Using the Fundamental Theorem of Prevision to Identify Coherency Conditions for Finite Exchangeable Inference

by

Frank Lad, John Deely and Andrea Piesse
Department of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand

No. 95 July, 1993

Keywords and phrases: Coherence, fundamental theorem of prevision, exchangeable, psi-function, extendible, de Finetti's representation theorem, completely monotone sequence, Hausdorff's theorem, Cantor distribution, Polya distribution, frequency, histogram, $A(n)$, $H(n)$.

SUMMARY

Suppose N+1 events are regarded exchangeably. We begin by identifying the coherent implications of asserting simultaneously the conditional previsions $P(E_{N+1} | (S_N = a)) = a/N$ for values of $a = 1, \dots, (N-1)$. We present a new proof that if you make these assertions for every value of N, then coherency requires you also to assert for every N that $P[(S_N = 0) + (S_N = N)] = 1$. The proof is based directly on de Finetti's fundamental theorem of prevision. We continue the proof scheme to identify all vectors $\mathbf{p}_{N+1}$ of coherent inferential assertions $P[E_{N+1} | (S_N = a)] = p_{a,N}$ for values of $a = 0, \dots, N$. We specify computable conditions under which such assertions are infinitely exchangeably extendible, and we provide a complete parameterization of this class of assertions, identifying the family of Polya inferences as a proper subclass. Corollary to this theorem, we specify a computable procedure for determining precisely the value of predictive probabilities $P(E_{N+1} | (S_N = a))$ when the probability distribution of S_{N+1} is representable as *any* Binomial mixture distribution. The algorithm is based on the moments of the mixing distribution. As an example of its application, we compute the relevant conditional probabilities for a Binomial-Cantor mixture distribution. Finally, we generalize the analysis to consider a sequence of N+1 discrete quantities that are regarded exchangeably, when each quantity may be any integer in the set $\{0, 1, \dots, K\}$. We establish the coherency of asserting conditional probability distributions that mirror every histogram for which there is at least one observation in each of the K+1 categories. However we also show that such inferences require several accompanying assertions which are not attractive in typical problems: if the histogram rule is followed for large N, it must also be followed for small N; strictly positive histograms must be accorded extremely small probabilities; and histogram categories must be regarded *subexchangeably* over strictly positive histograms, a concept that is defined precisely in the text.

0. Introduction and statement of the theorems. In a seminal paper, de Finetti (1937) presented two central theorems in subjective probability theory. One of these results, now commonly termed "de Finetti's representation theorem," showed that an infinite sequence of zero-one random variables is exchangeable if and only if the joint distribution of any finite sequence of them is a unique mixture of a joint distribution of conditionally independent variables, each having the same Bernoulli (θ) distribution, mixed by a distribution $F(\theta)$. This theorem has drawn considerable interest and received widespread attention in the literature which contains papers discussing various aspects and generalizations. See for example Heath and Sudderth (1976), Diaconis (1977), and Hill (1988) and numerous further references contained in these papers.

The other result, subsequently labeled by de Finetti as "the fundamental theorem of probability" (de Finetti, 1974, 3.10) has been apparently widely unnoticed, having only recently received new attention. See Bruno and Gilio (1980) and Lad, Dickey, and Rahman (1990). In its simplest form, this theorem states that when probabilities are asserted for any N events, then coherency identifies precise bounds on the concomitant assertion of probability for any other specifiable event. Its proof revolves on the fact that every coherent prevision vector must lie within the convex hull of the set of all possible vector measurements for the quantities under consideration. The mathematical statement of the theorem is so simple that it perhaps lacks immediate appeal. Yet the implications of the "fundamental theorem" are deep for statisticians of classical, Bayesian, or subjectivist persuasions, and are yet to be fully appreciated. One purpose of the present paper is to show how this theorem can be used in rather simple yet practical situations to obtain new and surprising results both in subjective probability theory and statistical methodology. Hence we illustrate the theoretical potential of the general fundamental theorem of prevision by using it to obtain three theorems that present serious questions for practical areas of statistical inference.

Specifically, we begin by showing that if one is committed for *every* N to use the predictor a/N for the conditional probability of success on an $(N+1)^{st}$ event given a record of "a" successes among N other events with which it is regarded exchangeably ($a = 1, 2, \ldots, N-1$) then this strategy is coherent if and only if it is concomitantly asserted that sequences containing all N+1 successes or all failures are accorded probability equal to 1. Thus, a frequentist strategy using only recorded history as a basis for inference about zero-one exchangeable quantities is incoherent unless it is also asserted with certainty that only all zeroes or all ones will be observed! It is clear that very few practicing statisticians would or could confidently make this latter assertion in many practical instances, and thus in these cases the "history only" strategy would be incoherent. The conceptual content of our first result was known already to Jeffreys (1939), but our analysis yields new and fuller algebraic detail.

Next, we replace the "history only" strategy of inference with a general expression which incorporates every coherent inference strategy. That is, let $p_{a,N} = P[E_{N+1} | (S_N = a)]$ denote the asserted conditional probability for an $(N+1)^{st}$ event given "a" successes among N observed events with which it is regarded exchangeably; and let $\mathbf{p}_{N+1}$ denote a vector $(p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$ of such inferential assertions. We show that any vector $\mathbf{p}_{N+1}$ lying in the (N+1)-dimensional hypercube represents coherent assertions, but that coherency also requires that such assertions be accompanied by a specific upper triangular matrix for all lower order conditional previsions based on fewer conditioning observations. In addition, we derive *computable* conditions under which the assertions embodied in $\mathbf{p}_{N+1}$ are infinitely exchangeably extendible. We show further that each component of the cohering upper triangular matrix can be parameterized as $p_{a,K} = (a + \alpha_{a,K})/(K + \alpha_{a,K} + \beta_{a,K})$ for a specific choice of $(\alpha_{a,K}, \beta_{a,K})$ determined by $p_{a,K}$ and $p_{a,K-1}$. When this parameterization specifies $\alpha_{a,K} = \alpha$ and $\beta_{a,K} = \beta$ for all "a" and K, it yeilds the well-known Polya inferences as a subfamily. But our second result establishes a *complete* parameterization of all coherent inferences concerning N+1 events regarded exchangeably. We display its range of applicability with a numerical example of the Binomial-Cantor mixture distribution for 25 events.

Finally, we study another companion problem to the first, extending its context to allow each of the N+1 considered quantities to assume any value in {0,1,...,K}. In this case we show that the "histogram only" prediction strategy, which uses any strictly positive histogram based on N quantities as a conditional probability distribution for an $(N+1)^{st}$ quantity, is coherent. However, we also find that coherency would impose some annoying concomitant assertions.

To begin with, if you assert that you would use any positive histogram based on a large number of observations as your conditional distribution for the next observation, coherency requires that you also follow the "histogram only" rule when conditioning upon any smaller number of observations as well! The prescription to use the histogram rule when conditioning on a large sample but not necessarily on a small sample is incoherent.

Perhaps even more startling, the "histogram only" strategy has embedded within it a coherency induced upper bound on the probability of observing a strictly positive histogram on the basis of any finite number of observed quantities. This bound decreases to 0 as N increases. In the extreme, if you would assert strictly positive histogram mimicking conditional probabilities for *every* N, then your probability for observing a strictly positive histogram from the first M observations must equal 0, no matter what be the value of M. The importance of this qualification is most easily recognized in measurement problems for large populations, where the categories of possible measurement are crude, crude enough that you expect to observe at least one instance of each category value within a practical number of opportunities. (Details of an example concerning milk yields from dairy cows are provided in

Section 4.) This very minimal expectation would preclude the systematic use of histogram mimicking conditional probabilities. The assertion of positive probability that all category values will be observed within a finite number of measurements requires that conditional probabilities involve some adjustment to conditioning histograms.

Still another odd requirement of the "histogram only" conditioning rule is that any two strictly positive histograms that are permutations of one another must be accorded identical probabilities. The formal specification of this result suggests a natural definition for the concept of subexchangeability of quantities, a concept distinct from that of partial exchangeability.

None of these assertions that are required to accompany the "histogram only" rule for inference seems warranted in many real problems. Thus, the practical statistician would be driven away from systematic use of the histogram rule except perhaps in very small scale problems.

The results proved in the present paper bear direct relation to Hill's analysis of the structures of inference that he terms A(n) and H(n). See Hill (1968, 1987, 1988), Berliner and Hill (1988), and Lane and Sudderth (1978). Whereas Hill's A(n) theory pertains to real valued quantities that do not involve ties, and his H(n) theory merely allows ties, there is a sense in which our theory for finite discrete quantities *requires* ties among them. All our analysis pertains to bounded discrete quantities that can take only a *specified* finite number of values, as in the standard case of a discretely calibrated measuring instrument. Details will be easier to discuss once our results have been stated formally and proved.

Our results are made precise in Theorems 1, 2, and 3 below. They are proved in Sections 1, 2, and 3, respectively. Section 2 includes a numerical example which establishes exact inferences based upon a Binomial-Cantor mixture distribution. Section 4 addresses the relevance of our theorems to applied problems, and discusses them in comparison with Hill's work.

The theorems are expressed in the formal language of de Finetti's operational subjective theory of probability, the context which motivated their development. We recognize that this language may not be too familiar, even to readers who are partially acquainted with de Finetti's general ideas. Since we would like to expose this language and the manner of thinking it embeds, our proofs of the three theorems are constructive and expository, supplemented by geometrical representations and examples. The logic of our proof scheme is based upon the central result of de Finetti (1937, 1949, 1974, 1975) that coherency requires a prevision vector $P(\mathbf{X}_N)$ to lie within the convex hull of the set of *possible* observation vectors for $\mathbf{X}_N$, called the realm of $\mathbf{X}_N$, and denoted by $\mathfrak{R}(\mathbf{X}_N)$. A detailed exposition of the technical language of the operational subjective theory and a generalization of de Finetti's fundamental theorem can be found in the article of Lad, Dickey, and Rahman (1990).

THEOREM 1. If $E_1, \ldots, E_{N+1}$ are logically independent events that are regarded exchangeably, then it is coherent to assert concomitantly the $(N-1)$ conditional previsions $P[E_{N+1} | (S_N = a)] = a/N$ for *every* conditioning event satisfying the restriction $S_N \equiv \sum_{i=1}^{N} E_i = a \in \{1, 2, \ldots, (N-1)\}$. But any cohering distribution assertion for the sum of all N+1 events, S_{N+1}, must be representable in the following two parameter form (with parameters denoted by $q_{0,N+1}$ and $q_{1,N+1}$):

$$P(S_{N+1} = 0) = q_{0,N+1},$$

$$P(S_{N+1} = 1) = q_{1,N+1},$$

$$P(S_{N+1} = a) = q_{a,N+1} = (1/a)[N/(N-a+1)]\, q_{1,N+1} \quad \text{for } a = 2, \ldots, N, \text{ and}$$

$$P(S_{N+1} = N+1) = q_{N+1,N+1} = 1 - q_{0,N+1} - q_{1,N+1}\, 2[N/(N+1)][1 + \Psi(N+1) - \Psi(2)],$$

where $0 \leq q_{0,N+1} \leq 1$, and

$$0 \leq q_{1,N+1} \leq (1 - q_{0,N+1})/\{2[N/(N+1)][1 + \Psi(N+1) - \Psi(2)]\} \leq 1.$$

Here $\Psi(\cdot)$ represents the psi-function, defined for $x \geq 2$ by $\Psi(x) = -\gamma + \sum_{a=1}^{x-1} a^{-1}$, where γ is Euler's constant. If the specified conditional previsions are asserted for *every* positive integer value of N, then coherency requires that for each value of N it also be asserted that $P[(S_N = 0) + (S_N = N)] = 1$. ◆

THEOREM 2. Suppose $E_1, \ldots, E_{N+1}$ are logically independent events that are regarded exchangeably. Any vector $\mathbf{p}_{N+1} \equiv (p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$ of conditional prevision assertions $p_{a,N} \equiv P[E_{N+1} | (S_N = a)]$ for $a = 0, 1, \ldots, N$ is coherent as long as it lies within or on the boundary of the (N+1)-dimensional unit hypercube. Coherency requires that such assertions be accompanied by the specific upper triangular matrix of all lower order conditional previsions, $\mathbf{P}_{N+1,N+1} \equiv (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N+1})$, that is generated by the recursive equations

$p_{a-1,K-1} = p_{a-1,K} / [1 - p_{a,K} + p_{a-1,K}]$ for $K \le N$. The assertions $\mathbf{p}_{N+1}$ are infinitely exchangeably extendible if and only if the associated main diagonal vector of $\mathbf{P}_{N+1,N+1}$, $(p_{0,0}, p_{1,1}, \ldots, p_{N,N})$, generates a sequence of successive products $\{1, p_{00}, p_{00}p_{11}, \ldots, \prod_{i=0}^{N} p_{ii}\}$ that begins a completely monotone sequence. Computable conditions expressed in terms of Hankel determinants are detailed in the text of our proof in Section 2. A necessary condition for infinite exchangeable extendibility of the assertions $\mathbf{p}_{N+1}$ is that every row of $\mathbf{P}_{N+1,N+1}$ be decreasing, and that every column and every diagonal be increasing. But this condition is not sufficient. Finally, each component $p_{a,K}$ of a coherent conditional prevision matrix $\mathbf{P}_{N+1,N+1}$ can be parameterized in the form

$$P[E_{K+1} \mid (S_K = a)] = (a + \alpha_{a,K})/(K + \alpha_{a,K} + \beta_{a,K}) \quad \text{for } 0 \le a \le K \le N,$$

where $\alpha_{a,K}$ and $\beta_{a,K}$ are identified for $0 \le a \le K-1$ by the equations

$$\alpha_{a,K} = -a + p_{a,(K-1)}\, p_{a,K} / [\, p_{a,(K-1)} - p_{a,K}\,] \ ,$$

and $$\beta_{a,K} = -(K-a) + p_{a,(K-1)}\,(1 - p_{a,K}) / [\, p_{a,(K-1)} - p_{a,K}\,] \ ,$$

while $\alpha_{K,K} = \alpha_{K-1,K}$, and $\beta_{K,K} = \beta_{K-1,K}$. ◆

The statement of Theorem 3 is made most succinctly if we first introduce the following definitions and notation.

Definition 1. Suppose $\mathbf{X}_N = (X_1, \ldots, X_N)^T$ denotes a vector of N quantities, and that the realm of possibility for each component quantity is the set of integers $\mathfrak{R}(X_i) = \{0, 1, \ldots, K\}$. The **(K+1)-category histogram** for these quantities is defined by the vector $\mathbf{H}_{K+1}(\mathbf{X}_N) \equiv (H_0(\mathbf{X}_N), H_1(\mathbf{X}_N), \ldots, H_K(\mathbf{X}_N))^T$, where the a^{th}

component, $H_a(\mathbf{X}_N) \equiv \sum_{i=1}^{N} (X_i = a)$ is the number of the quantities that equal the number $a \in \{0, 1, \ldots, K\}$. In general, the realm of possibility for the histogram is the set of all (K+1)-dimensional vectors of nonnegative integers whose component sum equals N:

$$\mathfrak{R}(\mathbf{H}_{K+1}(\mathbf{X}_N)) = \{\mathbf{h}_{K+1} \mid \mathbf{1}^T_{K+1}\mathbf{h}_{K+1} = N, \ \text{and } h_a \in \{0, 1, \ldots, N\}\}\ .$$ •

Since every histogram considered in this article contains K+1 categories, we shall suppress this dimension notation in all further references to histogram vectors $\mathbf{H}$ or $\mathbf{h}$.

Definition 2. A histogram $\mathbf{h}$ is said to be **strictly positive** if each of its components is greater than 0. We denote this by writing $\mathbf{h} > \mathbf{0}$. •

Definition 3. Suppose $\mathbf{Y}_N = (Y_1, \ldots, Y_N)^T$ denotes a vector of N quantities, and the realm of $\mathbf{Y}_N$ is $\mathfrak{R}(\mathbf{Y}_N) = \{\mathbf{y}_{N,1}, \mathbf{y}_{N,2}, \ldots, \mathbf{y}_{N,R}\}$. You are said to regard the quantities $Y_1, \ldots, Y_N$ **subexchangeably** over $\mathfrak{R}^*(\mathbf{Y}_N) \subseteq \mathfrak{R}(\mathbf{Y}_N)$ if your assertions $P(\mathbf{Y}_N = \mathbf{y}_N)$ are constant over permutations of the components of $\mathbf{y}_N$ whenever $\mathbf{y}_N \in \mathfrak{R}^*(\mathbf{Y}_N)$. If $\mathfrak{R}^*(\mathbf{Y}_N) = \mathfrak{R}(\mathbf{Y}_N)$, then you are said to regard the quantities exchangeably. •

Theorem 3. Suppose $\mathbf{X}_{N+1} = (X_1, \ldots, X_{N+1})^T$ denotes a vector of logically independent quantities, and that the realm of possibility for each component quantity is the set of integers $\mathfrak{R}(X_i) = \{0, 1, \ldots, K\}$. Let $\mathbf{X}_N$ denote the vector containing the first N components of $\mathbf{X}_{N+1}$. If you regard the quantities $X_1, \ldots, X_{N+1}$ exchangeably, then it is coherent to assert concomitantly the conditional probability distributions

$$P[(X_{N+1} = a) | (\mathbf{H}(\mathbf{X}_N) = \mathbf{h})] = h_a/N \in (0,1) \quad \text{for each } a \in \{0,\dots,K\}$$

for every strictly positive vector $\mathbf{h} \in \mathfrak{R}(\mathbf{H}(\mathbf{X}_N))$. However, if you do make these assertions, coherency would also require that

(a) you assert the same form of conditional probability distribution when conditioning on a positive histogram computed from any $M < N$ quantities; i.e.,

$$P[(X_{M+1} = a) | (\mathbf{H}(\mathbf{X}_M) = \mathbf{h})] = h_a/M \in (0,1) \quad \text{for each } a \in \{0,\dots,K\};$$

(b) you regard the components of the histogram based on all N+1 quantities, $\mathbf{H}(\mathbf{X}_{N+1}) = (H_0(\mathbf{X}_{N+1}), H_1(\mathbf{X}_{N+1}), \dots, H_K(\mathbf{X}_{N+1}))$, subexchangeably over the set of strictly positive histograms, $\{\mathbf{h} | \mathbf{h} \in \mathfrak{R}(\mathbf{H}(\mathbf{X}_{N+1}))$, and $\mathbf{h} > \mathbf{0}\}$;

(c) your $P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K, 1, 1, \dots, 1)] \geq P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}]$ for every strictly positive histogram $\mathbf{h}$; precisely, for any strictly positive $\mathbf{h}$, your

$$P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}] = [(N+1-K)/\prod_{a=0}^{K} h_a]\, P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K, 1, 1, \dots, 1)]; \quad \text{and}$$

(d) for any fixed integer $M \leq N$, your probability for achieving a strictly positive histogram from the first M quantities, $P[\mathbf{H}(\mathbf{X}_M) > \mathbf{0}]$, is bounded above by a computable number $\epsilon(M,N,K)$. As a function of N, with M and K fixed, this upper bound converges to 0 as $N \to \infty$. ◆

1. Proof of Theorem 1: on the coherency of conditional prevision assertions that avow $P[E_{N+1} | (S_N = a)] = a/N$ for each $a \in \{1, \dots, (N-1)\}$, and their surprising extendibility condition for every N. We shall begin with a general specification of notation, followed by a numerical example for the case N = 3. Then we shall proceed with the proof of Theorem 1, continuing the example for expository purposes.

Consider a vector of (N+1) completely logically independent events, $\mathbf{E}_{N+1} \equiv (E_1, \dots, E_{N+1})^T$, which can be represented as

$$\mathbf{E}_{N+1} = \mathbf{R}_{N+1, 2^{N+1}} \mathbf{C}_{2^{N+1}} . \tag{1.1}$$

The matrix $\mathbf{R}_{N+1,2^{N+1}}$ includes as its columns all the (N+1)-dimensional vectors that are *possible* numerical values of the vector $\mathbf{E}_{N+1}$. We presume that the columns of $\mathbf{R}_{N+1,2^{N+1}}$ are weakly ordered by the rule that the column vector $\mathbf{R}_{.j}$ precedes $\mathbf{R}_{.k}$ if the sum of its components is smaller: $\mathbf{1}^T\mathbf{R}_{.j} < \mathbf{1}^T\mathbf{R}_{.k}$. The symbol $\mathbf{1}$ denotes an appropriately dimensioned vector of 1's. Details of the ordering of columns whose sums are equal will be irrelevant to our argument. The components of the column vector $\mathbf{C}_{2^{N+1}}$ are the constituents of the partition specified by the events $(\mathbf{E}_{N+1} = \mathbf{R}_{.j})$, for values of $j = 1, \dots, 2^{N+1}$. These constituent events are the 2^{N+1} summands that appear in the expansion of the product $\prod_{i=1}^{N}(E_i + \tilde{E}_i)$, where $\tilde{E}_i \equiv (1 - E_i)$. This product necessarily equals 1.

The judgment to regard the events $E_1, \dots, E_{N+1}$ exchangeably means that you regard as equi-likely any sequence of these events that yield the same sum (the same number of "successes" among them). Notice that for each value of $a = 0, 1, \dots, N+1$, there are ${}^{(N+1)}C_a$ distinct sequences that yield the sum $S_{N+1} = a$. Thus, algebraically the judgment of exchangeability implies that your prevision for the constituent vector $\mathbf{C}_{2^{N+1}}$ is representable as

$$P(\mathbf{C}_{2^{N+1}}) \;=\; \mathbf{M}_{2^{N+1},(N+2)}\;\mathbf{q}_{N+2}\quad, \qquad (1.2)$$

where

$\mathbf{M}_{2^{N+1},(N+2)}$ is a matrix whose j^{th} column contains only ${}^{(N+1)}C_{(j-1)}$ nonzero components, each equal to $1/{}^{(N+1)}C_{(j-1)}$; these nonzero components appear in column j immediately following the first $\sum_{t=0}^{j-2} {}^{(N+1)}C_t$ components, which each equal 0; and

$\mathbf{q}_{N+2}$ is some vector in the (N+1)-dimensional simplex,

$$\{\, \mathbf{q}_{N+2} \equiv (q_{0,N+1}, q_{1,N+1}, \ldots, q_{N+1,N+1})^T \,|\, \mathbf{q}_{N+2} \geq \mathbf{0}_{N+2},\ \text{and}\ \mathbf{1}^T\mathbf{q}_{N+2} = 1\} \;.$$

The vector $\mathbf{q}_{N+2}$ would represent your previsions for the partition generated by the sum of the (N+1) events, $\{(S_{N+1}=0), (S_{N+1}=1), \ldots, (S_{N+1}=N+1)\}$, if only you would assert your previsions for these constituents.

The linearity of coherent prevision and your judgment to regard the events $\mathbf{E}_{N+1}$ exchangeably imply that your prevision for the event vector can be represented as

$$P(\mathbf{E}_{N+1}) \;=\; \mathbf{R}_{N+1,2^{N+1}}\;\mathbf{M}_{2^{N+1},(N+2)}\;\mathbf{q}_{N+2}\;,$$

for some vector $\mathbf{q}_{N+2}$ in the (N+1)-dimensional simplex. The goal of our analysis is to study what further restrictions would be placed on the allowable vectors $\mathbf{q}_{N+2}$ by coherency if you would assert concomitantly the (N-1) conditional previsions $P[E_{N+1} \,|\, (S_N = a)] = a/N$ for all values of $a \in \{1,2,\ldots,(N-1)\}$.

First, **a comment on our notation** and a numerical example will be helpful. Throughout this paper we subscript vectors and matrices by their dimensions. Your understanding of our argument will be aided by your explicit recognition now that the vector $\mathbf{q}_{N+2}$ has one component for each of the possible values of the sum of (N+1) events, viz., $\mathbf{q}_{N+2} \equiv (q_{0,N+1}, q_{1,N+1}, \ldots, q_{N+1,N+1})^T$.

EXAMPLE (N=3): In order to fix familiarity with our notation and construction, let us display the vectors $\mathbf{E}_{N+1}$ and $\mathbf{C}_{2^{N+1}}$, and the matrices $\mathbf{R}_{N+1,2^{N+1}}$ and $\mathbf{M}_{2^{N+1},(N+2)}$ for the case of N = 3. In this case,

$$\mathbf{E}_4 = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix}; \qquad \mathbf{R}_{4,16} = \begin{pmatrix} 0&1&0&0&0&1&1&1&0&0&0&1&1&1&0&1 \\ 0&0&1&0&0&1&0&0&1&1&0&1&1&0&1&1 \\ 0&0&0&1&0&0&1&0&1&0&1&1&0&1&1&1 \\ 0&0&0&0&1&0&0&1&0&1&1&0&1&1&1&1 \end{pmatrix};$$

$$\mathbf{C}_{16} = \begin{pmatrix} \tilde{E}_1\tilde{E}_2\tilde{E}_3\tilde{E}_4 \\ E_1\tilde{E}_2\tilde{E}_3\tilde{E}_4 \\ \tilde{E}_1E_2\tilde{E}_3\tilde{E}_4 \\ \tilde{E}_1\tilde{E}_2E_3\tilde{E}_4 \\ \tilde{E}_1\tilde{E}_2\tilde{E}_3E_4 \\ E_1E_2\tilde{E}_3\tilde{E}_4 \\ E_1\tilde{E}_2E_3\tilde{E}_4 \\ E_1\tilde{E}_2\tilde{E}_3E_4 \\ \tilde{E}_1E_2E_3\tilde{E}_4 \\ \tilde{E}_1E_2\tilde{E}_3E_4 \\ \tilde{E}_1\tilde{E}_2E_3E_4 \\ E_1E_2E_3\tilde{E}_4 \\ E_1E_2\tilde{E}_3E_4 \\ E_1\tilde{E}_2E_3E_4 \\ \tilde{E}_1E_2E_3E_4 \\ E_1E_2E_3E_4 \end{pmatrix}; \quad \text{and} \qquad \mathbf{M}_{16,5} = \begin{pmatrix} 1&0&0&0&0 \\ 0&1/4&0&0&0 \\ 0&1/4&0&0&0 \\ 0&1/4&0&0&0 \\ 0&1/4&0&0&0 \\ 0&0&1/6&0&0 \\ 0&0&1/6&0&0 \\ 0&0&1/6&0&0 \\ 0&0&1/6&0&0 \\ 0&0&1/6&0&0 \\ 0&0&1/6&0&0 \\ 0&0&0&1/4&0 \\ 0&0&0&1/4&0 \\ 0&0&0&1/4&0 \\ 0&0&0&1/4&0 \\ 0&0&0&0&1 \end{pmatrix} .$$

In particular, notice the satisfaction of the ordering property for the columns of the matrix $\mathbf{R}_{4,16}$. Column j precedes column k if the sum of its components is smaller. We shall continue this numerical example throughout our development.

•••

Having specified our notation and usage, let us proceed with the proof of Theorem 1, identifying what further restrictions would be placed on the vector $\mathbf{q}_{N+2}$ if you would assert the conditional previsions $P[E_{N+1} | (S_N = a)] = a/N$ for all values of $a \in \{1,2,\ldots,(N-1)\}$. The affirmations of the theorem are that such a collection of assertions is coherent, but that coherency implies specific further restrictions on the probability distribution asserted for the sum of the N+1 events, S_{N+1}. Corollary to the theorem we shall also identify interesting restrictions on the mere prevision for this sum, $P(S_{N+1})$, and on the probability distribution for the sum of any $M < N+1$ of the events, $P(S_M = 0)$, $P(S_M = 1)$, ... , $P(S_M = M)$.

Now the coherency of conditional prevision requires that for any quantity X and event E, asserted previsions must satisfy the relation

$$P(XE) = P(X|E)\ P(E).$$

Thus, when the conditional prevision $P(X|E)$ is asserted as a specific number, coherency again requires the concomitant assertion of

$$P[XE - P(X|E)E] = 0\ ,$$

where the symbol $P(X|E)$ in this expression is replaced by whatever number has been asserted. In fact, this statement is virtually the operational definition of a conditional prevision assertion, $P(X|E)$. In the context of the sequence of events we are considering, this coherency requirement means that the assertions we have specified in the form $P[E_{N+1} | (S_N = a)] = a/N$ for each $a \in \{1,2,\ldots,(N-1)\}$ are equivalent to the assertions $P[E_{N+1}(S_N = a) - (a/N)(S_N = a)] = 0$ for the same values of a. We can represent these assertions in our setup as follows.

To the matrix equation (1.1) which represents the (N+1) events as

$$\mathbf{E}_{N+1} = \mathbf{R}_{N+1,2^{N+1}}\ \mathbf{C}_{2^{N+1}}\ ,$$

let us append another matrix equation that identifies the (N-1) quantities whose previsions are restricted by coherency. We do this by defining a vector of quantities denoted by $\mathbf{A}_{N-1} = (A_{1,N},\ldots,A_{N-1,N})^T$, defined componentwise by the equations

$$A_{a,N} \equiv [E_{N+1}(S_N = a) - (a/N)(S_N = a)] \quad \text{for } a = 1,2,\ldots,(N-1). \qquad (1.3)$$

Since each of the quantities $A_{a,N}$ is defined functionally in terms of the event vector $\mathbf{E}_{N+1}$, we can write each quantity $A_{a,N}$ as a linear combination of the constituents of the partition they generate, $A_{a,N} = \mathbf{r_a}^T{}_{2^{N+1}}\ \mathbf{C}_{2^{N+1}}$, where the coefficient vector $\mathbf{r_a}^T{}_{2^{N+1}}$ identifies elements of $\mathbf{C}_{2^{N+1}}$ that represent event sequences for which the sum of the first N events equals a. That is, the symbol $\mathbf{r_a}^T{}_{2^{N+1}}$ denotes a row vector whose j^{th} component equals 0 if the sum of the first N components of the j^{th} column of $\mathbf{R}_{N+1,2^{N+1}}$ does not equal a, and whose j^{th} component equals the value of the $(N+1)^{st}$ component less a/N if the sum does equal a. Evidently, the possible values of the quantity $A_{a,N}$ are 0, (1−a/N), and (−a/N).

<u>EXAMPLE</u> (N=3) continued: Before considering a general algebraic expression for the vector $\mathbf{r_a}^T{}_{2^{N+1}}$ let us display the structure numerically in the application to N=3:

$\mathbf{r_1}^T{}_{16}$ = (0 −1/3 −1/3 −1/3 0 0 0 2/3 0 2/3 2/3 0 0 0 0 0)

$\mathbf{r_2}^T{}_{16}$ = (0 0 0 0 0 −2/3 −2/3 0 −2/3 0 0 0 1/3 1/3 1/3 0).

These vectors should be evident from applying their generating definition to the matrix $\mathbf{R}_{4,16}$ that was displayed in the introduction of this example. •••

Though this example is small, it should solidify your awareness of two general features of a vector $\mathbf{r_a}^T{}_{2^{N+1}}$. Since the sum of the first N components of the j^{th} column of $\mathbf{R}_{N+1,2^{N+1}}$ can equal a only if the sum of all (N+1) components equals either a or (a+1), the first $\sum_{j=0}^{a-1} {}^{(N+1)}C_j$ components of $\mathbf{r_a}^T{}_{2^{N+1}}$ must equal 0, and the last $\sum_{j=a+2}^{N+1} {}^{(N+1)}C_j$ components of $\mathbf{r_a}^T{}_{2^{N+1}}$ must equal 0. This is by virtue of the arbitrary ordering we have chosen for the column vectors of the matrix $\mathbf{R}_{N+1,2^{N+1}}$. Furthermore, of the first ${}^{(N+1)}C_a$ components following the first $\sum_{j=0}^{a-1} {}^{(N+1)}C_j$ components that equal 0, ${}^{N}C_a$ of them will equal -(a/N) and the remainder will equal 0. For of those columns of $\mathbf{R}_{N+1,2^{N+1}}$ that sum to a, only ${}^{N}C_a$ sum to a over their first N components, and of these, all exhibit 0 as their $(N+1)^{st}$ component. In similar fashion, of the next ${}^{(N+1)}C_{(a+1)}$ components of $\mathbf{r_a}^T{}_{2^{N+1}}$ just preceding the final $\sum_{j=a+2}^{N+1} {}^{(N+1)}C_j$ components that equal 0, ${}^{N}C_a$ of them will equal [1-(a/N)] and the remainder will equal 0. For of those columns of $\mathbf{R}_{N+1,2^{N+1}}$ that sum to (a+1), only ${}^{N}C_a$ sum to a over their first N components, and of these, all exhibit 1 as their $(N+1)^{st}$ component. For reasons that we shall now explain, further details of the ordering of components of $\mathbf{r_a}^T{}_{2^{N+1}}$ are irrelevant to the remainder of our argument.

Denoting by $\mathbf{R(A)}_{N-1,2^{N+1}}$ the matrix whose (N-1) rows are the vectors we have been describing as $\mathbf{r_a}^T{}_{2^{N+1}}$ for a = 1,2,...,(N-1), we can consolidate all the information about the quantities we have defined for this problem, writing

$$\begin{bmatrix} \mathbf{E}_{N+1} \\ \mathbf{A}_{N-1} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{N+1,2^{N+1}} \\ \mathbf{R(A)}_{N-1,2^{N+1}} \\ \mathbf{1}^T{}_{2^{N+1}} \end{bmatrix} \mathbf{C}_{2^{N+1}} \quad . \tag{1.4}$$

The three block components of this matrix equation (1.4) denote that

i) the events composing the vector $\mathbf{E}_{N+1}$ are logically independent;

ii) the (N-1) quantities composing $\mathbf{A}_{N-1}$ are logically dependent, specified by linearly independent funtions of the basic (N+1) events, $\mathbf{E}_{N+1}$; and

iii) the events composing the vector $\mathbf{C}_{2^{N+1}}$ are the constituents of the partition generated by the components of $\mathbf{E}_{N+1}$. Thus, they sum to 1.

Let us turn to the further structure imposed on the problem by the conditional prevision assertions presumed in Theorem 1. Since the quantities we have denoted by the vector $\mathbf{A}_{N-1}$ are linear functions of $\mathbf{C}_{2^{N+1}}$, to be explicit $\mathbf{R(A)}_{N-1,2^{N+1}}\mathbf{C}_{2^{N+1}}$, the linearity of coherent prevision requires that the presumed $P(\mathbf{A}_{N-1}) = \mathbf{0}_{N-1}$ be representable as

$$\mathbf{R(A)}_{N-1,2^{N+1}} P(\mathbf{C}_{2^{N+1}}) = \mathbf{R(A)}_{N-1,2^{N+1}} \mathbf{M}_{2^{N+1},(N+2)} \mathbf{q}_{N+2} = \mathbf{0}_{N-1} ,$$

with the vector $\mathbf{q}_{N+2}$ already constrained to the (N+1)-dimensional simplex. Thus, we have identified the coherency constraints on $\mathbf{q}_{N+2}$ as

$$P \begin{bmatrix} \mathbf{A}_{N-1} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{N-1} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R(A)}_{N-1,2^{N+1}} \mathbf{M}_{2^{N+1},(N+2)} \\ \mathbf{1}^T{}_{N+2} \end{bmatrix} \mathbf{q}_{N+2} \quad . \tag{1.5}$$

We can simplify our analysis by deriving an explicit form for the product matrix $\mathbf{R(A)}_{N-1,2^{N+1}} \mathbf{M}_{2^{N+1},(N+2)}$.

Notice that the matrix product $\mathbf{R(A)}_{N-1,2^{N+1}} \mathbf{M}_{2^{N+1},(N+2)}$ yields a matrix with dimension (N-1)×(N+2). According to the manner in which the multiplicand matrices have been constructed, the a^{th} row of the product matrix exhibits 0 in its first "a" places and in its last (N-a) places. As its $(a+1)^{st}$ component appears the number

$$[\mathbf{R(A)M}]_{a,(a+1)} = {}^{N}C_a \ (-a/N) \ / \ {}^{(N+1)}C_a = -a(N-a+1)/(N+1)N \ .$$

As the $(a+2)^{nd}$ component of the a^{th} row of the product matrix appears the number

$$[\mathbf{R(A)}\,\mathbf{M}]_{a,(a+2)} = {}^{N}C_a\ [1 - a/N]\ /\ {}^{(N+1)}C_{(a+1)} = (a+1)(N-a)/(N+1)N.$$

Thus, the a^{th} row of the product matrix has the form

$$[\mathbf{R(A)}\,\mathbf{M}]_{a\cdot} = (\ \mathbf{0}^T_a\quad -a(N-a+1)/(N+1)N\quad (a+1)(N-a)/(N+1)N\quad \mathbf{0}^T_{(N-a)}\ ).$$

<u>EXAMPLE</u> (N=3) continued: Our recurring numerical example displays this result. The matrix $\mathbf{R(A)}_{2,16}$ is the matrix whose two rows are the vectors $\mathbf{r}_1{}^T{}_{16}$ and $\mathbf{r}_2{}^T{}_{16}$ displayed in this example just above. Thus, simple multiplication with $\mathbf{M}_{16,5}$ shows that

$$\mathbf{R(A)}_{2,16}\,\mathbf{M}_{16,5} = \begin{bmatrix} 0 & -1/4 & 1/3 & 0 & 0 \\ 0 & 0 & -1/3 & 1/4 & 0 \end{bmatrix}.$$

Each of these rows exhibits the form we have just derived. •••

Thus, the conditional prevision assertions we are studying, $P[E_{N+1} \mid (S_N = a)] = a/N$ for values of $a = 1,2,\ldots,(N-1)$, place (N-1) more linear restrictions on the vector $\mathbf{q}_{N+2}$, which is already required to lie in the (N+1)-dimensional simplex. So there remain only 2 free components of $\mathbf{q}_{N+2}$ allowed by the satisfaction of all N linear constraints. And even these are not completely free. Surely, for example, their sum cannot exceed 1. But they are constrained even further. If we arbitrarily denote the free components by q_0 and q_1 (suppressing for the remainder of this Section their full denotations of $q_{0,N+1}$ and $q_{1,N+1}$) then the (N-1) restrictions represented by the equation

$$\mathbf{R(A)}_{N-1,2^{N+1}}\ \mathbf{M}_{2^{N+1},(N+2)}\ \mathbf{q}_{N+2} = \mathbf{0}_{N-1}$$

yield the recursive solution

$$q_a = [(a-1)/a][(N-a+2)/(N-a+1)]\,q_{a-1} \quad \text{for } a = 2, \ldots, N, \tag{1.6}$$

and $$q_{N+1} = 1 - \sum_{i=0}^{N} q_i\ ,$$

which comes from the simplicial requirement that $\sum_{i=0}^{N+1} q_i = 1$. This final equation identifies the restrictions on both q_0 and q_1.

The direct form of the recursive equations (1.6) shows all the components of $\mathbf{q}_{N+2}$ as functions of q_0 and q_1, viz.,

$$q_a = (1/a)\ [N/(N-a+1)]\ q_1 \quad \text{for} \quad a = 2,\ldots,N, \quad \text{and}$$
$$q_{N+1} = 1 - q_0 - q_1\,\{1 + N \sum_{a=2}^{N} [a(N-a+1)]^{-1}\}. \tag{1.7}$$

Now the summation that appears in this restriction on q_{N+1} can be rewritten as

$$\begin{aligned} {}_2\Sigma_a{}^N [a(N-a+1)]^{-1} &= (N+1)^{-1} [\sum_{a=2}^{N} a^{-1} + \sum_{a=2}^{N} (N-a+1)^{-1}] \\ &= (N+1)^{-1} [1 - N^{-1} + 2 \sum_{a=2}^{N} a^{-1}] \\ &= (N-1)/[N(N+1)] + 2\,[\Psi(N+1) - \Psi(2)]/(N+1), \end{aligned} \tag{1.8}$$

where $\Psi(\cdot)$ is the so-called "psi-function" or "digamma-function" defined for $x \geq 2$ by

$$\Psi(x) \equiv -\gamma + \sum_{a=1}^{x-1} a^{-1}\ ,$$

and γ is Euler's constant:

$$\gamma \equiv \lim_{x\to\infty} [\sum_{a=1}^{x} a^{-1} - \log x] = .5772156649\ldots\ .$$

See Abramovitz and Stegun (1964, p. 255). If you are not familiar with it, notice that the psi-function is the discrete (integer) analogue of the logarithm function. Now replacing the summation in (1.7) with this derived equivalent in (1.8), we arrive at the representation

$$q_{N+1} = 1 - q_0 - q_1 2[N/(N+1)]\{1 + \Psi(N+1) - \Psi(2)\}\ .$$

Thus, the only restrictions on q_0 and q_1 are that they be nonnegative and that $q_0 + q_1 2[N/(N+1)]\{1 + \Psi(N+1) - \Psi(2)\} \leq 1$; or equivalently, that

$$q_1 \leq (1-q_0)\ /\ \{2[N/(N+1)]\{1 + \Psi(N+1) - \Psi(2)\}\} \equiv B_{N+1}(q_0)\ , \tag{1.9}$$

where $\Psi(\cdot)$ is the unbounded increasing psi-function mentioned above. The function $B_{N+1}(q_0)$ expresses the upper bound on q_1 associated with the value of q_0. Notice that the limit of this upper bound on q_1 decreases to 0 as N becomes large: $\lim_{N\to\infty} B_{N+1}(q_0) = 0$.

Thus far we can conclude that it is coherent to assert the conditional previsions $P[E_{N+1} | (S_N = a)] = a/N$ for $a = 1,2,\ldots,(N-1)$. And we have precise algebraic expressions for the 2-dimensional space of cohering vectors $\mathbf{q}_{N+2}$ identified as a function of N, specified by equations (1.7). A further question we can address is what are the consequences of asserting conditional previsions of this type for *every* value of N. The answer is the surprising result stated in the final statement of our Theorem 1. Proving it requires a little more work.

Consider your previsions regarding the sum of only M of the (N+1) events that you regard exchangeably, where $M < (N+1)$ is a fixed number. It is well known from the work of Diaconis (1977) and others that coherency requires your previsions for the partition $(S_M = 0), (S_M = 1), \ldots, (S_M = M)$ must be related to the vector $\mathbf{q}_{N+2}$ by the hypergeometric mixture equations

$$P(S_M = j) = \sum_{i=0}^{N+1-M} [\,{}^{(j+i)}C_j \; {}^{(N+1-j-i)}C_{(M-j)} \,/\, {}^{(N+1)}C_M\,]\; q_{(j+i),(N+1)} \quad . \qquad (1.10)$$

We can replace the values of $q_{(j+i),(N+1)}$ in equation (1.10) by their coherency required equivalents, derived for values of $j = 1,\ldots,(M-1)$ from equation (1.7) as

$$q_{(j+i),(N+1)} = N \,/\, [(j+i)(N+1-j-i)]\; q_{1,N+1} \quad .$$

Performing some algebraic reduction after this substitution then yields the result

$$P(S_M = j) = {}^{M}C_j\,(N+1)^{-1}\, q_{1,N+1} \sum_{i=0}^{N+1-M} {}^{(N+1-M)}C_i \,/\, {}^{(N-1)}C_{(i+j-1)} \;, \qquad (1.11)$$

again for the values of $j = 1,\ldots,(M-1)$. Now the summands in equation (1.11) satisfy the inequality

$${}^{(N+1-M)}C_i \,/\, {}^{(N-1)}C_{(i+j-1)} \le 1 \quad \text{for } i = 0,1,\ldots,(N+1-M). \qquad (1.12)$$

This can be seen by substituting $I = N+1-M$, $S = M-2$, and $s = j-1$. Then (1.12) reduces to the well known form ${}^{I}C_i \le {}^{(I+S)}C_{(i+s)}$ for $i \le I$ and $0 \le s \le S$. Applying (1.12) to equation (1.11) yields the inequalities

$$P(S_M = j) \le {}^{M}C_j\; [(N+2-M)/(N+1)]\; q_{1,N+1} \quad \text{when } 1 \le j \le (M-1)\,. \qquad (1.13)$$

Since we have already seen in (1.9) that coherency requires $q_{1,N+1} \le B_{N+1}(q_0) \to 0$ as $N \to \infty$, we can now conclude that for any finite integer M, $P(1 \le S_M \le (M-1))$ must equal 0. This substantiates the final statement that completes Theorem 1: if you coherently assert the conditional previsions

$P[E_{N+1} | (S_N = a)] = a/N$ for values of $a \in \{1,2,\ldots,(N-1)\}$ for *every* value of N,

then coherency requires that for each value of N you also assert the prevision

$$P[(S_N = 0) + (S_N = N)] = 1.$$

Our proof of Theorem 1 is complete. We conclude this Section with a geometrical representation of an example of the theorem, and a recognition of two associated corollaries.

Figure 1.1 displays the set of pairs $(q_{0,N+1}, q_{1,N+1})$ that meet the specified restrictions of Theorem 1 for increasing values of N, and translates them into the set of M-tuples they would imply for cohering previsions for the partition $(S_M = 0), (S_M = 1), \ldots, (S_M = M)$ in the special case of $M = 2$, displayed in a barycentric coordinate system. In this case the transformation equations (1.11) reduce to yield $P(S_2 = 0), P(S_2 = 1), P(S_2 = 2))$ as

$$P(S_2 = 0) = T_0(q_0,q_1) = q_{0,N+1} + (N+1)^{-1}\{1 + N[\Psi(N) - \Psi(2)]\}\; q_{1,N+1} \quad ,$$

$$P(S_2 = 1) = T_1(q_0,q_1) = 2[N/(N+1)]\; q_{1,N+1} \quad , \quad \text{and}$$

$$P(S_2 = 2) = T_2(q_0,q_1) = 1 - q_{0,N+1} - (N+1)^{-1}\{(2N+1) + N[\Psi(N) - \Psi(2)]\}\; q_{1,N+1}\,.$$

Since the transformation from $(q_{0,N+1}, q_{1,N+1})$ to $\mathbf{T}(q_{0,N+1}, q_{1,N+1})$ is linear, the transformation of the three *extreme* feasible pairs of $(q_{0,N+1}, q_{1,N+1})$ for each value of N yields the three vertices of the sub-simplex of cohering values of prevision over the partition $\{(S_2=0),(S_2=1),(S_2=2)\}$. The extreme feasible pairs of $(q_{0,N+1}, q_{1,N+1})$ are the three extreme points (0,0), $(0, B_{N+1}(q_0))$, and (1,0), which satisfy equation (1.9). Notice in Figure 1.1 that the feasible set of prevision vectors $(P(S_2=0), P(S_2=1), P(S_2=2))$ diminishes as N increases, reducing in the limit to the line segment connecting the base vertices of the simplex, (1,0,0) and (0,0,1). This corresponds to our algebraic recognition in proving Theorem 1 that $P(S_M=j) \to 0$ as $N \to \infty$ for values of $j = 1, \ldots, (M-1)$.

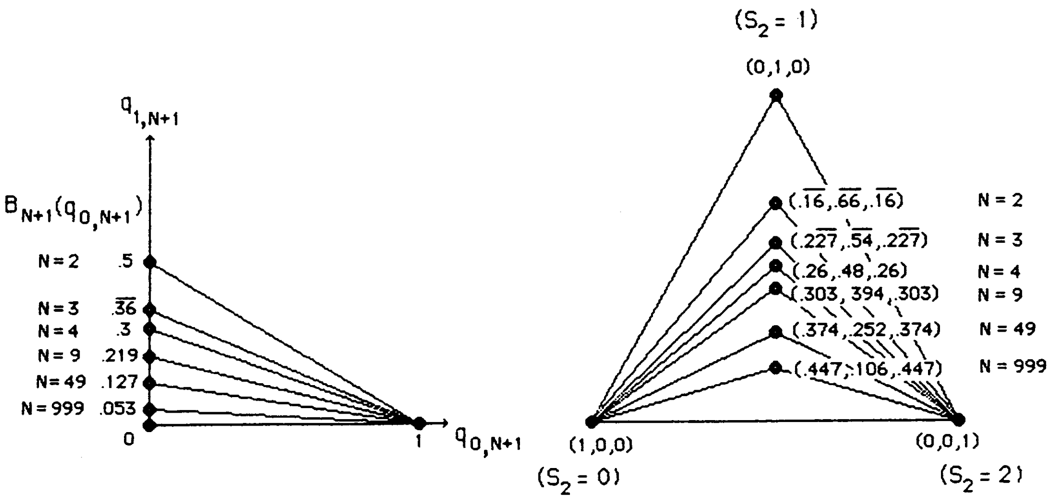

Figure 1.1. For increasing values of N, the set of pairs $(q_{0,N+1}, q_{1,N+1})$ that satisfy $0 \le q_0 \le 1$, and $0 \le q_1 \le (1-q_0)/\{2[N/(N+1)][1+\Psi(N+1)-\Psi(2)]\}$, is translated into the sub-simplex of cohering prevision assertions over the partition $\{(S_2=0),(S_2=1),(S_2=2)\}$.

The limiting result, which applies when the conditional prevision assertions $P(E_{N+1} | (S_N = a)) = a/N$ for $a = 1, \ldots, (N-1)$ are made *for every value of N*, relegates these conditional prevision assertions to assertions conditioned on events that are themselves assessed with probability 0. It is worth noticing that if one asserts $P(1 \le S_N \le N-1) = 0$, then *any* assertions $P(E_{N+1} | (S_N = a))$ for $a = 1, \ldots, (N-1)$ would be coherent.

Corollory to Theorem 1 are two further items of interest. Firstly, the conditions of the theorem allow great freedom in specifying one's prevision for the sum of the N+1 events under consideration. We prove this forthwith.

COROLLARY 1.1. Under the conditions of Theorem 1, coherency would allow the assertion of $P(S_{N+1})$ as any number within $[0, (N+1)]$.

PROOF: Using the coherency restriction that $P(S_{N+1}) = \sum_{a=0}^{N+1} a\, q_{a,N+1}$, and substituting the representations of $\mathbf{q}_{N+2}$ identified in equation (1.7) yields

$$P(S_{N+1}) = 0q_0 + 1q_1 + \sum_{a=2}^{N} [N/(N-a+1)]q_1$$
$$+ (N+1)\{1 - q_0 - q_1 2[N/(N+1)][1+\Psi(N+1)-\Psi(2)]\}$$

$$= (N+1)(1-q_0) - N q_1 [1 + \Psi(N+1) - \Psi(2)]. \qquad (1.14)$$

Thus, $P(S_{N+1})$ would equal a constant C for all allowable (q_0, q_1) pairs satisfying

$$q_{1,N+1} = [(N+1)(1-q_{0,N+1}) - C] \,/\, \{N[1+\Psi(N+1)-\Psi(2)]\}$$

$$= [(N+1)-C]/\{N[1+\Psi(N+1)-\Psi(2)]\} - q_{0,N+1}(N+1)/\{N[1+\Psi(N+1)-\Psi(2)]\}. \qquad (1.15)$$

This latter form is presented to highlight the awareness that the pairs (q_0, q_1) supporting the same assertion value of $P(S_{N+1}) = C$ constitute negatively sloped lines in (q_0, q_1) space. Figure 1.2 depicts such lines superimposed on the convex hull of allowable cohering values of the vector $(q_{0,N+1}, q_{1,N+1})$. Notice in the Figure that the slope of the hypotenuse of the convex hull of coherent assertion values of (q_0, q_1) is less sharply negative than are the slopes of lines supporting the assertion $P(S_{N+1}) = C$. For the hypotenuse portrays the boundary of the coherency restriction (1.9) that

$$q_1 \le (1-q_0)/\{2[N/(N+1)][1 + \Psi(N+1) - \Psi(2)]\} = B_{N+1}(q_0) \ .$$

Comparatively, the slope of this hypotenuse is precisely one-half the slopes of the lines representing constant values of $P(S_{N+1})$ in equation (1.15).

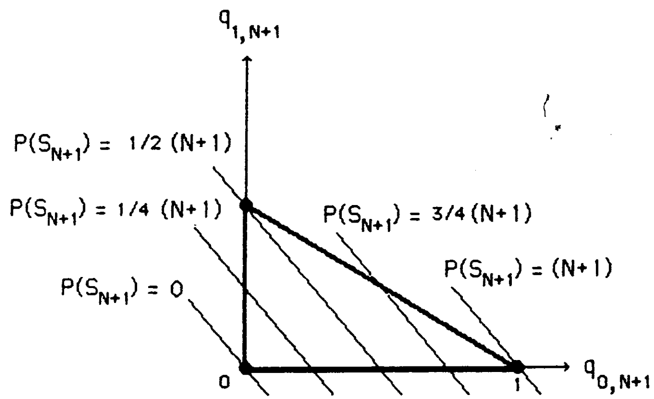

Figure 1.2. The bold outlined triangle identifies values of (q_0, q_1) cohering with conditional prevision assertions $P[E_{N+1} | (S_N = a)] = a/N$ for every $a \in \{1, 2, \ldots, N-1\}$. Lines depicted with slopes twice as sharply negative as the hypotenuse of this triangle constitute pairs of (q_0, q_1) that support constant values of $P(S_{N+1})$ at various levels within $[0, (N+1)]$. This Figure is drawn to scale for $N = 2$.

The geometrical analysis displays that if you assert some value for your $P(S_{N+1})$ in addition to the conditional prevision assertions that we have been studying, this would place one additional linear restriction on the

components of $\mathbf{q}_{N+2}$. That is, $P(S_{N+1}) = \sum_{a=0}^{N+1} a q_{a,N+1}$. This restriction identifies a line segment within the triangle of allowable values for $q_{0,N+1}$ and $q_{1,N+1}$, thus reducing your freedom of choice to specifying a value *either* for $q_{0,N+1} = P(S_{N+1} = 0)$ *or* for $q_{1,N+1} = P(S_{N+1} = 1)$. The pair of these values is then restricted to lying on the appropriate line segment.

Throughout this Section we have focused on the coherent extendibility of conditional probabilities asserting $P[E_{N+1} | (S_N = a)] = a/N$ for each value of $a \in \{1, 2, \ldots, N-1\}$. It is worth noting explicitly that coherency *requires* that such assertions be accompanied by the same form of assertion when conditioning on less than N events. This result constitutes a second Corollary to Theorem 1.

COROLLARY 1.2. Under the conditions of Theorem 1, coherency requires the concomitant assertion of conditional previsions for any $M < N$,

$$P[E_{M+1} | (S_M = a)] = a/M \quad \text{for } \textit{every} \quad a \in \{1, 2, \ldots, (M-1)\}.$$

PROOF: In our derived formula (1.11) for $P(S_M = j)$, merely replace M by (N+1) and j by a. The result is an expression identical to the representation for $P(S_{N+1} = a)$ in equation (1.7), as required by Theorem 1. Thus, the structure of the vector $\mathbf{q}_{M+2}$ is exactly that required to support the conditional probabilities specified. ∇

Corollary 1.2 constitutes the discrete counterpart to Hill's result (1968, 1988), using his terminology, that the assertion of A(n) implies the assertion of A(k) for every $k < n$.

The substance of Theorem 1 contains two intriguing aspects for statistical theorists. Firstly, for the Bayesian theorist who is unduly concerned about inference seemingly based upon "improper prior distributions" when the asserted distribution of events regarded exchangeably is viewed as a

Binomial-Beta mixture: the "problem" with inference based on the improper prior distribution is *not* that there is no coherent distribution that would support such assertions. Rather, the formal "problem" is that there are *many* distinct distributions that would cohere with such inference. But that amounts to no problem at all in the subjective theory of inference. Secondly, for the objectivist frequentist statistician who would like exclusively to use observed frequencies as conditional probability assertions, especially if the number of conditioning quantities is great: a startling result is that coherency would require the annoying concomitant assertion that $P[(S_N = 0) + (S_N = N)] = 1$ for every N. Our analysis shows that a *systematic* strategy to use conditioning frequencies as conditional probabilities would be coherent only when these extreme assertions also seem reasonable. In most cases, they would not.

Theorems 2 and 3, which we now address, generalize Theorem 1 in interesting ways. Section 2 identifies coherency conditions for inferences represented by conditional prevision assertions specified at values other than a/N. Section 3 analyses the situation that the quantities judged exchangeably contain many possible values (but a specific finite number) in their realm, not merely zero and one.

2. Proof of Theorem 2: on the coherency of conditional previsions asserting $P[E_{N+1} | (S_N = a)] \equiv p_{a,N} \in [0,1]$ for each $a \in \{0, \ldots, N\}$, and their infinite extendibility criterion. In formulating Theorem 2, we generalize the conditional prevision assertions we have studied in Section 1 to the full range of conditional prevision assertions that are allowable by coherency with the judgment to regard the (N+1) events exchangeably. That is, we consider the concomitant conditional prevision assertions of the form

$$P[E_{N+1} | (S_N = a)] \equiv p_{a,N} \in [0,1] \quad \text{for every } a \in \{0, \ldots, N\}.$$

Notice that this formulation extends the *number* of conditional previsions asserted by 2, since it specifies previsions for E_{N+1} conditional upon the events $(S_N = 0)$ and $(S_N = N)$ as well as upon the events $(S_N = a)$ for integer values of a between 1 and (N-1). This formulation also generalizes the numerical values of the conditional prevision assertions studied to allow any value of $p_{a,N} \in [0,1]$, rather than necessarily specifying the numbers $p_{a,N} = a/N$. We shall identify coherency conditions on concomitant inferential assertions of $p_{0,N}, p_{1,N}, \ldots, p_{N,N}$.

It will be helpful at this point to firm in your mind the recognition that the symbols $p_{a,N}$ denote values of *conditional* probability assertions, $P[E_{N+1} | (S_N = a)]$, as opposed to the symbols $q_{a,(N+1)}$ which continue to denote values of *unconditional* probability assertions, $P(S_{N+1} = a)$. Moreover, we shall use the symbol $\mathbf{p}_{N+1}$ to denote the (N+1)-dimensional vector $(p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$.

Our constructive program follows that of Section 1 identically, until we come to the point of defining the individual quantities denoted by $A_{a,N}$ in equation (1.3), which we now define more generally as

$$A_{a,N} \equiv [E_{N+1}(S_N = a) - p_{a,N}(S_N = a)] \quad \text{for } a = 0, 1, \ldots, N. \qquad (2.1)$$

Again, each of the quantities $A_{a,N}$ is defined functionally in terms of the event vector E_{N+1}. So again we can write $A_{a,N} = \mathbf{r_a}^T{}_{2^{N+1}} \mathbf{C}_{2^{N+1}}$. But now $\mathbf{r_a}^T{}_{2^{N+1}}$ is a row vector whose i^{th} component equals 0 if the sum of the first N components of the i^{th} column of $\mathbf{R}_{N+1,2^{N+1}}$ does not equal a, and equals the value of the $(N+1)^{st}$ component less $p_{a,N}$ if it does. The possible values of the quantity $A_{a,N}$ are 0, $(1-p_{a,N})$, and $(-p_{a,N})$.

Now defining the matrix $\mathbf{R(A)}_{N+1,2^{N+1}}$ as the matrix whose rows are the (N+1) vectors so defined, we can write

$$\mathbf{A}_{N+1} \; = \; \mathbf{R(A)}_{N+1,2^{N+1}} \; \mathbf{C}_{2^{N+1}} \quad .$$

Assertion of the conditional previsions $\mathbf{p}_{N+1} \equiv (p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$ is equivalent via coherency to the assertion $P(\mathbf{A}_{N+1}) = \mathbf{0}_{N+1}$. Thus, using the exchangeability condition (1.2) we can now write (N+2) linear restrictions on the components of $\mathbf{q}_{N+2}$, which must still lie in the (N+1)-dimensional simplex:

$$\mathbf{0}_{N+1} = P(\mathbf{A}_{N+1}) = \mathbf{R(A)}_{N+1,2^{N+1}} \, \mathbf{M}_{2^{N+1},(N+2)} \, \mathbf{q}_{N+2} \quad , \quad \text{and}$$

$$1 = \mathbf{1}_{N+2}{}^T \, \mathbf{q}_{N+2} \; .$$

Solving these simultaneous equations yields the following solution:

$$q_{0,N+1} = \{ 1 + \sum_{a=0}^{N} {}^{(N+1)}C_{(a+1)} \prod_{i=0}^{a} [p_{i,N}/(1-p_{i,N})] \}^{-1} ,$$

$$\text{and } q_{a,(N+1)} = {}^{(N+1)}C_a \prod_{i=0}^{a-1} [p_{i,N}/(1-p_{i,N})] \; q_{0,N+1} \text{ for } a = 1, \ldots, (N+1). \quad (2.2)$$

Equations (2.2) specify a nonlinear but invertible transformation from the vector of conditonal prevision assertion values $\mathbf{p}_{N+1} = (p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$ to the vector of cohering prevision assertion values $\mathbf{q}_{N+2} = (q_{0,N+1}, \ldots, q_{(N+1),(N+1)})^T$. Evidently,

the components of the vector $\mathbf{q}_{N+2}$ cohering with an asserted vector $\mathbf{p}_{N+1}$ are proportional to the sums of the various products of conditional odds ratios. The form of transformation equation (2.2) makes it evident that any vector $\mathbf{p}_{N+1}$ lying strictly within an (N+1)-dimensional unit hypercube would represent a coherent conditional prevision assertion, since it would transform into a vector $\mathbf{q}_{N+2}$ within the (N+1)-dimensional simplex. It may be clarifying to notice that the transformation formulas (2.2) from $\mathbf{p}_{N+1}$ to $\mathbf{q}_{N+2}$ yield the well known binomial solutions in the special case that you judge the events independently. This case would be identified by your asserting $P[E_{N+1} | (S_N = a)] = p \in [0,1]$ for every $a \in \{0, \ldots, N\}$.

The inverse transformation from $\mathbf{q}_{N+2}$ to $\mathbf{p}_{N+1}$ can be derived as that specified by the nonlinear equations, for $a = 0, 1, \ldots, N$,

$$p_{a,N} = {}^{(N+1)}C_a \, q_{(a+1),(N+1)} \, / \, [{}^{(N+1)}C_{(a+1)} \, q_{a,(N+1)} + {}^{(N+1)}C_a \, q_{(a+1),(N+1)}] \, . \quad (2.3)$$

Equation (2.3) which is also derivable directly from Bayes' Theorem, identifies the remaining coherent conditional prevision vectors $\mathbf{p}_{N+1}$. These are the vectors $\mathbf{p}_{N+1}$ lying on the surfaces of the (N+1)-dimensional simplex. That is, these are the vectors generated using (2.3) beginning with vectors $\mathbf{q}_{N+1}$ that are convex combinations of no more than N of the row vectors of an (N+1)-dimensional identity matrix. We have proved the first assertion of Theorem 2: any vector of conditional prevision assertions $P[E_{N+1} | (S_N = a)] = p_{a,N} \in [0,1]$ for $a = 0, \ldots, N$ is coherent as long as the vector $\mathbf{p}_{N+1}$ lies within or on the boundary of the (N+1)-dimensional hypercube.

The next substantive content of Theorem 2 identifies a structure on the components of the upper triangular matrix of conditional prevision assertions $\mathbf{P}_{N+1,N+1} \equiv (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N+1})$ that coherency requires of anyone who asserts the conditional prevision vector $\mathbf{p}_{N+1}$. We shall prove this part of the theorem by using the transformation we just developed between conditional prevision vectors $\mathbf{p}$ and unconditional prevision vectors $\mathbf{q}$, following the details of the transformation from $\mathbf{p}_{N+1}$ to $\mathbf{q}_{N+2}$ to $\mathbf{q}_{N+1}$ to $\mathbf{p}_N = (p_{0,N-1}, p_{1,N-1}, \ldots, p_{N-1,N-1})^T$.

First, having specified the conditional previsions $\mathbf{p}_{N+1} \equiv (p_{0,N}, \ldots, p_{N,N})^T$, the components of $\mathbf{q}_{N+2} \equiv (q_{0,N+1}, q_{1,N+1}, \ldots, q_{N+1,N+1})^T$ would be those specified by equations (2.2).

Now this vector $\mathbf{q}_{N+2}$ reduces to $\mathbf{q}_{N+1} = (q_{0,N}, q_{1,N}, \ldots, q_{N,N})^T$ by means of the finite exchangeability transformations studied by Diaconis (1977) and others, which specify

$$q_{a,N} = [{}^{N}C_a / {}^{(N+1)}C_a]\, q_{a,(N+1)} + [{}^{N}C_a / {}^{(N+1)}C_{(a+1)}]\, q_{(a+1),(N+1)},$$

for a = 0, 1,...,N. Upon substitution of $q_{a,(N+1)}$ and $q_{(a+1),(N+1)}$ with their representations identified in (2.2) these yield the components of $\mathbf{q}_{N+1}$ as

$$q_{0,N} = q_{0,(N+1)} / (1-p_{0,N}), \quad \text{and generally}$$

$$q_{a,N} = [q_{0,(N+1)} / (1-p_{a,N})]\ {}^{N}C_a \prod_{i=0}^{a-1} [p_{i,N}/(1-p_{i,N})] \quad \text{for } a = 1,2,\ldots,N.$$

Finally, transforming $\mathbf{q}_{N+1}$ to $\mathbf{p}_N$ via equations (2.3) yields the solution

$$p_{(a-1),(N-1)} = p_{(a-1),N} / [1 - p_{a,N} + p_{(a-1),N}], \quad \text{for } a = 1,\ldots,N, \qquad (2.4)$$

which appears in the second assertion of Theorem 2; equivalently,

$$p_{a,N} = 1 - p_{(a-1),N} [(1 - p_{(a-1),(N-1)})/ p_{(a-1),(N-1)}], \quad \text{for } a = 1,\ldots,N. \qquad (2.5)$$

These equivalent equations can be read in two ways. In the first place, equation (2.4) was derived to show the values of $p_{i,(N-1)}$ implied by coherency

with the assertion of the conditional previsions $\mathbf{p}_{N+1} \equiv (p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$. In these terms it is interesting to note that $p_{(a-1),(N-1)}$ depends only on the two assertion values $p_{(a-1),N}$ and $p_{a,N}$. It should be evident how equation (2.4) can be used iteratively to compute the entire vector $\mathbf{p}_N$ from $\mathbf{p}_{N+1}$, and then again to compute $\mathbf{p}_{N-1}$ from $\mathbf{p}_N$, and so on, down to the computation of $\mathbf{p}_1 = p_{0,0}$. This is the form of the structure on the upper triangular matrix of conditional previsions $\mathbf{P}_{N+1,N+1} \equiv (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N+1})$ that is identified in the second statement of Theorem 2. Alternatively, equation (2.5) can be read as specifying the allowable pairs of assertions $p_{(a-1),N}$ and $p_{a,N}$ that would provide coherent extensions of the assertion $p_{(a-1),(N-1)}$. Evidently, a coherent conditional prevision assertion vector $\mathbf{p}_{N+1}$ can be extended to a corresponding vector one dimension larger with some freedom. We can be more precise.

For the third assertion of Theorem 2 specifies necessary and sufficient computable conditions distinguishing the conditional prevision assertions $\mathbf{p}_{N+1}$ that are infinitely extendible from those that are only finitely extendible.

De Finetti's representation theorem for exchangeable distributions yields part of this distinction directly. It states that the assertions $\mathbf{p}_{N+1}$ are infinitely exchangeably extendible if and only if they are associated with a vector $\mathbf{q}_{N+2}$ that satisfies the following representation:

for each value of a = 0, 1, ..., N+1,

$$q_{a,(N+1)} = {}^{(N+1)}C_a \ {}_0\!\int^1 \theta^a (1-\theta)^{N+1-a} \, dF(\theta)$$

for some distribution function $F(\cdot)$.

Replacing the values of $q_{(a+1),(N+1)}$ and $q_{a,(N+1)}$ in equation (2.3) by their representation in this form, and applying a bit of algebraic reduction, yields the following equivalent statement of de Finetti's result, which we state as a Lemma.

Lemma 2.1. The conditional prevision assertions $\mathbf{p}_{N+1} = (p_{0,N}, p_{1,N}, \ldots, p_{N,N})^T$ are infinitely exchangeably extendible if and only if each component is representable in terms of the same distribution function $F(\cdot)$ as

$$p_{a,N} = \frac{{}_0\int^1 \theta^{a+1}(1-\theta)^{N-a}\, dF(\theta)}{{}_0\int^1 \theta^{a+1}(1-\theta)^{N-a}\, dF(\theta) + {}_0\int^1 \theta^{a}(1-\theta)^{N+1-a}\, dF(\theta)}$$

$$= \frac{{}_0\int^1 \theta^{a+1}(1-\theta)^{N-a}\, dF(\theta)}{{}_0\int^1 \theta^{a}(1-\theta)^{N-a}\, dF(\theta)} = {}_0\int^1 \theta\, dF(\theta\,|\,(S_N = a)) .$$

Generally, for any integers j and k for which $0 \le j \le a$ and $0 \le k \le N-a$,

$$p_{a,N} = \frac{{}_0\int^1 \theta^{j+1}(1-\theta)^{k}\ \theta^{a-j}\ (1-\theta)^{N-k-a}\, dF(\theta)}{{}_0\int^1 \theta^{j+1}(1-\theta)^{k+1}\ \theta^{a-j}\ (1-\theta)^{N-k-a}\, dF(\theta)} ,$$

$$= \frac{{}_0\int^1 \theta^{j+1}\,(1-\theta)^{k}\, dF(\theta\,|\,(S_{N-k} = a-j))}{{}_0\int^1 \theta^{j}\,(1-\theta)^{k}\, dF(\theta\,|\,(S_{N-k} = a-j))} ,$$

for some distribution function $F(\cdot)$, for each $a = 0, \ldots, (N+1)$. Δ

Lemma 2.1 is weaker than the claim made in Theorem 2. For the lemma provides merely an existential statement, as opposed to a constructive statement. Theorem 2 identifies *computable* characteristics of $\mathbf{p}_{N+1}$ which determine whether the assertions are infinitely extendible or not. However, an immediate usefulness of Lemma 2.1 is that, together with equation (2.4) it establishes a *necessary* computable condition for the infinite extendibility of the conditional previsions $\mathbf{p}_{N+1}$. We state this condition as Lemma 2.2.

Lemma 2.2. If the conditional previsions $\mathbf{p}_{N+1}$ are infinitely exchangeably extendible then they specify an upper triangular matrix of cohering conditional previsions

$$\mathbf{P}_{(N+1),(N+1)} = \begin{bmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots & p_{0,N} \\ & p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ & & p_{2,2} & & \\ & & & \ddots & \\ & & & & p_{N,N} \end{bmatrix} ,$$

computed sequentially using equation (2.4), that exhibits the properties

i) $p_{a,N}$ decreases weakly with N for each $a = 0, 1, \ldots, N$;

ii) $p_{a,K}$ increases weakly with a for each $K = 1, 2, \ldots, N$; and

iii) $p_{N-a,N}$ increases weakly with N for each $a = 0, 1, \ldots, N$.

Each of these three properties imply the other two.

Proof: Suppose the assertions $\mathbf{p}_{N+1}$ are infinitely exchangeably extendible.

i) Using the representations stated in Lemma 2.1, that

$$p_{a,N} = {}_0\int^1 \theta\, dF(\theta\,|\,(S_N = a)) ,$$

and

$$p_{a,(N+1)} = \frac{{}_0\int^1 \theta\,(1-\theta)\, dF(\theta\,|\,(S_N = a))}{{}_0\int^1 (1-\theta)\, dF(\theta\,|\,(S_N = a))} ,$$

it is evident that $p_{a,N} \ge p_{a,(N+1)}$ if and only if

$${}_0\int^1 \theta\, dF(\theta\,|\,(S_N = a))\ {}_0\int^1 (1-\theta)\, dF(\theta\,|\,(S_N = a)) \ \ge\ {}_0\int^1 \theta\,(1-\theta)\, dF(\theta\,|\,(S_N = a)) .$$

But this is true, for any value of N, by the fact that $[{}_0\int^1 \theta\, dF(\theta)]^2 \le {}_0\int^1 \theta^2\, dF(\theta)$ for any distribution function $F(\cdot)$.

ii) Similarly, using the representations from Lemma 2.1,

$$p_{a,N} = \frac{{}_0\int^1 \theta\,(1-\theta)\, dF(\theta \mid (S_{N-1} = a))}{{}_0\int^1 (1-\theta)\, dF(\theta \mid (S_{N-1} = a)}\,,$$

and

$$p_{(a+1),N} = \frac{{}_0\int^1 \theta^2\, dF(\theta \mid (S_{N-1} = a))}{{}_0\int^1 \theta\, dF(\theta \mid (S_{N-1} = a))}\,,$$

it is evident that $p_{a,N} \le p_{(a+1),N}$ if and only if

$${}_0\int^1 \theta\, dF(\theta \mid (S_{N-1} = a))\; {}_0\int^1 \theta\,(1-\theta)\, dF(\theta \mid (S_{N-1} = a)$$

$$\le\; {}_0\int^1 \theta^2\, dF(\theta \mid (S_{N-1} = a))\; {}_0\int^1 (1-\theta)\, dF(\theta \mid (S_{N-1} = a)\,.$$

But this is true for any value of N for the same reason, that for any distribution function $F(\cdot)$, $[{}_0\int^1 \theta\, dF(\theta)]^2 \le {}_0\int^1 \theta^2\, dF(\theta)$.

iii) Finally, using the representations from Lemma 2.1,

$$p_{(N-a),N} = {}_0\int^1 \theta\, dF(\theta \mid (S_N = N-a))\,,$$

and

$$p_{(N-a+1),(N+1)} = \frac{{}_0\int^1 \theta^2\, dF(\theta \mid (S_N = N-a))}{{}_0\int^1 \theta\, dF(\theta \mid (S_N = N-a))}\,,$$

it is evident that $p_{(N-a),N} \le p_{(N-a+1),(N+1)}$ if and only if

$$[{}_0\int^1 \theta\, dF(\theta \mid (S_N = N-a))]^2 \le {}_0\int^1 \theta^2\, dF(\theta \mid (S_N = N-a))\,,$$

which again is true for any N for any distribution function F. ∇

The equivalent conditions i, ii, and iii of Lemma 2.2 are *necessary* for the infinite extendibility of the conditional prevision assertions $\mathbf{p}_{N+1}$, but they are *not sufficient*. The stronger sufficiency conditions are derived from the known 1-1 relationship between the family of distribution functions on the interval [0, 1] and the family of completely monotone sequences, a result known as Hausdorff's Theorem. See Akhiezer (1965), Karlin and Studden (1966), and Daboni (1982). The argument for sufficiency conditions procedes as follows.

First, notice that our equation (2.2) specifies a 1-1 transformation of the matrix $\mathbf{P}_{(N+1),(N+1)}$ into the corresponding matrix

$$\mathbf{Q}_{(N+2),(N+2)} \equiv (\mathbf{q}_0, \mathbf{q}_1, \cdots, \mathbf{q}_{N+2})$$

$$= \begin{bmatrix} q_{0,0} & q_{0,1} & q_{0,2} & \cdots & q_{0,(N+1)} \\ & q_{1,1} & q_{1,2} & \cdots & q_{1,(N+1)} \\ & & q_{2(2)} & & \\ & & & \ddots & \\ & & & & q_{(N+1),(N+1)} \end{bmatrix}.$$

Now the infinite exchangeable extendibility of the conditional prevision assertions $\mathbf{p}_{N+1}$ means that every diagonal element of $\mathbf{Q}_{(N+2),(N+2)}$ must be representable as $q_{i,i} = {}_0\int^1 \theta^i\, dF(\theta)$ for the same distribution function $F(\cdot)$, according to de Finetti's theorem. Thus, these diagonal elements must be the moments of some distribution function $F(\cdot)$. Hausdorff's theorem implies then that the sequence of diagonal elements of $\mathbf{Q}$ must be infinitely extendible as a completely monotone sequence.

Computable conditions for a finite sequence $\{1, q_1, q_2, \ldots, q_{N+1}\}$ to be infinitely extendible as a completely monotone sequence are reported by Karlin and Studden (1966, pp. 106-107): for each $K = 1,\dots,N+1$,

$$L(q_1,\dots,q_{K-1}) = L \le q_K \le U = U(q_1,\dots,q_{K-1}) \quad ,$$

where, writing K as 2m+1 or as 2m, depending on whether K is odd or even,

$$L(q_1,\dots,q_{K-1}) = \sum_{i=0}^{m} (-1)^{i+1} q_{K-i} \, \mathrm{adj}_{m+1-i,m+1}(\Delta_K) \,/\, \Delta_{K-2} \ ,$$

and

$$U(q_1,\dots,q_{K-1}) = q_{K-1} - \sum_{i=0}^{m} (-1)^{i+1} (q_{K-i-1} - q_{K-i}) \, \mathrm{adj}^*_{m+1-i,m+1}(\overline{\Delta}_K) \,/\, \overline{\Delta}_{K-2} \ .$$

The symbol Δ_K denotes a "lower Hankel determinant":

for $K = 2m$; for $K = 2m+1$,

$$\Delta_K \equiv \begin{vmatrix} 1 & q_1 & q_2 & \cdots & q_m \\ q_1 & q_2 & q_3 & \cdots & q_{m+1} \\ \cdot & & & & \\ \cdot & & & & \\ q_m & q_{m+1} & & \cdots & q_K \end{vmatrix} \qquad \Delta_K \equiv \begin{vmatrix} q_1 & q_2 & q_3 & \cdots & q_{m+1} \\ q_2 & q_3 & q_4 & \cdots & q_{m+2} \\ \cdot & & & & \\ \cdot & & & & \\ q_{m+1} & q_{m+2} & & \cdots & q_K \end{vmatrix} .$$

The symbol $\overline{\Delta}_K$ denotes an "upper Hankel determinant":

for $K = 2m$; for $K = 2m+1$,

$$\overline{\Delta}_K \equiv \begin{vmatrix} q_1-q_2 & q_2-q_3 & \cdots & q_m - q_{m+1} \\ q_2-q_3 & q_3-q_4 & \cdots & q_{m+1}-q_{m+2} \\ \cdot & & & \\ \cdot & & & \\ q_m-q_{m+1} & & \cdots & q_{K-1}-q_K \end{vmatrix} \qquad \overline{\Delta}_K \equiv \begin{vmatrix} 1-q_1 & q_1-q_2 & \cdots & q_m - q_{m+1} \\ q_1-q_2 & q_2-q_3 & \cdots & q_{m+1}-q_{m+2} \\ \cdot & & & \\ \cdot & & & \\ q_m-q_{m+1} & & \cdots & q_{K-1}-q_K \end{vmatrix} .$$

The subscripts on the adjoint notation, $\mathrm{adj}^*_{m+1-i,m+1}(\overline{\Delta}_K)$, should read $(m-i,m)$ if K is an even number. (Notice that the dimension of the Hankel matrix $(\overline{\Delta}_K)$ is only m×m in that case.)

Finally, these conditions are related to the conditional probability matrix $\mathbf{P}_{(N+1),(N+1)}$ by the multiplication rule for conditional probabilities, which requires that $q_{i,i} = \prod_{j=0}^{i-1} p_{j,j}$ for $i = 1,\dots,N$. Thus, the computable condition for the infinite exchangeable extendibility of an asserted inference vector $(p_{0,N}, p_{1,N},\dots,p_{N,N})$ is that for each $K = 1,\dots,N$, its associated vector $(p_{0,0}, p_{1,1},\dots,p_{N,N})$ satisfies

$$\frac{L(p_{00}, p_{00}p_{11},\dots,\prod_{i=0}^{K-1} p_{ii})}{\prod_{i=0}^{K-1} p_{ii}} \le p_{K,K} \le \frac{U(p_{00}, p_{00}p_{11},\dots,\prod_{i=0}^{K-1} p_{ii})}{\prod_{i=0}^{K-1} p_{ii}} .$$

Although they have been ungainly to state, these conditions are simple to compute. The following example will be useful for reference subsequently.

<u>EXAMPLE</u> 2.1. Suppose that the conditional prevision vector $\mathbf{p}_6 = (p_{0,5}, p_{1,5},\dots,p_{5,5})$ is asserted as (.1429, .2857, .4286, .5714, .7143, .8571). These happen to be the conditional probabilities associated with a Binomial-Uniform mixture distribution for S_6 (equivalently, a Polya (6,1,1) distribution) rounded to the nearest fourth decimal place. Applying the infinite extendibility criterion just developed, the boundary conditions for the further assertion of p_{66}, that is, $P(E_7 | (S_6=6))$ amount to $L = .87428571 \le p_{6,6} \le .87571429 = U$, printed to the nearest eighth decimal place. The associated full $\mathbf{p}_7$ vectors corresponding to these extreme extended assertions are computed using our iterative equations (2.4 or 2.5), as $\mathbf{p}_7$(lower) = (.1243, .2543, .3643, .5143, .6143, .7543, .8743) , and $\mathbf{p}_7$(upper) = (.1257, .2457, .3857, .4857, .6357, .7457, .8757). Sixteen-place decimals were used in the computations. •••

Before moving on to the concluding statement of Theorem 2, we should note a very useful corollary to this argument: if you regard the events $E_1, \ldots, E_{N+1}$ exchangeably, asserting a mixture-binomial distribution for S_{N+1} (the infinitely extendible case) then every inferential probability assertion $P(E_{N+1} \mid (S_N = a))$ can be computed directly from the moments of the mixing distribution, which constitute the main diagonal elements of the matrix we have denoted by $\mathbf{Q}_{(N+2),(N+2)}$. The following example of the Binomial-Cantor mixture distribution provides a striking example of the power of this result.

EXAMPLE 2.2. Suppose that for a sequence of 25 events, opinions about S_N are representable in de Finetti's form $q_{a,N} = {}^N C_a \int_0^1 \theta^a (1-\theta)^{N-a} \, dF(\theta)$, where specifically $F(\theta)$ is the standard Cantor distribution. Using the direct moments of this distribution, derived and calculated by Lad and Taylor (1992) the corresponding matrix of inferential probabilities $\mathbf{P}_{24,24}$ were computed according to the procedure outlined in the text: calculate the diagonal elements of $\mathbf{P}_{24,24}$ from the moments of $F(\cdot)$ using the equation $p_{ii} = \mu_{i+1}/\mu_i$; then generate the entire matrix $\mathbf{P}$ from the recursive equations that were identified as equation (2.4). The first 25 moments of the Cantor distribution to 4 decimal places are printed in Table 2.1 in the column headed $\mu_a = q_{a,a}$. The associated vector of conditional probabilities $(p_{0,24}, p_{1,24}, \ldots, p_{24,24})^T$ appear in the column headed $P(E_{25} \mid S_{24} = a)$. Again, the moments were entered to 16 decimal places to generate these results. The columns headed $\alpha_{a,24}$ and $\beta_{a,24}$ will be interpreted shortly.

Table 2.1. Coherent inferences $P(E_{25} \mid (S_{24} = a))$ when the asserted probability distribution for S_{25} is the Binomial-Cantor mixture. The four columns list the vectors denoted in the text as the diagonal vector of $\mathbf{Q}_{26}$, $\mathbf{p}_{25}$, $\boldsymbol{\alpha}_{25}$, and $\boldsymbol{\beta}_{25}$.

a	$\mu_a = q_{a,a}$	$P(E_{25} \mid S_{24} = a)$	$\alpha_{a,24}$	$\beta_{a,24}$
0	1.0000	0.0247	0.6527	1.7852
1	0.5000	0.0625	0.7178	2.7613
2	0.3750	0.0989	0.1647	-2.2780
3	0.3125	0.1446	-0.3578	-5.3693
4	0.2719	0.1993	0.7342	-0.9827
5	0.2422	0.2414	5.3248	13.4414
6	0.2192	0.2648	13.9880	37.4935
7	0.2009	0.2780	22.2260	58.8827
8	0.1860	0.2878	22.0425	58.4281
9	0.1737	0.2970	7.9684	25.1371
10	0.1632	0.3149	-4.1733	-1.3088
11	0.1543	0.3688	-8.1859	-8.1806
12	0.1466	0.5000	-8.2225	-8.2173
13	0.1398	0.6312	-1.0743	-4.0538
14	0.1338	0.6851	23.5220	7.2613
15	0.1285	0.7030	62.7775	23.8321
16	0.1236	0.7122	56.1104	21.1381
17	0.1192	0.7220	38.4645	14.3467
18	0.1152	0.7352	13.2652	5.2663
19	0.1116	0.7586	-0.9560	0.7413
20	0.1082	0.8007	-5.3752	-0.3589
21	0.1050	0.8554	-2.2760	0.1649
22	0.1021	0.9011	2.7608	0.7178
23	0.0994	0.9375	1.7852	0.6527
24	0.0968	0.9753	1.7852	0.6527
25	0.0944			

●●●

Theorem 2 finally concludes by distinguishing assertions that are infinitely extendible as a Polya distribution from those that are infinitely extendible but defy representation in this form. This is achieved by determining a functional procedure that will parameterize the assertions $\mathbf{p}_{N+1}$ in the form

$$p_{a,N} = (a + \alpha_{a,N})/(N + \alpha_{a,N} + \beta_{a,N}), \qquad \text{for } a = 0, 1, \ldots, N. \qquad (2.6)$$

This determination involves some work. For the representation of a coherent $p_{a,N}$ by a pair $(\alpha_{a,N}, \beta_{a,N})$ satisfying (2.6) is not unique. In fact, the representation in form (2.6) for the conditional previsions we have denoted by $(p_{0,N}, \ldots, p_{N,N})$ would be satisfied by any vector of $(\alpha_{a,N}, \beta_{a,N})$ pairs for which

$$\beta_{a,N} = -(N-a) + (a+\alpha_{a,N})[(1-p_{a,N})/p_{a,N}] . \qquad (2.7)$$

Figure 2.1 displays three lines of $(\alpha_{a,N}, \beta_{a,N})$ pairs that would represent assertions of $p_{a,N} = 1/3$, $1/2$, and $2/3$. Their slopes are 2, 1, and 1/2, respectively. The coherency restriction of $\mathbf{p}_{N+1}$ to the unit cube requires that the slopes of these lines be nonnegative. Thus, for each value of a and N, either $[\,\alpha_{a,N} \geq -a$ *and* $\beta_{a,N} \geq -(N-a)\,]$ or $[\,\alpha_{a,N} \leq -a$ *and* $\beta_{a,N} \leq -(N-a)\,]$.

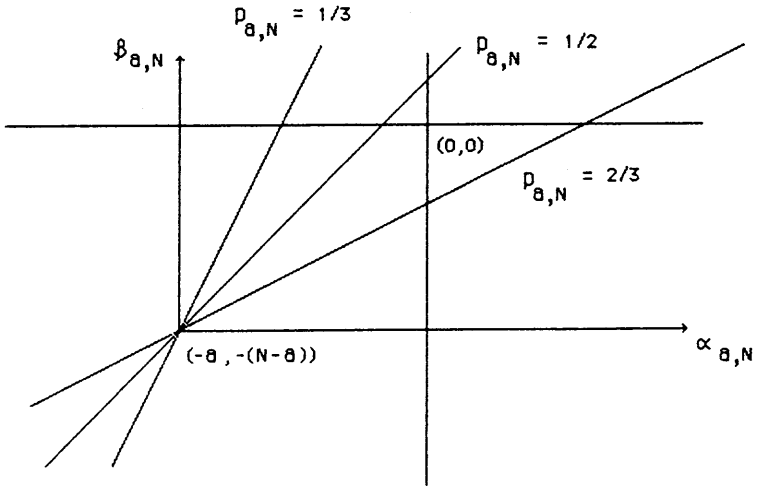

Figure 2.1. Lines of $(\alpha_{a,N}, \beta_{a,N})$ pairs that would represent assertions of $p_{a,N} = 1/3$, $1/2$, and $2/3$: $\beta_{a,N} = -(N-a) + (a+\alpha_{a,N})[(1-p_{a,N})/p_{a,N}]$.

Before progressing to the development of a parameterization of $\mathbf{p}_{N+1}$, notice the special case that previsions $\mathbf{p}_{N+1}$ supporting an infinite extension as a Polya distribution are those that can be represented as

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})_{N+1} = (\alpha_{0,N}, \beta_{0,N}, \ldots, \alpha_{N,N}, \beta_{N,N}) = (\alpha, \beta)\, \mathbf{1}_{N+1}$$

for some positive values of α and β. For these would yield the well known conditional probabilities of the Binomial-Beta, or Polya distribution. Figure 2.2 (printed on page 42) shows how such assertions can be identified geometrically by displaying the representation of one example of a conditional prevision vector that *can* be represented by the Polya distribution. Every (α_a, β_a) line associated with the Polya $(6, \alpha, \beta)$ distribution must intersect in the common point (α, β).

Thus, our second theorem identifies precisely how the Polya distribution is *properly* included within the class of all exchangeable infinitely extendible distributions. For an algebraic characterization criterion of this subclass of coherent infinitely extendible assertions, see the article of Hill, Lane, and Sudderth (1988) on exchangeable urn processes.

Of course, there is no *necessity* that coherent conditional previsions specify lines of $(\alpha_{a,N}, \beta_{a,N})$ pairs that all intersect in one common point. In contrast, Figure 2.3 (printed on page 43) suggests several different reasons why an allowable assertion vector might not be representable as a Polya distribution. As is displayed in the Figure, lines of $(\alpha_{a,N}, \beta_{a,N})$ pairs for neighboring values of "a" may intersect in any of the four quadrants of (α, β) space, or they may not intersect at all! This latter situation occurs whenever $p_{a,N} = p_{(a+1),N}$.

Finally, we can generate a unique representation of $\mathbf{p}_{N+1}$ in terms of a vector $(\alpha_{0,N}, \beta_{0,N}, \ldots, \alpha_{N,N}, \beta_{N,N})$ by using equation (2.4), repeated here, which related any three adjoining components of the conditional prevision matrix:

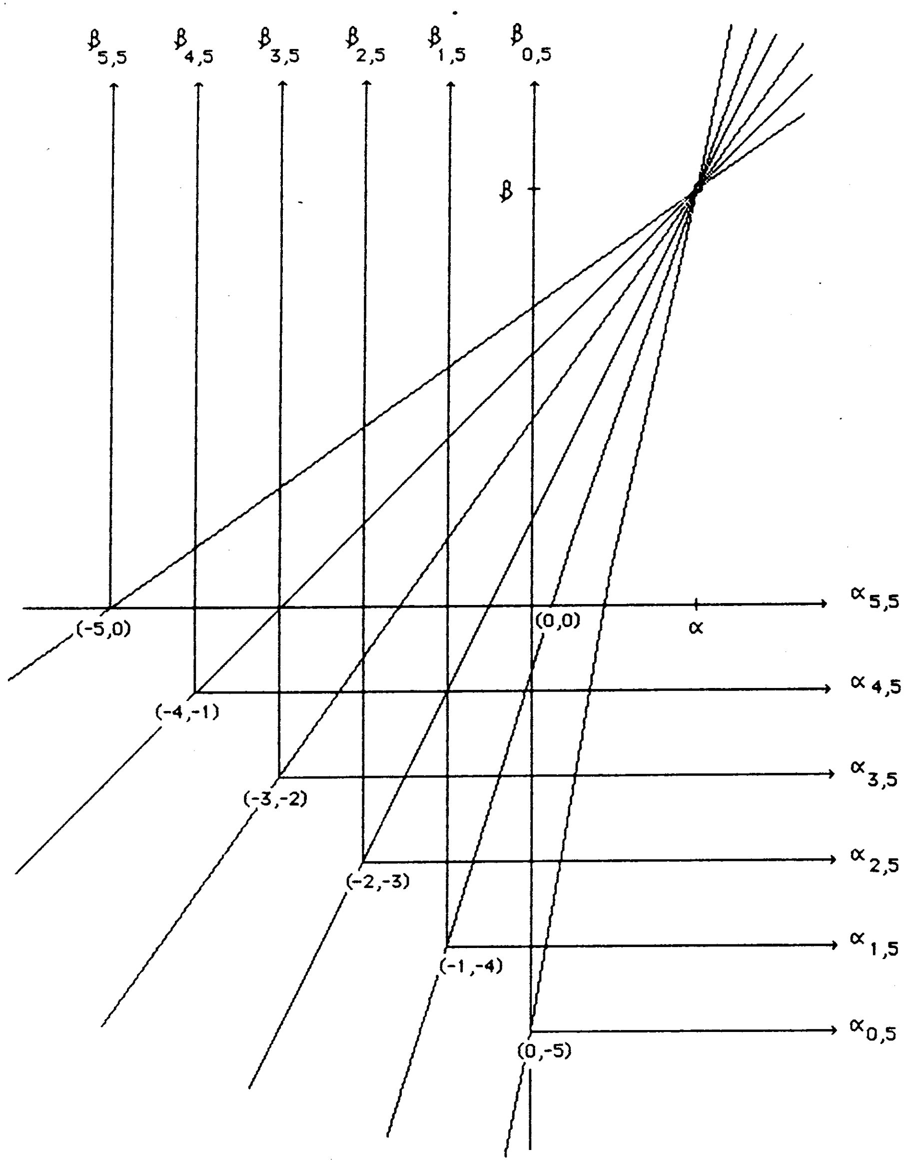

Figure 2.2. Conditional prevision assertions $\mathbf{p}_6$ that are exchangeably infinitely extendible as the Polya distribution, Polya(N,α,β). The parameters α and β are identified by the common intersection point of all 6 lines defined by the equations $\beta_{a,5} = -(5-a) + (a+\alpha_{a,5})[(1-p_{a,5})/p_{a,5}]$ for $a = 0,1,2,3,4,5$.

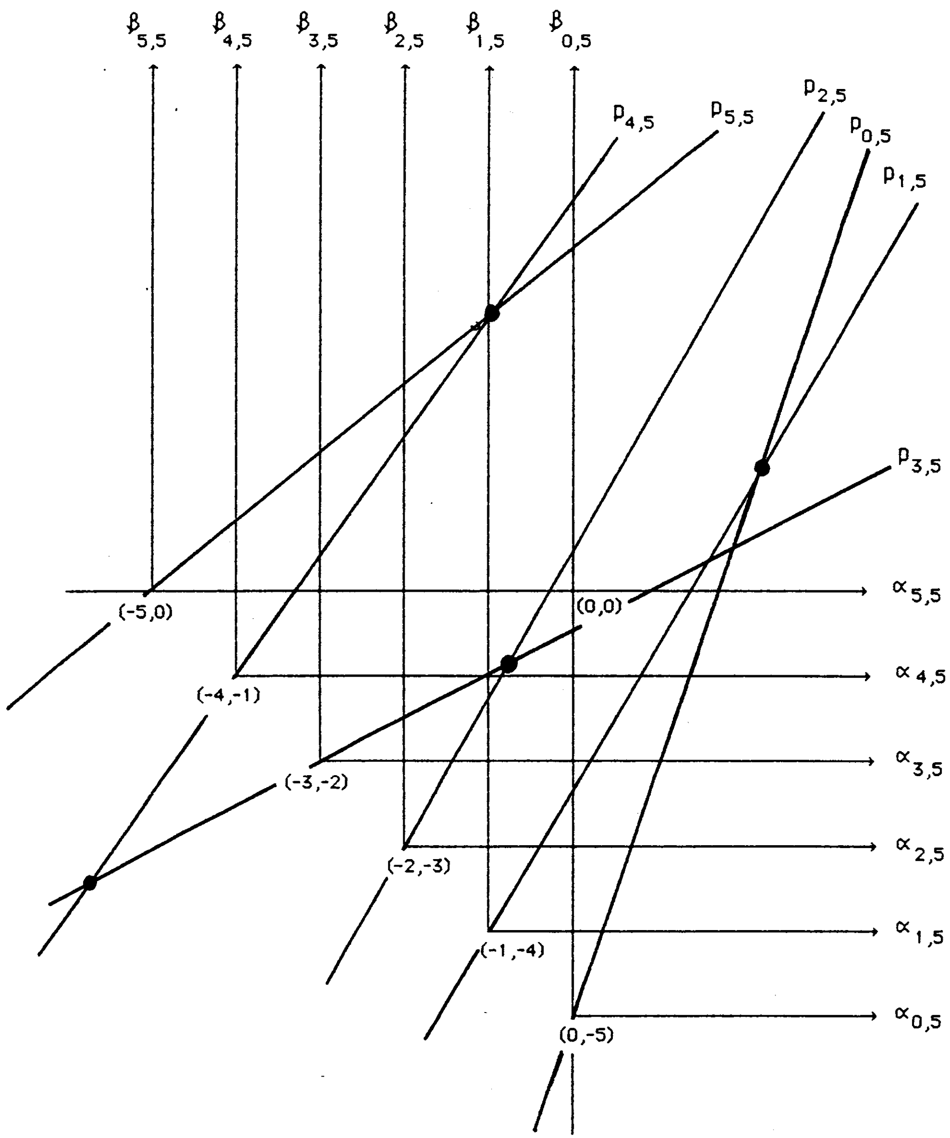

Figure 2.3. Conditional prevision assertions that are not representable as a Polya distribution. Circled points identify relevant intersections. Notice that $p_{1,5} = p_{2,5}$, since their associated lines do not intersect.

$$p_{(a-1),(N-1)} = p_{(a-1),N} / [1 - p_{a,N} + p_{(a-1),N}], \quad \text{for} \quad a = 1,\dots,N. \qquad (2.4)$$

Geometrically, equation (2.4) implies that the line of $(\alpha_{a,N}, \beta_{a,N})$ pairs supporting $p_{a,N}$ must intersect the lines of $(\alpha_{a,(N+1)}, \beta_{a,(N+1)})$ and $(\alpha_{(a+1),(N+1)}, \beta_{(a+1),(N+1)})$ pairs supporting $p_{a,(N+1)}$ and $p_{(a+1),(N+1)}$, respectively, in a common point shared by the three of them. This interpretation is achieved by replacing the expressions $p_{a,N}$, $p_{a,(N+1)}$, and $p_{(a+1),(N+1)}$ in (2.4) by their representations in terms of the pairs $(\alpha_{a,N}, \beta_{a,N})$, $(\alpha_{a,(N+1)}, \beta_{a,(N+1)})$, and $(\alpha_{(a+1),(N+1)}, \beta_{(a+1),(N+1)})$, respectively, applying equation (2.6). Equation (2.4) is found to be satisfied only when these three lines of pairs intersect in a common point. Algebraically, the solution equations for the common point (α, β) of the three lines, expressed as a function of $p_{a,N}$ and $p_{a,(N+1)}$, reads

$$\alpha = -a + p_{a,N}\, p_{a,(N+1)} / [\, p_{a,N} - p_{a,(N+1)} \,] \equiv \alpha^*_{a,(N+1)} ,$$

$$(2.8)$$

and $\beta = -(N+1-a) + p_{a,N} (1 - p_{a,(N+1)}) / [\, p_{a,N} - p_{a,(N+1)} \,] \equiv \beta^*_{a,(N+1)}$.

Thus, we can use these equations to define the parameters $\alpha^*_{a,(N+1)}$ and $\beta^*_{a,(N+1)}$ that would uniquely characterize $p_{a,(N+1)}$ as

$$p_{a,(N+1)} = (a + \alpha^*_{a,(N+1)}) / [N+1+ \alpha^*_{a,(N+1)} + \beta^*_{a,(N+1)}] ,$$

as specified in Theorem 2. Figure 2.4 displays the structure of this relation geometrically.

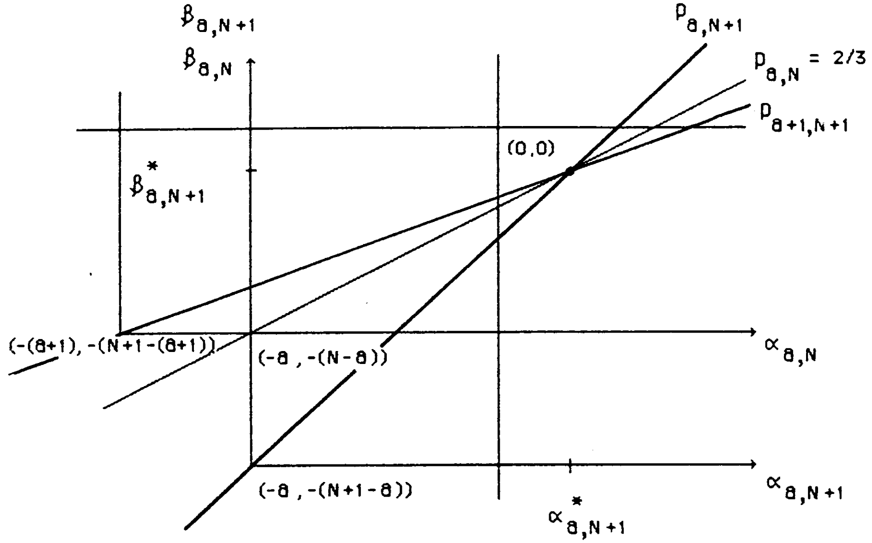

Figure 2.4. A pair of lines $(\alpha_{a,(N+1)}, \beta_{a,(N+1)})$ and $(\alpha_{(a+1),(N+1)}, \beta_{(a+1),(N+1)})$ that support the assertion $p_{a(N)} = 2/3$ are shown, along with the line $(\alpha_{a,N}, \beta_{a,N})$. The pair $(\alpha^*_{a,(N+1)}, \beta^*_{a,(N+1)})$ that we define to represent $p_{a,(N+1)}$ is the unique point where these three lines intersect. This point may appear anywhere in either the I or III quadrant relative to the point (-a, -(N-a)).

Once you understand Figure 2.4, you will also understand that Figure 2.2 can be "filled in" to show that all conditional probabilities $p_{a,N}$ for values of $N < 5$ are also representable by the same Polya distribution. For every line of $(\alpha_{a,N}, \beta_{a,N})$ pairs for $N < 5$ must also intersect in the same common point (α, β), as required by (2.4). Similarly, you can sequentially "fill in" the lines of Figure 2.3 associated with the conditional probabilities $p_{a,N}$ for values of $N < 5$. Beginning with $a=0$ and $N=4$, merely draw the line through $(0,-4)$ and the intersection point of the lines associated with $p_{0,5}$ and $p_{1,5}$. Then proceed to $a=1$ and $N=4$, drawing the line through $(-1,-3)$ and the intersection point of the lines associated with $p_{1,5}$ and $p_{2,5}$. (In fact, in this case as drawn, the lines do not intersect. So merely draw the line between the two parallels. These three lines then will all "intersect" at ∞.) Continue this process with $a = 2$ and $N = 4$, and so on until $a=4$. Then proceed to $N=3$.

On the other hand, in Example 2.1 we have already made computations showing that Figure 2.4 can be "filled out" by a whole family of inferences about E_7 that would agree with the Polya (6,1,1) inferences and all their associated lower order inferences.

Our proof of Theorem 2 is completed.

3. Proof of Theorem 3: on the coherency requirements for conditional probability distributions that mimic positive histograms.

The coherency of strictly positive histogram mimicking conditional probability distributions can be made evident by embedding them in a coherent exchangeable joint distribution for $\mathbf{X}_{N+1}$ that agrees with them. This is achieved as follows. Consider any strictly positive $\mathbf{h} \in \mathfrak{R}(\mathbf{H}(\mathbf{X}_{N+1}))$. There are ${}^{(N+1)}C_{\mathbf{h}}$ distinct vectors $\mathbf{x}_{N+1} \in \mathfrak{R}(\mathbf{X}_{N+1})$ that yield this histogram $\mathbf{h} = \mathbf{h}(\mathbf{x}_{N+1})$. Given the presumed assertion of exchangeability, the histogram conditioning rule implies that each of these vectors must be accorded the same probability, which is representable in several ways:

$$P(\mathbf{X}_{N+1} = \mathbf{x}_{N+1}) \;=\; [(h_*-1)/N]\; P[\mathbf{H}(\mathbf{X}_N) = \mathbf{h}(\mathbf{x}_{N+1})-\mathbf{e}_*] \,/\, {}^{N}C_{\mathbf{h}(\mathbf{x})-\mathbf{e}^*} \qquad (3.1)$$

where h_* is *any* component of $\mathbf{h}$ not equal to 1, and $\mathbf{e}_*$ is the corresponding echelon basis vector for (K+1)-dimensional space in that direction. It equals 1 in its * component, and 0 elsewhere. The factor $(h_*-1)/N$ assures the agreement of the probability distribution for $\mathbf{X}_{N+1}$ with the required histogram mimicking character of its conditional distributions based on strictly positive histograms. Now select a probability distribution for $\mathbf{H}(\mathbf{X}_N)$ that accords arbitrarily small probabilities to strictly positive histograms. With such a choice, the sum of equations (3.1) over $\mathbf{x}_{N+1}$ for which $\mathbf{h}(\mathbf{x}_{N+1}) > \mathbf{0}$ can be made less than 1. Ascribing the remainder to the probability of all the nonpositive histograms, we have identified a coherent exchangeable probability distribution for $\mathbf{X}_{N+1}$ that supports the conditional distributions specified in Theorem 3.

Mechanics of the proofs of the qualifications to Theorem 3 rest on use of the following three results:

R1) $P[(X_{N+1}=a)(\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N))] = [(h_a(N)+1)/(N+1)]\; P[\mathbf{H}(\mathbf{X}_{N+1})=\mathbf{h}(N)+\mathbf{e}_a]$, where $\mathbf{h}(N)$ represents any strictly positive histogram based on N quantities, $h_a(N)$ is its component for category $a \in \{0,\dots,K\}$, and $\mathbf{e}_a$ is the echelon vector with 1 in component a and 0 elsewhere;

R2) for every $a \in \{0,\dots,K\}$ and for every $\mathbf{h}(N) > \mathbf{0}$,

$$[h_a(N)/N]\, P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N)] \;=\; [(h_a(N)+1)/(N+1)]\, P[\mathbf{H}(\mathbf{X}_{N+1})=\mathbf{h}(N)+\mathbf{e}_a]\,; \quad \text{and}$$

R3) for any integers a and b, each within $\{0,\dots,K\}$, and for every $\mathbf{h}(N-1) > \mathbf{0}$,

$$P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_b] \;=\; C(a,b)\, P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_a] \;,$$

where the proportionality constant is

$$C(a,b) \;=\; [h_a(N-1)+1]\,[h_b(N-1)] \;/\; [h_a(N-1)]\,[h_b(N-1)+1] \;.$$

The first result, R1, follows directly from the exchangeability assertion regarding the components of $\mathbf{X}_{N+1}$. Thus, a corresponding statement is true if the value of N is replaced by any smaller integer as well. The second result, R2, follows from applying the histogram rule for X_{N+1} to the left-hand side of R1. The third result, R3, follows from applying R2 to the relations between $P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_b]$ and $P[\mathbf{H}(\mathbf{X}_{N+1})=\mathbf{h}(N-1)+\mathbf{e}_b+\mathbf{e}_a]$ on the left-hand side, and between $P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_a]$ and $P[\mathbf{H}(\mathbf{X}_{N+1})=\mathbf{h}(N-1)+\mathbf{e}_a+\mathbf{e}_b]$ on the right-hand side.

Qualification (a) of Theorem 3 constitutes a generalization of the Corollary 1.2 we proved in Section 1. Backward induction allows that we need prove only that the statement is true for M = N-1. That is, we need show only that for any strictly positive histogram based on N-1 quantities, $\mathbf{h}(N-1) > \mathbf{0}$,

$$P[(X_N=a)\,|\,\mathbf{H}(\mathbf{X}_{N-1})=\mathbf{h}(N-1)] \;=\; h_a(N-1)/(N-1) \quad \text{for each } a \in \{0,\dots,K\}.$$

Now for any positive histogram $\mathbf{h}(N-1)$ and any $a \in \{0,1,\dots,K\}$,

$$P[(X_N=a)\,|\,\mathbf{H}(\mathbf{X}_{N-1})=\mathbf{h}(N-1)] \;=\; \frac{[(h_a(N-1)+1)/N]\; P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_a]}{\sum_{b=0}^{K} [(h_b(N-1)+1)/N]\; P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_b]} \qquad (3.2)$$

Equation (3.2) derives from firstly applying R1 to $P[(X_N=a)(\mathbf{H}(\mathbf{X}_{N-1})=\mathbf{h}(N-1))]$ for the numerator. For the denominator, expand $P[\mathbf{H}(\mathbf{X}_{N-1})=\mathbf{h}(N-1)]$ to the summation $\sum_{b=0}^{K} P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_b]$ according to the theorem of total probability, and apply R1 to each of the summands.

Finally, applying R3 to each term in the denominator of (3.2) reduces the right-hand side to $[(h_a(N-1)+1)/N] \,/\, \sum_{b=0}^{K} C(a,b)$ since the expression $P[\mathbf{H}(\mathbf{X}_N)=\mathbf{h}(N-1)+\mathbf{e}_a]$ then cancels in the numerator and denominator. Simple cancellations and the recognition that $\sum_{b=0}^{K} h_b(N-1) = N-1$ yield the desired equality of qualification (a), that for every strictly positive histogram $\mathbf{h}(N-1)$, $P[(X_N=a)\,|\,\mathbf{H}(\mathbf{X}_{N-1})=\mathbf{h}(N-1)] \;=\; h_a(N-1)/(N-1)$ for each $a \in \{0,\dots,K\}$.

Qualifications (b) and (c) of Theorem 3 follow from the iterative application of qualification (a) to the probability of any histogram based upon N+1 quantities. Choose any $\mathbf{x}_{N+1}$ whose first K+1 components are the integers 0,1,...,K. The sufficiency of the histogram for exchangeable inferences requires that

$$P[\mathbf{H}(\mathbf{X}_{N+1})=\mathbf{h}(N+1)] \;=\; {}^{N+1}C_{\mathbf{h}(N+1)}\; P(\mathbf{X}_{N+1}=\mathbf{x}_{N+1}) \qquad (3.3)$$

as long as $\mathbf{h}(\mathbf{x}_{N+1}) = \mathbf{h}(N+1)$. Iteratively applying the histogram rule to $P(\mathbf{X}_{N+1}=\mathbf{x}_{N+1})$ on the basis of qualification (a) yields

$$P(\mathbf{X}_{N+1}=\mathbf{x}_{N+1}) = \prod_{a=0}^{K} [(h_a(N+1)-1)!\; K! \,/\, N!]\; P(\mathbf{X}_{K+1} = \mathbf{1}_{K+1}).$$

Substituting this into (3.3) and cancelling the appropriate factorial expressions yields for any strictly positive $\mathbf{h}(N+1)$,

$$P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)] \;=\; [(N+1)\, K! \,/\, \prod_{a=0}^{K} (h_a(N+1))]\;\; P(\mathbf{X}_{K+1} = \mathbf{1}_{K+1}) \;. \qquad (3.4)$$

In particular, $P[\mathbf{H}(\mathbf{X}_{N+1})=(N+1-K,1,1,\dots,1)] = [(N+1)\,K! \,/\, (N+1-K)]\; P(\mathbf{X}_{K+1} = \mathbf{1}_{K+1})$.

Inversely, $P(\mathbf{X}_{K+1} = \mathbf{1}_{K+1}) = [(N+1-K) / (N+1)\,K!]\, P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K,1,1,\ldots,1)]$. Substituting this expression into (3.4) yields the desired result, that for any strictly positive $\mathbf{h}(N+1)$,

$$P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)] \;=\; [(N+1-K) / \prod_{a=0}^{K} (h_a(N+1)]\; P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K,1,1,\ldots,1)] \;,$$

which was to be proved.

This algebraic result establishes the stated subexchangeability of the category components of histograms $\mathbf{H}(\mathbf{X}_{N+1})$ over strictly positive histograms, qualification (b). For the product of the components of a histogram is constant over permutations of the components. It is also evident from this equation that $P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K,1,1,\ldots,1)] \geq P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)]$ for any $\mathbf{h}(N+1) > \mathbf{0}$, which is qualification (c). Since this extreme probability will be used in what follows, we denote it hence by $p_{max} \equiv P[\mathbf{H}(\mathbf{X}_{N+1}) = (N+1-K,1,1,\ldots,1)]$.

Finally, we consider qualification (d) of Theorem 3. To begin, we will show that coherency requires a computable bound on p_{max} which vanishes as N increases. To see this, notice firstly that

$$P[\mathbf{H}(\mathbf{X}_{N+1}) > \mathbf{0}] \;=\; (N+1-K) \textstyle\sum^{*} [\prod_{a=0}^{K} h_a(N+1)]^{-1}\, p_{max} \;\leq\; 1 \quad ,$$

where the summation $\sum^{*}$ runs over all strictly positive $\mathbf{h}(N+1) > \mathbf{0}$. Thus, an upper bound for p_{max} is

$$p_{max} \;\leq\; \{\, (N+1-K) \textstyle\sum^{*} [\prod_{a=0}^{K} h_a(N+1)]^{-1} \,\}^{-1} \;\equiv\; \overline{\pi}(N,K)\,. \qquad (3.5)$$

Since the summation $\sum^{*}$ runs over all positive histograms $\mathbf{h}(N+1)$, it can be replaced by the partial sum over positive histograms that contain only two categories exceeding 1. This smaller sum equals ${}^{K+1}C_2\, [\sum_{a=2}^{N-K} a(N+2-K-a)\,]^{-1}$. (For each choice of two categories, the possible histogram component products are the summands of $\sum_{a=2}^{N-K} a(N+2-K-a)$, since all the other K-1 categories of these histograms have components equal to 1.) Using the identity $[a(N+2-K-a)]^{-1} = [a^{-1} + (N+2-K-a)^{-1}]/(N+2-K)$ and the psi-function notation introduced in Theorem 1, the partial summation $\sum_{a=2}^{N-K} [a(N+2-K-a)]^{-1} \;=\; 2\,[\varphi(N-K+1) - \varphi(2)]/(N+2-K)$. So replacing $\sum^{*} [\prod_{a=0}^{K} h_a(N+1)]^{-1}$ in equation (3.5) by the partial sum $\sum_{a=2}^{N-K} [a(N+2-K-a)]^{-1}$ yields $\overline{\pi}(N,K) < \{\, [(N+1-K)(K+1)K/(N+2-K)]\,[\varphi(N-K+1) - \varphi(2)]\,\}^{-1}$, which goes to 0 as $N \rightarrow \infty$. Thus, any coherent assertion of p_{max} must be small relative to N.

Now suppose we fix $M \leq N$, and consider $P[\mathbf{H}(\mathbf{X}_M) > \mathbf{0}]$. Firstly, notice

$$P[\mathbf{H}(\mathbf{X}_M) = \mathbf{h}(M)] \;=\; \textstyle\sum^{**} P[\mathbf{H}(\mathbf{X}_M) = \mathbf{h}(M) \,|\, \mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)]\; P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)]\,,$$

where the summation $\sum^{**}$ runs over all $\mathbf{h}(N+1) \geq \mathbf{h}(M) > \mathbf{0}$. On the basis of Bose-Einstein statistics, the number of these summands is thus ${}^{N+1-M+K}C_K$. Now apply the exchangeability induced multiple category hypergeometric distribution to the conditional probabilities in the summands of $\sum^{**}$, and apply qualification (c) to $P[\mathbf{H}(\mathbf{X}_{N+1}) = \mathbf{h}(N+1)]$. This yields

$$P[\mathbf{H}(\mathbf{X}_M) = \mathbf{h}(M)] \;=\; \textstyle\sum^{**} [{}^{\mathbf{h}(N+1)}C_{\mathbf{h}(M)} / {}^{N+1}C_M]\; [(N+1-K) / \prod_{a=0}^{K} (h_a(N+1)]\; p_{max}\;\cdot$$

which reduces algebraically to

$${}^{M}C_{\mathbf{h}(M)}\; [(N+1-K)!/(N+1)!]\; \textstyle\sum^{**} [{}^{N+1-M}C_{\mathbf{h}(N+1)-\mathbf{h}(M)} \,/\, {}^{N-K}C_{\mathbf{h}(N+1)-\mathbf{1}_{K+1}}]\; p_{max}\;\cdot$$

Standard combinatoric arguments provide that each summand of $\sum^{**} [{}^{N+1-M}C_{\mathbf{h}(N+1)-\mathbf{h}(M)} / {}^{N-K}C_{\mathbf{h}(N+1)-\mathbf{1}_{K+1}}]$ is less than 1. For the numerator expression of a summand is the number of ways that (N+1-M) items can be distributed into (K+1) categories in such a way that the components in the categories are restricted to equal $\mathbf{h}(N+1) - \mathbf{h}(M)$. Presuming $(K+1) \leq M \leq N$, the denominator equals the number of ways that a larger number of items can be distributed into (K+1) categories with the components restricted to be at least as large in every category. This denominator exceeds the numerator. So for any positive histogram $\mathbf{h}(M)$, we can identify an upper bound for $P[\mathbf{H}(\mathbf{X}_M)=\mathbf{h}(M)]$ as

$$P[\mathbf{H}(\mathbf{X}_M)=\mathbf{h}(M)] < \epsilon(M,N,K) \equiv {}^{M}C_{\mathbf{h}(M)} \, [(N+1-K)!/(N+1)!] \, {}^{N+1-M+K}C_K \; p_{max} .$$

The first multiplicand of this product, ${}^{M}C_{\mathbf{h}(M)}$, does not depend at all upon N. The second multiplicand, $[(N+1-K)!/(N+1)!]$, is approximated by N^{-K}. The third multiplicand, ${}^{N+1-M+K}C_K$, is approximated by $N^K/K!$ The final multiplicand, p_{max}, is known to converge to 0 as N increases. Thus, for any strictly positive $\mathbf{h}(M)$, the probability $P[\mathbf{H}(\mathbf{X}_M)=\mathbf{h}(M)]$ is required to be arbitrarily small when N is sufficiently large. Since there are only a finite number of such positive histograms, ${}^{M-1}C_K$, we can conclude the result stated as qualification (d).

We have completed the proof of Theorem 3. We conclude this article with some more specific remarks on the applicability of our results, and their relationship to Hill's theory of A(n) inference.

4. On the applicability of these three theorems, and their relevance to Hill's analysis of A(n) and H(n) inference. The theorems in the present article are related to the inferential theory developed by Hill (1968, 1987, 1988) which has come to be known as "A(n)" theory. Our comments in this concluding Section will presume the reader's familiarity with Hill's work.

The basic similarity of our work to Hill's lies in the shared approach of initially specifying a collection of inferences (conditional probability assertions) that would seem to be appropriate for a recognizable sampling situation. This is followed by analyzing the coherency of these assertions with an aim to uncovering any interesting concomitant assertions that coherency would require. As noted by Hill, this approach allows the explicit representation of very complex forms of reasoning which are not required to admit a straightforward simple characterization in terms of Bayes' theorem.

The basic difference of our mathematical setup from Hill's arises from a different attitude we presume toward the meaning of numerical observations and the possibility of so-called ties among observations. Differences in the scientific subject matter of applications can dictate different reasonable representations of the measurement process. To illustrate our ideas in a tangible way, we suggest the following practical situation.

Consider the milk yields during a standard 305-day milking period from the population of all grade (non-registered) Holstein heifers (mothers with first calf) living in the state of Wisconsin, none of which have been injected with the bovine somatatropine hormone. This population numbers on the order of 250,000 cows. Informed opinions about this group of cows might well regard their measured yields exchangeably. Healthy milk yields from such cows are on the order of 18,000 pounds. The annual measurement from a complete census of one cow's yield would be the sum of two weighings per day over the milking period. To a commercial dairy that may desire to buy a heifer or a group of heifers from a breeder, orders of magnitude that are needed to discriminate between heifers might be intervals of 100 pounds ranging from 15,000 pounds to 25,000 pounds. Thus, we might define a quantity X_i by coding the i^{th} cow's yield in discrete units that cover 100 pound interval possibilities. Say $X_i = 0$ represents a total yield less than 15000 pounds; $X_i = 1$ if the yield ranges from 15000 through 15099 pounds; ... ; and $X_i = 101$ if the total yield is 25,000 pounds or more. Although there do exist scales that can weigh liquids to much finer calibrations, it would be wasted expenditure to use them in defining quantities for milk yields from Holstein heifers, since finer distinctions that might be made would never be worth the overall cost of generating them.

However, if someone were interested in 305-day milk yields from tiny mammals (such as golden hamsters) beginning with their first litter, one would require a much more finely calibrated weight measurement procedure to perform a sensible discrimination exercise. Unit measurements calibrated to the nearest 100 pounds would be completely useless.

Whereas the mathematical setup of our theorems covers situations in which we desire measurements to distinguish particular dairy cows from one another, Hill's setup is meant to distinguish differences very finely not only among cows, but among cows *and* hamsters as well. His work pertains to situations that allow measured quantities to vary over *immense* ranges, with calibrated measurement distinctions defined finely enough to distinguish minute differences between units all across this wide range. To some extent, Hill's expansion of the A(n) theory to the H(n) theory which allows for ties through the theoretical filter of *agglutinated masses* and *splitting processes*, is generated with a view to recognizing small variations in measurements within species weights and very large differences between species weights. In fact, many of the conditional probabilities that are most interesting for Hill to compute involve the probability that the next weight measurement comes from a new species. Concern with this question has been central since his earliest work on

A(n) wherein the analysis was based like ours on measurements from a finite population that can exhibit exact ties (Hill, 1968).

Using Hill's notation for the general situation, one way to understand our setup is to identify his value of M (the number of distinct species or "types") as equal to 1. Our setup explicitly recognizes the realistic limitation in the number of possible measurements that are regarded as distinct when the subject population is restricted to a specialized, even though numerous, group. We are weighing milk yields from the population of grade Holstein heifers in Wisconsin, not milk yields from the population of all mammals on earth. Once this difference is realized, it makes sense that the conditional probabilities we specify to begin our problem involve conditioning observations from N subjects where N is larger than K, the number of distinct measurements that can be distinguished. In contrast, Hill's work would characterize K as the infinity of the continuum, with N some finite number. Seen in this context of a whole family of problems, our Theorem 3 constitutes an extremely interesting result. It allows and supports Hill's results on the coherency of A(n). But the fact that our upper bound on the probability of achieving a strictly positive histogram from a sizeable number of crude measurements is arbitrarily close to 0, places limits on one's willingness to assert this inferential strategy in a practical context that recognizes only bounded discrete measurements. The embarrassing concomitant requirement of subexchangeability of histogram categories over strictly positive histograms provides further grounds for suspicion.

A second way to compare our work with Hill's is to imagine that the distinct cow types we are considering are refinements of the general Holstein type, with the refinements *defined* in terms of discrete ranges of milk yields. Accordingly, each of our exhaustive possible measurement values for X_i (the integers 0, 1, ... , K) would identify a distinct type. In these terms, Hill's parameter M would equal our K+1, rather than an unknown quantity. Apparent "ties" among our measurements do mask real differences that are not worthwhile to notice. Hill has generated results relevant to this context, but in which real number observations are recorded for each unit measured. Since his individual measurements are proposed as differing mildly within a type, only the average measurement and the number of measurements of each type are recorded. The inferential probabilities that Hill derives from A(n) in this context constitute the predictive distribution of the true average measurements from many units of each type, conditioned on the observed measurements of a few from each type. In contrast, the probabilities that we have analyzed in the present paper represent exclusively the probability distribution for the value of the next observed measurement conditioned on specified measurement possibilities.

In a word, the results of our paper extend the type of analysis developed in Hill's A(n) theory to the purely finite and discrete realm found in many practical applications, according to our specific finite interpretation of the measurement process. The proof scheme for our theorems is based upon the unifying structure of de Finetti's fundamental theorem of probability, specified as a simple proposition in linear algebra when applied to a finite discrete problem.

Acknowledgments. Thanks to Bill Studden and to several readers for helpful suggestions. An earlier version of this paper was presented to the 40th meeting of the NBER-NSF Seminar on Bayesian Inference in Econometrics, in honor of Morris DeGroot, at George Washington University, Washington D.C. 4-5 May, 1990.

REFERENCES

Abromovitz, M. and Stegun, I. eds. (1964). *Handbook of Mathematical Functions.* National Bureau of Standards, Washington D.C. Also (1965) Dover, New York.

Akhiezer, N.I. (1965). *The Classical Moment Problem, and Some Related Questions in Analysis*. N. Kemmer, tr. Oliver and Boyd, London.

Berliner, L. M. and Hill, B. M. (1988). Bayesian nonparametric survival analysis, *J. Amer. Statist. Assoc.* **83** 772-784.

Bruno, G. and Gilio, A. (1980). Applicazione del metodo del simplesso al teorema fondamentale per le probabilita' nella concezione soggettivistica. *Statistica* **40** 337-344.

Daboni, L. (1982). Exchangeability and completely monotone functions. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 39-45. North Holland, Amsterdam.

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

de Finetti, B. (1937). Foresight: its logical laws, their subjective sources. H. Kyburg, tr. In *Studies in Subjective Probability* (H. Kyburg and H. Smokler, eds.) Second edition, 1980, Krieger, New York.

de Finetti, B. (1949). On the axiomatic theory of probability. G. Majone, tr. In *Probability, Induction, and Statistics*. (B. de Finetti, 1972) John Wiley, New York.

de Finetti, B. (1974, 1975). *Theory of Probability*, 2 volumes. A.F.M. Smith and A. Machi, trs. John Wiley, New York.

Heath, D. and Sudderth, W. (1976). de Finetti's theorem for exchangeable random variables. *Amer. Statist.* **30** 188-189.

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

Hill, B. (1987). Parametric models for A(n): splitting processes and mixtures. University of Michigan research report, Ann Arbor.

Hill, B. (1988). De Finetti's theorem, induction, and A(n) or Bayesian nonparametric predictive inference. In *Bayesian Statistics 3, Proceedings of the Third Valencia International Meeting* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, eds.) 211-229. Clarendon Press, Oxford.

Hill, B., Lane, D., and Sudderth, W. (1988). Exchangeable urn processes. *Ann. Prob.* **15** 1586-1592.

Jeffreys, H. (1939). *Theory of Probability*. University Press, Cambridge.

Karlin, S. and Studden, W. (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*. Wiley Interscience, New York.

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

Lad, F. and Taylor, W.F.C. (1992). The moments of the Cantor distribution. *Stats. Prob. Letts.* **13** 307-310.

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

DEPARTMENT OF MATHEMATICS and STATISTICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND