Prediction rules for exchangeable sequences related to species sampling[1]

by

Ben Hansen and Jim Pitman

Technical Report No. 520

Department of Statistics
University of California
367 Evans Hall # 3860
Berkeley, CA 94720-3860

May 1998

[1]Research supported in part by N.S.F. Grant DMS 97-03961

Prediction rules for exchangeable sequences related to species sampling[†]

Ben Hansen and Jim Pitman

May 18, 1998

Abstract

Suppose an exchangable sequence with values in a nice measurable space S admits a prediction rule of the following form: given the first n terms of the sequence, the next term equals the jth distinct value observed so far with probability $p_{j,n}$, for $j = 1, 2, \ldots$, and otherwise is a new value with distribution ν for some probability measure ν on S with no atoms. Then the $p_{j,n}$ depend only on the partitition of the first n integers induced by the first n values of the sequence. All possible distributions for such an exchangeable sequence are characterized in terms of constraints on the $p_{j,n}$ and in terms of their de Finetti representations.

1 Introduction

There are very few models for exchangeable sequences (X_n) with an explicit *prediction rule*, that is a formula for the conditional distribution of X_{n+1} given $X_1, \ldots, X_n$ for each $n = 0, 1, \ldots$. The Blackwell-MacQueen urn scheme [3] provides an example: given a probability measure $\nu(\cdot)$ on a nice measurable space $(S, \mathcal{S})$ and $\theta > 0$, the prediction rule

$$\mathbb{P}(X_{n+1} \in \cdot \,|\, X_1, \ldots, X_n) = \frac{1}{(n+\theta)} \sum_{i=1}^{n} 1(X_i \in \cdot) + \frac{\theta}{(n+\theta)} \, \nu(\cdot) \qquad (1)$$

determines an exchangeable sequence (X_n) whose directing random measure F has Dirichlet distribution with parameter $\theta\nu(\cdot)$. See [6] for background and applications of this model to non-parametric statistics. The subject of this paper is exchangeable sequences admitting a prediction rule of the more general form

$$\mathbb{P}(X_{n+1} \in \cdot \,|\, X_1, \ldots, X_n) = \sum_{i=1}^{n} r_{i,n} 1(X_i \in \cdot) + q_n \, \nu(\cdot) \qquad (2)$$

for some $r_{i,n}$ and q_n which are non-negative product-measurable functions of $(X_1, \ldots, X_n)$. As a minimal regularity condition on $(S, \mathcal{S})$, we suppose that the

[†]Research supported in part by N.S.F. Grant DMS 97-03961

diagonal $\{(x,y) : x = y\}$ is a product-measurable subset of $S \times S$. The rule (2) can then be rewritten as follows, by grouping terms with equal values of X_i:

$$\mathbb{P}(X_{n+1} \in \cdot \,|\, X_1, \ldots, X_n) = \sum_{j=1}^{K_n} p_{j,n} 1(\tilde{X}_j \in \cdot) + q_n \nu(\cdot) \tag{3}$$

where the $\tilde{X}_j$ for $1 \leq j \leq K_n$ are the distinct values among $X_1, \ldots, X_n$ in the order that they appear, and the $p_{j,n}$ and q_n are some non-negative product-measurable functions of $(X_1, \ldots, X_n)$. This paper provides a description of all prediction rules of this form which generate exchangeable sequences, assuming that the probability measure ν is *diffuse*, meaning $\nu\{x\} = 0$ for all points x of S.

Let Π denote the random partition of $\{1, 2, \ldots, \}$ generated by $X_1, X_2, \ldots$. So $\Pi = \{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$ where $\mathcal{A}_j$ is the random set of indices m such that $X_m = \tilde{X}_j$. Let Π_n be the restriction of Π to $\{1, \ldots, n\}$. So Π_n is a measurable function of $X_1, \ldots, X_n$ with values in the finite set of all partitions of the set $\{1, \ldots, n\}$. The main new result of this paper is the following theorem, which is proved in Section 2.

Theorem 1 *Suppose that an S-valued exchangeable sequence (X_n) admits a prediction rule of the form (3) for $p_{j,n}$ and q_n some product-measurable functions of $(X_1, \ldots, X_n)$, and ν a diffuse measure on S. Then for each n and $1 \leq j \leq K_n$ the $p_{j,n}$ and q_n are almost surely equal to some functions of Π_n, the partition of $\{1, \ldots, n\}$ generated by $(X_1, \ldots, X_n)$.*

While the focus of this paper is exchangeable sequences subject to a prediction rule of the form (3) for a diffuse measure ν, we note that a weakening of Theorem 1 holds for ν that is a mixture of diffuse and atomic measures. Then the $p_{j,n}$ and q_n are almost surely equal to some functions of Π_n and the collection of random sets

$$\{\{i \leq n : X_i = a\} : a \text{ an atom of } \nu\}. \tag{4}$$

This can be established by a slight variation of the proof of Theorem 1 given in Section 2.

The rest of this introduction shows how Theorem 1 combines with results obtained previously in [14] to yield a description of all possible functions $p_{j,n}$ and q_n that could be used to generate an exchangeable sequence (X_n) by a prediction rule of the form (3) for diffuse ν, and a corresponding description of the de Finetti representation of (X_n) in terms of sampling from a random distribution.

The assumption that (X_n) is exchangeable implies that Π is an *exchangeable random partition* of the set of positive integers, as considered by Kingman [9, 10] and subsequent authors [1, 12]. That is to say, for each $n = 1, 2, \ldots$ and each partition $\{A_1, \ldots, A_k\}$ of $\{1, \ldots, n\}$,

$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}) = p(\#A_1, \ldots, \#A_k) \tag{5}$$

for some non-negative symmetric function p of finite sequences of positive integers $\boldsymbol{n} := (n_1, \ldots, n_k)$. Here $\#A$ is the number of elements of A. Following [12, 14], call p the *exchangeable partition probability function* (EPPF) determined by Π. Write $k(\boldsymbol{n})$ for the length k of $\boldsymbol{n} := (n_1, \ldots, n_k)$. For each finite sequence $\boldsymbol{n}$ of positive integers and each $1 \leq j \leq k(\boldsymbol{n}) + 1$, a finite sequence $\boldsymbol{n}^{j+}$ of positive integers is defined by incrementing n_j by 1. From (5) and the addition rule of probability, an EPPF must satisfy

$$p(1) = 1 \text{ and } p(\boldsymbol{n}) = \sum_{j=1}^{k(\boldsymbol{n})+1} p(\boldsymbol{n}^{j+}), \text{ for all } \boldsymbol{n}. \tag{6}$$

Let

$$N_{j,n} := \sum_{m=1}^{n} 1[X_m = \tilde{X}_j] \tag{7}$$

which is the number of times that the jth distinct value $\tilde{X}_j$ appears among $X_1, \ldots, X_n$. So $N_{j,n}$ is the number of elements in the jth class of Π_n when classes are ordered by their least elements. If (X_n) is exchangeable and subject to a prediction rule of the form (3), with $p_{j,n}$ and q_n functions of Π_n, it is easily seen that almost surely for all $j \leq K_n$

$$p_{j,n} = p_j(N_{1,n}, \ldots, N_{K_n,n}); \quad q_n = q(N_{1,n}, \ldots, N_{K_n,n}) \tag{8}$$

for some non-negative functions p_j and q of finite sequences of positive integers. These functions p_j and q can be characterized as follows:

Theorem 2 [14, Prop. 13 and Thm. 14] *Suppose (X_n) is exchangeable and subject to a prediction rule of the form (3), with $p_{j,n}$ and q_n as in (8). Then the functions p_j and q can be expressed as follows in terms of the* EPPF *associated with the random partition Π generated by (X_n): provided $p(\boldsymbol{n}) > 0$,*

$$p_j(\boldsymbol{n}) = \frac{p(\boldsymbol{n}^{j+})}{p(\boldsymbol{n})} \text{ for } 1 \leq j \leq k(\boldsymbol{n}); \; q(\boldsymbol{n}) = \frac{p(\boldsymbol{n}^{\ell+})}{p(\boldsymbol{n})} \text{ for } \ell = k(\boldsymbol{n}) + 1. \tag{9}$$

Conversely, given a diffuse measure ν on $(S, \mathcal{S})$ and a non-negative symmetric function of finite sequences of positive integers subject to (6), the prediction rule (3) determined via (8) and (9) defines an exchangeable sequence (X_n). Such a sequence (X_n) may be constructed by first generating an exchangeable random partition $\Pi = \{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$ whose EPPF *is p, then setting $X_n = \tilde{X}_j$ for $n \in \mathcal{A}_j$ where the $\tilde{X}_j$ are i.i.d. with distribution ν, independent of Π.*

Following [14], call such an exchangeable sequence (X_n) a *species sampling sequence*. This terminology is used to suggest the interpretation of (X_n) as the sequence of species of individuals in a process of sequential random sampling from some hypothetical infinite population of individuals of various species. The species of the first individual to be observed is assigned a random tag $X_1 = \tilde{X}_1$ distributed according to ν. Given the tags $X_1, \ldots, X_n$ of the first n individuals observed, it is supposed that the next individual is one of the jth species observed so far with probability $p_{j,n}$, and one of a new species with probability q_n. Each distinct species is assigned an independent random tag with distribution ν as it appears in the sampling process. In this interpretation the random partition Π generated by the species sampling process is of primary importance: the allocation of i.i.d. random tags to distinct species is just a device to encode Π in a sequence of exchangeable random variables (X_n). As shown by Aldous [1], this device allows Kingman's representation of exchangeable random partitions to be immediately deduced from de Finetti's representation of exchangeable sequences. For this purpose, the choice of the space S of species tags and the diffuse measure ν on S is of no importance: one may as well take $S = [0,1]$ with Borel sets and ν the uniform distribution on $[0,1]$.

The de Finetti representation of a species sampling sequence (X_n) can be described as follows:

Theorem 3 [14] *Write $\tilde{P}_j$ for the limiting frequency of the jth species to appear in a species sampling sequence (X_n):*

$$\tilde{P}_j := \lim_{n\to\infty} \frac{N_{j,n}}{n} \tag{10}$$

which exists almost surely. Let F_n denote the conditional distribution of X_{n+1} given $X_1, \ldots, X_n$, as displayed in (3). *Then F_n converges in total variation norm almost surely as $n \to \infty$ to the random measure*

$$F(\cdot) := \sum_j \tilde{P}_j 1(\tilde{X}_j \in \cdot) + (1 - \sum_j \tilde{P}_j)\nu(\cdot). \tag{11}$$

Conditionally given F the X_n are independent and identically distributed according to F.

The joint law of the $\tilde{P}_j$ is determined by the EPPF of the partition Π generated by (X_n) via formulae described in [14]. See [12, 14] regarding the conditional distribution of Π given the sequence $(\tilde{P}_j)$, which is the same for all species sampling sequences. See [14] regarding the conditional distribution of F given $(X_1, \ldots, X_n)$. Theorem 3 yields also:

Corollary 4 [14] *A sequence (X_n) is a species sampling sequence with marginal distributions equal to ν if and only if (X_n) is conditionally i.i.d. (F) given some random probability distribution F on S of the form*

$$F := \sum_j P_j 1(\hat{X}_j \in \cdot) + (1 - \sum_j P_j)\nu(\cdot). \tag{12}$$

for some sequence of random variables $P_j \geq 0$ with $\sum_j P_j \leq 1$, and given (P_j) the $\hat{X}_j$ corresponding to j with $P_j > 0$ are i.i.d. (ν).

Example. *The Two-Parameter Model* [12]. Consider the prediction rule (3) defined by some diffuse measure ν and

$$p_{j,n} = \frac{N_{j,n} - \alpha}{n + \theta} \text{ for } 1 \le j \le K_n; \quad q_n = \frac{\theta + K_n \alpha}{n + \theta} \tag{13}$$

where α and θ are two real parameters and as before the $N_{j,n}, 1 \le j \le K_n$ are the numbers of representatives of the various distinct species $\tilde{X}_j, 1 \le j \le K_n$ among $X_1, \ldots, X_n$. To ensure that all relevant probabilities are non-negative and that the rule is not degenerate, it must be supposed that either

$$\alpha = -\kappa < 0 \text{ and } \theta = m\kappa \text{ for some } \kappa > 0 \text{ and } m = 2, 3 \ldots \tag{14}$$

or

$$0 \le \alpha < 1 \text{ and } \theta > -\alpha. \tag{15}$$

This prediction rule (13) is that determined by (9) for the function $p = p_{(\alpha,\theta)}$ defined by the formula

$$p_{(\alpha,\theta)}(n_1, \ldots, n_k) = \frac{\left(\prod_{\ell=1}^{k-1}(\theta + \ell\alpha)\right)\left(\prod_{i=1}^{k}[1-\alpha]_{n_i-1}\right)}{[1+\theta]_{n-1}} \tag{16}$$

where $n = \sum_i n_i$ and $[x]_m = \prod_{j=1}^{m}(x + j - 1)$. It is easily checked that $p_{(\alpha,\theta)}$ is an EPPF. So a sequence $(X_1, X_2, \ldots)$ defined by the prediction rule (13) is exchangeable, hence a species sampling sequence. The case with $\alpha = 0$ is the Blackwell-McQueen scheme. Then (16) is a variation of the Ewens sampling formula [4, 2, 5]. In the case (14), the distribution of the exchangeable sequence (X_n) is identical to that generated by sampling from $F := \sum_{i=1}^{m} P_i 1(\hat{X}_i \in \cdot)$, where $(P_1, \ldots, P_m)$ has a symmetric Dirichlet distribution with m parameters equal to κ, and the $\hat{X}_i$ are i.i.d. with distribution ν. This is Fisher's model for species sampling [7] with m species identified by i.i.d.(ν) tags. See [13, 16, 15, 17, 8, 18, 11] for further characterizations and applications of the two-parameter model.

2 Proof of Theorem 1

Suppose throughout this section that (X_n) is an S-valued exchangeable sequence subject to a prediction rule of the form (3) for $p_{j,n}$ and q_n some arbitrary measurable functions of $(X_1, \ldots, X_n)$, and ν a diffuse measure on S. Let Π_n be the partition of $\{1, \ldots, n\}$ generated by $X_1, \ldots, X_n$. In view of the last sentence of Theorem 2, to establish the conclusion of Theorem 1 that modulo null sets the $p_{j,n}$ and q_n depend only on Π_n, it suffices to show that conditionally given Π, the partition of all positive integers generated by (X_n), the random variables $\tilde{X}_j$ for $j = 1, 2, \ldots$ are independent and identically distributed according to ν. The following lemma provides a convenient reformulation of this condition:

Lemma 5 *For all $1 \le k \le n$, all partitions π of $\{1, \dots, n\}$ with k classes, and for all choices of measurable $B_j \subseteq S, 1 \le j \le k$*

$$\mathbb{P}(\Pi_n = \pi; \tilde{X}_j \in B_j, 1 \le j \le k) = \left(\prod_{j=1}^{k} \nu(B_j) \right) \mathbb{P}(\Pi_n = \pi) \tag{17}$$

Proof. This is the result of repeated application of the following formula, which is claimed to hold for all choices of $1 \le k \le n, \pi$ and $B_j, 1 \le j \le k$, as above, and all choices of i with $1 \le i \le n$:

$$\mathbb{P}(\Pi_n = \pi; \tilde{X}_j \in B_j \text{ all } j \le k) = \nu(B_i)\, \mathbb{P}(\Pi_n = \pi; \tilde{X}_j \in B_j \text{ all } j \le k, j \ne i) \tag{18}$$

If π is a partition of $\{1, \dots, n\}$ into k classes, write $A_1^\pi, \dots, A_k^\pi$ for the k classes, ordered such that $1 = \min A_1^\pi < \min A_2^\pi < \dots < \min A_k^\pi$. Let $n, \pi, k, B_1, \dots, B_k$ be as in (18). It follows immediately from the prediction rule (3) and the assumption that ν is diffuse that (18) holds if $i = k$ and $\# A_k^\pi = 1$. The assumed exchangeablility of (X_n) then yields (18) for any $1 \le i \le k$ with $\# A_i^\pi = 1$.

Now consider the inductive hypothesis, call it H_m, that (18) holds for all choices of $1 \le k \le n, \pi, B_j, 1 \le j \le k$ and $1 \le i \le k$ with $\# A_i^\pi = m$. We have just shown that H_1 holds. We now assume H_m for some $m = 1, 2, \dots$, and will complete the proof of the lemma by deducing H_{m+1}. As in the argument for $m = 1$, we first obtain a special case of H_{m+1}; but by exchangeability, the special case implies the general case of H_{m+1}. So consider partitions π' of $\{1, \dots, n+1\}$ for which $\# A_1^{\pi'} = m + 1$ and $n + 1 \in A_1^{\pi'}$. We prove H_{m+1} for these π' and for $i = 1$.

Fix such a π' partitioning $\{1, \dots, n+1\}$, and measurable $B_1, \dots, B_k \subseteq S$, and to avoid trivialities assume $B_1, \dots, B_k$ all have positive ν-measure. Write $\pi = \{A_1^\pi, \dots, A_k^\pi\}$ for the restriction of π' to $\{1, \dots, n\}$. For $\ell = 1, \dots, k$, write π^ℓ for the partition $\{A_1^\pi, \dots, A_\ell^\pi \cup \{n+1\}, \dots, A_k^\pi\}$ of $\{1, \dots, n+1\}$. Note that $\pi' = \pi^1$. Write π^{k+1} for the partition $\{A_1^\pi, \dots, A_k^\pi, \{n+1\}\}$ of $\{1, \dots, n+1\}$. By H_m, for each $\ell = 2, \dots, k+1$,

$$\mathbb{P}(\Pi_{n+1} = \pi^\ell, \tilde{X}_j \in B_j \text{ all } j \le k) = \nu(B_1)\mathbb{P}(\Pi_{n+1} = \pi^\ell; \tilde{X}_j \in B_j \text{ all } 2 \le j \le k)$$

since in each of the partitions $\pi^2, \dots, \pi^{k+1}$ the first class has size m. Similarly,

$$\mathbb{P}(\Pi_n = \pi, \tilde{X}_j \in B_j \text{ all } j \le k) = \nu(B_1)\mathbb{P}(\Pi_n = \pi; \tilde{X}_j \in B_j \text{ all } 2 \le j \le k).$$

The identity

$$\mathbb{P}(\Pi_n = \pi, \tilde{X}_j \in B_j \text{ all } j \le k) = \sum_{\ell=1}^{k+1} \mathbb{P}(\Pi_{n+1} = \pi^\ell, \tilde{X}_j \in B_j \text{ all } j \le k)$$

now implies that

$$\mathbb{P}(\Pi_{n+1} = \pi^1, \tilde{X}_j \in B_j \text{ all } j \le k) = \nu(B_1)\mathbb{P}(\Pi_{n+1} = \pi^1, \tilde{X}_j \in B_j \text{ all } 2 \le j \le k),$$

which is the identity required to establish H_{m+1}.□

References

[1] D.J. Aldous. Exchangeability and related topics. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XII, Springer Lecture Notes in Mathematics, Vol. 1117*. Springer-Verlag, 1985.

[2] C. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.*, 2:1152–1174, 1974.

[3] D. Blackwell and J.B. MacQueen. Ferguson distributions via Pólya urn schemes. *Ann. Statist.*, 1:353–355, 1973.

[4] W.J. Ewens. The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3:87 – 112, 1972.

[5] W.J. Ewens and S. Tavaré. The Ewens sampling formula. To appear in *Multivariate Discrete Distributions* edited by N.S. Johnson, S. Kotz, and N. Balakrishnan, 1995.

[6] T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.

[7] 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 an animal population. *J. Animal Ecol.*, 12:42–58, 1943.

[8] S. Kerov. Coherent random allocations and the Ewens-Pitman formula. PDMI Preprint, Steklov Math. Institute, St. Petersburg, 1995.

[9] J. F. C. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

[10] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13:235–248, 1982.

[11] A.Z. Mekjian and K.C. Chase. Disordered systems, power laws and random processes. *Phys. Letters A*, 229:340–346, 1997.

[12] J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Th. Rel. Fields*, 102:145–158, 1995.

[13] J. Pitman. Random discrete distributions invariant under size-biased permutation. *Adv. Appl. Prob.*, 28:525–539, 1996.

[14] J. Pitman. Some developments of the Blackwell-MacQueen urn scheme. In T.S. Ferguson et al., editor, *Statistics, Probability and Game Theory; Papers in honor of David Blackwell*, volume 30 of *Lecture Notes-Monograph Series*, pages 245–267. Institute of Mathematical Statistics, Hayward, California, 1996.

[15] J. Pitman. Coalescents with multiple collisions. Technical Report 495, Dept. Statistics, U.C. Berkeley, 1997. Available via http://www.stat.berkeley.edu/users/pitman.

[16] J. Pitman. Partition structures derived from Brownian motion and stable subordinators. *Bernoulli*, 3:79–96, 1997.

[17] J. Pitman and M. Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, 25:855–900, 1997.

[18] S.L. Zabell. The continuum of inductive methods revisited. In J. Earman and J. D. Norton, editors, *The Cosmos of Science*, Pittsburgh-Konstanz Series in the Philosophy and History of Science, pages 351–385. University of Pittsburgh Press/Universitätsverlag Konstanz, 1997.