J. R. Statist. Soc. B (1993)
55, *No.* 2, *pp.* 423–433

Parametric Models for A_n: Splitting Processes and Mixtures

By BRUCE M. HILL†

University of Michigan Ann Arbor, USA

[Received March 1990. Final revision January 1992]

SUMMARY

A class of parametric models, called splitting processes, is defined, by using de Finetti's concept of adherent mass. Such splitting processes give rise to complex mixtures of distributions. It is proved that the nonparametric Bayesian predictive procedure A_n, of Hill, holds exactly for a member of this class called a nested splitting process. The connection between A_n and the Dirichlet process is stated and proved. A multivariate version of A_n, based on splitting processes, is proposed.

Keywords: BAYESIAN NONPARAMETRIC STATISTICS; PREDICTION; SAMPLING FROM FINITE POPULATIONS

1. INTRODUCTION

A_n and H_n were proposed by Hill (1968, 1988) for prediction in the case of extremely vague *a priori* knowledge about the form of the underlying distribution. See Aitchison and Dunsmore (1975) for the theory underlying prediction. Let x_i, for $i=1, \ldots, n$, be the data values obtained in sampling from a finite population, and let the $x_{(i)}$ be their ordered values in increasing order of magnitude. Let X_i be the corresponding pre-data random quantities, so that the data consist of the realized values, $X_i = x_i$, for $i=1, \ldots, n$. In this paper A_n is defined as follows.

(a) The observable random quantities $X_1, \ldots, X_n$ are exchangeable. By this we mean that for each $k \leqslant n$ the joint distribution function of $X_{j_1}, \ldots, X_{j_k}$ does not depend on the (distinct) indices $j_1, \ldots, j_k$. (In Hill (1968, 1988) exchangeability was not included in the *definition* of A_n, to include more general situations, such as partial exchangeability.)

(b) Ties have probability 0.

(c) Given the data x_i, $i=1, \ldots, n$, the probability that the next observation falls in the open interval $I_i = (x_{(i)}, x_{(i+1)})$ is $1/(n+1)$, for each $i=0, \ldots, n$. By definition, $x_{(0)} = -\infty$, and $x_{(n+1)} = +\infty$, unless explicitly stated otherwise.

The weak prior knowledge underlying A_n might be described in terms of 'data on a rubbery scale'. For example, it is known that A_n can hold exactly when the observations are only simply ordered, i.e. when there is only an ordinal scale of measurement so that only relative magnitudes can be determined, as discussed in Hill (1968), p. 678, and Hill (1988), p. 224; this suggests that it might hold approximately even when there is something more than an ordinal scale of measurement. Earlier, Fisher (1939, 1948) had succinctly suggested a version of A_n from the fiducial point of view. Fisher (1948), p. 210, took as pivotal quantities the proportions of the unknown population exceeded by the largest, second-largest, etc., order statistic of the sample

†*Address for correspondence*: 1657 Glenwood, Ann Arbor, MI 48104, USA.

and derived the joint fiducial distribution of successive differences to be uniform on the appropriate subset of the unit cube. Hill (1968), p. 686, then proved that such a specification of the fiducial distribution for all n was equivalent to A_n for all n. Dempster (1963) elaborated and made more precise this fiducial insight of Fisher. Hill (1980a) showed that A_n yields a robust form of Bayesian inference and provides approximations to some real situations that arise in sampling from finite populations, such as in multidimensional contingency tables and the species sampling problem. Hill (1988) gave a new subjective Bayesian argument for A_n and, because of the minimal and realistic assumptions underlying it, proposed A_n as a possible solution to the problem of induction, as defined, for example, by Hume (1748). Lenk (1984) discusses the relationship between A_n and use of the empirical distribution function.

In this paper I shall attempt to provide further justification for A_n, showing that it arises from simple parametric models, called splitting processes. It can ordinarily be viewed as appropriate when the data arise from the process of sampling from *finite* populations that can be represented as complex mixtures of distributions. Although Hill (1968) proved that A_n cannot hold exactly for countably additive distributions, for any n, it is known from Jeffreys (1961), p. 171, that A_1 and A_2 hold for conventional parametric models with the usual improper *a priori* distributions for the parameters, and from the work of Lane and Sudderth (1978, 1984) that A_n is coherent in the sense of de Finetti for all n. Because of its practical importance for Bayesian statistics, it is essential also to understand precisely how A_n, for all n, can arise from simple conventional statistical models.

In Section 2 we define a nested splitting process and prove that this process satisfies A_n exactly. In Section 3 a multivariate version of A_n is proposed. The primary purpose of this paper is to provide further understanding of the nature of A_n by showing how it arises from a simple parametric model. I only briefly allude to the various practical applications of A_n. There are many such applications, e.g. the survival analysis of Berliner and Hill (1988), the results of Hill (1979, 1980a) for inference about the number of species or types in a finite population and the results of Hill (1968) for inference about the percentiles of a population. See Smith (1988) for a discussion of such applications. Perhaps because of the very subtle issues involved in the distinction between countable additivity and finite additivity, some practitioners have felt uneasy about such procedures.

My point of view is that these issues, although of theoretical interest, and of importance to clarify, are only tangential with respect to applications of A_n. In real problems the parameter space and the observation space can ordinarily be taken to be finite, with nothing of any practical importance lost in so doing, since any statistical analysis will necessarily be made on a computer with finite memory, or something equivalent. But, for finite spaces, finite additivity and countable additivity are the same, and so the important practical issues regarding finite *versus* countable additivity concern the methods of approximation to be used in dealing with large but finite sets. The two theories lead to quite different methods of approximation, with the countably additive theory ruling out certain approximations that are valid in the finitely additive theory. Thus A_n and the Dirichlet process provide different, although closely related, ways of making such approximations, but in the countably additive theory the methods based on A_n would automatically be ruled out. See Hill (1990a) for a general discussion of issues about finite *versus* infinite partitions, with regard to the validity of the frequentist admissibility principle and countable additivity. See Hill (1992), p. 90, for a simple example illustrating how A_n can arise as an approximation in finite spaces.

The finitistic approach that I am recommending is related to the intuitionist and constructivist positions regarding mathematics, as opposed to formalism. The 'intuitionist' school, founded by L. E. J. Brouwer, rejects the use of some of the standard logic in dealing with infinite collections; for example, it rejects the principle of *tertium non datur*, or the law of the excluded middle, and the axiom of choice. Several other distinguished mathematicians, such as H. Poincaré and H. Weyl, have supported parts of the intuitionist and/or constructivist thesis regarding the nature of mathematics and science. The distinguished mathematician–logician Ramsey (1950), pages 80, 183, 204 and 252, eventually adopted a finitist point of view that is very close to that of this paper. These considerations are perhaps even more important for statistics than for pure mathematics and logic, since statistics purports to deal with real problems. If statistical methods are to have validity, they must be based on sound logical grounds, rather than on hypothetical and highly questionable operations with infinite sets. The finitely additive theory of probability is compatible with this view, since it presumes nothing whatsoever about the validity of certain infinite operations, such as the conventional evaluation of the probability of a non-finite union of disjoint events, and is based on the logic of uncertainty for *finite* partitions. This theory was founded by de Finetti (1937, 1974), who emphasized that probabilistic evaluations must have a clear operational meaning. As observed by many distinguished mathematicians, operational meaning is often lost in attempting to deal with infinite collections as though the customary finitistic operations were automatically valid.

2. SPLITTING PROCESSES

We begin by recalling that A_1 and A_2 can be obtained by the use of improper prior distributions on the location, and on the location and scale parameters respectively of a normal distribution. See Jeffreys (1961), p. 171, and Hill (1968), p. 688. For example, for A_1, if μ has an improper prior distribution represented by Lebesgue measure, and if the distribution of the error is $N(0, 1)$, then given $X_1 = x_1$ the posterior distribution of μ is $N(x_1, 1)$, and the posterior predictive distribution for X_2 is $N(x_1, 2)$. Hence the posterior probability that $X_2 > x_1$ is $\frac{1}{2}$. Similarly, for unknown μ and σ, if these parameters are given the conventional improper joint prior distribution of Jeffreys, then A_2 holds. For $n > 2$, until now A_n had not been obtained, constructively, by means of parametric models and improper prior distributions. Lane and Sudderth (1978) proved an existence theorem to the effect that finitely additive distributions satisfying A_n for each n exist but did not explicitly model such distributions. Here we give explicit parametric representations that can hold for all n.

The first step in our construction is to introduce the concept of adherent mass at a point. This is an extremely simple and useful concept, due to de Finetti (1974), p. 240, that arises in the finitely additive theory of probability. Before making precise definitions, we shall motivate this concept in connection with the joint distribution of two variables X_1 and X_2 which will later represent the first stage in our iterative construction of a splitting process satisfying A_n. Let X_1 have distribution π, where π is any fixed distribution on the line. We now describe the conditional distribution of X_2, given $X_1 = x_1$. With probability $\frac{1}{2}$, X_2 is given the distribution π; with probability $\frac{1}{2}$, X_2 is *adherent* to x_1, with conditional probability $\frac{1}{2}$ of being larger than x_1, conditional probability $\frac{1}{2}$ of being smaller than x_1, conditional probability 0 of being equal to x_1, and with conditional probability 1 of being within any specified open interval containing x_1. Such adherence can be obtained as follows. Imagine that, given $X_1 = x_1$, $X_2 - x_1$ is equal to $1/K$ for some non-zero integer K, where K has a distribution symmetric about 0. If K has a diffuse finitely additive distribution on the integers, so that there is probability 1 that K is larger in absolute value than any finite constant, the result follows easily, since K must be some finite integer, so that $X_2 - x_1$ cannot be 0, and, with probability 1, $1/K$ will be smaller in absolute value than any positive constant. The concept of adherence does not depend on symmetry, although this is the primary case of interest in this paper.

Such distributions may at first sight appear rather exotic, but this is not really so. They correspond to a situation where no possible measurement can differentiate between a value and 0, for example, even though the quantity in question is known not to be equal to 0. See de Finetti (1974), p. 242. In this case neither empirically nor theoretically can we rule out such adherent distributions. Thus we may know that a particle has positive mass, but its mass may be so small that it is enormously beyond the powers of our technology to determine the exact value. It may only be possible in finite time to determine that the value is less than some specified positive ϵ. Compare also the discussion by de Finetti (1974), p. 242, of the indeterminacy of distribution functions at jump points.

It is not necessary that we view adherent mass distributions as holding exactly in real situations. Indeed, the primary purpose of the concept is merely to provide useful approximations and ways of thinking about very common situations involving large finite partitions. Clear understanding of the property of adherence is necessary to deal with ties and the grouping of data, such as in H_n, and in understanding the behaviour of the Dirichlet process. See Hill (1968), section 3, and Hill (1988) for the definition of H_n. For our purposes at present it suffices to observe that such finitely additive distributions are known to exist, so that the description that we have given for the generation of X_1 and X_2 is coherent in the sense of de Finetti, i.e. no Dutch book is possible. If desired, they can equally well be represented in terms of improper prior distributions. For example, a weight of 1 for each non-zero integer generates adherence at 0 for the reciprocal of an integer selected 'at random' over the non-zero integers. Mostly, we shall use the language of the finitely additive theory, which is fully rigorous, and whose foundations were developed by de Finetti (1974) and Savage (1972). See also Regazzini (1987), Schervish *et al.* (1984) and Hill and Lane (1985). Rényi (1970) and Hartigan (1983) provide rigorous theories of improper prior distributions and conditional probability spaces.

Since most probabilists and statisticians accept the countably additive framework, and therefore might immediately reject such concepts as that of adherent mass, it may also be useful to point out that according to Kolmogorov (1950) the axiom of countable additivity (or continuity) cannot be justified other than by expediency. Although expediency is important, it is hardly a matter of fundamental truth. For this reason I ask the indulgence of the reader to pursue further some of these ideas, even though at first glance they may seem unusual. The issues concerning countable additivity have some important implications for the theory and practice of statistics. For example, Ramakrishnan and Sudderth (1988) have shown that even in the simplest of all probability scenarios, that of flipping a fair coin, Borel's strong law does not hold in the finitely additive context. These researchers show that with exactly the same joint distributions for all finite sequences, i.e. probability $1/2^k$ for any k-tuple of 0s and 1s, we can have the average converge everywhere to 0, converge everywhere to 1 or fail to converge everywhere. This implies that *no finite experience* with a coin can determine what happens in the limit as the number of trials goes to infinity, except in terms of the weak law of large numbers. For the practice of statistics, ultimately it comes down to a question of choice as to approximations. See de Finetti (1974), p. 33, who describes countable additivity as a 'Procrustean bed', and Hill (1990a), p. 520, for further discussion of such issues.

We shall now make a few definitions which will enable us to operate with adherent distributions of mass, and to define a splitting process.

Definition 1. A probability distribution is said to have adherent mass at a point (finite or infinite) if the infimum of probabilities of all open intervals containing the point is greater than the probability of the point itself. It is said to have a *purely* adherent mass at a point if it has an adherent mass at the point and the probability of the point itself is 0. Such language is also used for random quantities with such distributions.

Definition 2. A random quantity is said to be negligible if the total mass either at, or adherent to, 0 is 1.

Definition 3. Two random quantities are said to be equivalent if their difference is negligible.

Definition 4. A distribution is said to be diffuse at ∞ if it has a purely adherent mass of $\frac{1}{2}$ at each of $+\infty$ and $-\infty$. (When π is diffuse at ∞, and a random quantity X has distribution π, we say that X splits from ∞, or is generated from ∞. When X has a distribution for which all the mass is adherent to a point x_1, we say that X splits from x_1.)

It follows immediately that a finite sum of negligible quantities is negligible, and that a diffuse distribution attaches probability 0 to any finite interval. Special diffuse distributions are used by some Bayesians to represent a form of ignorance. The improper uniform prior distribution for a location parameter, and for the logarithm of a scale parameter, as in Jeffreys (1961), are familiar special cases. These can be given a finitely additive interpretation as well. We can also strengthen the notion of diffuseness by requiring that the conditional probability for a particular value, conditional on a finite set of values, be uniform, as in Hill (1980b). There are some subtleties that arise in the finitely additive theory that are worth mentioning explicitly. Although a mass purely adherent to 0 is for many practical purposes indistinguishable from a mass exactly at 0, the two associated random quantities are not logically identical, since the first is certain *not* to be exactly 0. In dealing with such things we must therefore take greater care than is customary in the conventional countably additive theory.

We now construct a splitting process. Let X_1 and X_2 be defined as before. Given $X_1=x_1$ and $X_2=x_2$, we generate X_3 as follows. With conditional probability $\frac{1}{3}$, X_3 is generated according to π; with conditional probability $\frac{1}{3}$, X_3 is generated from a symmetrical distribution purely adherent at x_1; with conditional probability $\frac{1}{3}$, X_3 is generated from a symmetrical distribution purely adherent at x_2. This procedure can be continued iteratively. After $X_i = x_i$, for $i = 1, \ldots, n$, have been realized, X_{n+1} is equally likely, with common probability $1/(n+1)$, to be generated from π or to have a symmetrical distribution purely adherent to each of the n distinct values x_i already generated. The observations are generated sequentially in time, so that we can speak of X_i as the ith point generated. Finally, joint distributions of the X_i are defined to be strategic in that probabilities for future observations can be evaluated as expectations of conditional probabilities, given previous observations. See Lane and Sudderth (1984) and Regazzini and Zaboni (1988). We call such a sequence $X_1, \ldots, X_n$, for any fixed π, a nested splitting process.

We shall assume, for simplicity, that the distribution π and the adherent mass distributions have been defined for all subsets of the line. By virtue of de Finetti's fundamental theorem of probability it is always possible, in principle, coherently to extend any partially defined coherent evaluation of probability to all subsets; de Finetti (1974), p. 111, and Lad *et al.* (1990).

Theorem 1. For a nested splitting process with π diffuse at ∞, A_n holds exactly. If π is any distribution with neither adherent nor positive mass at finite points, then exchangeability holds, and ties have probability 0.

Proof. That ties have probability 0 follows immediately from the definition of pure adherence and the fact that π has no adherent or positive mass at finite points. That the conditional probabilities are in accord with A_n when π is diffuse may be seen as follows. Let $X_i = x_i$, for $i = 1, \ldots, n$, with all these values distinct, and consider the conditional distribution of X_{n+1}. (In the finitely additive theory all conditional distributions automatically satisfy the axioms of probability, as with full conditional probability distributions. See Hill and Lane (1985).) Now let I_i be the open interval between $x_{(i)}$ and $x_{(i+1)}$, for $i = 0, \ldots, n$. First take i to be between 1 and $n-1$, so that the I_i are finite intervals. Since I_i is finite and π is diffuse, if X_{n+1} is generated from π, then there is probability 0 that X_{n+1} will fall in I_i. Similarly, unless X_{n+1} splits from either $x_{(i)}$ or from $x_{(i+1)}$, there is probability 0 that X_{n+1} will fall in I_i. Conditionally on X_{n+1} splitting from $x_{(i)}$, the probability that it falls in I_i is $\frac{1}{2}$, and similarly if X_{n+1} splits from $x_{(i+1)}$. Since there are $n+1$ equally likely possible sources for X_{n+1}, including π, it follows that the probability that X_{n+1} falls in I_i is exactly $1/(n+1)$. When π is diffuse, this is also true if $i = 0$ or $i = n$, in which case the interval I_i is semi-infinite. For example, if $i = 0$, then (ignoring events of probability 0) for X_{n+1} to be in I_0 it must be the case that either X_{n+1} splits from $x_{(1)}$, or else that it is generated from π. In the latter case, because π is diffuse at ∞, there is probability $\frac{1}{2}$ that X_{n+1} will be smaller than $x_{(1)}$. This yields $1/(n+1)$, as before, for the posterior predictive probability that X_{n+1} will be in I_0, and similarly for $i = n$. This completes the proof that the conditional distribution for X_{n+1} is in accord with A_n when π is diffuse.

We now prove that $X_1, \ldots, X_{n+1}$ form an exchangeable sequence, for any π which has no adherent or positive mass at finite points.

By first conditioning on $X_1 = u$, and then using the fact that the distributions are strategic, to integrate with respect to u, we have, for $s_1 < s_2$,

$$\Pr\{X_1 \leqslant s_1, X_2 \leqslant s_2\} = \int_{-\infty}^{s_1} \Pr\{X_2 \leqslant s_2 | X_1 = u\} \, \pi(\mathrm{d}u)$$

$$= \int_{-\infty}^{s_1} \{\tfrac{1}{2}\,\pi(s_2)+\tfrac{1}{2}\}\,\pi(\mathrm{d}u)$$

$$= \tfrac{1}{2}\,\pi(s_1)\,\pi(s_2)+\tfrac{1}{2}\,\pi(s_1),$$

where $\pi(s)$ is the mass attached to the closed interval from $-\infty$ to s by π. With a similar evaluation for the case $s_1 \geqslant s_2$, we obtain the joint distribution

$$\Pr\{X_1 \leqslant s_1,\, X_2 \leqslant s_2\} = \tfrac{1}{2}\,\pi(s_1)\,\pi(s_2) + \tfrac{1}{2}\,\pi(s_1 \wedge s_2),$$

where $s_1 \wedge s_2$ is the smaller of s_1 and s_2. This function is symmetric in its arguments, proving that X_1 and X_2 are exchangeable.

By conditioning on the first k variables, similar evaluations can be made for the higher dimensional distributions. Let $F^{(k)}(s_1, \ldots, s_k)$ be the joint distribution function for the first k random quantities, for $k=1, \ldots, n+1$. Then it is easily verified that

$$F^{(k+1)}(s_1, \ldots, s_{k+1}) = \frac{1}{k+1}\left\{ \sum_{i=1}^{k} F^{(k)}(s_1, \ldots, s_{i-1}, s_i \wedge s_{k+1}, s_{i+1}, \ldots, s_k) + \pi(s_{k+1})\,F^{(k)}(s_1, \ldots, s_k)\right\}$$

where for $i=1$ in the above sum we take $(s_1, \ldots, s_{i-1}, s_i \wedge s_{k+1}, s_{i+1}, \ldots, s_k) = (s_1 \wedge s_{k+1}, s_2, \ldots, s_k)$.

Using the iterative character of such functions, it is easy to see that the joint distributions are symmetric functions of their arguments, which proves exchangeability. In the diffuse case, the joint distribution functions are in fact constant at finite points. For $k=1$ the constant is $\frac{1}{2}$; for $k=2$ it is $\frac{3}{8}$. If c_k is the constant for a k-dimensional joint distribution, then $c_{k+1} = c_k(k+\frac{1}{2})/(k+1)$.

The theorem shows that the probabilities specified by A_n can be realized exactly in theory. In our construction of the splitting process the time order was relevant to the realization of the process, or creation of the data. For example, X_2 could have split from ∞ or from the already realized x_1, which requires the existence of x_1 before the determination of X_2. But we have also proved that the process so engendered is exchangeable, which implies that *probabilistically* this time order is immaterial in so far as the joint distribution functions are invariant under permutations. For a related situation consider the discussion of the relationship between the Pólya urn model and the Bayes–Laplace model in de Finetti (1974), p. 220. Although two processes may be structurally different, the expression of our probabilistic knowledge about them can be precisely the same.

(The notion of exchangeability used in this paper, namely in terms of invariance under permutations of the joint distribution functions evaluated at finite points, is weaker in non-finite spaces than the definition in terms of such invariance for the joint distributions themselves. In the countably additive framework the two definitions are equivalent, but for merely finitely additive distributions such exchangeability is weaker than exchangeability in the sense of invariance of the distributions themselves, since in this context the probabilities for rectangle sets do not determine the joint distributions uniquely. At present, little seems to be known about exchangeability in the finitely additive case. There are some fairly subtle aspects to such exchangeability.

For example, a distribution with a purely adherent mass at a point would not be exchangeable with a distribution having all its mass exactly at that same point, even though for some practical purposes the two are indistinguishable. The weaker definition of exchangeability is the more useful when working in the finitely additive context. In any event, for finite spaces there is no distinction between the two definitions, and so from the finitistic perspective taken here the issue concerns how to pass to certain limits which are at best of mathematical interest.)

In a large but finite space, such as would be appropriate for applications of the model that I have proposed, the two concepts of exchangeability are identical, and so we can forget the time ordering for the purpose of statistical inference and prediction. Thus we can instead consider a population of values $X_1, \ldots, X_N$ that originated from a splitting process but now is simply an existing finite population of numbers. By construction these values are necessarily distinct, so that the ordered values in the population are $X_{(1)} < X_{(2)} < \ldots < X_{(N)}$. Before the process is realized, we can visualize the process as creating a random distribution, in which the probability attached to a set is simply the random proportion of these X_i that fall in the set. However, in the present context we shall imagine that the values have already been generated, but unobserved. In the subjective Bayesian theory, as long as there is no further information about the population values X_i, it is appropriate to use the same distribution after, just as before, they were generated. See Hill (1990b) for discussion. Now suppose that a simple random sample of size n, without replacement, is taken from such a population, and the observed ordered values in the data are $x_{(1)} < x_{(2)} < \ldots x_{(n)}$.

Because of the exchangeability, for prediction of the value of the next observation there is no harm in supposing that these values are the first n values created by the process, so that A_n, and indeed A_{N-1}, is automatically satisfied in sampling from a population $X_1, \ldots, X_N$ that is created by a nested splitting process. It was proved in Hill (1968), p. 688, that A_k implies A_j for $j < k$, so A_j can hold for any $j < N$. If we take seriously the notion of generating an infinite number of points, as for example in Lane and Sudderth (1978), then A_n holds for all n. There is a finitely additive version of de Finetti's theorem for infinite sequences of exchangeable random quantities, which suggests that the usual interpretation in terms of an 'unknown' distribution, representing the limiting frequency of points in various sets, may still be valid, although uniqueness of the de Finetti measure is lost, and in this context we must give serious attention to the distinction between the two versions of exchangeability. Note that the pre-data expected proportion of observations in a finite sample that fall in an open interval is simply the π-mass attached to that interval. See de Finetti (1937), Hewitt and Savage (1981), Lane and Sudderth (1978), Savage (1972), p. 53, Diaconis and Freedman (1980) and Hill (1988) for some related discussion. Although mathematically interesting, I do not think that the infinite population case is of great practical importance. The primary statistical problem considered in this paper is that of inference about a finite population based on a random sample, with the adherency assumption used to provide useful approximations.

We have seen that A_n holds exactly for a nested splitting process. It is an open question whether there is a *basically* different model that generates A_n exactly. Hill (1987) proves that H_n, which allows for ties, can also be realized exactly.

Finally, it is interesting to compare the analysis from splitting models, or from H_n, with that from the Dirichlet process. The Dirichlet process can be derived, as in

Blackwell and MacQueen (1973), as the limiting distribution of proportions obtained from a generalized Pólya process. In the notation of Blackwell and MacQueen, we have

$$\Pr\{X_{n+1}\in\mathscr{B}\mid X_1,\ldots,X_n\} = \mu_n(\mathscr{B})/\mu_n(\mathscr{X}), \tag{1}$$

where

$$\mu_n = \mu + \sum_{i=1}^{n} \delta(X_i),$$

$$P(X_i\in\mathscr{B}) = \mu(\mathscr{B})/\mu(\mathscr{X}),$$

$\delta(x)$ denotes the unit measure concentrating at x and $\mathscr{X}$ is the space of observations.

Now generalize the nested splitting process to include an additional parameter η_n, for the probability that the next observation is from π, and with equal probability $(1-\eta_n)/n$ that the next observation splits from each of the n already realized x_i. Then, for any open interval $\mathscr{B}$, for a nested splitting process we have

$$\Pr\{X_{n+1}\in\mathscr{B}\mid X_1,\ldots,X_n\} = \pi(\mathscr{B})\eta_n + \{C_n(\mathscr{B}) + \tfrac{1}{2}D_n(\mathscr{B})\}(1-\eta_n)/n, \tag{2}$$

where $C_n(\mathscr{B})$ is the number of observations among the first n that lie in $\mathscr{B}$ and $D_n(\mathscr{B})$ is the number of x_i that are on the boundary of $\mathscr{B}$. For $\pi(\mathscr{B}) = \mu(\mathscr{B})/\mu(\mathscr{X})$, and with $\eta_n = \mu(\mathscr{X})/\{n+\mu(\mathscr{X})\}$, then for intervals $\mathscr{B}$ with $D_n(\mathscr{B})=0$ this predictive probability is identical with the probability given by equation (2) of the Blackwell–MacQueen representation of the generalized Pólya process. If also $\mu(\mathscr{X})=1$, we obtain the original nested splitting process. If $\mu(\mathscr{X})=\infty$, then the above predictive probability is simply $\pi(\mathscr{B})$.

If we now make one further generalization then both the nested splitting process and the Dirichlet process become special cases of a single very general process. Define $\tau_{i,n}$ to be the probability that the next observation ties $x_{(i)}$, given that it splits from $X_{(i)}$, and given the first n observations. Given that X_{n+1} splits from $x_{(i)}$ but does not tie $x_{(i)}$, let the mass $1-\tau_{i,n}$ be symmetrically adherent to $x_{(i)}$. In my original construction $\tau_{i,n}=0$ for each n and $i=1,\ldots,n$, and $\eta_n=1/(n+1)$. To obtain the Dirichlet process of Ferguson (1973), p. 209, with Ferguson's parameters α and $M=\alpha(\mathscr{X})$, we need only to set $\tau_{i,n}=1$ for each n and $i=1,\ldots,n$, take $\pi = \alpha/M$ and $\eta_n = \alpha(\mathscr{X})/\{n+\alpha(\mathscr{X})\}$. In this case the term $D_n(\mathscr{B})$ drops out of equation (2); this equation then holds for all $\mathscr{B}$ and is identical with equation (2) of Blackwell and MacQueen (1973). If we thus choose the parameters to yield the Dirichlet process, and if further we assume countable additivity for the sequence of variables that are generated by the process, then the process is identical with that of Blackwell and MacQueen. Thus both the Dirichlet process and A_n can be seen to arise from such generalized splitting processes. We state these results as a theorem.

Theorem 2. Let X_i, for $i=1,\ldots,n,\ldots$, be a generalized splitting process with parameters η_n and $\tau_{i,n}$. Then for $\eta_n=1/(n+1)$, and $\tau_{i,n}=0$ for $i=1,\ldots,n$, the process is a nested splitting process. For $\eta_n = \alpha(\mathscr{X})/\{n+\alpha(\mathscr{X})\}$ and $\tau_{i,n}=1$ for $i=1,\ldots,n$, and under the assumption of countable additivity the process generates a Ferguson–Dirichlet process with parameter α.

3. MULTIVARIATE VERSIONS OF A_n AND CONCLUSIONS

The splitting processes that we have defined can immediately be generalized to higher dimensional spaces, e.g. to the surface of a sphere, three-dimensional Euclidean space, higher dimensional versions of these spaces and indeed to any surface or space whatsoever. We need only to generate points from an appropriate distribution π and then to define adherency in an appropriate way, using, for example, some metric in the space under consideration. Such generalizations lead to multivariate versions of A_n. For example, in two-dimensional Euclidean space, we can take π to be diffuse in the sense of attaching probability 1 to the complement of any bounded open sphere, and the adherent distribution of mass at a point can be taken as spherically symmetric about the point, giving mass $\frac{1}{4}$ to each of the four quadrants formed with the point as origin. In this case there would be probability 1 that the next observation will be within any open sphere about a point, given that it splits from that point. In n dimensions we would attach probability $1/2^n$ to each of the 2^n quadrants formed by a point as origin, given that a split occurs from that point, again using spherical symmetry, and take π to attach probability 1 to the complement of any bounded open sphere. We can proceed similarly on the surface of a sphere, except that now the symmetry must be restricted to the surface of the sphere. For more general surfaces and spaces there may be other notions of diffuseness and symmetry that are of interest. Also, in Bayesian survival analysis, as in Berliner and Hill (1988), there are a variety of ways to introduce a multivariate version of A_n to allow for covariates.

In conclusion, we have here constructed a splitting process that yields A_n exactly. A version of A_n was originally suggested from a fiducial point of view by Fisher. It also has a confidence or tolerance interpretation, as proposed by Dempster (1963), p. 110. It is simple, intuitive and coherent, and has several subjective Bayesian interpretations and justifications. I hope that it will be used more widely by practitioners in situations where there is a weak *a priori* knowledge than has hitherto been the case.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation under grant DMS 8901234.

REFERENCES

Aitchison, J. and Dunsmore, I. R. (1975) *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.

Berliner, L. M. and Hill, B. M. (1988) Bayesian nonparametric survival analysis (with discussion). *J. Am. Statist. Ass.*, **83**, 772–784.

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

Dempster, A. P. (1963) On direct probabilities. *J. R. Statist. Soc.* B, **25**, 100–110.

Diaconis, P. and Freedman, D. (1980) Finite exchangeable sequences. *Ann. Probab.*, **8**, 745–764.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, 2nd edn, vol. 2. New York: Wiley.

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

de Finetti, B. (1937) La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré*, **7**, 1-68.

——(1974) *Theory of Probability*, vol. 1. London: Wiley.

Fisher, R. A. (1939) Student. *Ann. Eugen.*, **9**, 1-9.

——(1948) Conclusions Fiduciare. *Ann. Inst. H. Poincaré*, **10**, 191-213.

Hartigan, J. (1983) *Bayes Theory*. New York: Springer.

Hewitt, E. and Savage, L. J. (1981) Symmetric measures on cartesian products. In *The Writings of Leonard Jimmie Savage—a Memorial Selection*, pp. 244-275. Washington DC: American Statistical Association.

Hill, B. M. (1968) Posterior distribution of percentiles: Bayes theorem for sampling from a finite population. *J. Am. Statist. Ass.*, **63**, 677-691.

——(1979) Posterior moments of the number of species in a finite population, and the posterior probability of finding a new species. *J. Am. Statist. Ass.*, **74**, 668-673.

——(1980a) Invariance and robustness of the posterior distribution of characteristics of a finite population, with reference to contingency tables and the sampling of species. In *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys* (ed. A. Zellner), pp. 383-395. Amsterdam: North-Holland.

——(1980b) On finite additivity, non-conglomerability, and statistical paradoxes (with discussion). In *Bayesian Statistics* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), pp. 39-66. Valencia: Valencia University Press.

——(1987) Parametric models for A_n: splitting processes and mixtures. Unpublished. Department of Statistics, University of Michigan, Ann Arbor.

——(1988) de Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference (with discussion). In *Bayesian Statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), pp. 211-241. Oxford: Oxford University Press.

——(1990a) Comment on "An ancillarity paradox that appears in multiple linear regression", by L. D. Brown. *Ann. Statist.*, **18**, 513-523.

——(1990b) A theory of Bayesian data analysis. In *Bayesian and Likelihood Methods in Econometrics and Statistics: Essays in Honor of George A. Barnard* (eds S. Geisser, J. Hodges, S. J. Press and A. Zellner), pp. 49-73. Amsterdam: North-Holland.

——(1992) Bayesian nonparametric prediction and statistical inference (with discussion). In *Bayesian Analysis in Statistics and Econometrics* (eds P. K. Goel and N. S. Iyengar), pp. 43-94. New York: Springer.

Hill, B. M. and Lane, D. (1985) Conglomerability and countable additivity. *Sankhya* A, **47**, 366-379.

Hume, D. (1748) *An Enquiry Concerning Human Understanding*. London.

Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Oxford: Clarendon.

Kolmogorov, A. N. (1950) *Foundations of Probability*, p. 15. New York: Chelsea Publishing.

Lad, F., Dickey, J. M. and Rahman, M. A. (1990) The fundamental theorem of prevision. *Statistica*, **50**, 19-38.

Lane, D. and Sudderth, W. (1978) Diffuse models for sampling and predictive inference. *Ann. Statist.*, **6**, 1318-1336.

——(1984) Coherent predictive inference. *Sankhya* A, **46**, 166-185.

Lenk, P. (1984) Bayesian nonparametric predictive distributions. *Doctoral Dissertation*. University of Michigan, Ann Arbor.

Ramakrishnan, S. and Sudderth, W. (1988) A sequence of coin-toss variables for which the strong law fails. *Am. Math. Mthly*, **95**, 939-941.

Ramsey, F. (1950) *The Foundations of Mathematics and Other Logical Essays* (ed. R. B. Braithwaite). New York: Humanities Press.

Regazzini, E. (1987) de Finetti's coherence and statistical inference. *Ann. Statist.*, **15**, 845-864.

Regazzini, E. and Zaboni, G. (1988) Strategies are coherent conditional probabilities. *Quaderno 28*. Università Degli Studi di Milano, Milan.

Rényi, A. (1970) *Probability Theory*. New York: Elsevier.

Savage, L. J. (1972) *The Foundations of Statistics*, 2nd revised edn. New York: Dover Publications.

Schervish, M., Seidenfeld, T. and Kadane, J. (1984) The extent of non-conglomerability. *Z. Wahrsch.*, **66**, 205-226.

Smith, J. Q. (1988) Discussion of "de Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference", by B. M. Hill. In *Bayesian Statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), p. 229. Oxford: Oxford University Press.