UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A martingale approach to Gaussian fluctuations and laws of iterated logarithm for Ewens-Pitman model

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/2042250 since 2024-12-27T20:04:50Z

Published version:

DOI:10.1016/j.spa.2024.104493

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

02 May 2026

A martingale approach to Gaussian fluctuations and laws of iterated logarithm for Ewens-Pitman model

Bernard Bercu[a,*], Stefano Favaro[b,**]

[a] *Institut de Mathématiques de Bordeaux, Université de Bordeaux, 351 Cours de la Libération, F33405 Bordeaux, France*
[b] *Department of Economics and Statistics, University of Torino and Collegio Carlo Alberto, Corso Unione Sovietica 218/bis, 10134 Torino, Italy*

Abstract

The Ewens-Pitman model refers to a distribution for random partitions of $[n] = \{1, \dots, n\}$, which is indexed by a pair of parameters $\alpha \in [0, 1)$ and $\theta > -\alpha$, with $\alpha = 0$ corresponding to the Ewens model in population genetics. The large n asymptotic properties of the Ewens-Pitman model have been the subject of numerous studies, with the focus being on the number K_n of partition sets and the number $K_{r,n}$ of partition subsets of size r, for $r = 1, \dots, n$. While for $\alpha = 0$ asymptotic results have been obtained in terms of almost-sure convergence and Gaussian fluctuations, for $\alpha \in (0, 1)$ only almost-sure convergences are available, with the proof for $K_{r,n}$ being given only as a sketch. In this paper, we make use of martingales to develop a unified and comprehensive treatment of the large n asymptotic behaviours of K_n and $K_{r,n}$ for $\alpha \in (0, 1)$, providing alternative, and rigorous, proofs of the almost-sure convergences of K_n and $K_{r,n}$, and covering the gap of Gaussian fluctuations. We also obtain new laws of the iterated logarithm for K_n and $K_{r,n}$.

Keywords: Almost-sure limit, Ewens-Pitman model, exchangeable random partition, Gaussian fluctuation, law of iterated logarithm, martingale, Mittag-Leffler distribution
2010 MSC: 60F15, 60F05

---

*Corresponding author.
**Also affiliated to IMATI-CNR "Enrico Magenes" (Milan, Italy).
Email addresses: `bernard.bercu@math.u-bordeaux.fr` (Bernard Bercu), `stefano.favaro@unito.it` (Stefano Favaro)

1. Introduction

The Ewens-Pitman (EP) model refers to a distribution for random partitions, introduced by Pitman [20] as a generalization of the celebrated Ewens model in population genetics [9]. For $n \in \mathbb{N}$, consider a random partition of $[n] = \{1, \ldots, n\}$ into $K_n \in \{1, \ldots, n\}$ partition subsets $\{A_1, \ldots, A_{K_n}\}$ of corresponding sizes $\mathbf{N}_n = (N_{1,n}, \ldots, N_{K_n,n})$, where $N_{k,n}$ is the cardinal number of set A_k, for $k = 1, \ldots, K_n$, such that

$$n = \sum_{k=1}^{K_n} N_{k,n}.$$

For any real numbers α in $[0, 1)$ and $\theta > -\alpha$, the EP model assigns to the random vector $(K_n, \mathbf{N}_n)$ the probability

$$\mathbb{P}(K_n = k, \mathbf{N}_n = (n_1, \ldots, n_k)) = \frac{n!}{k!} \frac{\left(\frac{\theta}{\alpha}\right)^{(k)}}{(\theta)^{(n)}} \prod_{i=1}^{k} \frac{\alpha(1-\alpha)^{(n_i-1)}}{n_i!} \tag{1}$$

where, for any $a \in \mathbb{R}$, $(a)^{(n)}$ stands for the rising factorial of a of order n, that is $(a)^{(n)} = a(a+1)\cdots(a+n-1)$. Denote by $K_{r,n}$ the number of partition subsets of size r, given for all $r = 1, \ldots, n$, by

$$K_{r,n} = \sum_{k=1}^{K_n} \mathrm{I}_{\{N_{k,n}=r\}}$$

where I is the indicator function. One can easily see that

$$n = \sum_{r=1}^{n} r K_{r,n} \qquad \text{and} \qquad K_n = \sum_{r=1}^{n} K_{r,n}.$$

The distribution of $\mathbf{K}_n = (K_{1,n}, \ldots, K_{n,n})$ is known as EP sampling formula or frequency-of-frequencies distribution, and it follows by means of a combinatorial rearrangement of (1). The Ewens model is recovered from (1) by setting $\alpha = 0$. We refer the reader to Pitman [20, 23] for more details. The joint distribution (1) admits several constructions, with the most common being a sequential or generative construction through the Chinese restaurant process [20, 14, 24] and a Poisson process construction through random sampling from the Pitman-Yor random probability measure [19, 22], see also Dolera and Favaro [7] for a construction through a class of negative Binomial compound Poisson sampling models [5]. The EP model plays a critical role in a variety of research fields, such as population genetics, Bayesian nonparametric statistics, excursion theory, combinatorics, machine learning and statistical physics. We refer to the monograph by Pitman [23, Chapter 3 and Chapter 4] for a comprehensive account on the EP model and generalizations thereof.

Under the EP model, there have been several studies on the large n asymptotic behaviour of K_n and the $K_{r,n}$'s; see Pitman [23, Chapter 3]. In particular, for $\alpha = 0$, Korwar and Hollander [17, Theorem 2.3] exploited the fact that K_n is a sum of n independent Bernoulli random variables, and showed that

$$\lim_{n\to+\infty} \frac{K_n}{\log n} = \theta \qquad \text{a.s.} \tag{2}$$

Furthermore, it follows directly from Lindeberg-Lévy central limit theorem that

$$\frac{K_n - \theta \log n}{\sqrt{\log n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \sqrt{\theta}\,\mathcal{N}(0,1), \tag{3}$$

with $\mathcal{N}(0,1)$ being a standard Gaussian random variable. See Feng and Hoppe [14] and references therein for some functional versions of both (2) and (3). As regard to the $K_{r,n}$'s, from Arratia et al. [1, Theorem 1] it follows that

$$K_{r,n} \xrightarrow[n\to+\infty]{\mathcal{L}} Z_r \tag{4}$$

and

$$(K_{1,n}, K_{2,n}, \ldots) \xrightarrow[n\to+\infty]{\mathcal{L}} (Z_1, Z_2, \ldots), \tag{5}$$

where the Z_r's are independent Poisson random variables with $\mathbb{E}[Z_r] = \theta/r$, for all $r \geq 1$. We refer to Arratia et al. [2] for various generalizations and refinements of these asymptotic results, and their interplay with combinatorics.

For any $\alpha \in (0,1)$, K_n is no longer a sum of independent Bernoulli random variables. However, Pitman [23, Theorem 3.8] made use of a martingale construction for K_n, based on a likelihood ratio defined through the distribution (1), and applied the strong law of large numbers for martingales to prove that

$$\lim_{n\to+\infty} \frac{K_n}{n^\alpha} = S_{\alpha,\theta} \qquad \text{a.s.,} \tag{6}$$

where $S_{\alpha,\theta}$ is a positive and almost surely finite random variable. If f_α is the positive α-Stable density function, then the distribution of $S_{\alpha,\theta}$ has density function

$$f_{S_{\alpha,\theta}}(s) = \frac{\Gamma(\theta+1)}{\alpha\Gamma(\theta/\alpha+1)} s^{\frac{\theta-1}{\alpha}-1} f_\alpha(s^{-1/\alpha}), \tag{7}$$

where Γ stands for the Euler Gamma function. Precisely, (7) is a generalization of the Mittag-Leffler density function, which is recovered by setting $\theta = 0$ [25]. As regard to the $K_{r,n}$'s, from Pitman [23, Lemma 3.11] it follows that

$$\lim_{n\to+\infty} \frac{K_{r,n}}{n^\alpha} = p_\alpha(r) S_{\alpha,\theta} \qquad \text{a.s.}, \tag{8}$$

where

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

The almost-sure convergences (6) and (8) are the natural counterparts of (2) and (4), respectively, though at different scales. Instead, it is still an open problem of obtaining a counterpart of (3), as well as Gaussian fluctuations for $K_{r,n}$.

1.1. *Our contributions*

For $\alpha \in (0,1)$ and $\theta > -\alpha$, we make use of martingales to develop a unified and comprehensive treatment of the large n asymptotic behaviours of K_n and $K_{r,n}$, providing alternative, and rigorous, proofs of the almost-sure limits (6) and (8), and covering the gap of Gaussian fluctuations. In particular, we propose a new martingale construction for K_n, simpler than that applied in Pitman [23, Theorem 3.8], which leads to an alternative proof of (6), still relying on the strong law of large numbers for martingales. Further, by exploiting our martingale construction, as well as (6), we show that

$$\sqrt{n^\alpha}\left(\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \sqrt{S'_{\alpha,\theta}}\,\mathcal{N}(0,1), \tag{9}$$

where $S'_{\alpha,\theta}$ is random variable independent of $\mathcal{N}(0,1)$ and sharing the same distribution as $S_{\alpha,\theta}$. We also prove the law of the iterated logarithm for K_n, i.e.,

$$\limsup_{n\to\infty} \frac{\left(K_n - n^\alpha S_{\alpha,\theta}\right)^2}{2n^\alpha \log\log n} = S_{\alpha,\theta} \qquad \text{a.s.} \tag{10}$$

Then, we extend our analysis to the $K_{r,n}$'s, for which only a sketch of the proof of (8) is given in Pitman [23, Lemma 3.11]. In particular, we introduce a martingale construction for $K_{r,n}$, providing a rigorous proof of (8) by means of the strong law of large numbers for martingales. Further, by exploiting our martingale construction and (8), we establish a Gaussian fluctuation as well as the law of iterated logarithm for $K_{r,n}$. Critical for our study is the work of Heyde [16], which provides sufficient conditions to achieve Gaussian fluctuations and laws of iterated logarithm from the martingale convergence theorem.

1.2. Organization of the paper

The paper is structured as follows. In Section 2, we present the martingale construction for K_n and the proofs of the almost-sure convergence, as well as the $\mathbb{L}^p$ convergence, the Gaussian fluctuation, and the law of iterated logarithm. Section 3 contains an analogous asymptotic analysis for $K_{r,n}$. In Section 4, we discuss some directions of future research. The appendices are devoted to the sequential construction of the EP model and an alternative proof, without relying on martingales, of the $\mathbb{L}^2$ convergence of K_n and $K_{r,n}$.

2. Asymptotic results for the number of partition sets

We start by introducing the keystone martingale construction for K_n. This is a critical tool at the basis of all the results in this section. From the sequential construction of the EP model [20, Proposition 9], for all $n \geq 1$,

$$\mathbb{P}(K_{n+1} = K_n + k \mid K_n) = \begin{cases} \dfrac{\alpha K_n + \theta}{n + \theta} & \text{if} \quad k = 1, \\[2ex] \dfrac{n - \alpha K_n}{n + \theta} & \text{if} \quad k = 0. \end{cases} \tag{11}$$

We refer the reader to Appendix A for more details on (11). Equation (11) simply means that

$$K_{n+1} = K_n + \xi_{n+1} \tag{12}$$

where the conditional distribution of the random variable ξ_{n+1}, given the σ-algebra $\mathcal{F}_n = \sigma(K_1, \ldots, K_n)$, is the Bernoulli $\mathcal{B}(p_n)$ distribution with parameter

$$p_n = \frac{\alpha K_n + \theta}{n + \theta}. \tag{13}$$

According to the above definition of the sequence (ξ_n), as $\mathbb{E}[\xi_{n+1} \,|\, \mathcal{F}_n] = p_n$, we have almost surely

$$\mathbb{E}[K_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[K_n + \xi_{n+1} \,|\, \mathcal{F}_n] = K_n + p_n = \beta_n K_n + \frac{\theta}{n+\theta} \tag{14}$$

where

$$\beta_n = 1 + \frac{\alpha}{n+\theta}. \tag{15}$$

Consequently, we obtain from (14) that almost surely

$$\mathbb{E}\left[K_{n+1} + \frac{\theta}{\alpha} \,\Big|\, \mathcal{F}_n\right] = \beta_n K_n + \frac{\theta}{n+\theta} + \frac{\theta}{\alpha} = \beta_n \left(K_n + \frac{\theta}{\alpha}\right). \tag{16}$$

Let (b_n) be the sequence defined by $b_1 = 1$ and for all $n \geq 2$,

$$b_n = \prod_{k=1}^{n-1} \beta_k^{-1} = \prod_{k=1}^{n-1} \left(\frac{k+\theta}{k+\alpha+\theta}\right) = \begin{cases} \left(\dfrac{\alpha+\theta}{\theta}\right) \dfrac{(\theta)^{(n)}}{(\alpha+\theta)^{(n)}} & \text{if} \quad \theta \neq 0, \\ \dfrac{\alpha(n-1)!}{(\alpha)^{(n)}} & \text{if} \quad \theta = 0. \end{cases} \tag{17}$$

Denote by (M_n) the sequence of random variables defined, for all $n \geq 1$, by

$$M_n = b_n K_n + \left(\frac{\alpha+\theta}{\alpha}\right) \frac{(\theta)^{(n)}}{(\alpha+\theta)^{(n)}} = b_n \left(K_n + \frac{\theta}{\alpha}\right). \tag{18}$$

Since $b_n = \beta_n b_{n+1}$, (16) immediately implies that for all $n \geq 1$,

$$\mathbb{E}[M_{n+1} \,|\, \mathcal{F}_n] = b_{n+1}\beta_n \left(K_n + \frac{\theta}{\alpha}\right) = b_n \left(K_n + \frac{\theta}{\alpha}\right) = M_n$$

almost surely. Moreover, (M_n) is square integrable as $K_n \leq n$. Consequently, (M_n) is a locally square integrable martingale. Moreover, noting that

$$M_{n+1} - M_n = b_{n+1}\big(\xi_{n+1} - \mathbb{E}[\xi_{n+1} \,|\, \mathcal{F}_n]\big),$$

we have almost surely

$$\mathbb{E}\big[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n\big] = b_{n+1}^2 \big(\mathbb{E}\big[\xi_{n+1}^2 \,|\, \mathcal{F}_n\big] - \mathbb{E}^2[\xi_{n+1} \,|\, \mathcal{F}_n]\big) = b_{n+1}^2 p_n(1-p_n),$$

which leads, via (13), to

$$\mathbb{E}\big[(M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n\big] = b_{n+1}^2 \left(\frac{(\theta+\alpha K_n)(n-\alpha K_n)}{(n+\theta)^2}\right) \qquad \text{a.s.} \tag{19}$$

2.1. An alternative approach for the almost-sure convergence

Based on the above martingale construction, we present an alternative proof of (2), by relying on the strong law of large numbers for martingale. Our proof is more natural and intuitive than that of Pitman [23, Theorem 3.8], which makes use of the exchangeable partition probability function induced by the EP model.

Theorem 2.1. *Let K_n be the number of partition sets in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then,*

$$\lim_{n\to+\infty} \frac{K_n}{n^\alpha} = S_{\alpha,\theta} \qquad a.s. \tag{20}$$

where $S_{\alpha,\theta}$ is a positive and almost surely finite random variable whose distribution has probability density function (7). This convergence holds in $\mathbb{L}^p$ for any integer $p \geq 1$,

$$\lim_{n\to+\infty} \mathbb{E}\left[\left|\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right|^p\right] = 0. \tag{21}$$

Proof. We already saw that (M_n) is a locally square integrable martingale. Denote by $\langle M\rangle_n$ its predictable quadratic variation,

$$\langle M\rangle_n = \sum_{k=0}^{n-1} \mathbb{E}[(M_{k+1} - M_k)^2 \,|\, \mathcal{F}_k].$$

We have from (19) that

$$\langle M\rangle_n = \left(\frac{\alpha+\theta}{\alpha}\right)^2 + \sum_{k=1}^{n-1} b_{k+1}^2 \left(\frac{(\theta+\alpha K_k)(k-\alpha K_k)}{(k+\theta)^2}\right) \qquad \text{a.s.} \tag{22}$$

Taking the expectation on both sides of (22), we obtain that

$$\mathbb{E}\big[\langle M\rangle_n\big] = \left(\frac{\alpha+\theta}{\alpha}\right)^2 + \sum_{k=1}^{n-1} b_{k+1}^2 \left(\frac{\theta k + \alpha(k-\theta)\mathbb{E}[K_k] - \alpha^2\mathbb{E}[K_k^2]}{(k+\theta)^2}\right).$$

Hence, it exists some constant $C > 0$ such that for n large enough,

$$\mathbb{E}\big[\langle M\rangle_n\big] \leq C \sum_{k=1}^{n} \frac{b_k^2 \mathbb{E}[K_k]}{k}. \tag{23}$$

Moreover, we have from (14) and (17) that if $\theta \neq 0$,

$$\mathbb{E}[K_n] = b_n^{-1} \sum_{k=0}^{n-1} \frac{(\theta)^{(k)}}{(\alpha+\theta+1)^{(k)}} = \frac{\theta}{\alpha} \left(\frac{(\alpha+\theta)^{(n)} - (\theta)^{(n)}}{(\theta)^{(n)}} \right), \tag{24}$$

whereas if $\theta = 0$,

$$\mathbb{E}[K_n] = \prod_{k=1}^{n-1} \Big(1 + \frac{\alpha}{k}\Big) = \frac{(\alpha)^{(n)}}{\alpha(n-1)!}. \tag{25}$$

In addition, it follows from standard results on the asymptotic behavior of the Euler Gamma function that for $\theta \neq 0$,

$$\lim_{n\to+\infty} \frac{1}{n^\alpha} \frac{(\alpha+\theta)^{(n)}}{(\theta)^{(n)}} = \lim_{n\to+\infty} \frac{\Gamma(\theta+1)\Gamma(n+\alpha+\theta)}{n^\alpha \theta \Gamma(\alpha+\theta)\Gamma(n+\theta)} = \frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha+\theta)}. \tag{26}$$

Hence, we obtain from (24), (25) and (26) that whatever the value of $\theta > -\alpha$,

$$\lim_{n\to+\infty} \frac{1}{n^\alpha} \mathbb{E}[K_n] = \frac{\Gamma(\theta+1)}{\alpha\Gamma(\alpha+\theta)}.$$

Consequently, we deduce from (23), (17) and (26) that

$$\sup_{n\geq 1} \mathbb{E}\big[\langle M\rangle_n\big] < +\infty. \tag{27}$$

Therefore, we deduce from (27) that (M_n) is bounded in $\mathbb{L}^2$. Then, it follows from Doob's martingale convergence theorem [15, Corollary 2.2], that the sequence (M_n) converges a.s. and in $\mathbb{L}^2$ to a square integrable random variable $M_{\alpha,\theta}$. In fact, we shall prove below that (M_n) is bounded in $\mathbb{L}^p$ for any integer $p \geq 1$ which means that (M_n) converges a.s. and in $\mathbb{L}^p$ to $M_{\alpha,\theta}$ for any integer $p \geq 1$. Hereafter, we obtain from (18) and (26) that

$$\lim_{n\to+\infty} \frac{K_n}{n^\alpha} = S_{\alpha,\theta} \qquad \text{a.s.},$$

where the random variables $M_{\alpha,\theta}$ and $S_{\alpha,\theta}$ are tightly related by means of the identity, which follows from (42) below,

$$M_{\alpha,\theta} = \Big(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\Big) S_{\alpha,\theta},$$

completing the proof of (20). We shall now proceed to the proof of the convergence in $\mathbb{L}^p$ for any integer $p \geq 1$. In particular, we are going to compute the falling factorial moments $\mathbb{E}[(K_n)_{(p)}]$ of K_n where, for any $a \in \mathbb{R}$, $(a)_{(p)} = a(a-1)\cdots(a-p+1)$ with $(a)_{(0)} = 1$. By an application of Vandermonde identity for the falling factorial together with (12), for any integer $p \geq 1$,

$$(K_{n+1})_{(p)} = (K_n + \xi_{n+1})_{(p)} = \sum_{k=0}^{p} \binom{p}{k} (K_n)_{(k)} (\xi_{n+1})_{(p-k)}.$$

By taking the conditional expectation on both sides of this identity, we obtain that

$$\mathbb{E}[(K_{n+1})_{(p)} \,|\, \mathcal{F}_n] = \sum_{k=0}^{p} \binom{p}{k} (K_n)_{(k)} \mathbb{E}[(\xi_{n+1})_{(p-k)} \,|\, \mathcal{F}_n] \qquad \text{a.s.,} \tag{28}$$

where $\mathbb{E}[(\xi_{n+1})_{(1)} \,|\, \mathcal{F}_n] = \mathbb{E}[\xi_{n+1} \,|\, \mathcal{F}_n] = p_n$, and $\mathbb{E}[(\xi_{n+1})_{(k)} \,|\, \mathcal{F}_n] = 0$ for all $k \geq 2$. Accordingly, we can deduce from (13) and (28) that for any integer $p \geq 1$,

$$\begin{aligned}
\mathbb{E}[(K_{n+1})_{(p)} \,|\, \mathcal{F}_n] &= (K_n)_{(p)} + p(K_n)_{(p-1)} \Big(\frac{\alpha K_n + \theta}{n + \theta} \Big) \qquad \text{a.s.} \\
&= (K_n)_{(p)} + \frac{\alpha p}{n+\theta} (K_n)_{(p-1)} (K_n - p + 1) + L_n(p) \qquad \text{a.s.} \\
&= \Big(1 + \frac{\alpha p}{n+\theta} \Big) (K_n)_{(p)} + L_n(p) \qquad \text{a.s.}
\end{aligned} \tag{29}$$

where

$$L_n(p) = \Big(\frac{p(\alpha(p-1) + \theta)}{n+\theta} \Big) (K_n)_{(p-1)}.$$

Denote $b_1(p) = 1$ and for all $n \geq 2$,

$$b_n(p) = \prod_{k=1}^{n-1} \Big(\frac{k + \alpha p + \theta}{k + \theta} \Big) = \begin{cases} \Big(\dfrac{\theta}{\alpha p + \theta} \Big) \dfrac{(\alpha p + \theta)^{(n)}}{(\theta)^{(n)}} & \text{if} \quad \theta \neq 0, \\[2ex] \dfrac{(\alpha p)^{(n)}}{\alpha p (n-1)!} & \text{if} \quad \theta = 0. \end{cases} \tag{30}$$

It follows from (29) and (30) that for all $n \geq 2$ and for any integer $p \geq 2$,

$$\mathbb{E}[(K_n)_{(p)}] = b_n(p) \sum_{k=1}^{n-1} \big(b_{k+1}(p) \big)^{-1} \mathbb{E}[L_k(p)]$$

leading, for $\theta \neq 0$, to

$$\mathbb{E}[(K_n)_{(p)}]=\Big(\frac{p(\alpha(p-1)+\theta)}{\alpha p+\theta}\Big)\Big(\frac{(\alpha p+\theta)^{(n)}}{(\theta)^{(n)}}\Big)\sum_{k=1}^{n-1}\Big(\frac{(\theta)^{(k)}\mathbb{E}[(K_k)_{(p-1)}]}{(\alpha p+\theta+1)^{(k)}}\Big) \quad (31)$$

while, for $\theta = 0$, to

$$\mathbb{E}[(K_n)_{(p)}] = \Big(\frac{(p-1)(\alpha p)^{(n)}}{(n-1)!}\Big)\sum_{k=1}^{n-1}\Big(\frac{(k-1)!\mathbb{E}[(K_k)_{(p-1)}]}{(\alpha p+1)^{(k)}}\Big). \quad (32)$$

We deduce from (24) and (31) that for $\theta \neq 0$,

$$\mathbb{E}[(K_n)_{(1)}] = \frac{\theta}{\alpha(\theta)^{(n)}}\big((\alpha+\theta)^{(n)} - (\theta)^{(n)}\big),$$

$$\mathbb{E}[(K_n)_{(2)}] = \frac{\theta(\theta+\alpha)}{\alpha^2(\theta)^{(n)}}\big((2\alpha+\theta)^{(n)} - 2(\alpha+\theta)^{(n)} + (\theta)^{(n)}\big),$$

and more generally, for all $n \geq 1$ and for any integer $p \geq 1$,

$$\begin{aligned}\mathbb{E}[(K_n)_{(p)}] &= \frac{\prod_{k=0}^{p-1}(k\alpha+\theta)}{\alpha^p(\theta)^{(n)}}\sum_{k=0}^{p}(-1)^{p-k}\binom{p}{k}(k\alpha+\theta)^{(n)}, \\ &= \frac{\big(\frac{\theta}{\alpha}\big)^{(p)}}{(\theta)^{(n)}}\sum_{k=0}^{p}(-1)^{p-k}\binom{p}{k}(k\alpha+\theta)^{(n)}. \end{aligned} \quad (33)$$

In addition, we have from (25) and (32) that for $\theta = 0$,

$$\mathbb{E}[(K_n)_{(1)}] = \frac{(\alpha)^{(n)}}{\alpha(n-1)!},$$

$$\mathbb{E}[(K_n)_{(2)}] = \frac{1}{\alpha(n-1)!}\big((2\alpha)^{(n)} - 2(\alpha)^{(n)}\big),$$

and more generally, for all $n \geq 1$ and for any integer $p \geq 1$,

$$\mathbb{E}[(K_n)_{(p)}] = \frac{(p-1)!}{\alpha(n-1)!}\sum_{k=1}^{p}(-1)^{p-k}\binom{p}{k}(k\alpha)^{(n)}. \quad (34)$$

We obtain from (30) together with (33) and (34) that it exists a positive constant $C(p,\alpha,\theta)$ such that for all $n \geq 1$ and $p \geq 1$,

$$\mathbb{E}[(K_n)_{(p)}] \leq C(p,\alpha,\theta)b_n(p). \quad (35)$$

Moreover, we also have the elementary inequality

$$\mathbb{E}[K_n^p] \leq p^p + p!\mathbb{E}[(K_n)_{(p)} \mathrm{I}_{K_n \geq p}]. \tag{36}$$

Consequently, we find from (18), (35) and (36) that it exists a positive constant $D(p, \alpha, \theta)$ such that for all $n \geq 1$ and $p \geq 1$,

$$\mathbb{E}[M_n^p] \leq D(p, \alpha, \theta) b_n(p) b_n^p. \tag{37}$$

Furthermore, it is easy to see from (17) and (30) that

$$b_n(p) b_n^p = \prod_{k=1}^{n-1} \left(1 + \frac{\alpha p}{k + \theta}\right) \left(1 + \frac{\alpha}{k + \theta}\right)^{-p} \leq 1. \tag{38}$$

Hence, we deduce from (37) and (38) that

$$\sup_{n \geq 1} \mathbb{E}[M_n^p] < +\infty,$$

which means that the martingale (M_n) is bounded in $\mathbb{L}^p$ for any $p \geq 1$. Therefore, it follows from Doob's martingale convergence theorem [15, Corollary 2.2], the sequence (M_n) converges almost surely and in $\mathbb{L}^p$ to a finite random variable $M_{\alpha,\theta}$. Our goal is now to compute all the moments of $M_{\alpha,\theta}$. We shall only carry out the proof for $\theta \neq 0$ inasmuch as the proof for $\theta = 0$ follows exactly the same arguments. First of all, as in (26),

$$\lim_{n \to +\infty} \frac{(\alpha p + \theta)^{(n)}}{n^{\alpha p} (\theta)^{(n)}} = \frac{\Gamma(\theta + 1)}{\theta \Gamma(\alpha p + \theta)}.$$

Hence, we obtain once again from (33) that for any integer $p \geq 1$,

$$\lim_{n \to +\infty} \frac{1}{n^{\alpha p}} \mathbb{E}[(K_n)_{(p)}] = \frac{\Gamma(\theta + 1)}{\theta \Gamma(\alpha p + \theta)} \left(\frac{\theta}{\alpha}\right)^{(p)}. \tag{39}$$

However, we have for all $p \geq 1$,

$$\mathbb{E}[K_n^p] = \sum_{k=0}^{p} \left\{ \begin{matrix} p \\ k \end{matrix} \right\} \mathbb{E}[(K_n)_{(k)}], \tag{40}$$

where the curly brackets are the Stirling numbers of the second kind given by

$$\left\{ \begin{matrix} p \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} i^p.$$

Since, $\left\{ {p \atop p} \right\} = 1$, we obtain from (39) and (40) that for all $p \geq 1$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha p}} \mathbb{E}[K_n^p] = \frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)} \Big(\frac{\theta}{\alpha}\Big)^{(p)}. \tag{41}$$

We also recall from (17) that

$$\lim_{n\to+\infty} n^\alpha b_n = \lim_{n\to+\infty} \frac{n^\alpha(\theta+1)_{(n)}}{(\alpha+\theta+1)_{(n)}} = \frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}. \tag{42}$$

Consequently, it follows from the conjunction of (18), (41) and (42) that

$$\lim_{n\to+\infty} \mathbb{E}[M_n^p] = \lim_{n\to+\infty} b_n^p \mathbb{E}\Big[\Big(K_n + \frac{\theta}{\alpha}\Big)^p\Big] = \lim_{n\to+\infty} b_n^p \sum_{k=0}^{p} \binom{p}{k} \mathbb{E}[K_n^k] \left(\frac{\theta}{\alpha}\right)^{p-k},$$

which drastically reduces to

$$\lim_{n\to+\infty} \mathbb{E}[M_n^p] = \lim_{n\to+\infty} b_n^p \mathbb{E}[K_n^p] = \left(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\right)^p \frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)} \left(\frac{\theta}{\alpha}\right)^{(p)}. \tag{43}$$

It immediately leads, for all $p \geq 1$, to

$$\mathbb{E}[M_{\alpha,\theta}^p] = \left(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\right)^p \frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)} \left(\frac{\theta}{\alpha}\right)^{(p)},$$

as well as

$$\mathbb{E}[S_{\alpha,\theta}^p] = \frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)} \left(\frac{\theta}{\alpha}\right)^{(p)}.$$

The distribution of the random variable $S_{\alpha,\theta}$ has probability density function given by (7). Finally, (21) follows from (20) and (41) together with Riesz-Scheffé theorem, which completes the proof of Theorem 2.1. □

Remark 2.2. *An alternative proof of the convergence in* $\mathbb{L}^2$ *given by* (21)*, which follows directly from the Poisson process construction of the EP model [21, Proposition 9] can be found in Appendix B. See also Pitman [23, Chapter 4].*

2.2. *The Gaussian fluctuation*

Our martingale approach for K_n suggests that, in addition to the almost-sure convergence, it is possible to establish a Gaussian fluctuations of K_n, properly normalized, around its almost-sure limit $S_{\alpha,\theta}$.

Theorem 2.3. *Let K_n be the number of partition sets in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then,*

$$\frac{K_n - n^\alpha S_{\alpha,\theta}}{\sqrt{K_n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1), \tag{44}$$

where $\mathcal{N}(0,1)$ denotes a standard Gaussian random variable. Moreover, we also have

$$\sqrt{n^\alpha}\left(\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \sqrt{S'_{\alpha,\theta}}\,\mathcal{N}(0,1), \tag{45}$$

where $S'_{\alpha,\theta}$ is a random variable independent of $\mathcal{N}(0,1)$ and sharing the same distribution as $S_{\alpha,\theta}$.

Proof. The proof relies on a beautiful result due to [16] concerning with Gaussian fluctuations for martingales. For all $n \geq 1$, denote $\Delta M_{n+1} = M_{n+1} - M_n$ and

$$\Lambda_n = \sum_{k=n}^{\infty} \mathbb{E}\big[\Delta M_{k+1}^2 | \mathcal{F}_k\big].$$

We immediately have from (19) that

$$\Lambda_n = \sum_{k=n}^{\infty} b_{k+1}^2 \left(\frac{(\theta + \alpha K_k)(k - \alpha K_k)}{(k+\theta)^2}\right). \tag{46}$$

On the one hand, we know from (20) that

$$\lim_{n\to+\infty} \frac{K_n}{n^\alpha} = S_{\alpha,\theta} \qquad \text{a.s.} \tag{47}$$

On the other hand, we already saw from (42) that

$$\lim_{n\to+\infty} n^\alpha b_n = \frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}. \tag{48}$$

Consequently, we deduce from (46), (47) and (48) that

$$\lim_{n\to+\infty} n^\alpha \Lambda_n = \left(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\right)^2 S_{\alpha,\theta} \qquad \text{a.s.} \tag{49}$$

By the same token, if

$$s_n^2 = \mathbb{E}[\Lambda_n] = \sum_{k=n}^{\infty} \mathbb{E}\big[\Delta M_{k+1}^2\big],$$

we also have

$$s_n^2 = \sum_{k=n}^{\infty} b_{k+1}^2 \left(\frac{\theta k + \alpha(k-\theta)\mathbb{E}[K_k] - \alpha^2 \mathbb{E}[K_k^2]}{(k+\theta)^2} \right).$$

However, we already saw from (41) that

$$\lim_{n \to +\infty} \frac{1}{n^\alpha} \mathbb{E}[K_n] = \frac{\Gamma(\theta+1)}{\alpha \Gamma(\alpha+\theta)} \tag{50}$$

and

$$\lim_{n \to +\infty} \frac{1}{n^{2\alpha}} \mathbb{E}[K_n^2] = \frac{(\alpha+\theta)\Gamma(\theta+1)}{\alpha^2 \Gamma(2\alpha+\theta)}. \tag{51}$$

It ensures that

$$\lim_{n \to +\infty} n^\alpha s_n^2 = \Big(\frac{\alpha+\theta}{\alpha}\Big) \frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}.$$

Hereafter, we claim that for any $\eta > 0$,

$$\lim_{n \to +\infty} n^\alpha \sum_{k=n}^{\infty} \mathbb{E}\big[\Delta M_{k+1}^2 \mathrm{I}_{\{|\Delta M_{k+1}| > \eta \sqrt{n^{-\alpha}}\}}\big] = 0. \tag{52}$$

As a matter of fact, we clearly have for any $\eta > 0$,

$$n^\alpha \sum_{k=n}^{\infty} \mathbb{E}\big[\Delta M_{k+1}^2 \mathrm{I}_{\{|\Delta M_{k+1}| > \eta \sqrt{n^{-\alpha}}\}}\big] \leq \frac{n^{2\alpha}}{\eta^2} \sum_{k=n}^{\infty} \mathbb{E}\big[\Delta M_{k+1}^4\big].$$

Moreover, it follows from tedious but straightforward calculations that for all $n \geq 1$,

$$\mathbb{E}\big[\Delta M_{n+1}^4 | \mathcal{F}_n\big] = b_{n+1}^4 \sum_{p=0}^{4} e_n(p) K_n^p \tag{53}$$

where

$$e_n(0) = \frac{n\theta}{(n+\theta)^4}(n^2 - n\theta + \theta^2),$$

$$\begin{aligned}
e_n(1) &= \frac{\alpha(n-\theta)}{(n+\theta)^4}(n^2 - 4n\theta + \theta^2),\\
e_n(2) &= \frac{-2\alpha^2(2n-\theta)(n-2\theta)}{(n+\theta)^4},\\
e_n(3) &= \frac{6\alpha^3(n-\theta)}{(n+\theta)^4},\\
e_n(4) &= -\frac{3\alpha^4}{n^4}.
\end{aligned}$$

Taking the expectation on both sides of (53), we obtain that for all $n \geq 1$,

$$\mathbb{E}\big[\Delta M_{n+1}^4\big] = b_{n+1}^4 \sum_{p=0}^{4} e_n(p)\mathbb{E}[K_n^p] \tag{54}$$

Furthermore, we already saw from (41) that

$$\lim_{n\to+\infty} \frac{1}{n^{3\alpha}}\mathbb{E}[K_n^3] = \frac{(\alpha+\theta)(2\alpha+\theta)\Gamma(\theta+1)}{\alpha^3\Gamma(3\alpha+\theta)} \tag{55}$$

and

$$\lim_{n\to+\infty} \frac{1}{n^{4\alpha}}\mathbb{E}[K_n^4] = \frac{(\alpha+\theta)(2\alpha+\theta)(3\alpha+\theta)\Gamma(\theta+1)}{\alpha^4\Gamma(4\alpha+\theta)}. \tag{56}$$

Hence, we deduce from (48) and (54) together with (50), (51), (55) and (56) that

$$\lim_{n\to+\infty} n^{1+3\alpha}\mathbb{E}\big[\Delta M_{n+1}^4\big] = (\alpha+\theta)\Big(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\Big)^3. \tag{57}$$

Consequently, we obtain from (57) that

$$\lim_{n\to+\infty} n^{3\alpha}\sum_{k=n}^{\infty}\mathbb{E}\big[\Delta M_{k+1}^4\big] = (\alpha+\theta)\Big(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\Big)^3$$

which clearly leads to (52). Finally, let (P_n) be the martingale defined by

$$P_{n+1} = \sum_{k=1}^{n} k^{\alpha}\big(\Delta M_{k+1}^2 - \mathbb{E}[\Delta M_{k+1}^2 \,|\, \mathcal{F}_k]\big). \tag{58}$$

Its predictable quadratic variation is given by

$$\langle P\rangle_{n+1} = \sum_{k=1}^{n} k^{2\alpha}\big(\mathbb{E}[\Delta M_{k+1}^4 \,|\, \mathcal{F}_k] - (\mathbb{E}[\Delta M_{k+1}^2 \,|\, \mathcal{F}_k])^2\big).$$

Therefore, we find from (53) that

$$\langle P\rangle_{n+1} \leq \sum_{p=0}^{4}\sum_{k=1}^{n} k^{2\alpha} b_{k+1}^4 e_k(p) K_k^p.$$

One can observe that we always have $e_n(4) < 0$. For n large enough, $e_n(2) < 0$. Furthermore, it exists some constant $C > 0$ such that for n large enough, $e_n(0) \leq C/n$, $e_n(1) \leq C/n$ and $e_n(3) \leq C/n^3$. Consequently, we get that for n large enough,

$$\langle P\rangle_n \leq C\left(\sum_{k=1}^{n}\frac{k^{2\alpha}b_k^4}{k} + \sum_{k=1}^{n}\frac{k^{2\alpha}b_k^4 K_k}{k} + \sum_{k=1}^{n}\frac{k^{2\alpha}b_k^4 K_k^3}{k^3}\right). \tag{59}$$

Hence, we obtain from (47), (48) and (59) that $\langle P\rangle_n$ converges a.s. to a finite random variable. Then, we deduce from the strong law of large numbers for martingales given by the first part of [8, Theorem 1.3.15] that (P_n) converges a.s. to a finite random variable. All the conditions of the second part of Theorem 1 and Corollaries 1 and 2 in [16] are satisfied, which leads to

$$\frac{M_n - M_{\alpha,\theta}}{\sqrt{\Lambda_n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1) \tag{60}$$

as well as to

$$\sqrt{n^{\alpha}}(M_n - M_{\alpha,\theta}) \xrightarrow[n\to+\infty]{\mathcal{L}} \frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\sqrt{S'_{\alpha,\theta}}\mathcal{N}(0,1), \tag{61}$$

where $M_{\alpha,\theta}$ is the almost-sure limit of the martingale (M_n) and $S'_{\alpha,\theta}$ is independent of the Gaussian $\mathcal{N}(0,1)$ random variable and $S'_{\alpha,\theta}$ shares the same distribution as $S_{\alpha,\theta}$. However, the random variables $M_{\alpha,\theta}$ and $S_{\alpha,\theta}$ are tightly related through

$$M_{\alpha,\theta} = \Big(\frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)}\Big) S_{\alpha,\theta}.$$

In addition, we obtain from the asymptotic behavior of the ratio of two Gamma functions that

$$b_n = \frac{\Gamma(\alpha+\theta+1)}{\Gamma(\theta+1)n^{\alpha}}\left(1 - \frac{\alpha(2\theta+\alpha-1)}{2n} + O\Big(\frac{1}{n^2}\Big)\right). \tag{62}$$

Finally, the Gaussian fluctuations (44) and (45) follow from (60) and (61) together with the almost-sure convergences (20), (49), (62) and Slutsky's lemma, which achieves the proof of Theorem 2.3. □

2.3. The law of the iterated logarithm

We conclude our asymptotic analysis of K_n by establishing the law of iterated logarithm.

Theorem 2.4. *Let K_n be the number of partition sets in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then,*

$$\limsup_{n\to\infty} \left(\frac{K_n - n^\alpha S_{\alpha,\theta}}{\sqrt{2K_n \log\log n}} \right) = -\liminf_{n\to\infty} \left(\frac{K_n - n^\alpha S_{\alpha,\theta}}{\sqrt{2K_n \log\log n}} \right) = 1 \qquad a.s. \quad (63)$$

Moreover, we also have

$$\limsup_{n\to\infty} \left(\frac{K_n - n^\alpha S_{\alpha,\theta}}{\sqrt{2n^\alpha \log\log n}} \right) = -\liminf_{n\to\infty} \left(\frac{K_n - n^\alpha S_{\alpha,\theta}}{\sqrt{2n^\alpha \log\log n}} \right) = \sqrt{S_{\alpha,\theta}} \qquad a.s.$$

In particular,

$$\limsup_{n\to\infty} \frac{\left(K_n - n^\alpha S_{\alpha,\theta}\right)^2}{2n^\alpha \log\log n} = S_{\alpha,\theta} \qquad a.s.$$

Proof. The proof is a direct application of [16]. Using the same notation as in the proof of Theorem 2.3, we have for any $\eta > 0$,

$$\sum_{n=1}^{\infty} n^{\alpha/2} \mathbb{E}\big[|\Delta M_n| \mathrm{I}_{\{|\Delta M_n| > \eta\sqrt{n^{-\alpha}}\}}\big] \leq \frac{1}{\eta^3} \sum_{n=1}^{\infty} n^{2\alpha} \mathbb{E}\big[\Delta M_n^4\big]. \qquad (64)$$

Hence, we deduce from (57) and (64) that it exists some constant $C > 0$ such that

$$\sum_{n=1}^{\infty} n^{\alpha/2} \mathbb{E}\big[|\Delta M_n| \mathrm{I}_{\{|\Delta M_n| > \eta\sqrt{n^{-\alpha}}\}}\big] \leq C \sum_{n=1}^{\infty} \frac{1}{n^{1+\alpha}} < \infty.$$

Moreover, we also have for any $\delta > 0$,

$$\sum_{n=1}^{\infty} n^{2\alpha} \mathbb{E}\big[\Delta M_n^4 \mathrm{I}_{\{|\Delta M_n| \leq \delta\sqrt{n^{-\alpha}}\}}\big] \leq \sum_{n=1}^{\infty} n^{2\alpha} \mathbb{E}\big[\Delta M_n^4\big] < \infty.$$

Furthermore, we already saw that the martingale (P_n), defined by (58), converges a.s. to a finite random variable. Consequently, all the conditions of the second part of Theorem 1 and Corollary 2 in [16] are satisfied, which ensures that

$$\limsup_{n\to\infty} \left(\frac{M_n - M_{\alpha,\theta}}{\sqrt{2\Lambda_n \log\log n}} \right) = -\liminf_{n\to\infty} \left(\frac{M_n - M_{\alpha,\theta}}{\sqrt{2\Lambda_n \log\log n}} \right) = 1 \qquad \text{a.s.} \quad (65)$$

where $M_{\alpha,\theta}$ is the almost-sure limit of the martingale (M_n). Finally, the law of iterated logarithm (63) follows from (65) together with the almost-sure convergences (20) and (49), which completes the proof of Theorem 2.4. $\square$

3. Asymptotic results for the partition subsets of size r

We start by introducing the keystone martingale construction for $K_{r,n}$. In particular, our approach is different from that developed in Section 2, since we are going to build a martingale that will not converge almost surely to a finite random variable. First, we consider the case $r = 1$. From the sequential construction of the EP model [20, Proposition 9], we have for all $n \geq 1$,

$$K_{1,n+1} = K_{1,n} + \xi_{1,n+1} \tag{66}$$

where the conditional distribution of the random variable $\xi_{1,n+1}$, given the σ-algebra $\mathcal{F}_n = \sigma(K_1, \ldots, K_n, K_{1,1}, \ldots, K_{1,n})$, is such that

$$\mathbb{P}(\xi_{1,n+1} = k \mid \mathcal{F}_n) = \begin{cases} p_{1,n} & \text{if} \quad k = 1, \\ q_{1,n} & \text{if} \quad k = -1, \\ 1 - p_{1,n} - q_{1,n} & \text{if} \quad k = 0, \end{cases} \tag{67}$$

where

$$p_{1,n} = \frac{\alpha K_n + \theta}{n + \theta} \qquad \text{and} \qquad q_{1,n} = \frac{(1-\alpha)K_{1,n}}{n + \theta}. \tag{68}$$

We refer the reader to Appendix A for more details on (67). From definition of the sequence $(\xi_{1,n})$, we clearly have $\mathbb{E}[\xi_{1,n+1} \mid \mathcal{F}_n] = p_{1,n} - q_{1,n}$. Consequently, we obtain from (66) that

$$\mathbb{E}[K_{1,n+1} \mid \mathcal{F}_n] = \mathbb{E}[K_{1,n} + \xi_{1,n+1} \mid \mathcal{F}_n] = K_{1,n} + p_{1,n} - q_{1,n} \qquad \text{a.s,}$$

which leads to

$$\mathbb{E}[K_{1,n+1} \mid \mathcal{F}_n] = \beta_{1,n} K_{1,n} + p_{1,n} \qquad \text{a.s.,} \tag{69}$$

where

$$\beta_{1,n} = 1 - \frac{1-\alpha}{n+\theta} = \frac{n - 1 + \theta + \alpha}{n + \theta}. \tag{70}$$

Let $(b_{1,n})$ be the sequence defined by $b_{1,1} = 1$ and for all $n \geq 2$,

$$b_{1,n} = \prod_{k=1}^{n-1} \beta_{1,k}^{-1} = \prod_{k=1}^{n-1} \left(\frac{k+\theta}{k-1+\theta+\alpha} \right) = \frac{(\theta+1)^{(n-1)}}{(\alpha+\theta)^{(n-1)}}. \tag{71}$$

Now, we introduce the sequence of random variables $(M_{1,n})$ that is defined, for all $n \geq 1$, by

$$M_{1,n} = b_{1,n} K_{1,n} - A_{1,n}, \tag{72}$$

where

$$A_{1,n} = \sum_{k=1}^{n-1} b_{1,k+1} \left(\frac{\alpha K_k + \theta}{k+\theta} \right). \tag{73}$$

One can observe that $A_{1,n+1} = A_{1,n} + p_{1,n} b_{1,n+1}$ and $b_{1,n} = \beta_{1,n} b_{1,n+1}$. Hence, we have from (69) and (70) that for all $n \geq 1$,

$$\begin{aligned}
\mathbb{E}[M_{1,n+1} \,|\, \mathcal{F}_n] &= b_{1,n+1} \mathbb{E}[K_{1,n+1} \,|\, \mathcal{F}_n] - A_{1,n+1} \\
&= b_{1,n+1}(\beta_{1,n} K_{1,n} + p_{1,n}) - A_{1,n} - p_{1,n} b_{1,n+1} \\
&= \beta_{1,n} b_{1,n+1} K_{1,n} - A_{1,n} \\
&= b_{1,n} K_{1,n} - A_{1,n} \\
&= M_{1,n}
\end{aligned}$$

almost surely. Moreover, $(M_{1,n})$ is square integrable as $K_{1,n} \leq K_n \leq n$. Consequently, $(M_{1,n})$ is a locally square integrable martingale. Because of (67), it is not hard to compute all the higher order conditional moments of $(M_{1,n})$. For example, as

$$\Delta M_{1,n+1} = M_{1,n+1} - M_{1,n} = b_{1,n+1}(K_{1,n+1} - \mathbb{E}[K_{1,n+1} \,|\, \mathcal{F}_n])$$

we clearly have from (66) and (69) that

$$\mathbb{E}[(\Delta M_{1,n+1})^2 \,|\, \mathcal{F}_n] = b_{1,n+1}^2 (\mathbb{E}[K_{1,n+1}^2 \,|\, \mathcal{F}_n] - (\mathbb{E}[K_{1,n+1} \,|\, \mathcal{F}_n])^2)$$

with

$$\mathbb{E}[K_{1,n+1}^2 \,|\, \mathcal{F}_n] = K_{1,n}^2 + 2K_{1,n}(p_{1,n} - q_{1,n}) + p_{1,n} + q_{1,n}$$

and

$$(\mathbb{E}[K_{1,n+1} \,|\, \mathcal{F}_n])^2 = (K_{1,n} + p_{1,n} - q_{1,n})^2$$

almost surely, which reduces to

$$\mathbb{E}[(\Delta M_{1,n+1})^2 \,|\, \mathcal{F}_n] = b_{1,n+1}^2 \big(p_{1,n} + q_{1,n} - (p_{1,n} - q_{1,n})^2 \big) \qquad \text{a.s.} \tag{74}$$

The above calculations extend easily to the case $r \geq 2$. From the sequential construction of the EP model [20, Proposition 9], we have for all $n \geq 1$,

$$K_{r,n+1} = K_{r,n} + \xi_{r,n+1} \tag{75}$$

where the conditional distribution of the random variable $\xi_{r,n+1}$, given the σ-algebra $\mathcal{F}_n = \sigma(K_{r-1,1}, \ldots, K_{r-1,n}, K_{r,1}, \ldots, K_{r,n})$, is such that

$$\mathbb{P}(\xi_{r,n+1} = k \,|\, \mathcal{F}_n) = \begin{cases} p_{r,n} & \text{if} \quad k = 1, \\ q_{r,n} & \text{if} \quad k = -1, \\ 1 - p_{r,n} - q_{r,n} & \text{if} \quad k = 0, \end{cases} \tag{76}$$

where

$$p_{r,n} = \frac{(r-1-\alpha)K_{r-1,n}}{n+\theta} \qquad \text{and} \qquad q_{r,n} = \frac{(r-\alpha)K_{r,n}}{n+\theta}. \tag{77}$$

We refer the reader to Appendix A for more details on (76). As before, $\mathbb{E}[\xi_{r,n+1} \,|\, \mathcal{F}_n] = p_{r,n} - q_{r,n}$, which implies that

$$\mathbb{E}[K_{r,n+1} \,|\, \mathcal{F}_n] = \beta_{r,n} K_{r,n} + p_{r,n} \qquad \text{a.s.}, \tag{78}$$

where

$$\beta_{r,n} = 1 - \frac{r-\alpha}{n+\theta} = \frac{n - r + \theta + \alpha}{n+\theta}. \tag{79}$$

Let $(b_{r,n})$ be the sequence defined by $b_{r,1} = 1$ and for all $n \geq 2$,

$$b_{r,n} = \prod_{k=r}^{n-1} \beta_{r,k}^{-1} = \prod_{k=r}^{n-1} \left(\frac{k+\theta}{k - r + \theta + \alpha} \right) = \frac{(\theta+1)^{(n-1)}}{(\alpha+\theta)^{(n-r)}}. \tag{80}$$

Hereafter, we introduce the sequence of random variables $(M_{r,n})$ that is defined, for all $n \geq 1$, by

$$M_{r,n} = b_{r,n} K_{r,n} - A_{r,n}, \tag{81}$$

where

$$A_{r,n} = \sum_{k=1}^{n-1} b_{r,k+1} \left(\frac{r-1-\alpha}{k+\theta} \right) K_{r-1,k}. \tag{82}$$

Since $A_{r,n+1} = A_{r,n} + p_{r,n}b_{r,n+1}$ and $b_{r,n} = \beta_{r,n}b_{r,n+1}$, we have from (78) and (79) that for all $n \geq 1$,

$$\begin{aligned}
\mathbb{E}[M_{r,n+1} \,|\, \mathcal{F}_n] &= b_{r,n+1}\mathbb{E}[K_{r,n+1} \,|\, \mathcal{F}_n] - A_{r,n+1} \\
&= b_{r,n+1}(\beta_{r,n}K_{r,n} + p_{r,n}) - A_{r,n} - p_{r,n}b_{r,n+1} \\
&= \beta_{r,n}b_{r,n+1}K_{r,n} - A_{r,n} \\
&= b_{r,n}K_{r,n} - A_{r,n} \\
&= M_{r,n}
\end{aligned}$$

almost surely. In addition, $(M_{r,n})$ is square integrable as $K_{r,n} \leq K_n \leq n$. Consequently, $(M_{r,n})$ is a locally square integrable martingale. Because of (76), it is not hard to compute all the higher order conditional moments of $(M_{r,n})$, as it was previously done for $r = 1$.

3.1. An alternative approach for the almost-sure convergence

Based on the above martingale construction, we present an alternative proof of (8), relying on the strong law of large numbers for martingale. Our proof is more rigorous than Pitman [23, Lemma 3.11], which contains only a sketch for the proof of (8).

Theorem 3.1. *Let $K_{r,n}$ be the number of partition subsets of size r in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then, for all $r \geq 1$*

$$\lim_{n\to+\infty} \frac{K_{r,n}}{n^{\alpha}} = p_{\alpha}(r)S_{\alpha,\theta} \qquad a.s. \tag{83}$$

where

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

and $S_{\alpha,\theta}$ is a positive and almost surely finite random variable whose distribution has probability density function (7). This convergence holds in $\mathbb{L}^p$ for any integer $p \geq 1$,

$$\lim_{n\to+\infty} \mathbb{E}\left[\left|\frac{K_{r,n}}{n^{\alpha}} - p_{\alpha}(r)S_{\alpha,\theta}\right|^p\right] = 0. \tag{84}$$

Proof. The proof is by induction on $r \geq 1$. In particular, we start by considering the case $r = 1$. We already saw that $(M_{1,n})$ is a locally square integrable martingale. We deduce from (74) that its predictable quadratic variation is given by

$$\langle M_1 \rangle_n = \sum_{k=1}^{n-1} b_{1,k+1}^2 \big(p_{1,k} + q_{1,k} - (p_{1,k} - q_{1,k})^2\big).$$

Consequently,

$$\langle M_1 \rangle_n \leq \sum_{k=1}^{n-1} b_{1,k+1}^2 \big(p_{1,k} + q_{1,k}\big).$$

On the one hand, it follows directly from the definition of $p_{1,n}$ in (68) together with the almost-sure convergence (20) that

$$\lim_{n \to +\infty} n^{1-\alpha} p_{1,n} = \alpha S_{\alpha,\theta} \qquad \text{a.s.} \tag{85}$$

On the other hand, we have from (71) that

$$\lim_{n \to +\infty} \frac{b_{1,n}}{n^{1-\alpha}} = \frac{\Gamma(\alpha + \theta)}{\Gamma(\theta + 1)}. \tag{86}$$

Accordingly, as $(n + \theta) q_{1,n} \leq K_n$, we deduce from (85) and (86) through a direct application of Toeplitz lemma that

$$\langle M_1 \rangle_n = O\left(\sum_{k=1}^{n} k^{1-\alpha} \right) = O(n^{2-\alpha}) \qquad \text{a.s.}$$

Therefore, it follows from the strong law of large numbers for martingales given by [8, Theorem 1.3.24] that

$$(M_{1,n})^2 = O(n^{2-\alpha} \log n) \qquad \text{a.s.},$$

which leads to

$$(b_{1,n} K_{1,n} - A_{1,n})^2 = O(n^{2-\alpha} \log n) \qquad \text{a.s.}$$

It clearly implies that

$$\lim_{n \to +\infty} \frac{b_{1,n} K_{1,n} - A_{1,n}}{n} = 0 \qquad \text{a.s.} \tag{87}$$

Furthermore, by combining (85) and (86) together with (73), we find that

$$\lim_{n \to +\infty} \frac{A_{1,n}}{n} = \frac{\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)} S_{\alpha,\theta} \qquad \text{a.s.} \tag{88}$$

Consequently, we obtain from (87) and (88) that

$$\lim_{n \to +\infty} \frac{b_{1,n} K_{1,n}}{n} = \lim_{n \to +\infty} \frac{A_{1,n}}{n} = \frac{\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)} S_{\alpha,\theta} \qquad \text{a.s.}$$

Then, we deduce from (86) that

$$\lim_{n \to +\infty} \frac{K_{1,n}}{n^{\alpha}} = \alpha S_{\alpha,\theta} \qquad \text{a.s.}$$

This completes the proof for $r = 1$. For $r \geq 2$, we shall proceed by induction, assuming that

$$\lim_{n \to +\infty} \frac{K_{r-1,n}}{n^{\alpha}} = \frac{\alpha(1-\alpha)^{(r-2)}}{(r-1)!} S_{\alpha,\theta} \qquad \text{a.s.} \tag{89}$$

We already saw that $(M_{r,n})$ is a locally square integrable martingale. As in the case $r = 1$, one can easily see that its predictable quadratic variation is given by

$$\langle M_r \rangle_n = \sum_{k=1}^{n-1} b_{r,k+1}^2 \big(p_{r,k} + q_{r,k} - (p_{r,k} - q_{r,k})^2 \big). \tag{90}$$

On the one hand, it follows from the definition of $p_{r,n}$ in (77) together with the induction hypothesis (89) that

$$\begin{aligned} \lim_{n \to +\infty} n^{1-\alpha} p_{r,n} &= \frac{(r-1-\alpha)\alpha(1-\alpha)^{(r-2)}}{(r-1)!} S_{\alpha,\theta} \qquad \text{a.s.} \\ &= \frac{\alpha(1-\alpha)^{(r-1)}}{(r-1)!} S_{\alpha,\theta} \qquad \text{a.s.} \end{aligned} \tag{91}$$

On the other hand, we have from (80) that

$$\lim_{n \to +\infty} \frac{b_{r,n}}{n^{r-\alpha}} = \frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}. \tag{92}$$

Accordingly, as $(n+\theta)q_{r,n} \leq rK_n$, we deduce from (90), (91) and (92) together with Toeplitz lemma that

$$\langle M_r \rangle_n = O\left(\sum_{k=1}^{n} k^{2r-1-\alpha}\right) = O(n^{2r-\alpha}) \qquad \text{a.s.}$$

Therefore, it follows from the strong law of large numbers for martingales that

$$(M_{r,n})^2 = O(n^{2r-\alpha} \log n) \qquad \text{a.s.},$$

which leads to

$$(b_{r,n}K_{r,n} - A_{r,n})^2 = O(n^{2r-\alpha} \log n) \qquad \text{a.s.}$$

As before, it implies that

$$\lim_{n \to +\infty} \frac{b_{r,n}K_{r,n} - A_{r,n}}{n^r} = 0 \qquad \text{a.s.} \tag{93}$$

Hereafter, by combining (91) and (92) together with (82), we obtain that

$$\lim_{n \to +\infty} \frac{A_{r,n}}{n^r} = \frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)} \frac{\alpha(1-\alpha)^{(r-1)}}{r!} S_{\alpha,\theta} \qquad \text{a.s.} \tag{94}$$

Finally, we find from (93) and (94) that

$$\lim_{n \to +\infty} \frac{b_{r,n}K_{r,n}}{n^r} = \lim_{n \to +\infty} \frac{A_{r,n}}{n^r} = \frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)} \frac{\alpha(1-\alpha)^{(r-1)}}{r!} S_{\alpha,\theta} \qquad \text{a.s.}$$

which ensures via (86) that

$$\lim_{n \to +\infty} \frac{K_{r,n}}{n^\alpha} = \frac{\alpha(1-\alpha)^{(r-1)}}{r!} S_{\alpha,\theta} \qquad \text{a.s.}$$

It only remains to prove the convergence in $\mathbb{L}^p$ given by (84). We have from Favaro et al. [12, Proposition 1] that for any integer $p \geq 1$,

$$\begin{aligned} \mathbb{E}[(K_{r,n})_{(p)}] &= (p_r(\alpha))^p \frac{n!}{(n-rp)!} \sum_{k=0}^{n-pr} \frac{\left(\frac{\theta}{\alpha}\right)^{(k+p)}}{(\theta)^{(n)}} \mathscr{C}(n-pr, k; \alpha) \qquad (95) \\ &= (p_r(\alpha))^p \frac{n!}{(n-rp)!} \left(\frac{\theta}{\alpha}\right)^{(p)} \sum_{k=0}^{n-pr} \frac{\left(\frac{\theta}{\alpha}+p\right)^{(k)}}{(\theta)^{(n)}} \mathscr{C}(n-pr, k; \alpha) \end{aligned}$$

where $\mathscr{C}(n,k;\alpha)$ is the generalized factorial coefficient [5, Chapter 2],

$$\mathscr{C}(n,k;\alpha)=\frac{1}{k!}\sum_{j=1}^{k}(-1)^j\binom{k}{j}(-j\alpha)^{(n)}.$$

An application of [5, Equation 2.46] allows to solve the summation over k in (95). More precisely,

$$\mathbb{E}[(K_{r,n})_{(p)}]=(p_r(\alpha))^p\frac{n!}{(n-rp)!}\left(\frac{\theta}{\alpha}\right)^{(p)}\frac{(\theta+\alpha p)^{(n-rp)}}{(\theta)^{(n)}}.$$

Consequently,

$$\lim_{n\to+\infty}\frac{1}{n^{\alpha p}}\mathbb{E}[(K_{r,n})_{(p)}]=\big(p_r(\alpha)\big)^p\left(\frac{\theta}{\alpha}\right)^{(p)}\frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)}\left(\frac{\theta}{\alpha}\right)^{(p)} \tag{96}$$

However, as in identity (40), we have

$$\mathbb{E}[K_{r,n}^p]=\sum_{k=0}^{p}\begin{Bmatrix}p\\k\end{Bmatrix}\mathbb{E}[(K_{r,n})_{(k)}]. \tag{97}$$

Then, we deduce from (96) and (97) that

$$\lim_{n\to+\infty}\frac{1}{n^{\alpha p}}\mathbb{E}[K_{r,n}^p]=\big(p_r(\alpha)\big)^p\frac{\Gamma(\theta+1)}{\theta\Gamma(\alpha p+\theta)}\left(\frac{\theta}{\alpha}\right)^{(p)}=\mathbb{E}\big[\big(p_r(\alpha)S_{\alpha,\theta}\big)^p\big]. \tag{98}$$

Finally, (84) follows from (83) and (98) together with Riesz-Scheffé theorem, which completes the proof of Theorem 3.1.

□

Remark 3.2. *An alternative proof of the convergence in* $\mathbb{L}^2$ *given by* (3.1)*, which follows directly from the Poisson process construction of the EP model [21, Proposition 9] can be found in Appendix B. See also Pitman [23, Chapter 4].*

3.2. The Gaussian fluctuation

We shall now focus on a Gaussian fluctuation of $K_{r,n}$, properly normalized, around an estimator of its almost-sure limit $p_\alpha(r)S_{\alpha,\theta}$.

Theorem 3.3. *Let $K_{r,n}$ be the number of partition subsets of size r in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then, for all $r \geq 1$*

$$\frac{1}{\sqrt{K_{r,n}}}\left(K_{r,n} - \frac{A_{r,n}}{b_{r,n}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1), \tag{99}$$

where $\mathcal{N}(0,1)$ denotes a standard Gaussian random variable. Moreover, we also have for all $r \geq 1$

$$\sqrt{n^{\alpha}}\left(\frac{K_{r,n}}{n^{\alpha}} - \frac{A_{r,n}}{b_{r,n}n^{\alpha}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \sqrt{p_\alpha(r)S'_{\alpha,\theta}}\,\mathcal{N}(0,1), \tag{100}$$

where $S'_{\alpha,\theta}$ is a random variable independent of $\mathcal{N}(0,1)$ and sharing the same distribution as $S_{\alpha,\theta}$.

Proof. We are going to carry out the proof for any $r \geq 1$. It follows from (83), (90), (91) and (92) together with Toeplitz lemma that the predictable quadratic variation $\langle M_r\rangle_n$ satisfies

$$\lim_{n\to+\infty} \frac{\langle M_r\rangle_n}{n^{2r-\alpha}} = p_\alpha(r)\left(\frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}\right)^2 S_{\alpha,\theta} \qquad \text{a.s.} \tag{101}$$

Moreover, we have from (84) with $p=1$ and $p=2$ that

$$\lim_{n\to+\infty} n^{1-\alpha}\mathbb{E}[p_{r,n}] = (r-1-\alpha)p_\alpha(r-1)\mathbb{E}[S_{\alpha,\theta}] \tag{102}$$

and

$$\lim_{n\to+\infty} n^{2-2\alpha}\mathbb{E}[p_{r,n}^2] = (r-1-\alpha)^2p_\alpha^2(r-1)\mathbb{E}[S_{\alpha,\theta}^2]. \tag{103}$$

In addition, we also have from (84) with $p=1$ and $p=2$ that

$$\lim_{n\to+\infty} n^{1-\alpha}\mathbb{E}[q_{r,n}] = (r-\alpha)p_\alpha(r)\mathbb{E}[S_{\alpha,\theta}] \tag{104}$$

and

$$\lim_{n\to+\infty} n^{2-2\alpha}\mathbb{E}[q_{r,n}^2] = (r-\alpha)^2p_\alpha^2(r)\mathbb{E}[S_{\alpha,\theta}^2]. \tag{105}$$

Therefore, denote $s_{r,n}^2 = \mathbb{E}[\langle M_r\rangle_n]$. We deduce from (90) and (92) together with (102), (103), (104) and (105) that

$$\lim_{n\to+\infty}\frac{s_{r,n}^2}{n^{2r-\alpha}} = \lim_{n\to+\infty}\frac{1}{n^{2r-\alpha}}\sum_{k=1}^{n-1} b_{r,k+1}^2\big(\mathbb{E}[p_{r,k}] + \mathbb{E}[q_{r,k}]\big),$$

$$= \frac{p_\alpha(r)}{\alpha}\left(\frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}\right). \tag{106}$$

Hence, we obtain from (101) and (106) that

$$\lim_{n\to+\infty} \frac{\langle M_r \rangle_n}{s_{r,n}^2} = \alpha\left(\frac{\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}\right) S_{\alpha,\theta} \qquad \text{a.s.} \tag{107}$$

We are now going to show that for any $\eta > 0$,

$$\lim_{n\to+\infty} \frac{1}{n^{2r-\alpha}} \sum_{k=1}^{n} \mathbb{E}\big[\Delta M_{r,k}^2 \mathrm{I}_{\{|\Delta M_{r,k}|>\eta s_{r,n}\}}\big] = 0. \tag{108}$$

We clearly have for any $\eta > 0$,

$$\sum_{k=1}^{n} \mathbb{E}\big[\Delta M_{r,k}^2 \mathrm{I}_{\{|\Delta M_{r,k}|>\eta s_{r,n}\}}\big] \le \frac{1}{\eta^2 s_{r,n}^2} \sum_{k=1}^{n} \mathbb{E}\big[\Delta M_{r,k}^4\big].$$

According to (106), in order to prove (108), it is only necessary to show that

$$\lim_{n\to+\infty} \frac{1}{n^{4r-2\alpha}} \sum_{k=1}^{n} \mathbb{E}\big[\Delta M_{r,k}^4\big] = 0. \tag{109}$$

As in the proof of (53), it follows from tedious but straightforward calculations that for all $n \ge 1$,

$$\mathbb{E}\big[\Delta M_{r,n+1}^4 | \mathcal{F}_n\big] = b_{r,n+1}^4 \sum_{\ell=1}^{4} e_{r,n}(\ell) \tag{110}$$

where

$$\begin{aligned}
e_{r,n}(1) &= (p_{r,n} + q_{r,n}),\\
e_{r,n}(2) &= -4(p_{r,n} - q_{r,n})^2,\\
e_{r,n}(3) &= 6(p_{r,n} + q_{r,n})(p_{r,n} - q_{r,n})^2,\\
e_{r,n}(4) &= -3(p_{r,n} - q_{r,n})^4.
\end{aligned}$$

Then, it follows from (110) that it exists some constant $C > 0$ such that for n large enough,

$$\mathbb{E}\big[\Delta M_{r,n+1}^4 | \mathcal{F}_n\big] \le C b_{r,n+1}^4 \big(p_{r,n} + q_{r,n}\big). \tag{111}$$

Taking the expectation on both sides of (111), we obtain that for n large enough,

$$\mathbb{E}\big[\Delta M_{r,n+1}^4\big] \leq C b_{r,n+1}^4 \big(\mathbb{E}[p_{r,n}] + \mathbb{E}[q_{r,n}]\big). \tag{112}$$

Consequently, we deduce from (92), (102), (104) and (112) that

$$\sum_{k=1}^{n} \mathbb{E}\big[\Delta M_{r,k}^4\big] = O\left(\sum_{k=1}^{n} k^{4r-1-3\alpha}\right) = O(n^{4r-3\alpha})$$

which leads to (109). Hence, we find from Corollary 1 in [16] that

$$\frac{M_{r,n}}{\sqrt{\langle M_r \rangle_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \tag{113}$$

and

$$\frac{M_{r,n}}{s_{r,n}} \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \sqrt{\frac{\alpha\Gamma(\alpha+\theta)}{\Gamma(\theta+1)}} \sqrt{S'_{\alpha,\theta}} \mathcal{N}(0,1) \tag{114}$$

where $S'_{\alpha,\theta}$ stands for a random variable independent of $\mathcal{N}(0,1)$ and sharing the same distribution as $S_{\alpha,\theta}$. Finally, as $M_{r,n} = b_{r,n}K_{r,n} - A_{r,n}$, the Gaussian fluctuations (99) and (100) follow from (92), (113) and (114) together with the almost-sure convergences (83), (101), (107) and Slutsky's lemma, which achieves the proof of Theorem 3.3. □

3.3. The law of the iterated logarithm

We conclude our asymptotic analysis of $K_{r,n}$ by establishing the law of the iterated logarithm.

Theorem 3.4. *Let $K_{r,n}$ be the number of partition subsets of size r in a random partition of $[n]$ distributed according to the EP model with $\alpha \in (0,1)$ and $\theta > -\alpha$. Then, for all $r \geq 1$*

$$\begin{aligned} &\limsup_{n\to\infty} \sqrt{\frac{n^\alpha}{2\log\log n}} \left(\frac{K_{r,n}}{n^\alpha} - \frac{A_{r,n}}{b_{r,n}n^\alpha}\right) \\ = & -\liminf_{n\to\infty} \sqrt{\frac{n^\alpha}{2\log\log n}} \left(\frac{K_{r,n}}{n^\alpha} - \frac{A_{r,n}}{b_{r,n}n^\alpha}\right) \\ = & \sqrt{p_\alpha(r) S_{\alpha,\theta}} \qquad a.s. \end{aligned} \tag{115}$$

In particular,

$$\limsup_{n\to\infty} \frac{n^\alpha}{2\log\log n} \left(\frac{K_{r,n}}{n^\alpha} - \frac{A_{r,n}}{b_{r,n}n^\alpha}\right)^2 = p_\alpha(r) S_{\alpha,\theta} \qquad a.s.$$

Proof. As in Section 2, the proof is a direct application of [16]. We have for any $\eta > 0$,

$$\sum_{n=1}^{\infty} \frac{n^{\alpha/2}}{n^r} \mathbb{E}\big[|\Delta M_{r,n}| \mathrm{I}_{\{\sqrt{n^\alpha}|\Delta M_{r,n}|>\eta n^r\}}\big] \leq \frac{1}{\eta^3} \sum_{n=1}^{\infty} \frac{n^{2\alpha}}{n^{4r}} \mathbb{E}\big[\Delta M_{r,n}^4\big]. \tag{116}$$

Hence, we deduce from (112) and (116) together with (92), (102) and (104) that it exists some constant $C > 0$ such that

$$\sum_{n=1}^{\infty} \frac{n^{\alpha/2}}{n^r} \mathbb{E}\big[|\Delta M_{r,n}| \mathrm{I}_{\{\sqrt{n^\alpha}|\Delta M_{r,n}|>\eta n^r\}}\big] \leq C \sum_{n=1}^{\infty} \frac{1}{n^{1+\alpha}} < \infty.$$

Moreover, we also have for any $\delta > 0$,

$$\sum_{n=1}^{\infty} \frac{n^{2\alpha}}{n^{4r}} \mathbb{E}\big[\Delta M_{r,n}^4 \mathrm{I}_{\{\sqrt{n^\alpha}|\Delta M_{r,n}|\leq \delta n^r\}}\big] \leq \sum_{n=1}^{\infty} \frac{n^{2\alpha}}{n^{4r}} \mathbb{E}\big[\Delta M_{r,n}^4\big] < \infty.$$

Hereafter, let $(P_{r,n})$ be the martingale defined by

$$P_{r,n+1} = \sum_{k=1}^{n} \frac{k^\alpha}{k^{2r}} \big(\Delta M_{r,k+1}^2 - \mathbb{E}[\Delta M_{r,k+1}^2 \,|\, \mathcal{F}_k]\big). \tag{117}$$

Its predictable quadratic variation is given by

$$\langle P_r \rangle_{n+1} = \sum_{k=1}^{n} \frac{k^{2\alpha}}{k^{4r}} \big(\mathbb{E}[\Delta M_{k+1}^4 \,|\, \mathcal{F}_k] - (\mathbb{E}[\Delta M_{k+1}^2 \,|\, \mathcal{F}_k])^2\big).$$

Hence, we obtain from (111) that it exists some constant $C > 0$ such that for n large enough,

$$\langle P_r \rangle_n \leq C \sum_{k=1}^{n} \frac{k^{2\alpha}}{k^{4r}} b_{r,k+1}^4 \big(p_{r,k} + q_{r,k}\big). \tag{118}$$

Consequently, we deduce from (90), (91) and (118) that $\langle P_r \rangle_n$ converges a.s. to a finite random variable. Then, it follows from the strong law of large numbers for martingales that $(P_{r,n})$ converges a.s. to a finite random variable. Therefore, all the conditions of the second part of Theorem 1 and Corollary 2 in [16] are satisfied, which ensures that

$$\limsup_{n\to\infty} \frac{M_{r,n}}{\sqrt{2\langle M_r \rangle_n \log\log n}} = -\liminf_{n\to\infty} \frac{M_{r,n}}{\sqrt{2\langle M_r \rangle_n \log\log n}} = 1 \quad \text{a.s.} \tag{119}$$

Finally, as $M_{r,n} = b_{r,n} K_{r,n} - A_{r,n}$, the law of iterated logarithm (115) follows from (92) together with the almost-sure convergence (101) and (119), which completes the proof of Theorem 3.4. □

4. Discussion

In this paper, we presented a unified and comprehensive treatment of the large n asymptotic behaviour of K_n and $K_{r,m}$ in the EP model with $\alpha \in (0, 1)$ and $\theta > -\alpha$. By means of a novel martingale construction for K_n and $K_{r,m}$, we obtained alternative, and rigorous, proofs of the almost-sure convergence of K_n and $K_{r,n}$, and also covered the gap of Gaussian fluctuations. We argue that our martingale approach may be further investigated to refine Theorem 2.1 and Theorem 3.1 in terms of Berry–Esseen theorems, as well as sharp large deviations and concentration inequalities for K_n and $K_{r,n}$. We refer to Feng and Hoppe [14], Favaro et al. [10], Favaro et al. [11], Dolera and Favaro [6] and Oliveira et al. [18] for early results along these lines of research. Investigating the large n asymptotic behaviour of functions that involve both K_n and the $K_{r,n}$'s is also a promising direction for future research. Our almost-sure limits imply

$$\lim_{n\to+\infty} \frac{K_{1,n}}{K_n} = \alpha \qquad \text{a.s.},$$

meaning that $K_{1,n}/K_n$ is a consistent estimator for the parameter α. Establishing a Gaussian fluctuation for $K_{1,n}/K_n$ is an interesting open problem, with potential applications in the context of Bayesian nonparametric statistics [3, 13]. A further problem of interest is to establish an almost-sure limit and a Gaussian fluctuation for $(K_{1,n}, K_{2,n}, \ldots)$, thus providing a counterpart of (5).

Appendix A. Sequential or generative construction of the EP model

The purpose of this appendix is to recall the sequential construction of the EP model given in the seminal work of Pitman [20, Proposition 9]. For any $\alpha \in [0, 1)$, $\theta > -\alpha$, an exchangeable random partition of $[n] = \{1, \ldots, n\}$ can be recursively constructed as follows: conditionally on the total size $K_n = k$ of the partition and on the partition subsets $\{A_1, \ldots, A_k\}$ of corresponding sizes $(n_1, \ldots, n_k)$, the partition $[n + 1]$ is an extension of $[n]$ such that the element $n + 1$ is attached to subset A_i for $1 \le i \le k$, with probability

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

or forms a new subset with probability

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

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

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

Hereafter, for $K_n \in \{1, \ldots, n\}$, denote by $\mathbf{N}_n = (N_{1,n}, \ldots, N_{K_n,n})$ the sizes of the partition subsets $\{A_1, \ldots, A_{K_n}\}$. Pitman [20, Proposition 9] shows that the above construction leads to the joint distribution (1) of the random vector $(K_n, \mathbf{N}_n)$. This is the reason with the above construction is referred to as the sequential or generative construction of the EP model. Equation (11) clearly follows from the above construction since

$$\mathbb{P}(K_{n+1} = K_n + 1 \,|\, K_n) = \frac{\alpha K_n + \theta}{n+\theta}.$$

Equation (67) also follows from the above construction as

$$\mathbb{P}(\xi_{1,n+1} = 1 \,|\, \mathcal{F}_n) = \frac{\alpha K_n + \theta}{n+\theta},$$

while for $n_1 = 1$,

$$\mathbb{P}(\xi_{1,n+1} = -1 \,|\, \mathcal{F}_n) = \frac{(1-\alpha)K_{1,n}}{n+\theta}.$$

Moreover, we also deduce Equation (76) from the above construction since for all $r \geq 2$ and for $n_i = r - 1$,

$$\mathbb{P}(\xi_{r,n+1} = 1 \,|\, \mathcal{F}_n) = \frac{(r-1-\alpha)K_{r-1,n}}{n+\theta},$$

whereas for $n_i = r$

$$\mathbb{P}(\xi_{r,n+1} = -1 \,|\, \mathcal{F}_n) = \frac{(r-\alpha)K_{r,n}}{n+\theta}.$$

Finally, it shows that Equations (11), (67) and (76) follow from the sequential or generative construction of the EP model.

Appendix B. Alternative proofs of the convergence in $\mathbb{L}^2$

B.1. $\mathbb{L}^2$ *convergence of the number of partition sets*

We propose an alternative proof, without martingale, of the $\mathbb{L}^2$ convergence of K_n,

$$\lim_{n\to+\infty} \mathbb{E}\left[\left(\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right)^2\right] = 0. \tag{B.1}$$

The proof follows from the Poisson process construction of the EP model [21, Proposition 9]. See also Pitman [23, Chapter 4]. From Pitman [23, Equation 3.11],

$$\mathbb{P}(K_n = k) = \frac{\left(\frac{\theta}{\alpha}\right)^{(k)}}{(\theta)^{(n)}} \mathscr{C}(n, k; \alpha). \tag{B.2}$$

Denote $T_{\alpha,\theta} = S_{\alpha,\theta}^{-1/\alpha}$. By combining Pitman [21, Equation 66] with Charalambides [4, Equation 2.61],

$$\mathbb{P}(K_n = k \,|\, T_{\alpha,\theta} = t) = V_{n,k}(t) \frac{\mathscr{C}(n, k; \alpha)}{\alpha^k}, \tag{B.3}$$

where

$$V_{n,k}(t) = \frac{1}{\Gamma(n - k\alpha)} \left(\frac{\alpha}{t^\alpha}\right)^k \int_0^1 v^{n-1-k\alpha} \frac{f_\alpha((1-v)t)}{f_\alpha(t)} \mathrm{d}v$$

with f_α being the positive α-Stable density function. Then, it follows from (7) and (B.3), as well as the tower property of the conditional expectation, that

$$\begin{aligned}
\mathbb{E}[K_n S_{\alpha,\theta}] &= \int_0^\infty \mathbb{E}[K_n S_{\alpha,\theta} | S_{\alpha,\theta} = s] f_{S_{\alpha,\theta}}(s) \mathrm{d}s, \\
&= \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} \int_0^{+\infty} \mathbb{E}[K_n | T_{\alpha,\theta} = t] t^{-\alpha-\theta} f_\alpha(t) \mathrm{d}t, \\
&= \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} \sum_{k=1}^n k \int_0^{+\infty} \mathbb{P}(K_n = k \,|\, T_{\alpha,\theta} = t) t^{-\alpha-\theta} f_\alpha(t) \mathrm{d}t, \\
&= \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} \sum_{k=1}^n \frac{k \mathscr{C}(n, k; \alpha)}{\Gamma(n - k\alpha)} I_{n,k}(t),
\end{aligned} \tag{B.4}$$

where

$$I_{n,k}(t) = \int_0^{+\infty} t^{-k\alpha-\alpha-\theta} \int_0^1 v^{n-1-k\alpha} f_\alpha((1-v)t) \mathrm{d}v \mathrm{d}t,$$

$$= \int_0^{+\infty} z^{-k\alpha-\alpha-\theta} \int_0^1 v^{n-1-k\alpha}(1-v)^{k\alpha+\alpha+\theta-1} f_\alpha(z) \mathrm{d}v \mathrm{d}z.$$

Since

$$\int_0^{+\infty} z^{-k\alpha-\alpha-\theta} f_\alpha(z) \mathrm{d}z = \frac{\Gamma\left(\theta/\alpha + k + 2\right)}{\Gamma(\alpha+\theta+k\alpha+1)},$$

and

$$\int_0^1 v^{n-k\alpha-1}(1-v)^{k\alpha+\alpha+\theta-1} \mathrm{d}v = \frac{\Gamma(n-k\alpha)\Gamma(k\alpha+\alpha+\theta)}{\Gamma(n+\alpha+\theta)},$$

the expectation (B.4) reduces to

$$\mathbb{E}[K_n S_{\alpha,\theta}] = \frac{\Gamma(\theta+1)}{\Gamma\left(\theta/\alpha+1\right)\Gamma(n+\alpha+\theta)} \sum_{k=1}^{n} \frac{k\mathscr{C}(n,k;\alpha)}{(\alpha+\theta+k\alpha)} \Gamma\left(\theta/\alpha + k + 2\right),$$

$$= \frac{\Gamma(\theta+n)}{\Gamma(n+\alpha+\theta)} \sum_{k=1}^{n} k \left(\frac{\theta}{\alpha} + k\right) \frac{\left(\frac{\theta}{\alpha}\right)^{(k)}}{(\theta)^{(n)}} \mathscr{C}(n,k;\alpha). \tag{B.5}$$

Hence, it follows from (B.2) and (B.5) that

$$\mathbb{E}[K_n S_{\alpha,\theta}] = \frac{\Gamma(\theta+n)}{\Gamma(n+\alpha+\theta)} \left(\frac{\theta}{\alpha}\mathbb{E}[K_n] + \mathbb{E}[K_n^2]\right). \tag{B.6}$$

We are now in the position to prove (B.1). We get from (41) with $p=1$ and $p=2$ that

$$\lim_{n\to+\infty} \frac{1}{n^\alpha}\mathbb{E}[K_n] = \mathbb{E}[S_{\alpha,\theta}] \qquad \text{and} \qquad \lim_{n\to+\infty} \frac{1}{n^{2\alpha}}\mathbb{E}[K_n^2] = \mathbb{E}[S_{\alpha,\theta}^2], \tag{B.7}$$

leading to

$$\lim_{n\to+\infty} \frac{1}{n^\alpha}\mathbb{E}[K_n S_{\alpha,\theta}] = \mathbb{E}[S_{\alpha,\theta}^2]. \tag{B.8}$$

Therefore, we deduce from (B.7) and (B.8) that

$$\lim_{n\to+\infty} \mathbb{E}\left[\left(\frac{K_n}{n^\alpha} - S_{\alpha,\theta}\right)^2\right] = \lim_{n\to+\infty} \frac{1}{n^{2\alpha}}\mathbb{E}[K_n^2] - \frac{2}{n^\alpha}\mathbb{E}[K_n S_{\alpha,\theta}] + \mathbb{E}[S_{\alpha,\theta}^2],$$

$$= \mathbb{E}[S_{\alpha,\theta}^2] - 2\mathbb{E}[S_{\alpha,\theta}^2] + \mathbb{E}[S_{\alpha,\theta}^2] = 0.$$

Similar arguments can be applied in order to show the convergence in $\mathbb{L}^p$ of $n^{-\alpha}K_n$ to $S_{\alpha,\theta}$, for any integer $p \geq 1$, thus providing an alternative proof of (21).

B.2. $\mathbb{L}^2$ convergence of the number of partition subsets of size r

We still rely on the Poisson process construction of the EP model [21, Proposition 9] to propose an alternative proof of the $\mathbb{L}^2$ convergence of $K_{r,n}$,

$$\lim_{n\to+\infty} \mathbb{E}\left[\left(\frac{K_{r,n}}{n^\alpha} - p_\alpha(r)S_{\alpha,\theta}\right)^2\right] = 0. \tag{B.9}$$

Denote $T_{\alpha,\theta} = S_{\alpha,\theta}^{-1/\alpha}$. By combining Pitman [21, Equation 66] with Favaro et al. [12, Proposition 1],

$$\mathbb{E}[K_{r,n} \,|\, T_{\alpha,\theta} = t] = p_\alpha(r)(n)_{(r)} \sum_{k=1}^{n} V_{n,k}(t) \frac{\mathscr{C}(n-r, k-1; \alpha)}{\alpha^k} \tag{B.10}$$

where

$$V_{n,k}(t) = \frac{1}{\Gamma(n-k\alpha)} \left(\frac{\alpha}{t^\alpha}\right)^k \int_0^1 v^{n-1-k\alpha} \frac{f_\alpha((1-v)t)}{f_\alpha(t)} \mathrm{d}v$$

with f_α being the positive α-Stable density function. Consequently, we can write that

$$\mathbb{E}[K_{r,n}S_{\alpha,\theta}] = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\alpha+1)} \int_0^{+\infty} \mathbb{E}[K_{r,n}|T_{\alpha,\theta} = t] t^{-\alpha-\theta} f_\alpha(t) \mathrm{d}t,$$

thus obtaining, along lines similar to the computation of $\mathbb{E}[K_n S_{\alpha,\theta}]$ in (B.5), that

$$\begin{aligned}
\mathbb{E}[K_{r,n}S_{\alpha,\theta}] &= \frac{\Gamma(\theta+1)}{\alpha\Gamma(n+\alpha+\theta)} p_\alpha(r)(n)_{(r)} \sum_{k=1}^{n} \left(\frac{\theta}{\alpha}+1\right)^{(k)} \mathscr{C}(n-r, k-1; \alpha) \\
&= \frac{\Gamma(\theta+1)}{\alpha\Gamma(n+\alpha+\theta)} p_\alpha(r)(n)_{(r)} \frac{\Gamma(\theta/\alpha+2)}{\Gamma(\theta/\alpha+1)} \sum_{k=0}^{n-r} \left(\frac{\theta}{\alpha}+2\right)^{(k)} \mathscr{C}(n-r, k; \alpha),
\end{aligned}$$

which leads to

$$\mathbb{E}[K_{r,n}S_{\alpha,\theta}] = \frac{\Gamma(\theta+1)}{\Gamma(n+\alpha+\theta)} p_\alpha(r)(n)_{(r)} \left(\frac{\alpha+\theta}{\alpha^2}\right) (\theta+2\alpha)^{(n-r)}. \tag{B.11}$$

We are now in position to prove (B.9). We find from (98) with $p = 2$ that

$$\lim_{n\to+\infty} \frac{1}{n^{2\alpha}} \mathbb{E}[K_{r,n}^2] = \big(p_r(\alpha)\big)^2 \mathbb{E}[\big(S_{\alpha,\theta}\big)^2]. \tag{B.12}$$

In addition, we deduce from (B.11) that

$$\lim_{n\to+\infty}\frac{1}{n^{\alpha}}\mathbb{E}[K_{r,n}S_{\alpha,\theta}]=p_r(\alpha)\left(\frac{\alpha+\theta}{\alpha^2}\right)\frac{\Gamma(\theta+1)}{\Gamma(2\alpha+\theta)}=p_r(\alpha)\mathbb{E}[S_{\alpha,\theta}^2]. \quad \text{(B.13)}$$

Therefore, it follows from (B.12) and (B.13) that

$$\begin{aligned}&\lim_{n\to+\infty}\mathbb{E}\left[\left(\frac{K_{r,n}}{n^{\alpha}}-p_r(\alpha)S_{\alpha,\theta}\right)^2\right]\\&=\lim_{n\to+\infty}\frac{1}{n^{2\alpha}}\mathbb{E}[K_{r,n}^2]-\frac{2p_r(\alpha)}{n^{\alpha}}\mathbb{E}[K_{r,n}S_{\alpha,\theta}]+\big(p_r(\alpha)\big)^2\mathbb{E}[\big(S_{\alpha,\theta}\big)^2],\\&=\big(p_r(\alpha)\big)^2\mathbb{E}[S_{\alpha,\theta}^2]-2\big(p_r(\alpha)\big)^2\mathbb{E}[S_{\alpha,\theta}^2]+\big(p_r(\alpha)\big)^2\mathbb{E}[\big(S_{\alpha,\theta}\big)^2]=0.\end{aligned}$$

Similar arguments can be applied to show the convergence in $\mathbb{L}^p$ of $n^{-\alpha}K_{r,n}$ to $p_\alpha(r)S_{\alpha,\theta}$, for any integer $p\geq 1$, thus leading to an alternative proof of (84).

Acknowledgements

The authors would like to thank two anonymous referees for their constructive comments which helped to improve the paper substantially. Stefano Favaro received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 817257. Stefano Favaro acknowledge the financial support from the Italian Ministry of Education, University and Research (MIUR), "Dipartimenti di Eccellenza" grant 2023-2027.

References

[1] 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

[2] Arratia, R., Barbour, A.D. and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics.

[3] Balocchi, C., Favaro, S. and Naulet, Z. (2023). Bayesian nonparametric inference for "species-sampling" problems. *Preprint arXiv:2203.06076*.

[4] Charalambides (2005) *Combinatorial methods in discrete distributions.* Wiley.

[5] Charalambides, C.A. (2007). Distributions of random partitions and their applications. *Methodology and Computing in Applied Probability* **9**, 163–193.

[6] Dolera, E. and Favaro, S. (2020). A Berry–Esseen theorem for Pitman's α–diversity. *The Annals of Applied Probability* **30**, 847–869.

[7] Dolera, E. and Favaro, S. (2021). A compound Poisson perspective of Ewens-Pitman sampling model. *Mathematics* **9**, 2820.

[8] Duflo, M. (1997) *Random iterative models.* Springer-Verlag.

[9] Ewens, W. (1972). The sampling theory or selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

[10] Favaro, S. and Feng, S. (2015). Large deviation principles for the Ewens-Pitman sampling model. *Electronic Journal of Probability* **20**, 1-26

[11] Favaro, S., Feng, S. and Gao. F. (2018). Moderate deviations for Ewens-Pitman sampling models. *Sankhya A* **80**, 330–341.

[12] Favaro, S., Lijoi, A. and Prünster, I. (2013). Conditional formulae for Gibbs-type exchangeable random partitions. *The Annals of Applied Probability* **23**, 1721–1754.

[13] Favaro, S. and Naulet, Z. (2023). Near-optimal estimation of the unseen under regularly varying tail populations. *Bernoulli* **29**, 3423–3442.

[14] 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.

[15] Hall, M. and Heyde, C.C. (1980) *Martingale limit theory and its applications.* Academic Press.

[16] Heyde, C.C. (1977). On central limit and iterated logarithm supplements to the martingale convergence theorem. *Journal of Applied Probability* **14**, 758–775.

[17] Korwar, R.M. and Hollander, M. (1973). Contribution to the theory of Dirichlet processes. *The Annals of Probability* **1**, 705–711.

[18] Oliveira, R.I., Pereira, A. and Ribeiro, R. (2022). Concentration in the generalized Chinese restaurant process. *Sankhya A* **80**, 628-670.

[19] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92**, 21–39.

[20] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields* **102**, 145–158.

[21] Pitman, J. (2003). Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed*, Goldstein, D.R. Eds. Institute of Mathematical Statistics.

[22] Pitman, J. and Yor, M. (1997). The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* **25**, 855–900.

[23] Pitman, J. (2006). *Combinatorial stochastic processes*. Lecture Notes in Mathematics, Springer Verlag.

[24] Zabell, S.L. (2005). *Symmetry and its discontents: essays on the history of inductive probability.* Cambridge University Press.

[25] Zolotarev, V.M. (1986). *One-dimensional Stable Distributions.* American Mathematical Society.