Exchangeable random partitions for statistical and economic modelling

Antonio Lijoi[1], Pietro Muliere[2], Igor Prünster[3] and Filippo Taddei[4]

[1] Dipartimento Economia Politica e Metodi Quantitatavi, Università degli Studi di Pavia, via San Felice 5, 27100 Pavia, and CNR-IMATI, Milano, Italy. *E-mail*: lijoi@unipv.it.

[2] Dipartimento di Scienze delle Decisioni, Università Bocconi, Via Roentgen 1, 20136 Milano, Italy. *E-mail*: pietro.muliere@unibocconi.it

[3] Dipartimento di Statistica e Matematica Applicata, ICER and Collegio Carlo Alberto, Università degli Studi di Torino, Corso Unione Sovietica 218/bis, Torino, Italy. *E-mail*: igor@econ.unito.it

[3] Collegio Carlo Alberto, via Real Collegio 30, 10024 Moncalieri, Italy. *E-mail*: filippo.taddei@carloalberto.org

1 Introduction

Random partitions occur in a variety of probabilistic and statistical problems. In applied probability, for example, they define models for population genetics, species sampling and processes of coagulation and fragmentation. See Pitman (2006) and references therein. They also represent a key ingredient for various inferential methods arising in Bayesian nonparametric statistics, machine learning and for Markov Chain Monte Carlo algorithms that are used for clustering and density estimation. See the monograph edited by Hjort et al. (2010) for a comprehensive review. Recently random partitions have been exploited also in macroeconomic modelling as a tool for describing the clustering dynamics of economic agents according to their decision strategies. This approach has been introduced by M. Aoki in a series of papers and is effectively summarized in Aoki and Yoshikawa (2007). The illustrations we focus on in the

present review concern Bayesian nonparametric inference and macroeconomic modelling.

To sum up the main theoretical framework, suppose $X^{(\infty)} = (X_n)_{n \geq 1}$ is a sequence of observations or types of economic agents, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with each X_i taking values in a complete and separable metric space $\mathbb{X}$ endowed with the Borel σ–algebra $\mathscr{X}$. It will be assumed that $X^{(\infty)}$ is *exchangeable* which implies that for any $n \geq 1$ and any permutation π of the indices $1, \ldots, n$, the probability distribution (p.d.) of the random vector $(X_1, \ldots, X_n)$ coincides with the p.d. of $(X_{\pi(1)}, \ldots, X_{\pi(n)})$. An important characterization is provided by the celebrated de Finetti's representation theorem: it states that a sequence $X^{(\infty)}$ is exchangeable if and only if there exists a probability measure Q on the space $\mathcal{P}_{\mathbb{X}}$ of all probability measures on $\mathbb{X}$ such that, for any $n \geq 1$ and $A = A_1 \times \cdots \times A_n \times \mathbb{X}^{\infty}$, one has

$$\mathbb{P}\left[X^{(\infty)} \in A\right] = \int_{\mathcal{P}_{\mathbb{X}}} \prod_{i=1}^{n} p(A_i) \, Q(\mathrm{d}p) \tag{1}$$

where $A_i \in \mathscr{X}$ for any $i = 1, \ldots, n$ and $\mathbb{X}^{\infty} = \mathbb{X} \times \mathbb{X} \times \cdots$. The probability Q is also termed the *de Finetti measure* of the sequence $X^{(\infty)}$. This particular form of the representation theorem can be found in de Finetti (1937), an article that contains a series of lectures delivered by de Finetti in Paris, at the Institut Henri Poincaré, in 1935.

The implication of (1) is apparent: conditional on a random probability measure $\tilde{p}$ from Q, the first n elements of the exchangeable sequence $X^{(\infty)}$ are independent and identically distributed and their common p.d. is $\tilde{p}$. When Q is concentrated on a set of elements in $\mathscr{P}_{\mathbb{X}}$ that are discrete, one can show that there might be ties among $X_1, \ldots, X_n$, *i.e.* $\mathbb{P}[X_i = X_j] > 0$ for $i \neq j$. Correspondingly, define Ψ_n to be a random partition of the integers $\{1, \ldots, n\}$ such that any two integers i and j belong to the same set in Ψ_n if and only if $X_i = X_j$. Let $k \in \{1, \ldots, n\}$ and suppose $\{C_1, \ldots, C_k\}$ is a partition of $\{1, \ldots, n\}$ into k sets C_i. Hence, $\{C_1, \ldots, C_k\}$ is a possible realization of Ψ_n. A common and sensible specification for the probability distribution of Ψ_n consists in assuming that it depends on the frequencies of each set in the partition. To

illustrate this point, introduce the set

$$\Delta_{n,k} := \left\{ (n_1, \ldots, n_k) : \ n_i \geq 1, \ \sum_{i=1}^{k} n_i = n \right\}.$$

Set $n_i = \mathrm{card}(C_i)$, then $(n_1, \ldots, n_k) \in \Delta_{n,k}$ and

$$\mathbb{P}[\Psi_n = \{C_1, \ldots, C_k\}] = \Pi_k^{(n)}(n_1, \ldots, n_k) \tag{2}$$

A useful and intuitive metaphor is that of species sampling: one is not much interested into the realizations of the X_i's, which stand as species labels thus being arbitrary, but rather in the probability of observing k distinct species with frequencies $(n_1, \ldots, n_k)$ in $n \geq k$ draws from a population. This leads us to state the following

Definition 1. *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence. Then, $\{\Pi_k^{(n)} : 1 \leq k \leq n, \ n \geq 1\}$ with $\Pi_k^{(n)}$ defined in (2) is termed exchangeable partition probability function (EPPF).*

Indeed, the EPPF defines an important tool which has been introduced in Pitman (1995) and it determines the distribution of a random partition of $\mathbb{N}$. From the above definition it follows that, for any $n \geq k \geq 1$ and any $(n_1, \ldots, n_k) \in \Delta_{n,k}$, $\Pi_k^{(n)}$ is a symmetric function of its arguments, namely

$$\Pi_k^{(n)}(n_1, \ldots, n_k) = \Pi_k^{(n)}(n_{\pi(1)}, \ldots, n_{\pi(k)})$$

for any permutation π of $(1, \ldots, k)$, and it satisfies the consistency property

$$\Pi_k^{(n)}(n_1, \ldots, n_k) = \Pi_{k+1}^{(n+1)}(n_1, \ldots, n_k, 1)$$
$$+ \sum_{j=1}^{k} \Pi_k^{(n+1)}(n_1, \ldots, n_j + 1, \ldots, n_k). \tag{3}$$

corresponding to the fact that the partition of $X_1, \ldots, X_n$ can be recovered from the partition of $X_1, \ldots, X_{n+1}$ by dropping X_{n+1}. On the other hand, as shown in Pitman (1995), every non–negative symmetric function satisfying (3) is the EPPF of some exchangeable sequence. See Pitman (1995, 2006) for a thorough and useful analysis of EPPFs.

In the following sections we will emphasize the role played by EPPFs in two different contexts: Bayesian nonparametric inference and economic modelling.

2 Bayesian nonparametric modelling

Bayesian nonparametric inference is a relatively young area of research that has recently undergone a strong development. Most of its success can be explained by the considerable degree of flexibility it ensures in statistical modelling, if compared to parametric alternatives, and by the emergence of new and efficient simulation techniques that make nonparametric models amenable to concrete use in a number of applied statistical problems.

2.1 The Dirichlet process

Even if the formal setting for Bayesian nonparametric inference had been laid out by de Finetti during the 30's, no tractable examples were given on how to construct a prior Q on $\mathscr{P}_\mathbb{X}$ so to make the nonparametric approach feasible in applied statistical problems. Such a task has been completed only 40 years later by T.S. Ferguson who defined in his 1973 paper, on the Annals of Statistics, the Dirichlet process. Nowadays it represents one of the most commonly used priors in Bayesian nonparametrics and its popularity can be explained by its mathematical tractability and by the recent development of Markov chain Monte Carlo (MCMC) techniques whose implementation allows a full Bayesian analysis of complex statistical models based on the Dirichlet process prior.

As highlighted in Ferguson (1973), the Dirichlet process can be defined in various ways. Here we will point out three different definitions in terms of the family of its finite-dimensional distributions, by means of a series representation and through its representation as a normalized completely random measure.

Definition 2. (Ferguson, 1973). *Let $c > 0$ be a constant and P_0 some probability measure on $(\mathbb{X}, \mathscr{X})$. Moreover, to any measurable partition $\{A_1, \ldots, A_{k+1}\}$ of $\mathbb{X}$, associate the vector $(p_1, \ldots, p_{k+1})$ with $p_i = P_0(A_i)$. Then, we say that $\tilde{p}$ is a Dirichlet process with parameter measure cP_0 if the vector $(\tilde{p}(A_1), \ldots, \tilde{p}(A_k))$ has density*

$$\frac{\Gamma(c)}{\prod_{i=1}^{k+1} \Gamma(cp_i)} \prod_{i=1}^{k} x_i^{cp_i - 1} (1 - x_1 - \cdots - x_k)^{cp_{k+1} - 1} \mathbb{1}_{\mathcal{S}_k}(x_1, \ldots, x_k)$$

where $\mathcal{S}_k := \{(x_1, \ldots, x_k) : x_i \geq 0, \ \sum_{i=1}^{k} x_i \leq 1\}$ is the k–dimensional simplex.

An alternative characterization of the Dirichlet process can be given in terms of a normalized completely random measure $\tilde{\mu}$. This is a random element defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in the space of boundedly finite measures $\mathbb{M}_{\mathbb{X}}$ with the property that if A and B are two sets in $\mathscr{X}$ such that $A \cap B = \varnothing$, then the random variables $\tilde{\mu}(A)$ and $\tilde{\mu}(B)$ are independent. See the Appendix for a short review on completely random measures.

Theorem 1. (Ferguson, 1973). *Suppose $\tilde{\mu}$ is a gamma completely random measure with parameter cP_0, namely for any set $A \in \mathscr{X}$*

$$\mathbb{P}[\tilde{\mu}(A) \leq x]$$
$$= \left\{ \frac{\mathbb{1}_{(0,\infty)}(P_0(A))}{\Gamma(cP_0(A))} \int_0^x s^{cP_0(A)-1} \, \mathrm{e}^{-s} \, \mathrm{d}s + \mathbb{1}_{\{0\}}(P_0(A)) \right\} \mathbb{1}_{[0,\infty)}(x).$$

If $\tilde{p}$ is a Dirichlet process with parameter measure cP_0, then

$$\frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})} \overset{d}{=} \tilde{p} \tag{4}$$

It is worth noting that $\tilde{\mu}$ is a jump process. If P_0 is non–atomic, this entails that $\tilde{\mu} = \sum_{i \geq 1} J_i \delta_{X_i}$ where the J_i's are independent and non–negative random variables and the X_i's are i.i.d from P_0. This suggests an important feature of the Dirichlet process that was first shown by Blackwell (1973): the Dirichlet process selects, almost surely, discrete probability distributions on $(\mathbb{X}, \mathscr{X})$. This property becomes even more apparent if one considers an alternative definition of the Dirichlet process that can be given in terms of the so–called stick–breaking construction.

Theorem 2. (Sethuraman, 1994). *Let $(V_i)_{i \geq 1}$ be a sequence of i.i.d. random variables with $V_i \sim Beta(1, c)$ and define*

$$w_1 = V_1, \qquad w_k = V_k \prod_{i=1}^{k-1} (1 - V_i) \quad k = 2, 3, \ldots \tag{5}$$

Then $\sum_{k \geq 1} w_k = 1$, almost surely, and if $(X_n)_{n \geq 1}$ is a sequence of i.i.d. random variables whose common p.d. P_0 is non–atomic, the random

probability measure

$$\tilde{p} = \sum_{k \geq 1} w_k \, \delta_{X_k} \tag{6}$$

coincides, in distribution, with the Dirichlet process with parameter measure cP_0.

The representation in (6) highlights an interpretation of the Dirichlet process $\tilde{p}$ as a species sampling model. This entails that the population can be thought of as split into an infinite number of species, w_k being the unknown proportion of the k–th species. As pointed out in the introduction, the discreteness of $\tilde{p}$ naturally leads one to analyze the partition structure it induces on a set $X_1, \ldots, X_n$ of the first n observations extracted from an infinite exchangeable sequence $(X_n)_{n \geq 1}$. The EPPF associated to the Dirichlet process is

$$\Pi_k^{(n)}(n_1, \ldots, n_k) = \frac{c^k}{(c)_n} \prod_{i=1}^{k} (n_i - 1)! \tag{7}$$

where $(c)_n = \Gamma(c + n)/\Gamma(c)$ is the n–th ascending factorial of c. See Antoniak (1974). It is worth noting that (7) is an equivalent form of the well-known Ewens sampling formula widely used in population genetics. Indeed, the formula introduced by Ewens (1972) represents the p.d. of the vector $(m_1, \ldots, m_n)$ of counts, where m_i is the number of clusters of size i, and it is given by

$$\Pi_{k,n}^*(m_1, \ldots, m_n) = \frac{n! c^k}{(c)_n} \prod_{i=1}^{n} \frac{1}{i^{m_i} m_i!}$$

for any vector of non–negative integers $(m_1, \ldots, m_n)$ such that $\sum_{i=1}^{n} m_i = k$ and $\sum_{i=1}^{k} i m_i = n$. When the EPPF is known, the determination of the corresponding predictive distribution is straightforward. Indeed, if one adheres to the species sampling interpretation for $\tilde{p}$, the probability of observing a new species, conditional on a sample $X_1, \ldots, X_n$ featuring k distinct species $X_1^*, \ldots, X_k^*$ with frequencies $n_1, \ldots, n_k$, is

$$\mathbb{P}[X_{n+1} = \text{new} \,|\, X_1, \ldots, X_n] = \frac{\Pi_{k+1}^{(n+1)}(n_1, \ldots, n_k, 1)}{\Pi_k^{(n)}(n_1, \ldots, n_k)} = \frac{c}{c + n}$$

On the other hand, the probability that X_{n+1} is from any of the species observed in the conditioning sample is

$$\mathbb{P}[X_{n+1} = \text{new} \,|\, X_1, \ldots, X_n]$$
$$= \frac{\sum_{j=1}^{k} \Pi_k^{(n+1)}(n_1, \ldots, n_j + 1, \ldots, n_k)}{\Pi_k^{(n)}(n_1, \ldots, n_k)} = \frac{n}{c+n}$$

for any $j = 1, \ldots, n$. These can be summarized in the following expression, known as predictive distribution,

$$\mathbb{P}[X_{n+1} \in A \,|\, X_1, \ldots, X_n] = \frac{c}{c+n} \, P_0(A) + \frac{n}{c+n} \, \hat{P}_n(A) \qquad \forall A \in \mathscr{X} \tag{8}$$

where $\hat{P}_n = \sum_{j=1}^{n} \delta_{X_i}/n$ is the empirical distribution.

Besides prediction, the EPPF is also a useful tool for studying distributional properties of the number K_n of clusters generated by an exchangeable sample of size n. If one marignalizes with respect to the frequencies $(n_1, \ldots, n_k)$ in (7), one obtains the p.d. of K_n

$$\mathbb{P}[K_n = k] = \frac{c^k}{(c)_n} \, |s(n,k)| \, \mathbb{1}_{\{1,\ldots,n\}}(k) \tag{9}$$

where $|s(n,k)|$ is the signless Stirling number of the first kind. Moreover, the asymptotic behaviour is readily available from results in Korwar and Hollander (1973), which state that

$$\frac{K_n}{\log n} \xrightarrow{\text{a.s}} c \tag{10}$$

as $n \uparrow \infty$. The rate of increase of K_n, as n increases, is an important quantity for assessing the implications of the use of the Dirichlet process in macroeconomic modelling and it will be compared to the behaviour associated to random probability measures that generalize the Dirichlet process.

2.2 The two–parameter Poisson–Dirichlet process

Despite the Dirichlet process has been a cornerstone in Bayesian non-parametrics, in some cases of interest for statistical applications it is not an adequate prior choice and alternative nonparametric models need to

be devised. An example is represented by survival analysis: if a Dirichlet prior is used for the survival time distribution, then the posterior, conditional on a sample containing censored observations, is not Dirichlet. It is, then, of interest to find an appropriate class of random distributions which contain, as a special case, the posterior distribution of the Dirichlet process given censored observations. Moreover, in survival problems one might be interested in modelling hazard rate functions or cumulative hazards and the Dirichlet process cannot be used in these situations. Also in clustering and prediction problems, which are of interest to the present paper, the predictive structure (8) induced by the Dirichlet process is sometimes not flexible enough to capture important aspects featured by the data. Indeed, the probabilities of generating a new observations and of re-observing one of the species that have appeared in the conditioning sample, $c/(c+n)$ and $n/(c+n)$, respectively, depend neither on the number k of clusters into which the data are grouped nor on the individual frequencies $n_1, \ldots, n_k$. An important piece of information for prediction is, then, neglected. This, and allied applied problems, have recently stimulated a number of contributions aiming at the definition of generalizations of the Dirichlet process that still preserve a reasonable amount of analytical tractability and that overcome some of the drawbacks inherent to modelling real phenomena with the Dirichlet process. Among these generalizations, a special role is played by the two–parameter Poisson–Dirichlet process introduced by Pitman (1995).

Definition 3. *Let (α, θ) be parameters such that either $\alpha \in [0, 1]$ and $\theta > -\alpha$ or $\alpha = -x < 0$ and $\theta = mx$ for some $m = 1, 2, \ldots$. Moreover, $(V_i)_{i \geq 1}$ is a sequence of independent random variables with $V_i \sim Beta(1 - \alpha, \theta + i\alpha)$ and $(w_i)_{i \geq 1}$ are random weights defined as*

$$w_1 = V_1, \qquad w_i = V_i \prod_{j=1}^{i-1}(1 - V_j) \quad i \geq 2.$$

If $(X_i)_{i \geq 1}$ is a sequence of i.i.d. random variables with non–atomic p.d. P_0, the random probability measure $\sum_{i \geq 1} w_i \, \delta_{X_i}$ is a two-parameter Poisson–Dirichlet process.

This definition points out an analogy to the Dirichlet process, namely

the realizations of a two–parameter Poisson–Dirichlet process are almost surely discrete. Note also that the Dirichlet process stands as a particular case for which $\alpha = 0$. Another useful definition can be given in terms of completely random measures as pointed out in Pitman and Yor (1997) for the case where $\alpha \in (0, 1)$. Let $\tilde{\mu}_\alpha$ be a α–stable completely random measure with parameter measure P_0. This means that

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \tilde{\mu}_\alpha(A)}\right] = \int_{\mathbb{M}_{\mathbb{X}}} \mathrm{e}^{-\lambda \mu(A)}\, \mathbb{P}_\alpha(\mathrm{d}\mu) = \mathrm{e}^{-P_0(A)\lambda^\alpha}$$

Let $\tilde{\mu}_{\alpha,\theta}$ be a random measure on $\mathbb{M}_{\mathbb{X}}$ with law $\mathbb{P}_{\alpha,\theta}$ such that $\mathbb{P}_{\alpha,\theta}$ is absolutely continuous with respect to $\mathbb{P}_\alpha$ and

$$\frac{\mathrm{d}\mathbb{P}_{\alpha,\theta}}{\mathrm{d}\mathbb{P}_\alpha}(\mu) = \{\mu(\mathbb{X})\}^{-\theta}.$$

Theorem 3. (Pitman and Yor, 1997). *The normalized random measure $\tilde{\mu}_{\alpha,\theta}/\tilde{\mu}_{\alpha,\theta}(\mathbb{X})$ coincides in distribution with a two–parameter Poisson–Dirichlet process.*

Among all generalizations of the Dirichlet process, the $\mathrm{PD}(\alpha, \theta)$ process stands out for its tractability. The EPPF, which characterizes the induced random partition, of a $\mathrm{PD}(\alpha, \theta)$ process is

$$\Pi_k^{(n)}(n_1, \ldots, n_k) = \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta + 1)_{n-1}} \prod_{j=1}^{k}(1-\alpha)_{n_j-1} \tag{11}$$

Now, denote by $m_j \geq 0$, $j = 1, \ldots, n$, the number of sets in the partition which contain j objects or, using again the species metaphor, the number of species appearing j–times in a sample of size n. Then, an alternative equivalent formulation of (11), known as *Pitman's sampling formula*, is given by

$$\Pi_{k,n}^{*}(m_1, \ldots, m_n) = n! \frac{\prod_{i=1}^{k-1}(\theta + i\alpha)}{(\theta + 1)_{n-1} \prod_{i=1}^{n} m_i!} \prod_{i=1}^{n}\left[\frac{(1-\alpha)_{i-1}}{i!}\right]^{m_i}$$

for any $n \geq 1$ and $m_1, \ldots, m_n$ such that $m_i \geq 0$, $\sum_{i=1}^{n} i\, m_i = n$ and $\sum_{i=1}^{n} m_i = k$. The above expression represents a two parameter generalization of the Ewens' sampling formula that can be recovered by

letting $\alpha \to 0$. The availability of the EPPF in (11) allows one to determine the system of predictive distributions associated with the $\mathrm{PD}(\alpha, \theta)$ process. Indeed, if $X_1, \ldots, X_n$ is a sample consisting of k distinct values $X_1^*, \ldots, X_k^*$ and n_j of them are equal to X_j^*, then

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x \mid X_1, \ldots, X_n] = \frac{\theta + k\alpha}{\theta + n} \, P_0(\mathrm{d}x) + \frac{1}{\theta + n} \sum_{j=1}^{k} (n_j - \alpha) \, \delta_{X_j^*}(\mathrm{d}x)$$

It can be noted that, unlike the Dirichlet process, the probability of observing a new species depends also on the number k of distinct observations. This is not the only remarkable difference from the Dirichlet process. Another important distinctive feature concerns the asymptotic behaviour of the number of distinct observations K_n detected in a sample of size n. For any n one has that

$$\mathbb{P}[K_n = k] = \frac{\prod_{i=1}^{k-1} (\theta + i\alpha)}{\alpha^k \, (\theta + 1)_{n-1}} \, \mathscr{C}(n, k; \alpha) \qquad k = 1, \ldots, n,$$

where

$$\mathscr{C}(n, k; \alpha) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\alpha)_n$$

is the generalized factorial coefficient. See Gnedin and Pitman (2005). In order to derive the asymptotic behaviour of K_n as n diverges, it is useful to first introduce a class of random variables, which will appear throughout the following developments. This class of random variables, which we term generalized Mittag–Leffler random variables, is defined as follows. Let f_α be the density function of a positive α–stable random variable and define Z_q to be, for any $q \geq 0$, a positive random variable with density function

$$f_{Z_q}(z) = \frac{\Gamma(q\alpha + 1)}{\alpha \, \Gamma(q + 1)} \, z^{q-1-1/\alpha} \, f_\alpha \left(z^{-1/\alpha} \right). \tag{12}$$

Then, by Theorem 3.8 in Pitman (2006), one has that

$$\frac{K_n}{n} \xrightarrow{\text{a.s.}} Z_{\theta/\alpha}. \tag{13}$$

Therefore, in the two–parameter case, one has that K_n increases at a rate of n^α (rather than the logarithmic rate of the Dirichlet process) and,

moreover, the normalized version of K_n converges to a strictly positive random variable (in contrast to the convergence to a constant of the Dirichlet case).

3 Uncertainty in macroeconomics models

3.1 Self–averaging phenomena

Aoki (2008) introduces the interesting concept of "self-averaging" in Economics in relation to stochastic growth models.

Definition 4. *A size–dependent random variable X_n is termed self–averaging if*

$$C.V.(X_n) = \frac{\sqrt{Var(X_n)}}{E(X_n)} \to 0 \quad as\ n \to \infty, \tag{14}$$

where C.V. *clearly denotes the coefficient of variation.*

Such a property typically holds for simple economic models, where some assumption of symmetry or homogeneity of the individuals underlies the whole model. The concept is best clarified by looking at an example: consider the popular Poisson model, in which for each "individual" an event (e.g. technical progress) occurs according to a Poisson process with parameter λ. Then, in the whole economy, which is based on n individuals, the number of events X_n follows a Poisson process with rate λn. Consequently, in a one–time period, we have $E(X_n) = \text{Var}(X_n) = \lambda n$ and it immediately follows that C.V.$(X_n) = \lambda n^{-1/2} \to 0$ as $n \to \infty$. Hence, the Poisson model is self–averaging. The same obviously holds for the Gaussian case.

In fact, the self-averaging condition (14) can be equivalently expressed as

$$\text{Var}\left(\frac{X_n}{E(X_n)}\right) \to 0 \qquad n \to \infty, \tag{15}$$

from which it becomes evident that for self–averaging macroeconomic phenomena, one can focus attention on the means of the involved variables since for sufficiently large n the residual variability of the normal-

ized X_n becomes negligible. On the other hand, the model is *non–self–averaging* if

$$\mathrm{Var}\left(\frac{X_n}{E(X_n)}\right) \not\to 0 \qquad n \to \infty.$$

In such a case, even if the number of agents diverges, the uncertainty about the "normalized" trajectories of X_n persists: clearly focusing solely on the mean behaviour is not enough for describing the phenomenon at issue. Some measure of the oscillations around the mean is essential for providing a clear picture.

In the following we introduce a simple endogenous growth model and show that it leads, under reasonable assumptions, to non–self–averaging phenomena. The model represents a rigorous development of some ideas presented in Aoki and Yoshikawa (2007). By deriving exact asymptotic results we show how the mean can be combined with measures of uncertainty represented by highest posterior density intervals.

3.2   A simple endogenous growth model

In this section we review the results of Lijoi, Muliere, Prünster and Taddei (2010). In accordance with the literature on endogenous growth, we assume that the economy grows by innovations, which are stochastic events of two types: the first type is represented by a productivity rise in an existing sector, whereas the second type is represented by the creation of a new sector. By the time the n–th innovation occurs, the economy will consist of a random number K_n of sectors, the i–th sector will have experienced n_i innovations and obviously $\sum_{i=1}^{K_n} n_i = n$.

Furthermore, we assume that the output of sector i is of the form

$$Y_i = \eta\, \gamma^{\frac{n_i}{n^{1-\sigma}}} \tag{16}$$

where $\gamma > 1$, $\eta > 0$ and $\sigma \in (0,1)$. Without loss of generality we can assume $\eta = 1$. Moreover, we will concentrate our attention on the case of γ close to 1, which is realistic in many situations. Therefore we can approximate (16) with

$$Y_i = 1 + \beta\, \frac{n_i}{n^{1-\sigma}}$$

where $\beta = \log(\gamma) > 0$. Hence, the aggregate output of the economy, which is the sum of the outputs of the K_n sectors, is given by

$$X_n = \sum_{i=1}^{n} Y_i = K_n + \beta\, n^\sigma \qquad (17)$$

which shows that K_n is the contribution to the aggregate output of the number of sectors and that n^σ is the contribution of the innovations within sectors.

Finally, the stochastic innovations are governed by a two–parameter Poisson–Dirichlet model with parameters $\alpha \in (0,1)$ and $\theta > 0$. This means that, given an economy with K_n sectors and the n innovations distributed as $(n_1, \ldots, n_{K_n})$, the probability that innovation $n+1$ will create a new sector is

$$\frac{\theta + \alpha\, K_n}{\theta + n},$$

whereas the probability that the $n+1$–th innovation will happen in sector i is

$$\frac{n_i - \alpha}{\theta + n} \qquad i = 1, \ldots, K_n.$$

Before proceeding it is worth to outline the reinforcement mechanism induced by the Poisson–Dirichlet process. The probability of having an innovation in one of the already existing sectors is $(n - \alpha K_n)/(\theta + n)$, but the mass is not allocated proportional to the number of innovations already observed in each sector. The probability of observing an innovation in sector i is determined by the size n_i of the cluster and by α. In fact, a reinforcement mechanism driven by α takes place. Indeed, one can see that the ratio of the probabilities assigned to any pair of sectors (i, j) is given by $(n_i - \alpha)/(n_j - \alpha)$. As $\alpha \to 0$, the previous quantity reduces to the ratio of the sizes of the two clusters, which characterizes the Dirichlet case and represents exactly the case of homogeneity of the sectors. If $n_i > n_j$, the ratio is an increasing function of α. Hence, as α increases the mass is reallocated from sector j to i. This means that the dynamics tends to reinforce, among the observed clusters, those having higher frequencies. Table 1 provides an idea of the magnitude of the reinforcement. See Lijoi et al. (2005, 2007) for details and more discussion on the reinforcement connected to such models.

Table 1: Ratio of the probabilities allocated to sector i observed n_i times and sector j observed only once for different choices of α.

	$n_i = 2$	$n_i = 10$	$n_i = 50$	$n_i = 100$
Dir	2	10	50	100
PD$(\theta, \alpha = 0.25)$	2.33	13	66.33	133
PD$(\theta, \alpha = 0.50)$	3	19	99	199
PD$(\theta, \alpha = 0.75)$	5	37	197	397
PD$(\theta, \alpha \to 1)$	$\to \infty$	$\to \infty$	$\to \infty$	$\to \infty$

The following result distinguishes various cases corresponding to different choices of the parameters of the model: it is worth noting that $\alpha > \sigma$ ($\alpha < \sigma$) means that the contribution to aggregate output from innovations represented by introduction of new sectors are more (less) relevant than those within an existing sector. Hence, our result essentially states that, when contributions to the economy given by the introduction of new sectors are at least as relevant as those given by the existing sectors, the economy presents a non–self–averaging behaviour.

Proposition 1. *Under the growth model* (17) *with innovations following a two parameter Poisson Dirichlet process, we have*

$$\mathbb{E}[X_n] = \frac{(\theta + \alpha)_n}{\alpha (\theta + 1)_{n-1}} - \frac{\theta}{\alpha} + \beta n^\sigma, \qquad (18)$$

where $(a)_n = a(a+1)\ldots(a+n-1)$ *is the ascending factorial. Moreover,*

(i) *If* $\alpha = \sigma = \upsilon$,

$$\frac{X_n}{n^\upsilon} \to Z_{\theta/\alpha} + \beta \quad a.s.$$

where Z_q *is a generalized Mittag–Leffler random variable defined in* (12), *and* X_n *is non–self-averaging.*

(ii) *If* $\alpha = \upsilon > \sigma$,

$$\frac{X_n}{n^\upsilon} \to Z_{\theta/\alpha} \quad a.s.$$

where Z_q *is a generalized Mittag–Leffler random variable defined in* (12), *and* X_n *is non–self-averaging.*

(iii) If $\sigma = \upsilon > \alpha$,

$$\frac{X_n}{n^\upsilon} \to \beta \quad a.s.$$

and X_n is self-averaging.

Proof. The proof follows by combining formula (3.13) in Pitman (2006), the asymptotics of K_n as recalled in (13) and standard limiting arguments. $\qquad\square$

From Proposition 1 it follows that the model can be described by the mean $\mathbb{E}[X_n]$, given in (18), in self–averaging situations: this is in agreement with the usual macroeconomic attitude to consider aggregate average quantities. The question is, what one should do in non–self–averaging cases, which as shown in Proposition 1 arise systematically in presence in of highly dynamic economies. In such cases, it seems essential to combine the study of the mean beahviour with a measure of uncertainty and the natural tool in this framework is represented by the asymptotic highest posterior density (HPD) intervals of the limiting random variable, which represent the Bayesian counterpart to frequentist confidence intervals.

As for the determination of the asymptotic HPD intervals, consider case (ii), case (i) follows then immediately: one can take the 95% HPD interval (z_1, z_2) of $Z_{\theta/\alpha}$ i.e. (z_1, z_2) such that $z_2 - z_1$ is minimal under the condition $\mathbb{P}(z_1 < Z_{\theta/\alpha} < z_2) \geq 0.95$. The asymptotic HPD interval for X_n is then given by $(z_1 n^\upsilon, z_2 n^\upsilon)$.

However, the determination of the quantiles of a generalized Mittag–Leffler random variable Z_q is cumbersome and, hence, we devise a simulation algorithm for generating values of Z_q by adapting arguments in Favaro et al. (2009) and one can then use the output to evaluate quantiles. The basic idea consists in setting $W_q = Z_q^{-1/\alpha}$ so that W_q has density function given by

$$f(w) = \frac{\alpha \Gamma(q\alpha)}{\Gamma(q)} w^{-q\alpha} f_\alpha(w) = \frac{\alpha}{\Gamma(q)} f_\alpha(w) \int_0^\infty u^{q\alpha-1} e^{-uw} \, du$$

Via augmentation, one then has

$$f(u,w) = \frac{\alpha}{\Gamma(q)} f_\alpha(w) u^{q\alpha-1} e^{-uw} = f(u) f_\alpha(w|u)$$

where $f(u)$ is the density function of a r.v. U_q such that $U_q^\alpha \sim \text{Gamma}(q, 1)$, and

$$f_\alpha(w|u) = f_\alpha(w)\, e^{-uw+u^\alpha}.$$

This means that, conditional on U_q, W_q is a positive tempered–stable random variable, according to the terminology adopted in Rosiński (2007). In order to draw samples from it, a convenient strategy is to resort to the series representation derived in Rosiński (2007), which, in our case, yields

$$W_q|U_q \overset{d}{=} \sum_{i=1}^\infty \min\left\{ (a_i \Gamma(1-\alpha))^{-1/\alpha}, e_i\, v_i^{1/\alpha} \right\} \tag{19}$$

where $e_i \overset{\text{iid}}{\sim} \text{Exp}(U_q)$, $v_i \overset{\text{iid}}{\sim} \text{U}(0,1)$ and $a_1 > a_2 > \cdots$ are the arrival times of a Poisson process with unit intensity. Other possibilities for simulating from a tempered stable random variable are the inverse Lévy measure method as described in Ferguson and Klass (1972) and a compound Poisson approximation scheme proposed in Cont and Tankov (2004). Summarizing the above considerations, an algorithm for simulating from the limiting random variable $Z_{\theta/\alpha}$ is as follows:

1. generate $X \sim \text{Ga}(\theta/\alpha, 1)$ and set $U = X^{1/\alpha}$;

2. for a given truncation N and U sampled in step 1., generate: $\{e_i\} \overset{\text{iid}}{\sim} \text{Exp}(U)$, $\{v_i\} \overset{\text{iid}}{\sim} \text{U}(0,1)$, $\xi_j \overset{\text{iid}}{\sim} \text{Exp}(1)$ and take $a_i = \sum_{j=1}^i \xi_j$, for $i = 1, \ldots, N$;

3. compute W according to (19) and set $Z = W^{-\alpha}$.

Recently two alternative and more efficient algorithms for drawing samples from the limiting distribution of K_n/n^α has been derived by Montagna (2009), who exploits results of Devroye (2009).

With such algorithm at hand, it is straightforward to describe the growth model via $\mathbb{E}(X_n)$ combined with the corresponding HPD intervals, which account for the persisting uncertainty due to the non–self–averaging nature of the phenomenon at issue.

The previous model can be seen as an unconditional model, where the economy starts from scratch. A more realistic model, would consider the status quo of the economy and analyze the contribution to the aggregate output of sectors which will emerge only in the future. From a

mathematical point of view this means predicting the future behaviour conditionally on a given state of the world. Therefore, we now assume the status quo as given (i.e. there are at present $K_n = k$ sectors where $n_1, \ldots, n_k$ innovations occurred) and study the aggregate output of new future sectors. By the time the m–th innovation occurs, there will be a random number $K_m^{(n)} = K_m - K_n$ of new sectors in the economy, where the i–th will have experienced s_i innovations. In this model, not all innovations will belong to the new sectors: in fact, $\sum_{i=1}^{K_m^{(n)}} s_i = L_m^{(n)}$ represents the number of innovations concerning the new sectors and $m - L_m^{(n)}$ innovations will concern the "old" sectors. Under the same assumptions of the unconditional case, the output of the i–th new sector is then of the form

$$(Y_i | K_n = j, n_1 \ldots, n_j) = 1 + \beta \, \frac{s_i}{m^{1-\sigma}} \qquad i = 1, \ldots, K_m^{(n)}$$

where $\beta > 0$ and $\sigma \in (0, 1)$. The aggregate output of the $K_m^{(n)}$ new sectors is then given by

$$(X_m | K_n = j, n_1 \ldots, n_j) = K_m^{(n)} + \beta \, \frac{L_m^{(n)}}{m^{1-\sigma}} \qquad (20)$$

Again, we model the stochastic innovations with a two–parameter Poisson–Dirichlet model with parameters $\alpha \in (0, 1)$ and $\theta > 0$. The following result provides a complete description of the model showing that non–self–averaging appears under any assumption on the innovation parameters α and σ highlighting how common such phenomena arise.

Proposition 2. *Under the growth model* (20) *with innovations following a two parameter Poisson Dirichlet process, we have*

$$\mathbb{E}[X_m | K_n = j, n_1 \ldots, n_j] = \left(j + \frac{\theta}{\alpha} \right) \left\{ \frac{(\theta + n + \alpha)_m}{(\theta + n)_m} - 1 \right\} + \beta \frac{\theta + j\alpha}{\theta + n} m^\sigma.$$
$$(21)$$

Moreover:

(i) *If* $\alpha = \sigma = \upsilon$,

$$\frac{(X_m | K_n = k, n_1 \ldots, n_k)}{m^\upsilon} \to U_{n,j} + \beta \, B_{\theta + \alpha j, n - \alpha j} \quad a.s.$$

where $U_{n,j} \stackrel{d}{=} B_{j+\theta/\alpha, n/\alpha-j} Z_{(\theta+n)/\alpha}$, Z_q is a generalized Mittag–Leffler random variable with density (12), $B_{a,b}$ is a beta random variable with parameters (a,b) and the random variables $B_{j+\theta/\alpha, n/\alpha-j}$ and $Z_{(\theta+n)/\alpha}$ are independent. Hence, the model is non–self-averaging.

(ii) If $\alpha = \upsilon > \sigma$,

$$\frac{(X_m | K_n = k, n_1 \ldots, n_k)}{m^\upsilon} \to U_{n,j} \quad a.s.$$

and the model is non–self-averaging.

(iii) If $\sigma = \upsilon > \alpha$,

$$\frac{(X_m | K_n = k, n_1 \ldots, n_k)}{m^\upsilon} \to \beta \, B_{\theta+\alpha j, n-\alpha j} \quad a.s..$$

PROOF. We start by considering the limiting behaviour of $K_m^{(n)}$, which is one of the two components the aggregate output (20) is made of. The proof strategy is as follows: we first mimick the arguments of Favaro et al. (2009) in order to establish that $K_m^{(n)}/m^\alpha$ converges a.s. and in the p–th mean for any $p > 0$, determine the moments of the limiting random variable and show that the limiting random variable is characterized by its moments. Then, the asymptotic behaviour of the second component of the aggregate output is studied and the two bits combined to achieved the desired result.

Let us start by computing the likelihood ratio

$$M_{\alpha,\theta,m}^{(n)} := \frac{\mathrm{d}P_{\alpha,\theta}^{(n)}}{\mathrm{d}P_{\alpha,0}^{(n)}} \bigg|_{\mathscr{F}_m^{(n)}} = \frac{q_{\alpha,\theta}^{(n)}(K_m^{(n)})}{q_{\alpha,0}^{(n)}(K_m^{(n)})}$$

where $\mathscr{F}_m^{(n)} = \sigma(X_{n+1}, \ldots, X_{n+m})$, $P_{\alpha,\theta}^{(n)}$ is the conditional probability distribution of a $\mathrm{PD}(\alpha, \theta)$ process given K_n and, by virtue Proposition 1 in Lijoi, Prünster and Walker (2008), $q_{\alpha,\theta}^{(n)}(k) = \alpha^{K_n}(\frac{\theta}{\alpha} + K_n)_k / (\theta + n)_m$ for any integer $k \geq 1$ and $q_{\alpha,\theta}^{(n)}(0) := 1/(\theta+n)_m$. Hence $(M_{\alpha,\theta,m}^{(n)}, \mathscr{F}_m^{(n)})_{m \geq 1}$ is a $P_{\alpha,0}^{(n)}$–martingale. By a martingale convergence theorem, $M_{\alpha,\theta,m}^{(n)}$ has a $P_{\alpha,0}^{(n)}$ almost sure limit, say $M_{\alpha,\theta}^{(n)}$, as $m \to \infty$. Convergence holds in the p–th mean as well, for any $p > 0$. One clearly has that $E_{\alpha,0}^{(n)}[M_{\alpha,\theta}^{(n)}] = 1$,

where $E_{\alpha,0}^{(n)}$ denotes the expected value w.r.t. $P_{\alpha,0}^{(n)}$. It can be easily seen that

$$M_{\alpha,\theta,m}^{(n)} \sim \frac{\Gamma(\theta+n)\Gamma(K_n)}{\Gamma(n)\Gamma\left(\frac{\theta}{\alpha}+K_n\right)} \left(\frac{K_m^{(n)}}{m^\alpha}\right)^{\theta/\alpha}$$

as $m \to \infty$. Hence $(K_m^{(n)}/m^\alpha)^{\theta/\alpha}$ converges $P_{\alpha,0}^{(n)}$–a.s. to a random variable, say $U_{n,j}$ such that

$$E_{\alpha,0}^{(n)}\left[U_{n,j}^{\theta/\alpha}\right] = \frac{\Gamma(n)\Gamma\left(\frac{\theta}{\alpha}+K_n\right)}{\Gamma(\theta+n)\Gamma(K_n)}.$$

In order to identify the distribution of the limiting random variable $U_{n,j}$ w.r.t. $P_{\alpha,\theta}^{(n)}$, we consider the asymptotic behaviour of $E[(K_m^{(n)})^r \mid K_n]$ as $m \to \infty$, for any $r \geq 1$. Hence, we first need to identify the moments $E[(K_m^{(n)})^r \mid K_n]$. Indeed, one has

$$E\left[(K_m^{(n)})^r \mid K_n = j,\, w\right] = \sum_{i=0}^m \binom{m}{i} w^i (1-w)^{m-i}\, E\left[K_i^r\right]$$

where the unconditional moment $E\left[K_i^r\right]$ is evaluated w.r.t. $\tilde{P}_{\alpha,\theta+j\alpha}$ prior. Such an expression is already available from Yamato and Sibuya (2000) and it is given by

$E\left[K_i^r\right]$
$$= \sum_{\nu=0}^r (-1)^{r-\nu} \left(1+\frac{\theta+j\alpha}{\alpha}\right)_\nu S\left(r,\nu;\frac{\theta+j\alpha}{\alpha}\right) \frac{(\theta+j\alpha+\nu\alpha+1)_{i-1}}{(\theta+1)_{i-1}}$$

where S is the non–central Stirling number of the second kind. Hence, one has

$E\left[(K_m^{(n)})^r \mid K_n = j\right]$
$$= \frac{\Gamma(\theta+n)\int_0^1 w^{\theta+j\alpha-1}(1-w)^{n-j\alpha-1} E\left[(K_m^{(n)})^r \mid K_n = j,\, w\right] \mathrm{d}w}{\Gamma(\theta+j\alpha)\Gamma(n-j\alpha)}$$

$$= \frac{\Gamma(\theta+n)\sum_{\nu=0}^r (-1)^{r-\nu}\left(1+\frac{\theta+j\alpha}{\alpha}\right)_\nu S\left(r,\nu;\frac{\theta+j\alpha}{\alpha}\right)}{\Gamma(\theta+j\alpha)\Gamma(n-j\alpha)}$$

$$\times \sum_{i=0}^m \binom{m}{i} \frac{(\theta+j\alpha+\nu\alpha+1)_{i-1}\int_0^1 w^{\theta+j\alpha+i-1}(1-w)^{n-j\alpha+m-i-1}\,\mathrm{d}w}{(\theta+1)_{i-1}}$$

$$
= \frac{1}{(\theta + n)_m} \sum_{\nu=0}^{r} (-1)^{r-\nu} \left(1 + \frac{\theta + j\alpha}{\alpha}\right)_\nu S\left(r, \nu; \frac{\theta + j\alpha}{\alpha}\right) \frac{\theta + j\alpha}{\theta + j\alpha + \nu\alpha}
$$

$$
\times \sum_{i=0}^{m} \binom{m}{i} (\theta + j\alpha + \nu\alpha)_i (n - j\alpha)_{m-i}
$$

$$
= \frac{1}{(\theta + n)_m} \sum_{\nu=0}^{r} (-1)^{r-\nu} \left(\frac{\theta}{\alpha} + j\right)_\nu S\left(r, \nu; \frac{\theta + j\alpha}{\alpha}\right) (\theta + n + \nu\alpha)_m,
$$

$$
\tag{22}
$$

where the last equality follows by an application of the Chu–Vandermonde formula. See, e.g., Charalambides (2005). Note, that for $r = 1$, we have

$$
E[K_m^{(n)} \mid K_n = j] = \left(j + \frac{\theta}{\alpha}\right) \left\{ \frac{(\theta + n + \alpha)_m}{(\theta + n)_m} - 1 \right\}, \tag{23}
$$

Now we can obtain the asymptotic moments by letting $m \to \infty$ in (22): using the Stirling formula we have

$$
\frac{1}{m^{r\alpha}} E\left[(K_m^{(n)})^r \mid K_n\right] \to \left(K_n + \frac{\theta}{\alpha}\right)_r \frac{\Gamma(\theta + n)}{\Gamma(\theta + n + r\alpha)} =: \mu_r^{(n)}. \tag{24}
$$

Such a moment sequence arises by taking $U_{n,j} \stackrel{d}{=} B_{j+\theta/\alpha,\, n/\alpha - j} Z_{(\theta+n)/\alpha}$, with the beta random variable $B_{j+\theta/\alpha,\, n/\alpha - j}$ independent from $Z_{(\theta+n)/\alpha}$, which has density (12). Hence, we are left with showing that the distribution of $U_{n,j}$ is uniquely characterized by the moment sequence $\{\mu_r^{(n)}\}_r$. In order to establish this, one can evaluate the characteristic function of $U_{n,j}$ which, at any $t \in \mathbb{R}$, coincides with

$$
\Phi(t) = \frac{\Gamma\left(\frac{\theta+n}{\alpha}\right)}{\Gamma\left(K_n + \frac{\theta}{\alpha}\right) \Gamma\left(\frac{n}{\alpha} - K_n\right)} \frac{\Gamma(\theta + n + 1)}{\Gamma\left(\frac{\theta+n}{\alpha} + 1\right)}
$$

$$
\times \int_0^\infty \mathrm{e}^{\mathrm{i}tz} z^{K_n + \frac{\theta}{\alpha} - 1} \int_z^\infty w \, (w - z)^{\frac{n}{\alpha} - K_n - 1} g_\alpha(w) \, \mathrm{d}w \, \mathrm{d}z
$$

$$
= \frac{\alpha \, \Gamma(\theta + n)}{\Gamma\left(K_n + \frac{\theta}{\alpha}\right) \Gamma\left(\frac{n}{\alpha} - K_n\right)} \int_0^\infty w \, g_\alpha(w)
$$

$$
\times \int_0^w \mathrm{e}^{\mathrm{i}tz} z^{K_n + \frac{\theta}{\alpha} - 1} (w - z)^{\frac{n}{\alpha} - K_n - 1} \, \mathrm{d}z \, \mathrm{d}w
$$

$$
= \frac{\Gamma(\theta + n + 1)}{\Gamma\left(\frac{\theta+n}{\alpha} + 1\right)} \sum_{r \geq 0} \frac{(\mathrm{i}t)^r}{r!} \frac{\left(K_n + \frac{\theta}{\alpha}\right)_r}{\left(\frac{\theta+n}{\alpha}\right)_r} \int_0^\infty w^{\frac{\theta+n}{\alpha} + r} g_\alpha(w) \, \mathrm{d}w
$$

$$= \sum_{r \geq 0} \frac{(\mathrm{i}t)^r}{r!} \frac{\left(K_n + \frac{\theta}{\alpha}\right)_r}{\left(\frac{\theta+n}{\alpha}\right)_r} \frac{\Gamma(\theta + n + 1)}{\Gamma\left(\frac{\theta+n}{\alpha} + 1\right)} \frac{\Gamma\left(\frac{\theta+n}{\alpha} + r + 1\right)}{\Gamma(\theta + n + 1 + r\alpha)}$$

$$= \sum_{r \geq 0} \frac{(\mathrm{i}t)^r}{r!} \mu_r^{(n)}$$

Hence, we have established that $(K_m^{(n)}|K_n = j)/m^\alpha$ converges a.s. and in p-th means to $U_{n,j}$.

As for the second component of the aggregate output (20), namely $\beta L_m^{(n)}/m^{1-\sigma}$, first note that by Proposition 2 in Lijoi, Prünster and Walker (2008), we have

$$\mathbb{E}[L_m^{(n)}] = m\frac{\theta + \alpha j}{\theta + n}.$$

This, combined with (23), yields immediately (21). The law of $L_m^{(n)}$ is given in Eq. (22) of Lijoi, Prünster and Walker (2008), which is easily seen to coincide with a Pólya distribution

$$\mathbb{P}(L_m^{(n)} = s|K_n = j) = \binom{m}{s} \frac{Be(m - s + n - j\alpha, s + \theta + j\alpha)}{Be(n - j\alpha, \theta + j\alpha)}, \quad (25)$$

for $s = 0, \ldots, m$, where $Be(a, b)$ denotes a beta function. Hence, the number of innovations within the new sectors follows a Pólya distribution. Therefore, by well–known martingale convergence arguments, it follows that $L_m^{(n)}/m$ converges a.s. and in the p–th mean to a beta random variable with parameters $\theta + j\alpha$ and $n - j\alpha$.

Now, combining this limit result with the previous concerning $K_m^{(n)}$ the asymptotic statements in (i), (ii) and (iii) follow immediately. $\square$

In order to associate the HPD intervals, which provide a measure of uncertainty of predictions based on the mean behaviour, to the limiting quantities of the conditional case one can easily extend the algorithm set forth for the unconditional case.

Some comments are in order at this point. The previous result shows how by complicating models so to adhere more closely to realistic assumptions non–self–averaging behaviours appear even more frequently. This represents a clear indicator that one cannot confine himself to studying mean behaviours but has to take the associated variability into account. This can be achieved in a quite straightforward way by

associating HPD intervals to the mean quantities. Therefore, the indication which clearly emerges from our analysis is that the usual way of proceeding in macroeconomics is legitimate as long as it is combined with suitable measures of uncertainty.

A   Appendix: Completely random measures

In this Appendix we provide a concise account on completely random measures, a concept introduced by Kingman (1967), which has the advantage of allowing to unify in an elegant way most classes of random probability measures dealt with in Bayesian Nonparametrics: indeed, all of them can be derived as suitable transformations of completely random measures. See Lijoi and Prünster (2010)

Let $(\mathbb{X}, \mathscr{X})$ be a Polish space equipped with the corresponding Borel σ–field and recall that a measure μ on $\mathbb{X}$ is said to be boundedly finite if $\mu(A) < +\infty$ for every bounded measurable set A. Denote by $(\mathbb{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ the space of boundedly finite measures endowed with the corresponding Borel σ–algebra. Let now $\tilde{\mu}$ be a measurable mapping from $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathbb{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ and such that for any $A_1, \ldots, A_n$ in $\mathscr{X}$, with $A_i \cap A_j = \varnothing$ for $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent. Then $\tilde{\mu}$ is termed *completely random measure* (CRM).

A CRM on $\mathbb{X}$ can always be represented as the sum of two components: a proper CRM $\tilde{\mu}_c = \sum_{i=1}^{\infty} J_i \delta_{Y_i}$, where both the positive jumps J_i's and the $\mathbb{X}$–valued locations Y_i's are random, and a measure with random masses at fixed locations in $\mathbb{X}$. Accordingly

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \, \delta_{z_i} \tag{26}$$

where the fixed jump points $z_1, \ldots, z_M$, with $M \in \{1, 2, \ldots, +\infty\}$, are in $\mathbb{X}$, the (non–negative) random jumps $V_1, \ldots, V_M$ are mutually independent and they are independent from $\tilde{\mu}_c$. Finally, $\tilde{\mu}_c$ is characterized by the Laplace functional

$$\mathbb{E}\left[\mathrm{e}^{- \int_{\mathbb{X}} f(x) \, \tilde{\mu}_c(\mathrm{d}x)} \right] = \exp\left\{ - \int_{\mathbb{R}^+ \times \mathbb{X}} \left[1 - \mathrm{e}^{-sf(x)} \right] \nu(\mathrm{d}s, \mathrm{d}x) \right\} \tag{27}$$

where $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $\int |f| \, \mathrm{d}\tilde{\mu}_c < \infty$ (almost surely) and ν is a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\int_B \int_{\mathbb{R}^+} \min\{s, 1\} \, \nu(\mathrm{d}s, \mathrm{d}x) < \infty \qquad (28)$$

for any B in $\mathscr{X}$. From (27), which provides a Lévy-Khintchine representation of CRMs, it is apparent that they are closely connected to Poisson processes. Indeed, $\tilde{\mu}_c$ can be represented as a linear functional of a Poisson process $\tilde{M}$ on $\mathbb{R}^+ \times \mathbb{X}$ with mean measure ν. To state this precisely, $\tilde{M}$ is a random subset of $\mathbb{R}^+ \times \mathbb{X}$ and if $\tilde{N}(A) = \mathrm{card}(\tilde{M} \cap A)$ for any $A \subset \mathscr{B}(\mathbb{R}^+) \otimes \mathscr{X}$ such that $\nu(A) < \infty$, then

$$\mathbb{P}[\tilde{N}(A) = k] = \frac{(\nu(A))^k \, \mathrm{e}^{-\nu(A)}}{k!} \qquad k = 0, 1, 2, \dots .$$

It can then be shown that

$$\tilde{\mu}_c(A) = \int_A \int_{\mathbb{R}^+} s \, \tilde{N}(\mathrm{d}s, \mathrm{d}x) \qquad \forall A \in \mathscr{X}. \qquad (29)$$

The measure ν characterizing $\tilde{\mu}_c$ is referred to as the *Lévy or Poisson intensity* of $\tilde{\mu}_c$: it contains all the information about the distributions of the jumps and locations of $\tilde{\mu}_c$. It is often useful to separate the jump and location part of ν by writing it as

$$\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(\mathrm{d}s) \, \gamma(\mathrm{d}x) \qquad (30)$$

where γ is a measure on $(\mathbb{X}, \mathscr{X})$ and ρ a transition kernel on $\mathbb{X} \times \mathscr{B}(\mathbb{R}^+)$, i.e. $x \mapsto \rho_x(A)$ is $\mathscr{X}$–measurable for any A in $\mathscr{B}(\mathbb{R}^+)$ and ρ_x is a measure on $(\mathbb{R}^+, \mathscr{B}(\mathbb{R}^+))$ for any x in $\mathbb{X}$. If $\rho_x = \rho$ for any x, then the distribution of the jumps of $\tilde{\mu}_c$ is independent of their location and both ν and $\tilde{\mu}_c$ are termed *homogeneous*. Otherwise, ν and $\tilde{\mu}_c$ are termed *non–homogeneous*.

Another important property of CRMs is their almost sure discreteness (Kingman, 1993), which means that their realizations are discrete measures with probability 1. This fact essentially entails discreteness of random probability measures obtained as transformations of CRMs.

The reader is referred to Kingman (1993) for a detailed treatment of the subject. Two important CRM to the present treatment are gamma CRM and the α–stable CRM, which we briefly outline here.

A homogeneous CRM $\tilde{\mu}$ whose Lévy intensity is given by

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{e}^{-s}}{s} \, \mathrm{d}s \, \gamma(\mathrm{d}x) \tag{31}$$

is a *gamma process* with parameter measure γ on $\mathbb{X}$. It is characterized by its Laplace functional which is given by

$$\mathbb{E}\left[\mathrm{e}^{-\int f \, \mathrm{d}\tilde{\mu}}\right] = \mathrm{e}^{-\int \log(1+f) \, \mathrm{d}\gamma} \tag{32}$$

for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\int \log(1 + |f|) \, \mathrm{d}\gamma < \infty$. Now set $f = \lambda \mathbb{1}_B$ with $\lambda > 0$, $B \in \mathscr{X}$ such that $\gamma(B) < \infty$ and $\mathbb{1}_B$ denoting the indicator function of set B. In this case one obtains

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \tilde{\mu}(B)}\right] = [1 + \lambda]^{-\gamma(B)},$$

from which it is apparent that $\tilde{\mu}(B)$ has a gamma distribution with scale and shape parameter equal to 1 and $\gamma(B)$, respectively.

As for the α–stable CRM, let $\alpha \in (0, 1)$, and γ be a boundedly finite measure on $\mathbb{X}$ and consider a CRM $\tilde{\mu}$ with Lévy intensity defined by

$$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\alpha}{\Gamma(1 - \alpha) \, v^{1+\alpha}} \, \mathrm{d}v \, \gamma(\mathrm{d}x). \tag{33}$$

The Laplace functional of such a CRM has the form

$$\mathbb{E}\left[\mathrm{e}^{-\int f \, \mathrm{d}\tilde{\mu}}\right] = \mathrm{e}^{-\int f^{\alpha} \mathrm{d}\gamma} \tag{34}$$

for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\int |f|^{\alpha} \mathrm{d}\gamma < \infty$. For instance, if $\alpha = 1/2$ and $f = \lambda \mathbb{1}_B$ with $\lambda > 0$, $B \in \mathscr{X}$ such that $\gamma(B) < \infty$ one obtains the well–known Laplace transform of a $1/2$–stable distribution

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \tilde{\mu}(B)}\right] = \mathrm{e}^{-\gamma(B)\sqrt{\lambda}}.$$

Acknowledgements

References

ANTONIAK, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1174.

AOKI, M. (2008). Thermodynamic limits of macroeconomic or financial models: one- and two-parameter Poisson-Dirichlet models. *J. Econom. Dynam. Control* **32**, 66–84.

AOKI, M. and YOSHIKAWA, H. (2007). Non-self-averaging Economic Growth: A Criticism of Endogenous Growth Theory. Technical Report, UCLA.

CHARALAMBIDES, C.A. (2005). *Combinatorial Methods in Discrete Distributions.* Hoboken, NJ: Wiley.

BLACKWELL, D. (1973). Discreteness of Ferguson selections. *Ann. Statist.* **1**, 356-358.

CONT, R. and TANKOV, P. (2004). *Financial modelling with jump processes.* Chapman & Hall/CRC, Boca Raton, FL.

DEVROYE, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation* **19**, N. 18.

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3** 87-112.

FAVARO, S., LIJOI A., MENA, R. H. and PRÜNSTER, I. (2009). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *J. R. Stat. Soc. Ser. B* **71**, 993–1008.

FERGUSON, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

FERGUSON, T.S. and KLASS, M.J. (1972). A representation of independent increments processes without Gaussian components. *Ann. Math. Statist.* **43**, 1634–1643.

GNEDIN, A. and PITMAN, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat.*

Inst. Steklov. (POMI) **325**, 83–102.

DE FINETTI, B. (1937). La prévision : ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68.

HJORT, N.L., HOLMES, C.C. MÜLLER, P., WALKER, S.G. (Eds.) (2010). *Bayesian Nonparametrics*. Cambridge University Press, Cambridge.

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

KINGMAN, J.F.C. (1993). *Poisson processes*. Oxford University Press, Oxford.

KORWAR, R.M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1**, 705–711.

LIJOI, A., MENA, R.H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *J. Amer. Statist. Assoc.* **100**, 1278–1291.

LIJOI, A., MENA, R.H. and PRÜNSTER, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Ser. B* **69**, 715–740.

LIJOI, MULIERE, PRÜNSTER and TADDEI (2010). Economic growth and averaging in macroeconomic models: a caveat. Working Paper.

LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. Eds.), pp. 80–136. Cambridge University Press, Cambridge.

LIJOI, A., PRÜNSTER, I., and WALKER, S.G. (2008). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18**, 1519–1547.

MONTAGNA, S. (2009). Random probability measures and their applications to Bayesian Nonparametrics. *Undergraduate thesis*, University of Torino, Torino.

PITMAN, J. (1995). Exchangeable and partially exchangeable random

partitions. *Probab. Theory Related Fields* **102**, 145–158.

PITMAN, J. (2006). *Combinatorial stochastic processes.* Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Math., Vol. **1875**. Springer, Berlin.

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

ROSIŃSKI, J. (2007). Tempering stable processes. *Stochastic Process. Appl.* **117**, 677–707.

SETHURAMAN, J. (1994). A constructive definition of the Dirichlet process prior. *Statist. Sinica* **2**, 639–650.

YAMATO, H. AND SIBUYA, M. (2000). Moments of some statistics of Pitman sampling formula. *Bull. Inform. Cybernet.* **32**, 1–10.