The Annals of Applied Probability
2015, Vol. 25, No. 2, 523–547
DOI: 10.1214/14-AAP1002

CENTRAL LIMIT THEOREMS FOR AN INDIAN BUFFET MODEL WITH RANDOM WEIGHTS

BY PATRIZIA BERTI, IRENE CRIMALDI[1], LUCA PRATELLI AND PIETRO RIGO

Università di Modena e Reggio-Emilia, IMT Institute for Advanced Studies, Accademia Navale and Università di Pavia

The three-parameter Indian buffet process is generalized. The possibly different role played by customers is taken into account by suitable (random) weights. Various limit theorems are also proved for such generalized Indian buffet process. Let L_n be the number of dishes experimented by the first n customers, and let $\overline{K}_n = (1/n)\sum_{i=1}^n K_i$ where K_i is the number of dishes tried by customer i. The asymptotic distributions of L_n and $\overline{K}_n$, suitably centered and scaled, are obtained. The convergence turns out to be stable (and not only in distribution). As a particular case, the results apply to the standard (i.e., nongeneralized) Indian buffet process.

1. Introduction. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space. Think of $\mathcal{X}$ as a collection of features potentially shared by an object. Such an object is assumed to have a finite number of features only and is identified with the features it possesses. To investigate the object, thus, we focus on the finite subsets of $\mathcal{X}$.

Each finite subset $B \subset \mathcal{X}$ can be associated to the measure $\mu_B = \sum_{x\in B}\delta_x$, where $\mu_\varnothing = 0$ and δ_x denotes the point mass at x. If B is random, μ_B is random as well. In fact, letting $F = \{\mu_B : B \text{ finite}\}$, there is a growing literature focusing on those random measures M satisfying $M \in F$ a.s. See [9] and most references quoted below in this section.

A remarkable example is the *Indian Buffet Process* (IBP) introduced by Griffiths and Ghahramani and developed by Thibaux and Jordan; see [17, 18, 33]. The objects are the customers which sequentially enter an infinite buffet $\mathcal{X}$ and the features are the dishes tasted by each customer. In this framework, each customer is modeled by a (completely) random measure M such that $M \in F$ a.s. The atoms of M represent the dishes experimented by the customer.

Our starting point is a three-parameter extension of IBP, referred to as *standard* IBP in the sequel, introduced in [9] and [32] to obtain power-law behavior. Fix $\alpha > 0$, $\beta \in [0, 1)$ and $c > -\beta$. Here, α is the mass parameter, β the discount parameter (or stability exponent) and c the concentration parameter. Also, let $\mathrm{Poi}(\lambda)$ denote the Poisson distribution with mean $\lambda \geq 0$, where $\mathrm{Poi}(0) = \delta_0$. The dynamics

Received April 2013; revised October 2013.
[1]Supported in part by CNR PNR project “CRISIS Lab.”
MSC2010 subject classifications. 60B10, 60F05, 60G09, 60G57, 62F15.
Key words and phrases. Bayesian nonparametrics, central limit theorem, conditional identity in distribution, Indian buffet process, random measure, random reinforcement, stable convergence.

of the standard IBP is as follows. Customer 1 tries $\mathrm{Poi}(\alpha)$ dishes. For each $n \geq 1$, let S_n be the collection of dishes experimented by the first n customers. Then:

- Customer $n+1$ selects a subset $S_n^* \subset S_n$. Each $x \in S_n$ is included or not into S_n^* independently of the other members of S_n. The inclusion probability is
$$\frac{\sum_{i=1}^n M_i\{x\} - \beta}{c+n},$$
where $M_i\{x\}$ is the indicator of the event {customer i selects dish x}.
- In addition to S_n^*, customer $n+1$ also tries $\mathrm{Poi}(\lambda_n)$ new dishes, where $\lambda_n = \alpha \frac{\Gamma(c+1)\Gamma(c+\beta+n)}{\Gamma(c+\beta)\Gamma(c+1+n)}$.

For $\beta = 0$, such a model reduces to the original IBP of [17, 18, 33].

IBP is a flexible tool, able to capture the dynamics of various real problems. In addition, IBP is a basic model in Bayesian nonparametrics; see [14] and [21]. In factor analysis, for instance, IBP works as an infinite-capacity prior over the space of latent factors; see [21]. In this way, the number of factors is not specified in advance but is inferred from the data. Such a number is also allowed to grow as new data points are observed. Among the other possible applications of IBP, we mention causal inference [35], modeling of choices [16], similarity judgements [26] and dyadic data [23].

Despite its prominent role, however, the asymptotics of IBP is largely neglected. To the best of our knowledge, the only known fact is the a.s. behavior of L_n (defined below) and some other related quantities for large n; see [9] and [32]. Nothing is known as regards limiting distributions.

In this paper, we aim to do two things:

First, we generalize the standard IBP. Indeed, the discount parameter β is allowed to take values in $(-\infty, 1)$ rather than in $[0, 1)$. More importantly, the possible different relevance of customers is taken into account by random weights. Let $R_n > 0$ be the weight attached to customer n. Then, for each $x \in S_n$, the inclusion probability becomes
$$\frac{\sum_{i=1}^n R_i M_i\{x\} - \beta}{c + \sum_{i=1}^n R_i}.$$
Similarly, the new dishes tried by customer $n+1$ are now $\mathrm{Poi}(\Lambda_n)$ rather than $\mathrm{Poi}(\lambda_n)$, where $\Lambda_n = \alpha \frac{\Gamma(c+1)\Gamma(c+\beta+\sum_{i=1}^n R_i)}{\Gamma(c+\beta)\Gamma(c+1+\sum_{i=1}^n R_i)}$. If $\beta \in [0, 1)$ and $R_n = 1$ for all n, the model reduces to the standard IBP.

Second, we investigate the asymptotics of the previous generalized IBP model. We focus on

$$L_n = \text{number of dishes experimented by the first } n \text{ customers} \quad \text{and}$$
$$\overline{K}_n = \frac{1}{n}\sum_{i=1}^n K_i \qquad \text{where } K_i = \text{number of dishes tried by customer } i.$$

Three results are obtained. Define $a_n(\beta) = \log n$ if $\beta = 0$ and $a_n(\beta) = n^\beta$ if $\beta \in (0, 1)$. Then, under some conditions on the weights R_n (see Theorems 4, 5, 8) it is shown that:

(i) if $\beta \in [0, 1)$, then $\frac{L_n}{a_n(\beta)} \xrightarrow{\text{a.s.}} \lambda$ where $\lambda > 0$ is a certain constant;

(ii) if $\beta \in [0, 1)$, then $\sqrt{a_n(\beta)}\{\frac{L_n}{a_n(\beta)} - \lambda\} \longrightarrow \mathcal{N}(0, \lambda)$ stably;

(iii) if $\beta < 1/2$, then $\overline{K}_n \xrightarrow{\text{a.s.}} Z$ and

$$\sqrt{n}\{\overline{K}_n - Z\} \longrightarrow \mathcal{N}(0, \sigma^2) \qquad \text{stably,}$$

$$\sqrt{n}\{\overline{K}_n - E(K_{n+1} \mid \mathcal{F}_n)\} \longrightarrow \mathcal{N}(0, \tau^2) \qquad \text{stably,}$$

where Z, σ^2, τ^2 are suitable *random* variables, and $\mathcal{F}_n$ is the sub-σ-field induced by the available information at time n.

Stable convergence is a strong form of convergence in distribution. The basic definition is recalled in Section 2.3. Further, $\mathcal{N}(0, a)$ denotes the Gaussian law with mean 0 and variance $a \geq 0$, where $\mathcal{N}(0, 0) = \delta_0$.

Among other things, the above results can be useful in making (asymptotic) inference on the model. As an example, suppose $\beta \in [0, 1)$. In view of (i),

$$\widehat{\beta}_n = \frac{\log L_n}{\log n}$$

is a strongly consistent estimator of β for each $\beta \in [0, 1)$. In turn, (ii) provides the limiting distribution of $\widehat{\beta}_n$ so that simple tests on β can be manufactured. Similarly, if $\beta < 1/2$, asymptotic confidence bounds for the random limit Z of $\overline{K}_n$ can be obtained by (iii); see Section 5.1.

Note also that, because of (iii), the convergence rate of $\overline{K}_n - E(K_{n+1} \mid \mathcal{F}_n)$ is at least $n^{-1/2}$. Therefore, $\overline{K}_n$ is a good predictor of K_{n+1} for large n and $\beta < 1/2$; see Section 5.1 again.

The results in (i)–(iii) hold in particular if $R_n = 1$ for all n. Thus, (ii) and (iii) provide the limiting distributions of L_n and $\overline{K}_n$ in the standard IBP model. Furthermore, in this case, (iii) holds for all $\beta < 1$ and not only for $\beta < 1/2$.

We close this section with some remarks on β and the R_n.

The discount parameter β. Roughly speaking, if $\beta < 0$, the inclusion probabilities are larger and the chances of tasting new dishes vanish very quickly; see Lemma 2. Define in fact

$$L = \sup_n L_n = \text{card}\{x \in \mathcal{X} : x \text{ is tried by some customer}\}.$$

Because of (i), L_n increases logarithmically if $\beta = 0$ while exhibits a power-law behavior if $\beta \in (0, 1)$. Accordingly, $L = \infty$ a.s. if $\beta \in [0, 1)$. On the contrary,

$$E(e^L) < \infty \qquad \text{if } \beta < 0;$$

see Lemma 3. In particular, $\beta < 0$ implies $L < \infty$ a.s., and this fact can help to describe some real situations.

Formally, the model studied in this paper makes sense whenever $R_n > \max(\beta, 0)$ for all n. Hence, one could also admit $\beta \geq 1$. However, $\beta = 1$ leads to trivialities. Instead, $\beta > 1$ could be potentially interesting, but it is hard to unify the latter case and $\beta < 1$. Accordingly, we will focus on $\beta < 1$.

Unless $R_n = 1$ for all n, the results in (iii) are available for $\beta < 1/2$ only. Certainly, (iii) can fail if $\beta \in [1/2, 1)$. Perhaps, some form of (iii) holds even if $\beta \in [1/2, 1)$, up to replacing $\sqrt{n}$ with some other norming constant and $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(0, \tau^2)$ with some other limit kernels. But we did not investigate this issue.

A last note is that β plays an analogous role to that of the discount parameter in the two-parameter Poisson–Dirichlet process. Indeed, such parameter regulates the asymptotic behavior of the number of distinct observed values, in the same way as β does for L_n. See, for example, [3, 28, 29] for the two-parameter Poisson–Dirichlet and [9, 32] for the standard IBP.

The weights R_n. Standard IBP has been generalized in various ways, mainly focusing on computational issues; see, for example, [13, 15, 25, 34]. In this paper, the possible need of distinguishing objects according to some associated random factor is dealt with. To this end, customer n is attached a random weight R_n. Indeed, it may be that different customers have different importance, due to some random cause, that does not affect their choices but is relevant to the choices of future customers. Analogous models occur in different settings, for instance in connection with Pólya urns and species sampling sequences; see [2–6, 27].

The model investigated in this paper, referred to as "weighted" IBP in the sequel, generally applies to evolutionary phenomena. In a biological framework, for instance, a newborn exhibits some features in common with the existing units with a probability depending on the latter's weights (reproductive power, ability of adapting to new environmental conditions or to compete for finite resources, and so on). The newborn also presents some new features that, in turn, will be transmitted to future generations with a probability depending on his/her weight. See, for example, [8] and [30].

Similar examples arise in connection with the evolution of language; see, for example, [12]. A neologism (i.e., a newly coined term, word, phrase or concept) is often directly attributable to a specific people (or journal, period, event and so on) and its diffusion depends on the importance of such a people. For instance, suppose we are given a sample of journals of the same type (customers) during several years. Each journal uses words (dishes), some of which have been previously used while some others are new. A word appearing for the first time in a journal has a probability of being reused which depends on the importance of the journal at issue.

Other applications of the weighted IBP could be found in Bayesian nonparametrics. Standard IBP is widely used as a prior on binary matrices with a fixed finite number of rows and infinitely many columns (rows correspond to objects and columns to features). The weighted IBP can be useful in all those settings where customers arrive *sequentially*. As an example, some dynamic networks present a competitive aspect, and not all nodes are equally successful in acquiring links. Suppose the network evolves in time, a node (customer) is added at every time step and some links are created with some of the existing nodes. The different ability of competing for links is modeled by a weight attached to each node; see for example, [7]. Following [24] and [31], each node could be described by a set of binary features (dishes) and the probability of a link is a function of the features of the involved nodes. A nonparametric latent feature model could be assessed at every time step, with the weighted IBP as a prior on the feature matrix.

A last remark concerns the probability distribution of the sequence (M_n), where M_n is the random measure corresponding to customer n. Because of the weights, unlike the standard IBP, (M_n) can fail to be exchangeable. Thus, the usual machinery of Bayesian nonparametrics cannot be automatically implemented, due to the lack of exchangeability, and this can create some technical drawbacks. On the other hand, the exchangeability assumption is often untenable in applications. In such cases, the weighted IBP is a realistic alternative to the standard IBP. We finally note that, when $\beta = 0$, (M_n) satisfies a weak form of exchangeability known as conditional identity in distribution; see Section 2.4 and Lemma 1.

2. Preliminaries.

2.1. *Basic notation.* Throughout, $\mathcal{X}$ is a separable metric space and $\mathcal{B}$ the Borel σ-field on $\mathcal{X}$. We let

$$\mathcal{M} = \{\mu : \mu \text{ is a finite positive measure on } \mathcal{B}\},$$

and we say that $\mu \in \mathcal{M}$ is *diffuse* in case $\mu\{x\} = 0$ for all $x \in \mathcal{X}$.

All random variables appearing in this paper, unless otherwise stated, are defined on a fixed probability space $(\Omega, \mathcal{A}, P)$. If $\mathcal{G} \subset \mathcal{A}$ is a sub-σ-field, and X and Y are random variables with values in the same measurable space, we write

$$X \mid \mathcal{G} \sim Y \mid \mathcal{G}$$

to mean that $P(X \in A \mid \mathcal{G}) = P(Y \in A \mid \mathcal{G})$ a.s. for each measurable set A.

2.2. *Random measures.* A *random measure* (r.m.) is a map $M : \Omega \to \mathcal{M}$ such that $\omega \mapsto M(\omega)(B)$ is $\mathcal{A}$-measurable for each $B \in \mathcal{B}$. In the sequel, we write $M(B)$ to denote the real random variable $\omega \mapsto M(\omega)(B)$. Similarly, if $f : \mathcal{X} \to \mathbb{R}$ is a bounded measurable function, $M(f)$ stands for

$$M(\omega)(f) = \int f(x) M(\omega)(dx).$$

A *completely* r.m. is an r.m. M such that $M(B_1), \dots, M(B_k)$ are independent random variables whenever $B_1, \dots, B_k \in \mathcal{B}$ are pairwise disjoint; see [20].

Let $\nu \in \mathcal{M}$. A *Poisson r.m. with intensity* ν is a completely r.m. M such that $M(B) \sim \mathrm{Poi}(\nu(B))$ for all $B \in \mathcal{B}$. Note that $M(B) = 0$ a.s. in case $\nu(B) = 0$. Note also that the intensity ν has been requested to be a finite measure (and not a σ-finite measure as it usually happens).

We refer to [10] and [20] for Poisson r.m.'s. We just note that a Poisson r.m. with intensity ν is easily obtained. Since ν has been assumed to be a finite measure, it suffices to let $M = 0$ if $\nu(\mathcal{X}) = 0$, and otherwise

$$M = I_{\{N>0\}} \sum_{j=1}^{N} \delta_{X_j},$$

where (X_j) is an i.i.d. sequence of $\mathcal{X}$-valued random variables with $X_1 \sim \nu/\nu(\mathcal{X})$, N is independent of (X_j) and $N \sim \mathrm{Poi}(\nu(\mathcal{X}))$.

As in Section 1, let $F = \{\mu_B : B \text{ finite}\}$ where $\mu_\varnothing = 0$ and $\mu_B = \sum_{x \in B} \delta_x$. Since $\mathcal{X}$ is separable metric and $\mathcal{B}$ the Borel σ-field, the set $\{M \in F\}$ belongs to $\mathcal{A}$ for every r.m. M. In this paper, we focus on those r.m.'s M satisfying $M \in F$ a.s. If M is a Poisson r.m. with intensity ν, then $M \in F$ a.s. if and only if ν is diffuse. Therefore, another class of r.m.'s is to be introduced.

Each $\nu \in \mathcal{M}$ can be uniquely written as $\nu = \nu_c + \nu_d$, where ν_c is diffuse and

$$\nu_d = \sum_j \gamma_j \delta_{x_j}$$

for some $\gamma_j \geq 0$ and $x_j \in \mathcal{X}$. (The case $\nu_d = 0$ corresponds to $\gamma_j = 0$ for all j.) Say that M is a *Bernoulli r.m. with hazard measure* ν, where $\nu \in \mathcal{M}$, if:

- $M = M_1 + M_2$ with M_1 and M_2 independent r.m.'s;
- M_1 is a Poisson r.m. with intensity ν_c;
- $M_2 = \sum_j V_j \delta_{x_j}$ where the V_j are independent indicators satisfying $P(V_j = 1) = \gamma_j$.

Some (obvious) consequences of the definition are the following:

– For each $B \in \mathcal{B}$, $E\{M(B)\} = \nu(B)$ and

$$E\{M(B)^2\} = \nu(B) + \nu(B)^2 - \sum_{x \in B} \nu\{x\}^2;$$

– $M = M_1$ a.s. if $\nu = \nu_c$ and $M = M_2$ a.s. if $\nu = \nu_d$;
– M is a completely r.m.;
– $M \in F$ a.s.

We will write

$$M \sim \mathrm{Be}\,P(\nu)$$

to mean that M is a Bernoulli r.m. with hazard measure ν.

2.3. *Stable convergence*. Stable convergence is a strong form of convergence in distribution. We just recall the basic definition and we refer to [11, 19] and references therein for more information.

An r.m. K such that $K(\omega)(\mathcal{X})=1$, for all $\omega\in\Omega$, is said to be a *kernel* or a *random probability measure*. Let K be a kernel and (X_n) a sequence of $\mathcal{X}$-valued random variables. Say that X_n *converges stably to* K if

$$E\{K(f)\mid H\}=\lim_n E\{f(X_n)\mid H\}$$

for all $H\in\mathcal{A}$ with $P(H)>0$ and all bounded continuous $f:\mathcal{X}\to\mathbb{R}$. (Recall that $\mathcal{A}$ denotes the basic σ-field on Ω.) For $H=\Omega$, stable convergence trivially implies convergence in distribution.

2.4. *Conditionally identically distributed sequences*. Let $(X_n:n\geq 1)$ be a sequence of random variables (with values in any measurable space) adapted to a filtration $(\mathcal{U}_n:n\geq 0)$. Say that (X_n) is *conditionally identically distributed* (c.i.d.) with respect to $(\mathcal{U}_n)$ in case

$$X_k\mid\mathcal{U}_n\sim X_{n+1}\mid\mathcal{U}_n\qquad\text{for all } k>n\geq 0.$$

Roughly speaking this means that, at each time $n\geq 0$, the future observations $(X_k:k>n)$ are identically distributed given the past $\mathcal{U}_n$. If $\mathcal{U}_0=\{\varnothing,\Omega\}$ and $\mathcal{U}_n=\sigma(X_1,\dots,X_n)$, the filtration $(\mathcal{U}_n)$ is not mentioned at all and (X_n) is just called c.i.d. Note that $X_k\sim X_1$ for all $k\geq 1$ whenever (X_n) is c.i.d.

The c.i.d. property is connected to exchangeability. Indeed, (X_n) is exchangeable if and only if it is stationary and c.i.d., and the asymptotic behavior of c.i.d. sequences is quite close to that of exchangeable ones. We refer to [6] for details.

3. The model. Let $(M_n:n\geq 1)$ be a sequence of r.m.'s and $(R_n:n\geq 1)$ a sequence of real random variables. The probability distribution of $((M_n,R_n):n\geq 1)$ is identified by the parameters m,α,β and c as follows:

- m is a diffuse probability measure on $\mathcal{B}$;
- α,β,c are real numbers such that $\alpha>0$, $\beta<1$ and $c>-\beta$;
- R_n independent of $(M_1,\dots,M_n,R_1,\dots,R_{n-1})$ and $R_n\geq u>\max(\beta,0)$, for some constant u and each $n\geq 1$;
- $M_{n+1}\mid\mathcal{F}_n\sim\operatorname{Be}P(\nu_n)$ for all $n\geq 0$, where

$$\mathcal{F}_0=\{\varnothing,\Omega\},\qquad \nu_0=\alpha m,\qquad \mathcal{F}_n=\sigma(M_1,\dots,M_n,R_1,\dots,R_n),$$

$$\nu_n=\sum_{x\in S_n}\frac{\sum_{i=1}^n R_iM_i\{x\}-\beta}{\sum_{i=1}^n R_i+c}\delta_x+\frac{\Gamma(c+1)\Gamma(c+\beta+\sum_{i=1}^n R_i)}{\Gamma(c+\beta)\Gamma(c+1+\sum_{i=1}^n R_i)}\alpha m\quad\text{and}$$

$$S_n=\{x\in\mathcal{X}:M_i\{x\}=1\text{ for some } i=1,\dots,n\}.$$

Our model is the sequence $((M_n,R_n):n\geq 1)$. It reduces to the standard IBP in case $\beta\in[0,1)$ and $R_n=1$ for all n. Note that M_1 is a Poisson r.m. with intensity αm. Note also that $M_n \in F$ a.s. for all $n \geq 1$, so that

$$S_n = \bigcup_{i=1}^{n} \text{Support}(M_i) \qquad \text{a.s.}$$

Formally, for such a model to make sense, β can be taken to be any real number satisfying $R_n > \max(\beta, 0)$ for all n. For the reasons explained in Section 1, however, in this paper we focus on $\beta < 1$. We also assume $R_n \geq u$, for all n and some constant $u > \max(\beta, 0)$, as a mere technical assumption. In the sequel, we let

$$\Lambda_0 = \alpha \quad \text{and} \quad \Lambda_n = \alpha \frac{\Gamma(c+1)\Gamma(c+\beta+\sum_{i=1}^{n} R_i)}{\Gamma(c+\beta)\Gamma(c+1+\sum_{i=1}^{n} R_i)}.$$

In this notation, the diffuse part of ν_n can be written as $\Lambda_n m$.

As remarked in Section 1, R_n should be regarded as the weight of customer n. Thus, the possibly different role played by each customer can be taken into account.

Apart from the possible negative values of β, the parameters m, α, β and c have essentially the same meaning as in the standard IBP. The probability measure m allows us to draw, at each step $n \geq 1$, an i.i.d. sample of new dishes. In fact, $m(\mathcal{X} \setminus S_n) = 1$ a.s. for m is diffuse and S_n finite a.s. The mass parameter α controls the total number of tried dishes per customer. The concentration parameter c tunes the number of customers which try each dish. The discount parameter β has been discussed in Section 1.

An r.m. can be seen as a random variable with values in $(\mathcal{M}, \Sigma)$, where Σ is the σ-field on $\mathcal{M}$ generated by the maps $\mu \mapsto \mu(B)$ for all $B \in \mathcal{B}$. In the standard IBP case, (M_n) is an exchangeable sequence of random variables. Now, because of the R_n, exchangeability is generally lost. In fact, the same phenomenon (loss of exchangeability) occurs in various other extensions of IBP; see [13, 15, 25, 34]. However, under some conditions, (M_n) is c.i.d. with respect to the filtration

$$\mathcal{G}_0 = \{\varnothing, \Omega\}, \qquad \mathcal{G}_n = \mathcal{F}_n \vee \sigma(R_{n+1}) = \sigma(M_1, \dots, M_n, R_1, \dots, R_n, R_{n+1}).$$

We next prove this fact. The c.i.d. property has been recalled in Section 2.4.

LEMMA 1. *(M_n) is c.i.d. with respect to $(\mathcal{G}_n)$ if and only if*

$$\Lambda_{n+1} = \Lambda_n \left(1 - \frac{R_{n+1} - \beta}{c + \sum_{i=1}^{n+1} R_i}\right) \qquad \textit{a.s. for all } n \geq 0. \tag{1}$$

In particular, (M_n) is c.i.d. with respect to $(\mathcal{G}_n)$ if $\beta = 0$ or if $R_n = 1$ for all $n \geq 1$. [*In these cases, in fact, condition* (1) *is trivially true.*]

PROOF. We just give a sketch of the proof. Suppose

$$M_{n+2}(B) \mid \mathcal{G}_n \sim M_{n+1}(B) \mid \mathcal{G}_n \qquad \text{for each } n \geq 0 \text{ and } B \in \mathcal{B}. \tag{2}$$

Conditionally on $\mathcal{G}_n$, the r.m.'s M_{n+1} and M_{n+2} are both *completely* r.m.'s. Hence, condition (2) implies

$$M_{n+2} \mid \mathcal{G}_n \sim M_{n+1} \mid \mathcal{G}_n \qquad \text{for each } n \geq 0.$$

In turn, given $n \geq 0$ and $A \in \Sigma$, the previous condition yields

$$\begin{aligned} P(M_{n+3} \in A \mid \mathcal{G}_n) &= E\{P(M_{n+3} \in A \mid \mathcal{G}_{n+1}) \mid \mathcal{G}_n\} \\ &= E\{P(M_{n+2} \in A \mid \mathcal{G}_{n+1}) \mid \mathcal{G}_n\} \\ &= P(M_{n+2} \in A \mid \mathcal{G}_n) = P(M_{n+1} \in A \mid \mathcal{G}_n) \qquad \text{a.s.} \end{aligned}$$

Hence, $M_{n+3} \mid \mathcal{G}_n \sim M_{n+1} \mid \mathcal{G}_n$ for each $n \geq 0$. Iterating this argument, one obtains $M_k \mid \mathcal{G}_n \sim M_{n+1} \mid \mathcal{G}_n$ for all $k > n \geq 0$. Therefore, condition (2) is equivalent to (M_n) being c.i.d. with respect to $(\mathcal{G}_n)$. We next prove that (1) $\Leftrightarrow$ (2).

Fix $n \geq 0$ and $B \in \mathcal{B}$. It can be assumed $m(B) > 0$. Since R_{n+1} is independent of $(M_1, \dots, M_n, M_{n+1}, R_1, \dots, R_n)$, then

$$P(M_{n+1} \in A \mid \mathcal{G}_n) = P(M_{n+1} \in A \mid \mathcal{F}_n) \qquad \text{a.s. for all } A \in \Sigma.$$

Thus, for each $t \in \mathbb{R}$,

$$\begin{aligned} &E\{e^{tM_{n+1}(B)} \mid \mathcal{G}_n\} \\ &\quad = E\{e^{tM_{n+1}(B)} \mid \mathcal{F}_n\} \\ &\quad = \exp\big(m(B)(e^t - 1)\Lambda_n\big) \prod_{x \in S_n \cap B} \left\{1 + (e^t - 1)\frac{-\beta + \sum_{i=1}^n R_i M_i\{x\}}{c + \sum_{i=1}^n R_i}\right\} \\ &\qquad \text{a.s.,} \end{aligned}$$

where the second equality is because $M_{n+1} \mid \mathcal{F}_n \sim \text{Be}\,P(\nu_n)$. Similarly,

$$\begin{aligned} &E\{e^{tM_{n+2}(B)} \mid \mathcal{G}_n\} \\ &\quad = E\{E(e^{tM_{n+2}(B)} \mid \mathcal{G}_{n+1}) \mid \mathcal{G}_n\} \\ &\quad = \exp\big(m(B)(e^t - 1)\Lambda_{n+1}\big) \\ &\qquad \times E\left\{\prod_{x \in S_{n+1} \cap B} \left(1 + (e^t - 1)\frac{-\beta + \sum_{i=1}^{n+1} R_i M_i\{x\}}{c + \sum_{i=1}^{n+1} R_i}\right) \,\middle|\, \mathcal{G}_n\right\} \qquad \text{a.s.} \end{aligned}$$

Finally, after some computations, one obtains

$$\begin{aligned} &E\left\{\prod_{x \in S_{n+1} \cap B} \left(1 + (e^t - 1)\frac{-\beta + \sum_{i=1}^{n+1} R_i M_i\{x\}}{c + \sum_{i=1}^{n+1} R_i}\right) \,\middle|\, \mathcal{G}_n\right\} \\ &\quad = \exp\left(m(B)(e^t - 1)\Lambda_n \frac{R_{n+1} - \beta}{c + \sum_{i=1}^{n+1} R_i}\right) \\ &\qquad \times \prod_{x \in S_n \cap B} \left\{1 + (e^t - 1)\frac{-\beta + \sum_{i=1}^n R_i M_i\{x\}}{c + \sum_{i=1}^n R_i}\right\} \qquad \text{a.s.} \end{aligned}$$

Thus, condition (1) amounts to $E\{e^{tM_{n+2}(B)} \mid \mathcal{G}_n\} = E\{e^{tM_{n+1}(B)} \mid \mathcal{G}_n\}$ a.s. for each $t \in \mathbb{R}$, that is, conditions (1) and (2) are equivalent. $\square$

4. Asymptotic behavior of L_n.

Let N_i be the number of new dishes tried by customer i, that is,

$$N_i = \operatorname{card}(S_i \setminus S_{i-1}) \qquad \text{with } S_0 = \emptyset.$$

Note that N_i is $\mathcal{F}_i$-measurable and $N_i \mid \mathcal{F}_{i-1} \sim \operatorname{Poi}(\Lambda_{i-1})$.

This section is devoted to

$$L_n = \operatorname{card}(S_n) = \sum_{i=1}^n N_i,$$

the number of dishes experimented by the first n customers. Our main tool is the following technical lemma.

LEMMA 2. *There is a function $h : (0, \infty) \to \mathbb{R}$ such that*

$$\sup_{x \ge c+u} |xh(x)| < \infty \quad \text{and} \quad \Lambda_n = \alpha \frac{\Gamma(c+1)}{\Gamma(c+\beta)} \frac{1 + h(c + \sum_{i=1}^n R_i)}{(c + \sum_{i=1}^n R_i)^{1-\beta}} \qquad \text{for all } n \ge 1.$$

In particular,

$$\Lambda_n \le \frac{D}{n^{1-\beta}} \quad \text{and} \quad |\Lambda_{n+1} - \Lambda_n| \le \frac{D}{n^{2-\beta}} \qquad \text{for all } n \ge 1, \tag{3}$$

where D is a suitable constant (nonrandom and not depending on n).

PROOF. Just note that $\frac{\Gamma(x+\beta)}{\Gamma(x+1)} = x^{\beta-1}(1 + h(x))$, with h as required, for all $x > \max(0, -\beta)$; see, for example, formula (6.1.47) of [1]. To prove (3), let $v = \min(u, c+u)$. Since $c + u > c + \beta > 0$, then $v > 0$. Hence, (3) follows from

$$c + \sum_{i=1}^n R_i \ge c + nu = c + u + (n-1)u \ge nv.$$

$\square$

Let $L = \sup_n L_n$ be the number of dishes tried by some customer. A first consequence of Lemma 2 is that $\beta < 0$ implies $L < \infty$ a.s.

LEMMA 3. *$P(N_i > \frac{1}{1-\beta}$ infinitely often$) = 0$. Moreover, $E(e^L) < \infty$ if $\beta < 0$.*

PROOF. Fix an integer $k \ge 1$. Since $N_{i+1} \mid \mathcal{F}_i \sim \operatorname{Poi}(\Lambda_i)$,

$$P(N_{i+1} \ge k) = E\{P(N_{i+1} \ge k \mid \mathcal{F}_i)\} = E\left\{e^{-\Lambda_i} \sum_{j \ge k} \frac{\Lambda_i^j}{j!}\right\} \le \frac{E(\Lambda_i^k)}{k!}.$$

By Lemma 2, $E(\Lambda_i^k) = \mathrm{O}(i^{-(1-\beta)k})$. Let

$$k = 1 + \max\{j \in \mathbb{Z} : j \le 1/(1-\beta)\}.$$

Since $k(1-\beta) > 1$, one obtains $\sum_i P(N_i > 1/(1-\beta)) = \sum_i P(N_i \ge k) < \infty$. Next, suppose $\beta < 0$. By Lemma 2, $\Lambda_n \le D n^{\beta-1}$ for some constant D. Letting $H = (e-1)D$ and noting that $E(e^{N_{n+1}} \mid \mathcal{F}_n) = e^{\Lambda_n(e-1)}$ a.s., one obtains

$$E(e^{L_{n+1}}) = E\{e^{L_n} E(e^{N_{n+1}} \mid \mathcal{F}_n)\} = E\{e^{L_n} e^{\Lambda_n(e-1)}\} \le E(e^{L_n}) e^{H n^{\beta-1}}$$
$$\le E(e^{L_{n-1}}) e^{H(n-1)^{\beta-1}} e^{H n^{\beta-1}} \le \cdots \le E(e^{L_1}) e^{H \sum_{j=1}^n j^{\beta-1}}.$$

Thus, $\beta < 0$ and $E(e^{L_1}) = E(e^{N_1}) < \infty$ yield

$$E(e^L) = \sup_n E(e^{L_n}) \le E(e^{L_1}) e^{H \sum_{j=1}^\infty j^{\beta-1}} < \infty. \qquad \square$$

In view of Lemma 3, if $\beta < 0$ there is a random index N such that $L_n = L_N$ a.s. for all $n \ge N$. The situation is quite different if $\beta \in [0, 1)$. In this case, the a.s. behavior of L_n for large n can be determined by a simple martingale argument.

In the rest of this section, we let $\beta \in [0, 1)$. Define

$$\overline{R}_n = \frac{1}{n} \sum_{i=1}^n R_i$$

and suppose that

$$\overline{R}_n \xrightarrow{\text{a.s.}} r \qquad \text{for some constant } r. \tag{4}$$

Since $R_i \ge u$ for all i, then $r \ge u > 0$. Define also

$$\lambda(\beta) = \frac{\alpha c}{r} \quad \text{if } \beta = 0 \quad \text{and} \quad \lambda(\beta) = \frac{\alpha \Gamma(c+1)}{\Gamma(c+\beta)} \frac{1}{\beta r^{1-\beta}} \quad \text{if } \beta \in (0, 1),$$
$$a_n(\beta) = \log n \quad \text{if } \beta = 0 \quad \text{and} \quad a_n(\beta) = n^\beta \quad \text{if } \beta \in (0, 1).$$

THEOREM 4. *If $\beta \in [0, 1)$ and condition (4) holds, then*

$$\frac{L_n}{a_n(\beta)} \xrightarrow{a.s.} \lambda(\beta).$$

PROOF. By Lemma 2, $\Lambda_j = \alpha \frac{\Gamma(c+1)}{\Gamma(c+\beta)} (c + \sum_{i=1}^j R_i)^{\beta-1} \{1 + h(c + \sum_{i=1}^j R_i)\}$ where the function h satisfies $|h(x)| \le (k/x)$ for all $x \ge c + u$ and some constant k. Write

$$\frac{\sum_{j=1}^{n-1} \Lambda_j}{a_n(\beta)} = \alpha \frac{\Gamma(c+1)}{\Gamma(c+\beta)} \frac{\sum_{j=1}^{n-1} j^{\beta-1} (c/j + \overline{R}_j)^{\beta-1}}{a_n(\beta)} + D_n,$$

$$\text{where } D_n = \alpha \frac{\Gamma(c+1)}{\Gamma(c+\beta)} \frac{\sum_{j=1}^{n-1} (c + \sum_{i=1}^j R_i)^{\beta-1} h(c + \sum_{i=1}^j R_i)}{a_n(\beta)}.$$

In view of (4), one obtains $D_n \xrightarrow{\text{a.s.}} 0$ and $\frac{\sum_{j=1}^{n-1} \Lambda_j}{a_n(\beta)} \xrightarrow{\text{a.s.}} \lambda(\beta)$. Next, define

$$T_0 = 0 \quad \text{and} \quad T_n = \sum_{j=1}^{n} \frac{N_j - E(N_j \mid \mathcal{F}_{j-1})}{a_j(\beta)} = \sum_{j=1}^{n} \frac{N_j - \Lambda_{j-1}}{a_j(\beta)}.$$

Then, (T_n) is a martingale with respect to $(\mathcal{F}_n)$ and

$$\begin{aligned} E(T_n^2) &= \sum_{j=1}^{n} \frac{E\{(N_j - \Lambda_{j-1})^2\}}{a_j(\beta)^2} = \sum_{j=1}^{n} \frac{E\{E((N_j - \Lambda_{j-1})^2 \mid \mathcal{F}_{j-1})\}}{a_j(\beta)^2} \\ &= \sum_{j=1}^{n} \frac{E(\Lambda_{j-1})}{a_j(\beta)^2}. \end{aligned}$$

Since $E(\Lambda_j) = \mathrm{O}(j^{-(1-\beta)})$, then $\sup_n E(T_n^2) = \sum_{j=1}^{\infty} \frac{E(\Lambda_{j-1})}{a_j(\beta)^2} < \infty$. Thus, T_n converges a.s., and Kronecker's lemma implies

$$\begin{aligned} \lim_n \frac{L_n}{a_n(\beta)} &= \lim_n \frac{\sum_{j=1}^{n} N_j}{a_n(\beta)} = \lim_n \frac{\sum_{j=1}^{n} \Lambda_{j-1}}{a_n(\beta)} \\ &= \lim_n \frac{\Lambda_0 + \sum_{j=1}^{n-1} \Lambda_j}{a_n(\beta)} = \lambda(\beta) \qquad \text{a.s.} \end{aligned}$$

□

In view of Theorem 4, as far as $\beta \in [0, 1)$ and the weights R_n meet the SLLN, L_n essentially behaves for large n as in the standard IBP model. The only difference is that the limit constant $\lambda(\beta)$ depends on r as well. (In the standard IBP one has $r = 1$.) Note also that, the R_n being independent, a sufficient condition for (4) is

$$\sup_n E(R_n^2) < \infty \quad \text{and} \quad \frac{\sum_{i=1}^{n} E(R_i)}{n} \longrightarrow r.$$

We next turn to the limiting distribution of L_n. To get something, stronger conditions on the R_n are to be requested.

THEOREM 5. *If* $\beta \in [0, 1)$ *and*

$$\overline{R}_n \xrightarrow{a.s.} r \quad \textit{and} \quad \frac{\sum_{j=1}^{n} j^{\beta-1} E|\overline{R}_j - r|}{\sqrt{a_n(\beta)}} \longrightarrow 0 \tag{5}$$

for some constant r, *then*

$$\sqrt{a_n(\beta)}\left\{\frac{L_n}{a_n(\beta)} - \lambda(\beta)\right\} \longrightarrow \mathcal{N}(0, \lambda(\beta)) \qquad \textit{stably}.$$

PROOF. We first prove that

$$\sqrt{a_n(\beta)}\left\{\frac{\sum_{j=1}^n \Lambda_{j-1}}{a_n(\beta)} - \lambda(\beta)\right\} \xrightarrow{P} 0. \tag{6}$$

By Lemma 2 and some calculations, condition (6) is equivalent to

$$Y_n := \frac{\sum_{j=1}^{n-1}\{(c+\sum_{i=1}^j R_i)^{\beta-1} - (rj)^{\beta-1}\}}{\sqrt{a_n(\beta)}} \xrightarrow{P} 0.$$

Let $v = \min(u, c+u)$. Then, $v > 0$, $r \geq u \geq v$ and $c + \sum_{i=1}^j R_i \geq vj$; see the proof of Lemma 2. Hence, one can estimates as follows:

$$\begin{aligned} E\left|(rj)^{\beta-1} - \left(c + \sum_{i=1}^j R_i\right)^{\beta-1}\right| &\leq \frac{E|(c+\sum_{i=1}^j R_i)^{1-\beta} - (rj)^{1-\beta}|}{(vj)^{2(1-\beta)}} \\ &\leq \frac{1}{(vj)^{2(1-\beta)}} \frac{1-\beta}{(vj)^\beta} E\left|c + \sum_{i=1}^j R_i - rj\right| \\ &\leq \frac{1-\beta}{v^{2-\beta}}\left\{\frac{|c|}{j^{2-\beta}} + \frac{E|\overline{R}_j - r|}{j^{1-\beta}}\right\}. \end{aligned}$$

Thus, condition (5) implies $E|Y_n| \to 0$. This proves condition (6).

Next, define

$$U_n = \sqrt{a_n(\beta)}\left\{\frac{L_n}{a_n(\beta)} - \frac{\sum_{j=1}^n \Lambda_{j-1}}{a_n(\beta)}\right\} = \frac{\sum_{j=1}^n (N_j - \Lambda_{j-1})}{\sqrt{a_n(\beta)}}.$$

In view of (6), it suffices to show that $U_n \longrightarrow \mathcal{N}(0, \lambda(\beta))$ stably. To this end, for $n \geq 1$ and $j = 1, \dots, n$, define

$$U_{n,j} = \frac{N_j - \Lambda_{j-1}}{\sqrt{a_n(\beta)}}, \qquad \mathcal{R}_{n,0} = \mathcal{F}_0 \quad \text{and} \quad \mathcal{R}_{n,j} = \mathcal{F}_j.$$

Then, $E(U_{n,j} \mid \mathcal{R}_{n,j-1}) = 0$ a.s., $\mathcal{R}_{n,j} \subset \mathcal{R}_{n+1,j}$ and $U_n = \sum_j U_{n,j}$. Thus, by the martingale CLT, $U_n \longrightarrow \mathcal{N}(0, \lambda(\beta))$ stably provided

$$\text{(i)} \quad \sum_{j=1}^n U_{n,j}^2 \xrightarrow{P} \lambda(\beta), \qquad \text{(ii)} \quad \max_{1\leq j\leq n} |U_{n,j}| \xrightarrow{P} 0,$$

$$\text{(iii)} \quad \sup_n E\left\{\max_{1\leq j\leq n} U_{n,j}^2\right\} < \infty;$$

see, for example, Theorem 3.2, page 58, of [19]. Let

$$H_j = (N_j - \Lambda_{j-1})^2 \quad \text{and} \quad D_n = \frac{\sum_{j=1}^n \{H_j - E(H_j \mid \mathcal{F}_{j-1})\}}{a_n(\beta)}.$$

By Kronecker's lemma and the same martingale argument used in the proof of Theorem 4, $D_n \xrightarrow{\text{a.s.}} 0$. Since $\overline{R}_j \xrightarrow{\text{a.s.}} r$ and $E(H_j \mid \mathcal{F}_{j-1}) = \Lambda_{j-1}$ a.s., then

$$\sum_{j=1}^{n} U_{n,j}^2 = \frac{\sum_{j=1}^n H_j}{a_n(\beta)} = D_n + \frac{\sum_{j=1}^n \Lambda_{j-1}}{a_n(\beta)} \xrightarrow{\text{a.s.}} \lambda(\beta).$$

This proves condition (i). As to (ii), fix $k \geq 1$ and note that

$$\max_{1\leq j\leq n} U_{n,j}^2 \leq \frac{\max_{1\leq j\leq k} H_j}{a_n(\beta)} + \max_{k<j\leq n} \frac{H_j}{a_j(\beta)} \leq \frac{\max_{1\leq j\leq k} H_j}{a_n(\beta)} + \sup_{j>k} \frac{H_j}{a_j(\beta)} \qquad \text{for } n > k.$$

Hence, $\limsup_n \max_{1\leq j\leq n} U_{n,j}^2 \leq \limsup_n \frac{H_n}{a_n(\beta)}$ and condition (ii) follows from

$$\frac{H_n}{a_n(\beta)} = \frac{\sum_{j=1}^n H_j}{a_n(\beta)} - \frac{\sum_{j=1}^{n-1} H_j}{a_n(\beta)} \xrightarrow{\text{a.s.}} 0.$$

Finally, condition (iii) is an immediate consequence of Lemma 2 and

$$E\Big\{\max_{1\leq j\leq n} U_{n,j}^2\Big\} \leq \frac{\sum_{j=1}^n E(H_j)}{a_n(\beta)} = \frac{\sum_{j=1}^n E(\Lambda_{j-1})}{a_n(\beta)}. \qquad \square$$

Note that, letting $R_n = 1$ for all n, Theorem 5 provides the limiting distribution of L_n in the standard IBP model.

For Theorem 5 to apply, condition (5) is to be checked. We now give conditions for (5). In particular, (5) is automatically true whenever $\sup_n E(R_n^2) < \infty$ and $E(R_n) = r$ for all n.

LEMMA 6. *Condition* (5) *holds provided* $\beta \in [0, 1)$ *and*

$$\sup_n E\big(R_n^2\big) < \infty \quad \textit{and} \quad \sqrt{n^\beta \log n}\{E(\overline{R}_n) - r\} \longrightarrow 0. \tag{7}$$

PROOF. Let $a = \sup_n E(R_n^2)$. Because of (7), $E(\overline{R}_n) \to r$. Thus, $\overline{R}_n \xrightarrow{\text{a.s.}} r$ since $a < \infty$ and (R_n) is independent. Moreover,

$$\begin{aligned} E|\overline{R}_j - r| &\leq E\big|\overline{R}_j - E(\overline{R}_j)\big| + \big|E(\overline{R}_j) - r\big| \leq \sqrt{\operatorname{var}(\overline{R}_j)} + \big|E(\overline{R}_j) - r\big| \\ &\leq \sqrt{a/j} + \big|E(\overline{R}_j) - r\big|. \end{aligned}$$

Hence, the second part of condition (5) follows from the above inequality and condition (7). $\square$

A last remark is in order. Fix a set $B \in \mathcal{B}$ and define

$$L_n(B) = \operatorname{card}(B \cap S_n)$$

to be the number of dishes, belonging to B, tried by the first n customers. The same arguments used for $L_n = L_n(\mathcal{X})$ apply to $L_n(B)$ and allow us to extend Theorems 4–5 as follows.

THEOREM 7. *Let $\beta \in [0, 1)$ and $B \in \mathcal{B}$. If condition* (4) *holds, then*

$$\frac{L_n(B)}{a_n(\beta)} \xrightarrow{a.s.} m(B)\lambda(\beta).$$

Moreover, under condition (5), *one obtains*

$$\sqrt{a_n(\beta)}\left\{\frac{L_n(B)}{a_n(\beta)} - m(B)\lambda(\beta)\right\} \longrightarrow \mathcal{N}\big(0, m(B)\lambda(\beta)\big) \qquad \textit{stably}.$$

PROOF. Let $N_i(B)$ denote the number of new dishes, belonging to B, tried by customer i. Then, $L_n(B) = \sum_{i=1}^n N_i(B)$ and $N_{i+1}(B) \mid \mathcal{F}_i \sim \mathrm{Poi}(m(B)\Lambda_i)$. Therefore, it suffices to repeat the proofs of Theorems 4–5 with $N_i(B)$ in the place of N_i and $m(B)\Lambda_i$ in the place of Λ_i. $\square$

5. Asymptotic behavior of $\overline{K}_n$.

5.1. *The result.* Let $K_i = M_i(\mathcal{X})$ be the number of dishes experimented by customer i and

$$\overline{K}_n = \frac{1}{n}\sum_{i=1}^n K_i$$

the mean number of dishes tried by each of the first n customers.

In IBP-type models, $\overline{K}_n$ is a meaningful quantity. One reason is the following. If the parameters m, α, β and c are unknown, $E(K_{n+1} \mid \mathcal{F}_n)$ cannot be evaluated in closed form. Then, $\overline{K}_n$ could be used as an empirical predictor for the next random variable K_{n+1}. Such prediction is consistent whenever

$$V_n := \overline{K}_n - E(K_{n+1} \mid \mathcal{F}_n) \xrightarrow{P} 0.$$

But this is usually true. For instance, $V_n \xrightarrow{\text{a.s.}} 0$ if the sequence (K_n) is c.i.d. with respect to $(\mathcal{F}_n)$; see [6] and [4]. In general, the higher the convergence rate of V_n, the better $\overline{K}_n$ as a predictor of K_{n+1}.

Under some conditions, $\overline{K}_n \xrightarrow{\text{a.s.}} Z$ for some real random variable Z. Thus, two random centerings for $\overline{K}_n$ should be considered. One (and more natural) is Z, while the other is $E(K_{n+1} \mid \mathcal{F}_n)$, to evaluate the performances of $\overline{K}_n$ as a predictor of K_{n+1}. Taking $\sqrt{n}$ as a norming factor, this leads to investigate

$$\sqrt{n}\{\overline{K}_n - Z\} \quad \text{and} \quad \sqrt{n}V_n.$$

The limiting distributions of these quantities are provided by the next result.

THEOREM 8. *Suppose $\beta < 1/2$ and*

$$\sup_n R_n \le b, \qquad E(R_n) \longrightarrow r, \qquad E(R_n^2) \longrightarrow q,$$

for some constants b, r, q. Then

$$\overline{K}_n \xrightarrow{a.s.} Z \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^n K_i^2 \xrightarrow{a.s.} Q,$$

where Z and Q are real random variables such that $Z^2 < Q$ a.s. Moreover,

$$\sqrt{n}\{\overline{K}_n - Z\} \longrightarrow \mathcal{N}(0, \sigma^2) \qquad \textit{stably and}$$

$$\sqrt{n}\{\overline{K}_n - E(K_{n+1} \mid \mathcal{F}_n)\} \longrightarrow \mathcal{N}(0, \tau^2) \qquad \textit{stably,}$$

$$\textit{where } \sigma^2 = \frac{2q - r^2}{r^2}(Q - Z^2), \tau^2 = \frac{q - r^2}{r^2}(Q - Z^2).$$

If $R_n = 1$ for all n, the previous results hold for $\beta < 1$ (and not only for $\beta < 1/2$).

Theorem 8 is a consequence of Theorem 1 of [5]. The proof, even if conceptually simple, is technically rather hard.

Theorem 8 fails, as it stands, for $\beta \in [1/2, 1)$. Let μ_n denote the probability distribution of the random variable $\sqrt{n}\{\overline{K}_n - Z\}$. The sequence (μ_n) might be not tight if $\beta \in (1/2, 1)$. For instance, (μ_n) is not tight if $\beta \in (1/2, 1)$ and $R_n = r$ for all n, where r is any constant such that $r \neq 1$. If $\beta = 1/2$, instead, (μ_n) is tight, but the possible limit laws are not mixtures of centered Gaussian distributions. Thus, even if $\sqrt{n}\{\overline{K}_n - Z\}$ converges stably, the limit kernel is not $\mathcal{N}(0, \sigma^2)$.

Since $q \ge r^2$ and $Q > Z^2$ a.s., then $\sigma^2 > 0$ a.s. Hence, $\mathcal{N}(0, \sigma^2)$ is a nondegenerate kernel. Instead, $\mathcal{N}(0, \tau^2)$ may be degenerate. In fact, if $q = r^2$ then $\mathcal{N}(0, \tau^2) = \mathcal{N}(0, 0) = \delta_0$. Thus, for $q = r^2$, Theorem 8 yields $\sqrt{n}V_n \xrightarrow{P} 0$.

The convergence rate of V_n is $n^{-1/2}$ when $q > r^2$. Such a rate is even higher if $q = r^2$, since $\sqrt{n}V_n \xrightarrow{P} 0$. Overall, $\overline{K}_n$ seems to be a good predictor of K_{n+1} for large n.

Among other things, Theorem 8 can be useful to get asymptotic confidence bounds for Z. Define in fact

$$\widehat{\sigma}_n^2 = \left\{\frac{(2/n)\sum_{i=1}^n R_i^2}{\overline{R}_n^2} - 1\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_i^2 - \overline{K}_n^2\right\}.$$

Since $\widehat{\sigma}_n^2 \xrightarrow{\text{a.s.}} \sigma^2$ and $\sigma^2 > 0$ a.s., one obtains

$$I_{\{\widehat{\sigma}_n > 0\}} \frac{\sqrt{n}\{\overline{K}_n - Z\}}{\widehat{\sigma}_n} \longrightarrow \mathcal{N}(0, 1) \qquad \text{stably.}$$

Thus, $\overline{K}_n \pm \frac{u_a}{\sqrt{n}}\widehat{\sigma}_n$ provides an asymptotic confidence interval for Z with (approximate) level $1 - a$, where u_a is such that $\mathcal{N}(0, 1)(u_a, +\infty) = a/2$.

Theorem 8 works if $\beta \in [0,1)$ and $R_n = 1$ for all n, that is, it applies to the standard IBP model. Also, in this case, the convergence rate of V_n is greater than $n^{-1/2}$ (since $q = r^2 = 1$). Hence, $\overline{K}_n$ is a good (asymptotic) predictor of K_{n+1}.

5.2. *The proof.* We begin with a couple of results from [5]. Let (X_n) be a sequence of real integrable random variables, adapted to a filtration $(\mathcal{U}_n)$, and let

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad Z_n = E(X_{n+1} \mid \mathcal{U}_n).$$

LEMMA 9. *If $\sum_n n^{-2} E(X_n^2) < \infty$ and $Z_n \xrightarrow{a.s.} Z$, for some real random variable Z, then*

$$\overline{X}_n \xrightarrow{a.s.} Z \quad \text{and} \quad n \sum_{k \geq n} \frac{X_k}{k^2} \xrightarrow{a.s.} Z.$$

PROOF. This is exactly Lemma 2 of [5]. $\square$

THEOREM 10. *Suppose (X_n^2) is uniformly integrable and*

(j) $n^3 E\{(E(Z_{n+1} \mid \mathcal{U}_n) - Z_n)^2\} \longrightarrow 0.$

Then $Z_n \xrightarrow{a.s.} Z$ and $\overline{X}_n \xrightarrow{a.s.} Z$ for some real random variable Z. Moreover,

$$\sqrt{n}\{\overline{X}_n - Z_n\} \longrightarrow \mathcal{N}(0, U) \qquad \textit{stably and}$$
$$\sqrt{n}\{\overline{X}_n - Z\} \longrightarrow \mathcal{N}(0, U + V) \qquad \textit{stably}$$

for some real random variables U and V, provided

(jj) $E\{\sup_{k \geq 1} \sqrt{k}|Z_{k-1} - Z_k|\} < \infty$,

(jjj) $\frac{1}{n}\sum_{k=1}^{n}\{X_k - Z_{k-1} + k(Z_{k-1} - Z_k)\}^2 \xrightarrow{P} U$,

(jv) $n \sum_{k \geq n}(Z_{k-1} - Z_k)^2 \xrightarrow{a.s.} V$.

PROOF. First note that (Z_n) is a quasi-martingale because of (j) and it is uniformly integrable for (X_n^2) is uniformly integrable. Hence, $Z_n \xrightarrow{\text{a.s.}} Z$. By Lemma 9, one also obtains $\overline{X}_n \xrightarrow{\text{a.s.}} Z$. Next, assume conditions (jj)–(jjj)–(jv). By Theorem 1 of [5] (and the subsequent remarks) it is enough to show that

$$\sqrt{n} E\Big\{\sup_{k \geq n} |Z_{k-1} - Z_k|\Big\} \longrightarrow 0 \quad \text{and} \quad \frac{1}{\sqrt{n}} E\Big\{\max_{1 \leq k \leq n} k|Z_{k-1} - Z_k|\Big\} \longrightarrow 0.$$

Let $D_k = |Z_{k-1} - Z_k|$. Because of (jv),

$$n D_n^2 = n \sum_{k \geq n} D_k^2 - \frac{n}{n+1}(n+1) \sum_{k \geq n+1} D_k^2 \xrightarrow{\text{a.s.}} 0.$$

Thus $\sup_{k\geq n}\sqrt{k}D_k \xrightarrow{\text{a.s.}} 0$, and condition (jj) implies

$$\sqrt{n}E\Big\{\sup_{k\geq n} D_k\Big\} \leq E\Big\{\sup_{k\geq n}\sqrt{k}D_k\Big\} \longrightarrow 0.$$

Further, for $1 \leq h < n$, one obtains

$$\begin{aligned} E\Big\{\max_{1\leq k\leq n} kD_k\Big\} &\leq E\Big\{\max_{1\leq k\leq h} kD_k\Big\} + \sqrt{n}E\Big\{\max_{h<k\leq n}\sqrt{k}D_k\Big\} \\ &\leq E\Big\{\max_{1\leq k\leq h} kD_k\Big\} + \sqrt{n}E\Big\{\sup_{k>h}\sqrt{k}D_k\Big\}. \end{aligned}$$

Hence, it suffices to note that

$$\limsup_n \frac{1}{\sqrt{n}}E\Big\{\max_{1\leq k\leq n} kD_k\Big\} \leq E\Big\{\sup_{k>h}\sqrt{k}D_k\Big\} \qquad \text{for all } h \quad \text{and}$$

$$\lim_h E\Big\{\sup_{k>h}\sqrt{k}D_k\Big\} = 0. \qquad \square$$

Note that condition (j) is automatically true in case (X_n) is c.i.d. with respect to the filtration $(\mathcal{U}_n)$. We are now able to prove Theorem 8.

PROOF OF THEOREM 8. We apply Theorem 10 with $X_n = K_n$ and $\mathcal{U}_n = \mathcal{F}_n$. Let

$$J_n(x) = \frac{\sum_{i=1}^n R_i M_i\{x\} - \beta}{\sum_{i=1}^n R_i + c} \qquad \text{for } x \in \mathcal{X}.$$

Note that

$$\sum_{x\in S_n} J_n(x) = \frac{\sum_{i=1}^n R_i \sum_{x\in S_n} M_i\{x\} - \beta L_n}{\sum_{i=1}^n R_i + c} = \frac{\sum_{i=1}^n R_i K_i - \beta L_n}{\sum_{i=1}^n R_i + c},$$

and recall the notation

$$\mathcal{G}_n = \mathcal{F}_n \vee \sigma(R_{n+1}) = \sigma(M_1, \ldots, M_n, R_1, \ldots, R_n, R_{n+1}).$$

Uniform integrability of (K_n^2). It suffices to show that $\sup_n E\{e^{tK_n}\} < \infty$ for some $t > 0$. In particular, (K_n^2) is uniformly integrable for $\beta < 0$, since Lemma 3 yields

$$\sup_n E\{e^{K_n}\} \leq \sup_n E\{e^{L_n}\} = E\{e^L\} < \infty \qquad \text{if } \beta < 0.$$

Suppose $\beta \in [0, 1/2)$. Define $g(t) = e^t - 1$ and

$$W_n = \frac{\sum_{i=1}^n R_i K_i}{\sum_{i=1}^n R_i + c}.$$

Arguing as in Lemma 1 and since $\Lambda_n \le Dn^{\beta-1}$ for some constant D, one obtains

$$
\begin{aligned}
E\{e^{tK_{n+1}} \mid \mathcal{G}_n\} &= e^{g(t)\Lambda_n} \prod_{x\in S_n} \{1 + g(t) J_n(x)\} \le \exp\left\{ g(t)\Lambda_n + g(t) \sum_{x\in S_n} J_n(x) \right\} \\
&\le \exp\left\{ \frac{Dg(t)}{n^{1-\beta}} + g(t) \frac{\sum_{i=1}^n R_i K_i - \beta L_n}{\sum_{i=1}^n R_i + c} \right\} \\
&\le \exp\left\{ \frac{Dg(t)}{n^{1-\beta}} + g(t) W_n \right\} \qquad \text{a.s.}
\end{aligned}
$$

Hence, it is enough to show that $\sup_n E\{e^{tW_n}\} < \infty$ for some $t > 0$. We first prove $E\{e^{tW_n}\} < \infty$ for all $n \ge 1$ and $t > 0$, and subsequently $\sup_n E\{e^{tW_n}\} < \infty$ for a suitable $t > 0$. Define $U_n = \frac{R_{n+1}}{\sum_{i=1}^{n+1} R_i + c}$. Since U_n is $\mathcal{G}_n$-measurable,

$$
\begin{aligned}
E(e^{tW_{n+1}}) &= E\{\exp(tW_n(1-U_n)) E(e^{tU_n K_{n+1}} \mid \mathcal{G}_n)\} \\
&\le E\left\{ \exp\left(\frac{Dg(tU_n)}{n^{1-\beta}} \right) \exp\left(tW_n + \left(g(tU_n) - tU_n\right) W_n\right) \right\}.
\end{aligned}
$$

On noting that $U_n \le b/(nu)$,

$$
E(e^{tW_{n+1}}) \le \exp\left(\frac{Dg(tb/(nu))}{n^{1-\beta}} \right) E\{e^{\{t+g(tb/(nu))\}W_n}\}.
$$

Iterating this procedure, one obtains

$$
E(e^{tW_{n+1}}) \le a_n(t) E(e^{b_n(t)W_1}) \qquad \text{for suitable constants } a_n(t) \text{ and } b_n(t).
$$

Since $K_1 \sim \mathrm{Poi}(\alpha)$ and $W_1 = \frac{R_1}{R_1+c} K_1 \le \frac{b}{u+c} K_1$, then $E(e^{b_n(t)W_1}) < \infty$. Hence, $E\{e^{tW_n}\} < \infty$ for all $n \ge 1$ and $t > 0$. Observe now that $g(z) \le 2z$ and $g(z) - z \le z^2$ for $z \in [0, 1/2]$. Since $U_n \le b/(nu)$, then $tU_n \le 1/2$ for $n \ge (2bt)/u$. Hence, if $t \in (0, 1]$ and $n \ge (2b)/u$, then

$$
\begin{aligned}
E(e^{tW_{n+1}}) &\le \exp\left\{ \frac{2D(b/u)t}{n^{2-\beta}} \right\} E\{\exp(tW_n + (tU_n)^2 W_n)\} \\
&\le \exp\left\{ \frac{D^* t}{n^{2-\beta}} \right\} E\left\{ \exp\left(tW_n\left(1 + \frac{D^*}{n^2}\right)\right)\right\},
\end{aligned} \tag{8}
$$

where $D^* = \max\{2D(b/u), (b/u)^2\}$. Take t and n_0 such that

$$
t \in (0, 1/2], \qquad n_0 \ge \frac{2b}{u}, \qquad \prod_{j\ge n_0} \left(1 + \frac{D^*}{j^2}\right) \le 2.
$$

Iterating inequality (8), one finally obtains

$$
E(e^{tW_{n+1}}) \le \exp\left\{ \sum_{j\ge n_0} \frac{2D^* t}{j^{2-\beta}} \right\} E(e^{2tW_{n_0}}) \qquad \text{for each } n \ge n_0.
$$

Therefore $\sup_n E\{e^{tW_n}\} < \infty$, so that (K_n^2) is uniformly integrable.

We now turn to condition (j). Since $M_{n+1} \mid \mathcal{F}_n \sim \text{Be}\, P(\nu_n)$,

$$Z_n = E(K_{n+1} \mid \mathcal{F}_n) = \nu_n(\mathcal{X}) = \Lambda_n + \sum_{x \in S_n} J_n(x) = \Lambda_n + \frac{\sum_{i=1}^n R_i K_i - \beta L_n}{\sum_{i=1}^n R_i + c}.$$

On noting that $L_n = L_{n-1} + N_n$, a simple calculation yields

$$Z_n - Z_{n-1} = \Lambda_n - \Lambda_{n-1} + \frac{R_n(K_n - Z_{n-1}) + R_n \Lambda_{n-1} - \beta N_n}{\sum_{i=1}^n R_i + c}.$$

Condition (j). Since R_{n+1} is independent of $(M_1, \dots, M_n, M_{n+1}, R_1, \dots, R_n)$,

$$E(K_{n+1} \mid \mathcal{G}_n) = E(K_{n+1} \mid \mathcal{F}_n) = Z_n \quad \text{and}$$
$$E(N_{n+1} \mid \mathcal{G}_n) = E(N_{n+1} \mid \mathcal{F}_n) = \Lambda_n \qquad \text{a.s.}$$

It follows that

$$\begin{aligned} E(Z_{n+1} - Z_n \mid \mathcal{F}_n) &= E\{E(Z_{n+1} - Z_n \mid \mathcal{G}_n) \mid \mathcal{F}_n\} \\ &= E\left\{\Lambda_{n+1} - \Lambda_n + \frac{(R_{n+1} - \beta)\Lambda_n}{\sum_{i=1}^{n+1} R_i + c} \,\Big|\, \mathcal{F}_n\right\} \qquad \text{a.s.} \end{aligned}$$

Hence,

$$\begin{aligned} E\{(E(Z_{n+1} \mid \mathcal{F}_n) - Z_n)^2\} &= E\{E(Z_{n+1} - Z_n \mid \mathcal{F}_n)^2\} \\ &\le 2E\{(\Lambda_{n+1} - \Lambda_n)^2\} + \frac{2(b + |\beta|)^2}{u^2} \frac{E(\Lambda_n^2)}{n^2}. \end{aligned}$$

By Lemma 2, $E\{(\Lambda_{n+1} - \Lambda_n)^2\} = O(n^{2\beta-4})$ and $E(\Lambda_n^2) = O(n^{2\beta-2})$. Hence, condition (j) follows from $\beta < 1/2$ (or equivalently $4 - 2\beta > 3$).

Having proved condition (j) and (K_n^2) uniformly integrable, Theorem 10 yields $Z_n \xrightarrow{\text{a.s.}} Z$ and $\overline{K}_n \xrightarrow{\text{a.s.}} Z$ for some Z. We next prove $(1/n)\sum_{i=1}^n K_i^2 \xrightarrow{\text{a.s.}} Q$ for some Q such that $Q > Z^2$ a.s. Recall that

$$\sup_n E(K_n^4) \le \frac{4!}{t^4} \sup_n E(e^{tK_n}) < \infty \qquad \text{for a suitable } t > 0.$$

Hence, by Lemma 9, $(1/n)\sum_{i=1}^n K_i^2 \xrightarrow{\text{a.s.}} Q$ provided $E(K_{n+1}^2 \mid \mathcal{F}_n) \xrightarrow{\text{a.s.}} Q$.

Almost sure convergence of $E(K_{n+1}^2 \mid \mathcal{F}_n)$ and $Q > Z^2$ a.s. Let $G_n = \sum_{x \in S_n} J_n(x)^2$. Since $M_{n+1} \mid \mathcal{F}_n \sim \text{Be}\, P(\nu_n)$, then

$$E(K_{n+1}^2 \mid \mathcal{F}_n) = \nu_n(\mathcal{X}) + \nu_n(\mathcal{X})^2 - \sum_{x \in \mathcal{X}} \nu_n\{x\}^2 = Z_n + Z_n^2 - G_n \qquad \text{a.s.;}$$

see Section 2.2. Thus, since $Z_n \xrightarrow{\text{a.s.}} Z$ and (G_n) is uniformly integrable, it suffices to prove that (G_n) is a sub-martingale with respect to $(\mathcal{G}_n)$.

Let us define the random variables $\{T_{n,r} : n, r \ge 1\}$, with values in $\mathcal{X} \cup \{\infty\}$, as follows. For $n = 1$, let $T_{1,r} = \infty$ for $r > L_1$. If $L_1 > 0$, define $T_{1,1}, \dots, T_{1,L_1}$ to be the dishes tried by customer 1. By induction, at step $n \geq 2$, let

$$T_{n,r} = T_{n-1,r} \qquad \text{for } 1 \leq r \leq L_{n-1} \quad \text{and} \quad T_{n,r} = \infty \qquad \text{for } r > L_n.$$

If $L_n > L_{n-1}$, define $T_{n,L_{n-1}+1}, \dots, T_{n,L_n}$ to be the dishes tried for the first time by customer n. Then, $\sigma(T_{n,r}) \subset \mathcal{G}_n$ for all $r \geq 1$. Letting $J_n(\infty) = 0$, one also obtains

$$G_n = \sum_r J_n(T_{n,r})^2.$$

For fixed r, since $\sigma(T_{n,r}) \subset \mathcal{G}_n$, it follows that

$$\begin{aligned}
&E\{J_{n+1}(T_{n,r}) \mid \mathcal{G}_n\} \\
&\quad = E\{I_{\{r \leq L_n\}} J_{n+1}(T_{n,r}) \mid \mathcal{G}_n\} \\
&\quad = I_{\{r \leq L_n\}} \frac{-\beta + \sum_{i=1}^n R_i M_i\{T_{n,r}\} + R_{n+1} E\{M_{n+1}\{T_{n,r}\} \mid \mathcal{G}_n\}}{c + \sum_{i=1}^{n+1} R_i} \\
&\quad = I_{\{r \leq L_n\}} \frac{J_n(T_{n,r})\{c + \sum_{i=1}^n R_i\} + R_{n+1} J_n(T_{n,r})}{c + \sum_{i=1}^{n+1} R_i} = J_n(T_{n,r}) \qquad \text{a.s.}
\end{aligned}$$

Then

$$\begin{aligned}
E\{G_{n+1} \mid \mathcal{G}_n\} &= E\left\{\sum_r J_{n+1}(T_{n+1,r})^2 \,\middle|\, \mathcal{G}_n\right\} \geq E\left\{\sum_r J_{n+1}(T_{n,r})^2 \,\middle|\, \mathcal{G}_n\right\} \\
&= \sum_r E\{J_{n+1}(T_{n,r})^2 \mid \mathcal{G}_n\} \geq \sum_r E\{J_{n+1}(T_{n,r}) \mid \mathcal{G}_n\}^2 \\
&= \sum_r J_n(T_{n,r})^2 = G_n \qquad \text{a.s.}
\end{aligned}$$

Therefore, (G_n) is a $(\mathcal{G}_n)$-sub-martingale, as required.

From now on, Q denotes a real random variable satisfying

$$E(K_{n+1}^2 \mid \mathcal{F}_n) \xrightarrow{\text{a.s.}} Q \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n K_i^2 \xrightarrow{\text{a.s.}} Q.$$

Let us prove $Q > Z^2$ a.s. Let $Y_n = J_n(T_{n,1})$. Since (Y_n) is a [0, 1]-valued sub-martingale with respect to $(\mathcal{G}_n)$, one obtains $Y_n \xrightarrow{\text{a.s.}} Y$ for some random variable Y. Thus

$$\begin{aligned}
Q - Z^2 &= \lim_n \{E(K_{n+1}^2 \mid \mathcal{F}_n) - Z_n^2\} = \lim_n \{Z_n - G_n\} \\
&= \lim_n \left\{\Lambda_n + \sum_r J_n(T_{n,r})(1 - J_n(T_{n,r}))\right\} \\
&\geq \lim_n Y_n(1 - Y_n) = Y(1 - Y) \qquad \text{a.s.}
\end{aligned}$$

Since $L_n \xrightarrow{\text{a.s.}} \infty$ (because of Theorem 4), then $T_{n,1} \neq \infty$ eventually a.s. Hence, arguing as in [2] and [22] (see also Section 4.3 of [5]) it can be shown that Y has a diffuse distribution. Therefore, $0 < Y < 1$ and $Q - Z^2 \geq Y(1 - Y) > 0$ a.s.

We next turn to conditions (jj)–(jjj)–(jv).

Condition (jj). Since $E(Z_{n-1}^4) = E\{E(K_n \mid \mathcal{F}_{n-1})^4\} \le E(K_n^4)$, then

$$\sup_n E\{K_n^4 + Z_{n-1}^4 + \Lambda_{n-1}^4 + N_n^4\} \le 2 \sup_n E\{K_n^4 + \Lambda_{n-1}^4 + N_n^4\} < \infty.$$

Therefore,

$$\begin{aligned}
E\Big\{\Big(\sup_{n\ge 1}\sqrt{n}|Z_n - Z_{n-1}|\Big)^4\Big\}& \\
&\le \sum_{n=1}^{\infty} n^2 E\{(Z_n - Z_{n-1})^4\} \\
&\le D_1 \sum_{n=1}^{\infty} n^2 \bigg\{E\{(\Lambda_n - \Lambda_{n-1})^4\} + \frac{E\{K_n^4 + Z_{n-1}^4 + \Lambda_{n-1}^4 + N_n^4\}}{n^4}\bigg\} \\
&\le D_2 \sum_{n=1}^{\infty} \bigg\{\frac{1}{n^{6-4\beta}} + \frac{1}{n^2}\bigg\} < \infty,
\end{aligned}$$

where D_1 and D_2 are suitable constants.

In order to prove (jjj)–(jv), we let

$$U = \frac{q - r^2}{r^2}(Q - Z^2) \quad \text{and} \quad V = \frac{q}{r^2}(Q - Z^2).$$

Condition (jjj). Let $X_n = \{K_n - Z_{n-1} + n(Z_{n-1} - Z_n)\}^2$. On noting that $\sum_n n^2 E\{(Z_n - Z_{n-1})^4\} < \infty$, as shown in (jj), one obtains $\sum_n n^{-2} E(X_n^2) < \infty$. Thus, by Lemma 9, it suffices to prove $E(X_n \mid \mathcal{F}_{n-1}) \xrightarrow{\text{a.s.}} U$. To this end, we first note that

$$E\{(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\} = E\{K_n^2 \mid \mathcal{F}_{n-1}\} - Z_{n-1}^2 \xrightarrow{\text{a.s.}} Q - Z^2.$$

We next prove

$$(*) \quad n^2 E\bigg\{\frac{R_n^2(K_n - Z_{n-1})^2}{(\sum_{i=1}^n R_i + c)^2} \,\Big|\, \mathcal{F}_{n-1}\bigg\} \xrightarrow{\text{a.s.}} V;$$

$$(**) \quad n E\bigg\{\frac{R_n(K_n - Z_{n-1})^2}{\sum_{i=1}^n R_i + c} \,\Big|\, \mathcal{F}_{n-1}\bigg\} \xrightarrow{\text{a.s.}} Q - Z^2.$$

In fact,

$$\begin{aligned}
n^2 E\bigg\{\frac{R_n^2(K_n - Z_{n-1})^2}{(\sum_{i=1}^n R_i + c)^2} \,\Big|\, \mathcal{F}_{n-1}\bigg\}& \\
&\le n^2 \frac{E\{R_n^2(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\}}{(\sum_{i=1}^{n-1} R_i)^2} \\
&= \Big(\frac{n}{n-1}\Big)^2 \frac{E(R_n^2) E\{(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\}}{(\overline{R}_{n-1})^2} \xrightarrow{\text{a.s.}} \frac{q(Q - Z^2)}{r^2} = V.
\end{aligned}$$

Since $R_n \le b$, one also obtains

$$n^2 E\left\{\frac{R_n^2(K_n - Z_{n-1})^2}{(\sum_{i=1}^n R_i + c)^2} \,\Big|\, \mathcal{F}_{n-1}\right\} \ge n^2 \frac{E\{R_n^2(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\}}{(\sum_{i=1}^{n-1} R_i + b + c)^2} \xrightarrow{\text{a.s.}} V.$$

This proves condition (*). Similarly, (**) follows from

$$\begin{aligned} nE\left\{\frac{R_n(K_n - Z_{n-1})^2}{\sum_{i=1}^n R_i + c} \,\Big|\, \mathcal{F}_{n-1}\right\} &\le \frac{n}{n-1} \frac{E(R_n)E\{(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\}}{\overline{R}_{n-1}} \\ &\xrightarrow{\text{a.s.}} Q - Z^2 \quad \text{and} \\ nE\left\{\frac{R_n(K_n - Z_{n-1})^2}{\sum_{i=1}^n R_i + c} \,\Big|\, \mathcal{F}_{n-1}\right\} &\ge n \frac{E(R_n)E\{(K_n - Z_{n-1})^2 \mid \mathcal{F}_{n-1}\}}{\sum_{i=1}^{n-1} R_i + b + c} \\ &\xrightarrow{\text{a.s.}} Q - Z^2. \end{aligned}$$

Finally, by Lemma 2 and after some calculations, one obtains

$$n^2 E\{(Z_{n-1} - Z_n)^2 \mid \mathcal{F}_{n-1}\} - n^2 E\left\{\frac{R_n^2(K_n - Z_{n-1})^2}{(\sum_{i=1}^n R_i + c)^2} \,\Big|\, \mathcal{F}_{n-1}\right\} \xrightarrow{\text{a.s.}} 0,$$

$$nE\{(K_n - Z_{n-1})(Z_{n-1} - Z_n) \mid \mathcal{F}_{n-1}\} + nE\left\{\frac{R_n(K_n - Z_{n-1})^2}{\sum_{i=1}^n R_i + c} \,\Big|\, \mathcal{F}_{n-1}\right\} \xrightarrow{\text{a.s.}} 0.$$

Therefore, $n^2 E\{(Z_{n-1} - Z_n)^2 \mid \mathcal{F}_{n-1}\} \xrightarrow{\text{a.s.}} V$ and

$$2nE\{(K_n - Z_{n-1})(Z_{n-1} - Z_n) \mid \mathcal{F}_{n-1}\} \xrightarrow{\text{a.s.}} -2(Q - Z^2),$$

which in turn implies $E(X_n \mid \mathcal{F}_{n-1}) \xrightarrow{\text{a.s.}} V - (Q - Z^2) = U$.

Condition (jv). Let $X_n = n^2(Z_n - Z_{n-1})^2$. Since $\sum_n n^2 E\{(Z_n - Z_{n-1})^4\} < \infty$ and $n^2 E\{(Z_{n-1} - Z_n)^2 \mid \mathcal{F}_{n-1}\} \xrightarrow{\text{a.s.}} V$, as shown in (jj) and (jjj), Lemma 9 yields

$$n \sum_{k \ge n} (Z_{k-1} - Z_k)^2 = n \sum_{k \ge n} \frac{X_k}{k^2} \xrightarrow{\text{a.s.}} V.$$

In view of Theorem 10, this concludes the proof of the first part.

Finally, suppose $R_n = 1$ for all n. Then, by Lemma 1, (M_n) is c.i.d. with respect to the filtration $(\mathcal{G}_n)$. Thus, (M_n) is c.i.d. with respect to $(\mathcal{F}_n)$ as well, and condition (j) (with $\mathcal{U}_n = \mathcal{F}_n$) is automatically true. To complete the proof, it suffices to note that $\beta < 1/2$ is only needed in condition (j). All other points of this proof are valid for each $\beta < 1$. □

Acknowledgments. Irene Crimaldi is a member of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). We also thank Paolo Boldi for suggesting us to investigate IBP-type models, and an anonymous referee for various helpful remarks.

REFERENCES

[1] ABRAMOWITS, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graph, and Mathematical Tables*. U.S. Department of Commerce. Washington, DC.

[2] ALETTI, G., MAY, C. and SECCHI, P. (2009). A central limit theorem, and related results, for a two-color randomly reinforced urn. *Adv. in Appl. Probab.* **41** 829–844. MR2571318

[3] BASSETTI, F., CRIMALDI, I. and LEISEN, F. (2010). Conditionally identically distributed species sampling sequences. *Adv. in Appl. Probab.* **42** 433–459. MR2675111

[4] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2009). Rate of convergence of predictive distributions for dependent data. *Bernoulli* **15** 1351–1367. MR2597596

[5] BERTI, P., CRIMALDI, I., PRATELLI, L. and RIGO, P. (2011). A central limit theorem and its applications to multicolor randomly reinforced urns. *J. Appl. Probab.* **48** 527–546. MR2840314

[6] BERTI, P., PRATELLI, L. and RIGO, P. (2004). Limit theorems for a class of identically distributed random variables. *Ann. Probab.* **32** 2029–2052. MR2073184

[7] BIANCONI, G. and BARABASI, A. L. (2001). Competition and multiscaling in evolving networks. *Europhys. Lett.* **54** 436–442.

[8] BIANCONI, G., FERRETTI, L. and FRANZ, S. (2009). Non-neutral theory of biodiversity *Europhys. Lett.* **87** 28001.

[9] BRODERICK, T., JORDAN, M. I. and PITMAN, J. (2012). Beta processes, stick-breaking and power laws. *Bayesian Anal.* **7** 439–475. MR2934958

[10] ÇINLAR, E. (2011). *Probability and Stochastics*. *Graduate Texts in Mathematics* **261**. Springer, New York. MR2767184

[11] CRIMALDI, I., LETTA, G. and PRATELLI, L. (2007). A strong form of stable convergence. In *Séminaire de Probabilités XL*. *Lecture Notes in Math.* **1899** 203–225. Springer, Berlin. MR2409006

[12] CROFT, W. A. (2006). The relevance of an evolutionary model to historical linguistics. In *Competing Models of Linguistic Change*: *Evolution and Beyond* (O. Nedergaard Thomsen, ed.) 91–132. Benjamins, Amsterdam.

[13] DOSHI-VELEZ, F. and GHAHRAMANI, Z. (2009). Correlated non-parametric latent feature models. In *Uncertainty in Artificial Intelligence* **25** 143–150. AUAI Press, Montréal, QC.

[14] GERSHMAN, S. J. and BLEI, D. M. (2012). A tutorial on Bayesian nonparametric models. *J. Math. Psych.* **56** 1–12. MR2903470

[15] GERSHMAN, S. J., FRAZIER, P. I. and BLEI, D. M. (2012). Distance dependent infinite latent feature models. Available at arXiv:1110.5454v2 [stat.ML].

[16] GORUR, D., JAKEL, F. and RASMUSSEN, C. E. (2006). A choice model with infinitely many latent features. In *International Conference on Machine Learning*, 23 361–368. Omni Press, New York.

[17] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2006). Infinite latent feature models and the Indian buffet process. In *Advances in Neural Information Processing Systems*, 18 475–482. MIT Press, Cambridge, MA.

[18] GRIFFITHS, T. L. and GHAHRAMANI, Z. (2011). The Indian buffet process: An introduction and review. *J. Mach. Learn. Res.* **12** 1185–1224. MR2804598

[19] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*: *Probability and Mathematical Statistics*. Academic Press, New York. MR0624435

[20] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185

[21] KNOWLES, D. and GHAHRAMANI, Z. (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *Ann. Appl. Stat.* **5** 1534–1552. MR2849785

[22] MAY, C. and FLOURNOY, N. (2009). Asymptotics in response-adaptive designs generated by a two-color, randomly reinforced urn. *Ann. Statist.* **37** 1058–1078. MR2502661

[23] MEEDS, E., GHAHRAMANI, Z., NEAL, R. M. and ROWEIS, S. T. (2007). Modeling dyadic data with binary latent factors. In *Advances in Neural Information Processing Systems*, 19 977–984. MIT Press, Cambridge, MA.

[24] MILLER, K. T., GRIFFITHS, T. L. and JORDAN, M. I. (2009). Nonparametric latent feature models for link prediction. In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a Meeting Held 7–10 December 2009, Vancouver, British Columbia, Canada* 1276–1284. Curran Associates, Red Hook, NY.

[25] MILLER, K. T., GRIFFITHS, T. L. and JORDAN, M. I. (2008). The phylogenetic Indian buffet process: A non-exchangeable nonparametric prior for latent features. In *Uncertainty in Artificial Intelligence* **24** 403–410. AUAI Press, Corvallis, OR.

[26] NAVARRO, D. J. and GRIFFITHS, T. L. (2008). Latent features in similarity judgments: A nonparametric Bayesian approach. *Neural Comput*. **20** 2597–2628. MR2463044

[27] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181

[28] PITMAN, J. (2006). *Combinatorial Stochastic Processes*. *Lecture Notes in Math*. **1875**. Springer, Berlin. MR2245368

[29] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann*. *Probab.* **25** 855–900. MR1434129

[30] RICE, S. H. (2004). *Evolutionary Theory. Mathematical and Conceptual Foundations*. Sinauer Associates, Sunderlang, MA.

[31] SARKAR, P., CHAKRABARTI, D. and JORDAN, M. I. (2012). Nonparametric link prediction in dynamic networks. In *Proceedings of the 29th International Conference on Machine Learning*. Edinburgh.

[32] TEH, Y. W. and GÖORÜUR, D. (2009). Indian buffet processes with power-law behaviour. In *Advances in Neural Information Processing Systems*, 22 1838–1846. MIT Press, Cambridge, MA.

[33] THIBAUX, R. and JORDAN, M. I. (2007). Hierarchical Beta processes and the Indian buffet process. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, *AISTATS* 2007, *San Juan*, *Puerto Rico*, *March* 21–24, 2007 564–571. San Juan, Puerto Rico.

[34] WILLIAMSON, S., ORBANZ, P. and GHAHRAMANI, Z. (2010). Dependent Indian buffet processes. In *International Conference on Artificial Intelligence and Statistics*, 9 924–931.

[35] WOOD, F., GRIFFITHS, T. L. and GHAHRAMANI, Z. (2006). A non-parametric Bayesian method for inferring hidden causes. In *Uncertainty in Artificial Intelligence*, 22 536–543. AUAI Press, Arlington, VA.

P. BERTI
DIPARTIMENTO DI MATEMATICA PURA
ED APPLICATA "G. VITALI"
UNIVERSITÀ DI MODENA E REGGIO-EMILIA
VIA CAMPI 213/B
41100 MODENA
ITALY
E-MAIL: patrizia.berti@unimore.it

I. CRIMALDI
IMT INSTITUTE FOR ADVANCED STUDIES
PIAZZA SAN PONZIANO 6
55100 LUCCA
ITALY
E-MAIL: irene.crimaldi@imtlucca.it

L. PRATELLI
ACCADEMIA NAVALE
VIALE ITALIA 72
57100 LIVORNO
ITALY
E-MAIL: pratel@mail.dm.unipi.it

P. RIGO
DIPARTIMENTO DI MATEMATICA "F. CASORATI"
UNIVERSITÀ DI PAVIA
VIA FERRATA 1
7100 PAVIA
ITALY
E-MAIL: pietro.rigo@unipv.it