CONCENTRATION OF DISTRIBUTIONS AT ENTROPY MAXIMA[†]

E. T. Jaynes

Department of Physics, Washington University

St. Louis, Missouri 63130, U.S.A.

1. INTRODUCTION

It has long been recognized, or conjectured, that the notion of entropy defines a kind of measure on the space of probability distributions such that those of high entropy are in some sense favored over others. The basis for this was stated first in a variety of intuitive forms: that distributions of high entropy represent greater "disorder," that they are "smoother," that they are "more probable," that they "assume less" according to Shannon's interpretation of entropy as an information measure, etc. While each of these doubtless expresses an element of truth, none seems explicit enough to lend itself to a quantitative demonstration. This alone, however, has not prevented the useful exploitation of this property of entropy.

In many statistical problems we have information which places some kind of restriction on a probability distribution without completely determining it. If, given two distributions that agree equally well with the information at hand, we prefer the one with greater entropy, then the distribution with the maximum entropy compatible with our information will be the most favored of all. Thus conversion of prior information into a definite prior probability assignment becomes a variational problem in which the prior information plays the role of constraint.

[†]To be presented at the 19'th NBER-NSF Seminar on Bayesian Statistics, Montreal, October 1979

But while this Principle of Maximum Entropy has an established usefulness in a variety of applications, it has left an unanswered question in the minds of many. Granted that the distribution of maximum entropy has a favored status, in exactly what sense, and how strongly, are alternative distributions of lower entropy ruled out?

Probably most information theorists have considered it obvious that, in some sense, the possible distributions are concentrated strongly near the one of maximum entropy; i.e., that distributions with appreciably lower entropy than the maximum are atypical of those allowed by the constraints.

Likewise, Schrödinger (1948) noted that thiS the reason why, in statistical mechanics, the Darwin-Fowler method and the Boltzmann "method of the most probable distribution" lead to the same result in the limit $N \to \infty$, where N is a suitable "size" parameter (i.e., in statistical mechanics the number of particles in a system; in communication theory the number of symbols in a message; in statistical inference the number of trials of a random experiment). A general proof of this limiting form (i.e., a generalized Darwin-Fowler theorem) is given by van Campenhout and Cover (1979).

But these results, pertaining only to the limiting distribution, leave us in the same unsatisfactory state as did the original limit theorem of Jacob Bernoulli (1713): (as $N \to \infty$, the observable frequency $f = r/N$ of successes converges in probability to p). This said nothing about how large N must be for a given accuracy. For applications one needed the more explicit de Moivre-Laplace theorem: (Asymptotically, $f \sim N(p,\sigma)$ where $\sigma^2 = N^{-1}\, p(1-p)$).

Similarly, in our present problem it would be desirable to have a quantitative demonstration of this entropy concentration phenomenon for finite N, so that one can see just how the limit is approached. This is so particularly because there are still some who, apparently unaware or unconvinced of the reality of the phenomenon, reject the Principle of Maximum Entropy as a method of inference.

This problem was discussed at the M.I.T. Maximum Entropy Formalism Conference of May 1978, in connection with some alternative solutions that had been proposed for maximum entropy problems. The result was a lengthy but awkward and unsatisfactory analysis (Jaynes, 1978) in which real insight into the problem had not yet been achieved. We give here a simpler, more accurate, and more general treatment of entropy concentration.

The general Principle of Maximum Entropy is applicable to any problem of inference with a well-defined hypothesis space but incomplete information, whether or not it involves a repetitive situation such as a random experiment. However, we consider below only the special applications where we use entropy as a criterion for (1) estimating frequencies in a random experiment about which incomplete information is available; or (2) testing hypotheses about systematic effects in experiments where frequency data are available.

The second application is illustrated by analyzing the famous dice data of R. Wolf. We show how entropy analysis enables one to draw conclusions about the specific physical imperfections that must have been present (not knowing whether those dice are still in existence, so that our conclusions might be checked directly).

2. ENTROPY CONCENTRATION THEOREM

A random experiment has n possible results at each trial; thus in N trials there are n^N conceivable outcomes (we use the word "result" for a single trial, while "outcome" refers to the experiment as a whole; thus one outcome consists of an enumeration of N results, including their order). Each outcome yields a set of sample numbers $\{N_i\}$ and frequencies $\{f_i = N_i/N,\ 1 \le i \le n\}$, with an entropy

$$H(f_1 \ldots f_n) = -\sum_{i=1}^{n} f_i \log f_i \quad . \tag{1}$$

Consider the subclass C of all possible outcomes that could be observed in N trials, compatible with m linearly independent constraints $(m < n)$ of the form

$$\sum_{i=1}^{n} A_{ji} f_i = D_j \quad , \quad (1 \le j \le m) \quad . \tag{2}$$

The conceptual interpretation is that m different "physical quantities" have been measured, the matrix A_{ji} defines their "nature," and D_j are the particular "data" for the case under study. These data tell us that the actual outcome must have been in class C, but are insufficient to determine the frequencies $\{f_i\}$. We examine the combinatorial basis for using--and the consequences of failing to use--the entropy (1) as a criterion for estimating the $\{f_i\}$.

Although it is not needed for this purpose, we note that in a real application one will wish, if possible, to choose the constraint matrix A_{ji} so that the resulting quantities D_j represent systematic physical influences, real or conjectured, (for example, eccentric position of the center of gravity of a die), which constrain the frequencies to be different from the uniform distribution of absolute maximum entropy $H_0 = \log n$. In using entropy analysis for hypothesis testing, the mathematical relations are used in the other direction, considering the $\{f_i\}$ as known experimentally. A successful hypothesis about the systematic influences is then one for which the experimentally observed entropy (1) is sufficiently close to the maximum H_{max} permitted by the assumed constraints (2), "sufficiently close" being defined by the following concentration theorem.

A certain fraction F of the outcomes in class C will yield an entropy in the range

$$H_{max} - \Delta H \leq H(f_1 \ldots f_n) \leq H_{max} \qquad (3)$$

where H_{max} may be determined by the following algorithm: define the partition function

$$Z(\lambda_1 \ldots \lambda_m) \equiv \sum_{i=1}^{n} \exp\left(-\sum_{j=1}^{m} \lambda_j A_{ji}\right) \quad . \qquad (4)$$

Then

$$H_{max} = \log Z + \sum_{j=1}^{m} \lambda_j D_j \qquad (5)$$

in which the Lagrange multipliers $\{\lambda_j\}$ are found from

$$\frac{\partial}{\partial \lambda_j} \log Z + D_j = 0 \quad , \qquad (1 \leq j \leq m) \qquad (6)$$

a set of m simultaneous equations for m unknowns. The frequency distribution which has this maximum entropy is then

$$f_i = Z^{-1} \exp\left(-\sum_j \lambda_j A_{ji}\right) \quad , \quad (1 \leq i \leq n) \quad . \qquad (7)$$

Other distributions $\{f_i'\}$ allowed by the constraints (2) will have various entropies less than H_{max}. Their concentration near this upper bound (i.e., the functional relation connecting F and ΔH) is given by the Concentration Theorem: Asymptotically, $2N\Delta H$ is distributed over class C as Chi-squared with $k = n - m - 1$ degrees of freedom, independently of the nature of the constraints. That is, denoting the critical Chi-squared for k degrees of freedom at the 100 P % significance level by $\chi_k^2(P)$, ΔH is given in terms of the upper tail area (1-F) by

$$2N\,\Delta H = \chi_k^2(1-F) \quad . \tag{8}$$

The proof is relegated to the Appendix, since it consists of little more than repeating mutatis mutandis Karl Pearson's original derivation of the Chi-squared distribution, taking note of the reduction of dimensionality due to constraints. Note that the theorem is combinatorial, expressing only a counting of the possibilities; it does not become a statement of probabilities unless one assigns equal probability to each outcome in class C.

3. EXAMPLES: FREQUENCY ESTIMATION

We illustrate the meaning and use of this result by a much-discussed example. Suppose a die is tossed $N = 1000$ times and we are told only that the average number of spots up was not 3.5 as we might expect from a "true" die, but 4.5, i.e.,

$$\sum_{i=1}^{6} i\, f_i = 4.5 \tag{9}$$

which is a special case of (2). Given this information and nothing else, (i.e., not making use of any additional information that you or I might get from inspection of the die or from past experience with dice in general), what estimates should we make of the frequencies $\{f_i\}$ with which the different faces appeared? This is a kind of caricature of a class of real problems that arises constantly in physical applications.

The distribution which has maximum entropy subject to the constraint (9) is given by (4)-(7) with $n = 6$, $m = 1$, $A_{j1} = i$, $Z(\lambda) = (e^{-\lambda} + \ldots + e^{-6\lambda})$, $\lambda = -0.37105$. The result, derived in more detail before (Jaynes, 1978), is

$$\{f_1 \ldots f_6\} = \{0.0543,\ 0.0788,\ 0.1142,\ 0.1654,\ 0.2398,\ 0.3475\} \tag{10}$$

and it has entropy

$$H_{max} = 1.61358 \tag{11}$$

as compared to the value $\log_e 6 = 1.79176$, corresponding to no constraint and a uniform distribution.

Applying the concentration theorem, we have $6-1-1=4$ degrees of freedom; entering the Chi-squared tables at the conventional 5% significance level, we find that 95% of all possible outcomes allowed by the constraint (9) have entropy in the range (3) of width $\Delta H = (2N)^{-1}\ \chi_4^2(0.05) = 0.00474$; or, to sufficient accuracy,

$$1.609 \leq H \leq 1.614 \quad . \tag{12}$$

Thus on the "null hypothesis" which supposes that no further systematic influence is operative in the experiment other than the one taken into account (i.e., which assigns equal probability to all outcomes in class C), there is less than a 5% chance of seeing a frequency distribution with entropy outside the interval (12).

A remarkable feature is that the "95% concentration range"

$$H_{max} - \frac{4.74}{N} \leq H \leq H_{max} \tag{13}$$

is valid asymptotically for any random experiment with four degrees of freedom, although the value of H_{max} may vary widely with other details.

More interesting numerical results are found at more extreme significance levels. Thus, in any experiment with 1000 trials and four degrees of freedom, 99.99% of all outcomes allowed by the constraints have entropy in a range of width $\Delta H = (2N)^{-1}\ \chi_4^2(0.0001) = 0.012$. In the above example this is

$$1.602 \leq H \leq 1.614 \tag{14}$$

and only in 10^8 of the possible outcomes has entropy below the range

$$1.592 \leq H \leq 1.614 \quad . \tag{15}$$

Thus, given certain incomplete information, the distribution of maximum entropy is not only the one that can be realized in the greatest number of ways; in fact, for large N the overwhelming majority of all possible distributions compatible with our information have entropy very close to the maximum.

Note that the width of this region of concentration goes down like N^{-1}; and not like $N^{-1/2}$ as one might have guessed. Thus, in 20,000 tosses agreeing with (9), 95 percent of the possible outcomes have entropy in the interval $(1.61334 < H < 1.61358)$ and only one in 10^8 has $H < 1.61253$. As $N \to \infty$, any frequency distribution other than the one of maximum entropy thus becomes highly atypical of those allowed by the constraints.

Even more interesting numbers are readily found. Rowlinson (1970) rejected the principle of maximum entropy for this problem, and proposed as an alternative solution in place of (10) the binomial distribution

$$f_i' = \binom{5}{i-1} p^{i-1} (1-p)^{6-i} \quad , \quad 1 \leq i \leq 6 \tag{16}$$

which also satisfies the constraint (9) if $p = 0.7$. But the distribution (16) has entropy $H' = 1.4136 = H_{max} - 0.200$, far below the limit (15). We now have $2N \, \Delta H = 400 = \chi_4^2(1-F)$; or from (A8),

$$1 - F = 2.94 \times 10^{-84} \quad . \tag{17}$$

This indicates that in 1000 tosses, less than one in 10^{83} of the outcomes compatible with the constraint (9) have entropy as low as H'.

But the concentration theorem is valid only asymptotically, because of the approximation (A4) made in its derivation; and even for $N = 1000$ we might distrust its numerical accuracy that far out in the tail of the distribution. However, we can check the magnitude of (17) by direct counting.

The number of ways W in which a specific set of sample numbers $\{N_1 \ldots N_6\}$ can be realized is given by the multinomial coefficient (A1). The asymptotic formula (A3) for the ratio W/W' (which is free from any errors that might result from the aforementioned approximation) says that, for every way in which the binomial distribution (16) can be realized, there are about $\exp(N\Delta H) \approx \exp(200)$, or more than 10^{86} ways, in which the maximum-entropy distribution (10) can be realized (about 10^{62} ways for every microsecond in the age of the universe). While this result does not take into account the volume element factors $(r^{k-1}\, dr)$ of the full concentration theorem, it does indicate that (17) did not mislead us.

Even if we come down to $N = 50$, we find the following. The sample numbers which agree most closely with (10), (16) while summing to $\Sigma N_k = 50$ are $\{N_k\} = \{3,4,6,8,12,17\}$ and $\{N_k'\} = \{0,1,7,16,18,8\}$ respectively. With such small numbers, we no longer need asymptotic formulas; for every way in which the distribution $\{N_k'\}$ can be realized, there are exactly $W/W' = (7!16!18!)/(3!4!6!12!17!) = 38{,}220$ ways in which the maximum-entropy distribution $\{N_k\}$ can be realized.

Such numbers illustrate rather clearly just what we are accomplishing when we maximize entropy. If our data do not fully determine a distribution $\{f_i\}$ it is prudent to adopt, for purposes of inference, that distribution which has maximum entropy subject to the data we do have.

4. HYPOTHESIS TESTING: WOLF'S DICE DATA

The Swiss astronomer Rudolph Wolf (1816-1893; best known today as the discoverer of the correlation between terrestrial magnetic disturbances and sunspot activity) performed a number of random experiments, conducted with great care, presumably to check the validity of statistical theory. An account with references is given by Czuber (1908).

In one of these experiments, a red and white die were tossed together 20,000 times in a way that precluded any systematic favoring of any face over any other. The resulting 36 joint sample numbers are given in Table 1 (taken from Czuber).

Table 1. Wolf's Dice Data

		White Die 1	2	3	4	5	5	Row Total
Red Die	1	547	587	500	462	621	690	3407
	2	609	655	497	535	651	684	3631
	3	514	540	468	438	587	629	3176
	4	462	507	414	413	509	611	2916
	5	551	562	499	506	658	672	3448
	6	563	598	519	487	609	646	3422
	Column Total	3246	3449	2897	2841	3635	3932	20000

These are the sample numbers $\{N_i, 1 \le i \le n\}$ of a random experiment with $n = 36$ possible results at each trial. On the null hypothesis which assigns uniform probabilities $p = n^{-1} = 1/36$, the expectation and standard deviation of any sample number are $Np = 555.55$, $\sigma = [Np(1-p)]^{\frac{1}{2}} = 23.24$ respectively.

Czuber, writing in the days when commonly understood statistical inference consisted of little more than fitting by least squares, compared σ with the observed mean-square deviation

$$\left[n^{-1}\ \Sigma(N_i - Np)^2\right]^{\frac{1}{2}} = 76.87 \tag{18}$$

and concluded only that the null hypothesis must have been wrong; "die Würfelseiten nicht als gleichmögliche Fälle sich darstellen."

Feller, writing 58 years later and extolling, in his Preface to Vol. I, the "success of the modern theory" that had evolved in the interval, did even less with the data. Noting only that agreement with prediction of the null hypothesis was atrocious, he castigated Wolf for having wasted his time on apparatus of such poor quality.

Neither seems to have seen in such "bad" data on opportunity for further analysis, that would have been lost had Wolf worked with perfect dice and produced the kind of data expected of him. To the best of the writer's knowledge, no statistician has ever attempted to draw any specific inferences about the imperfections in Wolf's dice from these data.

Yet to a physicist, Wolf's data stand there, telling us something very clear and simple about the condition of those dice; information that can be extracted from the data by a straightforward entropy analysis that does not require us to go into complicated mechanical details.

Ludwig Boltzmann, writing thirty years before Czuber and about six years before Wolf's experiment, had given the principle by which this analysis may be carried out; and J. Willard Gibbs, writing six years before Czuber, had developed the resulting mathematical apparatus to a high degree of perfection. Yet today, 100 years after Boltmann's work, it still seems generally believed that the principles of statistical mechanics apply only to molecules; and not to dice.

We do not expect, and Wolf's data do not give evidence for, any correlations between the results of the two dice. Therefore, the import of the data for our purposes is contained in the marginal totals. The observed frequencies $\{f_i\}$ and their deviations $\{\Delta_i = f_i - 1/6\}$ from the null hypothesis prediction are given in Table 2.

Table 2. Wolf's Marginal Frequencies

i	Red Die f_i	Red Die Δ_i	White Die f_i	White Die Δ_i
1	0.17035	+0.00368	0.16230	-0.00437
2	0.18155	+0.01488	0.17245	+0.00578
3	0.15880	-0.00787	0.14485	-0.02182
4	0.14580	-0.02087	0.14205	-0.02464
5	0.17240	+0.00573	0.18175	+0.01508
6	0.17110	+0.00443	0.19960	+0.02993

On the null hypothesis that the dice were true, the standard deviations of $\{f_i\}$ from $p = 1/6$ should be $\sigma = [p(1-p)/N]^{1/2} = 0.0026$. The observed deviations Δ_i are many times this amount.

Now let us judge the deviation by the entropy criterion, considering only the white die. The entropy of the observed distribution lies below the maximum, log 6, by

$$\begin{aligned} \log 6 &= 1.791\ 759 \\ H_{Wolf} &= 1.784\ 990 \\ \hline \Delta H &= 0.006\ 769 \end{aligned}$$

which looks rather small; but this is for N = 20,000 trials. As a "quick and dirty" estimate based on (A3) we find $\exp(N\Delta H) \approx 6 \times 10^{58}$, indicating a very large strong constraint (i.e., systematic influence) keeping the frequencies away from the uniform distribution that could happen in the greatest number of ways if the die were equally free to settle in all positions.

The more precise concentration theorem gives

$$2N\,\Delta H = 270.1 = \chi^2_5(1 - F) \tag{19}$$

and therefore, from (A8),

$$1 - F = 1.07 \times 10^{-56} \quad . \tag{20}$$

Only one in 10^{56} of the 6^N conceivable outcomes has an entropy as low as Wolf's data give.

In Jaynes (1978) we considered what specific imperfections one might expect to find in a die, that might tend to make the frequencies nonuniform. The two most obvious are (1) a shift of the center of gravity due to the mass of ivory excavated from the spots, which being proportional to the number of spots on any side, should make the quantity $(f_1(i) \equiv i - 3.5,\ 1 \leq i \leq 6)$ have a nonzero expectation; and (2) errors in trying to machine a perfect cube, which will tend to make one dimension (the last side cut) slightly different from the other two. It is clear from the data that Wolf's white die gave a lower frequency for the faces (3,4); and therefore that the (3-4) dimension was undoubtedly greater than the (1-6) or (2-5) ones. The effect of this is that the function

$$f_2(i) = \begin{Bmatrix} +1, & i = 1,2,5,6 \\ -2, & i = 3,4 \end{Bmatrix} \tag{21}$$

has a non-zero expectation. The strength of these two systematic influences is indicated by Wolf's measured averages for them:

$$\overline{f}_1 = 0.0983 , \qquad \overline{f}_2 = 0.1393 \qquad (22)$$

Now if these are the only two imperfections present, we expect that the die will be equally free to yield any outcome compatible with the constraints (22). Therefore the observed frequencies should be the ones that can be realized in the greatest number of ways while agreeing with (22); i.e., which has maximum entropy subject to these two constraints. On the other hand, if the entropy of the observed distribution is appreciably below the maximum allowed by (22), that would be evidence that there is still another imperfection present; i.e., a third systematic influence not yet taken into account.

The maximum entropy H_{max} allowed by (22) was calculated in Jaynes (1978) by the algorithm (4)-(7), with the result indicated below:

$$\begin{aligned} H_{max} &= 1.785\ 225 \\ H_{Wolf} &= 1.784\ 990 \\ \hline \Delta H &= 0.000\ 235 \end{aligned}$$

The discrepancy is reduced by nearly a factor of thirty. The concentration theorem now gives

$$2N\ \Delta H = 9.38 = \chi^2_3(1 - F) \qquad (23)$$

or

$$1 - F = 0.025 \qquad (24)$$

The result appears just barely significant. That is, 97.5 percent of all outcomes compatible with (22) have an entropy greater than observed by Wolf. To assume a further very tiny imperfection [the (2-3-6) corner chipped off] we could make even this discrepancy disappear; but in view of the great number of trials one will probably not consider the result (24) as sufficiently strong evidence for this.

5. CONCLUSION

In Jaynes (1978) we gave a much more lengthy analysis, using the conventional Chi-squared test but arriving at less detailed and less accurate conclusions. At that time, in ignorance of the concentration theorem, it was not realized that there is no need to carry out the laborious computation of Chi-squared from the observed deviations Δ_i; the discrepancy between the observed entropy and that allowed by the hypothesis is already a more precise measure of significance.

We now see that the single maximum entropy formalism defined by (1) - (7) provides not only the procedure for predicting frequencies when incomplete data are available, that is optimal by a certain well-defined criterion; but also the criterion for testing hypotheses about systematic influences when frequency data are at hand.

APPENDIX

In N trials of the aforementioned random experiment, the i'th result occurs $N_i = N f_i$ times, $1 \le i \le n$. Out of the n^N conceivable outcomes, the number which yield a particular set of frequencies $\{f_i\}$ is

$$W(f_1 \ldots f_n) \equiv \frac{N!}{(Nf_1!) \ldots (Nf_n)!} \tag{A1}$$

and as $N \to \infty$ we have by the Stirling approximation

$$N^{-1} \log W \longrightarrow H(f_1 \ldots f_n) \quad , \tag{A2}$$

the entropy function (1). Given two sets of frequencies $\{f_i\}$ and $\{f_i'\}$, the ratio (number of ways f_i can be realized)/(number of ways f_i' can be realized) is asymptotically

$$\frac{W}{W'} \sim A\, e^{N(H-H')}\left[1 + \frac{B}{12N} + O(N^{-2})\right] \tag{A3}$$

where

$$A \equiv \prod_i (f_i'/f_i)^{\frac{1}{2}}$$

$$B \equiv \sum_i (f_i - f_i')/f_i f_i' \tag{A4}$$

represent corrections from the higher terms in the Stirling approximation. Their variation with $\{f_i\}$ is, of course, completely overwhelmed by that of the factor $\exp N(H-H')$.

The conceivable frequencies $\{f_1 \ldots f_n\}$ may be regarded as cartesian coordinates of a point P in an n-dimensional space, restricted to $\{S: 0 \le f_i, \Sigma f_i = 1\}$, an (n-1)-dimensional convex set whose vertices are the n points

$\{f_i = 1, 1 \le i \le n\}$. On S, the entropy (1) varies continuously, taking on all values in $(0 \le H(P) \le \log n)$ as P moves from a vertex to the center.

But now we obtain information that imposes the m linearly independent constraints (2), which define an (n-m)-dimensional hyperplane M. P is now confined to the intersection $S' = M \cap S$, a closed set comprising a bounded portion of hyperplane of dimensionality $k = n - m - 1$.

On S' the entropy attains a maximum $H_{max} \le \log n$. That this is attained at a unique point of S' may be proved analytically, but is perhaps made obvious as follows. Since any "mixing" increases the entropy, the set $(S_x: P \in S, H(P) \ge x)$ is strictly convex. Entropy maximization with constraints linear in $\{f_i\}$ thus amounts to finding the value of $x = H_{max}$ for which S' is a supporting tangent plane to S_x.

After these preliminaries, our argument follows slavishly the original derivation by Karl Pearson, as recalled by Lancaster (1969). In S' we may define new coordinates $\{x_1 \ldots x_k\}$ as appropriate linear functions of $\{f_1 \ldots f_n\}$ such that the new origin is at the maximum-entropy point, and there is a distance $r = (\Sigma x_i^2)^{1/2}$ such that near the origin a power series expansion yields

$$H(P) = H_{max} - a\,r^2 + \ldots \quad , \qquad a > 0 \quad . \tag{A4}$$

We then have a volume element in S' proportional to $r^{k-1}\,dr$. The domain of all possible frequency distributions $\{f_1 \ldots f_n\}$ which satisfy the constraints and whose entropy is in the range (3) is a k-sphere of radius R, given by $aR^2 = \Delta H$.

In N trials this sphere contains a fraction F of all possible outcomes in class C. From (A2), (A4) this is given asymptotically by

$$F \sim I(R)/I(\infty) \tag{A5}$$

where

$$I(R) \equiv \int_0^R e^{-Nar^2} r^{k-1} dr \quad . \tag{A6}$$

But, setting $NaR^2 = N\Delta H = (1/2)\chi^2$, this is just the cumulative Chi-squared distribution with k degrees of freedom; in conventional notation the relation between ΔH and F is given by Eq. (4).

In our applications we are generally concerned with numerical values for large $N\ \Delta H$, beyond the range of tables. The Chi-squared distribution $F(N\ \Delta H)$ may be expressed analytically as

$$F(x) = \frac{1}{s!} \int_0^x t^s e^{-t} dt \tag{A7}$$

where $s = (k/2) - 1$. For large $x = N\ \Delta H$, this yields the asymptotic expansion

$$1 - F(x) \sim (s!)^{-1} x^s e^{-x} [1 + sx^{-1} + s(s-1)x^{-2} + \ldots] \tag{A8}$$

When s is an integer (k even) this terminates and gives the exact result. Most of the numerical results cited in the text have been obtained from (A8).

REFERENCES

E. Czuber (1908), Wahrscheinlichkeitsrechnung, Teubner, Berlin; Vol. I, pp. 149-151.

E. T. Jaynes (1978), in The Maximum Entropy Formalism, R. D. Levine and M. Tribus, Editors, M.I.T. Press, Cambridge, Mass.

H. O. Lancaster (1969), The Chi-squared Distribution, J. Wiley & Sons, Inc., N. Y. pp. 7-8.

J. S. Rowlinson (1970), "Probability, Information, and Entropy," Nature, 225, 1196.

E. Schrödinger (1948), Statistical Thermodynamics, Cambridge University Press.

J. van Campenhout & T. M. Cover, "Maximum Entropy and Conditional Probability," I.E.E.E. Information Theory Trans. (in press).