The Annals of Statistics
1980, Vol. 8, No. 6, 1198–1218

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

BY J. F. CROOK[2] AND I. J. GOOD[2]

Winthrop College and Virginia Polytechnic Institute and State University

This paper is a continuation of a paper in the *Annals of Statistics* (1976), **4** 1159–1189 where, among other things, a Bayesian approach to testing independence in contingency tables was developed. Our first purpose now, after allowing for an improvement in the previous theory (which also has repercussions on earlier work on the multinomial), is to give extensive numerical results for two-dimensional tables, both sparse and nonsparse. We deal with the statistics X^2, Λ (the likelihood-ratio statistic), a slight transformation G of the Type II likelihood ratio, and the number of repeats within cells. The latter has approximately a Poisson distribution for sparse tables. Some of the "asymptotic" distributions are surprisingly good down to exceedingly small tail-area probabilities, as in the previous "mixed Dirichlet" approach to multinomial distributions (*J. Roy. Statist. Soc. B*, 1967, **29** 399–431; *J. Amer. Statist. Assoc.* 1974, **69** 711–720).

The approach leads to a quantitative measure of the amount of evidence concerning independence provided by the marginal totals, and this amount is found to be small when neither the row totals nor the column totals are very "rough" and the two sets of totals are not both very flat.

For Model 3 (all margins fixed), the relationship is examined between the Bayes factor against independence and its tail-area probability.

TABLE OF CONTENTS

15. Introduction. In what we shall call Part I of a study of contingency tables, Good (1976) developed, among other things, a Bayesian approach for testing the null hypothesis H of independence of rows and columns. One practical, as distinct from philosophical, merit of the

Received March 1978; revised July 1978.

[1] Some of the new theory applies also to earlier multinomial investigations. The work was supported in part by H.E.W., N.I.H. Grant #R01 GM 18770.

[2] Most of the extensive programming was due to J.F. Crook and most of the new theory to I.J. Good.

AMS 1970 *subject classifications*. Primary 62F15; secondary 62G10; tertiary 05A15.

Key words and phrases. Contingency tables, multidimensional contingency tables, mixtures of Dirichlet distributions, Bayes factor, Bayes/non-Bayes synthesis, enumeration of arrays, independence, information in marginal totals, type II likelihood ratio, number of repeats in a contingency table, Fisher's "exact test", multinomial significance, log-Cauchy hyperprior, hierarchical Bayes.

purely Bayesian approach is that it does not depend on large samples, but we also consider non-Bayesian and partly Bayesian methods. The work is based on the assumption that the initial state of knowledge is symmetrical with respect to the rows and also with respect to the columns. The distinctive feature of the approach is that under both the null and nonnull hypotheses the prior distributions of the marginal probabilities are mixtures of symmetric Dirichlet distributions (which had given interesting results for the multinomial situation in Good, 1965, 1967, and Good and Crook, 1974). A number of the notations and terms of Part I will be adopted for Part II without further explanation, especially n_{ij}, $n_{i.}$, $n_{.j}$, N, p_{ij}, $p_{i.}$, $p_{.j}$, r, s, H, $\bar{H}$, $A\{(n_{i.}), (n_{.j})\}$, X^2, Λ, k, $\phi(k)$, $A(k)$, F_1, F_2, $F_{(2)}$, F_3, G, and FRACT (the Bayes factor against H provided by the row and column totals alone). We shall, however, modify $\phi(k)$ in a manner proposed by Good (1979a), and this will imply $F_1 = F_2$. See equation (15.4) and the discussion that follows it.

For convenience of reference we start Part II with Section 15. The term $2N \log N$ should of course be added to the right side of formula (13.2), which is in Part I, as the number indicates.

In the present paper we shall be concerned only with two-dimensional contingency tables. We shall report numerical results for some of the formulae in Part I based upon more than 5000 contingency tables chosen more or less at random but by no means haphazardly. We also include much new theoretical material.

Irrespective of one's approach to testing H there are three familiar procedures for sampling a contingency table, which we call Models 1, 2, and 3. In Model 1 the sample is taken at random from the population, in Model 2 the row (or column) totals are fixed before taking the sample, and in Model 3 both the row and column totals are fixed. The Bayes/non-Bayes compromise (or synthesis) used by Good (1967) for the equiprobable multinomial hypothesis, and further extended by Good and Crook (1974), is available in its simplest form only under Model 3 as explained in Section 16. For this and other reasons our primary concern will be with this model although Models 1 and 2 occur more frequently in practice. Model 3 does occur in practice, though at first sight this seems surprising, but the main practical justification for our preoccupation with it, apart from its considerable theoretical interest, depends not on its frequency of occurrence but on the following argument.

We find that FRACT, which is equal to F_1/F_3, depends on whether the row totals and column totals are rough or flat as measured, for example, by $r\sum n_{i.}^2/N^2$ and $s\sum n_{.j}^2/N^2$ where by convention we usually assume that no marginal total is zero. FRACT seems usually to lie between ½ and 2½ when neither of the two sets of marginal totals is very rough and the two sets are not both very flat. (One of us (Good, 1979a) has, however, conjectured that FRACT is not identically equal to 1 under *any* sensible Bayesian model.) When using Model 1 or 2 this condition will usually be satisfied in practice. As Jeffreys (1961, page 432) implies, a Bayes factor of 3 is of little importance. This judgement of Jeffreys, combined with the inequality for FRACT, to some extent justifies those statisticians who, following Fisher (1938, page 102; and 1956, pages 87–88), are prepared to assume FRACT ≈ 1 when the sampling is according to Model 1 or 2. (Fisher, 1935, page 48 is less dogmatic.)

The asymptotic distribution of X^2, given H, is the same for all three sampling models, as mentioned in Section 1 where references are given. But, from the point of view of both the Bayes and Neyman-Pearson theories, the use of X^2 alone strictly requires the assumption that the marginal totals contain no evidence for or against the null hypothesis (an assumption that is valid when sampling by Model 3). The use of Fisher's "exact test" and of our statistic G (the criterion based on the Type II Likelihood Ratio) also strictly depend on this assumption. If any of these three statistics is used when the sampling is by Model 1 or 2, then the (Bayesian) judgement is required that the evidence from the marginal totals is negligible.

A statistician using Model 1 or 2 and who is prepared to adopt a purely Bayesian approach, will of course not need to concern himself with the value of FRACT since he can simply calculate F_1 or F_2 (which are equal).

The Bayes factor F_3 against H is much more difficult to compute than the Bayes factors F_1 and F_2, and the computation can become impracticable. To meet this difficulty we are forced to investigate methods of approximating F_3 since Model 3 does occur. The problem is largely solved because of the inequality $\frac{1}{2} < F_1/F_3 < 2\frac{1}{2}$ which, as we just mentioned, is usually true in practice.

The formula for each of F_1, F_2 (and $F_{(2)}$), F_3, and $F_3(k)$, derived in Part I, can be restated in a form that is often more convenient for our computer programs when the marginal totals are not too large. For example,

$$F_3 = \frac{N! \int_0^\infty \frac{\prod_{i,j} \zeta(k, n_{ij})}{\zeta(rsk, N)} \phi(k)\, dk}{\prod_i n_{i\cdot}! \prod_j n_{\cdot j}! \sum^* \frac{1}{m_{ij}!} \int_0^\infty \frac{\prod_{i,j} \zeta(k, m_{ij})}{\zeta(rsk, N)} \phi(k)\, dk} \tag{15.1}$$

where

$$\zeta(\kappa, m) = \prod_{h=0}^{m-1} (h + \kappa) \qquad \text{if } m \geq 1,$$
$$= 1 \qquad \text{if } m = 0;$$

and $\sum^*$ denotes a summation over all tables (m_{ij}) having the assigned marginal totals. (The formulae involving gamma functions, derived from (2.7) and (5.1) to (5.4), and the function $A(k)$ of (5.6), are also of value.) The formula for $F_3(k)$ is obtained by taking $\phi(k)$ as a Dirac delta function in the expression for F_3. Thus

$$F_3(k) = \frac{N! \prod_{i,j} \zeta(k, n_{ij})}{\prod_i n_{i\cdot}! \prod_j n_{\cdot j}! \sum^* \left\{ \prod_{i,j} [\zeta(k, m_{ij})/m_{ij}!] \right\}}. \tag{15.2}$$

We shall need this formula in Section 16 on the Bayes/non-Bayes compromise.

Other matters discussed in this paper are (i) the distributions of G, X^2, and Λ both for nonsparse and for sparse tables; (ii) the statistic

$$R = \tfrac{1}{2} \sum_{ij} n_{ij}(n_{ij} - 1) \tag{15.3}$$

for sparse tables; (iii) the numerical analysis of FRACT, giving the support for our previous remarks on this subject; and (iv) the relationship between F_3 and its tail-area probability.

The statistics that we consider have various advantages and disadvantages. The purely Bayesian statistics have the advantages that (i) they allow explicitly for which of the three sampling models is used; (ii) they do not depend on asymptotic theory when used in a strictly Bayesian manner (cf. Good, 1967, page 400), and therefore when used in this manner they can be used for any sample size; (iii) they are likely to have good power because their construction depends on a precise formulation of the nonnull hypothesis (cf. Good and Crook, 1974, page 711); (iv) they lead to an indication of how much evidence concerning independence is contained in the marginal totals; and they have the disadvantages of (v) being at present more controversial than X^2 and (vi) requiring more calculation. The statistic G has the advantage of being based on Bayesian soil yet having a distribution that can be well approximated in nearly all circumstances (see Section 17). The statistics X^2, Λ, and R are all easy to calculate. The asymptotic theory for X^2 is good for tables that are not too sparse, and that for R is good for tables that are sparse enough. The asymptotic approximation for the tail-area probabilities of Λ are good for tables that are not sparse, and its derivation depends on Model 1 or 2 so that it can be used for these models without assuming that the evidence from the marginal totals is negligible, unlike X^2, G, and R. Some of these comments will be made more quantitative in Sections 17, 18, and 19.

We now give the reason for replacing (2.4) by (15.4):

$$\phi(k) = \phi(t, k) = \frac{1}{k\{\pi^2 + [\log_e(kt)]^2\}}, \tag{15.4}$$

where t is the number of categories in the corresponding Dirichlet distribution. (The argument was given briefly by Good, 1979a,) For example, in (15.1), $t = rs$.

If the symmetric Dirichlet density of t categories is denoted by $D(t, k)$, we know that the prior for the innards, given $\bar{H}$, is, for Model 1,

$$\int_0^\infty D(rs, k)\phi(rs, k)\, dk$$

if ϕ is permitted to be a function of both k and t, where t denotes the number of categories. Therefore, by the lumping property of the Dirichlet distribution, the prior for $(p_{i\cdot})$ is

$$\int_0^\infty D(r, sk)\phi(rs, k)\, dk = \frac{1}{s}\int_0^\infty D(r, l)\phi(rs, l/s)\, dl.$$

Now the row totals form a multinomial sample, and this sample could have been taken before we had decided to sample the contingency table. It is convenient for the statistician, and we believe it is usually a very good approximation, if the prior assumed for this multinomial is unaffected when we decide to sample a contingency table. If we insist on this convenient feature, then the prior must be independent of s. Therefore $\phi(rs, l/s)/s$ must be mathematically independent of s. Therefore (assuming differentiability of ϕ, though continuity may well be sufficient for the conclusion we shall soon reach) we have $rs^{-1}\phi_1 - s^{-2}\phi - ls^{-3} = 0$, where ϕ_1 and ϕ_2 denote the partial derivatives of ϕ with respect to its first and second arguments. Therefore $rs\phi_1 - ls^{-1}\phi_2 = \phi$. Let $rs = x$ and $ls^{-1} = y$. Then $x\phi_1 - y\phi_2 = \phi$. Let $\rho(u, v) = e^v\phi(e^u, e^v)$; then $\rho_1 = \rho_2$. Change to new independent variables $\xi = u + v$, $\eta = u - v$; then $\partial\rho/\partial\eta = 0$, that is, ρ is a function of $u + v$ alone and therefore $\phi(t, k)$ is of the form $\psi(tk)/k$, as in (15.4).

Note that when the distribution of $(n_{i.})$ does not depend on s, the row totals by themselves give no evidence for or against H, and the corresponding Bayes factor $F_1/F_2 = 1$, so that $F_1 = F_2$. Similarly, $F_1 = F_{(2)}$. Fisher usually assumed that FRACT = 1 and our implicit assumption that the row totals alone contain negligible evidence concerning H is a much weaker assumption.

The above argument has repercussions on the appropriate prior for any multinomial, *even in problems containing no reference to contingency tables.* Hence some adjustments are required to the results of Good (1967), and Good and Crook (1974), but we believe the adjustments will be small because the numerical adjustments in the present work are small when (2.4) is replaced by (15.4).

We wanted ϕ to behave approximately like the Jeffreys-Haldane density $1/k$ while maintaining propriety, and this was achieved by taking ϕ of the log-Cauchy form. The median and quartiles of the particular log-Cauchy density defined by (15.4) are $1/t$ and $e^{\pm\pi}/t$. It may be recalled that Perks (1947) first proposed the flattening constant $1/t$ (for multinomials) because, being inversely proportional to t, it leads to a desirable invariant property under pairing off of the t categories to reduce the number of categories to $\frac{1}{2}t$. He determined the constant of proportionality by taking $k = \frac{1}{2}$ for $t = 2$, that is, for $t = 2$ he adopted what is known as the Jeffreys-Perks invariant prior. Although Perks's suggestion proved to be untenable (Good, 1967, page 412; Perks, 1967, page 266), we have now arranged to make the median value of k equal to $1/t$ and the quantiles proportional to $1/t$. This resurrects Perks' idea at a "higher level". The upper quartile e^{π}/t seems to us to be at a reasonable place, but of course the more general log-Cauchy hyperprior

$$\frac{\lambda}{k\pi\{\lambda^2 + [\log(kt/\mu_1)]^2\}}, \tag{15.5}$$

whose median is $\mu = \mu_1/t$ and whose quartiles are $\mu e^{\pm\lambda}$, could be entertained. This density has no turning points if $\lambda > 1$, and we shall never refer to a value of $\lambda \le 1$. The nth percentile is

$$\mu(q/\mu)^{-\cot(\pi n/100)} \tag{15.6}$$

where q is the upper quartile of (15.5) (Good, 1969b, pages 45–46). The user, in any given application, could guess values for some quantiles of the flattening constant k so as to choose the hyperhyperparameters μ and λ. Good (1952, page 114) suggested a rough principle of "Bayesian robustness" for hierarchical Bayesian models and it is exemplified in Table 1 where the values of F_1 are given, for the 3×3 table with innards (2, 3, 0; 0, 1, 4; 0, 0, 5), corresponding to sixteen pairs (μ, q). The entry 12.6, where $\mu_1 = 1$ and $\lambda = \pi$, corresponds to (15.4). This table shows, for example, that large proportional changes in what could be the judged values of the median and upper quartile lead to very much smaller proportional changes in F_1.

For the main conclusions see Section 22 first.

16. The Bayes/non-Bayes compromise or synthesis. The idea of the Bayes-Fisher compromise is to use a Bayesian or somewhat Bayesian statistic to test a null hypothesis in a Fisherian manner. A discussion of the relationship between the compromise and Fisherian methods of hypothesis testing is given by Good and Crook (1974). If the null hypothesis H is a simple statistical hypothesis there is no difficulty in interpreting what is meant by the Bayes-Fisher compromise. If H is composite, then one could effectively "convert" it into a simple statistical hypothesis by assuming a prior distribution over the simple component hypotheses of which it is a logical disjunction. This procedure leans more towards a Bayesian philosophy than when H is simple, which is the case with which the present paper is concerned. For Model 3, H is indeed a simple statistical hypothesis because the probability of (n_{ij}) is then the Fisher-Yates probability

$$P\{(n_{ij}) \mid (n_{i.}), (n_{.j}), H\} = \prod n_{i.}! \prod n_{.j}!/(N! \prod n_{ij}!). \tag{16.1}$$

[This conditional probability is correct for Models 1 and 2 also, but then the marginal totals also contain some evidence about H.] It is especially convenient in this case to make use of the Type II likelihood-ratio statistic because H is a case ($k = \infty$) of the nonnull hypothesis, although it is on the boundary of the parameter space. The use of this statistic is a compromise between Bayesian, Fisherian, and Neyman-Pearsonian methods.

The Type II likelihood ratio for Model 3 is

$$\frac{\max_k P\{n_{ij}) \mid (n_{i.}), (n_{.j}), \bar{H}(k)\}}{P\{(n_{ij}) \mid (n_{i.}), (n_{.j}), H\}} = \max_k F_3(k), \tag{16.2}$$

where $\bar{H}(k)$ denotes a simple component of $\bar{H}$, and $F_3(k)$ is given by (15.2). We write, as in Part I,

$$G = \{2 \log_e \max_k F_3(k)\}^{1/2},$$

and we also consider the statistics X^2, Λ, and R.

It was conjectured in Section 10 that $F_3(k)$ as a function of k has at most one local maximum, and this conjecture has been well corroborated numerically and graphically. Accepting this, it can be proved that $\max_k F_3(k) = 1$ (attained at $k = \infty$) if and only if $X^2 \leq$

TABLE 1

Values of F_1 for the 3×3 table (2, 3, 0; 0, 1, 4; 0, 0, 5), depending on the hyperhyperparameters μ and q of the log-Cauchy distribution ($\mu = \mu_1/(rs) = \mu_1/9$; $q = \mu e^\lambda$)

The entry 12.6 corresponds to (15.4)

	λ	$\pi/2$	π	$3\pi/2$	2π
μ	q	0.53	2.6	12.4	59
1/36		27.9	15.8	11.1	8.8
1/18		24.5	14.2	10.2	8.2
1/9		21.1	12.6	9.3	7.5
2/9		17.7	11.0	8.3	6.3
4/9		14.0	9.4	7.4	

$(r-1)(s-1)$ (compare Good, 1965, page 37; Good, 1974; and Levin and Reeds, 1977). In this case $G = 0$, so that $P(G = 0) = P\{X^2 \le (r-1)(s-1)\} \approx 1 - c_\tau$ where $\tau = (r-1)(s-1) + 1$ and c_τ is the probability that χ^2 with $\tau - 1$ degrees of freedom exceeds $\tau - 1$. The values of c_τ, given in a table in Good (1967, page 411) increase from 0.317 when $\tau = 2$ to 0.5 when τ is large.

Wilks's theorem (1938) suggests that the asymptotic distribution of G^2 should be a chi-squared distribution with one degree of freedom, that is, that that of G is a standardized normal distribution, folded about its expectation. But, in view of the above comments combined with the analogous results for the multinomial, it is more natural to assume (cf. Section 10) that the exact tail-area probability $P(G)$, is approximated by $Q(G)$ where

$$Q(G) = c_\tau(2/\pi)^{1/2} \int_G^\infty e^{-1/2y^2}\, dy \qquad G > 0.$$

It turns out, however, though we do not know why, that after all $Q^*(G)$, defined as $c_\tau^{-1}\, Q(G)$, approximates $P(G)$ better in the extreme tail of the distribution. [A similar effect occurred in the multinomial results of Good and Crook (1974), but only when nearly all N "objects" fell into a single cell.] The theory of Wilks (1938) is not strictly applicable, because $k = \infty$ is on the boundary of the hyperparameter space, as we said before. A similar difficulty applies in the analogous multinomial problem in spite of the discussion in Good (1967, page 411).

Now Chernoff (1954) generalized Wilks's work to cover the case where the parameter lies on the boundary of the parameter space. He showed, under certain regularity conditions, that the asymptotic distribution of minus twice the natural logarithm of the likelihood ratio is that of a random number "which is zero half the time and behaves like χ^2 with one degree of freedom the other half of the time". The meaning of "asymptotic" in our application is that $rs \to \infty$ (while presumably $N/(rs)$ must not tend to zero). To see this, note that the likelihood $P\{(n_{ij}) \mid (n_{i.}), (n_{.j}), \bar{H}(k)\}$ is expressible in the form $\prod_{ij} f(n_{ij})$, as is required in the theory of Wilks and of Chernoff, because it is equal to

$$\prod_{ij} \frac{(n_{ij} + k - 1)^{(n_{ij})}}{n_{ij}!} \Big/ \Sigma^* \prod_{ij} \frac{(m_{ij} + k - 1)^{(m_{ij})}}{m_{ij}!}$$

in which the denominator does not depend on (n_{ij}). [Here $x^{(m)} = x(x-1)\cdots(x-m+1)$ and $x^{(0)} = 1$.]

When rs is not large we know that $P(G = 0)$ is $1 - c_\tau$ and we have already mentioned that this tends to ½ when $\tau \to \infty$, that is, when $rs \to \infty$. When τ is not large, Chernoff's theory could presumably be improved by introducing c_τ instead of ½. Our results appear to verify this presumption.

The above discussion of the asymptotic behavior of G applies also to the problem of significance tests for multinomials mutatis mutandis (Good, 1967; Good and Crook, 1974).

Before giving numerical results we note one advantage that G has over F_3 when we wish to calculate them both rather accurately. The advantage is that the calculation of F_3 requires that $F_3(k)$ or $A(k)$ be calculated for enough values of k to perform an approximate integration, whereas to calculate G (or max $F_3(k)$) we need to calculate $F_3(k)$ for only a few values of k, when $F_3'(k) = 0$ is solved by the Newton-Raphson method. The user of purely Bayesian methods should also note that the Bayes factors F_1 and F_2 are much easier to calculate than is F_3 and that Models 1 and 2 occur more often than Model 3.

17. The approximate distributions of G, X^2, and Λ, for nonsparse tables. Let S denote any of the three statistics G, X^2, or Λ; let $\tilde{S}$ denote the corresponding random number; and let $P(\tilde{S} \ge S)$, or $P(S)$ for short, be the exact right-hand tail-area probability of S. Let $Q(\tilde{S} \ge S)$ or $Q(S)$ be the approximation to $P(S)$, except that, as discussed above, we have two approximations, $Q(G)$ and $Q^*(G)$, for $P(G)$. The approximations $Q(X^2)$ and $Q(\Lambda)$ are the usual chi-squared asymptotic approximations. Furthermore, let the *error ratio* be the larger of $P(S)/Q(S)$ and $Q(S)/P(S)$. We shall use this ratio to measure the merit of the four approximate tail-area probabilities. Had it turned out that either $P(S)/Q(S)$ or $Q(S)/P(S)$ was always greater than 1 our measure would have been somewhat stronger, but this did not usually happen. [We did find, however, that, in a sample of 850 tables with distinct frequency counts, $P(G) > Q(G)$ 76% of the time while $Q^*(G) > P(G)$ 82% of the time. This suggests that $\{Q(G)Q^*(G)\}^{1/2}$ might be investigated but we have not done so.]

The exact distribution of S under Model 3 can be computed, assuming the null hypothesis, by tabulating the Fisher-Yates probability of each possible contingency table with the assigned marginal totals and then cumulating these probabilities in the appropriate manner.

Except in special cases, the cost of computing the exact distributions of the statistics can be great when either r, s, or the marginal totals are large, so in general we chose to keep rs no more than 16 and N no more than 36 (but see Section 18 and Table 10). Within these constraints we selected r, s, and N and then selected sets of row and column totals by each of the three sampling methods described in Appendix D, which are called Sampling Methods A, B, and C and should not be confused with Models 1, 2, and 3.

Table 2 shows the number of contingency-table margins selected for our contingency-table sample by Sampling Methods A, B, and C. In Section 18, the average cell entry E will play a role in our discussion of the merit of the approximation to the distribution of each statistic. [Another kind of average of the cell expectations is of interest, although we thought of it too late for inclusion in our computer programs. It is defined by

$$D = -\log\left\{\frac{1}{rs}\sum_{i,j}\exp(-n_{i.}n_{.j}/N)\right\}. \tag{17.1}$$

This is an average that gives the greatest weight to the small values of $n_{i.}n_{.j}/N$ and appears to be a better measure than E for the sparseness of a table. When all the expectations $n_{i.}n_{.j}/N$ are appreciably less than 1, then D is approximately equal to E.]

Let a *group* of contingency tables consist of all those tables whose "innards" sum to any fixed one of the sets of marginal totals in our sample. For each of the 504 groups of contingency tables in our sample, a program computed (i) the exact and approximate distributions of the statistics, and (ii) for each contingency table, the error ratio, which is max $[P(S)/Q(S), Q(S)/P(S)]$. The program then placed into the same set those error ratios that had an approximate tail-area, Q or Q^*, in the interval $[10^{-\nu}, 10^{-\nu+1})$ ($\nu = 1, 2, 3, \cdots$). Finally, the program categorized the error ratios for each ν according as they were in the interval [1, 2), [2, 5), [5, 10), or [10, ∞).

TABLE 2
Composition of the sample of contingency tables

The average value of E over the 504 groups of tables is 2.8 and the range is (0.75, 6.00)

$r \times s$	N	Number of groups of Contingency Tables Selected			
		Sampling Method A	Sampling Method B	Sampling Method C	Total
2 × 3	20	100	40	1	141
2 × 4	20	100	40	1	141
2 × 5	20	50	20	1	71
2 × 6	20	0	0	1	1
2 × 3	36	20	20	0	40
3 × 3	20	20	20	0	40
3 × 3	15	0	10	1	11
3 × 3	18	0	10	1	11
3 × 3	21	0	10	1	11
3 × 3	24	0	10	1	11
3 × 3	27	0	0	1	1
4 × 4	12	19	6	0	25
		309	186	9	504

Table 3 shows, for all the groups in our sample, the percentages of error ratios in the intervals given above whenever $10^{-\nu} \leq Q(\tilde{S} \geq S) < 10^{-\nu+1}$. This is shown for each of the approximations $Q^*(G)$, $Q(G)$, $Q(X^2)$, and $Q(\Lambda)$. For each approximation we show also the total number of error ratios for each value of ν. This number is also the number of distinct values of each statistic for a given value of ν. For example, for $Q^*(G)$ the number of error ratios when $\nu = 1$ is 410. Of these 410 error ratios 30% were less than 2, 70% were between 2 and 5 and none of them were greater than 5. For all of the statistics the sample of error ratios for values of ν greater than 8 was too small. Hence, our remarks are restricted to values of ν less than 9.

We shall give evidence in Section 18 that the accuracy of the approximate distribution depends upon E. Hence any rough conclusions that can be drawn from Table 3 about whether $Q^*(G)$, $Q(G)$, $Q(X^2)$, or $Q(\Lambda)$ approximates its tail-area the best are conditional upon the average value of E for the sample of contingency tables, which is 2.8. Within this sample the range of E was not large enough to detect any interesting dependence upon E.

One of the most striking facts that emerges from Table 3 is that, of the 4571 distinct values of $Q^*(G)$, not one of them had an error ratio that was more than 10, whereas, for small tail-area probabilities, all of the other approximate distributions did have some error ratios that were at least 10. But of the three approximations with some error ratios that were at least 10, $Q(\Lambda)$ had the least proportion on the average.

To obtain a single measure of the accuracy of each statistic's approximate distribution, for ν given, we compute a score, denoted by $\delta[Q(\tilde{S} \geq S) \mid \nu]$, by assigning weights to each of the intervals $[1, 2)$, $[2, 5)$, $[5, 10)$, and $[10, \infty)$. The weights we choose are 4, 2, 1, and 0 respectively. That is, we regard it as twice as good for the error ratio to be in the interval $[1, 2)$ as in the interval $[2, 5)$; four times as good in $[1, 2)$ as in $[5, 10)$; and of no value in $[10, \infty)$. A user could of course select his own set of weights. We write

$$\delta[Q(S) \mid \nu] = \tfrac{1}{4}\left(\textstyle\sum_i w_i m_i / \sum_i m_i\right) \qquad (i = 1, 2, 3, 4), \tag{17.2}$$

where m_1 is the number of error ratios in the interval $[1, 2)$ for ν given, m_2 is the number in the interval $[2, 5)$, m_3 is the number in $[5, 10)$, and m_4 is the number in $[10, \infty)$; and $w_1 = 4$, $w_2 = 2$, $w_3 = 1$, and $w_4 = 0$. As thus defined, δ has the merit of belonging to the interval $[0, 1]$, both endpoints being attainable. δ was computed with the help of Table 3.

Table 4 shows the scores of the approximate distributions for $\nu = 1(1)8$ and their rank orderings based on the scores. For example, when $\nu = 1$, $\delta[Q(G) \mid 1] > \delta[Q(\Lambda) \mid 1] > \delta[Q(X^2) \mid 1] > \delta[Q^*(G) \mid 1]$ which gives the rank ordering $Q(G) \succ Q(\Lambda) \succ Q(X^2) \succ Q(G^*)$, where the curly "greater than" symbol means "is preferable to". (The preference orderings are expressed in the table by means of the list $Q(G)$, $Q(\Lambda)$, $Q(X^2)$, $Q^*(G)$.) We see then that $Q(G)$ is a better approximation than the others for $1 \leq \nu \leq 4$, $Q(X^2)$ is the worst for $3 \leq \nu \leq 8$, and $Q(\Lambda)$ or $Q^*(G)$ is the best for $5 \leq \nu \leq 8$. Provided that a user accepts our values for the weights as appropriate he could use these rank orderings, for example, in the following way: suppose that for a contingency table that arises in practice $Q(\Lambda) = 1.5 \times 10^{-5}$. The exponent -5 implies that $\nu = 5$. Suppose also that $\nu = 5$ for the other approximate distributions. Then, since, for $\nu = 5$, $Q(\Lambda)$ has a higher score than any of the other approximations, the user can conclude that the approximate tail-area of any one of the other statistics would *probably* not be as good an approximation of its exact tail-area.

The recommendations that are implicit in Table 4 do not take into account (i) the convenience of always using the same statistic (in which connection note that $Q^*(G)$ is always a reasonable approximation for $1 \leq \nu \leq 8$), (ii) the costs of calculation (which seem to halve about every eighteen months), (iii) questions of power, and (iv) the possibility of sparse tables. Sparse tables are discussed in the next section.

Bayesian methods necessarily make implicit or explicit reference to the behavior of statistics when the null hypothesis is false. Therefore we conjecture that in general the statistic G will lead to more *powerful* significance tests than X^2 or Λ, even when $P(G)$ is replaced by $Q(G)$ or $Q^*(G)$. But we did not examine empirical evidence for this conjecture because our project was large enough already. We have begun to examine this question and we hope to report the results later.

TABLE 3

Percentages of error ratios which were in the intervals [1,2), [2, 5), [5, 10), and [10, ∞) *whenever* $10^{-\nu} \leq Q(\tilde{S} \geq S) < 10^{-\nu+1}$ *for the sample of contingency tables*

The percentages can be "improved" by smoothing but we have preferred to present the results as they occurred. [Owing to incomplete planning we had to repeat our runs for $\nu = 1$. Otherwise the first two totals, 4571 and 5371, would have been equal.]

	Where $Q(\tilde{S} \geq S) = Q^*(G)$					Where $Q(\tilde{S} \geq S) = Q(G)$					Where $S = X^2$					Where $S = \Lambda$				
ν \ Error ratio intervals	[1, 2)	[2, 5)	[5, 10)	[10, ∞)	Total	[1, 2)	[2, 5)	[5, 10)	[10, ∞)	Total	[1, 2)	[2, 5)	[5, 10)	[10, ∞)	Total	[1, 2)	[2, 5)	[5, 10)	[10, ∞)	Total
1	30	70	0	0	410	98	2	0	0	276	97	3	0	0	421	87	13	0	0	385
2	35	65	0	0	1999	93	6	1	0	2231	78	20	2	0	2350	69	30	1	0	1975
3	54	46	0	0	1121	82	15	3	0	1458	47	33	12	8	1319	54	43	2	1	1586
4	56	42	2	0	551	63	29	7	1	732	30	35	12	23	587	55	41	3	1	824
5	46	47	7	0	279	44	26	22	8	386	15	47	24	14	180	51	46	3	0	353
6	38	27	35	0	138	31	32	19	18	151	11	27	33	29	63	39	51	9	1	125
7	45	39	16	0	31	27	16	13	44	86	9	6	9	76	34	30	54	7	9	54
8	44	31	25	0	16	17	33	33	17	18	13	6	0	81	16	19	67	14	0	21
9	16	42	42	0	12	7	20	33	40	15	100	0	0	0	1	0	62	38	0	13
10	0	50	50	0	10	0	0	17	83	12	100	0	0	0	1	0	0	91	9	11
11	0	0	100	0	3	0	0	0	100	3	0	100	0	0	1	0	0	50	50	4
12	0	0	100	0	1	0	0	0	100	3					0	0	0	100	0	1
					4571					5371					4973					5352

TABLE 4

Scores of the approximate distributions for the sample of contingency tables. The last column gives the approximate probabilities in the rank order of their scores

Tables 3 and 4 do not apply to sparse contingency tables.

ν	$\delta[Q^*(G) \mid \nu]$	$\delta[Q(G) \mid \nu]$	$\delta[Q(X^2) \mid \nu]$	$\delta[Q(\Lambda) \mid \nu]$	Rank Orderings
1	.65	.99	.87	.94	$Q(G), Q(\Lambda), Q(X^2), Q^*(G)$
2	.67	.96	.88	.84	$Q(G), Q(X^2), Q(\Lambda), Q^*(G)$
3	.77	.90	.66	.77	$Q(G), Q^*(G), Q(\Lambda), Q(X^2)$
4	.78	.79	.50	.77	$Q(G), Q^*(G), Q(\Lambda), Q(X^2)$
5	.71	.62	.44	.75	$Q(\Lambda), Q^*(G), Q(G), Q(X^2)$
6	.61	.52	.33	.66	$Q(\Lambda), Q^*(G), Q(G), Q(X^2)$
7	.69	.38	.14	.58	$Q^*(G), Q(\Lambda), Q(G), Q(X^2)$
8	.66	.54	.16	.56	$Q^*(G), Q(\Lambda), Q(G), Q(X^2)$

18. Sparse tables. A sparse table may be defined as one containing a high proportion of cells for which the expectation $n_{i.}n_{.j}/N$ is small (say less than 0.5). Examples of sparse tables occur, for example, in Pearson (1905) (reproduced in Good, 1956), Eck (1961) (reproduced in Good, 1965, page 55), in Bishop, Fienberg, and Holland (1975), pages 191, 203, 326, 341, and in numerous texts dealing with the statistics of language, a guide to which is given by Good (1969a). Note that multidimensional tables are especially prone to be sparse. We regard this as an important forward-looking reason for an interest in sparse tables, although the reason is indirect because here we consider only ordinary (two-dimensional) contingency tables. In principle our methods can be extended to more dimensions, as explained in Part I. We shall concern ourselves mainly, in Sections 18 and 19, with those sparse contingency tables having a small value of E.

To investigate which of the approximate distributions is the best approximation for sparse contingency tables, we were forced to look at marginal totals that had special properties. Otherwise, by the time rs is large enough, for N fixed, to permit both a small value of E and small tail-area probabilities, it becomes impracticable to compute the exact distributions. (For tail-area probabilities greater than about 1/100, Monte Carlo methods would be practicable, but we have not yet tried this approach.)

Using one of the results of Good and Crook (1977) for square contingency tables that have all their marginal totals equal to 2, which can be conveniently called the all-twos case, we can deduce formulae for the exact distributions of G, X^2, and Λ. With these formulae one does not have to run through all possible $r \times r$ tables that have their marginal totals equal to 2 in order to compute the exact distributions.

When $r = s$ and all marginal totals are equal to n, where $n > 2$, it is practicable to evaluate the extreme right-hand tails of the distributions of X^2, Λ, and G. This can be done by means of specialized combinatorial arguments. We have considered the case $n = 3$ in some detail and our results are discussed at the end of the present section.

The number of tables having all marginal totals equal to 2, and having exactly t internal twos, is

$$\binom{r}{t}\binom{r}{t}t!\, A^*\{2, (r-t) \times (r-t)\}, \tag{18.1}$$

where $A^*(2, u \times u)$ is the number of $u \times u$ tables with all marginal totals 2 and with all interior entries equal to 0 or 1. We can deduce the value of $A^*(2, u \times u)$ from the formula

$$A(k, 2, u \times u) = (u!)^2 2^{-u} \textstyle\sum_{s=0}^{u} k^{u+s}\binom{2s}{s}[(u-s)!2^s]^{-1} \tag{18.2}$$

by putting $k = -1$ (Good and Crook, 1977, Formula (3.5)). Upon multiplying (18.1) by the Fisher-Yates probability we see that the probability of obtaining an interior of t twos and $2(r - t)$ ones is

$$\frac{(r!)^2 2^{2r-t} A^*\{2, (r-t) \times (r-t)\}}{[(r-t)!]^2 t!(2r)!}. \tag{18.3}$$

We used this formula in an obvious way to compute the exact distributions of G, X^2 and Λ. (The above theory for the all-twos case is an example of a more general approach mentioned in the penultimate paragraph of Section 20.)

We computed these exact distributions, the approximate tail-areas $Q^*(G)$, $Q(G)$, $Q(X^2)$, and $Q(\Lambda)$, and the error ratios. For $r = 3, 4, \cdots, 25$, Table 5 shows the largest values of ν for which the error ratio is less than 2, 5, and 10 *whenever* $10^{-\nu} \le Q(\tilde{S} \ge S)$. The results for $Q(G)$ are not given in the table because in every case $Q^*(G)$ approximated $P(G)$ more closely.

When $r = 15$, for example, the error ratio of $Q(X^2)$ is always less than 2 for $1 \le \nu \le 12$, less than 5 for $\nu = 13$ and 14, and less than 10 for $15 \le \nu \le 17$.

We also show, for each value of r in the table, the value of E, which is $2/r$, and the minimum value of $P(\tilde{S} \ge S)$, which is the most extreme exact tail-area probability that occurs for that value of r and is the same for all of the statistics. This most extreme case occurs when the innards of the contingency table consist of r twos arranged like rooks not mutually "en prise", in other words when the table resembles twice a permutation matrix.

From Table 5 we can see that $Q^*(G)$ approximates $P(G)$ within a factor of 2 almost down to the most extreme tail irrespective of the value of E. For example, (i) when $E = 0.182$, which is when $r = 11$, $Q^*(G)$ approximates $P(G)$ within a factor of 2 so long as $\nu \le 11$; (ii) when E

TABLE 5

Largest values of ν for which the error ratio is less than 2, 5, *and* 10, *whenever* $10^{-\nu} \le Q(\tilde{S} \ge S)$, *for* $n = 2$, $N = 2r$, *and* $r = 3(1)25$

An asterisk means that there is no such value of ν. (See also the text regarding Λ.) 6.7 (−2) for example, means 6.7×10^{-2}

	Where $Q(S) = Q^*(G)$			Where $S = X^2$			Where $S = \Lambda$				
r \ Error Ratio Intervals	2	5	10	2	5	10	2	5	10	E	Smallest $P(\tilde{S} \ge S)$
3	*	2	2	*	2	2	*	*	2	.667	6.7 (−2)
4	*	3	3	*	3	3	3	3	3	.5	9.5 (−3)
5	3	4	4	4	4	4	*	2	3	.4	1.1 (−3)
6	4	4	5	4	4	4	*	*	*	.333	9.6 (−5)
7	5	5	6	7	7	7	*	*	*	.286	7.4 (−6)
8	7	7	8	3	5	5	*	*	*	.25	4.9 (−7)
9	8	8	9	3	8	8	*	*	*	.222	2.9 (−8)
10	9	9	10	4	9	9	*	*	*	.2	1.5 (−9)
11	11	11	12	4	11	11	*	*	*	.182	7.3 (−11)
12	12	13	13	4	12	12	*	*	*	.167	3.2 (−12)
13	13	14	14	4	13	14	*	*	*	.154	1.3 (−13)
14	15	15	15	5	15	15	*	*	*	.143	4.7 (−15)
15	16	16	16	12	14	17	*	*	*	.133	1.6 (−16)
16	18	18	18	11	14	16	*	*	*	.125	5.2 (−18)
17	19	19	19	11	12	14	*	*	*	.118	1.6 (−19)
18	21	21	21	11	13	14	*	*	*	.111	4.5 (−21)
19	23	23	23	9	13	15	*	*	*	.105	1.2 (−22)
20	24	24	24	10	11	13	*	*	*	.1	3.1 (−24)
21	26	26	26	10	11	13	*	*	*	.095	7.6 (−26)
22	27	27	27	10	11	13	*	*	*	.091	1.8 (−27)
23	29	29	29	8	11	13	*	*	*	.087	3.9 (−29)
24	31	31	31	8	12	12	*	*	*	.083	8.4 (−31)
25	32	32	32	8	10	11	*	*	*	.08	1.7 (−32)

$= 0.118$ ($r = 17$), $Q^*(G)$ approximates $P(G)$ within a factor of 2 so long as $\nu \le 19$; and (iii) for the smallest value of E in Table 5, $E = 0.08$ ($r = 25$), $Q^*(G)$ also approximates $P(G)$ very well indeed since it is within factor of 2 of $P(G)$ so long as $\nu \le 32$. The most extreme tails of these three examples are 7.3×10^{-11}, 1.6×10^{-19}, and 1.7×10^{-32} respectively, which means that $Q^*(G)$ is an extraordinarily good approximation almost down to the most extreme tail (good enough for a cryptanalyst!). In fact, if the reader will compare the last column of Table 5 with the column headed 2 relating to $Q(S) = Q^*(G)$, he will see that the approximation $P(G) \approx Q^*(G)$ is amazingly good, even for the smallest probabilities, for all values of r up to $r = 25$. We recall that the statistic G for the equiprobable multinomial had an asymptotic distribution good down to such probabilities as 10^{-15} (Good and Crook, 1974).

The asymptotic distribution of X^2 approximates the exact distribution of X^2 within a factor of 5 almost down to the most extreme tail so long as $E \ge 0.143$ which is when $r \le 14$. When $E < 0.143$ ($r \ge 15$), $Q(X^2)$ approximates the more extreme tails of $P(X^2)$ less and less well as E decreases. We give the following two examples from the table: (i) When $E = 0.1$ ($r = 20$), $Q(X^2)$ approximates $P(X^2)$ within a factor of 5 so long as $\nu \le 11$ and within a factor of 10 so long as $\nu \le 13$, but, when $\nu \ge 14$, $Q(X^2)$ is a poor approximation to $P(X^2)$; and (ii) When $E = 0.08$ ($r = 25$), $Q(X^2)$ is within a factor of 10 of $P(X^2)$ for $\nu \le 11$, but it is out by factors from 13 to 10^{14} for $12 \le \nu \le 46$. This factor 10^{14}, which is not in the table, was determined from our computer output.

Table 5 shows that $Q(\Lambda)$ is a very poor approximation to $P(\Lambda)$ when $E < 0.35$. Also the computer run from which Table 5 was constructed demonstrates that, for $r \ge 6$, $Q(\Lambda)$ differs by at least a factor of 10 from $P(\Lambda)$ 98% of the time. And the approximation becomes, not surprisingly, very much worse as E decreases. For example, (i) when $E = 0.133$ ($r = 15$) our computer output showed that $Q(\Lambda)$ is out by factors between 10^5 and 10^{16} as $P(\Lambda)$ varies from 1.8×10^{-5} to 1.6×10^{-16}, and (ii) when $E = 0.083$ ($r = 24$) $Q(\Lambda)$ is out by factors from 10^9 to 10^{31} for $4.3 \times 10^{-9} > P(\Lambda) > 8.4 \times 10^{-31}$.

These results appear to establish that *both* $Q^*(G)$ and $Q(X^2)$ are much superior to $Q(\Lambda)$, for very small tail-area probabilities, when E is less than 0.35, at least when the marginal totals are flat.

To show that the remarkable accuracy of $Q^*(G)$ for sparse contingency tables, as obtained for $n = 2$, does not depend upon n being equal to 2, we constructed the most extreme part of the right-hand distributions when $n = 3$ for $r = 5, 10, 15, 20, 25$ and then computed the corresponding approximate probabilities to obtain the error ratios (see Appendix E). Table 6 shows the error-ratios for $r = 15, 20$, and 25, and also shows the "innards" of the tables to which the error ratios apply, and the probabilities $P(\tilde{S} \ge S)$. The approximation $Q^*(G)$ makes sense down to tail-area probabilities smaller than 10^{-40}.

19. The statistic *R* and a statistic of C.A.B. Smith. A statistic for sparse contingency tables that has belonged to folklore for some decades is $R = \frac{1}{2} \sum n_{ij}(n_{ij} - 1)$ which may be described as the total number of "repeats" within cells. For multinomial distributions, when $2tR = N(X^2 + N - t)$, the corresponding statistic was discussed by Good (1967, pages 400 and 418). This statistic dates back at least to Friedman (1922) and independently to Turing (1941), and was used by Eck (1961) for symmetrical tables. It is natural to assume that the distribution of R for sufficiently sparse tables is approximately Poissonian, given the null hypothesis, though Eck (1961) assumed a binomial distribution which is also reasonable. For $r \times r$ contingency tables having all the marginal totals equal to 2 we have found that the Poisson assumption is slightly better than the binomial assumption, and is extremely good. This is shown by Table 7 which compares the exact tail-area probabilities of R with the Poisson approximation, for a few values of r.

The example given by Eck was discussed by Good (1965, pages 55–56) where it was pointed out that the Poisson or binomial assumption is dangerous if there are *any* cells where the expectation is not small. The danger arises if the observed frequencies in such cells are high because they can then contribute too much to R.

In Appendix C, formulae are given for the expectation and variance of R in terms of the

TABLE 6

Error ratios for five of the six most extreme values in the right-hand tail of S when n = 3, N = 3r, and r = 15, 20, and 25

Innards	$P(\tilde{S} \geq S)$	Error Ratios For		
		$Q^*(G)$	$Q(X^2)$	$Q(\Lambda)$
r = 15				
3^{15}	5.1 (−33)			
$3^{13}2^21^2$	4.9 (−30)	2.4	1.1 (11)	5.5 (27)
$3^{12}2^31^3$	1.3 (−28)	1.3	3.8 (9)	3.0 (26)
$3^{12}2^21^5$	5.1 (−28)	4.9	5.4 (8)	10.0 (25)
$3^{11}2^41^4$	5.6 (−27)	1.1	2.3 (8)	10.0 (24)
$3^{11}2^31^6$	3.7 (−26)	2.9	6.0 (7)	1.9 (24)
r = 20				
3^{20}	1.1 (−48)			
$3^{18}2^21^2$	1.8 (−45)	2.9	1.9 (28)	3.5 (44)
$3^{17}2^31^3$	6.8 (−44)	1.5	9.8 (25)	1.0 (43)
$3^{17}2^21^5$	2.6 (−43)	5.8	4.7 (24)	2.7 (42)
$3^{16}2^41^4$	4.0 (−42)	1.3	9.0 (23)	1.8 (41)
$3^{16}2^31^6$	2.7 (−41)	3.5	8.0 (22)	2.8 (40)
r = 25				
3^{25}	2.3 (−67)			
$3^{23}2^21^2$	6.3 (−64)	23	Not run	Not run
$3^{22}2^31^3$	3.0 (−62)	45	Not run	Not run
$3^{22}2^21^5$	1.2 (−61)	492	Not run	Not run
$3^{21}2^41^4$	2.3 (−60)	50	Not run	Not run
$3^{21}2^31^6$	1.6 (−59)	297	Not run	Not run

marginal totals $(n_{i.})$ and $(n_{.j})$, assuming the null hypothesis. These formulae have a neat form for square tables when all the marginal totals are equal.

Smith (1951, 1952) (or see Kendall and Stuart, 1961, or 1973, Exercise 33.9) proposed a test for heterogeneity of proportions which can also be interpreted as a test for independence of the rows and columns of a contingency table. (It is so interpreted by Kendall and Stuart.) Smith's statistic, which we shall call X^2_{CABS}, is defined by the equation

$$X^2_{\text{CABS}} = \textstyle\sum_{i,j} \{(n_{ij} - n_{i.}n_{.j}N^{-1})^2/n_{.j}\}.$$

Smith states that this "test might be more powerful than χ^2 $[X^2]$, in certain circumstances, and also that it might be applicable when there were small expectations in some of the classes". Unfortunately it can also be much less powerful than X^2. Consider, for example, the two 3 × 2 contingency tables shown in Table 8. The contingency table on the left should clearly refute the null hypothesis and the one on the right should not, yet X^2_{CABS} is larger for the table on the right. So Smith's statistic may not have adequate power when the row or column totals are "rough". This difficulty might not arise for sparse contingency tables, but we have not examined this question.

20. The Bayes factor against *H* provided by the row and column totals. The main purpose of this section is to discuss FRACT, but first we make some brief comments concerning the factor F_1/F_2 provided by the row totals alone. As pointed out in Section 15, we have now decided that this factor should be equal to 1 so that F_1, F_2, and of course $F_{(2)}$, should all be equal and we have chosen a mathematical form of ϕ that achieves this. With the hyperprior ϕ that was used in Part I and elsewhere (see equation (2.4)) we have found that F_1/F_2 is merely *close* to 1 for 42 sets of $n_{i.}$'s with $N = 20$ and r and s between 2 and 5. For example, when $r = 2$, $s = 4$, and $N = 20$, and assuming the ϕ of Part I, we found that F_1/F_2 was well approximated by $1.26 - [(n_{1.} - 10)^2 + (n_{2.} - 10)^2]/300$; in fact, F_1/F_2 was 0.76 for $n_{1.} = 1$, $n_{2.} = 19$ and was 1.26 for $n_{1.} = n_{2.} = 10$. On the other hand, the "Bayes postulate" $k = 1$ (also expressible as $\phi(k) = \delta(k - 1)$) can lead, by (6.1) and (6.2), to much larger values of F_1'/F_2'; for example, when $r = s = 5$, and $n_{i.} = 4$ for $i = 1, 2, 3, 4$ and 5, we obtain $F_1'/F_2' = 10.1$.

The Bayes postulate had been independently proposed by Jeffreys (1936, page 427) and by Good (1950, page 99) (although Good proposed it in a tentative spirit), and they both withdrew it (Jeffreys, 1937, page 494; Good, 1965, Chapter 5). Jeffreys (1937, page 495) refers to "the revision of prior probabilities as knowledge in a subject advances." Instead of suggesting a new Bayesian model to replace the Bayes postulates for H and $\bar{H}$, he advocated testing new parameters one at a time rather than simultaneously. Hence, in his 1937 paper, he proposed a test only for 2 × 2 contingency tables. While agreeing that 2 × 2 tables are especially important we believe that tests for larger tables are also of value.

Any purely Dirichlet prior obtained by taking $\phi(k) = \delta(k - k_0)$ would lead to the conclusion that the row totals alone give nonzero weight of evidence for or against H, whether or not $k_0 = 1$ (which gives the Bayes postulate). It is fortunate that this objection to purely Dirichlet priors can be circumvented by using appropriate mixtures of them.

We now turn our attention to FRACT which, with our present hyperprior (15.4) [and indeed whenever $\phi(t, k)$ is of the form $\psi(tk)/k$] is equal to F_1/F_3 and to F_2/F_3. From (5.1) and (5.4) we have, with Φ as defined in Section 2.

$$\text{FRACT} = \frac{F_1}{F_3} = \frac{\sum^* \Phi\{(m_{ij}), rs, 1\}}{\Phi\{(n_{i.}), r, 1\}\Phi\{(n_{.j}), s, 1\}} \tag{20.1}$$

which of course does not depend on the innards of the table.

Nominally, FRACT applies to Models 1 and 2 and not to Model 3. But, if we want to use the criterion X^2 or G for testing H, although the sampling was done according to Models 1 or

TABLE 7

Some exact tail-area probabilities for R in the all-twos case, and a comparison with the Poisson approximation Q

The table is $r \times r$ and ν_2 denotes the number of n_{ij} equal to 2

r	$\mathscr{E}(R)$	$\frac{\text{var}(R)}{\mathscr{E}(R)}$	ν_2	P	P/Q
5	.556	1.016	5	1.058 (−3)	3.800
			3	2.222 (−2)	1.173
10	.526	1.003	10	1.53 (−9)	5.48
			8	1.39 (−7)	1.52
			6	2.09 (−5)	1.106
			4	2.17 (−3)	1.029
			2	9.85 (−2)	1.002
20	.512	1.00069	20	3.13 (−24)	7.83
			18	1.19 (−21)	2.06
			14	7.59 (−16)	1.23
			10	2.35 (−10)	1.078
			6	1.66 (−5)	1.019
			2	9.42 (−2)	1.00054
25	.510	1.0004	25	1.71 (−32)	8.78
			23	1.03 (−29)	2.29
			22	3.25 (−28)	1.60
			18	6.69 (−22)	1.26
			14	6.44 (−16)	1.12
			10	2.17 (−10)	1.045
			6	1.60 (−5)	1.012
			2	9.33 (−2)	1.00034

TABLE 8

Two 3 × 2 contingency tables, one highly significant, the other not

Only the innards are shown

10,083	10,013	10,168	9928
10,047	10,007	10,017	10,037
100	0	45	55

2, then, as argued in Section 15, we need to assume that these models are adequately approximated by Model 3 which implies that we need to find out if FRACT is close enough to unity. It may be recalled from Section 1 that Sokal and Rohlf (1969, page 589) said that X^2 and Λ seem to give similar results even when they are used for the wrong model. As they say, Λ is based on Model 1, though we claim it is equally valid to base it on Model 2 because, as we have argued, the row totals alone (with no other information) give no evidence about H.

Another reason for our interest in FRACT is that it leads to approximations to F_3 as we said in Section 15. In Section 7 it was conjectured that F_3 might be well approximated by $F_2F_{(2)}/F_1$, but under our new assumptions this expression is equal to F_1. From an inequality mentioned in the paragraph preceding (15.1), a good approximation to F_3 can be derived from the inequality $2F_1 > F_3 > \frac{2}{5}F_1$.

Before giving the results of our calculations concerning FRACT we consider some intuitive arguments that help to make the results appear reasonable. We begin by considering a 2 × 2 table sampled by Model 1 or Model 2, with $N = 2$ and row totals $n_{1\cdot} = n_{2\cdot} = 1$. The column totals are necessarily either (i) 0 and 2 (in some order), or (ii) 1 and 1. In Case (i), the columns of the table are necessarily proportional (in a trivial sense) so Case (i) must support H (slightly). In Case (ii) the innards must have two 1's on a diagonal, so Case (ii) is slightly against H. As a second example, consider a 3 × 3 table with row totals $n_{1\cdot} = n_{2\cdot} = n_{3\cdot} = 5$. Now suppose we are told that the column totals are (0, 0, 15). This will again force the columns to be proportional and so will support H. (Column totals of (1, 1, 13) would also support H though a little less clearly.) But if instead we are told that the column totals are also (5, 5, 5), then there are many possible innards that would undermine H, especially the six innards each with frequency count $0^6 5^3$ where the three 5's are necessarily not mutually en prise when regarded as rooks on a "chessboard".

In both these examples, *when the row totals are already known to be flat*, we find that very rough column totals somewhat support H and very flat column totals somewhat undermine H. We would thus expect the intermediate cases, where the column totals are neither very rough nor very flat, to supply less evidence for or against H.

The above argument is made more convincing when we notice that the various logically possible outcomes of an experiment cannot all support or all undermine a hypothesis. One way of seeing this is by recalling Turing's theorem that the expected Bayes factor against a true hypothesis is unity! (Good, 1950, page 72.) Although we have rejected the "Bayes postulate" ($k = 1$), it is instructive to use this postulate for verifying Turing's theorem for the first 2 × 2 example. The calculations can be done either by using (6.1) and (6.3) or by recalling that all ordered partitions of N into multinomial categories are equally probable under Bayes's postulate (Jeffreys, 1961, pages 133–134). The probability under Model 1 of getting all marginal totals equal to 1 is $\frac{1}{9}$ given H, and $\frac{2}{10}$ given $\bar{H}$. Hence $F_1'/F_3' = \frac{9}{5}$; while, for all other sets of marginal totals, $F_1'/F_3' = \frac{9}{10}$. Thus the expected value of FRACT, given H, is $\frac{1}{9} \cdot \frac{9}{5} + \frac{8}{9} \cdot \frac{9}{10} = 1$, in accordance with Turing's theorem.

In Table 9 we give some values of FRACT for a 3 × 3 contingency table when the hyperprior is given by (15.4). It bears out the intuitive reasoning just given. A new qualitative feature of Table 9 is the large value of FRACT when both row and column totals are very rough, namely when the marginal totals are (1, 1, 13; 1, 1, 13). We offer the following intuitive explanation of why this value of FRACT is so large.

Since the row totals (1, 1, 13) by themselves convey zero evidence concerning H, we regard them as permanently given, and we consider the evidence *then* provided by the column totals

TABLE 9
Some values of FRACT *when* $r = S = 3$ *and* $N = 15$, *using the hyperprior* (15.4)

Marginal totals	FRACT	Marginal totals	FRACT
(1, 1, 13; 1, 1, 13)	19.9	(5, 5, 5; 1, 1, 13)	0.140
(1, 1, 13; 1, 3, 11)	3.62	(5, 5, 5; 1, 3, 11)	0.538
(1, 1, 13; 1, 4, 10)	2.68	(5, 5, 5; 1, 4, 10)	0.948
(1, 1, 13; 1, 5, 9)	2.10	(5, 5, 5; 1, 5, 9)	1.34
(1, 1, 13; 2, 4, 9)	0.973	(5, 5, 5; 2, 4, 9)	1.77
(1, 1, 13; 3, 3, 9)	0.641	(5, 5, 5; 3, 3, 9)	1.92
(1, 1, 13; 2, 6, 7)	0.570	(5, 5, 5; 2, 6, 7)	2.29
(1, 1, 13; 3, 5, 7)	0.261	(5, 5, 5; 3, 5, 7)	2.65
(1, 1, 13; 4, 4, 7)	0.214	(5, 5, 5; 4, 4, 7)	2.72
(1, 1, 13; 5, 5, 5)	0.140	(5, 5, 5; 5, 5, 5)	3.12

TABLE 10
Values of FRACT for 2×2 *contingency tables with all marginal totals equal to n*
The last row gives the values of $0.0282(\log_2 n)^2 + 0.0071(\log_2 n) + 2.172$

n	2	4	8	16	32	64	128	256	512	1024
FRACT	2.215	2.269	2.444	2.654	2.918	3.227	3.602	4.035	4.523	5.065
	2.207	2.299	2.447	2.652	2.913	3.231	3.604	4.035	4.522	5.065

(1, 1, 13). There are just seven possible innards consistent with these marginal totals, two of which have the frequency count $0^6 1^2 13^1$. We call these the "most extreme cases". One of the tables with frequency count $0^6 1^2 13^1$ has [1, 1, 13] as its main diagonal. Assuming H, the expectations in the cells are 1/15 (four times), 13/15 (four times), and 169/15 (once). The total expectation in the "rare" cells is 4/15, and that in the "medium" cells is 52/15. Then X^2, with two degrees of freedom, for each of the most extreme cases, is

$$\tfrac{15}{4}\,(2 - \tfrac{4}{15})^2 + \tfrac{15}{52}\,(0 - \tfrac{52}{15})^2 + \tfrac{15}{169}\,(13 - \tfrac{169}{15})^2 = 15.$$

This corresponds to a tail-area probability of about 1/1800, or about 1/500 if we multiply by 7/2 to make rough allowance for there being two ways out of seven for obtaining the most extreme cases. This tail-area probability of 1/500 should be further increased somewhat because our specific lumping of the nine categories has a degree of arbitrariness, and because the χ^2 approximation is not good in this example, but the argument has removed any surprise that our Bayes factor is as large as 19.9.

For 2×2 tables, with all marginal totals equal to n, we can conveniently calculate FRACT from (20.1) and (5.5) by using also formula (4.1) of Good and Crook (1977) which was quoted as (B1.9), in Part I. Some results of the calculation are given in Table 10. It will be seen that FRACT appears to be tending very slowly to infinity (in fact FRACT is well approximated by a quadratic in log n, as shown in the table). This slow tendency to infinity is presumably also true for larger contingency tables with flat margins.

For 2×2 tables with margins $(1, N - 1; 1, N - 1)$, which are very rough when N is not small, we calculated FRACT for nine values of N between 4 and 50. We found that it appeared to be tending to infinity slower than $2N^{1/2}$ and was always less than $2N^{1/2}$.

As a final collection of examples of FRACT we took thirty sets of row and column totals, selected at random from the 504 sets mentioned in Section 17. The results are given in Table 11. We see that the largest value of FRACT was 5.7, and the smallest was 0.15. With the obsolete hyperprior ϕ, defined by (2.4), the values of FRACT were much the same as for the present ϕ. We again appear to have Bayesian robustness.

As a general guide for when one may neglect the evidence from the marginal totals we give our conclusions regarding the qualitative behavior of FRACT. In the following list we use the

TABLE 11

Values of FRACT *using the old and new hyperprior* ϕ, *for thirty sets of marginal totals. Also the values of* F_3 *and* $P(F_3)$ *for the most extreme innards*

(i) the number of the table; (ii) the marginal totals; (iii) and (iv), the values of FRACT when ϕ is given by (2.4) and (15.4) respectively; (v) the frequency count of the most extreme innards; (vi) the number of ways in which the most extreme innards can occur; (vii) $P(F_3)$, (viii) F_3; (ix) $N^{1/2}P(F_3)F_3$; (x) the natural logarithm of the entries in column (ix).

(i)	(ii) margins	FRACT (iii)	FRACT (iv)	(v)	(vi)	(vii)	F_3 (viii)	(ix)	(x)
CT1	1, 19; 4, 7, 9	0.13	0.15	$9^1 7^1 3^1 1^1 0^2$	1	0.20	1.7	1.54	.43
CT2	1, 19; 3, 8, 9	0.22	0.25	$9^1 8^1 2^1 1^1 0^2$	1	0.15	2.4	1.60	.47
CT3	7, 13; 1, 2, 17	0.7	0.7	$13^1 4^1 2^1 1^1 0^2$	1	3	68	0.93	−.07
CT4	9, 11; 1, 6, 13	1.0	1.0	$11^1 6^1 2^1 1^1 0^2$	1	4.6 (−4)	207	0.43	−.84
CT5	9, 11; 1, 7, 12	1.1	1.1	$11^1 7^1 1^2 0^2$	1	7.1 (−5)	1371	0.44	−.82
CT6	8, 12; 1, 8, 11	1.6	1.8	$11^1 8^1 1^1 0^3$	1	7.9 (−6)	3388	1.20	.18
CT7	5, 15; 1, 8, 11	1.8	1.8	$11^1 5^1 3^1 1^1 0^2$	1	3.6 (−3)	33	0.54	−.62
CT8	5, 15: 1, 7, 12	1.8	1.9	$12^1 5^1 2^1 1^1 0^2$	1	1.4 (−3)	101	0.61	−.49
CT9	7, 13; 4, 7, 9	2.1	2.0	$9^1 7^1 4^1 0^3$	1	1.3 (−5)	2251	0.13	−2.04
CT10	8, 12; 5, 6, 9	2.3	2.2	$8^1 6^1 5^1 1^1 0^2$	1	7.1 (−5)	85	0.027	−3.61
CT11	1, 19; 1, 4, 15	2.6	3.0	$15^1 4^1 1^1 0^3$	1	0.050	13	2.96	1.09
CT12	1, 19; 1, 2, 17	3.7	4.2	$17^1 2^1 1^1 0^3$	1	0.050	13	2.92	1.07
CT13	10, 10; 1, 1, 2, 16	0.4	0.5	$10^1 6^1 2^1 1^2 0^3$	2	0.087	2.3	0.87	−.14
CT14	8, 12; 1, 2, 3, 14	0.9	0.9	$12^1 3^1 2^2 1^1 0^3$	1	7.2 (−4)	92	0.30	−1.20
CT15	6, 14; 1, 1, 2, 16	1.1	1.2	$14^1 2^2 1^2 0^3$	1	3.1 (−3)	40	0.56	−.58
CT16	6, 14; 1, 1, 3, 15	1.4	1.5	$14^1 3^1 1^3 0^3$	1	3.9 (−4)	327	0.57	−.56
CT17	9, 11; 2, 2, 5, 11	1.6	1.6	$11^1 5^1 2^2 0^4$	1	6.0 (−6)	8501	0.23	−1.47
CT18	1, 19; 1, 1, 8, 10	1.5	1.8	$10^1 8^1 1^2 0^4$	2	0.100	7.7	3.46	1.24
CT19	7, 13; 1, 2, 5, 12	1.7	1.9	$12^1 5^1 2^1 1^1 0^4$	1	1.3 (−5)	10690	0.62	−.48
CT20	6, 14; 2, 2, 5, 11	2.0	2.1	$11^1 5^1 2^1 1^2 0^3$	2	1.03 (−4)	716	0.33	−1.11
CT21	6, 14; 1, 1, 5, 13	2.3	2.7	$13^1 5^1 1^2 0^4$	2	5.2 (−5)	8050	1.86	.62
CT22	2, 18; 1, 2, 3, 14	2.5	2.8	$14^1 3^1 2^1 1^1 0^4$	1	5.3 (−3)	62	1.45	.37
CT23	5, 15, 2, 2, 3, 6, 7	1.3	1.2	$7^1 6^1 3^1 2^2 0^5$	2	1.3 (−4)	212	0.12	−2.12
CT24	7, 13; 1, 1, 3, 6, 9	1.9	2.0	$9^1 6^1 3^1 1^2 0^5$	2	2.6 (−5)	2789	0.32	−1.14
CT25	6, 14; 1, 1, 3, 3, 12	2.2	2.4	$12^1 3^2 1^2 0^5$	6	2.6 (−5)	2858	0.33	−1.11
CT26	3, 17; 1, 1, 1, 2, 15	4.9	5.7	$15^1 2^1 1^3 0^5$	4	3.5 (−3)	149	2.33	.85
CT27	1, 9, 10; 4, 6, 10	1.4	1.5	$10^1 6^1 3^1 1^1 0^5$	1	2.2 (−6)	27528	0.27	−1.31
CT28	1, 2, 17; 2, 7, 11	1.2	1.5	$11^1 6^1 2^1 1^1 0^5$	1	2.0 (−3)	100	0.91	−.09
CT29	2, 5, 13; 1, 9, 10	1.9	2.0	$10^1 5^1 3^1 1^2 0^4$	1	3.1 (−4)	142	0.20	−1.61
CT30	1, 6, 13; 2, 3, 15	2.3	2.5	$13^1 3^1 2^1 1^2 0^4$	2	7.7 (−4)	183	0.63	−.46

terminology "Flat-Flat", for example, to mean that both the row totals and the column totals are flat:

(i) Flat-Flat; Somewhat undermines *H*.
(ii) Flat-Rough: Somewhat supports *H*.
(iii) Rough-Rough (with no zero marginal totals); Undermines *H* much more than Flat-Flat.
(iv) Flat-Medium: ½ < FRACT < 2½.
(v) Medium-Medium: ½ < FRACT < 2½.

Cases (iv) and (v) are the ones usually encountered in practice. This to a large extent justifies the assumption that Models, 1, 2, and 3 are approximately equivalent when testing for independence of the rows and columns of a contingency table.

The calculation of $A(k)$. When F_3 or FRACT is calculated with the help of (5.5), the calculation depends on the value of $A(k)$. Efficient algorithms for the calculation of $A(k)$ are available, especially when $r = s$ and when the row and column totals are completely flat (Good and Crook, 1977) and they apply also for the classical combinatorial problem of calculating $A(1)$, the number of arrays. (In Good and Crook (1977), page 39, line 1, an index −1 was omitted, and, in line 2, the symbol q_1 in the upper limit of summation should be q_0.) Approximations, for arbitrary margins, based on statistical and other arguments were given therein and in Part I. A further statistical argument for the case $k = 1$ was proposed by Gail and Mantel (1977) and a comparison of the statistical arguments is given by Good (1979b).

One of the methods for calculating $A(k)$ depends on the number $A(\boldsymbol{\nu}) = A(\boldsymbol{\nu}, (n_{i.}), (n_{.j}))$ of tables with given margins whose innards have a given frequency count $\boldsymbol{\nu}$ (meaning ν_0 zeros, ν_1 ones, $\cdots$). The times taken to compute the exact distributions of G, Λ, and R, when "running through all tables" with given margins, are proportional to $A(\boldsymbol{\nu})$ and not to $A(1)$. Only in special cases is there a known convenient closed form or simple finite series for calculating $A(k)$, $A(1)$, or $A(\boldsymbol{\nu})$. Some such special cases were relevant for Section 18 and for Table 6.

Other literature related to FRACT. A number of statisticians have held that the marginal totals contain evidence about H; for example, Barnard (1945, 1947), Pearson (1947), Good (1950, pages 99–101; 1965; 1976), McDonald, Davis, and Milliken (1977). Plackett (1977) made some remarks that "seem to confirm the intuitive view that the likelihood function provides little information ...". Perhaps only a Bayesian approach can offer quantitative answers, as in the present work.

21. The Bayes factor F_3 and its tail-area probability. To shed further light on the relationship between Bayesian and Fisherian significance we calculated $N^{1/2}P(F_3)F_3$, which relates the Bayes factor F_3 to $P(F_3)$, the right-hand tail-area probability of F_3. Model 3 is used here rather than Models 1 or 2 because, as we mentioned in Section 16, it is only for Model 3 that the null hypothesis is "simple". The relationship between a Bayes factor and its tail area was previously examined in the context of multinomial significance testing in Good (1967) and Good and Crook (1974) and found to be a strong one under fairly weak conditions. We would expect the relationship to fail when the Bayesian or Fisherian showed poor judgement.

To carry out a similar analysis for contingency tables would be very laborious. We therefore decided to look only at the innards of a table, with given margins, having the largest F_3, because for these innards we can easily calculate the tail-area probability $P(F_3)$.

For each of our thirty sets of row and column totals we computed $N^{1/2}P(F_3)F_3$ for the table innards with the largest F_3: see Table 11. The lower and upper quartiles of $N^{1/2}P(F_3)F_3$ in our sample of thirty tables were 0.32 and 1.45, so half the time $0.32 < N^{1/2}P(F_3)F_3 < 1.45$. (When the ϕ of Part I was used, the interquartile range was (0.25, 1.27), so we again have Bayesian robustness.) Over the thirty tables $\log_e[N^{1/2}P(F_3)F_3]$ appears to have a normal distribution with mean -0.5 and standard deviation 1.1, which we may express as $N^{1/2}P(F_3)F_3 = 0.6 \overset{\times}{\div} 3$. Although we have dealt somewhat cursorily with the relationship between F_3 and $P(F_3)$, we feel justified in saying that our observations are tolerably consistent with our experience in the multinomial situation.

22. Conclusions. We have stated some conclusions within each section of this paper, but we now summarise by mentioning a few salient points.

(i) The purely Bayesian methods appear to be robust with respect to variations in the hyperhyperparameters μ and q of the log-Cauchy hyperprior.

(ii) For tables that are not very sparse the approximation $P(G) \approx Q(G)$ is very good when $Q(G) > 10^{-4}$, and therefore $P(G) \approx Q^*(G)$ is a better approximation.

(iii) *For sparse tables*, at least when the marginal totals are flat, it seems that the approximation $P(G) \approx Q^*(G)$ is extraordinarily good for small tail-area probabilities, the chi-squared approximation for X^2 is also very good when $\nu \leq 8$ and $E \geq 0.08$, but that for Λ is very poor. The statistic R, the number of repeats, has closely a Poisson distribution (but not if any cell expectation is large). The exact expectation and variance for R are given in Appendix C.

(iv) Our conclusions concerning FRACT are given near the end of Section 20.

(v) A rough relationship between F_3 and $P(F_3)$ is given in Section 21 and it resembles the results found earlier for the corresponding multinomial problem.

APPENDIX C

The mean and variance of the number of repeats within cells

[Appendices A and B occurred in Part I]

It is simple to prove that

$$\text{(C1)} \qquad \mathscr{E}(R \mid H) = \frac{R_2 S_2}{2N^{(2)}}$$

and straightforward but heavy to prove that

$$\text{(C2)} \qquad \operatorname{Var}(R \mid H) = \frac{R_3 S_3}{N^{(3)}} + \frac{R_2 S_2}{2N^{(2)}} + \frac{(R_2^2 - 4R_3 - 2R_2)(S_2^2 - 4S_3 - 2S_2)}{4N^{(4)}} - \frac{R_2^2 S_2^2}{4(N^{(2)})^2},$$

where

$$R_\mu = \textstyle\sum_i n_{i.}^{(\mu)}, \qquad S_\mu = \sum_j n_{.j}^{(\mu)} \qquad \mu = 1, 2, 3, \cdots.$$

The method of proof is to make use of the familiar identity

$$\text{(C3)} \qquad \sum{}^* \frac{\prod n_{i.}! \prod n_{.j}!}{N! \prod m_{ij}!} = 1,$$

for a variety of values of N etc., where $\sum^*$ was defined in Section 15. (The combinatorial identity (C3) expresses the fact that the Fisher-Yates probabilities add up to 1.)

In particular, for square tables having all marginal totals equal to n, it is easy to see that

$$\text{(C4)} \qquad R = (X^2 - r^2 + rn)n/(2r),$$

and it can be shown that

$$\text{(C5)} \qquad \mathscr{E}(R \mid H) = \frac{(n-1)^2 nr}{2(nr-1)},$$

$$\text{(C6)} \qquad \operatorname{Var}(R \mid H) = \frac{n^3(n-1)^2(r-1)^2 r}{2(nr-1)^2(nr-3)},$$

and

$$\text{(C7)} \qquad \frac{\operatorname{Var}(R \mid H)}{\mathscr{E}(R)} = \frac{n^2(r-1)^2}{(nr-1)(nr-3)}.$$

It is easier to deduce (C5) from Smith (1951, 1952) or from Kendall and Stuart (1961, example 33.9) than from (C2). Presumably it can also be obtained from Welch (1938, page 158) or Haldane (1940, page 353).

Note that $\operatorname{Var}(R \mid H)/\mathscr{E}(R)$ is close to 1 when r is large, even if n is large. But we cannot expect R to have approximately a Poisson distribution unless n/r is small.

APPENDIX D

Description of the methods for sampling marginal totals when r, s, and N are fixed

Sampling Method A. The selection of a set of contingency-table marginal totals for assigned values of r, s, and N, by Sampling Method A was accomplished in the following manner: Think of the marginal totals as occupying boxes $B_{i.}$ ($i = 1, 2, \cdots, r$) and $B_{.j}$ ($j = 1, 2, \cdots, s$). Now, let $p_{i.}$ denote the probability that a ball "tossed at the row totals" will land in $B_{i.}$, and let $p_{.j}$ be similarly defined. Random numbers between 0 and 1 were chosen to simulate the tossing of balls at $(B_{i.})$ (and $(B_{.j})$), and $n_{i.}$ (and similarly $n_{.j}$) was increased by 1 whenever a random number was in the interval $(P_{i-1.}, P_{i.})$, where $P_{i.} = p_{1.} + \cdots + p_{i.}$, until each $n_{i.}$ was at least 1 and $\sum_i n_{i.} (= \sum_j n_{.j}) = N$. Each time we wished to determine a new set of marginal totals, $(p_{i.})$ and $(p_{.j})$ were chosen from a uniform distribution and then scaled to force $\sum_i p_{i.} = 1 = \sum_j p_{.j}$.

Sampling Method B. Sampling Method A does not give enough examples of flat row totals so we introduced a Method B in which the margins were selected (for given r, s, and N) by first setting all row totals equal to N/r, or, if N/r was not an integer, by making the set of totals as flat as possible. A computer program was then used to create a list of all possible sets of *column* totals (with $n_{.1} \geq n_{.2} \geq \cdots \geq n_{.s}$) which is a list of the possible partitions of N into exactly s parts. The margins were then "completed" by randomly selecting a previously unselected partition from the list.

Sampling Method C. In Method C, introduced to obtain both flat row and column totals, the row totals were chosen as in Method B. The column totals were then selected to be as flat as possible. For this method of sampling there is obviously only one set of margins (up to permutations of rows and of columns) that can be selected when r, s, and N are given.

APPENDIX E

The extreme right-hand tails of the distributions of X^2, Λ, and G

In Section 18 we mentioned that the most extreme part of the right-hand distributions of X^2, Λ, and G were calculated by specialized combinatorial arguments for square $r \times r$ tables with all marginal totals n equal to 3. The problem is clearly essentially solved by counting the frequencies of the frequencies of the frequencies, that is, by counting the tables with a given internal frequency count $\boldsymbol{\nu} = (\nu_0, \nu_1, \cdots, \nu_n)$ where ν_μ is the number of cells containing an entry of μ. Thus $\sum_\mu \nu_\mu = r^2$, $\sum_\mu \mu\nu_\mu = nr$. When $n = 3$, which is the case we are now considering, unlike the case $n = 2$, we do not know a neat general formula (like (18.2)) for the number $A(\boldsymbol{\nu}) = A(\boldsymbol{\nu}; n, r \times r)$ of tables. But note that ν_3 3's can be placed in $\nu_3!\binom{r}{\nu_3}^2$ ways and therefore

$$\text{(E1)} \qquad A(\nu_0, \nu_1, \nu_2, \nu_3; 3, r \times r) = \nu_3!\binom{r}{\nu_3}^2 A(\nu_0, \nu_1, \nu_2, 0; 3, (r - \nu_3) \times (r - \nu_3)).$$

If $r - \nu_3 = r'$ is small enough it is practicable to evaluate this expression by listing all essentially distinct $r' \times r'$ arrays of 0's, 1's, and 2s; and there are the following checks with say $r = 4$: (i) $\sum_{\boldsymbol{\nu}} A(\boldsymbol{\nu})$ is equal to the number of $r \times r$ tables with marginal totals all equal to 3, and (ii) $\sum_{\boldsymbol{\nu}} A(\boldsymbol{\nu})$ F.Y. $= 1$, where F.Y. denotes the Fisher-Yates probability $6^{2r-\nu_3}2^{-\nu_2}/(3r)!$ (see (4.1)).

For $n > 3$ the same method can be used but it becomes rapidly more laborious as n increases, for any given value of $r - \nu_n$.

REFERENCES

BARNARD, G. A. (1945). A new test for 2 × 2 tables. *Nature* **156** 177.

BARNARD, G. A. (1947). Significance tests for 2 × 2 tables. *Biometrika* **34** 123–138.

BISHOP, Y. Y., FIENBERG, S. E. and HOLLAND P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Massachusetts and London, England: M.I.T. Press.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578.

ECK, R. V. (1961). Non-randomness in amino-acid 'alleles'. *Nature* **191** 1284–1285.

FISHER, R. A. (1935). The logic of inductive inference. *J. Roy Statist. Soc.* **98** 39–54.

FISHER, R. A. (1938). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh and London.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference.* 7th Ed; Oliver and Boyd, Edinburgh and London.

FRIEDMAN, W. F. (1922). *The index of coincidences and its applications in cryptology.* Riverbank Laboratories, Geneva, Illinois.

GAIL, M. and MANTEL, N. (1977). Counting the number of $r \times c$ contingency tables. *J. Amer. Statist. Assoc.* **72** 859–862.

GOOD, I. J. (1950). *Probability and the Weighing of Evidence.* Griffin, London.

GOOD, I. J. (1952). Rational Decisions. *J. Roy. Statist. Soc. Ser. B* **14** 107–114.

GOOD, I. J. (1956). On the estimation of small frequencies in contingency tables. *J Roy. Statist. Soc.* **18** 113–124.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* M.I.T. Press. Cambridge, Massachusetts. (Paperback 1968).

GOOD, I. J. (1967). The Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B* **29** 399–431. (Corrigendum (1974) **36** 109.) (With discussion.)

GOOD, I. J. (1969a). Statistics of Language. In *Encyclopedia of Linguistics, Information, and Control* (A. R. Meetham, ed.). 567–581. Pergamon.

GOOD. I. J. (1969b). A subjective analysis of Bode's law and an 'objective' test for approximate numerical rationality, *J. Amer. Stat. Soc.* **64** 23–66 (with discussion).

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

GOOD, I. J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *Ann. Statist.* **4** 1159–1189.

GOOD, I. J. (1979a). The contributions of Jeffreys to Bayesian statistics. In *Studies in Bayesian Econometrics and Statistics in Honor of Harold Jeffreys* (A. Zellner, ed.). 21–34. North Holland, Amsterdam.

GOOD, I. J. (1979b). A comparison of some statistical estimates for the number contingency tables. *J. Statist. Comp. and Simulation* **8** 312–314.

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

GOOD, I. J. and CROOK, J. F. (1977). The enumeration of arrays and a generalization related to contingency tables. *Discrete Math.* **19** 23–45.

HALDANE, J. B. S. (1940). The mean and variance of χ^2, when used as a test for homogeneity, when expectations are small. *Biometrika* **31** 346–355.

JEFFREYS, H. (1936). Further significance tests, *Proc. Cambridge Philos. Soc.* **32** 416–445.

JEFFREYS, H. (1937). The tests for sampling differences and contingency. *Proc. Roy. Soc. Ser. A,* **162** 479–495.

JEFFREYS, H. (1961). *Theory of Probability.* 3rd ed.; Clarendon, Oxford.

KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics,* II. Griffin, London. (3rd Ed., 1973.)

LEVIN, B. and REEDS, J. (1977). Compound multinomial likelihood functions are unimodal; proof of a conjecture of I. J. Good. *Ann. Statist.* **5** 79–87.

MCDONALD, L. L., DAVIS, B. M., and MILLIKEN, G. A. (1977). A nonrandomized unconditional test for comparing two proportions in 2×2 contingency tables. *Technometrics* **19** 145–150.

PEARSON, E. S. (1947). The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table. *Biometrika* **34** 139–167.

PEARSON, K. (1905). On the theory of contingency and its relation to association and normal correlation. *Draper's Co. Res. Mem. Biometric series.* Reprinted in *Karl Pearson's Early Papers,* Cambridge Univ. Press, 1948.

PERKS, W. (1947). Some observations on inverse probability including a new indifference rule, *J. Inst. Actuar.* **73** Part II, 285–334 (with discussion).

PERKS, W. (1967). Contribution to the discussion of a paper by G. A. Barnard, "The Bayesian controversy in statistical inference", *J. Inst. Actuar.* **73** Part II, **395** 264–268.

PLACKETT, R. L. (1977). The marginal totals of a 2×2 table. *Biometrika* **64** 37–42.

SMITH, C. A. B. (1951). A test for heterogeneity of proportions. *Ann. Eugenics* **16** 16–25; continued in **17** (1952) 35–36.

SOKAL, R. R. and ROHLF, F. J. (1969). *Biometry: The Principles and Practice of Statistics in Biological Research.* Freeman, San Francisco.

TURING, A. M. (1941). Private communication to I. J. Good, E. H. Simpson and others.

WELCH, B. L. (1938). On tests for homogeneity. *Biometrika* **30** 149–158.

WILKS, S. S. (1938). The large-scale distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.

SCHOOL OF BUSINESS ADMINISTRATION
WINTHROP COLLEGE
ROCK HILL, SOUTH CAROLINA 29733

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