De FINETTI'S GENERALIZATIONS OF EXCHANGEABILITY

BY

PERSI DIACONIS and DAVID FREEDMAN

TECHNICAL REPORT NO. 109

MARCH 1978

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

De FINETTI'S GENERALIZATIONS OF EXCHANGEABILITY

by

Persi Diaconis
Stanford University

and

David Freedman
University of California - Berkeley

TECHNICAL REPORT NO. 109

MARCH 1978

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

De FINETTI'S GENERALIZATIONS OF EXCHANGEABILITY

Persi Diaconis and David Freedman

1. Introduction

De Finetti has written about partial exchangeability in his articles of 1938 (translated in this volume) and 1959, Section 9.6.2 (translated in de Finetti 1974). Both treatments are rich sources of ideas which seem to take many readings to digest. This article gives examples of partial exchangeability that we understand well enough to put into crisp mathematical terms. All results in Sections 2-5 involve random quantities taking only two values. We follow de Finetti by giving results for finite as well as infinite sequences. In Section 2 we review exchangeability. In Section 3 we present examples of partially exchangeable sequences--2×2 tables and Markov chains--and give general definitions. In Section 4 a finite form of de Finetti's theorem is presented. Section 5 gives some infinite versions of de Finetti's theorem and a counter-example which shows that the straightforward generalization from the exchangeable case sketched by de Finetti is not possible. The last section contains comments about the practical implications of partial exchangeability. We also discuss the closely related work of P. Martin-Löf on repetitive structures.

2. Exchangeability

We begin with the case of exchangeability. Consider the following experiment: A coin is spun on a table. We will denote heads by 1 and tails by 0. The experiment is to be repeated ten times. A subjectivist believes he can assign probabilities to the $2^{10} = 1024$ possible outcomes by looking at the coin and thinking about what he knows. There are no apriori restrictions on the probabilities assigned save that they are non-negative and sum to 1. Even with so simple a problem, assigning over a thousand probabilities is not a simple task. De Finetti has called attention to exchangeability as a possible simplifying assumption. With an exchangeable probability two sequences of length 10 with the same number of zeros are assigned the same probability. Thus, the probability assignment is symmetric, or invariant under changes in order. Another way to say this is that only the number of ones in the ten trials matters, not the location of the ones. If believed, the symmetry assumption reduces the number of probabilities to be assigned from 1024 to 11--the probability of no ones through the probability of ten ones.

It is useful to single out certain extreme exchangeable probability assignments. Though it is unrealistic, we might be sure there will be exactly one head in the ten spins. The assumption of exchangeability forces each of the ten possible sequences with a single one in it to be equally likely. The distribution is just like the outcome of ten draws without replacement from an urn with one ball marked 1 and nine balls marked 0--the ball marked 1 could be drawn at any stage and must be drawn at some time. There are 11 such extremal urns corresponding to sure knowledge of exactly i one outcomes in the ten draws where i is a fixed number between zero and ten. The first form of de Finetti's theorem follows from these observations:

(1) <u>Finite form of de Finetti's theorem</u>

Every exchangeable probability assignment on sequences of length N is an unique mixture of draws without replacement from the N + 1 extremal urns.

By a mixture of urns we simply mean a probability assignment over the N + 1 possible urns. Any exchangeable assignment on sequences of length N can be realized by first choosing an urn and then drawing without replacement until the urn is empty.

This form of the result has been given by de Finetti (1938), (1974), and by many other writers on the subject: Diaconis (1977), Ericson (1973), Heath and Sudderth (1976), Hewitt and Savage (1955), and Kendall (1967).

The extremal urns, representing certain knowledge of the total number of ones, seem like unnatural probability assignments in most cases. While such situations arise (for example, when drawing a sample from a finite population), the more usual situation is that of the coin. While we are only considering ten spins, in principle it seems possible to extend the ten to an arbitrarily large number. In this case there is a stronger form of (1) which restricts the probability assignment within the class of exchangeable assignments.

(2) Infinite form of de Finetti's theorem

Every exchangeable probability assignment which can be extended to a probability assignment on infinite sequences of zeros and ones has the following representation:

$$\text{(3)} \qquad P(\text{j ones in k trials}) = \binom{k}{j} \int_0^1 p^j (1-p)^{k-j}\, d\mu(p)$$

for a uniquely determined measure μ.

Note that (3) holds for every k with the same μ. Equation (3) says that probability assignments which can be extended have the special form of mixtures of independent and identically distributed coin tossing. We have more to say about this interpretation in Remark 1 of Section 6.

Not all exchangeable measures on sequences of length k can be extended to exchangeable sequences of length $n > k$. For example, sampling without replacement from an urn with k balls in it cannot be extended to $k + 1$ trials. Necessary and sufficient conditions for extension are found in de Finetti (1969), Crisma (1971), and Diaconis (1977). The requirement that an exchangeable sequence of length k be infinitely extendable seems out of keeping with de Finetti's general program of restricting attention to finite samples. An appropriate finite version of (3) is given by Diaconis and Freedman (1978a). We show that if an exchangeable probability on sequences of length k is extendable to an exchangeable probability on sequences of length $n > k$, then (3) almost holds in the sense that there is a measure μ such that for any set $A \subset \{0,1,2,\ldots,k\}$,

$$(4) \qquad \left| P\left\{ \begin{matrix} \text{number of ones in} \\ \text{k trials is in A} \end{matrix} \right\} - \sum_{j \in A} \binom{k}{j} \int_0^1 p^j (1-p)^{k-j} \, d\mu(p) \right| \leq \frac{2k}{n} \quad ,$$

uniformly in n, k, and A.

For example, it is easy to imagine the spins of the coin as the first ten spins in a series of 1000 spins. This yields a bound of .02 on the right side of (4).

Both of the results (3) and (4) imply that for many practical purposes instead of specifying a probability assignment on the number of ones in n trials it is equivalent to specify a prior measure μ on the unit interval. Much of de Finetti's discussion in his papers on partial exchangeability is devoted to reasonable choices of μ. The paper by A. Bruno (1964) is also devoted to this important problem. We will not discuss the choice of a prior further but rather restrict attention to generalization of (2), (3), and (4) to partially exchangeable probability assignments.

3. Examples of Partial Exchangeability

In many situations exchangeability is not believable or does not permit incorporation of other relevant data. Here are two examples which will be discussed further in Sections 4 and 5.

Example a (2×2 tables)

Consider zero/one outcomes in a medical experiment with n subjects. We are told each subject's sex and if each subject was given a treatment or was in a control group. A reasonable symmetry assumption is partial exchangeability: regard all the treated males as exchangeable with one another but not with the subjects in the other three categories. Thus, two sequences of zeros and ones of length n which had the same number of ones in each of the four categories would be assigned the same probability. For example, if n = 10, each of the three sequences below must be assigned the same probability.

Trial	1	2	3	4	5	6	7	8	9	10
Sex	M	M	M	F	F	M	M	F	F	F
Treatment/Control	T	T	T	T	T	C	C	C	C	C
1	1	0	1	0	1	1	1	0	1	0
2	0	1	1	0	1	1	1	0	0	1
3	0	1	1	1	0	1	1	1	0	0

In this example, there are three treated males, two treated females, two control males, and three control females. The data from each of the three sequences can be summarized in a 2×2 table which records the number of one outcomes in each group. Each of the three sequences leads to the same matrix T.

$$T = \begin{array}{c|c|c|} & T & C \\ \hline M & i & j \\ \hline F & k & \ell \\ \hline \end{array} = \begin{array}{c|c|c|} & T & C \\ \hline M & 2 & 2 \\ \hline F & 1 & 1 \\ \hline \end{array} \tag{5}$$

Example b (Markov dependence)

Consider an experiment in which a thumbtack is placed on the floor and given an energetic flick with the fingers. We record a one if the tack lands point upward and a zero if it lands point to the floor. For simplicity suppose the tack starts point to the floor. If after each trial the tack were reset to be point to the floor, exchangeability might be a tenable assumption. If each flick of the tack was given from the position in which the tack just landed, then the result of each trial may depend on the result of the previous trial. For example, there is some chance the tack will slide across the floor without ever turning. It does not seem reasonable to think of a trial depending on what happened two or more trials before. A natural notion of symmetry here is to say that if two sequences of zeros and ones of length n which both begin with zero have the same number of transitions: zero to zero, zero to one, one to zero, and one to one, they should both be assigned the same probability. For example, if n = 10, any of the following three sequences would be assigned the same probability.

Trial	1	2	3	4	5	6	7	8	9	10
1	0	1	0	1	1	0	0	1	0	1
2	0	1	0	1	0	0	1	1	0	1
3	0	0	1	0	1	0	1	0	1	1

Each sequence begins with a zero and has the same transition matrix

$$(6) \qquad T = \text{from} \begin{array}{c c} & \text{to} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1 & 4 \\ 3 & 1 \end{bmatrix} \end{array} = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix}$$

It turns out that there are 16 different sequences starting with zero that have this transition matrix T.

A general definition of partial exchangeability that includes these examples involves the notion of a Statistic: A function from the sequences of length n into a set X. A probability assignment P on sequences of length n is *partially exchangeable for a statistic T* if

$$T(x) = T(y) \quad \text{implies} \quad P(x) = P(y)$$

where x and y are sequences of length n. Freedman (1962) and Lauritzen (1974) have said that P was summarized by T in this situation.

In the case of exchangeability the statistic T is the number of ones. Two sequences with the same number of ones get assigned the same probability by an exchangeable probability. This definition is equivalent to the usual one of permutation invariance. In the case of 2×2 tables the statistic T is the 2×2 table (5). In the Markov example the statistic T is the matrix of transition counts (6) along with the outcome of the first trial.

Such examples can easily be combined and extended. For instance, two stage Markov dependence with additional information such as which experimenter reported the trial and the time of day being given. De Finetti (1938), (1974, Section 9.6.2), Martin-Löf (1970), (1974), and Diaconis and Freedman (1978a,b,c) give further examples.

4. Finite Forms of de Finetti's Theorem

There is a simple analog of Theorem (1) for partially exchangeable sequences of length n. We will write 2^n for the set of sequences of zeros and ones of length n. Let $T : 2^n \to X$ be a statistic taking values $t_1, t_2, \ldots, t_k$. Let $S_i = \{x \varepsilon 2^n : T(x) = t_i\}$ and suppose S_i contains n_i elements. Let P_i be the probability assignment on 2^n which picks a sequence $x \varepsilon 2^n$ by choosing an x from S_i uniformly--e.g., with probability $1/n_i$. Then P_i is partially exchangeable with respect to T. In terms of these definitions we now state

(7) Finite form of de Finetti's theorem

Every probability assignment P on 2^n which is partially exchangeable with respect to T is a unique mixture of the extreme measures P_i. The mixing weights are $w_i = P(T(x)) = t_i$.

In the language of convex sets, the set of partially exchangeable probabilities with respect to T forms a simplex with extreme points P_i. Theorem (7) seems trivial but in practice a more explicit description of the extreme measures P_i--like the urns in (1)--can be difficult. We now explore this for the examples of Section 3.

(8) Example a (2×2 tables)

Suppose we know there are: a treated males, b untreated males, c treated females, and d untreated females with $a + b + c + d = n$. The sufficient statistic is the 2×2 table with entries which are the number of ones in each of the four groups.

$$T = \begin{array}{c} \\ M \\ F \end{array}\!\!\begin{array}{c} \begin{array}{cc} T & C \end{array} \\ \begin{pmatrix} i & j \\ k & \ell \end{pmatrix} \end{array} \quad \text{where} \quad \begin{array}{l} 0 \le i \le a,\ 0 \le j \le b \\ 0 \le k \le c,\ 0 \le \ell \le d \end{array} .$$

There are $(a+1) \times (b+1) \times (c+1) \times (d+1)$ possible values of the sufficient statistic T. The extreme partially exchangeable probability can be thought of as follows: Fix a possible value of T. Make up four urns. In the first urn, labeled TM, put i balls marked one and $a-i$ balls marked zero. Similarly, construct urns labeled UM, TF, and UF. To generate a sequence of length n given the labels (Sex and Treated/Control) draw a ball without replacement from the appropriate urn. This scheme generates all sequences $x \varepsilon 2^n$ with $T(x) = \begin{pmatrix} i & j \\ k & \ell \end{pmatrix}$ equiprobably. Theorem (7) says that any partially exchangeable probability assignment on 2^n is a unique measure of such urn schemes. If a,b,c, and d all tend to infinity, the binomial approximation to the hypergeometric distribution will lead to the appropriate infinite version of de Finetti's theorem.

(9) <u>Example b (Markov chains)</u>

For simplicity, assume we have a sequence of length n + 1 that begins with a zero. The sufficient statistic is $T = \begin{pmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{pmatrix}$ where t_{ij} is the number of i to j transitions. A counting argument shows that there are $\binom{n}{2} + 1$ different values of T possible. Here is an urn model which generates all sequences of length n + 1 with a fixed transition matrix T equiprobably. Form two urns U_0 and U_1 as follows: Put t_{ij} balls marked j into urn U_i. It is now necessary to make an adjustment to make sure the process doesn't run into trouble. There are two cases possible:

Case 1. If $t_{01} = t_{10}$, remove a zero from U_1 and a one from U_0.

Case 2. If $t_{01} = t_{10} + 1$, remove a one from U_0.

To generate a sequence of length $n + 1$, let $X_1 = 0$. Let X_2 be the result of a draw from U_{X_1}; and, in general, let X_i be the result of a draw without replacement from urn $U_{X_{i-1}}$. If the sequence generated ever forces a draw from an empty urn, make a forced transition to the other urn. The adjustment made above guarantees that such a forced jump can only be made once from either urn and that the process generates all sequences of length $n + 1$ that start with zero and have transition matrix T with the same probability. Theorem (7) says that every probability assignment on sequences of length $n + 1$ which is partially exchangeable for the transition matrix T is a unique mixture of the $\binom{n}{2} + 1$ different urn processes described above. Again, the binomial approximation to the hypergeometric will lead to an infinite form of de Finetti's theorem in certain cases. This is further discussed in Sections 5 and 6.

Determining when a simple urn model can be found to describe the extreme partially exchangeable probabilities may be difficult. For instance, we do not know how to extend the urn model of Example b (Markov chains) to processes taking three values.

5. Infinite Forms of De Finetti's Theorem

Let us examine the results and problems for infinite sequences in the two examples.

(10) Example a (2×2 tables)

Let $X_1, X_2, X_3, \ldots$ be an infinite sequence of random variables each taking values 0 or 1. Suppose each trial i is labeled Male/Female and Treated/Untreated and that the number of labels in each of the four possible categories (M,U), (M,T), (F,U), (F,T) is infinite. Suppose that for each n the distribution of $X_1, X_2, \ldots, X_n$ is partially exchangeable with respect to the sufficient statistic that counts the number of zeros and the number of ones for each label. Thus, $T_n = (a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4)$ where, for example, a_1 is the number of ones labeled (M,U), b_1 is the number of zeros labeled (M,U), a_2 is the number of ones labeled (M,T), b_2 is the number of zeros labeled (M,T) and so on. Then there is a unique probability distribution μ such that for every n and each sequence $x_1, x_2, \ldots, x_n$

$$(11) \qquad P(X_1 = x_1, \ldots, X_n = x_n) = \int \prod_{i=1}^{4} p_i^{a_i} (1 - p_i)^{b_i} \, d\mu(p_1, p_2, p_3, p_4) \quad .$$

where a_i, b_i are the values of $T_n(x_1, x_2, \ldots, x_n)$. Details of a proof for (11) and a finite version paralleling (4) are given in Diaconis and Freedman (1978a). Basically, the result follows by passing to the limit in the urn model of Section 4.

(12) Example b (Markov chains)

Let $X_1, X_2, \ldots$ be an infinite sequence of random variables each taking values 0 or 1. For simplicity assume $X_1 = 0$.

Suppose that for each n the joint distribution of $X_1,\ldots,X_n$ is partially exchangeable with respect to the matrix of transition counts. Suppose that the following recurrence condition is satisfied

$$(13) \qquad P(X_1 = 0 \text{ infinitely often}) = 1 \quad .$$

Then there is a unique probability distribution μ such that for every n, and each sequence $x_2,x_3,\ldots,x_n$ of zeros and ones

$$(14) \qquad \begin{aligned} &P(X_1 = 0,\ X_2 = x_2,\ldots,X_n = x_n) \\ &= \int p_{11}^{t_{11}}(1-p_{11})^{t_{10}}\, p_{00}^{t_{00}}(1-p_{00})^{t_{01}}\, d\mu(p_{11},p_{00}) \quad . \end{aligned}$$

where t_{ij} are the four entries of the transition matrix of $x_1,x_2,\ldots,x_n$.

De Finetti appears to state (pp. 218-219 of de Finetti (1974)) that the representation (14) is valid for every partially exchangeable probability assignment in this case. Here is an example to show that the representation (14) need not hold in the absence of the recurrence condition (13). Consider the probability assignment which goes 001111111... (all ones after two zeros) with probability one. This assignment is partially exchangeable.

Write p_k for the probability that the last zero occurs at trial k ($k = 1,2,3,\ldots$). For a mixture of Markov chains the numbers p_k can be represented as

$$(15) \qquad p_k = c \int_0^1 p_{00}^k(1-p_{00})\, d\mu(p_{00}) \qquad k = 1,2,\ldots \quad ,$$

where c is the probability mass the mixing distribution puts on

$p_{11} = 1$. The representation (15) implies that the numbers p_k are completely monotone (Feller 1971, Chapter 7). In particular, the p_k are decreasing. For the example 001111111..., $p_1 = 0$, $p_2 = 1$, $p_3 = p_4 = \dots = 0$. So this example doesn't have a representation as a mixture of Markov chains. A detailed discussion of which partially exchangeable assignments are mixtures of Markov chains is in Diaconis and Freedman (1978b).

When can we hope for representations like (3), (11), and (14) in terms of averages over a naturally constructed "parameter space?" De Finetti sketches out what appears to be a general theory in terms of what he calls the type of an observation. In de Finetti (1938) he only gives examples which generalize our Example a. Here things are simple. In the example there are four types of observations depending on the labels (M,U), (M,T), (F,U), (F,T). In general, for each i we observe the value of another variable containing information like sex, time of day, and so on. An analog of (11) clearly holds in these cases. In de Finetti (1974), Section 9.6.2, the type of an observation is allowed to depend on the outcome of past observations like in our Example b. In this example there are three types of observations--observations X_i that follow a 0 are of type zero; observations X_i that follow a 1 are of type one; and the first observation X_1 is of type 2. For the original case of exchangeability there is only one type.

Such examples can easily be combined and extended, but a useful general definition of type is not known to us. One use for a definition of type is to say when the extreme point representation can be realized as an integral over a finite dimensional parameter space. We now construct a contrived example in which the number in each of two types tends to infinity, but the representation is not of the form expected.

A computer scientist is watching a sequence of zeros and ones, $X_1, X_2, X_3, \ldots$. He thinks trials $1,3,5,7,\ldots$ may not be random but be generated by reading off some fixed reference sequence x of zeros and ones or the compliment $\bar{x}$ (the compliment switches all ones for zeros and zeros for ones). He thus keeps track of two types of X_i. If i is odd and all the preceding X_j with j odd lead to a sequence which matches x or $\bar{x}$, he calls i of type 1. Otherwise, X_i is of type 2. He counts the number of ones and zeros in each type. Suppose that the reference sequence x is constructed so that there are increasingly longer runs of zeros and ones such that

(16) the proportion of ones in the first n places of x does not tend to a limit as n tends to infinity,

(17) there are always more ones than zeros in the first n places of x.

Condition (16) insures that the odd positions cannot be described as any mixture of finite parameter Markov chain or the like while condition (17) insures that he unambiguously knows if the odd positions come from x, $\bar{x}$ or a coin tossing procedure. Here the process is a mixture

$$w_1P_1 + w_2P_2 + w_3P_3 \tag{18}$$

where each of P_1, P_2, and P_3 puts a (possibly different) mixture of coin tossing on the even positions, P_1 yields x at the odd positions, P_2 yields $\bar{x}$ at the odd positions, and P_3 yields the same mixture of coin tossing at the odd positions as it does at the even positions. The representation (18) uses the number of ones and zeros of the first type in quite a different way from the counts for the second type.

Some cases, when parametric representation is possible, have been determined by Martin-Löf (1970, 1974) and Lauritzen (1976). A general theory and more examples are in Diaconis and Freedman (1978c).

6. Concluding Remarks

1. Some Bayesians are willing to talk about "tossing a coin with unknown p." For them, de Finetti's theorem can be interpreted as follows: If a sequence of events is exchangeable, then it is like the successive tosses of a p-coin with unknown p. Other Bayesians do not accept the idea of p coins with unknown p: de Finetti is a prime example. Writers on subjective probability have suggested that de Finetti's theorem bridges the gap between the two positions. We have trouble with this synthesis and the following quote (de Finetti 1976) indicates that de Finetti has reservations about it:

> "The sensational effect of this concept (which went well beyond its intrinsic meaning) is described as follows in

Kyburg and Smokler's preface to the collection *Studies in subjective probability* which they edited (pp. 13-14).

'In a certain sense the most important concept in the subjective theory is that of 'exchangeable events.' Until this was introduced (by de Finetti (1931)) the subjectivistic theory of probability remained little more than a philosophical curiosity. None of those for whom the theory of probability was a matter of knowledge or application paid much attention to it. But, with the introduction of the concept of '*equivalence*' or '*symmetry*' or '*exchangeability*', as it is now called, a way was discovered to link the notion of subjective probability with the classical problem of statistical inference.' "

It does not seem to us that the theorem explains the ideas of coins with unknown p. The theorem does show that calculations involving the notion of unknown p lead to exchangeable probability assignments and so are interpretable in de Finetti's framework. Conversely, if a subjectivist has an exchangeable prior opinion about an infinite sequence of events, then his opinion is represented by a probability μ on $[0,1]$ through formula (3). Having seen $n + m$ events with n successes and m failures the conditional chance that the next event will be a success is

$$\frac{\int p^{n+1} (1-p)^{n} \, d\mu(p)}{\int p^{n} (1-p)^{n} \, d\mu(p)} .$$

It is this calculation which the theorem interprets not the notion of "unknown p."

2. The exchangeable form of de Finetti's theorem (1) is a useful computational device for specifying probability assignments. The result is more complicated and much less useful in the case of real valued variables. Here there is no natural notion of a

parameter p. Instead, de Finetti's theorem says real valued exchangeable variables, $\{X_i\}_{i=1}^{\infty}$, are described as follows: There is a prior measure π on the space of all probability measures on the real line such that $P(X_1 \in A_1, \ldots, X_n \in A_n) = \int \prod_{i=1}^{n} p(X_i \in A_i) d\pi(p)$. The space of all probability measures on the real line is so large that no one claims to be able to describe a personally meaningful distribution on this space. Ferguson (1974) contains examples of the various attempts to choose such a prior. Thus, for real valued variables de Finetti's theorem is far from an explanation of the type of parametric estimation Bayesian statisticians from Bayes and Laplace to Lindley have been using in real statistical problems.

Useful supplements to exchangeability are needed. Here are some examples of possible supplements adapted from Freedman (1963), (1962).

Example: (Scale mixtures of normal variables)

When can a sequence of real valued variables $\{X_i\}$ $1 \leq i < \infty$ be represented as a scale mixture of normal variables

$$(19) \qquad P(X_1 \leq t_1, \ldots, X_n \leq t_n) = \int_0^{\infty} \prod_{i=1}^{n} \Phi(\sigma t_i) d\pi(\sigma)$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2}\, dt$?

In Freedman (1963) it is shown that a necessary and sufficient condition for (19) to hold is that for each n the joint distribution of $X_1, X_2, \ldots, X_n$ be rotationally symmetric. This result is related to the derivation of Maxwell's distribution for velocity in a

Monoatomic Ideal Gas (Khinchin 1949, Chapter VI).

<u>Example: (Poisson distribution)</u>

Let X_i $1 \leq i < \infty$ take integer values. In Freedman (1962) it is shown that a necessary and sufficient condition for X_i to have a representation as a mixture of Poisson variables,

$$P(X_1 = a_1, \ldots, X_n = a_n) = \int_0^\infty \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{a_i}}{a_i!} \, d\pi(\lambda)$$

is as follows: For every n the joint distribution of $X_1, X_2, \ldots, X_n$ given $S_n = \Sigma_{i=1}^n X_i$ must be multinomial; like the joint distribution of S_n balls dropped at random into n boxes.

Many further examples and some general theory are given in Diaconis and Freedman (1978c).

3. De Finetti's generalizations of partial exchangeability are closely connected to recent work of P. Martin-Löf. Martin-Löf does not seem to work in a Bayesian context, rather he takes the notion of sufficient statistic as basic and from this constructs the joint distribution of the process in much the same way as de Finetti. Martin-Löf connects the conditional distribution of the process with the microcannonical distributions of statistical mechanics.

Martin-Löf's work appears most clearly spelled out in a set of mimeographed lecture notes (Martin-Löf 1970), unfortunately available only in Swedish. A technical treatment of part of this may be found in Martin-Löf (1974). Further discussion of Martin-Löf ideas are in Lauritzen (1973), (1975) and the last third of Tjur (1974).

These references contain many new examples of partially exchangeable processes and their extreme points. We have tried to connect Martin-Löf's treatment to the general version of de Finetti's theorem we have derived in Diaconis and Freedman (1978c).

BIBLIOGRAPHY

Bruno, A. (1964). On the notion of partial exchangeability. Giorn. Ist. It. Attuari, 27, 174-196. Translated in Chapter 10 of de Finetti (1974).

Crisma, L. (1971). Sulla proseguibilità di processi scambiabili, Rend. Matem. Trieste 3, 96-124.

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeabilities, Synthese 36, 271-281.

Diaconis, P. and Freedman, D. (1978a). Finite exchangeable probability. Submitted to Ann. Prob.

Diaconis, P. and Freedman, D. (1978b). De Finetti's theorem for Markov chains. Submitted to Ann. Prob.

Diaconis, P. and Freedman, D. (1978c). Partial exchangeability and sufficiency. Submitted to Ann. Stat.

Ericson, W. (1973). A Bayesian approach to two-stage sampling, Technical Report No. 26, Department of Statistics, University of Michigan, Ann Arbor.

Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Wiley, New York.

Ferguson, T. (1974). Prior distributions on spaces of probability measures, Ann. Stat. 2, 615-629.

De Finetti, B. (1937). La prevision: ses Lois Logiques, ses sources subjectives, Ann. Inst. H. Poincaré (Paris), 7, 1-68. Translated in Kyburg and Smokler (1964).

De Finetti, B. (1938). Sur la condition d'equivalence partielle (Colloque Geneve, 1937). Act. Sci. Ind. 739, 5-18, Herman, Paris. Translated in this volume.

De Finetti, B. (1959). La probabilita e la statistica nei rapporti con l'induzione, secondo i dwersi punti di vista. In Corso C.I.M.E. su Induzione e Statistica, Cremonese, Rome. Translated as Chapter 9 of de Finetti 1974.

De Finetti, B. (1969). Sulla proseguibilità di processi aleatori scambiabili, Rend. Matem. Trieste, 1, 53-67.

De Finetti, B. (1974). Probability, Induction and Statistics, Wiley, New York.

De Finetti, B. (1976). Probability: Beware of falsifications!, Scientia 111, 283-303.

Freedman, D. (1962a). Mixtures of Markov processes, Ann. Math. Stat. 33, 114-118.

Freedman, D. (1962b). Invariants under mixing which generalize de Finetti's theorem, Ann. Math. Stat. 33, 916-923.

Heath, D. and Sudderth, W. (1976). De Finetti's theorem on exchangeable variables, American Statistician 30, 188-189.

Kendal, D. (1967). On finite and infinite sequences of exchangeable events, Studia Sci. Math. Hung. 2, 319-327.

Khinchin, A. (1949). Mathematical Foundations of Statistical Mechanics, Dover, New York.

Kyburg, H.E. and Smokler, H.E. (1964). Studies in Subjective Probability, Wiley, New York.

Lauritzen, S.L. (1973). On the interrelationships among sufficiency, total sufficiency, and some related concepts, Institute of Math. Stat., University of Copenhagen.

Lauritzen, S.L. (1975). General exponential models for discrete observations, Scand. J. Statist. 2, 23-33.

Martin-Löf, P. (1970). Statistiska Modeller, Anteckningar frän seminarer läsaret, 1969-70, uturbetade of R. Sundberg, Institutionen för Matematisk Statistik, Stockholms Universitet.

Martin-Löf, P. (1973). Repetitive structures and the relation between canonical and microcanonical distributions in statistics and statistical mechanics. In Proc. Conference on Foundational Questions in Statistical Inference, Department of Theoretical Statistics, University of Aarhus--Barndorf-Nielsen, P. Bluesild, G. Schon (eds.).

Tjur, T. (1974). Conditional probability distributions, Lecture Notes 2, Institute of Mathematical Statistics, University of Copenhagen.