Central limit theorems for an Indian buffet model with random weights

Irene Crimaldi

IMT Institute for Advanced Studies, Lucca, Italy

The Indian buffet model

The Indian Buffet Model (IBM) is related to the modelling of the feature structure.

Let E be an *unbounded* collection of possible features that an entity can exhibit.

Such an entity is assumed to have a *finite* number of features only.

Different entities *can share* some features.

The maximum number of features is **not specified a priori**.

The *culinary metaphor* is the following: the entities are the customers which sequentially enter a buffet with an unbounded collection of dishes (available in infinity quantities) and the features exhibited by each entity are the dishes tasted by each customer.

The standard Indian buffet model (Griffiths-Ghahramani 2006, Thibaux-Jordan 2007, Teh-Gorur 2009)

Given $\alpha > 0$, $0 \leq \beta < 1$ and $c > -\beta$, the dynamics is the following:

- Customer 1 tries Poi(α) dishes
- For each $t \geq 1$, set $S_t =$collection of dishes experimented by the first t customers. Then:
  - Customer $t+1$ selects a subset S_t^* of S_t. Each $x \in S_t$ is included or not into S_t^* *independently* of the other elements of S_t. The *inclusion probability* is
    $$\frac{m(t,x) - \beta}{c+t}$$
    where $m(t,x) =$number of previous customers who tried dish x.
  - In addition to S_t^*, customer $t+1$ also tries Poi(λ_t) *new* dishes, where $\lambda_t = \alpha \frac{\Gamma(c+1)\Gamma(c+\beta+t)}{\Gamma(c+\beta)\Gamma(c+1+t)}$.

The Indian buffet model: main effect of each parameter

- α is the *mass parameter* that controls the total number of new dishes tried by a customer
- c is the *concentration parameter* that tunes the number of customers which try each dish
- β is the *discount parameter* (or *stability exponent*) that regulates the asymptotic behaviour of the random variable $L_t =$overall number of different dishes experimented by the first t customers

The Indian buffet model with random weights (Berti-Crimaldi-Pratelli-Rigo 2015)

We introduced a generalized IBM where the **different relevance of the entities (customers) is taken into account by random weights**:

- For each $x \in S_t$, the inclusion probability becomes

$$\frac{\sum_{i=1}^{t} R_i M_i\{x\} - \beta}{c + \sum_{i=1}^{t} R_i}$$

  where $M_i\{x\}$ =indicator of the event $\{$customer i selects dish $x\}$ and R_i is the weight attached to customer i

- λ_t is replaced by $\Lambda_t = \alpha \frac{\Gamma(c+1)\Gamma(c+\beta+\sum_{i=1}^{t} R_i)}{\Gamma(c+\beta)\Gamma(c+1+\sum_{i=1}^{t} R_i)}$

We assume R_t independent of the previous weights and the dishes tasted by the previous customers and customer t.

Remark: The exact formulation of the model can be given using *random measures*.

The Indian buffet model with random weights: some applications

The "weighted" IBM generally applies in order to model evolutionary phenomena in which we need to **distinguish the entities (customers) according to some associated random factor (weight)**, that does not affect their features (dishes) but it is relevant for the features of the future entities. For example:

- *Biological framework*: A new born 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 new born also presents some new features that, in turn, will be transmitted to future generations with a probability depending on his/her weight.

The Indian buffet model with random weights: some applications

- *Evolution of a language*: A neologism is often directly attributable to a specific “entity” (person or journal, period, event and so on) and its diffusion depends on the relevance of such an entity. 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.

The Indian buffet model with random weights: some applications

- *Dynamics of a complex network*: Some 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, and some links are created with some of the existing nodes. The different ability of competing for links can be modeled by a weight (fitness parameter) attached to each node. Each node could be described by a set of features (dishes) and the probability of a link could be modeled by an increasing function of the number of the common features of the involved pair of nodes.

The Indian buffet model with random weights: asymptotic results

Set $L_t = \text{card}(S_t)$ =overall number of different dishes tried by the first t customers

Set $L = \sup_t L_t$ =overall number of dishes tried

- If $\beta < 0$, then $L < +\infty$ a.s.
- If $\beta \in [0, 1)$ and $\overline{R}_t = \frac{1}{t}\sum_{i=1}^{t} R_i \xrightarrow{a.s.} r$ for some constant r, then

$$\frac{L_t}{a_t(\beta)} \xrightarrow{a.s.} \lambda$$

where $a_t(\beta) = \ln(t)$ if $\beta = 0\,,\quad a_t(\beta) = t^{\beta}$ if $\beta \in (0, 1)$,

$\lambda = \frac{\alpha\, c}{r}$ if $\beta = 0\,,\quad \lambda = \frac{\alpha\,\Gamma(c+1)}{\Gamma(c+\beta)}\,\frac{1}{\beta\, r^{1-\beta}}$ if $\beta \in (0, 1)$

Thus $\widehat{\beta}_t = \ln(L_t)/\ln(t)$ is a **strongly consistent estimator** of β

The Indian buffet model with random weights: asymptotic results

If $\beta \in [0, 1)$ and

$$\overline{R}_t \xrightarrow{a.s.} r \quad \text{and} \quad \lim_t \frac{\sum_{j=1}^t j^{\beta-1}\,\mathrm{E}\left[|\overline{R}_j - r|\right]}{\sqrt{a_t(\beta)}} = 0$$

for some constant r, then

$$\sqrt{a_t(\beta)}\left\{\frac{L_t}{a_t(\beta)} - \lambda\right\} \xrightarrow{stably} \mathcal{N}(0,\, \lambda)$$

Remark: Since the R_t are independent, above conditions hold when $\beta \in [0, 1)$ and

$$\sup_t \mathrm{E}[R_t^2] < +\infty \quad \text{and} \quad \lim_t \sqrt{t^\beta \log t}\,\left(\mathrm{E}[\overline{R}_t] - r\right) = 0$$

The Indian buffet model with random weights: asymptotic results

K_i =number of dishes experimented by customer i

$\overline{K}_t = \frac{1}{t}\sum_{i=1}^{t} K_i$ = mean number of dishes tried by each of the first t customers

Remark: If the parameters of the model and/or the weights R_i are *unknown*, the conditional expectation

$$\mathrm{E}\left[K_{t+1} \mid \mathcal{F}_t\right] = \frac{\sum_{i=1}^{t} R_i K_i - \beta L_t}{c + \sum_{i=1}^{t} R_i} + \Lambda_t$$

(where $\mathcal{F}_t$ is the natural σ-field associated to the model at time t) can *not be evaluated*. Then, $\overline{K}_t$ could be used as an **empirical predictor** of K_{t+1}

The Indian buffet model with random weights: asymptotic results

If $\beta < 1/2$ and

$$\sup_t R_t \leq b, \quad \lim_t \mathrm{E}[R_t] = r, \quad \lim_t \mathrm{E}[R_t^2] = q,$$

for some constants b, r, q, then

$$\mathrm{E}\big[K_{t+1} \mid \mathcal{F}_t\big] \xrightarrow{a.s.} Z, \quad \overline{K}_t \xrightarrow{a.s.} Z \quad \text{and} \quad \frac{1}{t}\sum_{i=1}^{t} K_i^2 \xrightarrow{a.s.} Q$$

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

$$\sqrt{t}\,\{\overline{K}_t - Z\} \xrightarrow{stably} \mathcal{N}(0,\,\sigma^2),\ \sqrt{t}\,\{\overline{K}_t - \mathrm{E}\big[K_{t+1} \mid \mathcal{F}_t\big]\} \xrightarrow{stably} \mathcal{N}(0,\,\tau^2)$$

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

Remark: If $R_t = 1$ for all t, the previous results hold for $\beta < 1$ (and not only for $\beta < 1/2$).

The Indian buffet model with random weights: inference

The convergence rate of $\left\{\overline{K}_t - \mathrm{E}\left[K_{t+1} \mid \mathcal{F}_t\right]\right\}$ is $t^{-1/2}$ when $q > r^2$ and such a rate is even higher if $q = r^2$ (e.g. standard IBM). Hence, $\overline{K}_t$ seems to be a **good empirical predictor** of K_{t+1} for large t.

The Indian buffet model with random weights: inference

Define

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

Then

$$I_{\{\widehat{\sigma}_t > 0\}} \frac{\sqrt{t}\,\{\overline{K}_t - Z\}}{\widehat{\sigma}_t} \xrightarrow{stably} \mathcal{N}(0,\,1)$$

Thus, $\overline{K}_t \pm \frac{u_a}{\sqrt{t}}\,\widehat{\sigma}_t$, with $\mathcal{N}(0,1)(u_a,\,+\infty) = a/2$, provides an asymptotic **confidence interval** for Z

Formulation of the model: Poisson random measures

E separable metric space endowed with its Borel σ-field $\mathcal{B}(E)$.

Set $\mathcal{M} = \{\nu : \nu$ is a finite positive measure on $\mathcal{B}(E)\}$

A random measure (r.m.) is a map $M : \Omega \to \mathcal{M}$ s.t. $M(B) : \omega \mapsto M(\omega)(B)$ is measurable for each $B \in \mathcal{B}(E)$.

A completely r.m. is a r.m. M such that $M(B_1), \dots, M(B_k)$ are independent random variables whenever $B_1, \dots, B_k \in \mathcal{B}(E)$ are pairwise disjoint.

Let $\nu \in \mathcal{M}$. A *Poisson r.m.* with intensity ν is a completely r.m. M such that $M(B) \sim \mathrm{Poi}\big(\nu(B)\big)$ for all $B \in \mathcal{B}(E)$.

Formulation of the model: Bernoulli random measures

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 $x_j \in E$.
M is a *Bernoulli r.m.* with hazard measure $\nu \in \mathcal{M}$ (we write $M \sim BeP(\nu)$) 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 $P(V_j = 1) = \gamma_j$.

Properties:

- M is a completely r.m.
- $M \in F$ a.s. where $F = \{\nu_B = \sum_{x \in B} \delta_x : B \text{ finite}\} \subseteq \mathcal{M}$

The Indian buffet model with random weights: formulation of the model

Let $(M_t)_{t\geq 1}$ be a sequence of r.m. and $(R_t)_{t\geq 1}$ a sequence of real r.v. s.t. the probability distribution of $([M_t, R_t])_{t\geq 1}$ is identified by the parameters m, α, β and c as follows:

- m is a diffuse probability measure on $\mathcal{B}(E)$;
- α, β, c are real numbers such that $\alpha > 0$, $\beta < 1$ and $c > -\beta$;
- R_t independent of $(M_1, \ldots, M_t, R_1, \ldots, R_{t-1})$ and $R_t \geq u > \max(\beta, 0)$, for some constant u and each $t \geq 1$;
- $M_{t+1} \mid \mathcal{F}_t \sim BeP(\nu_t)$ for all $t \geq 0$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_t = \sigma(M_1, \ldots, M_t, R_1, \ldots, R_t)$, $\nu_0 = \alpha m, \quad \nu_t = \sum_{x\in S_t} \frac{\sum_{i=1}^t R_i M_i\{x\} - \beta}{c + \sum_{i=1}^t R_i} \delta_x + \Lambda_t m$, $S_t = \{x \in E : \ M_i\{x\} = 1 \text{ for some } i = 1, \ldots, t\}$.

Remark: m allows to draw, at each $t \geq 1$, an i.i.d. sample of new dishes.

The Indian buffet model with random weights: a remark

A 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 $B \in \mathcal{B}(E)$.

Because of the weights, unlike the standard IBM, (M_t) *can fail to be exchangeable*.

This can create some technical drawbacks. However, the exchangeability assumption is often untenable in applications. In such cases, the weighted IBM is a *more realistic* alternative to the standard IBM.

The Indian buffet model with random weights: a property

(M_t) is *conditionally identically distributed* (c.i.d.) with respect to a suitable filtration if

$$\Lambda_{t+1} = \Lambda_t \left(1 - \frac{R_{t+1} - \beta}{c + \sum_{i=1}^{t+1} R_i}\right) \quad \text{a.s. for all } t \geq 0.$$

In particular, (M_t) is c.i.d. if $\beta = 0$ or if $R_t = 1$ for all $t \geq 1$.

A general central limit theorem

For $t \geq 1$, $X_t : (\Omega, \mathcal{A}, P) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\sup_{t\geq 1} \mathrm{E}[X_t^2] < +\infty$,
$\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ increasing filtration,

$$Z_t = \mathrm{E}[X_{t+1} \,|\, \mathcal{F}_t]$$

Suppose

$$\lim_t t^3 \,\mathrm{E}\Big\{ \big(\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t] - Z_t \big)^2 \Big\} = 0 \,. \tag{1}$$

Then

$$Z_t \xrightarrow{a.s.} Z \quad \text{and} \quad \overline{X}_t = \frac{\sum_{i=1}^t X_i}{t} \xrightarrow{a.s.} Z$$

Note: condition (1) obviously holds when $(Z_t)_t$ is a $\mathcal{F}$-martingale (e.g. exchangeable seq. with natural filtration or, more generally, $\mathcal{F}$-conditionally identically distributed seq.).
There are also cases in which condition (1) holds but $(Z_t)_t$ is not a martingale.

A general central limit theorem (Berti-Crimaldi-Pratelli-Rigo 2011)

Assume (X_t) adapted to $\mathcal{F}$, $(X_t^2)_t$ uniformly integrable, condition (1) and

(a) $\lim_t \frac{1}{\sqrt{t}}\, \mathrm{E}\{\max_{1\leq i\leq t} i|Z_{i-1}-Z_i|\,\} = 0$

(b) $\frac{1}{t}\sum_{i=1}^{t}\{X_i - Z_{i-1} + i(Z_{i-1}-Z_i)\}^2 \xrightarrow{P} U$

(c) $\lim_t \sqrt{t}\, \mathrm{E}[\sup_{i\geq t}|Z_{i-1}-Z_i|\,] = 0$

(d) $t\sum_{i\geq t}(Z_{i-1}-Z_i)^2 \xrightarrow{P} V$

Then

$$[C_t, D_t] = \left[\sqrt{t}(\overline{X}_t - Z_t), \sqrt{t}(Z_t - Z)\right] \xrightarrow{stably} \mathcal{N}(0,U)\otimes\mathcal{N}(0,V)$$

In particular,

$$W_t = \sqrt{t}(\overline{X}_t - Z) \xrightarrow{stably} \mathcal{N}(0, U+V)$$

Other contexts

- Multicolor Randomly Reinforced Urn Models
  (without or with dominant colors)
- Poisson-Dirichlet model/Chinese restaurant model
  (which belongs to the class of species sampling sequences)

These models are suitable in order to describe evolutionary phenomena, such as the evolution of populations, or of languages or of complex networks.

They are *preferential attachment models*.

Preferential Attachment Models

The **Preferential Attachment Models** are stochastic models in which, along the time-steps, different individuals or objects or categories (colors) receive some quantity, called “weight” (number of balls), in such a way that the higher the total weight they already have until a certain time-step, the greater the probability of receiving an additional weight at the next time-step (i.e. a “self-reinforcing” property).

The preferential attachment is a key feature governing the dynamics of many biological, economic and social systems.

Preferential Attachment principle

The preferential attachment principle has been studied by many authors and it can be found in the scientific literature in different forms:

- *Urn models*: Pólya urn (G. Pólya, Ann. Inst. H. Poincaré 1931) and related generalizations
- *Rich get richer rule* (H. A. Simon, Biometrika 1955)
- *Matthew effect* (R. K. Merton, Science 1968)
- *Cumulative advantage* (D. J. Price, J. Amer. Soc. Inform. Sci. 1976)
- *Preferential attachment* (Barabási and Albert, Science 1999)

Some references

- P. Berti, I. Crimaldi, L. Pratelli, P. Rigo (2015)
Central limit theorems for an Indian buffet model with random weights.
The Annals of Appl. Probab. 25, 523-547.

- T. L. Griffiths, Z. Ghahramani (2011)
The Indian Buffet Process: An Introduction and Review. *J. Machine Lear. Res.* 12, 1185-1224.

- Y. W. Teh, D. Gorur (2009)
Indian Buffet Processes with Power-law Behaviour. *Adv. in Neural Infor. Proc. Sys.* 22, 1838-1846.

- R. Thibaux R., M. I. Jordan (2007)
Hierarchical Beta Processes and the Indian Buffet Process.
Internat. Conf. on Art. Intel. Stat. 11, 564-571.

Random probability measures

$(\Omega, \mathcal{A}, P)$ probability space, E separable metric space

Def. A *random probability measure*, or *kernel*, on E is a family $K = \{K(\omega, \cdot) : \omega \in \Omega\}$ of probability measures on $\mathcal{B}(E)$ such that, for each bounded Borel function f on E, the map

$$\omega \mapsto Kf(\omega) = \int f(x)\, K(\omega, \mathrm{d}x)$$

is $\mathcal{A}$-measurable.

For $H \in \mathcal{A}$ with $P(H) > 0$, we define a probability measure on $\mathcal{B}(E)$ by

$$P_H K(B) = \mathrm{E}\big[K(\omega, B) \mid H\big] = \frac{1}{P(H)} \int_H K(\omega, B)\, P(\mathrm{d}\omega) \tag{2}$$

Stable convergence (Rényi 1963)

$Y_t : (\Omega, \mathcal{A}, P) \longmapsto (E, \mathcal{B}(E))$, K kernel on E

Def. $(Y_t)_t$ converges *stably* to K if

$$Y_t \xrightarrow{d} P_H K \qquad \text{under } P_H = P(\cdot \mid H)$$

for all $H \in \mathcal{A}$ with $P(H) > 0$.
In other words,

$$\lim_t E[f(Y_t) \mid H] = E[Kf \mid H]$$

for all $f \in \mathcal{C}_b(E)$ and $H \in \mathcal{A}$ with $P(H) > 0$

It is intermediate between convergence in law and convergence in probability.

See: Aldous-Eagleson (Ann. Probab., 1978), Jacod-Memin (Sem. de Probab., 1981), Fegin (Stoch. Proc. Appl., 1985), Peccati-Taqqu (Elect. J. Probab., 2006) and Crimaldi-Letta-Pratelli (Sem. de Probab., 2007), Peccati-Taqqu (J. Theor. Probab., 2008)

Conditional identity in distribution (Berti-Pratelli-Rigo 2004)

For $t \geq 1$, X_t r.v. on $(\Omega, \mathcal{A}, P)$, $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ increasing filtration

Def. $(X_t)_{t \geq 1}$ is *conditionally identically distributed* with respect to $\mathcal{F}$ if (X_t) is adapted to $\mathcal{F}$ and for each t

$$X_{t+n} \overset{d}{=} X_{t+1} \qquad \text{under } P(\cdot \mid H) \tag{3}$$

for each $n \geq 1$ and each $H \in \mathcal{F}_t$ with $P(H) > 0$.

It is equivalent to require that

$$Z_t = \mathrm{E}[f(X_{t+1}) \mid \mathcal{F}_t] \qquad \text{for } t \geq 0$$

is a $\mathcal{F}$-martingale for each measurable real function f s.t. $\mathrm{E}[\,|f(X_1)|\,] < +\infty$

Quasi-martingale

Def. A sequence of integrable real random variables $(Z_t)_{t\geq 0}$ is a *quasi-martingale* with respect to an increasing filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ if (X_t) is adapted to $\mathcal{F}$ and

$$\sum_{t\geq 0} \mathrm{E}\big[\, |\mathrm{E}[Z_{t+1}|\mathcal{F}_t] - Z_t|\,\big] < +\infty \tag{4}$$

Complements regarding the general central limit theorem: useful lemma

$\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ increasing filtration
For $t \geq 1$, $X_t : (\Omega, \mathcal{A}, P) \longmapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $\sigma(X_t) \subseteq \mathcal{F}_t$ for each t,

$$\overline{X}_t = \frac{1}{t}\sum_{i=1}^{t} X_i \quad \text{and} \quad Z_t = \mathrm{E}\big[X_{t+1} \mid \mathcal{F}_t\big].$$

Lemma
If $\sum_t t^{-2} E[X_t^2] < +\infty$ and $Z_t \xrightarrow{a.s.} Z$, for some real random variable Z, then

$$\overline{X}_t \xrightarrow{a.s.} Z \quad \text{and} \quad t\sum_{i\geq t} \frac{X_i}{i^2} \xrightarrow{a.s.} Z.$$

Complements regarding the general central limit theorem: sufficient conditions

- (X_t) adapted to $\mathcal{F}$ and $(X_t^2)_{t\geq 1}$ uniformly integrable
- $\lim_t \, t^3 \, \mathrm{E}\big\{\big(\mathrm{E}[Z_{t+1} \mid \mathcal{F}_t] - Z_t\big)^2\big\} = 0$
- $\mathrm{E}\big[\sup_{t\geq 1} \sqrt{t}\,|Z_{t-1} - Z_t|\,\big] < +\infty$
- $\frac{1}{t}\sum_{i=1}^{t}\big\{X_i - Z_{i-1} + i(Z_{i-1} - Z_i)\big\}^2 \xrightarrow{P} U$
- $t\sum_{i\geq t}(Z_{i-1} - Z_i)^2 \xrightarrow{a.s.} V$