Georgian Mathematical Journal
Volume 10 (2003), Number 2, 319–324

WEAK CONVERGENCE OF A DIRICHLET-MULTINOMIAL PROCESS

PIETRO MULIERE AND PIERCESARE SECCHI

Abstract. We present a random probability distribution which approximates, in the sense of weak convergence, the Dirichlet process and supports a Bayesian resampling plan called a proper Bayesian bootstrap.

2000 Mathematics Subject Classification: 62G09, 60B10.
Key words and phrases: Dirichlet process, proper Bayesian bootstrap.

1. Introduction

The purpose of this paper is to throw light on a random probability distribution called the *Dirichlet-multinomial process* that approximates, in the sense of weak convergence, the Dirichlet process. A Dirichlet-multinomial process is a particular mixture of Dirichlet processes: in two previous works [11, 12] we showed that the process supports a Bayesian resampling plan which we called a *proper Bayesian bootstrap* suitable for approximating the distribution of functionals of the Dirichlet process and therefore being of interest in the context of Bayesian nonparametric inference.

Under different names, variants of the Dirichlet-multinomial model have been recently considered by other authors: see, for instance, [7] and the references therein. In fact, it has been pointed out that the Dirichlet-Multinomial model is equivalent to Fisher's species sampling model [5] recently reconsidered by Pitman among those extending the Blackwell and MacQueen urn scheme [13]. However none of these works allude to a connection between the Dirichlet-multinomial model and Bayesian bootstrap resampling plans. Recent applications of our proper Bayesian bootstrap include those in [3] for the approximation of the posterior distribution of the overflow rate in discrete-time queueing models.

In Section 2 we define the Dirichlet-multinomial process and we show that it can be used to approximate a Dirichlet process. Section 3 is dedicated to the proper Bayesian bootstrap algorithm and its connections with the Dirichlet-multinomial process.

2. A Convergence Result

Let $\mathcal{P}$ be the class of probability measures defined on the Borel σ-field $\mathcal{B}$ of $\Re$; for the reason of simplicity we work with $\Re$ but all the arguments below still hold if $\Re$ is replaced by a separable metric space. Endow $\mathcal{P}$ with the topology

ISSN 1072-947X / \$8.00 / © Heldermann Verlag *www.heldermann.de*

of weak convergence and write $\sigma(\mathcal{P})$ for the Borel σ-field in $\mathcal{P}$. With these assumptions $\mathcal{P}$ becomes a separable and complete metric space [14].

A useful random probability measure $P \in \mathcal{P}$ is the Dirichlet process introduced by Ferguson [4]. When α is a finite, nonnegative, nonnull measure on $(\Re, \mathcal{B})$ and P is a Dirichlet process with parameter α, we write $P \in \mathcal{D}(\alpha)$. We want to define a random element of $\mathcal{P}$ that is a mixture of Dirichlet processes; according to [1] we thus need to specify a transition measure and a mixing distribution.

Given $w > 0$, let $\alpha_w : \mathcal{P} \times \mathcal{B} \to [0, +\infty)$ be defined by setting, for every $P \in \mathcal{P}$ and $B \in \mathcal{B}$,

$$\alpha_w(P, B) = wP(B).$$

The function α_w is a transition measure. Indeed, for every $P \in \mathcal{P}$, $\alpha_w(P, \cdot)$ is a finite, nonnegative and nonnull measure on $(\Re, \mathcal{B})$ whereas, for every $B \in \mathcal{B}$, $\alpha_w(\cdot, B)$ is measurable on $(\mathcal{P}, \sigma(\mathcal{P}))$ since $\sigma(\mathcal{P})$ is a smallest σ-field in $\mathcal{P}$ such that the function $P \to P(B)$ is measurable, for every $B \in \mathcal{B}$.

Given a probability distribution P_0, let $X_1^*, \dots, X_m^*$ be an i.i.d. sample of size $m > 0$ from P_0. Assume $P_m^* \in \mathcal{P}$ to be the empirical distribution of $X_1^*, \dots, X_m^*$ defined by

$$P_m^* = \frac{1}{m} \sum_{i=1}^{m} \delta_{X_i^*},$$

where δ_x denotes the point mass at x. Write $\mathcal{H}_m^*$ for the distribution of P_m^* on $(\mathcal{P}, \sigma(\mathcal{P}))$.

Roughly, the following definition introduces a process P such that, conditionally on P_m^*, $P \in \mathcal{D}(wP_m^*)$.

Definition 2.1. A random element $P \in \mathcal{P}$ is called a Dirichlet-multinomial process with parameters (m, w, P_0) $(P \in \mathcal{DM}(m, w, P_0))$ if it is a mixture of Dirichlet processes on $(\Re, \mathcal{B})$ with mixing distribution $\mathcal{H}_m^*$ and transition measure α_w.

Remark 2.2. We call the process P defined above Dirichlet-multinomial since, as it will be seen in a moment, given any finite measurable partition $B_1, \dots, B_k$ of $\Re$, the distribution of $(P(B_1), \dots, P(B_k))$ is a mixture of Dirichlet distributions with multinomial weights. This process must not be confused with the Dirichlet-multinomial point process of Lo [9, 10] whose marginal distributions are mixtures of multinomial with Dirichlet weights.

It follows from the definition that if $P \in \mathcal{D}M(m, w, P_0)$, for every finite measurable partition $B_1, \dots, B_k$ of $\Re$ and $(y_1, \dots, y_k) \in \Re^k$,

$$\begin{aligned}
&\Pr\left(P(B_1) \le y_1, \dots, P(B_k) \le y_k\right) \\
&\quad = \int_{\mathcal{P}} D(y_1, \dots, y_k | \alpha_w(u, B_1), \dots, \alpha_w(u, B_k))\, d\mathcal{H}_m^*(u),
\end{aligned}$$

where $D(y_1, \dots, y_k | \alpha_1, \dots, \alpha_k)$ denotes the Dirichlet distribution function with parameters $(\alpha_1, \dots, \alpha_k)$ evaluated at $(y_1, \dots, y_k)$. With different notation, we may say that the vector $(P(B_1),\dots,P(B_k))$ has a distribution

$$\text{Dirichlet}\Big(w\frac{M_1}{m},\dots,w\frac{M_k}{m}\Big)\bigwedge_{(M_1,\dots,M_k)}\text{multinomial}\,(m,(P_0(B_1),\dots,P_0(B_k)))\,;$$

i.e., a mixture of Dirichlet distributions with multinomial weights.

For our purposes, the introduction of the Dirichlet-Multinomial process is justified by the following theorem.

Theorem 2.3. *For every $m > 0$, let $P_m \in \mathcal{P}$ be a Dirichlet-multinomial process with parameters (m, w, P_0). Then, when $m \to \infty$, P_m converges in distribution to a Dirichlet process with parameter wP_0.*

The result appears in [11] as well as in [13]. See also [8]. For ease of reference we sketch a simple argument, inspired by [16], that we consider as a nice didactic illustration of Prohorov's Theorem.

Proof. Given any finite measurable partition $B_1,\dots,B_k$ of $\Re$, the distribution of the vector $(P_m(B_1),\dots,P_m(B_k))$ weakly converges to a Dirichlet distribution with parameters $(wP_0(B_1),\dots,wP_0(B_k))$ when $m\to\infty$. In order to prove that P_m weakly converges to a Dirichlet process with parameter wP_0 it is therefore enough to show that the sequence of measures induced on $(\mathcal{P},\sigma(\mathcal{P}))$ by the processes P_m, $m = 1,2,\dots$, is tight. Given $\epsilon > 0$, let K_r, $r = 1,2,\dots$, be a compact set of $\Re$ such that $P_0(K_r^c) \le \epsilon/r^3$ and define

$$M_r = \Big\{P\in\mathcal{P} : P(K_r^c) \le \frac{1}{r}\Big\}.$$

The set $M = \bigcap_{r=1}^{\infty} M_r$ is compact in $\mathcal{P}$. For $m = 1,2,\dots$ and $r = 1,2,\dots$, $E[P_m(K_r^c)] = P_0(K_r^c)$ and thus

$$\Pr\Big(P_m(K_r^c) > \frac{1}{r}\Big) \le rP_0(K_r^c) \le \frac{\epsilon}{r^2}.$$

Hence, for every $m = 1,2,\dots$,

$$\Pr(P_m\in M) \ge 1-\sum_{r=1}^{\infty}\Pr\Big(P_m(K_r^c) > \frac{1}{r}\Big) \ge 1-\epsilon\sum_{r=1}^{\infty}\frac{1}{r^2}. \qquad \square$$

3. Connections with the Proper Bayesian Bootstrap

Let $T : \mathcal{P} \to \Re$ be a measurable function and $P \in \mathcal{D}(wP_0)$ with $w > 0, P_0 \in \mathcal{P}$. It is often difficult to work out analytically the distribution of $T(P)$ even when T is a simple statistical functional like the mean [6, 2]. However, when P_0 is discrete with finite support one may produce a reasonable approximation of the distribution of $T(P)$ by a Monte Carlo procedure that obtains i.i.d. samples from $\mathcal{D}(wP_0)$. If P_0 is not discrete, we propose to approximate the distribution of $T(P)$ by the distribution of $T(P_m)$, where P_m is a Dirichlet-multinomial process with parameters (m, w, P_0) and m is large enough.

Of course, since the Continuous Mapping Theorem does not apply to every function T, the fact that P_m converges in distribution to P does not always imply that the distribution of $T(P_m)$ is close to that of $T(P)$. However, we proved in [12] that this is in fact the case when T belongs to a large class of linear functionals or when T is a quantile. In [12] we also tested by means of a few numerical examples a bootstrap algorithm that generates an approximation of the distribution of $T(P)$ in the following steps:

(1) Generate an i.i.d sample $X_1^*, \ldots, X_m^*$ from P_0.
(2) Generate an i.i.d. sample $V_1, \ldots, V_m$ from a Gamma$(\frac{w}{m}, 1)$.
(3) Compute $T(P_m)$, where $P_m \in \mathcal{P}$ is defined by
$$P_m = \frac{1}{\sum_{i=1}^m V_i} \sum_{i=1}^m V_i \delta_{X_i^*}.$$
(4) Repeat steps (1)–(3) s times and approximate the distribution of $T(P)$ with the empirical distribution of the values $T_1, \ldots, T_s$ generated at step (3).

It is easily seen that the probability distribution P_m generated in step (3) is in fact a trajectory of the Dirichlet-multinomial process with parameters (m, w, P_0). We may therefore conclude that the previous algorithm aims at approximating the distribution of $T(P)$ by distribution of $T(P_m)$, where $P_m \in \mathcal{D}M(m, w, P_0)$, and approximates the latter by means of the empirical distribution of the values $T_1, \ldots, T_s$ generated in step (3).

Remark 3.1. Step (1) is useless when P_0 is discrete with finite support $\{z_1, \ldots, z_m\}$ and $P_0(z_i) = p_i, i = 1, \ldots, m$, with $\sum_{i=1}^m p_i = 1$. In fact, in this case one may generate at step (3) a trajectory of $P \in \mathcal{D}(wP_0)$, by taking
$$P_m = \frac{1}{\sum_{i=1}^m V_i} \sum_{i=1}^m V_i \delta_{z_i}$$
where $V_1, \ldots, V_m$, are independent and V_i has distribution Gamma$(wp_i, 1)$, $i = 1, \ldots, m$.

We call the algorithm (1)–(4) the *proper Bayesian bootstrap.* To understand the reason for this name consider the following situation. A sample $X_1, \ldots, X_n$ from a process $P \in \mathcal{D}(kQ_0)$, with $k > 0$ and $Q_0 \in \mathcal{P}$, has been observed and the problem is to compute the posterior distribution of $T(P)$ where T is a given statistical functional. Ferguson [4] proved that the posterior distribution of P is again a Dirichlet process with parameter $kQ_0 + \sum_{i=1}^n \delta_{X_i}$. In order to approximate the posterior distribution of $T(P)$ our algorithm generates an i.i.d. sample $X_1^*, \ldots, X_m^*$ from
$$\frac{k}{k+n} Q_0 + \frac{n}{k+n} \left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i} \right)$$
and then, in step (3), produces a trajectory of a process that, given $X_1^*, \ldots, X_m^*$, is Dirichlet with parameter$= (k+n)m^{-1} \sum_{i=1}^m \delta_{X_i^*}$ and evaluates T with respect to this trajectory. The algorithm is therefore a bootstrap procedure since it samples from a mixture of the empirical distribution function generated by

$X_1, \dots, X_n$ and Q_0 which, together with the weight k, elicits the prior opinions relative to P. Because it takes into account prior opinions by means of a proper distribution function, the procedure was termed proper.

The name proper Bayesian bootstrap also distinguishes the algorithm from the Bayesian bootstrap of Rubin [15] that approximates the posterior distribution of $T(P)$ by means of the distribution of $T(Q)$ with $Q \in \mathcal{D}(\sum_{i=1}^{n} \delta_{X_i})$. We already noticed in the previous work [12] that there are no proper priors for P which support Rubin's approximation and that the proper Bayesian bootstrap essentially becomes the Bayesian bootstrap of Rubin when k tends to 0 or n is very large.

Acknowledgements

We thank an anonymous referee for his helpful comments that greatly improved the paper. Both authors were financially supported by the Italian MIUR's research project *Metodi Bayesiani nonparametrici e loro applicazioni.*

References

1. C. Antoniak, Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**(1974), 1152–1174.
2. D. M. Cifarelli and E. Regazzini, Distribution functions of means of Dirichlet process. *Ann. Statist.* **18**(1990), No. 1, 429–442.
3. P. L. Conti, Bootstrap approximations for Bayesian analysis of Geo/G/1 discrete time queueing models. *J. Statist. Plann. Inference*, 2002 (in press).
4. T. S. Ferguson, A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**(1973), No. 2, 209–230.
5. R. A. Fisher, A. S. Corbet, and C. B. Williams, The relation between the number of species and the number of individuals in a random sample of animal population. *J. Animal Ecology* **12**(1943), 42–58.
6. R. C. Hannum, M. Hollander, and N. A. Langberg, Distributional results for random functionals of a Dirichlet process. *Ann. Prob.* **9**(1981), 665–670.
7. H. Ishwaran and L. F. James, Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**(2001), 161–173.
8. H. Ishwaran and M. Zarepour, Exact and approximate sum-representations for the Dirichlet process. *Canad. J. Statist.* **30**(2002), 269–283.
9. A. Y. Lo, Bayesian statistical inference for sampling a finite population. *Ann. Statist.* **14**(1986), No. 3, 1226–1233.
10. A. Y. Lo, A Bayesian bootstrap for a finite population. *Ann. Statist.* **16**(1988), No. 4, 1684–1695.
11. P. Muliere and P. Secchi, A note on a proper Bayesian mootstrap. *Technical Report #18, Dip. di Economia Politica e Metodi Quantitativi, Università di Pavia,* 1995.
12. P. Muliere and P. Secchi, Bayesian nonparametric predictive inference and bootstrap techniques. *Ann. Inst. Statist. Math.* **48**(1996), No. 4, 663–673.
13. J. Pitman, Some developments of the Blackwell–MacQueen urn scheme. *Statistics, probability and game theory,* 245–267, *IMS Lecture Notes Monogr. Ser.,* 30, *Inst. Math. Statist., Hayward, CA,* 1996.

14. Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1**(1956), 157–214.
15. D. M. Rubin, The Bayesian bootstrap. *Ann. Statist.* **9**(1981), No. 1, 130–134.
16. J. Sethuraman and R. C. Tiwari, Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical decision theory and related topics, III, Vol.* 2 (*West Lafayette, Ind.,* 1981), 305–315, *Academic Press, New York–London,* 1982.

(Received 3.10.2002; revised 23.01.2003)

Authors' addresses:

Pietro Muliere
Università L. Bocconi
Istituto di Metodi Quantitativi
Viale Isonzo 25, 20135 Milano
Italy
E-mail: pietro.muliere@uni-bocconi.it

Piercesare Secchi
Politecnico di Milano
Dipartimento di Matematica
Piazza Leonardo da Vinci 32, 20132 Milano
Italy
E-mail: secchi@mate.polimi.it