, nz). \tag{2.2.16}$$

Hereafter, we investigate the large n asymptotic behaviours of the random variables Z_n and $R_n(z)$, as well as their interplay with respect to the asymptotic expansions in (2.1.1) and (2.1.2).

2.2.3.1 LLN and CLT for Z_n

The next proposition provides a LLN and a CLT for the random variable Z_n defined in (2.2.15).

Proposition 1 (LLN and CLT for Z_n)**.** *For any $\alpha \in (0,1)$ and $\lambda > 0$, set*

$$z_0 := \frac{\lambda}{\alpha}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}$$

and

$$\Sigma^2 := \frac{\lambda}{\alpha}\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\left(1 - \frac{\alpha}{\lambda+1}\right).$$

As $n \to +\infty$, there hold:

i)

$$\mathbb{E}[Z_n] = nz_0 + O(1) \tag{2.2.17}$$

and

$$\mathrm{Var}(Z_n) = n\Sigma^2 + O(1); \tag{2.2.18}$$

ii)

$$\frac{Z_n}{n} \xrightarrow{p} z_0; \tag{2.2.19}$$

iii)

$$\frac{Z_n - nz_0}{\sqrt{n}} \xrightarrow{w} \mathcal{N}\left(0, \Sigma^2\right). \tag{2.2.20}$$

Proof. The proof, particularly that of (2.2.20), is technical, and it is therefore deferred to Appendix A.1. ∎

2.2.3.2 LLN and CLT for $R_n(z)$

The next proposition provides a LLN, a CLT for the random variable $R_n(z)$ defined in (2.2.16); in particular, the CLT is refined (or quantified) by means of a Berry-Esseen inequality.

Proposition 2 (LLN and Berry-Esseen inequality for $R_n(z)$)**.** *For any $\alpha \in (0,1)$ and $z > 0$, let $\mu : (0,+\infty) \to \mathbb{R}$ and $\sigma : (0,+\infty) \to \mathbb{R}$ be functions defined as*

$$\mu(z) := z\left(1 - \frac{1}{\tau(z)}\right)$$

and

$$\sigma^2(z) := z\left(1 - \frac{1}{\tau(z)} - \frac{\alpha}{\alpha z + (1-\alpha)\tau(z)}\right) > 0$$

where, for any $z > 0$, $\tau(z)$ denotes the unique real, positive solution to the equation

$$\tau(z)^{\frac{1}{\alpha}} = \frac{\tau(z)}{\alpha z} + 1. \tag{2.2.21}$$

As $n \to +\infty$, there hold:

i)

$$\mathbb{E}[R_n(z)] = n\mu(z) + O(1) \tag{2.2.22}$$

and

$$\mathrm{Var}(R_n(z)) = n\sigma^2(z) + O(1); \tag{2.2.23}$$

ii)

$$\frac{R_n(z)}{n} \xrightarrow{p} \mu(z). \tag{2.2.24}$$

Finally, set

$$W_n(z) := \frac{R_n(z) - n\mu(z)}{\sqrt{n\sigma^2(z)}}$$

and denote by $F_{W_n(z)}$ its distribution function. Then, there exists a continuous function $C : (0,+\infty) \to (0,+\infty)$ such that, for any choice of ζ_0 and ζ_1 that satisfy $0 < \zeta_0 < z_0 < \zeta_1 < +\infty$, with the same z_0 as in Proposition 1, the inequality

$$\left\| F_{W_n(z)} - \Phi \right\|_\infty \leq \frac{C(z)\,\log(n)}{n^{1/8}} \tag{2.2.25}$$

holds for every $z \in [\zeta_0, \zeta_1]$ and every $n \geq \bar{n}(\zeta_0, \zeta_1)$. In particular, the Berry-Esseen bound (2.2.25) implies that $W_n(z) \xrightarrow{w} N(0,1)$ as $n \to +\infty$, for every fixed $z > 0$.

Proof. The proof, particularly that of (2.2.25), is technical, and it is therefore deferred to Appendix A.2. ■

Proposition 2 is of independent interest in compound Poisson random partitions [17, Example 3.2].

2.2.3.3 Completing the proof of the CLT (2.1.4)

We start by showing that the LLNs and the CLTs established in Proposition 1 and Proposition 2 have a direct interplay with respect to the asymptotic expansions (2.1.1) and (2.1.2).

Proposition 3. *Under the assumptions of Proposition 1 and Proposition 2, there hold*

$$\mu(z_0) = \mathfrak{m}_{\alpha,\lambda}$$

and

$$\sigma^2(z_0) + \Sigma^2 \left(\mu'(z_0)\right)^2 = \mathfrak{s}^2_{\alpha,\lambda}$$

Proof. Set $\tau_0 := \left(\frac{\lambda+1}{\lambda}\right)^\alpha$. First, we prove that $\tau(z_0) = \tau_0$ by checking that τ_0 is a (and therefore the unique) positive solution to (2.2.21) when $z = z_0$. In particular, it holds

$$\frac{\tau_0}{\alpha z_0} + 1 = \frac{\alpha}{\alpha\lambda}\left(\frac{\lambda+1}{\lambda}\right)^\alpha \left(\frac{\lambda}{\lambda+1}\right)^\alpha + 1 = \frac{1}{\lambda} + 1 = \frac{\lambda+1}{\lambda} = \tau_0^{\frac{1}{\alpha}}.$$

The first identity then follows by evaluating μ at z_0. For the second identity, differentiate both sides of (2.2.21) with respect to z and rearrange to obtain that for $z > 0$

$$\tau'(z) = -\frac{\tau(z)}{zD(z)},$$

where

$$D(z) := z\tau(z)^{\frac{1-\alpha}{\alpha}} - 1 = \frac{\alpha z + (1-\alpha)\tau(z)}{\alpha\tau(z)}. \tag{2.2.26}$$

Whence,

$$\mu'(z) = z\frac{\tau(z)'}{\tau(z)^2} + 1 - \frac{1}{\tau(z)} = -\frac{1}{\tau(z)D(z)} + 1 - \frac{1}{\tau(z)}.$$

A simple computation yields

$$D(z_0) = -\left(1 - \frac{\lambda+1}{\alpha}\right)$$

and, in turn,

$$\mu'(z_0) = 1 - \frac{\lambda+1}{\tau_0(\lambda+1-\alpha)}.$$

Thus,we can write

$$\begin{aligned}
&\sigma^2(z_0) + \Sigma^2 \cdot \left(\mu'(z_0)\right)^2 \\
&\quad = \frac{\lambda}{\alpha}\tau_0\left(1 - \frac{1}{\tau_0} - \frac{\alpha}{(\lambda+1-\alpha)\tau_0}\right) - \frac{\lambda}{\alpha}\tau_0^2\left(1 - \frac{\alpha}{\lambda+1}\right)\left(1 - \frac{\lambda+1}{\tau_0(\lambda+1-\alpha)}\right)^2 \\
&\quad = \frac{\lambda}{\alpha}\left\{\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\left(1 - \frac{\alpha}{\lambda+1}\right) - \left(\frac{\lambda+1}{\lambda}\right)^\alpha\right\} \\
&\quad = \mathfrak{s}^2_{\alpha,\lambda}.
\end{aligned}$$

■

Based on Proposition 3, we show how Proposition 1 and Proposition 2 can be used to prove the CLT (2.1.4). Denoting by F_n the cumulative distribution function of the random variable

$$\frac{K_n^{\{n\}} - n\mathfrak{m}_{\alpha,\lambda}}{\sqrt{n\mathfrak{s}^2_{\alpha,\lambda}}},$$

we prove that, for any $x \in \mathbb{R}$,

$$\lim_{n\to+\infty} F_n(x) := \lim_{n\to+\infty} P\left[K_n^{\{n\}} \leq n\mathfrak{m}_{\alpha,\lambda} + \sqrt{n\mathfrak{s}^2_{\alpha,\lambda}}x\right] = \Phi(x).$$

Denote by $\mu_{\frac{Z_n}{n}}$ the probability distribution of Z_n/n. By conditional probability, we can rewrite (2.2.14) as

$$\begin{aligned} F_n(x) &= \int_0^{+\infty} P\left[R_n(z) \leq n\mathfrak{m}_{\alpha,\lambda} + \sqrt{n\mathfrak{s}^2_{\alpha,\lambda}}x\right] \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \\ &= \int_0^{+\infty} P\left[W_n(z) \leq \frac{\sqrt{n}\,[\mathfrak{m}_{\alpha,\lambda} - \mu(z)] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z)}\right] \mu_{\frac{Z_n}{n}}(\mathrm{d}z). \end{aligned}$$

Whence, $F_n(x) = \mathcal{I}_1^{(n)}(x) + \mathcal{I}_2^{(n)}(x)$, where

$$\mathcal{I}_1^{(n)}(x) := \int_0^{+\infty} \Phi\left(\frac{\sqrt{n}\,[\mathfrak{m}_{\alpha,\lambda} - \mu(z)] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z)}\right) \mu_{\frac{Z_n}{n}}(\mathrm{d}z)$$

and

$$\begin{aligned} \mathcal{I}_2^{(n)}(x) := \int_0^{+\infty} \Bigg\{ & F_{W_n(z)}\left(\frac{\sqrt{n}\,[\mathfrak{m}_{\alpha,\lambda} - \mu(z)] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z)}\right) \\ & - \Phi\left(\frac{\sqrt{n}\,[\mathfrak{m}_{\alpha,\lambda} - \mu(z)] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z)}\right)\Bigg\} \mu_{\frac{Z_n}{n}}(\mathrm{d}z). \end{aligned}$$

Based on the above identities, the proof of the CLT (2.1.4) is completed by showing that, for every $x \in \mathbb{R}$

$$\lim_{n\to+\infty} \mathcal{I}_1^{(n)}(x) = \Phi(x) \tag{2.2.27}$$

and

$$\lim_{n\to+\infty} \mathcal{I}_2^{(n)}(x) = 0. \tag{2.2.28}$$

We start with the identity (4.3.16); in particular, to prove (4.3.16), we premise two lemmas.

Lemma 2.2.1. *If Y is a Gaussian random variable with mean 0 and variance Σ^2, then*

$$\mathbb{E}\left[\Phi\left(\frac{\mu'(z_0)Y + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z_0)}\right)\right] = \Phi(x)$$

holds for every $x \in \mathbb{R}$.

Proof. Introduce a standard Gaussian random variable Z, independent of Y. By standard properties of conditional probability, it holds that

$$\Phi\left(\frac{\mu'(z_0)Y + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z_0)}\right) = P\left[Z \leq \frac{\mu'(z_0)Y + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z_0)} \;\middle|\; Y\right],$$

which implies

$$\mathbb{E}\left[\Phi\left(\frac{\mu'(z_0)Y + \mathfrak{s}_{\alpha,\lambda}x}{\sigma(z_0)}\right)\right] = P\left[\frac{\sigma(z_0)Z - \mu'(z_0)Y}{\mathfrak{s}_{\alpha,\lambda}} \leq x\right].$$

Notice that $\mathfrak{s}_{\alpha,\lambda}^{-1}[\sigma(z_0)Z - \mu'(z_0)Y]$ has Gaussian distribution with mean 0 and variance

$$\frac{\sigma(z_0)^2 + (\mu'(z_0))^2\Sigma^2}{\mathfrak{s}_{\alpha,\lambda}^2} = 1,$$

thanks to the last identity in Proposition 3. This completes the proof. ■

Lemma 2.2.2. *Let $\psi \in C^0([0,+\infty)) \cap C^1((0,+\infty))$, with bounded (first) derivative. Then, as $n \to +\infty$ there holds*

$$\sqrt{n}\left[\psi\left(\frac{Z_n}{n}\right) - \psi(z_0)\right] \xrightarrow{w} \mathcal{N}\left(0, (\psi'(z_0))^2\Sigma^2\right)$$

Proof. According to (2.2.20), $\sqrt{n}\left(Z_n/n - z_0\right) \xrightarrow{w} \mathcal{N}\left(0, \Sigma^2\right)$. The poof then follows by means of a straightforward application of the delta-method; see for example [85, Corollary 1.1].

■

The next proposition combines Lemma F.1 and Lemma F.2 to prove the identity (4.3.16).

Proposition 4. *For every $x \in \mathbb{R}$,*

$$\lim_{n\to+\infty} \mathcal{I}_1^{(n)}(x) := \lim_{n\to+\infty} \mathbb{E}\left[\Phi\left(\frac{\sqrt{n}\left[\mathfrak{m}_{\alpha,\lambda} - \mu\left(\frac{Z_n}{n}\right)\right] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma\left(\frac{Z_n}{n}\right)}\right)\right] = \Phi(x).$$

Proof. According to Proposition 3, we have $\mathfrak{m}_{\alpha,\lambda} = \mu(z_0)$. Now, it is useful to observe that the function μ considered in Proposition 2 belongs to $C^0([0,+\infty)) \cap C^1((0,+\infty))$, and has bounded (first) derivative. Therefore, we can apply Lemma F.2 for the choice $\psi = \mu$, to obtain that, as $n \to +\infty$

$$\sqrt{n}\left[\mu(z_0) - \mu\left(\frac{Z_n}{n}\right)\right] \xrightarrow{w} \mu'(z_0)Y \ ,$$

where the random variable Y is the same random variable as in Lemma F.1. Further, thanks to the continuity of σ and the mapping theorem for the convergence in probability, the LLN (2.2.19) entails that, as $n \to +\infty$

$$\sigma\left(\frac{Z_n}{n}\right) \xrightarrow{p} \sigma(z_0).$$

At this stage, Slutsky's theorem shows that, as $n \to +\infty$

$$\frac{\sqrt{n}\left[\mu\left(\frac{Z_n}{n}\right) - \mathfrak{m}_{\alpha,\lambda}\right] + \mathfrak{s}_{\alpha,\lambda}x}{\sigma\left(\frac{Z_n}{n}\right)} \xrightarrow{w} \frac{\mu'(z_0)Y + \mathfrak{s}_{\alpha,\lambda}x}{\sigma\left(z_0\right)}.$$

Since the (cumulative distribution) function Φ is bounded and continuous, the proof is completed by applying the Portmanteau theorem and recalling Lemma F.1. ■

Now, consider the identity (4.3.17); the next proposition proves (4.3.17) by an application of Proposition 2.

Proposition 5. *For every* $x \in \mathbb{R}$, *as* $n \to +\infty$, *it holds*

$$\left|\mathcal{I}_2^{(n)}(x)\right| \leq \int_0^{+\infty} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \to 0.$$

Proof. The inequality is straightforward, as it follows directly by taking the absolute value inside the integral in $\mathcal{I}_2^{(n)}$. Then, in order to prove convergence to zero, fix $\varepsilon > 0$ and choose $\delta = \delta(\varepsilon) > 0$ such that $\Phi(\delta) = 1 - \varepsilon/2$. Moreover, fix ζ_0 and ζ_1 as in Proposition 2. Set

$$\bar{m} := \bar{m}(\varepsilon, \zeta_0, \zeta_1) := \min\left\{m \in \mathbb{N} : \; z_0 - \frac{\delta\Sigma}{\sqrt{m}} > \zeta_0, \text{ and } z_0 - \frac{\delta\Sigma}{\sqrt{m}} < \zeta_1\right\},$$

which is well-defined since $z_0 \in (\zeta_0, \zeta_1)$. Set $\tilde{\zeta}_0 := z_0 - \frac{\delta\Sigma}{\sqrt{\bar{m}}}$ and $\tilde{\zeta}_1 := z_0 + \frac{\delta\Sigma}{\sqrt{\bar{m}}}$, and write

$$\begin{aligned}
&\int_0^{+\infty} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \qquad (2.2.29)\\
&= \int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) + \int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z).
\end{aligned}$$

Now, consider the terms on the right-hand side of (4.3.18) separately. With regard to the first term, inequality (2.2.25) entails that, for every $n \geq \bar{n}$, there hold

$$\int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \leq \frac{\log(n)}{n^{\frac{1}{8}}} \int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} C(z)\, \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \leq \frac{\log(n)}{n^{\frac{1}{8}}}\, \mathcal{M}_C, \qquad (2.2.30)$$

where $\mathcal{M}_C := \max_{z\in[\zeta_0,\zeta_1]} C(z)$ is finite since the function C is continuous. Concerning the second term on the right-hand side of (4.3.18), for any $n \geq \bar{m}$, there hold

$$\begin{aligned}
\int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) &\leq \int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \qquad (2.2.31)\\
&= P\left[\frac{Z_n}{n} \notin \left[z_0 - \frac{\delta\Sigma}{\sqrt{\bar{m}}}, z_0 + \frac{\delta\Sigma}{\sqrt{\bar{m}}}\right]\right]\\
&\leq P\left[\frac{Z_n}{n} \notin \left[z_0 - \frac{\delta\Sigma}{\sqrt{n}}, z_0 + \frac{\delta\Sigma}{\sqrt{n}}\right]\right]\\
&= P\left[\frac{Z_n - nz_0}{\sqrt{n}\Sigma} \notin [-\delta, \delta]\right].
\end{aligned}$$

Thus, combining identity (4.3.18) with inequalities (4.3.19)–(4.3.20) yields

$$\int_0^{+\infty} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z) \leq \frac{\log(n)}{n^{\frac{1}{8}}}\, \mathcal{M}_C + P\left[\frac{Z_n - nz_0}{\sqrt{n}\Sigma} \notin [-\delta, \delta]\right]$$

for any $n \geq \max(\bar{n}, \bar{m})$. Whence, from (2.2.20),

$$\begin{aligned}
0 &\leq \limsup_{n\to+\infty} \int_0^{+\infty} \left\|F_{W_n(z)} - \Phi\right\|_\infty \mu_{\frac{Z_n}{n}}(\mathrm{d}z)\\
&\leq \lim_{n\to+\infty} \left\{\frac{\log(n)}{n^{\frac{1}{8}}}\, \mathcal{M}_C + P\left[\frac{Z_n - nz_0}{\sqrt{n}\Sigma} \notin [-\delta, \delta]\right]\right\}\\
&= 0 + 2 - 2\Phi(\delta) = \varepsilon,
\end{aligned}$$

thanks to $\Phi(\delta) = 1 - \varepsilon/2$. The proof is completed by exploiting the arbitrariness of ε. ∎

2.3 Proofs for $\alpha = 0$

For $\alpha = 0$, the LLN (2.1.3) and the CLT (2.1.4) were established in [91, Proposition 2 and Theorem 2]. In this section, we present an alternative proof of the CLT (2.1.4), as well as the Berry-Esseen inequality (2.1.6). Technical lemmas and propositions are deferred to Appendix B.

2.3.1 An alternative proof of the CLT (2.1.4)

For $n \in \mathbb{N}$ let $(B_1, \dots, B_n)$ be independent random variables such that B_i is distributed according to a Bernoulli distribution with parameter $\theta/(\theta + i + 1)$, for $i = 1, \dots, n$. Then, it holds

$$K_n = \sum_{i=1}^{n} B_i. \tag{2.3.1}$$

[80, Chapter 3]. Let G_{K_n} be the probability generating function of K_n. From (2.3.1), for $s > 0$

$$G_{K_n}(s) = \frac{[s\theta]_{(n)}}{[\theta]_{(n)}} = \frac{\Gamma(s\theta + n)\Gamma(\theta)}{\Gamma(s\theta)\Gamma(\theta + n)},$$

so that, for $\theta = \lambda n$, with $\lambda > 0$,

$$G_{K_n^{\{n\}}}(s) = \frac{\Gamma(n\,(s\lambda + 1))\Gamma(\lambda n)}{\Gamma(s\lambda n)\Gamma(n\,(\lambda + 1))},$$

Let φ_n be the characteristic function of the random variable

$$\frac{K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda}}{\sqrt{n\mathfrak{s}_{0,\lambda}^2}}.$$

The proof of the CLT (2.1.4) follows by the application of the following Berry-Esseen lemma.

Lemma 2.3.1. *If $\xi \in \mathbb{R}$ satisfies*

$$|\xi| \le \mathcal{C}_0\, \mathfrak{s}_{0,\lambda}\, n^{\delta} \tag{2.3.2}$$

for a positive constant $\mathcal{C}_0$ and some $\delta \in (0, 1/6)$, then there exists a constant $\tilde{c}$ such that

$$\left| \varphi_n(\xi) - e^{-\frac{\xi^2}{2}} \right| \le \tilde{c}\, e^{-\frac{\xi^2}{2}}\, n^{3\delta - \frac{1}{2}}.$$

Proof. We start by recalling a well-known approximation for the Gamma function [69, Equation (5.11.10) and Equation (5.11.11) with $K = 1$ therein], which will be applied to G_{K_n}. For $w \in \mathbb{C}$ and $|\mathrm{ph}(w)| < \pi$, as $|w| \to +\infty$

$$\Gamma(w) = e^{-w}\, w^w \left(\frac{2\pi}{w}\right)^{\frac{1}{2}} [1 + \mathfrak{R}(w)] =: \mathcal{P}(w)\ [1 + \mathfrak{R}(w)]$$

where, for $|\mathrm{ph}(w)| \le \frac{\pi}{3}$,

$$|\mathfrak{R}(w)| \le \frac{1}{2\pi^2|w|}\,[1 + \min\{\sec(\mathrm{ph}(w)), 2\}] \le \frac{3}{2\pi^2|w|}. \tag{2.3.3}$$

Then,

$$\begin{aligned} G_{K_n^{\{n\}}}(s) &= \frac{\Gamma(n(s\lambda+1))}{\Gamma(ns\lambda)}\frac{\Gamma(n\lambda)}{\Gamma(n(\lambda+1))} \\ &= \frac{\mathcal{P}\left(n(s\lambda+1)\right)\,\left[1+\Re(n(s\lambda+1))\right]\;\mathcal{P}\left(n\lambda\right)\,\left[1+\Re(n\lambda)\right]}{\mathcal{P}(ns\lambda)\,\left[1+\Re(ns\lambda)\right]\;\mathcal{P}\left(n(\lambda+1)\right)\,\left[1+\Re(n(\lambda+1))\right]} \\ &= \left(\frac{s(\lambda+1)}{s\lambda+1}\right)^{\frac{1}{2}}\,\left[\frac{(s\lambda+1)^{s\lambda+1}\,\lambda^{\lambda}}{(s\lambda)^{s\lambda}\,(\lambda+1)^{\lambda+1}}\right]^{n}\frac{\left[1+\Re(n(s\lambda+1))\right]\,\left[1+\Re(n\lambda)\right]}{\left[1+\Re(ns\lambda)\right]\,\left[1+\Re(n(\lambda+1))\right]} \end{aligned}$$

for every $s\in\mathbb{C}$ such that $|\mathrm{ph}(s)|\leq\pi$. Based on the above expression for G_{K_n}, we set

$$\mathfrak{R}_1^{(n)}(s) := \frac{\left[1+\Re(n(s\lambda+1))\right]\,\left[1+\Re(n\lambda)\right]}{\left[1+\Re(ns\lambda)\right]\,\left[1+\Re(n(\lambda+1))\right]},$$

$$\mathfrak{R}_2^{(n)}(s) := \left(\frac{s(\lambda+1)}{s\lambda+1}\right)^{\frac{1}{2}}$$

and

$$f(s) := s\lambda\log\left(s\lambda\right) - (s\lambda+1)\log\left(s\lambda+1\right)\;,$$

so that $G_{K_n^{\{n\}}}(s) = \exp\{-n\left[f(s)-f(1)\right]\}\mathfrak{R}_1^{(n)}(s)\mathfrak{R}_2^{(n)}(s)$. Accordingly, we write

$$\begin{aligned} \varphi_n(\xi) &= \exp\left\{-\sqrt{n}\,\frac{i\,\xi\,\mathfrak{m}_{0,\lambda}}{\mathfrak{s}_{0,\lambda}}\right\}\,G_{K_n^{\{n\}}}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right) \\ &= \exp\left\{-\sqrt{n}\,\frac{i\,\xi\,\mathfrak{m}_{0,\lambda}}{\mathfrak{s}_{0,\lambda}}\right\}\,\exp\left\{-n\left[f\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right)-f(1)\right]\right\} \\ &\quad\times\mathfrak{R}_1^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right)\mathfrak{R}_2^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right). \end{aligned}$$

If $\xi\in\mathbb{R}$ satisfies (2.3.2), then [78, Chapter IV, Lemma 5] guarantees that

$$\left|e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1\right| \leq \left|\frac{\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}\right| \leq \mathcal{C}_0\,n^{\delta-\frac{1}{2}}.$$

Hence, by applying Taylor's formula to $f(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}})-f(1)$ and to $(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1)$, we obtain

$$\begin{aligned} \varphi_n(\xi) = \exp\left\{-\sqrt{n}\,\frac{i\,\xi}{\mathfrak{s}_{0,\lambda}}\left[\mathfrak{m}_{0,\lambda}+f'(1)\right]-\frac{\xi^2}{2\,\mathfrak{s}_{0,\lambda}^2}\left[f'(1)+f''(1)\right]\right\}\times \\ \times\,\mathfrak{R}_1^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right)\,\mathfrak{R}_2^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right)\,\mathfrak{R}_3^{(n)}\left(\xi\right), \end{aligned} \tag{2.3.4}$$

where

$$\begin{aligned} \mathfrak{R}_3^{(n)}\left(\xi\right) := \exp\Bigg\{ & n\,f'(1)\left[\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1\right)-\left(\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}-\frac{\xi^2}{n\mathfrak{s}_{0,\lambda}^2}\right)\right] \\ & +\frac{n}{2}\,f''(1)\left[\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1\right)^2+\frac{\xi^2}{n\mathfrak{s}_{0,\lambda}^2}\right] \\ & +\frac{n}{2}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1\right)^3\int_0^1 f'''\left(1+t\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}-1\right)\right)\,(1-t)^2\,\mathrm{d}t\Bigg\}. \end{aligned}$$

By direct computation,

$$f'(1) = -\lambda \log\left(\frac{\lambda+1}{\lambda}\right) = -\mathfrak{m}_{0,\lambda}$$

and

$$f'(1) + f''(1) = -\mathfrak{m}_{0,\lambda} + \frac{\lambda}{\lambda+1} = -\mathfrak{s}^2_{0,\lambda}.$$

Accordingly, for every $\xi \in \mathbb{R}$ that satisfies (2.3.2), we can rewrite φ_n in (2.3.4) as

$$\varphi_n(\xi) = \exp\left\{-\frac{\xi^2}{2}\right\} \mathfrak{R}_1^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right) \mathfrak{R}_2^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right) \mathfrak{R}_3^{(n)}(\xi)\,.$$

By combining (2.3.3) and [78, Chapter IV, Lemma 5], for any ξ satisfying (2.3.2), one has that

$$\mathfrak{R}^{(n)}(\xi) := \mathfrak{R}_1^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right) \mathfrak{R}_2^{(n)}\left(e^{\frac{i\,\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}}}\right) \mathfrak{R}_3^{(n)}(\xi)\,,$$

is continuous and satisfies, as $n \to +\infty$

$$\mathfrak{R}^{(n)}(\xi) = 1 + O\left(n^{3\delta-\frac{1}{2}}\right) \tag{2.3.5}$$

uniformly on compact sets. We refer to Appendix B.1 for a detailed proof of (2.3.5). Then, for every $n \in \mathbb{N}$, one can write

$$n^{-3\delta+\frac{1}{2}}\left|\mathfrak{R}^{(n)}(\xi) - 1\right| \le S(\xi),$$

with $S(\xi) := \sup_{n\in\mathbb{N}} n^{-3\delta+\frac{1}{2}}\left|\mathfrak{R}^{(n)}(\xi) - 1\right|$, and Lemma A.10 guarantees that the function S is continuous. Thus, for every $\xi \in \mathbb{R}$ satisfying (2.3.2), there exists $\tilde{c}$ such that $\left|\mathfrak{R}^{(n)}(\xi) - 1\right| \le \tilde{c}\, n^{3\delta-\frac{1}{2}}$, completing the proof. ■

2.3.2 Proof of the Berry-Esseen inequality (2.1.6)

We start by combining Lemma 2.3.1 with the well-known inequality [78, Chapter V, Theorem 2]

$$\begin{aligned}
\|F_n - \Phi\|_\infty &\le \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,n^\delta} \left|\frac{\varphi_n(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\right| \mathrm{d}\xi + \tilde{\mathcal{C}} n^{-\delta} \\
&\le \int_{-\frac{1}{n}}^{\frac{1}{n}} \left|\frac{\varphi_n(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\right| \mathrm{d}\xi + \frac{\tilde{c}}{n^{-3\delta+\frac{1}{2}}} \int_{\frac{1}{n}}^{+\infty} \frac{e^{-\frac{\xi^2}{2}}}{\xi}\,\mathrm{d}\xi + \tilde{\mathcal{C}} n^{-\delta} \\
&=: I_1 + I_2 + \tilde{\mathcal{C}} n^{-\delta}.
\end{aligned}$$

We consider separately the terms I_1 and I_2. For I_1, we combine the triangle inequality, [78, Chapter IV, Lemma 5] and [20, Section 8.4, Theorem 1, Equation (4)] to write that, in a neighborhood of 0, it holds

$$\begin{aligned}
&\left|\frac{\varphi_n(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\right| \\
&\le \left|\mathbb{E}\left[\frac{K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda}}{\sqrt{n\mathfrak{s}^2_{0,\lambda}}}\right]\right| + \frac{1}{2}\mathbb{E}\left[\left(\frac{K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda}}{\sqrt{n\mathfrak{s}^2_{0,\lambda}}}\right)^2\right]|\xi| + \frac{1}{2}|\xi| + \frac{1}{8}\left|\xi^4\right|.
\end{aligned}$$

From (2.1.1) and (2.1.2), we obtain

$$\left|\mathbb{E}\left[\frac{K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda}}{\sqrt{n\mathfrak{s}_{0,\lambda}^2}}\right]\right| = \frac{1}{\sqrt{n\mathfrak{s}_{0,\lambda}^2}}\left|\mathbb{E}\left[K_n^{\{n\}}\right] - n\mathfrak{m}_{0,\lambda}\right| \leq \frac{S_1}{\sqrt{n}},$$

where $S_1 := \frac{1}{\mathfrak{s}_{0,\lambda}} \sup_{n\in\mathbb{N}} \left|\mathbb{E}\left[K_n^{\{n\}}\right] - n\mathfrak{m}_{0,\lambda}\right|$. An analogous argument allows to prove

$$\mathbb{E}\left[\left(\frac{K_n^{\{n\}} - n\mathfrak{m}_{0,\lambda}}{\sqrt{n\mathfrak{s}_{0,\lambda}^2}}\right)^2\right] \leq \frac{S_2}{n}$$

where $S_2 := \frac{1}{\mathfrak{s}_{0,\lambda}} \sup_{n\in\mathbb{N}} \left|\mathrm{Var}\left(K_n^{\{n\}}\right) - n\left(\mathcal{D} + \mathfrak{m}_{0,\lambda} - 2\mathfrak{m}_{0,\lambda}\mathcal{B}\right)\right|$. Accordingly, we write

$$I_1 \leq \frac{2S_1}{n^{3/2}} + \frac{2S_2 + 1/2}{n^3} + \frac{1}{40\, n^5} \leq \mathcal{C} n^{-3/2}.$$

For I_2, argue as in the proof of the Berry-Esseen theorem for $R_n(z)$ to conclude that

$$I_2 \leq \tilde{c}_1 \log(n)\, n^{3\delta - \frac{1}{2}}.$$

In conclusion,

$$\|F_n - \Phi\|_\infty \leq C_1 n^{-\frac{3}{2}} + C_2 \log(n)\, n^{-\frac{1}{2}+3\delta} + \tilde{\mathcal{C}} n^{-\delta}$$

for some positive constants C_1 and C_2. To conclude, it is easy to see that for every $\delta \in (0, 1/6)$, $\min\left(\delta, -3\delta + \frac{1}{2}, \frac{3}{2}\right) \geq \frac{1}{8}$, which produces the rate $\log(n)\, n^{-1/8}$ in (2.1.6).

As a final remark, note that our proof was not designed to achieve rate optimality, but merely to provide a quantitative refinement of the convergence result of the CLT. This may account for the (notable) slowness of this rate, which we believe could be improved with the use of more refined techniques, at the price of an increasing complexity.

2.4 Proof of the strong LLN (2.1.5)

To prove the strong LLN, we show complete convergence of the sequence $(K_n^{\{n\}}/n)_{n\in\mathbb{N}}$. Consider the identity

$$\frac{K_n^{\{n\}} - n\mathfrak{m}_{\alpha,\lambda}}{n} = \frac{K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]}{n} + \frac{\mathbb{E}\left[K_n^{\{n\}}\right] - n\mathfrak{m}_{\alpha,\lambda}}{n}.$$

In particular, as $n \to +\infty$, the almost sure convergence to 0 of the first term on the right-hand side follows from a standard application of the Borel–Cantelli lemma, after proving that

$$\sum_{n=0}^{+\infty} P\left(\left|K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right| > n\varepsilon\right) < +\infty$$

holds for any $\varepsilon > 0$. This follows from Markov's inequality (with exponent $p = 4$) with the fact that, as $n \to +\infty$

$$\mathbb{E}\left[\left(K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right)^4\right] = O\left(n^2\right). \tag{2.4.1}$$

We refer to Appendix C.1 for details on (2.4.1). This completes the proof of the strong LLN (2.1.5).

Chapter 3

Strong law of large numbers and joint central limit theorem for the blocks of given size in the "large θ" regime

In this chapter we give an alternative proof of Theorem 2.1.1 based on the construction of a martingale related to the partition structure 1.3.1, and then extend the argument to establish a joint CLT for

$$\mathbf{K}_{d,n} = (K_n, K_{1,n}, ..., K_{d,n})$$

in the regime $\theta = \lambda n$. The proof adapts the arguments already used in [9] to the "large θ" setting, requiring the use of asymptotic results for triangular arrays of martingales, rather than a single martingale. Hence, the proof relies heavily on the sequential construction 1.3.1, and it calls for establishing asymptotic results not only for the variable $\mathbf{K}_n^{\{n\}}$ (which was the case in the previous chapter), but also for $\mathbf{K}_{d,h}^{\{n\}}$ for $d, h \in \{1, ..., n\}$. For this reason, in this chapter we return to the full notation of section 1.3.1, paying the price of a heavier notation to avoid confusion.

3.1 An alternative proof of Theorem 2.1.1 via a martingale construction

Recall that $\mathcal{F}_h^{\{n\}} = \sigma\left(\mathbf{K}_1^{\{n\}}, ..., \mathbf{K}_h^{\{n\}}\right)$. For what concerns K_n, the sequential construction of the partition of section 1.3.1 can be equivalently stated by saying that for every $n \in \mathbb{N}$ and $h \in \{1, ..., n\}$, there exist a random variable $\xi_h^{(n)}$ defined on the same probability space as $K_1^{\{n\}}, ..., K_{h-1}^{\{n\}}$ such that

$$K_h^{\{n\}} = K_{h-1}^{\{n\}} + \xi_h^{\{n\}} \tag{3.1.1}$$

and

$$P\left[\xi_h^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right] = \begin{cases} p_{h-1}^{\{n\}} & \text{if } k = 1 \\ 1 - p_{h-1}^{\{n\}} & \text{if } k = 0 \end{cases}$$

with

$$p_{h-1}^{\{n\}} = \frac{\alpha K_{h-1}^{\{n\}} + \lambda n}{\lambda n + h - 1}.$$

In particular,

$$\begin{aligned}\mathbb{E}\left[K_h^{\{n\}} \mid \mathcal{F}_h^{\{n\}}\right] &= K_{h-1}^{\{n\}} + p_{h-1}^{\{n\}} \\ &= \gamma_{h-1}^{\{n\}} K_{h-1}^{\{n\}} + \beta_{h-1}^{\{n\}}\end{aligned} \tag{3.1.2}$$

where

$$\gamma_{h-1}^{\{n\}} = \frac{\lambda n + h - 1 + \alpha}{\lambda n + h - 1} \quad \text{and} \quad \beta_{h-1}^{\{n\}} = \frac{\lambda n}{\lambda n + h - 1}.$$

The following lemma describes the asymptotic behavior of the variables $K_{\lfloor xn \rfloor}^{\{n\}}$ for $x \in [0,1]$ and is a cardinal point in the proof of the CLT.

Lemma 3.1.1. *For $x \in [0,1]$, define*

$$m_{\alpha,\lambda}(x) = \begin{cases} \frac{\lambda}{\alpha}\left[\left(\frac{\lambda+x}{\lambda}\right)^{\alpha} - 1\right] & \text{if } \alpha \in (0,1) \\ \lambda \log\left(\frac{\lambda+x}{\lambda}\right) & \text{if } \alpha = 0 \end{cases}$$

Then, as $n \to +\infty$

(i)

$$\frac{\mathbb{E}\left[K_{\lfloor xn \rfloor}^{\{n\}}\right]}{n} \to m_{\alpha,\lambda}(x)$$

uniformly on $[0,1]$.

(ii) the following SLLN holds uniformly for $x \in [0,1]$:

$$\frac{K_{\lfloor xn \rfloor}^{\{n\}}}{n} \xrightarrow{a.s.} m_{\alpha,\lambda}(x) \tag{3.1.3}$$

Proof of Lemma 3.1.1. The structure of the proof is the same for the cases $\alpha = 0$ and $\alpha \in (0,1)$, but it involves different calculations. For this reason, we present here only the general idea; the complete calculations, differentiated for the two cases, are deferred to appendix D.1.

We prove by direct calculation that

$$\lim_{n\to\infty} \sup_{x\in[0,1]} \frac{1}{n^{s-2}} \cdot \left| \mathbb{E}\left[(K_{\lfloor xn \rfloor}^{\{n\}})_{\downarrow s}\right] - \left(n^s \cdot (m_{\alpha,\lambda}(x))^s + n^{s-1} \cdot S_s(x)\right) \right| = c$$

for some constant c and some explicit function $S_s(x)$ not depending on n. In particular, for $s = 1$, this rewrites as

$$\lim_{n\to+\infty} \sup_{x\in[0,1]} \left| \frac{\mathbb{E}\left[K_{\lfloor xn \rfloor}^{\{n\}}\right]}{n} - m_{\alpha,\lambda}(x) \right| = 0$$

Proving point (i).

To prove the strong law of large numbers 3.1.3, we show complete convergence of the sequence $K^{\{n\}}_{\lfloor xn\rfloor}/n$ by showing that $\mathbb{E}\left[\left(K^{\{n\}}_{\lfloor xn\rfloor} - \mathbb{E}\left[K^{\{n\}}_{\lfloor xn\rfloor}\right]\right)^4\right] = O(n^2)$ uniformly for $x \in [0,1]$. Write

$$\begin{aligned}&\mathbb{E}\left[\left(K^{\{n\}}_{\lfloor xn\rfloor} - \mathbb{E}\left[K^{\{n\}}_{\lfloor xn\rfloor}\right]\right)^4\right]\\ &\quad= \sum_{k=0}^{4}(-1)^{4-k}\binom{4}{k}\left(\mathbb{E}\left[K^{\{n\}}_{\lfloor xn\rfloor}\right]\right)^{4-k}\sum_{s=0}^{k} S(k,s)\,\mathbb{E}\left[\left(K^{\{n\}}_{\lfloor xn\rfloor}\right)_{\downarrow s}\right],\end{aligned}$$

where $S(k,s)$ denote the Stirling number of the second kind. Expanding such expression and computing the expansion of powers of $\mathbb{E}\left[K^{\{n\}}_{r,\lfloor xn\rfloor}\right]$ yields, after tedious but straightforward calculations,

$$\lim_{n\to+\infty}\sup_{x\in[0,1]}\frac{1}{n^2}\left|\mathbb{E}\left[\left(K^{\{n\}}_{\lfloor xn\rfloor} - \mathbb{E}\left[K^{\{n\}}_{\lfloor xn\rfloor}\right]\right)^4\right] - \left(n^4\cdot\mathcal{A}_{\alpha,\lambda}(x) + n^3\cdot\mathcal{B}_{\alpha,\lambda}(x)\right)\right| = c$$

for some constant c, where

$$\mathcal{A}_{\alpha,\lambda}(x) = \left(m_{\alpha,\lambda}(x)\right)^4\cdot(1-4+6-3) = 0$$

$$\begin{aligned}\mathcal{B}_{\alpha,\lambda}(x) &= m^3_{\alpha,\lambda}(x)\left[(-12+12-4)S_1(x) + (6-12+6)\right] + 6\,m^2_{\alpha,\lambda}(x)\,S_2(x)\\ &\quad + m_{\alpha,\lambda}(x)\,S_3(x) + S_4(x)\\ &= 0\end{aligned}$$

where the last equality is proven in detail in appendix D.1. Therefore,

$$\lim_{n\to+\infty}\sup_{x\in[0,1]}\frac{1}{n^2}\,\mathbb{E}\left[\left(K^{\{n\}}_{\lfloor xn\rfloor} - \mathbb{E}\left[K^{\{n\}}_{\lfloor xn\rfloor}\right]\right)^4\right] = c$$

and the proof is concluded.

■

3.1.1 Proof of the expansion (2.1.1) and of the SLLN (2.1.5)

Upon noting that

$$\mathfrak{m}_{\alpha,\lambda} = m_{\alpha,\lambda}(1) = \begin{cases}\frac{\lambda}{\alpha}\left[\left(\frac{\lambda+1}{\lambda}\right)^{\alpha} - 1\right] & \text{if } \alpha\in(0,1)\\ \lambda\log\left(\frac{\lambda+1}{\lambda}\right) & \text{if } \alpha = 0\end{cases}$$

apply Lemma 3.1.1 for $x = 1$.

3.1.2 Proof of the CLT (2.1.4)

3.1.2.1 An interesting martingale

For $n\in\mathbb{N}$ and $h\in\{1,\dots,n\}$, define

$$a^{\{n\}}_h = \prod_{k=1}^{h-1}\left(\gamma^{\{n\}}_k\right)^{-1} \quad\text{and}\quad A^{\{n\}}_h = \sum_{k=1}^{h-1} a^{\{n\}}_{k+1}\beta^{\{n\}}_k. \tag{3.1.4}$$

and

$$M_h^{\{n\}} = a_h^{\{n\}} K_h^{\{n\}} - A_h^{\{n\}}. \tag{3.1.5}$$

Denote by

$$\Delta_h^{\{n\}} = M_h^{\{n\}} - M_{h-1}^{\{n\}}$$

the increments of $M^{\{n\}}$.

Lemma 3.1.2. *For every $n \in \mathbb{N}$, $\left(M_h^{\{n\}}\right)_{h\in\{1,\dots,n\}}$ is a martingale with respect to the filtration $\left(\mathcal{F}_h^{\{n\}}\right)_{h\in\{1,\dots n\}}$.*

Proof. It follows from definition (3.1.4) that

$$a_{h-1}^{\{n\}} = \gamma_{h-1}^{\{n\}} \cdot a_h^{\{n\}}$$

and

$$A_{h-1}^{\{n\}} = A_h^{\{n\}} - a_h^{\{n\}} \beta_{h-1}^{\{n\}}$$

Hence,

$$\begin{aligned}
\Delta_h^{\{n\}} &= a_h^{\{n\}} K_h^{\{n\}} - A_h^{\{n\}} - a_{h-1}^{\{n\}} K_{h-1}^{\{n\}} + A_{h-1}^{\{n\}} \\
&= a_h^{\{n\}} \left[K_{h-1}^{\{n\}} + \xi_h^{\{n\}}\right] - A_h^{\{n\}} - \gamma_{h-1} a_h^{\{n\}} K_{h-1}^{\{n\}} + A_h^{\{n\}} - \gamma_{h-1} a_h^{\{n\}} \beta_{h-1}^{\{n\}} \\
&= a_h^{\{n\}} \left[\left(1 - \gamma_{h-1}^{\{n\}}\right) K_{h-1}^{\{n\}} - \beta_{h-1}^{\{n\}} + \xi_h^{\{n\}}\right] \\
&= a_h^{\{n\}} \cdot \left[\xi_h^{(n)} - p_{h-1}^{\{n\}}\right].
\end{aligned} \tag{3.1.6}$$

Taking conditional expectations on both side we obtain

$$\mathbb{E}\left[\Delta_h^{\{n\}} \,|\, \mathcal{F}_{h-1}^{\{n\}}\right] = a_h^{\{n\}} \cdot \mathbb{E}\left[\xi_h^{(n)} \,|\, \mathcal{F}_{h-1}^{\{n\}}\right] - p_{h-1}^{\{n\}} = 0,$$

which concludes the proof. ∎

Now denote by

$$\langle M^{\{n\}} \rangle = \left(\langle M^{\{n\}} \rangle_1, ..., \langle M^{\{n\}} \rangle_n\right)$$

the increasing process of $M^{\{n\}}$, i.e.

$$\langle M^{\{n\}} \rangle_k = \sum_{h=1}^{k} \mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^2 \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right]. \tag{3.1.7}$$

the following lemma describes the asymptotic behavior of $\langle M^{\{n\}} \rangle_n$ as $n \to +\infty$.

Lemma 3.1.3. *As $n \to +\infty$,*

$$\frac{\langle M^{\{n\}} \rangle_n}{n} \xrightarrow{a.s.} s_{\alpha,\lambda}^2$$

where

$$s_{\alpha,\lambda}^2 = \begin{cases} \frac{\lambda}{\alpha} \cdot \left[\frac{\lambda+1-\alpha}{\lambda+1} - \left(\frac{\lambda}{\lambda+1}\right)^{\alpha}\right] & \text{if } \alpha \in (0,1) \\ \log\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda+1} & \text{if } \alpha = 0 \end{cases}$$

Proof. It follows from (3.1.6) that

$$\begin{aligned}\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^2 \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] &= \left(a_h^{\{n\}}\right)^2 \cdot p_{h-1}^{\{n\}}\left(1-p_{h-1}^{\{n\}}\right)\\ &= \left[\frac{(\lambda n)_{\uparrow h}\,(\lambda n+\alpha)}{(\lambda n+\alpha)_{\uparrow h}\,\lambda n}\right]^2 \cdot \frac{\alpha K_{h-1}^{\{n\}}+\lambda n}{\lambda n+h-1}\cdot\frac{h-1-\alpha K_{h-1}^{\{n\}}}{\lambda n+h-1}\\ &= \left[\frac{(\lambda n)_{\uparrow h-1}}{(\lambda n+\alpha+1)_{\uparrow h-1}}\right]^2 \cdot \frac{\alpha K_{h-1}^{\{n\}}+\lambda n}{\lambda n}\cdot\frac{h-1-\alpha K_{h-1}^{\{n\}}}{\lambda n}\end{aligned}$$

Now we split $\langle M^{\{n\}}\rangle_n$ into the sum of a deterministic term and a random term:

$$\begin{aligned}&\langle M^{\{n\}}\rangle_n =\\ &= \sum_{h=1}^{n}\left\{\left[\frac{(\lambda n)_{\uparrow h-1}}{(\lambda n+\alpha+1)_{\uparrow h-1}}\right]^2 \cdot \frac{\alpha K_{h-1}^{\{n\}}+\lambda n}{\lambda n}\cdot\frac{h-1-\alpha K_{h-1}^{\{n\}}}{\lambda n}\right\}\\ &= \sum_{h=1}^{n}\left\{\left[\frac{(\lambda n)_{\uparrow h-1}}{(\lambda n+\alpha+1)_{\uparrow h-1}}\right]^2 \cdot \left[\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)+1\right]\cdot\left[\frac{1}{\lambda}\frac{h-1}{n}-\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)\right]\right\}\\ &\quad+\sum_{h=1}^{n}\left\{\left[\frac{(\lambda n)_{\uparrow h-1}}{(\lambda n+\alpha+1)_{\uparrow h-1}}\right]^2 \cdot \left[\frac{\alpha}{\lambda}\left(\frac{K^{\{n\}}_{\frac{h-1}{n}\cdot n}}{n}-m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)\right)+1\right]\times\right.\\ &\qquad\left.\times\left[\frac{1}{\lambda}\frac{h-1}{n}-\frac{\alpha}{\lambda}\left(\frac{K^{\{n\}}_{\frac{h-1}{n}\cdot n}}{n}-m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)\right)\right]\right\}\\ &= \mathcal{S}_1^{\alpha,\lambda}(n)+\mathcal{S}_2^{\alpha,\lambda}(n)\end{aligned}$$

The term $\mathcal{S}_1^{\alpha,\lambda}(n)$ is deterministic and

$$\begin{aligned}\mathcal{S}_1^{\alpha,\lambda}(n) &= \sum_{h=1}^{n}\left\{\left[\frac{(\lambda n)_{\uparrow h-1}}{(\lambda n+\alpha+1)_{\uparrow h-1}}\right]^2 \cdot \left[\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)+1\right]\right.\\ &\qquad\left.\times\left[\frac{1}{\lambda}\frac{h-1}{n}-\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)\right]\right\}\\ &= \sum_{h=1}^{n}\left\{\left[\frac{\Gamma\left(\left(\lambda+\frac{h-1}{n}\right)n\right)}{\Gamma(\lambda n)}\cdot\frac{\Gamma(\lambda n+\alpha+1)}{\Gamma\left(\left(\lambda+\frac{h-1}{n}\right)n+\alpha+1\right)}\right]^2\right.\\ &\qquad\left.\times\left[\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)+1\right]\cdot\left[\frac{1}{\lambda}\frac{h-1}{n}-\frac{\alpha}{\lambda}m_{\alpha,\lambda}\left(\frac{h-1}{n}\right)\right]\right\}\\ &= n\cdot\sum_{h=1}^{n} f_{\alpha,\lambda}^{\{n\}}\left(\frac{h-1}{n}\right)\frac{1}{n},\end{aligned}$$

where, for $x\in[0,1]$,

$$f_{\alpha,\lambda}^{\{n\}}(x) = \left[\frac{\Gamma\left((\lambda+x)\,n\right)}{\Gamma(\lambda n)}\cdot\frac{\Gamma(\lambda n+\alpha+1)}{\Gamma\left((\lambda+x)\,n+\alpha+1\right)}\right]^2\cdot\left[\frac{\alpha}{\lambda}m_{\alpha,\lambda}(x)+1\right]\cdot\left[\frac{x}{\lambda}-\frac{\alpha}{\lambda}m_{\alpha,\lambda}(x)\right]$$

Thus, $\mathcal{S}_1^{\alpha,\lambda}(n)/n$ can be viewed as a Riemann sum for the function $f_{\alpha,\lambda}^{\{n\}}$ on $[0,1]$. Using again [89, Equation 1] we obtain that

$$\lim_{n\to+\infty}\sup_{x\in[0,1]}\left|\left[\frac{\Gamma\left((\lambda+x)\,n\right)}{\Gamma(\lambda n)}\cdot\frac{\Gamma(\lambda n+\alpha+1)}{\Gamma\left((\lambda+x)\,n+\alpha+1\right)}\right]-\left(\frac{\lambda}{\lambda+x}\right)^{\alpha+1}\right| = 0.$$

This immediately implies that, as $n \to +\infty$, $f_{\alpha,\lambda}^{\{n\}} \longrightarrow f_{\alpha,\lambda}$ uniformly on $[0,1]$, where

$$f_{\alpha,\lambda}(x) = \left(\frac{\lambda}{\lambda+x}\right)^{\alpha+1} - \left(\frac{\lambda}{\lambda+x}\right)^{2}$$

By definition and elementary properties of the Riemann integral,

$$\lim_{n\to+\infty} \frac{\mathcal{S}_1^{\alpha,\lambda}(n)}{n} = \int_0^1 f_{\alpha,\lambda}(x)\,\mathrm{d}x$$

and by a simple calculation

$$\int_0^1 f_{\alpha,\lambda}(x)\,\mathrm{d}x = \begin{cases} \frac{\lambda}{\alpha}\cdot\left[\frac{\lambda+1-\alpha}{\lambda+1} - \left(\frac{\lambda}{\lambda+1}\right)^{\alpha}\right] & \text{if } \alpha \in (0,1) \\ \lambda\log\left(\frac{\lambda+1}{\lambda}\right) - \frac{\lambda}{\lambda+1} & \text{if } \alpha = 0 \end{cases} = s_{\alpha,\lambda}^2.$$

It remains to prove that $\mathcal{S}_2^{\alpha,\lambda}(n)/n \xrightarrow{a.s.} 0$ as $n \to +\infty$. If $\alpha = 0$, $\mathcal{S}_2^{0,\lambda}(n)$ is deterministic and

$$\begin{aligned}\mathcal{S}_2^{0,\lambda}(n) &= \sum_{h=1}^{n} \frac{(\lambda n)^2}{(\lambda n + h - 1)^2}\cdot\frac{h-1}{\lambda n} \\ &= \psi((\lambda+1)n) - \psi(\lambda n+1) + \lambda n\cdot\left[\psi^{(1)}((\lambda+1)n) - \psi^{(1)}(\lambda n+1)\right]\end{aligned}$$

Applying the asymptotic expansions of the digamma [69, Equation (5.11.2)] and trigamma functions [69, Equation (5.15.8)] we obtain

$$\lim_{n\to+\infty}\sup_{x\in[0,1]}\left|\mathcal{S}_2^{0,\lambda}(n) - \left[\log\left(\frac{\lambda}{\lambda+1}\right) + \frac{1}{\lambda+1}\right]\right| = 0$$

which indeed implies

$$\frac{\mathcal{S}_2^{0,\lambda}(n)}{n} \longrightarrow 0$$

uniformly for $x \in [0,1]$. When $\alpha \in (0,1)$, apply the triangular inequality together with lemma 3.1.1, obtaining

$$\lim_{n\to+\infty}\sup_{x\in[0,1]}\left|\frac{\mathcal{S}_2^{\alpha,\lambda}(n)}{n}\right| \le \left|\frac{\mathcal{S}_2^{\alpha,\lambda}(n) - \mathcal{S}_2^{0,\lambda}(n)}{n}\right| + \left|\frac{\mathcal{S}_2^{0,\lambda}(n)}{n}\right| = 0$$

Hence, $\mathcal{S}_2^{\alpha,\lambda}(n)/n \longrightarrow 0$ uniformly for $x \in [0,1]$, concluding the proof. ■

3.1.2.2 Completing the proof

The following lemma is an immediate consequence of [55, Corollary 3.1]:

Lemma 3.1.4. *Consider an integer sequence $h_n \to +\infty$ as $n \to +\infty$. Assume that the following two conditions are satisfied,*

(a)

$$\frac{\langle M^{\{n\}}\rangle_{h_n}}{h_n} \xrightarrow{a.s.} g_{\alpha,\lambda}$$

for some positive $g_{\alpha,\lambda}$

(b) For every $\varepsilon > 0$,

$$\frac{1}{h_n}\sum_{h=1}^{h_n}\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^2\cdot\mathbb{1}_{\left\{\left(\Delta_h^{\{n\}}\right)^2>h_n\,\varepsilon\right\}}\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]\xrightarrow{a.s.}0$$

Then, we have

$$\frac{M_{h_n}^{\{n\}}}{\sqrt{h_n}}\xrightarrow{w}N(0,g_{\alpha,\lambda}) \tag{3.1.8}$$

We now show that, for the sequence $h_n = n$, conditions (a) and (b) of lemma 3.1.4 are satisfied with $g_{,\alpha,\lambda} = s_{\alpha,\lambda}$. Condition (a) is equivalent to $\frac{1}{n\,s_{\alpha,\lambda}^2}\sum_{h=1}^{n}\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^2\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]\xrightarrow{a.s.}1$ as $n\to+\infty$, and hence it is immediately implied by lemma 3.1.3. For condition (b), we can bound the sum by

$$\begin{aligned}
\mathcal{S}_n &:= \frac{1}{n\,s_{\alpha,\lambda}^2}\sum_{h=1}^{n}\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^2\cdot\mathbb{1}_{\left\{\left|\left(\Delta_h^{\{n\}}\right)^2\right|>n\,\mathfrak{s}_{\alpha,\lambda}^2\,\varepsilon\right\}}\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]\\
&\le \frac{1}{n\,s_{\alpha,\lambda}^2}\sum_{h=1}^{n}\mathbb{E}\left[\frac{\left(\Delta_h^{\{n\}}\right)^4}{\left(\Delta_h^{\{n\}}\right)^2}\cdot\mathbb{1}_{\left\{\left|\Delta_h^{\{n\}}\right|>\sqrt{n}\,\mathfrak{s}_{\alpha,\lambda}\,\varepsilon\right\}}\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]\\
&\le \frac{1}{n^2\,s_{\alpha,\lambda}^4\,\varepsilon^2}\sum_{h=1}^{n}\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^4\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]\\
&\le \frac{1}{n^2\,s_{\alpha,\lambda}^4\,\varepsilon^2}\sum_{h=1}^{n}\left(a_h^{\{n\}}\right)^4\cdot p_{h-1}^{\{n\}}\\
&=: \mathfrak{S}_n
\end{aligned}$$

where the last inequality holds because

$$\mathbb{E}\left[\left(\Delta_h^{\{n\}}\right)^4\,\middle|\,\mathcal{F}_{h-1}^{\{n\}}\right]=\left(a_h^{\{n\}}\right)^4\cdot p_{h-1}^{\{n\}}\left[1-4p_{h-1}^{\{n\}}+6\left(p_{h-1}^{\{n\}}\right)^2-3\left(p_{h-1}^{\{n\}}\right)^3\right]$$

and the function $x\mapsto 1-4x+6x^2-3x$ has its maximum on $[0,1]$ in $x=0$, where it takes value 1. Now,

$$\sum_{h=1}^{n}\left(a_h^{\{n\}}\right)^4\cdot p_{h-1}^{\{n\}}=n\cdot\sum_{h=1}^{n}f_n\left(\frac{h-1}{n}\right)\cdot\frac{1}{n}$$

with

$$f_n(x)=\left[\frac{\Gamma((\lambda+x)n)\,\Gamma(\lambda n+\alpha+1)}{\Gamma(\lambda n)\,\Gamma((\lambda+x)n+\alpha+1)}\right]^4\cdot\left[\frac{\alpha K_{\lfloor xn\rfloor}}{\lambda n}+1\right],$$

and lemma 3.1.1 entails that as $n\to+\infty$, $f_n\longrightarrow f$ uniformly on $[0,1]$, where

$$f(x)=\left(\frac{\lambda}{\lambda+x}\right)^{4(\alpha+1)}\cdot\left[\frac{\alpha}{\lambda}\,m_{\alpha,\lambda}(x)+1\right].$$

Then, letting $\mathcal{I}_f=\int_0^1 f(x)\,\mathrm{d}x$,

$$\mathfrak{S}_n=\frac{1}{n^2\,s_{\alpha,\lambda}^4\,\varepsilon^2}\cdot\left[n\,\mathcal{I}_f+O(1)\right]=O\left(\frac{1}{n}\right)$$

and this proves point (b).

It then follows from Lemma 3.1.4 that

$$\frac{M_n^{\{n\}}}{\sqrt{n \cdot s_{\alpha,\lambda}^2}} \xrightarrow{w} N(0,1)$$

Rewrite $M_n^{\{n\}}/\sqrt{n \cdot s_{\alpha,\lambda}^2}$ as

$$\frac{M_n^{\{n\}}}{\sqrt{n \cdot s_{\alpha,\lambda}^2}} = \frac{a_n^{\{n\}} K_n^{\{n\}} - A_n^{\{n\}}}{\sqrt{n \cdot s_{\alpha,\lambda}^2}} = \frac{K_n^{\{n\}} - \left(a_n^{\{n\}}\right)^{-1} A_n^{\{n\}}}{\sqrt{n \cdot \left(a_n^{\{n\}}\right)^{-2} s_{\alpha,\lambda}^2}}.$$

It is easy to show that for all $n \in \mathbb{N}$ and for all $h \in \{1, ..., n\}$,

$$\mathbb{E}\left[K_h^{\{n\}}\right] = \left(a_h^{\{n\}}\right)^{-1} \left[1 + A_h^{\{n\}}\right].$$

Further,

$$\left(a_n^{\{n\}}\right)^{-1} = \frac{(\lambda n + \alpha + 1)_{\uparrow(n-1)}}{(\lambda n + 1)_{\uparrow(n-1)}} = \frac{\Gamma((1+\lambda)n + \alpha + 1) \cdot \Gamma(\lambda n + 1)}{\Gamma(\lambda n + \alpha + 1) \cdot \Gamma((1+\lambda)n + 1}$$

and we can apply the standard asymptotics for ratios of Gamma functions [89, Equation 1], to obtain that, for every $n \in \mathbb{N}$,

$$\lim_{n \to +\infty} \left| \left(a_n^{\{n\}}\right)^{-1} - \left(\frac{\lambda+1}{\lambda}\right)^{\alpha} \right| \leq \lim_{n \to +\infty} \frac{1}{n} \cdot \frac{(\alpha^2 + \alpha)}{4\lambda} = 0$$

Hence,

$$\lim_{n \to +\infty} \left(a_n^{\{n\}}\right)^{-2} = \mathfrak{a}_{\alpha,\lambda}^2 = \left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}$$

By means of these identities and of lemma 3.1.1 we obtain

$$\lim_{n \to +\infty} \frac{1}{\sqrt{n}} \cdot \left| \left(a_n^{\{n\}}\right)^{-1} A_n^{\{n\}} - n \cdot \mathfrak{m}_{\alpha,\lambda} \right| = 0$$

and since it can be easily verified that

$$\mathfrak{a}_{\alpha,\lambda}^2 \cdot s_{\alpha,\lambda}^2 = \mathfrak{s}_{\alpha,\lambda}^2,$$

the proof is concluded by means of Slutsky's lemma.

3.2 A SLLN and a joint CLT for $\left(K_n^{\{n\}}, K_{1,n}^{\{n\}}, ..., K_{d,n}^{\{n\}}\right)$

The proof of the previous section can be generalized to the proof of a joint central limit theorem for the vector

$$\mathbf{K}_{d,n}^{\{n\}} = \left(K_n^{\{n\}}, K_{1,n}^{\{n\}}, ..., K_{d,n}^{\{n\}}\right)^T.$$

While this requires lengthier computations, and the use of multivariate analogues of the univariate theorems of the previous section, the conceptual lines of the proof remain identical.

3.2.1 Main result

For $r \in \mathbb{N}, r \geq 1$, define

$$\mathfrak{m}_{r;\alpha,\lambda} = \frac{p_\alpha(r)}{\alpha} \lambda^{1-\alpha}(1+\lambda)^{\alpha-r} \tag{3.2.1}$$

where, for $k \in \mathbb{N}$,

$$p_\alpha(k) = \begin{cases} 0 & \text{if } k = 0 \\ \frac{\alpha(1-\alpha)_{\uparrow k-1}}{k!} & \text{if } k \geq 1 \end{cases}$$

and for $d \in \mathbb{N}$ let

$$\mathfrak{M}_{d,\alpha,\lambda} = (\mathfrak{m}_{\alpha,\lambda}, \mathfrak{m}_{1;\alpha,\lambda}, ..., \mathfrak{m}_{d;\alpha,\lambda})^T \tag{3.2.2}$$

Further, let

$$(\Sigma_{d,\alpha,\lambda})_{i,j} = \left(\frac{\lambda+1}{\lambda}\right)^{i+j-2-2\alpha} (\Gamma_{d,\alpha,\lambda})_{i,j} \tag{3.2.3}$$

where $\Gamma_{d,\alpha,\lambda}$ is a symmetric matrix defined by

$$(\Gamma_{d,\alpha,\lambda})_{i,j} = \begin{cases} s^2_{i-1;\alpha,\lambda} & i = j \\ \frac{\lambda}{\alpha-1} - \frac{(\lambda+1)}{\alpha-1} \cdot \left(\frac{\lambda}{\lambda+1}\right)^\alpha - \left(\frac{\lambda}{\lambda+1}\right)^2 + \frac{1-\alpha}{2\,(\lambda+1)} \cdot H(1,2) & i = 1, j = 2 \\ \lambda^{2-j}\left[-c_{j-2}\frac{(\lambda+1)^{-3}}{j-1} \cdot H(j-3;j-1) + c_{j-1}\frac{(\lambda+1)^{-2}}{j} \cdot H(j-1;j)\right] & i = 1, j \geq 3 \\ -\lambda^{1+\alpha-2i} c_{i-1}\frac{(\lambda+1)^{i-1-\alpha}}{i} \cdot H(i-\alpha;i) - \lambda^{2-2i} \cdot \varrho_{\alpha,\lambda}(i-2,i-1) & i \geq 2, j = i+1 \\ -\lambda^{3-i-j} \cdot \varrho_{\alpha,\lambda}(i-2,j-2) & i \geq 2, j \geq i+2 \end{cases} \tag{3.2.4}$$

with

$$H(B,C) = {}_2F_1\left(1, B; C\,; -\frac{1}{\lambda}\right)$$

where ${}_2F_1(a,b;c;z)$ stands for the Gauss hypergeometric function defined, for all $z \in \mathbb{C}$ where $|z| < 1$, by

$${}_2F_1(a,b;c;z) = \sum_{r=0}^{\infty} \frac{(a)_{r\uparrow}(b)_{r\uparrow}}{(c)_{r\uparrow}} \frac{z^r}{r!};$$

see [69, Equation (15.2.1)]. Further, for $r, s \in \mathbb{N}$,

$$c_r = \frac{(1-\alpha)_{r\uparrow}}{r!} = \frac{r}{\alpha} \cdot p_\alpha(r)$$

and

$$\begin{aligned} \varrho_{\alpha,\lambda}(r,s) = {} & c_r c_s \frac{(\lambda+1)^{-4}}{r+s+1} H(r+s+2, r+s+1) \\ & - (c_{r+1}c_s + c_r c_{s+1}) \frac{(\lambda+1)^{-3}}{r+s+2} H(r+s+3, r+s+2) \\ & + c_{r+1}c_{s+1} \frac{(\lambda+1)^{-2}}{r+s+3} H(r+s+4, r+s+3). \end{aligned}$$

Finally,

$$s^2_{0;\alpha,\lambda} = \begin{cases} \frac{\lambda}{\alpha} \cdot \left[\frac{\lambda+1-\alpha}{\lambda+1} - \left(\frac{\lambda}{\lambda+1}\right)^{\alpha}\right] & \text{if } \alpha \in (0,1) \\ \log\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda+1} & \text{if } \alpha = 0, \end{cases} \tag{3.2.5}$$

and for $i \geq 2$,

$$\begin{aligned} s^2_{i-1;\alpha,\lambda} = \lambda^{\alpha+2-2i} \Bigg[& c_{i-2} \frac{(\lambda+1)^{i-3-\alpha}}{i-1} \cdot H(i-\alpha-2; i-1) \\ & + c_{i-1} \frac{(\lambda+1)^{i-2-\alpha}}{i} \cdot H(i-\alpha; i) \Bigg] - \lambda^{3-2i} \cdot \varrho_{\alpha,\lambda}(i-2, i-2). \end{aligned} \tag{3.2.6}$$

The following theorem is the main result of this chapter:

Theorem 3.2.1. *Fix $d \geq 1$. For $n \in \mathbb{N}$, let $\mathbf{K}_{d,n}^{\{n\}}$ be the vector containing the number of partition blocks and the number of partition blocks of sizes $1, ..., d$ under the Ewens-Pitman model with parameters $\alpha \in [0,1)$ and $\theta = \lambda n$, with $\lambda > 0$. Then, as $n \to +\infty$ there hold:*

i)

$$\mathbb{E}\left[\mathbf{K}_{d,n}^{\{n\}}\right] = n\,\mathfrak{M}_{d,\alpha,\lambda} + O(1)$$

ii) (SLLN)

$$\frac{\mathbf{K}_{d,n}^{\{n\}}}{n} \xrightarrow{a.s.} \mathfrak{M}_{d,\alpha,\lambda};$$

iii) (CLT)

$$\frac{\mathbf{K}_{d,n}^{\{n\}} - n\,\mathfrak{M}_{d,\alpha,\lambda}}{\sqrt{n}} \xrightarrow{w} \mathcal{N}(\mathbf{0}, \Sigma_{d,\alpha,\lambda}).$$

where $\mathfrak{M}_{d,\alpha,\lambda}$ is as in (3.2.2) and $\Sigma_{d,\alpha,\lambda}$ as in (3.2.4).

Theorem 2.1.1 can be recovered as a corollary of theorem 3.2.1, by marginalization, upon noting that $(\Sigma_{d,\alpha,\lambda})_{1,1} = \mathfrak{s}^2_{\alpha,\lambda}$.

We conclude this section by proving the following analogue of lemma 3.1.1 for $K_{r,\lfloor xn \rfloor}^{\{n\}}$ for $r \geq 1$. This result is an essential preliminary to the proof of theorem 3.2.1.

Lemma 3.2.2. *For $r \geq 1$ and $x \in [0,1]$, define*

$$m_{r;\alpha,\lambda}(x) = \frac{(1-\alpha)_{\uparrow r-1}}{r!}\, x^r \lambda^{1-\alpha} (x+\lambda)^{\alpha-r}$$

Then, as $n \to +\infty$

(i)

$$\frac{\mathbb{E}\left[K_{r,\lfloor xn \rfloor}^{\{n\}}\right]}{n} \to m_{r;\alpha,\lambda}(x)$$

uniformly on $[0,1]$.

(ii) the following SLLN holds uniformly for $x \in [0,1]$:

$$\frac{K_{r,\lfloor xn \rfloor}^{\{n\}}}{n} \xrightarrow{a.s.} m_{r;\alpha,\lambda}(x) \tag{3.2.7}$$

Proof. From [35, Proposition 1] we recover the following expression for the (falling) factorial moments of $K_{r,\lfloor xn\rfloor}^{\{n\}}$:

$$\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}}\right)_{\downarrow s}\right] = \left(\frac{(1-\alpha)_{\uparrow r-1}}{r!}\right)^s \frac{\Gamma(xn+1-\mathfrak{r}_x)}{\Gamma(xn-rs+1-\mathfrak{r}_x)} \frac{\Gamma(\lambda n+1)}{\Gamma((\lambda+x)n-\mathfrak{r}_x)} \times \\ \times \alpha^{s-1} \frac{\Gamma\left(\frac{\lambda n}{\alpha}+s\right)}{\Gamma\left(\frac{\lambda n}{\alpha}+1\right)} \frac{\Gamma((\lambda+x)n+s\alpha-sr-\mathfrak{r}_x)}{\Gamma(\lambda n+s\alpha)}$$

where we are denoting by $\mathfrak{r}_x = xn - \lfloor xn\rfloor$. Since $0 \le \mathfrak{r}_x < 1$ for all $x \in [0,1]$ and for all $n \in \mathbb{N}$, we can apply the usual asymptotic expansion for ratios of gamma functions to obtain

$$\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}}\right)_{\downarrow s}\right] = n^s \cdot m_{r;\alpha,\lambda}^s(x) + n^{s-1} \cdot m_{r;\alpha,\lambda}^s(x)\, S_s(x) + O\left(s^{s-2}\right)$$

where

$$\begin{aligned} S_s(x) &= \frac{1}{2x\lambda(x+\lambda)} \cdot \left\{s^2 \cdot \left[x^2\alpha(1-\alpha) + x[\lambda\alpha - 2r\alpha\lambda] - \lambda^2 r^2\right]\right. \\ &\quad \left. + s \cdot \left[-x^2\lambda\left[r(1-2\mathfrak{r}_x) - \alpha(1+2\mathfrak{r}_x)\right] x\lambda^2 r(1-2\mathfrak{r}_x)\right]\right\} \\ &= s^2 A(x) + sB(x) \end{aligned}$$

with $A(x)$ and $B(x)$ independent of s. Note that since

$$\lim_{x\to 0^+} x \cdot A(x),\ \lim_{x\to 0^+} x \cdot B(x) \in (0, +\infty)$$

and

$$\lim_{x\to 0^+} \frac{m_{r;\alpha,\lambda}(x)}{x^r} = \frac{(1-\alpha)_{\uparrow(r-1)}}{r!} \cdot \lambda^{1-r},$$

with the particular case

$$\lim_{x\to 0^+} \frac{m_{1;\alpha,\lambda}(x)}{x} = 0,$$

then for every $r, s \ge 1$,

$$\lim_{x\to 0^+} S_s(x) = 0.$$

These expansions hold uniformly for $x \in [0,1]$. In particular, for $s = 1$, this rewrites as

$$\lim_{n\to+\infty} \sup_{x\in[0,1]} \left| \frac{\mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]}{n} - m_{r;\alpha,\lambda}(x) \right| = 0$$

Proving point (i).

For the strong law of large numbers, we prove complete convergence of the sequence $K_{r,\lfloor xn\rfloor}^{\{n\}}/n$ by showing that, uniformly for $x \in [0,1]$, $\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}} - \mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]\right)^4\right] = O(n^2)$. In fact we can again write

$$\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}} - \mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]\right)^4\right] \\ = \sum_{k=0}^{4} (-1)^{4-k} \binom{4}{k} \left(\mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]\right)^{4-k} \sum_{s=0}^{k} S(k,s)\, \mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}}\right)_{\downarrow s}\right],$$

where $S(k,s)$ denote the Stirling number of the second kind. Expanding such expression and computing the expansion of powers of $\mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]$ yields, after tedious but straightforward calculations,

$$\lim_{n\to+\infty}\frac{1}{n^2}\left|\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}}-\mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]\right)^4\right]-\left(n^4\cdot\mathcal{A}_{\alpha,\lambda,r}(x)+n^3\cdot\mathcal{B}_{\alpha,\lambda,r}(x)\right)\right|=c$$

for some constant c and with

$$\mathcal{A}_{\alpha,\lambda,r}(x)=\left(m_{r;\alpha,\lambda}(x)\right)^4\cdot(1-4+6-3)=0$$

and

$$\begin{aligned}\mathcal{B}_{\alpha,\lambda,r}(x)&=\left(m_{r;\alpha,\lambda}(x)\right)^4\cdot\{-4\left[A(x)+B(x)\right]+6[4A(x)+2B(x)]\\&\quad-4\left[9A(x)+3B(x)\right]+\left[16A(x)+4B(x)\right]\}\\&=\left(m_{r;\alpha,\lambda}(x)\right)^4\cdot[A(x)\cdot(-4+24-36+16)+B(x)\cdot(-4+12-12+4)]\\&=0.\end{aligned}$$

Hence,

$$\lim_{n\to+\infty}\sup_{x\in[01,]}\mathbb{E}\left[\left(K_{r,\lfloor xn\rfloor}^{\{n\}}-\mathbb{E}\left[K_{r,\lfloor xn\rfloor}^{\{n\}}\right]\right)^4\right]=c$$

and the proof is concluded. ■

3.2.2 Proof of the expansion (3.2.1) and of the SLLN (3.2.1)

Upon noting that

$$\mathfrak{m}_{r;\alpha,\lambda}=m_{r;\alpha,\lambda}(1)=\frac{(1-\alpha)_{\uparrow r-1}}{r!}\,\lambda^{1-\alpha}(1+\lambda)^{\alpha-r}$$

apply Lemmas 3.1.1 and 3.2.2 for $x=1$.

3.2.3 Proof of the CLT (3.2.1)

Analogously to the case of (2.1.4), the proof of (3.2.1) relies on the construction of a martingale and on the use of the Lindeberg–Levy central limit theorem for arrays of martingales.

3.2.3.1 Construction of the martingale

Troughout this section we use the convention

$$K_{0,h}^{\{n\}}:=K_h^{\{n\}}.$$

The sequential construction of the partition of section 1.3.1 can be equivalently stated by saying that for every $n\in\mathbb{N}$ and $h\in\{1,...,n\}$, and for every $r\in\{0,...,n\}$, there exist a random variable $\xi_{r,h}^{(n)}$ defined on the same probability space as $\mathbf{K}_1^{\{n\}},...,\mathbf{K}_{h-1}^{\{n\}}$ such that

$$K_{r,h}^{\{n\}}=K_{r,h-1}^{\{n\}}+\xi_{r,h}^{\{n\}}\tag{3.2.8}$$

and

$$P\left[\xi_{r,h}^{\{n\}}=k\,|\,\mathcal{F}_{h-1}^{\{n\}}\right]=\begin{cases}p_{h-1}^{\{n\}} & \text{if } k=1\\ q_{h-1}^{\{n\}} & \text{if } k=-1\\ 1-p_{h-1}^{\{n\}}-q_{h-1}^{\{n\}} & \text{if } k=0\end{cases}$$

with

$$p_{r,h-1}^{\{n\}} = \begin{cases} \frac{\alpha K_{h-1}^{\{n\}} + \lambda n}{\lambda n + h - 1} & \text{if } r = 0, 1 \\ \frac{(r-1-\alpha) K_{r-1,h-1}^{\{n\}}}{\lambda n + h - 1} & \text{if } r \geq 2 \end{cases} \quad \text{and} \quad q_{r,h-1}^{\{n\}} = \begin{cases} 0 & \text{if } r = 0 \\ \frac{(r-\alpha) K_{r,h-1}^{\{n\}}}{\lambda n + h - 1} & \text{if } r \geq 1 \end{cases}$$

In particular,

$$\begin{aligned} \mathbb{E}\left[K_{r,h}^{\{n\}} \,|\, \mathcal{F}_h^{\{n\}}\right] &= K_{r,h-1}^{\{n\}} + p_{r,h-1}^{\{n\}} - q_{r,h-1}^{\{n\}} \\ &= \gamma_{r,h-1}^{\{n\}} K_{r,h-1}^{\{n\}} + \beta_{r,h-1}^{\{n\}} \end{aligned} \tag{3.2.9}$$

where

$$\gamma_{r,h-1}^{\{n\}} = \frac{\lambda n + h - 1 - r + \alpha}{\lambda n + h - 1}$$

and

$$\beta_{r,h-1}^{\{n\}} = \begin{cases} \frac{\lambda n}{\lambda n + h - 1} & \text{if } r = 0 \\ p_{r,h-1}^{\{n\}} & \text{if } r \geq 1 \end{cases}.$$

For $r \in \{0, ..., n\}$ - with the convention $\bullet_{0,h}^{\{n\}} := \bullet_h^{\{n\}}$ - and for $n \in \mathbb{N}$ and $h \in \{1, ..., n\}$, define

$$a_{r,h}^{\{n\}} = \prod_{k=1}^{h-1} \left(\gamma_{r,k}^{\{n\}}\right)^{-1} \quad \text{and} \quad A_{r,h}^{\{n\}} = \sum_{k=1}^{h-1} a_{r,k+1}^{\{n\}} \beta_{r,k}^{\{n\}}. \tag{3.2.10}$$

and for fixed $d \in \mathbf{N}$, let

$$\begin{aligned} \mathbf{\Xi}_{d,h}^{\{n\}} &= \left(\xi_h^{\{n\}}, \xi_{1,h}^{\{n\}}, ..., \xi_{d,h}^{\{n\}}\right)^T \\ \mathbf{A}_{d,h}^{\{n\}} &= \left(A_h^{\{n\}}, A_{1,h}^{\{n\}}, ..., A_{d,h}^{\{n\}}\right)^T \\ \mathbf{a}_{d,h}^{\{n\}} &= \left(a_h^{\{n\}}, a_{1,h}^{\{n\}}, ..., a_{d,h}^{\{n\}}\right)^T. \end{aligned}$$

We introduce the following notation: given a vector $v \in \mathbb{R}^{d+1}$, we denote by $\mathrm{Diag}(v)$ the $(d+1) \times (d+1)$ diagonal matrix defined by

$$\left(\mathrm{Diag}(v)\right)_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ v_i & \text{if } i = j \end{cases}$$

With this convention, let

$$\mathcal{A}_{d,h}^{\{n\}} = \mathrm{Diag}\left(\mathbf{a}_{d,h}^{\{n\}}\right)$$

and finally define

$$\mathbf{M}_{d,h}^{\{n\}} = \mathcal{A}_{d,h}^{\{n\}} \cdot \mathbf{K}_{d,h}^{\{n\}} - \mathbf{A}_{d,h}^{\{n\}} \tag{3.2.11}$$

and

$$\mathbf{\Delta}_{d,h}^{\{n\}} = \mathbf{M}_{d,h}^{\{n\}} - \mathbf{M}_{d,h-1}^{\{n\}} \tag{3.2.12}$$

Lemma 3.2.3. *For every $d \in \mathbb{N}$ and $n \in \mathbb{N}$, $\left(\mathbf{M}_{d,h}^{\{n\}}\right)_{h \in \{1,...,n\}}$ is a martingale with respect to the filtration $\left(\mathcal{F}_h^{\{n\}}\right)_{h \in \{1,...n\}}$.*

Proof. It follows from definition (3.2.10) that for every d, n, h,

$$\mathcal{A}_{d,h-1}^{\{n\}} = \text{Diag}\left(\left(\gamma_{r,h}^{\{n\}}\right)_{r\in\{0,...,d\}}\right) \cdot \mathcal{A}_{d,h}^{\{n\}}$$

and

$$\mathbf{A}_{d,h}^{\{n\}} = \mathbf{A}_{d,h-1}^{\{n\}} + \mathcal{A}_{d,h}^{\{n\}} \cdot \left(\beta_{r,h-1}^{\{n\}}\right)_{r\in\{0,...,d\}}$$

Hence,

$$\begin{aligned}
\mathbf{\Delta}_{d,h}^{\{n\}} &= \mathcal{A}_{d,h}^{\{n\}} \cdot \mathbf{K}_{d,h}^{\{n\}} - \mathbf{A}_{d,h}^{\{n\}} - \mathcal{A}_{d,h-1}^{\{n\}} \cdot \mathbf{K}_{d,h-1}^{\{n\}} + \mathbf{A}_{d,h-1}^{\{n\}} \\
&= \mathcal{A}_{d,h}^{\{n\}} \cdot \left[\mathbf{K}_{d,h-1}^{\{n\}} + \mathbf{\Xi}_{d,h}^{\{n\}}\right] - \mathbf{A}_{d,h-1}^{\{n\}} - \mathcal{A}_{d,h}^{\{n\}} \cdot \left(\beta_{r,h-1}^{\{n\}}\right)_{r\in\{0,...,d\}} \\
&\quad - \text{Diag}\left(\left(\gamma_{r,h}^{\{n\}}\right)_{r\in\{0,...,d\}}\right) \cdot \mathcal{A}_{d,h}^{\{n\}} \cdot \mathbf{K}_{d,h-1}^{\{n\}} + \mathbf{A}_{d,h-1}^{\{n\}} \\
&= \mathcal{A}_{d,h}^{\{n\}} \cdot \left\{\left[\mathbf{I}_{d+1} - \text{Diag}\left(\left(\gamma_{r,h}^{\{n\}}\right)_{r\in\{0,...,d\}}\right)\right] \cdot \mathbf{K}_{d,h-1}^{\{n\}} + \mathbf{\Xi}_{d,h}^{\{n\}} - \left(\beta_{r,h-1}^{\{n\}}\right)_{r\in\{0,...,d\}}\right\} \\
&= \mathcal{A}_{d,h}^{\{n\}} \cdot \left[\mathbf{\Xi}_{d,h}^{\{n\}} - \left(p_{r,h-1}^{\{n\}} - q_{r,h-1}^{\{n\}}\right)_{r\in\{0,...,d\}}\right], \qquad (3.2.13)
\end{aligned}$$

where the last equality follows from (3.2.9). Taking conditional expectations on both side we obtain

$$\mathbb{E}\left[\mathbf{\Delta}_h^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right] = \mathcal{A}_h^{\{n\}} \cdot \mathbb{E}\left[\mathbf{\Xi}_{d,h}^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right] - \left(p_{r,h-1}^{\{n\}} - q_{r,h-1}^{\{n\}}\right)_{r\in\{0,...,d\}} = \mathbf{0},$$

which concludes the proof. ■

3.2.3.2 Completing the proof

Let n be a sequence of integers such that $h_n \to +\infty$ as $n \to +\infty$. Denote by

$$\langle \mathbf{M}_d^{\{n\}} \rangle = \left(\langle \mathbf{M}_d^{\{n\}} \rangle_1, ..., \langle \mathbf{M}_d^{\{n\}} \rangle_n\right)^T$$

the increasing process of $\mathbf{M}_d^{\{n\}}$, i.e.

$$\langle \mathbf{M}_d^{\{n\}} \rangle_{h_n} = \sum_{h=1}^{h_n} \mathbb{E}\left[\mathbf{\Delta}_{d,h}^{\{n\}} \cdot \left(\mathbf{\Delta}_{d,h}^{\{n\}}\right)^T \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right]. \qquad (3.2.14)$$

The following lemma is an immediate consequence of [55, Corollary 3.1], and it is the multidimensional analogue of Lemma 3.1.4.

Lemma 3.2.4. *Assume that the following two conditions are satisfied,*

(a)

$$\frac{\langle \mathbf{M}_d^{\{n\}} \rangle_{h_n}}{h_n} \xrightarrow{a.s.} G_{d,\alpha,\lambda}$$

for some positive semi-definite $(d+1)\times(d+1)$ matrix $G_{d,\alpha,\lambda}$

(b) For every $\varepsilon > 0$,

$$\frac{1}{h_n}\sum_{h=1}^{h_n} \mathbb{E}\left[\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 \cdot \mathbb{1}_{\left\{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 > h_n\,\varepsilon\right\}} \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \xrightarrow{a.s.} 0$$

Then, we have

$$\frac{\mathbf{M}_{d,h_n}^{\{n\}}}{\sqrt{h_n}} \xrightarrow{w} N(\mathbf{0}, G_{d,\alpha,\lambda}) \tag{3.2.15}$$

We now show that, for the sequence $h_n = n$, conditions (a) and (b) of lemma 3.2.4 are satisfied with $G_{d,\alpha,\lambda} = \Gamma_{d,\alpha,\lambda}$ where $\Gamma_{d,\alpha,\lambda}$ is as in (3.2.4). To prove (a), we work component by component. It follows from (3.2.13) that

$$\begin{aligned}
&\left(\mathbf{\Delta}_{d,h}^{\{n\}} \cdot \left(\mathbf{\Delta}_{d,h}^{\{n\}}\right)^T\right)_{i,j} \\
&= a_{i-1,h}^{\{n\}} a_{j-1,h}^{\{n\}} \cdot \left(\xi_{i-1,h}^{\{n\}} - p_{i-1,h-1}^{\{n\}} + q_{i-1,h-1}^{\{n\}}\right)\left(\xi_{j-1,h}^{\{n\}} - p_{j-1,h-1}^{\{n\}} + q_{j-1,h-1}^{\{n\}}\right) \\
&= a_{i-1,h}^{\{n\}} a_{j-1,h}^{\{n\}} \cdot \left[\xi_{i-1,h}^{\{n\}}\xi_{j-1,h}^{\{n\}} - \left(p_{i-1,h-1}^{\{n\}} - q_{i-1,h-1}^{\{n\}}\right)\xi_{j-1,h}^{\{n\}}\right. \\
&\quad \left. - \left(p_{j-1,h-1}^{\{n\}} - q_{j-1,h-1}^{\{n\}}\right)\xi_{i-1,h}^{\{n\}} + \left(p_{i-1,h-1}^{\{n\}} - q_{i-1,h-1}^{\{n\}}\right)\left(p_{j-1,h-1}^{\{n\}} - q_{j-1,h-1}^{\{n\}}\right)\right]
\end{aligned}$$

so that

$$\left(\mathbb{E}\left[\mathbf{\Delta}_{d,h}^{\{n\}} \cdot \left(\mathbf{\Delta}_{d,h}^{\{n\}}\right)^T \middle| \mathcal{F}_{h-1}^{\{n\}}\right]\right)_{i,j} = a_{i-1,h}^{\{n\}} a_{j-1,h}^{\{n\}} \cdot \left[\left(P_{h-1}^{\{n\}}\right)_{i,j} - \left(R_{h-1}^{\{n\}}\right)_{i,j}\right]$$

where, for all $1 \le i \le j$,

$$\left(P_{h-1}^{\{n\}}\right)_{i,j} = \mathbb{E}\left[\xi_{i-1,h}^{\{n\}}\xi_{j-1,h}^{\{n\}} \middle| \mathcal{F}_{h-1}^{\{n\}}\right] = \begin{cases} 0 & \text{if } j \ge i+2 \\ -q_{i-1,h-1}^{\{n\}} & \text{if } j = i+1 \text{ and } i \ge 2 \\ p_{i-1,h-1}^{\{n\}} + q_{i-1,h-1}^{\{n\}} & \text{if } j = i \text{ or } i = 1, j = 2 \end{cases}$$

and

$$\left(R_{h-1}^{\{n\}}\right)_{i,j} = \left(p_{i-1,h-1}^{\{n\}} - q_{i-1,h-1}^{\{n\}}\right)\left(p_{j-1,h-1}^{\{n\}} - q_{j-1,h-1}^{\{n\}}\right).$$

This implies

$$\begin{aligned}
\left(\langle \mathbf{M}_d^{\{n\}} \rangle_n\right)_{i,j} &= \sum_{h=1}^{n} a_{i-1,h}^{\{n\}} a_{j-1,h}^{\{n\}} \cdot \left[\left(P_{h-1}^{\{n\}}\right)_{i,j} - \left(R_{h-1}^{\{n\}}\right)_{i,j}\right] \\
&= \sum_{h=1}^{n} a_{i-1,n\cdot\frac{h-1}{n}+1}^{\{n\}} a_{j-1,n\cdot\frac{h-1}{n}+1}^{\{n\}} \cdot \left[\left(P_{n\cdot\frac{h-1}{n}}^{\{n\}}\right)_{i,j} - \left(R_{n\cdot\frac{h-1}{n}}^{\{n\}}\right)_{i,j}\right] \\
&= n \cdot \sum_{h=1}^{n} a_{i-1,n\cdot\frac{h-1}{n}+1}^{\{n\}} a_{j-1,n\cdot\frac{h-1}{n}+1}^{\{n\}} \cdot \left[\left(P_{n\cdot\frac{h-1}{n}}^{\{n\}}\right)_{i,j} - \left(R_{n\cdot\frac{h-1}{n}}^{\{n\}}\right)_{i,j}\right] \cdot \frac{1}{n}
\end{aligned}$$

so that $\left(\langle \mathbf{M}_d^{\{n\}} \rangle_n\right)_{i,j} / n$ can be viewed as a Riemann sum for the function $F_{i,j}^{\{n\}}$ on $[0,1]$, where

$$F_{i,j}^{\{n\}}(x) = a_{i-1,\lfloor nx \rfloor+1}^{\{n\}} a_{j-1,\lfloor nx \rfloor+1}^{\{n\}} \cdot \left[\left(P_{\lfloor nx \rfloor}^{\{n\}}\right)_{i,j} - \left(R_{\lfloor nx \rfloor}^{\{n\}}\right)_{i,j}\right] \tag{3.2.16}$$

Making use of lemma 3.2.2 we are able to show that

$$F_{i,j}^{\{n\}}(x) \xrightarrow{a.s.} f_{i,j}(x)$$

uniformly for $x \in [0,1]$, where $f_{i,j}(x)$ is a deterministic, explicit function only depending on α, λ, i and j, and it is Riemann integrable on [0,1]. By definition and elementary properties of the Riemann integral,

$$\frac{\left(\langle \mathbf{M}_d^{\{n\}} \rangle_n\right)_{i,j}}{n} \xrightarrow{a.s.} \int_0^1 f_{i,j}(x)\,\mathrm{d}x.$$

and since we are able to prove

$$\int_0^1 f_{i,j}(x)\,\mathrm{d}x = \left(\Gamma_{d,\alpha,\lambda}\right)_{i,j},$$

the proof is concluded. See appendix D.2 for the proof of the convergence of $F_{i,j}^{\{n\}}$, the definition of $f_{i,j}$ and the computation of the integral.

For what concerns condition (b), we can write

$$\begin{aligned} S_n &:= \frac{1}{n}\sum_{h=1}^{n} \mathbb{E}\left[\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 \cdot \mathbb{1}_{\left\{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 > n\varepsilon\right\}} \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \\ &= \frac{1}{n}\sum_{h=1}^{n} \mathbb{E}\left[\frac{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^4}{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2} \cdot \mathbb{1}_{\left\{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 > n\varepsilon\right\}} \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \end{aligned}$$

and since

$$\frac{\mathbb{1}_{\left\{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2 > n\varepsilon\right\}}}{\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^2} \leq \frac{1}{n\varepsilon}$$

almost surely, then

$$\begin{aligned} S_n &\leq \frac{1}{n^2\,\varepsilon}\sum_{h=1}^{n} \mathbb{E}\left[\left\|\mathbf{\Delta}_{d,h}^{\{n\}}\right\|^4 \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \\ &= \frac{1}{n^2\,\varepsilon}\sum_{h=1}^{n} \mathbb{E}\left[\left\|\mathcal{A}_{d,h}^{\{n\}} \cdot \left[\mathbf{\Xi}_{d,h}^{\{n\}} - \mathbb{E}\left[\mathbf{\Xi}_{d,h}^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right]\right]\right\|^4 \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \\ &\leq \frac{\mathcal{C}}{n^2\,\varepsilon}\sum_{h=1}^{n} \mathbb{E}\left[\left\|\mathcal{A}_{d,h}^{\{n\}}\right\|^4 \cdot \left\|\mathbf{\Xi}_{d,h}^{\{n\}} - \mathbb{E}\left[\mathbf{\Xi}_{d,h}^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right]\right\|^4 \,\middle|\, \mathcal{F}_{h-1}^{\{n\}}\right] \end{aligned}$$

for some positive constant $\mathcal{C}$. Finally, observe that, since the $\xi_{r,h}^{\{n\}}$ take values in $\{-1,0,1\}$,

$$\left\| \mathbf{\Xi}_{d,h}^{\{n\}} - \mathbb{E}\left[\mathbf{\Xi}_{d,h}^{\{n\}} \mid \mathcal{F}_{h-1}^{\{n\}}\right] \right\| \leq 2$$

almost surely; furthermore, by definition of $\mathcal{A}_{d,h}^{\{n\}}$,

$$\left\|\mathcal{A}_{d,h}^{\{n\}}\right\| = \left\|\mathbf{a}_{d,h}^{\{n\}}\right\|.$$

Then, we can bound S_n by

$$\begin{aligned} S_n &\leq \frac{16\,\mathcal{C}}{n^2\varepsilon}\sum_{h=1}^{n}\left\|\mathbf{a}_{d,h}^{\{n\}}\right\|^4 \\ &\leq \frac{16\,\mathcal{C}}{n\varepsilon} \cdot \max_{r\in\{0,\dots,d\},\, x\in[0,1]}\left(a_{r,\lfloor nx\rfloor}^{\{n\}}\right)^4 \\ &=: \mathfrak{S}_n \end{aligned}$$

We know from (i) of lemma D.1 that for every $r \in \{1, ..., d\}$

$$a_{r,\lfloor nx \rfloor+1}^{\{n\}} = \left(\frac{\lambda + x}{\lambda}\right)^{r-\alpha} + O\left(\frac{1}{n}\right) \leq \left(\frac{\lambda+1}{\lambda}\right)^{d-\alpha} + O\left(\frac{1}{n}\right)$$

and for $r = 0$,

$$a_{r,\lfloor nx \rfloor+1}^{\{n\}} = \left(\frac{\lambda}{\lambda + x}\right)^{1+\alpha} + O\left(\frac{1}{n}\right) \leq 1 + O\left(\frac{1}{n}\right)$$

uniformly for $x \in [0, 1]$. In conclusion

$$\mathfrak{S}_n = \frac{16\,\mathcal{C}}{n\varepsilon} \cdot \left(\frac{\lambda+1}{\lambda}\right)^{4(d-\alpha)} + O\left(\frac{1}{n^2}\right) = O\left(\frac{1}{n}\right)$$

and this proves point (b).

Therefore, it follows from Lemma 3.2.4 that as $n \to +\infty$,

$$\frac{\mathbf{M}_{d,n}^{\{n\}}}{\sqrt{n}} \xrightarrow{w} N(\mathbf{0}, \Gamma_{d,\alpha,\lambda}) \tag{3.2.17}$$

By definition of $\mathbf{M}_{d,n}^{\{n\}}$ (3.2.11),

$$\begin{aligned}\mathbf{K}_{d,n}^{\{n\}} &= \left(\mathcal{A}_{d,n}^{\{n\}}\right)^{-1} \cdot \left[\mathbf{M}_{d,n}^{\{n\}} + \mathbf{A}_{d,n}^{\{n\}}\right] \\ &= \left(\mathcal{A}_{d,n}^{\{n\}}\right)^{-1} \cdot \mathbf{M}_{d,n}^{\{n\}} + \mathfrak{M}_{d,n}^{\{n\}} + \left(\mathcal{A}_{d,n}^{\{n\}}\right)^{-1} \cdot \mathbf{A}_{d,n}^{\{n\}} - \mathfrak{M}_{d,n}^{\{n\}}\end{aligned}$$

It follows from lemma 3.2.2 that

$$\left(\mathcal{A}_{d,n}^{\{n\}}\right)^{-1} \cdot \mathbf{A}_{d,n}^{\{n\}} \xrightarrow{a.s.} \mathfrak{M}_{d,n}^{\{n\}},$$

and by (i) of lemma D.1

$$\left(\mathcal{A}_{d,n}^{\{n\}}\right)^{-1} \xrightarrow{a.s.} \mathfrak{A}_{d,\alpha,\lambda}$$

where

$$\mathfrak{A}_{d,\alpha,\lambda} = \mathrm{Diag}\left(\left(\left[\frac{\lambda+1}{\lambda}\right]^{r-\alpha}\right)_{r\in\{0,...,d\}}\right).$$

In conclusion, apply Slutsky's lemma to show that

$$\frac{\mathbf{K}_{d,n}^{\{n\}} - \mathfrak{M}_{d,m}^{\{n\}}}{\sqrt{n}} \xrightarrow{w} N(\mathbf{0}, \mathfrak{A}_{d,\alpha,\lambda}^T \cdot \Gamma_{d,\alpha,\lambda} \cdot \mathfrak{A}_{d,\alpha,\lambda})$$

and since by definition (3.2.3)

$$\mathfrak{A}_{d,\alpha,\lambda}^T \cdot \Gamma_{d,\alpha,\lambda} \cdot \mathfrak{A}_{d,\alpha,\lambda} = \Sigma_{d,\alpha,\lambda}$$

the proof is concluded.

Chapter 4

Gaussian credible intervals in the BNP estimation of the unseen

In this chapter we present an application of the Ewens–Pitman model in the large θ regime to the problem of species sampling under the Bayesian nonparametrics (BNP) approach. More specifically, as already mentioned in section 1.3.1, the use of a sample-size dependent Pitman-Yor prior leads to considering the partition structure (1.3.1), and the statistical problem at hand requires the analysis of the the posterior distribution of such partition. The main theoretical result of this chapter is then a posterior counterpart of theorem 2.1.1.

4.1 Introduction

The estimation of the number of unseen species is a long-standing problem in statistics, dating back to the seminal work of [47] on "species extrapolation". It assumes that $n \geq 1$ random samples $(X_1, \ldots, X_n)$ are collected from an unknown discrete distribution P on $\mathbb{S}$, with $\mathbb{S}$ being a space of species' labels or symbols, and calls for estimating

$$\mathcal{K}_{n,m} = |\{X_{n+1}, \ldots, X_{n+m}\} \setminus \{X_1, \ldots, X_n\}|,$$

namely the number of hitherto unseen species that would be observed if $m \geq 1$ additional samples $(X_{n+1}, \ldots, X_{n+m})$ were collected from the same distribution P. First introduced in ecology [18; 19; 13], the unseen-species problem has more recently found applications in biological and physical sciences, where it poses significant challenges in handling large values of n and m [66; 49; 59; 24]. See [26] for an overview with emphasis on large-scale biological data. Further applications include, e.g., information theory and theoretical computer science [54; 68; 48; 12; 14], empirical linguistics and natural large language modeling [32; 90; 71; 70; 5; 61], and forensic DNA analysis [15; 39].

4.1.1 Background and motivation

A frequentist nonparametric approach to the unseen-species problem was proposed by [51] and [32], then developed rigorously by [72]. This is a distribution-free approach, in the sense that it does not rely on any assumption on P, leading to estimates of $\mathcal{K}_{n,m}$ that are minimax optimal for any n and $m \leq n \log n$, with such a range being the best (largest) possible [72; 94; 95]. From a Bayesian nonparametric (BNP) perspective, it is natural to specify a prior distribution for P, an approach that was first investigated in [67] by focussing on the Pitman-Yor prior [81], which is a prior indexed by $\alpha \in [0, 1)$ and $\theta > -\alpha$, with $\alpha = 0$ being the celebrated Dirichlet prior [46]. Under the Pitman-Yor

prior, [67] showed that the posterior distribution of $\mathcal{K}_{n,m}$, given $(X_1, \ldots, X_n)$, depends on the sampling information only through the sample size n and the number K_n of species in $(X_1, \ldots, X_n)$. Such a distribution is in closed-form, with the posterior mean estimate $\hat{K}_{n,m}$ that can be easily evaluated for any value of n and m [34]. See [4] for an up-to-date overview.

Uncertainty quantification for estimates of $\mathcal{K}_{n,m}$ has been addressed under the BNP approach, but remains an open problem under the distribution-free approach, especially when $m > n$ [72]. For $\alpha \in [0, 1)$, exact credible intervals for $\mathcal{K}_{n,m}$ can be derived, for any n and m, by Monte Carlo sampling the posterior distribution through the predictive distributions of the Pitman-Yor prior [4]. Further, if $\alpha \in (0, 1)$ then large m asymptotic credible intervals can be derived using the method proposed by [34]. Denoting by $K_m^{(n)}$ a random variable whose distribution is the posterior distribution of $\mathcal{K}_{n,m}$ given $K_n = j$, [34] showed that, as $m \to +\infty$

$$\frac{K_m^{(n)}}{(\theta + n + m)^\alpha - (\theta + n)^\alpha} \xrightarrow{\text{w}} S_{\alpha,\theta}^{(n,j)},$$

where $S_{\alpha,\theta}^{(n,j)}$ is a scaled Mittag-Leffler random variable [80, Chapter 0]. Given values of (n, j) and (α, θ), with (α, θ) estimated by means of empirical or fully Bayes procedures, for m sufficiently large the distribution of $K_m^{(n)}$ is approximated by the distribution of $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$, with $c_{\alpha,\theta,n}(m) = (\theta + n + m)^\alpha - (\theta + n)^\alpha$. Mittag-Leffler credible intervals for $\mathcal{K}_{n,m}$ are then derived by Monte Carlo sampling $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$. In particular, the scaling $c_{\alpha,\theta,n}(m)$ is determined in such a way that $\hat{K}_{n,m}$ coincides with the expected value of $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$, ensuring that intervals are centered on the BNP estimator $\hat{K}_{n,m}$ for any $m \geq 1$.

While the method proposed by [34] addresses uncertainty quantification for large values of m, it comes with notable limitations. Firstly, Monte Carlo sampling $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$ is computationally expensive, despite the recent advances on sampling scaled Mittag-Leffler distributions [83], which limits the practical (computational) benefits of asymptotic credible intervals over exact ones derived by sampling the posterior distribution, unless m is extremely large. Secondly, empirical analyses by [34] show that Mittag-Leffler credible intervals are shorter than the exact intervals, with the gap decreasing when m enters the regime $m \gg \theta + n$, namely m is much larger than $\theta + n$. Thirdly, the method of [34] fails to extend to the case $\alpha = 0$ due to a degenerate behaviour of the limiting posterior distribution [4]. These theoretical and empirical limitations highlight the motivation for this paper, which introduces an alternative methodology to uncertainty quantification for BNP estimates of $\mathcal{K}_{n,m}$.

4.1.2 Preview of our contributions

We propose a novel method to derive large m asymptotic credible intervals for $\mathcal{K}_{n,m}$, which allows to deal with $\alpha \in [0, 1)$ and avoids the use of Monte Carlo sampling. Under the Pitman-Yor prior, for $\alpha \in [0, 1)$ and $\theta > -\alpha$, we rely on the large m asymptotic behaviour of $K_m^{(n)}$ assuming that both the sampling information (n, j) and the parameter θ are large. In particular, we set $n = \nu m$, $j = \rho m$ and $\theta = \tau m$, with $\nu, \rho, \tau > 0$, and show that, as $m \to +\infty$

$$\frac{K_m^{(n)} - m\mathscr{M}_{\alpha,\tau,\nu,\rho}}{\sqrt{m\mathscr{S}_{\alpha,\tau,\nu,\rho}}} \xrightarrow{\text{w}} N(0, 1),$$

where

$$\mathbb{E}\left[K_m^{(n)}\right] = m\mathscr{M}_{\alpha,\tau,\nu,\varrho} + O(1)$$

and

$$\mathrm{Var}\left(K_m^{(n)}\right) = m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho} + O(1),$$

for some (explicit) functions $\mathscr{M}_{\alpha,\tau,\nu,\varrho}$ and $\mathscr{S}_{\alpha,\tau,\nu,\varrho}$ of $(\alpha,\tau,\nu,\varrho)$, respectively, and where $N(0,1)$ is the standard Gaussian random variable. Given values of $(\nu m, \rho m)$ and $(\alpha, \tau m)$, with $(\alpha, \tau m)$ estimated by means of empirical or fully Bayes procedures, for m sufficiently large the distribution of $K_m^{(n)}$ is approximated by a Gaussian distribution with mean $m\mathscr{M}_{\alpha,\tau,\nu,\varrho}$ and variance $m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}$. Gaussian credible intervals for $\mathcal{K}_{n,m}$, with a prescribed (asymptotic) level, are then derived from the Gaussian quantiles, enhancing computational efficiency.

Given the sampling information (n, j) and the parameter (α, θ), the proposed methodology and that of [34] employ different approximations of the posterior distribution of $\mathcal{K}_{n,m}$, for m sufficiently large. Both approximations lead to large m asymptotic credible intervals that are centered on the BNP estimator $\hat{K}_{n,m}$. Figure 4.1 shows that our method outperforms that of [34] on synthetic data generated from various natural distributions, namely Zipf, Dirichlet-Multinomial and Uniform distributions; the same synthetic datasets were analyzed in [72, Figure 3]. In particular, compared to Mittag-Leffler credible intervals, Gaussian credible intervals provide greater coverage of the exact credible intervals, for any $m \geq 1$. A comprehensive empirical validation of our methodology is presented in the paper, encompassing both synthetic and real datasets. For real data, we present an application to the Expressed Sequence Tags (ESTs) data considered in [34, Section 3], confirming the behaviour displayed in Figure 4.1.

4.1.3 Organization of the chapter

This chapter is organized as follows. In Section 2.2 we recall the BNP approach to the unseen-species problem, including the modeling assumptions and posterior inferences. Section 4.3 contains the Gaussian CLT for $K_m^{(n)}$, with corresponding large m asymptotic credible intervals for $\mathcal{K}_{n,m}$. The empirical performance of our methodology is investigated in Section 4.4. Section 5.1 concludes with a discussion and some directions for future research. Technical preliminary results, proofs and additional numerical illustrations are deferred to Appendices E, F and G.

4.1.4 A disclaimer on notation

In this chapter we consider random variables which depend on several parameters, making the overall notation quite heavy. While we occasionally refer to known results in the BNP framework in the standard, fixed θ regime, all original results are obtained in a novel asymptotical regime where all parameters depend on m. For this reason, and to improve local readability, we sacrifice notation uniformity with the rest of the work and introduce some simplifications

Start by noting that the assumptions $\theta = \tau m$ and $n = \nu m$ imply $\theta = \frac{\tau}{\nu+1}(n+m)$. For $h \in \{1, ..., n+m\}$, let $K_h^{\{n+m\}}$ denote the number of blocks in the large–θ Ewens–Pitman partition of $\{1, ..., h\}$, when $\theta = \frac{\tau}{\nu+1}(n+m)$. The following notations, without a superscript, will be used both in the context of the standard and the large θ regime.

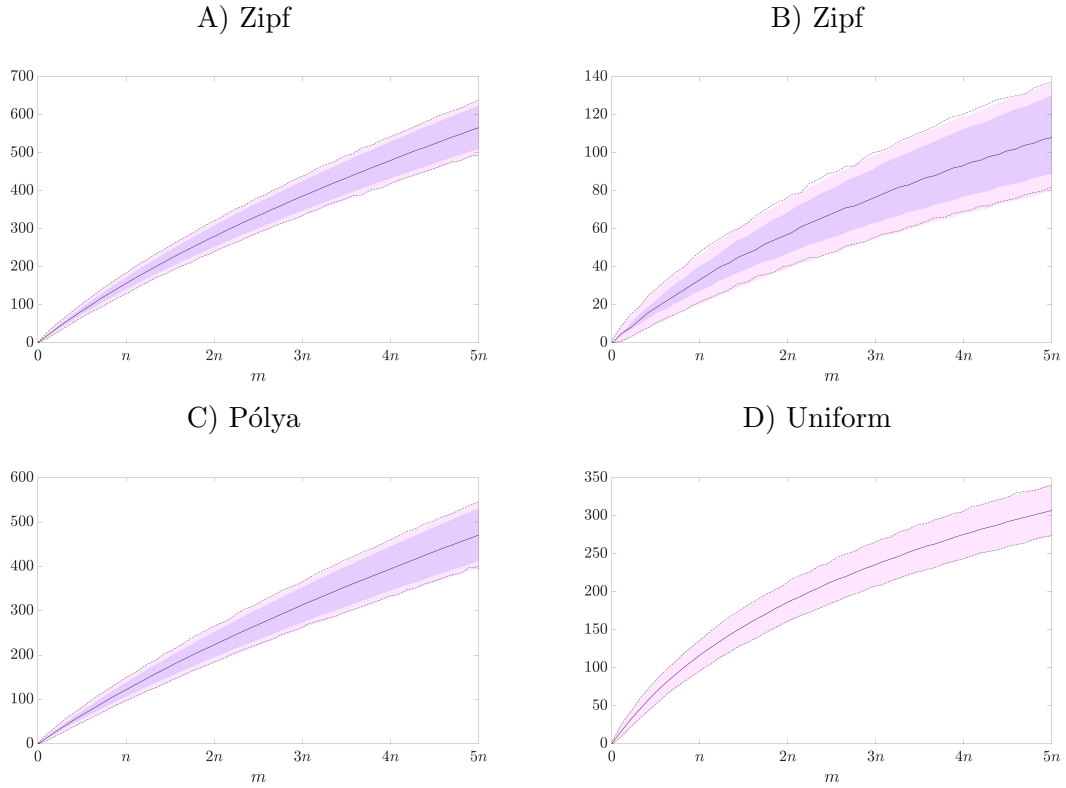

Figure 4.1: BNP estimates of $\mathcal{K}_{n,m}$ (solid line –) with 95% exact credible intervals (dashed line - -), Mittag-Leffler credible intervals (violet) and Gaussian credible intervals (pink), as a function of $m \in [0, 5n]$. Synthetic datasets generated from the following discrete distributions: A) Zipf distribution on $\{0, 1, \ldots, 300\}$ with parameter 2, $n = 977$, $j = 300$, and estimated $(\alpha, \theta) = (0.54, 26.67)$; B) Zipf distribution on $\{0, 1, \ldots, 100\}$ with parameter 1.5, $n = 1877$, $j = 100$, and estimated $(\alpha, \theta) = (0.38, 4.66)$; C) Pólya distribution on $\{0, 1, \ldots, 500\}$ with parameter (2, 2, 500, 500, ..., 500), $n = 2,000$, $j = 227$, and estimated $(\alpha, \theta) = (0.69, 1.80)$; D) Uniform distribution on $\{0, 1, \ldots, 500\}$, with $n = 2,000$, $j = 447$, and estimated $(\alpha, \theta) = (0, 178.48)$. The parameter (α, θ) is estimated through an empirical Bayes procedure [34, Section 3].

- $\mathcal{K}_{n,m} = K_{n+m} - K_n$ in the standard regime, $\mathcal{K}_{n,m} = K_{n+m}^{\{n+m\}} - K_n^{\{n+m\}}$ in the large–θ regime;
- $K_m^{(n)}$ for a variable with the following distribution,

$$P\left[K_m^{(n)} = k\right] = P\left[\mathcal{K}_{n,m} = k \,|\, K_n = j\right]$$

  in the standard regime; when considering the large–θ regime, we introduce the assumption $j = \varrho m$ and $n = \nu m$, so that

$$P\left[K_m^{(n)} = k\right] = P\left[\mathcal{K}_{\nu m,m} = k \,|\, K_{\nu m}^{\{(\nu+1)m\}} = \varrho m\right]$$

  and the distribution of $K_m^{(n)}$ only depends on ν, τ, ϱ and m.

Finally, $K_m^\star$ will be used to denote the number of blocks in the large–θ Ewens–Pitman partition of $\{1, ..., m\}$, when $\theta = (\tau + \nu)m = \lambda m$;

4.2 The BNP approach to the unseen-species problem

Following the BNP approach of [67], we consider $n \geq 1$ observations with values in the space of species' labels or symbols $\mathbb{S}$, modeled as random samples $\mathbf{X}_n = (X_1, \ldots, X_n)$ such

that

$$\begin{aligned} X_1, \ldots, X_n \,|\, P &\overset{\text{iid}}{\sim} P, \\ P &\sim \text{PYP}(\alpha, \theta), \end{aligned} \tag{4.2.1}$$

where $\text{PYP}(\alpha, \theta)$ is the Pitman-Yor prior indexed by $\alpha \in [0, 1)$ and $\theta > -\alpha$. For short, we say that $\mathbf{X}_n$ is a random sample from $\text{PYP}(\alpha, \theta)$. See 1.2.3 for details on the construction of the Pitman–Yor prior.

4.2.1 Estimates

Under the model (4.2.1), [67, Propostion 1] computes the posterior distribution of $\mathcal{K}_{n,m}$, given $\mathbf{X}_n$. For $\alpha \in (0, 1)$, this is expressed in terms of the generalized factorial coefficient

$$\mathscr{C}(u, v; a, b) := \frac{1}{v!} \sum_{i=0}^{v} (-1)^i \binom{v}{i} (-ia - b)_{(u)}$$

for $a > 0$, $b \geq 0$ and $u, v \in \mathbb{N}_0$ such that $v \leq u$, where $(a)_{(u)}$ denotes the u-th rising factorial of a, i.e. $(a)_{(u)} := \prod_{0 \leq i \leq u-1} (a + i)$; see Appendix **??**. If the sample $\mathbf{X}_n$ features $K_n = j$ species with (empirical) frequencies $(N_{1,n}, \ldots, N_{1,K_n}) = (n_1, \ldots, n_j)$ then for $k \in \{0, 1, \ldots, m\}$

$$\Pr[K_m^{(n)} = k] = \Pr[\mathcal{K}_{n,m} = k \,|\, \mathbf{X}_n] = \frac{\left(j + \frac{\theta}{\alpha}\right)_{(k)}}{(\theta + n)_{(m)}} \mathscr{C}(m, k; \alpha, -n + j\alpha). \tag{4.2.2}$$

From (4.2.2), [34, Proposition 1] provides a BNP estimator of $\mathcal{K}_{n,m}$ as posterior expectation, i.e.,

$$\hat{K}_{n,m} = \mathbb{E}[K_m^{(n)}] = \mathbb{E}[\mathcal{K}_{n,m} \,|\, \mathbf{X}_n] = \left(j + \frac{\theta}{\alpha}\right) \left(\frac{(\theta + n + \alpha)_{(m)}}{(\theta + n)_{(m)}} - 1 \right). \tag{4.2.3}$$

While the estimator (4.2.3) can be easily evaluated for any n and m, the computational burden for evaluating (4.2.2) becomes overwhelming as m increases, due the generalized factorial coefficients.

For $\alpha = 0$, the posterior distribution of $\mathcal{K}_{n,m}$, given $\mathbf{X}_n$, follows from (4.2.2) by taking the limit as $\alpha \to 0$ [67]. The resulting distribution is expressed in terms of Stirling numbers, whose evaluation remains computationally unfeasible for large values of m.The corresponding BNP estimator of $\mathcal{K}_{n,m}$ as posterior expectation is

$$\hat{K}_{n,m} = \mathbb{E}[K_m^{(n)}] = \mathbb{E}[\mathcal{K}_{n,m} \,|\, \mathbf{X}_n] = \sum_{i=1}^{m} \frac{\theta}{\theta + n + i - 1}, \tag{4.2.4}$$

which can be easily evaluated for any n and m. See [4] for further details on the case $\alpha = 0$.

4.2.2 Exact credible intervals

Exact credible intervals for $\mathcal{K}_{n,m}$ are derived by Monte Carlo sampling the posterior distribution of $\mathcal{K}_{n,m}$, given $\mathbf{X}_n$. By exploiting the closed-form expression of the posterior distribution (4.2.2), as well as the corresponding expression for $\alpha = 0$, one may implement the inverse transform algorithm to sample $K_m^{(n)}$. However, due to the evaluation of the

distribution of $K_m^{(n)}$, the inverse transform algorithm becomes computationally unfeasible for large values of m. As an alternative approach, one may consider to sample $K_m^{(n)}$ by relying on the predictive distribution, or generative scheme, of the Pitman-Yor prior [79].The use of the predictive distribution reduces the problem of Monte Carlo sampling $K_m^{(n)}$ to the problem of sampling m Bernoulli random variables, which can be easily performed for any value of n and m [4]; in particular, for $\alpha = 0$ the Bernoulli random variables are independent. This approach is reported in Algorithm 1.

Algorithm 1 Monte Carlo sampling $K_m^{(n)}$

Require: n, K_n, m, α, θ
  function MonteCarloK(n, K_n, m, α, θ)
    $K \leftarrow K_n$
    for $i \leftarrow 0$ to $m-1$ **do**
      $b \leftarrow$ Random sample from Bernoulli $\left(\frac{\theta+\alpha K}{\theta+n+i}\right)$
      $K \leftarrow K + b$
    end for
    return $K_m^{(n)} = K - K_n$
  end function

4.2.3 Large m asymptotic credible intervals

For $\alpha \in (0,1)$, [34] proposed a method to derive large m asymptotic credible intervals for $\mathcal{K}_{n,m}$. In particular, [34, Proposition 2] shows that, as $m \to +\infty$

$$\frac{K_m^{(n)}}{(\theta+n+m)^\alpha - (\theta+n)^\alpha} \xrightarrow{\text{w}} S_{\alpha,\theta}^{(n,j)} \overset{\text{d}}{=} B_{j+\theta/\alpha, n/\alpha - j} S_{\alpha,(\theta+n)/\alpha}, \tag{4.2.5}$$

where $B_{j+\theta/\alpha,n/\alpha-j}$ and $S_{\alpha,(\theta+n)/\alpha}$ are independent random variables such that: i) $B_{a,b}$ is Beta distributed with parameter $a, b > 0$; ii) $S_{\alpha,q}$ is Mittag-Leffler distributed with parameter $\alpha \in (0,1)$ and $q > 0$, i.e. with density function $f_{S_{\alpha,q}}(s) \propto s^{q-1/\alpha-1} f_\alpha(y^{1/\alpha})$, where f_α is the positive α-Stable density [97]. From (4.2.5), if $c_{\alpha,\theta,n}(m) = (\theta+n+m)^\alpha - (\theta+n)^\alpha$ then

$$K_m^{(n)} \overset{d}{\approx} c_{\alpha,\theta,n} S_{\alpha,\theta}^{(n,j)}, \tag{4.2.6}$$

namely for m sufficiently large the distribution of $K_m^{(n)}$ is approximated by means of the distribution of $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$. In particular, the scaling $c_{\alpha,\theta,n}(m)$ is determined in such a way that

$$\hat{K}_{n,m} = \mathbb{E}[K_m^{(n)}] = \mathbb{E}[c_{\alpha,\theta,n} S_{\alpha,\theta}^{(n,j)}].$$

Mittag-Leffler credible intervals for $\mathcal{K}_{n,m}$, centered on the BNP estimator $\hat{K}_{n,m}$, are then derived by Monte Carlo sampling $c_{\alpha,\theta,n}(m) S_{\alpha,\theta}^{(n,j)}$. We refer to [83] for recent developments on (exact) sampling $S_{\alpha,(\theta+n)/\alpha}$.

Instead, for $\alpha = 0$, it follows from [65, Theorem 2.3] that, as $m \to +\infty$

$$\frac{K_m^{(n)}}{\log m} \xrightarrow{\text{w}} \theta, \tag{4.2.7}$$

namely the posterior distribution has a large m limiting behaviour that is degenerate at $\theta > 0$. See also [4] for details. Such a degenerate limiting behaviour prevents from extending the Monte Carlo sampling procedure of [34] to the case $\alpha = 0$.

4.3 Gaussian credible intervals

For the BNP approach to the unseen-species problem, we present a new method to derive large m asymptotic credible intervals for $\mathcal{K}_{n,m}$, which improves over the method of [34]: firstly, it allows to deal with $\alpha \in [0,1)$, thus including the case $\alpha = 0$; secondly, it avoids the use of Monte Carlo sampling, enhancing computational efficiency. Following the notation of Section 4.2, for $\alpha \in [0,1)$ and $\theta > -\alpha$ let $\mathbf{X}_n$ by a collection of $n \geq 1$ random samples from PYP(α, θ), namely the model (4.2.1), such that $\mathbf{X}_n$ features $K_n = j$ species with (empirical) frequencies $(N_{1,n}, \ldots, N_{1,K_n}) = (n_1, \ldots, n_j)$. Our methodology relies on a large m approximation of the posterior distribution of $\mathcal{K}_{n,m}$, given $\mathbf{X}_n$, which is obtained assuming that both the sampling information (n, j) and the parameter θ are large. In particular, we set $n = \nu m$, $j = \rho m$ and $\theta = \tau m$, with $\nu, \rho, \tau > 0$, and provide a (weak) law of large numbers (LLN) and a Gaussian central limit theorem (CLT) for $K_m^{(n)}$, as $m \to +\infty$. The CLT is then applied to derive a large m Gaussian approximation of the distribution of $K_m^{(n)}$. Here, we describe the approach that leads to the Gaussian credible intervals for $\mathcal{K}_{n,m}$.

4.3.1 Preliminary results

For $\alpha \in [0,1)$ and $\theta > -\alpha$, let K_m^* be the (random) number of species in $m \geq 1$ random samples from PYP$(\alpha, \theta + n)$, such that $K_m^* \in \{1, \ldots, m\}$, and let $B_{a,b}$ be a Beta random variable with parameter $a, b > 0$. If we denote by $Q(n,p)$ a Binomial random variable with parameter $n \in \mathbb{N}$ and $p \in (0,1)$, then according to [4, Proposition 1] there hold:

i) for $\alpha \in (0,1)$

$$K_m^{(n)} \stackrel{\mathrm{d}}{=} Q\left(K_m^*, B_{\frac{\theta}{\alpha}+j, \frac{n}{\alpha}-j}\right); \tag{4.3.1}$$

ii) for $\alpha = 0$

$$K_m^{(n)} \stackrel{\mathrm{d}}{=} Q\left(K_m^*, \frac{\theta}{\theta+n}\right). \tag{4.3.2}$$

As discussed in [4], the compound Binomial representations (4.3.1)-(4.3.2) follow from the quasi-conjugacy and conjugacy properties of the Pitman-Yor and Dirichlet priors, respectively. See also [27; 28] for a more detailed account on (4.3.1).

The next result is an immediate corollary of theorem 2.1.1, and it provides a LLN and a Gaussian CLT for K_m^*, as $m \to +\infty$, under the assumption that $\theta + n$ increases linearly in m, i.e. $\theta + n = \lambda m$, with $\lambda > 0$.

Theorem 4.3.1. *For $m \in \mathbb{N}$ let $K_m^* \in \{1, \ldots, m\}$ be the (random) number of species in m random samples from PYP$(\alpha, \theta + n)$, such that: $\alpha \in [0,1)$ and $\theta + n = \lambda m$, for some $\lambda > 0$. If*

$$\mathfrak{m}_{\alpha,\lambda} = \begin{cases} \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{\alpha} - 1\right] & \text{for } \alpha \in (0,1) \\ \lambda \log\left(1+\frac{1}{\lambda}\right) & \text{for } \alpha = 0 \end{cases}$$

and

$$\mathfrak{s}^2_{\alpha,\lambda} = \begin{cases} \frac{\lambda}{\alpha}\left[\left(1+\frac{1}{\lambda}\right)^{2\alpha}\left(1-\frac{\alpha}{1+\lambda}\right) - \left(1+\frac{1}{\lambda}\right)^{\alpha}\right] & \text{for } \alpha \in (0,1) \\ \lambda \log\left(1+\frac{1}{\lambda}\right) - \frac{\lambda}{\lambda+1} & \text{for } \alpha = 0, \end{cases}$$

then, as $m \to +\infty$ there hold:

i)

$$\mathbb{E}\left[K_m^*\right] = m\mathfrak{m}_{\alpha,\lambda} + O(1)$$

and

$$\mathrm{Var}(K_m^*) = m\mathfrak{s}^2_{\alpha,\lambda} + O(1); \tag{4.3.3}$$

ii)

$$\frac{K_m^*}{m} \xrightarrow{p} \mathfrak{m}_{\alpha,\lambda}; \tag{4.3.4}$$

iii)

$$\frac{K_m^* - m\mathfrak{m}_{\alpha,\lambda}}{\sqrt{m\mathfrak{s}^2_{\alpha,\lambda}}} \xrightarrow{w} N(0,1). \tag{4.3.5}$$

Furthermore, for any $\lambda > 0$ *there hold that* $\mathfrak{m}_{0,\lambda} = \lim_{\alpha\to 0}\mathfrak{m}_{\alpha,\lambda}$ *and also that* $\mathfrak{s}^2_{0,\lambda} = \lim_{\alpha\to 0}\mathfrak{s}^2_{\alpha,\lambda}$.

4.3.2 Main result

By relying on the compound Binomial representations (4.3.1)-(4.3.2) and Theorem 4.3.1, the next theorem provides a LLN and a Gaussian CLT for $K_m^{(n)}$. See Appendix E and Appendix F for the proof.

Theorem 4.3.2. *For* $n, m \in \mathbb{N}$ *let* $K_m^{(n)} \in \{0, 1, \ldots, m\}$ *be distributed according to the posterior distribution of* $\mathcal{K}_{n,m}$, *given* $\mathbf{X}_n$ *from* $PYP(\alpha, \theta)$, *with* $\mathbf{X}_n$ *featuring* $K_n = j$ *species, such that:* $n = \nu m$, $j = \varrho m$ *and* $\theta = \tau m$, *for some* $\nu, \rho, \tau > 0$, *and* $\tau + \nu = \lambda$, *for some* $\lambda > 0$. *If*

$$\mathscr{M}_{\alpha,\tau,\nu,\varrho} = \begin{cases} \frac{\tau+\varrho\alpha}{\alpha}\left[-1 + \left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\right] & \text{for } \alpha \in (0,1) \\ \tau \log\left(1 + \frac{1}{\lambda}\right) & \text{for } \alpha = 0 \end{cases}$$

and

$$\mathscr{S}^2_{\alpha,\tau,\nu,\varrho} = \begin{cases} \frac{\tau+\varrho\alpha}{\lambda}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\left\{\frac{\lambda}{\alpha}\left[-1 + \left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\right] - \frac{\tau+\varrho\alpha}{\lambda+1}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\right\} & \text{for } \alpha \in (0,1) \\ \tau \log\left(1 + \frac{1}{\lambda}\right) - \frac{\tau^2}{\lambda(\lambda+1)} & \text{for } \alpha = 0, \end{cases}$$

then, as $m \to +\infty$ *there hold:*

i)

$$\mathbb{E}\left[K_m^{(n)}\right] = m\mathscr{M}_{\alpha,\tau,\nu,\varrho} + O(1)$$

and

$$\mathrm{Var}\left(K_m^{(n)}\right) = m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho} + O(1); \tag{4.3.6}$$

ii)

$$\frac{K_m^{(n)}}{m} \xrightarrow{p} \mathscr{M}_{\alpha,\tau,\nu,\varrho}; \tag{4.3.7}$$

iii)

$$\frac{K_m^{(n)} - m\mathscr{M}_{\alpha,\tau,\nu,\varrho}}{\sqrt{m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}}} \xrightarrow{w} N(0,1). \tag{4.3.8}$$

Furthermore, for any $\lambda > 0$ *there hold that* $\mathscr{M}_{0,\tau,\nu,\varrho} = \lim_{\alpha\to 0} \mathscr{M}_{\alpha,\tau,\nu,\varrho}$ *and also that* $\mathscr{S}^2_{0,\tau,\nu,\varrho} = \lim_{\alpha\to 0} \mathscr{S}^2_{\alpha,\tau,\nu,\varrho}$.

Remark 4.3.3. *In theorem 4.3.2, we are conditioning on an observed number of species* $K_n = j$ *with* j *proportional to* n *(and to* m*). While this scaling might appear as an additional assumption, it is in fact fully justified by the LLN (2.1.3) established in Chapter 2 , which shows that* $K_n^{\{n\}}/n$ *converges to a positive constant in the large* θ *regime.*

4.3.2.1 Sketch of the proof of Theorem 4.3.2

The asymptotic expansions (4.3.6) follow by combining (4.3.1) and (4.3.2), the asymptotic expansions (4.3.6) and standard properties of conditional expectation; see sections E.1 and E.2 for details. The proof of the LLN (4.3.7) relies on (4.3.6) and Chebychev inequality; see section E.3 for details.

The proof of the CLT (4.3.8) relies on Theorem 4.3.1, in combination with Proposition 6 and Proposition 7 below. In particular, the structure of the proof is the same for $\alpha \in (0,1)$ and for $\alpha = 0$, with some differences that are of technical nature, and can be appreciated in the proof of Proposition 6, which is deferred to Appendix F.1 (for $\alpha = 0$) and to Appendix F.2 (for $\alpha \in (0,1)$). From (4.3.1) and (4.3.2), which in the regime $\theta = \tau m$, $n = \nu m$, $j = \varrho m$, become

i) for $\alpha \in (0,1)$,

$$K_m^{(n)} \overset{\mathrm{d}}{=} Q\left(K_m^*,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)$$

ii) for $\alpha = 0$,

$$K_m^{(n)} \overset{\mathrm{d}}{=} Q\left(K_m^*,\, \frac{\tau}{\lambda}\right),$$

define

$$Q_m(z) = \begin{cases} Q\left(\lfloor mz \rfloor\,,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right) & \text{if } \alpha \in (0,1) \\ Q\left(\lfloor mz \rfloor\,,\, \frac{\tau}{\tau+\nu}\right) & \text{if } \alpha = 0 \end{cases} \tag{4.3.9}$$

where $\lfloor mz \rfloor := \max\left(k \in \mathbb{N} \,:\, k \le mz\right)$.

The next propositions are critical for the proof of Theorem 4.3.2, showing how Theorem 4.3.1 interplay with the asymptotic expansions (4.3.6). The proof is deferred to Appendix F.1 and to Appendix F.2.

Proposition 6 (Berry-Esseen theorem for $Q_m(z)$)**.** *Let* $\alpha \in [0,1)$ *and* $z > 0$*. Further, let* $\mu : (0,+\infty) \to \mathbb{R}$ *and ,* $\sigma : (0,+\infty) \to \mathbb{R}$ *be functions defined as*

$$\mu(z) = z \cdot \frac{\tau + \varrho\alpha}{\tau + \nu}$$

and

$$\sigma^2(z) = z \cdot \frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}\left[1 + \frac{\alpha z}{\tau+\nu}\right]$$

Then, as $m \to +\infty$,

$$\mathbb{E}[Q_m(z)] = m\,\mu(z) + O(1) \tag{4.3.10}$$

$$\mathrm{Var}(Q_m(z)) = m\,\sigma^2(z) + O(1) \tag{4.3.11}$$

If

$$V_m(z) := \frac{Q_m(z) - m\,\mu(z)}{\sqrt{m\,\sigma^2(z)}}.$$

then, for every $0 < \zeta_0 < \mathfrak{m}_{\alpha,\lambda} < \zeta_1 < +\infty$ *there exist* $\bar{m} = \bar{m}(\zeta_0, \zeta_1) \in \mathbb{N}$, *and a continuous function* $C = C_{\zeta_0,\zeta_1} : [\zeta_0, \zeta_1] \to (0, +\infty)$, *such that for every* $z \in [\zeta_0, \zeta_1]$ *and every* $m \geq \bar{m}$

$$\left\| F_{V_m(z)} - \Phi \right\|_\infty \leq C(z)\,\phi(m). \tag{4.3.12}$$

where

$$\phi(m) = \begin{cases} m^{-\frac{1}{6}} & \text{if } \alpha \in (0,1) \\ m^{-\frac{1}{2}} & \text{if } \alpha = 0 \end{cases}$$

This implies in particular that for every $z \in [\zeta_0, \zeta_1]$, $V_m(z) \xrightarrow{w} N(0,1)$ *as* $m \to +\infty$.

Proposition 7. *For any* $\alpha \in [0, 1)$, *the following two equalities hold:*

$$\mu\left(\mathfrak{m}_{\alpha,\lambda}\right) = \mathscr{M}_{\alpha,\tau,\nu,\varrho} \tag{4.3.13}$$

and

$$\sigma^2\left(\mathfrak{m}_{\alpha,\lambda}\right) + \mathfrak{s}^2_{\alpha,\lambda} \cdot \left(\mu'\left(\mathfrak{m}_{\alpha,\lambda}\right)\right)^2 = \mathscr{S}^2_{\alpha,\tau,\nu,\varrho}. \tag{4.3.14}$$

Now, we conclude the proof of the CLT (4.3.8). The line of reasoning is the same as in [21, Section 2.3], with K^*_m playing the role of Z_n and $Q_m(z)$ that of $R_n(z)$. Denoting by F_m the cumulative distribution function of the random variable $\frac{K^{(n)}_m - m\mathscr{M}_{\alpha,\tau,\nu,\varrho}}{\sqrt{m}\mathscr{S}_{\alpha,\tau,\nu,\varrho}}$, our aim is to prove

$$\lim_{m\to+\infty} F_m(x) := \lim_{m\to+\infty} P\left[K^{(n)}_m \leq m\mathscr{M}_{\alpha,\tau,\nu,\varrho} + \sqrt{m}\mathscr{S}_{\alpha,\tau,\nu,\varrho}\,x\right] = \Phi(x),$$

where Φ denotes the cumulative distribution function of the standard normal distribution. By resorting to the compound Binomial representations (4.3.1) (for $\alpha \in (0,1)$) and (4.3.2) (for $\alpha = 0$), and by means standard properties of conditional probability (see Appendix F for details), we write

$$F_m(x) = \mathcal{I}^{(m)}_1(x) + \mathcal{I}^{(m)}_2(x), \tag{4.3.15}$$

with

$$\mathcal{I}^{(m)}_1(x) := \int_0^{+\infty} \Phi\left(\frac{\sqrt{m}\,\left[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)\right] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\,x}{\sigma(z)}\right)\,\mu_{\frac{K^*_m}{m}}(\mathrm{d}z)$$

and

$$\begin{aligned}\mathcal{I}^{(m)}_2(x) := \int_0^{+\infty} \Bigg\{ & F_{V_m(z)}\left(\frac{\sqrt{m}\,\left[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)\right] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\,x}{\sigma(z)}\right) \\ & - \Phi\left(\frac{\sqrt{m}\,\left[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)\right] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\,x}{\sigma(z)}\right)\Bigg\}\,\mu_{\frac{K^*_m}{m}}(\mathrm{d}z).\end{aligned}$$

Accordingly, the proof of the CLT (4.3.8) is completed by showing that, for every $x \in \mathbb{R}$ there hold

$$\lim_{m\to+\infty} \mathcal{I}^{(m)}_1(x) = \Phi(x) \tag{4.3.16}$$

and

$$\lim_{m\to+\infty} \mathcal{I}^{(m)}_2(x) = 0. \tag{4.3.17}$$

This is done in Proposition 8 (for Equation (4.3.16)) and Proposition 9 (for Equation (4.3.17)) below.

Proposition 8. *For every $x \in \mathbb{R}$,*

$$\lim_{m\to+\infty} \mathcal{I}_1^{(m)}(x) := \lim_{m\to+\infty} \mathbb{E}\left[\Phi\left(\frac{\sqrt{m}\left[\mu\left(\frac{K_m^*}{m}\right) - \mathscr{M}_{\alpha,\tau,\nu,\varrho}\right] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\frac{K_m^*}{m}\right)}\right)\right] = \Phi(x).$$

Proof. It follows from Proposition 7 and (4.3.5) of Theorem 4.3.1, trough Lemma F.2, that as $m \to +\infty$

$$\sqrt{n}\left[\mu\left(\frac{K_m^*}{m}\right) - \mu(\mathfrak{m}_{\alpha,\lambda})\right] \xrightarrow{\text{w}} \mu'(\mathfrak{m}_{\alpha,\lambda})Y.$$

Further, since σ is a continuous function, by means of (4.3.4), as $m \to +\infty$, it holds

$$\sigma\left(\frac{K_m^*}{m}\right) \xrightarrow{\text{p}} \sigma(\mathfrak{m}_{\alpha,\lambda}).$$

Since $\mathscr{M}_{\alpha,\tau,\nu,\varrho} = \mu(\mathfrak{m}_{\alpha,\lambda})$ by Proposition 7, by means of Slutsky's theorem, as $m \to +\infty$

$$\frac{\sqrt{m}\left[\mu\left(\frac{K_m^*}{m}\right) - \mathscr{M}_{\alpha,\tau,\nu,\varrho}\right] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\frac{K_m^*}{m}\right)} \xrightarrow{\text{w}} \frac{\mu'(\mathfrak{m}_{\alpha,\lambda})Y + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\mathfrak{m}_{\alpha,\lambda}\right)}.$$

Since Φ is a bounded and continuous function, the proof is completed by Portmanteau theorem. ∎

Proposition 9. *For every $x \in \mathbb{R}$,*

$$\left|\mathcal{I}_2^{(m)}(x)\right| \le \int_0^{+\infty} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \to 0.$$

Proof. Fix $\varepsilon > 0$ and choose $\delta = \delta(\varepsilon) > 0$ such that $\Phi(\delta) = 1 - \varepsilon/2$. Now, let

$$\tilde{m} = \tilde{m}(\varepsilon, \zeta_0, \zeta_1) := \min\left\{m \in \mathbb{N} :\ z_0 - \frac{\delta \mathfrak{s}_{\alpha,\lambda}}{\sqrt{m}} > \zeta_0,\ \text{and}\ z_0 - \frac{\delta\Sigma}{\sqrt{m}} < \zeta_1,\right\},$$

which exists because $\mathfrak{m}_{\alpha,\lambda} \in [\zeta_0, \zeta_1]$. We set $\tilde{\zeta}_0 := z_0 - \frac{\delta \mathfrak{s}_{\alpha,\lambda}}{\sqrt{\tilde{m}}}$ and $\tilde{\zeta}_1 := z_0 + \frac{\delta \mathfrak{s}_{\alpha,\lambda}}{\sqrt{\tilde{m}}}$, and write

$$\begin{aligned}
&\int_0^{+\infty} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \qquad (4.3.18)\\
&= \int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) + \int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z)
\end{aligned}$$

and treat the terms on the right-hand side of (4.3.18) separately. For the first term on the right-hand side of (4.3.18), by means of (4.3.12), for every $m \ge \bar{m}$ it holds

$$\int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \le \phi(m) \int_{\tilde{\zeta}_0}^{\tilde{\zeta}_1} C(z)\ \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \le \phi(m)\, \mathcal{M}_C, \qquad (4.3.19)$$

where $\mathcal{M}_C := \max_{z\in[\zeta_0,\zeta_1]} C(z)$ exists because C is continuous; in particular, $\mathcal{M}_C$ is bounded on $[\tilde{\zeta}_0, \tilde{\zeta}_1]$. For the second term on the right-hand side of (4.3.18),

$$\begin{aligned}\int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) &\leq \int_{\mathbb{R}\setminus[\tilde{\zeta}_0,\tilde{\zeta}_1]} \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\ &= P\left[\frac{K_m^*}{m} \notin \left[\mathfrak{m}_{\alpha,\lambda} - \frac{\delta\mathfrak{s}_{\alpha,\lambda}}{\sqrt{\tilde{m}}}, \mathfrak{m}_{\alpha,\lambda} + \frac{\delta\mathfrak{s}_{\alpha,\lambda}}{\sqrt{\tilde{m}}}\right]\right] \\ &\leq P\left[\frac{K_m^*}{m} \notin \left[\mathfrak{m}_{\alpha,\lambda} - \frac{\delta\mathfrak{s}_{\alpha,\lambda}}{\sqrt{m}}, \mathfrak{m}_{\alpha,\lambda} + \frac{\delta\mathfrak{s}_{\alpha,\lambda}}{\sqrt{m}}\right]\right] \\ &= P\left[\frac{K_m^* - m\,\mathfrak{m}_{\alpha,\lambda}}{\sqrt{m}\mathfrak{s}_{\alpha,\lambda}} \notin [-\delta,\delta]\right].\end{aligned} \tag{4.3.20}$$

From (4.3.18), with (4.3.19) and (4.3.20), we obtained that, for every $n \geq \max(\bar{m}, \tilde{m})$ it holds

$$\int_0^{+\infty} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \leq \phi(m)\,\mathcal{M}_C + P\left[\frac{K_m^* - m\,\mathfrak{m}_{\alpha,\lambda}}{\sqrt{m}\mathfrak{s}_{\alpha,\lambda}} \notin [-\delta,\delta]\right].$$

Hence,

$$\begin{aligned}0 &\leq \limsup_{m\to+\infty} \int_0^{+\infty} \left\|F_{V_m(z)} - \Phi\right\|_\infty \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\ &\leq \lim_{m\to+\infty} \left\{\phi(m)\,\mathcal{M}_C + P\left[\frac{K_m^* - m\,\mathfrak{m}_{\alpha,\lambda}}{\sqrt{m}\,\mathfrak{s}_{\alpha,\lambda}} \notin [-\delta,\delta]\right]\right\} \\ &= 0 + 2 - 2\Phi(\delta) \\ &= \varepsilon,\end{aligned}$$

where the last identities follows from (4.3.5) and the definition of δ such that $\Phi(\delta) = 1-\varepsilon/2$, respectively. The proof is completed by the arbitrariness of ε. ■

This completes the proof of the CLT (4.3.8), and the proof of Theorem 4.3.2. See Appendix E and Appendix F.

4.3.3 Credible intervals

Theorem 4.3.2 provides a large m Gaussian approximation of the distribution of $K_m^{(n)}$; in particular the approximation is centered on the BNP estimator $\hat{K}_{n,m}$. From (4.3.8),

$$K_m^{(n)} \overset{d}{\approx} N\left(m\mathscr{M}_{\alpha,\tau,\nu,\varrho},\, m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}\right)$$

namely for m sufficiently large the distribution of $K_m^{(n)}$ is approximated by a Gaussian distribution with mean $m\mathscr{M}_{\alpha,\tau,\nu,\varrho}$ and variance $m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}$. This approximation is applied to construct Gaussian credible intervals for $\mathcal{K}_{n,m}$ with a prescribed (asymptotic) probability level for $K_m^{(n)}$. Given $\delta \in (0,1)$, a $(1-\delta)$-level symmetric credible interval centered at the mean is

$$\Pr\left[K_m^{(n)} \in \left[m\mathscr{M}_{\alpha,\tau,\nu,\varrho} - q_{\frac{\delta}{2}}\sqrt{m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}},\; m\mathscr{M}_{\alpha,\tau,\nu,\varrho} - q_{1-\frac{\delta}{2}}\sqrt{m\mathscr{S}^2_{\alpha,\tau,\nu,\varrho}}\right]\right] \approx 1-\delta, \tag{4.3.21}$$

where $q_\varepsilon := \Phi^{-1}(\varepsilon)$ is the ε-quantile of the standard Gaussian distribution. The Gaussian credible interval (4.3.21) is fully analytical, thus avoiding the use of Monte Carlo sampling algorithms.

4.4 Numerical illustrations

We validate our methodology on synthetic and real data, comparing its performance with that of state-of-the-art procedures for the construction of exact and asymptotic credible intervals for $K_{m,n}$.

Under the Pitman-Yor prior, the BNP approach to estimate $\mathcal{K}_{n,m}$ requires the specification of the prior's parameters $\alpha \in [0,1)$ and $\theta > -\alpha$. Here, we adopt an empirical Bayes approach to estimate (α, θ), which relies on the marginal distribution of $\mathbf{X}_n = (X_1, \ldots, X_n)$ from PYP(α, θ) [34, Section 3]. For $\mathbf{X}_n$ featuring K_n distinct species with (empirical) frequencies $\mathbf{N}_n = (N_{1,n}, \ldots, N_{K_n,n})$, such a distribution is Ewens-Pitman's formula [79],

$$Pr\left[K_n = j, \mathbf{N}_n = \mathbf{n}\right] = \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta+1)_{(n-1)}} \prod_{j=1}^{k} (1-\alpha)_{(n_j-1)}. \tag{4.4.1}$$

The empirical Bayes approach estimates the parameters (α, θ) with the values $(\hat{\alpha}_n, \hat{\theta}_n)$ that maximizes the distribution (marginal likelihood) (4.4.1) corresponding to the observed sample $(j, n_1, \ldots, n_j)$, i.e.,

$$(\hat{\alpha}_n, \hat{\theta}_n) = \arg \max_{\{(\alpha,\theta)\,:\,\alpha\in[0,1),\theta>-\alpha\}} \left\{ \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta+1)_{(n-1)}} \prod_{j=1}^{k} (1-\alpha)_{(n_j-1)} \right\}. \tag{4.4.2}$$

Alternatively, one may specify a prior on (α, θ) and pursue fully Bayes estimates of prior's parameters. Due to the large sample size of the datasets considered, we do not expect substantial differences between empirical Bayes and fully Bayes. See [4, Section 6] for a discussion of the estimation of (α, θ) and issues thereof, especially with respect to $\theta > 0$.

4.4.1 Synthetic data

We test our method on synthetic data generated from a collection of discrete distributions taken from [72, Figure 3]. More precisely, we generate four datasets, which correspond to the four panels of Figure 4.1: A) Zipf distribution on $\{0, 1, ..., 300\}$ with parameter 2; B) Zipf distribution on $\{0, 1, ..., 100\}$ with parameter 1.5; C) Pólya distribution on $\{0, 1, ..., 500\}$ with parameter $\beta = (\beta_1, ..., \beta_{500})$ with $\beta_1 = \beta_2 = 2$ and $\beta_i = 500$ for all $i \geq 3$; D) Uniform distribution on $\{0, 1, ..., 500\}$. For each dataset, Table 4.1 collects the sample size n, the number of distinct species j, and the empirical Bayes estimates $(\hat{\alpha}_n, \hat{\theta}_n)$ of (α, θ).

Dataset	n	j	$\hat{\alpha}_n$	$\hat{\theta}_n$
A) Zipf	977	300	0.54	26.67
B) Zipf	1877	100	0.38	4.66
C) Pólya	2000	215	0.64	2.39
D) Uniform	2000	447	0	178.48

Table 4.1: Sample size n, number of distinct species j in the sample, and empirical Bayes estimates of (α, θ) for datasets A, B, C and D.

For datasets A, B, C and D, Figure 4.1 displays the BNP estimates of $\mathcal{K}_{n,m}$ with the 95% exact credible intervals, Mittag-Leffler credible intervals and Gaussian credible intervals as a function of m, for $m \in [0, 5n]$. Table 4.2 contains the BNP estimates of $\mathcal{K}_{n,m}$ for $m = 2, 2n, 3n, 4n, 5n$, and the corresponding 95% exact credible intervals, Mittag-Leffler credible intervals with their coverages of exact intervals, and Gaussian credible

intervals with their coverages of exact intervals. The coverage is defined as the ratio between the length of the intersection of the (rounded to the nearest integer) exact credible interval with the (rounded) approximate credible interval, and the length of the (rounded) exact interval. Table 4.3 contains the same quantities as Table 4.2, for larger values of m. We refer to Appendix G for figures displaying the coverage of Mittag-Leffler and Gaussian credible intervals as a function of $m \in [0, 5n]$. Exact intervals in Table 4.2 are derived by Monte Carlo sampling, through the inverse transform algorithm, the posterior distribution of $\mathcal{K}_{n,m}$, given $\mathbf{X}_n$. Instead, due to the larger values of m, the exact intervals in Table 4.3 are derived by Monte Carlo sampling, through Algorithm 1. In both algorithms, we applied 2000 Monte Carlo samples.

Dataset	m	$\hat{K}_{n,m}$	95% Exact C.I.	Mittag-Leffler C.I.		Gaussian C.I.	
				95% C.I.	Coverage (%)	95% C.I.	Coverage (%)
A) Zipf, $n = 977$	n	156	(130, 184)	(141, 173)	59.3	(129, 183)	98.1
	$2n$	280	(241, 321)	(252, 309)	71.3	(239, 320)	98.8
	$3n$	386	(335, 439)	(348, 426)	75.0	(334, 437)	98.1
	$4n$	480	(423, 541)	(433, 530)	82.2	(419, 541)	100
	$5n$	566	(501, 638)	(511, 625)	83.2	(496, 636)	98.5
B) Zipf, $n = 1877$	n	33	(22, 47)	(28, 40)	48.0	(21, 46)	96.0
	$2n$	57	(40, 77)	(47, 69)	59.5	(39, 76)	97.3
	$3n$	77	(57, 102)	(63, 92)	64.4	(55, 99)	93.3
	$4n$	93	(69, 119)	(77, 112)	70.0	(68, 119)	100
	$5n$	108	(80, 137)	(89, 129)	70.2	(80, 136)	98.2
C) Pólya, $n = 2000$	n	122	(98, 149)	(107, 139)	62.7	(96, 149)	100
	$2n$	224	(185, 265)	(195, 254)	73.8	(183, 264)	98.8
	$3n$	313	(263, 369)	(273, 356)	78.3	(261, 366)	97.2
	$4n$	395	(334, 460)	(344, 449)	83.3	(332, 458)	98.4
	$5n$	471	(398, 549)	(410, 535)	82.8	(398, 544)	96.7
D) Uniform, $n = 2000$	n	116	(96, 137)	–	–	(96, 137)	100
	$2n$	186	(160, 211)	–	–	(160, 212)	100
	$3n$	236	(206, 265)	–	–	(207, 266)	98.3
	$4n$	275	(244, 309)	–	–	(243, 307)	96.9
	$5n$	307	(274, 341)	–	–	(273, 341)	100

Table 4.2: Additional sample m, BNP estimates of $\mathcal{K}_{n,m}$, 95% exact C.I., Mittag-Leffler C.I. with their coverages (of the the exact C.I.) and Gaussian credible intervals with their coverages (of the exact C.I.). All values are rounded to the nearest integer.

Dataset	m	$\hat{K}_{n,m}$	95% Exact C.I.	Mittag-Leffler C.I.		Gaussian C.I.	
				95% C.I.	Coverage (%)	95% C.I.	Coverage (%)
A) Zipf, $n = 977$	$10n$	923	(813, 1030)	(834, 1019)	85.3	(817, 1029)	97.7
	$50n$	2582	(2299, 2863)	(2332, 2851)	92.0	(2311, 2854)	96.3
	$100n$	3904	(3519, 4304)	(3526, 4311)	99.1	(3501, 4307)	100
	$1000n$	14493	(13064, 16031)	(13088, 16002)	98.2	(13038, 15949)	97.2
B) Zipf, $n = 1877$	$10n$	165	(127, 205)	(134, 196)	79.5	(125, 204)	98.7
	$50n$	381	(304, 467)	(309, 453)	88.3	(301, 460)	95.7
	$100n$	525	(421, 636)	(427, 625)	92.1	(419, 632)	98.1
	$1000n$	1400	(1131, 1667)	(1137, 1664)	98.1	(1132, 1668)	99.8
C) Pólya, $n = 2000$	$10n$	799	(685, 913)	(701, 908)	90.8	(682, 915)	100
	$50n$	2502	(2171, 2857)	(2195, 2845)	94.8	(2165, 2839)	97.4
	$100n$	3998	(3491, 4547)	(3508, 4547)	98.4	(3467, 4529)	98.3
	$1000n$	18139	(15824, 20559)	(15915, 20629)	98.1	(15776, 20501)	98.8
D) Uniform, $n = 2000$	$10n$	414	(375, 457)	–	–	(375, 453)	95.1
	$50n$	687	(636, 739)	–	–	(636, 738)	99.0
	$100n$	809	(757, 864)	–	–	(753, 864)	100
	$1000n$	1218	(1150, 1286)	–	–	(1150, 1286)	100

Table 4.3: Additional sample m, BNP estimates of $\mathcal{K}_{n,m}$, 95% exact C.I., Mittag-Leffler C.I. with their coverages (of the the exact C.I.) and Gaussian credible intervals with their coverages (of the exact C.I.). All values are rounded to the nearest integer.

Figure 4.1, as well as from Table 4.2 and the rows of Table 4.3 corresponding to

$m = 10n$ and $m = 50n$, show that Mittag-Leffler credible intervals have a smaller coverage than Gaussian credible intervals. As expected, the performance of Mittag-Leffler credible intervals improves as m grows. In particular, for $m = 100n$ and $m = 1000n$ the coverage of the Mittag-Leffler credible intervals eventually matches or outperforms that of the Gaussian credible intervals, which in any case maintain a coverage of at least 97%. Such a behaviour provides a further indication on the magnitude of the additional sample size m for which the Mittag-Leffler regime of approximation may be more suited than the Gaussian regime. Furthermore, the Gaussian intervals appear to be near-to-optimal, in the sense that their length is always comparable to that of the exact intervals even when they are larger. For the Uniform dataset, where the empirical Bayes estimate for α is 0, Gaussian credible intervals sill work, preserving their good performance in terms of coverage of the exact interval. Instead, the method of [34] does not apply for $\alpha = 0$, as discussed in Section 4.2.3.

4.4.2 Real data

For real data analysis, we consider the same Expressed Sequence Tags (EST) datasets previously analyzed in [34]. These datasets are generated by sequencing cDNA libraries consisting of millions of genes and one of the main quantities of interest is the number of distinct genes. Due to the cost of the sequencing procedure, only a small portion of the cDNA library is typically sequenced. Given the resulting sequenced sample of size n, it is required to estimate the number of new genes $\mathcal{K}_{n,m}$ to appear in an additional sample of size m. On the basis of such estimates, geneticists decide whether it is worth proceeding with sequencing and, if so, up to which additional sample size. The five libraries considered are: i) tomato flower cDNA library [84]; ii) two cDNA libraries of the amitochondriate protist *Mastigamoeba balamuthi*, one of which has undergone a normalization protocol [88]; iii)two *Naegleria gruberi* cDNA libraries from cells grown respectively in aerobic and anaerobic conditions [88].

For each EST dataset, Table 4.4 collects the sample size n, the number of distinct species j, and the empirical Bayes estimates $(\hat{\alpha}_n, \hat{\theta}_n)$ of (α, θ); Table 4.4 coincides with [34, Table 1], where the same datasets are analyzed. For these datasets, Table 4.5 contains the BNP estimates of $\mathcal{K}_{n,m}$ for $m = 2$, $2n$, $3n$, $4n$, $5n$, and the corresponding 95% exact credible intervals, Mittag-Leffler credible intervals with their coverages of exact intervals, and Gaussian credible intervals with their coverages of exact intervals. Table 4.6 contains the same quantities as Table 4.5, for larger values of m. We refer to Appendix G for figures displaying the coverage of Mittag-Leffler and Gaussian credible intervals as a function of $m \in [0, 5n]$.

Library	n	j	$\hat{\alpha}_n$	$\hat{\theta}_n$
Tomato flower	2586	1825	0.612	741.0
Mastigamoeba	715	460	0.770	46.0
Mastigamoeba–normalized	363	248	0.700	57.0
Naegleria aerobic	959	473	0.670	46.3
Naegleria anaerobic	969	631	0.660	155.5

Table 4.4: Sample size n, number of distinct species j in the sample, and empirical Bayes estimates of (α, θ) for the five EST datasets.

For the *Naegleria* aerobic dataset, Figure 4.2 displays BNP estimates of $\mathcal{K}_{n,m}$ with 95% exact credible intervals, Mittag-Leffler credible intervals and Gaussian credible intervals

Library	m	$\hat{K}_{n,m}$	95% Exact C.I.	Mittag-Leffler C.I.		Gaussian C.I.	
				95% C.I.	Coverage (%)	95% C.I.	Coverage (%)
Tomato flower	n	1281	(1223, 1339)	(1244, 1321)	66.4	(1222, 1340)	100
$n = 2586$	$2n$	2354	(2264, 2446)	(2287, 2427)	76.9	(2262, 2445)	99.5
	$3n$	3305	(3184, 3424)	(3211, 3409)	82.5	(3185, 3425)	99.6
	$4n$	4173	(4031, 4318)	(4054, 4304)	87.1	(4028, 4319)	100
	$5n$	4980	(4815, 5146)	(4838, 5136)	90.0	(4811, 5148)	100
Mastigamoeba	n	346	(312, 379)	(323, 369)	68.7	(312, 379)	100
$n = 715$	$2n$	654	(596, 706)	(610, 697)	79.1	(599, 708)	97.3
	$3n$	939	(866, 1014)	(875, 1001)	85.1	(865, 1012)	98.6
	$4n$	1208	(1119, 1301)	(1126, 1288)	89.0	(1116, 1299)	98.9
	$5n$	1465	(1357, 1578)	(1366, 1562)	88.7	(1356, 1573)	97.7
Mastigamoeba–norm.	n	180	(157, 202)	(164, 197)	73.3	(157, 203)	100
$n = 363$	$2n$	336	(299, 371)	(306, 367)	84.7	(299, 372)	100
	$3n$	477	(429, 525)	(435, 522)	90.6	(428, 526)	100
	$4n$	608	(546, 671)	(555, 666)	88.8	(548, 668)	96.0
	$5n$	732	(660, 803)	(668, 801)	93.0	(662, 803)	98.6
Naegleria aerobic	n	307	(272, 343)	(284, 331)	66.2	(272, 342)	98.6
$n = 959$	$2n$	566	(514, 621)	(524, 611)	81.3	(511, 622)	100
	$3n$	798	(730, 873)	(739, 861)	85.3	(726, 871)	98.6
	$4n$	1012	(921, 1099)	(937, 1091)	86.5	(923, 1101)	98.9
	$5n$	1212	(1108, 1319)	(1122, 1307)	87.7	(1109, 1315)	97.6
Naegleria anaerobic	n	439	(402, 476)	(415, 465)	67.6	(402, 476)	100
$n = 969$	$2n$	812	(753, 871)	(767, 860)	78.8	(754, 870)	98.3
	$3n$	1146	(1065, 1219)	(1083, 1213)	84.4	(1069, 1223)	97.4
	$4n$	1454	(1365, 1550)	(1373, 1538)	89.2	(1360, 1547)	98.4
	$5n$	1741	(1635, 1855)	(1645, 1843)	90.0	(1632, 1851)	98.2

Table 4.5: Additional sample m, BNP estimates of $\mathcal{K}_{n,m}$, 95% exact C.I., Mittag-Leffler C.I. with their coverages (of the the exact C.I.) and Gaussian credible intervals with their coverages (of the exact C.I.). All values are rounded to the nearest integer.

Library	m	$\hat{K}_{n,m}$	95% Exact C.I.	Mittag-Leffler C.I.		Gaussian C.I.	
				95% C.I.	Coverage (%)	95% C.I.	Coverage (%)
Tomato flower	$10n$	8432	(8171, 8705)	(8188, 8687)	93.4	(8164, 8700)	99.1
$n = 2586$	$50n$	25926	(25137, 26706)	(25176, 26712)	97.5	(25154, 26698)	98.4
	$100n$	40888	(39690, 42082)	(39705, 42128)	99.4	(39684, 42092)	100
	$1000n$	113848	(110620, 117142)	(110555, 117300)	100	(110540, 117157)	100
Mastigamoeba	$10n$	2634	(2442, 2825)	(2463, 2809)	90.3	(2448, 2819)	96.9
$n = 715$	$50n$	9718	(9010, 10370)	(9089, 10367)	94.0	(9063, 10372)	96.1
	$100n$	16797	(15664, 17928)	(15711, 17920)	97.6	(15674, 17921)	99.2
	$1000n$	58889	(54962, 62878)	(55079, 62824)	97.8	(54976, 62803)	98.9
Mastigamoeba–norm.	$10n$	1280	(1158, 1392)	(1169, 1393)	95.3	(1163, 1397)	97.9
$n = 363$	$50n$	4344	(3951, 4707)	(3966, 4729)	98.0	(3969, 4720)	97.6
	$100n$	7203	(6581, 7809)	(6576, 7840)	100	(6585, 7820)	99.7
	$1000n$	22759	(20826, 24856)	(20779, 24774)	98.0	(20825, 24694)	96.0
Naegleria aerobic	$10n$	2084	(1920, 2253)	(1926, 2246)	96.1	(1917, 2252)	99.7
$n = 959$	$50n$	6781	(6260, 7299)	(6265, 7306)	99.5	(6270, 7293)	98.5
	$100n$	11030	(10209, 11923)	(10190, 11883)	97.7	(10207, 11852)	95.9
	$1000n$	33286	(30777, 35851)	(30752, 35862)	100	(30833, 35740)	96.7
Naegleria anaerobic	$10n$	2994	(2809, 3175)	(2826, 3161)	91.5	(2817, 3171)	96.7
$n = 969$	$50n$	9679	(9145, 10213)	(9135, 10218)	100	(9140, 10218)	100
	$100n$	15671	(14809, 16530)	(14791, 16544)	100	(14807, 16535)	100
	$1000n$	46683	(44226, 49260)	(44062, 49283)	100	(44137, 49229)	99.4

Table 4.6: Additional sample m, BNP estimates of $\mathcal{K}_{n,m}$, 95% exact C.I., Mittag-Leffler C.I. with their coverages (of the the exact C.I.) and Gaussian credible intervals with their coverages (of the exact C.I.). All values are rounded to the nearest integer.

as a function of m, for $m \in [0, 5n]$. See Appendix G for similar plots for the other datasets described in Table 4.4.

Tables 4.5 and 4.6 confirm the behavior already observed on synthetic data, with Gaussian credible intervals providing a better coverage than Mittag-Leffler credible intervals for values of m corresponding to $m = n, 2n, 3n, 4n, 5n, 10n$ and $50n$, and displaying near-to-optimal length. Again, the Mittag-Leffler credible intervals display a substantial improvement in performance as m grows, with their coverage eventually matching or out-

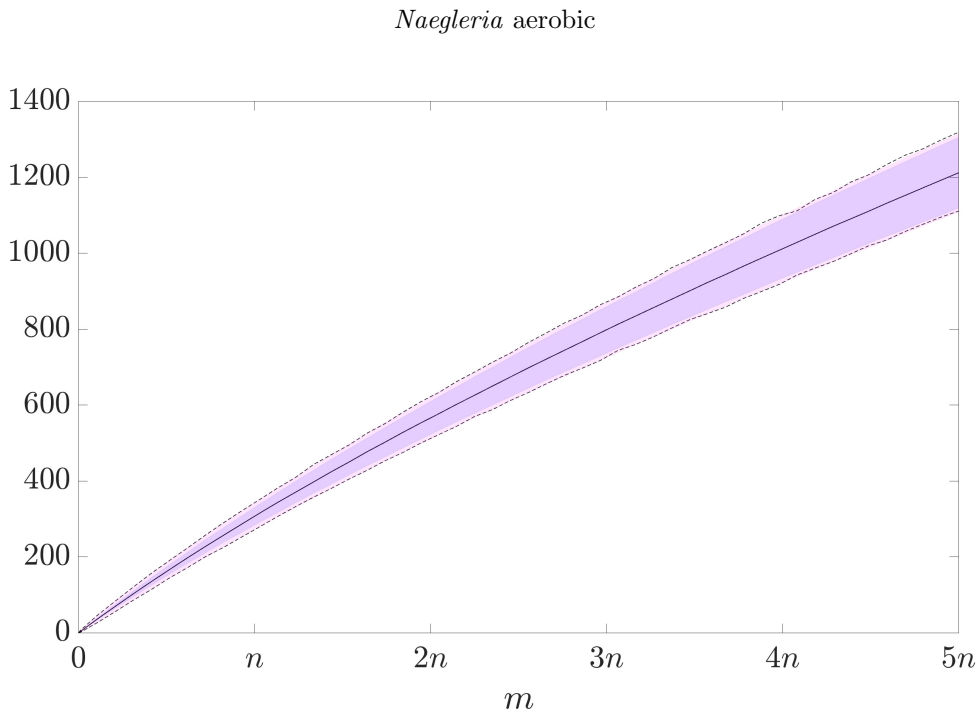

Figure 4.2: BNP estimates of $\mathcal{K}_{n,m}$ (solid line –) with 95% exact credible intervals (dashed line - -), Mittag-Leffler credible intervals (violet) and Gaussian credible intervals (pink), as a function of m, for $m \in [0, 5n]$.

performing that of the Gaussian credible intervals for values of $m = 100n, m = 1000n$. It shall be mentioned that, also when outperformed, the Gaussian intervals maintain in any case a coverage of at least 96%.

Chapter 5

Discussion and directions for future work

5.1 Results of chapter 2

Motivated by the interpretation of the Ewens model in population genetics, [40] first investigated the large n asymptotic behaviour of K_n in the regime $\theta = \lambda n$, with $\lambda > 0$, providing a large deviation principle [40, Theorem 4.1 and Theorem 4.4]. Subsequently, still for the Ewens model, a LLN and a CLT for K_n were established in [91, Proposition 2 and Theorem 2]; see also [92]. In this chapter, we extended the LLN and the CLT to the more general Ewens-Pitman model, and also established a Berry-Esseen theorem for $\alpha = 0$.

5.1.1 Extensions of Theorem 2.1.1

An interesting extension of Theorem 2.1.1 would be to consider the more general regime $\theta = \lambda n^{\beta}$, for $\lambda > 0$ and $\beta \in [0, 1]$. For $\alpha = 0$, this regime has been investigated in [40] and [91], yielding a large deviation principle, a LLN and a CLT. A natural continuation would be to establish analogous results for $\alpha \in (0, 1)$. This direction has interesting applications in the context of microclustering data analysis, where the number of clusters grows with the sample size while their average size remains small, a feature commonly observed in modern large-scale data settings such as entity resolution and network analysis [7].

5.2 Results of chapter 3

In this chapter we have established an alternative proof of theorem 2.1.1 by adapting the arguments already used in [9] to the large θ regime. More specifically, the proof consists in exploiting the sequential contruction presented in section 1.3.1 to construct a martingale whose relation with the variables $K_h^{\{n\}}$ is explicit. Upon establishing a CLT for such martingale, the CLT for $K_h^{\{n\}}$ is derived by means of standard arguments. Further, a $d-$dimensional generalization of this argument has allowed us to prove a strong LLN and a joint CLT (theorem 3.2.1) for the $d-$dimensional vector

$$\mathbf{K}_{d,n}^{\{n\}} = (K_n, K_{1,n}, ..., K_{d,n})$$

thereby providing a finer description of the partition structure.

Beyond what already discussed in section 5.1.1, which would be worth exploring for the whole partition structure, we note that the martingale argument could be adapted without too much effort to obtain analogous results for general partition models whose sequential construction is explicitly known. The martingale structure could be further explored in order to recover the Berry–Esseen theorem (2.1.6) proven in Chapter 2 and extend it to the case $\alpha \in (0,1)$; further asymptotic properties of the partition, such as LDPs, could be similarly recovered from refinements of our martingale argument.

5.2.1 Estimation of α and λ – consistency results.

In the standard Pitman–Yor setting, estimation of the parameter α poses in general no problems, with more than one consistent estimator available in the literature. The estimation of θ is instead more troublesome, with some commonly used estimators (e.g. the MLE) which have been proven to be inconsistent [4, Section 6].

It is of particular interest, then, that for the model under the large θ regime, theorem 3.2.1 immediately implies the following result via the use of the delta–method: let $g : \mathbb{R}^3 \to \mathbb{R}^2$ be defined as

$$g(x,y,z) = \left(1 - \frac{2y^2}{3xz - 2y^2}\,,\, -1 + \frac{xy}{3xz - 2y^2}\right).$$

also let $\tilde{\mathfrak{M}}_{3,\alpha,\lambda}$ denote the vector obtained removing the first entry of $\mathfrak{M}_{3,\alpha,\lambda}$ and $\tilde{\Gamma}_{3,\alpha,\lambda}$ the 3×3 minor obtained removing the first row and the first column of $\Gamma_{3,\alpha,\lambda}$.

Proposition 5.2.1. *Define*

$$\left(\hat{\alpha}, \hat{\lambda}\right) := g\left(K_{1,n}^{\{n\}}, K_{2,n}^{\{n\}}, K_{3,n}^{\{n\}}\right).$$

Then, as $n \to +\infty$,

(i)

$$\left(\hat{\alpha}, \hat{\lambda}\right) \xrightarrow{a.s.} g\left(\tilde{\mathfrak{M}}_{3,\alpha,\lambda}\right) = (\alpha, \lambda)$$

(ii)

$$\sqrt{n}\left(\hat{\alpha} - \alpha, \hat{\lambda} - \lambda\right) \xrightarrow{a.s.} \mathcal{N}\left(\mathbf{0}, S_{\alpha,\lambda}\right)$$

where $S_{\alpha,\lambda} = \nabla g(\tilde{\mathfrak{M}}_{3,\alpha,\lambda})\, \tilde{\Gamma}_{3,\alpha,\lambda}\, \nabla g(\tilde{\mathfrak{M}}_{3,\alpha,\lambda})^T$.

This result gives an estimator of the two parameters α and λ of the model which has a simple form, is consistent, and for which (asymptotic) confidence intervals are available via the CLT. While the limit variance $S_{\alpha,\lambda}$ has, in the specific case of the estimator of the proposition, an involute form which makes it less appealing for application, work is ongoing [8] to explore the possibility of deriving an analogous result for other simple estimators of α and λ. Finally, to avoid confusion, we stress that our consistency result does not imply the existence of a consistent estimator for $\theta = \lambda n$ in our model, due to its dependence on the sample size.

5.3 Results of chapter 4

The unseen-species problem is arguably the most popular example of "species-sampling" problem. Given $n \geq 1$ observed sample $\mathbf{X}_n = (X_1, \ldots, X_n)$ from a population of individuals belonging to different species $\mathbb{S}$, a broad class of "species-sampling" (extrapolation) problems call for estimating features of the unknown species composition of $m \geq 1$ additional unobservable samples $\mathbf{X}_m = (X_{n+1}, \ldots, X_{n+m})$ from the same population. If $(N_{s,n})_{s\in\mathbb{S}}$ and $(N_{s,m})_{s\in\mathbb{S}}$ denote the (empirical) frequencies of species in $\mathbf{X}_n$ and $\mathbf{X}_m$, respectively, for $u, v \geq$ we set

$$\mathcal{K}_{n,m}(u,v) = \sum_{s\in\mathbb{S}} I(N_{s,n} = u) I(N_{s,m} = v), \tag{5.3.1}$$

with $I(\cdot)$ being the indicator function, namely the number of species with frequency u in $\mathbf{X}_n$ and with frequency v in $\mathbf{X}_m$. The number of unseen species is recovered from (5.3.1) by taking

$$\mathcal{K}_{n,m} = \sum_{v=1}^{m} \mathcal{K}_{n,m}(0,v).$$

We refer to [4] for an up-to-date overview on "species-sampling" problems of the form (5.3.1), with applications to biological data. While the emphasis is on the BNP approach, [4] also discusses the most recent advances in the distribution-free approach.

5.3.1 Gaussian credible intervals for $\mathcal{K}_{n,m}(0,v)$

Among "species-sampling" problems of the form (5.3.1), the number of unseen rare species certainly stands out [26], with a rich literature under the distribution-free and BNP approach [35; 56]. It calls for estimating $\mathcal{K}_{n,m}(0,v)$, namely the number of hitherto unseen species that would be observed with frequency $v \geq 1$ in the m additional samples. Under the BNP approach with a Pitman-Yor prior, with $\alpha \in [0,1)$ and $\theta > -\alpha$, [4, Proposition 1] provides a compound Binomial representation of the posterior distribution of $\mathcal{K}_{n,m}(0,v)$, given $\mathbf{X}_n$. In particular, let $\mathcal{K}^{(n)}_{n,m}(0,v)$ be random variable with such a distribution. Following the notation of (4.3.1)-(4.3.2), and denoting by $K^*_m(v)$ the (random) number of species with frequency v in $m \geq 1$ random samples from PYP$(\alpha, \theta + n)$, such that $K^*_m(v) \in \{0, 1, \ldots, m\}$, there hold

i) for $\alpha \in (0,1)$

$$\mathcal{K}^{(n)}_{n,m}(0,v) \overset{\mathrm{d}}{=} Q\left(K^*_m(v),\, B_{\frac{\theta}{\alpha}+k,\, \frac{n}{\alpha}-k}\right); \tag{5.3.2}$$

ii) for $\alpha = 0$

$$\mathcal{K}^{(n)}_{n,m}(0,v) \overset{\mathrm{d}}{=} Q\left(K^*_m(v),\, \frac{\theta}{\theta+n}\right). \tag{5.3.3}$$

The representations (5.3.2)-(5.3.3) lead to posterior mean estimates of $\mathcal{K}_{n,m}(0,v)$ that can be easily evaluated for any value of any n and m [35, Equation 24 and Equation 30].

Exact credible intervals for $\mathcal{K}_{n,m}(0,v)$ can be derived by Monte Carlo sampling the posterior distribution of $\mathcal{K}_{n,m}(0,v)$, given $\mathbf{X}_n$. One may exploit the representations (5.3.2)-(5.3.3), as well as the predictive distributions of the Pitman-Yor prior, to sample $\mathcal{K}^{(n)}_{n,m}(0,v)$, though with a computational burden that becomes overwhelming as m increases. In particular, Algorithm 1 does not extend to $\mathcal{K}^{(n)}_{n,m}(0,v)$; see [4] for details. Large m asymptotic credible intervals for $\mathcal{K}_{n,m}(0,v)$ can derived by relying on the large m

asymptotic behaviour of $\mathcal{K}_{n,m}^{(n)}(0,v)$ in [35, Theorem 4], still involving the scaled Mittag-Leffler distribution. However, this approach would suffer from the same limitations as the approach of [34] for $\mathcal{K}_{n,m}$. Alternatively, one may consider to extend our approach to in order to derive Gaussian credible intervals for $\mathcal{K}_{n,m}(0,v)$, which is a promising direction for future work. This would require to extend Theorem 2.1.1 to $K_m^*(v)$, and then combine it with the representations (5.3.2)-(5.3.3) along the same lines of Theorem 4.3.2.

5.3.2 More directions for future work

Beyond the extension of our approach to other "species-sampling", this work opens several opportunities for future research. For example, one may consider the problem of deriving exact credible intervals for $\mathcal{K}_{n,m}$, namely for any n and m, by avoiding the use of Monte Carlo sampling the posterior distribution. By relying on the representations (5.3.2)-(5.3.3), this problem would require to refine the CLT (2.1.4) with a Berry-Esseen type inequality or, better, to develop a concentration inequality for K_m^*. See [21, Equation 11] for a preliminary result in this direction, though limited to the case $\alpha = 0$. Another research direction could involve the use of the Gaussian credible interval (4.3.21) as a confidence interval under the semi-parametric approach of [38], which assumes the tail of P to be regularly varying of index $\alpha \in (0,1)$. This problem would open to the study of frequentist properties of our approach to construct Gaussian credible intervals under regular variation.

Appendices

A Details on the proof of Theorem 2.1.1 for $\alpha \in (0,1)$

A.1 Proof of Proposition 1

A.1.1 Proof of the asymptotic expansions (2.2.17), (2.2.18)

The expression of $\mathbb{E}[S_{\lambda n,\alpha}]$ is known form (2.2.10). Furthermore, one has that

$$\mathbb{E}\left[G_{A,1}^{k\alpha}\right] = \frac{\Gamma(A+k\alpha)}{\Gamma(A)}.$$

By independence,

$$\mathbb{E}[Z_n] = \mathbb{E}[S_{\lambda n,\alpha}]\,\mathbb{E}\left[G_{(\lambda+1)n,1}^{\alpha}\right] = \frac{\lambda n}{\alpha}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+\alpha)}\frac{\Gamma((\lambda+1)n+\alpha)}{\Gamma((\lambda+1)n)}.$$

Identity (2.2.17) follows by applying [89, Equation 1] to the above expression. Analogously,

$$\mathbb{E}\left[Z_n^2\right] = \mathbb{E}\left[S_{\lambda n,\alpha}^2\right]\mathbb{E}\left[G_{(\lambda+1)n,1}^{2\alpha}\right] = \frac{\lambda n}{\alpha}\left(\frac{\lambda n}{\alpha}+1\right)\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+2\alpha)}\frac{\Gamma((\lambda+1)n+2\alpha)}{\Gamma((\lambda+1)n)}$$

which, upon rearrangement of the terms, entails

$$\begin{aligned}\mathrm{Var}[Z_n] &= \mathbb{E}\left[Z_n^2\right] - \left(\mathbb{E}[Z_n]\right)^2 \\ &= \frac{\lambda^2 n^2}{\alpha^2}\left[\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+2\alpha)}\frac{\Gamma((\lambda+1)n+2\alpha)}{\Gamma((\lambda+1)n)} - \frac{\Gamma(\lambda n)^2}{\Gamma(\lambda n+\alpha)^2}\frac{\Gamma((\lambda+1)n+\alpha)^2}{\Gamma((\lambda+1)n)^2}\right] \\ &\quad + \frac{\lambda n}{\alpha}\left[\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+2\alpha)}\frac{\Gamma((\lambda+1)n+2\alpha)}{\Gamma((\lambda+1)n)}\right].\end{aligned}$$

Again [89, Equation 1], simplified for the purpose of an approximation to the principal order, yields

$$\begin{aligned}\mathrm{Var}[Z_n] = \frac{\lambda^2 n^2}{\alpha^2}&\left[\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\left(1-\frac{\alpha(2\alpha-1)}{\lambda(\lambda+1)n}+O\left(\frac{1}{n^2}\right)\right)\right. \\ &\left.-\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\left(1-\frac{\alpha(\alpha-1)}{\lambda(\lambda+1)n}+O\left(\frac{1}{n^2}\right)\right)\right] \\ &+\frac{\lambda n}{\alpha}\left[\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}+O\left(\frac{1}{n}\right)\right].\end{aligned}$$

A direct computation shows that the coefficient of n^2 vanishes, while the coefficient of n rewrites as Σ^2, proving identity (2.2.18).

A.1.2 Proof of the LLN (2.2.19)

Fix $\varepsilon > 0$ and combine Chebyshev's inequality with the (variance) asymptotic expansion (2.2.18) to get

$$\begin{aligned} P\left[\left|\frac{Z_n - \mathbb{E}[Z_n]}{n}\right| > \varepsilon\right] &= P\left[|Z_n - \mathbb{E}[Z_n]| > n\varepsilon\right] \\ &\leq \frac{\mathrm{Var}(Z_n)}{n^2\varepsilon^2} \\ &= O\left(\frac{1}{n}\right) \to 0 \end{aligned}$$

as $n \to +\infty$. Since (2.2.17) implies that $n^{-1}\mathbb{E}[Z_n] \to z_0$ as $n \to +\infty$, the proof is concluded.

A.1.3 Proof of the CLT (2.2.20)

The proof of the CLT (2.2.20) consists of three main steps, which are outlined hereafter for clarity:

Step 1. In Lemma A.2, we provide an approximation of the distribution of $S_{\lambda n,\alpha}$ in terms of the distribution of the $(1-\alpha)$-th power of a Gamma random variable. More precisely, denote by d_{TV} the total variation distance between distributions on $(0,+\infty)$, namely

$$d_{\mathrm{TV}}(\mu,\nu) := \sup_{A \in \mathscr{B}((0,+\infty))} |\mu(A) - \nu(A)|.$$

We will show that

$$d_{\mathrm{TV}}\left(\mu_{S_{\lambda n,\alpha}}, \mu_{G^{1-\alpha}_{\rho n+\tau,B}}\right) = O\left(\frac{1}{n}\right),$$

where $G_{\rho n+\tau,B}$ has Gamma $(\rho n+\tau, B)$ distribution, for a suitable choice of parameters ρ, τ and B.

- *Step 2*. In Lemma A.3, we will establish a general CLT for sum of variables of the form

$$\left(\sum_{i=0}^{n} X_i\right)^{1-\alpha} \left(\sum_{i=1}^{n} Y_i\right)^{\alpha},$$

where $\{X_i\}$ and $\{Y_i\}$ are two independent i.i.d. sequences of positive random variables. As a consequence (Corollary A.4), we derive a CLT for products of powers of independent Gamma random variables, which we apply to the auxiliary variable $\hat{Z}_n := G^{1-\alpha}_{\rho n+\tau,B} G^{\alpha}_{(\lambda+1)n,1}$.

- *Step 3*. As a consequence of Lemma A.2, and of a further general result (Lemma A.5), we will show that it is possible to replace the approximating Gamma random variable with the random variable $S_{\lambda n,\alpha}$ in the CLT for $\hat{Z}_n$, obtaining the desired limit theorem for Z_n.

A.1.3.1 Step 1

Lemma A.1. *Let $E_\alpha : (0,+\infty) \to (0,+\infty)$ be defined by*

$$E_\alpha(y) := \frac{1}{\sqrt{2\pi\alpha(1-\alpha)}} \left(\frac{\alpha}{y}\right)^{\frac{2-\alpha}{2(1-\alpha)}} \exp\left\{-(1-\alpha)\left(\frac{\alpha}{y}\right)^{\frac{\alpha}{1-\alpha}}\right\}.$$

Then, for any $y \in (0,+\infty)$, there hold

$$\left| \frac{f_\alpha(y)}{E_\alpha(y)} - 1 \right| \leq C_\alpha y^{\frac{\alpha}{1-\alpha}} \tag{A.1}$$

$$\left| \frac{f_\alpha(y)}{E_\alpha(y)} - 1 - Q_\alpha y^{\frac{\alpha}{1-\alpha}} \right| \leq K_\alpha y^{\frac{2\alpha}{1-\alpha}} \tag{A.2}$$

for suitable positive constants C_α, Q_α, K_α depending only on α.

Proof. [97, Theorem 2.5.2] shows that, as $y \to 0$

$$\frac{f_\alpha(y)}{E_\alpha(y)} = 1 + O\left(y^{\frac{\alpha}{1-\alpha}}\right)$$

Furthermore, it is known that, as $y \to +\infty$,

$$f_\alpha(y) = O\left(y^{-(1+\alpha)}\right) \text{ and } E_\alpha(y) = O\left(y^{-\frac{2-\alpha}{2(1-\alpha)}}\right).$$

Whence, as $y \to +\infty$

$$\frac{f_\alpha(y)}{E_\alpha(y)} = O\left(y^{\frac{\alpha(2\alpha-1)}{1-\alpha}}\right).$$

Since $2\alpha - 1 < 1$ for every $\alpha \in (0,1)$, (A.1) is proven. To prove (A.2), use again [97, Theorem 2.5.2] to show that, as $y \to 0$

$$\frac{f_\alpha(y)}{E_\alpha(y)} = 1 + Q_\alpha y^{\frac{\alpha}{1-\alpha}} + O\left(y^{\frac{2\alpha}{1-\alpha}}\right)$$

holds with a suitable constant Q_α. The conclusion is obtained by considering the previous asymptotic behaviours as $y \to +\infty$. ∎

Lemma A.2. *Set $\rho = \lambda\frac{1-\alpha}{\alpha}$, $\tau = \frac{1}{2}$, $B = (1-\alpha)\,\alpha^{\frac{\alpha}{1-\alpha}}$. Then, letting $\mu_{S_{\lambda n,\alpha}}$ and $\mu_{G^{1-\alpha}_{\rho n+\tau,B}}$ stand for the distributions of $S_{\lambda n,\alpha}$ and $G^{1-\alpha}_{\rho n+\tau,B}$, respectively, as $n \to +\infty$ there holds*

$$d_{TV}\left(\mu_{S_{\lambda n,\alpha}}, \mu_{G^{1-\alpha}_{\rho n+\tau,B}}\right) = O\left(\frac{1}{n}\right).$$

Proof. Since $\mu_{S_{\lambda n,\alpha}}$ and $\mu_{G^{1-\alpha}_{\rho n+\tau,B}}$ have densities

$$f_{S_{\lambda n,\alpha}}(x) = \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} x^{\frac{\lambda n-1}{\alpha}-1} f_\alpha\left(x^{-\frac{1}{\alpha}}\right) \mathbb{1}_{(0,+\infty)}(x)$$

$$f_{G^{1-\alpha}_{\rho n+\tau,B}}(x) = \frac{\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}\right]^{\frac{1-\alpha}{\alpha}\lambda n+\frac{1}{2}}}{(1-\alpha)\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n+\frac{1}{2}\right)} \exp\left\{-(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}\right\} x^{\frac{\lambda n}{\alpha}+\frac{2\alpha-1}{2(1-\alpha)}} \mathbb{1}_{(0,+\infty)}(x)$$

respectively, the total variation distance at issue can be handled as follows:

$$\begin{aligned}
d_{\mathrm{TV}}\left(\mu_{S_{\lambda n,\alpha}}, \mu_{G_{\rho n+\tau,B}}\right) &= \frac{1}{2}\int_0^{+\infty} \left| f_{S_{\lambda n,\alpha}}(x) - f_{G^{1-\alpha}_{\rho n+\tau,B}}(x) \right| \mathrm{d}x \\
\text{(bound (A.1))} &\leq \frac{1}{2}\int_0^{+\infty} \left| \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} x^{\frac{\lambda n-1}{\alpha}-1} E_\alpha\left(x^{-\frac{1}{\alpha}}\right) - f_{G^{1-\alpha}_{\rho n+\tau,B}}(x) \right| \mathrm{d}x \\
&\quad + \frac{1}{2} C_\alpha \int_0^{+\infty} \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} x^{\frac{\lambda n-1}{\alpha}-1} E_\alpha\left(x^{-\frac{1}{\alpha}}\right) x^{-\frac{1}{1-\alpha}} \mathrm{d}x \\
&=: \frac{1}{2}\mathcal{J}_1^{(n)} + \frac{C_\alpha}{2}\mathcal{J}_2^{(n)}.
\end{aligned}$$

Thus, the proof is concluded if we show that

$$\mathcal{J}_1^{(n)} = O\left(\frac{1}{n}\right) \qquad \text{and} \qquad \mathcal{J}_2^{(n)} = O\left(\frac{1}{n}\right).$$

Simple algebraic manipulations yield

$$\begin{aligned}
\mathcal{J}_1^{(n)} &= \left| \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \frac{\alpha^{\frac{2-\alpha}{2(1-\alpha)}}}{\sqrt{2\pi\alpha(1-\alpha)}} - \frac{B^{\frac{1-\alpha}{\alpha}\lambda n + \frac{1}{2}}}{(1-\alpha)\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n + \frac{1}{2}\right)} \right| \times \\
&\quad \times \int_0^{+\infty} \exp\left\{-Bx^{\frac{1}{1-\alpha}}\right\} x^{\frac{\lambda n}{\alpha} + \frac{2\alpha-1}{2(1-\alpha)}} \, \mathrm{d}x \\
&= \left| \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \frac{\alpha^{\frac{2-\alpha}{2(1-\alpha)}}}{\sqrt{2\pi\alpha(1-\alpha)}} - \frac{B^{\frac{1-\alpha}{\alpha}+\frac{1}{2}}}{(1-\alpha)\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n + \frac{1}{2}\right)} \right| \times \\
&\quad \times \frac{(1-\alpha)\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n + \frac{1}{2}\right)}{B^{\frac{1-\alpha}{\alpha}+\frac{1}{2}}} \\
&= \left| \frac{\Gamma(\lambda n)\,\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n + \frac{1}{2}\right)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \left(\frac{1}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha}\lambda n} \left(\frac{1}{\alpha}\right)^{\lambda n - \frac{1}{2}} (2\pi)^{-\frac{1}{2}} - 1 \right|.
\end{aligned}$$

Moreover, for some suitable constants c_α, $\tilde{C}_\alpha$ depending only on α,

$$\begin{aligned}
\mathcal{J}_2^{(n)} &= \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \int_0^{+\infty} c_\alpha \exp\left\{-Bx^{\frac{1}{1-\alpha}}\right\} x^{\frac{\lambda n}{\alpha} + \frac{2\alpha-3}{2(1-\alpha)}} \, \mathrm{d}x \\
&= \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \frac{c_\alpha(1-\alpha)\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n - \frac{1}{2}\right)}{B^{\frac{1-\alpha}{\alpha}\lambda n - \frac{1}{2}}} \\
&= \tilde{C}_\alpha \frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)} \frac{\Gamma\left(\frac{1-\alpha}{\alpha}\lambda n - \frac{1}{2}\right)}{\left[(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}\right]^{\frac{1-\alpha}{\alpha}\lambda n}}.
\end{aligned}$$

Set $\lambda n =: t$. Then, we study the asymptotic behavior of

$$g_j(t) := \frac{\Gamma(t)\,\Gamma\left(\frac{1-\alpha}{\alpha}t + \frac{j}{2}\right)}{\Gamma\left(\frac{t}{\alpha}\right)}$$

as $t \to +\infty$, for $j \in \{-1, 1\}$. By combining Stirling's approximation to the first order, i.e.

$$\Gamma(t) = \sqrt{2\pi(t-1)} \left(\frac{t-1}{e}\right)^{t-1} \left[1 + O\left(t^{-1}\right)\right] \quad (t \to +\infty),$$

with the fact that

$$\left(1 + \frac{a}{t}\right)^t = e^a \left[1 + O\left(t^{-1}\right)\right] \quad (t \to +\infty),$$

we get, after straightforward computations,

$$\begin{aligned}
g_j(t) &= \begin{cases} (2\pi e)^{\frac{1}{2}} \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}t} \alpha^{\frac{t}{\alpha}-\frac{1}{2}} \left[1 - \frac{1}{2t} + O\left(\frac{1}{t^2}\right)\right]^t & \text{if } j = 1 \\ c_\alpha^\star (2\pi)^{\frac{1}{2}} e^{\frac{3}{2}} \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}t} \alpha^{\frac{t}{\alpha}-\frac{1}{2}} \frac{1}{t} \left[1 - \frac{3}{2t} + O\left(\frac{1}{t^2}\right)\right]^t & \text{if } j = -1 \end{cases} \\
&= \begin{cases} (2\pi)^{\frac{1}{2}} \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1-\alpha}{\alpha}t} \alpha^{\frac{t}{\alpha}-\frac{1}{2}} \left[1 + O\left(\frac{1}{t}\right)\right] & \text{if } j = 1 \\ O\left(\frac{1}{t}\right) & \text{if } j = -1 \end{cases}
\end{aligned}$$

where $c_\alpha^\star$ is another constant depending only on α. Substituting the first and second expression, respectively, in the expressions of $\mathcal{J}_1^{(n)}$ and $\mathcal{J}_2^{(n)}$, we obtain the desired results. ■

A.1.3.2 Step 2

Lemma A.3. *Let X_0 be a real random variable and $\{X_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ be two sequences of i.i.d. positive random variables. We further assume:*

a) $\mathbb{E}\left[X_1^4\right] < +\infty$ *and* $\mathbb{E}\left[Y_1^4\right] < +\infty$;

b) $\mu_X := \mathbb{E}[X_1]$, $\mu_Y := \mathbb{E}[Y_1]$, $\sigma_X^2 := \mathrm{Var}(X_1)$ *and* $\sigma_Y^2 := \mathrm{Var}(Y_1)$;

c) X_0, $\{X_i\}_{i\geq 1}$ *and* $\{Y_i\}_{i\geq 1}$ *are independent.*

Then, the following CLT holds:

$$\frac{1}{\sqrt{n}}\left[\left(\sum_{i=0}^{n} X_i\right)^{1-\alpha}\left(\sum_{i=1}^{n} Y_i\right)^{\alpha} - nm\right] \xrightarrow{w} \mathcal{N}\left(0, V^2\right)$$

as $n \to +\infty$, where

$$m := \mu_X^{1-\alpha}\mu_Y^{\alpha}$$

and

$$V^2 := \mu_X^{2(1-\alpha)}\mu_Y^{2\alpha}\left[\frac{(1-\alpha)^2\sigma_X^2}{\mu_X^2} + \frac{\alpha^2\sigma_Y^2}{\mu_Y^2}\right].$$

Proof. Start considering these trivial identities:

$$\frac{1}{n}\sum_{i=0}^{n} X_i = \frac{X_0}{n} + \mu_X + \frac{\sigma_X}{n}\sum_{i=1}^{n}\left(\frac{X_i - \mu_X}{\sigma_X}\right)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu_Y + \frac{\sigma_Y}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right).$$

From [86, Theorem 1.5], there exist two independent Brownian motions $\{B_t\}_{t\geq 0}$ and $\{B'_t\}_{t\geq 0}$ such that

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu_X}{\sigma_X}\right) \overset{a.s.}{=} B_n + R_n\,, \qquad R_n := \sum_{i=1}^{n}\left(\frac{X_i - \mu_X}{\sigma_X}\right) - B_n$$

$$\sum_{i=1}^{n}\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right) \overset{a.s.}{=} B'_n + R'_n\,, \qquad R'_n := \sum_{i=1}^{n}\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right) - B'_n$$

with

$$R_n = O\left((n\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}\right) = o\left(n^{\frac{1}{4}+\varepsilon}\right) \quad \forall \varepsilon > 0$$

$$R'_n = O\left((n\log\log n)^{\frac{1}{4}}(\log n)^{\frac{1}{2}}\right) = o\left(n^{\frac{1}{4}+\varepsilon}\right) \quad \forall \varepsilon > 0.$$

Recalling that, from the law of iterated logarithm

$$\frac{|B_n|}{n} = O\left(\sqrt{\frac{\log\log n}{n}}\right) \qquad \text{and} \qquad \frac{|B'_n|}{n} = O\left(\sqrt{\frac{\log\log n}{n}}\right),$$

one gets

$$\left(\frac{1}{n}\sum_{i=0}^{n} X_i\right)^{1-\alpha} \overset{a.s.}{=} \mu_X^{1-\alpha}\left[1+\frac{X_0}{n\mu_X}+\frac{\sigma_X}{n\mu_X}B_n+\frac{\sigma_X}{n\mu_X}R_n\right]^{1-\alpha}$$
$$= \mu_X^{1-\alpha}\left[1+\frac{(1-\alpha)\,\sigma_X}{n\mu_X}B_n+o\left(n^{-\frac{3}{4}+\varepsilon}\right)\right]$$

and

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^{\alpha} \overset{a.s.}{=} \mu_Y^{\alpha}\left[1+\frac{\sigma_Y}{n\mu_Y}B'_n+\frac{\sigma_Y}{n\mu_Y}R'_n\right]^{\alpha}$$
$$= \mu_Y^{\alpha}\left[1+\frac{\alpha\,\sigma_Y}{n\mu_Y}B'_n+o\left(n^{-\frac{3}{4}+\varepsilon}\right)\right]$$

for any $\varepsilon > 0$. This implies

$$\sqrt{n}\left[\left(\frac{1}{n}\sum_{i=0}^{n} X_i\right)^{1-\alpha}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^{\alpha}-\mu_X^{1-\alpha}\mu_Y^{\alpha}\right]$$
$$\overset{a.s.}{=} \mu_X^{1-\alpha}\mu_Y^{\alpha}\left[\frac{(1-\alpha)\,\sigma_X}{\mu_X}\frac{B_n}{\sqrt{n}}+\frac{\alpha\,\sigma_Y}{\mu_Y}\frac{B'_n}{\sqrt{n}}+o\left(n^{-\frac{1}{2}+\varepsilon}\right)\right].$$

Observe that, by elementary properties of the two independent Brownian motions $\{B_t\}_{t\geq 0}$ and $\{B'_t\}_{t\geq 0}$,

$$\mu_X^{1-\alpha}\mu_Y^{\alpha}\left[\frac{(1-\alpha)\,\sigma_X}{\mu_X}\frac{B_n}{\sqrt{n}}+\frac{\alpha\,\sigma_Y}{\mu_Y}\frac{B'_n}{\sqrt{n}}\right] \overset{d}{=} \mathcal{N}\left(0,V^2\right).$$

Finally, apply Slutsky's theorem to conclude the proof. ■

Corollary A.4. *Let $G_{\rho n+\tau,B} \sim \text{Gamma}(\rho n+\tau, B)$ and $G_{(\lambda+1)n,1} \sim \text{Gamma}((\lambda+1)n,1)$ be two independent random variables. Then, as $n \to +\infty$,*

$$\frac{1}{\sqrt{n}}\left[G_{\rho n+\tau,B}^{1-\alpha}G_{(\lambda+1)n,1}^{\alpha}-nm\right] \xrightarrow{w} \mathcal{N}(0,V^2),$$

where

$$m := \left(\frac{\rho}{B}\right)^{1-\alpha}(\lambda+1)^{\alpha}$$

and

$$V^2 := \left(\frac{\rho}{B}\right)^{2(1-\alpha)}(\lambda+1)^{2\alpha}\left[\frac{(1-\alpha)^2}{\rho}+\frac{\alpha^2}{\lambda+1}\right].$$

Proof. Introduce a random variable X_0 and two sequences of i.i.d. random variables $\{X_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ such that:

i) $X_1 \sim \text{Gamma}(\rho, B)$, $Y_1 \sim \text{Gamma}(\lambda+1,1)$ and $X_0 \sim \text{Gamma}(\tau, B)$;

ii) X_0, $\{X_i\}_{i\geq 1}$ and $\{Y_i\}_{i\geq 1}$ are independent.

Since $G_{\rho n+\tau,B} \overset{d}{=} \sum_{i=0}^{n} X_i$ and $G_{(\lambda+1)n,1} \overset{d}{=} \sum_{i=1}^{n} Y_i$, the statement immediately follows from Lemma A.3. The expressions of m and V^2 can be checked via direct computation. ■

A.1.3.3 Step 3

The next statement makes use of the so-called Ky-Fan distance between random variables, namely

$$d_{\mathrm{KF}}(X,Y) := \inf\{\varepsilon > 0 : P(|X-Y| > \varepsilon) \le \varepsilon\}.$$

Lemma A.5. *On a probability space $(\Omega, \mathcal{F}, P)$, consider a real random variable η, and two sequences of real random variables $\{\eta_n\}_{n\ge1}$, $\{(\xi_n, \xi'_n)\}_{n\ge1}$ such that:*

i) $\eta_n \xrightarrow{w} \eta$, *as* $n \to +\infty$;

ii)

$$d_{KF}(\xi_n, \xi'_n) = o\left(\frac{1}{\sqrt{n}}\right); \tag{A.3}$$

iii) there exist $c \in \mathbb{R}$ *and a real random variable* L *such that, as* $n \to +\infty$

$$\sqrt{n}\left(\xi'_n\eta_n - c\right) \xrightarrow{w} L. \tag{A.4}$$

Then, as $n \to +\infty$,

$$\sqrt{n}\left(\xi_n\eta_n - c\right) \xrightarrow{w} L,$$

with the same c *and* L *as in* (A.4).

Proof. Write

$$\sqrt{n}\left(\xi_n\eta_n - c\right) = \sqrt{n}\left(\xi'_n\eta_n - c\right) + \sqrt{n}\eta_n\left(\xi_n - \xi'_n\right).$$

By means of Slutsky's theorem, the proof is concluded if we show

$$\sqrt{n}\left(\xi_n - \xi'_n\right) \xrightarrow{p} 0$$

as $n \to +\infty$. Since the Ky-Fan distance metrizes convergence in probability, it is enough to note that

$$d_{\mathrm{KF}}\left(\sqrt{n}\left(\xi_n - \xi'_n\right), \mathbf{0}\right) = d_{\mathrm{KF}}\left(\sqrt{n}\xi_n, \sqrt{n}\xi'_n\right) \le \sqrt{n} d_{\mathrm{KF}}\left(\xi_n, \xi'_n\right) \xrightarrow{(A.3)} 0$$

where $\mathbf{0}$ denotes the degenerate random variable equal to 0 a.s.. ■

Proof of the CLT (2.2.20). Set $\rho = \lambda\frac{1-\alpha}{\alpha}$, $\tau = \frac{1}{2}$, $B = (1-\alpha)\,\alpha^{\frac{\alpha}{1-\alpha}}$ as in Lemma A.2 and introduce an auxiliary random variable $G_{\rho n+\tau,B} \sim \mathrm{Gamma}(\rho n + \tau, B)$, independent of $G_{(\lambda+1)n,1}$. Then, the assumptions of Corollary A.4 are met, with m and V^2 as follows:

$$m = \left(\frac{\rho}{B}\right)^{1-\alpha}(\lambda+1)^{\alpha} = \frac{\lambda}{\alpha}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha} = z_0$$

and

$$\begin{aligned} V^2 &= \left(\frac{\rho}{B}\right)^{2(1-\alpha)}(\lambda+1)^{2\alpha}\left[\frac{(1-\alpha)^2}{\rho} + \frac{\alpha^2}{\lambda+1}\right] \\ &= \frac{\lambda^2}{\alpha^2}\left(\frac{\lambda+1}{\lambda}\right)^{2\alpha}\left[\frac{(1-\alpha)\alpha}{\lambda} + \frac{\alpha^2}{\lambda+1}\right] = \Sigma^2. \end{aligned}$$

Corollary A.4 then entails that, as $n \to +\infty$

$$\frac{1}{\sqrt{n}}\left[G^{1-\alpha}_{\rho n+\tau,B}G^{\alpha}_{(\lambda+1)n,1} - nz_0\right] \xrightarrow{w} \mathcal{N}(0,\Sigma^2).$$

Now, let d_{Prok} denote the Prokhorov distance between distributions, defined as follows. Given $B \subseteq \mathbb{R}$ and $\varepsilon > 0$, let $B^\varepsilon := \{x \in \mathbb{R} : \inf_{y \in B} |x-y|^2 \leq \varepsilon\}$; then

$$d_{\text{Prok}}(\mu,\nu) := \inf\{\varepsilon > 0 : \mu(B) \leq \nu(B^\varepsilon) + \varepsilon, \forall B \in \mathscr{B}(\mathbb{R})\}.$$

Thanks to a well-known bound between probability metrics [58, Equation 4.13], Lemma A.2 implies that

$$d_{\text{Prok}}\left(\mu_{S_{\lambda n,\alpha}}, \mu_{G^{1-\alpha}_{\rho n+\tau,B}}\right) \leq d_{\text{TV}}\left(\mu_{S_{\lambda n,\alpha}}, \mu_{G^{1-\alpha}_{\rho n+\tau,B}}\right) = O\left(\frac{1}{n}\right).$$

Now, a theorem of Strassen [31, Corollary 11.6.4] allows, in turn, to find a suitable coupling between the random variables $S_{\lambda n,\alpha}$ and $G^{1-\alpha}_{\rho n+\tau,B}$ such that

$$d_{\text{KF}}\left(S_{\lambda n,\alpha}, G^{1-\alpha}_{\rho n+\tau,B}\right) = O\left(\frac{1}{n}\right).$$

To conclude, it is enough to invoke Lemma A.5 with the choices $\eta_n = G^{\alpha}_{(\lambda+1)n,1}$, $\xi'_n = G^{1-\alpha}_{\rho n+\tau,B}$, $\xi_n = S_{\lambda n,\alpha}$, $c = z_0$ and $L = \mathcal{N}(0,\Sigma^2)$. ■

A.1.4 On a property of uniform integrability of Z_n

We now prove a property of uniform integrability of Z_n which, in combination with convergence in distribution, guarantees convergence of the moments. Hence, from the CLT we will deduce convergence of the moments of $(Z_n - n\,z_0)/(\sqrt{n\Sigma^2})$ to those of the standard Gaussian.

Lemma A.6. *Under the assumption of Proposition 1, it holds*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left[\left(\frac{Z_n - nz_0}{\sqrt{n}}\right)^4\right] < +\infty.$$

Proof. We show that as $n \to +\infty$

$$\mathbb{E}\left[(Z_n - nz_0)^4\right] = O\left(n^2\right).$$

In fact,

$$\begin{aligned}
\mathbb{E}\left[(Z_n - nz_0)^4\right] &= \sum_{k=0}^{4}\binom{4}{k}\mathbb{E}[Z_n^k]\,(nz_0)^{4-k} \\
&= \sum_{k=0}^{4}\binom{4}{k}\frac{\Gamma(\lambda n)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)}\frac{\Gamma\left(k+\frac{\lambda n}{\alpha}\right)}{\Gamma(k\alpha+\lambda n)}\frac{\Gamma\left((\lambda+1)n+k\alpha\right)}{\Gamma\left((\lambda+1)n\right)}\times \\
&\quad \times (-1)^{4-k}\,n^{4-k}\left(\frac{\lambda n}{\alpha}\right)^{4-k}\left(\frac{\lambda+1}{\lambda}\right)^{4-k}.
\end{aligned}$$

First-order Stirling's approximation, combined with some simple computations, yields

$$\begin{aligned}
\mathbb{E}\left[(Z_n - nz_0)^4\right] &= n^4\left(\frac{\lambda}{\alpha}\right)^4\left(\frac{\lambda+1}{\lambda}\right)^4\sum_{k=0}^{4}\binom{4}{k}(-1)^{4-k}\times \\
&\quad \times\left[1 - \frac{k\alpha(k\alpha-1)}{2\lambda n} + \frac{k(k-1)\alpha}{2\lambda n} + \frac{k\alpha(k\alpha-1)}{2(\lambda+1)n} + O\left(\frac{1}{n^2}\right)\right] \\
&= c_0\,n^4\sum_{k=0}^{4}\binom{4}{k}(-1)^{4-k} + n^3\sum_{k=0}^{4}\binom{4}{k}(-1)^{4-k}\left(c_1\,k + c_2\,k^2\right) \\
&\quad + O(n^2)
\end{aligned}$$

for some constants c_0, c_1 , c_2 depending only on α and λ. Now, it can be easily seen that

$$\sum_{k=0}^{4} \binom{4}{k} (-1)^{4-k} k^i = 0$$

for $i = 0, 1, 2$, which concludes the proof. ■

By combining Lemma A.6 and Proposition 1 we obtain the following statement.

Proposition 10. *Let $\Psi : (0, +\infty) \to \mathbb{R}$ be such that $\Psi \in C^3(0, +\infty)$ with bounded derivatives. Then*

$$\mathbb{E}\left[\Psi\left(\frac{Z_n}{n}\right)\right] = \Psi(z_0) + \frac{1}{2n}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha} \frac{1-\alpha}{\lambda+1}\Psi'(z_0) + \frac{1}{2n}\Sigma^2\,\Psi''(z_0) + o\left(\frac{1}{n}\right).$$

Proof. Let $e_n := \mathbb{E}\left[\Psi\left(\frac{Z_n}{n}\right)\right]$. By means of [89, Equation 1], we write

$$\begin{aligned} e_n &= \frac{1}{n}\left[n\,\frac{\lambda}{\alpha}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\left(1 + \frac{\alpha(1-\alpha)}{2n\lambda(\lambda+1)}\right)\right] + o\left(\frac{1}{n}\right) \\ &= z_0 + \frac{1}{2n}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\frac{1-\alpha}{\lambda+1} + o\left(\frac{1}{n}\right). \end{aligned}$$

Taylor's formula with Bernstein's integral remainder [31, Appendix B] shows that

$$\begin{aligned} \Psi\left(\frac{Z_n}{n}\right) - \Psi(e_n) = \Psi'(e_n)\left(\frac{Z_n}{n} - e_n\right) + \frac{1}{2}\Psi''(e_n)\left(\frac{Z_n}{n} - e_n\right)^2 \\ + \frac{1}{2}\left(\frac{Z_n}{n} - e_n\right)^3 \int_0^1 \Psi'''\left(e_n + s\left(\frac{Z_n}{n} - e_n\right)\right)(1-s)^2\,\mathrm{d}s. \end{aligned}$$

Taking expectation on both sides we obtain

$$\begin{aligned} \mathbb{E}\left[\Psi\left(\frac{Z_n}{n}\right)\right] = \Psi(e_n) + \frac{1}{2}\Psi''(e_n)\,\mathrm{Var}\left(\frac{Z_n}{n}\right) \\ + \frac{1}{2}\mathbb{E}\left[\left(\frac{Z_n}{n} - e_n\right)^3 \int_0^1 \Psi'''\left(e_n + s\left(\frac{Z_n}{n} - e_n\right)\right)(1-s)^2\,\mathrm{d}s\right]. \end{aligned}$$

By applying again Taylor's formula to Ψ' and Ψ'', the last term becomes

$$\begin{aligned} \mathbb{E}\left[\Psi\left(\frac{Z_n}{n}\right)\right] = \Psi(z_0) + \frac{1}{2n}\left(\frac{\lambda+1}{\lambda}\right)^{\alpha}\frac{1-\alpha}{\lambda+1}\,\Psi'(z_0) + \frac{1}{2}\,\Psi''(z_0)\,\frac{n\Sigma^2 + O(1)}{n^2} \\ + \frac{1}{2}\mathbb{E}\left[\left(\frac{Z_n}{n} - e_n\right)^3 \int_0^1 \Psi'''\left(e_n + s\left(\frac{Z_n}{n} - e_n\right)\right)(1-s)^2\,\mathrm{d}s\right]. \end{aligned}$$

Since Ψ''' is bounded, to reach the conclusion it is enough to prove that

$$\lim_{n\to+\infty} n\,\mathbb{E}\left[\left|\frac{Z_n}{n} - e_n\right|^3\right] = \lim_{n\to+\infty} n\,\mathbb{E}\left[\left|\frac{Z_n}{n} - z_0\right|^3\right] = 0.$$

Now, by combining Lemma A.6 with Proposition 1, we get

$$\lim_{n\to+\infty} \mathbb{E}\left[\left|\frac{Z_n - nz_0}{\sqrt{n}}\right|^3\right] = \int_{\mathbb{R}} |x|^3\,\mathrm{d}\Phi(x)$$

in view of the uniform integrability. Since $n\left|\frac{Z_n}{n}-z_0\right|^3 = \frac{1}{\sqrt{n}}\left|\frac{Z_n-nz_0}{\sqrt{n}}\right|^3$, we finally obtain that

$$\lim_{n\to+\infty}\mathbb{E}\left[\frac{1}{\sqrt{n}}\left|\frac{Z_n-nz_0}{\sqrt{n}}\right|^3\right] = \lim_{n\to+\infty}\frac{1}{\sqrt{n}}\times\lim_{n\to+\infty}\mathbb{E}\left[\left|\frac{Z_n-nz_0}{\sqrt{n}}\right|^3\right] = 0.$$

This concludes the proof. ■

A.2 Proof of Proposition 2

Denote by $G_{R_n(z)}$ the probability generating function of $R_n(z)$. The proof of Proposition 2 is based on the study of the large n asymptotic behavior of $G_{R_n(z)}$, as well as its derivatives. In particular, this large n asymptotic analysis is based on the following steps:

1. *Step 1*. State the preparatory results in Lemma A.7, Lemma A.8, Lemma A.9, and Lemma A.10.
2. *Step 2*. Show how the functions $\mu(z)$ and $\sigma(z)$ in Proposition 2 are related to the moments of $R_n(z)$. In particular, we prove here the asymptotic expansions (2.2.22) and (2.2.23)
3. *Step 3*. Prove a Berry-Esseen inequality for $R_n(z)$;
4. *Step 4*. Conclude the proof of the Berry-Esseen inequality (2.2.25).

A.2.0.1 Step 1

The first lemma allows to write a quantity of interest in terms of the so-called Krätzel function (see [74, Equation 1.1]), namely

$$F_{p,\nu}(w) := \int_0^{+\infty} t^{\nu-1}\exp\left\{t^p-\frac{w}{t}\right\}\mathrm{d}t$$

for any $w\in\mathbb{C}^+ := \{w\in\mathbb{C} : Re(w)>0\}$. A standard check shows that $F_{p,\nu}$ is holomorphic on the whole of $\mathbb{C}^+$.

Lemma A.7. *For $\varepsilon\geq 0$, $\delta\geq 0$, $y\in\mathbb{C}^+$ define*

$$J^{(n)}_{\varepsilon,\delta}(y) := \int_0^{+\infty} x^{n+\varepsilon-\frac{2-\alpha}{2(1-\alpha)}+\frac{\delta\alpha}{1-\alpha}}\exp\left\{-x(ny)^{\frac{1}{\alpha}}-(1-\alpha)\left(\frac{\alpha}{x}\right)^{\frac{\alpha}{1-\alpha}}\right\}\mathrm{d}x. \tag{A.5}$$

Then, it holds that

$$J^{(n)}_{\varepsilon,\delta}(y) = \mathcal{C}(\alpha,n,y)\left(\frac{1}{ny}\right)^{\frac{\varepsilon}{\alpha}+\frac{\delta}{1-\alpha}}\mathcal{F}^{(n)}_{\varrho(\varepsilon,\delta)}(y),$$

where:

$$\mathcal{C}(\alpha,n,y) := \frac{1-\alpha}{\alpha}\left(\frac{1}{ny}\right)^{\frac{n}{\alpha}-\frac{2-\alpha}{2\alpha(1-\alpha)}+\frac{1}{\alpha}} \tag{A.6}$$

$$\varrho(\varepsilon,\delta) := \frac{1}{2}-\delta-\varepsilon\frac{1-\alpha}{\alpha} \tag{A.7}$$

$$\mathcal{F}^{(n)}_{\varrho}(y) := F_{\frac{1-\alpha}{\alpha},n\frac{1-\alpha}{\alpha}-\varrho}\left((1-\alpha)(\alpha^{\alpha}ny)^{\frac{1}{1-\alpha}}\right).$$

Proof. First, fix $y \in (0,+\infty)$ and prove the result by direct computation, using the substitution $t = \left[x(ny)^{\frac{1}{\alpha}}\right]^{\frac{\alpha}{1-\alpha}}$ in the integral on the right-hand side of (A.5). Notice that both sides of the identity at issue are analytic function of y, as y varies in $\mathbb{C}^+$. To be precise, we are considering the principal branch of each power of y, which is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$. Finally, since both sides of (A.5) coincide on $(0,+\infty)$, then they must coincide on the whole of $\mathbb{C}^+$ by uniqueness of the analytic continuation. ■

The next lemma is concerned with equation (2.2.21). To simplify its statement, we rewrite (2.2.21) in a simpler form. For fixed $v \in (0,+\infty)$, let $\xi(v)$ denote the only real, positive solution to the equation

$$\xi(v)^{\frac{1}{\alpha}} = v\xi(v) + 1. \tag{A.8}$$

Indeed, the regularity of the mapping $v \mapsto \xi(v)$, contained in the next statement, is an important issue in what follows.

Lemma A.8. *The mapping $v \mapsto \xi(v)$ admits an analytical continuation to the whole of $\mathbb{C}^+$. In particular, $v \mapsto \xi(v)$ is analytic on $(0,+\infty)$.*

Proof. First, the existence of the mapping $v \mapsto \xi(v)$ from $[0,+\infty)$ into $[1,+\infty)$ is checked by basic calculus tools. Then, a direct application of [6, Equation (11.5)]), with the same α as in (A.8), $\beta = \alpha$, $x = \xi(v)^{\frac{1}{\alpha}}$, and $y = v$, shows that the function

$$\xi_1(v) := \alpha \sum_{n=0}^{+\infty} \frac{\Gamma(\alpha n + \alpha)}{\Gamma\big((\alpha-1)n + \alpha + 1\big)} \frac{v^n}{n!}$$

provides a solution to (A.8). More precisely, the above series converges inside the disc $|v| < \alpha^{-\alpha}(1-\alpha)^{\alpha-1}$, and meets $\xi_1(v) = 1 + \alpha v + O(v^2)$ as $v \to 0$. Whence, $\xi(v) = \xi_1(v)$ for any $v \in [0, \alpha^{-\alpha}(1-\alpha)^{\alpha-1})$. To get closer to the goal, it is crucial to notice that ξ_1 is a special case of a so-called H-function (or Fox function), whose theory is exposed, e.g., in [62, Chapters 1–2]) and [73, Chapter 2]). Exploiting this, it is worth considering

$$\xi_2(v) := \frac{\alpha}{2\pi i} \int_{\gamma_2} \frac{\Gamma(\alpha s + \alpha)\Gamma(-s)}{\Gamma\big((\alpha-1)s + \alpha + 1\big)} (-v)^s \mathrm{d}s$$

where γ_2 is a path defined as follows: it stars at $-i\infty$; proceeds upward along the imaginary axis; before reaching the real axis, encircles the origin clock-wise, remaining on $\mathbb{C}^- := \{w \in \mathbb{C} : Re(w) < 0\}$ and leaving all the poles of the form $-1 - m\alpha^{-1}$, $m \in \mathbb{N}_0$, to its left; proceeds again along the imaginary axis towards $+i\infty$. See Fig 1 below.

At this stage, an application of [62, Theorem 1.1 and Equation 1.2.20] with $a^* = 1 + \alpha - (1-\alpha) = 2\alpha$ therein shows that ξ_2 is well-defined and holomorphic in the domain $\{v \in \mathbb{C} : |\mathrm{Arg}(v)| < \pi\alpha\}$, where Arg stands for the principal argument. Therefore, recalling that

$$Res\,(s \mapsto \Gamma(-s); n) = \frac{(-1)^{n+1}}{n!} \qquad (n \in \mathbb{N}_0),$$

the residue theorem entails that

$$\frac{1}{2\pi i} \oint_{\gamma_{2,N}} \frac{\Gamma(\alpha s + \alpha)\Gamma(-s)}{\Gamma\big((\alpha-1)s + \alpha + 1\big)} (-v)^s \mathrm{d}s = \sum_{n=0}^{N} \frac{\Gamma(\alpha n + \alpha)}{\Gamma\big((\alpha-1)n + \alpha + 1\big)} \frac{v^n}{n!}$$

holds for any $v \in \mathbb{C}$ and $N \in \mathbb{N}$, where $\gamma_{2,N}$ is the closed path that starts at $-i(N+1/2)$, proceeds exactly as γ_2 up to $+i(N+1/2)$, and comes back to $-i(N+1/2)$ along the

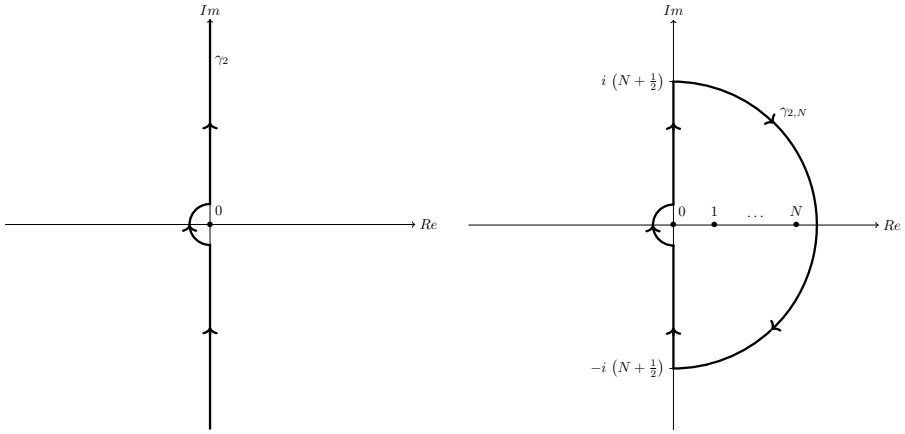

Figure 5.1: The paths γ_2 (left) and $\gamma_{2,N}$(right)

semicircle of radius $(N+1/2)$, centered at the origin, routed clock-wise. See Fig 1. Exploiting the estimates displayed in [62, Section 1.2], we let N go to infinity to conclude that $\xi_1(v) = \xi_2(v)$ on $\{w \in \mathbb{C} : |v| < \alpha^{-\alpha}(1-\alpha)^{\alpha-1}, |\mathrm{Arg}(v)| < \pi\alpha\}$. Whence, $\xi(v) = \xi_2(v)$ for any $v \in (0, \alpha^{-\alpha}(1-\alpha)^{\alpha-1})$. In order to include the half-line $(\alpha^{-\alpha}(1-\alpha)^{\alpha-1}, +\infty)$, we invoke again [6, Equation (11.5)]), with $-\alpha(1-\alpha)^{-1}$ in the place of α therein, $\beta = \alpha(1-\alpha)^{-1}$, $x = v^{-1}\xi(v)^{\frac{1-\alpha}{\alpha}}$, and $y = v^{\frac{1}{1-\alpha}}$. We get that the function

$$\xi_3(v) := \frac{\alpha}{1-\alpha} v^{\frac{\alpha}{1-\alpha}} \sum_{n=0}^{+\infty} \frac{\Gamma\left(\frac{n-\alpha}{1-\alpha}\right)}{\Gamma\left(\frac{\alpha n+1-2\alpha}{1-\alpha}\right)} \frac{(-1)^{n+1}}{n!} \left(v^{\frac{-1}{1-\alpha}}\right)^n$$

provides a solution to (A.8). More precisely, the above series converges if $|v| > \alpha^{-\alpha}(1-\alpha)^{\alpha-1}$, and represents a holomorphic function on the region $\{w \in \mathbb{C} : |v| > \alpha^{-\alpha}(1-\alpha)^{\alpha-1}, |\mathrm{Arg}(v)| < \pi\}$, provided that the powers of v are intended as their the principal branches. Since $\xi_3(v) = v^{\frac{\alpha}{1-\alpha}} + O(v^{-1})$ as $v \to +\infty$, we deduce that $\xi(v) = \xi_3(v)$ for any $v \in (\alpha^{-\alpha}(1-\alpha)^{\alpha-1}, +\infty)$. Now, we assume temporarily that α is not of the form $q(q+1)^{-1}$ for some $q \in \mathbb{N}$. This way, none of the values of the form $\alpha - m(1-\alpha)$, $m \in \mathbb{N}$, can coincide with the origin. Resorting once again to the theory of H-functions, we consider

$$\xi_4(v) := -\frac{\alpha}{1-\alpha} v^{\frac{\alpha}{1-\alpha}} \frac{1}{2\pi i} \int_{\gamma_4} \frac{\Gamma\left(\frac{s-\alpha}{1-\alpha}\right)\Gamma(-s)}{\Gamma\left(\frac{\alpha s+1-2\alpha}{1-\alpha}\right)} \left(v^{\frac{-1}{1-\alpha}}\right)^s \mathrm{d}s$$

where γ_4 is a path of the following form: it starts at $-i\infty$; proceeds upward along the imaginary axis before reaching the origin; turns right to cross upward the real axis between α and 1; avoids all the poles of the form $\alpha - m(1-\alpha)$, $m \in \mathbb{N}_0$, which lie at the right of the origin, by turning around them counter-clock-wise; crosses again the real axis downward between the origin itself and the least positive pole of the form $\alpha - m(1-\alpha)$; encircles the origin clock-wise, ranging in the fourth, third and second quadrant, respectively, by leaving all the negative poles of the form $\alpha - m(1-\alpha)$ to its left; proceeds again along the imaginary axis towards $+i\infty$. See the Fig 2 below.

Again, an application of [62, Theorem 1.1 and Equation 1.2.20] with $a^* = 1 - \frac{\alpha}{1-\alpha} + \frac{1}{1-\alpha} = 2$ therein shows that ξ_4 is well-defined and holomorphic in $\{v \in \mathbb{C} : |\mathrm{Arg}(v)| < \pi(1-\alpha)\}$. A similar argument, as the one above based of the residue theorem, leads us to conclude that $\xi_3(v) = \xi_4(v)$ on $\{w \in \mathbb{C} : |v| > \alpha^{-\alpha}(1-\alpha)^{\alpha-1}, |\mathrm{Arg}(v)| < \pi(1-\alpha)\}$.

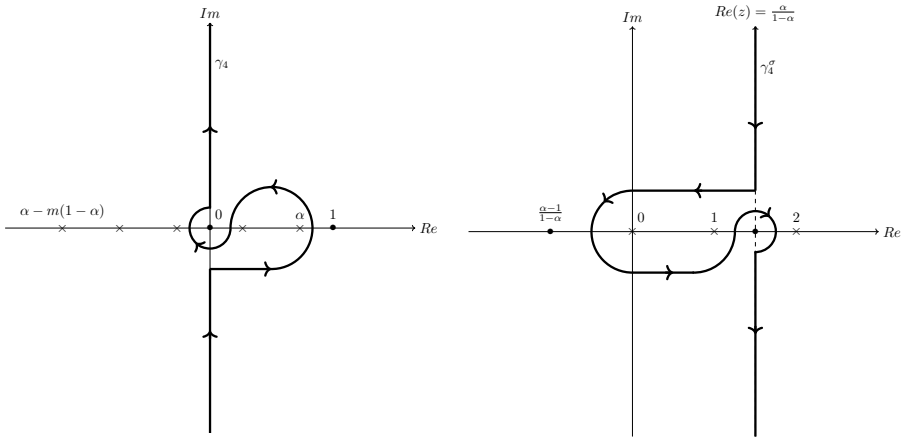

Figure 5.2: The path γ_4 (left) and its transformation γ_4^σ trough the map $s \mapsto \sigma(s) := \frac{\alpha - s}{1-\alpha}$ (right)

Whence, $\xi(v) = \xi_4(v)$ for any $v \in (\alpha^{-\alpha}(1-\alpha)^{\alpha-1}, +\infty)$. It remains to show that $\xi_2(v) = \xi_4(v)$ on $\{v \in \mathbb{C} \ : \ |\mathrm{Arg}(v)| < \pi\alpha_*\}$, where $\alpha_* := \min\{\alpha, 1-\alpha\}$. To this aim, we change the variable in the expression of ξ_4 by setting $\sigma = \frac{\alpha - s}{1-\alpha}$, to obtain

$$\xi_4(v) = \frac{\alpha}{2\pi i} \int_{\gamma_4^\sigma} \frac{\Gamma\big(1 - (\sigma(\alpha-1) + \alpha + 1)\big)\Gamma(-\sigma)}{\Gamma\big(1 - (\alpha + \alpha\sigma)\big)} v^\sigma \mathrm{d}\sigma$$

where γ_4^σ denotes the transformation of γ_4. See again Fig 2. Now, an application of the reflection formula of the gamma function yields

$$\xi_4(v) = \frac{\alpha}{2\pi i} \int_{-\gamma_4^\sigma} \frac{\Gamma(\alpha\sigma + \alpha)\Gamma(-\sigma)}{\Gamma\big((\alpha-1)\sigma + \alpha + 1\big)} \frac{\sin\big(\pi(\alpha\sigma + \alpha)\big)}{\sin\big(\pi((\alpha-1)\sigma + \alpha)\big)} v^\sigma \mathrm{d}\sigma$$

where the minus in front of γ_4^σ means a change of orientation. Finally, by resorting to a well-known Cauchy theorem, we can deform the path $-\gamma_4^\sigma$ into a new path that consists of the conjunction of γ_2 with a loop the encircles the point $\frac{\alpha}{1-\alpha}$ counter-clock-wise, without changing the value of the integral. See Fig 3 below.

Thus, setting

$$g(\sigma; v) := \frac{\Gamma(\alpha\sigma + \alpha)\Gamma(-\sigma)}{\Gamma\big((\alpha-1)\sigma + \alpha + 1\big)} \frac{\sin\big(\pi(\alpha\sigma + \alpha)\big)}{\sin\big(\pi((\alpha-1)\sigma + \alpha)\big)} v^\sigma$$

we have

$$\xi_4(v) = \frac{\alpha}{2\pi i} \int_{\gamma_2} g(\sigma; v)\mathrm{d}\sigma + \alpha \ Res\left(g(\cdot; v); \frac{\alpha}{1-\alpha}\right).$$

Repeating the above argument based on the approximation of γ_2 with $\gamma_{2,N}$, we get

$$\frac{1}{2\pi i} \oint_{\gamma_{2,N}} g(\sigma; v)\mathrm{d}\sigma = -\sum_{n=0}^{N} Res\left(g(\cdot; v); n\right) - \sum_{n=0}^{N_\alpha} Res\left(g(\cdot; v); \frac{n+\alpha}{1-\alpha}\right)$$

with $N_\alpha := [(N + 1/2)(1-\alpha) - \alpha]$. At this stage, it is enough to notice that

$$Res\left(g(\cdot; v); \frac{n+\alpha}{1-\alpha}\right) = 0$$

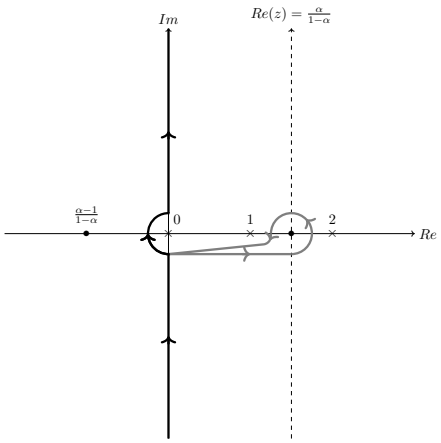

Figure 5.3: Modification of the path $-\gamma_4^\sigma$, equal to the conjunction of γ_2 (black in the figure) and a loop around the point $\frac{\alpha}{1-\alpha}$ (gray in the figure).

for any $n \in \mathbb{N}$. Therefore, taking the limit as $N \to +\infty$, we can conclude that $\xi_4(v) = \xi_1(v)$ on $\{w \in \mathbb{C} : |v| < \alpha^{-\alpha}(1-\alpha)^{\alpha-1}, |\mathrm{Arg}(v)| < \pi\alpha_*\}$ and, consequently, that $\xi_4(v) = \xi_2(v)$ on $\{w \in \mathbb{C} : |\mathrm{Arg}(v)| < \pi\alpha_*\}$. Finally, after proving the last identity, we can remove the constraint that $\alpha \neq q(q+1)^{-1}$ for some $q \in \mathbb{N}$ thanks to a straightforward approximation argument with respect to α, to conclude that $\xi_4(v) = \xi_2(v)$ on $\{w \in \mathbb{C} : |\mathrm{Arg}(v)| < \pi\alpha_*\}$ whichever $\alpha \in (0,1)$ may be. At this stage, the remark that $\mathbb{C}^+ \subseteq \{w \in \mathbb{C} : |\mathrm{Arg}(v)| < \pi \max\{\alpha, 1-\alpha\}\}$ leads to identify the desired analytical continuation of $v \mapsto \xi(v)$ with either $v \mapsto \xi_2(v)$ or $v \mapsto \xi_4(v)$ according to whether $\alpha \geq 1/2$ or $\alpha < 1/2$, respectively. This concludes the proof. ■

Now, we investigate the large n asymptotic behaviour of the function $\mathcal{F}_\varrho^{(n)}$ that appears in Lemma A.7. Our Lemma A.9 below can be regarded as a re-adaptation of [74, Equation 5.3]. We just premise the definition of a useful notation. Let $\mathcal{D}$ be some subset of $\mathbb{R}^d$ (or $\mathbb{C}$). Consider two sequences of functions $\{f_n\}_{n\in\mathbb{N}}$, $\{g_n\}_{n\in\mathbb{N}}$, with $f_n, g_n : \mathcal{D} \to \mathbb{C}$ for every n. If, for some $\beta \in \mathbb{R}$, there exists a function $h : \mathcal{D} \to \mathbb{C}$ such that

$$n^{-\beta}\,[f_n - g_n] \to h$$

uniformly on compact sets, we write

$$f_n = g_n + O_{u.c.}(n^\beta).$$

Let now recall some previous notation. For $y \in \mathbb{C}^+$, set $\tau(y) := \xi(1/(\alpha y))$, with the same function ξ as in Lemma A.8, and $D(y)$ be as in (2.2.26), with y in place of z. Moreover, set

$$g(y) := \frac{y}{\tau(y)} - \frac{1-\alpha}{\alpha}\log\tau(y) \tag{A.9}$$

where log is intended as the principal branch of the logarithm.

Lemma A.9. *With the above notation, for $y \in \mathbb{C}^+$ it holds that*

$$\mathcal{F}_\varrho^{(n)}(y) = \mathcal{P}_\varrho^{(n)}(y)\left[1 + \frac{1}{n}\mathcal{R}_\varrho(y) + r_\varrho^{(n)}(y)\right] \tag{A.10}$$

where:

i) the principal part $\mathcal{P}_\varrho^{(n)}$ *is defined as*

$$\mathcal{P}_\varrho^{(n)}(y) := (\alpha n y)^{n-\varrho\frac{\alpha}{1-\alpha}} \left(\frac{1}{\tau(y)}\right)^\varrho \exp\left\{-n\left[1+g(y)\right]\right\} \sqrt{\frac{2\pi\alpha}{n(1-\alpha)D(y)}};$$

ii) the first order remainder $\mathcal{R}_\varrho$ *is given by*

$$\mathcal{R}_\varrho(y) = \varrho^2\, c_1(y) + \varrho\, c_2(y) + c_3(y)$$

with

$$c_1(y) := \frac{\alpha}{2(1-\alpha)D(y)} \qquad \textit{and} \qquad c_2(y) := \frac{1-2\alpha}{2(1-\alpha)D(y)} + \frac{1}{2D^2(y)}$$

and another analytic function c_3 *independent of* ϱ*;*

iii) the second order remainder $r_\varrho^{(n)}$ *satisfies, as* $n \to +\infty$

$$r_\varrho^{(n)}(y) = O_{u.c.}\left(\frac{1}{n^2}\right).$$

Proof. Fix $y \in (0,+\infty)$. Following [74, Section 5], define

$$X(y) := (1-\alpha)\, ny$$

and

$$a_{n,\varrho}(y) := \frac{1}{\alpha y} - \frac{\varrho+1}{(1-\alpha)ny}.$$

Notice that $a_{n,\varrho}(y)X(y) = n^{\frac{1-\alpha}{\alpha}} - \varrho - 1$. Set

$$\tau_n(y) := \xi(a_{n,\varrho}(y)) \ ,$$

where ξ denotes the solution to (A.8). Now, a strict application of [74, Equation 5.3] yields

$$\begin{aligned}\mathcal{F}_\varrho^{(n)}(y) = \sqrt{\frac{2\pi}{(1-\alpha)ny}} \left[\frac{1}{\alpha\tau_n(y)} + \frac{1-\alpha}{\alpha}\frac{1}{\alpha y} - \frac{\varrho+1}{\alpha n y}\right]^{-\frac{1}{2}} (\alpha n y)^{n-\frac{\varrho\alpha}{1-\alpha}} \times \\ \times \left(\frac{1}{\tau_n(y)}\right)^\varrho \exp\left\{-n\left[1+\frac{y}{\tau_n(y)} - \frac{1-\alpha}{\alpha}\log(\tau_n(y))\right] + \frac{\alpha(\varrho+1)}{1-\alpha}\right\} \times \\ \times \left[1 + \frac{\mathscr{R}_1(y)}{n} + O_{u.c.}\left(\frac{1}{n^2}\right)\right]\end{aligned} \tag{A.11}$$

where $\mathscr{R}_1$ is a suitable analytical function which does not depend on ϱ. At this stage, to prove (A.10), we need further investigations on the quantity $\tau_n(y)$, obtained by exploiting the regularity of $v \mapsto \xi(x)$ state in Lemma A.8. Accordingly, set $a(y) := (\alpha y)^{-1}$, and notice that $\tau(y) = \xi(a(y))$. Moreover, notice that

$$a_{n,\varrho}(y) = a(y) - \left(\frac{\varrho+1}{(1-\alpha)y}\right)\frac{1}{n},$$

which shows that $a_{n,\varrho}(y)$ can be regarded as a small increment (as $n \to +\infty$) of $a(y)$. Thus, a Taylor expansion of the function ξ around $a(y)$ yields

$$\xi(a_{n,\varrho}(y)) = \xi(a(y))\left[1 + \frac{\mathscr{A}_\varrho(y)}{n} + \frac{\mathscr{B}_\varrho(y)}{n^2} + O_{u.c.}\left(\frac{1}{n^3}\right)\right],$$

where

$$\mathscr{A}_\varrho(y) := -\frac{\alpha(\varrho+1)}{(1-\alpha)D(y)}$$

and

$$\mathscr{B}_\varrho(y) := -\mathscr{A}_\varrho^2(y)\,\frac{D(y)\,(1-3\alpha)+1-\alpha}{2\alpha\,D(y)}$$

with $D(y)$ as in (2.2.26). Three consequences of the above Taylor expansion are:

i)
$$\left(\frac{1}{\tau_n(y)}\right)^\varrho = \left(\frac{1}{\tau(y)}\right)^\varrho \left[1+\frac{1}{n}\,\mathscr{R}_{\varrho,2}(y)+O_{u.c.}\left(\frac{1}{n^2}\right)\right]$$

ii)
$$\begin{aligned}&\left[\frac{1}{\alpha\tau_n(y)}+\frac{1-\alpha}{\alpha}\,\frac{1}{\alpha y}-\frac{\varrho+1}{\alpha n y}\right]^{-\frac{1}{2}}\\ &\quad=\left[\frac{1}{\alpha\tau(y)}+\frac{1-\alpha}{\alpha}\,\frac{1}{\alpha y}\right]^{-\frac{1}{2}}\left[1+\frac{1}{n}\,\mathscr{R}_{\varrho,3}(y)+O_{u.c.}\left(\frac{1}{n^2}\right)\right]\end{aligned}$$

iii)
$$\begin{aligned}&\exp\left\{-n\left[1+\frac{y}{\tau_n(y)}-\frac{1-\alpha}{\alpha}\,\log(\tau_n(y))\right]+\frac{\alpha(\varrho+1)}{1-\alpha}\right\}\\ &\quad=\exp\left\{-n\left[1+\frac{y}{\tau(y)}-\frac{1-\alpha}{\alpha}\,\log(\tau(y))\right]\right\}\left[1+\frac{1}{n}\,\mathscr{R}_{\varrho,4}(y)+O_{u.c.}\left(\frac{1}{n^2}\right)\right],\end{aligned}$$

with

$$\begin{aligned}\mathscr{R}_{\varrho,2}(y) &:= \varrho\,\mathscr{A}_\varrho(y)\\ \mathscr{R}_{\varrho,3}(y) &:= \frac{1}{2}\left(\frac{\mathscr{A}_\varrho(y)}{\alpha\tau(y)}+\frac{\varrho+1}{\alpha y}\right)\frac{\alpha y}{D(y)}\\ \mathscr{R}_{\varrho,4}(y) &:= \frac{y}{\tau(y)}\left[\mathscr{A}_\varrho(y)^2-\mathscr{B}_\varrho(y)\right].\end{aligned}$$

At this stage, identity (A.10) follows by just plugging the above expansions i)-ii)-iii) into (A.11) and patiently rearranging the terms. Finally, after establishing (A.10) for $y \in (0,+\infty)$, we conclude that it must hold also for $y \in \mathbb{C}^+$ by the uniqueness of the analytic continuation. ■

The following lemma states a basic calculus result, that guarantees the existence of a continuous function bounding from above any uniformly convergent sequence of functions on a compact set. Being unable to find a precise statement on some calculus textbook, we include here both statement and proof, to maintain the paper self-contained.

Lemma A.10. *Let $\mathbb{K} \subseteq \mathbb{R}^d$ (or $\mathbb{C}$) be a compact set. Consider a sequence of continuous functions $\{f_n\}_{n\in\mathbb{N}}$, $f_n : \mathbb{K} \to \mathbb{R}$ such that, as $n \to +\infty$, $f_n \to f$ uniformly for some (continuous) function $f : \mathbb{K} \to \mathbb{R}$. For each $x \in \mathbb{K}$, define $S(x) := \sup_{n\in\mathbb{N}} |f_n(x)|$. Then $S : \mathbb{K} \to \mathbb{R}$ is continuous.*

To prove Lemma A.10, we first need the following lemma.

Lemma A.11. *For $i = 1, 2$, let $(\mathbb{M}_i, d_{\mathbb{M}_i})$ be metric spaces, endowed with the metric topology. Let $\mathbb{K}_i \subseteq \mathbb{M}_i$ be compact subsets, and let $F : \mathbb{K}_1 \times \mathbb{K}_2 \to \mathbb{R}$ be a continuous function on the product space. Then, $G : \mathbb{K}_1 \to \mathbb{R}$ defined by $G(x) = \sup_{y \in \mathbb{K}_2} F(x, y)$ is continuous.*

Proof of lemma A.11. Consider a convergent sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq \mathbb{K}_1$ with $x_n \to \bar{x} \in \mathbb{K}_1$. We prove that

$$G(\bar{x}) \leq \liminf_{n\to+\infty} G(x_n) \leq \limsup_{n\to+\infty} G(x_n) \leq G(\bar{x}).$$

Since $\mathbb{K}_2$ is compact and the real-valued function $y \mapsto F(\bar{x}, y)$ on $\mathbb{K}_2$ is continuous, there exists $y^\star \in \mathbb{K}_2$ such that

$$G(\bar{x}) = \sup_{y\in\mathbb{K}_2} F(\bar{x}, y) = F(\bar{x}, y^\star).$$

By definition, $G(x_n) \geq F(x_n, y^\star)$ for every $n \in \mathbb{N}$. Furthermore, the real-valued function $x \mapsto F(x, y^\star)$ on $\mathbb{K}_1$ is continuous. Whence,

$$\liminf_{n\to+\infty} G(x_n) \geq \liminf_{n\to+\infty} F(x_n, y^\star) = F(\bar{x}, y^\star) = G(\bar{x}).$$

We prove the other inequality by contradiction. Assume there exists a subsequence $\{x_{n_j}\}_{j\in\mathbb{N}}$ of $\{x_n\}$ such that

$$G(x_{n_j}) \to L > G(\bar{x}).$$

For each $j \in \mathbb{N}$, since the real-valued function $y \mapsto F(x_{n_j}, y)$ on $\mathbb{K}_2$ is continuous, there exists $y_j \in \mathbb{K}_2$ such that

$$G(x_{n_j}) = \sup_{y\in\mathbb{K}_2} F(x_{n_j}, y) = F(x_{n_j}, y_j).$$

Now, consider the sequence $\{y_j\}_{j\in\mathbb{N}} \subseteq \mathbb{K}_2$. By sequential compactness, there exists a subsequence $\{y_{j_k}\}_{k\in\mathbb{N}}$ converging to some point $\bar{y} \in \mathbb{K}_2$. Hence, by continuity of F,

$$G(x_{n_{j_k}}) = F(x_{n_{j_k}}, y_{j_k}) \to F(\bar{x}, \bar{y}) \leq G(\bar{x})$$

which produces the contradiction. ■

Proof of lemma A.10. Let $\mathbb{N}^* = \mathbb{N} \cup \{\infty\}$ denote the Alexandroff compactification of $\mathbb{N}$ with the discrete topology. Further, define $F : \mathbb{K} \times \mathbb{N}^* \to [0, +\infty)$,

$$F(x, n) = \begin{cases} |f_n(x)| & \text{if } n \in \mathbb{N} \\ |f(x)| & \text{if } n = \infty. \end{cases}$$

If F is continuous, then Lemma A.11 entails that the function $G : \mathbb{K} \to [0, +\infty)$, defined by

$$G(x) := \sup_{n\in\mathbb{N}^*} F(x, n) = \max\{\sup_{n\in\mathbb{N}} |f_n(x)|, |f(x)|\} = \max\{S(x), |f(x)|\}$$

is continuous. Note that, for every $x \in \mathbb{K}$, $S(x)$ is either equal to the limit of $|f_n(x)|$ or it is a maximum, so that $S(x) \geq |f(x)|$. This implies that $S = G$, and that S is continuous. It remains to prove that F is continuous. Consider a convergent sequence $\{(x_k, n_k)\}_{k\in\mathbb{N}} \subseteq \mathbb{K} \times \mathbb{N}^*$, with $(x_k, n_k) \to (\bar{x}, \bar{n})$. We treat two cases separately:

i) if $\bar{n} \in \mathbb{N}$, then the sequence $\{n_k\}$ is eventually constant with $n_k = \bar{n}$. Hence, by continuity of $f_{\bar{n}}$,

$$F(x_k, n_k) = f_{n_k}(x_k) = f_{\bar{n}}(x_k) \to f_{\bar{n}}(\bar{x}) = F(\bar{x}, \bar{n});$$

ii) if $\bar{n} = \infty$, then by uniform convergence of $\{f_n\}_{n\geq 1}$,

$$F(x_k, n_k) = f_{n_k}(x_k) \to f(\bar{x}) = F(\bar{x}, \infty).$$

■

A.2.0.2 Step 2

The following proposition connects the quantities $\mu(z)$ and $\sigma^2(z)$ that appear in Proposition 2 to the mean and variance of $R_n(z)$, respectively.

Proposition 11. *As z varies in $(0, +\infty)$ and $n \to \infty$, one has*

$$\mathbb{E}[R_n(z)] = n\mu(z) + O_{u.c.}(1)$$

and

$$\mathrm{Var}(R_n(z)) = n\sigma^2(z) + O_{u.c.}(1).$$

Remark A.12. *The LLN (2.2.24) follows directly from the asymptotic expansions of Proposition 11 by a standard application of Chebychev inequality.*

Proof of Proposition 11. First, exploit [27, equation (12)] to obtain the expression of the probability generating function of $R_n(z)$:

$$\begin{aligned} G_{R_n(z)}(s) &= \frac{\sum_{k=1}^n \mathscr{C}(n,k;\alpha)\,(nsz)^k}{\sum_{k=1}^n \mathscr{C}(n,k;\alpha)\,(nz)^k} \\ &= e^{nz(s-1)}\, s^{\frac{n}{\alpha}} \frac{\int_0^{+\infty} x^n \exp\left\{-x(nsz)^{\frac{1}{\alpha}}\right\} f_\alpha(x)\,\mathrm{d}x}{\int_0^{+\infty} x^n \exp\left\{-x(nz)^{\frac{1}{\alpha}}\right\} f_\alpha(x)\,\mathrm{d}x} \\ &= e^{nz(s-1)}\, s^{\frac{n}{\alpha}} \frac{I_n(zs)}{I_n(z)}\ , \end{aligned} \tag{A.12}$$

where s can be considered as a real variable and, for any $y > 0$,

$$I_n(y) := \int_0^{+\infty} x^n \exp\left\{-x(ny)^{\frac{1}{\alpha}}\right\} f_\alpha(x)\,\mathrm{d}x\ .$$

Whence,

$$I_n'(y) = -\frac{n^{\frac{1}{\alpha}}}{\alpha} y^{\frac{1-\alpha}{\alpha}} \int_0^{+\infty} x^{n+1} \exp\left\{-x(ny)^{\frac{1}{\alpha}}\right\} f_\alpha(x)\,\mathrm{d}x =: -\frac{n^{\frac{1}{\alpha}}}{\alpha}\, y^{\frac{1-\alpha}{\alpha}}\, \mathcal{J}_n^{(1)}(y)$$

and

$$\begin{aligned} I_n''(y) &= \frac{1-\alpha}{\alpha}\frac{1}{y} I_n'(y) + \frac{n^{2\alpha}}{\alpha^2} y^{\frac{2(1-\alpha)}{\alpha}} \int_0^{+\infty} x^{n+2} \exp\left\{-x(ny)^{\frac{1}{\alpha}}\right\} f_\alpha(x)\,\mathrm{d}x \\ &=: \frac{1-\alpha}{\alpha}\frac{1}{y} I_n'(y) + \frac{n^{2\alpha}}{\alpha^2} y^{\frac{2(1-\alpha)}{\alpha}}\, \mathcal{J}_n^{(2)}(y). \end{aligned}$$

Differentiation and evaluation at $s = 1$ of (A.12) then yield

$$G'_{R_n(z)}(1) = nz + \frac{n}{\alpha} + z\frac{I_n'(z)}{I_n(z)} = nz + \frac{n}{\alpha} - \frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\frac{\mathcal{J}_n^{(1)}(z)}{I_n(z)} \tag{A.13}$$

and

$$G''_{R_n(z)}(1) = \left(nz + \frac{n}{\alpha}\right)^2 - \frac{n}{\alpha} + 2\left(nz + \frac{n}{\alpha}\right) z\,\frac{I'_n(z)}{I_n(z)} + z^2\,\frac{I''_n(z)}{I_n(z)} \tag{A.14}$$
$$= \left(nz + \frac{n}{\alpha}\right)^2 - \frac{n}{\alpha} + \left[2n\left(z + \frac{1}{\alpha}\right) + \frac{1-\alpha}{\alpha}\right] \times$$
$$\times \left[G'_{R_n(z)}(1) - nz - \frac{n}{\alpha}\right] + \frac{(nz)^{2\alpha}}{\alpha^2}\,\frac{\mathcal{J}_n^{(2)}(z)}{I_n(z)}.$$

Thanks to the approximation results given in Lemma A.1, we can utilize the function E_α therein to get the identities

$$\frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\,\frac{\mathcal{J}_n^{(1)}(z)}{I_n(z)} = \frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\,\frac{J_{1,0}^{(n)}(z) + Q_\alpha J_{1,1}^{(n)}(z) + R_1^{(n)}(z)/\chi_\alpha}{J_{0,0}^{(n)}(z) + Q_\alpha J_{0,1}^{(n)}(z) + R_0^{(n)}(z)/\chi_\alpha}$$

and

$$\frac{(nz)^{\frac{2}{\alpha}}}{\alpha^2}\,\frac{\mathcal{J}_n^{(2)}(z)}{I_n(z)} = \frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\,\frac{J_{2,0}^{(n)}(z) + Q_\alpha J_{2,1}^{(n)}(z) + R_2^{(n)}(z)/\chi_\alpha}{J_{0,0}^{(n)}(z) + Q_\alpha J_{0,1}^{(n)}(z) + R_0^{(n)}(z)/\chi_\alpha},$$

where Q_α is the same as in (A.2), χ_α is a positive constant depending only on α, and the $J_{\varepsilon,\delta}^{(n)}$'s are as in (A.5). Moreover, for $j = 0, 1, 2$, define

$$R_j^{(n)}(z) := \int_0^{+\infty} x^{n+j}\,\exp\left\{-x(nz)^{\frac{1}{\alpha}}\right\} E_\alpha(x)\left[-1 - Q_\alpha x^{\frac{\alpha}{1-\alpha}} + \frac{f_\alpha(x)}{E_\alpha(x)}\right] \mathrm{d}x.$$

These quantities represent remainders terms that satisfy

$$\left|\frac{R_j^{(n)}(z)}{\chi_\alpha}\right| \leq \mathcal{K}_\alpha J_{j,2}^{(n)}(z) \tag{A.15}$$

for some positive constant $\mathcal{K}_\alpha$. Now, to make some simplifications, set

$$\Psi(\alpha, n, z) := (\alpha n y)^n \sqrt{\frac{2\pi\alpha}{n(1-\alpha)D(z)}}$$

with the same $D(z)$ as in (2.2.26). Then, for $j = 0, 1, 2$, define

$$\mathfrak{R}_j^{(n)}(z) := \frac{R_j^{(n)}(z)\,\alpha^{\frac{\alpha}{2(1-\alpha)}-j}(nz)^{\frac{\alpha}{2(1-\alpha)}}\sqrt{\tau(z)}}{\chi_\alpha\,\mathcal{C}(\alpha, n, z)\,\Psi(\alpha, n, z)}$$

where $\tau(z)$ denotes the unique real, positive solution to equation (2.2.21), while $\mathcal{C}(\alpha, n, z)$ is the same as in (A.6). Combining Lemmata A.7 and A.9 with (2.2.21) allows to write, after some straightforward algebraic manipulations, the identities

$$\frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\,\frac{\mathcal{J}_n^{(1)}(z)}{I_n(z)} = \frac{N_1(z,n)}{1+\varepsilon(z,n)} \qquad \text{and} \qquad \frac{(nz)^{\frac{2}{\alpha}}}{\alpha^2}\,\frac{\mathcal{J}_n^{(2)}(z)}{I_n(z)} = \frac{N_2(z,n)}{1+\varepsilon(z,n)},$$

where:

$$\begin{aligned}
\varepsilon(z,n) &:= \frac{1}{n} R_{\varrho(0,0)} + r^{(n)}_{\varrho(0,0)} + \frac{1}{n}\, Q_\alpha \alpha^{\frac{\alpha}{1-\alpha}}\, \frac{\tau(z)}{z} \left[1 + \frac{1}{n} R_{\varrho(0,1)} + r^{(n)}_{\varrho(0,1)}\right] + \mathfrak{R}^{(n)}_0 \\
N_1(z,n) &:= n \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right) \left[1 + \frac{1}{n} R_{\varrho(1,0)} + r^{(n)}_{\varrho(1,0)}\right] \\
&\quad + \; Q_\alpha \alpha^{\frac{\alpha}{1-\alpha}}\, \frac{\tau(z)}{z} \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right) \left[1 + \frac{1}{n} R_{\varrho(1,1)} + r^{(n)}_{\varrho(1,1)}\right] + \mathfrak{R}^{(n)}_1 \\
N_2(z,n) &:= n^2 \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2 \left[1 + \frac{1}{n} R_{\varrho(2,0)} + r^{(n)}_{\varrho(2,0)}\right] \\
&\quad + n\, Q_\alpha \alpha^{\frac{\alpha}{1-\alpha}}\, \frac{\tau(z)}{z} \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2 \left[1 + \frac{1}{n} R_{\varrho(2,1)} + r^{(n)}_{\varrho(2,1)}\right] + \mathfrak{R}^{(n)}_2.
\end{aligned}$$

Now, consider the straightforward identity

$$\frac{1}{1+\varepsilon(z,n)} = 1 - \varepsilon(z,n) + \frac{\varepsilon(z,n)^2}{1+\varepsilon(z,n)}.$$

The bound (A.15) and Lemma A.9 guarantee that

$$\frac{\varepsilon^2(z,n)}{1+\varepsilon(z,n)} = O_{u.c.}\left(\frac{1}{n^2}\right).$$

Thus, rearranging of the terms and highlighting only the first orders in the asymptotical expansion in n, we can write

$$\begin{aligned}
&\frac{(nz)^{\frac{1}{\alpha}}}{\alpha}\, \frac{\mathcal{J}^{(1)}_n(z)}{I_n(z)} \\
&= N_1(z,n) - \varepsilon(z,n) N_1(z,n) + N_1(z,n) \frac{\varepsilon^2(z,n)}{1+\varepsilon(z,n)} \\
&= n \left[\frac{1}{\alpha} + \frac{z}{\tau(z)}\right] + \left[\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right) \left(R_{\varrho(1,0)} - R_{\varrho(0,0)}\right)\right] + O_{u.c.}\left(\frac{1}{n}\right) \\
&= n \left[\frac{1}{\alpha} + \frac{z}{\tau(z)}\right] + \left[\frac{1-\alpha}{\alpha} \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right) \left(\frac{1-2\alpha}{\alpha} c_1(z) - c_2(z)\right)\right] + O_{u.c.}\left(\frac{1}{n}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{(nz)^{\frac{2}{\alpha}}}{\alpha^2}\, \frac{\mathcal{J}^{(2)}_n(z)}{I_n(z)} \\
&= N_2(z,n) - \varepsilon(z,n) N_2(z,n) + N_2(z,n) \frac{\varepsilon(z,n)^2}{1+\varepsilon(z,n)} \\
&= n^2 \left[\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2\right] + n \left[\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2 \left(R_{\varrho(2,0)} - R_{\varrho(0,0)}\right)\right] + O_{u.c.}(1) \\
&= n^2 \left[\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2\right] + n \left[2\frac{1-\alpha}{\alpha} \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2 \left(\frac{2-3\alpha}{\alpha} c_1(z) - c_2(z)\right)\right] + O_{u.c.}(1).
\end{aligned}$$

Finally, substitute the above expansions in the expressions (A.13) and (A.14) $G'_{R_n(z)}(1)$ and $G''_{R_n(z)}(1)$, respectively, to obtain

$$G'_{R_n(z)}(1) = n\, \mathcal{A}(z) + \mathcal{B}(z) + O_{u.c.}\left(\frac{1}{n}\right)$$

and

$$G''_{R_n(z)}(1) = n^2\,\mathcal{C}(z) + n\,\mathcal{D}(z) + O_{u.c.}(1)$$

where:

$$\begin{aligned}
\mathcal{A}(z) &:= z - \frac{z}{\tau(z)} \\
\mathcal{B}(z) &:= -\frac{1-\alpha}{\alpha}\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)\left(\frac{1-2\alpha}{\alpha}c_1(z) - c_2(z)\right) \\
\mathcal{C}(z) &:= \left(z + \frac{1}{\alpha}\right)^2 - 2\left(z + \frac{1}{\alpha}\right)\left(\frac{z}{\tau(z)} + \frac{1}{\alpha}\right) + \left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2 = \left(z - \frac{z}{\tau(z)}\right)^2 \\
\mathcal{D}(z) &:= -\frac{1}{\alpha} + 2\left(z + \frac{1}{\alpha}\right)\mathcal{B}(z) - \frac{1-\alpha}{\alpha}\left(\frac{z}{\tau(z)} + \frac{1}{\alpha}\right) \\
&\quad + \frac{2(1-\alpha)}{\alpha}\left(\frac{1}{\alpha} + \frac{z}{\tau(z)}\right)^2\left(\frac{2-3\alpha}{\alpha}c_1(z) - c_2(z)\right)
\end{aligned}$$

with the same c_1 and c_2 as in Lemma A.9. At this stage, notice that, with reference to the quantities $\mu(z)$ and $\sigma^2(z)$ that appear in Proposition 2, one has that $\mathcal{A}(z) = \mu(z)$, $\mathcal{C}(z) = (\mu(z))^2$ and $\mathcal{D}(z) + \mu(z) - 2\mu(z)\mathcal{B}(z) = \sigma^2(z)$. Indeed, the first two identities are evident while the third can be verified via direct computation. Combining the above identities, one gets

$$\mathbb{E}[R_n(z)] = G'_{R_n(z)}(1) = n\,\mathcal{A}(z) + O_{u.c.}(1)$$

and

$$\begin{aligned}
\mathrm{Var}\,(R_n(z)) &= G''_{R_n(z)}(1) + G'_{R_n(z)}(1) - \left(G'_{R_n(z)}(1)\right)^2 \\
&= n^2\,\left[\mathcal{C}(z) - \mathcal{A}^2(z)\right] + n\,\left[\mathcal{D}(z) + \mathcal{A}(z) - 2\mathcal{A}(z)\mathcal{B}(z)\right] + O_{u.c.}(1)
\end{aligned}$$

and the proof is concluded. ■

A.2.0.3 Step 3

We prove the following Berry-Esseen lemma for $R_n(z)$.

Lemma A.13. *Fix ζ_0 and ζ_1 such that $0 < \zeta_0 < z_0 < \zeta_1 < +\infty$, with the same z_0 as in Proposition 1. If $\xi \in \mathbb{R}$ satisfies*

$$|\xi| \le \mathcal{C}\,\sigma(z)\,n^{\delta} \tag{A.16}$$

for every $z \in [\zeta_0, \zeta_1]$, for some positive constant $\mathcal{C}$, and some $\delta \in (0, 1/6)$, then there exists anoter constant $\tilde{c}$, depending on ζ_0, ζ_1, α, λ and $\mathcal{C}$ such that

$$\left|\varphi_{W_n(z)}(\xi) - e^{-\frac{\xi^2}{2}}\right| \le \tilde{c}\,e^{-\frac{\xi^2}{2}}\,n^{3\delta-\frac{1}{2}}.$$

Proof. Combining (A.12) with Lemma A.1, we get

$$\frac{I_n(zs)}{I_n(z)} = \frac{J^{(n)}_{0,0}(zs) + R_n(zs)/\chi_\alpha}{J^{(n)}_{0,0}(z) + R_n(z)/\chi_\alpha} = \frac{J^{(n)}_{0,0}(zs)}{J^{(n)}_{0,0}(z)}\,\frac{1 + R_n(zs)/\left(\chi_\alpha\,J^{(n)}_{0,0}(zs)\right)}{1 + R_n(z)/\left(\chi_\alpha\,J^{(n)}_{0,0}(z)\right)},$$

for $z \in (0, +\infty)$ and $s \in \mathbb{C}^+$, where: χ_α is a positive constant depending only on α, the $J^{(n)}_{\varepsilon,\delta}$'s are as in (A.5), and, for any $y \in \mathbb{C}^+$

$$R_n(y) := \int_0^{+\infty} x^n\,\exp\left\{-x(ny)^{\frac{1}{\alpha}}\right\}\,E_\alpha(x)\left[-1 + \frac{f_\alpha(x)}{E_\alpha(x)}\right]\,\mathrm{d}x$$

satisfies

$$|R_n(y)| \leq C_\alpha \, J_{0,1}^{(n)}\left(Re(y)\right) \tag{A.17}$$

with C_α as in (A.1). Then, combining Lemma A.7 and Lemma A.9, we have

$$\begin{aligned}\frac{I_n(zs)}{I_n(z)} &= s^{-\frac{n}{\alpha}+\frac{2-\alpha}{2\alpha(1-\alpha)}-\frac{1}{\alpha}} \frac{\mathcal{F}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} \, \mathfrak{R}_1^{(n)}(z,s) \\ &= s^{-\frac{n}{\alpha}+\frac{2-\alpha}{2\alpha(1-\alpha)}-\frac{1}{\alpha}} \frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)} \, \mathfrak{R}_1^{(n)}(z,s)\, \mathfrak{R}_2^{(n)}(z,s),\end{aligned}$$

where

$$\mathfrak{R}_1^{(n)}(z,s) := \frac{1+R_n(zs)/\left(\chi_\alpha \, J_{0,0}^{(n)}(zs)\right)}{1+R_n(z)/\left(\chi_\alpha \, J_{0,0}^{(n)}(z)\right)}$$

$$\mathfrak{R}_2^{(n)}(z,s) := \frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} \frac{\mathcal{F}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)}.$$

Now, combine the above identities with the expression of $\mathcal{P}_{\frac{1}{2}}^{(n)}$ in Lemma A.9, with g as in (A.9), to obtain

$$\begin{aligned}G_{R_n(z)}(s) &= e^{nz(s-1)} \, s^{\frac{1}{2(1-\alpha)}} \frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)} \, \mathfrak{R}_1^{(n)}(z,s)\, \mathfrak{R}_2^{(n)}(z,s) \\ &= \exp\left\{-n\,\left[(g(zs)-zs-\log(s))-(g(z)-z)\right]\right\} \mathfrak{R}_1^{(n)}(z,s)\, \mathfrak{R}_2^{(n)}(z,s)\, \mathfrak{R}_3^{(n)}(z,s),\end{aligned}$$

with

$$\mathfrak{R}_3^{(n)}(z,s) := s^{\frac{1}{2}} \sqrt{\frac{\alpha z + \tau(z)(1-\alpha)}{\alpha zs + \tau(zs)(1-\alpha)}}.$$

Now, let $\varphi_{W_n(z)}$ denote the characteristic function of the variable $W_n(z)$. That is,

$$\begin{aligned}&\varphi_{W_n(z)}(\xi) \\ &= \exp\left\{-\sqrt{n}\,\frac{i\,\xi\,\mu(z)}{\sigma(z)}\right\} G_{R_n(z)}\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right) \\ &= \exp\left\{-\sqrt{n}\,\frac{i\,\xi\,\mu(z)}{\sigma(z)}\right\} \exp\left\{-n\,\left[g\left(ze^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)-g(z)-z\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}-1\right)-\frac{i\,\xi}{\sqrt{n}\sigma(z)}\right]\right\} \times \\ &\quad \times \mathfrak{R}_1^{(n)}\left(z,e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right) \mathfrak{R}_2^{(n)}\left(z,e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right) \mathfrak{R}_3^{(n)}\left(z,e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right).\end{aligned}$$

If ξ satisfies (F.4), then [78, Chapter IV, Lemma 5] guarantees that

$$\left|e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}-1\right| \leq \left|\frac{\xi}{\sqrt{n}\sigma(z)}\right| \leq \mathcal{C}\, n^{\delta-\frac{1}{2}}.$$

We apply Taylor's formula to $g\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right) - g(z)$ and then to $\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}} - 1\right)$ to obtain

$$\begin{aligned}
&\varphi_{W_n(z)}(\xi) \\
&\quad = \exp\left\{\sqrt{n}\,\frac{i\,\xi}{\sigma(z)}\left[-\mu(z) + z + 1 - z\,g'(z)\right] - \frac{\xi^2}{2\,\sigma^2(z)}\left[z + zg'(z) - z^2 g''(z)\right]\right\} \times \\
&\qquad \times \mathfrak{R}_1^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_2^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_3^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_4^{(n)}(z,\xi)\,,
\end{aligned}$$

where

$$\begin{aligned}
&\mathfrak{R}_4^{(n)}(z,\xi) \\
&\quad := \exp\left\{nz\left[1 - g'(z)\right]\left[\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}} - 1\right) - \left(\frac{i\,\xi}{\sqrt{n}\sigma(z)} - \frac{\xi^2}{2n\sigma^2(z)}\right)\right]\right. \\
&\qquad + \frac{1}{2}\,nz^2 g''(z)\left[\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}} - 1\right)^2 + \frac{\xi^2}{n\sigma^2}\right] \\
&\qquad \left. + \frac{1}{2}nz^3\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}} - 1\right)^3 \int_0^1 g'''\left(z\left[1 + t\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}} - 1\right)\right]\right)(1-t)^2\,\mathrm{d}t\right\}.
\end{aligned}$$

Recalling the definitions of $\mu(z)$, $\sigma^2(z)$ and $g(z)$, we conclude that

$$\begin{aligned}
&\varphi_{W_n(z)}(\xi) \\
&\quad = \exp\left\{-\frac{\xi^2}{2}\right\}\mathfrak{R}_1^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_2^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_3^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_4^{(n)}(z,\xi)\,.
\end{aligned}$$

By combining Lemma A.1, Lemma A.9 and [78, Chapter IV, Lemma 5] we prove that the function

$$\mathfrak{R}^{(n)}(z,\xi) := \mathfrak{R}_1^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_2^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_3^{(n)}\left(z, e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}\right)\mathfrak{R}_4^{(n)}(z,\xi)\,,$$

defined for $z \in [\zeta_0, \zeta_1]$ and ξ satisfying (F.4), is continuous and satisfies

$$\mathfrak{R}^{(n)}(z,\xi) = 1 + O_{u.c.}\left(n^{-\frac{1}{2}+3\delta}\right). \tag{A.18}$$

We refer to Appendix A.2.1 for a detailed proof of (A.18). Lemma A.10 then guarantees that for every $z \in [\zeta_0, \zeta_1]$ and ξ satisfying (F.4)

$$n^{\frac{1}{2}-3\delta}\left|\mathfrak{R}^{(n)}(z,\xi) - 1\right| \leq S(z,\xi),$$

where $S(z,\xi) := \sup_{n\in\mathbb{N}}\left|\mathfrak{R}^{(n)}(z,\xi) - 1\right|$ is continuous, hence bounded on compact sets. This allows to conclude that there exists a constant $\tilde{c}$, depending on ζ_0, ζ_1, α, λ and $\mathcal{C}$ for which $\left|\mathfrak{R}^{(n)}(z,\xi) - 1\right| \leq \tilde{c}\,n^{-\frac{1}{2}+3\delta}$. This concludes the proof. ■

A.2.0.4 Step 4

Fix ζ_0 and ζ_1 such that $0 < \zeta_0 < z_0 < \zeta_1 < +\infty$, with the same z_0 as in Proposition 1. Then, combine Lemma A.13 with the well-known inequality [78, Chapter V, Theorem 2]

to obtain that

$$\begin{aligned}\left\|F_{W_n(z)}-\Phi\right\|_\infty &\leq \int_{|\xi|\leq \mathcal{C}\,\sigma(z)\,n^\delta}\left|\frac{\varphi_{W_n(z)}(\xi)-e^{-\frac{\xi^2}{2}}}{\xi}\right|\mathrm{d}\xi+\tilde{\mathcal{C}}n^{-\delta}\\ &\leq \int_{-\frac{1}{n}}^{\frac{1}{n}}\left|\frac{\varphi_{W_n(z)}(\xi)-e^{-\frac{\xi^2}{2}}}{\xi}\right|\mathrm{d}\xi+\frac{2\tilde{c}}{n^{-3\delta+\frac{1}{2}}}\int_{\frac{1}{n}}^{+\infty}\frac{e^{-\frac{\xi^2}{2}}}{\xi}\,\mathrm{d}\xi+\tilde{\mathcal{C}}n^{-\delta}\\ &=: I_1+I_2+\tilde{\mathcal{C}}n^{-\delta}\end{aligned}$$

hold for any $z\in[\zeta_0,\zeta_1]$, with $\tilde{\mathcal{C}}=\max_{z\in[\zeta_0,\zeta_1]}[\mathcal{C}\sigma(z)]^{-1}$.

To bound I_1, combine the triangle inequality, [78, Chapter IV, Lemma 5], [20, Section 8.4, Theorem 1, Equation (4)], and the elementary inequality $e^{-x}\geq 1-x$, to write

$$\begin{aligned}\left|\frac{\varphi_{W_n(z)}(\xi)-e^{-\frac{\xi^2}{2}}}{\xi}\right| &\leq \left|\frac{\varphi_{W_n(z)}(\xi)-1}{\xi}\right|+\left|\frac{1-e^{-\frac{\xi^2}{2}}}{\xi}\right|\\ &\leq \left|\mathbb{E}\left[W_n(z)\right]\right|+\frac{1}{2}\mathbb{E}\left[W_n(z)^2\right]|\xi|+\frac{1}{2}|\xi|\,.\end{aligned}$$

The combination of Proposition 11 with Lemma A.10 entails the existence of some positive constant M_1 for which, for any $z\in[\zeta_0,\zeta_1]$,

$$\left|\mathbb{E}\left[W_n(z)\right]\right|\leq\frac{\mathrm{Var}(R_n(z))}{n\sigma^2(z)}\leq M_1\ .$$

Analogously, there exists a positive constant M_1 for which, for any $z\in[\zeta_0,\zeta_1]$,

$$\mathbb{E}\left[W_n^2(z)\right]\leq M_2\ .$$

Whence,

$$I_1\leq\frac{2M_1}{n}+\frac{M_2+1}{2n^2}.$$

To bound I_2, notice that

$$I_2=\frac{\tilde{c}}{n^{-3\delta+\frac{1}{2}}}\,\Gamma\left(0,\frac{1}{2n^2}\right),$$

where $\Gamma(a,x):=\int_x^{+\infty}t^{a-1}e^{-t}\mathrm{d}t$ denotes the incomplete Gamma function. Now, recall that, as $n\to+\infty$

$$\Gamma\left(0,\frac{1}{2n^2}\right)\sim 2\log(n).$$

See, e.g., [69, Equation (8.4.4)] and Equation (6.6.2)]. Whence, there exists some positive constant $\tilde{c}_1$ for which

$$I_2\leq\tilde{c}_1\log(n)\,n^{-\frac{1}{2}+3\delta}.$$

To conclude, for some positive constants C_1,C_2, we can write

$$\left\|F_{W_n(z)}-\Phi\right\|_\infty\leq C_1n^{-1}+C_2\log(n)\,n^{-\frac{1}{2}+3\delta}+\tilde{\mathcal{C}}n^{-\delta}.$$

It is easy to see that for every $\delta\in(0,1/6)$, $\min\left(\delta,-3\delta+\frac{1}{2}\right)\geq\frac{1}{8}$, which produces (2.2.25) by choosing $\delta=1/8$.

A.2.1 Proof of Equation (A.18)

Fix ζ_0 and ζ_1 such that $0 < \zeta_0 < z_0 < \zeta_1 < +\infty$, with the same z_0 as in Proposition 1, and $z \in [\zeta_0, \zeta_1]$. Set $\tau(zs) := \xi(1/(\alpha z s))$, with the same function ξ as in Lemma A.8. Then, there exists some positive $\tilde{\delta}$, depending only on ζ_0, ζ_1, for which the map $s \mapsto \tau(zs)$ is holomorphic in the disc $|s-1| \leq \tilde{\delta}$. Preliminarily, note that if $s = e^{\frac{i\xi}{\sqrt{n}\sigma(z)}}$ with ξ satisfying (F.4), [78, Chapter IV, Lemma 5] entails

$$|s-1| = \left| e^{\frac{i\xi}{\sqrt{n}\sigma(z)}} - 1 \right| \leq \left| \frac{\xi}{\sqrt{n}\sigma(z)} \right| \leq \mathcal{C}\, n^{\delta - \frac{1}{2}}, \tag{A.19}$$

so that there exists $\bar{n} \in \mathbb{N}$ such that for every $n \geq \bar{n}$,

$$\left| e^{\frac{i\xi}{\sqrt{n}\sigma(z)}} - 1 \right| \leq \mathcal{C} n^{\delta - \frac{1}{2}} \leq \tilde{\delta}.$$

Concerning the behaviour of $\mathfrak{R}_1^{(n)}$, write

$$\mathfrak{R}_1^{(n)}(z,s) = \frac{1+\mathfrak{H}(s)}{1+\mathfrak{H}(1)} = 1 + [\mathfrak{H}(s) - \mathfrak{H}(1)]\,[1 - \mathfrak{H}(1)] + \frac{[\mathfrak{H}(s) - \mathfrak{H}(1)]\,\mathfrak{H}^2(1)}{1+\mathfrak{H}(1)},$$

with

$$\mathfrak{H}(s) := \frac{R_n(zs)}{\chi_\alpha\, J_{0,0}^{(n)}(zs)}.$$

The combination of (A.17), Lemma A.7 and Lemma A.9 yields, after straightforward computations,

$$\begin{aligned} |\mathfrak{H}(1)| &\leq \frac{C_\alpha\, J_{0,1}^{(n)}(z)}{\chi_\alpha\, J_{0,0}^{(n)}(z)} \\ &= \frac{1}{n} \frac{C_\alpha\, \alpha^{\frac{\alpha}{1-\alpha}}}{\chi_\alpha} \frac{\tau(z)}{z} \frac{1 + \frac{1}{n}\mathcal{R}_{-\frac{1}{2}}(z) + r_{-\frac{1}{2}}^{(n)}(z)}{1 + \frac{1}{n}\mathcal{R}_{\frac{1}{2}}(z) + r_{\frac{1}{2}}^{(n)}(z)} \\ &= \frac{1}{n}\phi(z)\,[1 + r_n]\,, \end{aligned}$$

where $\phi(z) := C_\alpha\, \alpha^{\frac{\alpha}{1-\alpha}} \chi_\alpha^{-1}\, [\tau(z)/z]$ is continuous and strictly positive on $[\zeta_0, \zeta_1]$. Moreover, from Lemma A.9,

$$n r_n := n \left[\frac{1 + \frac{1}{n}\mathcal{R}_{-\frac{1}{2}}(z) + r_{-\frac{1}{2}}^{(n)}(z)}{1 + \frac{1}{n}\mathcal{R}_{\frac{1}{2}}(z) + r_{\frac{1}{2}}^{(n)}(z)} - 1 \right]$$

converges uniformly to some continuous function on $[\zeta_0, \zeta_1]$. Then, we can conclude that

$$\mathfrak{H}(1) = O_{u.c.}\left(\frac{1}{n}\right)$$

and, by the same argument,

$$\frac{\mathfrak{H}^2(1)}{1+\mathfrak{H}(1)} = O_{u.c.}\left(\frac{1}{n^2}\right).$$

Now, by the fundamental theorem of calculus,

$$\mathfrak{H}(s) - \mathfrak{H}(1) = (s-1) \int_0^1 \mathfrak{H}'(1 + t(s-1))\, \mathrm{d}t.$$

It can be easily seen that $s \mapsto \mathfrak{H}(s)$ is holomorphic in $\{w \in \mathbb{C} : Re(w) > 0\}$. The combination of this fact with the bound (A.19) entails that there exists a continuous function $G : [\zeta_0, \zeta_1] \to (0, +\infty)$, not depending on s and n, for which

$$|\mathfrak{H}(s) - \mathfrak{H}(1)| \leq G(z)\, \mathcal{C}\, n^{\delta - \frac{1}{2}}.$$

Whence,

$$\mathfrak{R}_1^{(n)}(z,s) = 1 + O_{u.c.}\left(n^{\delta - \frac{1}{2}}\right).$$

Concerning the behaviour of $\mathfrak{R}_2^{(n)}$, note that $\mathfrak{R}_2^{(n)}(z,s)$ is the product of two terms. As to the former, apply Lemma A.9 to obtain

$$\frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} = \frac{1}{1 + \frac{1}{n}\mathcal{R}_{\frac{1}{2}}(z) + r_{\frac{1}{2}}^{(n)}(z)} =: \frac{1}{1+\varepsilon(z)} = 1 - \varepsilon(z) + \frac{\varepsilon(z)^2}{1+\varepsilon(z)}.$$

Whence,

$$\left|\frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} - 1\right| \leq \left|\frac{1}{n}\mathcal{R}_{\frac{1}{2}}(z) + r^{(n)}(z)\right| + \left|\frac{\varepsilon(z)^2}{1+\varepsilon(z)}\right|.$$

For the latter, Lemma A.9 again shows that

$$\left|\frac{\mathcal{F}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)} - 1\right| = \left|\frac{1}{n}\mathcal{R}_{\frac{1}{2}}(zs) + r^{(n)}(zs)\right|.$$

If $|s-1| < \tilde{\delta}$, we conclude that

$$\frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} = 1 + O_{u.c.}\left(\frac{1}{n}\right) \quad \text{and} \quad \frac{\mathcal{F}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)} = 1 + O_{u.c.}\left(\frac{1}{n}\right).$$

Whence,

$$\mathfrak{R}_2^{(n)}(z,s) = \frac{\mathcal{P}_{\frac{1}{2}}^{(n)}(z)}{\mathcal{F}_{\frac{1}{2}}^{(n)}(z)} \frac{\mathcal{F}_{\frac{1}{2}}^{(n)}(zs)}{\mathcal{P}_{\frac{1}{2}}^{(n)}(zs)} = 1 + O_{u.c.}\left(\frac{1}{n}\right).$$

Concerning the behaviour of $\mathfrak{R}_3^{(n)}$, rewrite $\mathfrak{R}_3^{(n)}(z,s)$ as

$$\mathfrak{R}_3^{(n)}(z,s) = s^{\frac{1}{2}} \left(1 + \frac{\alpha z(s-1) + (1-\alpha)\left[\tau(zs) - \tau(z)\right]}{\alpha z + (1-\alpha)\tau(z)}\right)^{-\frac{1}{2}} =: s^{\frac{1}{2}}\, (1 + H_z(s))^{-\frac{1}{2}}.$$

Now, if $z \in [\zeta_0, \zeta_1]$ and $|s-1| < \tilde{\delta}$, there exists some positive constant M_3 for which

$$\left|(1 + H_z(s))^{-\frac{1}{2}} - 1\right| \leq M_3 \left|H_z(s)\right|.$$

Moreover, Lemma A.8 entails

$$|\tau(zs)-\tau(z)| \le z\int_1^s \left|\frac{\mathrm{d}}{\mathrm{d}s}\tau(zs)\right| \le C_1(z)\,|s-1|$$

for some continuous function C_1 of z. Whence,

$$|H_z(s)| \le \left[\frac{\alpha + C_1(z)}{\alpha z + (1-\alpha)\tau(z)}\right] |s-1| =: C_2(z)\,|s-1|$$

where the function C_2 is continuous. Thus, if s fulfills (A.19), we get

$$|H_z(s)| \le C_2(z)\,\mathcal{C}\,n^{\delta-\frac{1}{2}}$$

and, consequently,

$$\mathfrak{R}_3^{(n)}(z,s) = s^{\frac{1}{2}}\,(1+H_z(s))^{-\frac{1}{2}} = 1 + O_{u.c.}\left(n^{\delta-\frac{1}{2}}\right).$$

Concerning the behaviour of $\mathfrak{R}_4^{(n)}$, elementary properties of the exponential combined with [78, Chapter IV, Lemma 5] yield

$$\mathfrak{R}_4^{(n)}(z,\xi) := \exp\left\{ nz\left[1-g'(z)\right]\frac{C\,\xi^3}{6\,n^{\frac{3}{2}}\sigma(z)^3} + \frac{1}{2}\,nz^2 g''(z)\,\frac{C\,\xi^3}{n^{\frac{3}{2}}\sigma(z)^3} \right.$$
$$\left. + \frac{1}{2}nz^3\,\frac{C^3\,\xi^3}{n^{\frac{3}{2}}\sigma(z)^3}\int_0^1 g'''\left(z\left[1+t\left(e^{\frac{i\,\xi}{\sqrt{n}\sigma(z)}}-1\right)\right]\right)(1-t)^2\,\mathrm{d}t\right\}$$

for a suitable $C \in \mathbb{C}$ with $|C|<1$. Note that, if $|s-1|<\tilde{\delta}$, it also holds that $|1+t(s-1)|<\tilde{\delta}$ for all $t\in[0,1]$. Moreover, the function $t \mapsto g'''(z[1+t(s-1)])$ is bounded on $[0,1]$, if $z \in [\zeta_0,\zeta_1]$. Therefore, we get

$$\left|\frac{1}{6}z\left[1-g'(z)\right] + \frac{1}{2}z^2 g''(z) + \frac{1}{2}z^3\int_0^1 g'''(z[1+t(s-1)])(1-t)^2\,\mathrm{d}t\right| \le G(z)$$

for some continuous function $G:[\zeta_0,\zeta_1]\to(0,+\infty)$ not depending on ξ and n. To conclude, apply again [78, Chapter IV, Lemma 5] to show that, for any ξ satisfying (F.4), it holds

$$\left|\mathfrak{R}_4^{(n)}(z,\xi)-1\right| \le G(z)\frac{|\xi|^3}{n^{\frac{1}{2}}\sigma(z)^3} \le G(z)\,\mathcal{C}^3\,n^{3\delta-\frac{1}{2}}.$$

Whence,

$$\mathfrak{R}_4^{(n)}(z,\xi) = 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right).$$

Altogether, we have proved that for $z\in[\zeta_0,\zeta_1]$, $s = e^{\frac{i\xi}{\sqrt{n}\sigma(z)}}$ with ξ satisfying (F.4)

$$\mathfrak{R}_1^{(n)}(z,s) = 1 + O_{u.c.}\left(n^{\delta-\frac{1}{2}}\right)$$
$$\mathfrak{R}_2^{(n)}(z,s) = 1 + O_{u.c.}\left(n^{-1}\right)$$
$$\mathfrak{R}_3^{(n)}(z,s) = 1 + O_{u.c.}\left(n^{\delta-\frac{1}{2}}\right)$$
$$\mathfrak{R}_4^{(n)}(z,\xi) = 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right).$$

Recalling that $\mathfrak{R}^{(n)}(z,\xi)$ is defined as the product of these four terms, and since $0 > 3\delta-\frac{1}{2} > \delta-\frac{1}{2} > -1$ for $\delta\in\left(0,\frac{1}{6}\right)$, we conclude that $\mathfrak{R}^{(n)}(z,\xi) = 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right)$.

B Details on the proof of (2.1.4) and (2.1.6) for $\alpha = 0$

B.1 Proof of Equation (2.3.5)

Preliminary note that if $s = e^{\frac{i\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}(z)}}$ with ξ satisfying (2.3.2), then [78, Chapter IV, Lemma 5] entails

$$|s-1| = \left| e^{\frac{i\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}(z)}} - 1 \right| \leq \left| \frac{\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}(z)} \right| \leq \mathcal{C}\, n^{\delta - \frac{1}{2}}.$$

In particular, there exists $\bar{n} \in \mathbb{N}$ such that, for every $n \geq \bar{n}$,

$$\left| \frac{\xi}{\sqrt{n}\mathfrak{s}_{0,\lambda}(z)} \right| \leq \frac{\pi}{3}.$$

From now on, we assume $n \geq \bar{n}$. Concerning the behaviour of $\mathfrak{R}_1^{(n)}$, rewrite $\mathfrak{R}_1^{(n)}(s)$ as

$$\mathfrak{R}_1^{(n)}(s) := 1 + \mathfrak{R}(n(s\lambda+1)) + \mathfrak{R}(n\lambda) - \mathfrak{R}(ns\lambda) - \mathfrak{R}(n(\lambda+1)) + \mathfrak{E}(s),$$

where

$$\mathfrak{E}(s) = \mathfrak{R}(n(s\lambda+1))\mathfrak{R}(n\lambda) - [\mathfrak{R}(n(s\lambda+1)) + \mathfrak{R}(n\lambda)]\, \varepsilon(s) + \frac{\varepsilon(s)^2}{1+\varepsilon(s)}$$

with

$$\varepsilon(s) := \mathfrak{R}(ns\lambda) + \mathfrak{R}(n(\lambda+1)) + \mathfrak{R}(ns\lambda)\mathfrak{R}(n(\lambda+1)).$$

From (2.3.3), it holds

$$|\mathfrak{R}(w)| \leq \frac{3}{2\pi^2\,|w|} \leq \frac{3}{2\pi^2\, n\lambda}$$

for every $w \in \{n(s\lambda+1), ns\lambda, n(\lambda+1), n\lambda\}$, since $|w| \geq n\lambda$. This implies

$$|\mathfrak{R}_1^{(n)}(s) - 1| \leq \frac{1}{n}\,\frac{6}{\pi^2\lambda} + |\mathfrak{E}(s)| \quad \text{and} \quad |\mathfrak{E}(s)| = O_{u.c.}\left(\frac{1}{n^2}\right).$$

Thus,

$$\mathfrak{R}_1^{(n)}(s) = 1 + O_{u.c.}\left(n^{-1}\right).$$

Concerning the behaviour of $\mathfrak{R}_2^{(n)}$, rewrite $\mathfrak{R}_2^{(n)}$ as

$$\mathfrak{R}_2^{(n)}(s) = \left(1 + \frac{s-1}{s\lambda+1}\right)^{\frac{1}{2}}.$$

Then,

$$\begin{aligned}\left|\mathfrak{R}_2^{(n)}(s) - 1\right| &= \left| \left(1 + \frac{s-1}{s\lambda+1}\right)^{\frac{1}{2}} - 1 \right| \\ &= \left| \frac{s-1}{s\lambda+1} \right| \sum_{k=0}^{+\infty} \binom{\frac{1}{2}}{k+1} \left[\frac{s-1}{s\lambda+1}\right]^k \leq \left| \frac{s-1}{s\lambda+1} \right| C_\varepsilon(s),\end{aligned}$$

where the last identity holds if $|\frac{s-1}{s\lambda+1}| \leq 1-\varepsilon$, for C_ε a continuous function only depending on $\varepsilon > 0$. Since, for ξ satisfying (2.3.2), $|s\lambda+1| > \lambda$, there exists $\bar{n} \in \mathbb{N}$ such that for every $n \geq \bar{n}$

$$\left| \frac{s-1}{s\lambda+1} \right| \leq \frac{\mathcal{C}}{\lambda} n^{\delta - \frac{1}{2}} \leq 1 - \varepsilon$$

and

$$\left|\mathfrak{R}_2^{(n)}(s) - 1\right| \leq C_\varepsilon(s)\,\frac{\mathcal{C}}{\lambda}\,n^{\delta-\frac{1}{2}}.$$

We conclude that

$$\mathfrak{R}_2^{(n)}(s) = 1 + O_{u.c.}\left(n^{\delta-\frac{1}{2}}\right).$$

Concerning the behaviour of $\mathfrak{R}_3^{(n)}$, upon noting that the function f is holomorphic in the disc of center 1 and radius 1, argue as for $\mathfrak{R}_4^{(n)}$ of the case $\alpha \in (0,1)$ to conclude that

$$\mathfrak{R}_3^{(n)}(s) = 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right).$$

Altogether, we proved that, for $s = e^{\frac{i\xi}{\sqrt{n}s_{0,\lambda}(z)}}$ with ξ satisfying (2.3.2) and for every $n \geq \bar{n}$,

$$\begin{aligned}
\mathfrak{R}_1^{(n)}(s) &= 1 + O_{u.c.}\left(n^{-1}\right)\\
\mathfrak{R}_2^{(n)}(s) &= 1 + O_{u.c.}\left(n^{\delta-\frac{1}{2}}\right)\\
\mathfrak{R}_3^{(n)}(s) &= 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right).
\end{aligned}$$

Recalling that $\mathfrak{R}^{(n)}(\xi)$ is defined as the product of these three terms, and since $0 > 3\delta-\frac{1}{2} > \delta-\frac{1}{2} > -1$ for any $\delta \in \left(0,\frac{1}{6}\right)$, we conclude that $\mathfrak{R}^{(n)}(z,\xi) = 1 + O_{u.c.}\left(n^{3\delta-\frac{1}{2}}\right)$.

C Details on the proof of the SLLN (1.2.9)

C.1 Proof of Equation (2.4.1) for $\alpha \in (0,1)$

By standard combinatorial relations,

$$\mathbb{E}\left[\left(K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right)^4\right] = \sum_{r=0}^{4}(-1)^{4-r}\binom{4}{r}\mathbb{E}\left[K_n^{\{n\}}\right]^{4-r}\sum_{j=1}^{r}S(r,j)\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow j}\right] \quad \text{(C.1)}$$

where $S(r,j)$ denotes the Stirling number of the second kind, and $(x)_{\downarrow j} := \prod_{i=0}^{j-1}(x-i)$ is the falling factorial of x. Now, rewrite the explicit expression of the falling factorial moments of $K_n^{\{n\}}$ already given in section 2.2.1 as

$$\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow j}\right] = \frac{\Gamma(\theta/\alpha+j)}{\Gamma(\theta/\alpha)}\sum_{i=0}^{r}(-1)^{j-i}\binom{j}{i}\frac{\Gamma(\theta+i\alpha+n)}{\Gamma(\theta+i\alpha)}\frac{\Gamma(\theta)}{\Gamma(\theta+n)}.$$

Setting $\theta = \lambda n$, resort again to [89, Equation 1] to obtain

$$\begin{aligned}
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 1}\right] &= n \cdot \frac{\lambda}{\alpha}(L-1) - L\frac{\alpha-1}{2(\lambda+1)} + O\left(n^{-1}\right)\\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] &= n^2 \cdot \frac{\lambda^2}{\alpha^2}(L-1)^2\\
&\quad + n \cdot \frac{\lambda}{\alpha}\left[(L-1)^2 + L\frac{\alpha-1}{\lambda+1} - L^2\frac{2\alpha-1}{\lambda+1}\right] + O(1)\\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] &= n^3 \cdot \frac{\lambda^3}{\alpha^3}(L-1)^3\\
&\quad + n^2 \cdot 3\frac{\lambda^2}{\alpha^2}\left[(L-1)^3 - L\frac{\alpha-1}{2(\lambda+1)} + L^2\frac{2\alpha-1}{\lambda+1} - L^3\frac{3\alpha-1}{2(\lambda+1)}\right]\\
&\quad + O(n)\\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 4}\right] &= n^4 \cdot \frac{\lambda^4}{\alpha^4}(L-1)^4\\
&\quad + n^3 \cdot 2\frac{\lambda^3}{\alpha^3}\left[3(L-1)^4 + L\frac{\alpha-1}{\lambda+1} - 3L^2\frac{2\alpha-1}{\lambda+1} + 3L^3\frac{3\alpha-1}{2(\lambda+1)}\right.\\
&\quad \left. - L^4\frac{4\alpha-1}{\lambda+1}\right] + O\left(n^2\right)
\end{aligned}$$

where $L := \left(\frac{\lambda+1}{\lambda}\right)^{\alpha}$. Since $\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 1}\right] = \mathbb{E}\left[K_n^{\{n\}}\right]$, the above identities lead to

$$\begin{aligned}
\mathbb{E}\left[K_n^{\{n\}}\right]^2 &= n^2 \cdot \frac{\lambda^2}{\alpha^2}(L-1)^2 - n \cdot \frac{\lambda}{\alpha}(L-1)L\frac{\alpha-1}{\lambda+1} + O(1)\\
\mathbb{E}\left[K_n^{\{n\}}\right]^3 &= n^3 \cdot \frac{\lambda^3}{\alpha^3}(L-1)^3 - n^2 \cdot 3\frac{\lambda^2}{\alpha^2}(L-1)^2 L\frac{\alpha-1}{2(\lambda+1)} + O(n)\\
\mathbb{E}\left[K_n^{\{n\}}\right]^4 &= n^3 \cdot \frac{\lambda^3}{\alpha^3}(L-1)^3\\
&\quad + n^2 \cdot 3\frac{\lambda^2}{\alpha^2}\left[(L-1)^3 - L\frac{\alpha-1}{2(\lambda+1)} + L^2\frac{2\alpha-1}{\lambda+1} - L^3\frac{3\alpha-1}{2(\lambda+1)}\right]\\
&\quad + O(n).
\end{aligned}$$

Therefore, (C.1) reduces to

$$\begin{aligned}
&\mathbb{E}\left[\left(K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right)^4\right]\\
&= -3\mathbb{E}\left[K_n^{\{n\}}\right]^4 + 6\mathbb{E}\left[K_n^{\{n\}}\right]^3 + 6\mathbb{E}\left[K_n^{\{n\}}\right]^2\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] - 12\mathbb{E}\left[K_n^{\{n\}}\right]\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right]\\
&\quad - 4\mathbb{E}\left[K_n^{\{n\}}\right]\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] + 6\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] + \mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 4}\right]\\
&= n^4 \cdot \mathfrak{A}(\alpha,\lambda) + n^3 \cdot \mathfrak{B}(\alpha,\lambda) + O\left(n^2\right)
\end{aligned}$$

with

$$\mathfrak{A}(\alpha,\lambda) = \frac{\lambda^4}{\alpha^4}(L-1)^4(-3+6-4+1)$$

and

$$\begin{aligned}\mathfrak{B}(\alpha,\lambda) = 2\,\frac{\lambda^3}{\alpha^3}\Bigg\{ & 3(L-1)^3 L\,\frac{\alpha-1}{\lambda+1} + 3(L-1)^3 \\ & +3(L-1)^2\left[(L-1)^2 + L\,\frac{\alpha-1}{\lambda+1} - L^2\,\frac{2\alpha-1}{\lambda+1} - L(L-1)\,\frac{\alpha-1}{\lambda+1}\right] \\ & -6(L-1)^3 + (L-1)^3 L\,\frac{\alpha-1}{\lambda+1} + 3(L-1)^3 \\ & -3(L-1)\left[2(L-1)^3 - L\,\frac{\alpha-1}{\lambda+1} + 2L^2\,\frac{2\alpha-1}{\lambda+1} - L^3\,\frac{3\alpha-1}{\lambda+1}\right] \\ & +3(L-1)^4 + L\,\frac{\alpha-1}{\lambda+1} - 3L^2\,\frac{2\alpha-1}{\lambda+1} + 3L^3\,\frac{3\alpha-1}{\lambda+1} - L^4\,\frac{4\alpha-1}{\lambda+1}\Bigg\}.\end{aligned}$$

Finally, straightforward algebraic computations show that

$$\mathfrak{A}(\alpha,\lambda) = \mathfrak{B}(\alpha,\lambda) = 0,$$

proving equation (2.4.1).

C.2 Proof of equation (2.4.1) for $\alpha = 0$

Arguing as in Appendix C.1, one gets

$$\begin{aligned}&\mathbb{E}\left[\left(K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right)^4\right] \\ &\quad = -3\mathbb{E}\left[K_n^{\{n\}}\right]^4 + 6\mathbb{E}\left[K_n^{\{n\}}\right]^3 + 6\mathbb{E}\left[K_n^{\{n\}}\right]^2\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] - 12\mathbb{E}\left[K_n^{\{n\}}\right]\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] \\ &\qquad - 4\mathbb{E}\left[K_n^{\{n\}}\right]\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] + 6\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] + \mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 4}\right].\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 1}\right] &= n\cdot\lambda\,\Phi_0(n,\lambda) \\ \mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] &= n^2\cdot\lambda^2\,\left[\Phi_0^2(n,\lambda) + \Phi_1(n,\lambda)\right] \\ \mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] &= n^3\cdot\lambda^3\,\left[\Phi_0^3(n,\lambda) + 3\Phi_0(n,\lambda)\Phi_1(n,\lambda) + \Phi_2(n,\lambda)\right] \\ \mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 4}\right] &= n^4\cdot\lambda^4\,\left[\Phi_0^4(n,\lambda) + 6\Phi_0^2(n,\lambda)\Phi_1(n,\lambda) + 4\Phi_0(n,\lambda)\Phi_2(n,\lambda)\right. \\ &\qquad \left. +3\Phi_1^2(n,\lambda) + \Phi_3(n,\lambda)\right]\end{aligned}$$

where

$$\Phi_i(n,\lambda) := \psi^{(i)}\left(n(\lambda+1)\right) - \psi^{(i)}\left(n\lambda\right)$$

and $\psi^{(i)}$ denotes the polygamma function [69, Section (5.15)]. Making use of the asymptotic expansions of the polygamma functions for large argument [69, Equation (5.11.2)], one can write

$$\begin{aligned}\Phi_0(n,\lambda) &= L + \frac{1}{2\lambda(\lambda+1)n} + O\left(n^{-2}\right) \\ \Phi_1(n,\lambda) &= -\frac{1}{\lambda(\lambda+1)n} + O\left(n^{-2}\right) \\ \Phi_2(n,\lambda) &= O\left(n^{-2}\right) \\ \Phi_3(n,\lambda) &= O\left(n^{-3}\right)\end{aligned}$$

where $L = \log\left(\frac{\lambda+1}{\lambda}\right)$. Whence,

$$\begin{aligned}
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 1}\right] &= n \cdot \lambda L + \frac{1}{2(\lambda+1)} + O\left(n^{-1}\right) \\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 2}\right] &= n^2 \cdot \lambda^2 L^2 + n \cdot \frac{\lambda}{\lambda+1}(L-1) + O(1) \\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 3}\right] &= n^3 \cdot \lambda^3 L^3 + n^2 \cdot 3\frac{\lambda^2}{\lambda+1}\left(\frac{1}{2}L^2 - L\right) + O(n) \\
\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 4}\right] &= n^4 \cdot \lambda^4 L^4 + n^3 \cdot 2\frac{\lambda^3}{\lambda+1}(L^3 - 3L^2) + O\left(n^2\right).
\end{aligned}$$

Since $\mathbb{E}\left[(K_n^{\{n\}})_{\downarrow 1}\right] = \mathbb{E}\left[K_n^{\{n\}}\right]$, this also gives

$$\begin{aligned}
\mathbb{E}\left[K_n^{\{n\}}\right]^2 &= n^2 \cdot \lambda^2 L^2 + n \cdot L\frac{\lambda}{\lambda+1} + O(1) \\
\mathbb{E}\left[K_n^{\{n\}}\right]^3 &= n^3 \cdot \lambda^3 L^3 + n^2 \cdot 3L^2\frac{\lambda^2}{2(\lambda+1)} + O(n) \\
\mathbb{E}\left[K_n^{\{n\}}\right]^4 &= n^4 \cdot \lambda^4 L^4 + n^3 \cdot 2L^3\frac{\lambda^3}{\lambda+1} + O\left(n^2\right).
\end{aligned}$$

It follows that

$$\mathbb{E}\left[\left(K_n^{\{n\}} - \mathbb{E}\left[K_n^{\{n\}}\right]\right)^4\right] = n^4 \cdot \mathfrak{A}(\lambda) + n^3 \cdot \mathfrak{B}(\lambda) + O\left(n^2\right)$$

with

$$\mathfrak{A}(\lambda) = \lambda^4 L^4 \left(-3 + 6 - 4 + 1\right)$$

and

$$\begin{aligned}
\mathfrak{B}(\lambda) = 2\,\lambda^3 L \Bigg\{ &-3\frac{L^2}{\lambda+1} + 3L^2 + 3L\left[\frac{2}{\lambda+1}L - \frac{1}{\lambda+1}\right] - 6L^2 \\
&-2\left[L^2\frac{1}{2(\lambda+1)} + \frac{3}{\lambda+1}\left(\frac{L^2}{2} - L\right)\right] + 3L^3 + \frac{1}{\lambda+1}(L^2 - 3L)\Bigg\}
\end{aligned}$$

Straightforward algebraic computations show that

$$\mathfrak{A}(\lambda) = \mathfrak{B}(\lambda) = 0,$$

proving equation (2.4.1).

D Proofs for Chapter 3

D.1 Details on the proof of lemma 3.1.1

D.1.1 $\alpha \in (0,1)$

Based on the distribution of K_n [80, Equation 3.11], a direct calculation shows that

$$\mathbb{E}\left[(K_h^{\{n\}})_{\downarrow s}\right] = \frac{\Gamma(\lambda n/\alpha + s)}{\Gamma(\lambda n/\alpha)} \sum_{i=0}^{s} (-1)^{s-i} \binom{s}{i} \frac{\Gamma(\theta + i\alpha + h)}{\Gamma(\lambda n + i\alpha)} \frac{\Gamma(\lambda n)}{\Gamma(\lambda n + h)}$$

so that

$$\mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow s}\right] = \frac{\Gamma\left(\frac{\lambda n}{\alpha}+s\right)}{\Gamma\left(\frac{\lambda n}{\alpha}\right)}\sum_{k=0}^{s}(-1)^{s-k}\binom{s}{k}\frac{\Gamma((\lambda+x)n+k\alpha-\mathfrak{r}_x)\,\Gamma(\lambda n)}{\Gamma(\lambda n+k\alpha)\,\Gamma((\lambda+x)n-\mathfrak{r}_x)}$$

where we are denoting by $\mathfrak{r}_x = xn - \lfloor xn\rfloor$. Since $0 \le \mathfrak{r}_x < 1$ for all $x \in [0,1]$ and for all $n \in \mathbb{N}$, we can apply the asymptotic expansion for ratios of gamma functions [89, Equation 1] to obtain, after calculations analogous to those of Appendix C,

$$\lim_{n\to\infty}\sup_{x\in[0,1]}\frac{1}{n^{s-2}}\cdot\left|\mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow s}\right] - \left(n^s\cdot(m_{\alpha,\lambda}(x))^s + n^{s-1}\cdot S_s(x)\right)\right| = c$$

for some constant c, with

$$S_s(x) = \begin{cases} \frac{(1-\alpha)x}{2(\lambda+x)}\left(\frac{\lambda+x}{\lambda}\right)^{\alpha} & \text{if } s=1 \\ \frac{s(s-1)\alpha}{\lambda}\,m^s_{\alpha,\lambda}(x) - \left(\frac{\lambda}{\alpha}\right)^2\frac{\alpha x s}{2\lambda(\lambda+x)}\left(\frac{\lambda+x}{\lambda}\right)^{\alpha}\times & \\ \quad\times\left[\left(\frac{\lambda+x}{\lambda}\right)^{\alpha}(\alpha s-1)+1-\alpha\right]m^{s-2}_{\alpha,\lambda}(x) & \text{if } s\ge 2. \end{cases}$$

It remains to show that $\mathcal{B}_{\alpha,\lambda}(x) = 0$:

$$\begin{aligned}\mathcal{B}_{\alpha,\lambda}(x) &= m^3_{\alpha,\lambda}(x)\left[(-12+12-4)S_1(x)+(6-12+6)\right]+6\,m^2_{\alpha,\lambda}(x)\,S_2(x)\\ &\quad + m_{\alpha,\lambda}(x)\,S_3(x)+S_4(x)\\ &= -4\,m^3_{\alpha,\lambda}(x)S_1(x)+6\,m^2_{\alpha,\lambda}(x)\,S_2(x)-4m_{\alpha,\lambda}(x)\,S_3(x)+S_4(x)\\ &= m^3_{\alpha,\lambda}(x)\cdot\left[-\frac{2(\alpha-1)}{\lambda+x}\left(\frac{\lambda+x}{\lambda}\right)^{\alpha}\right]\\ &\quad + m^2_{\alpha,\lambda}(x)\cdot\frac{x\lambda}{x+\lambda}\left[\left(\frac{\lambda+x}{\lambda}\right)^{2\alpha}\left(6-\frac{2(4\alpha-1)}{\alpha}\right)-\left(\frac{\lambda+x}{\lambda}\right)^{\alpha}\frac{2(1-\alpha)}{\alpha}\right]\\ &= 0\end{aligned}$$

Where the last equality can be easily verified by substituting the expression for $m_{\alpha,\lambda}(x)$.

D.1.2 $\alpha = 0$

Arguing as in appendix C, we obtain

$$\begin{aligned}\mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow 1}\right] &= n\cdot\lambda\,\Phi_0(n,\lambda,x)\\ \mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow 2}\right] &= n^2\cdot\lambda^2\,\left[\Phi^2_0(n,\lambda,x)+\Phi_1(n,\lambda,x)\right]\\ \mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow 3}\right] &= n^3\cdot\lambda^3\,\left[\Phi^3_0(n,\lambda,x)+3\Phi_0(n,\lambda,x)\Phi_1(n,\lambda,x)+\Phi_2(n,\lambda,x)\right]\\ \mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow 4}\right] &= n^4\cdot\lambda^4\,\left[\Phi^4_0(n,\lambda,x)+6\Phi^2_0(n,\lambda,x)\Phi_1(n,\lambda,x)+4\Phi_0(n,\lambda,x)\Phi_2(n,\lambda,x)\right.\\ &\qquad\left.+3\Phi^2_1(n,\lambda,x)+\Phi_3(n,\lambda,x)\right]\end{aligned}$$

where

$$\Phi_i(n,\lambda,x) := \psi^{(i)}\left(n(\lambda+x)\right)-\psi^{(i)}\left(n\lambda\right)$$

and $\psi^{(i)}$ denotes the polygamma function [69, Section (5.15)].

Making use of the asymptotic expansions of the polygamma functions for large argument [69, Equation (5.11.2)], one can write

$$\lim_{n\to\infty}\sup_{x\in[0,1]}\frac{1}{n^{s-2}}\cdot\left|\mathbb{E}\left[(K^{\{n\}}_{\lfloor xn\rfloor})_{\downarrow s}\right]-\left(n^s\cdot(m_{\alpha,\lambda}(x))^s+n^{s-1}\cdot S_s(x)\right)\right|=c$$

for some constant c, with

$$S_s(x)=\begin{cases}\frac{x}{2(\lambda+x)} & \text{if } s=1\\ \frac{x\lambda}{\lambda+x}(L(x)-1) & \text{if } s=2\\ 3\frac{x\lambda^2}{\lambda+x}\left(\frac{1}{2}L^2(x)-L(x)\right) & \text{if } s=3\\ 2\frac{x\lambda^3}{\lambda+x}(L^3(x)-3L^2(x)) & \text{if } s=4\end{cases}$$

where $L(x)=\log\left(\frac{\lambda+x}{\lambda}\right)$. Whence,

$$\begin{aligned}\mathfrak{B}(x)&=-4\lambda^3L^3(x)S_1(x)+6\lambda^2L^2(x)S_2(x)-4\lambda L(x)S_3(x)+S_4(x)\\&=\frac{x\lambda^3}{\lambda+x}\cdot\left[L^3(x)\cdot(-2+6-6+2)+L^2(x)\cdot(-6+12-6)\right]\\&=0\end{aligned}$$

D.2 Details on the proof of Lemma 3.2.4

We begin by proving the following lemma:

Lemma D.1. *As $n\to+\infty$, for all $i\in\{1,...,d+1\}$,*

(i)

$$\lim_{n\to+\infty}a^{\{n\}}_{i-1,\lfloor nx\rfloor+1}=\left(\frac{\lambda+x}{\lambda}\right)^{i-1-\alpha}$$

(ii)

$$p^{\{n\}}_{i-1,\lfloor nx\rfloor}\xrightarrow{a.s.}\begin{cases}\left(\frac{\lambda+x}{\lambda}\right)^{\alpha-1} & \text{if } i=1\\ \frac{(1-\alpha)_{(i-2)\uparrow}\,\lambda^{1-\alpha}}{(i-2)!}\cdot x^{i-2}(x+\lambda)^{\alpha-i-1} & \text{if } i\geq 2\end{cases}$$

(iii)

$$q^{\{n\}}_{i-1,\lfloor nx\rfloor}\xrightarrow{a.s.}\begin{cases}0 & \text{if } i=1\\ \frac{(1-\alpha)_{(i-1)\uparrow}\,\lambda^{1-\alpha}}{(i-1)!}\cdot x^{i-1}(x+\lambda)^{\alpha-i} & \text{if } i\geq 2\end{cases}$$

where convergence in (i), (ii) and (ii) is uniform for $x\in[0,1]$.

Proof. By definition of $a^{\{n\}}_{i-1,\lfloor nx\rfloor+1}$ (3.2.10),

$$\begin{aligned}a^{\{n\}}_{i-1,\lfloor nx\rfloor+1}&=\frac{(\lambda n+1)_{\lfloor nx\rfloor\uparrow}}{(\lambda n-i+2+\alpha)_{\lfloor nx\rfloor\uparrow}}\\&=\frac{\Gamma((\lambda+x)n+1-\mathfrak{r}_x)\cdot\Gamma(\lambda n-i+2+\alpha)}{\Gamma((\lambda+x)n-i+2+\alpha-\mathfrak{r}_x)\cdot\Gamma(\lambda n+1)}\end{aligned}$$

where we are denoting by $\mathfrak{r}_x=xn-\lfloor xn\rfloor$. Since $0\leq\mathfrak{r}_x<1$ for all $x\in[0,1]$ and for all $n\in\mathbb{N}$, we can apply the asymptotic expansion for ratios of gamma functions [89, Equation 1] to obtain the desired asymptotics.

For what concerns $p^{\{n\}}_{r,\lfloor nx\rfloor}$, recall

$$p^{\{n\}}_{r,h-1} = \begin{cases} \frac{\alpha K^{\{n\}}_{\lfloor nx\rfloor}+\lambda n}{\lambda n+\lfloor nx\rfloor} & \text{if } r=0,1 \\ \frac{(r-1-\alpha)K^{\{n\}}_{r-1,\lfloor nx\rfloor}}{\lambda n+\lfloor nx\rfloor} & \text{if } r\geq 2 \end{cases}$$
$$= \begin{cases} \frac{\alpha}{\left(x+\lambda-\frac{\mathfrak{r}_x}{n}\right)}\cdot\frac{K^{\{n\}}_{\lfloor nx\rfloor}}{n}+\frac{\lambda}{\left(x+\lambda-\frac{\mathfrak{r}_x}{n}\right)} & \text{if } r=0,1 \\ \frac{(r-1-\alpha)}{\left(x+\lambda-\frac{\mathfrak{r}_x}{n}\right)}\cdot\frac{K^{\{n\}}_{r-1,\lfloor nx\rfloor}}{n} & \text{if } r\geq 2 \end{cases}$$

since $\sup_x \mathfrak{r}_x/n < 1/n \to 0$, it sufficies to apply lemma 3.2.2 to obtain the desired asymptotics.

Analogously,

$$q^{\{n\}}_{r,h-1} = \begin{cases} 0 & \text{if } r=0 \\ \frac{(r-\alpha)K^{\{n\}}_{r,\lfloor nx\rfloor}}{\lambda n+\lfloor nx\rfloor} & \text{if } r\geq 1 \end{cases} = \begin{cases} 0 & \text{if } r=0 \\ \frac{(r-\alpha)}{\left(x+\lambda-\frac{\mathfrak{r}_x}{n}\right)}\cdot\frac{K^{\{n\}}_{r,\lfloor nx\rfloor}}{n} & \text{if } r\geq 1 \end{cases}$$

and the desired asymptotics follows from lemma 3.2.2. ■

Recalling the definitions of $\left(P^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j}$ and $\left(R^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j}$, lemma D.1 implies

$$\left(P^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j} \xrightarrow{a.s.} \begin{cases} 0 & \text{if } j\geq i+2 \\ -c_{i-1}\,\lambda^{1-\alpha}\,x^{\,i-1}(x+\lambda)^{\,\alpha-i} & \text{if } j=i+1,\ i\geq 2 \\ \left(\frac{x+\lambda}{\lambda}\right)^{\alpha-1} & \text{if } i=j=1 \text{ or } i=1, j=2 \\ c_{i-2}\,\lambda^{1-\alpha}\,x^{\,i-2}(x+\lambda)^{\,\alpha-i-1} & \\ \quad +c_{i-1}\,\lambda^{1-\alpha}\,x^{\,i-1}(x+\lambda)^{\,\alpha-i} & \text{if } i=j\geq 2. \end{cases}$$

and

$$\left(R^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j} \xrightarrow{a.s.} \begin{cases} \left(\frac{x+\lambda}{\lambda}\right)^{2(\alpha-1)} & \text{if } i=j=1, \\ \lambda^{2(1-\alpha)}\left[c_{j-2}\,x^{\,j-2}(x+\lambda)^{\,2\alpha-j-2}-c_{j-1}\,x^{\,j-1}(x+\lambda)^{\,2\alpha-j-1}\right] & \text{if } i=1,\ j\geq 2 \\ \lambda^{2(1-\alpha)}\Big\{c_{i-2}\,c_{j-2}\,x^{\,i+j-4}(x+\lambda)^{\,2\alpha-i-j-2} & \\ \quad -\left[c_{i-1}\,c_{j-2}+c_{i-2}\,c_{j-1}\right]x^{\,i+j-3}(x+\lambda)^{\,2\alpha-i-j-1} & \\ \quad +c_{i-1}\,c_{j-1}\,x^{\,i+j-2}(x+\lambda)^{\,2\alpha-i-j}\Big\} & \text{if } i,j\geq 2. \end{cases}$$

uniformly for $x\in[0,1]$, where

$$c_r = \frac{(1-\alpha)_{r\uparrow}}{r!}$$

Finally, recalling that

$$F^{\{n\}}_{i,j}(x) = a^{\{n\}}_{i-1,\lfloor nx\rfloor+1}a^{\{n\}}_{j-1,\lfloor nx\rfloor+1}\cdot\left[\left(P^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j}-\left(R^{\{n\}}_{\lfloor nx\rfloor}\right)_{i,j}\right]$$

all the above results entail $F_{i,j}^{\{n\}}(x) \xrightarrow{a.s.} f_{i,j}(x)$ uniformly for $x \in [0,1]$, where

$$
f_{i,j}(x) = \begin{cases}
\left(\frac{\lambda+x}{\lambda}\right)^{-1-\alpha} - \left(\frac{\lambda+x}{\lambda}\right)^{-2} & \text{if } i = j = 1 \\
\left(\frac{\lambda+x}{\lambda}\right)^{-\alpha} - \lambda \cdot (x+\lambda)^{-3} + \lambda\, c_1 \cdot x\,(x+\lambda)^{-2} & \text{if } i = 1, j = 2 \\
\lambda^{3-j} \cdot \left[-c_{j-2} \cdot x^{j-2}\,(x+\lambda)^{-3} + c_{j-1} \cdot x^{j-1}\,(x+\lambda)^{-2} \right] & \text{if } i = 1, j \geq 3 \\
\lambda^{\alpha+3-2i} \left[c_{i-2} \cdot x^{i-2}(x+\lambda)^{i-3-\alpha} + c_{i-1} \cdot x^{i-1}(x+\lambda)^{i-2-\alpha} \right] & \\
\quad -\lambda^{4-2i} \Big\{ c_{i-2}^2 \cdot x^{2i-4}(x+\lambda)^{-4} & \\
\qquad -2c_{i-1}\, c_{i-2} \cdot x^{2i-3}(x+\lambda)^{-3} & \\
\qquad +c_{i-1}^2 \cdot\, x^{2i-2}(x+\lambda)^{-2} \Big\} & \text{if } i = j \geq 2 \\
-c_{i-1}\lambda^{2+\alpha-2i} \cdot x^{i-1}(x+\lambda)^{i-1-\alpha} & \\
\quad -\lambda^{3-2i} \Big\{ c_{i-2}\, c_{i-1} \cdot x^{2i-3}(x+\lambda)^{-4} & \\
\qquad -\left[c_{i-1}\, c_{i-1} + c_{i-2}\, c_i \right] \cdot x^{2i-2}(x+\lambda)^{-3} & \\
\qquad +c_{i-1}\, c_i\, x^{2i-1}(x+\lambda)^{-2} \Big\} & \text{if } i \geq 2, j = i+1 \\
-\lambda^{4-i-j} \cdot \Big\{ c_{i-2}\, c_{j-2} \cdot x^{\,i+j-4}(x+\lambda)^{-4} & \\
\qquad -\left[c_{i-1}\, c_{j-2} + c_{i-2}\, c_{j-1} \right] \cdot x^{\,i+j-3}(x+\lambda)^{-3} & \\
\qquad +c_{i-1}\, c_{j-1} \cdot x^{\,i+j-2}(x+\lambda)^{-2} \Big\} & \text{if } i \geq 2, j \geq i+2 \\
f_{j,i}(x) & \text{if } i > j
\end{cases} \tag{D.1}
$$

All that remains to prove is that

$$
\mathcal{I}_{i,j} := \int_0^1 f_{i,j}(x)\,\mathrm{d}x = \left(\Gamma_{\alpha,\lambda}\right)_{i,j}.
$$

Exploiting the identities [52, Equation 3.197.8] and [1, Equation 15.3.3] we obtain

$$
\begin{aligned}
\int_0^1 x^A (x+\lambda)^B &= \frac{\lambda^B}{A+1} \cdot {}_2F_1\left(A+1, -B; A+2; -\frac{1}{\lambda}\right) \\
&= \frac{(\lambda+1)^B}{\lambda(A+1)} \cdot {}_2F_1\left(1, A+B+2; A+2; -\frac{1}{\lambda}\right)
\end{aligned}
$$

Applying such identity allows, after straightforward but tedious computations, to show that for every $i, j \in \{0, ..., d\}$

$$\mathcal{I}_{i,j}(x) = \begin{cases} s^2_{\alpha,\lambda} & i=j=1 \\ \frac{\lambda}{\alpha-1} - \frac{(\lambda+1)}{\alpha-1}\cdot\left(\frac{\lambda}{\lambda+1}\right)^{\alpha} - \left(\frac{\lambda}{\lambda+1}\right)^{2} + \frac{(1-\alpha)\lambda}{2\,(\lambda+1)}\cdot\log\left(\frac{\lambda+1}{\lambda}\right) & i=1, j=2 \\ \lambda^{2-j}\Big[- c_{j-2}\frac{(\lambda+1)^{-3}}{j-1}\,{}_2F_1(1,j-3;j-1;-1/\lambda) & \\ \quad + c_{j-1}\frac{(\lambda+1)^{-2}}{j}\,{}_2F_1(1,j-1;j;-1/\lambda)\Big] & i=1, j\geq 3 \\ \lambda^{\alpha+2-2i}\Big[c_{i-2}\frac{(\lambda+1)^{i-3-\alpha}}{i-1}\,{}_2F_1(1,i-\alpha-2;i-1;-1/\lambda) & \\ \quad + c_{i-1}\frac{(\lambda+1)^{i-2-\alpha}}{i}\,{}_2F_1(1,i-\alpha;i;-1/\lambda)\Big] & \\ \quad -\lambda^{3-2i}\Big[c^2_{i-2}\frac{(\lambda+1)^{-4}}{2i-3}\,{}_2F_1(1,2i-2;2i-3;-1/\lambda) & \\ \quad -2c_{i-1}c_{i-2}\frac{(\lambda+1)^{-3}}{2i-2}\,{}_2F_1(1,2i-1;2i-2;-1/\lambda) & \\ \quad + c^2_{i-1}\frac{(\lambda+1)^{-2}}{2i-1}\,{}_2F_1(1,2i;2i-1;-1/\lambda)\Big] & i=j\geq 2 \\ -c_{i-1}\lambda^{1+\alpha-2i}\frac{(\lambda+1)^{i-1-\alpha}}{i}\,{}_2F_1(1,i-\alpha;i;-1/\lambda) & \\ \quad -\lambda^{2-2i}\Big[c_{i-2}c_{i-1}\frac{(\lambda+1)^{-4}}{2i-2}\,{}_2F_1(1,2i-1;2i-2;-1/\lambda) & \\ \quad -(c^2_{i-1}+c_{i-2}c_i)\frac{(\lambda+1)^{-3}}{2i-1}\,{}_2F_1(1,2i;2i-1;-1/\lambda) & \\ \quad + c_{i-1}c_i\frac{(\lambda+1)^{-2}}{2i}\,{}_2F_1(1,2i+1;2i;-1/\lambda)\Big] & i\geq 2, j=i+1 \\ -\lambda^{3-i-j}\Big[c_{i-2}c_{j-2}\frac{(\lambda+1)^{-4}}{i+j-3}\,{}_2F_1(1,i+j-2;i+j-3;-1/\lambda) & \\ \quad -(c_{i-1}c_{j-2}+c_{i-2}c_{j-1})\frac{(\lambda+1)^{-3}}{i+j-2}\,{}_2F_1(1,i+j-1;i+j-2;-1/\lambda) & \\ \quad + c_{i-1}c_{j-1}\frac{(\lambda+1)^{-2}}{i+j-1}\,{}_2F_1(1,i+j;i+j-1;-1/\lambda)\Big] & i\geq 2, j\geq i+2 \\ f_{j,i}(x) & i>j \end{cases} \tag{D.2}$$

which, rewritten in terms of H (3.2.1), matches the definition of $\Gamma_{\alpha,\lambda}$.

E Proof of (4.3.6) and of the LLN (4.3.7)

The proof of (4.3.6) follows the same lines when $\alpha \in (0,1)$ or $\alpha = 0$, relying on representations (4.3.1) and (4.3.2) respectively; we split the two cases for the sake of clarity. The LLN (4.3.7) follows directly from (4.3.6) and its proof is carried out at the end of the section.

E.1 Proof of (4.3.6) in the case $\alpha \in (0,1)$

We resort to representation (4.3.1) which, in the regime $\theta = \tau m$, $n = \nu m$, $j = \varrho m$, becomes

$$K_m^{(n)} \stackrel{\mathrm{d}}{=} Q\left(K_m^*,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)$$

By the law of total expectation and recalling the expression for the mean of the beta-binomial distribution, the asymptotics for the mean of K_m^* in theorem 4.3.1 and the definitions of $\mathfrak{m}_{\alpha,\lambda}$ and $\mathscr{M}_{\alpha,\tau,\nu,\varrho}$ in the case $\alpha \in (0,1)$,

$$\begin{aligned}
\mathbb{E}\left[K_m^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[Q\left(K_m^*,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)\,\middle|\, K_m^*\right]\right] \\
&= \mathbb{E}\left[K_m^* \cdot \frac{\tau+\varrho\alpha}{\tau+\nu}\right] \\
&= m \cdot \mathfrak{m}_{\alpha,\lambda}\frac{\tau+\varrho\alpha}{\tau+\nu} + O(1) \\
&= m \cdot \mathscr{M}_{\alpha,\tau,\nu,\varrho} + O(1)
\end{aligned}$$

By the law of total variance, and recalling the expression for the moments of the beta-binomial distribution, the asymptotics in (4.3.6) and the definitions of $\mathfrak{m}_{\alpha,\lambda}, \mathfrak{s}_{\alpha,\lambda}$ and $\mathscr{S}_{\alpha,\tau,\nu,\varrho}$ in the case $\alpha \in (0,1)$,

$$\begin{aligned}
&\mathrm{Var}\left(K_m^{(n)}\right) \\
&\quad= \mathbb{E}\left[\mathrm{Var}\left(Q\left(K_m^*,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)\,\middle|\, K_m^*\right)\right] \\
&\qquad+ \mathrm{Var}\left(\mathbb{E}\left[Q\left(K_m^*,\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)\,\middle|\, K_m^*\right]\right) \\
&\quad= \mathbb{E}\left[K_m^*\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}\left(1-\frac{\alpha}{(\tau+\nu)m+\alpha}\right) + \frac{(K_m^*)^2}{m}\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}\frac{\alpha m}{(\tau+\nu)m+\alpha}\right] \\
&\qquad+ \mathrm{Var}\left(K_m^* \cdot \frac{\tau+\varrho\alpha}{\tau+\nu}\right) \\
&\quad= [m\cdot\mathfrak{m}_{\alpha,\lambda} + O(1)]\cdot\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}\left[1+O\left(\frac{1}{m}\right)\right] + m\cdot\mathfrak{s}_{\alpha,\lambda}\frac{(\tau+\varrho\alpha)^2}{(\tau+\nu)^2} + O(1) \\
&\qquad+ \frac{m^2\cdot\mathfrak{m}_{\alpha,\lambda}^2 + O(m)}{m}\cdot\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}\left[\frac{\alpha}{(\tau+\nu)} + O\left(\frac{1}{m}\right)\right] \\
&\quad= m\left[\mathfrak{m}_{\alpha,\lambda}\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)} + \mathfrak{s}_{\alpha,\lambda}\frac{(\tau+\varrho\alpha)^2}{(\tau+\nu)} + \mathfrak{m}_{\alpha,\lambda}^2\frac{\alpha(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^3}\right] + O(1) \\
&\quad= m\cdot\mathscr{S}_{\alpha,\tau,\nu,\varrho} + O(1)
\end{aligned}$$

E.2 Proof of (4.3.6) in the case $\alpha = 0$

We resort to representation (4.3.2) which, in the regime $\theta = \tau m$, $n = \nu m$, becomes

$$K_m^{(n)} \stackrel{\mathrm{d}}{=} Q\left(K_m^*,\, \frac{\tau}{\lambda}\right)$$

By the law of total expectation, and recalling the expression for the mean of the binomial distribution, the asymptotics for the mean of K_m^* in (**??**) and the definitions of $\mathfrak{m}_{0,\lambda}$ and

$\mathscr{M}_{0,\tau,\nu,\varrho}$,

$$\begin{aligned}\mathbb{E}\left[K_m^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[Q\left(K_m^*, \frac{\tau}{\lambda}\right) \,\middle|\, K_m^*\right]\right] \\ &= \mathbb{E}\left[K_m^* \cdot \frac{\tau}{\lambda}\right] \\ &= m \cdot \mathfrak{m}_{0,\lambda} \frac{\tau}{\lambda} + O(1) \\ &= m \cdot \mathscr{M}_{0,\tau,\nu,\varrho} + O(1)\end{aligned}$$

By the law of total variance, and recalling the expression for the moments of the binomial distribution, the asymptotics in theorem 4.3.1 and the definitions of $\mathfrak{m}_{0,\lambda}, \mathfrak{s}_{0,\lambda}$ and $\mathscr{S}_{0,\tau,\nu,\varrho}$,

$$\begin{aligned}\mathrm{Var}\left(K_m^{(n)}\right) &= \mathbb{E}\left[\mathrm{Var}\left(Q\left(K_m^*, \frac{\tau}{\lambda}\right) \,\middle|\, K_m^*\right)\right] + \mathrm{Var}\left(\mathbb{E}\left[Q\left(K_m^*, \frac{\tau}{\lambda}\right) \,\middle|\, K_m^*\right]\right) \\ &= \mathbb{E}\left[K_m^* \cdot \frac{\tau\nu}{\lambda^2}\right] + \mathrm{Var}\left(K_m^* \cdot \frac{\tau}{\lambda}\right) \\ &= m \cdot \mathfrak{m}_{0,\lambda} \cdot \frac{\tau\nu}{\lambda^2} + O(1) + m \cdot \mathfrak{s}_{0,\lambda} \frac{\tau^2}{\lambda^2} + O(1) \\ &= m \cdot \mathscr{S}_{0,\tau,\nu,\varrho} + O(1)\end{aligned}$$

E.3 Proof of the LLN (4.3.7)

To prove the LLN (4.3.7), we fix $\varepsilon > 0$ and combine Chebychev inequality with the (variance) asymptotic expansion (4.3.6). In particular, we write

$$\begin{aligned}P\left[\left|\frac{K_m^{(n)} - \mathbb{E}\left[K_m^{(n)}\right]}{m}\right| > \varepsilon\right] &= P\left[\left|K_m^{(n)} - \mathbb{E}\left[K_m^{(n)}\right]\right| > m\varepsilon\right] \\ &\leq \frac{\mathrm{Var}\left(K_m^{(n)}\right)}{m^2\varepsilon^2} \\ &= O\left(\frac{1}{m}\right) \to 0\end{aligned}$$

as $m \to +\infty$. Since the asymptotic expansion (4.3.6) of $\mathbb{E}\left[K_m^{(n)}\right]$ implies that $m^{-1}\mathbb{E}\left[K_m^{(n)}\right] \to \mathscr{M}_{\alpha,\tau,\nu,\varrho}$ as $m \to +\infty$, the proof is concluded by means of Slutsky's theorem.

F Proof of the CLT (4.3.8)

We begin by providing some detail on the sketch of proof presented in section 4.3. Equality (4.3.15) follows from the chain of equalities

$$\begin{aligned}
F_m(x) &= \int_0^{+\infty} P\left[Q_m(z) \le m\mathscr{M}_{\alpha,\tau,\nu,\varrho} + \sqrt{m}\mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x\right] \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\
&= \int_0^{+\infty} P\left[V_m(z) \le \frac{\sqrt{m}\,[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma(z)}\right] \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\
&= \int_0^{+\infty} \Phi\left(\frac{\sqrt{m}\,[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma(z)}\right) \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\
&\quad + \int_0^{+\infty} \left\{ F_{V_m(z)}\left(\frac{\sqrt{m}\,[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma(z)}\right)\right. \\
&\qquad \left. - \Phi\left(\frac{\sqrt{m}\,[\mathscr{M}_{\alpha,\tau,\nu,\varrho} - \mu(z)] + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma(z)}\right)\right\} \mu_{\frac{K_m^*}{m}}(\mathrm{d}z) \\
&=: \mathcal{I}_1^{(m)}(x) + \mathcal{I}_2^{(m)}(x).
\end{aligned}$$

The following two lemmas make use Proposition 7 and the CLT for K_m^*, i.e. (4.3.5) of theorem 4.3.1, and are instrumental to the proof of (4.3.16) trough Proposition 8.

Lemma F.1. *If Y is a Gaussian random variable with mean 0 and variance $\mathfrak{s}_{\alpha,\lambda}$, then*

$$\mathbb{E}\left[\Phi\left(\frac{\mu'(\mathfrak{m}_{\alpha,\lambda})Y + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\mathfrak{m}_{\alpha,\lambda}\right)}\right)\right] = \Phi(x)$$

for every $x \in \mathbb{R}$.

Proof. We introduce a standard Gaussian random variable Z, with Z independent from the random variable Y. By a standard property of conditional probability,

$$\Phi\left(\frac{\mu'(\mathfrak{m}_{\alpha,\lambda})Y + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\mathfrak{m}_{\alpha,\lambda}\right)}\right) = P\left[Z \le \frac{\mu'(\mathfrak{m}_{\alpha,\lambda})Y + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\mathfrak{m}_{\alpha,\lambda}\right)} \,\middle|\, Y\right],$$

which implies

$$\mathbb{E}\left[\Phi\left(\frac{\mu'(\mathfrak{m}_{\alpha,\lambda})Y + \mathscr{S}_{\alpha,\tau,\nu,\varrho}\, x}{\sigma\left(\mathfrak{m}_{\alpha,\lambda}\right)}\right)\right] = P\left[\frac{\sigma(\mathfrak{m}_{\alpha,\lambda})Z - \mu'(\mathfrak{s}_{\alpha,\lambda})Y}{\mathscr{S}_{\alpha,\tau,\nu,\varrho}} \le x\right].$$

To conclude, notice that $\frac{\sigma(\mathfrak{m}_{\alpha,\lambda})Z - \mu'(\mathfrak{s}_{\alpha,\lambda})Y}{\mathscr{S}_{\alpha,\tau,\nu,\varrho}}$ is a linear combination of independent Gaussian random variables, hence it is Gaussian with mean 0 and variance

$$\frac{\sigma^2\left(\mathfrak{m}_{\alpha,\lambda}\right) + \mathfrak{s}_{\alpha,\lambda}^2 \cdot \left(\mu'\left(\mathfrak{m}_{\alpha,\lambda}\right)\right)^2}{\mathscr{S}_{\alpha,\tau,\nu,\varrho}^2} = 1,$$

where the last identity follows from Proposition 7. This completes the proof. ∎

Lemma F.2. *Let $\psi : [0, +\infty) \to \mathbb{R}$ be a continuous function such that $\psi \in C^1((0, +\infty))$. If the function ψ has bounded (first) derivative, then as $n \to +\infty$*

$$\sqrt{n}\left[\psi\left(\frac{K_m^*}{m}\right) - \psi(\mathfrak{m}_{\alpha,\lambda})\right] \xrightarrow{w} \mathcal{N}\left(0, (\psi'(\mathfrak{m}_{\alpha,\lambda}))^2 \mathfrak{s}_{\alpha,\lambda}^2\right)$$

Proof. By means of the fundamental theorem of calculus, we can write that

$$\begin{aligned}\sqrt{n}\left[\psi\left(\frac{K_m^*}{m}\right)-\psi(\mathfrak{m}_{\alpha,\lambda})\right] &= \\ &= \sqrt{n}\left(\frac{K_m^*}{m}-\mathfrak{m}_{\alpha,\lambda}\right)\int_0^1 \psi'\left(\mathfrak{m}_{\alpha,\lambda}+t\left[\frac{K_m^*}{m}-\mathfrak{m}_{\alpha,\lambda}\right]\right)\mathrm{d}t. \end{aligned} \tag{F.1}$$

By (4.3.4), we have that $m^{-1}K_m^* - \mathfrak{m}_{\alpha,\lambda} \xrightarrow{\mathrm{p}} 0$, as $m \to +\infty$. Since ψ' is bounded, as $m \to +\infty$

$$\int_0^1 \psi'\left(\mathfrak{m}_{\alpha,\lambda}+t\left[\frac{K_m^*}{m}-\mathfrak{m}_{\alpha,\lambda}\right]\right)\mathrm{d}t \xrightarrow{\mathrm{p}} \psi'(\mathfrak{m}_{\alpha,\lambda}) \tag{F.2}$$

and, by (4.3.5), as $m \to +\infty$

$$\sqrt{n}\left(\frac{K_m^*}{m}-\mathfrak{m}_{\alpha,\lambda}\right) \xrightarrow{\mathrm{w}} \mathcal{N}\left(0,\mathfrak{s}^2_{\alpha,\lambda}\right). \tag{F.3}$$

From (F.1), with (F.2) and (F.3), the proof completed by means of Slutsky's theorem. ■

F.1 Proof of proposition 6 in the case $\alpha = 0$

Proof of (4.3.10) *and* (4.3.11). By a simple computation,

$$\begin{aligned}\mathbb{E}[Q_m(z)] &= \lfloor mz \rfloor \cdot \frac{\tau}{\tau+\nu} \\ &= mz \cdot \frac{\tau}{\tau+\nu} - (mz - \lfloor mz \rfloor)\cdot\frac{\tau}{\tau+\nu}\end{aligned}$$

Since $mz - \lfloor mz \rfloor < 1$ by definition, this proves the first equality. For the second equaility,

$$\begin{aligned}\mathrm{Var}\left(Q_m(z)\right) &= \lfloor mz \rfloor \cdot \frac{\tau\nu}{(\tau+\nu)^2} \\ &= mz \cdot \frac{\tau\nu}{(\tau+\nu)^2} - (mz - \lfloor mz \rfloor)\cdot\frac{\tau\nu}{(\tau+\nu)^2} \\ &= m\sigma^2(z) + O(1)\end{aligned}$$

so that the $O(1)$ again accounts for the fact that we are discarding the floor function. ■

Proof of (4.3.12). This is the standard Berry-Essen bound for the binomial distribution - see [78, Chapter V, Theorem 4]. ■

F.2 Proof of proposition 6 in the case $\alpha \in (0,1)$

Proof of (4.3.10) *and* (4.3.11). By the law of total expectation,

$$\begin{aligned}\mathbb{E}[Q_m(z)] &= \mathbb{E}\left[\mathbb{E}\left[Q_m(z)\,\middle|\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right]\right] \\ &= \lfloor mz \rfloor\, \mathbb{E}\left[B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right] \\ &= mz \cdot \frac{\tau+\varrho\alpha}{\tau+\nu} - (mz - \lfloor mz \rfloor)\cdot\frac{\tau+\varrho\alpha}{\tau+\nu}\end{aligned}$$

Since $mz - \lfloor mz \rfloor < 1$ by definition, this proves the first equality. By the law of total variance,

$$\begin{aligned}\operatorname{Var}\left(Q_m(z)\right) &= \mathbb{E}\left[\operatorname{Var}\left(Q_m(z)\,\middle|\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)\right] + \operatorname{Var}\left(\mathbb{E}\left[Q_m(z)\,\middle|\, B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right]\right)\\ &= \lfloor mz \rfloor\, \mathbb{E}\left[B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m} - B^2_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right] + \lfloor mz \rfloor \operatorname{Var}\left(B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right)\\ &= mz\,\frac{\tau+\varrho\alpha}{\tau+\nu}\left[1 - \frac{\tau+\varrho\alpha}{\tau+\nu} + \alpha z\,\frac{\nu-\varrho\alpha}{(\tau+\nu)^2}\right] + O(1)\end{aligned}$$

where the $O(1)$ accounts for the fact that we are discarding the floor function and that we are substituting the exact expression for the variance of $B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}$ with its asymptotic principal part:

$$\begin{aligned}m\,\operatorname{Var}\left(B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}\right) &= \frac{\alpha\,(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2(\tau+\nu+\alpha/m)}\\ &= \frac{\alpha\,(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^3\,[1+\alpha/(\tau m+\nu m)]}\\ &= \frac{\alpha\,(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^3}\left[1+O\left(\frac{1}{m}\right)\right]\end{aligned}$$

■

Proof of (4.3.12). The outline of the proof is as follows: fix $\delta \in (0, 1/4)$ and a constant $\mathcal{C}$ and start with the well-known inequality [78, Chapter V, Theorem 2]

$$\left\|F_{V_m(z)} - \Phi\right\|_\infty \le \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^\delta} \left|\frac{\varphi_{V_m(z)}(\xi) - e^{-\frac{\xi^2}{2}}}{\xi}\right| \mathrm{d}\xi + \tilde{\mathcal{C}} m^{-\delta}$$

where $\varphi_{V_m(z)}$ denotes the characteristic function of $V_m(z)$ and $\tilde{\mathcal{C}} = \max_{z\in[\zeta_0,\zeta_1]} \sigma(z)\,\mathcal{C}$. For notational convenience, let $B_m = B_{\left(\frac{\tau}{\alpha}+\varrho\right)m,\left(\frac{\nu}{\alpha}-\varrho\right)m}$ and define, for $z \in [\zeta_0, \zeta_1]$,

$$\begin{aligned}G_m(z) &= \sqrt{m}\,[\mu(z) - zB_m]\\ S_m(z) &= \frac{z}{\sigma^2(z)}\left[B_m - B_m^2\right]\end{aligned}$$

and

$$S(z) = \frac{z}{\sigma^2(z)} \cdot \frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^2}.$$

Further, let $G(z)$ denote a Gaussian random variable with mean 0 and variance

$$s^2(z) = z^2 \cdot \frac{\alpha(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^3},$$

independent of B_m. In steps 1–3 we rewrite the right-end side of the inequality as

$$\begin{aligned}&\left\|F_{V_m(z)} - \Phi\right\|_\infty\\ &\le \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^\delta} \left|\frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z) - \frac{\xi^2}{2}S_m(z)}\right] - \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G(z) - \frac{\xi^2}{2}S(z)}\right]}{\xi}\right| \mathrm{d}\xi\\ &\quad + \tilde{C}_2 m^{2\delta-\frac{1}{2}} + \tilde{\mathcal{C}} m^{-\delta}\end{aligned}$$

Then, making use of the triangular inequality, we further split the problem in two parts:

$$\begin{aligned}
&\left\|F_{V_m(z)} - \Phi\right\|_\infty \\
&\quad \le \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^\delta} \left| \frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S_m(z)}\right] - \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S(z)}\right]}{\xi} \right| \mathrm{d}\xi \\
&\qquad + \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^\delta} \left| \frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S(z)}\right] - \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G(z)-\frac{\xi^2}{2}S(z)}\right]}{\xi} \right| \mathrm{d}\xi \\
&\qquad + \tilde{C}_2 m^{2\delta-\frac{1}{2}} + \tilde{\mathcal{C}} m^{-\delta} \\
&\quad = \mathcal{I}_m^{(1)}(z) + \mathcal{I}_m^{(2)}(z) + \tilde{C}_2 m^{2\delta-\frac{1}{2}} + \tilde{\mathcal{C}} m^{-\delta}
\end{aligned}$$

In steps 4 and 5 we bound $\mathcal{I}_m^{(1)}(z)$ and $\mathcal{I}_m^{(2)}(z)$ for every $z \in [\zeta_0, \zeta_1]$ respectively by

$$\mathcal{I}_m^{(1)}(z) \le c_1\, m^{2\delta-\frac{1}{2}} + c m^{-3/2}$$

for some suitable constants $c_1, c > 0$ and

$$\mathcal{I}_m^{(2)}(z) \le c_2\, m^{-\gamma}$$

for any $\gamma \in \left(0, \frac{1}{2}\right)$ and some constant $c_2 > 0$ depending only on γ. Thus,

$$\left\|F_{V_m(z)} - \Phi\right\|_\infty \le c_1\, m^{2\delta-\frac{1}{2}} + c m^{-3/2} + c_2\, m^{-\gamma} + \tilde{C}_2 m^{2\delta-\frac{1}{2}} + \tilde{\mathcal{C}} m^{-\delta}.$$

To conclude, it is easy to check that for every $\delta \in \left(0, \frac{1}{4}\right)$ and every $\gamma \in \left(\frac{1}{6}, \frac{1}{2}\right)$,

$$\min\left(\frac{3}{2}, \delta, -2\delta + \frac{1}{2}, \gamma\right) \ge \frac{1}{6},$$

whence for every $z \in [\zeta_0, \zeta_1]$ it holds

$$\left\|F_{V_m(z)} - \Phi\right\|_\infty \le \bar{C}\, m^{-\frac{1}{6}}$$

for some constant $\bar{C} > 0$. This concludes the proof of (4.3.12).

Step 1. Let f_{B_m} denote the density of B_m and write

$$\begin{aligned}
\varphi_{V_m(z)}(\xi) &= e^{-i\xi\sqrt{m}\frac{\mu(z)}{\sigma(z)}} \cdot \varphi_{Q_m(z)}\left(\frac{\xi}{\sqrt{m}\sigma(z)}\right) \\
&= e^{-i\xi\sqrt{m}\frac{\mu(z)}{\sigma(z)}} \cdot \mathbb{E}\left[\varphi_{Q(\lfloor mz\rfloor, B_m)}\left(\frac{\xi}{\sqrt{m}\sigma(z)}\right)\right] \\
&= \int_0^1 \exp\left\{-i\xi\sqrt{m}\frac{\mu(z)}{\sigma(z)} + \lfloor mz\rfloor \log\left[1 + p\left(e^{i\xi\frac{1}{\sqrt{m}\sigma(z)}} - 1\right)\right]\right\} f_{B_m}(p)\,\mathrm{d}p
\end{aligned}$$

If ξ satisfies

$$|\xi| \le \mathcal{C}\,\sigma(z)\,m^\delta, \tag{F.4}$$

then [78, Chapter IV, Lemma 5] guarantees that, for every $p \in [0,1]$,

$$\left|p\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}} - 1\right)\right| \le p\left|\frac{\xi}{\sqrt{m}\sigma(z)}\right| \le C\, m^{\delta-\frac{1}{2}}.$$

Then, we can apply Taylor's formula to $\log\left[1+p\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)\right]$ and then to $\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)$ to obtain

$$\begin{aligned}\varphi_{V_m(z)}(\xi) &= \int_0^1 \exp\left\{-i\frac{\xi}{\sigma(z)}\sqrt{m}\left[\mu(z)-zp\right]-\frac{\xi^2}{2}\frac{z}{\sigma^2(z)}\left[p-p^2\right]\right\}\mathfrak{R}_m(p,z,\xi)\,f_{B_m}(p)\,\mathrm{d}p\\ &= \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S_m(z)}\cdot\mathfrak{R}_m\left(B_m,z,\xi\right)\right]\end{aligned}$$

where

$$\begin{aligned}\mathfrak{R}_m(p,z,\xi) = \exp\Bigg\{&(mz-\lfloor mz\rfloor)\,\log\left[1+p\left(e^{i\xi\frac{1}{\sqrt{m}\sigma(z)}}-1\right)\right]\\ &+mzp\left[\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)-\left(\frac{i\xi}{\sqrt{m}\sigma(z)}-\frac{\xi^2}{2m\sigma^2(z)}\right)\right]\\ &+\frac{1}{2}mzp^2\left[\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)^2+\frac{\xi^2}{m\sigma^2(z)}\right]\\ &+mz\,p^3\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)^3\int_0^1\left(2+tp\left(e^{\frac{i\xi}{\sqrt{m}\sigma(z)}}-1\right)\right)^{-2}(1-t)^2\,\mathrm{d}t\Bigg\}.\end{aligned}$$

Making use of elementary properties of the exponential together with [78, Chapter IV, Lemma 5], we can prove that there exists a constant $\tilde{C}_1$ such that for every $z\in[\zeta_0,\zeta_1]$, every $p\in[0,1]$ and every ξ satisfying (F.4)

$$\left|\mathfrak{R}_m(p,z,\xi)-1\right|\le\tilde{C}_1\left|\xi\right|m^{2\delta-\frac{1}{2}}.$$

It follows that

$$\varphi_{V_m(z)}(\xi)=\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S_m(z)}\right]+\mathcal{R}_1(m) \tag{F.5}$$

with

$$\begin{aligned}\left|\mathcal{R}_1(m)\right| &= \left|\varphi_{V_m(z)}(\xi)-\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S_m(z)}\right]\right|\\ &= \left|\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)-\frac{\xi^2}{2}S_m(z)}\cdot\left[\mathfrak{R}_m\left(B_m,z,\xi\right)-1\right]\right]\right|\\ &\le \int_0^1\left|\exp\left\{-i\frac{\xi}{\sigma(z)}\sqrt{m}\left[\mu(z)-zp\right]-\frac{\xi^2}{2}\frac{z}{\sigma^2(z)}\left[p-p^2\right]\right\}\right|\cdot\left|\mathfrak{R}_m\left(p,z,\xi\right)-1\right|\,f_{B_m}(p)\mathrm{d}p\\ &\le \int_0^1\tilde{C}_1\left|\xi\right|m^{2\delta-\frac{1}{2}}\,f_{B_m}(p)\mathrm{d}p=\tilde{C}_1\left|\xi\right|m^{2\delta-\frac{1}{2}}\end{aligned}$$

Step 2. Now note that

$$s^2(z)=\sigma^2(z)\cdot\frac{\alpha}{\alpha z+\tau+\nu}$$

and therefore

$$\begin{aligned}\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G(z)-\frac{\xi^2}{2}S(z)}\right] &= \varphi_{G(z)}\left(\frac{\xi}{\sigma(z)}\right)\cdot e^{\frac{\xi^2}{2}S(z)}\\ &= \exp\left\{-\frac{\xi^2}{2}\cdot\left[\frac{\alpha z}{\alpha z+\tau+\nu}+\frac{\tau+\nu}{\alpha z+\tau+\nu}\right]\right\}\\ &= e^{-\frac{\xi^2}{2}}\end{aligned} \tag{F.6}$$

Step 3. Using (F.5) and (F.6), write

$$\begin{aligned}
&\int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^{\delta}} \left| \frac{\varphi_{V_m(z)}(\xi) - e^{-\frac{\xi^2}{2}}}{\xi} \right| \mathrm{d}\xi \\
&\quad = \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^{\delta}} \left| \frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z) - \frac{\xi^2}{2}S_m(z)}\right] + \mathcal{R}_1(m) - \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G(z) - \frac{\xi^2}{2}S(z)}\right]}{\xi} \right| \mathrm{d}\xi \\
&\quad \le \int_{|\xi|\le \mathcal{C}\,\sigma(z)\,m^{\delta}} \left| \frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z) - \frac{\xi^2}{2}S_m(z)}\right] - \mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G(z) - \frac{\xi^2}{2}S(z)}\right]}{\xi} \right| \mathrm{d}\xi + \tilde{C}_2\, m^{2\delta-\frac{1}{2}}
\end{aligned}$$

where $\tilde{C}_2 = 2\,\mathcal{C}\,\tilde{C}_1\,\max_{z\in[\zeta_0,\zeta_1]}\sigma(z)$.

Step 4. By definition of $\mathcal{I}_m^{(1)}(z)$, Jensen's inequality and the triangular inequality,

$$\begin{aligned}
\mathcal{I}_m^{(1)}(z) &= \int_{-\mathcal{C}\sigma(z)m^{\delta}}^{\mathcal{C}\sigma(z)m^{\delta}} \left| \frac{\mathbb{E}\left[e^{-i\frac{\xi}{\sigma(z)}G_m(z)}\left(e^{-\frac{\xi^2}{2}S_m(z)} - e^{-\frac{\xi^2}{2}S(z)}\right)\right]}{\xi} \right| \mathrm{d}\xi \\
&\le \int_{-\mathcal{C}\sigma(z)m^{\delta}}^{\mathcal{C}\sigma(z)m^{\delta}} \frac{\mathbb{E}\left[\left|e^{-\frac{\xi^2}{2}S_m(z)} - e^{-\frac{\xi^2}{2}S(z)}\right|\right]}{|\xi|}\,\mathrm{d}\xi
\end{aligned}$$

Now note that, for every $z \in [\zeta_0,\zeta_1]$, $S_m(z) \ge 0$ a.s., and $S(z) \ge 0$. Since the function $[0,+\infty) \to (0,1]$; $x \mapsto e^{-x}$ is 1-Lipschitz, the last term of the above inequality can be bounded by

$$\begin{aligned}
\int_{-\mathcal{C}\sigma(z)m^{\delta}}^{\mathcal{C}\sigma(z)m^{\delta}} \frac{\mathbb{E}\left[\left|e^{-\frac{\xi^2}{2}S_m(z)} - e^{-\frac{\xi^2}{2}S(z)}\right|\right]}{|\xi|}\,\mathrm{d}\xi &\le \int_{-\mathcal{C}\sigma(z)m^{\delta}}^{\mathcal{C}\sigma(z)m^{\delta}} \frac{|\xi|}{2}\,\mathbb{E}\left[|S_m(z) - S(z)|\right]\,\mathrm{d}\xi \\
&\le \sqrt{\mathbb{E}\left[(S_m(z) - S(z))^2\right]}\cdot \mathcal{C}^2\sigma^2(z)\,m^{2\delta} \qquad \text{(F.7)}
\end{aligned}$$

A simple computation shows

$$\mathbb{E}[S_m(z)] = S(z)\left[1 + O\left(m^{-1}\right)\right],$$

which in turn entails

$$\mathbb{E}\left[(S_m(z) - S(z))^2\right] = \left[\mathbb{E}\left[S_m^2(z)\right] - S^2(z)\right]\cdot\left[1 + O\left(m^{-2}\right)\right].$$

The second moment of $S_m(z)$ can be written as follows

$$\mathbb{E}\left[S_m^2(z)\right] = \frac{z^2}{\sigma^4(z)}\cdot\left\{\mathbb{E}\left[B_m^2\right] - 2\mathbb{E}\left[B_m^3\right] + \mathbb{E}\left[B_m^4\right]\right\},$$

reducing the problem to the study of the moments of B_m. By standard results regarding the beta distribution it is known that, for $k \in \mathbb{N}$,

$$\mathbb{E}\left[B_m^k\right] = \frac{\Gamma\left(\frac{\tau+\varrho\alpha}{\alpha}\,m + k\right)\,\Gamma\left(\frac{\tau+\nu}{\alpha}\,m\right)}{\Gamma\left(\frac{\tau+\varrho\alpha}{\alpha}\,m\right)\,\Gamma\left(\frac{\tau+\nu}{\alpha}\,m + k\right)},$$

so that, for $m \to +\infty$, a straightforward application of [89, formula (1)] yields

$$\begin{aligned}\mathbb{E}\left[B_m^k\right] &= \left(\frac{\tau+\varrho\alpha}{\alpha}\,m\right)^k\left[1+\frac{k(k-1)\alpha}{2(\tau+\varrho\alpha)\,m}+O\left(m^{-2}\right)\right]\\ &\quad+\left(\frac{\tau+\varrho\alpha}{\alpha}\,m\right)^k\left[1-\frac{k(k-1)\alpha}{2(\tau+\nu)\,m}+O\left(m^{-2}\right)\right]\\ &= \left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^k\left[1+\frac{k(k-1)}{m}\,\eta+O\left(m^{-2}\right)\right]\end{aligned}$$

where $\eta=\eta(\tau,\nu,\varrho,\alpha):=\frac{\alpha(\nu-\varrho\alpha)}{2(\tau+\varrho\alpha)(\tau+\nu)}$ is a constant not depending on k or m. Then,

$$\begin{aligned}\mathbb{E}\left[S_m^2(z)\right] &= \frac{z^2}{\sigma^4(z)}\cdot\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\cdot\left\{\left[1+\frac{2\eta}{m}\right]-2\frac{\tau+\varrho\alpha}{\tau+\nu}\left[1+\frac{6\eta}{m}\right]\right.\\ &\qquad\left.+\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\left[1+\frac{12\eta}{m}\right]+O\left(m^{-2}\right)\right\}\\ &= \frac{z^2}{\sigma^4(z)}\cdot\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\cdot\left\{\left(\frac{\nu-\varrho\alpha}{\tau+\nu}\right)^2+\frac{1}{m}\cdot\left[2\eta-12\eta\frac{\tau+\varrho\alpha}{\tau+\nu}+12\eta\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\right]\right.\\ &\qquad\left.+O\left(m^{-2}\right)\right\}\\ &= S^2(z)+\frac{1}{m}\,g(z,\tau,\nu,\varrho,\alpha)+O\left(m^{-2}\right),\end{aligned}$$

with $g(z,\tau,\nu,\varrho,\alpha):=\frac{z^2}{\sigma^4(z)}\cdot\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\cdot\left[2\eta-12\eta\frac{\tau+\varrho\alpha}{\tau+\nu}+12\eta\left(\frac{\tau+\varrho\alpha}{\tau+\nu}\right)^2\right]$. We can conclude that

$$\sqrt{\mathbb{E}\left[\left(S_m(z)-S(z)\right)^2\right]}=\frac{1}{\sqrt{m}}\,g(z,\tau,\nu,\varrho,\alpha)\left[1+r_m\right]$$

with $r_m=O\left(m^{-1}\right)$. Since all the asymptotic expansions above hold uniformly on compact sets, in the sense of [21, Definition A.9], and the function $z\to g(z,\tau,\nu,\varrho,\alpha)$ is continuous, one can resort to [21, Lemma A.11] to obtain

$$|R_m|:=\left|\sqrt{\mathbb{E}\left[\left(S_m(z)-S(z)\right)^2\right]}-\frac{1}{\sqrt{m}}\,g(z,\tau,\nu,\varrho,\alpha)\right|\le\frac{c}{m^{3/2}}$$

fro some suitable positive constant c. To conclude, plug the above result in (F.7) to obtain that for all $z\in[\zeta_0,\zeta_1]$,

$$\mathcal{I}_m^{(1)}(z)\le c_1\,m^{2\delta-\frac{1}{2}}+\frac{c}{m^{3/2}}$$

for some suitable constant $c_1>0$.

Step 5. By definition,

$$\begin{aligned}\mathcal{I}_m^{(2)}(z) &= \int_{-\mathcal{C}\sigma(z)m^\delta}^{\mathcal{C}\sigma(z)m^\delta}e^{-\frac{\xi^2S(z)}{2}}\left|\int_{\mathbb{R}}\frac{e^{i\xi x}-1}{\xi}\left[f_{G_m(z)}(x)-f_{G(z)}(x)\right]\,\mathrm{d}x\right|\,\mathrm{d}\xi\\ &\le \int_{-\mathcal{C}\sigma(z)m^\delta}^{\mathcal{C}\sigma(z)m^\delta}e^{-\frac{\xi^2S(z)}{2}}\int_{\mathbb{R}}\left|\frac{e^{i\xi x}-1}{\xi}\right|\left|f_{G_m(z)}(x)-f_{G(z)}(x)\right|\,\mathrm{d}x\,\mathrm{d}\xi\\ &= \int_{\mathbb{R}}\left[\int_{-\mathcal{C}\sigma(z)m^\delta}^{\mathcal{C}\sigma(z)m^\delta}e^{-\frac{\xi^2S(z)}{2}}\left|\frac{e^{i\xi x}-1}{\xi}\right|\,\mathrm{d}\xi\right]\left|f_{G_m(z)}(x)-f_{G(z)}(x)\right|\,\mathrm{d}x\end{aligned}$$

Making use of [78, Chapter IV, Lemma 5] and of the elementary properties of the Gaussian density write the bound

$$\begin{aligned}\int_{-\mathcal{C}\sigma(z)m^\delta}^{\mathcal{C}\sigma(z)m^\delta} e^{-\frac{\xi^2 S(z)}{2}} \left|\frac{e^{i\xi x}-1}{\xi}\right| \mathrm{d}\xi &\le \int_{-\mathcal{C}\sigma(z)m^\delta}^{\mathcal{C}\sigma(z)m^\delta} e^{-\frac{\xi^2 S(z)}{2}} \, |x| \, \mathrm{d}\xi \\ &\le |x| \int_{\mathbb{R}} e^{-\frac{\xi^2 S(z)}{2}} \, \mathrm{d}\xi \\ &= |x| \cdot \sqrt{\frac{2\pi}{S(z)}}\end{aligned}$$

Hence, letting $c(z) = \sqrt{\frac{2\pi}{S(z)}}$,

$$\mathcal{I}_m^{(2)} \le c(z) \cdot \int_{\mathbb{R}} |x| \cdot \left| f_{G_m(z)}(x) - f_{G(z)}(x) \right| \mathrm{d}x \tag{F.8}$$

For $x \in \mathbb{R}$,

$$f_{G_m(z)}(x) = \begin{cases} 0 & \text{if } x \notin [-\sqrt{m}\mu(z),\, \sqrt{m}(z-\mu(z))] \\ \frac{1}{\sqrt{m}z} \frac{\Gamma\left(\frac{\tau+\nu}{\alpha} m\right)}{\Gamma\left(\frac{\tau+\varrho\alpha}{\alpha} m\right) \Gamma\left(\frac{\nu-\varrho\alpha}{\alpha} m\right)} \cdot \left(\frac{\tau+\varrho\alpha}{\tau+\nu} - \frac{x}{\sqrt{m}z}\right)^{\frac{\tau+\varrho\alpha}{\alpha} m-1} \cdot \left(\frac{\nu-\varrho\alpha}{\tau+\nu} + \frac{x}{\sqrt{m}z}\right)^{\frac{\nu-\varrho\alpha}{\alpha} m-1} \\ \quad =: \psi_m(x) & \text{if } x \notin [-\sqrt{m}\mu(z),\, \sqrt{m}(z-\mu(z))] \end{cases}$$

and

$$f_{G(z)}(x) = \frac{1}{\sqrt{2\pi s^2(z)}} \exp\left\{-\frac{x^2}{2s^2(z)}\right\}.$$

Introduce $\gamma \in (0, 1/2)$ and let $A_m = c\, m^\gamma$, with c a positive constant such that $-\mu(z)\sqrt{m} < -A_m < 0 < A_m < (z-\mu(z))\sqrt{m}$ for all $z \in [\zeta_0, \zeta_1]$, and write the integral in the RHS of (F.8) as

$$\begin{aligned}&\int_{\mathbb{R}} |x| \cdot \left| f_{G_m(z)}(x) - f_{G(z)}(x) \right| \mathrm{d}x \\ &\quad = \int_{-\infty}^{-\sqrt{m}\,\mu(z)} -x\, f_{G(z)}(x)\, \mathrm{d}x + \int_{-\sqrt{m}\mu(z)}^{-A_m} |x| \cdot \left|\psi_m(x) - f_{G(z)}(x)\right| \mathrm{d}x \\ &\qquad + \int_{A_m}^{-A_m} |x| \cdot \left|\psi_m(x) - f_{G(z)}(x)\right| \mathrm{d}x \\ &\qquad + \int_{A_m}^{\sqrt{m}\,(z-\mu(z))} |x| \cdot \left|\psi_m(x) - f_{G(z)}(x)\right| \mathrm{d}x + \int_{\sqrt{m}\,(z-\mu(z))}^{+\infty} x\, f_{G(z)}(x)\, \mathrm{d}x \\ &\quad \le \int_{-\infty}^{-A_m} -x\, f_{G(z)}(x)\, \mathrm{d}x + \int_{-\sqrt{m}\,\mu(z)}^{-A_m} |x|\, \psi_m(x)\, \mathrm{d}x \\ &\qquad + \int_{A_m}^{-A_m} |x| \cdot \left|\psi_m(x) - f_{G(z)}(x)\right| \mathrm{d}x \\ &\qquad + \int_{A_m}^{\sqrt{m}\,(z-\mu(z))} |x|\, \psi_m(x)\, \mathrm{d}x + \int_{\sqrt{m}\,(z-\mu(z))}^{+\infty} x\, f_{G(z)}(x)\, \mathrm{d}x \\ &\quad = \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)}\end{aligned}$$

The terms (I) and (V) vanish exponentially fast as $m \to +\infty$, in fact for any $\zeta > 0$

$$\int_{\zeta}^{+\infty} x\, f_{G(z)}(x)\, \mathrm{d}x = \frac{s(z)}{\sqrt{2\pi}}\, e^{-\frac{\zeta^2}{s^2(z)}},$$

whence

$$(\mathrm{I}) = \frac{s(z)}{\sqrt{2\pi}}\, e^{-\frac{c^2 m^{2\gamma}}{s^2(z)}}$$

and

$$(\mathrm{V}) = \frac{s(z)}{\sqrt{2\pi}}\, e^{-\frac{c^2 m^{2\gamma}}{s^2(z)}}.$$

We now study (II) and (IV). By a straightforward application of Cauchy-Schwartz inequality,

$$\begin{aligned}(\mathrm{II}) &= \mathbb{E}\left[|G_m(z)| \cdot \mathbb{1}_{\{G_m(z)\in[-\sqrt{m}\,\mu(z),-A_m]\}}\right] \\ &\le \sqrt{\mathrm{Var}\left(G_m(z)\right)} \cdot \sqrt{P\left[G_m(z) \in [-\sqrt{m}\,\mu(z), -A_m]\right]}.\end{aligned}$$

Now,

$$\mathrm{Var}\left(G_m(z)\right) = mz^2\,\mathrm{Var}\left(B_m\right) = z^2\,\frac{\alpha\,(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{(\tau+\nu)^3}\,\left[1+O\left(\frac{1}{m}\right)\right] \le z^2\,\tilde{c}_1$$

for some suitable constant $\tilde{c}_1$. By Chebichev's inequality

$$\begin{aligned}P\left(G_m(z) \in [-\sqrt{m}\,\mu(z), -A_m]\right) &\le P\left(G_m(z) \le -A_m\right) + P\left(G_m(z) \ge A_m\right) \\ &\le \frac{\mathrm{Var}(G_m(z))}{A_m^2} \\ &\le z^2\,\tilde{c}_2\, m^{-2\gamma}\end{aligned}$$

for a suitable constant $\tilde{c}_2$. Combining these results we obtain

$$(\mathrm{II}) \le z^2\,\tilde{c}\, m^{-\gamma}$$

for $\tilde{c} = \sqrt{\tilde{c}_1\,\tilde{c}_2}$. The same argument allows to prove

$$(\mathrm{IV}) \le z^2\,\tilde{c}\, m^{-\gamma}$$

To assess the behavior of (III), use the well–known asymptotic expansion of the Gamma function for large argument [69, Equation (5.11.3)] and some simple algebraic rearrangements to write that, as $m \to +\infty$,

$$\begin{aligned}&\psi_m(x) \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{z^2\alpha(\tau+\nu)}}\cdot\left(\frac{\tau+\varrho\alpha}{\tau+\nu}-\frac{x}{\sqrt{m}\,z}\right)^{-1}\cdot\left(\frac{\nu-\varrho\alpha}{\tau+\nu}+\frac{x}{\sqrt{m}\,z}\right)^{-1} \\ &\quad\cdot\left\{\left(\frac{\tau+\nu}{\tau+\varrho\alpha}\right)^{\frac{\tau+\varrho\alpha}{\alpha}}\left(\frac{\tau+\nu}{\nu-\varrho\alpha}\right)^{\frac{\nu-\varrho\alpha}{\alpha}}\cdot\left(\frac{\tau+\varrho\alpha}{\tau+\nu}-\frac{x}{\sqrt{m}\,z}\right)^{\frac{\tau+\varrho\alpha}{\alpha}}\cdot\left(\frac{\nu-\varrho\alpha}{\tau+\nu}+\frac{x}{\sqrt{m}\,z}\right)^{\frac{\nu-\varrho\alpha}{\alpha}}\right\}^m \\ &\quad\cdot\mathfrak{R}_1^{(m)} \\ &= \exp\left\{m\cdot\left[\frac{\tau+\varrho\alpha}{\alpha}\log\left(1-\frac{x}{\sqrt{m}\,\mu(z)}\right)+\frac{\nu-\varrho\alpha}{\alpha}\log\left(1+\frac{x}{\sqrt{m}\,(z-\mu(z))}\right)\right]\right\} \\ &\quad\cdot\frac{1}{\sqrt{2\pi}}\,\Psi_m(z,x)\cdot\mathfrak{R}_1^{(m)}\end{aligned}$$

where

$$\Psi_m(z,x) = \sqrt{\frac{(\tau+\varrho\alpha)(\nu-\varrho\alpha)}{z^2\,\alpha(\tau+\nu)}}\cdot\frac{1}{\left(\frac{\tau+\varrho\alpha}{\tau+\nu}-\frac{x}{\sqrt{m}\,z}\right)\cdot\left(\frac{\nu-\varrho\alpha}{\tau+\nu}+\frac{x}{\sqrt{m}\,z}\right)}$$

and

$$\mathfrak{R}_1^{(m)} = \frac{\Gamma^*\left(\frac{\tau+\nu}{\alpha}\, m\right)}{\Gamma^*\left(\frac{\tau+\varrho\alpha}{\alpha}\, m\right)\, \Gamma^*\left(\frac{\nu-\varrho\alpha}{\alpha}\, m\right)}.$$

We focus first on the behavior of these last two terms as $m \to +\infty$; simple algebraic rearrangements allow to write

$$\begin{aligned}\Psi_m(z,x) &= \sqrt{\frac{\alpha(\tau+\nu)^3}{z^2(\tau+\varrho\alpha)(\nu-\varrho\alpha)}} \cdot \frac{1}{1+\frac{x(\tau+\nu)}{\sqrt{m}\, z}\left(\frac{1}{\nu-\varrho\alpha}-\frac{1}{\tau+\varrho\alpha}\right)-\frac{x^2(\tau+\nu)^2}{mz(\tau+\varrho\alpha)(\nu-\varrho\alpha)}} \\ &= \frac{1}{s(z)}\left[1+O\left(\frac{x}{m^{1/2}}\right)\right].\end{aligned}$$

uniformly for $z \in [\zeta_0, \zeta_1]$. Resorting again to [69, Equations (5.11.3) and (5.11.4)], we also obtain

$$\mathfrak{R}_1^{(m)} = \frac{1+\frac{\alpha}{12\,(\tau+\nu)\, m}+O\left(m^{-2}\right)}{1+\frac{\alpha\,(\tau+\nu)}{12\,(\tau+\varrho\alpha)(\nu-\varrho\alpha)\, m}+O\left(m^{-2}\right)} = 1+O\left(m^{-1}\right).$$

uniformly for $z \in [\zeta_0, \zeta_1]$. For the exponential term, use Taylor's expansion of the logarithm around 1 to write

$$\begin{aligned}&\exp\left\{m\cdot\left[\frac{\tau+\varrho\alpha}{\alpha}\log\left(1-\frac{x}{\sqrt{m}\,\mu(z)}\right)+\frac{\nu-\varrho\alpha}{\alpha}\log\left(1+\frac{x}{\sqrt{m}\,(z-\mu(z))}\right)\right]\right\} \\ &= \exp\left\{m\cdot\left[\frac{\tau+\varrho\alpha}{\alpha}\left(-\frac{x}{\sqrt{m}\,\mu(z)}-\frac{x^2}{2m\,\mu^2(z)}\right)\right.\right. \\ &\qquad \left.\left.+\frac{\nu-\varrho\alpha}{\alpha}\left(\frac{x}{\sqrt{m}\,(z-\mu(z))}-\frac{x^2}{2m\,(z-\mu(z))^2}\right)\right]\right\}\cdot\mathfrak{R}_2^{(m)}(x,z) \\ &= \exp\left\{-\frac{x^2}{2}\cdot\frac{1}{s^2(z)}\right\}\cdot\mathfrak{R}_2^{(m)}(x,z)\end{aligned}$$

where

$$\begin{aligned}\mathfrak{R}_2^{(m)}(x,z) = \exp\Bigg\{m\cdot\Bigg[&\frac{\tau+\varrho\alpha}{\alpha}\left(\log\left(1-\frac{x}{\sqrt{m}\,\mu(z)}\right)+\frac{x}{\sqrt{m}\,\mu(z)}+\frac{x^2}{2m\,\mu^2(z)}\right) \\ &+\frac{\nu-\varrho\alpha}{\alpha}\left(\log\left(1+\frac{x}{\sqrt{m}\,(z-\mu(z))}\right)-\frac{x}{\sqrt{m}\,(z-\mu(z))}+\frac{x^2}{2m\,(z-\mu(z))^2}\right)\Bigg]\Bigg\}\end{aligned}$$

As $m \to +\infty$,

$$\mathfrak{R}_2^{(m)}(x,z) = 1+O\left(\frac{x^3}{m^{1/2}}\right)$$

uniformly for $z \in [\zeta_0, \zeta_1]$. Combining all the above results we can write

$$\psi_m(x) = \frac{1}{\sqrt{2\pi}\, s(z)}\exp\left\{-\frac{x^2}{2}\cdot\frac{1}{s^2(z)}\right\}\left[1+\mathcal{R}_m(x,z)\right] = f_{G(z)}(x)\left[1+\mathcal{R}_m(x,z)\right]$$

with

$$\mathcal{R}_m(x,z) = O\left(\frac{x^3}{m^{1/2}}\right)$$

uniformly for $z \in [\zeta_0, \zeta_1]$. Resorting again to [21, Lemma A.11], this in turn implies

$$\left|\psi_m(x) - f_{G(z)}(x)\right| \le \frac{C\, x^3}{m^{1/2}}$$

for every $x \in [-A_m, A_m]$ and $z \in [\zeta_0, \zeta_1]$, for some constant $C > 0$, whence

$$\begin{aligned}(\text{III}) &= \int_{A_m}^{-A_m} |x| \cdot \left|\psi_m(x) - f_{G(z)}(x)\right| \, \mathrm{d}x \\ &\leq \int_{A_m}^{-A_m} |x| \cdot \frac{C\, x^3}{m^{1/2}} \, \mathrm{d}x \\ &\leq \frac{\mathcal{C}\, \mathbb{E}\left[G(z)^4\right]}{m^{1/2}}\end{aligned}$$

In conclusion,

$$\mathcal{I}_m^{(2)}(z) \leq \frac{s(z)}{\sqrt{2\pi}} \left[e^{-\frac{c^2\, m^{2\gamma}}{s^2(z)}} + e^{-\frac{c^2\, m^{2\gamma}}{s^2(z)}} \right] + 2z^2\, c\, m^{-\gamma} + \mathcal{C}\, \mathbb{E}\left[G(z)^4\right]\, m^{-1/2}$$

Recalling that $\gamma \in \left(0, \frac{1}{2}\right)$, we conclude

$$\mathcal{I}_m^{(2)}(z) \leq C_2(z)\, m^{-\gamma} \leq c_2\, m^{-\gamma}$$

for some continuous function $C_2 : [\zeta_0, \zeta_1] \to (0, +\infty)$ depending only on γ and $c_2 = \max_{z \in [\zeta_0, \zeta_1]} C_2(z)$. ■

G Additional numerical illustrations

This section collects additional figures displaying the performance gap between the Mittag-Leffler credible intervals and the Gaussian confidence intervals, and its behavior with respect to the additional sample size m on the datasets, both synthetic and real, considered in Section 4.4.

G.1 Synthetic data

Figure 5.4 complements the analysis of the synthetic datasets in Section 4.4 (Table 4.2 and Table 4.3); in particular, it displays the coverage of the Mittag-Leffler credible interval (blue) and of the Gaussian credible interval (red) as a function of $m \in [0, 5n]$. The coverages are evaluated at a uniform mesh of 50 points over $[0, 5n]$, as for Figure 4.1. Monte Carlo algorithms to obtain exact credible intervals and Mittag-Leffler credible intervals apply 2000 Monte Carlo samples.

Figure 5.4 confirms the behavior of the coverage observed in Table 4.2: the coverage of the Gaussian credible intervals is nearly constant in m, with values oscillating between 95% and 100%. Instead, the coverage of the Mittag-Leffler credible intervals is, for all values of m, lower than that of the corresponding Gaussian credible intervals; such a coverage increases in m.

G.2 Real data

Figure 5.5 complements the analysis of the real EST datasets in Section 4.4 (Table 4.5 and Table 4.6); in particular, for the tomato flower, *Mastigamoeba*, *Mastigamoeba* normalized, and *Naegleria* anaerobic EST datasets, it displays BNP estimates of $\mathcal{K}_{n,m}$ with 95% exact credible intervals, Mittag-Leffler credible intervals and Gaussian credible intervals as a function of $m \in [0, 5n]$. Credible intervals are evaluated at a uniform mesh of 50 points

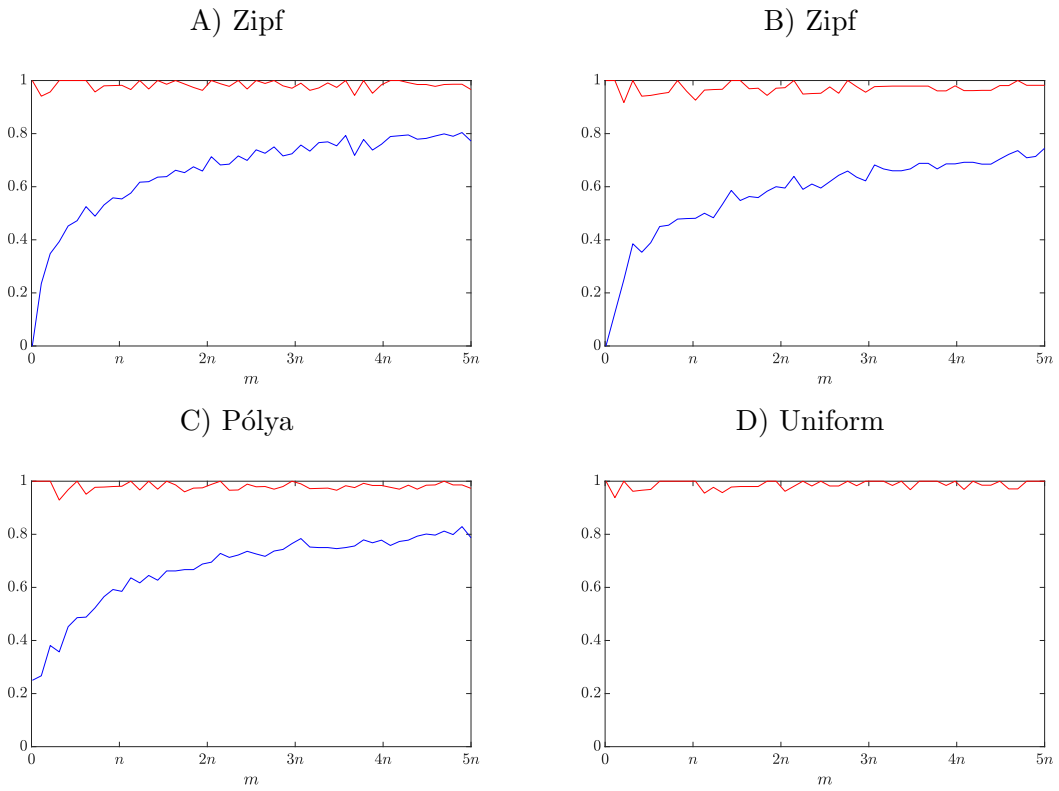

Figure 5.4: Coverage of the Mittag-Leffler credible interval (blue) and of the Gaussian credible interval (red) as a function of $m \in [0, 5n]$.

over $[0, 5n]$, as for Figure 4.2. Monte Carlo algorithms to obtain exact credible intervals and Mittag-Leffler credible intervals apply 2000 Monte Carlo samples.

For the tomato flower, *Mastigamoeba*, *Mastigamoeba* normalized, *Naegleria* aerobic and *Naegleria* anaerobic EST datasets, Figure 5.6 shows the coverage of the Mittag-Leffler credible interval (blue) and of the Gaussian credible interval (red) as a function of $m \in [0, 5n]$. The coverages are evaluated at a uniform mesh of 50 points over $[0, 5n]$, the same values of m considered for Figure 4.2 and Figure 5.5.

Figure 5.6 confirms the behavior of the coverage observed in Table 4.5: the coverage of the Gaussian credible intervals is nearly constant in m, with values oscillating between 95% and 100%. Instead, the coverage of the Mittag-Lefffler credible intervals is, for all values of m, lower than that of the corresponding Gaussian credible intervals; such a coverage increases in m.

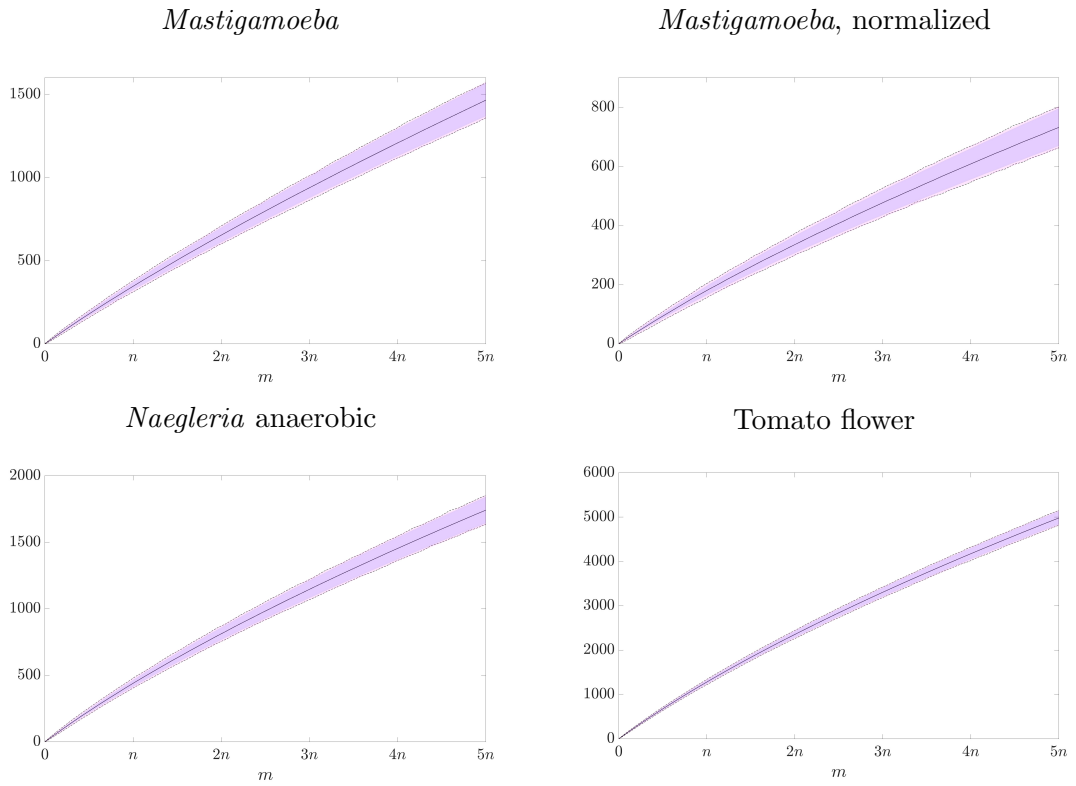

Figure 5.5: BNP estimates of $\mathcal{K}_{n,m}$ (solid line –) with 95% exact credible intervals (dashed line - -), Mittag-Leffler credible intervals (violet) and Gaussian credible intervals (pink), as a function of $m \in [0, 5n]$.

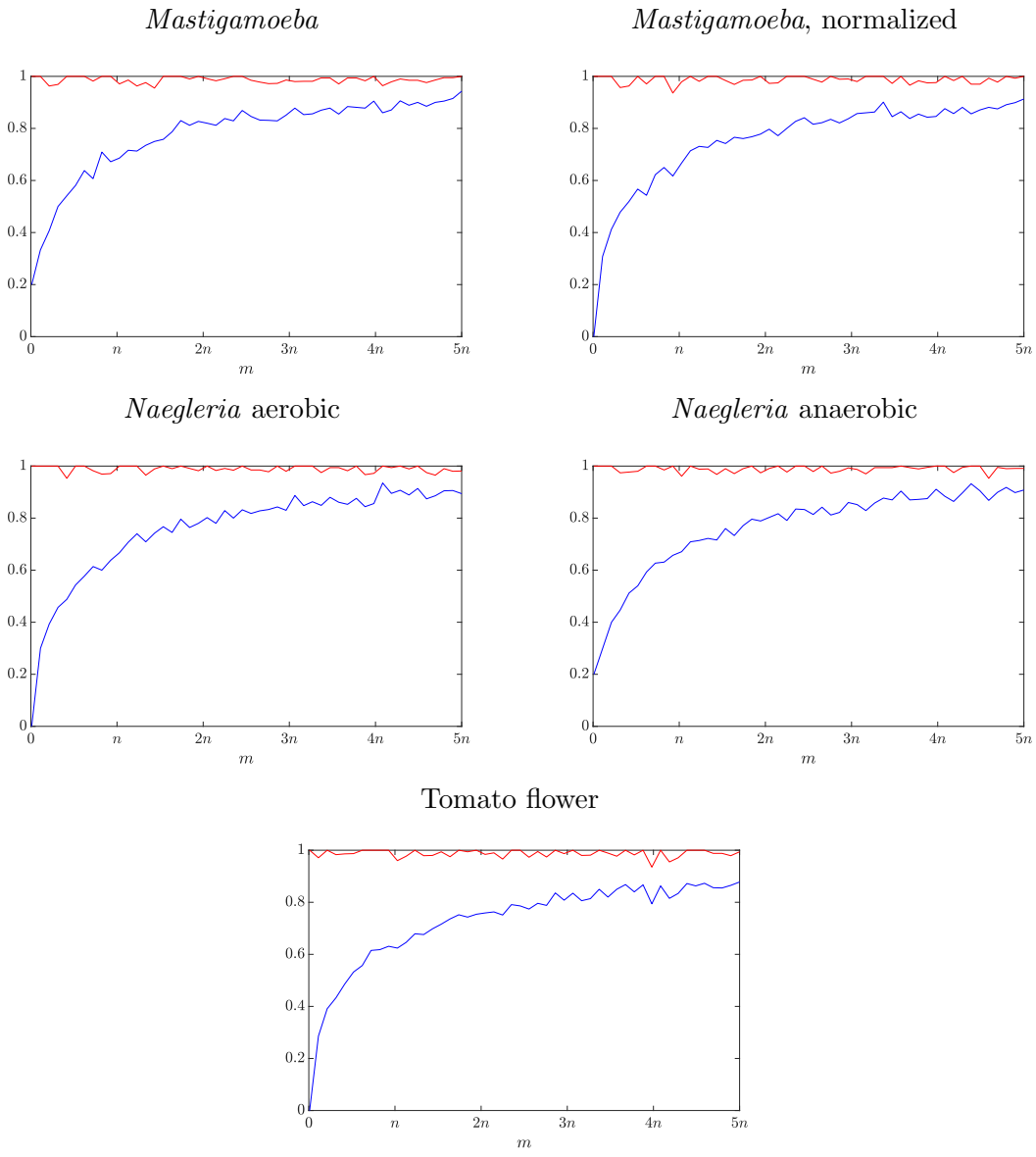

Figure 5.6: Coverage of the Mittag-Leffler credible interval (blue) and of the Gaussian credible interval (red) as a function of $m \in [0, 5n]$.

Bibliography

[1] Abramowitz, M. and Stegun, I.A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Applied Mathematics Series 55, Washington, D.C.

[2] Arratia, R., Barbour, A.D. and Tavaré, S. (1992). Poisson process approximations for the Ewens sampling formula. *The Annals of Applied Probability* **2**, 519–535

[3] Arratia, R., Barbour, A.D. and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics.

[4] Balocchi, C., Favaro, S. and Naulet, Z. (2024). Bayesian nonparametric inference for "species-sampling" problems. *Statistical Science*, to appear.

[5] Ben-Hamou, A., Boucheron, S. and Gassiat, E. (2018). Pattern coding meets censoring: (almost) adaptive coding on countable alphabets. *Preprint arXiv:1608.08367*.

[6] Belkiˇc, D. (2019) All the trinomial roots, their powers and logarithms from the Lambert series, Bell polynomials and Fox-Wright function: illustration for genome multiplicity in survival of irradiated cells. *Journal of Mathematical Chemistry* **57**, 59–106.

[7] Beraha, M. and Favaro, S. (2025). Large-scale entity resolution via microclustering Ewens-Pitman random partitions. *Preprint arXiv:2507.18101*.

[8] Bercu B., Contardi C., Dolera E., Favaro S. (2026). *A Central Limit Theorem for the Ewens-Pitman random partition in the large–θ regime via a martingale approach*. Preprint: *arXiv: 2601.18935*

[9] Bercu, B. and Favaro, S. (2024). A martingale approach to Gaussian fluctuations and laws of iterated logarithm for Ewens-Pitman model. *Stochastic Processes and their Applications* **178**, 104493.

[10] Bercu, B. and Favaro, S. (2025). A new look on large deviations and concentration inequalities for the Ewens-Pitman model. *Preprint arXiv:2503.06783*.

[11] Brown, B. M. (1971). Martingale central limit theorems, *Ann. Math. Statist.*, **42** , 59–66.

[12] Bubeck, S., Ernst, D., and Garivier, A. (2013). Optimal discovery with probabilistic expert advice: finite time analysis and macroscopic optimality. *Journal of Machine Learning Research* **14**, 601–623.

[13] Bunge, J. and Fitzpatrick, M. (1993) Estimating the number of species: a review. *Journal of the American Statistical Association* **88**, 364-373.

[14] Cai, D., Mitzenmacher, M. and Adams, R.P. (2018). A Bayesian nonparametric view on count–min sketch. *Advances in Neural Information Processing Systems*, **31**.

[15] Cereda, G. (2017) Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach). *Scandinavian Journal of Statistics* **44**, 230-248.

[16] Charalambides (2005) *Combinatorial methods in discrete distributions.* Wiley.

[17] Charalambides, C.A. (2007). Distributions of random partitions and their applications. *Methodology and Computing in Applied Probability* **9**, 163–193.

[18] Chao, A. (2017) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**, 256–270.

[19] Chao, A. and Lee, S. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* **87**, 210–217.

[20] Chow, Y.S. and Teicher, H. (1997) *Probability theory. Independence, interchangeability, martingales.* Springer.

[21] Contardi C., Dolera E., Favaro S. (2025) Laws of large numbers and central limit theorem for Ewens-Pitman model. *Electron. J. Probab.* **30** 1 - 51, 2025

[22] Contardi, C., Dolera, E. and Favaro, S. (2025). Gaussian credible intervals in Bayesian nonparametric estimation of the unseen. *Preprint arXiv:2501.16008.*

[23] Crane, H. (2016). The ubiquitous Ewens sampling formula. *Statistical Science* **31**, 1–19.

[24] Daley, T. and Smith, A.D. (2013). Predicting the molecular complexity of sequencing libraries. *Nature Methods* **10**, 325–327.

[25] Dawson, D.A. and Feng, S. (2006). Asymptotic behavior of Poisson-Dirichlet distribution for large mutation rate. *The Annals of Applied Probability* **16**, 562–582.

[26] Deng, C. Daley, T., De Sena Brandine, G. and Smith, A.D. (2019). Molecular heterogeneity in large-scale biological data: techniques and applications. *Annual Review of Biomedical Data Science* **2**, 39–67.

[27] Dolera, E. and Favaro, S. (2020). A Berry–Esseen theorem for Pitman's α–diversity. *The Annals of Applied Probability* **30**, 847–869.

[28] Dolera, E. and Favaro, S. (2021). A compound Poisson perspective of Ewens-Pitman sampling model. *Mathematics* **9**, 2820.

[29] Dolera, E. and Favaro, S. (2020b). Rates of convergence in de Finetti's representation theorem, and Hausdorff moment problem. *Bernoulli* **26**, 1294–1322.

[30] Duflo, M. (1997) *Random iterative models.* Springer-Verlag.

[31] Dudley, R.M. (2002). *Real Analysis and Probability.* Cambridge University Press.

[32] Efron, B. and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63**, 435–447.

[33] Ewens, W. (1972). The sampling theory or selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

[34] Favaro, S., Lijoi, A., Mena, R.H. and Prünster, I. (2009). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *Journal of the Royal Statistical Society Series B* **71**, 992–1008.

[35] Favaro, S., Lijoi, A. and Prünster, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *The Annals of Applied Probability* **23**, 1721–1754.

[36] Favaro, S. and Feng, S (2014). Asymptotics for the number of blocks in a conditional Ewens-Pitman sampling model. *Electronic Journal of Probability* **19**, 1–15.

[37] Favaro, S., Feng, S. and Gao. F. (2018). Moderate deviations for Ewens-Pitman sampling models. *Sankhya A* **80**, 330–341.

[38] Favaro, S., Naulet, Z. (2023). Near-optimal estimation of the unseen under regularly varying tail populations. *Bernoulli* **29**, 3423–3442.

[39] Favaro, S., Naulet, Z. (2024). Optimal estimation of high-order missing masses, and the rare-type match problem. *Preprint arXiv:2306.14998*.

[40] Feng, S. (2007). Large deviations associated with Poisson–Dirichlet distribution and Ewens sampling formula. *The Annals of Applied Probability***17**, 1570–1595.

[41] Feng, S. (2007). Large deviations for Dirichlet processes and Poisson-Dirichlet distribution with two parameters. *Electronic Journal of Probability***12**, 787–807.

[42] Feng, S. and Gao, F.Q. (2008). Moderate deviations for Poisson-Dirichlet distribution. *The Annals of Applied Probability***18**, 1794–1824.

[43] Feng, S. (2010). *The Poisson-Dirichlet distribution and Related Topics*. Springer.

[44] Feng, S. and Gao, F.Q. (2010). Asymptotic results for the two-parameter Poisson-Dirichlet distribution. *Stochastic Processes and their Applications***120**, 1159–1177.

[45] Feng, S. and Hoppe, F.M. (1998). Large deviation principles for some random combinatorial structures in population genetics and Brownian motion. *The Annals of Applied Probability***8**, 975–994.

[46] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

[47] Fisher, R.A., Corbet, A.S. and Williams, C.B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **12**, 42–58.

[48] Florencio, D. and Herley, C. (2007). A large-scale study of web password habits. *Proceedings of the International Conference on World Wide Web*.

[49] Gao, Z., Tseng, C.H., Pei, Z. an Blaser, M.J. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proceedings of the National Academy of Sciences of USA* **104**, 2927–2932.

[50] Good, I.J.(1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237-264.

[51] Good, I.J. and Toulmin, G.H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* **43**, 45–63.

[52] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, 7th ed. Academic Press, Amsterdam.

[53] Griffiths, R.C. (1979). On the distribution of allele frequencies in a diffusion model. *Theoretical Population Biology***15**, 140–158.

[54] Haas, P.J., Naughton, J.F., Seshadri, S. and Stokes, L. (1995). Sampling-based estimation of the number of distinct values of an attribute. *Proceedings of the Very Large Data Bases Conference*.

[55] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.

[56] Hao, Y. and Li, P. (2020). Optimal prediction of the number of unseen species with multiplicity. *Advances in Neural Information Processing Systems*, **33**.

[57] Heyde, C.C. (1977). On central limit and iterated logarithm supplements to the martingale convergence theorem. *Journal of Applied Probability* **14**, 758–775.

[58] Huber, P. (1981). *Robust statistics*. Wiley.

[59] Ionita-Laza, I., Lange, C. and Laird, N.M. (2009). Estimating the number of unseen variants in the human genome. *Proceedings of the National Academy of Sciences of USA* **106**, 5008–5013.

[60] Joyce, P., Krone, S.M. and Kurtz, T.G. (2002). Gaussian limits associated with the Poisson–Dirichlet distribution and the Ewens sampling formula. *The Annals of Applied Probability* **12**, 101–124.

[61] Kalai, A.T. and Vempala, S.S. (2024). Calibrated language models must hallucinate. *Proceeding of the Annual ACM Symposium on Theory of Computing*,

[62] Kilbas, A.A. and Saigo, M. (2004). *H-transforms. Theory and applications*. Chapman & Hall/CRC.

[63] Kingman, J.F.C. (1975). Random discrete distributions. *Journal of the Royal Statistical Society Series B* **37**, 1–15.

[64] Kingman, J.F.C. (1978). The representation of partition structures. *J. London Math. Soc.,* 18:374–380.

[65] Korwar, R.M. and Hollander, M. (1973). Contribution to the theory of Dirichlet processes. *The Annals of Probability* **1**, 705–711.

[66] Kroes, I., Lepp, P.W. and Relman, D.A. (1999). Bacterial diversity within the human subgingival crevice. *Proceeding of the National Academy of Sciences of USA* **96**, 14547–14552.

[67] Lijoi, A., Mena, R.H. and Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering new species. *Biometrika* **94**, 769–786.

[68] Motwani, S. and Vassilvitskii, S. (2006) Distinct value estimators in power law distributions. *Proceedings of Analytic Algorithms and Combinatorics*, **8**.

[69] *NIST Digital Library of Mathematical Functions.* https://dlmf.nist.gov/, Release 1.2.1.

[70] Ohannessian, M.I. and Dahleh, M.A. (2012). Rare probability estimation under regularly varying heavy tails. *Proceedings of the Conference on Learning Theory*, **23**.

[71] Orlitsky, A., Santhanam, N.P. and Zhang, J. (2004). Universal compression of memoryless sources over unknown alphabets. *IEEE Transaction on Information Theory* **50**, 1469–1481.

[72] Orlitsky, A., Suresh, A.T. and Wu, Y. (2017). Optimal prediction of the number of unseen species. *Proceeding of the National Academy of Sciences of USA* **113**, 13283–13288.

[73] Paris, R.B. and Kaminski, D. (2001). *Asymptotics and Mellin–Barnes integrales*. Encyclopedia of Mathematics and its Applications, 85.

[74] Paris, R.B. (2021). The asymptotic expansion of Kraetzel's integral and an integral related to an extension of the Whittaker function. *Preprint ArXiv: 2112.02928*

[75] Peng, Z. and Zhou, Y. (2025). Precise Deviations for the Ewens-Pitman Model. *Preprint ArXiv: 2512.12323*

[76] Pereira, A., Oliveira, R.I. and Ribeiro, R. (2022). Concentration in the generalized Chinese restaurant process. *Sankhya A* **80**, 628-670.

[77] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields* **92**, 21–39.

[78] Petrov, V.V. (1975). *Sums of independent random variables*. Springer.

[79] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102**, 145–158.

[80] Pitman, J. (2006). *Combinatorial stochastic processes*. Lecture Notes in Mathematics, Springer Verlag.

[81] Pitman, J. and Yor, M. (1997). The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* **25**, 855–900.

[82] Pollard, H. (1946). The representation of e^{-x^λ} as a Laplace integral. *Bulletin of the American Mathematical Society* **52**, 908–910.

[83] Qu, Y., Dassios, A. and Zhao, H. (2021). Random variate generation for exponential and Gamma tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation* **31**, 19.

[84] Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Pertea, G., Sultana, R. and White, J. (2000). The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Research* **29**, 159–164.

[85] Shao, J. (1999) Mathematical Statistics. Springer Texts in Statistics, Springer Verlag, New York.

[86] Strassen, V. (1967). Almost sure behavior of sums of independent random variables and martingales. In *Berkeley Symposium on Mathematical Statistics and Probability* **5**, 315–343.

[87] Teh, Y. W. (2010). Bayesian nonparametric lecture notes. *Technical report*. University of Oxford.

[88] Susko, E. and Roger, A.J. (2004). Estimating and comparing the rates of gene discovery and expressed sequence tag (EST) frequencies in EST surveys. *Bioinformatics* **20**, 2279–2287.

[89] Tricomi, F.G. and Erdèlyi, A. (1951). The asymptotic expansion of a ratio of Gamma functions. *Pacific Journal of Mathematics* **1**, 133–142.

[90] Thisted, R. and Efron, B. (1987). Did Shakespeare write a newly-discovered poem? *Biometrika* **74**, 445–455.

[91] Tsukuda, K. (2017). Estimating the large mutation parameter of the Ewens sampling formula. *Journal of Applied Probability* **54**, 42–54.

[92] Tsukuda, K. (2019). On Poisson approximations for the Ewens sampling formula when the mutation parameter grows with the sample size. *Annals of Applied Probability* **29**, 1188–1232

[93] Watterson. G.A. and Guess, H.A. (1977). Is the most frequent allele the oldest? *Theoretical Population Biology* **11**, 141–160.

[94] Wu, Y. and Yang, P. (2016). Minimax rates of entropy estimation on large alphabets via best polynomial approximation. *IEEE Transactions on Information Theory* **62**, 3702–3720.

[95] Wu, Y. and Yang, P. (2019). Chebyshev polynomials, moment matching, and optimal estimation of the unseen. *The Annals of Statistics* **47**, 857–883.

[96] Zabell, S.L. (2005). *Symmetry and its discontents: essays on the history of inductive probability*. Cambridge University Press.

[97] Zolotarev, V.M. (1986). *One-dimensional Stable distributions*. American Mathematical Society.