On the predictive use of insufficient statistics: an intriguing family of distributions

Frank LAD
Department of Mathematics and Statistics
University of Canterbury, Christchurch, New Zealand
E-mail: F.Lad@math.canterbury.ac.nz

Romano SCOZZAFAVA
Dipartimento Metodi e Modelli Matematici
Università “La Sapienza", 00161 Roma, Italy
E-mail: romscozz@dmmm.uniroma1.it

Keywords: Coherence, Exchangeability, Sequential Forecasting
AMS 1991 subject classifications: primary 62E10, 62L99; secondary 62B99, 62M20

ABSTRACT – We derive the most general relation that coherency requires of the simultaneous assertion of a probability mass function for the sum of $N+1$ ordered events, and a conditional probability function for the final event given each possible value of the sum of the first N events. We then use this relation to characterise the family of all distributions on $N+1$ events that support the well-known sequential forecasting equations (conditioning only on a sum) that are motivated by exchangeable assessments. Surprisingly, this intriguing family is much larger than the family of exchangeable distributions, but is included within the family of all pairwise exchangeable (and thus equiprobable) distributions. Of course the sum is not generally a sufficient statistic for this family. We display some small numerical examples, and we discuss the implication of this discovery for applied sequential forecasting.

RÉSUMÉ – Nous dérivons la relation la plus générale que la cohérence demande entre l'assertion simultanée de la loi de probabilité pour la somme de $N+1$ événements ordonnés et la probabilité conditionnelle pour l'événement final, etant donnés tous les possibles valeurs de la somme des premiers N événements. Ensuite nous employons cette relation pour caractériser la famille de toutes le lois de probabilité sur $N+1$ événements qui supportent les bien connues équations pour la prévision en série (en conditionnant seulement à une somme) qui dérive d'assignations échangeables. Avec surprise, cette famille "intrigante" est beaucoup plus grande que la famille de lois échangeables, mais elle est incluse dans la famille des lois deux à deux échangeables (et ainsi équiprobable). Naturellement la somme n'est pas en général une statistique suffisante pour cette famille. Nous montrons des petits exemples numériques et nous discutons les implications de cette découverte pour les applications à la prévision en série.

1 Introduction

In his seminal lectures at the Institute Henri Poincaré on the concept of coherent prevision, de Finetti (1937) introduced both his now-celebrated "representation theorem" for exchangeable distributions and the lesser known "fundamental theorem of prevision." The latter theorem identifies the bounding coherent implications of any finite list of prevision assertions (including perhaps conditional previsions) on the further assertion of prevision for any other specific quantity. For an exposition and extensive discussion, see Lad (1996).

The most important applications of the structure of exchangeable distributions using sufficient statistics (the sum in the case of events, the histogram in the case of more general quantities) have been to sequential forecasting. This is achieved in the case of events via the equations

$$P(E_{K+1}|S_K = a) \;=\; \frac{\binom{K+1}{a}\; P(S_{K+1} = a+1)}{\binom{K+1}{a+1}\; P(S_{K+1} = a) \;+\; \binom{K+1}{a}\; P(S_{K+1} = a+1)} \qquad (1)$$

for $a = 0, 1, ..., K$, at least in the "usual case" that $P(S_{K+1} = a) > 0$ for each value of $a = 0, 1, ..., K+1$. The subscripted symbol S_K denotes the sum of the first K events. When $N+1$ events are regarded exchangeably, all equations of the form of equation (1) hold for each value of $K = 1, 2, ..., N$. For this reason they are called "sequential forecasting equations." The article of de Finetti (1952) addresses the unusual case in which some of the probabilities $P(S_{K+1} = a)$ may equal zero. For a general study of conditional assessments allowing zero probability for the conditioning events, see Coletti and Scozzafava (1996).

In the present article we use the fundamental theorem of prevision to develop a surprising result: that the family of all distributional assertions that support equations (1) is much larger than merely the exchangeable distributions. We characterise this “intriguing family of distributions” and we study and exemplify some of its properties. In Section 2 we remind the reader of a preliminary inversion relation between sequential conditional forecasting probabilities and the distribution for a sum that is honoured by exchangeable distributions. We use this “intriguing” relation in subsequent algebraic details. We then begin our analysis in Section 3 with a derivation of the most general relation between a probability mass function for the sum of $N+1$ events and the inferential conditional probabilities $P(E_{N+1}|S_N = a)$ for $a = 0, 1, ..., N$ that is required by coherency. In Section 4 we characterise the structure of intriguing distributions by appending the inversion relation of Section 2 to these general coherency conditions. Section 5 presents small numerical examples of non-exchangeable but intriguing distributions. These provoke us in Section 6 to establish that intriguing distributions are properly contained within the family of all pairwise exchangeable distributions. We conclude in Section 7 with some commentary on the meaning and applicability of intriguing distributions to statistical problems of sequential forecasting.

The keen reader will have noticed something unusual in our syntax to this point, for example in our allusion to "the sum of specific events." Throughout this article, we use analytic conventions favoured by de Finetti in his construction of the operational subjective theory of probability. For the unfamiliar reader, the following “translations” should make both our syntax and semantics understandable. Our “events” are *numbers* that would correspond to the indicator function of events in the measure theoretic formulation of probability (so, in particular, the sum alluded to in the Abstract is nothing else that the number of “successes” out of N “trials”). “Prevision” corresponds to an expectation, and is denoted by a capital letter P. Of course, all previsions (expectations) for events are also probabilities, thus meriting the unified notation for these two operators. Other quantities that are not events correspond to general (non $0-1$ valued) random variables. Previsions for quantities that are not indicator events are merely expectations of random variables. Whenever parentheses surround an arithmetic expression that may be true and may be false, e.g. $(S_K = a)$, the entire parenthetical expression denotes an event that equals 1 if the interior expression proves true, and 0 if false.

2 A preliminary relation and the question it poses

Studying a characterisation of inference regarding exchangeable events, Lad, Deely and Piesse (1995) examined the inversion equations relating the probability mass function for the sum of $N+1$ events, $P(S_{N+1} =$

$a)$, for $a = 0, 1, ..., N+1$, to the $N+1$ conditional assertions of $P(E_{N+1}|S_N = a)$, for $a = 0, 1, ..., N$. Denoting the latter N+1 conditional probabilities by the vector $p_{N+1} \equiv (p_o, p_1, ..., p_N)^T$ and the unconditional probabilities for the sum by the vector $q_{N+2} \equiv (q_o, q_1, ..., q_{N+1})^T$, the inversion equations are

$$q_o = \frac{1}{1+\sum_{a=0}^{N} \binom{N+1}{a+1} \prod_{i=0}^{a} \frac{p_i}{1-p_i}} \quad ,$$

and

$$q_a = \binom{N+1}{a} \prod_{i=0}^{a-1} \frac{p_i}{1-p_i} \, q_o \ , \ \text{for } a = 1, ..., N+1. \tag{2}$$

That is, if a distribution over $N+1$ events is exchangeable, then the components of the vector q_{N+2} cohering with the asserted vector p_{N+1} are proportional to the products of sequentially expanding numbers of conditional odds ratios. These equations invert the sequential forecasting equations (1) which are expressed in the notation of p_{N+1} and q_{N+2} as

$$p_a = \frac{\binom{N+1}{a} q_{a+1}}{\binom{N+1}{a+1} q_a + \binom{N+1}{a} q_{a+1}} \ , \ \text{for } a = 0, 1, ..., N \tag{3}$$

at least in the usual case of strictly positive probability mass functions for S_N that we mentioned in the Introduction. We also restrict our attention to this case in the present article.

Although the inversion equations (2) are implied by the judgment to regard the $N+1$ events in question exchangeably, adherence to them does *not* require that the individual events are assessed exchangeably. For exchangeability forces a stronger requirement than (3): a similar equation would be required for the conditional probability of *each one* of the $N+1$ events E_i given that the sum of the other N equals a. It also forces the sufficiency of S_{N+1} for the distribution of E_{N+1}. Nonetheless, all distributions that satisfy equations (2) do share some of the properties of exchangeable distributions, and are intriguing for several reasons. We refer to them as "intriguing distributions". The most important of the shared properties are the sequential forecasting equations (3) conditioned on the sum of the ordered events. However we show that although intriguing distributions properly contain the exchangeable distributions, they are properly contained in the family of all pairwise exchangeable distributions.

The satisfaction of the intriguing equations (2) merely for a particular size of N does *not* imply via coherency the satisfaction of the corresponding equations of the same form for $N-1$. However, the assertion of the p.m.f. values for the sum of $N+1$ events (the vector q_{N+2}) concomitantly with equations (2) does imply the p.m.f. values for the sum of only the first N events, S_N. In defining the family of intriguing joint distributions over the *ordered* events $E_{N+1} \equiv (E_1, E_2, ..., E_{N+1})$, we require the satisfaction of the intriguing equations (2) based on $N+1$ events as it is stated, *and also* the satisfaction of the corresponding equations for all "smaller sizes of N." Let us state this formally as a definition.

Definition: A joint distribution over the ordered vector of events E_{N+1} is an **intriguing distribution** if, for every $n = 1, 2, ..., N$, the probability mass function for the sum, q_{n+2}, of the ordered vectors of events E_{n+1} and the corresponding conditional probability function, p_{n+1}, satisfy equations (2) (written with n in place of N).

Before investigating this family, let us develop the general coherent relations between q_{N+2} and p_{N+1} that govern *every* joint distribution over E_{N+1}.

3 Coherent relations between the distribution for a sum and conditional probabilities for the final event given the accumulated sum

Consider $N+1$ events about which are asserted both a probability distribution for their sum (via the vector q_{N+2}), and a conditional probability function for the final event given each possible value of the sum of the first N events (via the vector p_{N+1}). To avoid complications, we consider here only problems in which p_{N+1} lies strictly within the $(N+1)$-dimensional unit-hypercube, not on the boundary. The reason is evident in equation (2). On their own of course, coherency would restrict the vector q_{N+2} only to lie within the $(N+1)$-dimensional unit-simplex, while allowing p_{N+1} to lie anywhere within the unit-hypercube of the same dimension. But what are the strongest coherency conditions required of their *simultaneous assertion*, supposing nothing more than that p_{N+1} is strictly within the hypercube, not on its boundary? (This would imply, of course, that q_{N+2} is also strictly within the unit-simplex, not on its boundary).

In the first place, it should be recognised that the $2N+3$ quantities assessed via q_{N+2} and p_{N+1} can all be expressed as functions of only $N+2$ events: E_{N+1} and $(S_N = a)$ for $a = 0, 1, ..., N$. The smallest partition these events generate that is relevant to every probability presumed to be asserted via p_{N+1} and q_{N+2} is constituted by the $2N+2$ events we shall denote by

$$C_a^1 \equiv (S_N = a)E_{N+1} \quad \text{and} \quad C_a^o \equiv (S_N = a)\tilde{E}_{N+1}$$

for $a = 0, 1, ..., N$, where $\tilde{E}$ denotes the *negation* of the event E. Arithmetically, $\tilde{E} \equiv 1 - E$. Using this notation for the constituents of the partition, notice that

$$(S_{N+1} = a) = C_a^o + C_{a-1}^1 \qquad \text{and} \qquad (S_N = a) = C_a^o + C_a^1 \ .$$

The assertion of probabilities specified by q_{N+2} and p_{N+1} implies a unique probability distribution over these constituents via the $2N + 3$ linear restrictions it places on their cohering valuations. We shall denote the induced probability valuations for these constituents by $\alpha_a^i \equiv P(C_a^i)$ for the values of $a = 0, 1, ..., N$ and $i = 0, 1$, referring in short to the vector of all these α_a^i as α.

It is instructive to exhibit the details. Let us begin by constructing the algebraic restrictions on a probability distribution over this partition of events that are imposed merely by the assertion of values for the $N + 1$ conditional probabilities p_{N+1} strictly within the unit-hypercube, $(0, 1)^{N+1}$. These would imply the $N + 1$ equations

$$\alpha_a^1 \;=\; p_a\;(\alpha_a^1 + \alpha_a^0)\;\;,\;\;\text{for } a = 0, 1, ..., N \tag{4}$$

for the reason that $P[(S_N = a)E_{N+1}] = P[E_{N+1}|(S_N = a)]P(S_N = a)$. An additional universal restriction on the values of the α_a^i is the summation constraint over the partition:

$$\sum_{a=0}^{N}\sum_{i=0}^{1}\alpha_a^i \;=\; 1\;\;. \tag{5}$$

These $N + 2$ equations in $2N + 2$ unknowns exhaust the coherency restrictions entailed in the assertion of the conditional probabilities p_{N+1}.

Since it is well-known that any assertion of conditional probabilities of the type expressed by p_{N+1} within the unit-hypercube is coherent, there surely exist many vectors $\alpha > 0$ of solutions to equations (4) and (5). Particular solutions could be identified by arbitrarily selecting probabilities for N of the constituents, i.e., N components of α.

However there is a more informative way to understand the solutions to the iterative constraints. Notice that any one of the solutions for the vector α would determine the cohering values of q_{N+2}, since each $q_a = \alpha_a^0 + \alpha_{a-1}^1$. Thus, there surely exist many choices of values for numbers we denote by $\beta_1, \beta_2, ..., \beta_N$, each within the interval $(0, 1)$ for which

$$\alpha_a^1 = \beta_{a+1} q_{a+1}$$

for $a = 0, 1, ..., N - 1$, and

$$\alpha_a^0 = (1 - \beta_a) q_a$$

for $a = 1, 2, ..., N$. In the context of strictly positive probability mass functions q_{N+2}, these values of β_a then correspond to the conditional probabilities $P[E_{N+1}|(S_{N+1} = a + 1)]$, since $\alpha_a^1 \equiv P[E_{N+1}(S_N = a)]$ must also equal $P[E_{N+1}(S_{N+1} = a + 1)]$. Moreover for the special cases of $a = N$ and $a = 0$, these two equations require respectively that $\beta_{N+1} = 1$ and $\beta_0 = 0$. So the N free choices for components of α correspond to N component values of a vector $\beta_N = (\beta_1, \beta_2, ...\beta_N)$ each within $(0, 1)$.

Now restating the restrictions of equations (4) and (5) in terms of these values of β_a yields the system of N equations

$$\beta_a q_a \;=\; p_{a-1}[\beta_a q_a + (1-\beta_{a-1})q_{a-1}] \;,\; \text{ for } a = 1, ..., N \tag{6}$$

along with an $(N+1)^{st}$ equation

$$q_{N+1} \;=\; p_N[q_{N+1} + (1-\beta_N)q_N] \tag{7}$$

and of course the $(N+2)^{nd}$ summation equation that forces the components of q_{N+2} to sum to 1. Equation (7) corresponds to the case of (4) when $a = N$, recalling that $\alpha^1_N = q_{N+1}$.

Equations (6) and (7) amount to $N+1$ non-linear relations of q_{N+2} to p_{N+1}, expressed in only N unknowns, the components of the vector β_N. Now if a mass function vector q_{N+2} were actually asserted concomitantly with the p_{N+1}, the first N of these equations (6) yield solutions for the vector β_N, which coherence would require to lie within $(0,1)^N$. The $(N+1)^{st}$ equation (7) then specifies a further compatibility condition. This can be seen from the solution of this system in terms of the vector β_N. The first N equations (6) resolve to

$$\beta_a \;= \sum_{r=0}^{a-1}(-1)^r \frac{q_{a-1-r}}{q_a} \prod_{i=a-1-r}^{a-1} \frac{p_i}{1-p_i} \;,\; \text{ for } a = 1, ..., N \tag{8}$$

which coherency would require each to lie within the interval $(0,1)$, since $\alpha^1_{a-1} = \beta_a q_a$.

Inserting these solutions for β_a into the $(N+1)^{st}$ equation (7) then yields the following compatibility condition between the components of q_{N+2} and p_{N+1}:

$$q_{N+1}\,(\frac{1-p_N}{p_N}) \;=\; q_N \;-\; \sum_{r=0}^{N-1}(-1)^r q_{N-1-r} \prod_{i=N-1-r}^{N-1} \frac{p_i}{1-p_i} \quad . \tag{9}$$

Equation (9) together with the restrictions of equation (8) that each β_i lie within $(0,1)$, represent the general coherency conditions on the simultaneous assertion of q_{N+2} and p_{N+1} for which we have searched. It does *not* presume exchangeability. Let us state this result formally as a theorem.

Theorem 1: A strictly positive probability mass function for the sum of $N+1$ ordered events, q_{N+2}, coheres with a conditional probability function p_{N+1} if and only if together they specify values of $\beta_1, \beta_2, ..., \beta_N$ each within the interval $(0,1)$ via equation (8) and they also comply with the restriction of equation (9).

A simple example when $N = 2$ should suffice for the moment. Suppose $p_3 = (.25, .5, .75)$. The assertions of q_4 either as $(.3, .2, .2, .3)$ or as

$(3/11, 5/22, 5/22, 3/11)$ would both satisfy equations (8) with $\beta_2 = (.5, .5)$ and $\beta_2 = (.4, .6)$ respectively, and they both also satisfy equation (9). However the assertion of $q_4 = (.15, .1, .1, .65)$ would be incoherent since it satisfies (8) with $\beta_2 = (.5, .5)$ but not (9). The vector $q_4 = (.1, .2, .3, .4)$ on the other hand would not even satisfy equation (8).

4 Identifying the structure of intriguing distributions

The judgment of exchangeability regarding the events E_{N+1} implies a relation between the components of q_{N+2} and p_{N+1} that is stricter than equation (9). It is the relation that we expressed in Section 2 via the "intriguing equations" (2), and which we restate here more simply without the proportionality constant:

$$q_a \;\propto\; \binom{N+1}{a} \prod_{i=0}^{a-1} \frac{p_i}{1-p_i} \;, \;\text{ for } a = 0, 1, ..., N+1, \tag{10}$$

using the convention that $\prod_{i=0}^{-1} \equiv 1$. Of course the proportionality is determined by the sum of the specified values of q_a.

Since (9) is a more general condition than (10), one finds that equation (9) is satisfied when the components of q_{N+2} are replaced by their expressions in p_{N+1} stipulated by (10). Moreover, using the restrictions of (10) in the equations (8) yields the specific solutions

$$\beta_a \;\equiv\; P(E_{N+1}|S_{N+1} = a) \;=\; \frac{a}{N+1} \;, \;\text{ for } a = 1, 2, ..., N. \tag{11}$$

Let us state this constructed result formally as a theorem.

Theorem 2: An intriguing probability distribution on $N+1$ ordered events requires the specific conditional probability assertions

$$P(E_{N+1}|S_{N+1} = a) = \frac{a}{N+1} \;, \;\text{ for } a = 1, 2, ..., N \,.$$

It is interesting that for each value of a this is only one of the $N+1$ similar restrictions that must hold if the $N+1$ events were regarded exchangeably. For in that case the equation

$$P[(S_{N+1} = a)E_i] = \frac{a}{N+1} P(S_{N+1} = a)$$

must hold for every $i = 1, ..., N+1$. An event of this form can occur in $\binom{N}{a-1}$ different ways, each assessed under exchangeability with the probability $q_a / \binom{N+1}{a}$. Summing them yields this result.

It is also interesting that intriguing distributions allow great range of coherent assertion values for the conditional probabilities $P(E_{N+1}|S_N = a)$. In fact any vector p_{N+1} within the unit-hypercube coheres with many intriguing distributions. Nonetheless, while the seemingly appealing assertions

$$P(E_{N+1}|S_N = a) = \frac{a}{N}$$

for each $a = 1, 2, ..., N-1$ can be a coherent array of assertions for any finite value of N, coherency then would require the same structure of conditional assertions for all "smaller values of N", which is not so appealing. Moreover, the assertion of

$$P(E_{N+1}|S_N = a) = \frac{a}{N} \quad \text{for every value of N}$$

and $a = 1, 2, ..., N-1$ would be incoherent! See the article of Lad, Deely, and Piesse (1995) for details.

Notice now that the N conditions of equation (11) exhaust the coherent implications of the intriguing equations (2 or 10) for the joint distribution of E_{N+1}, for a specific value of N. These, then, are N linear conditions on probabilities for the constituents of the partition generated by E_{N+1}. Remember, however, that by definition, if the distribution on E_{N+1} is to be considered intriguing, the marginal distribution on E_N must be intriguing too: thus, conditions of the form of equation (11) must pertain to each "smaller value of N" as well! At this point let us address the implications of this reductive aspect of intriguing distributions on E_{N+1}.

4.1 Restrictions induced by the reductive aspect of intriguing distributions

It is evident that the intriguing conditions (11), together with the assertion of the p.m.f. for the sum of the $N+1$ events denoted by q_{N+2}, imply the p.m.f. for the sum of the first N events as well. For when equations (11) hold, we can compute

$$\begin{aligned} P(S_N = a) &= P[(S_N = a)E_{N+1}] + P[(S_N = a)\tilde{E}_{N+1}] \\ &= [\frac{a+1}{N+1}]\, P(S_{N+1} = a+1) \; + \; [1 - \frac{a}{N+1}]\, P(S_{N+1} = a) \end{aligned} \qquad (12)$$

So the feature of intriguing distributions over E_{N+1} that they must margin to an intriguing distribution over E_N implies $N-1$ more linear conditions on the partition probabilities in the form of equations (11) at this level. Continuing, then, the further reduction to a marginal distribution that is intriguing over E_{N-1} places $N-2$ more restrictions; and so on down to the final condition on the distribution for E_2 that $P[(S_2 = 1)E_2] = P(S_2 = 1)/2$. As a result of all these reductive implications of our definition then, the family of all intriguing distributions on

E_{N+1} is specified by

$$N + (N-1) + (N-2) + ... + 1 \;=\; \frac{N(N+1)}{2}$$

linearly independent linear conditions on the probabilities for the constituents of the partition generated by E_{N+1}.

4.2 The dimension of the intriguing family

Our constructive discussion has also resulted in a theorem on the dimensionality of the family of intriguing distributions. Of course the family of all coherent distributions over E_{N+1} has dimension $2^{N+1} - 1$. (We presume the logical independence of the events). The loss of 1 dimension from the size of the partition these events generate is due to the summation constraint on probabilities over a partition. Moreover, the dimension of the family of exchangeable distributions on E_{N+1} is only $N + 1$, recognised as the unit-simplex in this dimension. It is characterised by the distribution over the sum of the events, S_{N+1}. This result derives from the fact that exchangeability places

$$\binom{N+1}{1} - 1 + \binom{N+1}{2} - 1 + ... \binom{N+1}{N} - 1 = 2^{N+1} - N - 2$$

further linear restrictions on the family of all distributions. Thus,

$$2^{N+1} - 1 - (2^{N+1} - N - 2) = N + 1$$

is the dimension of the exchangeable family.

Using the same algebraic logic of coherency, we can recognise that the intriguing conditions specify only $N(N+1)/2$ conditions through the required conditions identified in Theorem 2, that

$$P(E_{K+1} | S_{K+1} = a) = \frac{a}{K+1},$$

for $a = 1, 2, ..., K$ and $K = 1, 2, ..., N$. We thus understand the following:

Theorem 3: The dimension of the family of all intriguing distributions over the logically independent ordered events E_{N+1} is $2^{N+1}-1-N(N+1)/2$.

Table 1 displays the relative sizes of $D(F), D(E), D(I)$, and $D(PE)$ for selected values of N, where F, E, I, and PE represent the families of all distributions, exchangeable distributions, intriguing distributions, and pairwise exchangeable distributions, respectively. Conditions of pairwise exchangeable distributions are equivalent to those of equiprobable distributions. Thus, their dimension over E_{N+1} arises from merely N linear restrictions additional to the summation constraint: $D(PE) = 2^{N+1} - 1 - N$. A complete analysis of PE is deferred until Section 6 after we study some suggestive examples for the cases of $N = 2$ and $N = 3$.

Table 1. Dimensions of various families of distributions.

N	$D(F)$	$D(E)$	$D(I)$	$D(PE)$
1	3	2	2	2
2	7	3	4	5
3	15	4	9	12
4	31	5	21	27
5	63	6	48	58
10	4095	11	4044	4085

5 Examples of special cases when $N = 2$ and $N = 3$

The case of $N = 1$ is trivial, involving only the events E_1 and E_2. They generate a partition of size 4. The only intriguing condition is $P[(S_2 = 1)E_2] = P(S_2 = 1)/2$. Supplemented by the p.m.f. assertion $q_3 = (q_0, q_1, q_2)$, this is enough to identify the intriguing distributions $(N = 1)$ as the exchangeable distributions over two events.

The simple algebra is worth displaying, since it formalises the logical process that will be followed in larger cases as well. The central equation required by coherency is

$$P\begin{pmatrix} E_1 \\ E_2 \\ \text{------} \\ 1 \\ \text{------} \\ (S_2 = 0) \\ (S_2 = 1) \\ (S_2 = 2) \\ \text{------} \\ (S_2 = 1)E_2 \end{pmatrix} = \begin{pmatrix} ? \\ ? \\ \text{---} \\ 1 \\ \text{---} \\ q_o \\ q_1 \\ q_2 \\ \text{---} \\ q_1/2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ \text{--} & \text{--} & \text{--} & \text{--} \\ 1 & 1 & 1 & 1 \\ \text{--} & \text{--} & \text{--} & \text{--} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \text{--} & \text{--} & \text{--} & \text{--} \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} . \tag{13}$$

The first equality identifies all the probability assertions that have been presumed in the problem. To the right of the second equal sign appears firstly the matrix whose columns list all the possibile observation values of the unknown column of quantities whose probabilities are being asserted (or not) listed after the P operator at the far left. This matrix is known as the *realm matrix* for the vector of quantities. The first two of the quantities in the vector, E_1 and E_2, generate the exhaustive list of possibilities under consideration. Nowhere in the specification of the problem are their probabilities asserted. Thus, their assertion values are denoted by a question mark “?”, in the vector of probability assertion values. The remaining events in the quantity vector are all defined by functions of E_1 and E_2. Their assertion values, expressed generally in terms of components of q_3 in keeping with the restriction on intriguing distributions, should be understood as numbers. The intriguing condition should be recognised in the final row of this linear equation: $P[(S_2 = 1)E_2] = P(S_2 = 1)/2$, in the context that $P(S_2 = 1)$ has been specified asserted at the value of q_1. It is the principle of coherency that requires the satisfaction of equation (13). For a coherent prevision (probability in this case) vector must lie within the convex hull of the columns of the realm matrix. The convexity is assured by the third line in the vector equation, which requires the components of c_4 to sum to 1.

The five linear restrictions on c_4 have rank 4 since the partition events defined by S_2 surely sum to 1. Thus, as long as the assertion vector $q_3 = (q_o, q_1, q_2)$ is nonnegative and sums to 1, these assertions are coherent. Moreover the restrictions identify precisely one cohering probability mass function for E_2 via the solution for c_4: $c_1 = q_o$, $c_2 = q_1/2$, $c_3 = q_1/2$, and $c_4 = q_2$. This, obviously, is the exchangeable distribution on E_2 specified by q_3, for it ensures that $P(E_1\tilde{E}_2) = P(\tilde{E}_1 E_2)$. So in the case of $N = 1$ the intriguing distributions are the same as the exchangeable distributions and the equiprobable distributions, and are characterised by the family of distributions on the sum, S_2.

5.1 The case of N=2

Now when $N = 2$, the central equation expands to

$$P\begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ ------ \\ 1 \\ ------ \\ (S_3 = 0) \\ (S_2 = 1) \\ (S_3 = 2) \\ (S_3 = 3) \\ ------ \\ (S_3 = 1)E_3 \\ (S_3 = 2)E_3 \\ ------ \\ (S_2 = 1)E_2 \end{pmatrix} = \begin{pmatrix} ? \\ ? \\ ? \\ ----- \\ 1 \\ ----- \\ q_o \\ q_1 \\ q_2 \\ q_3 \\ ----- \\ q_1/3 \\ 2q_2/3 \\ ----- \\ (q_1 + q_2)/3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ - & - & - & - & - & - & - & - \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ - & - & - & - & - & - & - & - \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ - & - & - & - & - & - & - & - \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_8 \end{pmatrix}. \tag{14}$$

The final row constraint in this equation comes from applying the sum reduction equation (12), which yields $P(S_2 = 1) = (2/3)P(S_3 = 2) + (2/3)P(S_3 = 1) = (2/3)(q_1 + q_2)$, and then realising that $P((S_2 = 1)E_2) = P(S_2 = 1)/2$. It is evident from the displayed equation that there are only seven linearly independent restrictions on the constituent probabilities c_8. Thus, there is one dimension of free choice in their solution in addition to the three dimensions of choice in the specification of the vector

q_4. These four dimensions of freedom represent the four dimensions of the family of intriguing distributions in the case of $N = 2$.

Algebraically, the solutions to the constituent probabilities are

$$c_8^T = (q_o,\ 2q_1/3 - c_3,\ c_3,\ q_2/3,\ q_1/3,\ (q_2 - q_1)/3 + c_3,\ (q_1 + q_2)/3 - c_3,\ q_3)\ . \tag{15}$$

For each choice of q_4 in the three-dimensional unit-simplex, there is a whole dimension of solutions to the cohering constituent probabilities, parameterised here by the value of c_3. Thus, the space of intriguing distributions is larger than that of the exchangeable distributions by one dimension, as we had seen in Table 1. Limitations on the value of c_3 (and thus on the length of this one-dimensional line segment) depend on the companion assertion of q_4. It turns out that if $q_1 > q_2$, then $c_3 \in ((q_1 - q_2)/3, (q_1 + q_2)/3)$; whereas if $q_1 < q_2$, then $c_3 \in (0, 2q_1/3)$.

Further features of the general solution can be seen in this example. It is easily determined, for example, from these solutions that the individual events must be regarded equiprobably, and thus pairwise exchangeably. For $P(E_1) = P(E_2) = P(E_3) = q_1/3 + 2q_2/3 + q3$. This feature is worth checking, because we know that the intriguing distributions include but exceed the exchangeable ones, which of course are equiprobable. Alternatively, this same result can be seen by checking that these intriguing distributions all respect the property of pairwise exchangeability. This results from the observation that $P(E_1(1 - E_2)) = c_2 + c_6 = (q_1 + q_2)/3$, while $P((1 - E_1)E_2) = c_3 + c_7 = (q_1 + q_2)/3$, the same value. Similarly, we find that $P(E_1(1 - E_3)) = P((1 - E_1)E_3) = 2q_1/3 + q_2/3 - c_3$, which can easily differ from $(q_1 + q_2)/3$. Although all intriguing family members are equiprobable (and equivalently pairwise exchangeable), complete exchangeability holds only for those members for which $c_3 = q_1/3$.

However, not all pairwise exchangeable distributions are intriguing! Consider, for examples, the pairwise exchangeable distributions identified by $c_8 = (1, 5, 5, 1, 1, 5, 5, 1)/24$ and $c_8 = (7, 1, 1, 7, 7, 1, 1, 7)/32$. (The pairwise exchangeability is checked more easily by noticing that they are equiprobable distributions). These distributions on E_3 support different distributions on the sum S_3, $q_4 = (1, 11, 11, 1)/24$, and $q_4 = (7, 9, 9, 7)/32$, respectively. Interestingly however, both of these distributions would imply $p_3 = (.5, .5, .5)$ by direct computation; whereas this specification of p_3 would imply the distribution $q_4 = (1, 3, 3, 1)/8$ via the inverted intriguing equations (2). So the distributions corresponding to these two specification of c_8 are both pairwise exchangeable, but neither is intriguing. We shall analyse this situation generally in Section 6 after a brief glance at the case of N = 3.

Despite the fact that intriguing distributions are defined by the exchangeable relation of the conditional forecasting probabilities to the distribution of the sum, the sum is not a sufficient statistic for the informative content of observations for the general intriguing family member. This property is commonly referred to as "predictive sufficiency." Observe in this example that

$$\begin{aligned} P(E_3|E_1\tilde{E}_2) &= c_6/(c_2+c_6) = (c_3 - q_1/3 + q_2/3)/[(q_1+q_2)/3] \ , \quad \text{and} \\ P(E_3|\tilde{E}_1 E_2) &= c_7/(c_3+c_7) = [(q_1+q_2)/3 - c_3)/[(q_1+q_2)/3] \ , \quad \text{while} \\ P(E_3|(S_2=1)) &= (c_6+c_7)/(c_2+c_3+c_6+c_7) \ = \ q_2/(q_1+q_2) \ . \end{aligned}$$

These three conditional probabilities would all be equal only in the case of $c_3 = q_1/3$, the exchangeable member of the intriguing family.

5.2 The case of N=3

When $N = 3$, the central equation expands to

$$P\begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ \text{------} \\ 1 \\ \text{------} \\ (S_4=0) \\ (S_4=1) \\ (S_4=2) \\ (S_4=3) \\ (S_4=4) \\ \text{------} \\ (S_4=1)E_4 \\ (S_4=2)E_4 \\ (S_4=3)E_4 \\ \text{------} \\ (S_3=1)E_3 \\ (S_3=2)E_3 \\ \text{------} \\ (S_2=1)E_2 \end{pmatrix} = \begin{pmatrix} ? \\ ? \\ ? \\ ? \\ \text{-----} \\ 1 \\ \text{-----} \\ q_o \\ q_1 \\ q_2 \\ q_3 \\ q_4 \\ \text{-----} \\ q_1/4 \\ 2q_2/4 \\ 3q_3/4 \\ \text{-----} \\ (3q_1/4 + q_2/2)/3 \\ (q_2/2 + 3q_3/4)(2/3) \\ \text{-----} \\ q_1/4 + q_2/3 + q_3/4 \end{pmatrix} = \begin{pmatrix} 0\;1\;0\;1\;0\;1\;0\;1\;0\;1\;0\;1\;0\;1\;0\;1 \\ 0\;0\;1\;1\;0\;0\;1\;1\;0\;0\;1\;1\;0\;0\;1\;1 \\ 0\;0\;0\;0\;1\;1\;1\;1\;0\;0\;0\;0\;1\;1\;1\;1 \\ 0\;0\;0\;0\;0\;0\;0\;0\;1\;1\;1\;1\;1\;1\;1\;1 \\ -------- \\ 1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1\;1 \\ -------- \\ 1\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0 \\ 0\;1\;1\;0\;1\;0\;0\;0\;1\;0\;0\;0\;0\;0\;0\;0 \\ 0\;0\;0\;1\;0\;1\;1\;0\;0\;1\;1\;0\;1\;0\;0\;0 \\ 0\;0\;0\;0\;0\;0\;0\;1\;0\;0\;0\;1\;0\;1\;1\;0 \\ 0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;1 \\ -------- \\ 0\;0\;0\;0\;0\;0\;0\;0\;1\;0\;0\;0\;0\;0\;0\;0 \\ 0\;0\;0\;0\;0\;0\;0\;0\;0\;1\;1\;0\;1\;0\;0\;0 \\ 0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;0\;1\;0\;1\;1\;0 \\ -------- \\ 0\;0\;0\;0\;1\;0\;0\;0\;0\;0\;0\;0\;1\;0\;0\;0 \\ 0\;0\;0\;0\;0\;1\;1\;0\;0\;0\;0\;0\;0\;1\;1\;0 \\ -------- \\ 0\;0\;1\;0\;0\;0\;1\;0\;0\;0\;1\;0\;0\;0\;1\;0 \end{pmatrix} \mathbf{c}_{16} \ . \tag{16}$$

Again, the final three constraint values of the previsions are determined by applying the reduction equation (12) to the p.m.f. values for S_4 yielding the p.m.f. on S_3, and then again on S_2. The eleven linearly independent constraints on the components of c_{16} generate the following solution with five degrees of freedom, parameterised by the choices of c_5, c_6, c_7, c_9, and c_{14}:

$$\begin{aligned} c_{16}^T \quad = \quad & (q_o,\ q_1/4,\ 3q_1/4 - c_5 - c_9,\ -q_1/2 + q_2/6 + c_5 + c_9, \\ & c_5,\ c_6,\ c_7,\ q_1/4 + q_2/3 + q_3/4 - c_5 - c_6 - c_7, \\ & c_9,\ q_1/2 + q_2/3 - c_5 - c_6 - c_9,\ q_2/3 - q_3/4 - c_7 + c_{14}, \end{aligned}$$

$$-q_1/4 - q_2/3 + q_3/2 + c_5 + c_6 + c_7 - c_{14},\ q_2/6 + q_3/4 - c_{14},$$
$$c_{14},\ q_3/4,\ q_4)\ .$$

The derivation of such results is conducted most easily with algebraic programming software such as MAPLE. The five free choices of c_i along with four components of q_5 exhaust the dimension of the intriguing family in this case.
We now turn to a proof that all intriguing distributions are equiprobable, and thus pairwise exchangeable.

6 Intriguing distributions are equiprobable distributions

In Section 5.1 we have observed examples of equiprobable distributions that are not intriguing. Nonetheless, every intriguing distribution is equiprobable. This can be proved by finite induction.

Firstly, we have observed that intriguing distributions are equiprobable when $N = 2$ and when $N = 3$. Now suppose that an intriguing distribution specified for some value of N is equiprobable for the implied intriguing distribution specified by $N-1$. We shall prove that the conditions imposed by the intriguing property of the full distribution at the N^{th} stage then imply that the distribution is also equiprobable through this stage.

Consider the equation

$$\begin{aligned} E_N - E_{N+1} \quad = \quad & (S_{N+1} = N) \\ & -(S_{N+1} = 1)E_{N+1} - (S_{N+1} = 2)E_{N+1} - \cdots - (S_{N+1} = N-1)E_{N+1} \\ & -2(S_{N+1} = N)E_{N+1} \\ & +(S_N = 1)E_N + (S_N = 2)E_N + \cdots + (S_N = N-1)E_N \end{aligned} \tag{17}$$

which we shall now prove by enumeration. Probabilities for each of the events on the right hand side of equation (17) have been specified through the presumed assertion of q_{N+2} and the conditions that make the distribution intriguing. Once we have certified that equation (17) holds, we shall notice that probabilities for all the right-hand side events sum to 0. Thus, we will have proved that $P(E_N) = P(E_{N+1})$, establishing that intriguing distributions for any finite vector of events are equiprobable.

As to equation (17), let us check its validity for each of the four possible observation pairs of (E_N, E_{N+1}), these being $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$.

If $(E_N, E_{N+1}) = (0,0)$, then every term on the right-hand side of equation (17) must equal 0: for every summand term in the second through fourth lines of the identity includes either E_N or E_{N+1} as a factor; moreover, the single term in the first line must also equal zero, since in this case the sum S_{N+1} can equal at most $N-1$. So the equation holds.

Now suppose $(E_N, E_{N+1}) = (1, 0)$. In this case every term among the summands on the second and third lines of the right-hand side equals 0, because of the factor E_{N+1} that occurs in each. In addition, in this case the numerical value of S_N may equal only the values $1, 2, ..., N-1$ or N. In the first $N-1$ of these cases, there is exactly one positive summand in the fourth line of the right-hand side corresponding to each of these possibilities, exactly one of which must then equal 1, as required for the equation to hold. Finally, in the case that $S_N = N$, under this scenario it must also be true that $S_{N+1} = N$. In this case, the only nonzero term on the right-hand side would be the first term, $(S_{N+1} = N)$, which equals 1. Thus, the equation must hold.

If $(E_N, E_{N+1}) = (0, 1)$, then all summands on the right-hand side that include the factor E_N must equal 0. All remaining terms contain as a factor exactly one of the events $(S_{N+1} = 1)$ through $(S_{N+1} = N)$, exactly one of which must equal 1. Of these, if S_{N+1} equals any of 1 through $N-1$, then only that relevant term appearing in line 3 will equal -1 (on account of the negative sign) while all other terms, including the two terms containing the event $(S_{N+1} = N)$, will equal 0. Both sides of the equation will then equal -1. Whereas, if $S_{N+1} = N$, then only those two terms including the factor $(S_{N+1} = N)$ are nonzero, the first term (in line 1) with a coefficient of $+1$ and the second term (in line 3) with coefficient -2. Thus, again, both sides of the equation (17) would equal -1, as required to establish that it holds.

Finally, suppose $(E_N, E_{N+1}) = (1, 1)$. In this case the value of S_N can only equal one of the values $1, 2, ..., N-1$, or N. If it does equal N, then S_{N+1} must equal $N+1$. In this case, every term on the right-hand side of equation (17) equals 0, so the equation holds. But if $S_N = N-1$, then only the terms $(S_{N+1} = N)$ and $(S_N = N-1)E_N$ equal 1, while the term $-2(S_{N+1} = N)E_{N+1}$ would then equal -2, forcing the entire right-hand side to equal 0 as required, since all other terms on the right-hand side would equal 0. Whereas if S_N equals any of 1 through $N-2$, then S_{N+1} must equal correspondingly 2 through $N-1$. There is exactly one positive and one negative term on the right-hand side in each of these cases, establishing that the equation holds. We have now covered all cases, so equation (17) is established.

Let us now consider the prevision required for both sides of equation (17) according to the linearity required by coherency. First consider the fact that for any $a = 1, ..., N-1$, $P((S_{N+1} = a)E_{N+1}) = aq_a/(N+1)$ as required by the intriguing property of the distribution we have established as Theorem 2. Furthermore,

$$P((S_N = a)E_N) = \frac{a}{N}P(S_N = a) = \frac{a}{N}\left(\frac{a+1}{N+1}\, q_{a+1} \;+\; [1 - \frac{a}{N+1}]\, q_a\right)$$

as we have established by the reduction equation (12) for $P(S_N = a)$.

Assessing prevision for both sides of equation (12) by substituting these values on the right-hand side then yields

$$
\begin{aligned}
P(E_N) - P(E_{N+1}) \quad = \quad & q_N \\
& - q_1/(N+1) - 2q_2/(N+1) - \cdots - (N-1)q_{N-1}/(N+1) \\
& - 2Nq_N/(N+1) \\
& + (1/N)[(2/(N+1))q_2 + (N/(N+1))q_1] \\
& \quad + (2/N)[(3/(N+1))q_3 + ((N-1)/(N+1))q_2] \\
& \qquad + \cdots + ((N-1)/N)[(N/(N+1))q_N + (2/(N+1))q_{N-1}] \\
= \quad & 0 \quad . \qquad (18)
\end{aligned}
$$

The fact that the right-hand side of equation (18) equals 0 is recognised by collecting terms with common factors.

7 Concluding remarks

The sequential forecasting equations conditional on the sum of preceding events that are motivated by exchangeability have been found to specify coherent inferences for a much larger family of distributions than merely the exchangeable ones. The single feature that characterises this larger intriguing family is the regularity of the conditional probability function that they honour: $P(E_{N+1}|S_{N+1} = a) = a/(N+1)$ for $a = 0, 1, ..., N+1$.

Our identification of this family of distributions can prove useful for two reasons. In some instances, a forecaster may specifically reject the constancy of probability over permutations of occurrences that exchangeability requires. A simple example would be the sequential order of males and females who enter a lecture hall. In many cultural contexts one might well assess sequences in which females arrive in clusters of a specific size as more likely than sequences in which every female is interspersed between two males. Nonetheless, one may yet recognise as appropriate the abovementioned conditional probabilities that characterise the intriguing distributions. In such a case one could coherently use the intriguing forecasting equations conditioned on a sum.

Secondly, it is not at all uncommon that inexperienced investigators routinely collect data, sometimes quite interesting, on a sum of events without recording the complete sequence of individual responses. For many reasons that an analyst might not want to regard exchangeably the actual sequence of responses that were not recorded, the coherent intriguing inferential equations may still prove appropriate and useful for inferential data summary, even if the sum is not formally sufficient.

It has long been known that the exchangeable distributions are characterised by the sufficiency of the sum of any N events for the conditional probability of any other event, equivalent to conditioning on the complete ordered sequence of occurrences and non-occurrences. (A completion of this result extends it by characterising forms of partial exchangeability in terms of the dimension of the sufficient statistic vector supporting a distribution. This is described in a beautiful article by Diaconis and Freedman (1984).) At any rate, for those intriguing distributions that are not exchangeable then, there will be a difference between the conditional probability for E_{N+1} given the sum of N events and the conditional probability given the complete ordered sequence that generates that sum. Our investigations into the systematic structure on bounds for the order of magnitude of these differences are proceeding.

References

Coletti, G., Scozzafava, R. (1996) Characterization of coherent conditional probabilities as a tool for their assessment and extension, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **4**, 103-127.

Diaconis, P., Freedman, D. (1984) Partial exchangeability and sufficiency, in *Statistics: Applications and New Directions*, J. K. Ghosh and J. Roy (eds.), Calcutta: Indian Statistical Institute, 205-236.

de Finetti, B. (1937) La prevision: ses lois logiques, ses sources subjective, *Annales de L'Institute Henri Poincaré*, **7**, 1-68.

de Finetti, B. (1952) Gli eventi equivalenti e il caso degenere, *Gior. Ist. Ital. Att.*, **I-2**, 40-64. (Reprinted with English translation, Equivalent events and the degenerate case, in *Probabilità e Induzione*, 1993, Bologna: CLUEB, Biblioteca di Statistica, 129-152, 389-413.)

Lad, F. (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.

Lad, F., Deely, J., and Piesse, A. (1995) Coherency conditions for finite exchangeable inference, *Journal of the Italian Statistical Society*, **4**, 195-213.

Acknowledgments: Research support was provided by the University of Canterbury through an Erskine Grant to R. Scozzafava during August-October, 1996, and by the University of Perugia to F. Lad during June, 1998.

Thanks to Andrea Piesse for many helpful comments.