95228

DEPARTMENT

OF

STATISTICS

Carnegie-Mellon University

PITTSBURGH, PENNSYLVANIA 15213

95228

IS VICTIMIZATION CHRONIC? A BAYESIAN ANALYSIS OF MULTINOMIAL MISSING DATA

J. B. Kadane

Technical Report #307

November 1983

Carnegie-Mellon University
Pittsburgh, PA 15213

Is Victimization Chronic? A Bayesian Analysis of Multinomial Missing Data.

by Joseph B. Kadane*

1. Introduction

This paper analyses, using Bayesian methods, a simple data set concerning successive criminal victimization drawn from the National Crime Survey. As in many surveys, not all the intended interviews could be conducted. Unlike many analyses, however, this paper takes that missing data explicitly into account, to see what difference it makes to the conclusions. This leads to very substantial (and apparently novel) computational difficulties. Hence this paper is partly addressed to victimization, and partly to statistical computation. However, the latter is addressed only to the extent necessary to support the former. Whether the computational methods used here are wise general strategies for such problems is left to further investigation.

2. The Data and a Preliminary Analyses Ignoring the fact that some Data are Missing.

The National Crime Survey is a national household longitudinal survey conducted by the Census Bureau. Households are revisited to see if they have been victims of crimes in the intervening six month period. The data used in this paper comes from interviews six months apart in a rotation design (see Griffin, 1983) and are as shown in Table 1.

A first analysis of this data set is undertaken ignoring the non-response. Thus we consider a "reduced" data set consisting of the upper-left 2x2 subtable of Table 1.

*Research supported in part by National Institutes of Justice Grant 81-IJ-CX-0087. I thank Lionel Galway and Diane Griffin for their help.

	2nd visit		
1st visit	Crime free	Victims	Non-response
Crime free	392	55	33
Victims	76	38	9
Non-response	31	7	115

Table 2-1: Victimization results from the National Crime Survey.

Now we suppose that the data in these four cells are multinomially distributed. The probability is u_1 that a household is crime-free in both periods, u_2 that it is crime free in period 1 and victimized in period 2, u_3 that it is victimized in period 1 not crime-free in period 2, and u_4 that it is victimized in both periods. Naturally we have

$$u_{\bullet} = \sum_{i=1}^{4} u_i = 1 \text{ and } u_i > 0 \text{ for } i = 1, \ldots, 4.$$

The likelihood function for the data set treated this way is proportional to

$$\prod_{i=1}^{4} u_i^{n_i} \tag{2.1}$$

where $n_1 = 392$, $n_2 = 55$, $n_3 = 76$ and $n_4 = 38$.

The Dirichlet distribution with parameter vector $\underline{b} = (b_1, \ldots, b_k)'$ has density

$$f(\underline{u};\underline{b}) = B(\underline{b})^{-1} \prod_{i=1}^{k} u_i^{b_i-1} \tag{2.2}$$

where $B(\underline{b}) = [\pi_1^k \Gamma(b_i)]/\Gamma(b.)$ over the simplex $S = \{\underline{u} : u_i > 0,\ u. = 1\}$. Suppose $\underline{u}$ has a prior distribution that is Dirichlet with parameter $\underline{b}$ (i.e., $u \sim \mathcal{D}(\underline{b})$), and suppose that the data counts $\underline{n}$ are multinomial. Then the posterior distribution $\underline{u} | \underline{n} \sim \mathcal{D}(\underline{b}+\underline{n})$. (This is the property of being closed under sampling, for which see Raiffa and Schlaifer (1961)). We write the general moment of u as $\pi_1^k u_i^{c_i}$. Under the above assumptions, the general prior moment of $\underline{u}$ is

$$\begin{aligned} g(\underline{c};\underline{b}) &= E_{u\,|\,b}\,\Pi_1^k u_i^{c_i} \\ &= B(\underline{b} + \underline{c})/B(\underline{b}) \qquad (2.3) \end{aligned}$$

Different choices of the vector $\underline{c}$ can give us the prior means, variances, covariances, etc. of the u's. Combining (2.3) and (2.2) can give similar facts about the posterior of $\underline{u}$ given $\underline{n}$.

There are various ways to measure association in a contingency table like ours. (See Bishop, Fienberg and Holland (1975), p. 13ff) for a discussion.) Our choice is

$$\phi = \frac{u_1 u_4}{u_2 u_3} \qquad (2.4)$$

called the odds ratio or cross-product ratio which gives how many times more likely it is a household will be victimized the second time if they were victimized the first time than if they were not. Values of ϕ greater than 1.0 indicate that victimization is "catching"; values less than one indicate that it is not. Hence we would like to know about ϕ, say its mean and variance, under various reasonable prior beliefs $\underline{b}$ about $\underline{u}$.

Happily, the expectation of ϕ is in the form (2.3) where $\underline{c}$ takes the special value c_1 = (1, -1, -1, 1). Furthermore, the second moment of ϕ is in the same form, where now $\underline{c}$ takes the value c_2 = (2, -2, -2, 2). Then we have

$$\begin{aligned} E(\phi^i) &= g(\underline{c}_i;\underline{b} + \underline{n}) = B(\underline{b} + \underline{c}_i + \underline{n})/B(\underline{b} + \underline{n}) \\ &= \frac{\prod_1^4 \Gamma(b_i' + c_i)}{\Gamma(\underline{b}' + \underline{c})_\bullet} \Big/ \frac{\prod_1^4 (b_i')}{\Gamma(\underline{b}')_\bullet}, \qquad i=1,2 \qquad (1) \end{aligned}$$

where $\underline{b}' = \underline{b} + \underline{n}$. Notice that

$$(\underline{b}' + \underline{c})_\bullet = \sum_1^4 (b_i' + c_i) = \sum_1^4 b_i' = b_\bullet' \text{ since } \sum_1^4 c_i = 0.$$

Then

$$E(\phi^i) = \prod_1^4 \left\{\frac{\Gamma(b_i'+c_i)}{\Gamma(b_i')}\right\} \qquad (2.6)$$

In particular,

$$\begin{aligned} E(\phi) &= \frac{\Gamma(b_1'+1)}{\Gamma(b_1')} \cdot \frac{\Gamma(b_4'+1)}{\Gamma(b_4')} \cdot \frac{\Gamma(b_2'-1)}{\Gamma(b_2')} \cdot \frac{\Gamma(b_3'-1)}{\Gamma(b_3')} \\ &= \frac{b_1' b_4'}{(b_2'-1)\,(b_3'-1)} \qquad (2.7) \end{aligned}$$

Similarly

$$\begin{aligned} E(\phi^2) &= \frac{\Gamma(b_2'+2)}{\Gamma(b_2')} \cdot \frac{\Gamma(b_4'+2)}{\Gamma(b_4')} \cdot \frac{\Gamma(b_2'-2)}{\Gamma(b_2')} \cdot \frac{\Gamma(b_3'-2)}{\Gamma(b_3')} \\ &= \frac{b_1'(b_1'+1) b_4'(b_4'+1)}{(b_2'-1)\,(b_2'-2)\,(b_3'-1)\,(b_3'-2)} \\ &= (E\phi)\,\frac{(b_1'+1)(b_4'+1)}{(b_2'-2)(b_3'-2)} \qquad (2.8) \end{aligned}$$

Thus computation of the mean and variance of ϕ under these assumptions is relatively simple.

One of the features of Bayesian statistics is the choice of a prior distribution. Given that we have decided to limit ourselves to the conjugate, Dirichlet family (2.2), this reduces to the choice of a vector $\underline{b} = (b_1,b_2,b_3,b_4)$, where $b_i > 0$ is necessary. One of the grails of Bayesian theory is a theory of ignorance. There are several ideas of what that would entail. Haldane (1945) proposes $b_i \rightarrow 0$. The Jeffreys' prior (1961) gives $b_i = ½$. The Savage personalistic position (1962) rejects the idea that "ignorance" is an especially desirable state to represent in a prior distribution, and asks instead that the prior represent the opinion of the analyst before seeing the data.

I find it useful to think of a choice of $\underline{b}$ in two parts; first the choice of $\underline{p} = (p_1 \ldots, p_k)$ where $p_i = b_i/b_{\bullet}$, and second a choice of $b_{\bullet}$. Roughly p_i represents how much of the data a priori I expect to fall in category i, and b. represents how much weight I think my prior should have in the analysis. In this case I suppose I believe that $p_1 = .75$, meaning I expect 75% of households to be crime-free in both periods, $p_2 = p_3 = .10$ to be crime free in one period but not the other, and, a fortiori, $p_4 = .05$ to be victimized in both periods. How sure am I of these assessments, that is, how much information do I think I have? I would choose $b_{\bullet} = 10$ for my assessment of that. Then my prior is represented by b = (7.5, 1, 1, .5). Formal assessment procedures for priors for this case are available in Chaloner and Duncan (1981).

Now the mean and variance of ϕ under each assumption can be computed, and are reported in Table 2.2.

	Haldane prior	Jeffreys prior	Information prior
$\underline{b}$ =	(0,0,0,0)	(1/2,1/2,1/2,1/2)	(7.5,1,1,.5)
$E(\phi)$ =	3.678	3.672	3.680
$SD(\phi)$ =	.920	.913	.912

Table 2-2: Expectation and standard deviation of ϕ under alternative prior distributions ignoring missing data.

Examining Table 2.2, we observe that the choice among prior distributions - at least among these three prior distributions in this circumstance - does not matter much to the mean and standard deviation of ϕ. There certainly are prior distributions for which this conclusion would not hold: as $b_{\bullet} \rightarrow \infty$, $E(\phi) \rightarrow p_1\ p_4/p_2p_3$, comes to dominate the analysis.

With this as background, the reader is invited to choose your own opinion $\underline{b}$, and to compute, using (2.6) and (2.7), your posterior expectation and standard deviation for ϕ. I expect that most of the numbers so computed will look much like those of Table 2.2.

For the three opinions reported in Table 2.3, we can address the question of whether victimization in one period tends to be associated with victimization in the next period. Thus we wish to compute the probability $p_r\{\phi \geq 1\}$.

A useful theorem of Bayesian analysis (Walker, 1969) states conditions under which the posterior distribution is asymptotically normal, as the sample size increases. I judge that here those conditions are satisfied, principally because the data information n. = 561 is so much larger than the prior information b. = 0,2 and 10 respectively for the three prior distributons under discussion here. Here we use the exact moments, calculated in Table 2, instead of the asymptotic moments, which are respectively the maximum likelihood estimate for ϕ and sampling its standard error calculated from the Fisher information. A more precise analysis of the distribution of ϕ could be conducted since the characteristic function of ϕ can be calculated using (2.3). However, the normal approximation is sufficient for these purposes.

	Haldane Prior	Jeffreys Prior	Informative Prior
$\underline{b}$	(0,0,0,0)	(½,½,½,½)	(7.5,1,1,.5)
$\frac{E(\phi)-1}{SD(\phi)}$	2.91	2.94	2.94
$p_r\{\phi<1\}$	.00181	.00164	.00164

Table 2-3: Probability that $\phi < 1$ under alternative prior distributions, using the normal approximation.

Thus I find that under all three alternative prior distributions under study here, the probability that ϕ is less than one is quite small. Under these assumptions, being victimized once does lead to being victimized again. The odds of being victimized again are about 3 2/3 times greater than if one had not been victimized in the previous year.

3. Data Missing at Random

To ignore information is against statistical principles, and in particular, Bayesian principles. Consequently we return to the analysis of Section 2, but no longer ignore the non-response. In this section we assume instead that the data is missing in such a way that the fact of being missing does not impinge on which category may be correct. Under this assumption, the likelihood is proportional to

$$\prod_{i=1}^{4} u_i^{n_i} (u_1+u_2)^{n_{12}} (u_3+u_4)^{n_{34}} (u_1+u_3)^{n_{13}} (u_2+u_4)^{n_{24}} \tag{3.1}$$

where $n_{12} = 33$, $n_{34} = 9$, $n_{13} = 31$ and $n_{24} = 8$. In a rough, qualitative way one can see that because the n_{ij}'s are smaller than the n_i's, and the information in them is less specific anyway, an analysis of (3.1) is likely to be similar to the analyses already given for (2.1).

Let us continue to consider the Dirichlet family of prior distributions (2.2) indexed by $\underline{b}$. When the likelihood is in the form (3.1), the posterior distribution of $\underline{u}$ given the data is no longer in Dirichlet form. Thus another, more general family of distributions must be introduced to discuss the posterior distribution. Such a family is introduced in Dickey, Jiang and Kadane (1983).

With the likelihood function (3.1) and the prior (2.2), the posterior distribution is proportional to

$$\prod_{i=1}^{4} u_i^{n_i+b_i-1} (u_1+u_2)^{n_{12}} (u_3+u_4)^{n_{34}} (u_1+u_3)^{n_{13}} (u_2+u_4)^{n_{24}} \tag{3.2}$$

In order to find the constant of proportionality, (3.2) must be integrated over the simplex. We consider this integral in two parts: $g(\underline{n}\,|\,\underline{b})\ h(\underline{y};\underline{b}+\underline{n})$, where $g(\underline{n}\,|\,\underline{b})$ is given in (2.3) and $h((n_{12},n_{34},n_{13},n_{24});\ \underline{b}+\underline{x})$ is the integral of

$$(u_1+u_2)^{n_{12}} (u_3+u_4)^{n_{34}} (u_1+u_3)^{n_{13}} (u_2+u_4)^{n_{24}} \tag{3.3}$$

under the Dirichlet distribution $(\mathcal{D}(\underline{b}+\underline{n})$. Thus we consider the integral (3.2) as if the complete data had come first, allowing updating to the Dirichlet distribution $\mathcal{D}(\underline{b}+n)$, and then the incomplete data, requiring the integral h. For brevity, we'll write $\underline{y} = (n_{12}, n_{34}, n_{13}, n_{24})$.

Once the integral h is understood, the general moment is

$$E_{u|\underline{b}+\underline{n},y} \prod u_i^{c_i} = g(\underline{c};\underline{b}+\underline{n})h(y;\underline{b}+\underline{n}+\underline{c})/h(\underline{y};\underline{b}+\underline{n}) \qquad (3.4)$$

(see Dickey et al., 1983, (2.13)).

The next result, Theorem 4.1 from Dickey et al. (1983), relates the integral h to the literature on special functions. The integral h can be represented as

$$h(\underline{d};\underline{b}) = \mathcal{R}_{d.}(\underline{b},\tilde{Z},-\underline{d}), \qquad (3.5)$$

where

$$\tilde{Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

is an indicator matrix, and $\mathcal{R}$ is Carlson's function (Carlson, (1977), Dickey (1983)). While this representation allows us to prove various facts about h, it apparently does not offer readily available computational methods.

To compute the expectation of ϕ and ϕ^2 from our data, we have, using (3.4) and

$$E(\phi^i) = g(c_i;\underline{b}+\underline{n})h(y;\underline{b}+\underline{n}+c_i)/h(y;\underline{b}+\underline{n}) \quad i = 1,2 \qquad (3.6)$$

Since we have previously computed $g(c_i;\underline{b}+n)$ (see (2.7) (2.8) and Table 2), the computational problem is to calculate the ratio of h functions. The method for doing this is given in Appendix A.

	Haldane Prior	Jeffreys Prior	Informative Prior
b =	(0,0,0,0)	(1/2,1/2,1/2,1/2)	(7.5,1,1,.5)
$E(\phi)$	3.667	3.661	3.669
$SD(\phi)$	.908	.902	.903

Table 3-1: Expectation and standard deviation of ϕ under alternative prior distribution assuming data missing at random.

The results of this computation are given in Table 3.1.

The import of these calculations are similar to those reported in Table 2.3. Notice that the standard deviations are slightly smaller, due to the gain in information from using the missing data. But the overall message is very similar: The odds of being victimized again are about 3 2/3 greater than if one had not been victimized in the previous year.

4. Informative Non-Response

The assumption of Section 3 that non-response is uninformative is not necessary to Bayesian analysis. In fact, one of the main advantages of the Bayesian viewpoint is its flexibility in incorporating varying beliefs about the data. In this section I explore the consequences of a belief that non-response is associated with victimization.

Let α be the probability that a household will not respond to a survey given that it was victimized in that period, and β be the probability that a household will not respond given that it was crime free. It is reasonable to suppose $\alpha > \beta$.

With this assumption, there are five configurations to analyze:

$$p(NR_2,CF_1) = P(NR_2 | CF_2,CF_1)P(CF_2,CF_1)+P(NR_2 | V_2,CF_1)P(V_2,CF_1)$$
$$= \beta u_1+\alpha u_2 \tag{4.1}$$

$$p(NR_2,V_1) = P(NR_2 | CF_2,V_1)P(CF_2,V_1)+P(NR_2 | V_2,V_1)P(V_1,V_2)$$
$$= \beta\ u_3 + \alpha\ u_2$$
$$P(NR_1,CF_2) = P(NR_1 | CF_2,CF_1)P(CF_2,CF_1)+P(NR_1 | CF_2,V_1)P(CF_2,V_1)$$
$$= \beta\ u_1 + \alpha\ u_3$$
$$P(NR_1,V_2) = P(NR_1 | CF_1,V_2)P(CF_1,V_2)+P(NR_1 | V_1,V_2)P(V_1,V_2)$$
$$= \beta\ u_2 + \alpha\ u_4$$

Finally, supposing the non-response on the two surveys is independent, conditonal on their victimization status, we have

$$P(NR_1,NR_2) = P(NR_1,NR_2 | CF_1,CF_2)P(CF_1,CF_2)+P(NR_1,NR_2 | CF_1,V_2)P(CF_1,V_2)$$
$$+ P(NR_1,NR_2 | V_1,CF_2)P(V_1,CF_2)+P(NR_1,NR_2 | V_1,V_2)P(V_1,V_2)$$
$$= P(NR_1 | CF_1,CF_2)P(NR_2 | CF_1,CF_2)P(CF_1,CF_2)+P(NR_1 | CF_1,V_2)P(NR_2 | CF_1,V_2)P(CF_1,V_2)$$
$$+ P(NR_1 | V_1CF_2)P(NR_2 | V_1,CF_2)P(V_1,CF_2)+P(NR_1 | V_1,V_2)P(NR_2 | V_1,V_2)P(V_1,V_2)$$
$$= \beta^2u_1+\beta\alpha u_2+\alpha\beta u_3+\alpha^2u_4 \tag{4.2}$$

Thus the new likelihood is proportional to

$$\prod_{i=1}^{4} u_i^{n_i}(\beta u_1+\alpha u_2)^{n_{12}} (\beta u_3+\alpha u_4)^{n_{34}} (\beta u_1+\alpha u_3)^{n_{13}}$$
$$(\beta u_2+\alpha u_4)^{n_{24}} (\beta^2u_1+\alpha\beta(u_2+u_3)+\alpha^2u_4)^{n_{1234}} \tag{4.3}$$

When $\alpha = \beta$, which is the case considered in the last section, the last term does not enter, since $u_\bullet = 1$. The likelihood (4.3) can be simplified by dividing and multiplying by $\beta^{n_{12}+n_{34}+n_{13}+n_{24}+n_{1234}}$, and substituting $\gamma = \alpha/\beta$. Then we have a likelihood proportional to

$$\prod_{i=1}^{4} u_i^{n_i}(u_1+\gamma u_2)^{n_{12}} (u_3+\gamma u_4)^{n_{34}} (u_1+\gamma u_3)^{n_{13}} (u_2+\gamma u_4)^{n_{24}} (u_1+\gamma u_2+\gamma u_4)^{n_{1234}} \tag{4.4}$$

It is reasonable, although not required, to suppose $\gamma > 1$. There are several special cases of (4.4) available already. When $\gamma = 1$, the analysis reduces to that of section 3. When $\gamma = 0$, we are saying that we are sure all non-reporting households were crime-free, which is not particularly reasonable.

In this case, this likelihood (4.4) becomes

$$u_1^{n_1+n_{12}+n_{13}+n_{1234}}\ u_2^{n_2+n_{24}}\ u_3^{n_3+n_{34}}\ u_4^{n_4} \tag{4.5}$$

Thus the data reduces to a 2x2 table, then as follows:

		2nd	
		cf	v
1st	cf	571	63
	v	85	38

and the analysis is similar to that given in Section 2. By analogy to Table 2-2, we have

	Haldane	Jeffreys	Informative
$\underline{b}$	(0,0,0,0)	(1/2,1/2,1/2,1/2)	(7.5,1,1,.5)
E(ϕ)	4.17	4.17	4.16
SD(ϕ)	.992	.987	.989

Table 4-1: Expectation and Standard Deviation of ϕ under alternative prior distributions assuming $\gamma = 0$.

A second easy special case to analyse is $\gamma = \infty$. In this case all the non-reporting households are assumed to be victimized, which is only slightly more reasonable than $\gamma = 0$. In this case, after dividing the likelihood in (4.4) by $\gamma^{n_{12}+n_{34}+n_{13}+n_{24}+n_{1234}}$, and we obtain the likelihood proportional to

$$\Pi^{n}_{i}{}_{1} \ (u_1/\gamma+u_2)^{12} \ (u_3/\gamma+u_4)^{34} \ (u_1/\gamma+u_3)^{13} \ (u_2/\gamma+u_4)^{24} \ (u_2/\gamma+u_4) \ (u_1/\gamma^2+u_2/\gamma+u_4)^{n_{12}} \quad (4.6)$$

Now allowing $(1/\gamma) \rightarrow 0$, we have a likelihood proportional to

$$u_1^{n_i} \ u_2^{n_2+n_{12}} \ u_3^{n_3+n_{13}} \ u_4^{n_4+n_{34}+n_{24}+n_{1234}} \ ,$$

which yields a 2x2 table,

		2nd	
		cf	v
1st	cf	392	88
	v	107	170

and estimates as follows:

	Haldane	Jeffreys	Informative
b	(0,0,0,0)	(1/2,1/2,1/2,1/2)	(7.5,1,1,.5)
$E(\phi)$	7.23	7.18	7.23
$SD(\phi)$	1.239	1.243	1.256

Table 4-2: Expectation and Standard Deviation of ϕ under alternative prior distributions assuming $\gamma = \infty$.

Thus we see that victimization appears to be catching even at the extreme assumptions $\gamma = 0$ and $\gamma = \infty$, although assumptions about α seem somewhat important to estimates $E(\phi)$ of how catching it is. However the computational used to obtain the numbers for Table 4 do not appear feasible at present for γ's other than 0,1, and ∞. Nonetheless, we may anticipate that the results would not be qualitatively different from those reported here.

5. Conclusions

To answer the question posed by the title, yes, victimization is chronic. This data set indicates a factor of about 3 2/3 greater odds of one household being victimized again if it was victimized before. A deeper analysis of victimization would look for household characteristics that "explain" victimization, in the sense that, given those characteristics, victimization is not chronic. Perhaps socio-economic status, or some surrogate of it, would be a good first variable to use.

We have also demonstrated a Bayesian approach to missing data that is computationally feasible for some, but not all, of the calculations the study of victimization led us to. This aspect will be more fully developed in subsequent work.

I. Appendix A. Computational Method

We consider integrals of the form (3.1) over the simplex S. Let

$$I(\underline{n},y) = \int_S \prod_{i=1}^{4} u_i^{n_i} (u_1+u_2)^{n_{12}} (u_3+u_4)^{n_{34}} (u_1+u_2)^{n_{13}} \quad (u_2+u_4)^{n_{24}} \, d\underline{u} \tag{A1}$$

Then the expectations sought can be rewritten

$$E(\phi^i) = I(\underline{b}+\underline{n}+c_i,y)/I(\underline{b}+\underline{n},y). \quad (i=1,2) \tag{A2}$$

Hence evaluation of ratios of I's is sufficient, and is equivalent to evaluation of ratios of h's.

The basic approach used here is a binomial expansion of each of the four sums in (A1), as follows:

$$I(\underline{n},\underline{y}) = \int_S \prod_{i=1}^{4} u_i^{n_i} \Sigma \binom{n_{12}}{r} u_1^r u_2^{n_{12}-r} \binom{n_{34}}{j} u_3^j u_4^{n_{34}-j}$$
$$\binom{n_{13}}{\ell} u_1^{\ell} u_3^{n_{13}-\ell} \binom{n_{24}}{m} u_2^m u_4^{n_{24}-m} \, du \tag{A3}$$

where the summation extends over the set $T = \{0 \le r \le n_{12},\ 0 \le j \le n_{34},\ 0 \le \ell \le n_{13}$ and $0 \le m \le n_{24}\}$. Since T has finitely many elements, the integral and summation can be interchanged, and each term in the summation evaluated as a Dirichlet integral:

$$I(\underline{n},\underline{y}) = \sum_T \binom{n_{12}}{r}\binom{n_{34}}{j}\binom{n_{13}}{\ell}\binom{n_{24}}{m} \int_S u_1^{n_1+r+\ell} \ u_2^{n_2+n_{12}-r+m} \ u_3^{n_3+j+n_{13}-\ell} \ u_4^{n_4+n_{34}-jn_{24-m}} \, d\underline{u}$$

$$=\sum_T \binom{n_{12}}{r}\binom{n_{34}}{j}\binom{n_{13}}{\ell}\binom{n_{24}}{m} B(n_1+r+\ell+1, n_2+n_{12}-r+1, n_3+j+n_{13}-\ell+1, n_4+n_{34}-j+n_{24}-m+1) \tag{A4}$$

Continuing (A4), we have

$$I(\underline{n},\underline{y}) = \sum_T \frac{\binom{n_{12}}{r}\binom{n_{34}}{j}\binom{n_{13}}{\ell}\binom{n_{24}}{m}\Gamma(n_1+r+\ell+1)\Gamma(n_{12}+1-r+m)}{\Gamma(n_\bullet+y_\bullet+4)!}$$

$$\Gamma(n_3+j+n_{13}-\ell 1+1)\Gamma(n_4+n_{34}-j+1+n_{24}-m)$$

$$= \frac{n_{12}!n_{34}!n_{13}!n_{24}!}{(n_\bullet+y_\bullet+3)!} \sum_T \frac{\Gamma(n_1+r+\ell+1)\Gamma(n_2+n_{12}-r+m+1)\Gamma(n_3+j+n_{13}-\ell+1)\Gamma(n_4+n_{34}-j+n_{24}-m+1)}{r!(n_{12}-r)!j!(n_{34}-j)!\ell!(n_{13}-\ell)!m!(n_{34}-m)!}$$

$$= K(\underline{n},y) \sum_T I^*(r,j,\ell,m) \tag{A5}$$

where

$$K(\underline{n},y) = \frac{n_{12}!n_{34}!n_{13}!n_{24}!}{\Gamma(n. + y. + 4)}$$

and

$$I^*(r,j,\ell,m) = \frac{\Gamma(n_1+r+\ell+1)\Gamma(n_2+n_{12}-r+m+1)\Gamma(n_3+j+n_{13}-\ell+1)\Gamma(n_4+n_{34}-j+n_{24}-m+1)}{r!(n_{12}-r)!j!(n_{34}-j)!\ell!(n_{13}-\ell)!m!(n_{24}-m)!}$$

and

$$I^*(r,j,\ell,m) = \frac{\Gamma(n_1+r+\ell+1)\Gamma(n_2+n_{12}-r+m+1)\Gamma(n_3+j+n_{13}-\ell+1)\Gamma(n_4+n_{34}-j+n_{24}-m+1)}{r!(n_{12}-r)!j!(n_{34}-j)!\ell!(n_{13}-\ell)!m!(n_{24}-m)!}$$

Direct evaluation of I is not attractive because it would involve the summation of $(n_{12}+1)(n_{34}+1)(n_{13}+1)(n_{24}+1)$ terms, in our case $34 \cdot 10 \cdot 32 \cdot 9$ = 97,920 terms, each of which involves 8 factorials and 4 gamma functions. Consequently a different strategy must be used. However, one aspect of (A5) less unpleasant is that

$$K\,(\underline{b}+\underline{n} + c_i, \underline{y}) = \frac{n_{12}!\,n_{34}!\,n_{13}!\,n_{24}!}{\Gamma(\Sigma n_i + \Sigma b_i + c_{i\bullet} + n_{12} + n_{34} + n_{24} + n_{13} + 4)} \tag{A6}$$

But $c_{i\bullet} = 0$, so

$$K(\underline{b}+\underline{n}+c_i, \underline{y}) = K(\underline{b}+\underline{n}; \underline{y}) \text{ for all } \underline{b}, \underline{n} \text{ and } \underline{y},\ i=1,2 \tag{A7}$$

Hence for computing expectations as in (A2), the constants K in (A5) cancel.

Each of the terms I^* is non-negative. Furthermore, they are likely to decrease exponentially fast from the maximum, so that only a few contribute nearly everything to I, and the vast bulk of them are negligible. Consequently I wish to find the largest term in T, and to study how it relates to other relatively large terms in T. While succssive differences of I^*'s are a mess, successive ratios are not. Thus we have

$$\frac{I^*(r+1,j,\ell,m)}{I^*(r,j,\ell,m)} = \frac{n_1+r+\ell+1}{r+1} \cdot \frac{n_{12}-r}{n_2+n_{12}-r+m} \tag{A8a}$$

$$\frac{I^*(r-1,j,\ell,m)}{I^*(r,j,\ell,m)} = \frac{r}{n_1+r+\ell} \cdot \frac{n_2+n_{12}-r+m+1}{n_{12}-r+1} \tag{A8b}$$

$$\frac{I^*(r,j+1,\ell,m)}{I^*(r,j,\ell,m)} = \frac{n_3+j+n_{13}-\ell+1}{j+1} \cdot \frac{n_{34}-j}{n_4+n_{34}-j+n_{24}-m} \tag{A8c}$$

$$\frac{I^*(r-1,j,\ell,m)}{I^*(r,j,\ell,m)} = \frac{j}{n_3+j+n_{13}-\ell} \cdot \frac{n_4+n_{34}-j+n_{24}-m+1}{n_{34}-j+1} \tag{A8d}$$

$$\frac{I^*(r,j,\ell+1,m)}{I^*(r,j,\ell,m)} = \frac{n_1+r+\ell+1}{\ell+1} \cdot \frac{n_{13}-\ell}{n_3+j+n_{13}-\ell} \tag{A8e}$$

$$\frac{I^*(r,j,\ell-1,m)}{I^*(r,j,\ell,m)} = \frac{\ell}{n_1+r+\ell} \cdot \frac{n_3+j+n_{13}-\ell+1}{n_{13}-\ell+1} \tag{A8f}$$

$$\frac{I^*(r,j,\ell,m+1)}{I^*(r,j,\ell,m)} = \frac{n_2+n_{12}-r+m+1}{m+1} \cdot \frac{n_{24}-m}{n_4+n_{34}-j+n_{24}-m} \tag{A8g}$$

$$\frac{I^*(r,j,\ell,m-1)}{I^*(rmj,\ell,m)} = \frac{m}{n_2+n_{12}-r+m} \cdot \frac{n_4+n_{34}-j+n_{24}-m+1}{n_{24}-m+1} \tag{A8h}$$

Consider the set S_0 of all points in T for which each of the eight expressions A8 is less than or equal to 1.0. S_0 is not empty, since the maximum of $I^*(r,j,\ell,m)$ over the finite set T is an element of S_0. We call S_0 the set of coordinate-wise local maxima.

The property of being the unique coordinate-wise local maximum is stronger than being the unique global maximum, as the following example shows:

<u>Example 1</u> Suppose on a 2x2 discrete grid, f takes the following values:

$f(1,1) = 3$, $f(0,0) = 2$, and $f(0,1) = f(1,0) = 1$.

Then (1,1) is the global maximum, but both (1,1) and (0,0) are coordinate-wise local maxima.

If f has a unique coordinate-wise local maximum, maximization coordinate-by-coordinate is sufficient to find the global maximum. This is so because such maximization can terminate only at a coordinate-wise local maximum, which, if it is unique, is the global maximum. The following theorem shows an important property of maximization that is particularly convenient if S_0 is small.

Theorem 1. Let S be a set of points containing S_0. Then the maximum of f(s) over the set $s \in S^c$, occurs at the boundary of S, that is at points of the form $s \pm e_k$ where $s \in S$ and e_k has a 1 in the kth coordinate position and is zero otherwise.

Proof: Suppose to the contrary that s^* maximizes f over the set S^c, where s^* is not on the boundary of S. Then s^* is a coordinate-wise local maximum, which contradicts the hypothesis. ■

Theorem 1 suggests an algorithm for computing I. We start with its set S_0 of all coordinate-wise local maxima. We show in Appendix B that in our case, this is the single point (29,6,26,5). We divide all terms in I by $I^*(29,6,26,5)$, that is

$$
\begin{aligned}
I &= \Sigma_T\, I^*(r,j,k,\ell) \\
&= I^*(29,6,26,5)\ \Sigma_T \left\langle \frac{I^*(r,j,k,\ell)}{I^*(29,6,26,5)} \right\rangle \\
&= I^*(29,6,26,5)\ \Sigma_T\, J(r,j,k,\ell)
\end{aligned}
$$

where $J(r,j,k,\ell) = I^*(r,j,k,\ell)/I^*(29,6,26,5)$. Now since (29,6,26,5) maximizes I^*, we have $0 < J \le 1$ for all coordinates, and J = 1 only at (29,6,26,5). Then we can at each stage add to a running sum S the largest term J not already in the sum, but in the set S, and add to the set S

the set of coordinate-wise neighbors of the point added to the sum. Theorem 1 assures us that the maximum term not in the sum at each stage will be on the coordinate-wise boundary of the set of terms included, and hence in S. Furthermore, since the value of the largest term excluded is always available, a bound of the accuracy of stopping at any given stage is available: the number of terms excluded from the current sum, times the maximum value excluded. If this criterion is not satisfied, the maximum term is added to the sum, new points not in S are joined to S, and the bound recomputed. When the criterion is first met, the algorithm stops and reports the sum obtained.

Note that when a term is added to the sum, the ratios in A8 give the value of J at the new boundary points. The only limitation on this method is the numerical accuracy of multiplying and dividing many numbers together. By keeping track of the path length of each new element of S from (29, 6, 26, 5), control can be kept on how many multiplication and divisions are involved, and hence whether the numerical accuracy of the result is satisfactory.

Note that I^* is a function both of $\underline{n}$ and of (r,j,k,ℓ). Consequently

$$
E(\phi^i) = \frac{I^*(\underline{n}+\underline{b}+c_i,29,6,26,5)}{I^*(\underline{n}+\underline{b},29,6,26,5} \quad \frac{\Sigma_T J(\underline{n}+\underline{b}+c_i;r,j,k.\ell)}{\Sigma_T J(\underline{n}+\underline{b};r,j,u,\ell}
$$

The sum of terms J over the set T are computed as above. However the ratio of I^* terms can be calculated directly:

$$
\frac{I^*(\underline{n}+\underline{b}+c_i,29,6,26,5)}{I^*(\underline{n}+\underline{b},29,6,26,5)} = \frac{\Gamma(n_1+b_1+c_{i1}+56)}{\Gamma(n_1+b_1+56)} \cdot \frac{\Gamma(n_2+b_2+c_{i2}+8)}{\Gamma(n_2+b_2+8)}
$$

$$\frac{\Gamma(n_3+b_3+c_{i3}+12)}{\Gamma(n_3+b_3+12)} \; \frac{\Gamma(n_4+b_4+c_{i4}+8)}{\Gamma(n_4+b_4+8)}$$

Hence

$$\frac{I^*(\underline{n}+\underline{b}+c_i;29,6,26,5)}{I^*(\underline{n}+\underline{b};29,6,26,5)} = \frac{\Gamma(448+b_1+1)}{\Gamma(448+b_1)} \cdot \frac{\Gamma(63+b_2-1)}{\Gamma(63+b_2)} \cdot \frac{\Gamma(88+b_3-1)\;.\;\Gamma(46+b_4+1)}{\Gamma(88+b_3)\;.\;\Gamma(46+b_4)}$$

$$= \frac{(448+b_1)\;(46+b_4)}{(63+b_2)\;(88+b_3)}$$

Similarly,

$$\frac{I^*(\underline{n}+\underline{b}+c_2;29,6,26,5)}{I^*(\underline{n}+\underline{b}+c_1;29,6,26,5)} = \frac{(449+b_1)\;(47+b_4)}{(62+b_2)\;(87+b_3)}.$$

II. Appendix B. Determination of Coordinate-Wise Local Maxima

We wish to study which points (r,j,ℓ,m) in T satisfy the inequalities generated by setting the eight quantities (A8) less than or equal to 1.

Each of the four pairs of inequalities has the same form:

$$\frac{a(s)+1+s}{s+1}\;\frac{k(s)-s}{b(s)+k(s)-s} \le 1 \qquad \text{(B1)}$$

$$\frac{s}{a(s)+s}\;\frac{b(s)+k(s)-s+1}{k(s)-s+1} \le 1 \qquad \text{(B2)}$$

where a(s), b(s) and k(s) are functions given below:

s	a(s)	b(s)	k(s)
r	$n_1+\ell$	n_2+m	n_{12}
j	$n_3+(n_{13}-\ell)$	$n_4+n_{24}-m$	n_{34}
ℓ	n_1+r	n_3+j	n_{13}
m	$n_2+(n_{12}-r)$	$n_4+n_{34}-j$	n_{24}

Working with inequalities (B1) and (B2), and suppressing the argument (s) for simplicity, we have

$$(a+s+1)\;(k-s) \le (s+1)\;(b+k-s) \qquad \text{(B3)}$$

and

$$(a+s)(k-s+1) \ge s\;(b+k-s+1) \qquad \text{(B4)}$$

Thus $ak-s+ks-s^2+k-s \le bs+ks-s^2+b+k-s$ (B5)

and $ak-as+a+ks-s^2+s \ge bs+ks-s^2+s$ (B6)

Then, simplifying, we have

$$ak-b \le s(a + b) \tag{B7}$$

and

$$ak + a \ge s(a + b) \tag{B8}$$

Now (B7) and (B8) can be joined together:

$$ak - b \le s(a + b) \le ak + a \tag{B9}$$

$$\text{or } \frac{ak-b}{a+b} \le s \le \frac{ak+a}{a+b} \tag{B10}$$

Let $\alpha(s)$ = ak-b and $\beta(s)$ = a + b.

Then

$$\alpha/\beta + 1 = \frac{\alpha+\beta}{\beta} = \frac{ak-b+a+b}{a+b} = \frac{ak+a}{a+b}.$$

Hence (B10) can be rewritten

$$\alpha/\beta \le s \le \alpha/\beta + 1 \tag{B11}$$

Hence (B11) leads us to be interested in the quantity

$$\alpha/\beta = \frac{ak-b}{a+b} = \frac{k(a/b)-1}{(a/b)+1} = \frac{kz-1}{z+1}, \tag{B12}$$

where $z(s) = \dfrac{a(s)}{b(s)}$.

Now we note that the quantities a(s) and b(s) depend on values of the other indices in the

set (r,j,ℓ,m) besides s. (See the table between (B2) and (B3)), but k(s) does not. Then (B11) and B(12) determine values for s only if z(s) is well-enough appproximated. However, bounds for z(s) can be generated, and are sharpened by bounds for $s' \neq s$. Thus recursively tighter bounds may be obtained for each s. I now leave the general case, and revert to the particulars of my data and feasible the priors specified above. I believe that the computational strategy used here can be generalized.

There are in fact nine values of I that we wish to calculate, each of which is characterized by one of the three priors listed in Table 2, and a value of c, c_0 = (0,0,0,0), c_1 = (1, -1, -1, 1), and c_2 = (2, -2, -2, 2).

Prior b	c	n_1	n_2	n_3	n_4
(0,0,0,0)	(0,0,0,0)	392	55	76	38
(.5,.5,.5,.5)		392.5	55.5	76.5	38.5
(7.5,1,1,.5)		399.5	56	77	38.5
(0,0,0,0)	(1,-1,-1,1)	393	54	75	39
(.5,.5,.5,.5)		393.5	54.5	75.5	39.5
(7.5,1,1,.5)		400.5	55	76	39.5
(0,0,0,0)	(2,-2,-2,2)	394	53	74	40
(.5,.5,.5,.5)		394.5	53.5	74.5	40.5
(7.5,1,1,.5)		401.5	54	75	40.5

Hence we have the following inequalities governing all nine calculations:

$$392 \le n_1 \le 401.5$$
$$53 \le n_2 \le 56$$
$$74 \le n_3 \le 77$$
$$38 \le n_4 \le 40.5$$

Now we calculate a (round 1) upperbound on z(s) and α/β:

$\underline{s}$	Upper bound $\underline{\text{for } a}$	Lower bound $\underline{\text{for } b}$	Upper bound $\underline{\text{for } z}$	Upper bound $\underline{\text{for } \alpha/\beta}$
r	431.5	53	8.16	29.29
j	108	38	2.84	6.40
ℓ	434.5	74	5.87	26.34
m	89	38	2.34	5.31

Similarly we calculate a (round 1) lower bound on z(s) and α/β

$\underline{s}$	Lower bound $\underline{\text{for } a}$	Upper bound $\underline{\text{for } b}$	Lower bound $\underline{\text{for } z}$	Lower bound $\underline{\text{for } \alpha/\beta}$
r	392	64	6.125	28.22
j	74	48.5	1.526	5.04
ℓ	392	86	4.558	25.24
m	53	49.5	1.070	3.65

From this calculation, we conclude that S_0 contains only points satisfying $29 \leq r \leq 30$, $6 \leq j \leq 7$, $26 \leq \ell \leq 27$ and $4 \leq m \leq 6$. Using those bounds, we obtain tighter round 2 estimates as follows:

$\underline{s}$	Upper bound $\underline{\text{for } a}$	Upper bound $\underline{\text{for } b}$	Upper bound $\underline{\text{for } z}$	Upper bound $\underline{\text{for } \alpha/\beta}$
r	428.5	57	7.52	29.01
j	82	40	2.05	5.72
ℓ	431.5	80	5.39	25.995
m	60	40	1.50	4.4

Similarly a (round 2) lower bound is calculated

$\underline{s}$	Lower bound $\underline{\text{for } a}$	Upper bound $\underline{\text{for } b}$	Lower bound $\underline{\text{for } z}$	Lower bound $\underline{\text{for } \alpha/\beta}$
r	418	62	6.74	28.61
j	78	44.5	1.75	5.36
ℓ	421	84	5.01	25.68
m	56	43.5	1.29	4.07

Hence we conclude: $29 \leq r \leq 30$, $j = 6$, $\ell = 26$, $m = 5$. Hence we do a round 3 upper bound for r only: a has an upper bound of 427.5, b has a lower bound of 58; hence the upper bound for z is 7.37 and the upper bound for α/β is 28.94. Consequently, we have $r = 29$. Thus for all nine calculations, S_0 consists only of the term (29, 6, 26, 5).

REFERENCES

1. Bishop, Y, S. Fienberg and P. Holland (1975) Discrete Multiariate Analysis: Theory and Practice, MIT Press, Cambridge.

2. Carlson, B. C. (1977) Special Functions of Applied Mathematics, Academic Press, New York.

3. Dickey, J. M. (1983) "Probabilistic interpretations and statistical uses of multiple hypergeometric functions", J.A.S.A., to appear.

4. Dickey, J. M., J. B. Kadane and J. M. Jiang (1983) "Multinomial Sampling with Missing Data and Multiple Hypergeometric Functions," Research Report #15, Dept. of Mathematics and Statistics, SUNY at Albany.

5. Griffin, D. (1983) "Estimation of Victimization Prevalence Using Data from the National Crime Survey," Ph.D. dissertation, Department of Statistics, Carnegie-Mellon University.

6. Haldane, J. B. S. (1945) "On a Method of Estimating Frequencies," Biometrika, 33, 222-225.

7. Jeffreys, H. (1961) Theory of Probability, Clarendon Press, Oxford, 3rd edition.

8. Raiffa, H and R. S. Schlaiffer (1961) Applied Statistical Decision Theory, Graduate School of Business Administration, Harvard University, Boston.

9. Savage, L. J. (1962) The Foundations of Statistical Inference: A Discussion, Metheuen, London.

10. Walker, A. M. (1969) "On the Asymptotic Behavior of Posterior Distribution," J. R. S. S. (ser. B), 31, 80-88.

Table of Contents

List of Tables

END