Information lost by conditioning on the random margins in a two-by-two table

By M. J. SCHERVISH

Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.

mark@cmu.edu

J. B. KADANE

Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.

kadane@stat.cmu.edu

T. SEIDENFELD

Department of Philosophy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.

ts1s@andrew.cmu.edu

and R. B. STERN

Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

rbstern@gmail.com

Summary

We study the amount of the Fisher information about the odds ratio parameter that one loses by conditioning on the random margin totals of a 2×2 table for the cases in which the data arise either as a multinomial sample or as two independent binomial samples. When there is a nusiance parameter, as in this problem, many authors have proposed the concept of partial information to quantify the amount of Fisher information about the parameter of interest. We show that, as the sample size of the table becomes infinite (in such a way that all margin totals also become infinite) the fraction of the partial information that one loses by conditioning on both margins goes to 0.

Some key words: Fisher information; partial information; fraction of information lost.

1. Introduction

Fisher (1935) introduced his exact test of equality of two success probabilities from two independent binomial samples when the data are displayed in a 2×2 table. That test is based on the conditional distribution of one cell count from the table given the counts in the row and column margins. The goal of this paper is to examine in detail the following carefully-worded statement from (Fisher, 1935, p. 48) which has generated decades of discussion in the statistical literature:

> If it be admitted that these marginal frequencies by themselves supply no information on the point at issue namely, as to the proportionality of the frequencies in the body of the table, we may recognize the information they supply as wholly ancillary; and therefore recognize that we are concerned only with the relative probabilities of occurrence of the different ways in which the table can be filled in, subject to these marginal frequencies.

When Fisher mentioned "information" in the above quote, he was referring to Fisher information, which he had introduced earlier in the article. Fisher did not assert that the marginal frequencies by themselves supply no information about proportionality, but rather he stated merely that one could justify the use of the exact test if the marginal frequencies provide no information about proportionality. Plackett (1977) examined the log-likelihood function based on samples from 2×2 tables and gave some numerical results concerning how much that log-likelihood varies with the (log) odds ratio. Yates (1984) presented a comprehensive history of significance testing based on 2×2 tables. In the appendix to that article he wrote "That the margins of a 2×2 table by themselves do not, except in extreme cases and in repeated sampling, contain any information on proportionality, is certainly true." Aitkin & Hinde (1984), in the discussion of Yates (1984), offered the following response to the previous quote: "the margins of a 2×2 table do in general contain information about the odds ratio parameter θ though in large samples this information may be negligible." Little (1989), while discussing inference about the odds ratio θ for the case in which each row margin total is fixed by design, offered one main argument and three supplemental ones in support of his view that the second margin "usually contains very little in formation about the odds ratio." His main argument is that the mean of the margin "tells us very little about the odds ratio." The supplementary arguments all involve changing the model in some way.

The history of inference about 2×2 tables given by Yates (1984) is augmented by a lively discussion which includes the following prediction by Barnard (1984) which this paper fulfills: "As Dr Yates points out, arguments about 2×2 tables have now gone on for 70 years, so perhaps it would be too much to hope to forestall a centenary." The intervening years have seen several papers, such as Amari (1987); Bhapkar (1989, 1991); Zhu & Reid (1994), that offer the concept of partial information about a parameter of interest in the presence of nuisance parameters (see Definition 1 in Section 2). Partial information can facilitate the calculation of how much information one loses by conditioning on part of the available data. The distribution of the cell counts in a 2×2 table depends not only on proportionality, i.e. the odds ratio, but also on at least one nuisance parameter. We compute the partial information about the odds ratio in the presence of the nuisance parameter(s) for two common sampling models of 2×2 tables both before and after conditioning on the margin totals. The two models we study are (a) independent binomial sampling, in which there is one nuisance parameter, one pair of fixed margin totals, and one pair of random margin totals, and (b) multinomial sampling, in which there are two nuisance parameters and two pairs of random margin totals. Chernoff (2004) examined similar issues about Kullback-Liebler information through numerical examples.

Our main result (Theorem 1 in Section 4) is that, under both sampling plans, the fraction of the partial information about the odds ratio in the presence of the nuisance parameter(s) that is lost by conditioning on the random margin(s) goes to zero as the number of observations in the table goes to infinity. In Section 5, we shed some light on Fisher's comments about information as to proportionality in the margins as well as Yates' comment about "extreme cases."

2. Information Lost by Conditioning

Let X be a random vector whose distribution is determined by a parameter vector λ and which will be used to make inference about one or more functions of λ. In our applications, X will be a one-to-one function of the vector of cell counts in a 2×2 table. Suppose that X is partitioned as (X_1, X_2). In our applications, X_1 will be either one or two of the margin totals of the table, and X_2 will be statistics such that the full table can be recovered from X_1 and X_2. Denote the Fisher information matrix based on the distribution of X by $I(\lambda)$. Also, let $I_1(\lambda)$ and $I_{2|1}(\lambda|X_1)$ denote the Fisher information matrices based respectively on the distribution of X_1 and on the conditional distribution of X_2 given X_1. Then

$$I(\lambda) = I_1(\lambda) + E_1[I_{2|1}(\lambda|X_1)], \tag{1}$$

where E_1 denotes expectation with respect to the distribution of X_1. An alternative way to compute $E_1[I_{2|1}(\lambda|X_1)]$ is to start with the score function based on the conditional distribution of X_2 given X_1, consider it as a function of both X_1 and X_2, and compute its covariance matrix with respect to the joint distribution of $X = (X_1, X_2)$, which we denote $I_{2|1}(\lambda)$. Then (1) becomes

$$I(\lambda) = I_1(\lambda) + I_{2|1}(\lambda), \tag{2}$$

where $I_{2|1}(\lambda)$ can be considered as an information matrix in the same way as $I(\lambda)$ and $I_1(\lambda)$. It is the average (with respect to the distribution of X_1) amount of information in the conditional distribution of X_2 given X_1.

Since all information matrices, including $I_{2|1}(\lambda)$, are positive semi-definite, overall measures of the information about λ (such as the trace or determinant of the matrix) are at least as large when applied to the left-hand side of (2) as when applied to either of the two summands on the right-hand side. This fact also verifies the intuition that neither conditioning on part of the data nor ignoring part of the data lead to more information. Based on (2) one might refer to

$$I(\lambda) - I_{2|1}(\lambda) = I_1(\lambda) \tag{3}$$

as the amount of information about λ that is lost by conditioning on X_1.

There are two reasons that the raw difference computed from (3) might be inadequate as a measure of how much information is lost by conditioning on part of the data. First, when λ contains both parameters of interest and nuisance parameters, it would be good to be able to measure the amount of information lost about the parameters of interest without confounding it with information lost about the nuisance parameters. Second, it would be good to have a measure of information loss that did not depend on which of many equivalent version of the parameters of interest is in use.

Begun et al. (1983); Amari (1987); Bhapkar (1989, 1991); Zhu & Reid (1994) study measures of information and information loss that focus on the parameters of interest.

Definition 1. *Suppose that the parameter vector is partitioned as $\lambda = (\lambda_1, \lambda_2)$, with a vector of parameters of interest λ_1 and a nuisance parameter vector λ_2. Suppose also that the Fisher information matrix is partitioned as*

$$I(\lambda) = \begin{pmatrix} J_{\lambda_1,\lambda_1} & J_{\lambda_1,\lambda_2} \\ J_{\lambda_2,\lambda_1} & J_{\lambda_2,\lambda_2} \end{pmatrix}, \tag{4}$$

with J_{λ_2,λ_2} being nonsingular. The partial information or P-information about λ_1 in the presence of λ_2 is

$$I(\lambda_1; \lambda_2) = J_{\lambda_1,\lambda_1} - J_{\lambda_1,\lambda_2} J^{-1}_{\lambda_2,\lambda_2} J_{\lambda_2,\lambda_1}. \tag{5}$$

Table 1. *Generic* 2×2 *table with row and column totals*

			Row Totals
	$N_{1,1,n}$	$N_{1,2,n}$	$N_{1,\cdot,n}$
	$N_{2,1,n}$	$N_{2,2,n}$	$N_{2,\cdot,n}$
Column Totals	$N_{\cdot,1,n}$	$N_{\cdot,2,n}$	n

The concept of partial information in Definition 1 applies to all information matrices, including $I_1(\lambda)$ and $I_{2|1}(\lambda)$. In Section 1 of the online supplement, we explain how to compute P-information when J_{λ_2,λ_2} is singular. The following is adapted from Zhu & Reid (1994)[Definition 2] into our notation.

Definition 2. *In the situation of equation (2), the P-information loss about λ_1 in the presence of λ_2 from conditioning on X_1 is*

$$L_{2|1}(\lambda_1;\lambda_2) = I(\lambda_1;\lambda_2) - I_{2|1}(\lambda).$$

Proposition 1 below is a special case of Zhu & Reid (1994)[Corollary 1].

Proposition 1. *Suppose that λ_2 does not appear in the conditional distribution of X_2 given X_1. Then $L_{2|1}(\lambda_1;\lambda_2) = I_1(\lambda_1;\lambda_2)$.*

Most of the cases we consider satisfy the condition in Proposition 1. The only case that does not satisfy the condition appears in Section 7 of the online supplement.

Zhu & Reid (1994)[pages 113–114] also show that partial information about λ_1 in the presence of λ_2 is invariant under reparameterizations that preserve the parameter of interest λ_1. For this reason, when λ_2 is understood from context, we often drop the clause "in the presence of λ_2" when referring to P-information (and P-information loss) about λ_1. Even though the P-information and loss thereof do not depend on reparameterizations that preserve the parameter of interest, they do depend on reparameterizations that transform only the parameter of interest. In Section 2 of the online supplement, we prove that if A and B are information matrices about the same parameters calculated under different conditions, (e.g., either conditional on different parts of the data or based on entirely different data sets or P-informations in the presence of the same nuisance parameters), the eigenvalues of $A^{-1}B$ are invariant under common differentiable invertible reparameterizaions of the parameter of interest, if A is nonsingular. In particular the determinant, trace, and largest eigenvalue are invariant under invertible parameterizations. For example, let $I(\lambda_1;\lambda_2)$ be as above, and let $L_{2|1}(\lambda_1;\lambda_2)$ be the P-information loss from conditioning on X_1. If $I(\lambda_1;\lambda_2)$ is nonsingular and if $\psi = g(\lambda_1)$ is an equivalent parameterization with corresponding P-information and loss of P-information matrices $I^\dagger(\psi;\lambda_2|X_1)$ and $L^\dagger(\psi;\lambda_2)$, then $I^\dagger(g(\lambda_1);\lambda_2)^{-1}L^\dagger_{2|1}(g(\lambda_1);\lambda_2)$ has the same eigenvalues as $I(\lambda_1;\lambda_2)^{-1}L_{2|1}(\lambda_1;\lambda_2)$ for each λ_1 and λ_2. For the case of a one-dimensional parameter of interest λ_1, the fraction of P-information that is lost by conditioning on X_1 is just the ratio $L_{2|1}(\lambda_1;\lambda_2)/I(\lambda_1;\lambda_2)$.

3. The Case of 2×2 Tables

3.1. *Information and Information Lost*

Consider a generic 2×2 table with row and column totals displayed as in Table 1. Assume that $n = N_{1,1,n} + N_{1,2,n} + N_{2,1,n} + N_{2,2,n}$ is known in advance of observing the cell counts. There are two common sampling plans that lead to such tables. One is *multinomial sampling* and the other is *independent binomial sampling*. In multinomial sampling, only n (the overall table total) is known in advance (usually by design). In independent binomial sampling the row totals $N_{1,\cdot,n}$ and $N_{2,\cdot,n}$ (or equivalently the column totals, but not both) are known in advance. There is a third sampling plan in which all four marginal totals are known in advance. We are not concerned with that case because one would be compelled to treat all margin totals as constant, so the question of whether or not to condition on them is moot. Barnard (1947) discusses all three sampling plans in the context of significance testing, while Pearson (1947); Pearson & Merrington (1948) discuss power functions of tests under the first two sampling plans. Other possible sampling plans, in which the total sample size n is not known in advance, are discussed in Section 3 of the online supplement.

Throughout the remainder of this paper we will add superscripts of b or m to symbols in order to distinguish calculations specific to the independent binomial and multinomial cases respectively. We also assume that all four cell probabilities in the 2×2 table are strictly positive. For independent binomial sampling we assume that the two sample sizes are strictly positive.

3.2. *Independent binomial sampling*

We start with the case in which one pair of margin totals is known in advance. Then an open question is whether or not to condition on the other margin totals. For definiteness, assume that the jth row margin total is fixed at $k_{j,\cdot,n}$ for $j = 1, 2$. (The somewhat awkward notation will become more useful when we turn to multinomial sampling and asymptotics.) Let $N_{j,1,n}$ in Table 1 have the binomial distribution with parameters $(k_{j,\cdot,n}, q_{j,1})$ for $j = 1, 2$. Let $q_{j,2} = 1 - q_{j,1}$ for $j = 1, 2$. Our interest is (as was Fisher's interest) in the odds ratio parameter

$$\theta = \frac{q_{1,1}(1 - q_{2,1})}{(1 - q_{1,1})q_{2,1}}.$$

The parameterization that we use is $\lambda^b = (\theta, \eta)$, where $\eta = q_{2,1}/(1 - q_{2,1})$ is the nusiance parameter. In the notation of (1) let $X_1^b = N_{\cdot,1,n}$, $X_2^b = N_{1,1,n}$, $\lambda_1^b = \theta$, $\lambda_2^b = \eta$, and E_1^b be expectation with respect to the distribution of X_1^b.

We compute the Fisher information matrix and the P-information about θ in the presence of η in Section 4.1 of the online supplement. We report the P-information here:

$$I^b(\theta; \eta) = \frac{k_{1,\cdot,n}\eta}{\theta(1 + \theta\eta)^2} - \left(\frac{k_{1,\cdot,n}}{(1 + \theta\eta)^2}\right)^2 \left(\frac{k_{1,\cdot,n}\theta}{\eta(1 + \theta\eta)^2} + \frac{n - k_{1,\cdot,n}}{\eta(1 + \eta)^2}\right)^{-1}. \tag{6}$$

Fisher recommended basing inference about θ on the conditional distribution of $N_{1,1,n}$ given $N_{\cdot,1,n}$. So, the following question arises: What is the P-information loss about θ in the presence of η from conditioning on $N_{\cdot,1,n}$? The distribution of X_1^b is that of the sum of two independent binomial random variables with respective parameters $(k_{1,\cdot,n}, \theta\eta/(1 + \theta\eta))$ and $(n - k_{1,\cdot,n}, \eta/(1 + \eta))$. The conditional distribution of X_2^b given X_1^b is a member of the family of noncentral hypergeometric distributions.

Definition 3. *A discrete random variable X has the noncentral hypergeometric distribution with parameters (θ, d, m_1, m_2) if its probability function is*

$$P_{\theta,d,m_1,m_2}(x) = \frac{\binom{m_1}{x}\binom{d - m_1}{m_2 - x}\theta^x}{\sum_{y=\max\{0,m_1+m_2-d\}}^{\min\{m_1,m_2\}} \binom{m_1}{y}\binom{d - m_1}{m_2 - y}\theta^y}, \tag{7}$$

for $x = \max\{0, m_1 + m_2 - d\}, \ldots, \min\{m_1, m_2\}$.

In the case we are considering, the conditional distribution of X_2^b given $X_1^b = k_{\cdot,1,n}$ is the noncentral hypergeometric distribution with parameters $(\theta, n, k_{1,\cdot,n}, k_{\cdot,1,n})$. The nuisance parameter η does not appear in this distribution. The P-information loss about θ from conditioning on $N_{\cdot,1,n}$ is

$$L_{2|1}^b(\theta;\eta) = I^b(\theta;\eta) - I_{2|1}^b(\theta;\eta) = I_1^b(\theta;\eta), \tag{8}$$

Where the second equality follows from Proposition 1. The ratio $L_{2|1}^b(\theta;\eta)/I^b(\theta;\eta)$ is the fraction of P-information that is lost by conditioning on the random margin.

Let Z_n be the conditional variance of $N_{1,1,n}$ given $N_{\cdot,1,n}$. Kou & Ying (1996)[Remark 3.2] stated (without proof) that the Fisher information in the conditional distribution of $N_{1,1,n}$ given $N_{\cdot,1,n}$ is Z_n/θ^2. Hence,

$$I_{2|1}^b(\theta;\eta) = E_1^b\left(\frac{Z_n}{\theta^2}\right). \tag{9}$$

Proofs of the results stated above, along with a formula for computing (9) appear in Section 4.1 of the online supplement. Although the formula for (9) is difficult to work with analytically, we give some asymptotic results in Section 4.

3.3. *Multinomial sampling*

Multinomial sampling is complicated by the presence of an additional nuisance parameter that governs the distribution of the row totals. A general multinomial sample of cell counts $N_{i,j,n}$ for $i, j = 1, 2$ has the multinomial distribution with parameters $(n, p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2})$, where $p_{i,j}$ is the probability that a single observation falls into row i and column j. The first row total $N_{1,\cdot,n}$ has the binomial distribution with parameters $(n, p_{1,\cdot})$, where $p_{1,\cdot} = p_{1,1} + p_{1,2}$. The second row total is n minus the first row total. The question arises as to whether or not to condition on both row and column totals. (In Section 7 of the online supplement we show that there is no P-information loss about θ from conditioning on only one margin total.)

We can parameterize the table in a manner similar to that in Section 3.2 by letting $\lambda^m = (\theta, \eta, p_{1,\cdot})$ with θ and η being identical to the parameters in Section 3.2. In the notation of (1) let $X_1^m = (N_{\cdot,1,n}, N_{1,\cdot,n})$, $X_2^m = N_{1,1,n}$, $\lambda_1^m = \theta$, $\lambda_2^m = (\eta, p_{1,\cdot})$, and E_1^m be expectation with respect to the distribution of X_1^m. We compute the Fisher information matrix and the P-information about θ in the presence of $(\eta, p_{1,\cdot})$ in Section 4.2 of the online supplement and report the P-information here:

$$I^m(\theta;(\eta,p_{1,\cdot})) = \frac{np_{1,\cdot}\eta}{\theta(1+\theta\eta)^2} - \left(\frac{np_{1,\cdot}}{(1+\theta\eta)^2}\right)^2\left(\frac{np_{1,\cdot}\theta}{\eta(1+\theta\eta)^2} + \frac{n(1-p_{1,\cdot})}{\eta(1+\eta)^2}\right)^{-1}. \tag{10}$$

We want to find the P-information loss from conditioning on both margins.

The conditional distribution of $N_{1,1,n}$ given both margins is the noncentral hypergeometric distribution with parameters $(\theta, n, k_{1,\cdot,n}, k_{\cdot,1,n})$, which depends solely on θ, so Proposition 1 applies and we obtain

$$L_{2|1}^m(\theta;(\eta,p_{1,\cdot})) = I^m(\theta;(\eta,p_{1,\cdot})) - I_{2|1}^m(\theta;(\eta,p_{1,\cdot})) = I_1^m(\theta;(\eta,p_{1,\cdot})). \tag{11}$$

The ratio $L_{2|1}^m(\theta;(\eta,p_{1,\cdot}))/I^m(\theta;(\eta,p_{1,\cdot}))$ is the fraction of P-information that is lost by conditioning on the random margins. In analogy to the independent binomial case, the Fisher information in the conditional distribution of $N_{1,1,n}$ given both margins is Z_n/θ^2, where Z_n is the conditional variance of $N_{1,1,n}$ given both margins. As a function of the two margins, Z_n has the same formula as in binomial sampling, and

$$I^m_{2|1}(\theta; (\eta, p_{1,\cdot})) = E^m_1\left(\frac{Z_n}{\theta^2}\right). \tag{12}$$

Proofs of the above results along with a formula for computing (12) appear in Section 4.2 of the online supplement.

4. The Main Result

In this section we present results about the fraction of the P-information about θ in the presence of nuisance parameter(s) that is lost by conditioning on the random margin(s) as the number of observations in the table goes to infinity.

Theorem 1. *In the case of independent binomial sampling, assume that both success probabilities are strictly between 0 and 1. Let $\{(k_{1,\cdot,n}, k_{2,\cdot,n})\}_{n=1}^{\infty}$ be the sequence of sample sizes that satisfy $k_{2,\cdot,n} = n - k_{1,\cdot,n}$ for each n. Assume that both $k_{1,\cdot,n}$ and $k_{2,\cdot,n}$ go to infinity with n. Then, for all θ and η,*

$$\frac{L^b_{2|1}(\theta;\eta)}{I^b(\theta;\eta)} \to 0 \text{ as } n \to \infty. \tag{13}$$

In the case of multinomial sampling, assume that all cell probabilities are strictly positive. Then, for all θ, η, and $p_{1,\cdot}$,

$$\frac{L^m_{2|1}(\theta;(\eta, p_{1,\cdot}))}{I^m(\theta;(\eta, p_{1,\cdot}))} \to 0 \text{ as } n \to \infty, \tag{14}$$

Note that the claims of convergence in Theorem 1 are pointwise rather than uniform in θ, η, and/or $p_{1,\cdot}$. The proof of Theorem 1 appears in Section 5 of the online supplement.

To put the conclusions of Theorem 1 into words, in both sampling plans, the fraction of the P-information about θ in the presence of the nuisance parameters that is lost by conditioning on the random margin(s) converges to 0 as $n \to \infty$. In the terminology of Fisher (1935)[p. 42], Theorem 1 shows that conditioning on both margins generates an *efficient* statistic about θ for the one remaining degree of freedom in the 2×2 table. Since Proposition 1 says that the numerators of (13) and (14) equal $I^b_1(\theta;\eta)$ and $I^m_1(\theta;(\eta, p_{1,\cdot}))$ respectively, Theorem 1 adds some weight to the comment of Aitkin & Hinde (1984), in the discussion of Yates (1984) (and quoted in Section 1), that the information about θ in the margins might be negligible in large samples.

5. How $\theta = 1$ is special

Zhu & Reid (1994)[Example 2] use independent binomial sampling as an example of their theory of P-information. They state that it can be shown that $L^b_{2|1}(\theta;\eta) \neq 0$ unless $\theta = 1$. In Section 6 of the online supplement, we show that $L^b_{2|1}(1;\eta) = 0$ and that both $L^b_{2|1}(\theta;\eta) > 0$ and $L^m_{2|1}(\theta;(\eta, p_{1,\cdot})) > 0$ for $\theta \neq 1$. We also show that $L^m_{2|1}(1;(\eta, p_{1,\cdot})) = I^m(1;(\eta, p_{1,\cdot}))/n > 0$ for all η and $p_{1,\cdot}$, hence one loses a fraction $1/n$ of the P-information about $\theta = 1$ by conditioning on the random margins in multinomial sampling. If Fisher's idea of "information on the point at issue namely, as to the proportionality of the frequencies in the body of the table" (from the quote in Section 1) had been P-information about $\theta = 1$ in the presence of a nuisance parameter (a concept not yet articulated in the literature at the time) in independent binomial sampling only, then the above analysis would support that interpretation. The interpretation, however, is undermined by the fact that, two pages later Fisher (1935)[p. 50] suggests using the conditional distribution of $N_{1,1,n}$ given both margins to draw inference about values of $\theta \neq 1$, despite the fact that one loses P-information about values of $\theta \neq 1$.

Zhu & Reid (1994)[Example 2] also gives a numerical example with a plot of $L^b_{2|1}(\theta;\eta)/I^b(\theta;\eta)$ showing where one loses more or less P-information in that example. The ratio is largest in regions where η and θ take extreme but different values. That is, where η is large and θ is small and vice versa. This example sheds some light on the quote of Yates (1984) mentioned in Section 1 about the margins not containing information on proportionality except in "extreme cases. . . "

Acknowledgement

The authors would like to thank Ruobin Gong of Rutgers University for helpful discussions. They also thank two referees for their careful reading of an earlier version and their thoughtful suggestions for improvements.

References

Aitkin, M. & Hinde, J. P. (1984). Discussion of "Tests of significance for 2×2 contingency tables" by F. Yates. *Journal of the Royal Statistical Society. Series A (General)* **147**, 453–454.

Amari, S.-I. (1987). Discussion of the paper by Cox and Reid. *Journal of the Royal Statistcal Society. Series B (Methodological)* **49**, 26–27.

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

Barnard, G. A. (1984). Discussion of "Tests of significance for 2×2 contingency tables" by F. Yates. *Journal of the Royal Statistical Society. Series A (General)* **147**, 449–450.

Begun, J. M., Hall, W. J., Huang, W.-M. & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics* **11**, 432–452.

Bhapkar, V. P. (1989). Conditioning on ancillary statistics and loss of information in the presence of nuisance parameters. *Journal of Statistical Planning and Inference* **21**, 139–160.

Bhapkar, V. P. (1991). Loss of information in the presence of nuisance parameters and partial sufficiency. *Journal of Statistical Planning and Inference* .

Chernoff, H. (2004). Information, for testing the equality of two probabilities, from the margins of the 2×2 table. *Journal of Statistical Planning and Inference* **121**, 209–214.

Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society* **98**, 39–82.

Kou, S. G. & Ying, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statistica Sinica* **6**, 809–829.

Little, R. J. A. (1989). Testing the equality of two independent binomial proportions. *The American Statistician* **43**, 283–288.

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

Pearson, E. S. & Merrington, M. (1948). 2×2 tables: The power function of the test on a randomized experiment. *Biometrika* **35**, 331–345.

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

Yates, F. (1984). Tests of significance for 2×2 contingency tables. *Journal of the Royal Statistical Society. Series A (General)* **147**, 426–449.

Zhu, Y. & Reid, N. (1994). Information, ancillarity, and sufficiency in the presence of nuisance parameters. *The Canadian Journal of Statistics* **22**, 111–123.

[*Received on* 16 *July* 2024. *Editorial decision on* 16 *July* 2024]

Biometrika (2024), **103**, 1, *p.* 1
Printed in Great Britain Advance Access publication on August 19, 2025

Online Supplement to "Information lost by conditioning on the random margins in a two-by-two table"

By M. J. SCHERVISH
Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.
mark@cmu.edu

J. B. KADANE
Department of Statistics and Data Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.
kadane@stat.cmu.edu

T. SEIDENFELD
Department of Philosophy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, U.S.A.
ts1s@andrew.cmu.edu

and R. B. STERN
Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil
rbstern@gmail.com

In this supplement, numerical references to the main article have numbers that begin with M-. The equations will reappear in this document to aid the reader.

Section 1 gives a general definition of P-information and explains how to compute P-information when the key parts of the information matrix are singular. Section 2 contains the proof that the fraction of P-information is invariant under certain transformations. Section 3 describes additional sampling plans that apply to 2×2 tables. Section 4 gives details of Fisher information and P-information calculations. Section 5 contains the proof of our main result, namely that as the sample size of the table goes to infinity, the fraction of the P-information about θ that is lost from conditioning on the random margins goes to 0. Section 6 proves some results on P-information loss that center around $\theta = 1$. Section 7 shows that there is no P-information loss from conditioning on only one margin.

1. General Definition of P-information

Equation (M-5)

$$I(\lambda_1; \lambda_2) = J_{\lambda_1,\lambda_1} - J_{\lambda_1,\lambda_2} J^{-1}_{\lambda_2,\lambda_2} J_{\lambda_2,\lambda_1}. \tag{1}$$

is the special case of a more general definition of P-information as presented by Zhu & Reid (1994)[Section 2.1]. Consider the linear space of random variables with zero-mean and finite variance that is the linear span of the coordinates of the score function ℓ (partial derivatives of the log-likelihood with respect to the parameters). This linear space is an inner-product space with the inner product being covariance. Partition the score function into the coordinates Y_1 that are derivatives with respect to the parameters of interest and Y_2 that are derivatives with respect to the nuisance parameters. Let P_{λ_2} be the orthogonal projection operator onto the subspace spanned by the coordinates of Y_2. The P-information about λ_1 in the presence of λ_2 is the covariance matrix of

$$Y_1 - P_{\lambda_2} Y_1. \tag{2}$$

Note that (2) is analogous to the residuals of a linear regression of Y_1 on Y_2 (multivariate if Y_1 has multiple coordinates, and/or multiple if Y_2 has multiple coordinates). The covariance matrix of (2) is commonly called the *partial covariance matrix of* Y_1 *given* Y_2. The following result is then straightforward.

PROPOSITION 1. *The partial covariance matrix of* Y_1 *given* Y_2 *is all zeros if and only if* Y_1 *is almost surely a linear function of* Y_2.

All of the cases of singular J_{λ_2,λ_2} that appear in this paper and the main document arise when some coordinates of λ_2 do not appear in the marginal or conditional likelihood being studied. They each have the property that the coordinates of Y_2 corresponding to the parameters that don't appear are identically 0, and the remaining coordinates (if any) are linearly independent. In these cases, (i) either the whole vector Y_2 is 0, in which case everything projects to 0, or (ii) the space spanned by the nonzero coordinates Y_2^* of Y_2 span the same space spanned by Y_2. In case (i) the P-information is J_{λ_1,λ_1}. In case (ii), let λ_2^* be the coordinates of λ_2 that do appear in the likelihood being studied. Then $P_{\lambda_2} = P_{\lambda_2^*}$ projects the jth coordinate of Y_1 to the jth coordinate of $J_{\lambda_1,\lambda_2^*} J^{-1}_{\lambda_2^*,\lambda_2^*} Y_2^*$. This discussion can be summarized by Proposition 2 below.

PROPOSITION 2. *In the situations and notation described above: If* J_{λ_2,λ_2} *is all zeros, P-information is* J_{λ_1,λ_1}. *If there are zero rows and columns in* J_{λ_2,λ_2} *but the matrix is not all zeros, P-information is*

$$J_{\lambda_1,\lambda_1} - J_{\lambda_1,\lambda_2^*} J^{-1}_{\lambda_2^*,\lambda_2^*} J_{\lambda_2^*,\lambda_1}.$$

More general cases of singular J_{λ_2,λ_2} are more complicated to describe, but none of them appear in this work.

2. PROOF THAT FRACTION OF P-INFORMATION IS INVARIANT UNDER CERTAIN TRANSFORMATIONS

If $I(\lambda)$ is an information matrix for a parameter λ and $\psi = g(\lambda)$ is a one-to-one reparameterization with differentiable inverse, then the information matrix about ψ is $JI(g^{-1}(\psi))J^\top$, where J is the Jacobian matrix containing the partial derivatives of the coordinates of g^{-1} with respect to the coordinates of ψ. If A and B are information matrices for a parameter λ, then the matrix corresponding to $A^{-1}B$ for the parameter ψ evaluated at $\psi = g(\lambda)$ is

$$(J^\top)^{-1} A^{-1} J^{-1} J B J^\top = (J^\top)^{-1} A^{-1} B J^\top.$$

Then $A^{-1}By = \alpha y$ for some vector y if and only if $(J^\top)^{-1}A^{-1}BJ^\top z = \alpha z$ for $z = (J^\top)^{-1} y$. Hence, the transformation g preserves the eigenvalues.

3. SOME OTHER SAMPLING PLANS

Independent binomial and multinomial sampling are not the only sampling plans that lead to 2×2 tables. For example, the total number of cases n might not be known in advance, that is N

might have a distribution on the positive integers while the conditional distribution of the table given $N = n$ is that of a multinomial sample of size n. In such a case, it is concievable that the distribution of N might depend in some way on the other parameters of the sampling plan. For example, Fisher (1935)[p. 370-371], in a genetic study of Lythrum salicaria plants, offers three models of random sample size. Some of the models introduce dependence between the sample size and the plant characteristics under study.

If the distribution of N is known in advance or depends only on parameters unrelated to the parameters of the 2×2 table, and if the conditional distribution of the table given $N = n$ is unrestricted multinomial with positive, but otherwise unrestricted cell probabilities, then there is no P-information loss from conditioning on $N = n$, just as in the case in which n is known in advance.

Another example would be two independent binomial samples with random sample sizes whose joint distribution is more general than the joint distributions that arise from the sampling plans already described. For example, the distribution of one of the random sample sizes given the total might not be binomial. Even in such cases, so long as the joint distribution of the random sample sizes is unrelated to the two binomial probability parameters, the analysis conditional on the sample sizes would be much like the case in which the sample sizes were known in advance.

4. Details of information calculations

In general, each information matrix is the covariance matrix of the corresponding score function, $\mathrm{var}(\ell)$. Calculating this variance will often be done using the formula:

$$\mathrm{var}(\ell) = E[\mathrm{var}(\ell|W)] + \mathrm{var}[E(\ell|W)], \tag{3}$$

where W is typically one of the margins. This is particularly useful for information matrices based on conditional distributions given one of the margins where $\mathrm{var}[E(\ell|W)] = 0$ because $E(\ell|W)$ is constant. It is also useful in multinomial sampling when the conditional score functions given the first margin match the score functions from binomial sampling.

4.1. *Independent Binomial Sampling*

Under independent binomial sampling with fixed row margin totals of $k_{1,\cdot,n}$ and $n - k_{1,\cdot,n}$, we can represent the 2×2 table as $(N_{\cdot,1,n}, N_{1,1,n})$ with probability function

$$f^b_{N_{\cdot,1,n},N_{1,1,n}}(k_{\cdot,1,n}, x) = \eta^{k_{\cdot,1,n}}(1+\theta\eta)^{-k_{1,\cdot,n}}(1+\eta)^{k_{1,\cdot,n}-n}\binom{k_{1,\cdot,n}}{x}\binom{n-k_{1,\cdot,n}}{k_{\cdot,1,n}-x}\theta^x, \tag{4}$$

for $k_{\cdot,1,n} \in \{0, \ldots, n\}$ and $x \in \{\max\{0, k_{1,\cdot,n} + k_{\cdot,1,n} - n, \ldots, \min\{k_{1,\cdot,n}, k_{\cdot,1,n}\}\}$.

From the above probability function we compute the Fisher information matrix $I^b(\lambda^b)$. The partial derivatives of the log-likelihood, which make up the score function, are

$$\begin{aligned}
\frac{\partial}{\partial\theta} &= \frac{N_{1,1,n}}{\theta} - \frac{k_{1,\cdot,n}\eta}{1+\theta\eta}, \\
\frac{\partial}{\partial\eta} &= \frac{N_{\cdot,1,n}}{\eta} - \frac{k_{1,\cdot,n}\theta}{1+\theta\eta} + \frac{k_{1,\cdot,n}-n}{1+\eta}.
\end{aligned}$$

Note that $N_{\cdot,1,n} = N_{1,1,n} + N_{2,1,n}$ with $N_{1,1,n}$ and $N_{2,1,n}$ being independent with binomial distributions having parameters $(k_{1,\cdot,n}, \theta\eta/(1+\theta\eta))$ and $(n - k_{1,\cdot,n}, \eta/(1+\eta))$ respectively. Hence

$\mathrm{cov}(N_{1,1,n}, N_{\cdot,1,n}) = \mathrm{var}(N_{1,1,n})$, and the covariance matrix of the score function is

$$I^b(\lambda^b) = \begin{pmatrix} \dfrac{\mathrm{var}(N_{1,1,n})}{\theta^2} & \dfrac{\mathrm{var}(N_{1,1,n})}{\theta\eta} \\ \dfrac{\mathrm{var}(N_{1,1,n})}{\theta\eta} & \dfrac{\mathrm{var}(N_{1,1,n}) + \mathrm{var}(N_{2,1,n})}{\eta^2} \end{pmatrix} = \begin{pmatrix} \dfrac{k_{1,\cdot,n}\eta}{\theta(1+\theta\eta)^2} & \dfrac{k_{1,\cdot,n}}{(1+\theta\eta)^2} \\ \dfrac{k_{1,\cdot,n}}{(1+\theta\eta)^2} & \dfrac{k_{1,\cdot,n}\theta}{\eta(1+\theta\eta)^2} + \dfrac{n-k_{1,\cdot,n}}{\eta(1+\eta)^2} \end{pmatrix}. \tag{5}$$

It is straightforward that (M-6)

$$I^b(\theta;\eta) = \frac{k_{1,\cdot,n}\eta}{\theta(1+\theta\eta)^2} - \left(\frac{k_{1,\cdot,n}}{(1+\theta\eta)^2}\right)^2 \left(\frac{k_{1,\cdot,n}\theta}{\eta(1+\theta\eta)^2} + \frac{n-k_{1,\cdot,n}}{\eta(1+\eta)^2}\right)^{-1}, \tag{6}$$

follows from (5).

For the P-information based on the conditional distribution, we use the fact that $N_{1,1,n}$ given the random margin has the probability function (M-7) with $d = n$, $m_1 = k_{1,\cdot,n}$ and $m_2 = k_{\cdot,1,.n}$:

$$f^b_{N_{1,1,n}|N_{\cdot,1,n}}(x|k_{\cdot,1,n}) = \frac{\binom{k_{1,\cdot,n}}{x}\binom{n-k_{1,\cdot,n}}{k_{\cdot,1,n}-x}\theta^x}{\sum_{y=\max\{0,k_{1,\cdot,n}+k_{\cdot,1,n}-n\}}^{\min\{k_{1,\cdot,n},k_{\cdot,1,n}\}} \binom{k_{1,\cdot,n}}{y}\binom{n-k_{1,\cdot,n}}{k_{\cdot,1,n}-y}\theta^y}, \tag{7}$$

for $x = \max\{0, k_{1,\cdot,n} + k_{\cdot,1,n} - n\}, \ldots, \min\{k_{1,\cdot,n}, k_{\cdot,1,n}\}$. The score function and its variance involve sums of the form

$$S_j = \sum_{x=\max\{0,k_{1,\cdot,n}+k_{\cdot,1,n}-n\}}^{\min\{k_{1,\cdot,n},k_{\cdot,1,n}\}} x^j \binom{k_{1,\cdot,n}}{x}\binom{n-k_{1,\cdot,n}}{k_{\cdot,1,n}-x}\theta^x, \tag{8}$$

for $j = 0, 1, 2$. Note that, as random variables, each S_j is a function of $k_{\cdot,1,n}$, the specific value of $N_{\cdot,1,n}$. Also, note that S_j/S_0 is the conditional mean of $N^j_{1,1,n}$ given $N_{\cdot,1,n} = k_{\cdot,1,n}$, and

$$\frac{\partial}{\partial\theta} S_j = \frac{S_{j+1}}{\theta}.$$

The derivative of the conditional log-likelihood with respect to θ is the conditional score function:

$$\frac{N_{1,1,n}}{\theta} - \frac{S_1}{\theta S_0}, \tag{9}$$

The variance of (9) is the upper-left diagonal entry of $I^b_{2|1}(\lambda^b)$, which we calculate using (3) with $W = N_{\cdot,1,n}$. Since the conditional mean of $N_{1,1,n}$ given $N_{\cdot,1,n}$ is S_1/S_0, the conditional mean of (9) is 0, which has variance 0. The conditional variance of $N_{1,1,n}$ given $N_{\cdot,1,n}$ is

$$Z_n = \frac{S_2}{S_0} - \frac{S_1^2}{S_0^2}, \tag{10}$$

hence the variance of (9) is the mean of the conditional variance, $E^b_1(Z_n)/\theta^2$. So,

$$I^b_{2|1}(\lambda^b) = \begin{pmatrix} \dfrac{E^b_1(Z_n)}{\theta^2} & 0 \\ 0 & 0 \end{pmatrix}.$$

From Proposition 2 in Section 1 above, the P-information about θ based on the conditional distribution of the table given $N_{\cdot,1,n}$ is the upper-left diagonal entry of $I^b_{2|1}(\lambda)$, because the second row and column are all zeros. That is

$$I^b_{2|1}(\theta;\eta) = \frac{E^b_1(Z_n)}{\theta^2}. \tag{11}$$

Since Z_n is a function (9) of $N_{.,1,n}$, a computational formula for (11) is $1/\theta^2$ times the mean of (9) with respect to the marginal probability function,

$$f^b_{N_{.,1,n}}(k_{.,1,n}) = \eta^{k_{.,1,n}}(1+\theta\eta)^{-k_{1,.,n}}(1+\eta)^{k_{1,.,n}-n}S_0, \tag{12}$$

for $k_{.,1,n} \in \{0,\dots,n\}$, of $N_{.,1,n}$. That is,

$$I^b_{2|1}(\theta;\eta) = \sum_{k_{.,1,n}=0}^{n} f^b_{N_{.,1,n}}(k_{.,1,n})\frac{1}{\theta^2}\left(\frac{S_2}{S_0} - \frac{S_1^2}{S_0^2}\right).$$

4.2. *Multinomial Sampling*

Under multinomial sampling, we can represent the 2×2 table by $(N_{1,.,n}, N_{.,1,n}, N_{1,1,n})$. In this case, the probability function of the table is

$$\begin{aligned} &f^m_{N_{1,.,n},N_{.,1,n},N_{1,1,n}}(k_{1,.,n}, k_{.,1,n}, x) \qquad (13)\\ &= \binom{n}{k_{1,.,n}} p_{1,.}^{k_{1,.,n}}(1-p_{1,.})^{n-k_{1,.,n}}\eta^{k_{.,1,n}}(1+\theta\eta)^{-k_{1,.,n}}(1+\eta)^{k_{1,.,n}-n}\binom{k_{1,.,n}}{x}\binom{n-k_{1,.,n}}{k_{.,1,n}-x}\theta^x, \end{aligned}$$

for $(k_{1,.,n}, k_{.,1,n}) \in \{0,\dots,n\}^2$ and $\max\{0, k_{1,.,n}+k_{.,1,n}-n\} \le x \le \min\{k_{1,.,n}, k_{.,1,n}\}$.

The partial derivatives of the log-likelihood with respect to θ and η are the coordinates of the score function ℓ:

$$\begin{aligned} \frac{\partial}{\partial\theta} &= \frac{N_{1,1,n}}{\theta} - \frac{N_{1,.,n}\eta}{1+\theta\eta}, \\ \frac{\partial}{\partial\eta} &= \frac{N_{.,1,n}}{\eta} - \frac{N_{1,.,n}\theta}{1+\theta\eta} + \frac{N_{1,.,n}-n}{1+\eta} \\ \frac{\partial}{\partial p_{1,.}} &= \frac{N_{1,.,n}-np_{1,.}}{p_{1,.}(1-p_{1,.})} \end{aligned}$$

We compute the covariance matrix of the score function using (3) with $W = N_{1,.,n}$. The upper-left diagonal 2×2 subtable of the conditional covariance matrix in (3) is just $I^b(\lambda^b)$ from (5) with $k_{1,.,n}$ replaced by $N_{1,.,n}$, and the mean of that submatrix is $I^b(\lambda^b)$ with $k_{1,.,n}$ replaced by $E_1(N_{1,.,n}) = np_{1,.}$. The rest of the mean of the conditional variance is all zeros because the third coordinate of ℓ is constant given $N_{1,.,n}$. The conditional mean part of (3) is two zeros followed by the third coordinate of ℓ. The covariance matrix of that vector is all zeros except the lower-right diagonal entry, which is $n/[p_{1,.}(1-p_{1,.})]$. So

$$I^m(\lambda^m) = \begin{pmatrix} \frac{np_{1,.}\eta}{\theta(1+\theta\eta)^2} & \frac{np_{1,.}}{(1+\theta\eta)^2} & 0 \\ \frac{np_{1,.}}{(1+\theta\eta)^2} & \frac{np_{1,.}\theta}{\eta(1+\theta\eta)^2} + \frac{n(1-p_{1,.})}{\eta(1+\eta)^2} & 0 \\ 0 & 0 & \frac{n}{p_{1,.}(1-p_{1,.})} \end{pmatrix}. \tag{14}$$

It is straightforward that (M-10)

$$I^m(\theta;(\eta,p_{1,.})) = \frac{np_{1,.}\eta}{\theta(1+\theta\eta)^2} - \left(\frac{np_{1,.}}{(1+\theta\eta)^2}\right)^2\left(\frac{np_{1,.}\theta}{\eta(1+\theta\eta)^2} + \frac{n(1-p_{1,.})}{\eta(1+\eta)^2}\right)^{-1}, \tag{15}$$

follows from (14).

For the P-information based on the conditional distribution, we use the fact that $N_{1,1,n}$ given the two margins is the same as in binomial sampling with first row total $k_{1,\cdot,n}$. That is

$$f^m_{N_{1,1,n}|N_{1,\cdot,n},N_{\cdot,1,n}}(x|k_{1,\cdot,n},k_{\cdot,1,n}) = f^b_{N_{1,1,n}|N_{\cdot,1,n}}(x|k_{\cdot,1,n}). \tag{16}$$

from (7). The corresponding score function as well as the covariance matrix are the same as in binomial sampling with first row total $k_{1,\cdot,n}$, with the understanding that both $N_{1,\cdot,n}$ and $N_{\cdot,1,n}$ are random variables in this case. So

$$I^m_{2|1}(\lambda^m) = \begin{pmatrix} \dfrac{E^m_1(Z_n)}{\theta^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where Z_n is the conditional variance of $N_{1,1,n}$ given $(N_{1,\cdot,n}, N_{\cdot,1,n})$, and has the same formula as (10) with the understanding that both $N_{1,\cdot,n}$ and $N_{\cdot,1,n}$ are random variables in this case. Since $I^m_{2|1}(\lambda^m)$ is all zeros except for the upper-left diagonal entry, Proposition 2 in Section 1 says that the P-information based on the conditional distribution is then

$$I^m_{2|1}(\theta;(\eta,p_{1,\cdot})) = \frac{E^m_1(Z_n)}{\theta^2}. \tag{17}$$

Because Z_n is a function of $(N_{1,\cdot,n}, N_{\cdot,1,n})$, a computational formula for (17) is $1/theta^2$ times the mean of that function with respect to the joint probability function,

$$f^m_{N_{1,\cdot,n},N_{\cdot,1,n}}(k_{1,\cdot,n},k_{\cdot,1,n}) = \binom{n}{k_{1,\cdot,n}} p_{1,\cdot}^{k_{1,\cdot,n}}(1-p_{1,\cdot})^{n-k_{1,\cdot,n}}\eta^{k_{\cdot,1,n}}(1+\theta\eta)^{-k_{1,\cdot,n}}(1+\eta)^{k_{1,\cdot,n}-n}S_0, \tag{18}$$

for $(k_{1,\cdot,n}, k_{\cdot,1,n}) \in \{0,\ldots,n\}^2$, of $(N_{1,\cdot,n}, N_{\cdot,1,n})$. That is,

$$I^m_{2|1}(\theta;(\eta,p_{1,\cdot})) = \sum_{k_{1,\cdot,n}=0}^{n}\sum_{k_{\cdot,1,n}=0}^{n} f^m_{N_{1,\cdot,n},N_{\cdot,1,n}}(k_{1,\cdot,n},k_{\cdot,1,n})\frac{1}{\theta^2}\left(\frac{S_2}{S_0}-\frac{S_1^2}{S_0^2}\right).$$

5. PROOF OF THEOREM M-1

The proof of Theorem M-1 relies on the following four results.

LEMMA 1. (Kou & Ying, 1996, Corollary 3.4) *For each $n = 1,2,\ldots$, let W_n have the noncentral hypergeometric distribution with parameters $(\theta, n, m_{1,n}, m_{2,n})$. If*

$$\lim_{n\to\infty}\frac{m_{1,n}(n-m_{1,n})m_{2,n}(n-m_{2,n})}{n^3} = \infty, \tag{19}$$

then

$$\mathrm{var}(W_n)\left[\frac{1}{W_n}+\frac{1}{W_n+n-m_{1,n}-m_{2,n}}+\frac{1}{m_{1,n}-W_n}+\frac{1}{m_{2,n}-W_n}\right]\to 1, \tag{20}$$

in probability as $n\to\infty$.

LEMMA 2. *Let $(\Omega,\mathcal{B},P)$ be a probability space. Let $A\in\mathcal{B}$ be such that $P(A)=1$. Define the probability space (A,C,Q) by $C=\{A\cap B : B\in\mathcal{B}\}$ and $Q(A\cap B)=P(B)$. For each random variable Y defined on Ω, define Y^A on A as the restriction of Y to A. Let $\{Y_n\}_{n=1}^\infty$ and Y be random variables defined on Ω. If $Y_n^A\to Y^A$ in probability under Q, then $Y_n\to Y$ in probability under P.*

Proof. For each $\epsilon > 0$ there exists m_ϵ such that $n \geq m_\epsilon$ implies that $Q(|Y_n^A - Y^A| > \epsilon) < \epsilon$. Since $P(|Y_n - Y| > \epsilon) = Q(|Y_n^A - Y^A| > \epsilon)$, the result follows. □

Lemma 3. (Kou & Ying, 1996, Corollary 3.1)

$$Z_n \leq \frac{k_{1,.n}(n - k_{1,.,n})k_{.,1,n}(n - k_{.,1,n})}{n^2(n-1)} \max\left\{\theta, \frac{1}{\theta}\right\}.$$

Lemma 4. *Assume that $\{V_n\}_{n=1}^{\infty}$ is a sequence of random variables that are uniformly bounded and such that $V_n \to v$ in probability, where v is a constant. Let $\mu_n = E(V_n)$. Then μ_n converges to v.*

Proof. Let c be a uniform bound on $|V_n|$. Since the $|\mu_n|$ are also uniformly bounded by c, $2c$ is a uniform bound for $|V_n - \mu_n|$. For each $\epsilon > 0$,

$$|\mu_n - v| \leq E|V_n - v| \leq \epsilon \mathrm{pr}(|V_n - v| \leq \epsilon) + 2c\mathrm{pr}(|V_n - v| > \epsilon),$$

which converges to ϵ as $n \to \infty$. Hence $\limsup_n |\mu_n - v| \leq \epsilon$ for every positive ϵ, and μ_n converges to v. □

5.1. *Both sampling plans*

Let $\theta > 0$. For each positive integer n, let $0 \leq m_{1,n}, m_{2,n} \leq n$ and let $P_{\theta,n,m_{1,n},m_{2,n}}$ denote the distribution of a random variable that has the noncentral hypergeometric distribution with parameters $(\theta, n, m_{1,n}, m_{2,n})$. In both sampling plans, denote the cell counts of the nth table $(N_{1,1,n}, N_{1,2,n}, N_{2,1,n}, N_{2,2,n})$. Let $\{k_{1,.,n}\}_{n=1}^{\infty}$ and $\{k_{.,1,n}\}_{n=1}^{\infty}$ denote the sequence of first row and first column margin totals of the nth table respectively. The conditional distribution of $N_{1,1,n}$ given the random margin total(s) is $P_{\theta,n,k_{1,.,n},k_{.,1,n}}$ in both sampling plans. With $W_n = N_{1,1,n}$, if (19) holds for $P_{\theta,n,k_{1,.,n},k_{.,1,n}}$, (20) can be rewritten as

$$Z_n \left[\frac{1}{N_{1,1,n}} + \frac{1}{N_{2,2,n}} + \frac{1}{N_{1,2,n}} + \frac{1}{N_{2,1,n}}\right] \to 1, \tag{21}$$

in probability, where Z_n stands for the conditional variance of $N_{1,1,n}$ given the random margin(s). The distribution corresponding to the "in probability" convergence is the conditional distribution of $N_{1,1,n}$ given the random margin total(s). (The other three cell counts are determined by $N_{1,1,n}$ and the row and column totals.)

Let $u_n = k_{1,.,n}/n$ for each n, and note that

$$nu_n(1 - u_n) = \frac{k_{1,.,n}(n - k_{1,.,n})}{n}.$$

Combine each pair of terms in (21) that have a common first supscript to obtain

$$\frac{1}{N_{1,1,n}} + \frac{1}{N_{1,2,n}} = \frac{nu_n}{N_{1,1,n}N_{1,2,n}},$$
$$\frac{1}{N_{2,1,n}} + \frac{1}{N_{2,2,n}} = \frac{n(1-u_n)}{N_{2,1,n}N_{2,2,n}}.$$

Next, divide the first factor of (21) by $nu_n(1 - u_n)$ and multiply the second factor by $nu_n(1 - u_n)$ to conclude that the left-hand side of (21) equals

$$\frac{Z_n}{nu_n(1-u_n)} \left(\frac{n^2u_n^2(1-u_n)}{N_{1,1,n}N_{1,2,n}} + \frac{n^2u_n(1-u_n)^2}{N_{2,1,n}N_{2,2,n}}\right). \tag{22}$$

Also, as $n \to \infty$,

$$\frac{n^2 u_n^2}{N_{1,1,n} N_{1,2,n}} \to \frac{(1+\theta\eta)^2}{\theta\eta} \text{ almost surely, and}$$
$$\frac{n^2(1-u_n)^2}{N_{2,1,n} N_{2,2,n}} \to \frac{(1+\eta)^2}{\eta} \text{ almost surely.}$$

It follows that (22), and hence the left-hand side of (21), is asymptotic to

$$\frac{Z_n}{n u_n(1-u_n)} \left(\frac{(1-u_n)(1+\theta\eta)^2}{\theta\eta} + \frac{u_n(1+\eta)^2}{\eta} \right), \tag{23}$$

as $n \to \infty$.

The remainders of the proofs for the two sampling plans begin by showing that (19) occurs with probability 1 under each sampling plan. Lemma 2 then allows us to conclude that (20) holds with $W_n = N_{1,1,n}$.

5.2. *Independent binomial sampling*

Assume that both $k_{1,\cdot,n}$ and $n - k_{1,\cdot,n}$ are increasing and go to ∞. Under the conditions of Theorem 1, $N_{i,j,n}/k_{i,\cdot,n}$ converges almost surely to $q_{i,j}$ for all i, j as $n \to \infty$. In particular, with probability 1, both the first-column total $N_{\cdot,1,n} = N_{1,1,n} + N_{2,1,n}$ and the second-column total $N_{\cdot,2,n} = N_{1,2,n} + N_{2,2,n}$ are eventually larger than

$$\min\{q_{i,j} : i, j = 1, 2\}(k_{1,\cdot,n} + k_{2,\cdot,n})/2 = n \min\{q_{i,j} : i, j = 1, 2\}/2.$$

Also, for each n at least one of $k_{1,\cdot,n}$ or $k_{2,\cdot,n}$ is at least $n/2$ while the smaller of the two goes to ∞ at an unspecified rate. Hence, with probability 1 at least three of the factors in the numerator of (19) eventually exceed n times positive constants and the fourth goes to infinity at an unspecified rate. It follows that $P_{\theta,\eta}(A) = 1$, where A is the event that

$$\lim_{n\to\infty} \frac{k_{1,\cdot,n}(n - k_{1,\cdot,n}) N_{\cdot,1,n}(n - N_{\cdot,1,n})}{n^3} = \infty.$$

It now follows from Lemma 1 that (21) holds. Then (23) converges in probability to 1 as $n \to \infty$

Next, we rearrange the formula (6) for $I^b(\theta;\eta)$ as follows:

$$\begin{aligned}
I^b(\theta;\eta) &= \frac{k_{1,\cdot,n}\eta}{\theta(1+\theta\eta)^2} - \left(\frac{k_{1,\cdot,n}}{(1+\theta\eta)^2}\right)^2 \left(\frac{k_{1,\cdot,n}\theta}{\eta(1+\theta\eta)^2} + \frac{n-k_{1,\cdot,n}}{\eta(1+\eta)^2}\right)^{-1} \\
&= \frac{\dfrac{k_{1,\cdot,n}^2}{(1+\theta\eta)^4} + \dfrac{k_{1,\cdot,n}(n-k_{1,\cdot,n})}{\theta(1+\theta\eta)^2(1+\eta)^2} - \left(\dfrac{k_{1,\cdot,n}}{(1+\theta\eta)^2}\right)^2}{\theta\dfrac{k_{1,\cdot,n}}{\eta(1+\theta\eta)^2} + \dfrac{n-k_{1,\cdot,n}}{\eta(1+\eta)^2}} \\
&= \frac{\dfrac{k_{1,\cdot,n}(n-k_{1,\cdot,n})}{\theta(1+\theta\eta)^2(1+\eta)^2}}{\theta\dfrac{k_{1,\cdot,n}}{\eta(1+\theta\eta)^2} + \dfrac{n-k_{1,\cdot,n}}{\eta(1+\eta)^2}} \\
&= \frac{k_{1,\cdot,n}(n-k_{1,\cdot,n})}{n\theta^2}\left(\frac{k_{1,\cdot,n}(1+\eta)^2}{n\eta} + \frac{(n-k_{1,\cdot,n})(1+\theta\eta)^2}{n\theta\eta}\right)^{-1} \\
&= \frac{n u_n(1-u_n)}{\theta^2}\left(\frac{u_n(1+\eta)^2}{\eta} + \frac{(1-u_n)(1+\theta\eta)^2}{\theta\eta}\right)^{-1}.
\end{aligned} \tag{24}$$

Hence,

$$\frac{\theta^2 I^b(\theta;\eta)}{nu_n(1-u_n)}\left(\frac{u_n(1+\eta)^2}{\eta}+\frac{(1-u_n)(1+\theta\eta)^2}{\theta\eta}\right)=1. \tag{25}$$

Dividing (23), which converges to 1 in probability, by the left-hand side of (25) gives

$$\frac{Z_n}{\theta^2 I^b(\theta;n)} \to 1, \text{ in probability}, \tag{26}$$

as $n \to \infty$.

Apply Lemma 4 with V_n being the ratio on the left-hand side of (26). We show that the left-hand side of (26) is uniformly (in $n > 1$) bounded for each θ and η. Lemma 3 implies that, for all θ and $n > 1$, $Z_n/[nu_n(1-u_n)] \le \max\left\{\theta, \theta^{-1}\right\}$, because $k_{.,1,n}(n-k_{.,1,n})/[n(n-1)] \le 1$ if $n > 1$, Next we find a strictly positive uniform lower bound on $\theta^2 I^b(\theta;\eta)/[nu_n(1-u_n)]$. It is straightforward to see from (24) that $\theta^2 I^b(\theta,\eta)/[nu_n(1-u_n)]$ is eventually greater than $c_b = 1/\max\{(1+\eta)^2/\eta, (1+\theta\eta)^2/(\theta\eta)\}$. Hence, the left-hand side of (26) is uniformly (in $n > 1$) bounded by $\max\{\theta, 1/\theta\}/c_b$. This proves that

$$\frac{E_1^b(Z_n)}{\theta^2 I^b(\theta;\eta)} \to 1, \text{ as } n \to \infty. \tag{27}$$

Recall that

$$\frac{E_1^b(Z_n)}{\theta^2} = I_{2|1}^b(\theta;\eta) = I^b(\theta;\eta) - L_{2|1}^b(\theta;\eta), \tag{28}$$

by (M-8):

$$L_{2|1}^b(\theta;\eta) = I^b(\theta;\eta) - I_{2|1}^b(\theta;\eta).$$

Combining (27) with (28) implies that (M-13) holds:

$$\frac{L_{2|1}^b(\theta;\eta)}{I^b(\theta;\eta)} \to 0 \text{ as } n \to \infty.$$

5.3. *Multinomial sampling*

Each of the four margin totals divided by n converges almost surely to the positive sum of the two corresponding cell probabilities. It follows that $P_{\theta,\eta}(A) = 1$, where A is the event that

$$\lim_{n\to\infty} \frac{N_{1,.,n}(n-N_{1,.,n})N_{.,1,n}(n-N_{.,1,n})}{n^3} = \infty.$$

It now follows from Lemma 1 that (21) holds. Then (23) converges in probability to 1 as $n \to \infty$ Hence (20) occurs with probability 1.

Rearrange the formula (15) for $I^m(\theta;(\eta, p_{1,.}))$ in the same manner as we did for $I^b(\theta;\eta)$. The result is that

$$I^m(\theta;(\eta,p_{1,.})) = \frac{np_{1,}(1-p_{1,.})}{\theta^2}\left(\frac{p_{1,.}(1+\eta)^2}{\eta}+\frac{(1-p_{1,.})(1+\theta\eta)^2}{\theta\eta}\right)^{-1}. \tag{29}$$

Hence

$$\frac{\theta^2 I^m(\theta;(\eta,p_{1,.}))}{np_{1,}(1-p_{1,.})}\left(\frac{p_{1,.}(1+\eta)^2}{\eta}+\frac{(1-p_{1,.})(1+\theta\eta)^2}{\theta\eta}\right)=1. \tag{30}$$

Also, since u_n converges almost surely to $p_{1,.}$ in multinomial sampling, (23), and hence the left-hand side of (21), is asymptotic to

$$\frac{Z_n}{np_{1,.}(1-p_{1,.})}\left(\frac{(1-p_{1,.})(1+\theta\eta)^2}{\theta\eta}+\frac{p_{1,.}(1+\eta)^2}{\eta}\right), \tag{31}$$

which converges to 1 in probability as $n\to\infty$. Dividing (31) by the left-hand side of (30) gives

$$\frac{Z_n}{\theta^2 I^m(\theta;(\eta,p_{1,.}))}\to 1,\ \text{in probability}, \tag{32}$$

as $n\to\infty$.

Apply Lemma 4 with V_n being the ratio on the left-hand side of (32). We show that the left-hand side of (32) is uniformly bounded by showing that Z_n/n is uniformly bounded above (for $n>1$) and $\theta^2 I^m(\theta;(\eta,p_{1,.}))/n$ is uniformly bounded below for each θ and η. Lemma 3 says that $Z_n/n\le\max\{\theta,1/\theta\}$ for $n>1$. It is straightforward from (29) that $\theta^2 I^m(\theta;(\eta,p_{1,.}))/n$ is eventually greater than $c_m=p_{1,.}(1-p_{1,.})/\max\{(1+\eta)^2/\eta,(1+\theta\eta)^2/(\theta\eta)\}$. Hence, the left-hand side of (32) is uniformly (in $n>1$) bounded by $\max\{\theta,1/\theta\}/c_m$. This proves that

$$\frac{E_1^m(Z_n)}{\theta^2 I^m(\theta;(\eta,p_{1,.}))}\to 1,\ \text{as } n\to\infty. \tag{33}$$

Recall that

$$\frac{E_1^m(Z_n)}{\theta^2}=I_{2|1}^m(\theta;(\eta,p_{1,.}))=I^m(\theta;(\eta,p_{1,.}))-L_{2|1}^m(\theta;(\eta,p_{1,.})), \tag{34}$$

by (M-11):

$$L_{2|1}^m(\theta;(\eta,p_{1,.}))=I^m(\theta;(\eta,p_{1,.}))-I_{2|1}^m(\theta;(\eta,p_{1,.})). \tag{35}$$

Combining (33) with (34) implies that (M-14) holds:

$$\frac{L_{2|1}^m(\theta;(\eta,p_{1,.}))}{I^m(\theta;(\eta,p_{1,.}))}\to 0 \text{ as } n\to\infty.$$

6. P-INFORMATION LOSS ABOUT $\theta=1$ AND OTHER VALUES

Proposition M-1 applies in both independent binomial and multinomial sampling when conditioning on the random margin(s). In particular, Proposition M-1 says that the P-information loss from conditioning on the random margin(s) is the P-information calculated from the distribution of the random margin(s). In this section, we apply Proposition 1 in Section 1 above by closer examination of the appropriate score functions. We will denote the coordinates of the score functions corresponding to θ by Y_1^b and Y_1^m. Those coordinates corresponding to the nuisance parameters will be denoted Y_2^b and Y_2^m. We will find those cases in which Y_1^b and Y_1^m are or are not almost surely linear functions of Y_2^b and Y_2^m respectively.

6.1. *Independent biniomial sampling*

We will prove that there is no P-information loss about $\theta=1$ from conditioning on the random margin $N_{.,1,n}$. We also prove that there is strictly positive P-information loss about $\theta\neq1$ from conditioning on $N_{.,1,n}$. The distribution of the random margin $N_{.,1,n}$ has probability function (12) in Section 4.1. In the notation of that section, the coordinates of the score function are

$$Y_1^b = \frac{\partial}{\partial\theta} = -\frac{k_{1,\cdot,n}\eta}{1+\theta\eta} + \frac{S_1}{\theta S_0},$$
$$Y_2^b = \frac{\partial}{\partial\eta} = \frac{N_{\cdot,1,n}}{\eta} - \frac{k_{1,\cdot,n}\theta}{1+\theta\eta} + \frac{k_{1,\cdot,n}-n}{1+\eta}.$$

If $\theta = 1$, $S_1/S_0 = k_{1,\cdot,n}N_{\cdot,1,n}/n$, hence both Y_1^b and Y_2^b are linear functions of $N_{\cdot,1,n}$ with nonzero coefficients, and it is clear that Y_1^b is a linear function of Y_2^b. It follows that the P-information about $\theta = 1$ in the distribution of $N_{\cdot,1,n}$ is 0. So there is no P-information loss about $\theta = 1$ from conditioning on $N_{\cdot,1,n}$.

If $\theta \neq 1$,

$$\frac{S_1}{S_0} = \frac{\sum_{x=\max\{0,k_{1,\cdot,n}+N_{\cdot,1,n}-n\}}^{\min\{k_{1,\cdot,n},N_{\cdot,1,n}\}} x\binom{k_{1,\cdot,n}}{x}\binom{n-k_{1,\cdot,n}}{N_{\cdot,1,n}-x}\theta^x}{\sum_{x=\max\{0,k_{1,\cdot,n}+N_{\cdot,1,n}-n\}}^{\min\{k_{1,\cdot,n},N_{\cdot,1,n}\}} \binom{k_{1,\cdot,n}}{x}\binom{n-k_{1,\cdot,n}}{N_{\cdot,1,n}-x}\theta^x}. \tag{36}$$

Then Y_1^b is a linear function of Y_2^b if and only if S_1/S_0 is a linear function of $N_{\cdot,1,n}$. Since $k_{1,\cdot,n}$ and θ are not random, the only random variable involved in S_1/S_0 is $N_{\cdot,1,n}$. Let $g(y)$ be the value of S_1/S_0 when $N_{\cdot,1,n} = y$. To show that S_1/S_0 is not a linear function of $N_{\cdot,1,n}$, it suffices to find three values y_1, y_2, y_3 of $N_{\cdot,1,n}$ with positive probability such that $(y_j, g(y_j))$ for $j = 1, 2, 3$ do not lie on a straight line. It is straightforward to calculate $g(0) = 0$, $g(1) = k_{1,\cdot,n}\theta/(k_{1,\cdot,n}\theta + n - k_{1,\cdot,n})$, $g(n) = k_{1,\cdot,n}$. Since $k_{1,\cdot,n} \neq 0$, the points $(0, 0)$, $(1, k_{1,\cdot,n}\theta/[k_{1,\cdot,n}\theta + n - k_{1,\cdot,n}])$, and $(n, k_{1,\cdot,n})$ are not on a line if $\theta \neq 1$.

6.2. *Multinomial sampling*

We will prove that, for all θ, there is strictly positive P-information loss about θ from conditioning on the random margins $(N_{1,\cdot,n}, N_{\cdot,1,n})$. The joint distribution of the random margins has probability function (12) from Section 4.2. In the notation of that section, the coordinates of the score function are

$$Y_1^m = \frac{\partial}{\partial\theta} = -\frac{N_{1,\cdot,n}\eta}{1+\theta\eta} + \frac{S_1}{\theta S_0},$$
$$Y_{2,1}^m = \frac{\partial}{\partial\eta} = \frac{N_{\cdot,1,n}}{\eta} - \frac{N_{1,\cdot,n}\theta}{1+\theta\eta} + \frac{N_{1,\cdot,n}-n}{1+\eta},$$
$$Y_{2,2}^m = \frac{\partial}{\partial p_{1,\cdot}} = \frac{N_{1,\cdot,n}-np_{1,\cdot}}{p_{1,\cdot}(1-p_{1,\cdot})}.$$

Note that Y_2^m has two coordinates unlike in independent binomial sampling. The first summand in the formula for Y_1^m, i.e. $-N_{1,\cdot,n}\eta/(1+\theta\eta)$, is a linear function of $Y_{2,2}^m$. So, Y_1^m is a linear function of Y_2^m if and only if S_1/S_0 is a linear function of Y_2^m. As a random variable,

$$\frac{S_1}{S_0} = \frac{\sum_{x=\max\{0,N_{1,\cdot,n}+N_{\cdot,1,n}-n\}}^{\min\{N_{1,\cdot,n},N_{\cdot,1,n}\}} x\binom{N_{1,\cdot,n}}{x}\binom{n-N_{1,\cdot,n}}{N_{\cdot,1,n}-x}\theta^x}{\sum_{x=\max\{0,N_{1,\cdot,n}+N_{\cdot,1,n}-n\}}^{\min\{N_{1,\cdot,n},N_{\cdot,1,n}\}} \binom{N_{1,\cdot,n}}{x}\binom{n-N_{1,\cdot,n}}{N_{\cdot,1,n}-x}\theta^x}$$

is a function of $(N_{1,\cdot,n}, N_{\cdot,1,n})$ because θ is not random. Let $h(v)$ denote the value of S_1/S_0 when $(N_{1,\cdot,n}, N_{\cdot,1,n}) = v$. To show that there is no parameter value such that Y_1^m is almost surely a linear function of $Y_2^m = (Y_{2,1}^m, Y_{2,2}^m)$, it suffices to find four possible values v_j, $j = 1, 2, 3, 4$ of $(N_{1,\cdot,n}, N_{\cdot,1,n})$ (with positive probability) such that the four (three-dimensional) points $(v_j, h(v_j))$, for $j = 1, 2, 3, 4$ are not coplanar. It is straightforward to calculate $h(0, 0) = h(n, 0) = h(0, n) = 0$

and $h(n,n) = n$. The four corresponding points, (0,0,0), $(n,0,0)$, $(0,n,0)$, and (n,n,n) are not coplanar.

Finally, we find how much P-information loss about $\theta = 1$ there is from conditioning on both margins. We will calculate $L^m_{2|1}(1;(\eta, p_{1,\cdot}))$ according to (35) with $\theta = 1$:

$$L^m_{2|1}(1;(\eta, p_{1,\cdot})) = I^m(1;(\eta, p_{1,\cdot})) - I^m_{2|1}(1|(\eta, p_{1,\cdot}).$$

From (29), we get

$$I^m(1;(\eta, p_{1,\cdot})) = \frac{np_{1,\cdot}(1-p_{1,\cdot})\eta}{(1+\eta)^2}. \tag{37}$$

From (37), we see that

$$I^m_{2|1}(1;(\eta, p_{1,\cdot})) = E^m_1\left[\frac{N_{1,\cdot,n}(n-N_{1,\cdot,n})N_{\cdot,1,n}(n-N_{\cdot,1,n})}{n^2(n-1)}\right],$$

where $N_{1,\cdot,n}$ and $N_{\cdot,1,n}$ are independent binomial random variables with respective parameters $(n, p_{1,\cdot})$ and $(n, \eta/(1+\eta))$. It is straightforward to see that, if W is a binomial random variable with parameters (m,r), then $E[W(1-W)] = m(m-1)r(1-r)$. Hence,

$$I^m_{2|1}(1;(\eta, p_{1,\cdot})) = \frac{(n-1)p_{1,\cdot}(1-p_{1,\cdot})\eta}{(1+\eta)^2} \tag{38}$$

Subtracting (38) from (37) yields the P-information loss about $\theta = 1$ from conditioning on both margins:

$$L^m_{2|1}(1;(\eta, p_{1,\cdot})) = \frac{p_{1,\cdot}(1-p_{1,\cdot})\eta}{(1+\eta)^2},$$

which is fraction $1/n$ of the P-information without conditioning and is strictly positive.

7. Conditioning on only one margin

We consider only multinomial sampling because we have already dealt with conditioning on the only random margin in independent binomial sampling. Since P-information is invariant under transformations that preserve the parameters of interest, we examine the P-information loss from conditioning on only the row margin totals. The answer is the same for the column margin totals because there is an obvious dual parameterization in which $(\theta, \eta, p_{1,\cdot})$ is replaced by $(\theta, \gamma, p_{\cdot,1})$ with $\gamma = p_{2,1}/(p_{\cdot,1} - p_{2,1})$.

For the case of conditioning on one margin, in the notation of (M-2)

$$I^m(\lambda^m) = I^m_1(\lambda^m) + I^m_{2|1}(\lambda^n),$$

we now have $X_1 = N_{1,\cdot,n}$, $X_2 = (N_{1,1,n}, N_{\cdot,1,n})$, $\lambda_1 = \theta$, and $\lambda_2 = (\eta, p_{1,\cdot})$. The conditional distribution of X_2 given $X_1 = k_{1,\cdot,n}$ is identical to the distribution of $(N_{1,1,n}, N_{\cdot,1,n})$ in independent binomial sampling with first row margin total $k_{1,\cdot,n}$. In particular, the condition of Proposition M-1 does not hold because the conditional distribution of X_2 given X_1 involves the nusiance parameter, specifically η. The information matrix based on the conditional distribution is

$$I^m_{2|1}(\lambda^m) = \begin{pmatrix} \dfrac{np_{1,\cdot}\eta}{\theta(1+\theta\eta)^2} & \dfrac{np_{1,\cdot}}{(1+\theta\eta)^2} & 0 \\ \dfrac{np_{1,\cdot}}{(1+\theta\eta)^2} & \dfrac{np_{1,\cdot}\theta}{\eta(1+\theta\eta)^2} + \dfrac{n(1-p_{1,\cdot})}{\eta(1+\eta)^2} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{39}$$

The P-information loss from conditioning on $N_{1,\cdot,n}$ is $I^m(\theta; (\eta, p_{1,\cdot}))$ from (15) minus the result of calculating the P-information from (39). Proposition 2 in Section 1 above applies with $d = 1$, hence we can calculate P-information using only the upper-left 2×2 diagonal submatrix of (39):

$$I^m_{2|1}(\theta; (\eta, p_{1,\cdot})) = \frac{np_{1,\cdot}\eta}{\theta(1+\theta\eta)^2} - \frac{\left(\dfrac{np_{1,\cdot}}{(1+\theta\eta)^2}\right)^2}{\dfrac{np_{1,\cdot}\theta}{\eta(1+\theta\eta)^2} + \dfrac{n(1-p_{1,\cdot})}{\eta(1+\eta)^2}}, \tag{40}$$

which equals $I^m(\theta; (\eta, p_{1,\cdot}))$ from (15). So, the P-information loss from conditioning on only one margin is 0.

References

Fisher, R. A. (1935). On the selective consequences of East's (1927) theory of heterostylism in lythrum. *Journal of Genetics* **30**, 369–382.

Kou, S. G. & Ying, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statistica Sinica* **6**, 809–829.

Zhu, Y. & Reid, N. (1994). Information, ancillarity, and sufficiency in the presence of nuisance parameters. *The Canadian Journal of Statistics* **22**, 111–123.

[*Received on* 16 *July* 2024. *Editorial decision on* 16 *July* 2024]