EVALUATING APPROXIMATIONS OF BAYESIAN SOLUTIONS: THE CASE OF FISHER TEST

by

Michel MOUCHART*and Eliana SCHEIHING†

July 1993

Abstract

In this paper, we evaluate from a Bayesian point of view how much information is lost when the sampling process for the 2x2 contingency table is specified conditionally on the two margins as in the exact test of Fisher. We first analyse the general problem of admissible conditioning and next consider the evaluation of the loss of information when a non-admissible conditioning is used for an approximation of the exact posterior distribution. Turning to the Fisher test, three different sampling models are considered and three comparisons are designed between the exact and the approximate posterior distributions. The numerical results obtained through simulation indicate that for a specific range of parameters the loss of information increases with the sample size and decreases with the precision of the *a priori* distribution.

Keywords: Approximate Bayesian Solutions, Admissible conditioning, contingency tables, Fisher Test.

AMS-MOS : Primary : 62F15, Secondary: 62A15, 65C05

Acknowledgement : Financial help from the Belgian Fund for Scientific Research (FNRS) partially supported the work of the first author and is gratefully acknowledged. This version has benefited from many useful discussions with J.M. Rolin.

*Institut de Statistique, Université Catholique de Louvain, Belgium

†Institut de Statistique, Université Catholique de Louvain, Belgium and Instituto de Informática, Universidad Austral de Chile

1 Introduction

In the analysis of 2 x 2 contingency tables, the exact test of Fisher is based on the sampling distribution of the table conditional on the two margins. When learning from the data concerns only a test of independence, Fisher test has been shown to enjoy reasonable sampling properties (such as UMPU test, see Kendall and Stuart(1967), Vol.2 p. 550-552). However when inference is considered more generally, it is also known that the two margins do contain information on the parameter (λ) characterizing the independence (see, for instance, Haber(1989)).

The object of this paper is to evaluate from a Bayesian point of view the admissibility of the conditioning on the two margins. More specifically we want to evaluate how much information is lost when the sampling process is specified only conditionally on the two margins rather than jointly with the two margins. From a Bayesian point of view, this loss of information may be evaluated either through a divergence (or a distance) between the posterior distributions corresponding respectively to a complete and to an incomplete specification of the sampling process, through a comparison of the distributions of the two Bayesian estimators corresponding respectively to a complete and to an incomplete specification of the sampling process, or, finally, through a comparison of the Bayesian risks associated to optimal decision rules based respectively on the complete and on the incomplete specification of the sampling process. Note that these three comparisons correspond to three different levels of the specificity of the underlying decision problem.

It so happens that these comparisons require, as a starting point, a precise specification of the complete sampling process and of the conditions that allow one to separate the inference about the various parameters characterizing the data generating process. Indeed, one of the conclusions of this analysis is that insisting either on the multinomial sampling of the table or on independent binomial samplings implies a loss of information on λ and insisting on no loss of information on λ implies a sampling different from a multinomial one or an independent binomial one, which in turn implies that λ, the parameter characterizing the row/column independence, is likely to be not anymore a parameter sufficient for the process conditional on the margins.

In next section, we analyse the problem of admissible conditioning in general and the evaluation of the loss of information when a not admissible conditioning is specified, while section 3 considers the problem in the particular case of the 2x2 contingency table, with special emphasis on the role of the model specification. Finally section 4 presents numerical results obtained through simulation, that illustrate the loss of information involved by using a non-admissible conditioning on the two margins.

2 Admissible Conditioning in Bayesian Methods

2.1 Exact Admissibility

In this section we shortly summarize the Bayesian approach to the admissibility of the reductions by conditioning. This is the question whether it is *admissible*, in the sense of losing no information, to specify the sampling process only partially by treating part of the data "as if" it were not random. This summary is short and concentrates on the issues relevant for the problem of conditioning on the two margins. The topic of conditioning is treated most systematically in chapter 3 of Florens, Mouchart and Rolin (1990), see also, Florens and Mouchart (1977,1985) and Dawid (1980).

As usual, we consider that a Bayesian model is simply a (unique) probability measure jointly on the parameters and the observations. If we write x for the observations and θ for the parameters, the Bayesian model may accordingly be represented in terms of densities (with respect to suitable measures) as follows:

$$q(x, \theta) = m(\theta)\, p(x \mid \theta) = p(x)\, m(\theta \mid x) \tag{1}$$

where as usually

$$m(\theta \mid x) \propto m(\theta)\, p(x \mid \theta) \tag{2}$$

represents the posterior distribution of θ, i.e. the Bayesian learning rule from the data.

Let us now decompose the observations into two components y and z, i.e. $x = (y, z)$ and decompose accordingly the sampling distributions:

$$p(x \mid \theta) = p(z \mid \theta)\, p(y \mid z, \theta) \tag{3}$$

and therefore

$$m(\theta \mid x) \propto m(\theta)\, p(z \mid \theta)\, p(y \mid z, \theta) \tag{4}$$

We are interested in the following question: how far can we "forget" the term $p(z \mid \theta)$ for the inference on θ? Clearly if z is *ancillary* (i.e. $z \perp\!\!\!\perp \theta$, so that $p(z \mid \theta) = p(z)$ and $m(\theta \mid z) = m(\theta)$) we have:

$$m(\theta \mid x) \propto m(\theta)\, p(y \mid z, \theta) \tag{5}$$

so that the posterior distribution of θ remains unaffected. We say that this reduction of the sampling process by conditioning is *admissible*.

This means that when z is ancillary, inference may proceed "as if" z were not random in the sense that the sampling process needs be specified up to the distribution of the data *conditionally* on z only. Note that Sampling theory, at least under the likelihood principle, and Bayesian theory here coincide on the essential.

Suppose now that we also decompose θ into two components α and β, i.e. $\theta = (\alpha, \beta)$, where α is a nuisance parameter and β is a parameter of interest (i.e. the underlying loss function depends on β only). Therefore we only have interest in the posterior distribution of β which may be evaluated indifferently as:

$$\begin{aligned} m(\beta \mid x) &= \int m(\alpha, \beta \mid x)\, d\alpha && (6) \\ &\propto m(\beta)\, p(x \mid \beta) && (7) \end{aligned}$$

where

$$p(x \mid \beta) = \int p(x \mid \alpha, \beta)\, m(\alpha \mid \beta)\, d\alpha \tag{8}$$

may be called a *marginal data density*. Considering again the decomposition of the observations into $x = (y, z)$, we also have:

$$m(\beta \mid x) \quad \propto \quad m(\beta)\, p(z \mid \beta)\, p(y \mid z, \beta) \tag{9}$$

and raise a question similar to the previous one: how far can we forget the term $p(z \mid \beta)$ for the inference on β. Formally, the answer is also similar to the previous answer, namely that if $z \perp\!\!\!\perp \beta$, i.e. $p(z \mid \beta) = p(z)$, equivalently $m(\beta \mid z) = m(\beta)$, we also have:

$$m(\beta \mid x) \quad \propto \quad m(\beta)\, p(y \mid z, \beta) \tag{10}$$

and say that z and β are *mutually ancillary*. Clearly, if z is ancillary, z is also mutually ancillary with any function of θ, but the converse is evidently false. Note however that the use of formula (10) implies integrating the conditional sampling distribution of $(y \mid z)$ with respect to the conditional posterior distribution of α given z and β, more specifically:

$$p(y \mid z, \beta) \quad = \quad \int p(y \mid z, \alpha, \beta)\, m(\alpha \mid \beta, z) d\alpha \tag{11}$$

unless β is a sufficient parameter of the conditional sampling distribution of $(y \mid z)$ (i.e. $y \perp\!\!\!\perp \alpha \mid \beta, z$), in which case $p(y \mid z, \beta) = p(y \mid z, \alpha, \beta)$ for any prior distribution and therefore $m(\alpha \mid \beta, z)$ (which would also be equal to $m(\alpha \mid \beta, y, z)$), need not be specified for performing the integration underlying (11). When both $z \perp\!\!\!\perp \beta$ and $y \perp\!\!\!\perp \alpha \mid \beta, z$ are verified, β and z are said to be *mutually exogenous*. When β is a sufficient parameter of the sampling distribution conditional on z, a rather natural way of obtaining the mutual exogeneity of β and z (i.e. thus a stronger condition than the mutual ancillarity condition), is through the structure of a *Bayesian cut* which means that α is furthermore a sufficient parameter of the sampling marginal process of z (i.e. $z \perp\!\!\!\perp \beta \mid \alpha$) and is also, a priori independent of β; this is so simply because $z \perp\!\!\!\perp \beta \mid \alpha$ and $\alpha \perp\!\!\!\perp \beta$ trivially imply $z \perp\!\!\!\perp \beta$. The advantage of that structure is that the first two conditions ($y \perp\!\!\!\perp \alpha \mid z, \beta$ and $z \perp\!\!\!\perp \beta \mid \alpha$) only depend on the sampling process and the last condition ($\alpha \perp\!\!\!\perp \beta$) only depends on the prior specification so that the structure of a Bayesian cut is robust with respect to any modification of the prior specification, provided it keeps the prior independence of α and β. Note however that these two requirements (of sufficiency of α and β for the marginal and the conditional sampling processes) do relate the structure of the loss function (depending only on β) and of the sampling process, a not always realistic feature.

2.2 Approximate Admissibility

Suppose now that z and β are not mutually ancillary but that one decides, for some (good or bad) reason to "forget" $p(z \mid \beta)$ in equation (9), i.e. using (10) as an "approximation" of (9). Thus let us write $m_A(\beta \mid x)$, as an approximation to $m(\beta \mid x)$, defined as:

$$m_A(\beta \mid x) \propto m(\beta)\, p(y \mid z, \beta) \tag{12}$$

without having $z \perp\!\!\!\perp \beta$, (in next section, β would also be a sufficient parameter of the sampling conditional process -i.e. $y \perp\!\!\!\perp \alpha \mid z, \beta$- but this is not essential for what follows).

A natural question is now "how much information do we lose by overlooking the term $p(z \mid \beta)$?"

We will consider two ways of evaluating this loss of information. The first one, that is more inference-oriented, is based on the evaluation of the expected divergence (or distance) between the posterior distribution of parameter β, $m(\beta \mid x)$ and its approximation, $m_A(\beta \mid x)$, more specifically one may measure a loss of information L as follows:

$$L = E[D(m_A(\beta \mid x),\, m(\beta \mid x))] = \int D(m_A(\beta \mid x),\, m(\beta \mid x))\, p(x)\, dx \tag{13}$$

where $D(\,,\,)$ is a divergence (or a distance) between probability measures. Note that $D(m_A(\beta \mid x),\, m(\beta \mid x))$ is a statistic, i.e. a function of the data only and the expectation is taken with respect to the predictive distribution of the exact model. For the sake of interpretation it may be more convenient to use a divergence (or a distance) bounded in the interval [0,1]. The second way, that is more decision-oriented, compares the risk functions corresponding to the optimal decision rules under the exact posterior distribution of β and under the approximate posterior distribution of β. Under the loss function $l(\beta, a)$, the optimal decision rule under the exact posterior distribution of β is:

$$a^*_E(x) = arg \inf_a E_{\beta|x}\, l(\beta, a) \tag{14}$$

and the corresponding risk function is:

$$\rho(a^*_E) = E_{x,\beta}\, l(\beta, a^*_E(x)) \tag{15}$$

Similarly the optimal decision rule under the approximate posterior distribution of β is:

$$a^*_A(x) = arg \inf_a E^A_{\beta|x}\, l(\beta, a) \tag{16}$$

where

$$E^A_{\beta|x} l(\beta, a) = \int l(\beta, a)\, m_A(\beta \mid x)\, d\beta \tag{17}$$

and the corresponding risk function is:

$$\rho(a^*_A) = E_{x,\beta}\, l(\beta, a^*_A(x)) \tag{18}$$

Clearly

$$0 \leq \rho(a^*_E) \leq \rho(a^*_A) \tag{19}$$

Therefore we propose as a measure of the loss of information the following comparison between the two risk functions, which may be read as the "Relative Efficiency of the approximation m_A":

$$RelEff(m_A) = \frac{\rho(a^*_E)}{\rho(a^*_A)} \;\in\; [0, 1] \tag{20}$$

If this ratio takes the value $\frac{1}{k}$ this means that the risk of a_A^* is k times greater than the risk of a_E^*, so this expression is a measure of the relative increase of the risk due to the approximation on the posterior distribution.

As a particular case, consider the square loss function:

$$l(\beta, a) = (a - \beta)^2 \tag{21}$$

then

$$\begin{aligned} a_E^*(x) &= E(\beta \mid x) && (22) \\ a_A^*(x) &= E^A(\beta \mid x) && (23) \end{aligned}$$

and the corresponding risk functions are:

$$\begin{aligned} \rho(a_E^*) &= E_x\, V(\beta \mid x) && (24) \\ \rho(a_A^*) &= E_{x,\beta}\, [E^A(\beta \mid x) - \beta]^2 && (25) \\ &= E_x\, [\, V(\beta \mid x) + (\, E(\beta \mid x) - E^A(\beta \mid x)\,)^2\,] && (26) \end{aligned}$$

Therefore:

$$RelEff(m_A) = \left(1 + \frac{E_x[(\, E(\beta \mid x) - E^A(\beta \mid x)\,)^2]}{E_x\, V(\beta \mid x)}\right)^{-1} \tag{27}$$

which shows that under a quadratic loss function, an approximation will be evaluated according to the ratio between the (predictively) expected exact posterior variance and the (predictively) expected square bias of the approximate posterior expectation. An evaluation finer than $RelEff(m_A)$ would be an analysis of the predictive distribution of the statistics $D(m_A(\beta \mid x),\, m(\beta \mid x))$. A possible interest of such an analysis is to provide a natural background for a bayesian testing of the mutual exogeneity (see e.g. Florens and Mouchart [1989], [1993]).

3 Conditioning on the two margins

3.1 Introduction

Let us now consider that X, the data, are in the form of a 2x2 contingency table, *viz.* $X = (X_{11}\, X_{12}\, X_{21}\, X_{22})$. We shall make use of the dot notation for the summation, for instance $X_{1.} = X_{11} + X_{12}$, or $X_{..} = X_{11} + X_{12} + X_{21} + X_{22}$. We want to consider how far it is admissible to realize the inference conditionally on the margins, i.e., in the notation of previous section, to assume the exogeneity of $Z = (X_{1.}, X_{.1})$. This analysis requires some care when specifying the sampling process and the parameter of interest. We shall deal successively with these problems.

It is usual, in the analysis of the contingency tables, to exogeneize the grand total $X_{..}$ on the ground of a non-informative stopping rule for the sample size. Formally, this may be obtained as follows. Let α_0 and θ be sufficient parameters for the sampling distribution of $X_{..}$ and of $(X \mid X_{..})$ respectively, i.e.:

$$p(X \mid \theta, \alpha_0) = p(X_{..} \mid \alpha_0)\, p(X \mid X_{..}, \theta) \tag{28}$$

or equivalently

$$X_{..} \perp\!\!\!\perp \theta \mid \alpha_0 \tag{29}$$

and

$$X \perp\!\!\!\perp \alpha_0 \mid X_{..}, \theta \tag{30}$$

The mutual exogeneity of θ and $X_{..}$, i.e. $X_{..} \perp\!\!\!\perp \theta$ along with (30) is usually obtained by an hypothesis of a Bayesian cut, i.e. the conditions (29),(30) and the prior independence between θ and α_0, i.e.

$$\theta \perp\!\!\!\perp \alpha_0 \tag{31}$$

This implies that once the parameter of interest is a function of θ, the inference may proceed conditionally on $X_{..}$, a situation we systematically assume from now on. In other words, in what follows, we never specify the way the sample size is generated, we only assume that any unknown parameter embodied in this process is a priori independent of the parameter characterizing the process generating the table conditionally on the sample size. We shall see later that this assumption might be less innocent than the first appearance might suggest.

In order to evaluate the exogeneity of the margins $(X_{1.}, X_{.1})$ in the process conditional on $X_{..}$ one has to be more specific on the sampling process and on the parameter of interest. In what follows, the parameter of interest, denoted $\lambda = l(\theta)$, is the parameter characterizing the statistical association between the row and column criteria, a particular value of which would imply independence. We now consider successively several sampling alternatives conditional on $X_{..}$.

3.2 Multinomial sampling

A multinomial sampling distribution may be obtained as the sampling distribution of the sufficient statistic under a generalized Bernoulli sampling with non-informative stopping. More specifically let

$$W_l = (W_{11,l}, W_{12,l}, W_{21,l}, W_{22,l}) \in S_4 = \{w \in \{0,1\}^4 \mid \sum_j w_j = 1\} \tag{32}$$

be distributed as independent Bernoulli of order 4, $W_l \sim I.B_4(\theta)$, i.e.

$$p(W_l = w \mid \theta) = \theta_{11}^{w_{11}}\, \theta_{12}^{w_{12}}\, \theta_{21}^{w_{21}}\, \theta_{22}^{w_{22}}\, \mathbb{1}_{\{w \in S_4\}} \tag{33}$$

with $\theta \in \mathcal{S}_4 = \{\theta \in [0,1]^4 \mid \sum_j \theta_j = 1\}$. Clearly under independent sampling with non-informative stopping, the statistic:

$$X = (X_{ij})_{1 \le i,j \le 2} \qquad X_{ij} = W_{ij.} = \sum_{1 \le l \le n} W_{ij,l} \tag{34}$$

where $n = X_{..}$ is the sample size, is a sufficient statistic (conditionally on $X_{..}$), the distribution of which is:

$$p(X = x \mid X_{..} = n, \theta) = \frac{n!}{\prod_{1 \le i,j \le 2} x_{ij}!} \prod_{1 \le i,j \le 2} \theta_{ij}^{x_{ij}} \tag{35}$$

The parameter of interest is defined as the usual cross product ratio:

$$\lambda = \frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}} \ , \tag{36}$$

the value 1 of which characterizes the row-column independence.

In order to evaluate the admissibility of the conditioning on $Z = (X_{1.}, X_{.1})$ for the inference on λ, we notice that $(X_{11}, Z, X_{..})$ characterizes the table. The sampling distribution conditional on Z and $X_{..}$ boils therefore down to the sampling distribution of $(X_{11} \mid Z, X_{..})$, a non central hypergeometric distribution of parameter λ (see e.g. Agresti [1991], Cornfield [1956], Fisher [1935a]):

$$p(X_{11} = x_{11} \mid x_{1.}, x_{.1}, x_{..}, \theta) = \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \left(\sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \right)^{-1} \tag{37}$$

$$= \lambda^{x_{11}} \begin{pmatrix} x_{1.} \\ x_{11} \end{pmatrix} \begin{pmatrix} x_{..} - x_{1.} \\ x_{.1} - x_{11} \end{pmatrix} \left[\sum_{u \in V} \lambda^u \begin{pmatrix} x_{1.} \\ u \end{pmatrix} \begin{pmatrix} x_{..} - x_{1.} \\ x_{.1} - u \end{pmatrix} \right]^{-1}$$

where

$$V = \{x \in I\!N \mid max\{0, x_{1.} + x_{.1} - x_{..}\} \le x \le min\{x_{1.}, x_{.1}\}\} \tag{38}$$

Note that in the equation (37), each x_{ij} is a function of $(x_{11}, x_{1.}, x_{.1}, x_{..})$; therefore this formula *can not* be simplified by $\prod_{i,j} x_{ij}!$. Thus λ is a sufficient parameter for the distribution of the table conditional on the grand total $X_{..}$ and the margins $Z = (X_{1.}, X_{.1})$:

$$X \perp\!\!\!\perp \theta \mid \lambda, Z, X_{..} \tag{39}$$

However, conditionally on $X_{..}$, a Bayesian cut between the parameter of interest, λ, and the margins $Z = (X_{1.}, X_{.1})$ is impossible because the sampling distribution of the margins identifies *all* the parameters θ, indeed:

$$p(X_{1.} = x_{1.}, X_{.1} = x_{.1} \mid X_{..}, \theta) = X_{..}!\, \theta_{12}^{x_{1.}}\, \theta_{21}^{x_{.1}}\, \theta_{22}^{X_{..}-x_{1.}-x_{.1}} \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \tag{40}$$

Therefore θ is a minimal sufficient parameter of the process generating $(Z \mid X_{..})$ and no prior distribution can satisfy the condition of a (conditional) Bayesian cut, namely $\lambda \perp\!\!\!\perp \theta \mid X_{..}$, because λ is a function of θ.

As the Bayesian cut is not possible, one may ask whether some prior distribution would allow Z and λ to be mutually exogenous (conditionally on $X_{..}$), this is condition (39) along with the condition of (conditional) mutual ancillarity:

$$Z \perp\!\!\!\perp \lambda \mid X_{..} \tag{41}$$

Note that (41) implies

$$Z \perp\!\!\!\perp \lambda \tag{42}$$

because of the first cut (equations (29), (30) et (31)).
Condition (41) means that

$$p(Z \mid \lambda, X_{..}) = \int p(Z \mid \theta, X_{..})\, m(\theta_1 \mid \lambda) d\theta_1 \tag{43}$$

where $p(Z \mid \theta, X_{..})$ is given in (40) and θ_1 is defined for instance as $\theta_1 = \left(\theta_{22}, \frac{\theta_{21}}{\theta_{21}+\theta_{22}}\right)$, does not depend on λ.

As an example, let us assume that θ is distributed a priori as a Dirichlet distribution with parameter $a = (a_{ij})$ i.e.

$$m(\theta) = \frac{\Gamma(a_{..})}{\prod_{i,j} \Gamma(a_{ij})} \prod_{i,j} \theta^{a_{ij}-1} \mathbb{1}_{\{\theta \in \mathcal{S}_4\}} \tag{44}$$

It is shown, in Proposition 1 of the appendix, that in such a cse:

$$p(Z = (x_{1.}, x_{.1}) \mid \lambda, X_{..}) = c\, X_{..}! \, \frac{E[(1-Y+\lambda Y)^{-(a_{1.}+x_{1.})}]}{E[(1-W+\lambda W)^{-a_{1.}}]} \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \tag{45}$$

where

$$\begin{aligned} c &= \frac{\Gamma^2(a_{..})}{\Gamma^2(a_{..}+x_{..})} \frac{\Gamma(a_{1.}+x_{1.})\Gamma(a_{2.}+x_{2.})\Gamma(a_{.1}+x_{.1})\Gamma(a_{.2}+x_{.2})}{\Gamma(a_{1.})\Gamma(a_{2.})\Gamma(a_{.1})\Gamma(a_{.2})} \\ Y &\sim \beta(a_{.1}+x_{.1}, a_{.2}+x_{.2}) \\ W &\sim \beta(a_{.1}, a_{.2}) \end{aligned}$$

Therefore under a Dirichlet prior specification for θ, it may be seen that no specification of a is likely to make $p(Z \mid \lambda, X_{..})$ independent of λ. The computations, in the appendix, suggest that $p(Z \mid \lambda, X_{..})$ could be independent of λ only under very particular prior (non Dirichlet) specification, a theme not to be pursued in this paper leaving open the question whether there exists a prior specification $m(\theta_1 \mid \lambda)$ implying (41). The conclusion is therefore that under multinomial sampling, conditioning on the margins is generally not an admissible reduction.

3.3 Independent Binomial Sampling

In this case, the 2x2 contingency table is generated by two independent binomial samplings corresponding, say, to each column. More specifically, we again assume the first cut (29), (30), (31) and consider accordingly the grand total $X_{..}$ as exogenous. Let us write again θ for the parameter sufficient for the process generating the 2x2 contingency table, conditionally on $X_{..}$ and decompose now θ as follows:

$$\theta = (\alpha, \beta) = (\alpha, \beta_1, \beta_2) \tag{46}$$

where α is the parameter sufficient for the process generating $(X_{.1} \mid X_{..})$, more specifically α is defined by the condition:

$$X_{.1} \perp\!\!\!\perp \theta \mid \alpha, X_{..} \tag{47}$$

and β is the parameter sufficient for the process generating the table conditionally on $(X_{.1}, X_{..})$, i.e.:

$$(X_{11}, X_{12}) \perp\!\!\!\perp \theta \mid \beta, X_{.1}, X_{..} \tag{48}$$

Finally, we shall also assume the prior independence of α and β:

$$\alpha \perp\!\!\!\perp \beta \tag{49}$$

or, equivalently under the first cut:

$$\alpha \perp\!\!\!\perp \beta \mid X_{..} \tag{50}$$

This hypothesis ensures a Bayesian cut between β and the column total $X_{.1}$ conditionally on $X_{..}$. Thus for an inference on any function of β, $(X_{.1}, X_{..})$ may be considered as exogenous. The first cut, along with (48) and (49), also implies:

$$\beta \perp\!\!\!\perp (X_{.1}, X_{..}) \tag{51}$$

The independent sampling of columns may now be written as:

$$X_{11} \perp\!\!\!\perp X_{12} \mid X_{.1}, X_{..}, \theta \tag{52}$$

and β_j, $j = 1, 2$ are now the parameters sufficient for the process generating the jth column conditionally on its total (remember that $X_{.2} = X_{..} - X_{.1}$: conditionally on $X_{..}$, knowing $X_{.1}$ is therefore equivalent to knowing $X_{.2}$), more specifically the β_j's are defined by the conditions:

$$X_{1j} \perp\!\!\!\perp \theta \mid X_{.1}, X_{..}, \beta_j \quad j = 1, 2 \tag{53}$$

Finally, we assume that each column is generated as an independent binomial sampling:

$$(X_{1j} \mid X_{.j}, X_{..}, \theta) \quad \sim \quad Bi(X_{.j}, \beta_j) \quad j = 1, 2 \tag{54}$$

i.e.

$$P(X_{1j} = x_{1j} \mid X_{.j}, X_{..}, \theta) = \begin{pmatrix} X_{.j} \\ x_{1j} \end{pmatrix} \beta_j^{x_{1j}} (1 - \beta_j)^{X_{.j} - x_{1j}} \tag{55}$$

where $X_{.2} = X_{..} - X_{.1}$.

Remark: If furthermore $(X_{.1} \mid X_{..}, \alpha)$ is also assumed to follow a binomial distribution, then $(X \mid X_{..}, \theta)$ has also a multinomial distribution as in the previous section, but this is not important for what follows.

The parameter of interest is now defined as

$$\lambda = \frac{\beta_1 (1 - \beta_2)}{\beta_2 (1 - \beta_1)} \tag{56}$$

The value $\lambda = 1$, meaning $\beta_1 = \beta_2$, is often read as a condition of "homogeneity" and implies, again, a row-column independence of the 2x2 table for whatever process generating the column totals (i.e. for any distribution of $(X_{.1} \mid X_{..}, \alpha)$).

The problem is again to check whether the inference on λ, a function of β only, may be done, without loss of information, conditionally on the two totals $Z = (X_{.1}, X_{1.})$. We follow the same route as in the previous section and first check whether the sampling distribution generating $(X_{11} \mid Z, X_{..}, \theta)$ depends on λ only. This is easily seen by noticing that the sampling distribution of the 2x2 table conditionally on $(X_{.1}, X_{..})$ is an independent product of two binomial distributions with parameters $(\beta_1, X_{.1})$ and $(\beta_2, X_{..} - X_{.1})$ respectively, from which one concludes that the sampling distribution of the 2x2 table conditionally on $(Z, X_{..})$ is again a non-central hypergeometric distribution of parameter λ, exactly as in (37); thus condition (39) is again satisfied even if the column total $(X_{.1} \mid X_{..})$ is not binomially generated.

Conditionally on $X_{..}$, a Bayesian cut between the parameter of interest, λ, and the margins $Z = (X_{.1}, X_{1.})$ is again impossible. Indeed, the sampling distribution generating $(X_{1.} \mid X_{.1}, X_{..})$ depends only on β and is given by

$$\begin{aligned} p(X_{1.} = x_{1.} \mid X_{.1}, X_{..}, \beta) &= X_{.1}!\,(X_{..} - X_{.1})!(1-\beta_1)^{X_{.1}}\,\beta_2^{x_{1.}} \\ &\cdot \quad (1-\beta_2)^{X_{..} - X_{.1} - x_{1.}} \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \end{aligned} \tag{57}$$

Thus, the marginal distribution of the margins, $Z = (X_{1.}, X_{.1})$ (given the grand total $X_{..}$) identifies both β_1 and β_2 so that $\theta = (\alpha, \beta)$ is a minimal sufficient parameter of the process generating $(Z \mid X_{..})$ and no prior distribution can satisfy the condition of a Bayesian cut, namely $\lambda \perp\!\!\!\perp \theta \mid X_{..}$, because λ is a function of β and therefore of θ.

Given the impossibility of a Bayesian cut, one may still ask whether some prior specification could allow Z and λ to be mutually exogenous, i.e. (39) along with (41), or equivalently (42) under the first cut. Now (43) may be decomposed as follows

$$p(Z \mid \lambda, X_{..}) = p(X_{.1} \mid \lambda, X_{..}) p(X_{1.} \mid \lambda, X_{..}, X_{.1}) \tag{58}$$

The question whether $p(Z \mid \lambda, X_{..})$ can possibly be independent of λ may be considered in two steps. The first term, $p(X_{.1} \mid \lambda, X_{..})$ does not depend on λ under (47) and (49) because λ is a function of β only, i.e. (47) and (49) jointly imply $X_{.1} \perp\!\!\!\perp \lambda \mid X_{..}$. Let us consider the second term

$$p(X_{1.} \mid \lambda, X_{..}, X_{.1}) = \int p(X_{1.} \mid \beta_0, X_{..}, X_{.1}) m(\beta_0 \mid \lambda) d\beta_0 \tag{59}$$

where β_0 is a function of λ: $\beta_0 = b(\lambda)$ such that (λ, β_0) is a (smoothly) bijective transformation of β.

As an example, when $\beta_j' s$ are a priori distributed as independent Beta variates with parameters (a_j, b_j) (as is the case when θ has a priori a Dirichlet distribution), it is shown in Proposition 2 of the appendix that

$$\begin{aligned} p(X_{1.} = x_{1.} \mid \lambda, X_{..}, X_{.1}) &= c\, X_{.1}! X_{.2}! \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \\ &\quad \cdot \frac{E[(1 - Y + \lambda Y)^{-(a_2 + b_2 + X_{.2})}]}{E[(1 - W + \lambda W)^{-(a_2+b_2)}]} \end{aligned} \tag{60}$$

where

$$\begin{aligned} c &= \frac{\Gamma(a_. + b_.)}{\Gamma(a_.)\Gamma(b_.)} \frac{\Gamma(a_. + x_{1.})\Gamma(b. + x_{2.})}{\Gamma(a_. + b_. + X_{..})} \\ Y &\sim \beta(a_. + x_{1.}, b_. + x_{2.}) \\ W &\sim \beta(a_., b_.) \end{aligned}$$

For the same reason as for the multinomial sampling, one may conclude that under independent binomial sampling, conditioning on the margins is generally not an admissible reduction.

3.4 Sampling Conditional on the Margins

A natural idea is to consider a sampling model where the table is actually generated conditionally on the two margins $Z = (X_{.1}, X_{1.})$. This situation is exemplified in the celebrated example of tea-tasting where a lady is tested for her ability to judge whether tea or milk was poured in the cup first (see Fisher (1935b)).

We again suppose the first cut (29),(30) and (31), making $X_{..}$ exogenous for the inference on any function of θ, the parameter (minimal) sufficient for the process generating $(X \mid X_{..})$. Conditioning on the two margins would be admissible under a second cut, namely under the condition

$$\delta \perp\!\!\!\perp \lambda \mid X_{..} \tag{61}$$

or equivalently, under the first cut:

$$\delta \perp\!\!\!\perp \lambda \tag{62}$$

where δ is a parameter sufficient for the $(Z \mid X_{..})$-DGP:

$$Z \perp\!\!\!\perp \theta \mid \delta, X_{..} \tag{63}$$

and λ is a parameter sufficient, as in (37), for the $(X \mid Z, X_{..})$-DGP. Under these definitions, $\theta = (\lambda, \delta)$ but the analysis of the multinomial sampling and of the independent binomial samplings show that under (61), the sampling distribution of $(X \mid X_{..})$ may not be multinomial, and the sampling distributions of $(X_{1j} \mid X_{1.}, X_{..})$ may not be independent binomial. It is therefore fitting to ensure a proper understanding of the sampling procedure.

Let us consider, along with the tea-tasting experiment, another example. Let $X_{..}$ be the number of students in a classroom and let these students be classified into {failure, success} in a mathematics exam (row criterion) and into {type A, type B} for the schooling system followed the year before (column criterion). Here, the $(X_{1.} \mid X_{..})$-DGP would reflect a recruitment structure of the school and the $(X_{.1} \mid X_{..})$-DGP would reflect the teacher's academic excellence policy. Note that in this case one could accept that $X_{1.}$ and $X_{.1}$ be independent (conditionally on $X_{..}$) whereas in the tea-testing experiment the joint $(X_{1.}, X_{.1} \mid X_{..})$-DGP would be degenerate, in the sense that $P(X_{1.} = X_{.1} \mid X_{..}) = 1$, once the tea-taster is told, before guessing, the number of cups of each type. In both examples, the row-column independence should be characterized as a property of the $(X_{11} \mid X_{1.}, X_{.1}, X_{..})$-DGP without any restriction on the $(X_{1.}, X_{.1} \mid X_{..})$-DGP. In this conditional process, the condition of row-column independence is a condition that each individual is assigned to the row and to the column criterion as if he had been selected randomly, in the sense of a hypergeometric sampling. This is equation (37) with $\lambda = 1$, i.e. (see Plackett (1974)):

$$p(X_{11} = x_{11} \mid x_{1.}, x_{.1}, x_{..}, \theta) = \frac{\prod_{i,j} \frac{1}{x_{ij}!}}{\sum_{x_{11} \in V} \prod_{i,j} \frac{1}{x_{ij}!}} \tag{64}$$

$$= \frac{\binom{x_{1.}}{x_{11}}\binom{x_{2.}}{x_{21}}}{\binom{x_{..}}{x_{.1}}} = \frac{\binom{x_{.1}}{x_{11}}\binom{x_{.2}}{x_{12}}}{\binom{x_{..}}{x_{1.}}}$$

The nature of this biconditional sampling is such that the joint process of the table can not anymore be a counting process of independent individual units, as was the case of the multinomial or of the independent binomial samplings. For instance, in the tea-testing case, a possible sampling scheme for a tea-taster aware of his (her) absolute inability of recognizing which liquid has been poured first, would be to assign randomly each cup selected in the order of a random permutation until he faces the constraint $X_{.1} = X_{1.}$. Note that in such a case, which particular probability is used in the randomization is irrelevant as far as it is independent of the true type of the cup but the joint probability of all cups cannot be independent, if only because of the constraint $X_{.1} = X_{1.}$ (i.e. the assignation of the $n - X_{1.}$ last ones is determined by the joint assignation of the $X_{1.}$ first ones).

Thus, in the case of this conditional sampling, where the $(X_{11} \mid X_{.1}, X_{1.}, X_{..})$−distribution represents an actual sampling process, the hypergeometric distribution (64) happens to represent the only distribution displaying the row-column independence; in other words, this hypothesis of independence is a simple (a one-point) hypothesis in the space of $(X_{11} \mid X_{.1}, X_{1.}, X_{..})$-distributions but leaves completely free the $(X_{.1}, X_{1.}, X_{..})$-distribution. Furthermore, the alternative hypothesis should be specific to any actual situation. For instance, the alternative hypothesis could model the (optimal control) behavior of a tea-taster trying sequentially each cup with a view to exert her guessing ability or the teacher's policy when he would be willing to differentiate his evaluation according to the origin of his student. In such circonstances there is no reason to believe that the non-central hypergeometric distibution (37) would be the only reasonable alternative. Also, one could expect that the sampling model under the alternative hypothesis would be parametrized with more than one parameter.

In conclusion, the only situation where the sampling model conditional on the two margins does not lead to lose information is a case where the modelling of the non-independent alternative should be context specific, because the framework of counting independent units is not anymore relevant. In particular, the non-central hypergeometric distribution does not appear anymore as a natural modelling of the $(X_{11} \mid X_{.1}, X_{1.}, X_{..})$-DGP. Therefore the prior information on that parameter and the form of the critical region should also be context specific.

4 Some numerical Results

4.1 Simulation Design and Methods

In this section, we want to evaluate numerically how much information is lost when the inference on λ is based on the sampling process conditional on the margins, considered as an approximation, rather than on the actual sampling process, considered as an exact one. More specifically we want to evaluate the respective roles of the sample size and the relative precision of the prior information.

As mentioned in section 2, one may compare either the two posterior distributions, the exact one and the approximate one, by means of a (predictively) expected divergence (or distance) or the two Bayesian risk functions, by means of the relative efficiency. As those quantities are not available analytically one has to resort to numerical comparisons through simulation methods. As a matter of fact, the numerical and the statistical problems involved in the comparison of the posterior distributions are vastly different from those involved in the comparison of the risks functions. In order to keep this paper within a reasonable size we have found more useful to concentrate on the presentation of numerical results relative to the comparison of posterior distributions only, leaving for a "further work in progress" a deeper comparison of the two risks functions.

Thus, we want to evaluate an expected discrepancy between both distributions, as written in (13):

$$L = E[D(m_A(\lambda \mid x),\, m(\lambda \mid x))] \tag{65}$$

where the expectation is taken with respect to the exact predictive distribution and $D(\cdot,\cdot)$ is a distance or a divergence on the space of the distributions on λ. We consider two specifications for $D(\cdot,\cdot)$. The first one is the entropy divergence defined by:

$$\begin{aligned} D_e(m_A(\lambda \mid x),\, m(\lambda \mid x))] &= \int ln(\frac{m(\lambda \mid x)}{m_A(\lambda \mid x)})\, m(\lambda \mid x) d\lambda \qquad (66) \\ &\in (0,\infty) \end{aligned}$$

The second one is the Hellinger distance:

$$\begin{aligned} D_H(m_A(\lambda \mid x),\, m(\lambda \mid x))] &= \int (\sqrt{m(\lambda \mid x)} - \sqrt{m_A(\lambda \mid x)})^2\, d\lambda \qquad (67) \\ &= 2(1 - \int \sqrt{m(\lambda \mid x)\, m_A(\lambda \mid x)}\, d\lambda) \qquad (68) \\ &\in [0,2] \end{aligned}$$

In these experiments we simulate both cases of a multinomial sampling with Dirichlet prior and independent binomial samplings with independent Beta priors. For the multinomial sampling, the simulation consider 1000 drawings of $\theta = (\theta_{ij})$, $j = 1,2$ distributed as a Dirichlet distribution of parameter $n_0\, p_0$, where p_0 is fixed to $(0.25, 0.25, 0.25, 0.25)$ and for each drawing of θ, we draw a multinomial distribution of parameter (n, θ), which constitutes the table X.

For the independent binomial sampling, the simulation consider 1000 drawings of two independent Beta distributions with the same parameter $n_0\, p_1$, where p_1 is fixed to $(0.5, 0.5)$ and for each one of them, let θ_1 and θ_2 be their values, a draw of two independent Binomial distributions of parameters (n, θ_1) and (n, θ_2) respectively, from which we constitute the table X.

For the two measures of comparison, we consider their behavior in relation to the values of two parameters: n the sample size and n_0 the relative precision of the prior information. We consider the following values of these parameters:

Multinomial Sampling	n	10, 15, 20, 25, 30, 35, 40, 45, 50
	n_0	10, 15, 20, 25, 30, 35, 40, 45, 50
Ind. Binomial Sampling	n	5, 10, 15, 20, 25, 30, 40, 50
	n_0	5, 10, 15, 20, 25, 30, 40, 50

In the multinomial sampling it is shown in Proposition 3 of the appendix that:

$$m(\lambda|x) = \frac{\lambda^{b_4-1} I_1(\lambda, x)}{\int_0^\infty \lambda^{b_4-1} I_1(\lambda, x) d\lambda} \tag{69}$$

where

$$\begin{aligned}
I_1(\lambda, x) &= \int_0^1 \delta^{b_1-1}(1-\delta)^{b_2-1}(1-\delta+\lambda\delta)^{-b_3} d\delta \\
b_1 &= x_{.1} + a_{.1} \\
b_2 &= x_{.2} + a_{.2} \\
b_3 &= x_{1.} + a_{1.} \\
b_4 &= x_{11} + a_{11} \\
(a_{11}, a_{12}, a_{21}, a_{22}) &= n_0 p_0
\end{aligned}$$

and

$$m_A(\lambda|x) = \frac{\lambda^{c_4-1} I_2(\lambda)\, p(y|z, \lambda)}{\int_0^\infty \lambda^{c_4-1} I_2(\lambda)\, p(y|z, \lambda) d\lambda} \tag{70}$$

where

$$
\begin{aligned}
p(y|z,\lambda) &= p(x_{11}|x_{1.},x_{.1},\lambda) \\
&= \frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!}\left(\sum_{x_{11}\in V}\frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!}\right)^{-1} \\
I_2(\lambda) &= \int_0^1 \delta^{c_1-1}(1-\delta)^{c_2-1}(1-\delta+\lambda\delta)^{-c_3}d\delta \\
c_1 &= a_{.1} \\
c_2 &= a_{.2} \\
c_3 &= a_{1.} \\
c_4 &= a_{11} \\
(a_{11},a_{12},a_{21},a_{22}) &= n_0p_0
\end{aligned}
$$

Furthermore, in the independent Binomial sampling, we have from the Proposition 4 of the appendix that:

$$
m(\lambda|x) = \frac{\lambda^{d_4-1}I_3(\lambda,x)}{\int_0^\infty \lambda^{d_4-1}I_3(\lambda,x)d\lambda} \tag{71}
$$

where

$$
\begin{aligned}
I_3(\lambda,x) &= \int_0^1 \beta_0^{d_1-1}(1-\beta_0)^{d_2-1}(1-\beta_0+\lambda\beta_0)^{-d_3}d\beta_0 \\
d_1 &= x_{1.}+a_{.} \\
d_2 &= x_{2.}+b_{.} \\
d_3 &= X_{.1}+a_1+b_1 \\
d_4 &= x_{11}+a_1 \\
(a_1,b_1) &= (a_2,b_2) = n_0p_1
\end{aligned}
$$

and

$$
m_A(\lambda|x) = \frac{\lambda^{e_4-1}I_4(\lambda)\,p(y|z,\lambda)}{\int_0^\infty \lambda^{e_4-1}I_4(\lambda)\,p(y|z,\lambda)d\lambda} \tag{72}
$$

where

$$
p(y|z,\lambda) = p(x_{11}|x_{1.},x_{.1},\lambda)
$$

$$
\begin{aligned}
&= \frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!}\left(\sum_{x_{11}\in V}\frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!}\right)^{-1} \\
I_4(\lambda) &= \int_0^1 \delta^{e_1-1}(1-\delta)^{e_2-1}(1-\delta+\lambda\delta)^{-e_3}d\delta \\
e_1 &= a_. \\
e_2 &= b_. \\
e_3 &= a_1+b_1 \\
e_4 &= a_1 \\
(a_1,b_1) &= (a_2,b_2) = n_0 p_1
\end{aligned}
$$

Consequently, the evaluation of the divergences (or distances) between the two distributions ((66) and (67)) require the numerical integration on the rectangle $(0,1)\times(0,\infty)$. The numerical integrations are carried out by the gaussian quadrature method applied in a first step to the interval $(0,1)$ and in a second step to the interval $(0,\infty)$. Because of the high cost of these double numerical integrations, we have considered a (n,n_0)-grid just fine enough for providing insights about the general behavior of the information loss in the (n,n_0)-space.

All the simulation have been performed with programs written in C language which call subroutines in Fortran language for the pseudo-random number generation (Ranlib package, Brown and Lovato (1991)) and for the numerical integration (Quadpack routines, Piessens et al (1983)). The figures were made in Splus 3.1. Some illustrative results along with the main conclusions are presented in next section, more results are presented in appendix B.

4.2 Results

4.2.1 Evaluating the loss of information

For both the multinomial sampling and the independent binomial sampling, the evaluation of the Hellinger distance and the Entropy divergence gave roughly similar results for the range of sample sizes and precision of prior information we have simulated. More specifically let us consider the loss of information $L(n,n_0)$, defined in (65), as a function of n, the sample size, and n_0, the precision of the posterior distribution.

Figures 1, 2, 3 and 4 (b) suggest that $L(n,n_0)$ behaves approximately as:

$$L(n,n_0) \sim h(\frac{n}{n_0}) \tag{73}$$

where $h : I\!R_+ \longrightarrow I\!R_+$ is such that $h' > 0$ or equivalently: $L(n, \alpha(c)n_0) = c$ with $\alpha : I\!R_+ \rightarrow I\!R_+$ is such that $\alpha' > 0$. This suggests that $L^{'}_{n} > 0$ and $L^{'}_{n_0} < 0$, implying in particular that the loss of information does *not* vanish asymptotically (in n). The graphs also suggest that $L^{''}_{n,n} < 0$, i.e. for any n_0, $L(n, n_0)$ is concave in n (Figures 1,2,3 and 4 (c)) and $L^{''}_{n_0,n_0} > 0$ i.e. for any n, $L(n, n_0)$ is convex in n_0 (Figures 1,2,3 and 4 (d)) and eventually tends to 0 when $n_0 \longrightarrow \infty$ (in which case the prior distribution becomes dogmatic and remains unrevised by the sample, whether the revision is "exact" or "approximate"). Thus, in this case, the approximation (12) can be qualified as a "small sample approximation" in the sense that the error of approximation tends to vanish when the prior precision tends to dominate the sample precision (more precisely $(n/n_0) \longrightarrow 0$) or when $n_0 \longrightarrow \infty$ for fixed n whereas this error does not tend to zero, but rather tends to increase when the sample size $n \longrightarrow \infty$ with fixed prior precision n_0.

4.2.2 Remarks on the simulation accuracy

Let us also mention that the graphs relative to the Hellinger distance appears to display a more regular pattern than those relative to the entropy divergence, particularly for the case of independent binomial sampling. These irregularities are due to simulation problems. These graphs are based on simulation of size $N = 1000$; previous trials with $N = 100$ revealed dramatic irregularities. In general terms, one may suspect that the multinomial sampling generate more variations than the independent binomial sampling because this one may be obtained as a conditioning (on one margin) of the first one. One may also suspect more variation for the Entropy divergence because it is not compact-valued as is the Hellinger distance.

Tables 1 to 4 in appendix B, along with figure 5 may also help getting more precise an insight on the simulation problem. Let us write D for the simulated divergence or distance. Tables 1 to 4 give the main characteristics of the simulation for each configuration of n and n_0. The mean values ("E(D)") have been plotted and commented in figures 1 to 4. The column "Gamma(D)", defined as:

$$\gamma = \frac{E(D - \bar{D})^3}{(E(D - \bar{D})^2)^{\frac{3}{2}}} \tag{74}$$

indicates a systematic positive skewness.The penultimate column gives a simulation error estimated by the ratio $\sqrt{(\frac{V(D)}{N})}$. This suggest confidence intervals of width roughly between 0.01 and 0.02, wider for the Entropy than for the

Hellinger distance. The noise to signal ratio, or coefficient of variation, is given in the last column and illustrated in figure 5 in the form of scatter diagrams between the standard deviations and averages with the main diagonal representing a ratio equal to 1. Note that there is a clear positive association between the standard deviation and the average (i.e. higher average is associated to higher dispersion). Thus one should expect less precise (and more irregular simulations) contour curves for regions of higher values of D. Furthermore, the graphs present a concave rather than linear relationship between the average and the dispersion, with a narrower and closer to the main diagonal relationship for low than for large values and for the Hellinger distance than for the Entropy divergence. Note also that the coefficient of variation is, in general, higher for the Entropy than for the Hellinger distance, particularly for the independent binomial sampling.

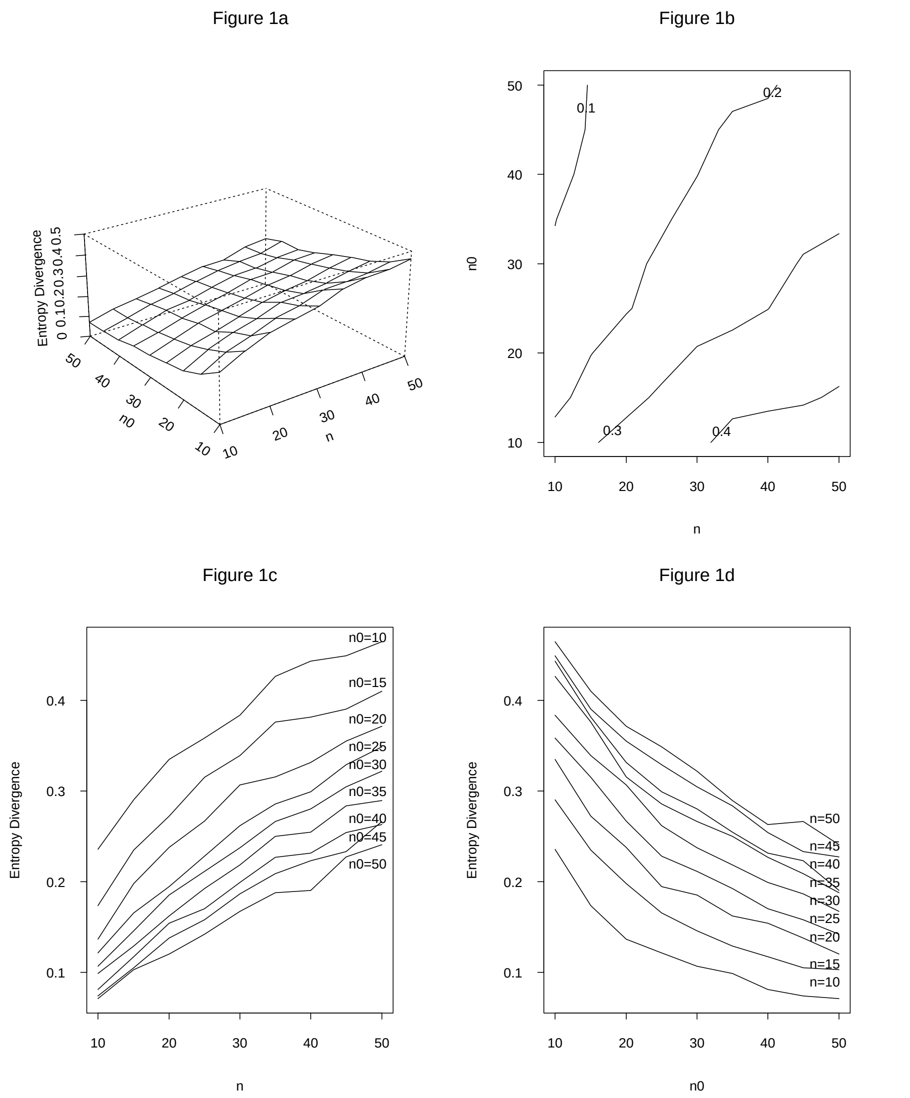

Figure 1: Expectation of the Entropy divergence in the multinomial sampling as a function of the size sample n and the a priori distribution weight n_0. Perspective (fig. 1a), contour (fig. 1b) and one-dimensional plots (fig. 1c and 1d).

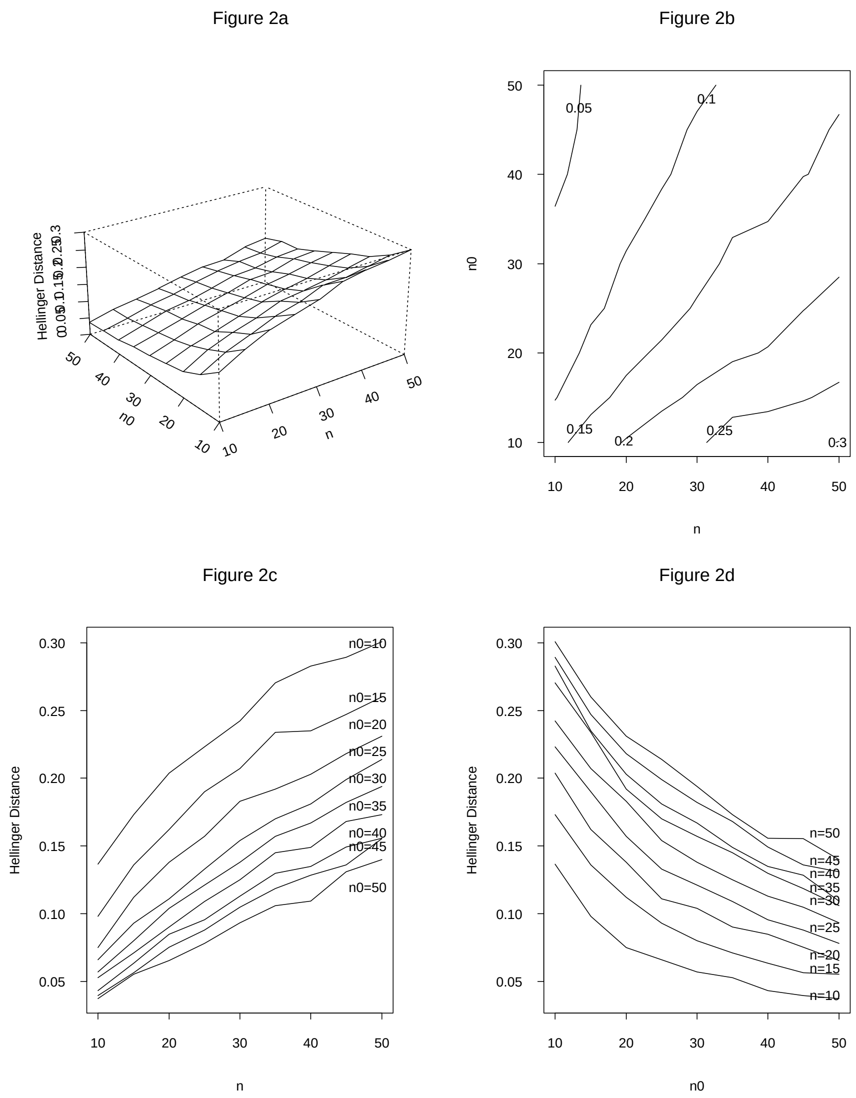

Figure 2: Expectation of the Hellinger distance in the multinomial sampling as a function of the size sample n and the a priori distribution weight n_0. Perspective (fig. 2a), contour (fig. 2b) and one-dimensional plots (fig. 2c and 2d).

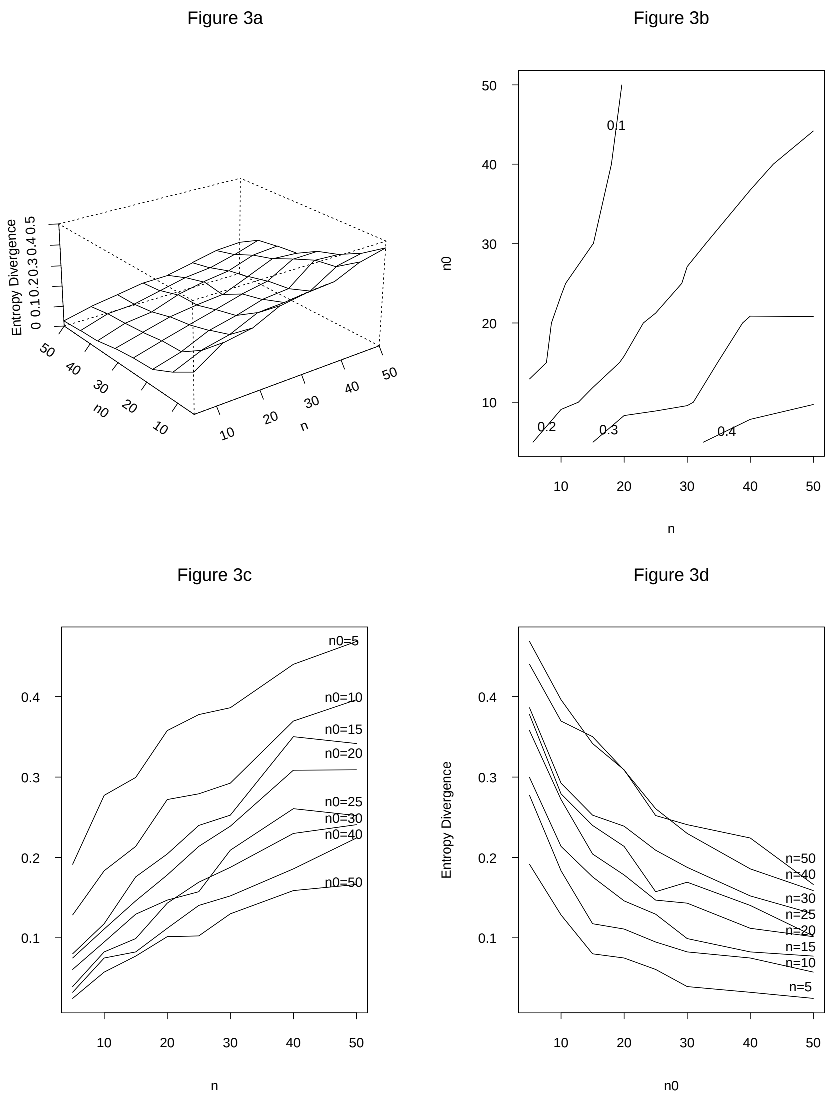

Figure 3: Expectation of the Entropy divergence in the independent binomial sampling as a function of the size sample n and the a priori distribution weight n_0. Perspective (fig. 3a), contour (fig. 3b) and one-dimensional plots (fig. 3c and 3d).

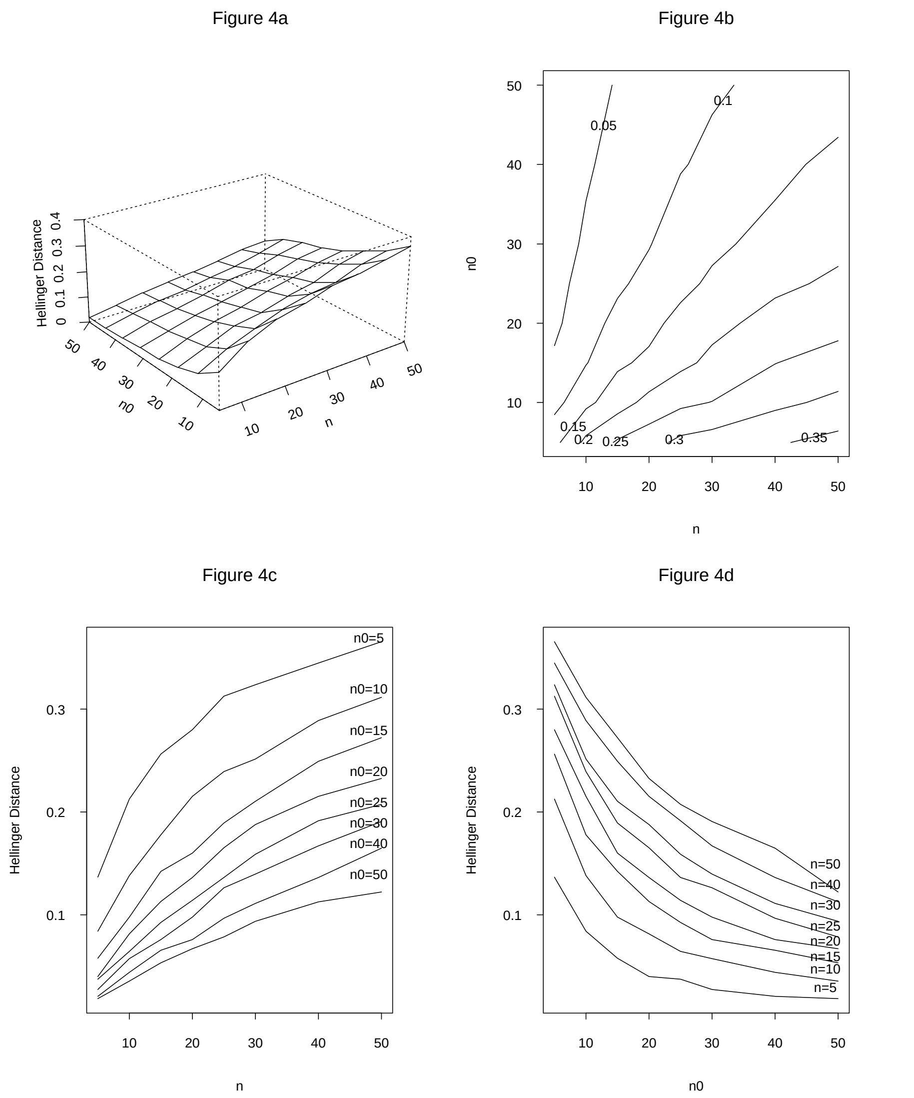

Figure 4: Expectation of the Hellinger distance in the independent binomial sampling as a function of the size sample n and the a priori distribution weight n_0. Perspective (fig. 4a), contour (fig. 4b) and unidimensional plots (fig. 4c and 4d).

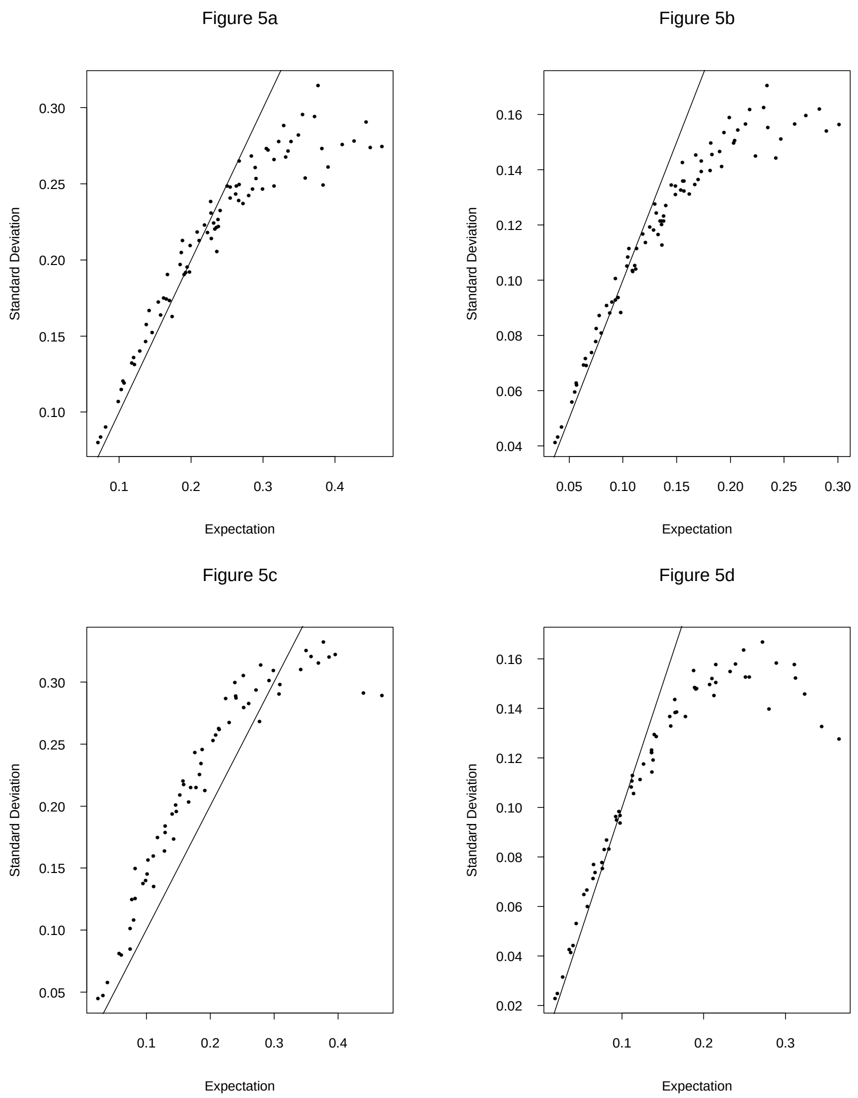

Figure 5: Expectation versus Standard Deviation of the divergence or distance D for the multinomial sampling (figures 5a and 5b) and for the independent binomial sampling (figures 5c and 5d), where D is the Entropy divergence (fig. 5a and 5c) or the Hellinger distance (fig. 5b and 5d).

References

[1] Agresti, A. (1992). A survey of exact inference for contingency tables *Statistical Science*, Vol. 7(1), 131-177.

[2] Altham, P.M.E.(1969) Exact bayesian analysis of a 2x2 contingency table and Fisher's "exact" significance test. *J. Roy. Statist. Soc. Ser.B*, Vol. 31, 261-269.

[3] Altham, P.M.E.(1971) Exact Bayesian analysis of an intraclass 2x2 table *Biometrika* Vol. 84, 679-680.

[4] Brown B.W. and J. Lovato (1991) Ranlib: Library of Fortran Routines for Random Number Generation. (bwb@odin.mda.uth.tmc.edu).

[5] Cornfield, J. (1956) A statistical problem arising from retrospectives studies. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, Vol.4, 135-148, Univ. California Press, Berkeley.

[6] Dawid, A.P.(1980). A Bayesian look at the nuisance parameters. In: *Bayesian Statistics.* See J.M. Bernardo, M.H. de Groot, D.V. Lindley, and A.F.M Smith,(1980), 167-184.

[7] Fisher, R.A. (1935a). The logic of inductive inference (with discussion). *J. Roy. Statist. Soc. Ser.A* Vol. 38, 39-82.

[8] Fisher R.A. (1935b). *The Design of Experiments* (8th ed. 1966). Oliver and Boyd, Edinburgh.

[9] Florens, J.-P. and M. Mouchart (1977). Reduction of Bayesian experiments. CORE Discussion Paper 7737, Université Catholique de Louvain, Louvain-la-Neuve, Belgium(revised July 1979).

[10] Florens, J.-P. and M. Mouchart (1985). Conditioning in dynamics models. *Journal of Time Series Analysis*, Vol. 53(1), 15-35.

[11] Florens, J.-P. and M.Mouchart (1989) Bayesian Specification Tests, in *Contributions to Operations Research and Economics* (B. Cornet and H. Tulkens, editors), Cambridge: The MIT Press, 467-490.

[12] Florens, J.-P. and M.Mouchart (1993) Bayesian Testing and Testing Bayesians, *Handbook of Statistics* Vol. 11, G.S. Maddala, C.R. Rao and H.D. Vinod (editors), Amsterdam : Elsevier Science Publishers B.V.

[13] Florens, J.-P., Mouchart, M., and J.-M. Rolin (1990). *Elements of Bayesian Statistics*, Marcel Dekker, inc., New York.

[14] Haber, M. (1989). Do the marginal totals of a 2x2 contingency table contain information regarding the table proportions? *Communications in Statistics - Theory and Methods*, Vol.18(1) 147-156.

[15] Kendall, M.G. et A. Stuart, (1967). *The Advanced Theory of Statistics*, Vol. 2, Griffin, London.

[16] Lindley, D. V. (1964). The Bayesian analysis of contingency tables. *Ann. Math. Statist.* Vol. 35, 1622-1643.

[17] Monfort A. (1982). *Cours de Statistique mathématique*, Economica, Paris.

[18] Piessens R., de Donker E. and D. Kahaner (1983) Quadpack: A package for the Numerical Computation of Definite Univariate Integrals. Appl. Math. and Prog. Div. - K.U. Leuven, Belgium.

[19] Plackett, R.L.,(1974). *The Analysis of Categorical Data*, Monograph n° 35, Griffin, London.

[20] Rolin, J.-M. (1992) Some Useful Properties of the Dirichlet Process. CORE Discussion Paper 9207.

[21] Scheihing, E. (1991) Une étude sur l'inférence dans les tables de contigence 2x2. Unpublished essay written for the DEA in mathematics, Louvain-la-Neuve, Belgium.

[22] Simar, L. (1979). Tables de Contingence à deux critères. Problèmes Particuliers des Tables (2x2). Chapter 4 in *Analyse des Données Discrètes* edited by Gérard G. et J.M. Rolin, Série Recyclage en Statistique, Vol. 3, Comité de Statistique, Louvain-la-Neuve.

[23] Tocher, K.D.(1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. *Biometrika*, Vol. 37, 130.

[24] Wilks, S.S (1962) *Mathematical Statistics.* Wiley, New York.

5 Appendix A: Proofs

Proposition 1: Consider the multinomial sampling with parameter $\theta = (\theta_{ij})$ for the 2x2 contingency table (section 3.2). Let $Z = (X_{1.}, X_{.1})$ be the total margins of the table and θ distributed a priori as a Dirichlet distribution with parameter $a = (a_{ij})$, then

$$p(Z = (x_{1.}, x_{.1}) \mid \lambda, X_{..}) = c\, X_{..}!\, \frac{E[(1 - Y + \lambda Y)^{-(a_{1.}+x_{1.})}]}{E[(1 - W + \lambda W)^{-a_{1.}}]} \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!}$$

where

$$\begin{aligned} c &= \frac{\Gamma^2(a_{..})}{\Gamma^2(a_{..} + X_{..})} \frac{\Gamma(a_{1.} + x_{1.})\Gamma(a_{2.} + x_{2.})\Gamma(a_{.1} + x_{.1})\Gamma(a_{.2} + x_{.2})}{\Gamma(a_{1.})\Gamma(a_{2.})\Gamma(a_{.1})\Gamma(a_{.2})} \\ Y &\sim \beta(a_{.1} + x_{.1}, a_{.2} + x_{.2}) \\ W &\sim \beta(a_{.1}, a_{.2}) \end{aligned}$$

and λ is the usual cross ratio (equation (36)).

Proof: Consider the one-to-one transformation of parameter space:

$$\begin{aligned} g : \Theta &\rightarrow \Gamma \\ (\theta_{22}, \theta_{21}, \theta_{12}) &\rightarrow \gamma = (\delta_1, \delta_2, \lambda) \end{aligned}$$

where

$$\begin{aligned} \Theta &= \{(\theta_{22}, \theta_{21}, \theta_{12}) \in (0,1)^3 / \theta_{22} + \theta_{21} + \theta_{12} < 1\} \\ \Gamma &= \{(\delta_1, \delta_2, \lambda) \in (0,1)^2 \times I\!R^+ / \delta_1 + \delta_2 < 1\} \\ &= \Delta \times I\!R^+ \end{aligned}$$

$$\begin{aligned} \delta_1 &= \theta_{22} \\ \delta_2 &= \frac{\theta_{21}}{\theta_{21} + \theta_{22}} \\ \lambda &= \frac{(1 - \theta_{22} - \theta_{21} - \theta_{12})\theta_{22}}{\theta_{12}\theta_{21}} \end{aligned}$$

Observe that

$$g^{-1}(\delta_1, \delta_2, \lambda) = (\delta_1, \frac{\delta_1 \delta_2}{1 - \delta_2}, \frac{1 - \delta_1 - \delta_2}{1 - \delta_2 + \lambda\delta_2}) \tag{75}$$

and

$$|Dg^{-1}| = \frac{\delta_1\delta_2(1-\delta_1-\delta_2)}{(1-\delta_2)^2(1-\delta_2+\lambda\delta_2)^2} \tag{76}$$

therefore

$$p(Z=(x_{1.},x_{.1}) \mid \lambda, X_{..}) = \int_\Delta p(Z=(x_{1.},x_{.1}) \mid \gamma, X_{..})\, dm(\gamma \mid \lambda) \tag{77}$$

it follows from the equation (40) that

$$\begin{aligned} p(Z=(x_{1.},x_{.1}) \mid \gamma, X_{..}) &= X_{..}! \sum_{x_{11}\in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \\ &\cdot \left(\frac{1-\delta_1-\delta_2}{1-\delta_2+\lambda\delta_2}\right)^{x_{1.}} \left(\frac{\delta_2}{1-\delta_2}\right)^{x_{.1}} \delta_1^{x_{2.}} \end{aligned} \tag{78}$$

Furthermore,

$$m(\gamma \mid \lambda) = \frac{m(\delta_1,\delta_2,\lambda)}{\int_\Delta m(\delta_1,\delta_2,\lambda)\, d\delta_1 d\delta_2} \tag{79}$$

where

$$m(\delta_1,\delta_2,\lambda) \propto m(g^{-1}(\delta_1,\delta_2,\lambda))|Dg^{-1}|$$

if $\theta \sim Di(a)$, equations (75) and (76) imply

$$\begin{aligned} m(\delta_1,\delta_2,\lambda) &= \frac{\delta_1\delta_2(1-\delta_1-\delta_2)}{(1-\delta_2)^2(1-\delta_2+\lambda\delta_2)^2} \left(\frac{\lambda\delta_2(1-\delta_1-\delta_2)}{(1-\delta_2)(1-\delta_2+\lambda\delta_2)}\right)^{a_{11}-1} \\ &\cdot \left(\frac{1-\delta_1-\delta_2}{1-\delta_2+\lambda\delta_2}\right)^{a_{12}-1} \left(\frac{\delta_1\delta_2}{1-\delta_2}\right)^{a_{21}-1} \delta_1^{a_{22}-1} \end{aligned} \tag{80}$$

$$\begin{aligned} &= \lambda^{a_{11}-1}\, \delta_1^{a_{2.}-1}\, (1-\delta_1-\delta_2)^{a_{1.}-1} \\ &\cdot\ \delta_2^{a_{.1}-1}\, (1-\delta_2)^{-a_{.1}}\, (1-\delta_2+\lambda\delta_2)^{-a_{1.}} \end{aligned} \tag{81}$$

which implies

$$I_1(\lambda) = \int_\Delta m(\delta_1,\delta_2,\lambda)\, d\delta_1 d\delta_2 \tag{82}$$

$$\begin{aligned} &= \lambda^{a_{11}-1} \int_0^1 \int_0^{1-\delta_2} \delta_1^{a_{2.}-1}\, ((1-\delta_2)-\delta_1)^{a_{1.}-1}\, d\delta_1 \\ &\cdot\ \delta_2^{a_{.1}-1}\, (1-\delta_2)^{-a_{.1}}\, (1-\delta_2+\lambda\delta_2)^{-a_{1.}}\, d\delta_2 \end{aligned} \tag{83}$$

$$= \lambda^{a_{11}-1} \frac{\Gamma(a_{1.})\Gamma(a_{2.})}{\Gamma(a_{..})}$$

$$\cdot \int_0^1 \delta_2^{a_{.1}-1}(1-\delta_2)^{a_{.2}-1}(1-\delta_2+\lambda\delta_2)^{-a_{1.}}\,d\delta_2 \tag{84}$$

$$= \lambda^{a_{11}-1}\frac{\Gamma(a_{1.})\Gamma(a_{2.})\Gamma(a_{.1})\Gamma(a_{.2})}{\Gamma^2(a_{..})}E[(1-W+\lambda W)^{-a_{1.}}] \tag{85}$$

where $W \sim \beta(a_{.1}, a_{.2})$.
Furthermore, by the equations (77), (78), (79) and (82),

$$p(Z=(x_{1.}, x_{.1}) \mid \lambda, X_{..}) = X_{..}! \sum_{x_{11}\in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!}\frac{I_2(\lambda)}{I_1(\lambda)} \tag{86}$$

where

$$\begin{aligned} I_2(\lambda) &= \lambda^{a_{11}-1}\int_\Delta \delta_1^{x_{2.}+a_{2.}-1}(1-\delta_1-\delta_2)^{x_{1.}+a_{1.}-1} \\ &\cdot\ \delta_2^{x_{.1}+a_{.1}-1}(1-\delta_2)^{-(x_{.1}+a_{.1})}(1-\delta_2+\lambda\delta_2)^{-(x_{1.}+a_{1.})}\,d\delta_1 d\delta_2 \end{aligned}$$

In the analogous way to (85)

$$\begin{aligned} I_2(\lambda) &= \frac{\Gamma(x_{1.}+a_{1.})\Gamma(x_{2.}+a_{2.})\Gamma(x_{.1}+a_{.1})\Gamma(x_{.2}+a_{.2})}{\Gamma^2(x_{..}+a_{..})} \\ &\cdot\ E[(1-Y+\lambda Y)^{-(x_{1.}+a_{1.})}]\,\lambda^{a_{11}-1} \end{aligned} \tag{87}$$

where $Y \sim \beta(x_{.1}+a_{.1}, x_{.2}+a_{.2})$.
Thus the result follows from the equations (85),(86) and (87).

Proposition 2: Consider the independent binomial sampling with parameter $\beta = (\beta_1, \beta_2)$ for the 2x2 contingency table (section 3.3). Let β_1 and β_2 be distributed a priori as two independent beta distribution with parameters (a_1, b_1) and (a_2, b_2) respectively, then

$$p(X_{1.} = x_{1.} \mid \lambda, X_{..}, X_{.1}) = c\,X_{.1}!X_{.2}! \sum_{x_{11}\in V}\frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!}\frac{E[(1-Y+\lambda Y)^{-(a_2+b_2+X_{.2})}]}{E[(1-W+\lambda W)^{-(a_2+b_2)}]}$$

where

$$\begin{aligned} c &= \frac{\Gamma(a_.+b_.)}{\Gamma(a_.)\Gamma(b_.)}\frac{\Gamma(a_.+x_{1.})\Gamma(b_.+x_{2.})}{\Gamma(a_.+b_.+X_{..})} \\ Y &\sim \beta(a_.+x_{1.}, b_.+x_{2.}) \\ W &\sim \beta(a_., b_.) \end{aligned}$$

and λ is the usual cross ratio (equation (56)).

Proof: Consider the one-to-one transformation of parameter space:

$$\begin{aligned} g : [0,1]^2 &\rightarrow [0,1] \times I\!R^+ \\ (\beta_2, \beta_1) &\rightarrow (\beta_0, \lambda) \end{aligned}$$

where

$$\begin{aligned} \beta_0 &= \beta_2 \\ \lambda &= \frac{\beta_1(1-\beta_2)}{\beta_2(1-\beta_1)} \end{aligned}$$

Observe that

$$g^{-1}(\beta_0, \lambda) = (\beta_0, \frac{\lambda\beta_0}{1-\beta_0+\lambda\beta_0}) \tag{88}$$

and

$$|Dg^{-1}| = \frac{\beta_0(1-\beta_0)}{(1-\beta_0+\lambda\beta_0)^2} \tag{89}$$

therefore

$$p(X_{1.} = x_{1.} \mid \lambda, X_{..}, X_{.1}) = \int_0^1 p(X_{1.} = x_{1.} \mid \beta_0, \lambda, X_{..}, X_{.1})\, dm(\beta_0, \lambda \mid \lambda) \tag{90}$$

it follows from the equation (57) that

$$\begin{aligned} p(X_{1.} = x_{1.} \mid \beta_0, \lambda, X_{..}, X_{.1}) &= X_{.1}! X_{.2}! \sum_{x_{11} \in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \left(\frac{1-\beta_0}{1-\beta_0+\lambda\beta_0} \right)^{X_{.1}} \\ &\cdot \quad \beta_0^{x_{1.}} (1-\beta_0)^{X_{..}-X_{.1}-x_{1.}} \end{aligned} \tag{91}$$

Furthermore,

$$m(\beta_0, \lambda \mid \lambda) = \frac{m(\beta_0, \lambda)}{\int_0^1 m(\beta_0, \lambda)\, d\beta_0} \tag{92}$$

where

$$m(\beta_0, \lambda) = m(g^{-1}(\beta_0, \lambda))|Dg^{-1}|$$

if $\beta_1 \sim \beta(a_1, b_1)$ and $\beta_2 \sim \beta(a_2, b_2)$, equations (88) and (89) imply

$$\begin{aligned} m(\beta_0, \lambda) &\propto \frac{\beta_0(1-\beta_0)}{(1-\beta_0+\lambda\beta_0)^2} \beta_0^{a_2-1} (1-\beta_0)^{b_2-1} \\ &\cdot \left(\frac{\lambda\beta_0}{1-\beta_0+\lambda\beta_0} \right)^{a_1-1} \left(\frac{1-\beta_0}{1-\beta_0+\lambda\beta_0} \right)^{b_1-1} \qquad (93) \\ &= \lambda^{a_1-1} \beta_0^{a_.-1} (1-\beta_0)^{b_.-1} (1-\beta_0+\lambda\beta_0)^{-(a_1+b_1)} \qquad (94) \end{aligned}$$

which implies

$$\begin{aligned} I_3(\lambda) &= \int_0^1 m(\beta_0, \lambda)\, d\beta_0 && (95)\\ &= \lambda^{a_1-1} \int_0^1 \beta_0^{a_.-1}(1-\beta_0)^{b_.-1}(1-\beta_0+\lambda\beta_0)^{-(a_1+b_1)}\, d\beta_0 && (96)\\ &= \lambda^{a_1-1} \frac{\Gamma(a_.)\Gamma(b_.)}{\Gamma(a_.+b_.)} E[(1-W+\lambda W)^{-(a_1+b_1)}] && (97) \end{aligned}$$

where $W \sim \beta(a_.,b_.)$.
Furthermore, by the equations (90), (91), (92) and (95),

$$p(X_{1.} = x_{1.} \mid \lambda, X_{..}, X_{.1}) = X_{.1}!X_{.2}! \sum_{x_{11}\in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \frac{I_4(\lambda)}{I_3(\lambda)} \qquad (98)$$

where

$$I_4(\lambda) = \lambda^{a_1-1} \int_0^1 \beta_0^{a_.+x_{1.}-1}(1-\beta_0)^{b_.+x_{2.}-1}(1-\beta_0+\lambda\beta_0)^{-(a_1+b_1+X_{.1})}\, d\beta_0 \qquad (99)$$

In the analogous way to (97)

$$I_4(\lambda) = \lambda^{a_1-1} \frac{\Gamma(a_.+x_{1.})\Gamma(b_.+x_{2.})}{\Gamma(a_.+b_.+X_{..})} E[(1-Y+\lambda Y)^{-(a_1+b_1+X_{.1})}]$$

where $Y \sim \beta(a_.+x_{1.}, b_.+x_{2.})$.
Thus the result follows from the equations (97), (98) and (100).

Proposition 3: Consider the multinomial sampling with parameter $\theta = (\theta_{ij})$ for the 2x2 contingency table $X = (X_{ij}), i,j = 1,2$(section 3.2). Let θ distributed a priori as a Dirichlet distribution with parameter $a = (a_{ij})$ and let λ be the usual cross ratio (equation (36)) then

$$m(\lambda|x) = \frac{\lambda^{b_4-1} I_1(\lambda,x)}{\int_0^\infty \lambda^{b_4-1} I_1(\lambda,x) d,\lambda}$$

where

$$\begin{aligned} I_1(\lambda,x) &= \int_0^1 \delta^{b_1-1}(1-\delta)^{b_2-1}(1-\delta+\lambda\delta)^{-b_3} d\delta\\ b_1 &= x_{.1}+a_{.1}, \quad b_2 = x_{.2}+a_{.2}\\ b_3 &= x_{1.}+a_{1.}, \quad b_4 = x_{11}+a_{11} \end{aligned}$$

and

$$m_A(\lambda|x) = \frac{\lambda^{c_4-1} I_2(\lambda, x)\, p(y|z,\lambda)}{\int_0^\infty \lambda^{c_4-1} I_2(\lambda, x)\, p(y|z,\lambda) d\lambda}$$

where

$$\begin{aligned}
p(y|z,\lambda) &= p(x_{11}|x_{1.}, x_{.1}, \lambda) \\
&= \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \left(\sum_{x_{11}\in V} \frac{\lambda^{x_{11}}}{\prod_{i,j} x_{ij}!} \right)^{-1} \\
I_2(\lambda, x) &= \int_0^1 \delta^{c_1-1}(1-\delta)^{c_2-1}(1-\delta+\lambda\delta)^{-c_3} d\delta \\
c_1 &= a_{.1}, \quad c_2 = a_{.2} \\
c_3 &= a_{1.}, \quad c_4 = a_{11} \\
(a_{11}, a_{12}, a_{21}, a_{22}) &= n_0 p_0
\end{aligned}$$

Proof: Consider the one-to-one transformation of parameter space defined in the proof of the proposition 1:

$$\begin{aligned}
g : \Theta &\rightarrow \Gamma \\
(\theta_{22}, \theta_{21}, \theta_{12}) &\rightarrow \gamma = (\delta_1, \delta_2, \lambda)
\end{aligned}$$

where

$$\begin{aligned}
\Theta &= \{(\theta_{22}, \theta_{21}, \theta_{12}) \in (0,1)^3 / \theta_{22} + \theta_{21} + \theta_{12} < 1\} \\
\Gamma &= \{(\delta_1, \delta_2, \lambda) \in (0,1)^2 \times I\!R^+ / \delta_1 + \delta_2 < 1\} \\
&= \Delta \times I\!R^+
\end{aligned}$$

$$\begin{aligned}
\delta_1 &= \theta_{22} \\
\delta_2 &= \frac{\theta_{21}}{\theta_{21} + \theta_{22}} \\
\lambda &= \frac{(1 - \theta_{22} - \theta_{21} - \theta_{12})\theta_{22}}{\theta_{12}\theta_{21}}
\end{aligned}$$

Then from (8) we can write:

$$m(\lambda \mid x) \propto \int_\Delta p(x \mid \gamma, X_{..}) m(\gamma) d(\delta_1, \delta_2) \qquad (100)$$

Furthermore, from (35) and (75) we have:

$$
\begin{aligned}
p(x \mid \gamma, X_{..}) &= \frac{X_{..}!}{\prod\limits_{i,j} x_{ij}!} \lambda^{x_{11}} \delta_1^{x_{2.}} \delta_2^{x_{.1}} (1-\delta_1-\delta_2)^{x_{1.}} (1-\delta_2)^{-x_{.1}} \\
&\cdot \quad (1-\delta_2+\lambda\delta_2)^{-x_{1.}} \qquad (101)
\end{aligned}
$$

it follows from (81) and (101) that:

$$
\begin{aligned}
m(\lambda \mid x) &\propto \lambda^{x_{11}+a_{11}} \int_\Delta \delta_1^{x_{2.}+a_{2.}-1} \delta_2^{x_{.1}+a_{1.}-1} (1-\delta_1-\delta_2)^{x_{1.}+a_{1.}-1} \\
&. \quad (1-\delta_2)^{-(x_{.1}+a_{.1})} (1-\delta_2+\lambda\delta_2)^{-(x_{1.}+a_{1.})} d(\delta_1,\delta_2) \qquad (102) \\
&\propto \lambda^{x_{11}+a_{11}} \int_0^1 \delta_2^{x_{.1}+a_{1.}-1} (1-\delta_2)^{x_{.2}+a_{.2}-1} \\
&. \quad (1-\delta_2+\lambda\delta_2)^{-(x_{1.}+a_{1.})} d\delta_2 \qquad (103)
\end{aligned}
$$

Thus the first result is obtained. For the second one, we have from (12) that

$$m_A(\lambda \mid x) \propto m(\lambda) p(y \mid z, \lambda)$$

Furthermore

$$m(\lambda) = \int_\Delta m(\delta_1, \delta_2, \lambda) d(\delta_1, \delta_2) \qquad (104)$$

In analogous way to (103) we obtain

$$m(\lambda) \propto \lambda^{a_{11}} \int_0^1 \delta_2^{a_{1.}-1} (1-\delta_2)^{a_{.2}-1} (1-\delta_2+\lambda\delta_2)^{a_{1.}} d\delta_2 \qquad (105)$$

Thus the result is obtained.

Proposition 4: Consider the independent binomial sampling with parameter $\beta = (\beta_1, \beta_2)$ for the 2x2 contingency table (section 3.3). Let β_1 and β_2 be distributed a priori as two independent beta distribution with parameters (a_1, b_1) and (a_2, b_2) respectively and let λ be the usual cross ratio (equation (56)) then

$$m(\lambda | x) = \frac{\lambda^{d_4-1} I_3(\lambda, x)}{\int_0^\infty \lambda^{d_4-1} I_3(\lambda, x) d\lambda}$$

where

$$
\begin{aligned}
I_3(\lambda, x) &= \int_0^1 \beta_0^{d_1-1} (1-\beta_0)^{d_2-1} (1-\beta_0+\lambda\beta_0)^{-b_3} d\beta_0 \\
d_1 &= x_{1.} + a_{.}, \quad d_2 = x_{2.} + b_{.} \\
d_3 &= X_{.1} + a_1 + b_1, \quad d_4 = x_{11} + a_1
\end{aligned}
$$

and

$$m_A(\lambda|x) = \frac{\lambda^{e_4-1} I_4(\lambda,x)\, p(y|z,\lambda)}{\int_0^\infty \lambda^{e_4-1} I_4(\lambda,x)\, p(y|z,\lambda) d\lambda}$$

where

$$\begin{aligned}
p(y|z,\lambda) &= p(x_{11}|x_{1.},x_{.1},\lambda) = \frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!} \left(\sum_{x_{11}\in V} \frac{\lambda^{x_{11}}}{\prod\limits_{i,j} x_{ij}!} \right)^{-1} \\
I_4(\lambda,x) &= \int_0^1 \delta^{e_1-1}(1-\delta)^{e_2-1}(1-\delta+\lambda\delta)^{-e_3} d\delta \\
e_1 &= a_{.}, \quad e_2 = b_{.} \\
e_3 &= a_1 + b_1, \quad e_4 = a_1
\end{aligned}$$

Proof: Consider the one-to-one transformation of parameter space defined in the proof of the proposition 2:

$$\begin{aligned}
g : [0,1]^2 &\rightarrow [0,1] \times \mathbb{R}^+ \\
(\beta_2, \beta_1) &\rightarrow (\beta_0, \lambda)
\end{aligned}$$

where

$$\begin{aligned}
\beta_0 &= \beta_2 \\
\lambda &= \frac{\beta_1(1-\beta_2)}{\beta_2(1-\beta_1)}
\end{aligned}$$

Then from (8) we can write:

$$m(\lambda \mid x) \propto \int_0^1 p(x \mid \beta_0, \lambda, X_{..}, X_{.1}) m(\beta_0,\lambda) d\beta_0 \tag{106}$$

Furthermore, from (55) and (88) we have:

$$p(x \mid \beta_0, \lambda, X_{..}, X_{.1}) = \frac{X_{.1}! X_{.2}!}{\prod\limits_{i,j} x_{ij}!} \lambda^{x_{11}} \beta_0^{x_{1.}} (1-\beta_0)^{x_{2.}} (1-\beta_0+\lambda\beta_0)^{X_{1.}} \tag{107}$$

The results follow from (94) and (107).

6 Appendix B: Further Numerical Results

Table 1: Entropy Divergence for the Multinomial Sampling

n	n0	Min	Max	E(D)	V(D)	Gamma(D)	Sim.E.	D.Coeff.
10	10	0.03882	1.11658	0.23543	0.04232	1.86719	0.00650	0.87389
15	10	0.05256	1.62332	0.29042	0.06428	2.31587	0.00802	0.87307
20	10	0.07857	2.03331	0.33471	0.07389	2.44357	0.00860	0.81216
25	10	0.07336	1.86783	0.35845	0.06454	2.32254	0.00803	0.70884
30	10	0.10641	1.81620	0.38330	0.06209	2.32287	0.00788	0.65008
35	10	0.10588	2.02086	0.42648	0.07732	2.16384	0.00879	0.65199
40	10	0.10070	3.08560	0.44298	0.08466	3.04028	0.00920	0.65681
45	10	0.08432	2.96753	0.44909	0.07494	3.11539	0.00866	0.60954
50	10	0.12171	3.40366	0.46471	0.07544	3.04294	0.00869	0.59106
10	15	0.02726	1.01012	0.17339	0.02657	1.78970	0.00515	0.94001
15	15	0.04145	1.60685	0.23492	0.04892	2.24831	0.00699	0.94155
20	15	0.05698	2.13187	0.27216	0.05628	2.35402	0.00750	0.87154
25	15	0.07310	2.09344	0.31524	0.07080	2.45951	0.00841	0.84417
30	15	0.10014	2.38897	0.33888	0.07712	2.46251	0.00878	0.81945
35	15	0.08975	2.44115	0.37613	0.09918	2.73256	0.00996	0.83736
40	15	0.10323	2.28049	0.38155	0.07476	2.49041	0.00865	0.71669
45	15	0.09900	2.14344	0.39004	0.06828	2.35951	0.00826	0.67003
50	15	0.10656	2.30952	0.40984	0.07624	2.72154	0.00873	0.67378
10	20	0.02013	0.91479	0.13655	0.02146	2.22331	0.00463	1.07308
15	20	0.03368	1.12427	0.19778	0.03693	2.01811	0.00608	0.97159
20	20	0.04702	1.66070	0.23790	0.04931	2.19342	0.00702	0.93337
25	20	0.05267	2.17037	0.26660	0.06230	2.57687	0.00789	0.93623
30	20	0.06800	2.58077	0.30677	0.07419	3.15254	0.00861	0.88783
35	20	0.07460	1.80831	0.31532	0.06172	2.11355	0.00786	0.78795
40	20	0.07369	2.13200	0.33141	0.07168	2.63806	0.00847	0.80789
45	20	0.10843	2.36202	0.35494	0.08741	2.94072	0.00935	0.83305
50	20	0.14446	2.80820	0.37129	0.08664	2.94413	0.00931	0.79277

10	25	0.01546	0.82560	0.12143	0.01723	2.11174	0.00415	1.08112
15	25	0.02745	1.06626	0.16566	0.03044	2.38587	0.00552	1.05360
20	25	0.03790	1.62721	0.19442	0.03819	2.56720	0.00618	1.00529
25	25	0.05386	2.04708	0.22803	0.04586	2.65631	0.00677	0.93923
30	25	0.06870	2.27344	0.26156	0.05922	2.60464	0.00770	0.93028
35	25	0.07498	1.99169	0.28553	0.06075	2.43350	0.00779	0.86331
40	25	0.08763	1.54029	0.29925	0.06096	2.14412	0.00781	0.82517
45	25	0.09662	1.95112	0.32887	0.08326	2.55804	0.00912	0.87731
50	25	0.10758	2.16320	0.34884	0.07974	2.61776	0.00893	0.80958
10	30	0.01224	0.74641	0.10668	0.01420	2.20253	0.00377	1.11669
15	30	0.02239	0.99844	0.14558	0.02323	2.20848	0.00482	1.04689
20	30	0.03141	1.56769	0.18509	0.03885	2.34151	0.00623	1.06481
25	30	0.03946	1.76353	0.21112	0.04532	2.70968	0.00673	1.00847
30	30	0.05342	1.88692	0.23707	0.05136	2.46722	0.00717	0.95581
35	30	0.07293	1.80787	0.26631	0.05711	2.12999	0.00756	0.89742
40	30	0.07441	1.89949	0.28016	0.05864	2.32349	0.00766	0.86425
45	30	0.09008	2.25708	0.30454	0.07474	2.57739	0.00864	0.89783
50	30	0.10178	2.58655	0.32169	0.07722	2.52395	0.00879	0.86383
10	35	0.00992	0.68628	0.09880	0.01149	1.88012	0.00339	1.08484
15	35	0.01844	0.94215	0.12891	0.01962	2.33483	0.00443	1.08675
20	35	0.02730	1.50168	0.16205	0.03068	2.79090	0.00554	1.08116
25	35	0.03457	1.68280	0.19255	0.03682	2.22214	0.00607	0.99685
30	35	0.04111	1.84492	0.21854	0.04970	2.63724	0.00705	1.02026
35	35	0.05276	1.75209	0.24979	0.06176	2.31045	0.00786	0.99482
40	35	0.05937	2.16320	0.25450	0.06157	2.81156	0.00785	0.97499
45	35	0.06506	2.32328	0.28361	0.07204	2.63873	0.00849	0.94644
50	35	0.08565	2.38022	0.28926	0.06798	2.67558	0.00825	0.90126

10	40	0.00821	0.62948	0.08115	0.00813	2.13392	0.00285	1.11200
15	40	0.01569	1.19348	0.11728	0.01755	2.44749	0.00419	1.12928
20	40	0.02282	1.13055	0.15422	0.02972	2.28367	0.00545	1.11805
25	40	0.03146	1.05765	0.16993	0.03010	2.05160	0.00549	1.02108
30	40	0.03832	1.84795	0.19904	0.04387	2.64478	0.00662	1.05256
35	40	0.04573	1.74031	0.22690	0.05695	2.51588	0.00755	1.05173
40	40	0.05581	2.16078	0.23157	0.05028	2.83415	0.00709	0.96823
45	40	0.06132	1.98368	0.25409	0.05792	2.64997	0.00761	0.94709
50	40	0.07504	2.31741	0.26298	0.06193	2.76199	0.00787	0.94656
10	45	0.00690	0.58335	0.07492	0.00697	2.14822	0.00264	1.12844
15	45	0.01370	0.83659	0.10501	0.01453	2.49344	0.00381	1.14789
20	45	0.01947	1.08089	0.13763	0.02481	2.44381	0.00498	1.14478
25	45	0.02555	1.01124	0.15783	0.02689	2.02832	0.00518	1.03914
30	45	0.03031	1.51385	0.18661	0.04200	2.62980	0.00648	1.09821
35	45	0.04033	1.46385	0.20855	0.04776	2.52807	0.00691	1.04810
40	45	0.05032	1.83925	0.22302	0.04765	2.49290	0.00690	0.97886
45	45	0.05782	1.69289	0.23318	0.04854	2.25157	0.00697	0.94513
50	45	0.06645	1.94704	0.26644	0.07020	2.55808	0.00838	0.99458
10	50	0.00592	0.54342	0.07072	0.00642	2.01478	0.00253	1.13304
15	50	0.01221	0.79113	0.10292	0.01317	2.34264	0.00363	1.11518
20	50	0.01789	0.79385	0.11990	0.01847	2.22239	0.00430	1.13339
25	50	0.02384	1.52548	0.14182	0.02783	2.96170	0.00528	1.17643
30	50	0.03261	1.47364	0.16733	0.03625	2.65482	0.00602	1.13809
35	50	0.03618	1.93243	0.18788	0.04535	3.14820	0.00673	1.13329
40	50	0.04343	1.58856	0.19031	0.03625	2.74663	0.00602	1.00044
45	50	0.05333	1.77418	0.22744	0.05330	2.55191	0.00730	1.01520
50	50	0.06010	1.93625	0.24047	0.05405	2.35858	0.00735	0.96672

Table 2: Hellinger Distance for the Multinomial Sampling

n	n0	Min	Max	E(D)	V(D)	Gamma(D)	Sim.E.	D.Coeff.
10	10	0.02146	0.58454	0.13655	0.01274	1.61683	0.00357	0.82690
15	10	0.02943	0.81635	0.17308	0.01940	1.94000	0.00441	0.80475
20	10	0.04570	0.98198	0.20388	0.02271	1.94904	0.00477	0.73914
25	10	0.04172	0.94801	0.22326	0.02104	1.85279	0.00459	0.64957
30	10	0.06347	0.94573	0.24238	0.02082	1.86580	0.00456	0.59533
35	10	0.06312	1.02625	0.27044	0.02547	1.75038	0.00505	0.59025
40	10	0.05842	1.32218	0.28282	0.02627	2.03965	0.00513	0.57310
45	10	0.04818	1.27850	0.28945	0.02376	2.11136	0.00487	0.53264
50	10	0.07358	1.40329	0.30089	0.02448	1.97877	0.00495	0.51998
10	15	0.01496	0.51806	0.09788	0.00783	1.62562	0.00280	0.90276
15	15	0.02316	0.78515	0.13601	0.01443	1.94733	0.00380	0.88336
20	15	0.03250	0.98744	0.16175	0.01726	1.91800	0.00415	0.81085
25	15	0.04240	0.99755	0.18999	0.02150	1.97688	0.00464	0.77168
30	15	0.05958	1.09864	0.20696	0.02388	1.97346	0.00489	0.74654
35	15	0.05347	1.20956	0.23396	0.02913	2.10555	0.00540	0.72936
40	15	0.06225	1.10196	0.23541	0.02412	1.96604	0.00491	0.66082
45	15	0.05872	1.06393	0.24680	0.02285	1.85466	0.00478	0.61199
50	15	0.06464	1.11872	0.26027	0.02451	2.11169	0.00495	0.60218
10	20	0.01095	0.46502	0.07543	0.00606	2.05235	0.00246	1.03752
15	20	0.01871	0.57778	0.11231	0.01082	1.80149	0.00329	0.92879
20	20	0.02668	0.80933	0.13805	0.01472	1.88863	0.00384	0.87932
25	20	0.03010	0.99990	0.15717	0.01850	2.13616	0.00430	0.86634
30	20	0.03928	1.13340	0.18330	0.02118	2.25155	0.00460	0.79525
35	20	0.04398	0.90891	0.19205	0.01994	1.76646	0.00447	0.73543
40	20	0.04299	1.02960	0.20338	0.02242	2.08598	0.00474	0.73765
45	20	0.06533	1.10999	0.21840	0.02620	2.26406	0.00512	0.74244
50	20	0.08941	1.23553	0.23063	0.02645	2.20821	0.00514	0.70406

10	25	0.00834	0.41788	0.06610	0.00477	1.95104	0.00218	1.04644
15	25	0.01517	0.54192	0.09254	0.00862	2.14447	0.00294	0.99855
20	25	0.02124	0.78136	0.11114	0.01111	2.23759	0.00333	0.94963
25	25	0.03089	0.93708	0.13278	0.01359	2.10749	0.00369	0.87658
30	25	0.04026	1.02987	0.15404	0.01761	2.10827	0.00420	0.86161
35	25	0.04398	0.95894	0.17070	0.01866	1.97989	0.00432	0.80347
40	25	0.05190	0.80829	0.18109	0.01952	1.85499	0.00442	0.77186
45	25	0.05763	0.96670	0.19919	0.02525	2.10445	0.00502	0.79844
50	25	0.06471	1.05096	0.21376	0.02453	2.06317	0.00495	0.73186
10	30	0.00655	0.37690	0.05742	0.00386	2.05604	0.00197	1.09026
15	30	0.01226	0.50428	0.08045	0.00657	2.01033	0.00256	1.01304
20	30	0.01743	0.74651	0.10392	0.01105	2.05150	0.00332	1.01076
25	30	0.02222	0.83388	0.12076	0.01294	2.26690	0.00360	0.94004
30	30	0.03061	0.89405	0.13777	0.01520	2.05254	0.00390	0.89345
35	30	0.04270	0.87607	0.15686	0.01754	1.78197	0.00419	0.84346
40	30	0.04359	0.92411	0.16711	0.01813	1.95723	0.00426	0.80634
45	30	0.05361	1.04971	0.18231	0.02245	2.05736	0.00474	0.82326
50	30	0.06137	1.15986	0.19429	0.02359	1.99939	0.00486	0.79177
10	35	0.00528	0.34610	0.05276	0.00312	1.77216	0.00177	1.05790
15	35	0.01002	0.47359	0.07050	0.00546	2.14207	0.00234	1.04035
20	35	0.01513	0.71152	0.09016	0.00850	2.40180	0.00292	1.02433
25	35	0.01932	0.79243	0.10911	0.01074	1.91193	0.00328	0.95086
30	35	0.02322	0.86592	0.12519	0.01423	2.21094	0.00377	0.95445
35	35	0.03027	0.84807	0.14452	0.01811	1.97479	0.00426	0.92802
40	35	0.03433	0.99875	0.14919	0.01796	2.24646	0.00424	0.89955
45	35	0.03777	1.05637	0.16749	0.02114	2.13097	0.00460	0.86543
50	35	0.05081	1.08739	0.17294	0.02052	2.14116	0.00453	0.82808

10	40	0.00435	0.31720	0.04314	0.00220	2.01216	0.00148	1.08752
15	40	0.00847	0.56943	0.06354	0.00480	2.20834	0.00219	1.09117
20	40	0.01254	0.55997	0.08482	0.00828	2.07959	0.00288	1.07292
25	40	0.01750	0.54277	0.09547	0.00880	1.88213	0.00297	0.98245
30	40	0.02150	0.85381	0.11291	0.01247	2.21480	0.00353	0.98922
35	40	0.02608	0.83285	0.12981	0.01629	2.17572	0.00404	0.98327
40	40	0.03233	0.98539	0.13482	0.01473	2.26274	0.00384	0.90023
45	40	0.03577	0.93795	0.14914	0.01716	2.19082	0.00414	0.87866
50	40	0.04416	1.04745	0.15560	0.01847	2.21832	0.00430	0.87352
10	45	0.00364	0.29382	0.03960	0.00187	2.03749	0.00137	1.09142
15	45	0.00737	0.41835	0.05655	0.00395	2.31217	0.00199	1.11251
20	45	0.01064	0.53322	0.07518	0.00683	2.23880	0.00261	1.09915
25	45	0.01409	0.51657	0.08791	0.00777	1.88338	0.00279	1.00307
30	45	0.01687	0.72597	0.10473	0.01175	2.29851	0.00343	1.03545
35	45	0.02276	0.72166	0.11854	0.01364	2.19036	0.00369	0.98557
40	45	0.02887	0.86998	0.12854	0.01397	2.10151	0.00374	0.91990
45	45	0.03335	0.82797	0.13599	0.01477	1.94252	0.00384	0.89368
50	45	0.03879	0.92546	0.15531	0.02035	2.13888	0.00451	0.91857
10	50	0.00312	0.27363	0.03717	0.00171	1.92730	0.00131	1.11363
15	50	0.00654	0.39496	0.05509	0.00356	2.17541	0.00189	1.08256
20	50	0.00974	0.40575	0.06528	0.00515	2.08553	0.00227	1.09930
25	50	0.01314	0.71229	0.07812	0.00762	2.58629	0.00276	1.11778
30	50	0.01816	0.70677	0.09328	0.01013	2.33044	0.00318	1.07892
35	50	0.02030	0.87972	0.10576	0.01242	2.63862	0.00352	1.05353
40	50	0.02466	0.77297	0.10924	0.01066	2.33595	0.00327	0.94558
45	50	0.03067	0.84975	0.13087	0.01545	2.17754	0.00393	0.94950
50	50	0.03476	0.91209	0.13986	0.01616	1.97787	0.00402	0.90872

Table 3: Entropy Divergence for the Independent Binomial Sampling

n	n0	Min	Max	E(D)	V(D)	Gamma(D)	Sim.E.	D.Coeff.
5	5	0.03882	1.11658	0.19155	0.04511	2.04732	0.00672	1.10853
10	5	0.05697	2.03640	0.27719	0.07193	1.78860	0.00848	0.96752
15	5	0.06599	1.81620	0.29932	0.09583	1.83040	0.00979	1.03428
20	5	0.13229	1.97523	0.35779	0.10276	1.74861	0.01014	0.89592
25	5	0.17843	2.56440	0.37748	0.11060	1.64947	0.01052	0.88099
30	5	0.18338	2.38126	0.38630	0.10243	1.27345	0.01012	0.82848
40	5	0.25924	2.49616	0.44000	0.08465	1.44272	0.00920	0.66124
50	5	0.29146	2.74809	0.46859	0.08366	1.05970	0.00915	0.61725
5	10	0.02013	0.91479	0.12859	0.02687	2.10357	0.00518	1.27458
10	10	0.03744	1.31204	0.18360	0.05081	2.10506	0.00713	1.22770
15	10	0.06932	1.82855	0.21389	0.06856	2.11085	0.00828	1.22417
20	10	0.11529	2.21683	0.27183	0.08634	2.07971	0.00929	1.08108
25	10	0.15184	2.24398	0.27891	0.09843	2.16340	0.00992	1.12493
30	10	0.17488	2.20518	0.29228	0.09081	2.02519	0.00953	1.03094
40	10	0.21473	2.43288	0.36957	0.09932	1.85149	0.00997	0.85269
50	10	0.24870	2.87660	0.39607	0.10385	1.78558	0.01019	0.81356
5	15	0.01224	0.75007	0.08057	0.01176	2.36700	0.00343	1.34718
10	15	0.03823	1.56768	0.11748	0.03056	2.90736	0.00553	1.48783
15	15	0.06206	2.27746	0.17590	0.05910	3.27917	0.00769	1.38210
20	15	0.07740	1.97938	0.20444	0.06397	2.52389	0.00800	1.23740
25	15	0.11407	1.80482	0.24004	0.08351	2.31656	0.00914	1.20408
30	15	0.12897	2.21983	0.25267	0.07821	2.20676	0.00884	1.10669
40	15	0.16237	3.12996	0.35013	0.10597	2.51153	0.01029	0.92982
50	15	0.19094	2.47539	0.34155	0.09608	2.13820	0.00980	0.90740
5	20	0.00821	0.63208	0.07508	0.00725	2.40484	0.00269	1.13370
10	20	0.02282	1.43462	0.11122	0.02559	2.97433	0.00506	1.43851
15	20	0.03996	1.84366	0.14586	0.04049	3.02849	0.00636	1.37910
20	20	0.06076	1.87408	0.17818	0.04637	2.83854	0.00681	1.20837
25	20	0.07689	2.16282	0.21376	0.06895	2.34196	0.00830	1.22815
30	20	0.08197	2.46983	0.23899	0.08979	2.55555	0.00948	1.25378

40	20	0.13371	2.29134	0.30828	0.08437	2.15048	0.00918	0.94215
50	20	0.16001	2.37305	0.30906	0.08886	2.00697	0.00943	0.96439
5	25	0.00589	0.54542	0.06096	0.00641	2.16266	0.00253	1.31240
10	25	0.01683	1.30664	0.09501	0.01888	3.22499	0.00435	1.44652
15	25	0.03261	1.77252	0.12965	0.03195	3.23835	0.00565	1.37915
20	25	0.04375	1.35481	0.14722	0.03822	2.41549	0.00618	1.32819
25	25	0.06238	1.69455	0.15766	0.04851	2.68292	0.00696	1.39662
30	25	0.06908	2.01020	0.20903	0.06622	2.42604	0.00814	1.23128
40	25	0.11014	1.71432	0.26062	0.07990	1.92745	0.00894	1.08469
50	25	0.13493	2.44972	0.25199	0.09311	2.28382	0.00965	1.21085
5	30	0.00442	0.47954	0.03966	0.00333	2.63910	0.00182	1.45290
10	30	0.01398	1.19226	0.08277	0.01579	3.39414	0.00397	1.51752
15	30	0.02375	1.14433	0.09922	0.01958	2.80986	0.00442	1.41068
20	30	0.03495	1.31552	0.14331	0.03010	2.26477	0.00549	1.21080
25	30	0.04836	1.68722	0.16953	0.04624	2.50004	0.00680	1.26864
30	30	0.06179	1.98296	0.18766	0.06030	2.63045	0.00776	1.30827
40	30	0.08399	2.60042	0.22965	0.07154	2.37095	0.00846	1.16441
50	30	0.11120	1.78723	0.24075	0.08265	2.08420	0.00909	1.19388
5	40	0.00259	0.38637	0.03235	0.00224	2.33057	0.00150	1.45913
10	40	0.00882	1.00590	0.07497	0.01028	3.01904	0.00321	1.35167
15	40	0.01662	1.23874	0.08267	0.02244	3.17552	0.00474	1.81157
20	40	0.02556	1.03101	0.11195	0.01836	2.31369	0.00428	1.20998
25	40	0.03422	1.43698	0.14058	0.03755	2.47933	0.00613	1.37828
30	40	0.04550	1.80534	0.15239	0.04369	2.86688	0.00661	1.37155
40	40	0.06092	2.02967	0.18579	0.05507	2.53534	0.00742	1.26298
50	40	0.08463	2.27720	0.22425	0.08215	2.65089	0.00906	1.27838
5	50	0.00169	0.32376	0.02470	0.00199	2.94358	0.00141	1.80832
10	50	0.00608	0.86538	0.05747	0.00664	3.07698	0.00258	1.41694
15	50	0.01181	1.10352	0.07767	0.01562	2.99427	0.00395	1.60834
20	50	0.01809	1.32613	0.10174	0.02111	2.97664	0.00459	1.42857
25	50	0.02501	1.14828	0.10272	0.02453	2.68232	0.00495	1.52506
30	50	0.03379	1.52314	0.12995	0.03394	2.79833	0.00583	1.41708
40	50	0.04860	1.67236	0.15869	0.04721	2.46629	0.00687	1.36917
50	50	0.06660	1.36681	0.16639	0.04140	2.07568	0.00643	1.22278

Table 4: Hellinger Distance for the Independent Binomial Sampling

n	n0	Min	Max	E(D)	V(D)	Gamma(D)	Sim.E.	D.Coeff.
5	5	0.02146	0.58454	0.13649	0.01308	1.76984	0.00362	0.83770
10	5	0.03193	0.98322	0.21252	0.02111	1.66603	0.00459	0.68367
15	5	0.03718	0.94573	0.25605	0.02328	1.94219	0.00482	0.59601
20	5	0.08054	1.02116	0.27996	0.01951	1.86675	0.00442	0.49888
25	5	0.11094	1.21586	0.31254	0.02320	2.05699	0.00482	0.48740
30	5	0.11414	1.18153	0.32359	0.02129	1.98845	0.00461	0.45088
40	5	0.16703	1.22988	0.34470	0.01765	2.37851	0.00420	0.38546
50	5	0.19034	1.30688	0.36546	0.01632	2.18902	0.00404	0.34947
5	10	0.01095	0.46502	0.08413	0.00693	1.90346	0.00263	0.98971
10	10	0.02088	0.67551	0.13812	0.01424	1.89148	0.00377	0.86397
15	10	0.04030	0.90748	0.17777	0.01872	1.94102	0.00433	0.76952
20	10	0.06966	1.06370	0.21485	0.02266	1.85722	0.00476	0.70048
25	10	0.09413	1.09283	0.23910	0.02497	1.98342	0.00500	0.66088
30	10	0.10981	1.09029	0.25129	0.02328	2.01527	0.00483	0.60722
40	10	0.13751	1.18842	0.28874	0.02512	2.11733	0.00501	0.54902
50	10	0.16143	1.32346	0.31123	0.02487	2.00946	0.00499	0.50671
5	15	0.00655	0.37878	0.05764	0.00361	1.96390	0.00190	1.04282
10	15	0.02154	0.74651	0.09771	0.00938	2.27958	0.00306	0.99109
15	15	0.03586	1.01562	0.14210	0.01656	2.32693	0.00407	0.90571
20	15	0.04549	0.94553	0.15967	0.01768	2.19471	0.00421	0.83272
25	15	0.06136	1.10287	0.18908	0.02206	2.24658	0.00470	0.78553
30	15	0.06468	1.06983	0.21031	0.02315	1.98100	0.00481	0.72354
40	15	0.09788	1.48364	0.24895	0.02679	2.15415	0.00518	0.65735
50	15	0.09558	1.10943	0.27206	0.02789	2.11870	0.00528	0.61373
5	20	0.00435	0.31851	0.03997	0.00196	2.27134	0.00140	1.10736
10	20	0.01254	0.67783	0.08132	0.00757	2.43880	0.00275	1.07011
15	20	0.02256	0.85543	0.11271	0.01274	2.24247	0.00357	1.00168
20	20	0.03516	0.89302	0.13634	0.01495	2.18331	0.00387	0.89719
25	20	0.04528	1.00912	0.16487	0.01917	1.94517	0.00438	0.83953
30	20	0.04843	1.11665	0.18774	0.02416	2.09917	0.00492	0.82814

40	20	0.08199	1.09670	0.21509	0.02490	1.97858	0.00499	0.73366
50	20	0.09998	1.14295	0.23230	0.02400	1.98186	0.00490	0.66695
5	25	0.00309	0.27461	0.03733	0.00171	1.98528	0.00131	1.10961
10	25	0.00914	0.61620	0.06437	0.00511	2.65115	0.00226	1.11000
15	25	0.01816	0.81240	0.09237	0.00930	2.47672	0.00305	1.04391
20	25	0.02480	0.68323	0.11406	0.01118	1.96121	0.00334	0.92653
25	25	0.03627	0.83107	0.13628	0.01494	2.07239	0.00386	0.89662
30	25	0.04051	0.95712	0.15862	0.01873	2.03023	0.00433	0.86289
40	25	0.06667	0.87896	0.19118	0.02189	1.83808	0.00468	0.77388
50	25	0.08320	1.14630	0.20733	0.02243	2.21301	0.00474	0.72245
5	30	0.00231	0.24133	0.02735	0.00100	2.13086	0.00100	1.15527
10	30	0.00755	0.56279	0.05712	0.00445	2.63124	0.00211	1.16814
15	30	0.01308	0.56747	0.07560	0.00607	2.10084	0.00246	1.03064
20	30	0.01959	0.65509	0.09756	0.00883	2.09749	0.00297	0.96262
25	30	0.02776	0.81295	0.12624	0.01382	1.93732	0.00372	0.93166
30	30	0.03584	0.92247	0.13955	0.01681	2.16135	0.00410	0.92867
40	30	0.04974	1.14766	0.16704	0.01921	2.13638	0.00438	0.82992
50	30	0.06710	0.91131	0.19054	0.02189	2.01357	0.00468	0.77662
5	40	0.00134	0.19433	0.02061	0.00062	2.19663	0.00079	1.20970
10	40	0.00469	0.47801	0.04402	0.00284	2.67286	0.00168	1.21096
15	40	0.00903	0.58772	0.06525	0.00591	2.57819	0.00243	1.17768
20	40	0.01411	0.52036	0.07589	0.00569	1.93793	0.00239	0.99375
25	40	0.01918	0.69514	0.09637	0.00966	2.24543	0.00311	1.01972
30	40	0.02602	0.84436	0.11114	0.01175	2.39624	0.00343	0.97567
40	40	0.03530	0.94894	0.13607	0.01518	2.13299	0.00390	0.90515
50	40	0.05017	1.04787	0.16463	0.02066	2.30368	0.00455	0.87320
5	50	0.00087	0.16276	0.01817	0.00052	2.45050	0.00072	1.24932
10	50	0.00320	0.41309	0.03528	0.00182	2.69121	0.00135	1.20987
15	50	0.00634	0.52402	0.05309	0.00423	2.73507	0.00206	1.22425
20	50	0.00981	0.62575	0.06682	0.00546	2.56801	0.00234	1.10606
25	50	0.01382	0.56857	0.07824	0.00692	2.25048	0.00263	1.06346
30	50	0.01892	0.72614	0.09335	0.00902	2.35710	0.00300	1.01668
40	50	0.02773	0.80446	0.11231	0.01224	2.24022	0.00350	0.98521
50	50	0.03540	0.92492	0.12203	0.01237	2.21719	0.00352	0.91157