The Annals of Statistics
1976, Vol. 4, No. 6, 1159–1189

ON THE APPLICATION OF SYMMETRIC DIRICHLET DISTRIBUTIONS AND THEIR MIXTURES TO CONTINGENCY TABLES[1]

BY I. J. GOOD

Virginia Polytechnic Institute and State University
Blacksburg

Bayes factors against various hypotheses of independence are proposed for contingency tables and for multidimensional contingency tables. The priors assumed for the nonnull hypothesis are linear combinations of symmetric Dirichlet distributions as in some work of 1965 and later. The results can be used also for probability estimation. The evidence concerning independence, provided by the marginal totals alone, is evaluated, and preliminary numerical calculations suggest it is small. The possibility of applying the Bayes/non-Bayes synthesis is proposed because it was found useful for an analogous problem for multinomial distributions. As a spin-off, approximate formulae are suggested for enumerating "arrays" in two and more dimensions.

TABLE OF CONTENTS

Received November 1974; revised March 1976.
[1] This work was supported in part by H.E.W., N.I.H. Grant # R01-GM18870.
AMS 1970 *subject classifications.* Primary 62F15; Secondary 62G10; Tertiary 05A15.
Key words and phrases. Contingency tables, multidimensional contingency tables, combinations of Dirichlet distributions, Bayes factor, Bayes/non-Bayes synthesis, enumeration of arrays, independence, multidimensional asymptotic expansions.

1. Introduction. Consider an $r \times s$ contingency table (n_{ij}) $(i = 1, 2, \cdots, r;$ $j = 1, 2, \cdots, s)$ where n_{ij} denotes the frequency in cell (i, j) and let p_{ij} denote the corresponding physical probability in that cell. We write $n_{i\cdot} = \sum_j n_{ij}$ and $n_{\cdot j} = \sum_i n_{ij}$ for the marginal totals, and $p_{i\cdot}$ and $p_{\cdot j}$ for the corresponding marginal probabilities. The hypothesis

$$H: p_{ij} = p_{i\cdot}p_{\cdot j} \qquad i = 1, 2, \cdots, r; j = 1, 2, \cdots, s$$

is called the null hypothesis of independence of rows and columns, and its negation, the nonnull hypothesis, is denoted by $\bar{H}$.

This paper is concerned primarily with a Bayesian approach to testing the hypothesis H, with extensions to multidimensional tables. Non-Bayesian methods also receive some attention. Apart from questions of the foundations of statistics, Bayesian methods for testing null hypotheses concerning multinomial distributions and contingency tables have the practical merit that they are apt to be applicable even when many cell expectations are small. Cell expectations are often small in large two-dimensional tables and in multidimensional tables. The theory also estimates how much evidence concerning H is contained in the (one-dimensional) marginal totals alone.

The hypotheses H and $\bar{H}$ will be fully specified, so that if either of them is accepted they can be used for estimating the cell probabilities, p_{ij}. Moreover this is especially convenient for the hypotheses entertained here (see Good (1967), page 407). Thus, although the main thrust of the work concerns significance tests, it also sheds light on probability estimation. Nevertheless, if H is rejected, the statistician would be well advised to look for deeper structure in the contingency table (for example, Woolf (1955); Good (1956), (1963), (1965a), page 61; Birch (1963); Bishop, Fienberg and Holland (1975)).

To test independence by Bayesian methods, it is necessary to assume prior densities for the physical probabilities (p_{ij}), both on the null hypothesis H of independence and on the rival (much more composite) hypothesis $\bar{H}$. These priors are made to depend on mixtures of symmetric Dirichlet distributions because such priors were previously found to be logical and useful for the analogous problem of testing equiprobability of multinomials (Good (1965), (1967), (1974); Good and Crook (1974)). Once the priors are chosen it is possible to compute $P((n_{ij}) \mid \bar{H})$ and $P((n_{ij}) \mid H)$ and the ratio of these probabilities is the *Bayes factor against H*, that is, the factor by which the initial *odds* of $\bar{H}$ are multiplied in the light of the evidence. This factor is not a Neyman–Pearson likelihood ratio when either H or $\bar{H}$ is composite, but it *can* be regarded as a simple likelihood ratio under a suitable physical model. The reason for this is that any prior can represent a physical distribution, in a superpopulation, given an appropriate non-Bayesian model. That is, composite hypotheses with associated priors can be *converted* into simple hypotheses at the level of superpopulations. In fact all Bayesian models can be given a non-Bayesian interpretation in this manner.

The three models for sampling. Regardless of the Bayesian approach, there are three familiar methods for sampling a contingency table, which we call Models 1, 2 and 3. In Model 1, the total N is alone fixed, in Model 2 the row (or column) totals are fixed, and in Model 3, both the row *and* column totals are fixed. We therefore discuss three Bayesian factors F_1, F_2 and F_3, corresponding to these three cases. Of these, the most difficult one to evaluate numerically is F_3 and much of our discussion is concerned with this problem.

Although Model 3 is less common than Models 1 and 2, it does occur in a variety of circumstances, usually for 2×2 tables; for example, in psychophysical experiments such as Fisher's famous tea-tasting experiment (Fisher (1949), Chapter 2), or in matching experiments for detecting extra-sensory perception (for example, Soal and Bateman (1954), pages 40, 41, 50), and in the experiment relating to weevil larvae and species of bean described by Sokal and Rohlf (1969), page 588. These examples can be easily generalized, at least conceptually, to $r \times s$ tables with $rs > 4$.

It is often judged by statisticians that the marginal totals by themselves convey little evidence about whether H is true. For example, it seems that Fisher never accepted the need to distinguish between models 1, 2 and 3 (see Fisher (1956), pages 87–88). Again Sokal and Rohlf ((1969), page 589), judge that the X^2 test, which is based on Model 3, and the likelihood-ratio test, which is based on Model 1, "provide rather similar results even when applied to the inappropriate model," so we should not expect F_1, F_2 and F_3 to be very unequal in ratio for most ordinary sets of marginal totals. We investigate this question by asking whether the Bayes factor against H, arising from the marginal totals alone, is close to unity. This factor is equal to F_1/F_3.

Tocher (1950) showed, with a slight gloss, that Fisher's exact test for the 2×2 table is uniformly most powerful and unbiased even for Models 2 and 3, and S. N. Roy and Mitra (1956) showed that X^2 for the general $r \times s$ table is asymptotically distributed as χ^2 with $(r-1)(s-1)$ degrees of freedom for all three models. But the power functions are not the same for Models 1, 2 and 3 (see, for example, Kendall and Stuart (1973), Section 33.25, for references) so one should not expect the Bayes factors F_1, F_2 and F_3 to be precisely equal. Our assumptions imply that they are in fact unequal but that they do not seem to differ very greatly in normal circumstances.

The Bayes/non-Bayes synthesis. In accordance with a synthesis or compromise between Bayesian and non-Bayesian methods, any Bayes factor can be reinterpreted as a "Fisherian" (non-Bayesian) statistic, and its distribution examined under the null hypothesis when the null hypothesis is simple. Under Model 3, H has been known for forty years to be a simple statistical hypothesis, as we shall explain, so this "synthesis" can be applied to F_3. The main numerical investigation of this matter, together with many other numerical details, is delayed to Part II of the present work, which will be jointly authored with Dr. J. F. Crook, but some preliminary results will be reported herein, especially in Section 13. For the multinomial distribution the corresponding results were reported by Good and Crook (1974).

The enumeration of arrays and allied problems. The problems in this work lead to a generalization of the classical combinatorial problem of enumerating rectangular arrays of nonnegative integers, having given marginal totals, a problem that is closely connected with the theory of symmetric polynomials. Some of the methods for attacking this generalization of the array problem also shed some light on the ordinary array problem. Some new exact and asymptotic results are given in the appendices, but the solution is incomplete. An approximate formula for the number of arrays is related to a conjecture that approximates F_3 in terms of the other Bayes factors.

We shall discuss two-dimensional and three-dimensional tables in turn, and the extension to higher dimensions will then be obvious.

2. Dirichlet distributions andtheir mixtures. All the work in this paper is based on Dirichlet distributions and their mixtures, so it will be convenient to introduce some relevant notation and terminology.

A Dirichlet density with say t categories is of the form

$$\frac{\Gamma(\sum_\nu k_\nu)}{\prod_\nu \Gamma(k_\nu)} \prod_{\nu=1}^t q_\nu^{k_\nu - 1} \qquad (k_\nu > 0;\ \nu = 1, 2, \cdots, t;\ \sum q_\nu = 1)\,, \tag{2.1}$$

where the q_ν's are physical probabilities. The k_ν's may be called hyperparameters (that is, parameters in a prior distribution). They may also be called flattening constants in the sense that the posterior expectation of q_ν given $(n_\nu) = (m_\nu)$ is

$$(m_\nu + k_\nu)/\sum_\nu (m_\nu + k_\nu)\,, \tag{2.2}$$

and conversely, if (2.2) is the Bayesian expectation whatever the m_ν's may be, then the only possible prior density is (2.1); see, for example, Good (1965a), page 23. Hence, if cells are lumped together in any way, the prior for the reduced collection of new cells is still Dirichletian with the corresponding addition of the flattening constants. Another proof of this result is given by Wilks (1962), page 181.

We denote the above-mentioned Dirichlet distribution by $D(t; k_1, k_2, \cdots, k_t)$ and, if $k_1 = k_2 = \cdots = k_t = k$, by $D(t, k)$ which we call the *symmetric Dirichlet distribution* with flattening constant (or hyperparameter) k.

In the earlier multinomial researches we used a mixture of $D(t, k)$'s

$$\int_0^\infty D(t, k)\phi(k)\, dk \,, \tag{2.3}$$

where ϕ is the specific log-Cauchy density

$$\phi(k) = \frac{1}{k[\pi^2 + (\log_e k)^2]} \,. \tag{2.4}$$

Previous to 1965, various writers had used fixed values of k, especially $k = 1$ and $k = \frac{1}{2}$, but I believe it is necessary to assume a distribution for k. The history and related matters are discussed by Good (1965a), (1967).

We shall later need to generalize (2.3) to

$$\int_0^\infty D(t, t'k)\phi(k)\, dk \,,$$

and we shall denote this mixture of symmetric Dirichlet priors by $D^*(t, t')$, the arguments q_ν being again taken for granted. The prior density (2.3) is $D^*(t, 1)$.

Mixtures of the more general Dirichlet density $D(t; k_1, \cdots, k_t)$ could be used if any specific real-world problem seemed to require it, but the more special density $D^*(t, 1)$ appeared to be adequate in the multinomial work for various values of t from 2 to 100. There is no point in choosing a more complicated prior if the simpler one is adequate. A more complicated model would be required if some of the q_ν's had larger (subjective) prior variances than others.

The log-Cauchy "hyperprior" (2.4) was chosen because it is noncommittal or "uniformative" in the sense, for example, of giving low prior density only to very large values of k and because it approximates the improper Jeffreys–Haldane density $1/k$ which was proposed for "representing ignorance" of the value of a positive variable. This Jeffreys–Haldane density has an interesting interpretation for our present problem, which will be mentioned soon. But an improper hyperprior cannot be strictly used in this work because it would lead to the total annihilation of the evidence as indicated by Good (1965a), page 38. The choice of the parameters (or hyper-hyperparameters) of the log-Cauchy distribution was made by Good (1967) by considering six different sets of these parameters and finding (2.4) to have a slight advantage over the others in the sense of being more consistent with non-Bayesian procedures. Moreover a graph of (2.4) looks like a rectangular hyperbola, whereas with some other values of the parameters, the graph has a kink in it.

In the present work we shall assume the density $D^*(t, 1)$ as a prior, with t taking the values r, s and rs, and with q_ν the values $p_{i\cdot}$, $p_{\cdot j}$ and p_{ij}.

We shall require the following formula (for example, Good (1965a), page 36)

$$P((m_\nu)\,|\,D(t, k)) = \frac{\Gamma(tk)N!\prod \Gamma(m_\nu + k)}{\Gamma(k)^t\Gamma(N + tk)\prod m_\nu!} \quad (\textstyle\sum m_\nu = N) \tag{2.5}$$

where $\Gamma(k)^t$ means $[\Gamma(k)]^t$. (This formula reduces to 1 when $N = 0$ and to $1/t$ when $N = 1$ as it should.) Consequently

$$P((m_\nu)\,|\,D^*(t, t')) = \Phi((m_\nu), t, t') \,, \tag{2.6}$$

where

$$\Phi((m_\nu), t, t') = \frac{N!}{\prod m_\nu!} \int_0^\infty \frac{\Gamma(tt'k)\prod\Gamma(m_\nu + t'k)}{\Gamma(t'k)^t\Gamma(N + tt'k)}\,\phi(k)\,dk$$
$$= \frac{N!}{\prod m_\nu!} \int_0^\infty \frac{\Gamma(tk)\prod\Gamma(m_\nu + k)}{\Gamma(k)^t\Gamma(N + tk)}\,\phi\left(\frac{k}{t'}\right)\frac{dk}{t'}\,. \tag{2.7}$$

Since $\phi(k)$ is roughly proportional to $1/k$, when k is not very large, we can expect that $\Phi((m_\nu), t, t')$ will not be very different from $\Phi((m_\nu), t, 1)$. Although this is intuitively to be expected from the form of the integral, we cannot replace $\phi(k)$ by $1/k$ (nor any other improper density) for it would make the integral in (2.7) diverge.

There is a much more general principle involved in the above argument than we have stated. Suppose for the moment that k is a positive hyperparameter with *any* hyperprior. Then if k is replaced by $t'k$ and the result is multiplied by $1/k$ and integrated the result is formally independent of t'.

3. The choice of priors. When the row totals $(n_{i\cdot})$ are not fixed in advance, then given H we shall need to assume a prior for $(p_{i\cdot})$. This prior should not involve s because the row categories could usually be combined with various numbers of column categories in different experiments. We shall adopt $D^*(r, 1)$ and $D^*(s, 1)$ for the prior densities of $(p_{i\cdot})$ and $(p_{\cdot j})$, given H, when the corresponding marginal totals are not fixed in advance. For Model 1 we need both these priors, and we assume them to be independent. Similarly, when $\bar{H}$ is given, we assume $D^*(rs, 1)$ for the prior density of (p_{ij}). In other words, wherever we need to assume a prior we assume the multinomial prior that was found useful by Good (1967) and by Good and Crook (1974).

In virtue of the "lumping" property of Dirichlet distributions, which was mentioned in Section 2, the priors for $(p_{i\cdot})$ and $(p_{\cdot j})$, given $\bar{H}$, do not need to be assumed for they must be $D^*(r, s)$ and $D^*(s, r)$ respectively.

4. The probabilities of the marginal totals and of the interior of the table. It was pointed out independently by Fisher (1934) and Yates (1934) that

$$P((n_{ij}) \mid (n_{i\cdot}), (n_{\cdot j}), H) = \frac{\prod n_{i\cdot}!\prod n_{\cdot j}!}{N!\prod n_{ij}!}\,, \tag{4.1}$$

a formula that does not depend on any priors because $(p_{i\cdot})$ and $(p_{\cdot j})$ have conveniently been eliminated. We denote this Fisher–Yates probability by F.Y. It shows, that, under Model 3, H is a simple statistical hypothesis.

For calculating the Bayes factors F_1, F_2 and F_3, we need some formulae for further probabilities resembling the left side of (4.1).

In virtue of the priors assumed in Section 3, we have

(4.2) $P((n_{i\cdot}) \mid H) = \Phi((n_{i\cdot}), r, 1)$ when row totals are not fixed;

(4.3) $P((n_{\cdot j}) \mid H) = \Phi((n_{\cdot j}), s, 1)$ when column totals are not fixed;

and

$$P((n_{ij}) \mid \bar{H}) = \Phi((n_{ij}), rs, 1) . \tag{4.4}$$

From (4.1), (4.2) and (4.3) we have

$$P((n_{ij}) \mid H) = \Phi((n_{i\cdot}), r, 1)\Phi((n_{\cdot j}), s, 1)\ \text{F.Y.} \tag{4.5}$$

Again, if the column totals are not fixed, we have

$$\begin{aligned} P((n_{ij}) \mid (n_{i\cdot}), H) &= P((n_{ij}), (n_{\cdot j}) \mid (n_{i\cdot}), H) \quad \text{(trivially)} \\ &= P((n_{ij}) \mid (n_{i\cdot}), (n_{\cdot j}), H)P((n_{\cdot j}) \mid H) \\ &= \Phi((n_{\cdot j}), s, 1)\ \text{F.Y.} , \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} P((n_{ij}) \mid (n_{i\cdot}), \bar{H}) &= P((n_{ij}) \mid \bar{H})/P((n_{i\cdot}) \mid \bar{H}) \\ &= \Phi((n_{ij}), rs, 1)/\Phi((n_{i\cdot}), r, s) . \end{aligned} \tag{4.7}$$

Further, since (n_{ij}) implies $(n_{i\cdot})$ and $(n_{\cdot j})$, we have

$$\begin{aligned} P((n_{ij}) \mid (n_{i\cdot}), (n_{\cdot j}), \bar{H}) &= P((n_{ij}) \mid \bar{H})/P((n_{i\cdot}), (n_{\cdot j}) \mid \bar{H}) \\ &= \Phi((n_{ij}), rs, 1)/P((n_{i\cdot}), (n_{\cdot j}) \mid \bar{H}) \end{aligned} \tag{4.8}$$

where, finally,

$$\begin{aligned} P((n_{i\cdot}), (n_{\cdot j}) \mid \bar{H}) &= \textstyle\sum^* P((m_{ij}) \mid \bar{H}) \\ &= \textstyle\sum^* \Phi((m_{ij}), rs, 1) , \end{aligned} \tag{4.9}$$

and $\sum^*$ denotes a summation over all tables (m_{ij}) having the "right" marginal totals, that is, for which $m_{i\cdot} = n_{i\cdot}$ and $m_{\cdot j} = n_{\cdot j}$ (in a self-explanatory notation).

5. The Bayes factors F_1, F_2 and F_3. Using the formulae of Section 4 we see that

$$\begin{aligned} F_1 &= P((n_{ij}) \mid \bar{H})/P((n_{ij}) \mid H) \\ &= \Phi((n_{ij}), rs, 1)/\{\Phi((n_{i\cdot}), r, 1)\Phi((n_{\cdot j}), s, 1)\ \text{F.Y.}\} ; \end{aligned} \tag{5.1}$$

$$\begin{aligned} F_2 &= P((n_{ij}) \mid (n_{i\cdot}), \bar{H})/P((n_{ij}) \mid (n_{i\cdot}), H) \\ &= \Phi((n_{ij}), rs, 1)/\{\Phi((n_i), r, s)\Phi((n_{\cdot j}), s, 1)\ \text{F.Y.}\} ; \end{aligned} \tag{5.2}$$

$$F_{(2)} = \Phi((n_{ij}), rs, 1)/\{\Phi((n_{i\cdot}), r, 1)\Phi((n_{\cdot j}), s, r)\ \text{F.Y.}\} , \tag{5.3}$$

by symmetry, where $F_{(2)}$ denotes the Bayes factor against H when the *column* totals are fixed. In view of the penultimate paragraph in Section 2, we can expect F_1, F_2 and $F_{(2)}$ to be not very different.

Finally,

$$\begin{aligned} F_3 &= P((n_{ij}) \mid (n_{i\cdot}), (n_{\cdot j}), \bar{H})/P((n_{ij}) \mid (n_{i\cdot}), (n_{\cdot j}), H) \\ &= \Phi((n_{ij}), rs, 1)/\{\text{F.Y.} \textstyle\sum^* \Phi((m_{ij}), rs, 1)\} . \end{aligned} \tag{5.4}$$

As pointed out in Good ((1965a), page 52), $\sum^*$ can here be written

$$N! \int_0^\infty \frac{\Gamma(rsk)}{\Gamma(N+rsk)} A(k)\phi(k)\,dk \tag{5.5}$$

where

$$A(k) = A(k, (n_{i\cdot}), (n_{\cdot j})) = \mathscr{C}(\textstyle\prod_{ij} x_i^{n_{i\cdot}} y_j^{n_{\cdot j}}) \prod_{ij} (1 - x_i y_j)^{-k} \tag{5.6}$$

where $\mathscr{C}(\cdots)$ denotes "the coefficient of $\cdots$ in." (It is convenient to define $A(k)$ as 1 when $N = 0$.) *Methods for calculating or approximating $A(k)$ are discussed in Appendix* B.

The integrals that occur in the formulae for F_1, F_2 and F_3 are all of the form that was used in Good (1967) and in Good and Crook (1974) and can therefore be handled by similar computer programs. But the summation that occurs in the denominator of F_3 is troublesome if handled directly, unless N is small, because the number of terms in it is equal to the number $A((n_{i\cdot}), (n_{\cdot j}))$ of "arrays" having the assigned marginal totals. This number is in fact $A(1, (n_{i\cdot}), (n_{\cdot j}))$, and it is useful to know this number when programming the calculation of F_3. The enumeration of arrays is a classical combinatorial problem and some of the literature is referenced in Appendix A.

6. The Bayes postulate. It is interesting to consider what happens to the Bayes factors F_1, F_2 and F_3 if $\phi(k)$ is the Dirac delta function $\delta(k - 1)$. In this case the prior for (p_{ij}), under $\bar{H}$, is given by the uniform multidimensional "Bayes postulate." We denote the factors in this case by F_1', F_2' and F_3'. We have

$$F_1' = \frac{\Gamma(rs)\Gamma(N+r)\Gamma(N+s) \prod n_{ij}!}{\Gamma(N+rs)\Gamma(r)\Gamma(s) \prod n_{i\cdot}! \prod n_{\cdot j}!}, \tag{6.1}$$

$$F_2' = \frac{\Gamma(s)^{r-1}\Gamma(N+s) \prod n_{ij}!}{\prod \Gamma(n_{i\cdot} + s) \prod n_{\cdot j}!}, \tag{6.2}$$

$$F_3' = \frac{N! \prod n_{ij}!}{\prod n_{i\cdot}! \prod n_{\cdot j}!} \frac{1}{A((n_{i\cdot}), (n_{\cdot j}))}. \tag{6.3}$$

The Bayes factors F_1' and F_2' agree with the values given by Good ((1950), pages 99 and 100) for Models 1 and 2 so we have obtained a verification of formulae (5.1) and (5.2). Moreover, F_3', the factor against the null hypothesis under Model 3, assuming the Bayes postulate, can be obtained directly. We need only invoke the Fisher-Yates formula (4.1), combined with the remark that, given $\bar{H}$ combined with the Bayes postulate, all contingency tables with assigned marginal totals have equal probability (which is therefore the reciprocal of the number of arrays). We have here made use of (2.5) with $k = 1$.

A conjecture was made by Good ((1950), page 100) that for Model 3 the factor against the null hypothesis "may reasonably be taken" as $F_2'F_{(2)}'/F_1'$, where $F_{(2)}'$ is F_2' with "rows and columns interchanged." Doubt was expressed in Good ((1965a), page 51), but we are going to see by calculations that the conjecture appears to be confirmed. The conjecture implies the following approximation to the number of arrays:

(6.4) $$A((n_{i\cdot}), (n_{\cdot j})) \approx B((n_{i\cdot}), (n_{\cdot j}))$$

where

(6.5) $$B((n_{i\cdot}), (n_{\cdot j})) = \frac{\prod \binom{n_{i\cdot} + s - 1}{n_{i\cdot}} \prod \binom{n_{\cdot j} + r - 1}{n_{\cdot j}}}{\binom{N + rs - 1}{N}}$$

which is equal to the product of the numbers of ways of distributing the row totals in the rows and the column totals in the columns, divided by the number of arrays of total N. (For an intuitive interpretation of (6.5) see Section A2; and for another asymptotic formula for the number of arrays, for further justification, and for a generalization, see Appendix B.)

The ratios of B/A for various marginal totals are shown in Table 1. We have taken row totals equal in each example because a simple general formula for $A((n_{i\cdot}), (n_{\cdot j}))$, is not known even for $r = s = 3$. We also assume that no row or column total vanishes since this would effectively reduce the size of the array. The ratios A/B are fairly close to 1 except when the $n_{\cdot j}$'s are very rough, and even for the table (10, 10, 10; 1, 1, 28) we have $B/A < 2.6$. The approximation can be improved still further by introducing a correction for "roughness," namely by writing

(6.6) $$A \approx C, \qquad \text{where} \quad C = 1.3N^2B/(r \textstyle\sum n_{\cdot j}^2),$$

TABLE 1

The numbers of arrays for various marginal totals, the approximations B and C, and the ratios A/B and A/C

Row totals	Column totals	A	B	A/B	C	A/C
1, 1, 1	1, 1, 1	6	4.42	1.35	5.7	1.05
2, 2, 2	2, 2, 2	21	15.54	1.35	20	1.05
2, 2, 2	1, 1, 4	9	9.71	0.93	8.4	1.07
3, 3, 3	3, 3, 3	55	41.14	1.33	53	1.04
4, 4, 4	4, 4, 4	120	90.4	1.33	118	1.02
4, 4, 4	2, 2, 8	36	43.4	0.83	38	0.95
4, 4, 4	1, 1, 10	9	15.91	0.56	9.7	0.93
6, 6, 6	6, 6, 6	406	308.5	1.32	401	1.01
10, 10, 10	10, 10, 10	2211	1690.1	1.32	2197	1.01
10, 10, 10	2, 6, 22	168	272.6	0.62	203	0.83
10, 10, 10	1, 10, 19	195	244.4	0.80	206	0.95
10, 10, 10	5, 5, 20	441	598.9	0.74	519	0.85
10, 10, 10	2, 2, 26	36	80.0	0.45	46	0.78
10, 10, 10	1, 1, 28	9	23.0	0.39	11.4	0.79
20, 20, 20	20, 20, 20	26796	20555	1.30	26722	1.00
20, 20, 20	10, 10, 40	4356	6254	0.69	5420	0.80
4, 4, 4, 4	4, 4, 4, 4	10147	7493	1.35	9741	1.04
6, 6, 6, 6, 6	6, 6, 6, 6, 6	1.642(8)	1.189(8)	1.38	1.546(8)	1.06

in fact

$$(6.7) \qquad 0.75 < A/C < 1.1$$

in all the examples tried (with $n_{i.}$ mathematically independent of i). If the row and column totals are both "rough," then perhaps

$$(6.8) \qquad A \approx D \quad \text{where} \quad D = 1.3N^4B/(rs \sum n_{i.}^2 n_{.j}^2)$$

but this is only a guess. In any case the approximation $A \approx B$ is likely to be quite good except for very rough marginal totals. Thus the statistical conjecture of Good ((1950), page 100) leads to an approximate solution to the purely combinatorial problem of enumeration of arrays, and conversely the approximation $A \approx B$ tends to corroborate the conjecture. See also Tables 2 and 3.

TABLE 2

Comparison of $A(k)$, $A_0(k)$ and $B(k)$ ($k = 1, 2$) for some square tables having flat margins

k	r	N	$A(k)$	$A_0(k)$	$B(k)$	$A_0(k)/A(k)$	$B(k)/A(k)$	$A_0(k)/B(k)$	D/A
1	3	3	6	7.268	4.418	1.21	0.74	1.65	.957
1	3	21	666	605.4	507.2	0.91	0.76	1.19	.989
1	3	60	26796	24118	20556	0.90	0.77	1.17	1.00
1	3	300	13268976	11924803	10192241	0.90	0.77	1.17	1.00
1	4	4	24	32.71	16.91	1.36	0.71	1.93	0.92
1	4	16	10147	9540	7493	0.94	0.74	1.27	0.96
1	4	32	981541	882107	730929	0.90	0.75	1.21	0.97
1	5	5	120	187	82	1.56	0.68	1.21	0.89
2	3	3	48	65.476	40.926	1.36	0.85	1.60	1.11
2	3	21	—	9569627	8575176	—	—	1.12	—
2	3	300	—	1.4648(20)	1.3536(20)	—	—	1.08	—
2	4	8	53784	64193	45854	1.19	0.85	1.40	1.11
0.5	4	8	2.022	1.647	1.020	0.82	0.50	1.61	0.65

TABLE 3

Values of A, A_0/A, B/A and D/A for some tables with flat column totals

$(n_{i.})$	$(n_{.j})$	A	A_0/A	B/A	D/A
15, 15	10, 10, 10	91	0.68	0.84	1.09
5, 25	10, 10, 10	21	1.25	1.44	1.30
5, 5, 20	10, 10, 10	441	1.17	1.36	1.18
5, 5, 40	10, 10, 10, 10, 10	15876	1.42	3.55	2.33

In connection with these "corrections for roughness" it is interesting to recall the following judgment made by Yates (1934) in relation to the accuracy of the χ^2 approximation to the distribution of X^2 for contingency tables: "Cases where some of the marginal totals are large and others are small $\cdots$ may be expected to give much more unfavourable results."

7. A conjectured approximation for F_3. Since F_3 is much more troublesome to compute than F_1 and F_2 it would be useful if we could extend the conjectures of the previous section to the models of the present paper, for example, with the conjecture

$$F_3 \approx F_2 F_{(2)}/F_1 . \tag{7.1}$$

This conjecture will be tested numerically and the results reported in Part II. (See also Section 13.) It may be possible to improve the approximation by allowing for the roughness of the row and column totals as in (6.6) and (6.8).

8. The evidence from the row and column totals. Let us denote by FRACT the Bayes factor against H provided by the row and column totals alone. Then

$$\text{FRACT} = \frac{P((n_{i\cdot}), (n_{\cdot j}) \mid \bar{H})}{P((n_{i\cdot}), (n_{\cdot j}) \mid H)} = \frac{F_1}{F_3} . \tag{8.1}$$

In Part II we shall investigate the conjecture that FRACT is not very far from 1 except for very rough marginal totals. When combined with (7.1), this conjecture asserts that F_1 and F_3 are of the same order of magnitude as the geometric mean of F_2 and $F_{(2)}$; and it would not be surprising to find that F_1, F_2, $F_{(2)}$ and F_3 are usually all of much the same magnitude. This would be consistent with the judgment made in the introduction related to some opinions of Fisher and of Sokal and Rohlf.

9. An interpretation of F_1 in terms of multinomials. It is known (Good (1965a), (1967)), or easily deducible from (2.5), that the factor against the null hypothesis of equiprobability provided by a multinomial sample $(m_1, m_2, \cdots, m_t)$ is

$$t^N \int_0^\infty \frac{\Gamma(tk) \prod \Gamma(m_\nu + k)}{\Gamma(k)^t \Gamma(N + tk)} \phi(k)\, dk = \frac{t^N \prod m_\nu !}{N!} \Phi((m_\nu), t, 1) . \tag{9.1}$$

Let H_{00}, $H_{0\cdot}$ and $H_{\cdot 0}$ denote the null hypotheses of equiprobability for the whole table, and for the row and column totals, respectively. For example: H_{00} means that $p_{ij} = 1/rs$ $(i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s)$. Then from (5.1) and (9.1) we have

$$F_1 = \frac{F(H_{00} : (n_{ij}))}{F(H_{0\cdot} : (n_{i\cdot}))F(H_{\cdot 0} : (n_{\cdot j}))} , \tag{9.2}$$

where the colons are read "provided by." A direct proof of (9.2) can be given that does not depend on the Dirichlet priors, so that *it could be used under other assumptions*. For we have

$$P((n_{ij}) \mid H_{00}) = P((n_{i\cdot}) \mid H_{0\cdot})P((n_{\cdot j}) \mid H_{\cdot 0}) \text{ F.Y.}$$
$$P((n_{ij}) \mid H) = P((n_{i\cdot}) \mid \bar{H}_{0\cdot})P((n_{\cdot j}) \mid \bar{H}_{\cdot 0}) \text{ F.Y.}$$

Hence

$$\frac{P((n_{ij}) \mid \bar{H})}{P((n_{ij}) \mid H)} = \frac{P((n_{ij}) \mid \bar{H})}{P((n_{ij}) \mid H_{00})} \cdot \frac{P((n_{i\cdot}) \mid H_{0\cdot})}{P((n_{i\cdot}) \mid \bar{H}_{0\cdot})} \cdot \frac{P((n_{\cdot j}) \mid H_{\cdot 0})}{P((n_{\cdot j}) \mid \bar{H}_{\cdot 0})}$$

and this expresses equation (9.2).

10. The Bayes/non-Bayes compromise or synthesis. The Bayes/non-Bayes compromise for significance testing is a technique that is illustrated especially by Good and Crook (1974) for purely multinomial problems. The main idea is to treat a Bayes factor as a non-Bayesian criterion and to seek its distribution given the null hypothesis. We consider this approach first for Model 3 because, for this model, H is a simple statistical hypothesis. It might be difficult to find the asymptotic distribution of F_3 so we proceed by analogy with the multinomial case. We define the factor $F_3(k)$ by taking $\phi(k)$ as a Dirac delta function in the expression for F_3 so that

$$F_3(k) = \frac{N! \prod \{(\Gamma(n_{ij} + k)/\Gamma(k)\}}{A(k) \prod n_{i.}! \prod n_{.j}!} . \tag{10.1}$$

(Of course F_3 is *not* obtained from $F_3(k)$ by multiplying it by the actual $\phi(k)$ and integrating from 0 to ∞.) Then $\max_k F_3(k)$ is a Type II Likelihood Ratio in the terminology of Good (1965a) and Good and Crook (1974): perhaps it should be called a *hyperlikelihood ratio*. Hence the asymptotic distribution of

$$G = (2 \log_e \max_k F_3(k))^{\frac{1}{2}} \tag{10.2}$$

will be proportional to a standardized normal distribution as in these two references. The corresponding result for the multinomial case was surprisingly accurate into the extreme tail. (See also Section 13.) It may be conjectured that $F_3(k)$ is unimodal, both on the basis of numerical examples and by analogy with the multinomial problem [see Good (1974) and some unpublished work of Bruce Levin and James Reeds].

It is not clear whether a hyperlikelihood ratio can be conveniently defined for Models 1 and 2.

11. Measures of association. It may be conjectured that when any of the hypotheses of independence is false, then $(\log F)/N$ tends in probability to a limit, where F is the Bayes factor against the hypothesis. If this is so, the limit would be a natural measure of association which could be called "the weight of evidence per item against independence." It is intuitively obvious, when F is F_1, F_2 or F_3, that this limit is what was called W_1 by Good ((1969), page 572), in other words "the expected mutual information between rows and columns,"

$$W_1 = \sum p_{ij} \log \frac{p_{ij}}{p_{i.}p_{.j}} . \tag{11.1}$$

The corresponding three-dimensional measure of association is

$$\sum p_{hij} \log (p_{hij}/(p_{h..}p_{.i.}p_{..j})) . \tag{11.2}$$

12. Extensions to three or more dimensions. Multidimensional contingency tables are important in sociology and in medical diagnosis, up to dimensionalities as high as a hundred or so. But readers interested only in two-dimensional tables could jump to Section 13.

When using a Bayesian approach to significance testing, for multidimensional tables, a succession of decisions must be made: (i) whether to test only for independence (rather than for the vanishing of second or higher-order interactions); (ii) if so, which hypotheses of independence to test; (iii) which sampling model to be used; (iv) what priors to assume for the null and nonnull hypotheses. We have agreed to test only for independence in this paper. If a table can be broken up into (approximately) independent parts, its analysis is greatly simplified, so much so that tables are often broken up even if the parts are dependent. We must now consider decisions (ii), (iii) and (iv).

If the dimensionality of a table is m, then the number of ways of breaking the table down, by grouping the dimensions, is $b_m - 1$, where b_m is the number of ways of partitioning a set of m objects into one or more subsets. The numbers b_m ($m = 0, 1, 2, \cdots$) are known as exponential or Bell numbers; see, for example, Gould (1971). The sequence begins 1, 1, 2, 5, 15, 52, 203, 877, 4140, 21127, 115975, $\cdots$, and has been tabulated up to $m = 74$ by Levine and Dalton (1962). The number γ_m of hypotheses of independence is even greater if allowance is made also for conditional independence, such as $p_{hi\cdot}p_{h\cdot j} = p_{hij}p_{h\cdot\cdot}$ for all h, i and j (in a self-explanatory notation); in fact the sequence γ_m ($m = 0, 1, 2, \cdots$) begins 0, 0, 1, 10, 70, 431, 2534, 14820, 88267, 542912, 3475978, $\cdots$. The exponential generating function (Good (1975)) is

$$\sum_{m=0}^{\infty} \frac{\gamma_m z^m}{m!} = \exp(e^z + 2z - 1) - e^{3z} . \tag{12.1}$$

These large numbers should be taken into account when judging prior probabilities of the various hypotheses or when setting thresholds on significance criteria. For example, when $m = 5$, if the *only* one of the 431 null hypotheses that appeared to be rejectable (or acceptable) was say

$$p_{\cdot\cdot i_3 \cdot i_5} p_{\cdot i_2 \cdot i_4 i_5} = p_{\cdot i_2 i_3 i_4 i_5} p_{\cdot\cdot\cdot\cdot i_5} \quad (\text{for all } i_2, i_3, i_4 \text{ and } i_5) ,$$

then perhaps we should "pay a factor" of 431 for "special selection," either on the prior probability of dependence or on the tail-area probability. (A fairly early published example of allowing for special selection, and the way it leads to the need to consider very small tail-area probabilities, is given by Fisher (1949), page 58; and it would be surprising if this phenomenon did not date back at least to the nineteenth century.) Thus, we might insist on a tail-area probability as small as say 1/10,000. (For $m = 10$, the corresponding tail-area would be say 10^{-8}.) When there are many apparent cases of dependence some theory of "substantialism" must be invoked, as defined by Good (1965b), page 124. In accordance with that idea one might test the hypothesis of complete independence of all m dimensions by using as a criterion the harmonic mean of the tail-area probabilities corresponding to all the γ_m null hypotheses described above. In a Bayesian analysis one might average the γ_m Bayes factors but a weighted average might be judged more appropriate for a specific problem. This Bayes criterion can in principle again be reinterpreted as a non-Bayesian significance criterion: another example of the Bayes/non-Bayes synthesis.

The difficulty in determining a significance test, when there is a hypothesis corresponding to each subset of some set of arguments, is especially well known in connection with multiple regression: see, for example, Aitkin (1974). The difficulty is even more acute for our problem of multidimensional contingency tables, and no attempt will be made here to settle the matter.

The remainder of this section is concerned with the generalization of formulae from the case $m = 2$ to larger values of m. Let us take $m = 3$ at first, and consider a $q \times r \times s$ table, having q Queen's highways, r rows or roads, and s shafts, streets or columns.

Consider the hypothesis $H(1, 23)$ of the independence of the Queen's highways on the one hand and the roads and streets on the other. (A sufficiently self-explanatory notation for the hypotheses is here introduced.) By treating the roads and streets combined as a single "vector" we might hope to treat the problem as being only two-dimensional. This requires that the prior assumed for roads and streets combined is a weighted combination of symmetric Dirichlets, the weight function being ϕ. This assumption will be consistent with our procedure provided that the hypothesis $H(2, 3)$ that roads and streets are independent has already been decisively rejected. Thus the hypothesis $H(1, 23)$ can be tested by two applications of the two-dimensional theory already discussed.

Next consider the hypothesis $H(\cdot, 2, 3)$ that roads and streets are independent conditional on each Queen's highway h. This is mainly a matter of performing q separate "two-dimensional" tests. This may not always be accurate enough for a strictly Bayesian interpretation because the priors can change as successive values of h are considered, a phenomenon called "learning from experience" in Good (1965a), page 51. I believe this criticism is much less serious when the present priors are used (linear combinations of symmetric Dirichlet's) than when straight symmetric Dirichlet priors are used. Let us assume for the present that this specific example of "learning from experience" is unimportant, at least if each of the q Bayes factors is interpreted as a non-Bayesian criterion, in accordance with the Bayes/non-Bayes synthesis. We should still need to apply some theory of "simultaneous inference," but to consider this intricate matter in detail would take us too far afield.

Then of the ten null hypotheses, $H(1, 23)$, $H(2, 31)$, $H(3, 12)$, $H(\cdot, 2, 3)$, $H(\cdot, 3, 1)$, $H(\cdot, 1, 2)$, $H(2, 3)$, $H(3, 1)$, $H(1, 2)$ and $H(1, 2, 3)$, only the last one requires a genuine three-dimensional analysis.

The above argument for the case $m = 3$ extends to all higher dimensionalities, so that apparently, of the γ_m null hypotheses, only one really requires an m-dimensional analysis, namely the hypothesis that all m dimensions are statistically independent, the hypothesis $H(1, 2, 3, \cdots, m)$. But the analysis of this hypothesis itself splits into several "models," just as the consideration of the hypothesis $H(1, 2)$ or H splits into three models leading to the Bayes factors F_1, F_2 and F_3.

For m-dimensional tables with $m \geqq 3$ an example where all one-dimensional marginal totals are fixed in advance occurs in the "matching" of d decks of cards, where each deck has a known number of cards of each suit. For discussions of the matching problem see Wilks (1946), page 212, Barton (1958), and David and Barton (1962). But, as in the two-dimensional case, the main interest in the corresponding Bayes factor is in the light it can shed on whether the marginal totals contain much evidence about the hypothesis H of independence. This will have repercussions on the use of X^2, defined for $m = 3$ by

$$(12.2) \qquad X^2 = \sum_{h=1}^{q} \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{N^2}{n_{h\cdot\cdot}\, n_{\cdot i\cdot}\, n_{\cdot\cdot j}} \left(n_{hij} - \frac{n_{h\cdot\cdot}\, n_{\cdot i\cdot}\, n_{\cdot\cdot j}}{N^2}\right)^2$$

whose χ^2 theory is based primarily on the assumption that the marginal totals are fixed.

We shall concentrate on the case $m = 3$.

Complete independence in three-dimensional tables. Let us redefine H to mean independence of Queen's highways, roads and streets in a $q \times r \times s$ contingency table, and we consider the following models. We call the first model number 4 to avoid confusion with the models for two-dimensional tables.

MODEL 4. Multinomial sampling with qrs categories, cell probabilities p_{hij}, and sample size N.

MODEL 5. Deciding the totals $n_{h\cdot\cdot}$ (or $n_{\cdot i\cdot}$ or $n_{\cdot\cdot j}$) in advance, and taking q independent multinomial samples, each with rs categories, and with cell probabilities $p_{hij}/p_{h\cdot\cdot}$.

MODEL 6. Deciding the totals $n_{\cdot ij}$ (or $n_{h\cdot j}$ or $n_{hi\cdot}$) in advance, and taking rs independent multinomial samples, each with q categories, and with cell probabilities $p_{hij}/p_{\cdot ij}$.

MODEL 7. Deciding the totals $n_{\cdot i\cdot}$ and $n_{\cdot\cdot j}$ in advance.

MODEL 8. Deciding the totals $n_{h\cdot\cdot}$ and $n_{\cdot ij}$ in advance.

MODEL 9. Deciding the totals $n_{h\cdot\cdot}$, $n_{\cdot i\cdot}$, and $n_{\cdot\cdot j}$ in advance.

Thus, for $m = 3$, there are six kinds of models for H, or fourteen in all if we allow comprehensively for permutations of the subscripts. [For $m = 1, 2, 3, 4, \cdots$, the numbers of models are $1, 3, 6, 11, 18, \cdots$, which are 1 less than the "numbers of partitions into parts of two kinds" (Sloane (1973), page 59), and are $b_{m+1} - 1$ if the comprehensive count is used, where the b_m's are again the exponential numbers.] We have ignored models in which there is "overlapping" of the subscripts, for example, where $n_{\cdot ij}$, $n_{h\cdot j}$ and $n_{hi\cdot}$ are all given in advance. For such models, if they arise, it is perhaps more appropriate to consider the hypotheses of maximum entropy subject to the given marginal totals (for example, Good (1963)) rather than hypotheses of ordinary independence.

As in Section 3, given H, we assume the priors $D^*(q, 1)$, $D^*(r, 1)$ and $D^*(s, 1)$ for the respective one-dimensional marginal totals, when they are not fixed; and, given $\bar{H}$, we assume the prior $D^*(qrs, 1)$ for (n_{hij}). It follows that, given $\bar{H}$, the priors for the one-dimensional marginal totals are respectively $D^*(q, rs)$, $D^*(r, sq)$ and $D^*(s, qr)$.

We introduce the notation

$$\Psi((m_\nu), t, t') = (N!)^{-1}\Phi((m_\nu), t, t') \prod m_\nu! \,. \tag{12.3}$$

Then

$$P(n_{h\cdot\cdot}) \mid H) = \Phi((n_{h\cdot\cdot}), q, 1) \tag{12.4}$$

and

$$P((n_{h\cdot\cdot}) \mid \bar{H}) = \Phi((n_{h\cdot\cdot}), q, rs) \tag{12.5}$$

$$P((n_{\cdot ij}) \mid \bar{H}) = \Phi((n_{\cdot ij}), rs, q) \tag{12.6}$$

$$P((n_{hij}) \mid \bar{H}) = \Phi((n_{hij}), qrs, 1)\,. \tag{12.7}$$

An analogue of the Fisher–Yates formula (4.1), readily proved, is

$$P((n_{hij}) \mid (n_{h\cdot\cdot}), (n_{\cdot i\cdot}), (n_{\cdot\cdot j}), H) = \frac{\prod n_{h\cdot\cdot}! \prod n_{\cdot i\cdot}! \prod n_{\cdot\cdot j}!}{(N!)^2 \prod n_{hij}!}\,. \tag{12.8}$$

From (12.4) and (12.7) we infer that

$$F_4 = \frac{\Psi((n_{hij}), qrs, 1)}{\Psi((n_{h\cdot\cdot}), q, 1)\Psi((n_{\cdot i\cdot}), r, 1)\Psi((n_{\cdot\cdot j}), s, 1)} \tag{12.9}$$

which gives the Bayes factor against independence of Queen's highways, roads and streets, when sampling N objects at random, without constraints. An obvious analogue of (9.2) could be jotted down.

By using (12.8) and arguments resembling those for (4.6) and (4.7) we obtain

$$F_5 = \frac{\Psi((n_{hij}), qrs, 1)}{\Psi((n_{h\cdot\cdot}), q, rs)\Psi((n_{\cdot i\cdot}), r, 1)\Psi((n_{\cdot\cdot j}), s, 1)}\,. \tag{12.10}$$

By means of the relation (compare (4.1)),

$$P((n_{hij}) \mid (n_{h\cdot\cdot}), (n_{\cdot ij}), H) = \frac{\prod n_{h\cdot\cdot}! \prod n_{\cdot ij}!}{N! \prod n_{hij}!}\,, \tag{12.11}$$

we can prove that

$$F_6 = \frac{\Psi((n_{hij}), qrs, 1)}{\Psi((n_{h\cdot\cdot}), q, 1)\Psi((n_{\cdot ij}), rs, q)}\,. \tag{12.12}$$

Again, by using (12.8) and arguments similar to those for F_3, we obtain

$$F_7 = \frac{N!}{\prod n_{\cdot i\cdot}! \prod n_{\cdot\cdot j}!} \cdot \frac{\Psi((n_{hij}), qrs, 1)}{\Psi((n_{h\cdot\cdot}), q, 1) \sum^{\cdot,2,3}\left[\dfrac{\Psi((m_{\cdot ij}), rs, q)}{\prod m_{\cdot ij}!}\right]} \tag{12.13}$$

where $\sum^{\cdot,2,3}$ denotes a summation over all $m_{\cdot ij}$ for which $m_{\cdot i\cdot} = n_{\cdot i\cdot}$, $m_{\cdot\cdot j} = n_{\cdot\cdot j}$ (for all i and j).

By using (12.11) we can prove that

$$F_8 = \frac{N!}{\prod n_{h\cdot\cdot}!\prod n_{\cdot ij}!}\cdot\frac{\Psi((n_{hij}), qrs, 1)}{\sum^{1,23}[\Psi((m_{hij}), qrs, 1)/\prod m_{hij}!]}, \tag{12.14}$$

where $\sum^{1,23}$ denotes a summation over all m_{hij} for which $m_{h\cdot\cdot} = n_{h\cdot\cdot}$ and $m_{\cdot ij} = n_{\cdot ij}$ (for all h, i and j).

By using (12.8) we can prove that

$$F_9 = \frac{(N!)^2}{\prod n_{h\cdot\cdot}!\prod n_{\cdot i\cdot}!\prod n_{\cdot\cdot j}!}\cdot\frac{\Psi((n_{hij}), qrs, 1)}{\sum^{1,2,3}[\Psi((m_{hij}), qrs, 1)/\prod m_{hij}!]}, \tag{12.15}$$

where $\sum^{1,2,3}$ denotes a summation over all m_{hij} for which $m_{h\cdot\cdot} = n_{h\cdot\cdot}$, $m_{\cdot i\cdot} = n_{\cdot i\cdot}$, $m_{\cdot\cdot j} = n_{\cdot\cdot j}$ (for all h, i and j).

The factors F_4 to F_9 all reduce to 1 when $N = 0$ or 1, as they should.

The summations $\sum^{\cdot,2,3}$, $\sum^{1,23}$ and $\sum^{1,2,3}$ can all be expressed in a form similar to (5.5). For example,

$$\sum\nolimits^{1,2,3} = \int_0^\infty \frac{\Gamma(qrsk)}{\Gamma(N+qrsk)}\mathscr{C}\left(\prod w_h^{n_{h\cdot\cdot}}\prod x_i^{n_{\cdot i\cdot}}\prod y_j^{n_{\cdot\cdot j}}\right) (1 - w_h x_i y_j)^{-k}\phi(k)\,dk\,, \tag{12.16}$$

while the expressions for $\sum^{\cdot,2,3}$ and $\sum^{1,23}$ can be directly inferred from (5.5) because they are also "two-dimensional." The number of terms in $\sum^{1,2,3}$, if computed directly instead of by complex integration, is equal to the number $A((n_{h\cdot\cdot}), (n_{\cdot i}), (n_{\cdot\cdot j}))$ of three-dimensional arrays having the assigned one-dimensional marginal totals. Apart from the obvious generating function $\prod(1 - w_h x_i y_j)^{-1}$, there is perhaps little known about the enumeration of such arrays. As clarified at the beginning of Section A2, a conjecture analogous to (6.4) would be

$$A((n_{h\cdot\cdot}), (n_{\cdot i\cdot}), (n_{\cdot\cdot j})) \approx B((n_{h\cdot\cdot}), (n_{\cdot i\cdot}), (n_{\cdot\cdot j}))\,, \tag{12.17}$$

where

$$B((n_{h\cdot\cdot}), (n_{\cdot i\cdot}), (n_{\cdot\cdot j})) = \frac{\prod\binom{n_{h\cdot\cdot}+rs-1}{n_{h\cdot\cdot}}\prod\binom{n_{\cdot i\cdot}+sq-1}{n_{\cdot i\cdot}}\prod\binom{n_{\cdot\cdot j}+qr-1}{n_{\cdot\cdot j}}}{\binom{N+qrs-1}{N}^2}. \tag{12.18}$$

There is the possibility of improving the conjecture along the lines of (6.6) to (6.8), perhaps by multiplying B by a quantity resembling

$$1.3^2N^6/(qrs\sum n_{h\cdot\cdot}^2\sum n_{\cdot i\cdot}^2\sum n_{\cdot\cdot j}^2)\,, \tag{12.19}$$

which reduces to 1.69 when $q = r = s = 2$, $n_{h\cdot\cdot} = n_{\cdot i\cdot} = n_{\cdot\cdot j} = 1$ (for all h, i, j) so that $N = 2$; and in this case $A/B = 1.75$.

Likewise an analogue of conjecture (7.1) would be, in a sufficiently self-explanatory notation,

$$(12.20) \qquad F_9 \approx F_5^{(1)}F_5^{(2)}F_5^{(3)}/(F_4)^2 = \frac{\Psi((n_{hij}), qrs, 1)}{\Psi((n_{h}..), qrs, 1)\Psi((n_{.i.}), qrs, 1)\Psi((n_{..j}), qrs, 1)}.$$

I hope we shall test these conjectures numerically in due course.

Under Model 9, H_9 could in principle be used as an "orthodox" (non-Bayesian or "Fisherian") statistic, by the Bayes/non-Bayes synthesis. Moreover $F_9(k)$ could be defined in a manner similar to that of $F_3(k)$ in (10.1), and the asymptotic distribution, given H, of

$$(12.21) \qquad (2 \log_e \max_k F_9(k))^{\frac{1}{2}},$$

under Model 9, will be proportional to a standardized normal distribution. This distribution may be more accurately attained than that of X^2, especially when many cell entries are small.

Extensions of the results to more than three dimensions can now easily be made, apart from computational difficulties.

13. Some provisional numerical examples on the distribution of G, etc. and the values of the Bayes factors. Since Part II of this work will involve many distinct calculations and may be delayed for some time, I give here some provisional results, programmed by Dr. J. F. Crook and reported here with his permission.

Let

$$(13.1) \qquad X^2 = \sum \frac{(n_{ij} - n_{i.}n_{.j}/N)^2}{n_{i.}n_{.j}/N}$$

and the ordinary likelihood-ratio statistic

$$(13.2) \qquad \Lambda = 2 \sum n_{ij} \log n_{ij} - 2 \sum n_{i.} \log n_{i.} - 2 \sum n_{.j} \log n_{.j},$$

which both have asymptotically the χ^2 distribution with $(r - 1)(s - 1)$ degrees of freedom when H is true. Let $t - 1 = (r - 1)(s - 1)$, and let c_t be the probability that χ^2 with $t - 1$ degrees of freedom exceeds $t - 1$ (as in Good and Crook (1974), or in Good (1967) where c_t is tabulated). Let $F_3(k)$ and G be defined by (10.1) and (10.2).

Then, asymptotically, as in the two papers just mentioned,

$$P(G > x) \approx c_t \times \text{double tail of the normal "sigmage" } x.$$

We have examined the complete exact distributions of G, X^2 and Λ, assuming independence, for the eleven contingency tables with row and column totals:

[8, 16; 8, 8, 8] [5, 5, 20; 5, 5, 20] [4, 6; 2, 3, 5] [5, 10, 15; 5, 5, 20]
[4, 8, 18; 5, 9, 16] [6, 6, 6; 6, 6, 6] [4, 5, 9; 5, 6, 7] [5, 6, 7; 5, 6, 7]
[5, 6, 7; 1, 2, 15] [2, 3, 13; 1, 2, 15] and [1, 2, 2, 3; 1, 1, 3, 3].

These totals are small enough so that all possible interiors (n_{ij}) could be examined.

Let P denote an exact tail probability and Q the corresponding asymptotic tail probability. Let the larger of P/Q and Q/P be called an "error ratio." A summary of cases where the error ratio was less than 2 was compiled in the following manner. Choose a range $[10^{-\nu-1}, 10^{-\nu})$ of values of Q ($\nu = 0, 1, 2, \cdots$) and record for each of the eleven sets of contingency tables the proportion of times that the error ratio is less than 2 among the cases where Q is in the chosen range (for a specific statistic such as Λ). For the statistic Λ, and the range $10^{-3} \leqq Q < 10^{-2}$ we obtained, for the eleven contingency tables, the ratios:

$$0.33, 1, 0.50, 1, 1, 0.33, 0.20, 0, —, 1, — .$$

We crudely summed these figures to get the entry 5.36/9 of the summary, Table 4. This table is itself crudely summarized by adding up the results for $\nu = 0, 1, \cdots, 8$. The theoretical asymptotic distribution of G, allowing for c_t, thus appears to approximate its exact distribution better than those for X^2 and Λ, although the number of experiments needs to be increased.

TABLE 4

Fraction of cases where $10^{-\nu-1} \leqq Q < 10^{-\nu}$ *and* $\frac{1}{2} < P/Q < 2$

ν	0	1	2	3	4	5	6	7	8	*Total*
G	2/2	9.67/11	7/10	6/8	5/6	4.59/5	2.37/6	1.5/3	1/4	39.1/55
X^2	2.67/3	8.25/11	6.21/8	4.76/8	3.33/6	1.85/5	1.5/5	0.67/3	1/2	30.2/51
Λ	1/2	4.42/11	5.36/9	4.72/6	4.8/7	3.36/6	2.14/5	1.67/4	0/1	27.5/51

In the output of another program we computed various statistics for all possible tables with eleven sets of row and column totals, such as [15, 15; 10, 10, 10]. Here is the output for the 2×3 table with $n_{11} = 8$, $n_{12} = 7$, $n_{13} = 0$; $n_{21} = 2$, $n_{22} = 3$ and $n_{23} = 10$:

$$F_1 = 47\,, \quad F_2 = 39\,, \quad F_{(2)} = 41\,, \quad F_3 = 17.5 \quad \text{(see Section 5)}\,.$$

$$F_1' = 1360\,, \quad F_2' = 770\,, \quad F_{(2)}' = 669\,, \quad F_3' = 316 \quad \text{(see Section 6)}\,.$$

$$B/A = F_3'/\hat{F}_3' = 0.833 \quad \text{where} \quad \hat{F}_3' = F_2'F_{(2)}'/F_1' \quad \text{(see Section 6)}\,.$$

$$\hat{F}_3 = F_2F_{(2)}/F_1 = 33.8\,, \qquad F_3/\hat{F}_3 = 0.517\,,$$

$$X^2 = 15.2 \quad (2 \text{ d.f.})\,, \qquad Q(X^2) = 1/2000\,,$$

$$\Lambda = 19.4 \quad (2 \text{ d.f.})\,, \qquad Q(\Lambda) = 1/16000\,.$$

Total of F.Y. probabilities for tables at least as extreme $= 1/1700 = P$, say.

$$\text{FRACT} = 2.71 \quad \text{(see Section 8)}\,.$$

Thus F_1, F_2, $F_{(2)}$ are very close, and F_3 is of the same order of magnitude; and F_3 and $\hat{F}_3$ are within a factor of 2 of each other. We have further $F_3PN^{\frac{1}{2}} = 0.056$, which is the kind of relationship that occurred in the earlier research on multinomial significance tests (Good (1967); Good and Crook (1974)).

The big discrepancy between $Q(\Lambda)$ and P is consistent with the dubious asymptotic χ^2 approximation for Λ in many cases for small P-values, where the χ^2 approximation for X^2 is more often still satisfactory.

The factors F_1' etc. are bad overestimates, and show the need to use *mixtures* of Dirichlet distribution, as advocated in Good (1965a), and as were used in the more comprehensive work on multinomial significance tests.

Note that, as we expected, FRACT (or rather the absolute value of its logarithm) is not large, i.e., the row and column totals do not give much evidence against H.

14. New literature on the Dirichlet approach. After this paper was submitted, a paper on the "Dirichlet" approach to two-dimensional contingency tables was published by Gûnel and Dickey (1974). It generalized the model of Good (1950, 1965) by using general(but unmixed) Dirichlet distributions instead of symmetric ones. My reason for using symmetric Dirichlets, for the most part, is the same as was mentioned in the analogous work on the multinomial distribution (Good (1967), page 430), namely that "my aim was to use the simplest model that makes reasonable sense." The use of mixtures of general Dirichlet distributions would be somewhat complex. This is not by any means to deny that the general Dirichlet distributions (and their mixtures) are of use, so that Gûnel and Dickey's paper may be regarded as complementary to the present one.

Acknowledgment. I am indebted to James F. Crook for interaction and fellowship which kept these problems in continual focus during a critical phase of the work.

APPENDIX A

The enumeration of arrays

A1. Brief review of the literature. The problem of enumerating arrays with given marginal totals, that is, of evaluating $A((n_{i\cdot}), (n_{\cdot j}))$, in the notation of Section 5, has been attacked in the literature of combinatorics. The generating function (5.6) with $k = 1$ is easily seen, but simple "explicit" formulae for the number of arrays are known only in a few special cases. We shall discuss (5.6) again in Appendix B.

It may be noted that $A((n_{i\cdot}), (n_{\cdot j}))$ is equal to the number of ways that r types of objects can be distributed in s boxes, so that there are altogether $n_{i\cdot}$ objects of type i, and there are $n_{\cdot j}$ objects in box j. This interpretation is mentioned, for example, for the case where $r = s$ and all marginal totals equal n, by Nath and Iyer (1972).

When $r = s$ and all the marginal totals are equal to n, we write $A(n, r \times r)$ for $A((n_{i\cdot}), (n_{\cdot j}))$. Obviously

$$A(n, 1 \times 1) = 1\,, \qquad A(n, 2 \times 2) = n + 1 \tag{A1.1}$$

and it was proved by MacMahon ((1915–1960), **2** 161) that

$$A(n, 3 \times 3) = \binom{n+2}{2} + 3\binom{n+3}{4}\,. \tag{A1.2}$$

These numbers were tabulated in 1856 (see Sloan (1973), page 142) and were called "doubly triangular numbers," because they happen to be equal to the triangular numbers of triangular numbers. The interpretation in terms of arrays was not then known.

Gupta, in Anand, Dumir and Gupta (1966), conjectured that

$$A(n, r \times r) = \sum_{\nu=0}^{(r-1)(r-2)/2} c_\nu^{(r)} \binom{n+r-1+\nu}{r-1+2\nu} \tag{A1.3}$$

where the coefficients $c_\nu^{(r)}$ are independent of n. This conjecture was proved by Stanley (1973). Based on this conjecture, Anand, Dumir and Gupta obtained the formula

$$A(n, 4 \times 4) = \binom{n+3}{3} + 20\binom{n+4}{5} + 152\binom{n+5}{7} + 352\binom{n+6}{9}. \tag{A1.4}$$

Stein and Stein (1970) also assumed Gupta's conjecture, and used a branching algorithm, the idea of which they attribute to MacMahon, to evaluate $A(n, r \times r)$ for enough values of n to obtain the coefficients for the cases $r = 5$ and 6, namely:

(A1.5) 1, 115, 5390, 101275, 858650, 3309025 and 4718075

and

(A1.6) 1, 714, 196677, 18941310, 809451144, 17914693608, 223688514048, 1633645276848, 6907466271384, 15642484909560 and 1466561365176.

In addition, Stein and Stein give tables of the exact values of $A(n, r \times r)$ for $r = 4$, 5 and 6 with $n = 1(1)11$; and for $n = 2$, 3, 4 and 5 with $r = 1(1)15$. For example, $A(5, 15 \times 15) = 1.9208\cdots \times 10^{50}$.

Anand, Dumir and Gupta (1966), give the formula

$$\sum_r A(1, 2, r \times r)x^r/(r!)^2 = e^{x/2}(1 - x)^{-\frac{1}{2}}, \tag{A1.7}$$

which is of the form called a "double exponential generating function" by Stanley (1975).

Some other results can be inferred from Abramson and Moser (1973). For example, the number of $r \times 3$ arrays with all row totals equal to n, and column totals μ, ν, $rn - \mu - \nu$ ($n \geqq \mu$, $n \geqq \nu$, $r \geqq 2$).

A2. Approximations to the number of arrays. Approximations to $A((n_{i\cdot}), (n_{\cdot j}))$ have already been mentioned or suggested in (6.5), (6.6) and (6.8). An intuitive interpretation of (6.5) can be given: Imagine each row total to be partitioned into the s cells in its row giving a table T_1, and each column total partitioned into the r cells in its column giving a table T_2. From the point of view of someone who did not know the marginal totals, each table T_1 and T_2 could be regarded as more or less a random table of sample size N. Thus the "probability" that T_1 and T_2 are identical might be roughly equal to the reciprocal of the number of ordered partitions of N into all rs cells.

If we know that the row totals are rough this intuitive argument is somewhat undermined, so a "correction for roughness" appears to be indicated.

A similar argument explains the approximation (12.17) for the number of three-dimensional arrays.

A further discussion of approximations is delayed to Appendix **B** where a more general problem is discussed.

APPENDIX B

The calculation of $A(k, (n_{i\cdot}), (n_{\cdot j}))$

B1. Some exact formulae. It will be recalled that to calculate $F_3(k)$ and F_3 we need to compute $A(k) = A(k, (n_{i\cdot}), (n_{\cdot j}))$, as defined by (5.6), for positive values of the real number k, unless we succeed in satisfying ourselves that (7.1) or (B2.1), below, give adequate approximations.

I here mention some formulae for $A(k)$ the proofs of which I hope to publish in Good and Crook (1976). We write $\mathbf{m} = (n_{i\cdot})$, $\mathbf{n} = (n_{\cdot j})$. Let

(B1.1) $$h_m(k) = \mathscr{C}(x^m) \prod_{j=1}^{s} (1 - xy_j)^{-k},$$

so that $h_m(k)$ is a generalization of the "homogeneous product sum" $h_m(1)$. Then

(B1.2) $$A(k, \mathbf{m}, \mathbf{n}) = \mathscr{C}(\mathbf{y}^{\mathbf{n}})\{h_{m_1}(k)\, h_{m_2}(k) \cdots h_{m_s}(k)\}$$

and this is a generalization of a formula for $A(\mathbf{m}, \mathbf{n})$ given by MacMahon (1915–1960 **1**, 234). It can be proved that

(B1.3) $$ih_i(k) = k[h_{i-1}(k)s_1 + h_{i-2}(k)s_2 + \cdots + h_1(k)s_{i-1} + s_i],$$

where $s_1 = y_1 + y_2 + \cdots + y_s$, $s_2 = y_1^2 + y_2^2 + \cdots + y_s^2, \cdots$, of which the case $k = 1$ gives the Newton–Crocchi formulae (see Vahlen (1898–1904), page 465). (B1.3) can be used recursively to express the $h_i(k)$'s in terms of the power sums $s_1, s_2, \cdots$. Then (B1.2) can be used to evaluate $A(k, \mathbf{m}, \mathbf{n})$, a method used with $k = 1$ by Stein and Stein (1970) for some of their results.

One of the main formulae in Good and Crook (1976) is

(B1.4) $$A(k, \mathbf{m}, \mathbf{n}) = \frac{1}{\prod t_j} \sum_{\boldsymbol{\nu}} \boldsymbol{\omega}^{(\mathbf{t}-\mathbf{n})\boldsymbol{\nu}} \prod_{i=1}^{r} C(k, (\omega_j^{\nu_j}), m_i),$$

where

(B1.5) $$C(k, (\omega_j^{\nu_j}), m_i) = \mathscr{C}(x^{m_i}) \prod_{j=1}^{s} (1 - x\omega_j^{\nu_j})^{-k},$$

t_j is any integer exceeding n_j $(j = 1, 2, \cdots, s)$, $\omega_j = \exp(2\pi(-1)^{\frac{1}{2}}/t_j)$, $\boldsymbol{\omega}^{\boldsymbol{\mu\nu}}$ denotes $\omega_1^{\mu_1\nu_1}\omega_2^{\mu_2\nu_2}\cdots$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \cdots, \nu_s)$, and $\nu_j = 0, 1, \cdots, t_j - 1$ $(j = 1, 2, \cdots, s)$.
Also

(B1.6) $$C(k, (\omega_j^{\nu_j}), m_i) = \sum_{\alpha} \frac{k^{\alpha_1+\alpha_2+\cdots} S_1^{\alpha_1} S_2^{\alpha_2} \cdots}{\alpha_1!\, \alpha_2!\, \alpha_3! \cdots 1^{\alpha_1} 2^{\alpha_2} 3^{\alpha_3} \cdots}$$

where $\alpha_1, \alpha_2, \alpha_3, \cdots$ run through all nonnegative integer solutions of

$$\alpha_1 + 2\alpha_2 + 3\alpha_3 + \cdots = m_i,$$

and

$$S_1 = \sum_j \omega_j^{\nu_j}, \qquad S_2 = \sum_j \omega_j^{2\nu_j}, \cdots .$$

For example,

$$\text{(B1.7)} \qquad A(k, n, r \times r) = \frac{1}{(n+1)^r} \sum_\nu \omega^{|\nu|}[C(k, (\omega^{\nu_j}), n]^r$$

where $\omega = \exp[2\pi(-1)^{\frac{1}{2}}/(n+1)]$, each of $\nu_1, \nu_2, \cdots, \nu_r$ runs from 0 to n, and $|\boldsymbol{\nu}| = \nu_1 + \nu_2 + \cdots + \nu_r$. For example,

$$\text{(B1.8)} \qquad A(k, 3, r \times r) = \frac{k^r}{24^r} \sum_\nu (-1)^{\frac{1}{2}|\nu|}(k^2 S_1^3 + 3kS_1S_2 + 2S_3)^r$$

with the appropriate definitions of S_1, S_2 and S_3, and where each of $\nu_1, \nu_2, \cdots, \nu_r$ runs from 0 to 3. Note that, given k and n, $C(k, (\omega^{\nu_j}), n)$ depends only on the frequency count of (ν_j) (the numbers of zeros, ones, etc., among its components), a fact that leads to a big reduction in computer time.

We have used (B1.7) for computing $A(1, n, r \times r)$ to 13 significant figures for $n = 3(1)8$, $r = 5(1)8$ and the results agree with those of Stein and Stein where they overlap. (B1.7) has also been used for values of $k \neq 1$. The branching method used by Stein and Stein to count arrays can be generalized to provide another method for calculating $A(k, n, r \times r)$ but it is much more difficult to program and we believe it is slower.

Some corresponding formulae for three-dimensional arrays are given in Good and Crook (1976). In addition a few more special results are proved there, such as

$$\text{(B1.9)} \qquad A(k, n, 2 \times 2) = \sum_{\nu=0}^{n} \binom{\nu+k-1}{\nu}^2 \binom{n-\nu+k-1}{n-\nu}^2 ,$$

$$\text{(B1.10)} \qquad A(k, 2, r \times r) = (r!)(k/2)^{2r} \sum_{\nu=0}^{r} \frac{(2/k)^\nu}{\nu!} \binom{2r-2\nu}{r-\nu}$$

of which the case $k = 1$ is given by Stein and Stein (1970). Further

$$\text{(B1.11)} \qquad \sum_{r=0}^{\infty} \frac{A(k, 2, r \times r)x^r}{(r!)^2} = \frac{e^{kx/2}}{(1 - k^2x)^{\frac{1}{2}}} ,$$

of which the case $k = 1$ is (A1.7).

Very explicit formulae for $A(k, n, 3 \times 3)$ ($n = 1, 2, 3$ and 4) are given in Good and Crook (1976), for example, $A(k, 1, 3 \times 3) = 6k^3$ and

$$\text{(B1.12)} \qquad A(k, 3, 3 \times 3) = 6\binom{k+2}{3}^3 + 12\binom{k+1}{2}^3 k^3 + 18\binom{k+2}{3}\binom{k+1}{2}^2 k^2 + 18\binom{k+1}{2}^2 k^5 + k^9 .$$

B2. Approximations to $A(k)$. We now discuss approximations to $A(k)$ from which F_3 can be approximated via equations (5.4) and (5.5). One of these approximations arises from an asymptotic expansion for large k, but the first term of the expression appears to be a good approximation even when k is as small as 1, that is, for the enumeration of arrays. Unfortunately when $k = 1$, unlike

Stirling's formula for the factorial, the approximation is less good when the second term is taken into account. If $F_3(k)$ takes its maximum value when $k \ll 1$ this asymptotic expression will not be useful, and then one may need to rely on the approximation (7.1), or, if practicable, to use one of the accurate methods described in Section B1.

Just as $F_3(k)$ was defined in Section 10 by taking ϕ as a Dirac delta function in the expression for F_3, we define $F_1(k)$, $F_2(k)$ and $F_{(2)}(k)$ in the obvious analogous manner, so that $F_1(1) = F_1'$, $F_2(1) = F_2'$, and $F_{(2)}(1) = F'_{(2)}$, where F_1' etc. are defined in Section 6. The obvious generalization of the approximation for F_3' (for which special case we have mentioned numerical support) is

(B2.1) $$F_3(k) \approx F_2(k)F_{(2)}(k)/F_1(k) \qquad (k \geqq 1)\,.$$

This implies, after some straightforward manipulations, that

(B2.1A) $$A(k) \approx B(k) \qquad (k \geqq 1)\,,$$

where

(B2.2) $$B(k) = B(k \mid (n_{i\cdot}), (n_{\cdot j})) = \frac{\prod \binom{n_{i\cdot} + sk - 1}{n_{i\cdot}} \prod \binom{n_{\cdot j} + rk - 1}{n_{\cdot j}}}{\binom{N + rsk - 1}{N}}\,,$$

which generalizes (6.4) and (6.5). For example, for square tables with flat marginal totals n, we see from (B1.10) and (B1.12) that the ratio $B(k)/A(k)$ for $(k, n, r) = (2, 2, 4)$ is 0.853, for $(4, 2, 4)$ is 0.923, for $(2, 3, 3)$ is 0.876, for $(4, 3, 3)$ is 0.938, but for $(\frac{1}{2}, 2, 4)$ is 0.504, and for $(\frac{1}{4}, 2, 4)$ is 0.231. See also Table 2.

Some of this appendix is concerned with a partial justification of (B2.2) along mathematical lines. To this end we consider a different way of approximating $A(k)$ (defined by (5.6)) based on a multivariate saddlepoint theorem. The solution is given in compact form in formula (B2.13) which requires the evaluation of a determinant and, for terms after the first, the inversion of a matrix. The determinant and inverse take simple forms when the column totals are equal, and we give a more explicit solution for this case in formula (B2.20).

Note first that $A(k)$ is unchanged if we replace y_s by 1. This is because in each term in the expansion of $\prod (1 - x_i y_j)^{-k}$, the sum of the powers of the y's must be equal to that of the x's. For this reason $A(k)$ can be expressed as a contour integral in $r + s - 1$ dimensions rather than $r + s$ dimensions.

We now invoke a theorem for approximating coefficients in a power of $f(\mathbf{z}) = f(z_1, z_2, \cdots, z_{r+s-1})$, where $f(\mathbf{z})$ is a power series having nonnegative coefficients: see, for example, Good (1957, 1961). For the present application

(B2.3) $$f(\mathbf{z}) = \prod_{i=1}^{r} \prod_{j=1}^{s} (1 - x_i y_j)^{-1}\,,$$

where $y_s = 1$, and $z_h = x_h$ $(1 \leqq h \leqq r)$, $z_h = y_{h-r}$ $(r + 1 \leqq h \leqq r + s - 1)$. Similarly, let $m_h = n_{h\cdot}$ $(1 \leqq h \leqq r)$, $m_h = n_{\cdot, h-r}$ $(r + 1 \leqq h \leqq r + s - 1)$. The theorem gives an approximation for $A(k)$, officially when k is large, but its first term appears to give good results in the present application even when k is as small as 1, that is, it appears to lead to a good approximation to the number of arrays, judging by numerical results for a limited class of tables for which the results could be calculated without using a large-scale computer. Further numerical results will be reported in Part II.

To apply the theorem we first write down equations satisfied by the saddle-point $\boldsymbol{\rho} = (\rho_1, \rho_2, \cdots, \rho_{r+s-1})$. It will be convenient to write $\rho_{r+j} = \tau_j$ ($j = 1, 2, \cdots, s$). We may assume, without real loss of generality, that $n_{\cdot s} \geqq n_{j\cdot}$ for all j. We have

(B2.4)
$$\frac{n_{i\cdot}}{k} = \rho_i \frac{\partial}{\partial \rho_i} \log f(\boldsymbol{\rho}) = \textstyle\sum_{j=1}^{s} \frac{\rho_i \tau_j}{1 - \rho_i \tau_j} \qquad (1 \leqq i \leqq r),$$
$$\frac{n_{\cdot j}}{k} = \tau_j \frac{\partial}{\partial \tau_j} \log f(\boldsymbol{\rho}) = \textstyle\sum_{i=1}^{r} \frac{\rho_i \tau_j}{1 - \rho_i \tau_j} \qquad (1 \leqq j \leqq s - 1)$$

(this is also true when $j = s$, but is not then an independent equation), and, as can be seen,

$$0 \leqq \rho_i < 1 \quad (i = 1, 2, \cdots, r); \qquad 0 \leqq \tau_j \leqq 1 \quad (j = 1, 2, \cdots, s), \qquad \tau_s = 1.$$

Owing to the convexity of a function called f^* by Good ((1957), page 874), these equations must be uniquely solvable by a method of steepest ascent, but the following iterative method has been found to work. When $\rho_1, \rho_2, \cdots, \rho_r$ have assigned values, then *each* of $\tau_1, \cdots, \tau_{s-1}$ can be obtained by Newton's iterative method, because the right sides of equations (B2.4) are increasing in all variables. We can then solve for $\rho_1, \rho_2, \cdots, \rho_r$ again; and so on, in a higher-level iteration. It can be seen that, at every stage of the iteration, we have (i) $\rho_i < \rho_{i'}$ if $n_{i\cdot} < n_{i'\cdot}$, (ii) $\tau_j < \tau_{j'}$ if $n_{\cdot j} < n_{\cdot j'}$, (iii) $\rho_i/n_{i\cdot} > \rho_{i'}/n_{i'\cdot}$ if $n_{i\cdot} < n_{i'\cdot}$, (iv) $\tau_j/n_{\cdot j} > \tau_{j'}/n_{\cdot j'}$ if $n_{j\cdot} < n_{j'\cdot}$, (v) $\rho_i < n_{i\cdot}/(n_{i\cdot} + k)$.

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_{r+s-1})$ be a vector of nonnegative integers, where $|\boldsymbol{\beta}| = \beta_1 + \beta_2 + \cdots, \beta_{r+s-1} \geqq 2$, and let

(B2.5)
$$\kappa_\beta = \left\{\textstyle\prod_{h=1}^{r+s-1} \left(\rho_h \frac{\partial}{\partial \rho_h}\right)^{\beta_h}\right\} \log f(\boldsymbol{\rho}),$$

where the ρ's are regarded as variables until after the differentiations have been performed. Let ν_β's be defined by the identity

(B2.6)
$$\textstyle\sum_{\beta}^{|\beta|\geq 0} \frac{\nu_\beta}{\boldsymbol{\beta}!} \boldsymbol{\xi}^\beta = \exp\left\{k \textstyle\sum_{\beta}^{|\beta|\geq 3} \frac{\kappa_\beta}{\boldsymbol{\beta}!} \boldsymbol{\xi}^\beta\right\}$$

where the multivariate notations used, for example, by Good (1957) are adopted. In particular

(B2.7)
$$\nu_0 = 1, \qquad \nu_\beta = 0 \quad \text{if} \quad |\boldsymbol{\beta}| = 1 \quad \text{or} \quad 2,$$
$$\nu_\beta = k\kappa_\beta \quad \text{if} \quad |\boldsymbol{\beta}| = 3, 4 \quad \text{or} \quad 5.$$

Let $\mathbf{K}$ denote the symmetric matrix formed from the κ_β's having $|\boldsymbol{\beta}| = 2$, that is,

(B2.8)
$$\mathbf{K} = \left\{\rho_h \frac{\partial}{\partial\rho_h}\left(\rho_{h'} \frac{\partial}{\partial\rho_{h'}}\right) \log f(\boldsymbol{\rho})\right\}$$

$(h, h' = 1, 2, 3, \cdots, r + s - 1)$. In our application $\mathbf{K}$ can be expressed as a block matrix,

(B2.9)
$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix},$$

where

(B2.10)
$$\mathbf{K}_{11} = \operatorname{diag}\left\{\sum_{j=1}^{s} \frac{\rho_i \tau_j}{(1 - \rho_i \tau_j)^2}\right\}_{i=1,2,\cdots,r},$$

(B2.11)
$$\mathbf{K}_{22} = \operatorname{diag}\left\{\sum_{i=1}^{r} \frac{\rho_i \tau_j}{(1 - \rho_i \tau_j)^2}\right\}_{j=1,2,\cdots,s-1},$$

(B2.12)
$$\mathbf{K}_{12} = \mathbf{K}'_{21} = \left\{\frac{\rho_i \tau_j}{(1 - \rho_i \tau_j)^2}\right\}_{i=1,2,\cdots,r;\, j=1,2,\cdots,s-1}$$

where the prime denotes transposition. ($\mathbf{K}$ would be singular if we had allowed j to range from 1 to s. This would have prevented the application of the saddlepoint theorem.) Let Δ denote the determinant of $\mathbf{K}$. Then, leaving aside finer points of rigor, the saddlepoint theorem gives the asymptotic formula (in which $|\boldsymbol{\beta}| = 2$ contributes nothing):

(B2.13)
$$A(k) \sim A_0(k) \sum_{\boldsymbol{\beta}}^{|\boldsymbol{\beta}| \text{ even}} \frac{\nu_{\boldsymbol{\beta}}}{(\frac{1}{2}|\boldsymbol{\beta}|)!} \left(-\frac{1}{2k}\right)^{\frac{1}{2}|\boldsymbol{\beta}|} \mathscr{C}(\mathbf{u}^{\boldsymbol{\beta}})(\mathbf{u}'\mathbf{K}^{-1}\mathbf{u})^{\frac{1}{2}|\boldsymbol{\beta}|},$$

where the m_h/k and k/m_h are bounded and k is "large," and

(B2.14)
$$A_0(k) = A_0(k \mid (n_{i\cdot}), (n_{\cdot j})) = \frac{[f(\boldsymbol{\rho})]^k}{(2\pi k)^{(r+s-1)/2} \boldsymbol{\rho}^{\mathbf{m}} \Delta^{\frac{1}{2}}}.$$

Stirling's formula for factorials is a special case of the one-dimensional form of (B2.13) (e.g., Good (1957), page 869), and the first and second Stirling approximations to 1! are 0.922 and 1.0060, so it is worth considering (B2.13) even when k is as small as 1. It may be difficult to obtain useful bounds on the errors, so that it may be necessary to generalize from numerical examples for the time being.

We recall that Δ can be expressed as

(B2.15)
$$\Delta = |\mathbf{K}_{11}|\,|\mathbf{K}_{22} - \mathbf{K}_{21}\mathbf{K}_{11}^{-1}\mathbf{K}_{12}|$$

(for example, Aitken (1956), page 67), and this is helpful because $\mathbf{K}_{11}$ and $\mathbf{K}_{22}$ are diagonal matrices.

We now consider the evaluation of $A_0(k)$ for the case where the column totals are equal ("flat"). (Of course the case of equal row totals is similar.) For this case it is readily verified that the solution of equations (B2.4) is given by

$$\text{(B2.16)} \qquad \rho_i = \frac{n_{i\cdot}}{n_{i\cdot} + ks} \qquad i = 1, 2, \cdots, r$$
$$\tau_j = 1 \qquad j = 1, 2, \cdots, s\,.$$

K can now be written in the form

$$\text{(B2.17)} \qquad \mathbf{K} = \begin{bmatrix} s\mathbf{\Theta} & \mathbf{\Theta J}_{r\times(s-1)} \\ \mathbf{\Theta J}_{(s-1)\times r} & \mathbf{I}_{s-1}\operatorname{tr}\mathbf{\Theta} \end{bmatrix}$$

where $\mathbf{\Theta} = \operatorname{diag}[\rho_i/(1-\rho_i)^2]$ and **J** is an all-ones matrix. The inverse turns out to be

$$\text{(B2.18)} \qquad \mathbf{K}^{-1} = \frac{1}{s\operatorname{tr}\mathbf{\Theta}}\begin{bmatrix} \mathbf{\Theta}^{-1}\operatorname{tr}\mathbf{\Theta} + (s-1)\mathbf{J}_{r\times r}\,, & -s\mathbf{J}_{r\times(s-1)} \\ -s\mathbf{J}_{(s-1)\times r} & , \quad s\mathbf{I}_{s-1} + s\mathbf{J}_{(s-1)\times(s-1)} \end{bmatrix}$$

as may be verified by multiplying it by **K**. The determinant of **K**, by (B2.15) combined with the evaluation of a circulant (e.g., Muir (1933–1960), pages 442 and 445), is

$$\text{(B2.19)} \qquad \Delta = s^{r-1}(\operatorname{tr}\mathbf{\Theta})^{s-1}\det\mathbf{\Theta}$$
$$= s^{r-1}\left\{\sum \frac{\rho_i}{(1-\rho_i)^2}\right\}^{s-1}\prod\frac{\rho_i}{(1-\rho_i)^2}\,.$$

It now follows from (B2.13) and (B2.7) that

$$\text{(B2.20)} \qquad A(k\,|\,n_{\cdot j}\text{'s equal}) \sim A_0(k)\left\{1 + \frac{1}{8k}\textstyle\sum_{|\beta|=4}\kappa_\beta\,\mathscr{E}(\mathbf{u}^\beta)(\mathbf{u}'\mathbf{K}^{-1}\mathbf{u})^2 + \cdots\right\},$$

where $\mathbf{K}^{-1}$ is given by (B2.18), κ_β by (B2.5), and

$$\text{(B2.21)} \qquad 1/A_0(k) = \prod \rho_i^{n_{i\cdot}+\frac{1}{2}}\prod(1-\rho_i)^{ks-1}(2\pi k)^{(r+s-1)/2}s^{(r-1)/2}$$
$$\times\left[\sum\frac{\rho_i}{(1-\rho_i)^2}\right]^{(s-1)/2}$$

where ρ_i is given by (B2.16). (When we write $A(k)$, $A_0(k)$, etc., we often leave it to the context to determine whether there is any constraint on the marginal totals.)

In the more special case of a table with both row totals equal and column totals equal ("flat margins") we have

$$\text{(B2.22)} \qquad A_0(k) = \frac{r^{r/2}s^{s/2}g(N+rsk)}{g(N)g(rsk)}\left[\frac{rsk}{2\pi N(N+rsk)}\right]^{(r+s-1)/2}$$

where g is the function defined by $g(x) = x^x$. In this case

$$\text{(B2.23)} \qquad \mathbf{K}^{-1} = \frac{rsk^2}{N(N+rsk)}\begin{bmatrix} r\mathbf{I}_r + (s-1)\mathbf{J}_{r\times r}\,, & -s\mathbf{J}_{r\times(s-1)} \\ -s\mathbf{J}_{(s-1)\times r} & , \quad s\mathbf{I}_{s-1} + s\mathbf{J}_{(s-1)\times(s-1)} \end{bmatrix}.$$

Let us define B_{SA} as the "Stirling approximation" to $B(k)$, obtained by replacing all the gamma functions and factorials in (B2.2) by the first terms of the Stirling formula. Then

$$B_{SA}(k) = B_{SA}(k\,|\,(n_{i\cdot}), (n_{\cdot j}))$$

$$(B2.24) \qquad = \frac{g(N) \prod g(n_{i\cdot} + sk) \prod g(n_{\cdot j} + rk)}{k^{rsk} g(N + rsk) \prod g(n_{i\cdot}) \prod g(n_{\cdot j})} \times \left\{ \frac{N(N + rsk)k^{r+s-1} r^{s-1} s^{r-1}}{(2\pi)^{r+s-1} \prod n_{i\cdot}(n_{i\cdot} + sk) \prod n_{\cdot j}(n_{\cdot j} + rk)} \right\}^{\frac{1}{2}}$$

and it can be seen from (B2.21) that when the column totals are flat,

$$(B2.25) \qquad A_0(k\,|\,n_{\cdot j}\text{'s equal}) = B_{SA}(k) \left[r \sum \frac{n_{i\cdot}}{N} \cdot \frac{n_{i\cdot} + sk}{N + rsk} \right]^{-(s-1)/2}.$$

By the Cauchy–Schwarz inequality it follows that

$$(B2.26) \qquad A_0(k) \geqq B_{SA}(k)\,,$$

equality holding if and only if the row totals are also flat. If the column totals are flat and the row totals are not very rough, and if s is small, we can see from (B2.25) that $B_{SA}(k)$, and hence $B(k)$, is not very different from $A_0(k)$. We therefore have a theoretical explanation of why $B(k)$ should often be a fair approximation to $A(k)$, in other words why $F_3(k) \approx F_2(k)F_{(2)}(k)/F_1(k)$, if k is not too small. We also have partly explained why the approximation can be improved by allowing for the roughness of the row totals and of the column totals, but the occurrence of the power $(s - 1)/2$ suggests that a better approximation than D might be found.

Even when k is as small as 1 the first term A_0 of the asymptotic expression gives a reasonable approximation to A, the number of arrays, for the square tables with flat margins mention in Table 2. Apart from the sparsest tables, the errors are of the order of 10% which for many purposes is negligible when estimating a Bayes factor, this being the application we have in mind.

Table 3 gives the values of A, A_0/A, B/A and D/A for a few tables having flat column totals but not flat row totals.

The asymptotic formula (B2.20) is based on regarding k as large, combined with the condition following (B2.13), so it is not surprising that the proportional errors do not tend to zero when k is fixed and $N \to \infty$. But the following results indicate that the approximation remains good for large N. When $r = s$, and when the tables have flat margins, we have

$$(B2.27) \qquad A_0(1) \sim e^{r^2} N^{(r-1)^2} r^{-(r-1)(2r-1)} (2\pi)^{-r+\frac{1}{2}}\,,$$

so that, from (A1.1) to (A1.6) when $N \to \infty$,

$$(B2.28) \qquad A_0(1)/A \to 0.867\,(r = 2)\,, \quad 0.899\,(r = 3)\,, \quad 0.878\,(r = 4)\,, \quad 0.857\,(r = 5)\,, \quad 0.840\,(r = 6)\,,$$

from which incidentally we can estimate the final coefficients in the Gupta–Stanley formula (A1.3) as 1.77×10^{22} for $r = 7$, 1.28×10^{34} for $r = 8$, and 7.6×10^{48} for $r = 9$.

Under the same circumstances, from (B2.2), we have

$$B(1) \sim (r^2-1)!\, N^{(r-1)^2} r^{-2r^2+2r}[(r-1)!]^{-2r}, \tag{B2.29}$$

so that

$$B(1)/A \to \tfrac{3}{4}\,(r=2)\,,\quad 0.768\,(r=3)\,,\quad 0.747\,(r=4)\,, \quad 0.728\,(r=5)\,,\quad 0.693\,(r=6)\,. \tag{B2.30}$$

On the evidence so far, $A_0(1)$ appears to be a somewhat better approximation to A than is $B(1)$, as well as having better theoretical support.

The second term of the asymptotic expansion. Think of formula (B2.13) as $A_0(k) + A_1(k) + \cdots$, where $|\boldsymbol{\beta}| = 4$ in $A_1(k)$. We shall give a formula for $A_1(k)$, but with most of the proof omitted. Let

$$T_{i,j,l} = \textstyle\sum_{\alpha=1}^{l-1} \alpha!\, \mathscr{S}_{l-1}^{(\alpha)}(\rho_i\tau_j)^{\alpha}(1-\rho_i\tau_j)^{-\alpha-1} \qquad (i \leqq r, j \leqq s, l = 1, 2, \cdots) \tag{B2.31}$$

where we have used the notation of Abramowitz and Stegun ((1964), page 835) for Stirling numbers of the second kind. Then

$$\begin{aligned}
\kappa_{\beta} &= \textstyle\sum_{j=1}^{s} T_{i,j,|\beta|} && \text{if } \beta_i = |\boldsymbol{\beta}| \text{ for some } i \leqq r\\
&= \textstyle\sum_{i=1}^{r} T_{i,j,|\beta|} && \text{if } \beta_{r+j} = |\boldsymbol{\beta}| \text{ for some } j\\
&= T_{i,j,|\beta|} && \text{if } \beta_i + \beta_{r+j} = |\boldsymbol{\beta}|\,,\ \beta_i > 0\,,\ \beta_{r+j} > 0\\
& && \text{for some } i \text{ and } j\ (i \leqq r)\\
&= 0 && \text{otherwise}\,,
\end{aligned} \tag{B2.32}$$

and

$$\frac{A_1(k)}{A_0(k)} = \frac{1}{8k} \textstyle\sum_{i=1}^{r} \sum_{j=1}^{s} T_{ij4}(k^{ii} + 2k^{i,r+j} + k^{r+j,r+j})^2 \tag{B2.33}$$

where $\mathbf{K}^{-1} = (k^{\mu\nu})$ $(\mu, \nu = 1, 2, \cdots, r+s-1)$.

When the column totals are flat, this reduces to

$$\frac{A_1(k)}{A_0(k)} = \frac{1}{8k} \textstyle\sum_{i=1}^{r} \sum_{j=1}^{s} T_{ij4} \left\{ \frac{(1-\rho_i)^2}{s\rho_i} + \frac{s-1}{s}\left[\sum_{i=1}^{r} \frac{\rho_i}{(1-\rho_i)^2}\right]^{-1} \right\}^2 . \tag{B2.34}$$

For $r \times r$ tables with all row and column totals equal to n we have

$$\frac{A_1(k)}{A_0(k)} = \frac{(2r-1)^2(6n^2 + 6nrk + r^2k^2)}{8r^2n(n+rk)k} . \tag{B2.35}$$

We cannot let $k \to \infty$ while keeping n constant because of the condition following (B2.13). It appears that k needs to exceed 10 for the second term of (B2.13) to be useful.

B3. Approximation to $A(k)$ for multidimensional tables. The approximation (B2.1) can be extended in a natural way to m-dimensional tables. For $m = 3$ we would assume by analogy with (B2.1) and (12.20) that, if k is not too small,

$$F_9(k) \approx F_5^{(1)}(k)F_5^{(2)}(k)F_5^{(3)}(k)/[F_1(k)]^2\,, \tag{B3.1}$$

where the notation is self-explanatory; and by analogy with (B2.2) and (12.18) that

$$\text{(B3.2)} \qquad \mathscr{C}\left(\prod w_h^{n_{h\cdot\cdot}} \prod x_i^{n_{\cdot i\cdot}} \prod y_j^{n_{\cdot\cdot j}}\right) \prod (1 - w_h x_i y_j)^{-k} \approx B(k)$$

where $k \geqq 1$ and

$$\text{(B3.3)} \qquad B(k) = \frac{\prod \binom{n_{h\cdot\cdot} + krs - 1}{n_{h\cdot\cdot}} \prod \binom{n_{\cdot i\cdot} + ksq - 1}{n_{\cdot i\cdot}} \prod \binom{n_{\cdot\cdot j} + kqr - 1}{n_{\cdot\cdot j}}}{\binom{N + kqrs - 1}{N}^2}.$$

The power in the denominator is $m - 1$.

REFERENCES

[1] Abramson, M. and Moser, W. O. J. (1973). Arrays with fixed row and column sums. *Discrete Mathematics* **6** 1–14.

[2] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions.* U. S. Government Printing Office, Washington, D.C.

[3] Aitken, A. C. (1956). *Determinants and Matrices*, 9th ed. Oliver and Boyd, Edinburgh.

[4] Aitkin, M. A. (1974). Simultaneous inference and the choice of variable subsets in multiple regression. *Technometrics* **16** 221–227.

[5] Anand, H., Dumir, V. C. and Gupta, H. (1966). A combinatorial distribution problem. *Duke Math. J.* **33** 757–769.

[6] Barton, D. E. (1958). The matching distributions: Poisson limiting forms and derived methods of approximation. *J. Roy. Statist. Soc. Ser. B* **20** 73–92.

[7] Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **25** 220–233.

[8] Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis.* M.I.T. Press, Cambridge.

[9] David, F. N. and Barton, D. E. (1962). *Combinatorial Chance.* Griffin, London.

[10] Fisher, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed. Oliver and Boyd, Edinburgh.

[11] Fisher, R. A. (1949). *The Design of Experiments.* Oliver and Boyd, Edinburgh.

[12] Fisher, R. A. (1956). *Statistical Methods and Scientific Inference.* Oliver and Boyd, Edinburgh.

[13] Good, I. J. (1950). *Probability and the Weighing of Evidence.* Griffin, London.

[14] Good, I. J. (1956). On the estimation of small frequencies in contingency tables. *J. Roy. Statist. Soc. Ser. B* **18** 113–124.

[15] Good, I. J. (1957). Saddle-point methods for the multinomial distribution. *Ann. Math. Statist.* **28** 861–881.

[16] Good, I. J. (1961). The multivariate saddle-point method and chi-squared for the multinomial distribution. *Ann. Math. Statist.* **32** 535–548.

[17] Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.

[18] Good, I. J. (1965 a). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* M.I.T. Press, Cambridge.

[19] Good, I. J. (1965 b). Categorization of classification. In *Mathematics and Computer Science in Biology and Medicine.* Her Majesty's Stationery Office, London.

[20] Good, I. J. (1967). A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B* **29** 399–431 (with discussion).

[21] Good, I. J. (1969). Statistics of language. In *Encyclopedia of Linguistics, Information, and Control* (A. R. Meetham, ed.) 567–581. Pergamon Press.

[22] GOOD, I. J. (1974). The Bayes factor against equiprobability of a multinomial population assuming a symmetric Dirichlet prior. *Ann. Statist.* **3** 246–250.

[23] GOOD, I. J. (1975). The number of hypotheses of independence for a random vector or for a multidimensional contingency table, and the Bell numbers. *Iranian J. Sci. Technology* **4** 77–83.

[24] GOOD, I. J. and CROOK, J. F. (1974). The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Assoc.* **69** 711–720.

[25] GOOD, I. J. and CROOK, J. F. (1976). The enumeration of arrays and a generalization related to contingency tables. *Discrete Math.* To appear.

[26] GOULD, H. W. (1971). Research bibliography of two special number sequences. *Math. Mongongaliae*, No. 12, 25 pages.

[27] GÛNEL, E. and DICKEY, J. (1974). Bayes factors for independence in contingency tables. *Biometrika* **61** 545–557.

[28] KENDALL, M. G. and STUART, ALAN (1973). *The Advanced Theory of Statistics* **3**, 3rd ed. Griffin, London.

[29] LEVINE, J. and DALTON, R. E. (1962). Minimum periods, modulo p, of first-order Bell exponential integers. *Math. Comp.* **16** 416–423.

[30] MACMAHON, P. A. (1915/1960). *Combinatory Analysis.* University Press, Cambridge, England.

[31] MUIR, T. (1933/1960). *A Treaties on the Theory of Determinants.* Constable, London.

[32] NATH, G. B. and IYER, P. V. K. (1972). Note on the combinatorial formula for ${}_nH_r$. *J. Austral. Math. Soc.* **14** 264–268.

[33] ROY, S. N. and MITRA, S. K. (1956). An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis. *Biometrika* **43** 361–376.

[34] SLOAN, N. J. A. (1973). *A Handbook of Integer Sequences.* Academic Press, New York.

[35] SOAL, S. G. and BATEMAN, F. (1954). *Modern Experiments in Telepathy.* Faber and Faber, London.

[36] SOKAL, R. R. and ROHLF, F. J. (1969). *Biometry: the Principles and Practice of Statistics in Biological Researh.* Freeman, San Francisco.

[37] STANLEY, R. P. (1973). Linear homogeneous diophantine equations and magic labellings of graphs. *Duke Math. J.* **40** 607–632.

[38] STANLEY, R. P. (1975). Generating functions. Unpublished typescript.

[39] STEIN, M. L. and STEIN, P. R. (1970). Enumeration of stochastic matrices with integer elements. Los Alamos Scientific Laboratory Report LA-4434.

[40] TOCHER, K. D. (1950). Extension of the Neyman–Pearson theory to discontinuous variates. *Biometrika* **37** 130–144.

[41] VAHLEN, K. TH. (1898–1904). Rationale Funktionen der Wurzeln, Symmetrische und Affektfunktionen. In *Enc. der Math. Wiss. Band* 1, 449–479. (Vahlen cites L. Crocchi *Gionarle di Matemat.* **17** (1879), 218–231, 380; **18** (1880), 377–380; and **20** (1882), 301–320.)

[42] WILKS, S. S. (1946). *Mathematical Statistics.* University Press, Princeton.

[43] WILKS, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

[44] WOOLF, B. (1955). On estimating the relation between blood groups and disease. *Ann. Human. Genetics* **19** 251–253.

[45] YATES, F. (1934). Contingency tables involving small numbers and the χ^2 test. *J. Roy. Statist. Soc. Suppl.* **1** 217–239.

DEPARTMENT OF STATISTICS
VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA 24061