May 27, 1982

SOME APPLICATIONS AND EXTENSIONS OF THE DE FINETTI REPRESENTATION THEOREM*

E. T. Jaynes

Arthur Holly Compton Laboratory of Physics

Washington University

St. Louis, Missouri 63130

Abstract. The de Finetti representation, when extended by a trivial modification to cover finite as well as infinite exchangeable sequences, becomes a powerful analytical tool for dealing with real statistical problems.

*To appear in Bayesian Inference and Decision Techniques with Applications: Essays in Honor of Bruno de Finetti, Prem K. Goel and Arnold Zellner, Editors, North-Holland Publishers, Amsterdam.

1. INTRODUCTION

The contributions of Bruno de Finetti to the general philosophy of probability theory, and to understanding of mathematical puzzles (finite additivity) and pathologies (non-conglomerability) that can arise when we try to apply it on infinite sets, have been well recognized and appreciated, here and elsewhere. To show the other side of the coin, we note some of the ways in which de Finetti's work has contributed to the technical solution of real problems.

Probability distributions that are symmetric under permutations of a finite number of variables arise constantly in applications. For example, it is known that a bank has n depositors, with total deposits T. Our joint probability distribution for the accounts will be a function $p(x_1 \ldots x_n)$, where x_i, the size of the i'th account, is presumably in (0,T). Whatever other information we may possess about the distribution of wealth, banking habits, etc. in the community, in the absence of any information that could distinguish one depositor from another $p(x_1 \ldots x_n)$ will be symmetric under permutations of the x_i.

One might then think that any such distribution could be represented in the integral form [Eq. (1) below] given by de Finetti, as generalized by Hewitt and Savage (1955); however, this turns out not to be the case because of the finiteness of n. The original de Finetti representation theorem holds only for subsets of an infinitely long exchangeable sequence, and not every finite symmetric sequence is such a subset. While one may feel intuitively that the difference cannot be very serious, in principle there is still unfinished business concerning the technical question of applicability of the representation (1) in real problems, which always involve finite sequences.

More generally, the variables x_i might be defined on almost any set X, as Hewitt and Savage showed in far greater generality than we need. But the phenomenon we wish to show arises already for the binary set X = {1,0} considered in the original theorem of de Finetti (Kyburg and Smokler, 1981). What we demonstrate here for the binary set will hold, mutatis mutandis, for any set X that is likely to arise in a real problem.

That there are difficulties with finite sequences was noted by Feller (1971). He proved de Finetti's representation theorem by reducing it to the Hausdorff moment problem, and then gave examples which "show that the theorem fails for finite sequences". But he offered no alternative for the finite case.

Heath and Sudderth (1976) gave such an alternative as part of a completely different proof in which the representation (1) emerged as the limit of a set of "Urn model" distributions for finite sequences. Their argument not only gives a much deeper understanding of the result; it is so simple that it seems futile to look for a still simpler one. But again, the integral representation (1) was proved only for subsets of an infinitely long sequence, and Heath and Sudderth also show by example that it can fail for finite sequences.

Of course, merely exhibiting examples of failure does not prove that (1) fails for all finite sequences; this question and also the question whether, in cases where it fails, another equally convenient integral representation may still hold, were left open by these arguments.

A closely related difficulty was discovered about twenty years ago in physics. Here a joint probability distribution for the positions of n particles, $p_n(x_1 \ldots x_n)$, arises as the marginal distribution of a symmetric distribution function P_N for a larger number $N > n$ of particles. Since typically n was 2 or 3, while N was 10^{23}, one sought to represent p_n without bringing in the details of the larger distribution P_N. Some anomalous results were obtained before it was realized that not every symmetric n-particle function can be obtained by marginalization from a symmetric N-particle one.

The problem of finding the necessary and sufficient conditions on an n-point distribution which guarantee that it has this parentage, is called the N-representability problem in the literature of Statistical Mechanics. That it is related to the theory of convex sets was realized at once (Coleman, 1963), but to the best of the writer's knowledge an explicit solution has not been given.

In the following we augment these discussions by showing that (A) the original de Finetti representation (1) does indeed fail for a wide class of finite sequences; (B) nevertheless, with a trivial modification (dropping the non-negativity condition) it is resurrected, and the de Finetti representation then holds for all finite sequences, leading to solution of the N-representability problem.

2. THE GENERATING FUNCTION

Following the terminology advocated by Jimmie Savage in 1960's, a sequence $\{x_1, x_2, \ldots x_N\}$ whose probability distribution remains symmetric as $N \to \infty$ is called <u>exchangeable</u>. The probability $p(x_1 \ldots x_N)$ can then depend only on the total number R of successes. The original theorem

asserts that, given any subset of n trials from an exchangeable sequence, the probability that it contains exactly r successes can be written in the form

$$p(r|n) = \int_0^1 \binom{n}{r} z^r (1-z)^{n-r} \, g(z)dz \qquad (1)$$

where it is necessary that $g(z) \geq 0$ and that

$$\int_0^1 g(z)dz = 1 \quad . \qquad (2)$$

Thus an exchangeable sequence is characterized completely by a single generating function g(z).

Mathematically, (1) is a weighted average of binomial distributions, and on that intuitive ground the representation (1) was introduced by Laplace in 1774. He obtained many useful results--in particular the Rule of Succession--by interpreting g(z) as the posterior density of a parameter z, and therefore (1) as a predictive distribution for future observations. But for over a century Laplace's method was in disrepute because of conceptual difficulties [g(z)dz seemed to be a "probability of a probability"].

In 1937 de Finetti obtained the representation (1) by a completely different argument, which showed that Laplace's fault lay not in the alleged metaphysics of his method, but only in his failure to establish its essential assumptions and generality. The results Laplace obtained are now reinstated as important parts of Bayesian statistics (Lindley, 1976). Henceforth we shall call (1) the Laplace-de Finetti (LdF) representation; its apparent failure for finite sequences remains as a puzzle to resolve.

3. AN EXAMPLE

As soon as we know what to look for, it is easy to produce almost trivial-looking examples of this failure and its relation to the N-representability problem; it arises already in the comparison of two and three tosses of a coin. A two-point symmetric probability distribution $p(x_1,x_2) = p(x_2,x_1)$ on the set $X = \{0,1\}$ is determined by three numbers:

$$\begin{aligned} p(11) &= A \\ p(01) = p(10) &= B \\ p(00) &= C \end{aligned} \tag{3}$$

and the possible distributions are represented by all non-negative A, B, C satisfying

$$A + 2B + C = 1 \quad . \tag{4}$$

Thus any point on or within the triangle PQR of Fig. 1 represents a possible two-point symmetric probability distribution.

But now suppose our two-point function is to be derived from a symmetric three-point one. For this we may write

$$\begin{aligned} p(111) &= a \\ p(011) = p(101) = p(110) &= b \\ p(100) = p(010) = p(001) &= c \\ p(000) &= d \quad . \end{aligned} \tag{5}$$

The possible three-point symmetric functions are given by all non-negative (a,b,c,d) satisfying

$$a + 3b + 3c + d = 1 \quad . \tag{6}$$

But these distributions are related by marginalization:

$$A = a + b$$

$$B = b + c$$

$$C = c + d$$

and now the normalization (6) imposes the constraint

$$1/3 \leq (A + C) \leq 1 \tag{7}$$

since $(b+c)$ cannot exceed 1/3. Therefore for a two-point symmetric function to be derivable from a symmetric three-point one, its representative point must lie on or within the quadrilateral PQST in Fig. 1. For it to be derivable from a symmetric 4-point function, we find a further restriction to a five-sided polygon inside PQST, etc. N-representability involves much intricate detail.

What two-point functions are representable in the original LdF form? This requires

$$A = \int_0^1 z^2 \, g(z)dz$$

$$B = \int_0^1 z(1-z) \, g(z)dz$$

$$C = \int_0^1 (1-z)^2 \, g(z)dz \quad . \tag{8}$$

After a little algebra, we find that this implies

$$2(A+C) - (A-C)^2 - 1 = 2\int_0^1 dy \int_0^1 dz(y-z)^2 \, g(y) \, g(z) \tag{9}$$

and so if $g(z) \geq 0$ we can generate only those two-point functions for which (9) is non-negative. This limits us to points on or above the tilted parabola in Fig. 1. The two-point distributions representable in LdF form are therefore limited to the parabolic slice PQU.

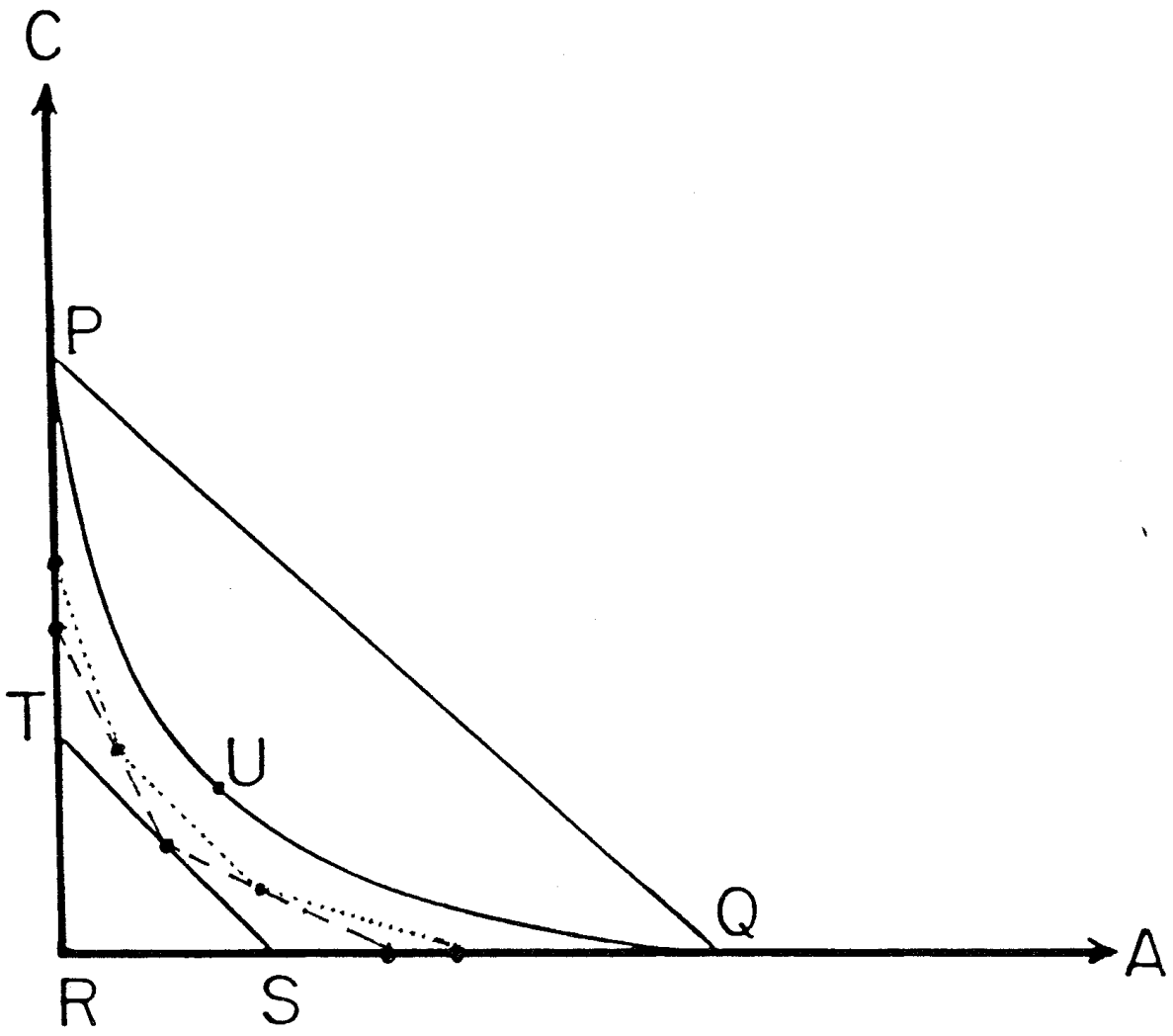

Figure 1. Exchangeable 2-point distributions derivable from N-point ones. A = P(2 heads), C = P(2 tails). PQR = all possible distributions. Derivability from a 3-point function truncates this to PQST. 4-point and 5-point parentage further truncate PQST as indicated by the dashed and dotted lines. As $N \to \infty$ these nested polygons tend to the parabola, representing independent Bernoulli trials. The original Laplace-de Finetti representation yielded only distributions on or above the parabola, where negative correlations are impossible.

We can understand this restriction intuitively by noting that the right-hand side of (9) is just four times the covariance of (x_1,x_2). The limitation to the parabolic slice therefore signifies that the LdF representation is incapable of giving negative correlations. Yet in our opening example of the bank depositors, knowledge of the total deposits T clearly does impose a negative correlation in $p(x_1 \ldots x_n)$. Thus the bank distribution has no generalized LdF representation.

Of course, if we wish only to represent a symmetric distribution of finite length, the Heath-Sudderth Urn representation is available. But while their proof was very simple, their result (a probability mixture of hypergeometric distributions) is not. As an analytical tool, the integral form (1) would be much easier to use. Also, a simpler Urn model may be defined, in which we put in (1)

$$g(z) = \sum_{k=0}^{n} U_k \, \delta\left(z - \frac{k}{n}\right) \tag{10}$$

(here the k'th Urn, containing a fraction k/n of red balls, is chosen with probability U_k, and n balls are drawn from it with replacement). However, an argument like (9) applies; if $U_k \geq 0$ not every symmetric distribution $p(x_1 \ldots x_n)$ can be generated from (10).

As this example shows, the general N-representability problem is nontrivial and threatens to become very complicated. Marginalization from an N-point function to an n-point one involves a summation

$$p(x_1 \ldots x_n) = \sum p(x_1 \ldots x_N) \tag{11}$$

over $\{x_{n+1} \ldots x_N\}$. The resulting intricate details suggest that we are using an awkward set of variables; for abstractly, marginalization is nothing but a projection operation. All its properties would be retained, and the problem should become easy, if we could find a "natural" coordinate

system in which this marginalization reduces to an elementary geometrical projection, parallel to the axes, of a point rather than a distributed mass.

The LdF representation comes close to accomplishing this. If we view the marginalization (11) in the LdF representation, what is happening is that the part $\hat{g}(z)$ of $g(z)$ that affects only details of sequences longer than n drops out from the expression for $p(r|n)$. Therefore, $\hat{g}(z)$ must be orthogonal to the coefficients of $g(z)$ in (1). These coefficients are Bernstein polynomials of degree n; for $r = (0,1,\ldots,n)$ they form a complete set for expansion of any polynomial of degree n. The projection property we seek is therefore one of orthogonality to all polynomials of a given degree.

4. ORTHOGONAL EXPANSIONS

The polynomials that are orthogonal with respect to uniform weight over a finite interval are not the Bernstein polynomials, but the Legendre polynomials $P_n(x)$ scaled to the interval:

$$\int_0^1 P_n(2z-1)\, P_m(2z-1)dz = (2n+1)^{-1}\, \delta_{nm} \quad . \tag{12}$$

Since $P_n(2z-1)$ is itself a polynomial in z of degree n, $P_k(2z-1)$ is orthogonal to every polynomial of degree $n < k$:

$$\int_0^1 P_k(2z-1)\, z^n\, dz = 0 \quad , \quad 0 \leq n < k \quad . \tag{13}$$

Therefore, if we expand the Laplace-de Finetti generating function in the form

$$g(z) = \sum_{k=0}^{\infty} (2k+1)\, a_k\, P_k(2z-1) \tag{14}$$

where the factor (2k+1) is inserted to simplify later formulas, the LdF representation reduces to a finite sum:

$$p(r|n) = \sum_{k=0}^{n} A_{rk}^{(n)} a_k \quad , \quad 0 \leq r \leq n \qquad (15)$$

where the coefficients $A_{rk}^{(n)}$, evaluated explicitly in the Appendix, represent the transformation from Bernstein to Legendre polynomials:

$$\binom{n}{r} z^r (1-z)^{n-r} = \sum_{k=0}^{n} A_{rk}^{(n)} P_k(2z-1) \quad , \quad 0 \leq r \leq n \quad . \qquad (16)$$

Now the crucial point is the observation that for r = (0,1,...,n) the Bernstein polynomials, although not orthogonal, are linearly independent; therefore the (n+1) x (n+1) matrix $A^{(n)}$ is nonsingular and (15) is uniquely invertible:

$$a_k = \sum_{r=0}^{n} B_{kr}^{(n)} p(r|n) \quad , \quad 0 \leq k \leq n \quad , \qquad (17)$$

where $B^{(n)}$, the inverse matrix to $A^{(n)}$, is also given explicitly in the Appendix.

The expansion coefficients $\{a_0, a_1, a_2, \ldots\}$ are therefore the "natural coordinates" that we sought. For any n, the probabilities $f_r \equiv p(r|n)$, $0 \leq r \leq n$ defining an n-point symmetric probability distribution $p(x_1 \ldots x_n)$, are determined uniquely by the first (n+1) expansion coefficients $\{a_0 \ldots a_n\}$; but these coefficients are in turn determined uniquely by $\{f_0 \ldots f_n\}$. This makes both the finiteness puzzle and the N-representability problem simple.

5. FINITE SEQUENCES

Comparing (15) and (9), we see that the LdF representation was unable to give all finite symmetric sequences because the allowable expansion coefficients a_i had been restricted by the non-negativity condition $g(z) \geq 0$. But this was in fact the only reason; for (14) and (17) constitute a proof by construction that for any finite symmetric distribution $p(x_1 \ldots x_n)$ we can find a set of expansion coefficients, and therefore a generating function $g(z)$, for which the representation (1) will hold; with the only difference that if there are negative correlations, then $g(z)$ must be allowed to become negative.

In our example of $n = 2$, we find from the value of $B^{(2)}$ given in the Appendix that any point in the triangle PQR of Fig. 1 can now be reached by the generating function

$$g(z) = 1 + 3(A - C)(2z - 1) + 5(A - 4B + C)(6z^2 - 6z + 1) \qquad (18)$$

but when the point (A,C) is outside a weird region bounded by two straight lines and the arc of an ellipse, $g(z)$ becomes negative somewhere in (0,1).

With hindsight, it is easy to see heuristically how non-negativity got into the original theorem. When $n \to \infty$ the binomial distribution in (1) goes into a delta-function, $(n+1)^{-1}\, \delta(z-r/n)$; and so the probability that the frequency (r/n) lies in $(z < r/n < z + dz)$ tends to $g(z)dz \geq 0$, in agreement with Laplace's intuitive meaning of the generating function $g(z)$.

At first glance, one would think that extending to $n \to \infty$ should increase the generality of the representation; but in view of (14), (17) it was actually restricting its generality! The properties of Legendre polynomials give us a new proof of the representation, in which this is obvious.

However, the important new content of de Finetti's theorm did not concern the length of the sequence, but rather the demonstration of something that Laplace probably never suspected; namely that every exchangeable sequence can be so represented.

Conclusion: we have only to drop the non-negativity condition, and then every finite exchangeable sequence has a de Finetti integral representation. This gives us the means to solve many real problems, including N-representability.

We find similarly that the simpler Urn model (10) can represent any finite exchangeable sequence, if we allow some of the U_k to become negative.

6. N-REPRESENTABILITY

Given an exchangeable N-point distribution $p(x_1 \ldots x_N)$, the n-point distribution $n < N$, obtained from it by marginalization is the one whose first $(n+1)$ expansion coefficients $\{a_0, a_1, \ldots a_n\}$ are the same (but $a_0 = 1$ is the normalization condition always satisfied). Conversely, given an n-point distribution, what is the condition that it be derivable from some N-point one? We cannot just assign new expansion coefficients $\{a_{n+1}, \ldots, a_N\}$ arbitrarily, because in general this would not lead to non-negative probabilities for the N-point distribution:

$$p(R|N) = \sum_{k=0}^{N} A_{Rk}^{(N)} a_k \geq 0 \quad , \quad (0 \leq R \leq N) \tag{19}$$

So our expansion coefficients have been freed from one non-negativity condition $g(z) \geq 0$ only to become entangled in a new one (19).

But because of the invertibility (17), this new condition can be dealt with at once. Think of the of the quantities $\{f_R = p(R|N),\ 0 \leq R \leq N\}$ as cartesian coordinates of a point in an (N+1)-dimensional space, restricted by $\Sigma f_R = 1$, $f_R \geq 0$ to an N-dimensional convex set F, a generalization of the familiar simplex triangle in three dimensions. The N+1 vertices of F (the "points of the triangle") are determined by $\{f_R = 1,\ f_m = 0,\ m \neq R;\ 0 \leq R \leq N\}$. Therefore, in the inversion

$$a_K = \sum_{R=0}^{N} B_{KR}^{(N)} f_R \quad , \quad 0 \leq K \leq N \tag{20}$$

the point $\{a_0 \ldots a_N\}$ is restricted to a convex set (polyhedron) S_N whose R'th vertex has coordinates

$$\{a_0 \ldots a_N\}_R = \{B_{0R}^{(N)}, \ldots, B_{NR}^{(N)}\} \quad , \quad 0 \leq R \leq N \quad . \tag{21}$$

From any point $\{a_0 \ldots a_N\}$ of S_N, marginalization to a lower sequence of length n consists simply of parallel projection, throwing away the coordinates $\{a_{n+1} \ldots a_N\}$ and retaining $\{a_0 \ldots a_n\}$.

The solution of the N-representability problem is therefore: the necessary and sufficient condition that a symmetric n-point distribution $p(x_1 \ldots x_n)$ can be obtained by marginalization from a symmetric N-point one is that its representative point $\{a_0 \ldots a_n\}$ must lie on the projection of S_N. This is a polyhedron S_n whose vertices are contained in the set of projected vertices of S_N(some of these may be interior points of S_n).

In our example of Fig. 1, for the 2-point distribution be derivable from an N-point one, $\{a_1, a_2\}$ must lie on the (N+1)-sided polygon whose

R'th vertex is

$$a_1 = B_{1R}^{(N)} = \frac{2R - N}{N}$$

$$a_2 = B_{2R}^{(N)} = \frac{6R^2 - 6NR + N^2 - N}{N(N-1)} \quad . \tag{22}$$

Four of these polygons are imaged in Fig. 1, which is an oblique projection of the (a_1,a_2)-plane. $N = 2,3,4,5$ gives the regions PQR, PQST, and the further truncations of the dashed and dotted lines. As $N \to \infty$ they approach the parabola.

Likewise, for a symmetric 3-point distribution to be derivable from some n-point one, it is necessary and sufficient that its expansion coefficients lie on or in the polyhedron whose vertices are

$$\left\{a_1,a_2,a_3\right\} = \left\{B_{1r}^{(n)},B_{2r}^{(n)},B_{3r}^{(n)}\right\} \quad , \quad (0 \leq r \leq n) \quad . \tag{23}$$

But as $n \to \infty$, we state without proof that $B^{(n)}$ goes asymptotically into

$$B_{kr}^{(n)} \sim P_k\left[\frac{2r}{n} - 1\right] + O(n^{-1}) \quad , \tag{24}$$

and (23) goes into a twisted smooth curve (the generalization of the parabola) whose parametric equations are

$$\left\{a_1,a_2,a_3\right\} = \left\{P_1(x),P_2(x),P_3(x)\right\} \ , \ (-1 \leq x \leq 1) \quad . \tag{25}$$

The possible 3-point distributions are then defined by the convex hull of the curve (25); in this limit we regain the result of the original LdF representation.

For many applications we need to generalize these results, in the nammer of Hewitt-Savage, to larger sets than the binary one $X = \{0,1\}$. One can, of course, find new functions which generalize the properties

of Legendre functions to higher dimensions. A more powerful and abstract approach, which does not require us to go into all that detail, was discovered by Dr. Eric Mjolsness while he was a student of the writer's. We hope that, with its publication, the useful results of this representation will become more readily obtainable.

7. CONCLUSION

With the clearing up of one small technical detail, de Finetti's famous representation theorem is valid for all exchangeable sequences, finite or infinite. This enables us to make a significant reduction in many real problems; we note briefly how the appearance of kinetic theory is changed by this result.

A great deal of work has been done on what is called the Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy, a set of distribution functions for the positions x_i (in classical theory, also the momenta g_i) of one particle, two particles, --- on up to perhaps 10^{23} particles. The n-particle distribution $p_n = p(x_1 \ldots x_n;t)$ is not only symmetric and derived by marginalization from all higher distributions; it depends explicitly on the time t because of Schrödinger's (or Newton's) equations of motion.

Now it turns out that almost all observable macroscopic properties of a system--reversible or irreversible in the sense of thermodynamics--could be predicted if we could calculate how the two-particle function evolves in time. But in this endeavor we have been frustrated for decades by the fact that the recursion relations go the wrong way; i.e., the time evolution of p_2 depends on p_3, which is unknown. The evolution of p_3 depends in turn on p_4, which is even more unknown--and so on. In other words, the evolution of p_2 depends on all the correlations in all the higher order distributions.

The problem appears hopeless without making drastic simplifying assumptions that nobody believes.

But now, let us note the de Finetti representation. The distribution function for any order n is determined by a single generating functional g[f(x)], the Hewitt-Savage generalization of g(z):

$$p(x_1 \ldots x_n) = \int f(x_1) f(x_2) \ldots f(x_n)\ g[f(x)]df \qquad (26)$$

which is a functional integral over functions f(x), generalizing Eq. (1) above. Furthermore, any physical prediction we wish to make can be extracted directly from g[f(x)] by another functional integral like (26).

Therefore, we can reformulate kinetic theory in a way that sidesteps the recursion problem by never introducing the hierarchy of distributions at all. The object of our study becomes the generating functional g[f(x)], which contains full information about all correlation effects of arbitrarily high order. We seek the time evolution of g:

$$\frac{dg}{dt} = Kg \qquad (27)$$

where K is an operator that we need not write down here, but that physicists know how to find.

This idea is still new in physics and not well explored. But its superior precision and simplicity make us predict that, fifty years from now, kinetic theory will be based on the idea of the de Finetti functional, and will be an incomparably more powerful tool for prediction than our present one.

APPENDIX

From (16), the transformation matrix $A^{(n)}$ is given by

$$A_{rk}^{(n)} = (2k+1) \int_0^1 \binom{n}{r} z^r (1-z)^{n-r} P_k(2z-1)dz \qquad (A1)$$

The integral is a ${}_3F_2$ hypergeometric function whose power series terminates, giving the result

$$A_{rk}^{(n)} = (2k+1) \sum_{m=0}^{k} (-1)^m \frac{n!(k+m)!(n-r+m)!}{(n-r)!(k-m)!(n+m+1)!(m!)^2} \quad . \qquad (A2)$$

The columns of $A^{(n)}$ are orthogonal vectors, but not normalized; one can show by induction that

$$\sum_{r=0}^{n} A_{rk}^{(n)} A_{rs}^{(n)} = \frac{(2k+1)(n!)^2}{(n-k)!(n+k+1)!} \delta_{rs} \quad , \quad 0 \leq r, s \leq n \quad . \qquad (A3)$$

Therefore, the inverse of $A^{(n)}$ is

$$B_{kr}^{(n)} = \frac{(n-k)!(n+k+1)!}{(2k+1)(n!)^2} A_{rk}^{(n)} \quad , \; 0 \leq k, r \leq n \quad . \qquad (A4)$$

The first few of these matrices are:

$$B^{(1)} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

$$B^{(2)} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

$$B^{(3)} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1/3 & 1/3 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix} .$$

REFERENCES

A. J. Coleman (1963), "Structure of Fermion Density Matrices", Revs. Mod. Phys. 35, pp. 668-687.

Wm. Feller (1971), An Introduction to Probability Theory and Its Applications, Vol. I; J. Wiley & Sons, Inc., N. Y.; pp. 228-230.

D. Heath and Wm. Sudderth (1976), "de Finetti's Theorem on Exchangeable Variables" Am. Stat. 30, pp. 188-189.

E. Hewitt and L. J. Savage (1955); "Symmetric Measures on Cartesian Products", Trans. Am. Math. Soc. 80, pp. 470-501.

H. E. Kyburg and H. E. Smokler (1981), Studies in Subjective Probability 2nd edition, John Wiley & Sons, Inc., New York.

D. V. Lindley (1976), "Inference for a Bernoulli Process (a Bayesian View)", Amer. Stat. 30, pp. 112-119.