Association and Estimation in Contingency Tables

Authors(s): Frederick Mosteller

Source: *Journal of the American Statistical Association*, Vol. 63, No. 321 (Mar., 1968), pp. 1-28
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2283825
Accessed: 29-03-2016 07:41 UTC

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 321 MARCH, 1968 *Volume 63*

ASSOCIATION AND ESTIMATION IN CONTINGENCY TABLES

FREDERICK MOSTELLER[1,2]
Harvard University

OUR 1967 Committee on Publications, chaired by David L. Wallace, found that many American Statistical Association members desired more review and survey papers. These have been hard to come by, and so as my last act before leaving office, I decided to provide a short survey paper on some related ideas in a field where nearly all of us sometimes work—that of contingency tables. These ideas are largely available in the literature, and yet they have not often been put together, though I. J. Good's monograph [9] and Leo Goodman's many papers form good sources. But my paper is not intended as a review of the literature, only as a survey of one set of ideas about estimation in the analysis of contingency tables.

I fear that the first act of most social scientists upon seeing a contingency table is to compute chi-square for it. Sometimes this process is enlightening, sometimes wasteful, but sometimes it does not go quite far enough.

EXAMPLE 1. CONTEXTUALITY OF PRESENT-DAY AMERICAN ENGLISH

In my first and only chi-square example, I draw on Kučera and Francis's attractive and informative 1967 book entitled *Computational analysis of present-day American English* [13].

The authors collected 500 samples of writing published about 1961, each sample of length about 2,000 words (but not exactly). These selections come from 15 different genres as follows

	Number of Samples		Number of Samples
A. Press reportage	44	J. Learned and scientific	80
B. Press editorial	27	K. Fiction: general	29
C. Press reviews	17	L. Fiction: mystery and detective	24
D. Religion	17	M. Fiction: science	6
E. Skills and Hobbies	36	N. Fiction: adventure and western	29
F. Popular lore	48	P. Fiction: romance and love story	29
G. Belles lettres, biography, etc.	75	R. Humor	9
H. Miscellaneous (gov'tese)	30		
		Total number of samples	500

[1] Presidential Address delivered at the annual meeting of the American Statistical Association, December 29, 1967, Shoreham Hotel, Washington, D. C.

[2] The preparation of this paper was facilitated by National Science Foundation grant GS-341.

To see whether high-frequency words were distributed nearly independent of context as defined by the 15 genres, the authors computed chi-square on the frequencies of occurrences of the 100 most used words. Table 1 exemplifies this for the word *the*. They found, with 14 degrees of freedom, that all the words are distributed in a manner significantly different from random according to the size of the samples from the several genres—that is, all 100 words were somewhat contextual as based upon the multinomial model. The smallest chi-square was 46, which is about 6 standard deviations away from the expected value of 14, and the largest is over 5,000, or about 1,000 standard deviations from the null.

TABLE 1. EXAMPLE OF DISTRIBUTION OF 69961 *THE'S* INTO 15 GENRES (KUČERA AND FRANCIS [13])

	Genres							
	A	B	C	D	E	F	G	H
Frequency	6385	3961	2370	2480	4757	6976	10758	4621
Expected	6122.5	3764.8	2441.6	2386.1	5009.1	6711.0	10500.6	4311.5
Per cent	7.19	7.26	6.70	7.17	6.55	7.17	7.07	7.39
	J	K	L	M	N	P	R	Total
Frequency	12536	3792	2817	723	3780	2988	1027	69971
Expected	11183.3	4030.3	3331.3	831.9	4033.7	4050.9	1261.5	
Per cent	7.73	6.49	5.83	5.99	6.46	5.09	5.62	

$\chi^2_{14}=688.7,\ 1000\chi^2_{14}/69971=9.8$

To those of us who are used to straining for statistical significance, these strong results will be refreshing. Some may say that they did not expect the words to be distributed randomly, but that is not the point, the real point is that the authors have a lot of data. They can therefore magnify modest deviations. For example, the number of uses of the word *the* in their list is 69,971. Distribute that into 15 categories nearly evenly and one can still get a huge value of chi-square. As Table 1 shows they got 688. Having established that the distributions are different from random Kučera and Francis's question is answered.

Suppose that we wish to go further and ask which words are more contextual and which words less. For this we do not want chi-square itself, but chi-square divided by sample size. The reason is that in the chi-square formula the term

$$\frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

doubles when observed and expected are doubled, and is multiplied by k when both are multiplied by k. Thus the size of chi-square for a given departure from a random distribution is approximately proportional to sample size, a cause for rejoicing when we want to detect differences, but an interfering variable when we want to compare departures from proportional distributions.

I have divided chi-square by n to give us a suitable measure of contextuality in this problem, a measure that makes it possible to compare different words. For convenience, I have multiplied this by 1,000 for making comparisons. Among the 100 most frequently used words the three with the least values of χ^2/n were the preposition, the conjunction, and the article *to*, *and*, and *the*, with contextuality scores 6.2, 6.9, and 9.6. Nevertheless, their chi-square values were 33rd, 43rd, and 79th largest counting from the bottom. Thus chi-square alone can be misleading if we want to estimate degree of contextuality. The three most contextual words were the pronouns *her*, *your*, and *she*, a finding which agrees with the experience Wallace and I had in our analysis of *The Federalist* [21] where we were forced to toss out all pronouns as too contextual. However, I am not primarily discussing chi-square tonight, and I introduced it here to emphasize estimation in contrast to testing for significance.

EXAMPLE 2. MEASURING ASSOCIATION BY THE CROSS-PRODUCT RATIO

Let us turn first to the classical 2×2 table, as illustrated in Table 2. The letters a, b, c, and d can be thought of as counts or unnormalized proportions, but think of them as population values rather than samples. For concreteness, we might think of A as a high grade in a course and not A as a not so high grade, of B as male and not B as female. Then the tables illustrate some possible degrees of association between grades and sex. Table 2-1 is the general table.

TABLE 2. EXAMPLES OF 2×2 TABLES, WITH COURSE GRADE VERSUS SEX AS ILLUSTRATIVE

Table 2-1. Prototype 2×2

[Course grade]		[Sex] [Male] B	[Female] not B
[High]	A	a	b
[Low]	not A	c	d

Table 2-2. Equivalent 2×2 after row and column multiplications

	B	not B
A	r_1c_1a	r_1c_2b
not A	r_2c_1c	r_2c_2d

Table 2-3

	B	not B
A	2	3
not A	1	4

Table 2-4

	B	not B
A	4	30
not A	2	40

Table 2-5

	B	not B
A	12	90
not A	1	20

	Table 2-3	Table 2-4	Table 2-5
Bivariate normal	$\rho=0.35$	$\rho=0.29$	$\rho=0.25$
Cross-product ratio	$\alpha=8/3$	$\alpha=8/3$	$\alpha=8/3$

Table 2-6. 2×2 probabilities formed by cutting a bivariate distribution of X and Y at x and y

		Low $X\le x$	High $X>x$
High	$Y>y$	b	a
Low	$Y\le y$	d	c

Table 2-6 illustrates how we might obtain a 2×2 table by cutting a bivariate distribution F by lines parallel to its x and y axes. If we refer Tables 2-3, 2-4, and 2-5 to a bivariate normal to discover the correlations required to produce them, we find from Pearson's Tables [22], the Pearsonian correlation coefficients $\rho=0.35$, 0.29, and 0.25, in that order.

Let us turn to an entirely different way of thinking about the association in the 2×2 table. Note that Table 2-4 is obtained from Table 2-3 by doubling the number of men and by multiplying the number of women by 10. In other words, the men are no better on the average than they were before, and the women are no better either, there are just more of them. Why should that change the degree of association? It is, of course, easy to call attention to the bivariate normal as an explanation. But that is not the point, the point is that it is reasonable to think of the association between sex and grades as independent of the relative numbers of men and of women, and indeed of the numbers of A's and not A's. Why should tripling the number of A's and halving the number of not A's in each category (to get from Table 2-4 to 2-5) have any effect upon the basic association in the table? One reasonable answer is that it should not. We might instead think of a contingency table as having a *basic nucleus* which describes its association and think of all tables formed by multiplying elements in rows and columns by positive numbers as forming an equivalence class—a class of tables with the same degree of association. Thus Tables 2-3, 2-4, and 2-5 fall into the same equivalence class under this definition, and Table 2-2 for any positive values of the r's and the c's is in the same equivalence class (has the same degree of association) as Table 2-1. Table 2-2 can be obtained by multiplying elements in row 1 of Table 2-1 by r_1, row 2 by r_2, column 1 by c_1, and column 2 by c_2. We shall suppose that a, b, c, and d are all strictly positive.

An index of association that is invariant under these row and column multiplications is the cross-product ratio

$$\alpha = \frac{ad}{bc} = \frac{(r_1c_1a)(r_2c_2d)}{(r_1c_2b)(r_2c_1c)}, \qquad r_i > 0,\ c_i > 0,\ i = 1, 2. \tag{1}$$

In our numerical examples, Tables 2-3, 2-4, and 2-5 give $\alpha=8/3$.

The quantity α runs from 0 to ∞ and $\alpha=1$ corresponds to independence. (If one wanted a correlation-like quantity he might first take the logarithm of α to get some symmetry, since α and $1/\alpha$ clearly correspond to the same degree of association in opposite directions and then to map onto the interval from -1 to $+1$, for example by using the arc tangent. A good set of references on the cross-product ratio is to be found in Goodman [11].

By writing

$$\frac{a}{b} = \alpha\frac{c}{d} \quad \text{and} \quad \frac{a}{c} = \alpha\frac{b}{d},$$

we can interpret α as the number by which to multiply the odds on one side of a 2×2 table (c/d or b/d) in order to equate them to those on the other side (a/b or a/c, respectively).

Plackett [23] has developed continuous bivariate families of distributions for which association measured in the cross-product manner is natural. It turns out that for given continuous margins G for X and H for Y and a given α one can always construct a joint cumulative distribution F having the property that when it is cut anywhere by lines parallel to the X and Y axes the probabilities in the four quadrants viewed as a contingency table have the invariance property.

Following the notation in Table 2-6 let $d=F$, $c=H-F$, $b=G-F$, $a=1-G-H+F$, then

$$\alpha = \frac{(1-G-H+F)F}{(G-F)(H-F)}, \qquad 0 < F < 1,\ F < G < 1,\ F < H < 1.$$

And for convenience write

$$Z = 1 - (1-\alpha)(G+H).$$

Then the required cumulative distribution having the invariance property is

$$F = \frac{-Z + \sqrt{Z^2 + 4\alpha(1-\alpha)GH}}{2(1-\alpha)}, \qquad \alpha \neq 1. \tag{2}$$

Does the family of bivariate normal distributions have the property? Alas, no. To help show this we computed Table 3 which gives the values of α for the bivariate normals with correlation coefficient $\rho=.25$, .50, .75, and standard deviations $\sigma_x=\sigma_y=1$ at the cutting points where x and y take values 0, .5, 1, 1.5, 2. These α's were readily obtained from Pearson's tables of the bivariate normal [22]. As the correlation increases, the variation in values of α increases.

Another way to appreciate the difference between the normal and the invariant distributions is to develop the invariant distribution with unit normal margins and then find the value of ρ that makes the cumulative of a normal

TABLE 3. VALUES OF α FOR THE BIVARIATE NORMALS WITH $\rho=.25$, .50, .75 FOR SELECTED POINTS (x, y), WHERE x AND y ARE MEASURED IN STANDARD DEVIATIONS FROM THEIR MEANS

y \ x	$\rho=.25$					$\rho=.50$				
	0	.5	1.0	1.5	2.0	0	.5	1.0	1.5	2.0
0	1.91	1.96	2.10	2.36	2.75	4.00	4.26	5.13	6.98	10.63
.5	1.96	1.98	2.10	2.33	2.68	4.26	4.23	4.76	6.02	8.45
1.0	2.10	2.10	2.21	2.43	2.79	5.13	4.76	5.04	6.03	8.00
1.5	2.36	2.33	2.43	2.68	3.10	6.98	6.02	6.03	6.90	8.84
2.0	2.75	2.68	2.79	3.10	3.62	10.63	8.45	8.00	8.84	11.11

y \ x	$\rho=.75$				
	0	.5	1.0	1.5	2.0
0	11.20	13.36	22.86	56.67	200.84
.5	13.36	12.03	15.38	27.70	69.65
1.0	22.86	15.38	15.04	20.66	38.91
1.5	56.67	27.70	20.66	22.30	33.17
2.0	200.84	69.65	38.91	33.17	40.31

TABLE 4. VALUES OF ρ THAT MAKE THE SAME CONTINGENCY TABLE FOR THE BIVARIATE NORMAL THAT THE INVARIANT DISTRIBUTION WITH NORMAL MARGINS HAS FOR $\alpha=2$

y \ x	0	.5	1.0	1.5	2.0
0	.27	.26	.23	.20	.18
.5	.26	.25	.23	.21	.18
1.0	.23	.23	.22	.19	.17
1.5	.20	.21	.19	.17	.15
2.0	.18	.18	.17	.15	.13

equal to an invariant F for a given α. For $\alpha=2$, our Table 4 shows ρ for a variety of cutting points. Note that different positions of the cuts require different correlations, instead of identical ones as they would if the invariant distribution were a bivariate normal.

Next let us illustrate one form of the invariant probability density function. We computed Table 5 to show selected ordinates for the joint probability density with uniform margins on the interval from 0 to 1 that produces $\alpha=2$. Note that the ordinates at the origin (0, 0) and at (1, 1) are high, at the other corners (0, 1) and (1, 0) the ordinates are low, and there is a saddle point in the middle, $(\frac{1}{2}, \frac{1}{2})$.

STANDARDIZING A 2×2 TABLE

We do not have to think about continuous distributions for this sort of situation, however. We can use the idea of association as a way to understand our data better. For example, we might especially arrange the table to have uniform margins on each side in the case of a two-way table so as to get a clearer look at the association that is actually occurring. If appropriate it could be viewed as a doubly stochastic matrix of transition probabilities, or condi-

TABLE 5. ORDINATES OF THE INVARIANT DENSITY FUNCTION WITH UNIFORM MARGINS, $\alpha=2$ AT SELECTED VALUES OF (x, y)

y \ x	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
0	2.00	1.65	1.39	1.18	1.02	.89	.78	.69	.62	.55	.50
.1	1.65	1.49	1.33	1.19	1.06	.94	.84	.76	.68	.61	.55
.2	1.39	1.33	1.26	1.17	1.08	.99	.91	.82	.75	.68	.62
.3	1.18	1.19	1.17	1.14	1.09	1.03	.96	.89	.82	.76	.69
.4	1.02	1.06	1.08	1.09	1.08	1.05	1.01	.96	.91	.84	.78
.5	.89	.94	.99	1.03	1.05	1.06	1.05	1.03	.99	.94	.89
.6	.78	.84	.91	.96	1.01	1.05	1.08	1.09	1.08	1.06	1.02
.7	.69	.76	.82	.89	.96	1.03	1.09	1.14	1.17	1.19	1.18
.8	.62	.68	.75	.82	.91	.99	1.08	1.17	1.26	1.33	1.39
.9	.55	.61	.68	.76	.84	.94	1.06	1.19	1.33	1.49	1.65
1.0	.50	.55	.62	.69	.78	.89	1.02	1.18	1.39	1.65	2.00

tional probabilities. We can either get a formula to do this or do it by successive iterations. Since other related problems require iterations, let us iterate even for a 2×2.

Let us start with Table 6-1. We might as well try to reduce it to a table with unit totals in each row and column. We can convert to probabilities at the end, if we desire. Multiply the elements of the first row by 1/110, of the second row by 1/7. This gives, to three decimals, Table 6-2.

TABLE 6. ITERATING TO A STANDARD 2×2 BY ROW AND COLUMN MULTIPLICATIONS

Table 6-1. Original table

		B	$\overline{B}$	Actual	Required
	A	100	10	110	1
	$\overline{A}$	5	2	7	1
Actual totals		105	12	117	
Required totals		1	1		

Table 6-2. 1st iteration

		B	$\overline{B}$	Actual	Required
	A	.909	.091	1.000	1.000
	$\overline{A}$	.714	.286	1.000	1.000
Actual totals		1.623	.377		
Required totals		1.000	1.000		

Table 6-3. 2nd iteration - completing 1 cycle

		B	$\overline{B}$		
	A	.560	.241	.801	1.000
	$\overline{A}$	.440	.759	1.199	1.000
Actual totals		1.000	1.000		
Required totals		1.000	1.000		

Table 6-4. After 4 cycles

		B	$\overline{B}$		
	A	.667	.333	1.000	1.000
	$\overline{A}$	.333	.667	1.000	1.000
Actual totals		1.000	1.000		
Required totals		1.000	1.000		

Next we multiply columns by 1/1.623 and 1/0.377 and get as our second iteration Table 6-3. This messed up the row totals and so we repeat the whole operation. After 4 cycles, we get Table 6-4. This is the table with uniform margins we required. If we divide cell values of Table 6-4 by 2 to change to numbers like probabilities we get

	B	$\overline{B}$	Total
A	.333	.167	.500
$\overline{A}$	.167	.333	.500
Total	.500	.500	1.000

a nucleus of association that is scarcely suggested by visual inspection of the original Table 6-1.

A formula for adjusted d, the $(\overline{A}, \overline{B})$ cell, when row totals are required to be R and $\overline{R}$, column totals are required to be C and $\overline{C}$ and $R+\overline{R}=C+\overline{C}=N$, is, from Mosteller [18]

$$\text{adjusted } d = \frac{1}{2(\alpha-1)}\{N+(\alpha-1)(\overline{R}+\overline{C}) - \sqrt{[N+(\alpha-1)(\overline{R}+\overline{C})]^2 - 4(\alpha-1)\alpha\overline{R}\overline{C}}\}. \quad (2^*)$$

In our example $N=2$, $\bar{R}=\bar{C}=1$, $\alpha=100\times 2/(5\times 10)=4$. Consequently, adjusted $d=2/3$, as shown in the lower right-hand cell of Table 6-4. Formula (2*) or the quadratic leading to it is probably known to many, and it is readily related to Plackett's formula (2) above.

EXAMPLE 3. LEVINE'S STUDY OF OCCUPATIONAL MOBILITY

Let us extend this idea of equivalence classes of association obtained by starting with a given table and generating others by multiplying levels of a given variable by positive constants. We wish to extend it not only to the two-way $r\times c$ contingency table but also to the k-way $r_1\times r_2\times \cdots \times r_k$ table.

The same technique of standardizing works in these larger tables. We shall illustrate its use for the two-way table by results kindly supplied by Joel H. Levine, [14, 15] who has recently developed and exposited the cross-product idea and applied the standardizing technique to Glass's British and Svalastoga's Danish occupational mobility data.

In Panel 7-1 of Table 7 the upper numbers in the cells are the British counts showing the joint distributions of father's and son's occupation distributed into 5 categories, the lower numbers are the Danish counts in corresponding categories. The two sets of row data in Panel 7-1, quite apart from their differing totals, even in good sized cells show ratios of British to Danish counts running from as high as 3 down to 2/5. Thus the ratios are far from constant. Levine has standardized the two sets separately to have row totals of 100 and column totals of 100, achieved by row and column multiplications. This standardizing shown in Panel 7-2 of Table 7 equates for Britain and Denmark the father's original distributions of occupational status and equates the son's distributions as well. Again, we can interpret the resulting numbers as transitional or conditional probabilities expressed in per cents—either son's distribution given the father's category, or father's given the son's.

When we look at the result in Panel 7-2, we observe that standardizing has brought a great deal of order into the lines of the British table—for example, the slump in the middle of father's category 2 is now just part of a smooth descent. Furthermore, where the original sets of data appeared rather incompatible, Panel 7-2 shows that the numbers in a cell are very similar for the two nationalities. In the sense of having a common nucleus of association, as we have discussed above, it would be fair to say that the two occupational tables are nearly equivalent. It is a great economy of description to find societies having much the same intergenerational mobility, needing only the row and column frequencies to finish the description.

EXAMPLE 4. ADJUSTING A JOINT DISTRIBUTION TO GIVEN MARGINS WHILE PRESERVING THE INTERACTIONS

Even though the heading above could be applied to Examples 2 and 3, the ideas involved in the next example, seem enough different to be worth special mention. The problem was originally called to my attention by Ithiel de Sola Pool in connection with a study of readership.

Pool had separate marginal distributions of several variables, such as sex, age, and education for a certain group of readers, but wished to use the joint distribution given by census data to reconstruct as best one can the joint dis-

TABLE 7. LEVINE'S ANALYSIS [15] OF GLASS'S BRITISH AND SVALASTOGA'S DANISH DATA. UPPER NUMBER OF CELL IS BRITISH, LOWER DANISH

Panel 7-1. ORIGINAL COUNTS

Status category of Son's Occupation

Status Category of Father's Occupation	1	2	3	4	5	
1	50 18	45 17	8 16	18 4	8 2	129 57
2	28 24	174 105	84 109	154 59	55 21	495 318
3	11 23	78 84	110 289	223 217	96 95	518 708
4	14 8	150 49	185 175	714 348	447 198	1510 778
5	3 6	42 8	72 69	320 201	411 246	848 530
	106 79	489 263	459 658	1429 829	1017 562	

Panel 7-2. STANDARDIZED COUNTS

Status category of Son's Occupation

Status Category of Father's Occupation	1	2	3	4	5	
1	68.5 58.6	20.9 25.0	4.6 12.0	3.7 2.6	2.3 1.8	100
2	17.8 21.1	37.5 41.6	22.5 21.9	14.7 10.3	7.5 5.1	100
3	8.0 11.7	19.2 19.3	33.7 33.7	24.3 21.9	14.9 13.5	100
4	4.1 4.1	14.7 11.4	22.6 20.7	31.1 35.5	27.6 28.4	100
5	1.6 4.5	7.8 2.7	16.6 11.8	26.2 29.8	47.8 51.2	100
	100	100	100	100	100	

tribution in his group of readers. In doing this, Pool wanted to "preserve the original interactions" in the census data. He was willing to suppose it reasonable that row, column, and layer multiplications would preserve the original interactions. It is important not to fool oneself here but to recognize that:

1) the concept of "preserving the original interactions" does not uniquely define the method of reconstructing cells for the joint distribution (indeed, for the 2×2, we have suggested in Example 2 another way to do it by doubly

dichotomizing the bivariate normal distribution for fixed ρ, and we noticed that the result would differ from that of the multiplicative model we plan to use here), and

2) making the assumption that when a subgroup is formed from the parent population the multiplicative invariance is preserved does not make it true (as a vehicle for getting an estimate in ignorance, it has the same status as the use of linear regression in the absence of knowledge of the shape of the function).

Our two-variable example of Table 6 could serve as a starter here. Read the table in 1,000's. Suppose that the original population was composed of 110 thousand men, 7 thousand women, 12 thousand high school graduates, and 105 thousand nongraduates. Then suppose that we had a subgroup with margins of 1 thousand men, 1 thousand women, and 1 thousand graduates and 1 thousand nongraduates. On the basis of the original joint distribution we wish to make a table (preserving the original association) that estimates the joint distribution in the subgroup. Our previous analyses showed that the estimate would be given by Table 6-4 which estimates 667 men-nongraduates and 667 women-graduates, 333 men-graduates and 333 women-nongraduates.

Far from being new, this iterative technique was suggested in 1940 by Deming and Stephan [5, 6] for the adjustment of tables to make margins fit properly. They did not have in mind the model of the invariance of association under multiplication of rows and columns and so on by positive constants, but thought of the method as an approximation to a least squares process that they were proposing. They extended the iterative scheme to more than two variables.

As Deming and Stephan describe [5], if one has three dimensions, then he would begin with some margin, say variable 1, and adjust every cell in a given layer by the same multiplicative factor so that the layer added up to the desired total. Next get the adjusted marginal totals on variable 2 and adjust each level of variable 2 by multiplying by the factor that makes them add up properly. This will mess up the margin for variable 1, but proceed to variable 3. Now start the cycle over again with variable 1. The convergence seems to be quite rapid.

Naturally, if one had more information about the subgroup, he could use it. For example one might have, as in Table 8, a joint distribution for the subgroup on a pair of variables, say x and y, and only marginal information on a third, say z. Then one would adjust every cell in a column for fixed x, y values to fit, and then adjust the layers for the levels of z and cycle. For example, suppose that the joint parent population is represented by the $2\times3\times2$ table shown in Panel 1 of Table 8. Let the levels of x, y, and z be x_1, x_2, y_1, y_2, y_3, z_1, z_2. Suppose that we know for the subgroup the joint distribution of x and y as shown in Panel 2 of Table 8, and the subgroup marginal distribution of z as shown in the bottom line of Panel 2—the numbers 9 and 7. (Let me emphasize that z is not thought to be independent of x and y in the subgroup, I am just describing the state of our knowledge of the subgroup.) Let us estimate the joint distribution of x, y, z in the subgroup, admitting now that the x, y association may not be preserved in the subgroup, but otherwise acting as if the multiplicative invariance held.

We begin the iteration by adjusting within each x, y column. For example, in the x_1, y_1 cell, the population margin is 15 and the subgroup margin 5, and

TABLE 8. EXAMPLE OF THREE-VARIABLE PARENT POPULATION

Panel 1: Population joint distribution

	z_1			z_2			Population x, y margin		
	y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
x_1	10	5	3	5	7	20	15	12	23
x_2	6	4	1	3	4	6	9	8	7
population z margin		29			45		grand total 74		

Panel 2: Desired subgroup joint distribution

	z_1			z_2			Subgroup x, y margin		
	y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
x_1							5	3	1
x_2			Cells to be filled in				1	2	4
Subgroup z margin		9			7		grand total 16		

Panel 3: Result of adjusting the x, y margin

	3.33	1.25	0.13	1.67	1.75	0.87
	0.67	1.00	0.57	0.33	1.00	3.43
Totals		6.95			9.05	
Subgroup margin		9			7	

Panel 4: Result of adjusting z margin—end of first cycle

4.32	1.62	0.17	1.29	1.35	0.67
0.86	1.29	0.74	0.26	0.77	0.65

Panel 5: Final subgroup table after 4 cycles carried with more decimals

3.97	1.73	0.22	1.04	1.27	0.78
0.79	1.31	0.97	0.21	0.69	3.03

so we adjust the 10 and 5 of the x_1, y_1 cells in layers z_1 and z_2 by multiplying each by 5/15. Similarly in the x_1, y_2 column we adjust 5 and 7 by multiplying by 3/12, and so on all through the x, y margin. The result is shown in Panel 3 of Table 8. The sum of the adjusted values in level z_1 is 6.95 and the sum of those in level z_2 is 9.05. These are used with the subpopulation margins 9 and 7 to adjust each value in the z_1 layer by multiplying by 9/(total of adjusted values in z level 1) $=9/6.95$ and by multiplying the adjusted values in z level 2 by 7/(total of adjusted values in z level 2) $=7/9.05$. This completes one cycle as shown in Panel 4 of Table 8. The iteration continues by readjusting the x, y table, and after 4 cycles the final result is shown in Panel 5 of Table 8.

EXAMPLE 5. BASELINE MODEL FOR PREFERENTIAL MARRIAGES

The anthropologist A. Kimball Romney [24], together with his colleagues, has been developing and analyzing a stochastic model for marriage in societies where there are rules or preferences for marriages between certain subgroups (perhaps clans).

THE BASIC MODEL

Categories. Romney conceives of men and women in the community as belonging to different categories, and, while he does not distinguish the categories for men from those for women, let us do that since the model might be useful for other problems. The men belong to categories $M_1, M_2, \cdots, M_K$, the women to $W_1, W_2, \cdots, W_K$. Making the numbers of categories equal for the two sexes loses no generality, because we can always add categories with no members.

Numbers. Let category M_i have m_i members and category W_j have w_j members, where $\Sigma m_i = N_{men}$, $\Sigma w_i = N_{women}$. These individuals are regarded as the pool from which marriages are made. For simplicity here we shall assume equal numbers of men and women in the pool, $N_{men} = N_{women} = N$.

Marriages. Romney conceives of the marriage process as having two-stages. In the first stage male-female pairs "meet"; in the second stage those who "meet" either marry or do not marry. (Do not read too much into the word "meet"; it is just a label for a stage. The notion of "meeting" gives a basis for the population distribution into the categories to have a selection effect. The second stage enables the differential probabilities (preferences) of the various categories for marrying to have effect.

Probabilities of meeting. At any instant there are m_i unmarried men and w_i unmarried women in categories M_i and W_j. The probability of an M_i and a W_j meeting is taken to be $m_i w_j / N^2$. This is a formal assumption. (Had there been unequal numbers of men and women we would use $m_i w_j / N_{men} N_{women}$.)

Probabilities of marriage given that a meeting of type M_i, W_j has occurred. Assume that the conditional probability of marriage is

$$P(\text{marriage} \mid \text{meeting } M_i, W_j) = \alpha_{ij}.$$

Here α_{ij} are given numbers, or numbers to be estimated from data.

In some examples Romney assumes that all α_{ii} have one value, and that all α_{ij}, $i \neq j$ have another value. If $\alpha_{ii} > \alpha_{ij}$, then marriage within the same category is preferred. The assumption of equality of the α_{ij}'s and the α_{ii}'s is mainly made to simplify the estimation problem that I do not discuss further here.

Process. A man-woman pair is drawn according to the probabilities $m_i w_j / N^2$. They marry or not according to the probability appropriate to their cell α_{ij}. If they marry, the number of individuals in their categories is reduced by unity. Thus, if M_i, W_j is the type of the chosen pair, the new numbers of men and women are

$$m_1, m_2, \cdots, m_i - 1, \cdots, m_K$$

$$w_1, w_2, \cdots, w_j - 1, \cdots, w_K.$$

And the totals are now reduced by unity, leaving $N-1$ unmarried men and $N-1$ unmarried women. A count of 1 (marriage) is assigned to cell M_i, W_j. The α_{ij} are unchanged.

If the couple does not marry, the meeting probabilities and the numbers of members in the cells are unchanged, as are the α_{ij}.

Next, draw another man-woman pair and continue in this manner until all the individuals are married. At the close there will have been some number of marriages x_{ij} in cell M_i, W_j.

THE PROBLEM

The general problem is to discover the properties of the process and this might include knowing the joint distribution of X_{ij} (the random variables for the number of marriages in the cells) at any stage in the process. The more immediate and much more modest goal is to find the expected number of marriages in a cell when the process runs until all are married.

Special case of two categories for each sex. Limiting the discussion to two categories after all the notation above seems like a great comedown, but the basic idea of a solution can be obtained from this case. It will also clarify ideas.

Illustration. Let us study a small example with α_{ij}:

	W_1	W_2
M_1	.8	.4
M_2	.4	.8

and initial population

	W_1	W_2	
M_1			1
M_2			1
	1	1	2

After the first draw resulting in a marriage the situation is one of the four shown in Figure 1, where the number in the cell is the number of marriages in that cell so far, and the margins indicate the numbers unmarried. The fractions at the left of the tables are the probabilities of getting the first marriage in the indicated cell. We explain the calculation of these probabilities shortly.

$\frac{1}{3}$

1		0
		1
0	1	

$\frac{1}{6}$

	1	0
		1
1	0	

$\frac{1}{6}$

		1
1		0
0	1	

$\frac{1}{3}$

		1
	1	0
1	0	

FIG. 1. Situation after first marriage.

The requirements of the margins now force the unmarried couple to fall into the cell diagonally opposite to the position of the first marriage. Before we compute the expected number of marriages in a cell, we need the probabilities corresponding to the four tables.

$$P(\text{cell } 1,\, 1 \text{ and marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times .8.$$
$$P(\text{cell } 1,\, 2 \text{ and marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times .4.$$
$$P(\text{cell } 2,\, 1 \text{ and marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times .4.$$
$$P(\text{cell } 2,\, 2 \text{ and marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times .8.$$
$$P(\text{marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times 2.4.$$

Then the conditional probabilities are:

$$P(\text{cell } 1,\, 1 \mid \text{marriage}) = \tfrac{1}{2} \times \tfrac{1}{2} \times .8/\tfrac{1}{2} \times \tfrac{1}{2} \times 2.4 = \tfrac{1}{3}.$$
$$P(\text{cell } 1,\, 2 \mid \text{marriage}) = \tfrac{1}{6}.$$
$$P(\text{cell } 2,\, 1 \mid \text{marriage}) = \tfrac{1}{6}.$$
$$P(\text{cell } 2,\, 2 \mid \text{marriage}) = \tfrac{1}{3}.$$

We compute the expected number in the upper left-hand cell; it is

$$\tfrac{1}{3} \times 1 + \tfrac{1}{6} \times 0 + \tfrac{1}{6} \times 0 + \tfrac{1}{3} \times 1 = \tfrac{2}{3}.$$

The final $\frac{1}{3} \times 1$ corresponds to the lower right-hand table, where the second marriage added a 1 to its upper left-hand corner. The other mean values can be gotten by subtraction and they are:

$\frac{2}{3}$	$\frac{1}{3}$	1
$\frac{1}{3}$	$\frac{2}{3}$	1
1	1	2

We can develop a recurrence relation for the mean number of marriages in a cell for the 2×2 problem. We need some notation.

Recurrence relation for the mean. Consider the marriage probabilities

α_{12}	α_{12}
α_{21}	α_{22}

and the initial population

	W_1	W_2	
M_1			m
M_2			$\bar{m}$
	w	$\bar{w}$	N

In the 2×2 case the notation m and its complement with respect to N, $\bar{m}(=N-m)$, is especially convenient, similarly w and $\bar{w}=N-w$.

After the first marriage, we have one of the following:

1		$m-1$
		$\bar{m}$
$w-1$	$\bar{w}$	$N-1$

	1	$m-1$
		$\bar{m}$
w	$\bar{w}-1$	$N-1$

		m
1		$\bar{m}-1$
$w-1$	$\bar{w}$	$N-1$

		m
	1	$\bar{m}-1$
w	$\bar{w}-1$	$N-1$

The probabilities of achieving these four tables are computed as follows:

$$P(\text{cell } 1,\, 1 \text{ and marriage}) = mw\alpha_{11}/N^2.$$
$$P(\text{cell } 1,\, 2 \text{ and marriage}) = m\bar{w}\alpha_{12}/N^2.$$
$$P(\text{cell } 2,\, 1 \text{ and marriage}) = \bar{m}w\alpha_{21}/N^2.$$
$$P(\text{cell } 2,\, 2 \text{ and marriage}) = \bar{m}\bar{w}\alpha_{22}/N^2.$$
$$P(\text{marriage}) = \text{total of above} = \Sigma/N^2, \quad \text{where}$$
$$\Sigma = mw\alpha_{11} + m\bar{w}\alpha_{12} + \bar{m}w\alpha_{21} + \bar{m}\bar{w}\alpha_{22}.$$

Then the conditional probabilities are

$$P(\text{cell } 1,\, 1 \mid \text{marriage}) = mw\alpha_{11}/\Sigma = P_{11}.$$
$$P(\text{cell } 1,\, 2 \mid \text{marriage}) = m\bar{w}\alpha_{12}/\Sigma = P_{12}.$$
$$P(\text{cell } 2,\, 1 \mid \text{marriage}) = \bar{m}w\alpha_{21}/\Sigma = P_{21}.$$
$$P(\text{cell } 2,\, 2 \mid \text{marriage}) = \bar{m}\bar{w}\alpha_{22}/\Sigma = P_{22}.$$

In making the calculation for the expected value in the upper left-hand cell (1, 1) we could use the notation $E(X \mid m, w, N)$ or, more briefly, $E(m, w, N)$, since m, w, and N are enough to determine the initial population.

Then we can immediately write:

$$E(m, w, N) = P_{11}\{1 + E(m-1, w-1, N-1)\} + P_{12}E(m-1, w, N-1) + P_{21}E(m, w-1, N-1) + P_{22}E(m, w, N-1). \tag{3}$$

The first marriage for the upper left-hand corner table adds 1 to the number of marriages to be expected from the table with index $(m-1, w-1, N-1)$. None of the other tables get additions.

The basic formula can be used to generate expected values for any 2×2 initial population. A computer program has been written and expected values to four decimals for all possible initial populations up to and including $N=49$ have been computed for a few sets of α's.

The basic formula can readily be generalized to more categories, but in such generalization, one must have more formulas. Here we have, in principle, four formulas for a two by two table because $2\times 2=4$. But since the table has fixed margins, there is but one degree of freedom. Therefore, computing for the (1, 1) cell is enough. In a 3×3 table there would be four formulas needed corresponding to the $(3-1)(3-1)=4$ degrees of freedom in a 2×2 table with fixed margins. More generally, $(k-1)^2$ formulas would be required.

Approximation. Naturally, it would be good to have an easier way to compute the expected value than a recurrence relation, because it requires a great deal of calculation just to get the answer for one table. One approximation which Romney has used regards the α's as the basic numbers determining the degree of association in a contingency table. We then find what multipliers for rows and for columns will produce the required margins either by the iterative scheme in large tables or by equation (2*). In that formula we replace $\overline{R}$ and $\overline{C}$ by $\bar{m}$ and $\bar{w}$, and $\alpha=\alpha_{11}\alpha_{22}/\alpha_{12}\alpha_{21}$. In an example:

.8	.4
.4	.8

$1\frac{1}{3}$	$\frac{2}{3}$	2
$\frac{2}{3}$	$1\frac{1}{3}$	2
2	2	4

We compare the iterative answer with the exact result from the recurrence relation for the same α's for initial populations of size 40. Although we have the whole printout both for the iterative scheme and for the exact recurrence relation, a more modest table will spare the eyes and increase understanding.

Table 9 compares the exact results with the solution of quadratic equation (2*) for a grid of 81 tables from the possible 1681 $(=41^2)$ tables with $N=40$. On the edges of the square, the two formulas agree perfectly. Note that the largest deviation is only 0.21.

EXAMPLE 6. REPORTING DATA BASED ON CONTINGENCY TABLES AS IN THE NATIONAL HALOTHANE STUDY

In thinking about reporting on his data, the investigator asks how to summarize them, but there are other issues. For example, is it possible that there are better reports than the data itself? Not in the usual sense of completeness. But suppose that three somewhat comparable surgical operations are repeatedly

TABLE 9. A COMPARISON OF EXACT AND APPROXIMATE EXPECTED VALUES IN UPPER LEFT-HAND CORNER OF 2×2 TABLE FOR ROMNEY'S MARRIAGE MODEL WITH $N=40$, VARIOUS COMPOSITIONS OF MEN AND WOMEN INTO THE TWO CATEGORIES, m AND w BEING THE NUMBERS OF MEN AND OF WOMEN IN CATEGORY 1, AND MARRIAGE PROBABILITIES

.8	.4
.4	.8

m \ w		0	5	10	15	20	25	30	35	40
	Exact	0								
0	Approx.	0								
	Diff.	0								
	Exact	0	1.50							
5	Approx.	0	1.52							
	Diff.	0	.01							
	Exact	0	2.52	4.54						
10	Aprox.	0	2.58	4.64						
	Diff.	0	.06	.10						
	Exact	0	3.23	6.08	8.47					
15	Approx.	0	3.33	6.22	8.64					
	Diff.	0	.10	.14	.17					
	Exact	0	3.74	7.24	10.40	13.13				
20	Approx.	0	3.86	7.42	10.60	13.33				
	Diff.	0	.11	.18	.20	.20				
	Exact	0	4.15	8.15	11.92	15.40	18.47			
25	Approx.	0	4.26	8.33	12.13	15.60	18.64			
	Diff.	0	.11	.18	.20	.20	.17			
	Exact	0	4.48	8.87	13.15	17.24	21.08	24.54		
30	Approx.	0	4.57	9.02	13.33	17.42	21.22	24.64		
	Diff.	0	.08	.15	.18	.18	.14	.10		
	Exact	0	4.76	9.48	14.15	18.74	23.23	27.52	31.50	
35	Approx.	0	4.80	9.57	14.26	18.86	23.33	27.58	31.52	
	Diff.	0	.04	.08	.11	.11	.10	.06	.01	
	Exact	0	5.	10.	15.	20.	25.	30.	35.	40.
40	Aprox.	0	5.	10.	15.	20.	25.	30.	35.	40.
	Diff.	0	0	0	0	0	0	0	0	0

(1) Exact values were computed using recurrence formula.

(2) Approximate values were computed from quadratic formula (2*).

(3) Difference = Approximation minus Exact. The values above the main diagonal are not shown because the Table is symmetrical with respect to this main diagonal. All computations, including differences, were made to 4 decimals and truncated to 2.

performed under several different conditions, is it possible that we can do better than to report the observed death rate for each operation under each condition? Should we make a distinction between the table of observed death rates and the table that we provide for the interested physician? A typical weakness of observed death rates is that some cell—an operation performed under a specific condition—will turn out to have very few cases compared to the numbers of operations performed under the other conditions. The investigator hesitates to report a death rate based on a very few cases, especially if the other cells are based upon many. Could we provide a better report for this cell than either a blank or the observed death rate? Although we have first couched the question in terms of one weak cell, once the idea is clear that there may be information in the other cells relevant to the particular cell (Mosteller [17]), then the more general question arises, should we not let the data from one cell support the data in another? And if so what are good ways to do this?

In this discussion there is danger of unintentionally leading the reader off into irrelevant problems of ethics, of destroying data, or of misleading the user. To avoid such distractions, we shall suppose that the original data are not only not lost, but that they are also presented to the user. The question is What report in addition to the original data would be especially useful?, for example, to the physician. (We do, however, need to consider how to warn him that the report is not simply based upon the observed counts in the cell.)

As my example, I shall draw on The National Halothane Study's analysis of death rates following surgery [4]. That study compared the safety of the major anesthetics (halothane, nitrous oxide-barbiturate, cyclopropane, ether) as used in surgery in 34 institutions.

In every field of statistical activity where we have contingency tables involving many variables, we find that our data thins out so that we cannot make cell-to-cell comparisons at as fine a level as we could desire. The reasons are well known. The product of the number of levels associated with each variable is the number of cells that has to be considered, and this number gets large very rapidly. For example, in The National Halothane Study we could have had 34 institutions, 5 anesthetics, 2 sexes, 7 ages, 8 levels of physical status, and these alone (without even considering the 75 different operations) would have accounted for 19,000 cells. Yet the study of surgical deaths, a very large one, had only 16,000 deaths based on 856,000 surgical procedures. Obviously, one death per cell could scarcely lead to reliable rates for making all the desirable comparisons. Naturally, one wants to control for variables such as those we have mentioned—different institutions have quite different death rates, there is almost a factor of 2 in the ratio of death rates for males as compared to females, and so on. Consequently, when we come to compare death rates following various anesthesias and surgery, it is important that we take account of the different frequencies with which different classes of patients participate in the use of the different anesthetics.

In the same manner in studies in the social sciences we need to control for many variables in order to make our results more meaningful or more precise. Several different ways of handling the problem of the small numbers in the cells have been developed in the course of the halothane study. One of these methods has been developed by Yvonne Bishop in her doctoral dissertation [3]. Part of this work appears in the full halothane report [4, Chapter 3]. The fundamental idea has been developed by Birch [2].

The general idea is as follows. When we fit a multivariate normal distribution to a set of data, we are fitting means, variances, and covariances. This is equivalent to fitting some parameters of several two-dimensional marginal distributions. Frequently this amount of fitting can be used to represent the multivariate data well. In contingency tables, we could not ordinarily make assumptions of normality (though Jerome Cornfield tells me that sometimes it works unexpectedly well), but we could consider fitting all two-dimensional and even all or some three-dimensional margins to the data. This approach would borrow strength for the cell estimation from the margins by giving up the contributions of the higher order effects. For example, if we had 6 variables and fitted three-dimensional margins, then we would make sure that all 20 sets of three variables were properly represented by their frequencies at each level of each variable, by their two-dimensional associations as represented by all 15 two-way marginal tables, and even their second-order interactions as represented by their 20 three-dimensional margins, one for every triple of variables. What would be sacrificed would be the third- and higher-order effects arising from the special interrelations of sets of 4 or more variables. This is not to deny the possible existence of such third-order interactions. Rather the thought is that it may be profitable for cell estimation purposes to ignore them. A simple analogy arises in ordinary regression where we often fit a straight line or a parabola when we know very well that the true state of affairs is more complicated, and sometimes even when we have a clear picture of how the true state of affairs should be represented. We do this because our estimation may be better when using a simple model in a limited range, even though the model is wrong, than when fitting the correct model.

Bishop has explored the fitting of two- and three-dimensional margins to these large sets of data, and has developed ways of assessing what models (what sets of margins) give the better representation.

These ideas applied to a contingency table relate to the multiplicative model, and therefore to the idea of taking logarithms of cell entries and treating these as measurements through the usual analysis of variance methods—not of testing but of estimation. If we think of the cell entries as the true probabilities and suppose that these are not zero, then the case of complete independence leads us to think of replacing the observed entries by their logarithms, which would just be the sum of the logarithms of the marginal probabilities, whatever the number of variables of the table.

Corresponding to this idea for populations, suppose that we want to fit the cells of an observed two-way contingency table but neglect the possible "interaction" effects. All we do is compute the usual table of "expected values" either by multiplication or by logarithmic methods, using the two one-variable margins.

In a similar vein, we can, if we wish, construct a logarithmic model with some but not necessarily all possible interactions, for three or more variables. Fitting less than the full model, but more than that for complete independence arises only when we have three or more variables. Therefore we do not think much about it as long as we deal with one- and two-way tables.

Many practical applications were made of this idea in The National Halothane Study to help with the comparison of anesthetics. In addition by its use we were able to come much closer to representing the unique case than usual. Ordinarily when a physician asks about death rates associated with operations, he has only single variables or at most pairs upon which to appraise the patient's chances. In this study we were able to construct tables that use sex, age, physical condition, and operation all simultaneously. Thus a physician considering an operation would be able to consider past results for a class of patients that represented his patient's properties better than usual because more variables would be fitted. (If I were redoing the Study now, I would use both a finer classification and more variables.)

Table 10 shows a portion of the death rates in the halothane study for four frequently-performed operations that have high death rates. The rates were obtained by fitting all 6 two-dimensional margins for the 4 variables mentioned above. The tables of deaths and the tables of estimated exposed were fitted separately and then the death rates were estimated from 100×fitted deaths/(fitted estimated exposed+fitted deaths) to reduce the mean square error (see [4, Part IV, Chapter 2, Appendix 3]).

I shall describe the fitting of the cells of a three variable contingency table by fitting all the two-variable margins but allowing no third order effects. The method is closely related to the one that we used earlier in Example 4 to adjust the cells so that we got a certain two-way margin and another one-variable margin. This time, the cell values are available for the determination of the three two-way margins say xy, xz, and yz, but we wish to reconstitute the body of the table omitting the effects of three dimensional interaction.

TABLE 10. SMOOTHED DEATH RATES FOR THE FOUR FREQUENTLY-PERFORMED OPERATIONS WITH HIGH DEATH RATES[1]

Operation[2]		Male				Female			
		1	2	3	4	1	2	3	4
	10–49	1.4	1.7	4.0	9.2	1.2	1.2	4.8	9.7
Risk A[3]	Cases[4]	17	24	29	8	17	21	28	9
	50–69	2.4	3.6	6.1	16.8	2.7	5.5	9.5	22.2
	Cases	27	21	17	1	25	17	15	1
	10–49	8.7	9.2	13.5	16.8	7.5	10.4	15.8	17.6
Risk B[3]	Cases	3	8	8	11	3	7	8	12
	50–69	11.4	14.6	15.8	23.8	12.8	20.8	23.1	30.6
	Cases	7	10	7	2	7	8	6	2
	10–49	64	38	57	37	44	27	46	25
Risk C[3]	Cases	—	.4	.4	.1	.1	.9	.8	.3
	50–69	60	40	51	37	48	35	47	31
	Cases	.1	.6	.3	—	.2	1.1	.7	—

[1] Estimated death rate = 100×deaths/(estimated exposed+deaths).

[2] 1: Large bowel; 2: exploratory laparotomy; 3: craniotomy; 4: heart and great vessel with pump.

[3] A: Low risk, anesthetic ratings 1, 2, 5; B: higher risk, anesthetic ratings 3, 4, 6; C: moribund, anesthetic rating 7.

[4] Estimated in hundreds.

If we replace every cell in the body of the table by a 1, we have a starting table that represents complete independence. (Alternatively, we could start with a table estimated by the product of the 3 one-dimensional marginal cells. In the usual notation the estimate for cell ijk would be $NX_{i++}X_{+j+}X_{++k}/N^3$, the usual generalization of the "expected value" for a cell based on marginal totals and complete independence. Nothing much seems to be gained by using these "expected values" instead of 1's as initial values; but the point of the 1's is that they stand for independence.)

Then we compute all the two-way margins for this starting table. Again we apply the iteration scheme. For example, suppose that we want to fill in a $2\times2\times2$ table so that all the two-way margins fit but there is no three-variable effect in the table. Then we pick one two-way marginal table, say the xy table and adjust for each xy pair all the z's (here 1's) by multiplying them by the correct marginal total from the original table, divided by the sum obtained from the starting table just as we did in Example 4. After this has been done for every xy pair, one continues in the same way with the xz pairs and then the yz pairs. Usually totals will not add up properly at the close of this operation and so one will need to continue the iteration, cycling through each two-way table systematically. Again the convergence is usually rather rapid. In the special case of the $2\times2\times2$ table there is a check that one can make by multiplying certain cells together. If the layout is

Layer 1

a	b
c	d

Layer 2

e	f
g	h

then $adgf=cbeh$ when there is no three-variable effect (Bartlett's formula [1]). Table 11 sketches an example.

This idea extends smoothly to several variables with various numbers of levels for each variable. It is also not restricted to the fitting of all two-way margins. In the presence of several variables, one can fit all three-way margins, for example, and we have done this for some tables in The National Halothane Study.

We are not restricted to fitting all margins of one size. For example in a five-variable problem with variables numbered 1, 2, 3, 4, 5 we could fit the three-way margin for variables 1, 2, 3, fit the two-way margin of variable 4 with each of 1, 2, and 3, and independently fit the one variable margin of variable 5. Bishop's work [3] on fitting and testing in such complicated problems is instructive.

EXAMPLE 7. BAYESIAN APPROACH TO ESTIMATION

The idea of "general purpose" reporting of data has been emphasized above. It may occur to you that the method just described is rather discontinuous from the point of view of dimensionality. For example, when we fit the two-way margins we allow all one-variable margins to have their full effect, and we allow all two-variable effects to be fully represented, but then we suddenly and

TABLE 11. ITERATING A 2×2×2 TABLE SO AS TO HAVE NO THREE-VARIABLE EFFECT

STARTING TABLE

z +

	y +	y −
x +	1	1
x −	1	1

z −

	y +	y −
x +	1	1
x −	1	1

All starting margins are 2's.

DESIRED MARGINS

	z +	z −
y +	.30	.20
y −	.15	.35

	z +	z −
x +	.25	.30
x −	.20	.25

	y +	y −
x +	.35	.20
x −	.15	.30

STARTING 2×2 MARGINS ARE ALL 2'S. APPLY *yz* FIRST AND GET

z +

	y +	y −
x +	.15	.075
x −	.15	.075

z −

	y +	y −
x +	.10	.175
x −	.10	.175

new *xz* margin

	z +	z −
x +	.225	.275
x −	.225	.275

APPLYING *xz* GIVES

z +

	y +	y −
x +	.1667	.0833
x −	.1333	.0667

z −

	y +	y −
x +	.1091	.1909
x −	.0909	.1591

new *xy* margin

	y +	y −
x +	.2758	.2742
x −	.2242	.2258

END OF FIRST FULL CYCLE GIVES:

z +

	y +	y −
x +	.2115	.0608
x −	.0892	.0886

z −

	y +	y −
x +	.1385	.1392
x −	.0608	.2114

AFTER 4 CYCLES, ROUNDING TO 2 DECIMALS GIVES

z +

	y +	y −
x +	.20	.05
x −	.10	.10

z −

	y +	y −
x +	.15	.15
x −	.05	.20

completely disallow effects of three or more variables simultaneously. This is not the only possibility. One ought to be able to allow effects of different order to contribute with differential weights. Ideas like this have been suggested by Stein [25] and Lindley [16], and of course, I. J. Good [9] has explored various approaches to cell estimation in contingency tables.

Recently Fienberg and Holland [8] have added to this estimation literature. Let us take it in two steps, first for the multinomial and then for the two-way table.

Multinomial. Let there be t multinomial categories with probabilities $p_1, p_2, \cdots, p_t$ and observations X_i corresponding to these categories, where $\Sigma X_i = N$. Then if T_i is the estimate of p_i and we use loss function $\Sigma(T_i - p_i)^2$ and have symmetric Dirichlet priors, the Bayes estimate for category i is

$$B_k = \frac{N}{N+k}\left(\frac{X_i}{N}\right) + \frac{k}{N+k}\left(\frac{1}{t}\right) \tag{4}$$

and the value of k which minimizes the risk for a specific set of p's is

$$k^* = \frac{1 - \Sigma p_i^2}{\Sigma p_i^2 - \dfrac{1}{t}}.$$

We can estimate k^* by

$$\hat{k} = \frac{1 - \Sigma(X_i/N)^2}{\Sigma(X_i/N)^2 - \dfrac{1}{t}}, \tag{5}$$

and perhaps we could jackknife it [20, Chapter E] to unbias it and to get some notion of its uncertainty. This estimate is closely related to Good's result [9].

Application 1. A sample of size $N=51$ was drawn from 6 multinomial categories as shown in Table 12. The estimated value of k was 1.65, and the jackknifed value 1.61. Using it we get the estimate in the last line of Table 12. Although the usual estimate of 0 was increased to 0.5% and 65% was reduced to 63%, the total change from X_i/N to B_k is small. The standard error of $\hat{k}$ is 0.55, according to the jackknife.

Application 2. To see what happens when data are very sparse consider Table 13 which has several 0's and few data otherwise (a fragment from Doll

TABLE 12. OUTCOME OF SAMPLING EXPERIMENT

Category	1	2	3	4	5	6	Total
Probabilities p_i	0.001	0.01	0.049	0.20	0.24	0.50	
Sample	0	1	1	4	12	33	51
x_i/n	0	0.020	0.020	0.078	0.235	0.646	1.000
$B_{1.61}$	0.005	0.024	0.024	0.081	0.233	0.632	1.000

$s_{\hat{k}} = 0.55$

and Hill's study of the mortality of British doctors [7, p. 240]). This time the value of $\hat{k}=11$ and the effect on the estimates is more substantial, the 0's increased to 2.3% and the 19% reduced to 16%.

TABLE 13. DEATHS FOR MALE PIPE AND CIGAR SMOKERS CONTINUING TO SMOKE 25 OR MORE g/DAY.

Age	25 29	30 34	35 39	40 44	45 49	50 54	55 59	60 64	65 69	70 74	75 79	80 84	85 ∞	Total
Deaths	0	0	0	0	1	1	5	0	3	5	5	3	3	26
x_i/n	0	0	0	0	.038	.038	.192	0	.115	.192	.192	.115	.115	1.000
B_{11}	.023	.023	.023	.023	.050	.050	.158	.023	.104	.158	.158	.104	.104	1.000

$\hat{k}=11$

Two-way tables. Estimation in two-way tables is more in the spirit of the rest of our discussion. One might suppose that an estimate in a two-way table would be composed of two components, the first based upon the "expected value" as computed in the usual chi-square analysis, and the second based on the actual observation. Thus we would take advantage both of the strength of estimation available when we have independence and of the two-variable effect present in the cell itself. Using the quadratic loss mentioned earlier, Fienberg and Holland [8] get the following weighted sum as an optimum estimate

$$\frac{N}{N+k}\left(\frac{X_{ij}}{N}\right)+\frac{k}{N+k}\left(\frac{X_{i+}X_{+j}}{N^2}\right);$$

$$X_{i+}=\sum_j X_{ij},\quad X_{+j}=\sum_i X_{ij},\quad N=\sum_{ij} X_{ij}. \tag{6}$$

Table 14 gives observed counts and fitted counts based on this formula (multiplied by N) for Kahn's data from Dorn's study of smoking and mortality among U.S. Veterans [12, p. 119]. In the present instance, the counts are not adjusted much—$k/(N+k)$ is about 1/14 and so we move about 7% toward the "expected value" estimate from the "observed value".

EXAMPLE 8. REGRESSION ESTIMATES FOR CELL ENTRIES IN CONTINGENCY TABLES

If you think of estimating probabilities in a table containing several empty cells as a hazardous occupation, let me introduce you to some work of Leo Goodman [10] who estimates *all* the cells in a two-way contingency table on the basis only of the margins, without assuming independence.

He performs this magic by making certain assumptions about a set of two-way tables, such as that they all have the same transition probabilities, and then he uses the variation in the margins from one table to another as a lever for prying out information through a regression approach.

If n_{ij} is the entry in the i^{th} row and j^{th} column of a 2×2 table and only the marginal distributions are known, then Goodman writes the sum of the first column as

$$n_{+1}=n_{11}+n_{21}.$$

Multiply and divide the first term on the right by the first row total n_{1+}, and the second term by the second row total n_{2+} to get:

$$n_{+1} = (n_{11}/n_{1+})n_{1+} + (n_{21}/n_{2+})n_{2+}. \tag{7}$$

Let $p_{11}=n_{11}/n_{1+}$ and $p_{21}=n_{21}/n_{2+}$ be the transitional probabilities leading from row 1 to column 1 and from row 2 to column 1 respectively. Substituting into (7) gives:

$$n_{+1} = p_{11}n_{1+} + p_{21}n_{2+}. \tag{8}$$

Since n_{+1}, n_{1+}, and n_{2+} are known in any table, equation (8) can be regarded as a relation between p_{11} and p_{21}, or the data n_{1+}, n_{2+}, and n_{+1} can be regarded as data for use in estimating p_{11}, p_{21}. The top panel of Table 15 gives an example of the relation between position in labor force and type of residence in three states. Pretend that we do not know the cell entries in the body of the table. Then using equation (8) we estimate

$$\hat{p}_{11} = 0.54 \qquad \hat{p}_{21} = 0.49$$

$$\hat{\sigma}_{\hat{p}_{11}} = 0.004 \qquad \hat{\sigma}_{\hat{p}_{21}} = 0.005 \qquad \hat{\sigma}^2 = 6.63 = \text{residual sum of squares.}$$

TABLE 14. OBSERVED AND FITTED COUNTS FOR DEATHS FROM CORONARY HEART DISEASE IN VETERANS AGED 65–74.

ORIGINAL OBSERVATIONS

AGE BEGAN	YEARS SINCE STOPPED SMOKING 1–4	5–9	10–14	15+	
<15	—	4	3	5	12
15–19	4	34	28	47	113
20–24	7	20	27	30	84
25+	1	26	21	21	69
	12	84	79	103	278

FITTED VALUES $\hat{k}=22.98$

	1–4	5–9	10–14	15+	
<15	.04	3.97	3.03	4.96	12.
15–19	4.07	34.01	28.31	46.61	113.
20–24	6.74	20.41	26.76	30.09	84.
25+	1.15	25.61	20.89	21.35	69.
	12.	84.	79.	103.	278.

If only the marginal distributions had been known, the estimates of p_{11} and p_{21} would have led to the estimated tables shown in the bottom panel of Table 15. Note that these fitted values agree well with the observed values, which will not be known in problems where we do such fitting.

The regression equation for each table can be represented as a line in the p_{11}, p_{12} plane. The reason the regression model works so well in this example is that the three lines intersect almost at the same point and that the true parameter values all lie close to the points of intersection. But the true point could

TABLE 15. RELATION (IN 000'S) BETWEEN LABOR FORCE AND RESIDENCE IN THREE STATES, U. S. CENSUS 1950

OBSERVED TABLES

	Alabama Labor Force	*Alabama* Not Labor Force		*Maine* Labor Force	*Maine* Not Labor Force		*Missouri* Labor Force	*Missouri* Not Labor Force	
Urban	539	441	980	194	164	358	1040	870	1910
Rural	545	582	1127	152	164	316	536	576	1112
	1084	1023	2107	346	328	674	1576	1446	3022

FITTED TABLES

	Alabama Labor Force	*Alabama* Not Labor Force	*Maine* Labor Force	*Maine* Not Labor Force	*Missouri* Labor Force	*Missouri* Not Labor Force
Urban	529	451	193	165	1031	879
Rural	552	575	155	161	545	567

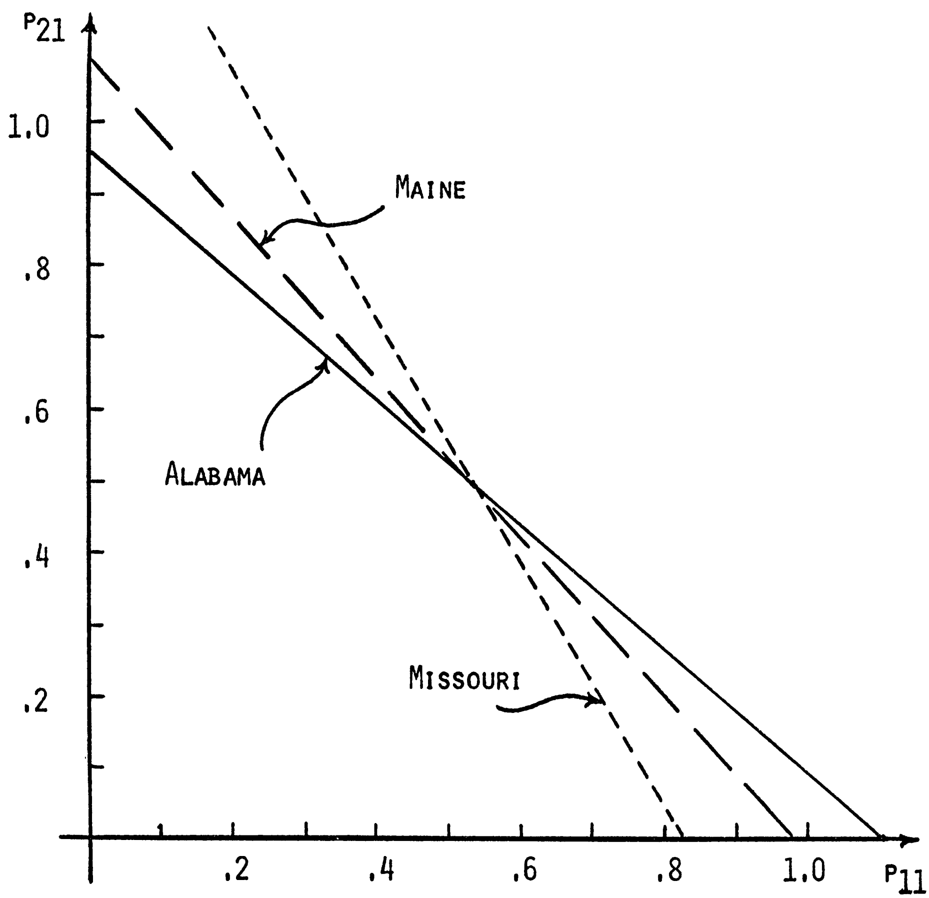

FIG. 2. The three regression lines (equation (7)) corresponding to the margins of the tables for Alabama, Maine, and Missouri shown in Table 15.

lie anywhere on the lines, consequently a common intersection is *no guarantee* that the true values have been found.

We can use considerably more empirical research on the applicability of this method.

What with Bayesian help, perhaps we will be able to get along without the margins too!

SUMMARY

In this paper I have stressed estimation rather than testing in the analysis of complex contingency tables and I have called attention to an attitude toward association different from the usual. I especially illustrated a statistically neglected notion that I call "reporting" (perhaps I should call it "statistical reporting") for probabilities and rates obtained from contingency tables. I feel that further research, both mathematical and empirical, can improve statistical reporting and make it more useful to practitioners such as social scientists and physicians by emphasizing estimation for very detailed classes of individuals. Although the statistician cannot hope to match the unique case, I believe that he can take steps to come much closer to it in the future than he has in the past. Although I have illustrated the reporting problem here for contingency tables involving several variables, I do not restrict it to them. The problem is that of estimating for general purposes the values of a function of several variables.

ACKNOWLEDGMENT

I wish to express my appreciation to the following who have contributed directly to the preparation of this paper: Yvonne M. M. Bishop, Gudmund Iversen, Stephen Fienberg, Paul W. Holland, Joel H. Levine, William S. Mosteller, A. Kimball Romney, and Cleo Youtz. I must also note that the contribution of many authors to this field of work have been regretfully omitted here in the interests of brevity and unity. I hope readers will not hesitate to send me additional references.

REFERENCES

[1] Bartlett, M. S., "Contingency table interactions," *Journal of the Royal Statistical Society*, Series B, 2(1935), 248–52.

[2] Birch, M. M., "Maximum likelihood in three-way contingency tables," *Journal of the Royal Statistical Society*, Series B, 25(1963), 220–33.

[3] Bishop, Yvonne M. M., *Multidimensional Contingency Tables: Cell Estimates*, Ph.D. Thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts, February, 1967.

[4] Bunker, J. P., Forrest, W. H., Jr., Mosteller, Frederick, and Vandam, L. D. (editors), *The National Halothane Study. A study of the possible association between halothane anesthesia and postoperative hepatic necrosis*. Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia, National Academy of Sciences—National Research Council. In press, will be available from Division of Medical Sciences, National Research Council, 2101 Constitution Avenue, Washington, D. C. 20418.

[5] Deming, W. Edwards, and Stephan, Frederick F., "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *Annals of Mathematical Statistics*, 11(1940), 427–44.

[6] Deming, W. Edwards, and Stephan, Frederick F., "The sampling procedure of the

1940 population census," *Journal of the American Statistical Association*, 35(1940), 615–30.

[7] Doll, R., and Hill, A. B., "Mortality of British doctors in relation to smoking: Observations on coronary thrombosis," in *Epidemiological approaches to the study of cancer and other chronic diseases*, National Cancer Institute Monograph 19, edited by William Haenszel. Washington, D. C., Superintendent of Documents, U. S. Government Printing Office, January, 1966, 205–68.

[8] Fienberg, Stephen E., and Holland, Paul W., "Multinomial estimation and two-way contingency tables," Memorandum NS-95, Department of Statistics, Harvard University, December 5, 1967.

[9] Good, I. J., *The Estimation of Probabilities*, Cambridge, Mass., Massachusetts Institute of Technology Press, 1965.

[10] Goodman, Leo, "Some alternatives to ecological correlation," *American Journal of Sociology*, 64(1959), 610–25.

[11] Goodman, Leo A., "Simultaneous confidence limits for cross-product ratios in contingency tables," *Journal of the Royal Statistical Society*, Series B, 26(1964), 86–102.

[12] Kahn, Harold A., "The Dorn study of smoking and mortality among U. S. veterans: Report on eight and one-half years of observation," in National Cancer Institute Monograph 19, 1–125, especially table on p. 119. See [7].

[13] Kučera, Henry, and Francis, W. Nelson, *Computational Analysis of Present-Day American English*, Providence, Rhode Island, Brown University Press, 1967.

[14] Levine, Joel Harvey, *Measurement in the study of intergenerational status mobility*, Ph.D. Thesis, Department of Social Relations, Harvard University, Cambridge, Massachusetts, March, 1967.

[15] Levine, Joel H., "A measure of association for intergenerational status mobility," Department of Social Relations, Harvard University, unpublished manuscript, December, 1967.

[16] Lindley, D. V., Discussion of Stein's paper. See [25].

[17] Mosteller, Frederick, "On pooling data," *Journal of the American Statistical Association*, 43(1948), 231–42.

[18] Mosteller, Frederick, "Analysis of some baseline models for preferential marriages." Memorandum NS-40, Department of Statistics, Harvard University, June 9, 1966.

[19] Mosteller, Frederick, "Statistical comparisons of anesthetics: the National Halothane Study," In press, *Bulletin of the International Statistical Institute*.

[20] Mosteller, Frederick, and Tukey, John W., "Data analysis, including statistics," Chapter for revised *Handbook of Social Psychology*, Lindzey, Gardner and Aronson, Eliot (editors). In press, Addison-Wesley.

[21] Mosteller, Frederick, and Wallace, David L., *Inference and Disputed Authorship: The Federalist*, Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1964.

[22] Pearson, Karl (editor), *Tables for Statisticians and Biometricians*, Part II. Issued by the Biometric Laboratory, University College, London, 1931. Table VIII and IX, 78–137.

[23] Plackett, R. L., "A class of bivariate distributions," *Journal of the American Statistical Association*, 60(1965), 516–22.

[24] Romney, A. Kimball, "A model for the analysis of endogomy," in Kay, Paul (editor), *Explorations in Mathematical Anthropology*, In press, Massachusetts Institute of Technology Press.

[25] Stein, C. M., "Confidence sets for the mean of a multivariate normal population," with discussion, *Journal of the Royal Statistical Society*, Series B, 24(1962), 265–85.