Statistics & Probability Letters 46 (2000) 251–256

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

# Prediction rules for exchangeable sequences related to species sampling ☆

Ben Hansen[a], Jim Pitman[b,*]

[a] Group in Logic and Methodology of Science, 731 Evans Hall, Berkeley, CA 94720-3840, USA
[b] Department of Statistics, University of California, 367 Evans Hall, Berkeley, CA 94720-3860, USA

Received February 1999

**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 $j$th distinct value observed so far with probability $p_{j,n}$, for $j=1,2,\dots$, and otherwise is a new value with distribution $\nu$ for some probability measure $\nu$ 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. © 2000 Elsevier Science B.V. All rights reserved

*Keywords:* Exchangeable sequence; Prediction rule; Species sampling

## 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 $X_1,\dots,X_n$ for each $n=0,1,\dots$ . The Blackwell–MacQueen urn scheme (Blackwell and MacQueen (1973)) provides an example: given a probability measure $\nu(\cdot)$ on a nice measurable space $(S,\mathscr{S})$ and $\theta>0$, the prediction rule

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

determines an exchangeable sequence $(X_n)$ whose directing random measure $F$ has Dirichlet distribution with parameter $\theta\nu(\cdot)$. See Ferguson (1973) 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

☆ Research supported in part by N.S.F. Grant DMS 97-03961.

* Corresponding author.
*E-mail address:* pitman@stat.berkeley.edu (J. Pitman)


PII: S0167-7152(99)00109-1

form

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

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

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

where the $\tilde{X}_j$ for $1 \leqslant j \leqslant K_n$ 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 $v$ is *diffuse*, meaning $v\{x\} = 0$ for all points $x$ of $S$.

Let $\Pi$ denote the random partition of $\{1, 2, \ldots,\}$ generated by $X_1, X_2, \ldots$ . So $\Pi = \{\mathscr{A}_1, \mathscr{A}_2, \ldots\}$ where $\mathscr{A}_j$ is the random set of indices $m$ such that $X_m = \tilde{X}_j$. Let $\Pi_n$ be the restriction of $\Pi$ to $\{1, \ldots, n\}$. So $\Pi_n$ 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 form* (3) *for* $p_{j,n}$ *and* $q_n$ *some product-measurable functions of* $(X_1, \ldots, X_n)$, *and* $v$ *a diffuse measure on S. Then for each n and* $1 \leqslant j \leqslant K_n$ *the* $p_{j,n}$ *and* $q_n$ *are almost surely equal to some functions of* $\Pi_n$, *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 $v$, we note that a weakening of Theorem 1 holds for $v$ 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 $\Pi_n$ and the collection of random sets

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

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 Pitman (1996b) 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 form (3) for diffuse $v$, 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 $\Pi$ is an *exchangeable random partition* of the set of positive integers, as considered by Kingman (1978, 1982) 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 Pitman (1995, 1996b), call $p$ the *exchangeable partition probability function* (EPPF) determined by $\Pi$. 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 $1 \leqslant j \leqslant k(\boldsymbol{n}) + 1$, 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

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

Let

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

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

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

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** (Pitman, 1996b, Proposition 13 and Theorem 14)**.** *Suppose* $(X_n)$ *is exchangeable and subject to a prediction rule of 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* $\Pi$ *generated by* $(X_n)$: *provided* $p(\boldsymbol{n}) > 0$,

$$p_j(\boldsymbol{n}) = \frac{p(\boldsymbol{n}^{j+})}{p(\boldsymbol{n})} \quad \textit{for } 1 \leqslant j \leqslant k(\boldsymbol{n}), \qquad q(\boldsymbol{n}) = \frac{p(\boldsymbol{n}^{l+})}{p(\boldsymbol{n})} \quad \textit{for } l = k(\boldsymbol{n}) + 1. \tag{9}$$

*Conversely*, *given a diffuse measure* $v$ *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* $\Pi = \{\mathscr{A}_1, \mathscr{A}_2, \ldots\}$ *whose EPPF is p*, *then setting* $X_n = \tilde{X}_j$ *for* $n \in \mathscr{A}_j$ *where the* $\tilde{X}_j$ *are i.i.d. with distribution* $v$, *independent of* $\Pi$.

Following Pitman (1996b), 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 $v$. 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 $j$th 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 $v$ as it appears in the sampling process. In this interpretation the random partition $\Pi$ 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 $\Pi$ in a sequence of exchangeable random variables $(X_n)$. As shown by Aldous (1985), 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 $v$ on $S$ is of no importance: one may as well take $S = [0,1]$ with Borel sets and $v$ the uniform distribution on $[0,1]$.

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

**Theorem 3** (Pitman, 1996b)**.** *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 $X_1,\dots,X_n$, as displayed in* (3). *Then $F_n$ converges in total variation norm almost surely as $n \to \infty$ to the random measure*

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

*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 $\Pi$ generated by $(X_n)$ via formulae described in Pitman (1996b). See Pitman (1995, 1996b) regarding the conditional distribution of $\Pi$ given the sequence $(\tilde{P}_j)$, which is the same for all species sampling sequences. See Pitman (1996b) regarding the conditional distribution of $F$ given $(X_1,\dots,X_n)$. Theorem 3 yields also:

**Corollary 4** (Pitman, 1996b). *A sequence $(X_n)$ is a species sampling sequence with marginal distributions equal to $v$ if and only if $(X_n)$ is conditionally i.i.d. $(F)$ given some random probability distribution $F$ on $S$ of the form*

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

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

**Example.** *The two-parameter model* (*Pitman*, 1995). Consider the prediction rule (3) defined by some diffuse measure $v$ and

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

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

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

or

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

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

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

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 $(X_1, X_2, \dots)$ defined by 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 (Ewens, 1972; Antoniak, 1974; Ewens and Tavaré, 1995) 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 $(P_1,\dots,P_m)$ has a symmetric Dirichlet distribution with $m$ parameters equal to $\kappa$, and the $\hat{X}_i$ are i.i.d. with distribution $v$. This is Fisher's model for species sampling (Fisher et al., 1943) with $m$ species identified by i.i.d.$(v)$ tags. See Kerov (1995); Mekjian and Chase (1997); Pitman (1996a, 1997a,b); Pitman and Yor (1997); and Zabell (1997) 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 $(X_1,\dots,X_n)$, and $v$ a diffuse measure on $S$. Let $\Pi_n$ be the partition of $\{1,\dots,n\}$ generated by $X_1,\dots,X_n$. 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 $\Pi_n$, it suffices to show that conditionally given $\Pi$, the partition of all positive integers generated by $(X_n)$, the random variables $\tilde{X}_j$ for $j=1,2,\dots$ are independent and identically distributed according to $v$. The following lemma provides a convenient reformulation of this condition:

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

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

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

$$\mathbb{P}(\Pi_n=\pi;\tilde{X}_j\in B_j \text{ all } j\leqslant k)=v(B_i)\,\mathbb{P}(\Pi_n=\pi;\tilde{X}_j\in B_j \text{ all } j\leqslant k, j\neq i). \tag{18}$$

If $\pi$ 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 $1=\min A_1^\pi<\min A_2^\pi<\cdots<\min A_k^\pi$. 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 $v$ is diffuse that (18) holds if $i=k$ and $\#A_k^\pi=1$. The assumed exchangeablility of $(X_n)$ then yields (18) for any $1\leqslant i\leqslant k$ with $\#A_i^\pi=1$.

Now consider the inductive hypothesis, call it $H_m$, that (18) holds for all choices of $1\leqslant k\leqslant n,\pi,B_j$, $1\leqslant j\leqslant k$ and $1\leqslant i\leqslant k$ 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 $\pi'$ 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 $\pi'$ and for $i=1$.

Fix such a $\pi'$ 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 $v$-measure. Write $\pi=\{A_1^\pi,\dots,A_k^\pi\}$ for the restriction of $\pi'$ to $\{1,\dots,n\}$. For $l=1,\dots,k$, write $\pi^l$ for the partition $\{A_1^\pi,\dots,A_l^\pi\cup\{n+1\},\dots,A_k^\pi\}$ of $\{1,\dots,n+1\}$. Note that $\pi'=\pi^1$. Write $\pi^{k+1}$ for the partition $\{A_1^\pi,\dots,A_k^\pi,\{n+1\}\}$ of $\{1,\dots,n+1\}$. By $H_m$, for each $l=2,\dots,k+1$,

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

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

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

The identity

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

now implies that

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

which is the identity required to establish $H_{m+1}$. $\square$

## References


Aldous, D.J., 1985. Exchangeability and related topics. In: P.L. Hennequin (Ed.), École d'Été de Probabilités de Saint-Flour XII, Springer Lecture Notes in Mathematics, Vol. 1117. Springer, Berlin.

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

Blackwell, D., MacQueen, J.B., 1973. Ferguson distributions via Pólya urn schemes. Ann. Statist. 1, 353–355.

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

Ewens, W.J., Tavaré, S., 1995. The Ewens sampling formula. In: N.S. Johnson, S. Kotz, N. Balakrishnan (Eds.), Multivariate Discrete Distributions, to apper.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209–230.

Fisher, R.A., Corbet, A.S., Williams, C.B., 1943. 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.

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

Kingman, J.F.C., 1978. The representation of partition structures. J. London Math. Soc. 18, 374–380.

Kingman, J.F.C., 1982. The coalescent. Stochast. Process. Appl. 13, 235–248.

Mekjian, A.Z., Chase, K.C., 1997. Disordered systems, power laws and random processes. Phys. Lett. A 229, 340–346.

Pitman, J., 1995. Exchangeable and partially exchangeable random partitions. Probab. Theory Rel. Fields 102, 145–158.

Pitman, J., 1996a. Random discrete distributions invariant under size-biased permutation. Adv. Appl. Probab. 28, 525–539.

Pitman, J., 1996b. Some developments of the Blackwell–MacQueen urn scheme. In: Ferguson, T.S. et al. (Eds.), Statistics, Probability and Game Theory; Papers in honor of David Blackwell, Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California. vol. 30. pp. 245–267.

Pitman, J., 1997a. Coalescents with multiple collisions. Technical Report 495, Department of Statistics, U.C. Berkeley. Available via http://www.stat.berkeley.edu/users/pitman. To appear in Ann. Probab.

Pitman, J., 1997b. Partition structures derived from Brownian motion and stable subordinators. Bernoulli 3, 79–96.

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

Zabell, S.L., 1997. The continuum of inductive methods revisited. In: Earman, J., Norton, J.D. (Eds.), The Cosmos of Science. Pittsburgh–Konstanz Series in the Philosophy and History of Science, University of Pittsburgh Press/Universitätsverlag Konstanz, pp. 351–385.