MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 32, No. 9, pp. 1737–1752, 2003

The Maximum Probability $2 \times c$ Contingency Tables and the Maximum Probability Points of the Multivariate Hypergeometric Distribution

F. Requena[1,*] and N. Martín Ciudad[2]

[1]Department of Statistics,
University of Granada, Granada, Spain
[2]Department of Mathematics,
University of Extremadura, Cáceres, Spain

ABSTRACT

The problem of obtaining the maximum probability $2 \times c$ contingency table with fixed marginal sums, $\boldsymbol{R} = (R_1, R_2)$ and $\boldsymbol{C} = (C_1, \ldots, C_c)$, and row and column independence is equivalent to the problem of obtaining the maximum probability points (mode) of the multivariate hypergeometric distribution MH(R_1; $C_1, \ldots, C_c$). The most simple and general method for these problems is Joe's (Joe, H. (1988). Extreme probabilities for contingency tables under row and column independence with application to Fisher's exact test. *Commun. Statist. Theory Meth.* 17(11):3677–3685.) In this article

*Correspondence: F. Requena, Dpto. de Estadistica (Bioestadistica), Facultad de Medicina, Universidad de Granada, 18071 Granada, Spain; Fax: 34-58-246117; E-mail: fcoreque@ugr.es.

DOI: 10.1081/STA-120022706 0361-0926 (Print); 1532-415X (Online)

we study a family of MH's in which a connection relationship is defined between its elements. Based on this family and on a characterization of the mode described in Requena and Martín (Requena, F., Martín, N. (2000). Characterization of maximum probability points in the multivariate hypergeometric distribution. *Statist. Probab. Lett*. 50:39–47.), we develop a new method for the above problems, which is completely general, non recursive, very simple in practice and more efficient than the Joe's method. Also, under weak conditions (which almost always hold), the proposed method provides a simple explicit solution to these problems. In addition, the well-known expression for the mode of a hypergeometric distribution is just a particular case of the method in this article.

Key Words: $2 \times c$ contingency tables; Multivariate hypergeometric distribution; Mode.

1. INTRODUCTION

Let $x_1, \ldots\ldots, x_c$ denote the first row of a $2 \times c$ contingency table with fixed row marginals $\boldsymbol{R} = (R_1, R_2)$ and fixed column marginals $\mathbf{C} = (C_1, \ldots, C_c)$ and $N = \Sigma R_i = \Sigma C_j$, and let Z_0 denote the set of nonnegative integers. The set of possible $2 \times c$ tables given $\boldsymbol{R}$ and $\boldsymbol{C}$ can be represented by

$$F_0 = \{(x_1, \ldots, x_c) \in Z_0^c | x_1 + \cdots + x_c = R_1,\\ 0 \le x_h \le C_h \quad h = 1, \ldots, c\}$$

If we further assume the hypothesis of row and column independence, it is known that the random vector $\boldsymbol{x} = (x_1, \ldots\ldots, x_c)$ has the multivariate hypergeometric distribution (MH)

$$\mathrm{MH}(R_1\,;\ C_1, \ldots., C_c) \tag{1}$$

whose reference set is F_0. Thus, obtaining the maximum probability $2 \times c$ table for fixed marginals $\boldsymbol{R}$ and $\boldsymbol{C}$ and row and column independence is equivalent to obtaining the mode of distribution (1).

These problems appear in different applications, for example, as part of the best-known and most efficient algorithm for Fisher's exact test in contingency tables: the Mehta and Patel's network algorithm (Mehta and Patel, 1980; 1983). The application of this algorithm to an observed $2 \times c$ table requires, for many nodes of the network, the calculation of the longest subpath from each node to the terminal node. This involves many applications of calculation of the maximum probability $2 \times k$

table ($k < c$) for fixed marginals (or, equivalently, of the mode of corresponding MH); see (Mehta and Patel, 1980; 1983; Joe 1988).

For $c = 2$ we have the particular case of the hypergeometric distribution $H(R_1, C_1, N)$. It is known that the mode of this distribution is obtained as an integer value x_1^* subject to

$$R_1'(C_1 + 1) - 1 \leq x_1^* \leq R_1'(C_1 + 1) \tag{2}$$

with $R_1' = (R_1 + 1)/(N + 2)$; see Johnson et al. (1993). In the same way, each marginal $x_j\,(j = 1, \ldots\ldots, c)$ of the MH (1) is distributed as $H(R_1, C_j, N)$ and its mode will be an integer z_j subject to

$$R_1'(C_j + 1) - 1 \leq z_j \leq R_1'(C_j + 1) \tag{3}$$

In the case of the MH, Boland and Proschan (1987) propose a simple method of obtaining the mode in the particular case in which $C_j = mk_j$, $j = 1, \ldots, c$, and $R_1 = k_1 + \cdots + k_c$ where m and $k_1, \ldots, k_c$ are integer values. More general algorithms, like those proposed in (Mehta and Patel, 1980, 1983; Joe, 1988), are found in the context of the above-mentioned application and in its equivalent version of obtaining the maximum probability $2 \times c$ table. The simplest and most general is Joe's (1988) (valid for $r \times c$ tables), which is based on a necessary condition that generally involves the construction (by a recursive way) of a subset of F_0 and the inspection of the probability of its tables in order to obtain the maximum probability $2 \times c$ table, which is contained in that subset.

In this article we propose a completely general, non recursive and very simple method of calculating the mode of the MH (or, equivalently, the maximum probability $2 \times c$ table for fixed marginals), which is based on a characterization of the mode in terms of a necessary and sufficient condition described in Requena and Martín (2000) and it provides (under weak conditions, which hold in the vast majority of cases) explicit expressions for the mode. In fact, this method is so simple that it can be carried out on paper or with a pocket calculator in a very short time, for any value of c and for any $\boldsymbol{R}$ and $\boldsymbol{C}$. Moreover, the proposed method is more efficient than the Joe's method, especially when c becomes large. In practice, large values of c can arise when, for example, we consider the number of patients with one specific characteristic (illness, symptom, ...) that arrive at a casualty department in each of c independent periods. So if we take, say, 24 one-week periods, we will have one MH (or one $2 \times c$ table) with $c = 24$. Finally, the methodology developed in this article will provide a substantial reduction in the amount of computing time required in the application of Fisher's exact test to a $2 \times c$ table. See Sec. 5 for further details.

The next result summarizes the above-mentioned characterization of the mode of the MH.

Theorem 1.1. *The necessary and sufficient condition so that* $\boldsymbol{x}^* = (x_1^*, \ldots\ldots, x_c^*)$ *is a mode of the* MH (1) *is that* $\boldsymbol{x}^* \in S_0$, *with* S_0 *defined as*

$$S_0 = \left\{\boldsymbol{x} = (x_1, \ldots, x_c) \in F_0 | x_i \leq m_{ij}(\boldsymbol{x})\ i \neq j\right\} \tag{4}$$

where for any point $\boldsymbol{x} \in Z_0^c$ *we denote*

$$m_{ij}(\boldsymbol{x}) = (x_i + x_j + 1)(C_i + 1)/(C_i + C_j + 2) \tag{5}$$

We will also make use of the following property: *if* $\boldsymbol{x} = (x_1, \ldots, x_c)$ *and* $\boldsymbol{x}' = (x_1', \ldots, x_c')$ *are two arbitrary points belonging to* S_0, *then it is true that*

$$|x_h' - x_h| \leq 1 \quad 1 \leq h \leq c \tag{6}$$

This article is organized as follows: in Sec. 2 we define a particular family $\mathscr{F}$ of MH's to which the MH (1) belongs and study a connection between the modes of its elements. The idea behind the method of calculation, developed in Sec. 3, is to start with the modes of one of the MH's of $\mathscr{F}$, determined by the modes of the hypergeometric marginal distributions of the MH (1), and arrive at the modes of the MH (1) via this connection. In Sec. 4, there are several examples to illustrate the simplicity and speed of the method.

2. RELATION BETWEEN THE MODES OF THE MH'S OF A PARTICULAR FAMILY

Given the parameters $R_1, C_1, \ldots, C_c$ of the MH (1), let $\mathscr{F}$ be the family formed by all the MH's $H_e = \mathrm{MH}(R_1 + e;\ C_1, \ldots, C_c)$, $-R_1 \leq e \leq N - R_1$, and let F_e be the reference set of H_e, i.e.,

$$F_e = \left\{\boldsymbol{x} = (x_1, \ldots, x_c) \in Z_0^c \middle| \sum_h x_h = R_1 + e,\ 0 \leq x_h \leq C_h\ h = 1, \ldots, c\right\}$$

Note that the extreme distributions of $\mathscr{F}$, H_{-R_1}, and H_{N-R_1}, are reduced to the degenerated distributions in $(0, \ldots, 0)$ and in $(C_1, \ldots, C_c)$, respectively. On the other hand, from Sec. 1, for each H_e

we will have a subset of F_e (which we will denote by S_e) defined as in expression (4), i.e.,

$$S_e = \left\{ \boldsymbol{x} = (x_1, \ldots, x_c) \in F_e | x_i \leq m_{ij}(\boldsymbol{x}) \ i \neq j \right\} \tag{7}$$

S_e has the same properties as the S_0, in particular it contains and only contains the modes of H_e and inequality (6) is also satisfied. Moreover, it is obvious that for $e = 0$, H_0 coincides with the MH (1), and so this distribution belongs to $\mathscr{F}$.

If $\boldsymbol{x}$ has distribution (1), the family $\mathscr{F}$ is the same type as the family of the conditional distributions of $(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_c / x_k)$, which are MH($R_1 - x_k$; $C_1, \ldots, C_{k\text{-}1}, C_{k+1}, \ldots, C_c$), for all possible x_k values of the kth coordinate of $\boldsymbol{x}$. The latter family is discussed in Requena and Martín (2000), where a connection relationship between the modes of its MH's is defined and some results are obtained. For the family $\mathscr{F}$ here we are going to define similar relations between the modes of its elements. The study of these relations will allow us to develop a method of calculation of the mode of the MH (1).

Definition 2.1. We say that two points, $\boldsymbol{x} \in S_e$ and $\boldsymbol{x}' \in S_{e+1} (-R_1 \leq e < N - R_1)$ are connected by the vth coordinate (or connected by v) when $x'_v - x_v = 1$ and $x'_j = x_j$ for all $j \neq v$ are satisfied.

That is to say, two points connected by v coincide in all the coordinates except in the vth, in which they differ by one unit.

Definition 2.2. We say that two points, $\boldsymbol{x} \in S_e$ and $\boldsymbol{x}' \in S_{e+1} (-R_1 \leq e < N - R_1)$ and, are *connected* when there is a vth coordinate so that these points are connected by v.

Definition 2.3. Given two integer values, p and p', such that $-R_1 \leq p < p' \leq N - R_1$, we say that the finite succession of points

$$C(p, p') = \left\{ \boldsymbol{x}_{[e]} \in S_e \right\}_{e=p, p+1, \ldots, p'}$$

is a *chain* when for every e ($p \leq e < p'$) $\boldsymbol{x}_{[e]}$ and $\boldsymbol{x}_{[e+1]}$ are connected. Also, given any two points of the chain $C(p, p')$, we always say that these points are *communicated* by this chain.

Thus for each value of e, we will have a point $\boldsymbol{x}_{[e]}$ (link) of the chain and each link is connected to the previous one and the following one. Moreover, given two arbitrary points of the chain, $\boldsymbol{x}_{[a]}$ and $\boldsymbol{x}_{[b]}$ $(a < b)$, we will be able to go from $\boldsymbol{x}_{[a]}$ to $\boldsymbol{x}_{[b]}$ (or vice versa) through the successive intermediate links of the chain, which implies executing *b-a* steps from $\boldsymbol{x}_{[e]}$ to $\boldsymbol{x}_{[e+1]}$ (or from $\boldsymbol{x}_{[e+1]}$ to $\boldsymbol{x}_{[e]}$) $a \leq e < b$.

The following results (Theorem 2.1 and Theorem 2.2) are presented without proof because they are similar to those presented in Requena and Martín (2000). Firstly, given two contiguous MH's of $\mathscr{F}$, H_e, and H_{e+1}, the following theorem characterizes the connection between their modes.

Theorem 2.1. *For any pair of distributions, H_e and H_{e+1}, of $\mathscr{F}$, given $\mathbf{x} \in S_e$ there will be another point $\mathbf{x}' \in S_{e+1}$, connected to it by v, if and only if v is a value of j $(1 \leq j \leq c)$ which minimizes the expression $(x_j+1)/(C_j+1)$. Reciprocally, given $\mathbf{x}' \in S_{e+1}$ another point $\mathbf{x} \in S_e$ will exist, connected to it by u, if and only if u is a value of j $(1 \leq j \leq c)$ which minimizes the expression $-x'_j/(C_j+1)$.*

*v*th and *u*th coordinates with the condition of Theorem 2.1 always exist, although they are not necessarily unique. Therefore we can immediately deduce the following result.

Corollary 2.1. *Given two sets, S_e and S_{e+1}, each point of one set is connected to at least one point of the other, i.e., given the distributions H_e and H_{e+1} of $\mathscr{F}$, each mode of one of them is connected to at least one mode of the other.*

Theorem 2.2. *Let S_a and S_b $(a < b)$ correspond to two MH's of $\mathscr{F}$. Given a point of one of the sets, in the other set a point will always exist such that the two points are communicated by a chain.*

This theorem states that given two MH's of $\mathscr{F}$, each one of the modes of one of them is always communicated by a chain with at least one mode of the other. Finally, the following theorem states the relationship between the coordinates of two points communicated by a chain.

Theorem 2.3. *If $\mathbf{x}_{[a]} \in S_a$ (with coordinates x_{aj}) and $\mathbf{x}_{[b]} \in S_b$ (with coordinates x_{bj}) $(b = a + d, d > 0)$ are two points communicated by a chain $C(a, b)$, then $x_{bj} = x_{aj} + q_j$ holds, where q_j is the number of pairs of points ($\boldsymbol{x}_{[e]}$ and $\boldsymbol{x}_{[e+1]}$, $a \leq e < b$) of $C(a,b)$ in which both points are connected by j. Moreover, $0 \leq q_j \leq d$ and $\Sigma q_j = d$ always hold.*

Proof. This theorem follows from the fact that given two points $\boldsymbol{x}_{[a]}$ and $\boldsymbol{x}_{[b]}$, communicated by a chain $C(a, b)$, we can go from one point to the other via the intermediate points (links) of the chain. This implies performing $d = b - a$ steps from $\boldsymbol{x}_{[e]}$ to $\boldsymbol{x}_{[e+1]}$ $a \leq e < b$, where in each one of these steps the jth coordinate through which the two points are connected increases by 1 when passing from $\boldsymbol{x}_{[e]}$ to $\boldsymbol{x}_{[e+1]}$ with the rest of the coordinates remaining constant. Therefore $x_{bj} = x_{aj} + q_j$ with $q_j \geq 0$. As the jth coordinate is not necessarily the same in all the steps (in fact it is usually different in each step) $q_j \leq d$ holds. Finally, as we have stated before, it is immediate that $\Sigma q_j = d$.

3. CALCULATING THE MAXIMUM PROBABILITY POINTS OF THE MH

Let D be the set of points whose coordinates are modes of the marginals x_j of distribution (1), i.e.,

$$D = \left\{ \boldsymbol{z} = (z_1, \ldots, z_c) \in Z_0^c \mid R_1'(C_j + 1) - 1 \leq z_j \leq R_1'(C_j + 1) \quad j = 1, \ldots, c \right\} \tag{8}$$

Given a $\boldsymbol{z} \in D$, let $\delta = \Sigma z_j - R_1$. From expression (8) we can directly obtain a maximum and a minimum for Σz_j and we can easily deduce that $-\{(1 - R_1')(c - 2) + 1\} \leq \delta \leq R_1'(c - 2) + 1$. However as $z_j \geq 0$ we have -$R_1 \leq \delta$, and as $z_j \leq C_j$ we have $\delta \leq N - R_1$. This is summarized in the following result.

Lemma 3.1. *Given the* MH (1) *and any point* $\boldsymbol{z} \in D$,

$$\max(-R_1, R_1^* - c) \leq \delta \leq \min(N - R_1, R_1^*) \tag{9}$$

always holds, where $R_1^* = R_1'(c - 2) + 1$.

On the other hand, for a particular value of δ satisfying inequality (9), we define the set

$$D_\delta = \left\{ \boldsymbol{z} \in D \,\middle|\, \sum z_j - R_1 = \delta \right\}$$

We will see how each nonempty D_δ is used as a starting point to get the modes of the MH (1).

Theorem 3.1. *Given the* MH (1) *and the family* $\mathscr{F}$, *if* δ *is an integer such that* D_δ *is nonempty, then* D_δ *coincides with the set* S_δ *corresponding to* H_δ *of* $\mathscr{F}$.

Proof. Let $\boldsymbol{z} \in D_\delta$. Then we have

$$0 \leq z_j \leq C_j, \quad \sum z_j = R_1 + \delta \tag{10}$$

and inequality (3) holds for $j = 1, \ldots, c$. From expression (3) we can deduce

$$z_i/(C_i + 1) \le R_1' \le (z_j + 1)/(C_j + 1) \quad i \ne j \tag{11}$$

and therefore $z_i \le m_{ij}(\mathbf{z})$, $i \ne j$ holds. Finally, from Eq. (10) we get $\mathbf{z} \in S_\delta$. Thus S_δ contains the points of D_δ. In order to demonstrate that S_δ only contains the points of D_δ, consider that S_δ contains more than one point (the proof would be trivial if S_δ only contains one point). Let $\mathbf{z} \in S_\delta$ such that $\mathbf{z} \in D_\delta$, and let $\mathbf{z}' \in S_\delta$.To prove that $\mathbf{z}' \in D_\delta$, it is sufficient that

$$R_1'(C_j + 1) - 1 \le z_j' \le R_1'(C_j + 1) \tag{12}$$

holds for $j = 1, \ldots, c$. Suppose that $\mathbf{z}$ and $\mathbf{z}'$ only differ in t coordinates. As $\mathbf{z} \in D_\delta$, expression (12) holds for the remaining c-t coordinates of $\mathbf{z}'$. On the other hand, from expression (6) and since $\Sigma z_i = \Sigma z_i'$, t will be an even number and these t coordinates will be structured in $t/2$ pairs, such that if $(z_i\ ,\ z_j)$ is one of these pairs in $\mathbf{z}$, its corresponding pair (z_i', z_j') in $\mathbf{z}'$ will satisfy $z_i' = z_i + 1$ and $z_j' = z_j - 1$. Moreover, as $\mathbf{z}' \in S_\delta$, for each one of these pairs we have $z_i' \le m_{ij}(\mathbf{z}')$, from which

$$z_i'/(C_i + 1) \le (z_j' + 1)/(C_j + 1) \tag{13}$$

follows. Finally, from inequalities (11) and (13), we have $z_i'/(C_i + 1) = (z_j' + 1)/(C_j + 1) = R_1'$. Therefore, the t coordinates of $\mathbf{z}'$ in which $\mathbf{z}'$ and $\mathbf{z}$ do not coincide also satisfy inequality (12) and this concludes the proof.

From this theorem and as we already know that S_e contains and only contains the modes of H_e, we can directly deduce the following result.

Corollary 3.1. *Given the* MH (1), *the family* $\mathscr{F}$ *and a nonempty* D_δ, *it is true that* D_δ *contains and only contains the modes of the* MH($R_1 + \delta$; $C_1, \ldots, C_c$). *In particular if* D_0 *is nonempty*, D_0 *contains and only contains the modes of the* MH (1).

As we have seen: (from Theorem 1.1) the set S_0 contains and only contains the modes of the MH (1), (from Theorem 3.1) D_δ coincides with S_δ if it is nonempty, and (from Theorem 2.2) starting from all the points of D_δ (or S_δ) we will arrive by different chains at all the points of S_0. From now on we will use δ^* to denote the value δ of the set D_δ that we start from, and $m = |\delta^*|$ will be the number of steps to arrive at S_0. It is obvious that in practice we will take as δ^* the value of δ that is nearest to 0. In the particular case where $\delta^* = 0$ it is because D_0 is nonempty and we will not have to execute any intermediate steps, D_0 (or S_0) would just be the set of all the modes of the MH (1), as we said in Corollary 3.1. Likewise, we will use $\boldsymbol{z}^*$ to denote the point of D_{δ^*} that we start from to arrive at a mode of the MH (1) which we will denote by $\boldsymbol{x}^*$.

If the modes of all the marginals x_j are unique, D will have only one point $\boldsymbol{z}$ and therefore only one nonempty D_δ will exist. In these cases it is clear that $\delta^* = \delta$ and $\boldsymbol{z}^* = \boldsymbol{z}$. However, if there are q marginals x_j $(q \leq c)$ whose modes are not unique, we can see from expression (3) that for each one of these x_j, $R_1'(C_j + 1)$ will be an integer and there will be two modes: $M_j = R_1'(C_j + 1)$ and $M_j' = R_1'(C_j + 1) - 1$. In this situation, the set D will have 2^q points and we will have several nonempty sets D_δ. Moreover, taking $z_j = M_j'$ and $z_j = M_j$ for all these q marginals we would obtain respectively the minimum value (δ_1) and the maximum value (δ_2) of δ satisfying that D_δ is nonempty. And for any δ such that $\delta_1 \leq \delta \leq \delta_2, D_\delta$ will be nonempty. Thus, if $\delta_1 > 0$, then $\delta^* = \delta_1$; if $\delta_2 < 0$, then $\delta^* = \delta_2$; otherwise, $\delta^* = 0$. Finally, $\boldsymbol{z}^* \in D_{\delta^*}$ will be a point $\boldsymbol{z} \in D$ in which we take $z_j = M_j'$ for $\delta_2 - \delta^*$ of the above q marginals and we take $z_j = M_j$ for the remaining $q - (\delta_2 - \delta^*)$. Note that D_{δ^*} has a unique point $\boldsymbol{z}^*$ except when $\delta^* = 0$.

For $\delta^* \neq 0$ (i.e., $m > 0$), the procedure to go from the point $\boldsymbol{z}^* \in D_{\delta^*}$ to a point $\boldsymbol{x}^* \in S_0$, and to get the coordinates of $\boldsymbol{x}^*$ from the coordinates of $\boldsymbol{z}^*$, is described as follows. Let us define

$$\lambda = \begin{cases} -1 & \text{if } \delta^* > 0 \\ 1 & \text{if } \delta^* < 0 \end{cases} \tag{14}$$

(1) Start with $\boldsymbol{z}^* = (z_1^*, \ldots, z_c^*) \in D_{\delta^*}$. Set $k = 1$ and $\boldsymbol{z}' = z^*$.
(2) Step k:

  (i) By application of Theorem 2.1, take a point $\boldsymbol{z}'' \in S_{\delta^* + \lambda k}$ connected to $\boldsymbol{z}'$ by, say, v.
  (ii) Set $e_{vk} = 1$ and $e_{jk} = 0,\ 1 \leq j \leq c,\ j \neq v$.
  (iii) For $j = 1, \ldots, c,$ calculate

$$q_{jk} = \sum\nolimits_{h=1}^{k} e_{jh} \tag{15}$$

  (iv) The point $\boldsymbol{z}''$ will have the coordinates $z_j^* + \lambda q_{jk}$, $j = 1, \ldots, c$.

(3) If $k < |\delta^*|$, then set $k = k + 1, \boldsymbol{z}' = \boldsymbol{z}''$ and go to (2). Otherwise, continue.
(4) The last $\boldsymbol{z}''$ (which we denote by $\boldsymbol{x}^*$) is just a point of S_0, i.e., a mode of the MH (1).

Furthermore, as every point of S_0 is necessarily communicated with some point of $D_{\delta *}$, its coordinates will always be expressed in the previous way. This process demonstrates the following theorem, which summarizes the expression of the mode of the MH (1).

Theorem 3.2. *Given* $\mathbf{z}^* \in D_{\delta^*}$ *for* δ^* *defined above,* $\boldsymbol{x}^*$ *with coordinates*

$$x_j^* = \begin{cases} z_j^* & \text{if } \delta^* = 0 \\ z_j^* + \lambda q_{jm} & \text{if } \delta^* \neq 0 \end{cases} \quad j = 1, \ldots, c \tag{16}$$

is a mode of the MH (1), *where* $m = |\delta^*|$ *and* λ *and* q_{jm} *are given in the expressions* (14) *and* (15) *respectively. Reciprocally, any mode of distribution* (1) *will be expressed in the form* (16) *for some* $\mathbf{z}^* \in D_{\delta^*}$.

3.1. Practical Algorithm

From a practical point of view and given the point $\mathbf{z}'$ prior to each step, let us redefine the expressions in Theorem 2.1 as

$$w_j = \begin{cases} -z_j'/(C_j + 1) & \text{if } \delta^* > 0 \\ (z_j' + 1)/(C_j + 1) & \text{if } \delta^* < 0 \end{cases}$$

for $j = 1, \ldots, c$. At the first step $\mathbf{z}' = \mathbf{z}^*$ and we will denote its corresponding w_j's by w_j^*'s. On the other hand, let J be a set of s integer values, $j_1, \ldots, j_r, j_{r+1}, \ldots, j_s$ such that

$$w_{j_1}^* \leq \cdots \leq w_{j_r}^* < w_{j_{r+1}}^* = \cdots = w_{j_s}^* < w_j^*, \quad j \neq j_1, \ldots, j \neq j_s \tag{17}$$

where $0 \leq r < m$, s is the smallest integer such that $m \leq s$ and we assume that the order between the elements of J is determined by expression (17), though the order is indifferent between the elements h and h' when $w_h^* = w_{h'}^*$. Also, for $0 \leq k \leq s$, let J_k be the subset formed by the first k elements, $j_1, \ldots, j_k$, of J, with the understanding that J_0 is the empty set. From the above procedure it is easy to see that the only coordinates involved, in order to get all the modes of the MH (1), are the j_1th, $\ldots, j_s$th corresponding to the elements of J, though in order to get one mode it suffices to consider the coordinates corresponding to the elements of J_m. Moreover, at each step the w_j's involved in the application of Theorem 2.1 are obtained from the w_j's of the previous step (starting from the w_j^*'s at the step 1). In this regard, let us define

$$W_i(j) = w_j^* + i/(C_j + 1) \tag{18}$$

for $0 \leq i \leq m - h$ when $j = j_h \in J_m$ and for $i = 0$ when $j \in J - J_m$.

In order to get one mode, in the first application of Theorem 2.1 (step 1 of (2) above) $w_j = W_0(j), j \in J_m$, and we have to take the minimum of these values. If this minimum is reached on the, say, vth coordinate ($v \in J_m$), then z' and z'' are connected by v. So the w_j's, $j \in J_m$, in the second application (step 2 of (2) above) are the same as in the previous step except that $W_0(v)$ is replaced by $W_1(v)$. Next, we have to take the minimum of these w_j's, $j \in J_m$, again, and so on. After the m steps, it is easily seen that it suffices to know the m lower values of the $W_i(j)$'s and their corresponding values j's. Thus, the method of calculating one mode $\boldsymbol{x}^*$of the MH (1) can be implemented as follows.

(1) From expression (3), obtain a vector $\boldsymbol{z} = (z_1, \ldots, z_c) \in D_\delta$ with $\delta = \Sigma z_j - R_1$.
(2) If $\boldsymbol{z}$ is unique, then $\delta^* = \delta$ and $\boldsymbol{z}^* = \boldsymbol{z}$. Otherwise, get δ^* and $\boldsymbol{z}^*$ as shown above.
(3) If $\delta^* = 0$, then $\boldsymbol{x}^* = \boldsymbol{z}^*$ and the algorithm terminates. Otherwise, continue.

Remarks. For $\delta^* = 0$, $\boldsymbol{z}^*$ is not necessarily unique. So the set of all $\boldsymbol{z}^*$'s will be the set of all modes of the MH (1). For $\delta^* \neq 0$, $\boldsymbol{z}^*$ is unique.

(4) Calculate w_j^* for $j = 1, \ldots, c$ and get J and J_m.
(5) From Eq. (18), calculate the necessary $W_i(j)$'s in order to get the m lower of these values $W_i(j)$'s, $j \in J_m$ and any i. Also make a note of the value j associated with each one of these m values.

Remarks. To do this, note that for each j, $W_i(j) < W_{i'}(j)$ according as $i < i'$ and all $W_0(j)$'s are already calculated and ordered since $W_0(j) = w_j^*$. Because of this, and since in the majority of cases m is very close to 0 (see Sec. 5 below), very few of the $W_i(j)$'s will be necessary to calculate. Also, note that the above set of m lower values is not necessarily unique.

(6) Count how many of the m lower values $W_i(j)$'s obtained in (5) are associated with the jth coordinate ($j \in J_m$). Denote this count by n_j (obviously $\Sigma n_j = m$). Then obtain $\boldsymbol{x}^* = (x_1^*, \ldots, x_c^*)$ as

$$x_j^* = \begin{cases} z_j^* + \lambda n_j & \text{if } j \in J_m \\ z_j^* & \text{if } j \notin J_m \end{cases}$$

In order to get all the modes of the MH (1) (if the mode is not unique) we can use the same algorithm described above with J_m replaced by J. Then, each one of the sets containing m lower values $W_i(j)$'s, $j \in J$, will yield one mode. Moreover, note that if we have already obtained one mode (applying the algorithm with J_m), then we do not need any additional calculations in order to get the remaining modes. So, a new general expression of the mode $\boldsymbol{x}^* = (x_1^*, \ldots, x_c^*)$ of the MH (1) can be written as

$$x_j^* = \begin{cases} z_j^* + \lambda n_j & \text{if } m > 0 \text{ and } j \in J \\ z_j^* & \text{otherwise} \end{cases}$$

for $\boldsymbol{z}^* \in D_{\delta^*}$ and some set of m lower values $W_i(j)$'s, $j \in J$.

However, in the vast majority of cases, it is not necessary to implement the complete algorithm, it suffices to know $\boldsymbol{z}^*$, the w_j^*'s and J. Note that if $\delta^* = 0$, then $\boldsymbol{x}^* = \boldsymbol{z}^*$as shown at the step (3) above. Moreover, for $\delta^* \neq 0$, , we can get more explicit expressions for the mode of the MH (1) in terms of $\boldsymbol{z}^*$ and J if certain simple conditions are satisfied. This conditions are easily checked and one of them almost always holds (see Sec. 5 below). In this sense, we next present some results which follow directly from the Theorem 3.2 and from the algorithm described above.

Corollary 3.2. *For* $m > 1$, *given* $\boldsymbol{z}^* \in D_{\delta^*}$ *and* $J_2 = \{j_1, j_2\}$, $\boldsymbol{x}^*$ *with coordinates*

$$x_j^* = \begin{cases} z_j^* + \lambda m & \textit{if} \quad j = j_1 \\ z_j^* & \textit{otherwise} \end{cases}$$

$(j = 1, \ldots, c)$ *is a mode of the* MH (1) *if and only if the following condition holds*:

$$(w_{j_2}^* - w_{j_1}^*)(C_{j_1} + 1) \geq m - 1 \tag{19}$$

Moreover, if the condition (19) with strict inequality holds, then the mode of the MH (1) is unique.

Corollary 3.3. *Given* $\boldsymbol{z}^* \in D_{\delta^*}$ and $J = \{j_1, \ldots, j_s\}$, *if* $\boldsymbol{x}^*$ *has the coordinates*

$$x_j^* = \begin{cases} z_j^* + \lambda & \textit{if } j \in J' \\ z_j^* & \textit{otherwise} \end{cases} \tag{20}$$

$(j = 1, \ldots, c)$ *where* J' *is a subset of* J, *then the following results are true.*

(i) For $m = 1$, the MH (1) has s modes, which (for $h = 1, \ldots, s$) are given by Eq. (20) with $J' = \{j_h\}$ $j_h \in J$.

(ii) For $m > 1$, $\boldsymbol{x}^*$ given by Eq. (20) with $J' = J_m$ is a mode of the MH (1) if and only if $r = 0$ or

$$(w^*_{j_m} - w^*_{j_h})(C_{j_h} + 1) \leq 1 \qquad (21)$$

holds for $h = 1, \ldots, r$ ($r > 0$), where r is the integer value arising from the definition of J.

(iii) For $m > 1$, if $r = 0$ or the condition (21) with strict inequality holds for $h = 1, \ldots, r$ ($r > 0$), then the MH (1) has $\binom{s-r}{m-r}$ modes, which are given by Eq. (20) with J' being a subset formed by the elements of J_r and m-r elements of $J - J_r$.

4. EXAMPLES

As the application of the proposed method is fairly straightforward when $|\delta^*| = 0$ or $|\delta^*| = 1$ (even for $|\delta^*| = 2$, it is very easy), let us consider three examples with $|\delta^*| > 2$ in order to illustrate our method in a better way. Here we will use the notation $a|t$ to indicate that the value a is repeated t times.

Example 1. Consider the MH(224; 8, 12|2, 13, 14, 17, 19, 21, 24, 27|3, 28, 31) or the 2×14 table with $\boldsymbol{R} = (224, 56)$ and $\boldsymbol{C}$=(8, 12|2, 13, 14, 17, 19, 21, 24, 27|3, 28, 31). Firstly we obtain $\boldsymbol{z} = $(7, 10|2, 11|2, 14, 15, 17, 19, 22|3, 23, 25) with $\delta = 4$. As $\boldsymbol{z}$ is unique, we have $\delta^* = 4$, $m = 4$, and $\boldsymbol{z}^* = \boldsymbol{z}$. Next we calculate the values w^*_j's, giving −0.778, −0.769|2, −0.786, −0.733, −0.778, −0.75, −0.773, −0.76, −0.786|3, −0.793, −0.781, for $j = 1, \ldots, 14$, respectively. Therefore $J = \{13, 4, 10, 11, 12\}$, $J_4 = \{13, 4, 10, 11\}$ and $r = 1$. As the condition (21) with strict inequality holds for $h = 1$, from Eq. (20) with $J' = J_4$ we have that $\boldsymbol{x}^* = $(7, 10|3, 11, 14, 15, 17, 19, 21|2, 22|2, 25) is one mode of the above MH. Likewise, $\boldsymbol{x}^*$ is the first row of one maximum probability 2×14 table for the given marginals. In addition, if we want to get all the modes (or all the maximum probability 2×14 tables), from the Corollary 3.3 the total number of modes is 4. One of them is $\boldsymbol{x}^*$ and the three remaining can be written directly from Eq. (20) by taking {13, 4, 10, 12}, {13, 4, 11, 12} and {13, 10, 11, 12} as J'.

Example 2. Consider the MH(28; 4|4, 7|4, 8, 9, 12|3, 43) or the 2×14 table with $\boldsymbol{R} = (28, 112)$ and $\boldsymbol{C} = $(4|4, 7|4, 8, 9, 12|3, 43). Firstly we obtain

$z = (1|9,\ 2|4,\ 8)$ with $\delta = -3$, and this z is unique. Thus, $\delta^* = -3$, $m = 3$, $z^* = z$, the values w_j^*'s ($j = 1, \ldots, 14$) are, respectively, 0.4|4, 0.25|4, 0.222, 0.3, 0.231|3, 0.205 and $J = \{14, 9, 11, 12, 13\}$. Since conditions (19) and (21) do not hold, we have to implement the complete algorithm. So, if we need only one mode, then we consider $J_3 = \{14, 9, 11\}$ and it suffices to calculate $W_1(14) = 0{,}227$ and $W_1(9) = 0.333$ in order to get the three lower values $W_i(j)$'s, $j \in J_3$, which are $W_0(14) = 0.205$, $W_0(9) = 0.222$, and $W_1(14) = 0{,}227$, corresponding to the 14th, 9th, and 14th coordinates, respectively. Thus $n_{14} = 2$, $n_9 = 1$, $n_{11} = 0$ and $\boldsymbol{x}^* = (1|8,\ 2|5,\ 10)$ is one mode of the above MH. On the other hand, note that $W_0(14)$, $W_0(9)$, and $W_1(14)$ above constitute the only set of three lower values $W_i(j)$'s, $j \in J$ and, therefore, the mode $\boldsymbol{x}^*$ is unique. Likewise, $\boldsymbol{x}^*$ is the first row of the only maximum probability 2×14 table for the given marginals.

Example 3. For the MH(24; 4|4, 5|5, 8, 9, 17, 69) or the 2×13 table with $\boldsymbol{R} = (24, 120)$ and $\boldsymbol{C} = (4|4,\ 5|5,\ 8,\ 9,\ 17,\ 69)$, as in the previous examples, we obtain $z = (0|4,\ 1|7,\ 3,\ 11)$, $\delta = -3$, $\delta^* = -3$, $m = 3$, $z^* = z$, the w_j^*'s ($j = 1, \ldots, 13$) are 0.2|4, 0.333|5, 0.222, 0.2, 0.222, 0.171 and $J = \{13, 1, 2, 3, 4, 11\}$. As the condition (19) holds (but not with strict inequality), $\boldsymbol{x}^* = (0|4,\ 1|7,\ 3,\ 14)$ is one mode, though it is not unique. If we want to get all the modes, then we implement the algorithm considering J. We calculate $W_1(13) = 0.186$ and $W_2(13) = 0.2$. Note that the set of three lower values $W_0(13) = 0.171$, $W_1(13) = 0.186$, and $W_2(13) = 0.2$ yields the preceding mode $\boldsymbol{x}^*$ and that there exist five additional sets of three lower values $W_i(j)$'s, $j \in J$, which can be obtained directly by replacing $W_2(13)$ by one of the $W_0(1)$, $W_0(2)$, $W_0(3)$, $W_0(4)$, $W_0(11)$ (all equal to 0.2) in the set before. So we can write directly the five corresponding (additional) modes. Likewise, these six modes above just correspond with the first rows of the six unique maximum probability 2×13 tables for the given marginals.

5. DISCUSSION

In order to explore the characteristics and the performance of the proposed method, we have carried out a simulation study in which we have generated (and inspected) more than 1.5×10^{11} MH's, without repetition of equivalent MH's, with $3 \leq c \leq 14$ and values of N depending on c, but always less than 200 (we say the same for the corresponding $2 \times c$ tables). From the previous sections and based on the valuable experience provided by the above study, we can to point out the following aspects.

In spite of the bounds for δ (see Lemma 3.1), m (or δ^*) is usually very close to zero. In fact, we have verified it by above study. Globally we have obtained that 45.0, 72.0, 96.8, and 99.9% of the inspected MH's have the value m less than or equal to 1, 2, 4, and 6, respectively, though these percentages depend on c, for example, they are 63.0, 92.4, 100, and 100% for $c = 7$ and 26.4, 48.4, 90.0, and 99.7% for $c = 13$.

In the vast majority of cases it is not necessary to implement the complete algorithm, because we can use an explicit expression for the mode in terms of z^* and J as shown in Corollaries 3.2 and 3.3 (or simply because $\boldsymbol{x}^* = z^*$ when $m = 0$). In this sense, an explicit expression always exists for $m < 3$ and, in general, it almost always exists. From our study about 98% of the inspected MH's were solved by an explicit expression. Moreover, when it is necessary to carry out the complete algorithm, from our study we have seen that values m tend to be even smaller than usual. In addition, this algorithm does not involve a recursive process and it does not require checking or calculating the probability of any point of the MH (or $2 \times c$ table).

Because of these reasons, the proposed method is very easy and very fast in practice. In fact, it can always be carried out on paper or with a pocket calculator in an easy way and in a short time, for any value of c and for any $\boldsymbol{R}$ and $\boldsymbol{C}$. Instead, implementation of the Joe's method with a pocket calculator could be very tedious, especially when value c becomes relatively large, because it could require many recursive applications of Theorem 6 of Joe (1988) and the inspection of the probability of the $2 \times c$ tables which result from this recursive process. This is so because it is based only on a necessary condition, but our method is based on a necessary and sufficient condition. Moreover, though the computation times of the Joe's method (when c is not too large) are not important if we use a computer, according to our experience (from the above study) the proposed method is always faster than Joe's, and the ratio of times "*Joe's method/proposed method*" increases as c becomes larger, with this ratio being very large in many cases when c takes a large value. So, in order to get the mode(s) of the MH (or equivalent problem in $2 \times c$ tables) we can say that the proposed method is more efficient than Joe's.

On the other hand, using the methodology developed in this article we can straightforwardly construct new practical algorithms which, in a recursive way and taking the mode of the MH as a starting point, allow us to determine new maximum probability points, like the maximum probability points of the MH with fixed value x_k at the kth coordinate (for each one of the possible values x_k) and the mode of the conditional MH given the value x_k of the kth coordinate. Likewise, if we apply the results described in this article to the Mehta and Patel's network algorithm for a $2 \times c$ table, we will have two advantages. First, we will compute the longest subpaths (from all nodes) in a recursive way and in a negligible time. Second, we will obtain a drastic reduction in the number of inequalities that we have to check in each node. Thus, we will achieve a substantial reduction in the amount of computing time required in the network algorithm. Moreover, the results in this article can be used as a starting point to get an extension to the case of $r \times c$ tables ($r > 2$), although more investigation is needed in this regard.

Finally, note that for $c = 2$ the proposed method is reduced to the known expression (2) of the mode of the hypergeometric distribution. Also, it is easily seen that this method gives us exactly the result given in Boland and Proschan (1987) for the particular case of MH proposed there.

ACKNOWLEDGMENTS

The authors would like to thank the referee for helpful comments which improved this article. This research was supported in part by the DGES, Spain, PB96-1443.

REFERENCES

Boland, P. J., Proschan, F. (1987). Schur convexity of the maximum likelihood function for the multivariate hypergeometric and multinomial distributions. *Statist. Probab. Lett.* 5:317–322.

Joe, H. (1988). Extreme probabilities for contingency tables under row and column independence with application to Fisher's exact test. *Commun. Statist. Theory Meth.* 17(11):3677–3685.

Johnson, N. L., Kotz, S. A., Kemp, W. (1993). *Univariate Discrete Distributions*. 2nd ed. New York: Wiley, 253–255.

Mehta, C. R., Patel, N. R. (1980). A network algorithm for the exact treatment of the 2 × k contingency table. *Commun. Statist. Simula. Comput.* 9(6):649–664.

Mehta, C. R., Patel, N. R. (1983). A network algorithm for performing Fisher's exact test in r × c contingency tables. *J. Amer. Statist. Assoc.* 78:427–434.

Requena, F., Martín, N. (2000). Characterization of maximum probability points in the multivariate hypergeometric distribution. *Statist. Probab. Lett.* 50:39–47.