

***APORTACIONES A LA MEJORA DEL
ALGORITMO DEL TEST DE FISHER EN
TABLAS 2xc. UNA EXTENSIÓN A TABLAS rxc.***

TESIS DOCTORAL

**NATIVIDAD MARTÍN CIUDAD
1998**

***Edita: Universidad de Extremadura
Servicio de Publicaciones***

Caldereros 2. Planta 3^a
Cáceres 10071
Correo e.: publicac@unex.es
<http://www.unex.es/publicaciones>

APORTACIONES A LA MEJORA DEL ALGORITMO DEL TEST DE FISHER EN
TABLAS $2 \times c$. UNA EXTENSIÓN A TABLAS $r \times c$.

Memoria para la obtención del Grado
de Doctor realizada por:

NATIVIDAD MARTÍN CIUDAD.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO I: Algoritmos de enumeración para resolver el Test exacto de Fisher. Algoritmos más notables de enumeración incompleta.....	8
I.1: Algoritmos de enumeración para resolver el Test exacto de Fisher	9
I.2: Algoritmos más notables de enumeración incompleta.....	15
I.2.1: Algoritmo de Marcelo Pagano y K.T. Halvorsen.....	15
I.2.2: Algoritmo de Mehta y Patel.....	19
I.3: Mejoras al algoritmo de red de Mehta y Patel.....	45
I.3.1: Probabilidades extremas de Harry Joe para tablas de contingencia $r \times c$	45
I.3.2: Mejoras computacionales de Clarkson, Fan y Joe.....	53
I.4: Discusión de los algoritmos de enumeración incompleta. Planteamientos de posibles mejoras. Concreción de los objetivos de esta tesis.....	54
CAPÍTULO II: Aportaciones sobre el máximo de una distribución hipergeométrica multivariante. Mejoras en el algoritmo del test exacto de Fisher para tablas $2 \times c$	57
II.1: Introducción.....	58
II.2: Desarrollo de un método general para el cálculo de la moda de una distribución hipergeométrica multivariante.....	59
II.2.1: Introducción.....	59
II.2.2: Notación.....	61
II.2.3: Caracterización de la moda de una MHD.....	61
II.2.4: Relación entre las modas de las MHD de una	

familia particular. Conexión entre modas. Cadena de modas.....	64
II.2.5: Nuevo método de cálculo de la moda de una MHD.....	75
II.2.6: Comparación del nuevo método de cálculo de la moda de una MHD con los métodos previamente existentes.....	84
II.2.7: Ejemplos y tiempos de cómputo	90
II.3: Relaciones entre las probabilidades extremas de tablas correspondientes a los nodos de la red de Mehta – Patel.....	93
II.3.1: Introducción.....	93
II.3.2: MHD asociadas a los nodos del retículo de Mehta-Patel. Notación.....	95
II.3.3: Relación entre las modas de las MHD asociadas a los nodos hijos de un nodo madre en la construcción reticulada de M-P para una tabla 2xc. Relación entre sus probabilidades.....	96
II.3.4: Relación entre la moda de la MHD asociada a un nodo de un estado y la moda de la MHD asociada a un nodo hijo especial del estado siguiente. Nodo hijo principal.....	99
II.3.5: Relaciones de conexión en una MHD asociada a un nodo madre a través de la conexión de las modas de las MHD asociadas a sus nodos hijos . Cadena de puntos.....	100
II.3.6: Subcadenas de puntos de la MHD asociada a un nodo madre con eslabón conocido.....	104
II.3.7: Ordenación entre las probabilidades de los puntos de una cadena. Ordenación entre las longitudes	

INTRODUCCIÓN

En los últimos años está aumentando considerablemente el interés por mejorar el alcance de los métodos exactos para tablas $r \times c$; La razón podríamos decir que es doble, en primer lugar la presencia de ordenadores, cada vez más potentes, está aminorando el principal impedimento de su uso: dificultad computacional. En segundo lugar, y aparte de otras consideraciones, cabe citar la no existencia de una guía simple que nos indique cuándo las aproximaciones asintóticas para grandes muestras son adecuadas (Cressie and Read (1989), Koehler and Larntz, (1980)), pues aún cuando el tamaño muestral es grande, recientes trabajos (por ejemplo, Haberman 1988) han mostrado que las aproximaciones pueden ser muy pobres cuando las tablas de contingencia contienen frecuencias esperadas, tanto pequeñas como grandes.

Fisher (1934,1935), Irwin (1935) y Yates (1934) fueron los primeros que propusieron un test exacto para el caso de tablas 2×2 , que conocemos como test exacto de Fisher, aunque eran conscientes de que el procedimiento podría ser usado igualmente en tablas $r \times c$. Este test requiere comparar la tabla 2×2 observada con la familia de tablas 2×2 con las mismas frecuencias marginales que la tabla observada, lo que supone condicionar en los marginales de la tabla que no son fijos, esto es, se trata de un test condicionado.

No es ni mucho menos la única opción de un test exacto para tablas 2×2 , Barnard (1945, 1947) propuso un test exacto para comparar 2 proporciones con muestras independientes (tablas 2×2) sin condicionar en las frecuencias marginales que no estaban fijadas. Esto originó una controversia, que se extiende hasta nuestros días, sobre la conveniencia de utilizar un test condicionado o un test incondicionado en tablas 2×2 . En esta controversia (que en realidad no se circunscribe únicamente al ámbito de las tablas 2×2) han participado una gran cantidad de autores y se han

propuesto diferentes versiones de estos test en tablas 2x2, (Pearson (1947), Berkson (1978), Barnard (1982), Upton (1982 y 1992), Yates (1984), Little (1989), Cormack and Mantel (1991), Greenland (1991), Martin Andres (1991), Silva (1993), Berger and Boos (1994), Helland (1995).....). Sin embargo en el caso general de tablas rxc esta controversia no se ha presentado, el argumento condicionado es casi exclusivamente el que permanece, pues aparte de consideraciones que también pueden ser utilizadas en el caso 2x2, tales como: mucho menor tiempo de cómputo, no aparecen parámetros perturbadores, los marginales contienen poca información sobre el problema (parámetros) de interés (Yates 1984, Lloyd 1988, Haber 1989), se pueden señalar otras más específicas, y es que cuando r y/o c aumentan, el número de posibles valores en las distribuciones exactas condicionales tiende a aumentar; Por tanto, el argumento de que los tests tipo Fisher son conservativos para testear a niveles de significancia fijados pierde su fuerza. De hecho Mehta and Hilton (1993) han probado que hay muy poca diferencia entre ambas metodologías (condicionada e incondicionada) para tablas mas grandes de 2x2. Así en esta tesis adoptaremos el principio condicionado en tablas rxc, trabajando fundamentalmente con tablas 2xc.

Una tabla X de contingencia $r \times c$ con valores casillas x_{ij} , totales filas R_i , totales columnas C_j , $i=1,2,\dots,r$, $j=1,2,\dots,c$, y total de la tabla N, puede surgir esencialmente a partir de cuatro tipos de muestreo:

- 1) Al considerar c distribuciones multinomiales, cada una con r casillas. En tal caso los totales columnas están fijados. Este modelo es llamado modelo multinomial producto.
- 2) Cuando solamente la suma de todas las entradas de la tabla están fijadas. En este caso el modelo es llamado modelo multinomial completo.

- 3) Ni el total ni los marginales fijados; Muestreo de Poisson.
- 4) Los dos marginales fijados; Muestreo hipergeométrico.

Para el modelo multinomial producto la hipótesis de interés es la homogeneidad.

Para el modelo multinomial completo, y también para el modelo de Poisson, la hipótesis de interés es la independencia.

Para el modelo hipergeométrico simplemente la aleatoriedad.

Para los tres primeros esquemas muestrales anteriores, y bajo las hipótesis nulas señaladas, la distribución exacta de X contiene parámetros desconocidos, parámetros “perturbadores”, estos parámetros son eliminados condicionando sobre sus estadísticos suficientes, los totales marginales. Con lo que la distribución exacta de X , para cualquiera de los esquemas muestrales de que procede, es lo que podríamos llamar una probabilidad hipergeométrica multivariante múltiple definida por:

$$P(X) = \frac{\prod_{i=1}^r R_i! \prod_{j=1}^c C_j!}{N! \prod_{i,j} x_{ij}!} \quad , \quad i=1,2,\dots,r \ ; \ j=1,2,\dots,c. \quad (1)$$

Que ya no contiene parámetros desconocidos y la inferencia exacta es posible.

El dominio de $P(X)$ es el conjunto $(r-1)(c-1)$ –dimensional de todas las tablas $r \times c$ con marginales filas R_i , $i=1,2,\dots,r$, y marginales columnas C_j , $j=1,2,\dots,c$. A este conjunto le llamamos familia isomarginal para dichos marginales. Si denotamos por

\mathbf{x} el vector aleatorio $(r-1)(c-1)$ con términos x_{ij} , $i=1,2,\dots,r-1$, $j=1,2,\dots,c-1$, entonces **(1)** es la función de probabilidad de este vector aleatorio.

Para tablas 2×2 **(1)** se simplifica a la hipergeométrica simple, y para tablas $2 \times c$ a la hipergeométrica multivariante.

Con este enfoque, el procedimiento usual para resolver un test exacto sería:

- Establecer un estadístico, o medida de discrepancia que permita ordenar las tablas de la familia.
- Determinar el valor del estadístico para la tabla .
- Enumerar todas las tablas posibles de la familia isomarginal.
- Computar el estadístico para cada tabla.
- Sumar las probabilidades de todas las tablas tan extremas al menos como la observada, con lo que tendremos el p-valor del test.

Una posible alternativa a la enumeración es el computar la función característica y después invertir esta (usando transformaciones rápidas de Fourier) para obtener las probabilidades excedentes.

A partir de aquí, podríamos decir que se abren dos líneas de investigación para examinar y mejorar test exactos condicionados para tablas de contingencia.

En la primera de ellas, enumeración, y usando distintas medidas de discrepancia, estadístico X^2 de Pearsón, estadístico razón de verosimilitud G^2 , Probabilidad Hipergeométrica P , son numerosas las investigaciones llevadas a efecto: Yates (1934 – X^2), Freeman and Halton (1951 - P), Bennett and Nakamura (1963 –

X^2 - G^2 - P), Lewontin and Felsenstein (1965 – X^2), March (1972 – P), Boulton(1974 – P), Hancock (1975 – P), Howell and Gordon (1976 – P), Agresti and Wackerly (1977 – X^2), Baker (1977 – X^2 – G^2 – P), Larntz (1978 – X^2 – G^2 – P), Verbeek and Kroonenberg (1979 – X^2), Cantor (1979 – P), Ngugen (1985 – Relaciones de concordancia), Pagano and Taylor-Halvorsen (1981 – P), Mehta and Patel (1980-1983-1986 – P con extensiones a X^2 y G^2), Joe (1985-1988 – P), Clarkson and Fan (1989-1993 – P), Shao (1997 – P). La atención en los últimos años, como puede apreciarse, se ha prestado a los test que usan como medida de discrepancia la Probabilidad hipergeométrica P, pues la elección de P origina algoritmos más rápidos, ya que P tiene que ser computado siempre y aporta ventajas en algunos puntos de la enumeración incompleta. Se acepta como mejor el test basado en el criterio de “menos probables”, conocido como test de Fisher, y como mejor método de resolverlo el de Mehta and Patel(1986), y a partir de aquí, los esfuerzos se centran en encontrar algoritmos que acorten aun más el tiempo de cómputo para la resolución del test (esto es, para el cálculo del p-valor).

La segunda línea de investigación para el análisis de tablas rxc se ha centrado en el uso de la transformación rápida de Fourier (FFT), Verbeek y Kroonenberg (1985) mencionaron la posibilidad de aplicar el método FFT general de Pagano y Trichler (1983) al análisis exacto de una tabla rxc. Una formulación completa del método FFT fue dado por Baglivo y otros (1992).

Realizado nuestro propio análisis de los métodos existentes, apreciando la eficiencia del algoritmo de Mehta and Patel, y la posibilidad de aportar sobre él nuestras propias mejoras, hemos optado por la primera línea. Ya iniciada nuestra investigación, esta opción se ha visto avalada por los estudios de Hirji y Johnson

(1996) que han realizado una comparación entre los dos métodos más desarrollados por una línea y por otra (M-P por la primera y Baglivo y otros por la segunda). Y, aparte de poner de manifiesto que ambos métodos tienen el mismo fundamento matemático, obtienen también resultados empíricos que muestran que el algoritmo de M-P es superior al algoritmo FFT con respecto a velocidad y exactitud de computación.

La idea de mejorar los métodos exactos sobre la base del mejor método existente, algoritmo de Mehta y Patel, es la que ha guiado el desarrollo de esta tesis, que se articula en varios capítulos. El capítulo I con un análisis general de los algoritmos de enumeración y específicamente de los algoritmos más notables de enumeración incompleta, así como de las propiedades matemáticas que los sustentan, con especial hincapié en el análisis del algoritmo de Mehta y Patel y de las mejoras aparecidas sobre este algoritmo, lo que nos permitirá concretar los objetivos de nuestro trabajo. El capítulo II, con una aportación personal sobre el máximo de la función de probabilidad de una hipergeométrica multivariante y su implementación en este tipo de algoritmos, y un estudio de la relación entre las probabilidades extremas en determinadas familias de distribuciones hipergeométricas multivariantes, lo que nos permitirá introducir importantes modificaciones en el algoritmo de red de Mehta and Patel que representan significativas mejoras en el algoritmo de resolución del test exacto de Fisher para tablas $2 \times c$. Y el capítulo III, con una generalización a tablas $r \times c$ de las metodologías desarrolladas en el capítulo II para tablas $2 \times c$.

**CAPÍTULO I: Algoritmos de enumeración para
resolver el Test exacto de Fisher. Algoritmos más notables de
enumeración incompleta.**

I.1: Algoritmos de enumeración para resolver el test exacto de Fisher.

Dada una tabla de contingencia $r \times c$ observada

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1c} \\ X_{21} & X_{22} & \dots & X_{2c} \\ \dots & \dots & \dots & \dots \\ X_{r1} & X_{r2} & \dots & X_{rc} \end{bmatrix} \quad (1.1)$$

con entradas no negativas x_{ij} , R_i la suma de entradas en la fila i , C_j la suma de entradas en la columna j , N la suma total de elementos de la tabla. Definimos el conjunto referencial de todas las posibles tablas con los mismos totales marginales que la dada (familia isomarginal) como

$$\Omega = \{ Y: Y \text{ es } r \times c, \sum_{j=1}^c y_{ij} = R_i, i=1, \dots, r, \sum_{i=1}^r y_{ij} = C_j, j=1, \dots, c \} \quad (1.2)$$

Para calcular en el test exacto de Fisher el p-valor de la tabla X dada, es preciso calcular la suma

$$P_F = \sum_Y P(Y) \quad (1.3)$$

Donde el sumatorio es sobre todas las tablas $Y \in \Omega$ que verifican $P(Y) \leq P(X)$ y donde

$$P(Y) = \frac{\{\prod_{j=1}^c C_j!\} \{\prod_{i=1}^r R_i!\}}{N! \prod_{i=1}^r \prod_{j=1}^c y_{ij}!} \quad \forall Y \in \Omega \quad (1.4)$$

Para computar este p-valor de forma exacta han sido elaborados una variedad de algoritmos que requieren enumeración total o parcial del conjunto referencial (1.2), y en aquellos casos en que las dimensiones de la tabla imposibilitan la

computación exacta se han conseguido estimaciones de este p-valor por el método de Monte Carlo.

La guía para la construcción de estos algoritmos es en primer lugar una concepción adecuada de una tabla de contingencia y en segundo lugar, para mejor conocimiento de la estructura de la familia isomarginal y de las relaciones que se pueden establecer entre sus miembros para simplificar los procedimientos implicados en la resolución de los algoritmos, una representación apropiada de (1.2).

Una tabla de contingencia tiene una concepción trivial como una tabla con r filas y c columnas con $\delta = (r-1) \times (c-1)$ grados de libertad o con δ casillas libres en las primeras $r-1$ filas y $c-1$ columnas. Las $r+c-1$ casillas bordes, esto es las adyacentes a los marginales, quedan fijadas dando los valores de las casillas libres y los valores marginales, y la última casilla libre está situada en la casilla sur-este. Una concepción menos trivial es la forma vectorizada: todas las casillas se disponen en un vector columna $t = (t_f, t_b)$ donde t_f es el vector de las casillas libres y t_b el de las casillas bordes, $t_f = (t_{11}, t_{21}, \dots, t_{12}, t_{22}, \dots, t_{(r-1), (c-1)})$.

Para la familia isomarginal se utilizan (Verbeek and Kroonenberg, 1985) tres tipos de representación: Arbol, Red o puntos de un subconjunto convexo.

En la representación de árbol la familia isomarginal se contempla como el conjunto de todas las hojas de un árbol dirigido, donde cada nivel del árbol corresponde a una casilla libre. En cada uno de los niveles tenemos solamente una tabla llena parcialmente. Las casillas de un segmento inicial de la tabla en forma vectorizada tienen valores, las casillas siguientes están en blanco. En la raíz la tabla entera está en blanco. En las hojas tenemos las tablas llenas por completo. Por tanto cada hoja corresponde a una tabla generada, es decir a un miembro de la familia isomarginal.

En la representación de red, que detallaremos con mayor profundidad al analizar el algoritmo reticulado de Mehta and Patel, la familia isomarginal se contempla como la enumeración de todos los posibles caminos a través de la red, desde el nodo inicial al nodo final. La red es un grafo dirigido con una fuente universal (representando una tabla vacía) y un sumidero universal (representando cualquier tabla llena). Este grafo no contiene ciclos (dirigidos). Todo camino maximal (dirigido) comienza en la fuente, termina en el sumidero, y tiene el mismo número de nodos (= número de columnas mas uno). La familia isomarginal se representa por estos caminos maximales; cada uno de los nodos a la distancia k del sumidero está en correspondencia uno-a-uno con los posibles pares de marginales de la subtabla formada por las k columnas sin rellenar. Cada avance desde un nodo a la distancia k a un nodo a la distancia $k-1$ corresponde a una posible columna (esta correspondencia entre nodos y posibles columnas es muchos-a-uno).

En la tercera representación la familia isomarginal se contempla como un subconjunto convexo de la rejilla Z_0^d en R^d con $d = rxc$ o en R^δ con $\delta = (r-1)x(c-1)$ donde $\{Z_0 = 0, 1, 2, \dots\}$ es el conjunto de enteros no negativos. Cualquier miembro de la familia isomarginal está en correspondencia uno a uno con un punto en Z_0^d . Representando por m el vector de los marginales columnas y filas $m = (C_1, \dots, C_c, R_1, \dots, R_r)$, el vector t_b de las casillas bordes es una función lineal de t_f y m , $t_b = -At_f + Bm$ (todos los elementos de las matrices A y B son cero o uno). Por tanto la familia isomarginal corresponde a $\{t / t_f \in Z_0^\delta \text{ y } t_b = -At_f + Bm \geq 0\}$, evidentemente esto es un subconjunto convexo de la rejilla Z_0^d .

En los primeros algoritmos de enumeración diseñados, Freeman and Halton (1951), March (1972), se necesita para computar el p-valor de la tabla (1.1) enumerar todas las tablas de (1.2). Utilizan la concepción trivial señalada para una tabla de

contingencia, y el interés principal se centra en el algoritmo de generación de las tablas con los mismos marginales que la dada y con valores enteros no negativos en las casillas. Concretamente, el algoritmo de March asigna a cada una de las $(r-1) \times (c-1)$ casillas el mínimo de sus correspondientes totales marginales filas y columnas que es el número máximo posible para las casillas. Estas casillas se varían a continuación en todas las posibles combinaciones con cada casilla variada entre su máximo y cero. Las casillas de los bordes se completan restando los valores generados para las casillas de los totales marginales. Se obtienen de esta forma algunas soluciones que no son válidas pues para respetar los totales marginales algunos de estos valores casillas deben ser negativos. Se generan no solo todas las tablas de la familia sino algunas tablas más, este problema de sobrecompletitud aparece ya eliminado en los algoritmos de Howell and Gordon (1976) y Cantor (1979) al utilizar representaciones no triviales para las tablas, pero se sigue necesitando de la enumeración de todas las tablas de la familia isomarginal, y esta enumeración podría no ser dificultosa si el conjunto no fuera demasiado grande, ahora bien, según Gail and Mantel (1977) el número de tablas aumenta con el tamaño muestral y aumenta exponencialmente con el número de grados de libertad, con lo que la enumeración completa es, a menudo, imposible, incluso con los ordenadores más rápidos.

Es posible abreviar la determinación del p-valor de una tabla, y en consecuencia desarrollar algoritmos para el test exacto de Fisher más rápidos y eficaces, sirviéndose de herramientas matemáticas, o sea propiedades matemáticas relativas a la función de probabilidad de Y que permiten alcanzar el objetivo con enumeración incompleta de las tablas implicadas en el proceso.

En esencia, aunque como veremos con distintos matices, hay dos tipos de propiedades usadas comúnmente en las metodologías existentes, las que hacen referencia a:

- 1) Descomposición de $P(Y)$ en un producto de factores que puedan ser calculados recursivamente.
- 2) Posibilidad de expresar la suma de las probabilidades de todas las tablas que tienen un segmento (conjunto de elementos: casillas individuales, columnas, filas,...) inicial común, como la probabilidad correspondiente a ese segmento, resultado de un proceso de colapsación sobre los restantes elementos. En el caso de que el segmento inicial común fueran las c' primeras columnas la suma de las probabilidades de todas las tablas rx de la familia isomarginal con unos valores comunes dados en las primeras c' columnas es la probabilidad de la tabla $rx(c'+1)$ obtenida al colapsar las columnas desde las $c'+1$ a la c en una sola.

Aparte de estas propiedades Mehta y Patel (1983) introducen además propiedades referentes al máximo y al mínimo de $P(Y)$, y complementan el trabajo elaborando la forma computacional del algoritmo (1986).

Las mejoras de Harry Joe (1985,1988) se orientan precisamente a establecer nuevas propiedades referentes al máximo y al mínimo de $P(Y)$ que pueden ser útiles en aquellas situaciones no solucionadas satisfactoriamente por los teoremas de Mehta y Patel, y que sirven para mejorar dicho algoritmo.

Clarkson y Yuan-an Fan (1989,1993) se centran en las mejoras computacionales del algoritmo, verificando, entre otras cosas, como las propiedades de Harry Joe son realmente eficaces al implementarlas en el algoritmo reticulado de M-P.

Por supuesto, en la elaboración de las propiedades anteriores y en otras situaciones en que se hace necesario, se utiliza asimismo el mayor conocimiento matemático de la hipergeométrica simple, en especial las propiedades referentes al cálculo recursivo de esta función y a su máximo.

Para tipos de tablas especiales también se han tratado de incorporar algunos acortamientos específicos. Saunders (1984) indicó un camino para evitar el enumerar un número de tablas en el caso especial que dos o más marginales filas (columnas) sean iguales. Más recientemente Shao (1997) ha vuelto a incidir sobre esta cuestión. No obstante este tema de acortamientos del p-valor para tablas especiales está tan abierto en la actualidad que merecería por si solo una investigación completa y nosotros haremos nuestro estudio sobre tablas generales.

En determinadas condiciones, dimensiones muy elevadas de la tabla, dispersión de sus valores, para las que la computación del p-valor de forma exacta se hace prácticamente imposible incluso a través de los algoritmos de enumeración incompleta, un buen recurso es estimarlo usando muestreo Monte Carlo de tablas en el conjunto (1.2), simulando muestreo de una hipergeométrica (Boyett (1979), Patefield (1981), StatXact (1991)). Cada tabla de la muestra proporciona una variable aleatoria Bernoulli, que indica si la discrepancia es por lo menos tan grande como la de la tabla observada. El p-valor estimado es la media muestral de esas variables aleatorias Bernoulli, esta media muestral es la proporción (\hat{p}) de las tablas de la muestra que tienen discrepancias al menos tan grande como la de la observada. La precisión de la estimación esta determinada por la varianza muestral estimada $\hat{p}(1 - \hat{p})/n$ donde n es el número de tablas de la muestra. Con muestras de 17.000 tablas se ha conseguido un error en la estimación menor de 0.01 con una confianza de 0.99. Una ventaja del método Monte Carlo es que el entramado de la computación

depende mucho menos del total y de las dimensiones de la tabla que para métodos de análisis exacto. Aunque se tarda mas en generar cada una de las tablas con este método, solamente se necesita generar un número relativamente pequeño.

Por último cabe señalar que otra alternativa al test exacto de Fisher antes de recurrir a métodos asintóticos puede ser la elaboración de un test híbrido, que mezcla teoría exacta y asintótica en un único camino (Baglivo, Olivier and Pagano (1985, 1988), Mehta and Patel (1986 a). Si en los algoritmos exactos más perfeccionados para realizar los acortamientos se realiza un test de máximo y caso de que este no nos permita resolver se realiza un test de mínimo, una tercera opción seria ver si se dan las condiciones de Cochran y resolver por aproximación asintótica. No obstante en esta metodología mixta, al igual que en los métodos puramente asintóticos falta una guía que nos indique la adecuación de los p-valores obtenidos a los p-valores exactos.

I.2: Algoritmos más Notables de Enumeración Incompleta.

I.2.1.-Algoritmo de Marcelo Pagano y Katherine Taylor Halvorsen.

Base Teórica

El algoritmo de P-H descansa en dos teoremas:

Teorema 1.1.- $P(Y)$, dado en (1.4), se puede reescribir como un producto de probabilidades hipergeométricas:

$$P(Y) = \prod_{j=1}^{c-1} \prod_{i=1}^{r-1} p_{ij} \quad (1.5)$$

Donde

$$p_{ij} = \frac{\binom{C_j - s_{ij}}{y_{ij}} \binom{N - v_{ij} - C_j + s_{ij}}{R_i - t_{ij} - y_{ij}}}{\binom{N - v_{ij}}{R_i - t_{ij}}},$$

$$s_{ij} = \sum_{g=1}^{i-1} y_{gj} \quad t_{ij} = \sum_{h=1}^{j-1} y_{ih} \quad (1.6)$$

$$v_{ij} = \sum_{g=1}^{i-1} R_g + \sum_{h=1}^{j-1} C_h - \sum_{g=1}^{i-1} \sum_{h=1}^{j-1} y_{gh}$$

y se usa la convención de que un sumatorio es nulo si su límite superior es inferior que su límite inferior. p_{ij} es la probabilidad condicional de obtener y_{ij} en la casilla (i,j)-ésima dados los valores observados en todas las casillas previas, donde las casillas previas son y_{kl} para $1 \leq k \leq r$ y $1 \leq l \leq j-1$ y y_{kj} para $1 \leq k \leq i-1$.

Este teorema resulta de mucha utilidad para la construcción del algoritmo de Pagano and Halvorsen pues la utilización de (1.5) en lugar de (1.4) le aporta numerosas ventajas, aparte de que los p_{ij} son fáciles de calcular recursivamente y todos están entre cero y uno, se puede sacar partido de la unimodalidad de la distribución hipergeométrica, le facilita encontrar un método de generación de tablas que al ser implementado en el algoritmo hace innecesario el generar la familia completa. A partir de (1.6) la cota superior, u_{ij} , e inferior, l_{ij} , para cada y_{ij} puede expresarse mediante las cantidades:

$$\text{cota inferior (i,j)} = u_{ij} = \max(0, v_{ij} + C_j - s_{ij} + R_i - t_{ij} - N)$$

$$\text{cota superior (i,j)} = l_{ij} = \min(C_j - s_{ij}, R_i - t_{ij})$$

v_{ij} , s_{ij} , t_{ij} , expresados en términos de las casillas previas. Con lo que realizando una linealización de subíndices a través de la correspondencia

$$y_{ij} = z_k \quad ; \quad p_{ij} = p_k \quad ; \quad u_{ij} = u_k \quad ; \quad l_{ij} = l_k$$

donde $k = (j-1)(r-1) + i$, generar una tabla equivale a generar un vector \mathbf{z} donde la componente k -ésima de este vector puede variar entre l_k y u_k . Cada vector $\mathbf{z} = (z_1, z_2, \dots, z_m)$, donde $m = (r-1)(c-1)$, origina una tabla en Ω (familia isomarginal) si la fila y la columna últimas se completan de tal forma que satisfagan las restricciones de los marginales filas y columnas, y (1.5) se convierte en

$$P(Y) = \prod_{k=1}^m p_k \quad (1.7)$$

La forma (1.7) de expresar $P(Y)$, y el tomar como guía una representación en árbol de la familia isomarginal permiten a Pagano and Halvorsen intuir herramientas de acortamiento en el cálculo del p-valor de la tabla X que quedan sustentadas formalmente en el siguiente teorema.

Teorema 1.2.- Supongamos que para algún $Y \in \Omega$ existe un $d < m$ tal que

$$\prod_{k=1}^d p_k < P(X) \quad (1.8)$$

y los correspondientes \mathbf{z} son z_1, z_2, \dots, z_d . Denotemos por D :

$$D = \{Y: Y \in \Omega \text{ y las primeras } d \text{ componentes de } Y \text{ son } z_1, z_2, \dots, z_d\}.$$

Entonces
$$\sum_{Y \in D} P(Y) = \prod_{k=1}^d p_k$$

y D es un subconjunto de la región crítica.

El testear (1.8) sobre vectores \mathbf{z} , con sus d primeras componentes comunes y determinadas, z_1, \dots, z_d , constituye la herramienta de acortamiento del algoritmo. Cuanto menor sea el valor de d para el que (1.8) sea cierta mayor serán los acortamientos que se producirán en la computación del p-valor. Con referencia a la representación en árbol, d representa el nivel del árbol hasta el que se tiene que extender el proceso, el nivel hasta el que se van a llenar las casillas de las tablas.

Implementación de las propiedades anteriores en la construcción del algoritmo de P-H

La linealización de los elementos de las tablas y consiguiente identificación de cada tabla Y con un vector $\mathbf{z} = (z_1, z_2, \dots, z_m)$ convierte el propósito inicial en identificar y sumar las probabilidades de todos los vectores \mathbf{z} tales que $P(\mathbf{z}) \leq P(X)$. Esta selección se llevará a cabo por niveles, o sea sobre el conjunto de todos los vectores \mathbf{z} que presentan las d primeras componentes comunes e iguales a unos valores determinados z_1, \dots, z_d , entonces $\prod_{k=1}^d p_k$ denota la parte de la probabilidad de estos vectores \mathbf{z} correspondiente a las d primeras componentes que son comunes.

Aunque el resto de las componentes sean diferentes podemos evitar su enumeración explícita si

$$\prod_{k=1}^d p_k \leq P(X)$$

pues por el *teorema 1.2* sabemos que la suma de las probabilidades de todos estos vectores sería, precisamente, $\prod_{k=1}^d p_k$. Si esta condición no se satisface, generamos una nueva componente del vector y discutimos de nuevo la condición.

Por consiguiente para computar el p-valor se comienza con el análisis de las tablas en el nivel 1, o sea se decide para cada una de las entradas posibles en la primera componente del vector \mathbf{z} , si (1.8) es cierta o no para $d=1$, en caso afirmativo esta entrada no necesitaría extensiones al nivel siguiente, y el valor del producto, $\prod_{k=1}^d p_k$, se incorporaría al p-valor. En otro caso para esa entrada se hace una extensión al nivel 2, o sea se incorpora otro valor a la componente siguiente y se

vuelve a testear la condición (1.8), repitiéndose el proceso anterior. Si hay que llegar al nivel m no se produciría ningún acortamiento pues no estaríamos determinando suma de grupos de tablas sino de tablas individuales.

I.2.2.-Algoritmo de Mehta y Patel

Introducción

Este algoritmo descansa también sobre una serie de propiedades relativas a la función de probabilidad que tienen validez propia independientemente del uso que se les dé. Pero, como el propio Mehta señala en sus comentarios al resumen de Agresti (1992) sobre Inferencia Exacta, fueron ideadas bajo el objetivo de encontrar algoritmos eficaces para ejecutar el test exacto de Fisher en tablas de contingencia $r \times c$, y quizás ni ellas, ni su implementación adecuada en un algoritmo tal, hubiesen sido realidad sin una representación geométrica de red de la familia isomarginal. Por ello, en el análisis de este algoritmo, para seguir con la misma metodología trazada en el algoritmo anterior, que posibilita la comparación de ambos procesos, estudiaremos también las propiedades que le sustentan y su forma de implementación, pero terminaremos además incidiendo en la construcción de la red que guió la elaboración de estas tareas y analizando la forma de computación de este algoritmo de red.

Una comparación con el algoritmo de P-H nos permitirá apreciar de una forma más clara y explícita los puntos de conexión existentes entre los dos algoritmos, así como los avances aportados por el algoritmo de M-P con respecto al de P-H.

Base teórica

La probabilidad de cualquier $Y \in \Omega$ se expresa en la forma

$$P(Y) = \left(\prod_{j=1}^c \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right) D^{-1}$$

donde D se define como

$$D = \frac{N!}{R_1! R_2! \dots R_r!}$$

y surgen dos tipos de propiedades ligadas a esta nueva forma de expresión, el primero

referente a $\sum \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$ siendo la sumatoria sobre todas las subtablas $r \times k$ de

marginales filas dados $R_{1k}, R_{2k}, \dots, R_{rk}$, y marginales columnas dados C_1, C_2, \dots, C_k . Y el

segundo tipo de propiedades referente al máximo y mínimo ($Y \in \Omega$) de $\prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$.

Propiedad 1.1.-

$\sum \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$ siendo la sumatoria sobre todas las subtablas $r \times k$ de

marginales filas R_{1k}, \dots, R_{rk} y columnas C_1, \dots, C_k está dado por

$$\frac{(C_1 + C_2 + \dots + C_k)!}{R_{1k}! R_{2k}! \dots R_{rk}!}$$

El segundo tipo de propiedades queda expresado por un grupo de teoremas dados por M-P(1983), que nos proporcionan condiciones de probabilidad máxima y mínima para una familia isomarginal a través del máximo y del mínimo de

$$FO = \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!} \quad (1.9)$$

Los teorema 1.3, teorema 1.4, y teorema 1.5 de M-P asumen que las tablas $r \times k$ en el conjunto referencial sobre el que se desea buscar un máximo y un mínimo tienen totales columnas iguales, total comun denotado por C.

El teorema 1.6 de M-P muestra que, ya que la situación anterior no es general, se puede cambiar el conjunto referencial para contener tablas $(r+1) \times k$ con totales columnas iguales.

El problema de determinar el máximo de FO para $C_j = C$ es sencillo, el teorema 1.3 de M-P proporciona una expresión determinada para el máximo.

Teorema 1.3 de M-P.-

El máximo de (1.9) sujeto a

$$\sum_{j=1}^k y_{ij} = R_{ik}, \quad i = 1, 2, \dots, r \quad (1.10)$$

$$\sum_{i=1}^r y_{ij} = C, \quad j = 1, 2, \dots, k \quad (1.11)$$

$$y_{ij} \geq 0 \text{ y enteros para todo } i, j. \quad (1.12)$$

está dado por

$$\left(\frac{C!}{d_1! d_2! \dots d_r!} \right)^k / \left((d_1 + 1)^{h_1} (d_2 + 1)^{h_2} \dots (d_r + 1)^{h_r} \right) \quad (1.13)$$

donde $d_i = \lfloor \frac{R_{ik}}{k} \rfloor$ ($\lfloor x \rfloor$ denota el mayor entero menor o igual que x), y $h_i = R_{ik} - k \cdot d_i$.

El mínimo de FO es más difícil de obtener. Se debe minimizar la probabilidad hipergeométrica de una matriz $r \times k$, (una función cóncava) sujeta a totales filas fijados y totales columnas fijados pero iguales. Los teoremas 1.4 y 1.5 de M-P reducen el problema a enumerar las soluciones básicas posibles (aquellas que tienen los ceros suficientes para que el resto de las incógnitas estén determinadas) de un sistema de $r \times k'$ ecuaciones lineales en k' variables, donde $k' < r$.

La enumeración se simplifica considerablemente utilizando el hecho de que los marginales columnas son todos iguales en la matriz original $r \times k$.

En la demostración de los teoremas se utiliza el concepto de ciclo, entendiéndose que una matriz cualquiera A , $r \times k$, con entradas a_{ij} en la fila i y columna j contiene un ciclo si existen p entradas distintas de cero $a_{i_1 j_1}, a_{i_2 j_2}, \dots, a_{i_p j_p}$ tales que p es número par y $j_1 = i_2, i_2 = j_3, j_3 = i_4, i_4 = j_5, \dots, i_p = i_1$.

La presencia de un ciclo en A implica que es posible el comenzar en una casilla no nula (i_1, j_1) y moviéndose alternativamente vertical y horizontalmente a otras casillas no nulas, regresar a la casilla inicial.

Teorema 1.4 de M-P.-

Sea $A = \{a_{ij}, i=1,2,\dots,r ; j=1,2,\dots,k\}$ la solución al problema minimizar (1.9) sujeto a las restricciones (1.10), (1.11) y (1.12). Entonces las entradas no nulas de A no pueden formar un ciclo.

Como consecuencia de este teorema, y ya que A no contiene ciclos, es bien conocido (Hadley 1962) que A debe ser alguna de las soluciones básicas posibles de las restricciones (1.10) y (1.11). Esta caracterización nos permite el minimizar (1.9) examinando sistemáticamente todas las soluciones básicas posibles de (1.10) y (1.11). Pero este proceso se puede simplificar, el teorema 1.5 de M-P nos permite reducir considerablemente el número de soluciones básicas posibles para ser examinadas, veamos en qué consiste.

Teorema 1.5 de M-P.-

Si $R_{1k} \geq C$, el mínimo valor de FO sujeto a las restricciones (1.10), (1.11) y (1.12) es igual al mínimo valor de la función objetivo del siguiente problema:

$$\text{Minimizar } \prod_{j=1}^{k-1} \left(\frac{C!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right) \quad (1.14)$$

sujeto a (1.12) y

$$\sum_{j=1}^{k-1} y_{ij} = R_{ik}, \quad i=2,3,\dots,r \quad (1.15)$$

$$\sum_{i=1}^r y_{ij} = C, \quad j=1,2,\dots,k-1 \quad (1.16)$$

$$\sum_{j=1}^{k-1} y_{1j} = R_{1k} - C \quad (1.17)$$

Este teorema es útil siempre que $k \geq r$. Pues en esta situación se demuestra de forma casi inmediata que por lo menos una fila, digamos i_0 , tiene $R_{i_0 k} \geq C$. Y se puede aplicar repetidamente, eliminando cada vez r variables del problema de optimización resultante, hasta que no haya totales filas que excedan a C . Esto origina el siguiente corolario:

Corolario 1.1.- El mínimo de FO sujeto a las restricciones (1.10), (1.11) y (1.12) es igual al mínimo valor de la función objetivo del siguiente problema:

$$\text{- Minimizar } \prod_{j=1}^{k'} \left(\frac{C!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right) \quad (1.18)$$

sujeto a (1.12) y

$$\sum_{j=1}^{k'} y_{ij} = R_{ik} - \left[\frac{R_{ik}}{C} \right] \cdot C; \quad i=1,2,\dots,r \quad (1.19)$$

$$\sum_{i=1}^r y_{ij} = C ; j=1,2,\dots,k' \quad (1.20)$$

donde $k'=k - \sum_{i=1}^r \left[\frac{R_{ik}}{C} \right]$.

Con lo que se ha reducido el problema de determinar el mínimo de la FO original a un problema de enumerar todas las soluciones básicas posibles de las restricciones (1.19) y (1.20), y seleccionar las que minimizan (1.18). Además, cada una de las soluciones básicas posibles se obtienen fácilmente resolviendo un sistema triangular de ecuaciones lineales.

Los teoremas 1.3, 1.4, y 1.5 de M-P nos permiten computar el mínimo y el máximo de la función objetivo cuando todos los totales columna son iguales a C, el teorema 1.6 de M-P, como hemos señalado anteriormente, nos permite, cuando no se cumple esta restricción, el usar estos teoremas para obtener una cota inferior para el mínimo, y una cota superior para el máximo; Consiste en lo siguiente:

Teorema 1.6 de M-P.-

Sean SP y LP los valores mínimos y máximos, respectivamente, de

$$OF = \prod_{j=1}^k \left(\frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right) \quad (1.21) \quad \text{donde los } y_{ij} \text{ verifican las restricciones}$$

(1.12), y

$$\sum_{i=1}^r y_{ij} = C_j ; j=1,2,\dots,k \quad (1.22)$$

$$\sum_{j=1}^k y_{ij} = R_{ik} ; i=1,2,\dots,r \quad (1.23)$$

y SP' y LP' los valores mínimo y máximo, respectivamente, de

$$OF' = \prod_{j=1}^k \left(\frac{C!}{z_{1j}! z_{2j}! \dots z_{r+1j}!} \right) \quad (1.24)$$

donde los z_{ij} verifican las restricciones (1.12), y

$$\sum_{i=1}^{r+1} z_{ij} = C \quad ; j=1,2,\dots,k \quad (1.25)$$

$$\sum_{j=1}^k z_{ij} = R_{ik} \quad ; i=1,2,\dots,r \quad (1.26)$$

$$\sum_{j=1}^k z_{r+1,j} = kC - \sum_{i=1}^r R_{ik} \quad (1.27)$$

y $C = \max(C_1, C_2, \dots, C_k)$.

Entonces

$$SP \geq \frac{SP'}{\prod_{j=1}^k \binom{C}{C_j}} \quad (1.28)$$

y

$$LP \leq \frac{LP'}{\prod_{j=1}^k \binom{C}{C_j}} \quad (1.29)$$

Por tanto, podemos computar SP' y LP' con la ayuda de los teoremas 1.3, 1.4, y 1.5, y siempre podemos acotar SP y LP cuando los C_j no son iguales.

Implementación de estas propiedades en el algoritmo de determinación del p-valor

La forma de expresión dada por M-P para la función de probabilidad, permite, dejando aparte el factor D^{-1} , la descomposición de $P(Y)$, dado k , $k \in \{1, 2, \dots, c\}$ como:

$$\left(\prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right) \left(\prod_{j=k+1}^c \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!} \right)$$

con lo que sí se denota por Q el conjunto de todas las tablas de la familia Ω que ofrecen entradas idénticas y determinadas desde la columna c hasta la columna $k+1$, tendremos, para todas las tablas de Q :

1.- Será común el segundo factor de la expresión anterior.

2.- Todas las entradas en las k primeras columnas tendrán los mismos totales marginales filas. Luego, si partiendo de $(R_{1c}, R_{2c}, \dots, R_{rc}) \equiv (R_1, R_2, \dots, R_r)$, generamos de forma conveniente estos totales marginales filas que podemos denotar por $(R_{1k}, R_{2k}, \dots, R_{rk})$, podemos utilizar los teoremas 1.3, 1.4, 1.5, y 1.6 de M-P para determinar el máximo y el mínimo del primer factor.

Con lo que denotando

$$PAST = \prod_{j=k+1}^c \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$$

$$SP(k; R_{1k}, \dots, R_{rk}) = \text{mínimo de } \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$$

$$LP(k; R_{1k}, \dots, R_{rk}) = \text{máximo de } \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$$

Lo único que haría falta para resolver el algoritmo de determinación del p -valor de una tabla X , $X = (x_{ij})_{i=1, \dots, r, j=1, \dots, c}$, sería testear sí, o bien

$$PAST \cdot LP \leq \prod_{j=1}^c \frac{C_j!}{x_{1j}! x_{2j}! \dots x_{rj}!} \quad (1.30)$$

o

$$PAST \cdot SP > \prod_{j=1}^c \frac{C_j!}{x_{1j}! x_{2j}! \dots x_{rj}!} \quad (1.31)$$

Si se verifica la desigualdad (1.30), la probabilidad de cualquier tabla de Q será menor o igual que $P(X)$. Por consiguiente, las probabilidades de todas estas tablas contribuyen al p -valor, pero, por la propiedad 1.1 de M-P,

$$\sum_{j=1}^k \prod \frac{C_j!}{y_{1j}!y_{2j}!\dots y_{kj}!} = \frac{(C_1 + C_2 + \dots + C_k)!}{R_{1k}!R_{2k}!\dots R_{rk}!}$$

Por tanto la contribución al p-valor del conjunto de tablas de Q será

$$\sum_{Y \in Q} P(Y) = \frac{(C_1 + C_2 + \dots + C_k)!}{R_{1k}!R_{2k}!\dots R_{rk}!} \text{PAST} \cdot D^{-1} \quad (1.32)$$

con lo que no es necesaria la enumeración explícita de todas estas tablas. Y, si se satisface (1.31), la probabilidad de cualquier tabla de Q excedería a P(X), por consiguiente ninguna de estas tablas contribuiría al p-valor; No necesitarían más atención, y su enumeración explícita tampoco sería necesaria.

Si no se satisfacen ni (1.30) ni (1.31) entonces no pueden ser enumeradas implícitamente las tablas de Q, y habría que extender las entradas idénticas, y construir un nuevo conjunto Q' formado por todas las tablas de la familia Ω que ofrecen entradas idénticas desde la columna c hasta la columna k, y repetir, de nuevo, el proceso anterior.

Tarea básica del Algoritmo, como ha quedado claramente de manifiesto, es el computar LP y SP para cada $(k; R_{1k}, \dots, R_{rk})$.

Hemos analizado hasta aquí los elementos esenciales del algoritmo de M-P, sin hacer referencia a redes pero, como indicábamos al principio, y por las razones allí aludidas, vamos a completar el análisis de este algoritmo indicando la construcción de la red, la computación del algoritmo con esta base de red, y una breve comparación con el algoritmo de P-H.

Construcción de la red de M-P

M-P representan el conjunto Ω como una red de **nodos y arcos**, donde cada uno de los caminos a través de la red representa una, y solamente una, tabla $Y \in \Omega$, y

la longitud de este camino es $D \cdot P(Y)$. El problema original se reduce, por consiguiente, a identificar y sumar las longitudes de todos los caminos que no son más largos que $D \cdot P(X)$. La red se construye en $c+1$ estados sucesivos: $c, c-1, \dots, 0$. En cualquier estado k , existe un conjunto de nodos, cada uno de ellos identificado por un vector único (k, \mathbf{R}_k) donde $\mathbf{R}_k \equiv (R_{1k}, R_{2k}, \dots, R_{rk})$. De cada uno de los nodos en cada estado k emanan arcos, y cada arco se dirige exactamente a un nodo en el estado $k-1$. La red se define recursivamente especificando todos los nodos de la forma $(k-1, \mathbf{R}_{k-1})$, que suceden al nodo (k, \mathbf{R}_k) , y están conectados a él por arcos. El nodo (k, \mathbf{R}_k) es conocido como nodo madre, y los nodos $(k-1, \mathbf{R}_{k-1})$ como nodos hijos. El rango de las componentes del vector $\mathbf{R}_{k-1} \equiv (R_{1,k-1}, R_{2,k-1}, \dots, R_{r,k-1})$ para estos nodos hijos viene dado por

$$\begin{aligned} & \text{Máx } (0, R_{ik} - C_k + \sum_{\ell=1}^{i-1} (R_{\ell k} - R_{\ell, k-1})) \leq R_{i, k-1} \leq \\ & \leq \text{Mín } (R_{ik}, \sum_{\ell=1}^{k-1} C_{\ell} - \sum_{\ell=1}^{i-1} R_{\ell, k-1}) \end{aligned} \quad (1.33)$$

donde un sumatorio es nulo si su límite inferior excede a su límite superior. Hay sólo un nodo en el estado c , el nodo inicial, que se denota por (c, \mathbf{R}_c) , donde $\mathbf{R}_c \equiv (R_1, R_2, \dots, R_r)$. Al aplicar de manera recursiva la expresión anterior de generación de nodos a los estados $c, c-1, \dots, 1$, obtenemos exactamente un nodo en el estado 0, denotado por $(0, \mathbf{0})$. Este es el nodo terminal. La longitud de un arco que conecta el nodo (k, \mathbf{R}_k) al nodo $(k-1, \mathbf{R}_{k-1})$ es igual a

$$\frac{C_k!}{(R_{1k} - R_{1, k-1})! \dots (R_{rk} - R_{r, k-1})!}$$

Un camino completo a través del retículo se define como una sucesión de arcos conectados, dirigidos del nodo inicial al nodo terminal.

Un subcamino inicial es una sucesión de arcos conectados dirigidos desde el nodo inicial (c, \mathbf{R}_c) a algún nodo intermedio (k, \mathbf{R}_k) .

Un subcamino continuación es una sucesión de arcos conectados dirigidos desde cualquier nodo intermedio (k, \mathbf{R}_k) al nodo terminal $(0, \mathbf{0})$.

La longitud de cualquier camino (o subcamino) es el producto de las longitudes de sus arcos.

Se verifica de forma inmediata que para una red construida como acabamos de hacer:

1.- Cada camino completo de la forma

$$(c, \mathbf{R}_c) \rightarrow (c-1, \mathbf{R}_{c-1}) \rightarrow \dots \rightarrow (0, \mathbf{0})$$

corresponde a una tabla de contingencia $r \times c$, $Y \in \Omega$, donde

$$y_{ij} = R_{ij} - R_{i,j-1}; i=1, 2, \dots, r; j=1, 2, \dots, c$$

Además, la longitud del camino es igual a $D \cdot P(Y)$.

2.- Otra propiedad de utilidad de la construcción de red es el hecho de que cualquier subcamino continuación de la forma

$$(k, \mathbf{R}_k) \rightarrow (k-1, \mathbf{R}_{k-1}) \rightarrow \dots \rightarrow (0, \mathbf{0})$$

corresponde a una tabla Z de contingencia $r \times k$ con

$$z_{ij} = R_{ij} - R_{i,j-1}; i=1, 2, \dots, r; j=1, 2, \dots, k$$

Definiendo la probabilidad hipergeométrica para una tabla de contingencia $r \times k$ como

$$P_k(\mathbf{Z}) = D^{-1} \cdot \prod_{j=1}^k \frac{C_j!}{y_{1j}! y_{2j}! \dots y_{rj}!}$$

la longitud de este subcamino es $D \cdot P_k(\mathbf{Z})$.

3.- Finalmente, el conjunto de todos los posibles subcaminos-continuación desde un nodo intermedio (k, \mathbf{R}_k) al nodo terminal $(0, \mathbf{0})$ está en correspondencia uno a uno con el conjunto de todas las tablas de contingencia $r \times k$

$$\Omega(k, \mathbf{R}_k) = \{ \mathbf{Z}: \mathbf{Z} \text{ es } r \times k, \sum_{j=1}^k z_{ij} = R_{ik}, \sum_{i=1}^r z_{ij} = C_j \}$$

Un sencillo ejemplo nos puede servir para ilustrar esta construcción de la red. Supongamos que Ω es el conjunto de todas las tablas de contingencia, 3×3 , con $R_1 = 8$, $R_2 = 3$, $R_3 = 2$, y $C_1 = 6$, $C_2 = 5$, $C_3 = 2$. Los caminos de la red (Figura 1.1) están en correspondencia uno- a- uno con las tablas de Ω . Por ejemplo, el camino

$$(3; 8,3,2) \rightarrow (2; 8,2,1) \rightarrow (1; 5,0,1) \rightarrow (0; 0,0,0)$$

corresponde a la tabla
$$\begin{bmatrix} 5 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

La longitud del arco que conecta el nodo inicial $(3; 8,3,2)$ al nodo $(2; 8,2,1)$

es:
$$\frac{C_3!}{(R_{13} - R_{12})!(R_{23} - R_{22})!(R_{33} - R_{32})!} = 2$$

La longitud del arco que conecta el nodo $(2; 8,2,1)$ al nodo $(1; 5,0,1)$ es:

$$\frac{C_2!}{(R_{12} - R_{11})!(R_{22} - R_{21})!(R_{32} - R_{31})!} = 10$$

La longitud del arco que conecta el nodo $(1; 5,0,1)$ al nodo terminal $(0; 0,0,0)$

es:
$$\frac{C_1!}{(R_{11} - R_{10})!(R_{21} - R_{20})!(R_{31} - R_{30})!} = 6$$

Por consiguiente la longitud del camino reseñado es: $2 \times 10 \times 6 = 120$.

En el nodo inicial $(3; 8,3,2)$ todos los caminos son posibles, las tablas están en blanco, en el nodo $(2; 8,2,1)$ hay un subcamino recorrido, y subcaminos por

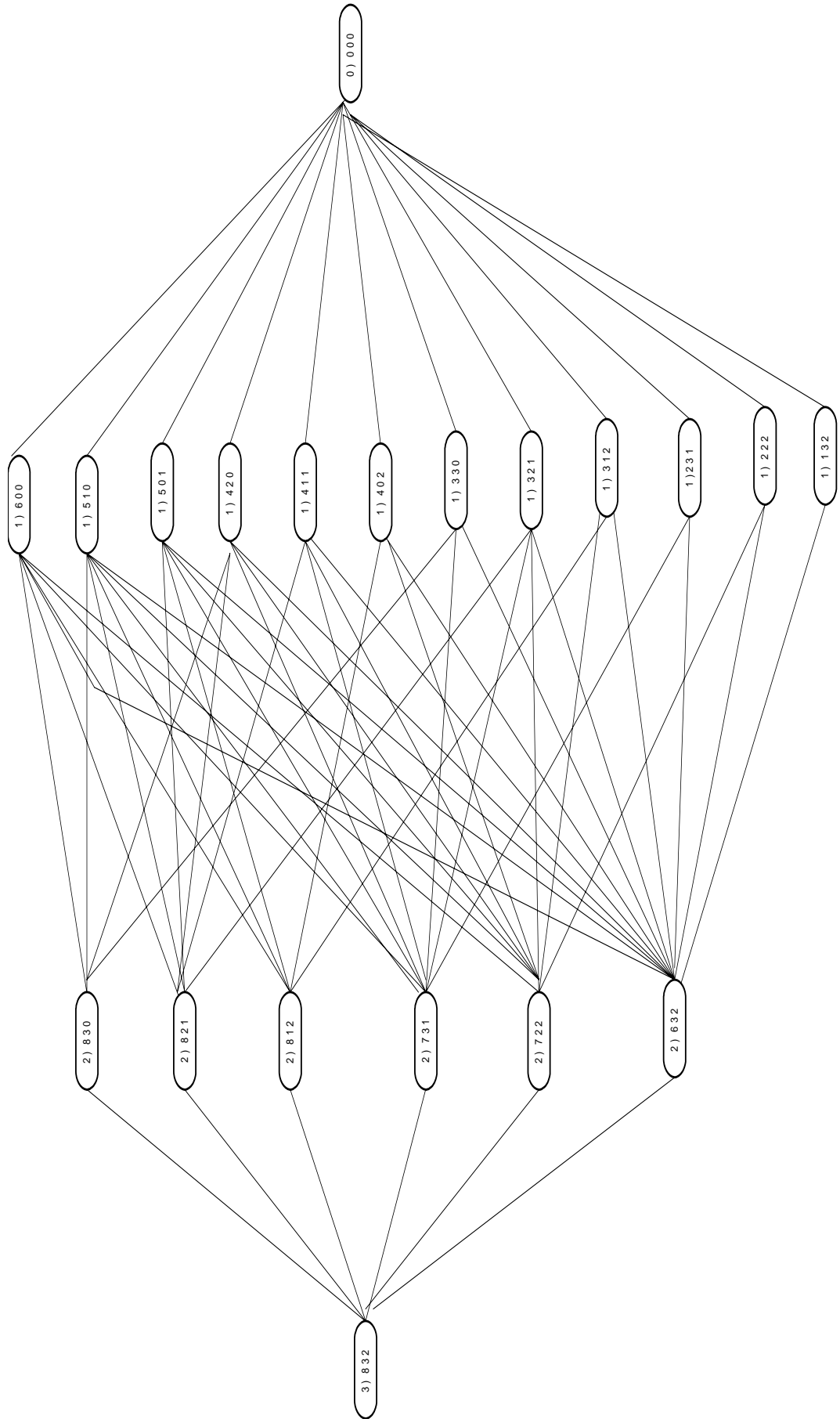


Figura 1.1

recorrer los que parten del nodo (2; 8,2,1), correspondería a todas las tablas con entradas determinadas en la columna 3, $y_{13}=(R_{13}-R_{12})=0$, $y_{23}=(R_{23}-R_{22})=1$, $y_{33}=(R_{33}-R_{32})=1$, y entradas por determinar en las columnas 2 y 1, que corresponderían a todas las subtablas 3x2 de marginales filas $R'_1=R_{12}=8$, $R'_2=R_{22}=2$, $R'_3=R_{32}=1$, y marginales columnas C_1 y C_2 ; siguiendo con este camino, en el nodo (1; 5,0,1) se habría ampliado el subcamino recorrido, correspondería a todas las tablas con entradas determinadas en las columnas 3 y 2, $y_{12}=3$, $y_{22}=2$, $y_{32}=0$; en el nodo (0; 0,0,0) el camino está totalmente recorrido, la tabla está completamente llena, $y_{11}=5$, $y_{21}=0$, $y_{31}=1$.

Con esta construcción de red como base, el algoritmo de M-P consiste en ir examinando cada nodo sucesivamente, comenzando con el nodo inicial (c, \mathbf{R}_c) , y decidir si es o no necesario examinar sus nodos hijos. Supongamos, por ejemplo, que hemos alcanzado un nodo intermedio (k, \mathbf{R}_k) a través de un subcamino inicial

$$(c, \mathbf{R}_c) \rightarrow (c-1, \mathbf{R}_{c-1}) \rightarrow \dots \rightarrow (k, \mathbf{R}_k)$$

cuya longitud es L.

Denotemos por Q el conjunto de todos los caminos completos desde el nodo inicial al nodo terminal que tienen el subcamino inicial de arriba, de longitud L, osea que comparten camino hasta el nodo (k, \mathbf{R}_k) , pero que después divergen, alcanzando el nodo terminal $(0, \mathbf{0})$ por cualquier otro subcamino-continuación en el conjunto $\Omega(k, \mathbf{R}_k)$.

Dependiendo del valor de L y de la longitud del subcamino-continuación más largo y más corto en $\Omega(k, \mathbf{R}_k)$ se puede, a menudo, eliminar el conjunto entero de caminos Q de más consideraciones, bien porque todos ellos contribuyan al p-valor, o

bien porque ninguno lo haga. La equivalencia con la terminología general usada anteriormente para resolver el algoritmo será:

La longitud del subcamino más corto desde el nodo $(k; R_{1k}, \dots, R_{rk}) = SP(k, R_{1k}, \dots, R_{rk})$.

La longitud del subcamino más largo desde el nodo $(k, R_{1k}, \dots, R_{rk}) = LP(k, R_{1k}, \dots, R_{rk})$.

$L = PAST$.

Habría que discutir para resolver el algoritmo lo que sucede en cada nodo con (1.30) y (1.31).

Además, cuando el conjunto entero Q contribuye, existe una expresión determinada que nos da la magnitud de la contribución, la expresión (1.32).

Retomando el ejemplo anteriormente señalado, y considerando el problema

de determinar el p-valor de la tabla $X = \begin{bmatrix} 5 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, la longitud del camino

correspondiente a esta tabla ya lo hemos obtenido, $L(X)=D$. $P(X)=120$, luego se trata de seleccionar los caminos cuya longitud no excede a 120 y computar la contribución de sus tablas correspondientes al p-valor.

Conforme a lo anteriormente expuesto se procedería del siguiente modo:

- Comenzamos por determinar el valor de D

$$D = \frac{N!}{R_1!R_2!R_3!} = \frac{13!}{8!3!2!} = 12870$$

- Establecemos el nodo inicial: $(3; 8,3,2)$

- Determinamos $LP(3; 8,3,2)$, $LP(3; 8,3,2) = 1200$, como es falso que $LP(3; 8,3,2) \leq L(X)$, hacemos p-valor = 0, y extendemos el recorrido a sus nodos hijos.

- Nodos hijos de (3; 8,3,2)

(2; 8,3,0)

(2; 8,2,1)

(2; 8,1,2)

(2; 7,3,1)

(2; 7,2,2)

(2; 6,3,2)

- Para el primer nodo hijo, (2; 8,3,0):

$$\left. \begin{array}{l} PAST = 1 \\ LP = 75 \end{array} \right\} \begin{array}{l} PAST \cdot LP = 75 \\ PAST \cdot LP \leq L(X) \text{ Verdadero} \end{array}$$

todas las tablas correspondientes a ese nodo hijo, con ese valor PAST, contribuyen al p-valor, contribución al p-valor = $165.1/12870 = 0.0128205$, y ese camino iniciado no necesita subcamino continuación.

- Para el siguiente nodo hijo, (2; 8,2,1):

$$\left. \begin{array}{l} PAST = 2 \\ LP = 150 \end{array} \right\} \begin{array}{l} PAST \cdot LP = 300 \\ PAST \cdot LP \leq L(X) \text{ Falso} \end{array}$$

$$\left. \begin{array}{l} PAST = 2 \\ SP = 30 \end{array} \right\} \begin{array}{l} PAST \cdot SP = 60 \\ PAST \cdot SP > L(X) \text{ Falso} \end{array}$$

el camino iniciado, $(3; 8,3,2) \rightarrow (2; 8,2,1)$, necesita continuación a través de los nodos hijos del nodo (2; 8,2,1), para los que habrá que repetir el examen anterior. Análogamente habría que decidir para los restantes nodos hijos del estado 2 si es o no necesario examinar sus nodos hijos.

Algoritmo 643 FEXACT de M-P

La resolución del algoritmo de M-P exige un complemento que indique la forma computacional del mismo, o sea, una adecuación en lenguaje informático que posibilite la computación mediante un ordenador, y este complemento es el que proporcionan M-P en FEXACT, una subrutina FORTRAN, que incorpora importantes detalles que contribuyen al éxito del algoritmo, y donde, precisamente, el mayor interés de FEXACT se centra en la lógica seguida para procesar los nodos en el retículo; Se procesan todos los nodos del estado k antes de continuar al estado $k-1$; Hecho que complica el programa de computación pero presta mayor eficacia al algoritmo.

Esta lógica de procesamiento precisa de cinco matrices: KEY, IPOIN, STP, IFREQ, y NPOIN, de llamadas a las subrutinas GET, PUT, LONGP, y SHORTP, y una función TABLE.

KEY es una matriz de almacenamiento de nodos, donde para realizar el almacenamiento se identifica cada nodo con un número entero.

STP es una matriz de almacenamiento de valores PAST.

IFREQ es una matriz de almacenamiento de copias (frecuencias) de los valores PAST.

IPOIN almacena indicadores y nos permite localizar para un nodo el primer valor PAST de STP asociado con dicho nodo.

NPOIN almacena también indicadores, y nos permite, una vez localizado en STP el primer valor PAST asociado a un nodo, el localizar en STP el siguiente valor PAST asociado al nodo.

La subrutina GET extrae nodos y valores PAST de las matrices correspondientes.

La subrutina PUT inserta nodos y valores PAST en las matrices correspondientes.

Las subrutinas LONGP y SHORTP nos permiten computar, respectivamente, las longitudes de los subcaminos más largos y más cortos de un nodo al nodo terminal.

La función TABLE (I,J) devuelve el logaritmo del coeficiente binomial

$$\frac{I!}{J!(I-J)!}. \text{ Y es usado para evitar } \textit{overflow} \text{ cuando estos coeficientes son}$$

demasiado grandes.

En el almacenamiento y recuperación de nodos se hace también uso del teorema de simetría de M-P, que permite establecer la contribución futura de todos los nodos, que son simples permutaciones de otros nodos, una sola vez, y sencillamente considerar el valor PAST asociado con todos esos nodos. Este teorema dice: *La contribución al p-valor de todos los subcaminos del nodo $(k, R_{1k} \dots R_{rk})$ al nodo $(0, 0 \dots 0)$ es igual a la contribución de todos los subcaminos del nodo $(k, R'_{1k}, R'_{2k}, \dots, R'_{rk})$ al nodo $(0, 0 \dots 0)$ supuesto que $(R'_{1k}, R'_{2k}, \dots, R'_{rk})$ es una permutación de $(R_{1k}, R_{2k}, \dots, R_{rk})$.*

El sentido y uso de este teorema, así como las ventajas que aporta, quedan de manifiesto en el ejemplo práctico que estamos considerando. Los nodos $(2; 8, 2, 1)$ y $(2; 8, 1, 2)$ requieren procesamiento, haciendo uso del teorema de simetría, podemos integrar ambos nodos en un único nodo, por ejemplo el $(2; 8, 2, 1)$ e incorporar a los PAST de este nodo los PAST del $(2; 8, 1, 2)$ con lo que solo tendremos que llevar a cabo el almacenaje y análisis de solo uno de ellos: el $(2; 8, 2, 1)$.

La mejor forma de resumir la lógica de FEXACT viene dada por el diagrama de la figura 1.2.

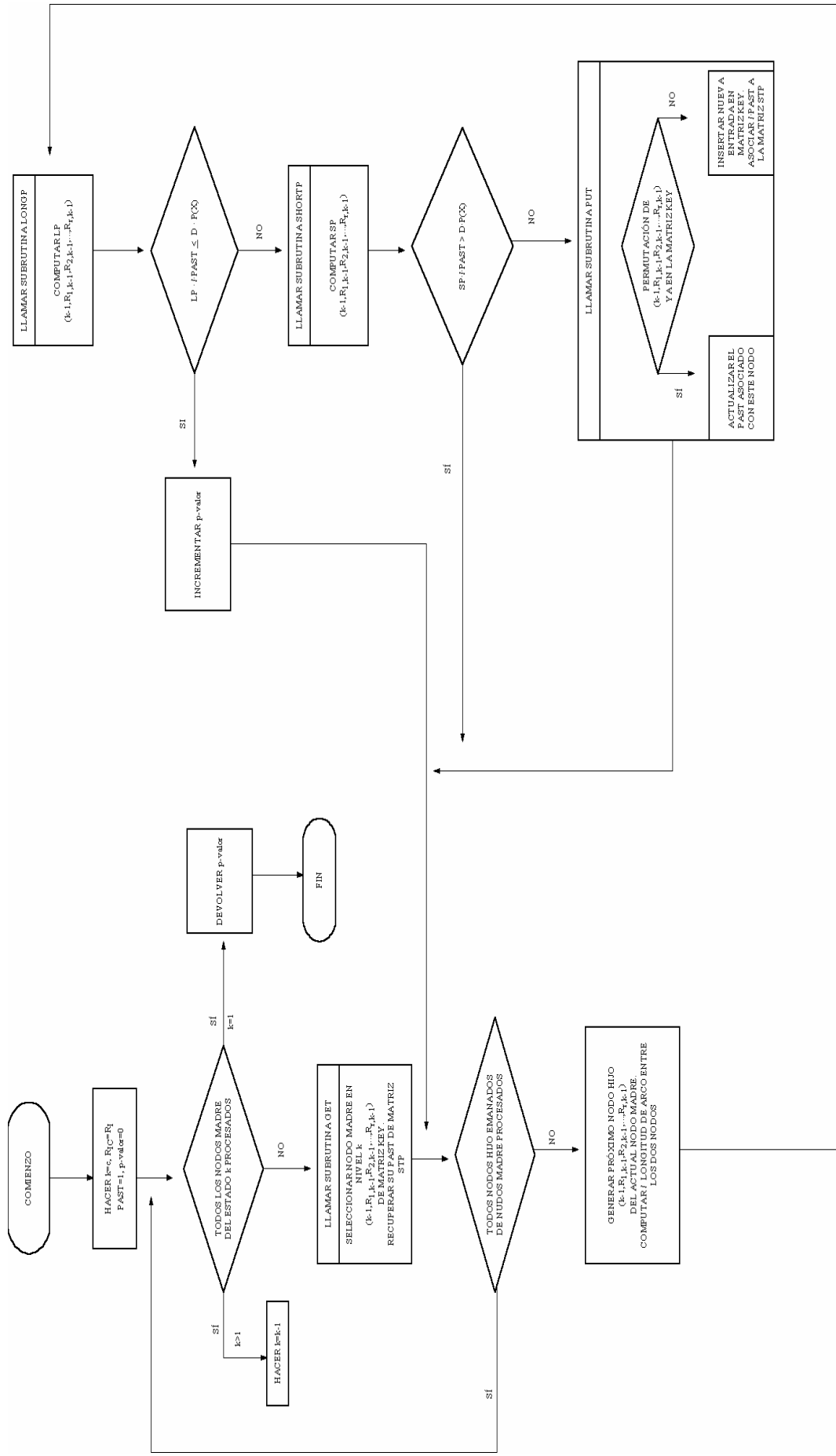


Figura 1.2

Es conveniente poner de manifiesto que por razones de computación (mas rapidez y ahorro de memoria a la hora de almacenar los diferentes valores PAST) las inspecciones con los subcaminos más largos y más cortos que se realizan en los distintos nodos, no se hace exactamente como se indica en (1.30) y (1.31), sino como se indica en el diagrama, esto es, la inspección se hace en un nodo hijo (estado k-1), testeando

$$\text{PAST} \cdot \ell \cdot \text{LP} \leq D \cdot P(X) \quad (1.34)$$

ó

$$\text{PAST} \cdot \ell \cdot \text{SP} > D \cdot P(X) \quad (1.35)$$

Donde PAST es la longitud del subcamino recorrido hasta el nodo madre, ℓ , la longitud del arco que une al nodo hijo con el nodo madre, LP la longitud del subcamino más largo a partir del nodo hijo, SP la longitud del subcamino más corto a partir del nodo hijo.

Con esta forma de entender el algoritmo de computación del p-valor de una tabla, aplicado a nuestro ejemplo, tendríamos, para el nodo madre (3; 8,3,2)

Nodos hijos	PAST	ℓ	LP	$\text{¿PAST} \cdot \ell \cdot \text{LP} \leq L(X)?$	SP	$\text{¿PAST} \cdot \ell \cdot \text{SP} > L(X)?$
(2; 8,3,0)	1	1	75	Si		
				C.p - valor = 165.1/12870		
				$p - \text{valor} = 0.0128205$		
(2; 8,2,1)	1	2	150	No	30	No → se almacena
(2; 8,1,2)	1	1	150	No	30	No → se almacena
(2; 7,3,1)	1	2	300	No	20	No → se Almacena
(2; 7,2,2)	1	2	600	No	30	No → se almacena
(2; 6,3,2)	1	1	1200	No	10	No → se almacena

y, siguiendo con los nodos almacenados :

Nodo madre PAST

(2; 8,2,1) 1
2

Nodos hijos	PAST	ℓ	LP	$\text{¿PAST}.\ell.LP \leq L(X)?$	SP	$\text{¿PAST}.\ell.SP > L(X)?$
(1; 3,2,1)	1	1	60	Sí C.p - valor = $60/12870 = 0.004662$ $p - \text{valor} = 0.0174825$		
	2	"	"	Sí C.p - valor = $120/12870 = 0.009324$ $p - \text{valor} = 0.0268065$		
(1; 4,1,1)	1	5	30	No	30	Sí → Fuera R.Crítica.
	2	"	"	No	"	Sí → Fuera R.Crítica
(1; 4,2,0)	1	5	15	Sí C.P - valor = $75/12870 = 0.0058275$ $p - \text{valor} = 0.032634$		
	2	"	"	No	15	Sí → Fuera R.Crítica.
(1; 5,0,1)	1	10	6	Sí C.p - valor = $60/12870 = 0.004662$ $p - \text{valor} = 0.037296$		
	2	"	"	Sí C.p - valor = $120/12870 = 0.009324$ $p - \text{valor} = 0.04662$		
(1; 5,1,0)	1	20	6	Sí C.p - valor = $120/12870 = 0.009324$ $p - \text{valor} = 0.055944$		
	2	"	"	No	6	Sí → Fuera R.Crítica.
(1; 6,0,0)	1	30	1	Sí C.p - valor = $30/12870 = 0.002331$ $p - \text{valor} = 0.058275$		
	2	"	"	Sí C.p - valor = $60/12870 = 0.004662$ $p - \text{valor} = 0.062937$		

Nodo Madre PAST
(2; 7,3,1) 2

Nodos hijos	PAST	ℓ	LP	$\text{¿PAST}.\ell.\text{LP} \leq L(X)?$	SP	$\text{¿PAST}.\ell.\text{SP} > L(X)?$
(1; 2,3,1)	2	1	60	Sí		
				<i>C.p - valor</i> = 120/12870 = 0.009324		
				<i>p - valor</i> = 0.072261		
(1; 3,2,1)	2	5	60	No	60	Sí → Fuera R.Crítica.
(1; 3,3,0)	2	5	20	No	20	Sí → Fuera R.Crítica.
(1; 4,1,1)	2	10	30	No	30	Sí → Fuera R.Crítica.
(1; 4,2,0)	2	20	15	No	15	Sí → Fuera R.Crítica.
(1; 5,0,1)	2	10	6	Sí		
				<i>C.p - valor</i> = 120/12870 = 0.009324		
				<i>p - valor</i> = 0.081585		
(1; 5,1,0)	2	30	6	No	6	Sí → Fuera R.Crítica.
(1; 6,0,0)	2	20	1	Sí		
				<i>C.p - valor</i> = 40/12870 = 0.003108		
				<i>p - valor</i> = 0.084693		

Nodo Madre PAST
(2; 7,2,2) 2

Nodos hijos	PAST	ℓ	LP	$\text{¿PAST}.\ell.\text{LP} \leq L(X)?$	SP	$\text{¿PAST}.\ell.\text{SP} > L(X)?$
(1; 2,2,2)	2	1	90	No	90	Sí → Fuera R.Crítica.
(1; 3,1,2)	2	5	60	No	60	Sí → Fuera R.Crítica.
(1; 3,2,1)	2	5	60	No	60	Sí → Fuera R.Crítica.
(1; 4,0,2)	2	10	15	No	15	Sí → Fuera R.Crítica.
(1; 4,1,1)	2	20	30	No	30	Sí → Fuera R.Crítica.
(1; 4,2,0)	2	10	15	No	15	Sí → Fuera R.Crítica.
(1; 5,0,1)	2	30	6	No	6	Sí → Fuera R.Crítica.
(1; 5,1,0)	2	30	6	No	6	Sí → Fuera R.Crítica.
(1; 6,0,0)	2	30	1	Sí		
				<i>C.p - valor</i> = 60/12870 = 0.004662		
				<i>p - valor</i> = 0.089355		

Nodo Madre PAST
(2; 6,3,2) 1

Nodos hijos	PAST	ℓ	LP	$\zeta_{\text{PAST}.\ell.\text{LP}} \leq L(X)?$	SP	$\zeta_{\text{PAST}.\ell.\text{SP}} > L(X)?$
(1;1,3,2)	1	1	60	Sí		
				C.p - valor = $60/12870 = 0.004662$		
				$p - \text{valor} = 0.094017$		
(1;2,2,2)	1	5	90	No	90	Sí → Fuera R.Crítica.
(1;2,3,1)	1	5	60	No	60	Sí → Fuera R.Crítica.
(1;3,1,2)	1	10	60	No	60	Sí → Fuera R.Crítica.
(1;3,2,1)	1	20	60	No	60	Sí → Fuera R.Crítica.
(1;3,3,0)	1	10	20	No	20	Sí → Fuera R.Crítica.
(1;4,0,2)	1	10	15	No	15	Sí → Fuera R.Crítica.
(1;4,1,1)	1	30	30	No	30	Sí → Fuera R.Crítica.
(1;4,2,0)	1	30	15	No	15	Sí → Fuera R.Crítica.
(1;5,0,1)	1	20	6	Sí		
				C.p - valor = $120/12870 = 0.009324$		
				$p - \text{valor} = 0.103341$		
(1;5,1,0)	1	30	6	No	6	Sí → Fuera R.Crítica.
(1;6,0,0)	1	10	1	Sí		
				C.p - valor = $10/12870 = 0.000777$		
				$p - \text{valor} = 0.104118$		

Comparación del algoritmo de M-P con el algoritmo de P-H

- 1) En ambos se realiza una descomposición de $P(Y)$ como producto de factores.

En el caso de P-H, cada factor es una hipergeométrica simple. En el caso de M-P, cada factor es la longitud de un arco desde un nodo de un estado a un nodo del estado siguiente.

- 2) El objetivo del teorema 1.2 de P-H y de la propiedad 1.1 de M-P, esencialmente, es el mismo: determinar la suma de las probabilidades de todas las tablas que presentan un segmento inicial común, o sea, presentan un grupo inicial común de factores en su descomposición.

Tanto en un caso como en otro el problema queda reducido, evidentemente, a la suma de los subproductos restantes no comunes para todas las tablas que se encuentran en las condiciones indicadas. En P-H se

demuestra que esta suma queda reducida a 1, con lo que la suma buscada coincide con el producto de los factores comunes, y queda expresada al igual que sucedía con las probabilidades individuales como un producto de hipergeométricas simples; y en el de M-P toma el aspecto de la longitud de un arco con un contenido algo especial, y también la suma buscada queda expresada, al igual que sucedía con las probabilidades individuales, como un producto de longitudes de arcos.

Con lo que, si la propiedad de P-H se aplica sobre columnas completas, se ve, de forma inmediata, que (no puede ser de otra manera) tanto el teorema 1.2 de P-H como la propiedad 1.1 de M-P confluyen en el mismo resultado para la suma de las probabilidades: La probabilidad hipergeométrica de una tabla que tiene todas sus columnas, excepto una, rellena con los elementos comunes, y esa otra columna la rellena con una colapsación de los elementos restantes.

- 3) En el algoritmo de P-H el resultado anteriormente obtenido se aprovecha directamente como elemento testeador; si la suma de las probabilidades de las tablas, con la situación indicada para estas tablas, fuese menor o igual que $P(X)$, todas ellas estarían en la región crítica, y su aportación al p-valor de X vendría dado por la suma obtenida.

Sin embargo, en el algoritmo de M-P, el resultado obtenido para la suma no se utiliza directamente para testear. En este caso el testeo se realiza incorporando nuevos resultados, aportados por los restantes teoremas de M-P, derivados del máximo y el mínimo de los subproductos restantes no coincidentes de la descomposición de las tablas con la situación indicada; y en el caso de que el producto total (subproducto

común y no común) correspondiente al máximo de los subproductos no comunes sea menor o igual que el producto de las longitudes de los arcos de $P(X)$, todas las tablas bajo consideración estarían en la región crítica, y, es entonces, cuando el resultado referente a la suma de sus probabilidades es utilizado para incorporarlo al p-valor de X .

- 4) La condición de testeo de M-P puede detener el proceso restante de computación, y evitar más extensiones, en muchas ocasiones antes que la condición de P-H; en todas aquellas en que no se verifica la condición de P-H y, sin embargo, sí se verifica la condición de M-P para máximo. Pues la verificación de la condición de P-H implica la verificación de la condición de M-P referente al máximo, pero no recíprocamente.
- 5) P-H, al no disponer de propiedades referentes a la probabilidad mínima, no pueden utilizarlas, como hacen M-P, para determinar el rechazo de grupos de tablas en bloque.
- 6) En el algoritmo de P-H, el proceso de computación se realiza de la forma más natural; trabajando recursivamente sobre casillas.

En el proceso de computación del algoritmo de M-P, se podía haber optado también por la forma más natural, pasando de cada nodo a un nodo del estado siguiente, sin embargo, M-P han optado por un proceso recursivo sobre estados, lo que se llama un procesamiento *stagewise*; se procesan todos los nodos de un estado antes de pasar al estado siguiente. Esta forma de procesamiento hace que el programa de computación del algoritmo de M-P sea más complejo que si se realizará de la forma más natural, pero, sin embargo, con ella gana en eficiencia el algoritmo.

- 7) El algoritmo de P-H, al utilizar las probabilidades hipergeométricas simples como base de la expresión de $P(Y)$, facilita el uso de las propiedades referentes a la moda y a la recursividad de estas hipergeométricas para acortar el propio algoritmo. Sin embargo, en el algoritmo de M-P, no se le saca provecho a estas propiedades en algunas situaciones en las que podrían resultar efectivas.
- 8) La primacía computacional del algoritmo de M-P sobre el de P-H se presenta de forma evidente en la comparación que mostramos en la siguiente tabla, realizada por M-P (1983) para una variedad de tablas de contingencia $r \times c$:

Tamaño	Segundos	CPU	Segundos	CPU	Razón	CPU
De la tabla $r \times c$	Mehta and Patel		Pagano-Halvorsen			
2x4	0'32		0'82		2'56	
2x4	0'27		0'73		2'70	
3x3	0'32		1'05		3'28	
3x3	0'31		0'81		2'61	
2x5	0'40		1'73		4'33	
3x4	0'36		1'25		3'47	
3x4	0'89		3'84		4'31	
3x5	2'38		34'02		14'29	
4x5	2'31		41'12		17'80	
4x5	0'35		4'39		12'54	
4x5	6'10		1263'40		207'11	
5x5	1'30		87'45		67'27	

5x5	2'04	815'36	399'69
5x5	14'22	4910'74	345'34

Esta tabla muestra que el algoritmo reticulado funciona uniformemente mucho mejor para estos problemas, y la razón CPU aumenta desmesuradamente con el tamaño de los problemas.

I.3: Mejoras al Algoritmo de red de Mehta y Patel.

I.3.1.-Probabilidades extremas de Harry Joe para tablas de contingencia rx_c .

Harry Joe (1988) enuncia y demuestra una serie de teoremas que le permiten encontrar las tablas que minimizan (o maximizan) $P(Y)$; Estas propiedades de maximización y minimización incorporadas al algoritmo de M-P proporcionan a este algoritmo mayor rapidez y menor exigencia de capacidad de memoria, sobre todo en el caso en que los marginales columna de la tabla que estamos manejando sean distintos y muy diferentes entre sí, que es el caso en que el algoritmo de M-P trabaja con acotaciones más pobres para la determinación del máximo y el mínimo requerido.

Hay un grupo de teoremas (teoremas 1.7, 1.12 y 1.13) que se obtienen trabajando directamente sobre la función de probabilidad; sin embargo, para conseguir el resto de sus resultados (teoremas 1.8, 1.10 y 1.11), Harry Joe se basa en los resultados previamente conseguidos por él (Harry Joe 1985), relativos a una ordenación de dependencia para tablas de contingencia rx_c .

En el teorema 1.7, H.J. establece como condición necesaria para que Y maximice (minimice) $P(Y)$ que todas las posibles subtablas de Y , maximicen (minimicen) la función de probabilidad correspondiente a todas las tablas con los mismos marginales que ellas. O sea,

Teorema 1.7 de H.J.-

Si $Y = (y_{ij})$ maximiza (minimiza) la expresión (1.4) sobre Ω ,

Entonces, para cada pareja de subconjuntos $I \subset \{1,2,\dots,r\}$, $J \subset \{1,2,\dots,c\}$, al menos de cardinal 2, la subtabla $Y^{IJ} = (y_{ij})_{i \in I, j \in J}$, maximiza (minimiza) $P(Y^{IJ})$ sobre Ω^{IJ} , donde Ω^{IJ} es el conjunto de tablas $I \times J$ con marginales filas $R_i^I = \sum_{j \in J} y_{ij}$ $i \in I$, y

marginales columnas $C_j^J = \sum_{i \in I} y_{ij}$ $j \in J$.

Aparte de este teorema, H.J. proporciona una condición necesaria para mínimo y dos condiciones necesarias y una suficiente para máximo. Hemos incorporado a estas condiciones otra condición necesaria de mínimo, Teorema 1.9, que se deriva igualmente de sus resultados de 1985, y que aunque es menos fuerte que la dada nos resultará de utilidad para conseguir una idea mas completa sobre sus aportaciones al comparar sus resultados sobre probabilidades extremas con los de M-P.

La condición necesaria para mínimo de Joe(1988) viene dada por el siguiente teorema:

Teorema 1.8.-

Supongamos $Y=(y_{ij})$ minimiza (1.4) sobre Ω . Sea $R_k = \max \{R_1, R_2, \dots, R_r\}$ y $C_v = \max \{C_1, C_2, \dots, C_c\}$

- a) Si $R_k \geq C_v$, y $r \leq c$, entonces $y_{iv} = \min \{R_i, C_v\}$ para algún $1 \leq i \leq r$.

b) Si $R_k \geq C_v$, y $r \geq c$, entonces $y_{kj} = C_j$ para algún $1 \leq j \leq c$, $j \neq m$, donde

$$C_m = \min \{C_1, C_2, \dots, C_c\}.$$

Y un resultado simétrico al anterior también es válido cuando $R_k \leq C_v$.

La condición necesaria de mínimo menos fuerte que la anterior pero útil para el objetivo señalado viene dada por:

Teorema 1.9.- Si Y minimiza (1.4) sobre Ω entonces para cualquier subtabla existe una línea (fila o columna) donde, a lo sumo, un elemento es estrictamente positivo.

De los teoremas 1.7 y 1.8 se deriva que si $r \geq c$, hay a lo sumo, $c^{r-c} \{(c-1)!\}^2$ tablas en Ω que satisfacen esa condición necesaria, y el mínimo de $P(Y)$ ocurre en una de ellas. Esto se consigue aplicando repetidamente las condiciones necesarias en el teorema 1.8 a las subtablas.

Con respecto al máximo H.J., como hemos indicado al principio, establece, a parte del teorema 1.7, dos teoremas: el teorema 1.10, donde da una condición necesaria, y el teorema 1.11, donde da una condición suficiente, apoyándose en los resultados obtenidos para matrices minimales; Y una condición necesaria desdoblada en los teoremas 1.12 y 1.13 más fuerte que la establecida en el teorema 1.10, obtenida usando, explícitamente, la fórmula de $P(Y)$.

Teorema 1.10.- (Condición necesaria de máximo)

Supongamos que $Y = (y_{ij})$ maximiza (1.4) sobre Ω , entonces a) $y_{ij} \geq y_{kj} - 1$, $j=1,2,\dots,c$, si $R_i \geq R_k$, y b) $y_{ij} \geq y_{iv} - 1$, $i=1,2,\dots,r$ si $C_j \geq C_v$.

Teorema 1.11.- (Condición suficiente de máximo)

Si existe $Y \in \Omega$ tal que (a) $|y_{ij}-y_{iv}| \leq 1$ para todo i y $j \neq v$, ó (b) $|y_{ij}-y_{kj}| \leq 1$ para todo j y $i \neq k$, entonces Y maximiza (1.4) sobre Ω .

Conviene observar que las condiciones de este teorema se satisfacen solamente si C_1, C_2, \dots, C_c (ó R_1, R_2, \dots, R_r) no difieren mucho. Para el caso de que estas condiciones no se den H.J. establece la siguiente condición necesaria:

Si Y maximiza (1.4) sobre Ω el teorema 1.12 nos proporciona unas cotas que delimitan superior e inferiormente los posibles valores (y_{ij}) de las casillas de Y . Si estas cotas delimitan un único valor entero, este valor puede considerarse ya como elemento conocido de Y , con lo que la posible variación de los restantes elementos podrá ajustarse más; Estas cotas mejoradas son las proporcionadas por el teorema 1.13.

Teorema 1.12.- (Condición necesaria de máximo)

Si $Y = (y_{ij})$ maximiza (1.4) sobre Ω , entonces

$$l_{ij} \leq y_{ij} \leq u_{ij}, \quad i=1,2,\dots,r ; j=1,2,\dots,c$$

donde

$$u_{ij} = (R_i + c - 1)(C_j + r - 1) / (N + r + c - 2) \quad (1.36)$$

y

$$l_{ij} + 1 = (R_i + 1)(C_j + 1) / (N + (r - 1)(c - 1) + 1) \quad (1.37)$$

Teorema 1.13.- (Condición necesaria de máximo)

Si $Y = (y_{ij})$ maximiza (1.4) sobre Ω , entonces

$$\frac{(R_k'+1)(C_j+1)}{N'+(r-1)(d-1)+1} - 1 \leq y_{kj} \leq \frac{(R_k'+d-1)(C_j+r-1)}{(N'+r+d-2)}$$

$$j \in J, \quad k=1,2,\dots,r$$

donde J es un subconjunto no vacío de $\{1,2,\dots,c\}$, $N'=N-\sum_{j \in J} C_j$, $R_k'=R_k-\sum_{j \in J} y_{kj}$, y d

es el cardinal de J . Una condición similar es válida para $I \subset \{1,2,\dots,r\}$, $N'=N-\sum_{i \in I} R_i$,

$$C_v'=C_v-\sum_{i \in I} y_{iv}.$$

En esta condición necesaria queda claramente de manifiesto al considerar los valores de la cota superior, u_{ij} , e inferior, l_{ij} , que la diferencia $u_{ij} - l_{ij}$ depende de la magnitud de los correspondientes valores marginales R_i , C_j , así como de las dimensiones r y c de las tablas de Ω , por consiguiente para tablas con dimensiones grandes, esta diferencia proporciona un rango de valores cuya combinación suministra un conjunto de tablas donde aparte de encontrarse las tablas de máxima probabilidad, se encuentran otras muchas que no lo son, e incluso un elevado número de tablas imposibles (sus marginales no son los de la familia Ω). El tamaño de este conjunto de tablas de donde hay que seleccionar las de máxima probabilidad, puede alcanzar tal magnitud que prácticamente esta metodología se hace inviable. Algunos ejemplos pueden servirnos para evidenciar esta situación:

Ejemplo 1. Sea Ω la familia de tablas 2×15 con marginales filas $R_1=63$, $R_2=140$, y marginales columnas $C_1=C_2=19$, $C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}=16$, $C_{12}=C_{13}=5$, $C_{14}=9$, $C_{15}=2$. Aplicando la condición necesaria de Joe a este ejemplo llegaríamos (dejando aparte las tablas imposibles) a un conjunto de mas de 2.8×10^6 tablas 2×15 entre las que habría que encontrar las de máxima probabilidad, que en este caso solo son tres.

Ejemplo 2. Sea Ω la familia de tablas 2×22 con marginales filas $R_1=78$, $R_2=58$, y marginales columnas $C_1=C_2=9$, $C_3=C_4=C_5=C_6=C_7=C_8=C_9=C_{10}=C_{11}=C_{12}=C_{13}=C_{14}=C_{15}=7$, $C_{16}=C_{17}=C_{18}=6$, $C_{19}=C_{20}=1$, $C_{21}=4$, $C_{22}=3$. Aplicando la condición necesaria de Joe llegaríamos (dejando aparte las tablas imposibles) a un conjunto de más de 1.8×10^9 tablas 2×22 , entre las que habría que encontrar las de máxima probabilidad, que en este caso son solo tres.

No obstante, y a pesar de las importantes limitaciones que acabamos de señalar, la incorporación de la metodología de determinación de las probabilidades extremas de Joe al algoritmo de M-P para la computación del p-valor de una tabla, representa como queda de manifiesto en el siguiente apartado mejoras que ocasionan una sensible disminución de los tiempos de cómputo.

Relación entre las Probabilidades extremas de Harry Joe y las de Mehta-Patel. Computación del p-valor.

1.-) Los teoremas 1.3, 1.4 y 1.5 de M-P referentes a probabilidades extremas de una tabla $r \times c$ cuando todos los marginales columnas son iguales, están fuertemente ligados a los teoremas 1.11, 1.9 y 1.8, respectivamente, de H.J., incluso se pueden deducir de ellos.

- a) Al establecer H.J. en el teorema 1.11 que si existe Y , $Y \in \Omega$, tal que $|y_{ij} - y_{ik}| \leq 1$ para todo i , y $j \neq k$, entonces Y es máximo de $P(Y)$, una tabla Y que tenga todos sus marginales columnas iguales, con h_i entradas en la fila i iguales a $d_i + 1$ y el resto de las entradas iguales a d_i ($d_i = [R_i/c]$, $h_i = R_i - cd_i$) estaría en las condiciones de este teorema, y, por consiguiente, en ellas se

alcanzaría el máximo, que es el resultado indicado por M-P en su teorema 1.3.

b) Si Y minimiza $P(Y)$ entonces H.J. establece (teorema 1.9) que para cualquier subtabla existe una línea (fila o columna) donde, a lo sumo, un elemento es positivo, con lo que, evidentemente, las entradas no nulas de Y no pueden formar un ciclo, que es lo que establece M-P en su teorema 1.4.

c) Si en el teorema 1.8 de H.J. hacemos todos los marginales columnas iguales, o sea, $C_j=C$ para todo $j=1,2,\dots,c$ obtenemos el mismo resultado que M-P en su teorema 1.5, pues investigar soluciones básicas posibles de

$$\sum_{j=1}^c y_{ij} = R_i \quad i=1,2,\dots,r$$

$$\sum_{i=1}^r y_{ij} = C \quad j=1,2,\dots,c$$

es un problema análogo al de encontrar tablas Y , con los marginales filas y columnas dados cuyos elemento minimicen $P(Y)$ en el sentido dado en el teorema 1.8.

2.-) En el caso de que los marginales columnas sean distintos, M-P lo que consigue son probabilidades extremas exactas para una tabla ampliada de la tabla objetivo, ampliada con una fila adicional de forma tal que la tabla resultante de la ampliación tenga todos sus marginales columnas iguales, y por los teoremas 1.3, 1.4 y 1.5 puedan determinarse sus probabilidades extremas. A partir de aquí, y por el teorema 1.6, se obtienen cotas para las probabilidades extremas de la tabla objetivo.

En el caso de H.J. se consiguen valores exactos en lugar de cotas para las probabilidades extremas de la tabla objetivo.

Por la misma forma de establecerse las cotas de M-P, y, además, porque se ha constatado experimentalmente, estas cotas son bastante buenas cuando existe poca diferencia entre los marginales columnas de la tabla objetivo; Sin embargo, no sucede así cuando estas diferencias son grandes. Por ejemplo, si $(R_1, R_2, R_3) = (8, 4, 3)$ y $(C_1, C_2, C_3) = (6, 5, 4)$ el mínimo de $P(Y)$ sobre la familia con los marginales anteriores es 0'0000444 y el máximo es 0'0639; Las cotas son 0'0000148 y 0'12787 respectivamente.

El hecho de que las cotas no sean buenas influye decisivamente en el proceso de computación del algoritmo, pues al descansar éste sobre las condiciones (1.34) y (1.35) hace que si el valor que se le atribuye a LP (SP) es mayor (menor) que el que realmente tiene deje de verificarse la (1.34) ((1.35)), con lo que se hace necesario el procesamiento de más nodos en el retículo de los que serían necesarios, y es precisamente en el procesamiento de los nodos del retículo donde se emplea la mayor parte del tiempo y de memoria del proceso global.

Para computar el p-valor del test exacto de Fisher, H.J. utiliza un programa que, en la parte esencial, que es el procesamiento del retículo, es el mismo que el de M-P (1986), pero donde una subrutina basada en el teorema 1.8 nos permite encontrar el mínimo de $P(Y)$. Y una subrutina basada en los teorema 1.11, 1.12 y 1.13 nos permite encontrar el máximo de $P(Y)$.

La disminución en el tiempo de computación mediante las probabilidades extremas de H.J. se pone claramente de manifiesto en la siguiente tabla:

Tabla	P-valor	CPU (minutos) Mehta-Patel	CPU (minutos) Harry Joe	Razón
22 13 5 4 5 3 2 1 7 1 4 3 1 2 3 4	0'0360	1'8	0'3	5'9

20364766222	0'0825	69'8	4'0	17'3
8845210211				
247381	0'0199	2'3	0'8	3'0
95503				
20201				
1854422	0'0036	68'4	35'5	1'9
5334310				
10140000				

I.3.2.- Mejoras computacionales de Clarkson, Fan y Joe

Las mejoras de estos autores se centran en el FEXACT, y las mas interesantes consisten en lo siguiente:

- 1) En las subrutinas LONGP y SHORTP para computar respectivamente las longitudes del camino más largo y más corto de un nodo al nodo terminal cuando los marginales columna son diferentes, en vez de utilizar las cotas aproximadas de M-P se utilizan las cotas exactas de Joe (1988).

Joe, como hemos visto, ya había verificado la ventaja de cotas exactas cuando los marginales columnas son muy diferentes, pero C-F-J extienden esta medida incluso cuando esos marginales columnas no son muy diferentes, pues a parte de la mayor exactitud que proporcionan, el coste computacional añadido de computar cotas exactas es prácticamente despreciable en el cómputo final del algoritmo resultante de la combinación de esta modificación con las modificaciones restantes.

- 2) En el algoritmo de M-P siempre que se genera un nodo hijo en el estado $k-1$, se computan la longitud del camino más largo, ó más corto (ó los

dos), C-F-J almacenan los valores de estas longitudes en una tabla auxiliar y los recuperan cuando los necesitan. Esto proporciona un ahorro considerable de tiempo computacional porque usualmente son varios los nodos madre del estado k que originan el mismo nodo hijo del estado $k-1$.

El esfuerzo computacional en recobrar las cotas de la tabla auxiliar no es, usualmente, una carga extra, porque la clave usada para recobrar las cotas puede usarse para almacenar los valores PAST. La memoria extra necesaria para almacenar las cotas tampoco es grande porque el número de nodos en cada estado del retículo es mucho más pequeño que el número de valores PAST que llegan a estos nodos.

- 3) Se obtienen considerables ahorros si la tabla auxiliar de localización de un nodo se computa una sola vez y se reutiliza cada vez que se necesita. El algoritmo original recomputa la tabla auxiliar de localización cada vez que se requiere.

I.4: Discusión de los Algoritmos de Enumeración incompleta.

Planteamiento de posibles mejoras. Concreción de los objetivos de esta tesis.

En el análisis que hemos realizado en el apartado anterior ha quedado de manifiesto en primer lugar la supremacía del algoritmo de M-P como algoritmo de enumeración incompleta pero también se ha evidenciado que sobre él se pueden y se deben realizar modificaciones que lo mejoren extendiendo el alcance de los métodos exactos a mayor número de tablas.

En el algoritmo de M-P para la computación del p-valor de una tabla la decisión de las tablas que se incorporan o no a la región crítica se realiza sobre nodos

para los que se discute la verificación o no de (1.34) y (1.35), que proporcionan los acortamientos de la enumeración completa de las tablas de Ω . Hay dos tareas básicas inmediatamente reconocibles en este algoritmo para facilitar su aplicabilidad:

- Obtención de las subtrayectorias extremas desde los nodos de la red hasta el nodo terminal de forma sencilla y rápida.
- Reducción del número de nodos sobre los que hay que realizar la discusión.

Admitiendo que el problema de determinación de las subtrayectorias mas cortas está adecuadamente resuelto por Joe (1988), no podemos admitirlo así para las subtrayectorias mas largas pues como ha quedado constatado la metodología de Joe se hace inviable para tablas con dimensiones grandes, sería de interés no solo para agilizar esta parte del algoritmo de M-P sino también cualquier otro proceso donde estuviera involucrado el cálculo del máximo de $P(Y)$, el profundizar en el análisis de esta cuestión y encontrar un algoritmo lo mas eficiente posible para calcularlo, considerando, como punto de partida, que este problema solo está adecuadamente resuelto para el caso de una Hipergeométrica $H(R_1, C_1, N)$, para la que la máxima probabilidad se alcanza en la moda \hat{x}_1 , que se obtiene como un valor entero tal que

$$\frac{(R_1+1)(C_1+1)}{N+2} - 1 \leq \hat{x}_1 \leq \frac{(R_1+1)(C_1+1)}{N+2} \quad (1.38)$$

En cuanto a la segunda tarea, reducción del número de nodos sobre los que hay que comprobar la verificación o no de (1.34) y (1.35), habría que estudiar la posibilidad de encontrar relaciones entre las probabilidades extremas de las tablas correspondientes a los nodos de la red que evitaran el proceso de verificación para determinados nodos por quedar predeterminada esta decisión por lo que sucede sobre otros nodos analizados.

Una resolución adecuada de las dos cuestiones que acabamos de plantear mejoraría de forma importante la resolución del Test exacto de Fisher para tablas rxc. La investigación exhaustiva de este problema y su solución idónea para tablas 2xc es el objetivo concreto y fundamental de esta tesis, que se completa con el objetivo de extender las mejoras a tablas rxc abriendo también vías de solución al problema.

**C A P Í T U L O II: Aportaciones sobre el máximo de una
distribución hipergeométrica multivariante. Mejoras en el algoritmo
del Test exacto de Fisher para tablas 2xc.**

II.1: Introducción

En el capítulo anterior hemos analizado el algoritmo de Mehta –Patel para la resolución del test exacto de Fisher para tablas de contingencia $r \times c$, contrastado sus ventajas sobre otros algoritmos de enumeración incompleta y hemos revisado las mejoras de mayor interés que sobre este algoritmo han aparecido. La reflexión sobre los procedimientos, y su base matemática, implicados en dicho algoritmo en el caso particular de tablas $2 \times c$ nos ha llevado a investigar dos cuestiones bien diferenciadas en su formulación pero cuyos resultados conjuntamente nos permiten elaborar un algoritmo, modificación del algoritmo de Mehta y Patel, que mejora en varios aspectos este algoritmo de referencia, lo que en definitiva representa una mejora importante en tiempos de cómputo de este algoritmo:

- Profundizar en el estudio de la Moda de una Distribución Hipergeométrica Multivariante, tratando de obtener un método general y rápido para su cálculo.
- Aprovechar la estructura reticulada de M-P que nos representa una familia isomarginal de tablas como nodos y trayectorias (camino) de un retículo y establece un fuerte vínculo en cuanto a construcción entre los nodos de un estado y los nodos del estado siguiente, para tratar de establecer también relaciones entre las subtrayectorias máximas correspondientes a estos nodos. Con ello haremos absolutamente despreciable el tiempo de cómputo para las subtrayectorias máximas y reduciremos drásticamente el número de veces que se inspecciona el cumplimiento de las condiciones (1.34) y (1.35).

II.2: Desarrollo de un método general para el cálculo de la moda de una distribución hipergeométrica multivariante.

II.2.1: Introducción.-

Sea T una tabla de contingencia $2 \times c$ ($c \geq 2$) con marginales filas fijos, R_1 y R_2 , y marginales columnas fijos, C_1, C_2, \dots, C_c , y gran total $N = \sum R_i = \sum C_j$. Si x_1, x_2, \dots, x_c es la primera fila de la tabla, podemos representar a T mediante el punto (x_1, x_2, \dots, x_c) . El conjunto de tablas posibles (familia isomarginal) dados los marginales R_1, R_2 y C_1, \dots, C_c , puede ser representado, como ya vimos en el capítulo anterior, mediante el subconjunto convexo de la "rejilla" Z_0^c

$$F_0 = \{(x_1, x_2, \dots, x_c) \in Z_0^c / x_1 + x_2 + \dots + x_c = R_1, 0 \leq x_h \leq C_h \quad h=1, 2, \dots, c\}$$

Donde Z_0 es el conjunto de los enteros no negativos. Existiendo, pues, una correspondencia uno a uno entre la familia isomarginal de una tabla $2 \times c$ con marginales fijos y el conjunto de puntos de F_0 .

Por otra parte, como ya hemos reflejado en la introducción de este trabajo dados los marginales filas y columnas y bajo la hipótesis de independencia entre filas y columnas de la tabla, el vector (x_1, x_2, \dots, x_c) sigue una distribución Hipergeométrica Multivariante (MHD)

$$MH(R_1; C_1, C_2, \dots, C_c) \quad (2.1)$$

cuyo conjunto de referencia es el conjunto F_0 . Siendo

$$P(T) = P(x_1, x_2, \dots, x_c) = \frac{R_1!(N - R_1)! \prod_{i=1}^c C_i!}{N!} \cdot \frac{1}{\prod_{i=1}^c x_i!(C_i - x_i)!} \quad (2.2)$$

Obviamente, para cualquier subtabla de T, el vector $(x_{\alpha 1}, \dots, x_{\alpha c})$ $2 \leq k \leq c$ seguirá también una MHD. Para el caso $c=2$ (tabla 2×2) tendremos el caso particular de una $MH(R_1; C_1, C_2)$, que en realidad es la distribución Hipergeométrica

representada mas comúnmente como $H(R_1, C_1, N)$. Igualmente, cada marginal x_j $j=1,2,\dots,c$ de (2.1) sigue una distribución Hipergeométrica $H(R_1, C_j, N)$.

Así pues, el problema de obtener la tabla $2 \times c$ de máxima probabilidad dentro de la familia isomarginal (para unos marginales fijados) y la probabilidad de esta tabla, es equivalente a obtener el máximo(modal) de la MHD(2.1), es decir, obtener (x_1, x_2, \dots, x_c) que maximice (2.2) y dicha probabilidad máxima.

Este problema surge, por ejemplo, como ha quedado de manifiesto en el capítulo anterior ,como parte del algoritmo reticulado de M-P. La aplicación de este algoritmo a una tabla $2 \times c$ observada requiere múltiples aplicaciones de la obtención de la subtrayectoria mas larga desde determinados nodos de la red hasta el nodo terminal. Cada una de estas aplicaciones equivale, como sabemos, a la obtención de la subtabla $2 \times k$ ($k \leq c$) de máxima probabilidad que corresponde a ese nodo y esto, a su vez, es equivalente, como hemos visto antes, a la obtención del máximo para la MHD correspondiente a la subtabla $2 \times k$. Así pues, nos pareció que sería de interés no solo para agilizar esta parte del algoritmo de M-P, sino también para cualquier otro proceso donde estuviera involucrado el máximo de una MHD , el profundizar en el estudio de esta cuestión y encontrar un método lo mas eficiente posible para calcularlo.

Como para el caso de una Hipergeométrica $H(R_1, C_1, N)$ el cálculo de su moda es conocido (1.38), en particular, la moda de cada marginal x_j $j=1,\dots,c$ de (2.1) será un entero, que llamaremos z_j , tal que

$$R(C_j+1) - 1 \leq z_j \leq R(C_j+1) \quad (2.3)$$

Con $R = (R_1+1)/(N+2)$

II.2.2: Notación.-

En adelante llamaremos $(i, j, k_1, \dots, k_{c-2})$ a una permutación del vector de enteros $(1, 2, \dots, c)$ tal que $k_1 < k_2 < \dots < k_{c-2}$. Y para cualquier punto $\mathbf{x} \in Z_0^c$ denotaremos

$$R_{ij}(\mathbf{x}) = R_1 - \sum_{h=1}^{c-2} x_{k_h} = x_i + x_j \quad (2.4)$$

$$m_{ij}(\mathbf{x}) = \frac{(R_{ij}(\mathbf{x}) + 1)(C_i + 1)}{C_i + C_j + 2} \quad (2.5)$$

Donde x_i, x_j, x_{k_h} representan, respectivamente, las coordenadas i, j, k_h de \mathbf{x} .

II.2.3: Caracterización de la Moda de una MHD.-

Consideremos la MHD(2.1) y una permutación del tipo $(i, j, k_1, \dots, k_{c-2})$. Entonces para cada conjunto de valores fijos de las coordenadas x_{k_h} $1 \leq h \leq c-2$ del vector \mathbf{x} tendremos que $(x_i / x_{k_1}, \dots, x_{k_{c-2}})$ sigue la distribución Hipergeométrica

$$H(R_{ij}(\mathbf{x}), C_i, C_i + C_j) \quad (2.6)$$

y su moda, según (2.3) y (2.5), será un valor entero \hat{x}_i que cumple

$$m_{ij}(\mathbf{x}) - 1 \leq \hat{x}_i \leq m_{ij}(\mathbf{x}) \quad (2.7)$$

Además, sabemos que siempre se cumple la siguiente propiedad, que podemos expresar como,

Lema 2.1. Dada cualquier permutación del tipo $(i, j, k_1, \dots, k_{c-2})$ si $\hat{\mathbf{x}} \in F_0$ maximiza la probabilidad (2.2) de la MHD (2.1) ($\hat{\mathbf{x}}$ es una moda de (2.1)) y $\hat{x}_i, \hat{x}_{k_1}, \dots, \hat{x}_{k_{c-2}}$ son sus coordenadas i, k_1, \dots, k_{c-2} , respectivamente, entonces \hat{x}_i maximiza la $P(x_i / \hat{x}_{k_1}, \dots, \hat{x}_{k_{c-2}})$ en la Hipergeométrica (2.6) (\hat{x}_i es una moda de (2.6) dados los valores $\hat{x}_{k_1}, \dots, \hat{x}_{k_{c-2}}$).

De hecho esta propiedad no es mas que una particularización del Teorema 1.7 de Harry Joe(1988) que hemos analizado en el capítulo anterior.

Definimos ahora el subconjunto S_0 como

$$S_0 = \{\mathbf{x} = (x_1, \dots, x_c) \in F_0 / m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad i \neq j\} \quad (2.8)$$

Este subconjunto será el conjunto fundamental para la caracterización de la moda de la MHD(2.1).

Lema 2.2. El subconjunto S_0 contiene a todos los puntos de F_0 tales que para cualquier permutación $(i, j, k_1, \dots, k_{c-2})$, el valor de su coordenada i es una moda de la Hipergeométrica(2.6) dados los valores de sus coordenadas k_1, \dots, k_{c-2} .

Demostración.-

Se sigue de forma inmediata de (2.7) y de la propia definición de S_0 .

El siguiente lema nos caracteriza los puntos de S_0 respecto a las distancias en sus coordenadas.

Lema 2.3. Dados dos puntos arbitrarios $\mathbf{x} = (x_1, \dots, x_c)$ y $\mathbf{x}' = (x'_1, \dots, x'_c)$ que pertenecen a S_0 siempre se cumple que $|x'_h - x_h| \leq 1 \quad 1 \leq h \leq c$.

Demostración .-

Sea

$$u = \max_{1 \leq h \leq c} |x'_h - x_h|$$

y sin pérdida de generalidad, sea i una coordenada para la que $x'_i = x_i + u$.

Como $\sum x_h = \sum x'_h = R_1$, tendrá que existir otra coordenada j para la que

$x'_j = x_j - u$ con $0 < u' \leq u$. Además, como $\mathbf{x} \in S_0$ se cumplirá

$$m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad (2.9)$$

igualmente, como $\mathbf{x}' \in S_0$ podemos escribir

$$m_{ij}(\mathbf{x}') - 1 \leq x'_i \leq m_{ij}(\mathbf{x}') = m_{ij}(\mathbf{x}) + (u - u') \cdot G$$

con

$$G = \frac{C_i + 1}{C_i + C_j + 2}$$

de donde podemos deducir

$$x_i \leq m_{ij}(\mathbf{x}) - [u(1-G) + u'G] \quad (2.10)$$

Tenemos que demostrar que $u=1$. Si lo contrario fuese cierto (esto es, si $u > 1$) y como u y u' son números enteros y $0 < G < 1$, se tendría

$$u(1-G) + u'G > 1$$

y según (2.10) se cumpliría

$$x_i < m_{ij}(\mathbf{x}) - 1$$

Lo que contradice (2.9) y la hipótesis inicial, por tanto $u=1$.

Basados en los lemas anteriores llegamos a un teorema que nos proporciona una condición necesaria y suficiente para caracterizar la moda de la MHD(2.1). Este será la base para desarrollar un método simple para el cálculo de dicha moda.

Teorema 2.1. La condición necesaria y suficiente para que $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_c)$ sea una moda de la MHD (2.1) es que $\hat{\mathbf{x}} \in S_0$.

Demostración.-

Es fácil demostrar que si $\hat{\mathbf{x}}$ es una moda de (2.1) entonces $\hat{\mathbf{x}} \in S_0$. En efecto, si $\hat{\mathbf{x}}$ es una moda de la MHD(2.1) obviamente $\hat{\mathbf{x}} \in F_0$ y, por el lema 2.1, se cumplirá que para cualquier permutación del tipo $(i, j, k_1, \dots, k_{c-2})$ \hat{x}_i será una moda de (2.6) dados $\hat{x}_{k_1}, \dots, \hat{x}_{k_{c-2}}$, de donde, por el lema 2.2, tenemos que $\hat{\mathbf{x}} \in S_0$. Esto nos prueba, además, que S_0 contiene al menos un punto.

Veamos ahora la condición suficiente. El caso en que S_0 contenga un solo punto es trivial. Supondremos, pues, que existen al menos dos puntos en S_0 . Como la

moda de la MHD (2.1) pertenece a S_0 y todo punto de S_0 es un punto de la MHD (2.1), bastará demostrar que dados dos puntos arbitrarios de S_0 , \mathbf{x} y \mathbf{x}' , se cumplirá que $P(\mathbf{x}) = P(\mathbf{x}')$.

Supongamos que \mathbf{x} y \mathbf{x}' se diferencian en d coordenadas y coinciden en las $c-d$ restantes. Entonces, por el lema 2.3 y como $\sum x_i = \sum x'_i = R_1$, d será un número par y estas d coordenadas se estructurarán en $d/2$ pares de coordenadas tales que si (x_k, x_h) es uno de estos pares en \mathbf{x} , su correspondiente par (x'_k, x'_h) en \mathbf{x}' cumplirá $x'_k = x_k + 1$ y $x'_h = x_h - 1$.

Para demostrar que $P(\mathbf{x}) = P(\mathbf{x}')$ bastará con probar que la contribución a la expresión (2.2) de cualquier par del tipo (x_k, x_h) es la misma que la de su correspondiente par (x'_k, x'_h) . Y esto es cierto, pues según lo anterior,

$x_k + x_h = x'_k + x'_h$ y como $\mathbf{x} \in S_0$ y $\mathbf{x}' \in S_0$ se cumplirá

$$m_{hk}(\mathbf{x}) - 1 \leq x_h \leq m_{hk}(\mathbf{x})$$

$$m_{hk}(\mathbf{x}) - 1 \leq x'_h = x_h - 1 \leq m_{hk}(\mathbf{x})$$

De donde $x_h = m_{hk}(\mathbf{x})$ y teniendo en cuenta (2.5) se llega a

$$x_h (C_k - x_k) = (x_k + 1)(C_h - x_h + 1)$$

Que puede escribirse como

$$x_h! x_k! (C_h - x_h)! (C_k - x_k)! = x'_h! x'_k! (C_h - x'_h)! (C_k - x'_k)!$$

Y el teorema queda demostrado.

II.2.4: Relación entre las Modas de las MHD de una familia particular. Conexión entre modas. Cadena de Modas.

Dados los parámetros R_1, C_1, \dots, C_c de la MHD (2.1), sea \mathfrak{S} la familia formada por todas las MHD

$$H_e \equiv MH(R_1 + e, C_1, \dots, C_c) \quad -R_1 \leq e \leq N - R_1,$$

Con conjunto de referencia

$$F_e = \{ \mathbf{x} = (x_1, \dots, x_c) \in Z_0^c / \sum x_h = R_1 + e, 0 \leq x_h \leq C_h \quad h=1, \dots, c \}$$

Donde las distribuciones extremas de la familia, H_{-R_1} y H_{N-R_1} se reducen a las distribuciones degeneradas en $(0, \dots, 0)$ y en (C_1, \dots, C_c) , respectivamente.

Por otro lado, según la sección anterior, para cada H_e tendremos un subconjunto de F_e , que llamaremos S_e , definido como en (2.8), esto es,

$$S_e = \{ \mathbf{x} = (x_1, \dots, x_c) \in F_e / m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad i \neq j \} \quad (2.11)$$

y con las mismas propiedades que el S_0 de (2.8), en particular, S_e será no vacío y contiene y solo contiene las modas de H_e . Entenderemos que los subconjuntos extremos S_{-R_1} y S_{N-R_1} solo contienen los puntos $(0, \dots, 0)$ y (C_1, \dots, C_c) , respectivamente. Además, es obvio que para $e=0$ la H_0 coincide con la MHD (2.1), con lo que esta distribución pertenece a \mathfrak{S} . Y lógicamente su subconjunto S_0 definido en (2.11) coincide con el (2.8).

Conexión entre las modas de las MHD de \mathfrak{S} .

El estudio de la relación entre las modas de las MHD de \mathfrak{S} nos permitirá desarrollar un método de cálculo de la moda de la MHD (2.1), que veremos en la próxima sección, y también la relación existente entre las modas de las MHD asociadas a los nodos del retículo de M-P, que veremos en el apartado siguiente que constituye este capítulo. En este sentido, vamos a definir una relación de conexión entre las modas de dos MHD contiguas, H_e y H_{e+1} .

Definición 2.1. Decimos que dos puntos, $\mathbf{x} \in S_e$ y $\mathbf{x}' \in S_{e+1}$, están conectados a través de la coordenada v cuando para sus coordenadas se cumple $x_v' - x_v = 1$ y $x_j' = x_j$ para todo $j \neq v$.

Esto es, dos puntos conectados a través de v coinciden en todas sus coordenadas salvo en la coordenada v , en la que se diferencian en una unidad.

Definición 2.2. Decimos que dos puntos, $\mathbf{x} \in S_e$ y $\mathbf{x}' \in S_{e+1}$, están conectados cuando exista una coordenada v tal que ambos puntos estén conectados a través de v .

Dadas dos MHD de \mathfrak{S} , H_e y H_{e+1} , el siguiente teorema nos caracteriza la conexión entre sus modas.

Teorema 2.2. Para cualquier par de distribuciones, H_e y H_{e+1} , pertenecientes a \mathfrak{S} , dado $\mathbf{x} \in S_e$ existirá otro punto $\mathbf{x}' \in S_{e+1}$, ambos conectados a través de la coordenada v , si y solo si v es un valor de j ($1 \leq j \leq c$) que minimiza la expresión $(x_j+1)/(C_j+1)$. Recíprocamente, dado $\mathbf{x}' \in S_{e+1}$ existirá otro punto $\mathbf{x} \in S_e$, ambos conectados a través de v , si y solo si v es un valor de j ($1 \leq j \leq c$) que maximiza la expresión $x_j/(C_j+1)$.

Demostración.-

Veamos la primera parte. Partimos de $\mathbf{x} \in S_e$ y de una coordenada v tal que

$$(x_v+1)/(C_v+1) \leq (x_j+1)/(C_j+1) \quad \forall j \quad (2.12)$$

Consideremos el punto \mathbf{x}' de coordenadas $x'_v = x_v+1$ y $x'_j = x_j \quad \forall j \neq v$.

Tendremos que demostrar que $\mathbf{x}' \in S_{e+1}$ (con lo que \mathbf{x} y \mathbf{x}' estarán conectados a través de v). En efecto, como $\mathbf{x} \in S_e$ se cumple

$$\sum_h x_h = R_1 + e \quad (2.13)$$

$$m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad i \neq j \quad (2.14)$$

de donde podemos deducir, respectivamente

$$\sum x'_h = R_1 + e + 1 \quad (2.15)$$

$$m_{ij}(\mathbf{x}') - 1 \leq x'_i \leq m_{ij}(\mathbf{x}') \quad i \neq j \quad i, j \neq v \quad (2.16)$$

pero de (2.12) tendremos

$$x_v + 1 \leq \frac{(x_j + 1)(C_v + 1)}{C_j + 1}$$

y como $x_v + 1 = x_v'$, $x_j = x_j'$ $j \neq v$

$$x_v' \leq \frac{(x_j' + 1)(C_v + 1)}{C_j + 1}$$

de donde se sigue

$$x_v' \leq \frac{(x_j' + x_v' + 1)(C_v + 1)}{C_v + C_j + 2}$$

o lo que es equivalente

$$x_v' \leq m_{vj}(\mathbf{x}')$$

y por otra parte de (2.14) haciendo $i=v$, se cumple $m_{vj}(\mathbf{x}) - 1 \leq x_v$, con lo que también se cumplirá

$$m_{vj}(\mathbf{x}) + \frac{C_v + 1}{C_v + C_j + 2} - 1 \leq x_v + 1, \quad j \neq v$$

o lo que es equivalente

$$m_{vj}(\mathbf{x}') - 1 \leq x_v', \quad j \neq v$$

con lo que habriamos probado que

$$m_{vj}(\mathbf{x}') - 1 \leq x_v' \leq m_{vj}(\mathbf{x}') \quad v \neq j$$

pero también de (2.12) tendremos

$$(x_i + 1) \geq \frac{(x_v + 1)(C_i + 1)}{C_v + 1}$$

de donde se sigue

$$(x_i + 1) \geq \frac{(x_v + 1 + x_i + 1)(C_i + 1)}{C_i + C_v + 2}$$

o sea $x_i \geq m_{iv}(\mathbf{x}') - 1$, o lo que es equivalente,

$$m_{iv}(\mathbf{x}') - 1 \leq x_i'$$

y por otra parte de (2.14) haciendo $j=v$ se cumple $x_i \leq m_{iv}(\mathbf{x})$, con lo que también se cumplirá $x_i \leq m_{iv}(\mathbf{x}) + (C_i + 1)/(C_i + C_v + 2)$, o lo que es equivalente,

$$x'_i \leq m_{iv}(\mathbf{x}')$$

con lo que habriamos probado que también

$$m_{iv}(\mathbf{x}') - 1 \leq x'_i \leq m_{iv}(\mathbf{x}') \quad i \neq v$$

Con lo que en definitiva la expresión (2.16) vale para todo $i \neq j$ y por tanto de (2.15) tendremos $\mathbf{x}' \in S_{e+1}$.

Por otro lado, para demostrar que $\mathbf{x} \in S_e$ y $\mathbf{x}' \in S_{e+1}$ solo están conectados a través de v si v es un valor de j que minimiza $(x_j+1)/(C_j+1)$, supongamos que es cierto lo contrario, esto es, que están conectados a través de una v tal que $(x_v+1)/C_v+1 > (x_j+1)/C_j+1$ para algún j . Entonces tendremos $(x_v+1)(C_v+C_j+2) > (x_v+x_j+2)(C_v+1)$ y, como ambos puntos están conectados a través de v , esta desigualdad se puede escribir como $x'_v > m_{vj}(\mathbf{x}')$ con lo que \mathbf{x}' no pertenece a S_{e+1} , en contra de la hipótesis inicial.

La segunda parte del teorema (recíproco) se demuestra de forma similar, veámoslo.

Partimos de $\mathbf{x}' \in S_{e+1}$ y de una coordenada v tal que

$$\frac{x'_v}{C_v+1} \geq \frac{x'_j}{C_j+1} \quad j \neq v \quad (2.17)$$

consideremos el punto \mathbf{x} de coordenadas

$$x_v = x'_v - 1 \quad y \quad x_j = x'_j \quad j \neq v$$

tendremos que demostrar que $\mathbf{x} \in S_e$ (con lo que \mathbf{x}' y \mathbf{x} estarán conectados a través de v). En efecto, como $\mathbf{x}' \in S_{e+1}$ se cumple

$$\sum_h x'_h = R_1 + e + 1 \quad (2.18)$$

$$m_{ij}(\mathbf{x}') - 1 \leq x'_i \leq m_{ij}(\mathbf{x}') \quad i \neq j \quad (2.19)$$

de donde podemos deducir , respectivamente

$$\sum_h x_h = R_1 + e \quad (2.20)$$

$$m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad i \neq j \quad i, j \neq v \quad (2.21)$$

pero de (2.17) tendremos

$$x_v' \geq \frac{x_j'(C_v + 1)}{C_j + 1}$$

de donde se implica que

$$x_v' - 1 \geq \frac{x_j'(C_v + 1)}{C_j + 1} - 1$$

y como $x_v = x_v' - 1$, $x_j = x_j'$ $j \neq v$, tendremos

$$x_v \geq \frac{x_j(C_v + 1)}{C_j + 1} - 1$$

de donde se sigue

$$x_v \geq \frac{(x_v + x_j + 1)(C_v + 1)}{C_j + C_v + 2} - 1$$

o sea

$$x_v \geq m_{vj}(\mathbf{x}) - 1$$

y por otra parte, haciendo $i=v$ en (2.19)

$$x_v' \leq m_{vj}(\mathbf{x}')$$

con lo que también se cumplirá

$$x_v' + \frac{C_v + 1}{C_v + C_j + 2} \leq m_{vj}(\mathbf{x}') + 1$$

o sea

$$x_v' + \frac{C_v + 1}{C_v + C_j + 2} \leq \frac{(x_v' + x_j' + 1)(C_v + 1)}{C_v + C_j + 2} + 1$$

de donde se implica

$$x_v' - 1 \leq \frac{(x_v' + x_j' + 1 - 1)(C_v + 1)}{C_v + C_j + 2}$$

$$x_v \leq \frac{(x_v + x_j + 1)(C_v + 1)}{C_v + C_j + 2}$$

lo que equivale a

$$x_v \leq m_{vj}(\mathbf{x})$$

con lo que habíamos probado

$$m_{vj}(\mathbf{x}) - 1 \leq x_v \leq m_{vj}(\mathbf{x}) \quad j \neq v$$

pero también de (2.17) tendremos

$$x_i \leq \frac{(x_v + 1)(C_i + 1)}{C_v + 1}$$

con lo que se cumplirá

$$x_i \leq \frac{(x_v + x_i + 1)(C_i + 1)}{C_i + C_v + 2}$$

o lo que es equivalente

$$x_i \leq m_{iv}(\mathbf{x}) \quad v \neq i$$

y por otra parte de (2.19) haciendo $j=v$ tenemos

$$m_{iv}(\mathbf{x}') - 1 \leq x'_i$$

con lo que también

$$m_{iv}(\mathbf{x}') - (C_i + 1)/(C_i + C_v + 2) - 1 \leq x'_i, \text{ o sea } (x_i + x_v + 1)(C_i + 1)/(C_i + C_v + 2) - 1 \leq x_i$$

o lo que es equivalente

$$m_{iv}(\mathbf{x}) - 1 \leq x_i$$

con lo que habíamos probado que también

$$m_{iv}(\mathbf{x}) - 1 \leq x_i \leq m_{iv}(\mathbf{x}) \quad v \neq i$$

Con lo que en definitiva la expresión (2.21) vale para todo $i \neq j$ y por tanto de (2.20)

tendremos $\mathbf{x} \in S_e$.

Por otro lado, para demostrar que $\mathbf{x}' \in S_{e+1}$ y $\mathbf{x} \in S_e$ solo están conectados a través de v si v es un valor de j que maximiza $\frac{x_j'}{C_j+1}$, supongamos que es cierto lo contrario,

esto es, que están conectados a través de una v tal que

$$\frac{x_v'}{C_v+1} < \frac{x_j'}{C_j+1} \quad \text{para algún } j$$

entonces tendremos

$$\frac{x_v'}{C_v+1} < \frac{x_v' + x_j'}{C_v + C_j + 2}$$

de donde se implica

$$x_v' < \frac{(x_v' - 1 + x_j + 1)(C_v + 1)}{C_v + C_j + 2}$$

$$x_v' - 1 < \frac{(x_v' - 1 + x_j' + 1)(C_v + 1)}{C_v + C_j + 2} - 1$$

lo que nos permite escribir

$$x_v < m_{vj}(\mathbf{x}) - 1$$

con lo que $\mathbf{x} \notin S_e$ en contra de la hipótesis inicial.

De este teorema deducimos los siguientes corolarios.

Corolario 2.1. Dados dos subconjuntos, S_e y S_{e+1} , cada punto de uno de ellos está conectado con al menos un punto del otro, es decir, dadas las distribuciones H_e y H_{e+1} de \mathfrak{S} , cada moda de una de ellas está conectada con al menos una moda de la otra.

Demostración.-

Si $\mathbf{x} = (x_1, \dots, x_c)$ es una moda de H_e siempre existirá (aunque no tiene por qué ser único) un valor v de j ($1 \leq j \leq c$) que minimice la expresión $(x_j + 1)/(C_j + 1)$, con lo

que por el teorema anterior $\mathbf{x}' = (x'_1, \dots, x'_c)$ $x'_j = x_j$ $j \neq v$, $x'_v = x_v + 1$, será una moda de H_{e+1} conectada con la moda de H_e a través de la coordenada v .

También si $\mathbf{x}' = (x'_1, \dots, x'_c)$ es una moda de H_{e+1} siempre existirá (aunque no tiene por qué ser único) un valor v de j ($1 \leq j \leq c$) que maximice la expresión $x_j / (C_j + 1)$, con lo que por el teorema anterior $\mathbf{x} = (x_1, \dots, x_c)$ $x_j = x'_j$ $j \neq v$, $x_v = x'_v - 1$, será una moda de H_e conectada con la moda \mathbf{x}' de H_{e+1} a través de la coordenada v .

Corolario 2.2. Dos puntos $\mathbf{x} \in S_e$ y $\mathbf{x}' \in S_{e+1}$ conectados a través de v , siempre cumplirán para todo j

$$\frac{x_j}{C_j + 1} \leq \frac{x'_v}{C_v + 1} \leq \frac{x_j + 1}{C_j + 1}$$

Demostración.-

Para $\mathbf{x} \in S_e$ tenemos

$$m_{ij}(\mathbf{x}) - 1 \leq x_i \leq m_{ij}(\mathbf{x}) \quad i \neq j$$

luego para $i=v$, $m_{vj}(\mathbf{x}) - 1 \leq x_v$, o sea $\frac{(x_v + x_j + 1)(C_v + 1)}{C_v + C_j + 2} \leq x_v + 1$, pero esto equivale

$$\frac{x_j(C_v + 1)}{C_j + 1} \leq x_v + 1, \text{ de donde obtenemos, al estar } \mathbf{x} \text{ conectado con } \mathbf{x}' \text{ a través de } v \text{ y}$$

verificarse $x'_v = x_v + 1$, que $\frac{x_j}{C_j + 1} \leq \frac{x'_v}{C_v + 1}$.

Y por otra parte, para $\mathbf{x}' \in S_{e+1}$ tenemos

$$m_{ij}(\mathbf{x}') - 1 \leq x'_i \leq m_{ij}(\mathbf{x}') \quad i \neq j$$

luego para $i=v$, $x'_v \leq m_{vj}(\mathbf{x}')$, o sea $x'_v \leq \frac{(x'_v + x'_j + 1)(C_v + 1)}{C_v + C_j + 2}$, pero esto equivale a

$$x'_v \leq \frac{(x'_j + 1)(C_v + 1)}{C_j + 1}, \text{ de donde obtenemos al estar } \mathbf{x}' \text{ conectado con } \mathbf{x} \text{ a través de } v \text{ y}$$

verificarse $x'_j = x_j$ $j \neq v$, que $\frac{x'_v}{C_v + 1} \leq \frac{x_j + 1}{C_j + 1}$

Definición 2.3. Dados dos valores enteros, w y w' , tales que

$$-R_1 \leq w < w' \leq N-R_1$$

y la sucesión finita de puntos

$$C(w,w') = \{x_e \in S_e, e = w, w+1, \dots, w'\}$$

decimos que $C(w,w')$ es una cadena cuando para cada e ($e=w, w+1, \dots, w'-1$) x_e y x_{e+1} están conectados. También, dados dos puntos cualesquiera de la cadena $C(w,w')$, siempre diremos que ambos puntos están comunicados a través de dicha cadena. Esto es, para cada valor de e tendremos un punto x_e (o eslabón) de la cadena, que será un punto de S_e , y cada eslabón está conectado con el anterior y el posterior. Además, dados dos puntos arbitrarios de la cadena, x_a y x_b ($a < b$), podremos pasar de x_a a x_b (y viceversa) a través de los sucesivos eslabones intermedios de la cadena, lo que implica realizar $b-a$ pasos de x_e a x_{e+1} (o de x_{e+1} a x_e) $a \leq e < b$.

Teorema 2.3. Sean los subconjuntos S_w y $S_{w'}$ ($w < w'$) correspondientes a dos MHD de \mathfrak{S} . Dado un punto arbitrario de uno de los subconjuntos, siempre existirá un punto del otro subconjunto tal que ambos estén comunicados a través de una cadena.

Demostración.-

Como cada MHD H_e de \mathfrak{S} tiene un subconjunto S_e , tendremos la serie de subconjuntos $S_w, S_{w+1}, \dots, S_{w'}$ y, según el corolario 2.1, partiendo de un $x_w \in S_w$ siempre podremos tomar un punto de cada subconjunto que forman una sucesión finita $x_w, x_{w+1}, \dots, x_{w'}$, tal que cada uno de ellos está conectado con el anterior y el posterior, con lo que formarán una cadena $C(w,w')$. Ello demuestra que dado un $x_w \in S_w$ siempre existirá un $x_{w'} \in S_{w'}$, ambos comunicados a través de una cadena. Igualmente se ve partiendo de un $x_{w'} \in S_{w'}$.

Corolario 2.3. Dadas dos MHD arbitrarias de \mathfrak{S} , cada una de las modas de una de ellas está siempre comunicada a través de una cadena con al menos una moda de la otra.

Demostración.-

Sean H_w y $H_{w'}$ ($w < w'$) dos MHD arbitrarias de \mathfrak{S} , y respectivamente $S_w, S_{w'}$ sus subconjuntos correspondientes, una moda de H_w es un punto de S_w y una moda de $H_{w'}$ es un punto de $S_{w'}$, luego aplicando el teorema anterior obtenemos el resultado buscado.

Teorema 2.4. Dados $\mathbf{x}_w \in S_w$, de coordenadas x_{wj} , y $\mathbf{x}_{w'} \in S_{w'}$, de coordenadas $x_{w'j}$, ($w' = w + d$ $d > 0$) ambos comunicados a través de una cadena $C(w, w')$, entonces se cumple que $x_{w'j} = x_{wj} + q_{jd}$, y esto se cumple tanto si pasamos de \mathbf{x}_w a $\mathbf{x}_{w'}$ como si pasamos de $\mathbf{x}_{w'}$ a \mathbf{x}_w , donde q_{jd} es el número de pasos o parejas de puntos $\mathbf{x}_e, \mathbf{x}_{e+1}$ ($w \leq e < w'$) de $C(w, w')$ en los que ambos puntos están conectados a través de j , cumpliéndose siempre $0 \leq q_{jd} \leq d$ y $\sum_j q_{jd} = d$.

Demostración.-

Este teorema se deduce del hecho de que dados dos puntos, \mathbf{x}_w y $\mathbf{x}_{w'}$, comunicados por una cadena $C(w, w')$ podemos pasar de uno de ellos al otro a través de los eslabones intermedios de la cadena. Y ello implica el realizar $d = w' - w$ pasos de \mathbf{x}_e a \mathbf{x}_{e+1} (o de \mathbf{x}_{e+1} a \mathbf{x}_e) $w \leq e < w'$, donde en cada uno de estos pasos (y por la definición de puntos conectados) la coordenada j a través de la cual están conectados ambos puntos \mathbf{x}_e y \mathbf{x}_{e+1} aumenta en una unidad al pasar de \mathbf{x}_e a \mathbf{x}_{e+1} (o disminuye una unidad al pasar de \mathbf{x}_{e+1} a \mathbf{x}_e) permaneciendo constantes el resto de las coordenadas. Por ello $x_{w'j} = x_{wj} + q_{jd}$ con $q_{jd} \geq 0$. Y como la coordenada j no es

necesariamente la misma en todos los pasos (de hecho suele ser distinta en cada paso) se tiene que $q_{jd} \leq d$. Finalmente, por lo dicho antes, es inmediato que $\sum_j q_{jd} = d$.

II.2.5: Nuevo método de cálculo de la moda de una MHD.

El método que proponemos en este trabajo para el cálculo de la moda de la MHD (2.1) está basado en las modas de las distribuciones Hipergeométricas marginales x_j $j=1, \dots, c$, que vienen dadas por la bien conocida expresión (2.3).

Sea M el conjunto de puntos cuyas coordenadas son las modas de las distribuciones Hipergeométricas marginales de (2.1), esto es,

$$M = \{z = (z_1, \dots, z_c) \in Z_0^c / R(C_j+1) - 1 \leq z_j \leq R(C_j+1) \ j=1, \dots, c\} \quad (2.22)$$

Este conjunto tendrá un solo punto cuando la moda de cada marginal sea única. En otro caso contendrá 2^r puntos, donde r es el número de marginales x_j cuya moda no es única. Dado un $z \in M$, notaremos $\delta = \sum z_j - R_1$. El siguiente Lema nos proporciona cotas para este δ .

Lema 2.4. Dada la MHD (2.1) y cualquier punto $z \in M$, siempre se cumple que

$$\max(-R_1, R'-c) \leq \delta \leq \min(N-R_1, R') \quad (2.23)$$

donde $R' = R(c-2) + 1$

Demostración.-

De (2.22) obtenemos directamente un máximo y un mínimo para $\sum z_j$ y de ello deducimos fácilmente que

$$-[(1-R)(c-2) + 1] \leq \delta \leq R(c-2) + 1$$

Pero al ser $z_j \geq 0$, tendremos que $\sum z_j \geq 0$, de donde $-R_1 \leq \delta$. Igualmente, al ser según (2.22) $z_j \leq C_j$, tendremos $\sum z_j \leq N$ y por tanto $\delta \leq N - R_1$.

Por otro lado, para un valor particular de δ que cumpla (2.23), definimos el subconjunto

$$M_\delta = \{z \in M / \sum z_j - R_1 = \delta\}$$

En el siguiente teorema y corolario daremos significado a estos subconjuntos dentro de la familia \mathfrak{S} . Y veremos como cada M_δ que sea no vacio nos puede servir como punto de partida para el cálculo de las modas de (2.1).

Teorema 2.5. Dada la MHD (2.1) y la familia \mathfrak{S} , para cada entero δ tal que M_δ es no vacio, se cumple que M_δ coincide con el subconjunto S_δ correspondiente a la MHD H_δ de \mathfrak{S} .

Demostración.-

Sea $z \in M_\delta$, entonces se cumplirá

$$0 \leq z_j \leq C_j, \quad \sum z_j = R_1 + \delta \quad (2.24)$$

y

$$R(C_{j+1}) - 1 \leq z_j \leq R(C_{j+1}) \quad j=1, \dots, c \quad (2.25)$$

de esta expresión (2.25) se deduce

$$z_i / (C_i + 1) \leq R \leq (z_j + 1) / (C_j + 1) \quad i \neq j \quad (2.26)$$

Igualmente se deduce (cambiando los subíndices)

$$z_j / (C_j + 1) \leq R \leq (z_i + 1) / (C_i + 1) \quad j \neq i \quad (2.27)$$

Ahora de (2.26) y (2.27) se obtienen, respectivamente

$$z_i \leq m_{ij}(z) \quad i \neq j \quad (2.28)$$

y

$$m_{ij}(z) - 1 \leq z_i \quad i \neq j \quad (2.29)$$

y finalmente, de (2.24), (2.28) y (2.29) tendremos que $\mathbf{z} \in S_\delta$, por tanto S_δ contiene a los puntos de M_δ .

Nos queda por probar que S_δ solo contiene a los puntos de M_δ .

Esto es trivial si S_δ solo contiene un punto \mathbf{z} , pues \mathbf{z} sería de M_δ . Consideremos que contiene mas de un punto. Sea $\mathbf{z}' \in S_\delta$, tendremos que demostrar que $\mathbf{z}' \in M_\delta$.

Para ello habrá que probar que

$$R(C_j+1) - 1 \leq z_j' \leq R(C_j+1) \quad j=1, \dots, c \quad (2.30)$$

Por lo dicho anteriormente, siempre existirá un punto $\mathbf{z} \in S_\delta$ tal que \mathbf{z} sea un punto de M_δ . Supongamos que \mathbf{z} y \mathbf{z}' se diferencian en d coordenadas y coinciden en las $c-d$ restantes. Obviamente, para estas $c-d$ coordenadas de \mathbf{z}' , como \mathbf{z} es un punto de M_δ , se cumplirá (2.30). Por otro lado, por el lema 2.3 y como $\sum z_j = \sum z_j'$, d será un número par y esas d coordenadas se estructurarán en $d/2$ pares de coordenadas tales que si (z_i, z_j) es uno de estos pares en \mathbf{z} , su correspondiente par (z_i', z_j') en \mathbf{z}' cumplirá $z_i' = z_i+1$ y $z_j' = z_j-1$. Ahora, para cada uno de estos pares, por ser $\mathbf{z}' \in S_\delta$, se cumplirá

$$m_{ij}(\mathbf{z}')-1 \leq z_i' \leq m_{ij}(\mathbf{z}')$$

de donde deducimos

$$z_i'/(C_i+1) \leq (z_j'+1)/(C_j+1)$$

o lo que es igual

$$(z_i+1)/(C_i+1) \leq z_j'/(C_j+1) \quad (2.31)$$

y como \mathbf{z} es un punto de M_δ se cumplirá (2.27). Entonces de (2.27) y (2.31) se obtiene

$$\frac{z_i+1}{C_i+1} = \frac{z_j}{C_j+1} = R$$

con lo que finalmente tendremos $z_i' = R(C_i+1)$ y $z_j' = R(C_j+1)-1$.

Así pues, las d coordenadas de \mathbf{z}' en las que \mathbf{z}' y \mathbf{z} no coinciden también verifican (2.30) y el teorema queda demostrado.

De este teorema, se deduce el siguiente corolario.

Corolario 2.4. Dada la MHD (2.1), la familia \mathfrak{S} y un M_δ no vacío, se cumple que M_δ contiene y solo contiene las modas de la $MH(R_1+\delta; C_1, \dots, C_c)$ de \mathfrak{S} . Y, en particular, si M_0 es no vacío, M_0 contiene y solo contiene las modas de la MHD (2.1).

Demostración.-

M_δ por el teorema anterior coincide con el subconjunto S_δ correspondiente a la $MH(R_1+\delta; C_1, \dots, C_c)$, y como ya ha quedado probado en el apartado II.2.4 que S_e contiene y solo contiene las modas de H_e , $H_e \equiv (R_1+e; C_1, \dots, C_c)$ $-R_1 \leq e \leq N-R_1$, en definitiva M_δ contiene y solo contiene las modas de la $MH(R_1+\delta; C_1, \dots, C_c)$. Y en particular, si M_0 es no vacío coincide con el subconjunto S_0 correspondiente a la MH (2.1), y por el teorema 2.1 sabemos que S_0 contiene y solo contiene las modas de la MH (2.1), con lo que en definitiva M_0 contiene y solo contiene las modas de la MHD(2.1)

Como hemos visto, por una parte (según teorema 2.1) el subconjunto S_0 contiene y solo contiene las modas de la MHD (2.1) y por otra parte (según teorema 2.5) para un entero δ tal que M_δ sea no vacío, este subconjunto coincide con S_δ . También sabemos (según teorema 2.3) que cada punto de S_0 está comunicado con algún punto de S_δ y viceversa, lo que nos permite partir de un punto de S_δ (o M_δ) y llegar a través de una cadena hasta un punto de S_0 y, en definitiva, partiendo de todos los puntos de M_δ llegaremos, a través de diferentes cadenas, a todos los

puntos de S_0 , es decir, al conjunto de todas las modas de la MHD (2.1). Y como ya hemos visto, el paso de un punto de S_δ (o M_δ) a un punto de S_0 se realiza a través de los sucesivos pasos intermedios (o eslabones) de la cadena y cada uno de estos pasos está regulado por el teorema 2.2.

Como dijimos al principio de esta sección, si las modas de todas las marginales x_j son únicas, M tendrá un solo punto y por tanto solo existirá un M_δ no vacío. En estos casos es claro que hay que partir de este M_δ para llegar a S_0 . Pero si, por el contrario, las modas de las marginales no son todas únicas, tendremos varios subconjuntos M_δ no vacíos. Por el teorema 2.3 es claro que podemos partir de cualquiera de estos M_δ , llegando siempre al conjunto de todas las modas de la MHD (2.1). En cualquier caso, en adelante llamaremos δ^* al valor δ del subconjunto M_δ de partida.

Es obvio que en la práctica tomaremos como δ^* el valor de δ que esté más próximo a 0, así tendremos que realizar menos pasos intermedios. En particular si $\delta^*=0$ es porque M_0 es no vacío, y no tendríamos que hacer ningún paso intermedio, M_0 sería ya el conjunto de todas las modas de la MHD (2.1), como ya dijimos en el corolario 2.4.

Supongamos que hay r marginales x_j ($r \leq c$) cuyas modas no son únicas, entonces de (2.3) tendremos para cada una de estas x_j que $R(C_j+1)$ será entero y habrá dos modas: $R(C_j+1)$ y $R(C_j+1)-1$. En esta situación, tomando como modas para estas r marginales $z_j = R(C_j+1)-1$ y $z_j = R(C_j+1)$ obtendríamos, respectivamente, el mínimo valor (δ_1) y el máximo valor (δ_2) de δ tal que M_δ es no vacío. Y para cualquier δ tal que $\delta_1 \leq \delta \leq \delta_2$ será M_δ no vacío.

Si $\delta_1 \leq 0$ y $\delta_2 \geq 0$ entonces M_0 será no vacío y ya tendremos el conjunto de modas de la MHD (2.1). Si $0 < \delta_1 < \delta_2$ tomaremos $\delta^* = \delta_1$ y si $\delta_1 < \delta_2 < 0$ tomaremos $\delta^* = \delta_2$.

Consideremos ya $M_{\delta^*} (\equiv S_{\delta^*})$ para $\delta^* \neq 0$ el subconjunto de partida. Además definiremos

$$\lambda = \begin{cases} 1 & \text{si } \delta^* < 0 \\ -1 & \text{si } \delta^* > 0 \end{cases} \quad (2.32)$$

Si partimos de un punto $\dot{\mathbf{z}} \in M_{\delta^*}$ llegaremos a un punto $\dot{\mathbf{x}} \in S_0$ después de realizar $m = |\delta^*|$ pasos a través de una cadena. Las coordenadas de $\dot{\mathbf{x}}$ (moda de (2.1)) se obtienen a partir de las coordenadas de $\dot{\mathbf{z}}$, aplicando el teorema 2.4.

Describamos ahora este proceso paso a paso. Para cada paso k definiremos $d_{jk} = 1$ si los dos puntos que corresponden al paso k están conectados a través de la coordenada j y $d_{jk} = 0$ si ambos puntos están conectados a través de otra coordenada $j' \neq j$. Además

$$q_{je} = \sum_{k=1}^e d_{jk} \quad (2.33)$$

Observemos que este q_{je} es el mismo que el q_{jd} del teorema 2.4. En el primer paso, aplicando el teorema 2.2, tomaremos un punto de $S_{\delta^* + \lambda}$ conectado con $\dot{\mathbf{z}}$. Este punto tendrá de coordenadas $\dot{z}_j + \lambda q_{j1}$. Así, después de realizado el paso e ($e < m$) llegaremos a un punto de $S_{\delta^* + e\lambda}$ de coordenadas $\dot{z}_j + \lambda q_{je}$. Ahora en el paso $e+1$, aplicando el teorema 2.2, tomaremos un punto de $S_{\delta^* + (e+1)\lambda}$ conectado al anterior, que tendrá de coordenadas $\dot{z}_j + \lambda q_{j(e+1)}$. Y finalmente, después de realizado el paso m , tendremos $\delta^* + m\lambda = 0$ y habremos llegado a un punto $\dot{\mathbf{x}} \in S_0$ de coordenadas $\dot{z}_j + \lambda q_{jm}$, que será una moda de la MHD (2.1). Desde el punto de vista práctico y de forma más simple, en cada paso partimos del punto correspondiente al

paso anterior (llamemos z'_j a sus coordenadas) y, aplicando el teorema 2.2, sumamos 1 a una coordenada que tenga mínimo valor de $(z'_j + 1)/(C_j + 1)$ si $\delta^* < 0$ ó restamos 1 a una coordenada que tenga máximo valor de $z'_j/(C_j + 1)$ si $\delta^* > 0$. Además, como todo punto de S_0 está necesariamente comunicado con algún punto de M_{δ^*} , sus coordenadas quedarán siempre expresadas en la forma anterior.

Este proceso descrito nos demuestra realmente el siguiente teorema, que nos resume la expresión de las modas de la MHD (2.1).

Teorema 2.6. Dado $\dot{z} \in M_{\delta^*}$ para el entero δ^* definido anteriormente, el punto \dot{x} de coordenadas

$$\dot{x}_j = \begin{cases} \dot{z}_j & \text{si } \delta^* = 0 \\ \dot{z}_j + \lambda q_{jm} & \text{en otro caso} \end{cases} \quad (2.34)$$

($j=1, \dots, c$) es un máximo(modal) de la MHD (2.1), donde $m = |\delta^*|$ y λ y q_{jm} están definidos en (2.32) y (2.33), respectivamente. Y, recíprocamente, cualquier máximo(modal) de la MHD (2.1) vendrá expresado en la forma (2.34) para algún $\dot{z} \in M_{\delta^*}$.

Es fácil ver que para $c=2$ este método de cálculo de la moda se reduce a la bien conocida expresión de la moda de una Hipergeométrica, dada en (1.38), pues para $c=2$

$$M = \{z = (z_1, z_2) \in Z_0^2 / R(C_j+1)-1 \leq z_j \leq R(C_j+1) \quad j=1,2\}$$

con lo que los puntos de este conjunto verifican $\sum z_j = R_1$, coinciden con la moda de la hipergeométrica y el proceso ha terminado.

Los resultados obtenidos en los puntos anteriores son extrapolables directamente a tablas de contingencia $2 \times c$, y sus hechos mas esenciales pueden ser

resumidos utilizando el lenguaje específico para tablas de contingencia bajo la siguiente forma.

Sea

$$\Omega = \{X / X \text{ es } 2 \times c, \sum_{j=1}^c x_{ij} = R_i \ i=1,2, \sum_{i=1}^2 x_{ij} = C_j \ j=1,\dots,c\}$$

Denotemos por

$$\Omega_0 = \{X \in \Omega / \text{todas las subtablas } 2 \times 2 \text{ de } X \text{ son de máxima probabilidad}\}$$

Lema 2.2*.- Si X es una tabla de máxima probabilidad de Ω entonces $X \in \Omega_0$.

Lema 2.3*.- Si $X = (x_{ij})_{i=1,2, j=1,\dots,c}$ e $X' = (x'_{ij})_{i=1,2, j=1,\dots,c}$ son dos tablas pertenecientes a Ω_0 entonces $|x_{ij} - x'_{ij}| \leq 1, i=1,2, j=1,\dots,c$.

Teorema 2.1*.- La condición necesaria y suficiente para que una tabla de Ω sea de máxima probabilidad es que pertenezca a Ω_0 . O sea, es condición necesaria y suficiente para que una tabla perteneciente a una familia isomarginal de tablas $2 \times c$ sea de máxima probabilidad que todas sus subtablas 2×2 sean de máxima probabilidad.

Teorema 2.2*.- Sean Φ y Φ' dos familias de tablas $2 \times c$ con marginales filas consecutivos dados y marginales columnas coincidentes dados. Φ familia de tablas $2 \times c$ con marginales filas H_1 y H_2 y marginales columnas C_1, \dots, C_c , y Φ' familia de tablas $2 \times c$ con marginales filas H'_1 y H'_2 y marginales columnas C'_1, \dots, C'_c , $C'_j = C_j \ \forall j$, $H'_1 = H_1 + 1, H'_2 = H_2 - 1$. Si $X = (x_{ij})_{i=1,2, j=1,\dots,c}$ es una tabla de máxima probabilidad de Φ , entonces $X' = (x'_{ij})_{i=1,2, j=1,\dots,c}$, $x'_{ij} = x_{ij} \ j \neq v$, $x'_{1v} = x_{1v} + 1, x'_{2v} = x_{2v} - 1$ es una tabla de máxima probabilidad de Φ' si y solo si v es un valor de j que minimiza $\{x_{1j} + 1 / C_j + 1\}_{j=1,\dots,c}$. Y recíprocamente, si $X' = (x'_{ij})_{i=1,2, j=1,\dots,c}$ es una tabla de máxima probabilidad de Φ' entonces $X = (x_{ij})_{i=1,2, j=1,\dots,c}$

$j=1,\dots,c$, $x_{ij} = x'_{ij}$ $j \neq v$, $x_{1v} = x'_{1v} - 1$, $x_{2v} = x'_{2v} + 1$ es una tabla de máxima probabilidad de Φ si y solo si v es un valor de j que maximiza $\{x'_{ij} / C_{j+1}\}_{j=1,\dots,c}$.

Si tuviéramos dos familias de tablas $2 \times c$ con marginales columnas coincidentes dados, C_1, \dots, C_c , y marginales filas dados respectivamente por H_1, H_2 y $H_1 + \delta, H_2 - \delta$, $\delta \in \mathbb{Z}_0^+$ el teorema anterior aplicado repetidamente nos permite a partir de una tabla de máxima probabilidad de la primera familia acceder a una tabla de máxima probabilidad de la segunda familia y recíprocamente.

Obtenida la caracterización de las tablas de máxima probabilidad de una familia de tablas $2 \times c$ de marginales dados y la relación entre las tablas de máxima probabilidad para familias $2 \times c$ con marginales columnas coincidentes, abordamos la posibilidad de aplicar los resultados obtenidos a la elaboración de un método que nos permita determinar las tablas de máxima probabilidad de la familia Ω , para ello tomamos, como elemento básico de partida, el conjunto de tablas cuyas entradas son máximos de las distribuciones hipergeométricas marginales x_{ij}

$$H = \left\{ Z = (z_{ij})_{i=1,2, j=1,\dots,c} / \frac{(R_i + 1)(C_j + 1)}{N + 2} - 1 \leq z_{ij} \leq \frac{(R_i + 1)(C_j + 1)}{N + 2} \right\}$$

Las tablas pertenecientes a este conjunto siguen teniendo por marginales columnas

C_1, \dots, C_c , y por marginales filas $\left\{ \sum_{j=1}^c z_{ij} \right\}$. Para cada valor de estos marginales filas, y

con los marginales columnas considerados, tenemos una familia de tablas $2 \times c$, las tablas de máxima probabilidad de cada una de esas familias coinciden con las tablas de H que tienen esos totales marginales filas. Por consiguiente si para una de estas familias sus marginales filas coinciden con los de la familia dada Ω , las tablas de máxima probabilidad buscadas serían las correspondientes del conjunto H , y en todo caso, una cualquiera de esas familias y la dada tendrían los mismos totales

marginales columnas y las tablas de máxima probabilidad de las dos familias estarían relacionadas como se ha indicado anteriormente, y a partir de la correspondiente tabla de H se podría acceder a la buscada. Eligiendo la familia cuyos marginales estén mas próximos a los de la familia dada resolveremos de forma mas rápida el problema.

II.2.6: Comparación del nuevo método de cálculo de la Moda de una MHD con los métodos previamente existentes.

Tomaremos como método referencial para realizar la comparación el de Harry Joe por ser, como ha quedado apuntado con anterioridad, el reconocido como más eficaz de los previamente existentes.

El método de Harry Joe consiste en obtener cotas que delimitan superior e inferiormente los posibles valores (x_{ij}) de una tabla de máxima probabilidad, pero esta acotación como ha quedado reflejado en el capítulo I, en numerosas ocasiones es excesivamente laxa y deja filtrar un elevado número de tablas que en muchos casos no pertenecen ni tan siquiera a la familia, y en otros no verifican la condición de que todas sus subtablas 2x2 sean de máxima probabilidad, y nos conduce a un conjunto con un elevado número de tablas donde determinando individualmente sus probabilidades hay que seleccionar la de probabilidad máxima. La eficacia del método está fuertemente ligada a que los marginales de la tabla proporcionen en algún caso cotas que delimiten un único valor entero, con lo que los valores de estas entradas quedan determinados, y hay posibilidad de incorporarlos en una etapa sucesiva para ajustar más la posible variación de los restantes, y también y de manera decisiva está ligado a que c no sea grande.

Respecto a las características y aplicabilidad del nuevo método se pueden reseñar los siguientes aspectos:

- En general es este un método recursivo que consta de $m=|\delta^*|$ pasos, aunque esto es un poco engañoso, pues en la mayoría de los casos es en realidad un método directo (no recursivo) o pseudo-directo. En este sentido, el Lema 2.4 da cotas para δ que no están muy lejos de 0. Además, la experiencia nos muestra que rara vez el δ se acerca a las cotas, mas bien tiende a estar próximo a 0. Por todo ello y como δ^* se toma como el mas próximo a 0 de los valores de δ que se obtengan, en una mayoría de los casos m será 0 o tomará un valor muy próximo a 0. Si $m \leq 1$ el método es en realidad un método directo (como en los ejemplos 1 y 2 que desarrollamos mas abajo). Finalmente, cuando $m > 1$, en muchos casos los m pasos pueden realizarse simultáneamente (como en el ejemplo 3 que desarrollamos mas abajo) y sería, por llamarlo así, un método pseudo-directo.

- Este nuevo método es tan simple que incluso puede resolverse a mano o con una calculadora de bolsillo en muy pocos minutos. Además, es completamente general y se aplica con la misma facilidad para cualquier valor de c y para cualesquiera valores de los marginales o cualesquiera valores de los parámetros de la MHD.

- El nuevo método propuesto constituye una generalización natural al caso multivariante ($c > 2$) de la bien conocida expresión de la moda de una distribución Hipergeométrica. Igualmente, es fácil ver que este método nos proporciona exactamente el resultado dado en Boland and Proschan (1987) para el caso particular de MHD allí propuesto, y el dado en el teorema 1.3 de M-P para el caso de C_j iguales.

Boland and Proschan establecen que la distribución hipergeométrica multivariante con parámetros $R_1 = k_1 + \dots + k_c$, y $C_j = mk_j$, $j=1, \dots, c$, m entero positivo, k_j entero positivo para todo j , tiene su moda en $\mathbf{x} = (k_1, \dots, k_c)$, para probar que nuestro método nos conduce a este mismo resultado bastará con verificar que los k_j son modas respectivamente de las hipergeométricas $H_j(R_1, mk_j, N)$ para cada j ,

$N = \sum_{j=1}^c mk_j$, o sea que

$$\frac{(R_1 + 1)(mk_j + 1)}{N + 2} - 1 \leq k_j \leq \frac{(R_1 + 1)(mk_j + 1)}{N + 2}$$

o lo que es equivalente

$$\frac{\left(\sum_{i=1}^c k_i + 1\right)(mk_j + 1)}{m \sum_{i=1}^c k_i + 2} - 1 \leq k_j \leq \frac{\left(\sum_{i=1}^c k_i + 1\right)(mk_j + 1)}{m \sum_{i=1}^c k_i + 2}$$

y esto se deduce de la siguiente forma:

En primer lugar tenemos

$$k_j \leq \frac{\left(\sum_{i=1}^c k_i + 1\right)(mk_j + 1)}{m \sum_{i=1}^c k_i + 2} \Leftrightarrow 2k_j \leq k_j(m + 1) + \sum_{i \neq j} k_i + 1$$

y esta última desigualdad es cierta evidentemente pues $m+1 \geq 2$ y $\sum_{i \neq j} k_i + 1 > 0$.

y por otra parte

$$\frac{\left(\sum_{i=1}^c k_i + 1\right)(mk_j + 1)}{m \sum_{i=1}^c k_i + 2} - 1 \leq k_j \Leftrightarrow \frac{\left(\sum_{i=1}^c k_i + 1\right)(mk_j + 1)}{m \sum_{i=1}^c k_i + 2} \leq k_j + 1 \Leftrightarrow \sum_{i \neq j} k_i \leq m \sum_{i \neq j} k_i + k_j + 1$$

y esta última desigualdad es cierta evidentemente pues $m \geq 1$ y $k_j + 1 > 0$.

Para probar que nuestro método también conduce, para el caso de C_j iguales, al mismo resultado dado en el teorema 1.3 de Mehta and Patel podemos seguir el proceso que figura a continuación.

Si consideramos la MHD de parámetros R_1 y $C_j = C$ para todo $j, j=1, \dots, c, c \geq 2$. Para todo x_j tenemos idéntica distribución marginal, la hipergeométrica $H(R_1, C, cC)$,

cuya moda viene dada por $z_j, \frac{(R_1+1)(C+1)}{cC+2} - 1 \leq z_j \leq \frac{(R_1+1)(C+1)}{cC+2}$, para

conseguir el resultado buscado, comencemos por probar que

$$\frac{(R_1+1)(C+1)}{cC+2} - 1 \leq \frac{R_1}{c} \leq \frac{(R_1+1)(C+1)}{cC+2}$$

en primer lugar,

$$\frac{R_1}{c} \leq \frac{(R_1+1)(C+1)}{cC+2} \Leftrightarrow R_1 c C + 2R_1 \leq c(R_1 C + 1 + R_1 + C) \Leftrightarrow 2R_1 \leq cR_1 + c(C+1)$$

y esta última desigualdad es cierta evidentemente pues $c \geq 2$ y $c(C+1) > 0$.

y, por otra parte

$$\frac{(R_1+1)(C+1)}{cC+2} - 1 \leq \frac{R_1}{c} \Leftrightarrow \frac{(R_1+1)(C+1)}{cC+2} \leq \frac{R_1+c}{c} \Leftrightarrow (c-1)R_1 \leq R_1 + (c-1)cC + c$$

y esta última desigualdad es cierta pues al ser $R_1 \leq cC$ resulta $(c-1)R_1 \leq (c-1)cC$, y también tenemos $R_1 + c > 0$.

El resultado que acabamos de establecer nos permite concluir en primer lugar que si R_1/c es entero sería la moda de cada una de las $H(R_1, C, cC)$, con lo que para el vector $\mathbf{z} = (z_1, \dots, z_c)$ formado con las modas de estas hipergeométricas se verifica $\sum z_j = R_1$, y el método que hemos propuesto nos conduce a que la moda de $MH(R_1;$

$C, \dots, C)$ vendría dada por $(\frac{R_1}{c}, \dots, \frac{R_1}{c})$ que es también la moda dada por Mehta and

Patel.

Y para otro caso, o sea si $\frac{R_1}{c}$ no es entero, caso de que $\frac{(R_1+1)(C+1)}{cC+2}$ sea entero,

$H(R_1, C, cC)$ tendría dos modas, y denotando por $[x]$ la parte entera de x , tendríamos

$$\frac{(R_1+1)(C+1)}{cC+2} = \left[\frac{R_1}{c} \right] + 1, \quad \text{y} \quad \frac{(R_1+1)(C+1)}{cC+2} = \left[\frac{R_1}{c} \right]$$

con lo que podemos conseguir $\sum z_j = R_1$ haciendo d coordenadas iguales a $\left[\frac{R_1}{c} \right] + 1$

, $d = R_1 - c \left[\frac{R_1}{c} \right]$ y el resto de las coordenadas iguales a $\left[\frac{R_1}{c} \right]$ que también coincide con

el resultado dado por Mehta and Patel.

Y caso de que siendo $\frac{R_1}{c}$ no entero, $\frac{(R_1+1)(C+1)}{cC+2}$ también sea no entero, la

moda de $H(R_1, C, cC)$ vendría dada por $\left[\frac{(R_1+1)(C+1)}{cC+2} \right]$ que coincide con $\left[\frac{R_1}{c} \right]$ o

con $\left[\frac{R_1}{c} \right] + 1$. En la primera situación $R_1 - c \left[\frac{R_1}{c} \right] = d > 0$, con lo que d componentes del

vector formado con las modas de las marginales tendrían que incrementar su valor en

una unidad. Y en la segunda situación,

$\sum z_j = c \left[\frac{(R_1+1)(C+1)}{cC+2} \right] = c \left[\frac{R_1}{c} \right] + c = R_1 - d + c$, con lo que $c-d$ componentes del

vector formado con las modas de las marginales tendrían que disminuir su valor en

una unidad, o sea tendrían que ser iguales a $\left[\frac{R_1}{c} \right]$. En ambas situaciones también el

resultado coincide con el dado por Mehta and Patel.

Por último y en cuanto a comparación con los métodos anteriormente existentes para la determinación de máximos podemos constatar que cuando los dos marginales filas de la familia de tablas $2 \times c$, por cuya tabla de máxima probabilidad

estamos interesados, son iguales nuestro método siempre es directo y nos lleva a los mismos resultados que los obtenidos también a través del teorema 1.3 de M-P. Un sencillo razonamiento nos confirma este hecho.

$$\text{Sea } \Omega = \{X / X \text{ es } 2 \times c, \sum_{j=1}^c x_{ij} = A, A \text{ entero positivo dado, } \sum_{i=1}^2 x_{ij} = C_j \}$$

Tenemos $\sum_{j=1}^c C_j = 2.A$, o sea $\sum_{j=1}^c C_j$ es un número par, de donde se implica que para

los C_j solo existen dos posibilidades:

- Todos los C_j sean pares.
- Un número par de C_j son números impares y el resto son números pares.

Por otra parte, los valores z_{ij} que maximizan las distribuciones hipergeométricas marginales son

$$\frac{(A+1)(C_j+1)}{N+2} - 1 \leq z_{ij} \leq \frac{(A+1)(C_j+1)}{N+2}$$

o sea

$$\frac{(A+1)(C_j+1)}{2(A+1)} - 1 \leq z_{ij} \leq \frac{(A+1)(C_j+1)}{2(A+1)}$$

y simplificando

$$\frac{C_j+1}{2} - 1 \leq z_{ij} \leq \frac{C_j+1}{2}$$

ahora bien con $C_j+1 / 2$ pueden suceder dos casos, que sea entero o que no lo sea. En el primer caso que evidentemente equivale a C_j impar, para cada elemento z_{ij} tendremos dos valores posibles: $C_j+1 / 2$ y $(C_j+1/2)-1$. En el segundo caso que equivale a C_j sea par, para cada elemento z_{ij} tendremos un único valor posible:

$$E[C_j+1/2] = C_j/2.$$

Con lo que teniendo en cuenta las dos posibilidades planteadas al principio para los C_j , si se da la primera posibilidad, todos los C_j pares, las entradas de la tabla vienen

dadas por $C_j/2$ y su suma para cada una de las filas es A , y si se da la segunda, número par de C_j impares y el resto pares, como los C_j impares se dan por parejas bastaría con tomar para las entradas correspondientes a cada una de estas parejas respectivamente cada una de las dos soluciones apuntadas anteriormente, y también en este caso la suma de las entradas por filas será A . Y esta solución es evidentemente la dada también por M-P.

II.2.7: Ejemplos y tiempos de cómputo.

Algunos ejemplos, y una atención a los tiempos de cómputo nos pueden servir para evidenciar, con mayor claridad aún, las ventajas del nuevo método.

Utilizaremos en los 3 ejemplos propuestos para ilustrar el nuevo método la notación $a|t$ para indicar que el valor a aparece repetido t veces.

Ejemplo 1. Sea la familia de tablas 2×10 con marginales filas (15, 23) y marginales columnas (5|5, 2|3, 3, 4) o la MH(15; 5|5, 2|3, 3, 4). En primer lugar, de (2.3) calculamos las modas de las distribuciones marginales. La última tiene una moda doble (1 y 2), con lo que el conjunto M estará constituido por dos puntos z : (2|5, 1|4, 2) con $\delta=1$ y (2|5, 1|4, 1) con $\delta=0$. Entonces $\delta^*=0$ y este último punto (2|5, 1|4, 1) es la única moda de la MHD anterior. Igualmente será la primera fila de la única tabla 2×10 de máxima probabilidad (bajo la hipótesis de independencia entre filas y columnas) para los marginales dados. Para este ejemplo, aplicando el método de Harry Joe(1988) a la primera fila de la tabla 2×10 (la segunda fila se obtiene por diferencia con los marginales columnas, así ahorramos mucho tiempo de cómputo) se llega a un conjunto de 7776 tablas 2×10 , de donde se obtendría (mediante la inspección de las probabilidades de cada una de esas tablas) el mismo resultado anterior para la tabla de máxima probabilidad o para la moda de la MHD.

Previamente habría que eliminar de ese conjunto todas las tablas que no cumplan los marginales filas-columnas propuestos al principio.

Ejemplo 2. Sea la MH(63; 19|2, 16|9, 5|2, 9, 2) o la familia de tablas 2x15 con marginales filas (63, 140) y marginales columnas (19|2, 16|9, 5|2, 9, 2). Mediante (2.3) obtenemos $\mathbf{z}=(6|2, 5|9, 1|2, 3, 0)$ y será $\delta=-1$. Entonces, $\delta^*=\delta=-1$, $\dot{\mathbf{z}}=\mathbf{z}$ y $\lambda=1$. Para llegar a la moda $\hat{\mathbf{x}}$ solo hay que realizar $m=1$ paso. Para aplicar el teorema 2.2, como $\delta^*<0$, calculamos $(z_j+1)/(C_j+1)$ para las coordenadas de \mathbf{z} dando 0.35|2, 0.353|9, 0.333|2, 0.4, 0.333 y sumamos 1 en una de las coordenadas que alcanza el mínimo valor 0.333. Estas son las coordenadas 12, 13 y 15 de \mathbf{z} . Así, la MHD anterior tendrá solamente tres modas que son (6|2, 5|9, 1|2, 3, 1), (6|2, 5|9, 2, 1, 3, 0) y (6|2, 5|9, 1, 2, 3, 0). Y estas serán las primeras filas de las tres únicas tablas 2x15 de máxima probabilidad para los marginales dados. En el mismo sentido del ejemplo anterior, aplicando el método de Harry Joe a este ejemplo llegaríamos a un conjunto de mas de 2.8×10^6 tablas 2x15, entre las que habría que encontrar las tres modas anteriores.

Ejemplo 3. Para la MH(78; 9|2, 7|13, 6|3, 1|2, 4, 3) o para la familia de tablas 2x22 con marginales filas (78, 58) y marginales columnas (9|2, 7|13, 6|3, 1|2, 4, 3) obtenemos $\mathbf{z}=(5|2, 4|16, 1|2, 2|2)$ y será $\delta=2$, $\delta^*=2$, $\dot{\mathbf{z}}=\mathbf{z}$ y $\lambda=-1$. Para llegar a la moda $\hat{\mathbf{x}}$ solo hay que realizar $m=2$ pasos. Para aplicar el teorema 2.2, como $\delta^*>0$, calculamos $z_j/(C_j+1)$ para las coordenadas de \mathbf{z} , dando 0.50|15, 0.57|3, 0.5|2, 0.4, 0.5. El máximo valor es 0.57 que es alcanzado en las coordenadas 16, 17 y 18 de \mathbf{z} . Así, en el primer paso se restará 1 a una de estas coordenadas y en el segundo paso se restará 1 a una de las dos coordenadas restantes de este grupo. Obsérvese que en estos casos nunca se podría restar 1 en ambos pasos a la misma coordenada, y por tanto ambos pasos pueden realizarse simultáneamente tomando una pareja de

coordenadas de las tres anteriores y restando 1 a cada una de ellas. Así hemos obtenido las $\binom{3}{2} = 3$ únicas modas que tiene la MHD anterior: (5|2, 4|13, 3|2, 4, 1|2, 2|2), (5|2, 4|13, 3, 4, 3, 1|2, 2|2) y (5|2, 4|14, 3|2, 1|2, 2|2) Igualmente, estas serán las primeras filas de las tres únicas tablas 2x22 de máxima probabilidad para los marginales dados. En el mismo sentido del ejemplo 1, aplicando aquí Harry Joe llegaríamos a un conjunto de mas de 1.8×10^9 tablas 2x22, entre las que habría que encontrar las tres anteriores de máxima probabilidad.

Por último y respecto a tiempos de cómputo entre el nuevo método y el método propuesto por Harry Joe, ambos métodos están a unos niveles de tiempos de cómputo absolutamente diferentes. Aunque para pequeños valores de c los tiempos de cómputo de ambos métodos son despreciables para un ordenador, el método propuesto aquí es mucho mas rápido que el de H,J. En el ejemplo 1 (haciendo un bucle de repetición de cálculos para poder medir tiempos de cómputo) obtenemos una ratio “*método de Harry Joe/ método propuesto aquí*” del orden de 1.4×10^3 . Pero esta ratio crece exponencialmente cuando c aumenta. Ya en el ejemplo 2 la ratio es superior a 2×10^5 . Y para valores algo mayores de c el tiempo real de cómputo con el método de Joe es ya muy elevado (mas de 70 minutos para el ejemplo 3 en un PC Pentium 120 Mhz, y este tiempo se multiplica por 2, 3, 4 ó 5 en muchos otros ejemplos con el mismo $c=22$). Podemos comprobar que para valores de c superiores a 25 ó 28 empiezan a ser completamente irresolubles la mayoría de los ejemplos por el método de Joe. En cambio, el tiempo de cómputo del método propuesto aquí permanece practicamente constante al aumentar c , siendo absolutamente despreciable

para cualquier valor de c , incluso para valores de c extremadamente grandes, y para cualesquiera valores de $R_1, R_2, C_1, \dots, C_c$.

II.3: Relaciones entre las probabilidades extremas de tablas correspondientes a los nodos de la red de Mehta-Patel.

II.3.1: Introducción

Nuestro interés inicial por mejorar el proceso para el cálculo del máximo de una MHD nos ha obligado, como queda de manifiesto en el apartado anterior, a desarrollar una serie de propiedades referentes a la relación entre las modas de una familia determinada de MHD, propiedades que muestran su utilidad y eficacia para resolver el proceso mencionado; pero también estas propiedades nos dan acceso para establecer, como haremos, ciertas relaciones que pueden incidir de manera importante en la eficacia de otros procesos igualmente implicados en un algoritmo para la determinación del p-valor de una tabla $2 \times c$.

En el algoritmo de M-P la decisión sobre las tablas que se incorporan o no a la región crítica se realiza sobre nodos. Para los nodos del estado $k-1$, hijos del nodo madre $(k; R_{1k})$ del estado k , hay que decidir la verificación o no de las desigualdades (1.34) y (1.35), que para tablas $2 \times k$ toman la forma

$$\text{PAST. } l. \text{ LMA}(k-1; R_{1,k-1}) \leq D. P(X) \quad (2.35)$$

ó

$$\text{PAST. } l. \text{ LMI}(k-1; R_{1,k-1}) > D. P(X) \quad (2.36)$$

donde PAST es la longitud del camino recorrido hasta el nodo madre, l la longitud del arco que une el nodo hijo al nodo madre, LMA y LMI respectivamente la

longitud del subcamino mas largo y mas corto a partir del nodo hijo, $D = N! / R_1! R_2!$, $P(X)$ la probabilidad de la tabla observada dada por (2.2).

Además como para un nodo madre pueden existir varios PAST, $PAST_1, \dots, PAST_M$, y un nodo en general tiene varios nodos hijos, hay que considerar la discusión de (2.35) y (2.36) para todas las posibles combinaciones resultantes.

En el algoritmo de M-P por una parte no se establece ningun tipo de relación entre los valores de las combinaciones anteriores lo que exige que no se pueda excluir el análisis de ninguna de ellas, y por otra parte la determinación del camino mas largo y su longitud correspondiente a un nodo madre no se aprovecha para la determinación de los caminos mas largos y sus longitudes correspondientes a los nodos hijos.

Pero hay una serie de observaciones que la representación de red de la familia isomarginal de la tabla X nos permite apreciar, y que nos pueden evitar el proceso de determinación de máximo para nodos diferentes al nodo inicial, y también la discusión de (2.35) y (2.36) para muchos nodos de la red; los nodos hijos recogen y se distribuyen las trayectorias del nodo madre, hay un nodo hijo que recoge la trayectoria máxima del nodo madre y por consiguiente su trayectoria máxima coincide con la de él; la subtrayectoria máxima a partir de este nodo hijo quedará determinada por la subtrayectoria máxima existente a partir del nodo madre. Para el resto de los nodos hijos las trayectorias máximas serán aquellas trayectorias del nodo madre que tienen subtrayectorias máximas a partir del nodo hijo. Como todas tienen subtrayectoria común hasta el nodo madre, también se puede tratar de relacionarlas. Además en los valores de las combinaciones anteriormente mencionadas se pueden intuir dos tipos de relaciones de ordenación, se aprecia de forma casi inmediata una ordenación entre los valores de las combinaciones correspondientes a un

determinado nodo hijo con los diferentes valores PAST como consecuencia de una ordenación de estos, y de forma menos inmediata, una ordenación de los valores de las combinaciones resultantes de un PAST determinado con los diferentes nodos hijos como consecuencia de su mayor o menor proximidad al nodo que recoge la trayectoria máxima del nodo madre.

Todas las cuestiones que acabamos de exponer quedan formalizadas matemáticamente a continuación.

II.3.2: MHD asociadas a los nodos del retículo de M-P. Notación.

Todo nodo del retículo de M-P tiene asociada en el siguiente sentido una MHD. Si (k, R_{1k}) es un nodo del estado k , con un R_{1k} determinado, las subtrayectorias posibles desde este nodo al nodo terminal vienen dadas por el conjunto de puntos de la $MH(R_{1k}; C_1, \dots, C_k)$ y en consecuencia la subtrayectoria mas larga a partir de este nodo se corresponde con la moda de esta MHD.

Siendo

$$H \equiv MH(R_{1k}; C_1, \dots, C_k) \quad (2.37)$$

la MHD asociada a (k, R_{1k}) notaremos por F_k su conjunto referencial

$$F_k = \{(x_1, \dots, x_k) \in \mathbb{Z}_0^k / x_1 + \dots + x_k = R_{1k} \quad 0 \leq x_h \leq C_h \quad h=1, \dots, k\}$$

La generación de los nodos hijos viene dada, al realizar una particularización a tablas $2 \times k$ de (1.33), por:

$$\text{Max}(0, R_{1k} - C_k) \leq R_{1, k-1} \leq \text{Min}(R_{1k}, \sum_{l=1}^{k-1} C_l) \quad (2.38)$$

Con lo que los valores de estos nodos hijos forman una sucesión de enteros no negativos que podemos denotar por $\{R_{1, k-1}(i) \quad i=1, \dots, m\}$

donde

$$R_{1,k-1}(1) = \text{Max}(0, R_{1k} - C_k)$$

$$m = [\text{Min}(R_{1k}, \sum_{l=1}^{k-1} C_l) - \text{Max}(0, R_{1k} - C_k)]$$

$$R_{1,k-1}(m) = \text{Min}(R_{1k}, \sum_{l=1}^{k-1} C_l)$$

$$R_{1,k-1}(i+1) = R_{1,k-1}(i) + 1$$

y las MHD asociadas a estos nodos hijos constituyen una familia de MHD, denotémosla por \mathfrak{Z}_{k-1} , familia formada por todas las MHD

$$H_i \equiv \text{MH}(R_{1,k-1}(i); C_1, \dots, C_{k-1}) \quad i=1, \dots, m$$

que son las distribuciones condicionadas de las primeras $k-1$ variables de (2.37) condicionando en las posibles entradas de la coordenada k . Mas concretamente y según la propia construcción del retículo de M-P, si (2.37) es la distribución del vector (x_1, \dots, x_k) entonces H_i es la distribución condicionada de $(x_1, \dots, x_{k-1} / x_k = R_{1k} - R_{1,k-1}(i))$

Denotemos por $F_{k-1}(i)$ al conjunto referencial de H_i

$$F_{k-1}(i) = \{(x_1, \dots, x_{k-1}) \in Z_0^{k-1} / x_1 + \dots + x_{k-1} = R_{1,k-1}(i) \quad 0 \leq x_h \leq C_h \quad h=1, \dots, (k-1)\}$$

$$i=1, \dots, m$$

II.3.3: Relación entre las modas de las MHD asociadas a los nodos hijos de un nodo madre en la construcción reticulada de M-P para una tabla 2xc. Relación entre sus probabilidades.

La relación establecida en el punto II.2.4. entre las modas de las MHD de una familia particular puede ser utilizada aquí sobre la familia de distribuciones asociadas a los nodos hijos.

Para cualquier par de distribuciones H_i y H_{i+1} pertenecientes a \mathfrak{Z}_{k-1} :

$$H_i \equiv \text{MH}(R_{1,k-1}(i); C_1, \dots, C_{k-1})$$

$$H_{i+1} \equiv MH(R_{1,k-1}(i+1); C_1, \dots, C_{k-1})$$

siempre existe una conexión entre sus modas que queda explícita de la siguiente forma:

A) Si denotamos por $\dot{\mathbf{x}} = (\dot{x}_1, \dots, \dot{x}_{k-1})$ una moda de H_i siempre existe una moda de H_{i+1} conectada a ella a través de una coordenada v , donde v es un valor de j ($1 \leq j \leq k-1$) que minimiza la expresión $(\dot{x}_j+1)/(C_j+1)$, y por consiguiente una moda de H_{i+1} viene dada por un punto de coordenadas $(\dot{x}'_1, \dots, \dot{x}'_{k-1})$ donde

$$\begin{aligned} \dot{x}'_i &= \dot{x}_i \quad \forall i \neq v, \quad i = 1, \dots, (k-1) \\ \dot{x}'_v &= \dot{x}_v + 1 \end{aligned}$$

B) Recíprocamente si denotamos por $\dot{\mathbf{y}} = (\dot{y}_1, \dots, \dot{y}_{k-1})$ una moda de H_{i+1} siempre existe una moda de H_i conectada a ella a través de una coordenada v , donde v es un valor de j ($1 \leq j \leq k-1$) que maximiza la expresión $(\dot{y}_j)/(C_j+1)$ y por consiguiente una moda de H_i viene dada por un punto de coordenadas $(\dot{y}'_1, \dots, \dot{y}'_{k-1})$ donde

$$\begin{aligned} \dot{y}'_i &= \dot{y}_i \quad \forall i \neq v, \quad i = 1, \dots, (k-1) \\ \dot{y}'_v &= \dot{y}_v - 1 \end{aligned}$$

Las relaciones que acabamos de establecer nos permiten que la determinación de la moda para una única distribución de la familia \mathfrak{S}_{k-1} sirva como punto de partida para la construcción de una cadena con eslabones correspondientes a modas de las restantes distribuciones de dicha familia.

Así mismo las relaciones anteriores nos permiten establecer de forma casi inmediata el siguiente teorema que nos proporciona una relación recurrente entre las probabilidades máximas de las MHD asociadas a nodos hijos consecutivos de un nodo madre $(k; R_{1k})$:

Teorema 2.7. Sean $(k-1; R_{1,k-1}(i))$ y $(k-1; R_{1,k-1}(i+1))$ nodos hijos consecutivos del nodo $(k; R_{1k})$, H_i y H_{i+1} respectivamente sus MHD asociadas. Si $\dot{\mathbf{x}} = (\dot{x}_1, \dots, \dot{x}_{k-1})$ y $\dot{\mathbf{x}}' = (\dot{x}'_1, \dots, \dot{x}'_{k-1})$ son modas respectivamente de H_i y H_{i+1} conectadas a través de la coordenada v , y denotamos por $PMA(k-1; R_{1,k-1}(i))$ y $PMA(k-1; R_{1,k-1}(i+1))$ sus probabilidades entonces

$$PMA(k-1; R_{1,k-1}(i)) = PMA(k-1; R_{1,k-1}(i+1)) \cdot \frac{[N_{k-1} - (R_{1,k-1}(i+1) - 1)]}{R_{1,k-1}(i+1)} \cdot \frac{\dot{x}'_v}{C_v - (\dot{x}'_v - 1)} \quad (2.39)$$

siendo $N_{k-1} = \sum_{j=1}^{k-1} C_j$, y \dot{x}'_v la coordenada v -ésima de $\dot{\mathbf{x}}'$.

Y recíprocamente

$$PMA(k-1; R_{1,k-1}(i+1)) = PMA(k-1; R_{1,k-1}(i)) \cdot \frac{R_{1,k-1}(i) + 1}{N_{k-1} - R_{1,k-1}(i)} \cdot \frac{C_v - x_v}{x_v + 1} \quad (2.40)$$

Siendo x_v la coordenada v -ésima de $\dot{\mathbf{x}}$.

Demostración. Basta con considerar los valores de (2.2) para $\dot{\mathbf{x}}$ y $\dot{\mathbf{x}}'$.

Observación 2.1. Como una subtrayectoria de longitud máxima a partir de un nodo se corresponde con una moda de la MHD asociada a ese nodo, y la longitud de dicha subtrayectoria se corresponde con la probabilidad de la moda, podemos establecer las relaciones correspondientes a (2.39) y (2.40) para las longitudes máximas de los subcaminos a partir de nodos consecutivos, que quedarían en la forma:

$$LMA(k-1; R_{1,k-1}(i)) = LMA(k-1; R_{1,k-1}(i+1)) \cdot \frac{\dot{x}'_v}{C_v - (\dot{x}'_v - 1)} \quad (2.41)$$

$$LMA(k-1; R_{1,k-1}(i+1)) = LMA(k-1; R_{1,k-1}(i)) \cdot \frac{C_v - x_v}{x_v + 1} \quad (2.42)$$

Observación 2.2. Por ahorros de computación y aprovechando las posibilidades de mejoras apuntadas por Clarkson, Fan and Joe (capítulo I., punto I.3.2), ya que un

nodo hijo puede tener diferentes nodos madres, una vez que se dispone del conjunto de nodos madres (nodos del estado k) que necesitan extensión, y llamando $\min\{R_{1k}\}$ y $\max\{R_{1k}\}$ al mínimo y al máximo respectivamente de sus valores R_{1k} , determinaremos el conjunto de los nodos del estado siguiente, cuyos valores $R_{1,k-1}$ vendrán determinados, según 2.38 por

$$\text{Max} (0, \min\{R_{1k}\}-C_k) \leq R_{1,k-1} \leq \text{Min} (\max\{R_{1k}\}, \sum_{\ell=1}^{k-1} C_\ell) \quad (2.43)$$

Después, utilizando las relaciones de recurrencia (2.41) y (2.42) computaremos sus LMA y LMI, que podrán ser utilizados cada vez que se precise, y para estos LMA seguirán siendo ciertas evidentemente las relaciones de recurrencia que acabamos de considerar.

II.3.4: Relación entre la Moda de la MHD asociada a un nodo de un estado y la moda de la MHD asociada a un nodo hijo especial del estado siguiente. Nodo hijo principal.

Sea (k, R_{1k}) un nodo del estado k , con un R_{1k} determinado, y sea (2.37) la MHD asociada a este nodo; siguiendo con la misma notación introducida hasta el momento, las MHD de \mathfrak{F}_{k-1} son las distribuciones condicionadas correspondientes a las $k-1$ primeras coordenadas de (2.37) para cada entrada fija de la coordenada k -sima, y si $\dot{\mathbf{x}} = (\dot{x}_1, \dots, \dot{x}_k)$ es una moda de (2.37), entonces podemos concluir por el teorema 1.7 de Harry Joe referente a la condición necesaria de máximo, que $(\dot{x}_1, \dots, \dot{x}_{k-1})$ es la moda de un elemento H_i de la familia \mathfrak{F}_{k-1} , el correspondiente a condicionar en $x_k = \dot{x}_k$. Elemento que evidentemente existe y que, según lo visto anteriormente, es aquel cuyo $R_{1,k-1}(i)$ coincide con $\sum_{j=1}^{k-1} \dot{x}_j$.

Este nodo hijo (que no tiene que ser único) para el que el máximo de su MHD asociada viene determinado por el máximo de la MHD asociada al nodo madre le denominaremos nodo hijo principal.

Observación 2.3. Si denotamos por $(k-1; R_{1,k-1}(i^*))$ un nodo hijo principal es inmediata la siguiente relación:

$$\text{LMA}(k-1; R_{1,k-1}(i^*)) = \frac{\text{LMA}(k; R_{1k})}{\begin{pmatrix} C_k \\ \dot{X}_k \end{pmatrix}} \quad (2.44)$$

II.3.5: Relaciones de conexión en una MHD asociada a un nodo madre a través de la conexión de las modas de las MHD asociadas a sus nodos hijos . Cadena de puntos.

Notación.

Denotaremos por $\mathbf{t}_{k,k-1(i)}$ un punto cualquiera de F_k con entradas $R_{1k}-R_{1,k-1}(i)$ en la coordenada k . Y por $\mathbf{t}_{k,k-1(i)}^{(k)}$ el punto resultante de suprimir en $\mathbf{t}_{k,k-1(i)}$ su coordenada k . Lógicamente, $\mathbf{t}_{k,k-1(i)}^{(k)}$ será un punto de $F_{k-1}(i)$.

Para

$$\mathbf{t} = (x_1, \dots, x_k), \quad \mathbf{t} \in F_k$$

denotaremos por $\mathbf{t}^{(v+k)}$ el vector $(k-1)$ -dimensional cuyas $k-2$ primeras coordenadas son las del vector $\mathbf{t}^{(vk)}$ (resultado de suprimir las coordenadas v y k de \mathbf{t}) y cuya coordenada $k-1$ es x_v+x_k .

Igualmente representaremos por $\mathbf{C}^{(v+k)}$ el conjunto de $k-1$ parámetros obtenidos a partir de C_1, \dots, C_k , quitando C_v y C_k y añadiendo el parámetro $C_v + C_k$.

Además, será

$$R_{vk} = R_{1k} - \sum_{i \neq v, k} x_i = x_v + x_k \quad (2.45)$$

Observación 2.4. Siguiendo la notación anterior, es claro que tenemos

$$\mathbf{t}^{(v+k)} \rightarrow \text{MH}(R_{1k}, \mathbf{C}^{(v+k)}) \quad (2.46)$$

$$(x_v/\mathbf{t}^{(vk)}) \rightarrow \text{H}(R_{vk}, C_v, C_v+C_k) \quad (2.47)$$

$$P(\mathbf{t}) = P(\mathbf{t}^{(v+k)}) \cdot P(x_v/\mathbf{t}^{(vk)}) \quad (2.48)$$

Definición 2.4. Dos puntos $\mathbf{t}_{k,k-1(i)}$, $\mathbf{t}_{k,k-1(i+1)}$ de F_k decimos que están conectados respecto de la coordenada k , si $\mathbf{t}_{k,k-1(i)}^{(k)}$ y $\mathbf{t}_{k,k-1(i+1)}^{(k)}$ son modas respectivamente de H_i y H_{i+1} conectadas a través de una coordenada v .

Cuando podamos referirnos a ellos sin necesidad de mencionar las coordenadas que establecen la conexión, lo haremos simplemente como puntos conectados de F_k o de su hipergeométrica.

Teorema 2.8. (Propiedad de recurrencia)

Sean $\mathbf{t}_{k,k-1(i)}$ y $\mathbf{t}_{k,k-1(i+1)}$ dos puntos de F_k conectados respecto de la coordenada k a través de una coordenada v , sus probabilidades se relacionan en la forma recurrente indicada en los apartados **A)** y **B)**.

A)

$$P(\mathbf{t}_{k,k-1(i)}) = P(\mathbf{t}_{k,k-1(i+1)}) \cdot \frac{\dot{x}_v}{C_v - \dot{x}_v + 1} \cdot \frac{C_k - (R_{1k} - R_{1,k-1(i)}) + 1}{R_{1k} - R_{1,k-1(i)}} \quad (2.49)$$

Siendo $(\dot{x}_1, \dots, \dot{x}_{k-1})$ las $k-1$ primeras coordenadas de $\mathbf{t}_{k,k-1(i+1)}$ y v un valor de l que maximiza $\left\{ \frac{\dot{x}_l}{C_l + 1} \right\} \quad l=1, \dots, (k-1)$

B)

$$P(\mathbf{t}_{k,k-1(i+1)}) = P(\mathbf{t}_{k,k-1(i)}) \cdot \frac{C_v - \dot{x}_v}{\dot{x}_v + 1} \cdot \frac{R_{1k} - R_{1,k-1(i)}}{C_k - (R_{1k} - R_{1,k-1(i)}) + 1} \quad (2.50)$$

Siendo $(\dot{x}_1, \dots, \dot{x}_{k-1})$ las $k-1$ primeras coordenadas de $\mathbf{t}_{k,k-1(i)}$ y v un valor de l que minimiza $\left\{ \frac{\dot{x}_l + 1}{C_l + 1} \right\} \quad l=1, \dots, (k-1)$

Demostración.

A) $\mathbf{t}_{k,k-1(i)}$ y $\mathbf{t}_{k,k-1(i+1)}$ son dos puntos cuyas coordenadas son coincidentes excepto en la coordenada v y en la coordenada k ; conforme a la notación, a las hipótesis del apartado A), y a la conexión entre las modas de H_{i+1} y H_i , podemos escribir

$$\mathbf{t}_{k,k-1(i+1)} = (\dot{x}_1, \dots, \dot{x}_v, \dots, \dot{x}_{k-1}, R_{1k} - R_{1,k-1}(i+1))$$

$$\mathbf{t}_{k,k-1(i)} = (\dot{x}_1, \dots, (\dot{x}_v - 1), \dots, \dot{x}_{k-1}, R_{1k} - R_{1,k-1}(i))$$

y aplicando a estos puntos la expresión (2.48) de la observación 2.4, obtenemos

$$P(\mathbf{t}_{k,k-1(i+1)}) = P(\mathbf{t}_{k,k-1(i+1)}^{(v+k)}) \cdot P(\dot{x}_v / \mathbf{t}_{k,k-1(i+1)}^{(vk)}) \quad (2.51)$$

$$P(\mathbf{t}_{k,k-1(i)}) = P(\mathbf{t}_{k,k-1(i)}^{(v+k)}) \cdot P(\dot{x}_v - 1 / \mathbf{t}_{k,k-1(i)}^{(vk)}) \quad (2.52)$$

Como

$$P(\mathbf{t}_{k,k-1(i+1)}^{(v+k)}) = P(\mathbf{t}_{k,k-1(i)}^{(v+k)})$$

ya que

$$\mathbf{t}_{k,k-1(i+1)}^{(v+k)} = \mathbf{t}_{k,k-1(i)}^{(v+k)}$$

las posibles relaciones entre las probabilidades de estos dos puntos conectados queda reducida a las relaciones existentes entre

$$P(\dot{x}_v / \mathbf{t}_{k,k-1(i+1)}^{(vk)}) \quad \text{y} \quad P(\dot{x}_v - 1 / \mathbf{t}_{k,k-1(i)}^{(vk)})$$

pero

$$P(\dot{x}_v / \mathbf{t}_{k,k-1(i+1)}^{(vk)}) =$$

$$\frac{R_{kv}!(C_v + C_k - R_{vk})!C_v!C_k!}{(C_v + C_k)!\dot{x}_v!(C_v - \dot{x}_v)!(R_{1k} - R_{1,k-1}(i+1))!(C_k - (R_{1k} - R_{1,k-1}(i+1)))!}$$

y

$$P(\dot{x}_v - 1 / \mathbf{t}_{k,k-1(i)}^{(vk)}) = \frac{R_{kv}!(C_v + C_k - R_{vk})!C_v!C_k!}{(C_v + C_k)!(\dot{x}_v - 1)!(C_v - (\dot{x}_v - 1))!(R_{1k} - R_{1,k-1(i)})!(C_k - (R_{1k} - R_{1,k-1(i)}))!}$$

con lo que

$$P(\dot{x}_v - 1 / \mathbf{t}_{k,k-1(i)}^{(vk)}) = P(\dot{x}_v / \mathbf{t}_{k,k-1(i+1)}^{(vk)}) \cdot \frac{\dot{x}_v}{C_v - \dot{x}_v + 1} \frac{C_k - (R_{1k} - R_{1,k-1(i)}) + 1}{R_{1k} - R_{1,k-1(i)}}$$

Y nos proporciona la relación (2.49) c.q.d.

Demostración B)

$\mathbf{t}_{k,k-1(i)}$ y $\mathbf{t}_{k,k-1(i+1)}$ son dos puntos cuyas coordenadas son coincidentes excepto en la coordenada v y en la coordenada k ; conforme a las hipótesis establecidas para este apartado y a la conexión de las modas de H_{i+1} y H_i , podemos escribir

$$\mathbf{t}_{k,k-1(i)} = (\dot{x}_1, \dots, \dot{x}_v, \dots, \dot{x}_{k-1}, R_{1k} - R_{1,k-1(i)})$$

$$\mathbf{t}_{k,k-1(i+1)} = (\dot{x}_1, \dots, (\dot{x}_v + 1), \dots, \dot{x}_{k-1}, R_{1k} - R_{1,k-1(i+1)})$$

y aplicando la expresión (2.48) de la observación 2.4 a estos puntos de manera análoga a como hemos hecho en el apartado anterior obtenemos la expresión (2.50) buscada.

Definición 2.5. Dados dos valores enteros w y w'

$$1 \leq w < w' \leq m \quad m = [\text{Min}(R_{1k}, \sum_{l=1}^{k-1} C_l) - \text{Max}(0, R_{1k} - C_k)]$$

y la sucesión finita de puntos

$$C_k(w, w') = \{\mathbf{t}_{k,k-1(s)} \in F_k \quad s = w, w+1, \dots, w' / \mathbf{t}_{k,k-1(s)}^{(k)} \text{ es una moda de } H_s\}$$

Decimos que $C_k(w, w')$ es una cadena respecto a la coordenada k cuando para cada $s (s = w, w+1, \dots, w'-1)$, $\mathbf{t}_{k,k-1(s)}$ y $\mathbf{t}_{k,k-1(s+1)}$ están conectados respecto a la coordenada k .

También, dados dos puntos cualesquiera de la cadena $C_k(w, w')$ siempre diremos que ambos puntos están comunicados a través de dicha cadena.

Dados dos puntos arbitrarios de la cadena $\mathbf{t}_{k,k-1(a)}$ y $\mathbf{t}_{k,k-1(b)}$ ($a < b$) podemos pasar de $\mathbf{t}_{k,k-1(a)}$ a $\mathbf{t}_{k,k-1(b)}$ y viceversa a través de los sucesivos eslabones intermedios de la cadena.

II.3.6: Subcadenas de puntos de la MHD asociada a un nodo madre con eslabón conocido.

Conforme a lo establecido en el punto II.3.4. la moda de la MHD asociada a un nodo madre, determina la moda del nodo hijo principal y, por la conexión establecida en el punto II.3.3. para las modas de los nodos hijos y la cadena que implican en la MHD, señalada en el punto II.3.5., si denotamos por $\mathbf{t}_{k,k-1(i^*)}$ el máximo de la hipergeométrica asociada al nodo madre, y por $\mathbf{t}_{k,k-1(i^*)}^{(k)}$ el máximo de la hipergeométrica asociada al nodo hijo principal, podemos definir las siguientes subcadenas $C_k^-(1, i^*), C_k^+(i^*, m)$ con eslabón conocido, $\mathbf{t}_{k,k-1(i^*)}$,

$$C_k^-(1, i^*) = \{ \mathbf{t}_{k,k-1(l)} \in F_k \quad l=1, \dots, i^* / \mathbf{t}_{k,k-1(l)}^{(k)} \text{ es una moda de } H_l \} \quad (2.53)$$

$$C_k^+(i^*, m) = \{ \mathbf{t}_{k,k-1(l)} \in F_k, \quad l = i^*, \dots, m / \mathbf{t}_{k,k-1(l)}^{(k)} \text{ es una moda de } H_l \} \quad (2.54)$$

Propiedad 2.1. Las probabilidades de los puntos de la cadena $C_k^-(1, i^*)$ se determinan de forma recurrente utilizando la expresión (2.49), tomando como inicio la probabilidad del punto $\mathbf{t}_{k,k-1(i^*)}$ o lo que es igual, la probabilidad del máximo de la Hipergeométrica multivariante asociada al nodo madre.

Propiedad 2.2. Las probabilidades de los puntos de la cadena $C_k^+(i^*, m)$ se determinan de forma recurrente utilizando la expresión (2.50), tomando como inicio la probabilidad del punto $\mathbf{t}_{k,k-1(i^*)}$ o lo que es igual la probabilidad del máximo de la Hipergeométrica multivariante asociada al nodo madre.

Definición 2.6. Llamamos cadena de puntos de la $MH(R_k; C_1, \dots, C_c)$ respecto a la coordenada k a los puntos de esta MHD con entradas comunes en las coordenadas $c, c-1, \dots, (k+1)$ y cuyas entradas restantes que se corresponden con los puntos de la $MH(R_k; C_1, \dots, C_k)$ forman una cadena respecto a la coordenada k .

Evidentemente la relación entre los puntos de esta cadena o sus probabilidades vienen determinadas por las establecidas para la cadena de puntos de la $MH(R_k; C_1, \dots, C_k)$.

II.3.7: Ordenación entre las probabilidades de los puntos de una cadena.

Ordenación entre las longitudes extremas de los caminos de los nodos hijos.

Teorema 2.9.(Propiedad de ordenación vertical).

Sea $\mathbf{t}_{k,k-1}(i^*) = (x_1^*, \dots, x_k^*)$ un máximo de la MHD asociada al nodo (k, R_{1k})

A) Si $\mathbf{t}_{k,k-1}(s)$ y $\mathbf{t}_{k,k-1}(s+1)$ son puntos (evidentemente consecutivos) de la cadena

$C_k^-(1, i^*)$, entonces se verifica

$$P(\mathbf{t}_{k,k-1}(s)) \leq P(\mathbf{t}_{k,k-1}(s+1)) \quad (2.55)$$

B) Si $\mathbf{t}_{k,k-1}(l)$ y $\mathbf{t}_{k,k-1}(l+1)$ son puntos (evidentemente consecutivos) de la cadena

$C_k^+(i^*, m)$ entonces se verifica

$$P(\mathbf{t}_{k,k-1}(l+1)) \leq P(\mathbf{t}_{k,k-1}(l)) \quad (2.56)$$

Demostración.- A) Supongamos que los puntos $\mathbf{t}_{k,k-1}(s)$ y $\mathbf{t}_{k,k-1}(s+1)$ de la cadena $C_k^-(1, i^*)$, están conectados respecto de la coordenada k a través de una coordenada v , sus coordenadas pueden expresarse en la forma:

$$\mathbf{t}_{k,k-1}(s+1) = (x_1, \dots, x_v, \dots, x_k)$$

$$\mathbf{t}_{k,k-1}(s) = (x_1, \dots, x_{v-1}, \dots, x_{k+1})$$

al aplicar, la expresión (2.48) de la observación 2.4, a estos puntos obtenemos

$$P(\mathbf{t}_{k,k-1(s+1)}) = P(\mathbf{t}_{k,k-1(s+1)}^{(v+k)}) \cdot P(x_k / \mathbf{t}_{k,k-1(s+1)}^{(vk)})$$

$$P(\mathbf{t}_{k,k-1(s)}) = P(\mathbf{t}_{k,k-1(s)}^{(v+k)}) \cdot P(x_{k+1} / \mathbf{t}_{k,k-1(s)}^{(vk)})$$

como

$$P(\mathbf{t}_{k,k-1(s+1)}^{(v+k)}) = P(\mathbf{t}_{k,k-1(s)}^{(v+k)})$$

ya que

$$\mathbf{t}_{k,k-1(s+1)}^{(v+k)} = \mathbf{t}_{k,k-1(s)}^{(v+k)}$$

el comparar

$$P(\mathbf{t}_{k,k-1(s+1)}) \text{ con } P(\mathbf{t}_{k,k-1(s)})$$

se reduce a comparar

$$P(x_k / \mathbf{t}_{k,k-1(s+1)}^{(vk)}) \text{ con } P(x_{k+1} / \mathbf{t}_{k,k-1(s)}^{(vk)})$$

o sea se reduce a comparar las probabilidades de dos puntos de la $H(R_{vk}; C_v, C_v+C_k)$, el punto (x_v, x_k) con el punto (x_{v-1}, x_{k+1}) .

Pero de la hipótesis se obtiene que (x_v^*, x_k^*) es un máximo de la Hipergeométrica $H(R_{vk}^*; C_v, C_v+C_k)$ y puede construirse una cadena que comunique el máximo de la $H(R_{vk}^*; C_v, C_v+C_k)$ con el máximo de la $H(R_{vk}; C_v, C_v+C_k)$ y este máximo, tal y como se ha establecido que se construyen las cadenas, será un punto del conjunto referencial de esta hipergeométrica cuyo valor de la coordenada k sea el primero accesible desde x_k^* o sea su diferencia será la menor posible con x_k^* , pero como por la propia construcción de $C_k^-(1, i^*)$, $x_k^* \leq x_k < x_{k+1}$, si denotamos por (x_v^M, x_k^M) el máximo de $H(R_{vk}; C_v, C_v+C_k)$ también se verificará $x_k^M \leq x_k < x_{k+1}$ con lo que aplicando la propiedad de ordenación de las hipergeométricas simples se obtiene

$$P(x_{k+1} / \mathbf{t}_{k,k-1(s)}^{(vk)}) \leq P(x_k / \mathbf{t}_{k,k-1(s+1)}^{(vk)})$$

y en definitiva

$$P(\mathbf{t}_{k,k-1(s)}) \leq P(\mathbf{t}_{k,k-1(s+1)}) \quad \text{c.q.d.}$$

B) La demostración es análoga a la del apartado A.

Supongamos que los puntos $\mathbf{t}_{k,k-1(l)}$ y $\mathbf{t}_{k,k-1(l+1)}$, de la cadena $C_k^+(i^*, m)$, están conectados respecto de la coordenada k a través de una coordenada v , sus coordenadas pueden expresarse en la forma:

$$\mathbf{t}_{k,k-1(l)} = (x_1, \dots, x_v, \dots, x_k)$$

$$\mathbf{t}_{k,k-1(l+1)} = (x_1, \dots, x_{v+1}, \dots, x_{k-1})$$

al aplicar la expresión (2.48) de la observación 2.4 a estos puntos

$$P(\mathbf{t}_{k,k-1(l)}) = P(\mathbf{t}_{k,k-1(l)}^{(v+k)}) \cdot P(x_k / \mathbf{t}_{k,k-1(l)}^{(vk)})$$

$$P(\mathbf{t}_{k,k-1(l+1)}) = P(\mathbf{t}_{k,k-1(l+1)}^{(v+k)}) \cdot P(x_{k-1} / \mathbf{t}_{k,k-1(l+1)}^{(vk)})$$

como

$$P(\mathbf{t}_{k,k-1(l)}^{(v+k)}) = P(\mathbf{t}_{k,k-1(l+1)}^{(v+k)})$$

ya que

$$\mathbf{t}_{k,k-1(l)}^{(v+k)} = \mathbf{t}_{k,k-1(l+1)}^{(v+k)}$$

el comparar

$$P(\mathbf{t}_{k,k-1(l)}) \text{ con } P(\mathbf{t}_{k,k-1(l+1)})$$

Se reduce a comparar

$$P(x_k / \mathbf{t}_{k,k-1(l)}^{(vk)}) \text{ con } P(x_{k-1} / \mathbf{t}_{k,k-1(l+1)}^{(vk)})$$

o sea se reduce a comparar las probabilidades de dos puntos de la $H(R_{vk}; C_v, C_{v+1}, C_k)$

el punto (x_v, x_k) con el punto (x_{v+1}, x_{k-1}) .

Pero de la hipótesis se obtiene que (x_v^*, x_k^*) es un máximo de la hipergeométrica $H(R_{vk}^*; C_v, C_{v+1}, C_k)$ y puede construirse una cadena que comunique

el máximo de la $H(R_{vk}^* ; C_v, C_v + C_k)$ con el máximo de la hipergeométrica $H(R_{vk}; C_v, C_v + C_k)$ y, este máximo, tal y como se ha establecido que se construyen las cadenas, será un punto del conjunto referencial de esta hipergeométrica cuyo valor de la coordenada k sea el primero accesible desde x_k^* o sea su diferencia será la menor posible con x_k^* , pero como por la propia construcción de $C_k^+(i^*, m)$, $x_k^* \geq x_k > x_{k-1}$, si denotamos por (x_v^M, x_k^M) el máximo de $H(R_{vk}; C_v, C_v + C_k)$ también se verificará que $x_k^M \geq x_k > x_{k-1}$, con lo que aplicando la propiedad de ordenación de las hipergeométricas simples se obtiene

$$P(x_{k-1} / \mathbf{t}_{k,k-1(l+1)}^{(vk)}) \leq P(x_k / \mathbf{t}_{k,k-1(l)}^{(vk)})$$

y en definitiva

$$P(\mathbf{t}_{k,k-1(l+1)}) \leq P(\mathbf{t}_{k,k-1(l)}) \quad \text{c.q.d.}$$

Observación 2.5.

Las probabilidades de los puntos de la cadena $C_k^-(1, i^*)$ son las que deciden las probabilidades máximas de las tablas que comparten subcamino hasta el nodo $(k; R_{1k})$ y cuya trayectoria continua respectivamente por los nodos hijos $\{(k-1; R_{1,k-1}(s)), s=i^*, \dots, 1\}$ y las que en definitiva nos permiten decidir si las probabilidades de todas estas tablas pueden incorporarse al p-valor de una tabla dada. Por consiguiente si para un punto $\mathbf{t}_{k,k-1(s)}$ de esta cadena puede decidirse que las probabilidades de todas las tablas que le corresponden en el sentido anteriormente fijado se integran en el p-valor, esta misma decisión puede adoptarse para todas las colecciones de tablas correspondientes a puntos posteriores de dicha cadena.

Las probabilidades de los puntos de la cadena $C_k^+(i^*, m)$ son las que deciden las probabilidades máximas de las tablas que comparten subcamino hasta el nodo

$(k; R_{1k})$ y cuya trayectoria continua respectivamente por los nodos hijos $\{(k-1; R_{1,k-1}(l)), l=i^*, \dots, m\}$ y las que en definitiva nos permiten decidir si las probabilidades de todas estas tablas pueden incorporarse al p-valor de una tabla dada. Por consiguiente si para un punto $t_{k,k-1}(l)$ de esta cadena puede decidirse que las probabilidades de todas las tablas que le corresponden en el sentido anteriormente fijado se integran en el p-valor, esta misma decisión puede adoptarse para todas las colecciones de tablas correspondientes a puntos posteriores de dicha cadena.

Observación 2.6. Teniendo en cuenta como se realiza el procesamiento de los nodos en el algoritmo de red de M-P y con vistas a la implementación de la última propiedad desarrollada podemos extraer en lenguaje de redes lo siguiente:

Sea $(k; R_{1k})$ un nodo madre, si $(k-1; R_{1,k-1}(i^*))$ es un nodo hijo principal y consideramos las subsucesiones ascendentes y descendentes:

$$SS_1 = \{(k-1; R_{1,k-1}(i^*)), \dots, (k-1; R_{1,k-1}(m))\}$$

$$SS_0 = \{(k-1; R_{1,k-1}(i^*)), \dots, (k-1; R_{1,k-1}(1))\}$$

Expresando el teorema 2.9 en terminos de longitudes de caminos, podemos afirmar que para todo par de nodos hijos consecutivos de la subsucesión SS_1

$$l_{k,k-1(s+1)} \cdot LMA(k-1; R_{1,k-1}(s+1)) \leq l_{k,k-1(s)} \cdot LMA(k-1; R_{1,k-1}(s)) \quad s=i^*, \dots, (m-1) \quad (2.57)$$

y, para todo par de nodos hijos consecutivos de la subsucesión SS_0

$$l_{k,k-1(s)} \cdot LMA(k-1; R_{1,k-1}(s)) \leq l_{k,k-1(s+1)} \cdot LMA(k-1; R_{1,k-1}(s+1)) \quad s=(i^*-1), \dots, 1 \quad (2.58)$$

donde $l_{k,k-1(s)}$ y $l_{k,k-1(s+1)}$ representan, respectivamente, las longitudes de los arcos desde el nodo madre $(k; R_{1k})$ a los nodos hijos $(k-1; R_{1,k-1}(s))$ y $(k-1; R_{1,k-1}(s+1))$.

Por último y con el objetivo de introducir en el algoritmo de M-P simplificaciones adicionales a las que se derivan de la implementación de todas las propiedades desarrolladas con anterioridad podemos establecer la siguiente observación que es inmediata.

Observación 2.7. Sea $(k; R_{1k})$ un nodo madre, y $PAST_1, \dots, PAST_M$ las longitudes de los subcaminos recorridos hasta ese nodo, ordenados de menor a mayor. Si $(k-1; R_{1,k-1}(h))$, $1 \leq h \leq m$, es un nodo hijo de dicho nodo, entonces

$$PAST_i \cdot l \cdot LMA(k-1; R_{1,k-1}(h)) \leq PAST_{i'} \cdot l \cdot LMA(k-1; R_{1,k-1}(h)) \Leftrightarrow i \leq i' \quad (2.59)$$

y

$$PAST_i \cdot l \cdot LMI(k-1; R_{1,k-1}(h)) \leq PAST_{i'} \cdot l \cdot LMI(k-1; R_{1,k-1}(h)) \Leftrightarrow i \leq i' \quad (2.60)$$

siendo i e i' elementos del conjunto T , $T = \{1, \dots, M\}$ y $l = l_{k,k-1}(h)$.

II.4: Algoritmo Modificado de M – P para determinar el p-valor de una tabla $2 \times c$.

II.4.1.- Introducción.

Sea una tabla X , $2 \times c$, con marginales filas R_1, R_2 , marginales columnas C_1, \dots, C_c , total global N y entradas (x_1, \dots, x_c) en la primera fila. El Algoritmo que hemos desarrollado para determinar el p-valor de X es una modificación del algoritmo reticulado de M-P, considerado ya este algoritmo con las modificaciones introducidas por Harry Joe (1988) para el cálculo de las probabilidades extremas, que sigue utilizando la estructura reticulada para representar la familia isomarginal de la tabla dada, y al incorporar los resultados teóricos anteriormente obtenidos queda concretado en el proceso que detallamos a continuación; proceso en el que denotaremos por:

- $L(X)$ la longitud del camino correspondiente a la tabla X .
- $LMA(k; R_{1k})$ la longitud del subcamino más largo a partir del nodo $(k; R_{1k})$.
- $LMI(k; R_{1k})$ la longitud del subcamino más corto a partir del nodo $(k; R_{1k})$.

- PMAX el PAST máximo que hay que considerar al verificar el cumplimiento de la condición sobre el camino mas largo correspondiente a un nodo hijo que se esté analizando (paso 8.II del algoritmo).
- PMIN el PAST mínimo que hay que considerar al verificar el cumplimiento de la condición sobre el camino mas corto correspondiente a un nodo hijo que se esté analizando (paso 8.III del algoritmo)

II.4.2. Algoritmo Modificado.

- 1- Determinar $L(X)$

$$L(X) = \binom{C_1}{x_1} \binom{C_2}{x_2} \dots \binom{C_c}{x_c}$$

- 2- Determinar D

$$D = \frac{N!}{R_1!.R_2!}$$

- 3- Establecer el nodo inicial: $(c; R_{1c})$, $R_{1c} \equiv R_1$. Este nodo tendrá $PAST=1$
- 4- Determinar para la MHD asociada al nodo inicial su máximo por el nuevo método que hemos desarrollado para calcular el máximo de una MHD, y utilizarlo para computar $LMA(c; R_{1c})$.
- 5- Contrastar si $LMA(c; R_{1c}) \leq L(X)$, en caso afirmativo el proceso habría terminado y p-valor de X sería igual a uno pues todas las tablas de la familia se encontrarían dentro de la región crítica. En otro caso (lo más usual) se hace

preciso continuar el proceso para seleccionar las tablas que contribuyen al p-valor de X; y para todo estado k, k=c,.....1, continuamos de la siguiente manera:

6- Determinar los nodos del estado siguiente k-1, como se vió en (2.43)

$$\text{Max} (0, \min\{R_{1k}\}-C_k) \leq R_{1,k-1} \leq \text{Min} (\max\{R_{1k}\}, \sum_{\ell=1}^{k-1} C_{\ell})$$

7- Determinar para todos los nodos del estado k-1, LMA(k-1; R_{1,k-1}) y LMI(k-1; R_{1,k-1}).

Para la determinación de las LMA utilizamos el siguiente procedimiento :

i) Tomamos el nodo madre de menor rango, $R_{1k}^* = \min\{R_{1k}\}$, y utilizamos la moda de su MHD asociada

$$\text{Moda MH}(R_{1k}^* ; C_1, \dots, C_k) \equiv (x_1^*, \dots, x_k^*)$$

y su LMA conocida para determinar, la moda y LMA de un nodo hijo principal(k-1; R_{1,k-1}(i*))

$$\text{Moda MH}(R_{1,k-1}; C_1, \dots, C_{k-1}) \equiv (x_1^*, \dots, x_{k-1}^*)$$

$$\text{LMA}(k-1; R_{1,k-1}(i^*)) = \frac{\text{LMA}(k; R_{1k}^*)}{\begin{pmatrix} C_k \\ x_k^* \end{pmatrix}}$$

ii) Las LMA para el resto de los nodos del estado k-1 se determinan recursivamente considerando las dos subsucesiones que se establecen a partir del nodo hijo principal anterior; Para la subsucesión descendente si (x₁,.....x_{k-1}) es el máximo del nodo hijo (k-1; R_{1,k-1}(s)) con LMA(k-1; R_{1,k-1}(s)) ya calculado, se calcula, de acuerdo a (2.41)

$$\text{LMA}(k-1; R_{1,k-1}(s-1)) = \text{LMA}(k-1; R_{1,k-1}(s)) \cdot \frac{x_v}{C_v - (x_v - 1)}$$

siendo v el valor de j que maximiza $\{x_j/C_{j+1}\}_{j=1,\dots,(k-1)}$.

Igualmente para la subsucesión ascendente si (x_1, \dots, x_{k-1}) es el máximo del nodo hijo $(k-1; R_{1,k-1}(s))$ con $LMA(k-1; R_{1,k-1}(s))$ ya calculado, se calcula de acuerdo a (2.42)

$$LMA(k-1; R_{1,k-1}(s+1)) = LMA(k-1; R_{1,k-1}(s)) \cdot \frac{C_v - x_v}{x_v + 1}$$

siendo v el valor de j que minimiza el $\{x_{j+1}/C_{j+1}\}_{j=1,\dots,(k-1)}$.

Este proceso de la subsucesión ascendente lo podemos prolongar hasta

el nodo hijo con $R_{1,k-1} = \text{Min}(\max\{R_{1k}\}, \sum_{\ell=1}^{k-1} C_{\ell})$.

Para la determinación de las LMI correspondientes utilizamos el Teorema 1.8 de Joe (1988).

- 8-** Recuperar uno a uno los nodos del estado k sobre los que es necesario decidir sobre la longitud de sus caminos. Para cada nodo $(k; R_{1k})$ del estado k , acompañado de sus valores PAST ordenados de menor a mayor, $PAST_1, \dots, PAST_M$, y con las frecuencias correspondientes, f_1, \dots, f_M , a estos PAST, procedemos del siguiente modo:

I) Generamos sus nodos hijos $(k-1; R_{1,k-1})$, denotamos por $R_{1,k-1}(1), \dots, R_{1,k-1}(m)$ ordenados de menor a mayor los valores de los $R_{1,k-1}$, computamos para cada uno de ellos la longitud del arco que lo une con su nodo madre, $l_{k,k-1}(s) = \left(\begin{array}{c} C_k \\ R_{1k} - R_{1,k-1}(s) \end{array} \right)$, que notaremos simplemente por l cuando no haya lugar a confusión.

II) Analizamos la longitud del camino mas largo para cada uno de estos nodos hijos combinado con los distintos valores PAST del nodo madre, decidiendo si

$$\text{LMAG}(k-1; R_{1,k-1}) = \text{PAST} \cdot l \cdot \text{LMA}(k-1; R_{1,k-1}) \leq L(X) \quad (2.61)$$

es cierta o no.

Para ello:

- a) Determinamos un nodo hijo principal: $(k-1; R_{1,k-1}(i^*))$
- b) Establecemos dos subsucesiones, una ascendente, otra descendente a partir del nodo hijo principal,

$$SS_1 = \{(k-1; R_{1,k-1}(i^*)), \dots, (k-1; R_{1,k-1}(m))\}$$

$$SS_0 = \{(k-1; R_{1,k-1}(i^*)), \dots, (k-1; R_{1,k-1}(1))\}$$

- c) Para cada una de las dos subsucesiones de nodos hijos (excluido el nodo hijo principal), y comenzando con el nodo hijo más extremo, o sea para cada uno de los elementos de

$$SS'_1 = \{(k-1; R_{1,k-1}(m)), \dots, (k-1; R_{1,k-1}(i^*-1))\}$$

$$SS'_0 = \{(k-1; R_{1,k-1}(1)), \dots, (k-1; R_{1,k-1}(i^*-1))\}$$

hacemos lo siguiente:

- c1) Seleccionar el menor PAST del nodo madre, PAST_1 .
- c2) Determinar la longitud del camino más largo correspondiente a ese nodo hijo con el PAST seleccionado

$$\text{LMAG}(k-1; R_{1,k-1}) = \text{PAST} \cdot l \cdot \text{LMA}(k-1; R_{1,k-1})$$

- c3) Decidir si (2.61) es cierta o no.

A) Si (2.61) es falsa para el PAST más pequeño, PAST_1 , del nodo madre ya no hay que analizar esa condición para ese nodo hijo con los restantes

valores PAST del nodo madre, pues también sería falsa (por la observación 2.7), ni tampoco hay que analizar esa condición para los restantes nodos hijos de la subsucesión pues igualmente también sería falsa (por el teorema 2.9 ó las expresiones 2.57 y 2.58) con lo que esos nodos con todos los $PAST_1, \dots, PAST_M$ quedan pendientes para analizar la condición de mínimo. Para todos esos nodos hijos sería $P_{MIN} = PAST_1$.

B) Si (2.61) es cierta para el PAST más pequeño, $PAST_1$, del nodo madre, todas las tablas correspondientes a ese nodo hijo con ese PAST quedarían dentro de la región crítica y como eso mismo puede ocurrir con otros $PAST_i$ superiores al $PAST_1$, pasaríamos a seleccionar el mayor PAST del nodo madre, P_{MAX} (para el nodo hijo más extremo será $P_{MAX} = PAST_M$ y para los otros nodos hijos se irá obteniendo recursivamente), repitiendo para el nodo hijo con este nuevo PAST lo indicado en c2) y c3).

Si (2.61) es cierta, todas las tablas correspondientes a ese nodo hijo con ese PAST máximo quedan dentro de la región crítica, y (por la observación 2.7) también quedarían dentro de la región crítica las tablas correspondientes a ese nodo hijo para los

restantes valores PAST del nodo madre sin necesidad de comprobarlo individualmente para ellos, y se determinarían las sumas de las longitudes de sus caminos y pasaríamos a reiniciar el proceso desde c1) con el siguiente nodo hijo de la subsucesión.

Si (2.61) no es cierta para el valor PAST máximo del nodo madre, iríamos considerando a partir de ese valor PAST y en orden decreciente los restantes valores de los PAST del nodo madre ($k; R_{1k}$) hasta llegar a uno, $PAST_F$, para el que la condición (2.61) sea cierta, con lo que (por la observación 2.7) ya sería cierta para los restantes valores PAST sin necesidad de comprobaciones individuales, se determinarían las sumas de las longitudes de sus caminos y pasaríamos a reiniciar el proceso desde c1) con el siguiente nodo hijo de la subsucesión. El nodo hijo se queda pendiente de analizar la condición de mínimo con todos los PAST para los que no es cierta la condición (2.61), $PAST_{F+1}, \dots, PAST_M$. Para todos los PAST que no satisfacen la condición (2.61) sucedería lo mismo (teorema 2.9) para los restantes nodos hijos de la subsucesión en los que quedan por analizar la condición de máximo, (denotaremos

respectivamente por $SS'_1(R)$ y $SS'_0(R)$ el conjunto de nodos de la subsucesión SS'_1 y SS'_0 que quedan pendientes del análisis de la condición de mínimo). Conservaremos la información de que al analizar la condición de mínimo para este nodo hijo el mínimo PAST a considerar será $PMIN = PAST_{F+1}$, y al seguir analizando la condición de máximo para los restantes nodos hijos de la subsucesión el máximo PAST a considerar será $PMAX = PAST_F$.

Respecto a la incorporación al p-valor, como la suma de las longitudes de los subcaminos a partir del nodo hijo viene dado por $\binom{C_1 + \dots + C_{k-1}}{R_{1,k-1}}$, la suma de todos los caminos para un $PAST_i$ determinado será $\binom{C_1 + \dots + C_{k-1}}{R_{1,k-1}} \cdot \ell \cdot PAST_i$; Así, si al p-valor hay que incorporarle los caminos correspondientes a los $PAST_1, \dots, PAST_F$, $1 \leq F \leq M$, la suma de todos los caminos para estos PAST, con sus frecuencias respectivas será

$$SL = \binom{C_1 + \dots + C_{k-1}}{R_{1,k-1}} \cdot \ell \cdot \sum_{i=1}^F PAST_i \cdot f_i$$

Y al p-valor habrá que sumarle SL/D .

III) Análizamos la longitud del camino mínimo para los nodos hijos del nodo $(k; R_{1k})$ que han quedado pendientes en II), tanto en la subsucesión ascendente como en la descendente ($SS'_1(R)$ y $SS'_0(R)$), comenzando por los nodos hijos mas extremos e incorporando ahora el nodo hijo principal en una de estas subsucesiones (en la ascendente, por ejemplo). Vamos a intentar

decidir para cada uno de esos nodos hijos combinado con los diferentes PAST del nodo madre si

$$LMIG(k-1; R_{1,k-1}) = PAST \cdot l \cdot LMI(k-1; R_{1,k-1}) > L(X) \quad (2.62)$$

es cierta o no.

Para cada subsucesión y comenzando por el nodo hijo mas extremo pendiente de analizar, procedemos del siguiente modo:

- d1) Seleccionamos el mayor PAST del nodo madre, $PAST_M$.
- d2) Determinamos la longitud del camino más corto correspondiente a ese nodo hijo con el PAST seleccionado

$$LMIG(k-1; R_{1,k-1}) = PAST \cdot l \cdot LMI(k-1; R_{1,k-1})$$

- d3) Decidimos si (2. 62) es cierta o no.

C) Si (2.62) es cierta para ese valor PAST máximo, todas las tablas correspondientes a ese nodo hijo con ese PAST resultante quedarían fuera de la región crítica, y no habría que determinar las longitudes de sus caminos. Y como esto mismo puede ocurrir con otros PAST menores, pasariamos al valor PAST más pequeño del nodo madre para el que es necesario analizar la condición de mínimo, $PMIN$, y repetimos lo indicado en d2) y d3).

Si (2.62) es cierta para ese $PMIN$ también lo sería para los restantes valores PAST del nodo madre (observación 2.7), y sin necesidad de mas comprobaciones todas las tablas correspondientes a ese nodo hijo para esos valores PAST quedarian fuera de la región crítica; pasariamos a repetir desde d1) el proceso con otro nodo hijo.

En otro caso, o sea si para el PAST mas pequeño PMIN, la condición (2.62) no fuese cierta, a partir de él y en orden creciente iriamos analizando los sucesivos PAST hasta llegar a uno para el que la condición (2.62) sea cierta, con lo que para todos los valores restantes sin necesidad de mas comprobaciones también lo sería y todas sus tablas quedarían así mismo fuera de la región crítica, quedando este nodo hijo con sus valores PAST para los que (2.62) no es cierta pendientes del análisis en el estado siguiente. Para ello habrá que almacenarlos.

Seguimos el proceso desde d1) con el siguiente nodo hijo de la subsucesión.

D) En otro caso, o sea si (2.62) no es cierta para el valor PAST máximo del nodo madre, tampoco sería cierta para los restantes valores PAST quedando este nodo hijo con sus valores PAST (desde PMIN hasta $PAST_M$) pendientes del análisis en el estado siguiente, se almacena, y reiniciamos el proceso desde d1) con otro nodo hijo.

Almacenamos los nodos sobre los que no se ha podido decidir bien la incorporación de todas sus tablas a la región crítica, o bien la eliminación de las mismas, con su información característica: valores PAST, frecuencia de estos PAST; para el nodo de menor rango almacenamos también Moda de su MHD asociada y LMA. Finalmente, si para un nodo que se va a almacenar ya existe una

permutación de él almacenada, entonces no lo almacenamos, simplemente se añade al nodo almacenado los PAST del nodo pendiente de almacenar (teorema de simetría de M-P, analizado en el capítulo I, punto I.2.2).

Todos los nodos hijos almacenados se convertirán para el estado siguiente en nodos madres para los que se repetirá el proceso desde 6).

II.4.3: Ejemplo de Aplicación del Algoritmo Modificado.

Determinación del p-valor de la tabla

$$X = \begin{bmatrix} 0 & 0 & 1 & 4 & 1 \\ 4 & 2 & 3 & 0 & 1 \end{bmatrix}$$

Tabla 2x5 con totales marginales filas $R_1=6$, $R_2=10$, marginales columnas $C_1=4$, $C_2=2$, $C_3=4$, $C_4=4$, $C_5=2$, total global $N=16$.

1- Determinamos $L(X)$

$$L(X) = \binom{4}{0} \binom{2}{0} \binom{4}{1} \binom{4}{0} \binom{2}{1} = 8$$

2- Determinamos D

$$D = \frac{16!}{6!10!} = 8008$$

3- Nodo inicial $(c; R_{1c}) = (5; 6)$

4- Moda MHD asociada al nodo $(5; 6) = (1,1,2,1,1)$

$$LMAG(5; 6) = LMA(5; 6) = 384$$

5- $\text{LMAG}(5; 6) \leq L(X)$ FALSO

Hacemos p-valor = 0 y continuamos.

6- Determinamos nodos del estado 4

$$4 \leq R_{14} \leq 6$$

7- Determinamos nodo hijo principal del nodo madre mas pequeño, moda de su MHD asociada, su LMA, y a partir de aquí, determinamos LMA de los demás nodos hijos.

También determinamos LMI para todos los nodos hijos.

Nodo	Moda MHD	LMA	LMI
(4; 6)		288	1
↑	↑	↑	
(4; 5)	(1,1,2,1)	192	2
↓	↓	↓	
(4; 4)		128	1

8- Análisis de los nodos del estado 5

Recuperación del primer nodo madre (en este caso único nodo del estado): (5; 6)

Nodo	PAST	f
(5; 6)	1	1

	Nodos hijos	ℓ
I)	(4; 6)	1
	(4; 5)	2
	(4; 4)	1

II) Análisis de la condición de máximo: $\text{LMAG} \leq L(X)$ (2.61)

a) Nodo hijo principal: (4; 5)

b) Subsucesión ascendente $SS_1 = \{(4; 5), (4; 6)\}$; $SS'_1 = \{(4; 6)\}$

Subsucesión descendente $SS_0 = \{(4; 5), (4; 4)\}$; $SS'_0 = \{(4; 4)\}$

	Nodo h.	ℓ	PAST	f	ℓ .	PAST	LMAG	ℓ LMAG \leq L(X)?	
SS'1 :	(4; 6)	1	1	1	1	1	288	No	Pendiente C. Min. \rightarrow SS'1(R)
SS'0 :	(4; 4)	1	1	1	1	1	128	No	"

III) Análisis de la condición de mínimo: $LMIG > L(X)$ (2.62)

	Nodo h.	ℓ	PAST	f	ℓ .	PAST	LMIG	ℓ LMIG $>$ L(X)?	
SS'1(R) :	(4; 6)	1	1	1	1	1	1	No	Se Almacena
Nodo P :	(4; 5)	2	1	1	2	2	4	No	Se Almacena
SS'0(R) :	(4; 4)	1	1	1	1	1	1	No	Se Almacena

Para el estado siguiente ha quedado almacenado

Nodo	PAST	f	Moda MHD	LMA
(4; 6)	1	1		
(4; 5)	2	1		
(4; 4)	1	1	(1,1,1,1)	128

Reiniciamos proceso desde 6)

6- Nodos del estado 3

$$0 \leq R_{13} \leq 6$$

Nodo	Moda MHD	LMA	LMI
(3; 6)		48	1
		↑	
(3; 5)		72	2
		↑	
(3; 4)		48	1
	↑	↑	
7- $\langle(3; 3)\rangle$	(1,1,1)	32	4
	↓	↓	
(3; 2)		16	1
		↓	
(3; 1)		4	2
	↓	↓	
(3; 0)		1	1

8- Análisis de los nodos del estado 4

Recuperación del primer nodo madre:(4; 4)

Nodo PAST f
(4; 4) 1 1

Nodos hijos ℓ
 (3; 4) 1
 I) (3; 3) 4
 (3; 2) 6
 (3; 1) 4
 (3; 0) 1

II) Análisis de la condición de máximo $LMAG \leq L(X)$

a) Nodo hijo principal: (3; 3)

b) Subsucesión ascendente $SS_1 = \{(3; 3), (3; 4)\}$; $SS'_1 = \{(3; 4)\}$

Subsucesión descendente $SS_0 = \{(3; 3), (3; 2), (3; 1), (3; 0)\}$;

$SS'_0 = \{(3; 0), (3; 1), (3; 2)\}$

	Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\zeta LMAG \leq L(X)$?	
SS'_1 :	(3; 4)	1	1	1	1	48	No	Pendiente C.Min. $\rightarrow SS'_1(R)$
SS'_0 :	(3; 0)	1	1	1	1	1	Si	Se incorpora al p - valor
								$SL = \binom{10}{0} \cdot 1.1 = 1; SL/D = 1/8008$
								$p = 0.000124875$
	(3; 1)	4	1	1	4	16	No	Pendiente C.Min. $\rightarrow SS'_0(R)$
	(3; 2)	6	1	1	6	\rightarrow	\rightarrow	„

III) Análisis de la condición de mínimo: $LMIG > L(X)$

	Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMIG	$\zeta LMIG > L(X)$?	
$SS'_1(R)$:	(3; 4)	1	1	1	1	1	No	Se Almacena
Nudo P :	(3; 3)	4	1	1	4	16	Si	Fuera R.C
$SS'_0(R)$:	(3; 1)	4	1	1	4	8	No	Se Almacena
	(3; 2)	6	1	1	6	6	No	Se Almacena

Recuperación del siguiente nodo madre: (4; 5)

Nodo PAST f
(4; 5) 2 1

Nodos Hijos		ℓ
I)	(3;5)	1
	(3;4)	4
	(3;3)	6
	(3;2)	4
	(3;1)	1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (3; 4)

b) Subsucesión ascendente $SS_1 = \{(3; 4), (3; 5)\}$; $SS'_1 = \{(3; 5)\}$

Subsucesión descendente $SS_0 = \{(3; 4), (3; 3), (3; 2), (3; 1)\}$

$SS'_0 = \{(3; 1), (3; 2), (3; 3)\}$

Nodo h.	ℓ	PAST	f	ℓ .PAST	LMAG	ℓ LMAG $\leq L(X)$?		
SS'1 :	(3;5)	1	2	1	2	144	No	Pendiente C.Min. \rightarrow SS'1(R)
SS'0	(3;1)	1	2	1	2	8	Si	Se incorpora al p - valor
								$SL = \binom{10}{1} \cdot 1.2 = 20; SL / D = 20 / 8008$
								$p = 0.00262237$
	(3;2)	4	2	1	8	128	No	Pendiente C.Min. \rightarrow SS'0(R)
	(3;3)	6	2	1	12	\rightarrow	\rightarrow	„

III) Análisis de la condición de mínimo: $LMIG > L(X)$

Nodo hijo	ℓ	PAST	f	ℓ .PAST	LMIG	ℓ LMIG $> L(X)$?		
SS'1(R) :	(3;5)	1	2	1	2	4	No	Se Almacena
Nodo P :	(3;4)	4	2	1	8	8	No	Se Almacena
								(Ya alm. se agrega PAST)
SS'0(R)	(3;2)	4	2	1	8	8	No	Se Almacena
								(Ya Alm. se agrega PAST)
	(3;3)	6	2	1	12	48	Si	Fuera R.C.

- Recuperación del siguiente nodo madre: (4; 6)

Nodo	PAST	f
(4;6)	1	1

	Nodos hijos	ℓ
	(3;6)	1
I)	(3;5)	4
	(3;4)	6
	(3;3)	4
	(3;2)	1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (3; 4)

b) Subsucesión ascendente $SS_1 = \{(3; 4), (3; 5), (3; 6)\};$

$$SS'_1 = \{(3; 6), (3; 5)\}$$

Subsucesión descendente $SS_0 = \{(3; 4), (3; 3), (3; 2)\}$

$$SS'_0 = \{(3; 2), (3; 3)\}$$

	Nodo hijo	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\ell LMAG \leq L(X)?$	
SS'1 :	(3;6)	1	1	1	1	48	No	Pendiente C.Min \rightarrow SS'1(R)
	(3;5)	4	1	1	4	\rightarrow	\rightarrow	,,
SS'0 :	(3;2)	1	1	1	1	16	No	Pendiente C.Min. \rightarrow SS'0(R)
	(3;3)	4	1	1	4	\rightarrow	\rightarrow	,,

III) Análisis de la condición de mínimo: $LMIG > L(X)$

	Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMIG	$\ell LMIG > L(X)?$	
SS'1(R) :	(3;6)	1	1	1	1	1	No	Se Almacena (Integrado con (3; 4))
	(3;5)	4	1	1	4	8	No	Se Almacena (Ya alm.se agrega PAST)
Nodo P :	(3;4)	6	1	1	6	6	No	Se Almacena (Ya alm.se agrega PAST)
SS'0(R) :	(3;2)	1	1	1	1	1	No	Se Almacena (Ya alm.se incorpora PAST)
	(3;3)	4	1	1	4	16	Si	Fuera R.C.

Para el estado siguiente ha quedado almacenado:

Nodo	PAST	f	Moda MHD	LMA
(3;5)	2	1		
	4	1		
(3;4)	1	2		
	6	1		
	8	1		
(3;2)	1	1		
	6	1		
	8	1		
(3;1)	4	1	(1,0,0)	4

Reiniciamos proceso desde 6)

6- Nodos del estado 2

$$0 \leq R_{12} \leq 5$$

Nodo	Moda MHD	LMA	LMI
(2;5)		4	2
		↑	
(2;4)		8	1
		↑	
7- (2;3)		6	4
		↑	
(2;2)		8	1
	↑	↑	
⟨(2;1)⟩	(1,0)	4	2
	↓	↓	
(2;0)		1	1

8- Análisis de los nodos del estado 3

- Recuperación del primer nodo madre: (3; 1)

Nodo	PAST	f
(3;1)	4	1

Nodos hijos	ℓ
I) (2;1)	1
(2;0)	4

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (2; 1)

b) Subsucesión ascendente $SS_1 = \{(2; 1)\}$; $SS'_1 = \emptyset$

Subsucesión descendente $SS_0 = \{(2; 1), (2; 0)\}$; $SS'_0 = \{(2; 0)\}$

	Nodo h.	ℓ	PAST	f	ℓ .PAST	LMAG	ℓ LMAG \leq L(X)?	
SS'0 :	(2;0)	4	4	1	16	16	No	Pendiente C.Min. \rightarrow SS'0(R)

III) Análisis de la condición de mínimo: $LMIG > L(X)$

	Nodo h.	ℓ	PAST	f	ℓ .PAST	LMIG	ℓ LMIG $>$ L(X)?	
Nodo P :	(2;1)	1	4	1	4	8	No	Se Almacena
SS'0(R) :	(2;0)	4	4	1	16	16	Si	Fuera R.C.

Recuperación del siguiente nodo madre: (3; 2)

Nodo	PAST	f
(3;2)	1	1
	6	1
	8	1

	Nodos hijos	ℓ
I)	(2;2)	1
	(2;1)	4
	(2;0)	6

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a)Nodo hijo principal: (2; 1)

b) Subsucesión ascendente $SS_1 = \{(2; 1), (2; 2)\}$; $SS'_1 = \{(2; 2)\}$

Subsucesión descendente $SS_0 = \{(2; 1), (2; 0)\}$; $SS'_0 = \{(2; 0)\}$

Nodo h.	ℓ	PAST	f	ℓ .PAST	LMAG	ζ LMAG \leq L(X)?	
SS'1 :	(2;2)	P1 = 1	1	1	8	Si	Se incorpora al p - valor;
		PM = P3 = 8	1	8	64	No	Pendiente C.Min. \rightarrow SS'1(R)
		P2 = 6	1	6	48	No	,,; PF + 1 = P2
							$SL = \binom{6}{2} \cdot 1 \cdot (1.1) = 15$
							$SL / D = 15 / 8008; p = 0.00449555$
SS'0 :	(2;0)	P1 = 1	1	6	6	Si	Se incorpora al p - valor
		PM = P3 = 8	1	48	48	No	Pendiente C.Min. \rightarrow SS'0(R)
		P2 = 6	1	36	36	No	,,; PF + 1 = P2
							$SL = \binom{6}{0} \cdot 6 \cdot (1.1) = 6$
							$SL / D = 6 / 8008; p = 0.00524475$

III) Análisis de la condición de mínimo: LMIG > L(X)

Nodo h.	ℓ	PAST	f	ℓ .PAST	LMIG	ζ LMIG > L(X)?	
SS'1(R) :	(2;2)	PM = P3 = 8	1	8	8	No	Se Almacena
		PF + 1 = P2 = 6	1	6	\rightarrow	\rightarrow	,,
Nodo P :	(2;1)	PM = P3 = 8	1	32	64	Si	Fuera R.C.
		P1 = 1	1	4	8	No	Se Almacena
		P2 = 6	1	24	48	Si	Fuera R.C.
SS'0(R) :	(2;0)	PM = P3 = 8	1	48	48	Si	Fuera R.C.
		PF + 1 = P2 = 6	1	36	36	Si	Fuera R.C.

Recuperación del siguiente nodo madre: (3; 4)

Nodo	PAST	f
(3;4)	1	2
	6	1
	8	1

Nodos hijos	ℓ
(2;4)	1
(2;3)	4
I) (2;2)	6
(2;1)	4
(2;0)	1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (2; 2)

b) Subsucesión ascendente $SS_1 = \{(2; 2), (2; 3), (2; 4)\}$; $SS'_1 = \{(2; 4), (2; 3)\}$

Subsucesión descendente $SS_0 = \{(2; 2), (2; 1), (2; 0)\}$; $SS'_0 = \{(2; 0), (2; 1)\}$

Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\zeta LMAG \leq L(X)?$	
SS'1 :	(2; 4)	P1=1	2	1	8	Si	Se incorpora al p - valor;
		PM = P3 = 8	1	8	64	No	Pendiente C.Min. \rightarrow SS'1(R)
		P2 = 6	1	6	48	No	,,; PF + 1 = P2
							$SL = \binom{6}{4} \cdot 1 \cdot (1.2) = 30$
							$SL / D = 30 / 8008; p = 0.008991$
(2; 3)	4	P1=1	2	4	24	No	Pendiente C.Min. \rightarrow SS'1(R)
		P2 = 6	1	24	\rightarrow	\rightarrow	,,
		P3 = PM = 8	1	32	\rightarrow	\rightarrow	,,
SS'0 :	(2; 0)	P1=1	2	1	1	Si	Se incorpora al p - valor;
		PM = P3 = 8	1	8	8	Si	,,
		P2 = 6	1	6	\rightarrow	\rightarrow	,,
							$SL = \binom{6}{0} \cdot 1 \cdot (1.2 + 6.1 + 8.1) = 16$
							$SL / D = 16 / 8008; p = 0.010989$
(2; 1)	4	P1=1	2	4	16	No	Pendiente C.Min. \rightarrow SS'0(R)
		P2 = 6	1	24	\rightarrow	\rightarrow	,,
		P3 = 8	1	32	\rightarrow	\rightarrow	,,

III) Análisis de la condición de mínimo: $LMIG > L(X)$

Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMIG	$\zeta LMIG > L(X)?$	
SS'1(R) :	(2; 4)	PM = P3 = 8	1	8	8	No	Se Almacena
		PF + 1 = P2 = 6	1	6	\rightarrow	\rightarrow	(Integrado con (2; 2))
(2; 3)	4	PM = P3 = 8	1	32	128	Sí	Fuera R.C.
		P1 = 1	2	4	16	Sí	Fuera R.C.
		P2 = 6	1	24	\rightarrow	\rightarrow	,,
Nodo P :	(2; 2)	PM = P3 = 8	1	48	48	Sí	Fuera R.C.
		P1 = 1	2	6	6	No	Se Almacena
		P2 = 6	1	36	36	Sí	(Ya Alm.se agrega PAST)
SS'0(R) :	(2; 1)	PM = P3 = 8	1	32	64	Sí	Fuera R.C.
		P1 = 1	2	4	8	No	Se Almacena
		P2 = 6	1	24	48	Sí	(Ya Alm.se agrega PAST)
							Fuera R.C.

Recuperación del siguiente nodo madre: (3; 5)

Nodo	PAST	f
(3;5)	2	1
	4	1

I)	Nodos hijos	ℓ
	(2;5)	1
	(2;4)	4
	(2;3)	6
	(2;2)	4
	(2;1)	1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (2; 3)

b) Subsucesión ascendente $SS_1 = \{(2; 3), (2; 4), (2; 5)\}$; $SS'_1 = \{(2; 5), (2; 4)\}$

Subsucesión descendente $SS_0 = \{(2; 3), (2; 2), (2; 1)\}$; $SS'_0 = \{(2; 1), (2; 2)\}$

	Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\ell LMAG \leq L(X)?$	
SS'1 :	(2;5)	1	2	1	2	8	Sí	Se incorpora al p - valor
			4	1	4	16	No	Pendiente C.Mín. $\rightarrow SS'1(R)$
								$SL = \binom{6}{5} \cdot 1 \cdot (2) = 12$
								$SL / D = 12 / 8008; P = 0.0124875$
	(2;4)	4	2	1	8	64	No	Pendiente C.Min. $\rightarrow SS'1(R)$
			4	1	16	\rightarrow	\rightarrow	„
SS'0 :	(2;1)	1	2	1	2	8	Sí	Se incorpora al p - valor
			4	1	4	16	No	Pendiente C.Min. $\rightarrow SS'0(R)$
								$SL = \binom{6}{1} \cdot 1 \cdot (2) = 12$
								$SL / D = 12 / 8008; p = 0.013986$
	(2;2)	4	2	1	8	64	No	Pendiente C.Min. $\rightarrow SS'0(R)$
			4	1	8	\rightarrow	\rightarrow	„

III) Análisis de la condición de mínimo: $LMIG > L(X)$

	Nodo	h	ℓ	PAST	f	ℓ.PAST	LMIG	¿LMIG > L(X)		
SS'1(R) :	(2;5)	1	4	4	1	4	8	No	Se Almacena	
	(2;4)	4	4	4	1	16	16	Sí	Fuera R.C.	
		2	1	8	8	8	No	Se Almacena		
(Ya Alm.se agrega PAST)										
Nodo P :	(2;3)	6	4	4	1	24	96	Sí	Fuera R.C.	
		2	1	12	48	Sí	Fuera R.C.			
SS'0(R) :	(2;1)	1	4	4	1	4	8	No	Se Almacena	
	(Ya Alm.se agrega PAST)									
	(2;2)	4	4	4	1	16	16	Sí	Fuera R.C.	
2		1	8	8	8	No	Se Almacena			
(Ya Alm.se agrega PAST)										

Para el estado siguiente ha quedado almacenado

Nodo	PAST	f	Moda MHD	LMA
(2;5)	4	1		
(2;2)	6	4		
	8	4		
(2;1)	4	5	(1,0)	4

Reiniciamos el proceso desde 6)

6- Nodos del estado 1

$$0 \leq R_{11} \leq 4$$

	Nodo	LMA	LMI
7-	(1;4)	1	1
	(1;3)	4	4
	(1;2)	6	6
	(1;1)	4	4
	(1;0)	1	1

8- Análisis de los nodos del estado 2

Recuperación del primer nodo madre: (2; 1)

Nodo	PAST	f
(2;1)	4	5

	Nodos hijos	ℓ
I)	(1;1)	1
	(1;0)	2

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (1; 1)

b) Subsucesión ascendente $SS_1 = \{(1; 1)\}; SS'_1 = \emptyset$

Subsucesión descendente $SS_0 = \{(1; 1), (1; 0)\}; SS'_0 = \{(1; 0)\}$

Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\ell LMAG \leq L(X)?$		
SS'0	(1;0)	2	4	5	8	8	Sí	Se incorpora al p - valor

$$SL = \binom{4}{0} \cdot 2 \cdot (4 \cdot 5) = 40$$

$$SL / D = 40 / 8008; p = 0.018981$$

III) El Análisis de la condición de mínimo no es necesario realizarlo pues las tablas que no se han incorporado a la región crítica ya quedan fuera.

Recuperación del siguiente nodo madre: (2; 2)

Nodo	PAST	f
(2; 2)	6	4
	8	4

Nodos hijos	ℓ
I) (1; 2)	1
(1; 1)	2
(1; 0)	1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (1; 1)

b) Subsucesión ascendente $SS_1 = \{(1; 1), (1; 2)\}; SS'_1 = \{(1; 2)\}$

Subsucesión descendente $SS_0 = \{(1; 1), (1; 0)\}; SS'_0 = \{(1; 0)\}$

Nodo h.	ℓ	PAST	f	$\ell \cdot PAST$	LMAG	$\ell LMAG \leq L(X)?$		
SS'1 :	(1; 2)	1	6	4	6	36	No	Fuera R.C
			8	4	8	→	→	..
SS'0 :	(1; 0)	1	6	4	6	6	Sí	Se incorpora al p - valor
			8	4	8	8	Sí	..

$$SL = \binom{4}{0} \cdot 1 \cdot (6 \cdot 4 + 8 \cdot 4) = 56$$

$$SL / D = 56 / 8008; p - valor = 0.025974$$

Recuperación del siguiente nodo madre: (2; 5)

Nodo PAST f
(2;5) 4 1

Nodos hijos ℓ
I) (1;4) 2
(1;3) 1

II) Análisis de la condición de máximo: $LMAG \leq L(X)$

a) Nodo hijo principal: (1; 3)

b) Subsucesión ascendente $SS_1 = \{(1; 3), (1; 4)\}$; $SS'_1 = \{(1; 4)\}$

Subsucesión descendente $SS_0 = \{(1; 3)\}$; $SS'_0 = \emptyset$

Nodo h.	ℓ	PAST	f	ℓ .PAST	LMAG	ζ LMAG \leq L(X)?	
SS'1:	(1;4)	2	4	1	8	8	Sí

Se incorpora al p - valor

$$SL = \binom{4}{4} \cdot 2 \cdot (4.1) = 8$$

$$SL / D = 8 / 8008; p - valor = 0.026973$$

Así pues, para la tabla observada en este ejemplo se obtiene finalmente p-valor = 0.026973.

II.5. Discusión de las mejoras aportadas al Algoritmo de resolución del Test de Fisher para tablas 2xc.

La discusión la realizaremos comparando nuestro algoritmo modificado (II.4.2), al que aludiremos de aquí en adelante simplemente como algoritmo modificado, con el algoritmo, que por ser el más eficaz entre los existentes nos ha servido de referencia y base de nuestra propia investigación, el algoritmo de Mehta and Patel (1986) restringido a tablas 2xc, con las modificaciones introducidas por Harry Joe (1988) para determinar las probabilidades extremas, al que aludiremos abreviadamente como algoritmo M-P-H.

Los ocho pasos considerados en la configuración del algoritmo modificado se corresponden con los que pueden establecerse para el algoritmo de M-P-H, pero nuestro método de resolución incorpora importantes mejoras para algunos de estos pasos.

En el algoritmo modificado hay que utilizar el método de determinación de la moda de una MHD solo para la MHD asociada al nodo inicial de la red, mientras que en el algoritmo de M-P-H hay que utilizarlo para calcular las modas de todas las MHD asociadas a nodos hasta los que se extiende el proceso.

El cálculo de una moda de la MHD asociada al nodo inicial, utilizando el nuevo método de determinación de la moda de una MHD que hemos elaborado (II.2.5) en lugar del método de Harry Joe también nos proporciona como ha quedado ampliamente constatado (II.2.6) importantes mejoras sobre este punto del proceso (paso 4).

En el algoritmo modificado, (paso 8), para cada nodo $(k; R_{1k})$ del estado k , con una serie de valores PAST, y con una serie de nodos hijos $(k-1; R_{1,k-1})$, evitamos el comprobar (2.61) o en su caso (2.62) para muchas combinaciones posibles de PAST y de nodos hijos. En el algoritmo de M-P-H hay que comprobar (2.61) o en su caso (2.62) para todas las combinaciones posibles de PAST y de nodos hijos. En el ejemplo analizado, y a pesar de la reducida dimensión de la tabla, aparecen ya algunas combinaciones para las que evitamos la comprobación. Evitamos comprobación de (2.61) para todos los nodos hijos principales y además para:

Nodo madre	PAST	Nodo hijo
(4;4)	1	(3;2)
(4;5)	2	(3;3)
(4;6)	1	(3;5)
(4;6)	1	(3;3)
(3;5)	4	(2;4)
(3;5)	4	(2;2)
(3;4)	6	(2;3)
(3;4)	8	(2;3)
(3;4)	6	(2;1)
(3;4)	8	(2;1)
(2;2)	8	(1;2)

y evitamos la comprobación de (2 .62) para

Nodo madre	PAST	Nodo hijo
(3;2)	6	(2;2)
(3;4)	1	(2;4)
(3;4)	6	(2;4)
(3;4)	6	(2;3)

Los tiempos de cómputo del p-valor, como es obvio, son siempre menores en el algoritmo modificado, aunque a efectos prácticos eso se empieza a notar para valores de c de 8 ó 10 en adelante, y el ahorro, como es lógico, también dependerá de la tabla observada. Para valores grandes de c los ahorros de tiempos de cómputo son muy importantes, como siempre dependiendo de la tabla observada.

Una atención a los tiempos de cómputo del p-valor para algunas tablas nos pueden servir para evidenciar con mayor claridad aún, las ventajas de las aportaciones realizadas. Los tiempos CPU usados por el algoritmo modificado han sido comparados con los correspondientes tiempos CPU usados por el algoritmo de M-P-H. Ambos algoritmos han sido programados en un PC Pentium120 Mhz. La tabla 2.1 muestra los tiempos de ahorro conseguidos.

Tabla 2.1. Tiempos de cómputo del p-valor

Tamaño de la tabla	Tabla de contingencia	p - valor	Tiempos A.M - P - H	Tiempos A.Modificado
2x10	20 3 6 4 7 6 6 2 2 2 8 8 4 5 2 1 0 2 1 1	0.082538	3seg.24c.	1seg.42c
-	-	-	-	-
2x11	0 12 3 11 8 8 5 2 2 1 0 15 3 12 4 7 7 10 3 3 4 2	0.0000325	4 min.49seg.	1 min.0 seg.
-	-	-	-	-
2x13	2 2 2 2 3 6 2 2 1 5 5 9 11 0 0 0 0 4 1 5 6 7 4 4 0 0	0.00001864	32 seg.46c.	8seg.79c
-	-	-	-	-
2x13	0 0 1 2 7 7 6 3 2 3 4 8 9 2 2 1 0 0 0 1 5 6 6 5 1 2	0.000137	30 seg.98c.	9 seg.7c.
-	-	-	-	-
2x16	1 3 3 4 5 5 5 5 4 4 8 8 6 6 4 5 1 2 5 5 11 11 11 11 12 12 8 8 10 10 15 15	0.8388	59 min.8seg.	11 min.28 seg.
-	-	-	-	-
2x20	8 13 10 9 11 12 10 11 9 12 8 7 7 6 6 5 4 1 3 1 11 4 7 8 6 5 7 6 8 5 9 10 10 11 11 12 1 2 0 2	0.044348	5h.49min.	32min.23seg
-	-	-	-	-
2x22	3 7 4 4 3 3 4 3 4 5 5 4 2 4 2 5 6 6 3 1 0 0 6 2 3 3 4 4 3 4 3 2 2 3 5 3 5 1 0 0 1 2 1 1	0.10699	2min.33seg.	36seg.19c.
-	-	-	-	-
2x23	1 1 2 3 3 6 5 5 6 6 6 7 7 4 4 3 3 6 5 4 7 8 4 2 2 1 2 4 4 5 6 5 5 5 4 4 7 7 8 8 7 8 9 6 5 4	0.8856	4h.1min.5seg.	8min.46seg.

**C A P Í T U L O III: Una extensión a tablas rxc de las mejoras
obtenidas en el algoritmo del Test de Fisher para tablas 2xc.**

III.1: Introducción

En el capítulo anterior hemos desarrollado una serie de cuestiones, que mejoran el Algoritmo del Test de Fisher para tablas $2 \times c$ en cuanto que simplifican y permiten ejecutar con mayor rapidez algunos de los procesos implicados en el Algoritmo reticulado de Mehta y Patel, considerado en la actualidad como el Algoritmo más eficaz para ejecutar dicho test. Estas mejoras se han orientado, como ha quedado señalado, en una doble vertiente:

- Obtención de la trayectoria más larga desde determinados nodos de la red al nodo terminal.
- Relaciones entre las trayectorias extremas correspondientes a nodos de la red.

Este mismo interés nos guiará también en el caso $r \times c$, pero si en $2 \times c$ la solución satisfactoria a la primera cuestión equivale a determinar un método eficaz para calcular el máximo de una Distribución Hipergeométrica Multivariante (MHD), en $r \times c$ equivale a determinar el máximo de lo que podríamos llamar una Distribución Hipergeométrica Multivariante Múltiple (MMHD). También y, al igual que sucedía en $2 \times c$, esto exige el desarrollo de una serie de cuestiones que inciden en la relación entre las subtrayectorias más largas correspondientes a los nodos madres y a los nodos hijos, y que permiten acortar el proceso de determinación de las subtrayectorias más largas correspondientes a nodos diferentes al inicial una vez que se dispone de la trayectoria más larga para este nodo.

Iniciaremos la extensión con $r = 3$ abriendo la vía de generalización para cualquier valor de r .

III.2: Planteamiento y resolución de cuestiones de interés sobre la tabla de máxima probabilidad de una familia isomarginal de tablas 3xc.

III.2.1: Introducción.

Sea T una tabla de contingencia $r \times c$, con entradas x_{ij} en la fila i y columna j , $R_i = \sum_{j=1}^c x_{ij}$ la suma de todas las entradas en la fila i , y $C_j = \sum_{i=1}^r x_{ij}$ la suma de todas las entradas en la columna j , y donde tanto las x_{ij} como sus posibles sumas parciales son enteros no negativos.

Denotemos por \mathfrak{T} la familia de todas las posibles tablas de contingencia con los mismos totales marginales que T

$$\mathfrak{T} = \{X / X \text{ es } r \times c, \sum_{j=1}^c x_{ij} = R_i, i=1, \dots, r, \sum_{i=1}^r x_{ij} = C_j, j=1, \dots, c\}$$

Bajo la hipótesis nula de independencia de filas y columnas a la distribución de probabilidad de las tablas de esta familia le llamaremos Distribución Hipergeométrica Múltiple multivariante (MMHD)

$$\text{MMH}(R_1, \dots, R_r; C_1, \dots, C_c) \quad (3.1)$$

El problema de determinar la tabla de máxima probabilidad de la familia \mathfrak{T} es equivalente al problema de obtener el máximo de la MMHD(3.1), siendo

$$P(X) = \frac{\left\{ \prod_{j=1}^c C_j! \right\} \left\{ \prod_{i=1}^r R_i! \right\}}{N! \prod_{j=1}^c \prod_{i=1}^r x_{ij}!} \quad (3.2)$$

donde $N = \sum_{i=1}^r \sum_{j=1}^c x_{ij}$

Obviamente, para cada par de subconjuntos $I \subset \{1, \dots, r\}$, $J \subset \{1, \dots, c\}$, con cardinales por lo menos de 2, la subtabla $X^{IJ} = (x_{ij})_{i \in I, j \in J}$ seguirá también una

MMHD. Para el caso de que $\text{Card}(I) = 2$, tendremos el caso particular de una MHD, $MH(R_i; C_1, \dots, C_c) \quad i \in I$. Y para el caso mas particular aun de que tanto $\text{Card}(I)$ como $\text{Card}(J)$ sean iguales a 2, tendremos una distribución hipergeométrica $H(R_i; C_j; N) \quad i \in I, j \in J$.

Igualmente, cada marginal fila $X^i = (x_{i1}, \dots, x_{ic})$ de (3.1) sigue una $MH(R_i; C_1, \dots, C_c)$ para todo $i=1, \dots, r$, cada marginal columna ${}^jX = (x_{1j}, \dots, x_{rj})$ de (3.1) sigue una $MH(C_j; R_1, \dots, R_r)$ para todo $j=1, \dots, c$, y cada $x_{ij}, i=1, \dots, r, j=1, \dots, c$, de (3.1) sigue una $H(R_i; C_j; N)$.

Y, según la expresión de la moda de una Hipergeométrica (1.38), la moda de cada x_{ij} $i=1, \dots, r, j=1, \dots, c$ de (3.1) es un entero, llamemosle z_{ij} tal que

$$\frac{(R_i + 1)(C_j + 1)}{N + 2} - 1 \leq z_{ij} \leq \frac{(R_i + 1)(C_j + 1)}{N + 2}$$

y, para el caso de una Distribución Hipergeométrica multivariante, el máximo puede determinarse por el método desarrollado en el capítulo anterior y puede expresarse en la forma indicada en el teorema 2.6 de dicho capítulo. En particular la moda de cada marginal fila o cada marginal columna puede determinarse por dicho procedimiento.

III.2.2: Moda de una MMHD

Teorema 3.1. Sean

$$T = \{x_{ij}, i=1, \dots, r, j=1, \dots, c\} \text{ y } T' = \{x'_{ij}, i=1, \dots, r, j=1, \dots, c\}$$

dos tablas de la familia \mathfrak{T} tales que para i_0, t_0 dos valores de i diferentes, p_0, q_0 dos valores de j diferentes

$$x'_{i_0 p_0} = x_{i_0 p_0} + 1 \quad x'_{i_0 q_0} = x_{i_0 q_0} - 1$$

$$x'_{t_0 p_0} = x_{t_0 p_0} - 1 \quad x'_{t_0 q_0} = x_{t_0 q_0} + 1$$

y,

$x'_{ij} = x_{ij}$ para el resto de valores i, j

entonces las probabilidades de T y T' quedan relacionadas en la siguiente forma

$$P(T') > P(T) \Leftrightarrow \frac{x_{i_0 p_0} + 1}{x_{i_0 p_0}} < \frac{x_{i_0 q_0}}{x_{i_0 q_0} + 1} \quad (3.3)$$

Cumplendose también dicha equivalencia con el signo $=$.

Demostración.- Sean $P(T)$ y $P(T')$ los valores de (3.2) para T y T' , entonces

$$\frac{P(T')}{P(T)} = \frac{x_{i_0 q_0}}{x_{i_0 p_0} + 1} \cdot \frac{x_{i_0 p_0}}{x_{i_0 q_0} + 1}$$

con lo que

$$P(T') > P(T) \Leftrightarrow \frac{x_{i_0 q_0}}{x_{i_0 p_0} + 1} \cdot \frac{x_{i_0 p_0}}{x_{i_0 q_0} + 1} > 1 \Leftrightarrow \frac{x_{i_0 p_0} + 1}{x_{i_0 p_0}} < \frac{x_{i_0 q_0}}{x_{i_0 q_0} + 1}$$

La equivalencia con el signo $=$ es trivial.

Corolario 3.1. Sea $T = \{x_{ij} \ i=1, \dots, r, \ j=1, \dots, c\}$ una tabla de la familia \mathfrak{S} , entonces la condición

$$\frac{x_{ip} + 1}{x_{ip}} \geq \frac{x_{iq}}{x_{iq} + 1} \quad \text{para todo } i \neq t, \ p \neq q \quad (3.4)$$

es condición necesaria para que T sea de máxima probabilidad.

Demostración.- Si $T = \{x_{ij} \ i=1, \dots, r, \ j=1, \dots, c\}$ es una tabla de máxima probabilidad de la familia \mathfrak{S} entonces la condición (3.4) es cierta pues si fuese falsa existirían i_0, t_0 , dos valores diferentes de i , y existirían p_0, q_0 , dos valores diferentes de j tales que

$$\frac{x_{i_0 p_0} + 1}{x_{i_0 p_0}} < \frac{x_{i_0 q_0}}{x_{i_0 q_0} + 1}$$

y entonces, podríamos construir la tabla $T' = \{x'_{ij} \ i=1, \dots, r, \ j=1, \dots, c\}$ en la forma

$$x'_{i_0 p_0} = x_{i_0 p_0} + 1 \quad x'_{i_0 q_0} = x_{i_0 q_0} - 1$$

$$x'_{t_0 p_0} = x_{t_0 p_0} - 1 \quad x'_{t_0 q_0} = x_{t_0 q_0} + 1$$

y,

$$x'_{ij} = x_{ij} \text{ para el resto de valores } i, j$$

con lo que aplicando el teorema 3.1 a las tablas T y T' tendríamos $P(T') > P(T)$ lo que es absurdo.

Corolario 3.2. Sea $T = \{x_{ij} \ i=1, \dots, r, j=1, \dots, c\}$ una tabla de la familia \mathfrak{S} , entonces (3.4) es condición necesaria y suficiente para que todas las subtablas $2 \times c$ de T sean de máxima probabilidad.

Demostración.- Para cada subconjunto $I \subset \{1, \dots, r\}$ de cardinal dos es condición necesaria y suficiente (teorema 2.1 ó teorema 2.1*) para que la subtabla $X^{IJ} = (x_{ij})_{i \in I, j \in J}$, $J=1, \dots, c$, sea de máxima probabilidad que todas sus subtablas 2×2 sean de máxima probabilidad, lo que equivale a que se verifique (3.4), pues si todas las subtablas 2×2 son de máxima probabilidad

$$\frac{(x_{ip} + x_{iq} + 1)(x_{ip} + x_{tp} + 1)}{x_{ip} + x_{iq} + x_{tp} + x_{tq} + 2} - 1 \leq x_{ip} \leq \frac{(x_{ip} + x_{iq} + 1)(x_{ip} + x_{tp} + 1)}{x_{ip} + x_{iq} + x_{tp} + x_{tq} + 2} \text{ para todo } i \neq t, p \neq q$$

y al simplificar esto equivale a

$$\frac{x_{ip} + 1}{x_{tp}} \geq \frac{x_{iq}}{x_{tq} + 1} \text{ para todo } i \neq t, p \neq q$$

o sea a (3.4).

Teorema 3.2. Sean los subconjuntos $I = \{i_1, i_2, i_3\} \subset \{1, \dots, r\}$ y $J = \{j_1, j_2, j_3\} \subset \{1, \dots, c\}$, y sean $T = \{x_{ij} \ i=1, \dots, r, j=1, \dots, c\}$ y $T' = \{x'_{ij} \ i=1, \dots, r, j=1, \dots, c\}$ dos tablas de la familia \mathfrak{S} tales que

$$x'_{i_u j_v} = x_{i_u j_v} + d_{uv} \quad u=1,2,3, v=1,2,3$$

donde d_{uv} son los elementos de un cuadrado latino 3×3 con los enteros $-1, 0, +1$, y

$$x'_{ij} = x_{ij} \text{ para } i \notin I, j \notin J$$

Entonces para las probabilidades de T y T' siempre se cumple que

$$P(T') > P(T) \Leftrightarrow \prod_{v=1}^3 \frac{x_{a_j v} + 1}{x_{b_j v}} < 1 \quad (3.5)$$

Donde, dado un v , a es el valor de i_u tal que $d_{uv}=1$ y b es el valor de i_u tal que $d_{uv}=-1$.

Además la equivalencia (3.5) también se cumple con el signo $=$.

Demostración.- Al determinar los valores de (3.2) para T y T' el cociente $\frac{P(T')}{P(T)}$

queda reducido con la notación anterior a $\prod_{v=1}^3 \frac{X_{bjv}}{X_{ajv} + 1}$, de donde se deduce de forma

inmediata el teorema.

Corolario 3.3. Sea $T = \{x_{ij} \ i=1, \dots, r, \ j=1, \dots, c\}$ una tabla de la familia \mathfrak{S} . Una condición necesaria para que T sea de máxima probabilidad es que para cualquier par de subconjuntos $I \subset \{1, \dots, r\}$ $\text{Card}(I) = 3$, $J \subset \{1, \dots, c\}$ $\text{Card}(J) = 3$, la subtabla $T^{IJ} = (x_{ij})_{i \in I, j \in J}$ sea tal que al sumarle los elementos de un cuadrado latino 3×3 con los enteros $-1, 0$, y $+1$ se verifique la condición

$$\prod_{j \in J} \frac{X_{aj} + 1}{X_{bj}} \geq 1 \quad (3.6)$$

donde para un valor de j dado ($j \in J$), a es el valor de i ($i \in I$) para el que al valor de la coordenada x_{ij} se le suma el $+1$ del cuadrado latino y, b es el valor de i para el que al valor de la coordenada x_{ij} se le suma el -1 del cuadrado latino.

Demostración.- Si $T = \{x_{ij} \ i=1, \dots, r, \ j=1, \dots, c\}$ es una tabla de máxima probabilidad de la familia \mathfrak{S} entonces la condición (3.6) es cierta pues si fuese falsa existirían subconjuntos $I = \{i_1, i_2, i_3\} \subset \{1, \dots, r\}$ y $J = \{j_1, j_2, j_3\} \subset \{1, \dots, c\}$ para los que la subtabla $T^{IJ} = (x_{ij})_{i \in I, j \in J}$ sea tal que al sumarle los elementos de un cuadrado latino 3×3 con los

enteros $-1, 0, +1$ se verificaría $\prod_{j \in J} \frac{X_{aj} + 1}{X_{bj}} < 1$, donde para un valor de j dado, a es el

valor de i_u ($i_u \in I$) para el que al valor de la coordenada x_{ij} se le suma el $+1$ del cuadrado latino y, b es el valor de i_u para el que al valor de la coordenada x_{ij} se le suma el -1 del cuadrado latino; denotemos por $T'^{IJ} = (t'_{ij})_{i \in I, j \in J}$ la tabla 3×3 resultante

de someter T^{IJ} a la operación anterior . Podríamos construir la tabla $T'=(x'_{ij})_{i=1,\dots,r, j=1,\dots,c}$ en la forma

$$x'_{ij} = t'_{ij} \quad i \in I, j \in J$$

$$x'_{ij} = x_{ij} \quad i \notin I, j \notin J$$

con lo que aplicando el teorema 3.2 a las tablas T y T' tendríamos $P(T') > P(T)$ lo que es absurdo.

Teorema 3.3. Sea $T = \{x_{ij} \quad i=1,2,3, \quad j=1,\dots,c\}$ una tabla de la familia \mathfrak{T}'

$$\mathfrak{T}' = \{X / X \text{ es } 3 \times c, \sum_{j=1}^c x_{ij} = R_i, \sum_{i=1}^3 x_{ij} = C_j\}$$

Si T es tal que todas sus subtablas $2 \times c$ son de máxima probabilidad y, para $I=\{1,2,3\}$ y cualquier subconjunto $J \subset \{1,\dots,c\}$ $\text{Card}(J) = 3$, al sumar a la subtabla $T^{IJ} = (x_{ij})_{i \in I, j \in J}$ los elementos de un cuadrado latino 3×3 con los enteros $-1, 0$, y $+1$, se verifica la condición (3.6) , entonces T es una tabla de máxima probabilidad de la familia \mathfrak{T}' .

Demostración.- Tenemos que probar que $P(T) \geq P(T')$ para toda T' de la familia \mathfrak{T}' .

Sea $T' = \{x'_{ij} \quad i=1,2,3, \quad j=1,\dots,c\}$, como $\sum_{j=1}^c x'_{ij} = \sum_{j=1}^c x_{ij} = R_i, i = 1,2,3$ y

$\sum_{i=1}^3 x'_{ij} = \sum_{i=1}^3 x_{ij} = C_j, j = 1,\dots,c$, los elementos x'_{ij} se pueden expresar en la forma

$$x'_{ij} = x_{ij} + \delta_{ij} \quad i=1,2,3, \quad j=1,\dots,c$$

donde $\sum_{j=1}^c \delta_{ij} = 0, \sum_{i=1}^3 \delta_{ij} = 0$, y los δ_{ij} constituyen una tabla $3 \times c$.

Dejemos aparte el caso en que todos los δ_{ij} son nulos para el que la demostración es trivial, y, para j dado, designemos por

$$I_j^+ = \{i \in \{1,2,3\} / \delta_{ij} > 0\}$$

$$I_j^- = \{i \in \{1,2,3\} / \delta_{ij} < 0\}$$

$$I_j^0 = \{i \in \{1,2,3\} / \delta_{ij}=0\}$$

para estos subconjuntos se verifica

$$\sum_{i \in I_j^+} |\delta_{ij}| = \sum_{i \in I_j^-} |\delta_{ij}|, \text{ llamemos } H_j \text{ a este valor común, y sea } H = \max\{H_j\}_{j=1,\dots,c}$$

y, para i dado designemos por

$$J_i^+ = \{j \in \{1,\dots,c\} / \delta_{ij} > 0\}$$

$$J_i^- = \{j \in \{1,\dots,c\} / \delta_{ij} < 0\}$$

$$J_i^0 = \{j \in \{1,\dots,c\} / \delta_{ij} = 0\}$$

Para estos subconjuntos se verifica

$$\sum_{j \in J_i^+} |\delta_{ij}| = \sum_{j \in J_i^-} |\delta_{ij}|, \text{ llamemos } E_i \text{ a este valor común, y sea } E = \max\{E_i\}_{i=1,2,3}.$$

Siendo $M = \max(H, E)$, la tabla $\delta = (\delta_{ij})_{i=1,2,3, j=1,\dots,c}$ se puede descomponer en suma de M tablas $3 \times c$, $\delta_1, \dots, \delta_M$, donde para cada una de estas tablas sus elementos son nulos excepto en una subtabla 3×3 cuyos elementos son los elementos de un cuadrado latino con los enteros $-1, 0, +1$, ó en una subtabla 2×2 cuyos elementos son los elementos de un cuadrado latino 2×2 con los enteros $-1, +1$, (la descomposición se puede conseguir como se indica a continuación. Cualquier δ_{ij} positivo se puede descomponer obviamente como suma de $|\delta_{ij}|$ elementos iguales a $+1$, y cualquier δ_{ij} negativo como suma de $|\delta_{ij}|$ elementos iguales a -1 . Para conseguir el objetivo buscado hay que asignar las unidades correspondientes a la descomposición de los δ_{ij} a las M tablas de forma adecuada. Comenzamos seleccionando el elemento de mayor valor absoluto, sea δ_{ij} , y se va asignando a $|\delta_{ij}|$ tablas de las M , en la casilla (i, j) una unidad de las resultantes de la descomposición de este elemento ($+1$ si δ_{ij} es positivo, -1 si es negativo); para proseguir con la

formación de las tablas que tienen asignada una unidad en la casilla (i,j) se considera un δ_{ih} y un δ_{tj} , $t \neq i, h \neq j$, de signo opuesto a δ_{ij} , de cada uno de los cuales se separa una unidad que se asigna a las casillas correspondientes, si existe δ_{th} del mismo signo que δ_{ij} con unidades de su descomposición por asignar se asigna una de estas unidades en el lugar correspondiente y al resto de las casillas se las asigna el valor cero, en estos casos la tabla resultante tendría todos sus elementos nulos excepto los de una subtabla 2×2 cuyos elementos son los de un cuadrado latino 2×2 con los enteros $-1, +1$; si no existe δ_{th} del mismo signo que δ_{ij} con unidades de su descomposición por asignar habrá que tomar una unidad de un δ_{tv} del mismo signo que δ_{ij} , $v \neq j, h$ (esto siempre es posible), y una unidad de un δ_{wh} de signo opuesto a δ_{ih} , $w \neq i, t$ (esto siempre es posible), y asignar al resto de las casillas el valor cero, la tabla resultante tendría por construcción todos sus elementos nulos exceptuando los de una subtabla 3×3 cuyos elementos son los de un cuadrado latino 3×3 con los enteros $-1, 0, +1$; se continua hasta que para todas las tablas con el primer elemento asignado, elemento (i,j) , se ha completado el proceso anterior, y lo mismo se hace para los restantes elementos de la misma fila que δ_{ij} y del mismo signo hasta terminar con la descomposición de todos ellos).

Si denotamos ahora por T_h , $h=1, \dots, M$, la tabla resultante de sumar a los elementos de T los elementos de δ_h , T y T_h para todo h son dos tablas de la familia \mathfrak{T}' cuyos elementos son coincidentes, evidentemente, excepto o bien por una subtabla 2×2 o bien por una subtabla 3×3 . En el caso de la subtabla 2×2 , sea $T^{I_h J_h} = \{x_{ij}, i \in I_h, j \in J_h / I_h \subset \{1,2,3\}$ de cardinal 2, $J_h \subset \{1, \dots, c\}$ de cardinal 2} la subtabla de T y $T_h^{I_h J_h}$ la correspondiente subtabla de T_h cuyos elementos son el resultado de sumar a los elementos de $T^{I_h J_h}$ los de un cuadrado latino con los enteros $-1, +1$. En el caso de la

subtabla 3x3, sea $T^{I_h J_h} = \{x_{ij}, i \in I_h, j \in J_h / I_h = 1, 2, 3, J_h \subset \{1, \dots, c\} \text{ de cardinal } 3\}$ la subtabla de T y $T_h^{I_h J_h}$ la correspondiente subtabla de T_h cuyos elementos son el resultado de sumar a los elementos de $T^{I_h J_h}$ los de un cuadrado latino 3x3 con los enteros $-1, 0, +1$. Tanto en la primera situación, por el teorema 3.1 y corolario 3.2 (ya que por hipótesis la tablas T tiene todas sus subtablas 2xc de máxima probabilidad), como en la segunda situación, por el teorema 3.2 y por la segunda hipótesis de este teorema, podemos concluir que $P(T) \geq P(T_h)$, $h=1, \dots, M$. O sea

$$\frac{P(T)}{P(T_h)} \geq 1 \text{ para todo } h, \text{ de donde podemos deducir que } \prod_{h=1}^M \frac{P(T)}{P(T_h)} \geq 1, \text{ pero este}$$

producto puede expresarse en la forma

$$\prod_{h=1}^M \frac{P(T)}{P(T_h)} = \prod_{h=1}^M \prod_{j \in J_h} \frac{x_{aj} + 1}{x_{bj}} \quad (3.7)$$

donde dado un $j \in J_h$, a es el valor de $i, i \in I_h$, para el que el elemento (ij) de T_h es $x_{ij} + 1$, y b es el valor de $i (i \in I_h)$ para el que el elemento (ij) de T_h es $x_{ij} - 1$. Y agrupando los factores que aparecen repetidos como consecuencia de aparecer para diferentes tablas, el producto anterior se puede expresar en la forma

$$\prod_{j=1}^c \frac{\prod_{i \in I_j^+} (x_{ij} + 1)^{|\delta_{ij}|}}{\prod_{i \in I_j^-} x_{ij}^{|\delta_{ij}|}}, \text{ y podemos sacar como consecuencia}$$

$$\prod_{j=1}^c \frac{\prod_{i \in I_j^+} (x_{ij} + 1)^{|\delta_{ij}|}}{\prod_{i \in I_j^-} x_{ij}^{|\delta_{ij}|}} \geq 1 \quad (3.8)$$

Pero, si por otra parte obtenemos los valores de (3.2) para T y T' , y formamos su cociente, este cociente quedará en la forma

$$\frac{P(T)}{P(T')} = \prod_{j=1}^c \frac{\prod_{i \in I_j^+} x'_{ij}! \prod_{i \in I_j^-} x'_{ij}! \prod_{i \in I_j^0} x'_{ij}!}{\prod_{i \in I_j^+} x_{ij}! \prod_{i \in I_j^-} x_{ij}! \prod_{i \in I_j^0} x_{ij}!} = \prod_{j=1}^c \frac{\prod_{i \in I_j^+} (x_{ij} + \delta_{ij})(x_{ij} + \delta_{ij} - 1) \dots (x_{ij} + 1)}{\prod_{i \in I_j^-} x_{ij} \cdot (x_{ij} - 1) \dots (x_{ij} - \delta_{ij} + 1)} \geq$$

$$\prod_{j=1}^c \frac{\prod_{i \in I_j^+} (x_{ij} + 1)^{|\delta_{ij}|}}{\prod_{i \in I_j^-} (x_{ij})^{|\delta_{ij}|}} \quad (3.9)$$

y, de (3.8) y(3.9) podemos concluir

$$\frac{P(T)}{P(T')} \geq 1, \text{ o lo que es equivalente } P(T) \geq P(T') \text{ que era lo que se}$$

queria demostrar.

Corolario 3.4. Es condición necesaria y suficiente para que una tabla perteneciente a una familia isomarginal de tablas $3 \times c$ sea de máxima probabilidad que todas sus subtablas 3×3 sean de máxima probabilidad.

Demostración.- La condición necesaria es consecuencia directa del teorema 1.7 de Harry Joe. Para la condición suficiente, si todas las subtablas 3×3 son de máxima probabilidad por el teorema 1.7 de Harry Joe todas sus subtablas 2×2 serian de maxima probabilidad y como esto es condición necesaria y suficiente (teorema 2.1*) para que todas sus subtablas $2 \times c$ sean también de máxima probabilidad se verificaria la condición primera del teorema anterior, pero además por ser las subtablas 3×3 de máxima probabilidad, según el corolario 3.3 al sumar a cualquiera de esta subtablas los elementos de un cuadrado latino 3×3 con los enteros $-1, 0$ y $+1$ se verificaria la segunda condición exigida en el teorema anterior, y la tabla $3 \times c$ seria de máxima probabilidad.

Las propiedades que acabamos de establecer nos permiten que a partir de una tabla cualquiera perteneciente a una familia de tablas $3 \times c$ con marginales dados

siempre se pueda acceder a la tabla de máxima probabilidad de dicha familia, pues bastaría en primer lugar a partir de la tabla escogida y conforme a lo indicado en el teorema 3.1 ,siempre que se verifique la condición (3.3) ir construyendo en la forma indicada nuevas tablas de la familia con una probabilidad mayor hasta obtener una tabla para la que se verifique la condición (3.4), con lo que habríamos conseguido una tabla con todas sus subtablas $2 \times c$ de máxima probabilidad, y a partir de aquí y conforme a lo indicado en el teorema 3.2, siempre que se verifique la condición (3.5), tratar de construir en la forma indicada nuevas tablas de probabilidad mayor hasta obtener una tabla para la que se verifique la condición (3.6) que sería ya según el teorema 3.3 de máxima probabilidad. Evidentemente este proceso sería más eficaz en cuanto que la tabla de partida necesite del menor número posible de construcciones intermedias hasta acceder a una tabla para la que se verifique la condición (3.4) y, así mismo necesite a partir de aquí del menor número de construcciones intermedias hasta acceder a una tabla para la que se verifique la condición (3.6), o sea la tabla de partida tenga una probabilidad cercana a la máxima.

En el caso de tablas $3 \times c$ si queremos determinar la tabla de máxima probabilidad de la familia \mathfrak{S}' un buen punto de partida sería una tabla T' construida de la siguiente forma

$T' = \{x_{ij} \ i=1,2,3 \ j=1,\dots,c \ / \text{ para un valor } i_1 \text{ de } i \ (x_{i_1 1}, \dots, x_{i_1 c}) \text{ es máximo de la } MH(R_{i_1}; C_1, \dots, C_c), \text{ y para los restantes valores } i_2, i_3 \text{ de } i, \text{ la subtabla } (x_{ij})_{i=i_2, i_3, j=1, \dots, c} \text{ sea de máxima probabilidad para la familia de tablas } 2 \times c \text{ con marginales filas } R'_1 = R_{i_2}, R'_2 = R_{i_3}, \text{ y marginales columnas } C'_1 = C_1 - x_{i_1 1}, \dots, C'_c = C_c - x_{i_1 c} \}$

Ejemplo 3.1. Determinar la tabla de máxima probabilidad de la familia 3×4 de marginales filas $R_1=3, R_2=3, R_3=4,$ y marginales columnas $C_1=1, C_2=1, C_3=1, C_4=7.$

Determinamos el máximo de la MH(3;1,1,1,7), resultando

$$(x_{11}, x_{12}, x_{13}, x_{14}) = (0, 0, 0, 3)$$

Determinamos la tabla de máxima probabilidad para la familia de tablas 2x4 de marginales filas $R'_1 = R_2 = 3$, $R'_2 = R_3 = 4$, y marginales columnas $C'_1 = C_1 - x_{11} = 1$, $C'_2 = C_2 - x_{12} = 1$, $C'_3 = C_3 - x_{13} = 1$, $C'_4 = C_4 - x_{14} = 4$, que es

$$\begin{bmatrix} 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 \end{bmatrix}$$

Establecemos como tabla T' de partida

$$T' = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 2 \end{bmatrix}$$

y, en este caso particular, para esta tabla se verifica tanto la condición (3.4) como la condición (3.6), con lo que no precisa de construcciones intermedias pues ya es de máxima probabilidad.

A pesar de la validez para tablas 3xc del algoritmo que acabamos de exponer trataremos de elaborar a continuación otro algoritmo para estas tablas que aprovechando igualmente los resultados de los teoremas anteriores nos permita que el algoritmo desarrollado en el capítulo anterior para tablas 2xc se vea como un caso particular y al mismo tiempo él se pueda ver como un caso particular para tablas rxc

Definición 3.1. Sean

$$\mathfrak{S} = \left\{ Y / Y \text{ es } r \times c, \sum_{j=1}^c y_{ij} = R_i, \sum_{i=1}^r y_{ij} = C_j \right\}$$

$$\mathfrak{R} = \left\{ Y' / Y' \text{ es } r \times c, \sum_{j=1}^c y'_{ij} = R'_i, \sum_{i=1}^r y'_{ij} = C'_j \right\}$$

decimos que \mathfrak{S} y \mathfrak{R} son familias de marginales consecutivos respecto a los marginales columnas s y t , si

$$R'_i = R_i \quad \text{para todo } i=1,\dots,r$$

$$C'_j = C_j \quad \text{para todo } j=1,\dots,c, j \neq s,t$$

$$C'_s = C_s + 1$$

$$C'_t = C_t - 1$$

Igualmente definiríamos dos familias de marginales consecutivos respecto a los marginales filas s y t .

Se puede establecer una relación entre los máximos de dos familias $3 \times c$ de marginales consecutivos .

Sean \mathfrak{S} y \mathfrak{R} dos familias de marginales consecutivos respecto a los marginales columnas s y t , o sea ambas con marginales filas R_1, R_2, R_3 y marginales columnas respectivos

$$C_1, \dots, C_s, \dots, C_t, \dots, C_c \quad \text{y} \quad C_1, \dots, C_s + 1, \dots, C_t - 1, \dots, C_c$$

Si $T = (x_{ij})_{i=1,2,3, j=1,\dots,c}$ es una tabla de máxima probabilidad de la familia \mathfrak{S} , entonces se puede acceder a una tabla de máxima probabilidad de la familia \mathfrak{R} de la siguiente forma; consideramos la subtabla A , 3×2 de T de marginales columnas C_s y C_t

$$A = (x_{ij})_{i=1,2,3 \quad j=s,t}$$

esta subtabla A es de máxima probabilidad y por el procedimiento desarrollado en el capítulo anterior podemos determinar una subtabla A' , 3×2 , de marginales columnas $C_s + 1$, $C_t - 1$, conectada a la anterior a través de una coordenada h , $1 \leq h \leq 3$. Si sustituimos en T la subtabla A por la subtabla A' la tabla resultante $T' = (x'_{ij})$ queda definida en la forma

$$x'_{hs} = x_{hs} + 1$$

$$x'_{ht} = x_{ht} - 1$$

$$x'_{ij} = x_{ij} \text{ en otro caso}$$

y, para esta tabla T' se verifica de forma inmediata:

- Es una tabla de la familia \mathfrak{R} .
- T' puede ser de máxima probabilidad pero si no lo es está cercana a la

tabla de máxima probabilidad pues todas las subtablas que se generan con elementos que solo pertenecen a A' ,o con elementos de T que no han sido sustituidos, son de máxima probabilidad y T' es un buen punto de partida para obtener la tabla de máxima probabilidad; La tabla T' podría aumentar su probabilidad sustituyendo solo sus subtablas 2x2 que fuesen de la forma

$$\begin{bmatrix} x'_{is} & x'_{ij} \\ x'_{hs} & x'_{hj} \end{bmatrix} \quad \text{ó} \quad \begin{bmatrix} x'_{it} & x'_{ij} \\ x'_{ht} & x'_{hj} \end{bmatrix} \quad j=1,\dots,c, \quad j \neq s,t, \quad i=1,2,3, \quad i \neq h$$

por subtablas 2x2 respectivamente de la forma

$$\begin{bmatrix} x'_{is} + 1 & x'_{ij} - 1 \\ x'_{hs} - 1 & x'_{hj} + 1 \end{bmatrix} \quad \text{ó} \quad \begin{bmatrix} x'_{it} - 1 & x'_{ij} + 1 \\ x'_{ht} + 1 & x'_{hj} - 1 \end{bmatrix}$$

y, siempre que se verificase la condición (3.3), o sea

$$\frac{x'_{hs}}{x'_{hj} + 1} > \frac{x'_{is} + 1}{x'_{ij}} \quad \text{ó} \quad \frac{x'_{ht} + 1}{x'_{hj}} < \frac{x'_{it}}{x'_{ij} + 1}$$

Conseguiríamos de esta forma una tabla perteneciente a la familia \mathfrak{R} con todas sus subtablas 2xc de máxima probabilidad , y a partir de aquí, y utilizando el teorema 3.2, iríamos realizando las sustituciones adecuadas hasta conseguir una tabla para la que se verifique la condición (3.6) para cualquier cuadrado latino 3x3 con los elementos -1, 0, +1. Esta tabla ya seria de maxima probabilidad dentro de la familia \mathfrak{R} .

(Ninguna otra sustitución de tablas 2x2 diferente a las que hemos señalado anteriormente valdria para aumentar la probabilidad de T'. Así, no valdria la sustitución de las tablas 2x2 consideradas por, respectivamente,

$$\begin{bmatrix} x'_{is} - 1 & x'_{ij} + 1 \\ x'_{hs} + 1 & x'_{hj} - 1 \end{bmatrix} \quad \text{ó} \quad \begin{bmatrix} x'_{it} + 1 & x'_{ij} - 1 \\ x'_{ht} - 1 & x'_{hj} + 1 \end{bmatrix}$$

pues no se cumpliría la condición (3.3) ya que al ser T una tabla de máxima probabilidad de la familia \mathfrak{S} se verificará

$$x_{is} \cdot x_{hj} \leq (x_{ij} + 1)(x_{hs} + 1)$$

y

$$x_{ht} \cdot x_{ij} \leq (x_{it} + 1)(x_{hj} + 1)$$

de donde se sigue respectivamente, al ser $x_{hs} + 1 = x'_{hs}$, $x_{ht} - 1 = x'_{ht}$, $x_{hj} = x'_{hj}$, $x_{ij} = x'_{ij}$,

$$x_{it} = x'_{it}$$

$$x'_{is} \cdot x'_{hj} \leq (x'_{ij} + 1) \cdot x'_{hs}$$

y

$$(x'_{ht} + 1) \cdot x'_{ij} \leq (x'_{it} + 1)(x'_{hj} + 1)$$

y, por consiguiente

$$x'_{is} \cdot x'_{hj} < (x'_{ij} + 1)(x'_{hs} + 1) \Leftrightarrow \frac{x'_{is}}{x'_{hs} + 1} < \frac{x'_{ij} + 1}{x'_{hj}}$$

y

$$x'_{ht} \cdot x'_{ij} < (x'_{it} + 1)(x'_{hj} + 1) \Leftrightarrow \frac{x'_{ht}}{x'_{hj} + 1} < \frac{x'_{it} + 1}{x'_{ij}} \quad)$$

III.2.3: Algoritmo para determinar el máximo de tablas 3xc (generalizable a tablas rxc).

En el caso 2xc el método que hemos propuesto para determinar la tabla de máxima probabilidad está basado en las modas de las distribuciones hipergeométricas marginales, y en el hecho de que estas modas siempre determinan una tabla de máxima probabilidad para la familia isomarginal de marginales coincidentes con los suyos, en cuyo caso si la tabla generada de esta forma pertenece a la familia cuyo máximo se pretende determinar, ella proporciona este máximo, y en otro caso, este hecho puede ser aprovechado para determinar la tabla de máxima probabilidad de una familia de marginales consecutivos respecto a las filas 1 y 2, y así, y en un número determinado de pasos, acceder al máximo de la familia que nos interesa. Trataremos a continuación de extender esta metodología en la medida de lo posible al caso 3xc.

El Algoritmo que proponemos para determinar el máximo para tablas 3xc está basado en las modas de las distribuciones hipergeométricas multivariantes correspondientes a las filas de la tabla y en los resultados obtenidos en los teoremas desarrollados en la sección anterior.

Sea

$$D = \{ \mathbf{z} = (z_{ij})_{i=1, \dots, r, j=1, \dots, c} / \mathbf{z}^i = (z_{i1}, \dots, z_{ic}) \text{ es} \\ \text{moda de MH}(R_i; C_1, \dots, C_c) \}_{i=1, \dots, r} \} \quad (3.10)$$

Este conjunto tendrá un solo punto cuando la moda de cada marginal fila \mathbf{z}^i sea única, algo que evidentemente no tiene por qué suceder.

Por otro lado, dado $\mathbf{z} \in D$ notaremos por $\boldsymbol{\delta} = (\delta_1, \dots, \delta_c)$ el vector c -dimensional cuyas componentes quedan definidas en la forma

$$\delta_j = \sum_{i=1}^r z_{ij} - C_j \quad j=1, \dots, c$$

Por la propia construcción realizada, para todo \mathbf{z} perteneciente a D se verifica

$$\sum_{j=1}^c z_{ij} = R_i, \quad (3.11)$$

$$\sum_{i=1}^r z_{ij} = C_j + \delta_j, \quad \sum_{j=1}^c \delta_j = 0 \quad (3.12)$$

por otro lado, para un valor particular de $\boldsymbol{\delta}$ en la forma anteriormente introducida, definimos el subconjunto

$$D_\delta = \{\mathbf{z} \in D / \sum_{i=1}^r z_{ij} - C_j = \delta_j \quad j=1, \dots, c\} \quad (3.13)$$

Y, consideramos la familia \mathfrak{F}_δ

$$\mathfrak{F}_\delta = \{Y / Y \text{ es } r \times c, \sum_{j=1}^c y_{ij} = R_i, \sum_{i=1}^r y_{ij} = C_j + \delta_j\} \quad (3.14)$$

Teorema 3.4. Dada la familia \mathfrak{F}_δ se verifica que $\mathbf{z} \in D_\delta$ es una tabla de dicha familia con todas sus subtablas $2 \times c$ de máxima probabilidad.

Demostración.- Por hipótesis $\mathbf{z}^i = (z_{i1}, \dots, z_{ic})$ es moda de la

$$MH(R_i; C_1, \dots, C_c) \quad \text{para todo } i=1, \dots, r$$

Por consiguiente, siendo u y v dos valores diferentes de i , \mathbf{z}^u es moda de $MH(R_u; C_1, \dots, C_c)$ y \mathbf{z}^v es moda de $MH(R_v; C_1, \dots, C_c)$, lo que nos permite establecer la siguientes desigualdades

$$z_{uk} \leq \frac{(z_{ut} + 1)(C_k + 1)}{C_t + 1} \quad k, t = 1, \dots, c, \quad k \neq t \quad (3.15)$$

$$z_{uk} \geq \frac{z_{ut}(C_k + 1)}{C_t + 1} - 1 \quad k, t = 1, \dots, c, \quad k \neq t \quad (3.16)$$

$$z_{vk} \leq \frac{(z_{vt} + 1)(C_k + 1)}{C_t + 1} \quad k, t = 1, \dots, c, \quad k \neq t \quad (3.17)$$

$$z_{vk} \geq \frac{z_{vt}(C_k + 1)}{C_t + 1} - 1 \quad k, t = 1, \dots, c, \quad k \neq t \quad (3.18)$$

de (3.16) y de (3.17) se deduce

$$(z_{uk} + 1) \geq \frac{z_{ut} \cdot z_{vk}}{(z_{vt} + 1)} \quad u \neq v, \quad k \neq t$$

con lo que queda verificada la condición (3.4) que era lo que se quería demostrar.

El resultado que hemos obtenido es válido para una tabla $r \times c$, pero de las propiedades que hemos desarrollado para tablas $3 \times c$ pueden obtenerse otros resultados para el caso de tablas $3 \times c$. En adelante particularizaremos para tablas $3 \times c$.

Teorema 3.5. Dada la familia \mathfrak{S}_δ se verifica que $\mathbf{z} \in D_\delta$ es una tabla de máxima probabilidad de dicha familia.

Demostración.- Como por el teorema anterior tenemos que todas las subtablas $2 \times c$ de \mathbf{z} son de máxima probabilidad, según el teorema 3.3 sería suficiente con probar que para cualquier subconjunto $J \subset \{1, \dots, c\}$ $\text{Card}(J) = 3$, y para $I = \{1, 2, 3\}$, al sumar a la subtabla $\mathbf{z}^{IJ} = (z_{ij})_{i \in I, j \in J}$ un cuadrado latino 3×3 con los elementos $-1, 0, +1$ se verifica

$$\prod_{j \in J} \frac{z_{aj} + 1}{z_{bj}} \geq 1$$

donde para un valor de j dado, a es el valor de i para el que la coordenada de la columna j aumenta en una unidad, y , b es el valor de i para el que la coordenada de la columna j disminuye en una unidad.

Si para un valor k perteneciente a J , a es el valor de i para el que la coordenada correspondiente a esa columna aumenta, forzosamente para el mismo valor de i anterior tiene que haber una coordenada de esa fila que disminuya y esta evidentemente no puede ser otra que $z_{bk'}$ donde k' es un valor perteneciente a J diferente a k .

z_{ak} y $z_{bk'}$ denotan pues a elementos de una misma fila, y al ser \mathbf{z}^i máximo de $MH(R_i; C_1, \dots, C_c)$ para todo $i=1,2,3$

$$z_{ik} + 1 \geq z_{ik} \cdot \frac{(C_k + 1)}{(C^{k'} + 1)}$$

por consiguiente

$$z_{ak} + 1 \geq z_{bk'} \cdot \frac{(C_k + 1)}{C^{k'} + 1}$$

pero ya que k y k' son una pareja de valores diferentes de j , estableciendo la desigualdad anterior para todas las parejas de valores posibles diferentes que se puedan formar con los elementos del conjunto J , y formando su producto, obtenemos

$$\prod_{j \in J} \frac{z_{aj} + 1}{z_{bj}} \geq \frac{\prod_{k \in J} (C_k + 1)}{\prod_{k' \in J} (C^{k'} + 1)} = 1$$

de donde se deduce directamente el resultado buscado.

Corolario 3.5. Sea

$D = \{ \mathbf{z} = (z_{ij})_{i=1,2,3, j=1, \dots, c} / \mathbf{z}^i = (z_{i1}, \dots, z_{ic}) \text{ es moda de } MH(R_i; C_1, \dots, C_c) \}$
 $i=1,2,3 \}$

Si existe $\mathbf{z} \in D$ para el que se verifique

$$\sum_{i=1}^3 z_{ij} - C_j = \delta_j = 0 \text{ para todo } j=1, \dots, c$$

entonces \mathbf{z} es una moda de la distribución hipergeométrica multivariante múltiple

$$\text{MMH}(R_1, R_2, R_3; C_1, \dots, C_c)$$

Demostración.- Si existe $\mathbf{z} \in D$ para el que se verifique

$$\sum_{i=1}^3 z_{ij} - C_j = \delta_j = 0 \text{ para todo } j=1, \dots, c$$

entonces $\mathbf{z} \in D_0$, y por el teorema 3.5, \mathbf{z} es una tabla de máxima probabilidad de la familia \mathfrak{T}_0 , pero \mathfrak{T}_0 coincide con \mathfrak{T} , lo que equivale a que \mathbf{z} es una moda de la distribución hipergeométrica multivariante múltiple $\text{MMH}(R_1, R_2, R_3; C_1, \dots, C_c)$.

En caso de no cumplirse la hipótesis de este corolario, un $\mathbf{z} \in D_\delta$ nos puede servir como punto de partida para determinar el máximo de $\text{MMH}(R_1, R_2, R_3; C_1, \dots, C_c)$ o lo que es equivalente, la tabla de máxima probabilidad de la familia \mathfrak{T}'

$$\mathfrak{T}' = \{Y / Y \text{ es } 3 \times c, \sum_{j=1}^c y_{ij} = R_i, \sum_{i=1}^3 y_{ij} = C_j\}$$

pues al ser, para un δ dado, $\delta = (\delta_1, \dots, \delta_c)$ $\sum_{j=1}^c \delta_j = 0$, \mathbf{z} máximo de la familia

$$\mathfrak{T}_\delta = \{Y / Y \text{ es } 3 \times c, \sum_{j=1}^c y_{ij} = R_i, \sum_{i=1}^3 y_{ij} = C_j + \delta_j\}$$

se trata de ir realizando sustituciones de la tabla definida por \mathbf{z} , por tablas también de máxima probabilidad que teniendo los mismos marginales filas, para sus marginales columnas se vayan haciendo nulas las diferencias existentes con los marginales columnas de la familia \mathfrak{T}' . Para ello aplicaremos de forma repetida el procedimiento desarrollado para pasar de la tabla de máxima probabilidad de una familia a la tabla

de máxima probabilidad de una familia de marginales consecutivos. Consideramos una subtabla 3x2 cualesquiera de \mathbf{z} con sus δ_j de distinto signo, $\delta_s < 0$, $\delta_t > 0$.

A partir, de $C_s + \delta_s$ y $C_t + \delta_t$ se pueden establecer dos sucesiones de valores consecutivos, una de ellas creciente y otra decreciente

$$C_s + \delta_s, C_s + \delta_s + 1, \dots$$

$$C_t + \delta_t, C_t + \delta_t - 1, \dots$$

Lo que nos determina la sucesión de familias consecutivas $\mathfrak{T}_\delta^0, \mathfrak{T}_\delta^1, \mathfrak{T}_\delta^2, \dots$, donde $\mathfrak{T}_\delta^0 = \mathfrak{T}_\delta$ y \mathfrak{T}_δ^k y $\mathfrak{T}_\delta^{k-1}$ son familias de marginales consecutivos respecto a los marginales columnas s y t , siendo $C_s + \delta_s + k$ y $C_t + \delta_t - k$ los marginales columnas s y t , respectivamente, para la familia \mathfrak{T}_δ^k . Y el máximo para cada una de estas familias puede determinarse a partir del máximo de la familia anterior en la forma ya indicada.

Si $|\delta_s| = |\delta_t| = h$, en h aplicaciones del método indicado para pasar del máximo de una familia al máximo de una familia consecutiva, habríamos pasado al máximo de una familia con los mismos marginales que los de \mathfrak{T}_δ excepto en las columnas s y t que ya son iguales a C_s y C_t .

Si $|\delta_s| \neq |\delta_t|$ y $\min\{|\delta_s|, |\delta_t|\} = |\delta_s| = h'$, en h' aplicaciones del método habríamos pasado al máximo de una familia con los mismos marginales que los de \mathfrak{T}_δ excepto en las columnas s y t , pues ya el marginal columna s es C_s y el marginal columna t se ha aproximado en h' unidades al valor C_t .

Si $|\delta_s| \neq |\delta_t|$ y $\min\{|\delta_s|, |\delta_t|\} = |\delta_t| = h''$, también en h'' aplicaciones del método indicado habríamos pasado al máximo de una familia con los mismos marginales que

los de \mathfrak{S}_δ excepto en las columnas s y t , pues ya el marginal de la columna t es C_t y el marginal columna s se ha aproximado en h'' unidades al valor C_s .

El proceso de sustitución se repetiría hasta lograr el máximo de una familia con todos sus marginales columnas coincidentes con C_1, \dots, C_c , o sea hasta conseguir el máximo de la familia \mathfrak{S}' .

Evidentemente, los conceptos de familias de marginales columnas consecutivos y los procesos desarrollados para acceder del máximo de una familia al máximo de una familia de marginales columnas consecutivos, así como el algoritmo construido a partir de los máximos de los marginales filas son directamente extrapolables a familias de marginales filas consecutivos, a procesos para acceder al máximo de una familia a través del máximo de una familia de marginales filas consecutivos, y a un algoritmo basado en los máximos de los marginales columnas.

Ejemplo 3.2. Determinar la tabla de máxima probabilidad para la familia de tablas 3×3 con marginales filas $R_1=12$, $R_2=5$, $R_3=2$, y marginales columnas $C_1=8$, $C_2=7$, $C_3=4$.

- Determinamos:

máximo de la MH(12; 8,7,4)

$$(x_{11}, x_{12}, x_{13}) = (5, 4, 3)$$

máximo de la MH(5; 8,7,4)

$$(x_{21}, x_{22}, x_{23}) = (2, 2, 1)$$

máximo de la MH(2; 8,7,4)

$$(x_{31}, x_{32}, x_{33}) = (1, 1, 0)$$

construimos la tabla 3x3 cuyas filas se corresponden respectivamente con los máximos anteriores

$$T = \begin{bmatrix} 5 & 4 & 3 \\ 2 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Esta tabla, por los teoremas 3.4 y 3.5, sin necesidad de mas comprobaciones, es la tabla de máxima probabilidad para la familia isomarginal a la que pertenece, que es precisamente la familia cuya tabla de máxima probabilidad pretendíamos determinar.

Ejemplo 3.3. Determinar la tabla de máxima probabilidad para la familia de tablas 3x3 , de marginales filas $R_1= 10, R_2=3, R_3= 2$, y marginales columnas , $C_1= 8, C_2=4, C_3=3$.

Determinamos :

Máximo de la MH(10; 8, 4,3)

$$(x_{11}, x_{12}, x_{13}) = (5, 3, 2)$$

Máximo de la MH(3; 8, 4, 3)

$$(x_{21}, x_{22}, x_{23}) = (2, 1, 0)$$

Máximo de la MH(2; 8, 4, 3)

$$(x_{31}, x_{32}, x_{33}) = (1, 1, 0)$$

construimos la tabla 3x3 cuyas filas se corresponden con los máximos anteriores

$$T = \begin{bmatrix} 5 & 3 & 2 \\ 2 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Esta tabla , por los teoremas 3.4 y 3.5 es de máxima probabilidad para la familia de tablas 3x3, con marginales filas

$$R'_1 = R_1 = 10, R'_2 = R_2 = 3, R'_3 = R_3 = 2$$

Y marginales columnas

$$C'_1 = 8 = C_1, C'_2 = 5 = C_2 + 1, C'_3 = 2 = C_3 - 1$$

Es de máxima probabilidad para una familia de marginales consecutivos con la familia dada a través de las columnas segunda y tercera; Aplicamos a continuación el procedimiento para pasar de la tabla de máxima probabilidad de una familia a la tabla de máxima probabilidad de una familia consecutiva a ella.

Sustituimos la subtabla, A, 3x2 de T, correspondientes a los marginales columnas C'_2 y C'_3

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Por la subtabla 3x2, con los mismos marginales filas que la anterior y marginales columnas C_2, C_3 , y conectada a ella

$$A' = \begin{bmatrix} 3 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

La tabla resultante de la sustitución

$$T' = \begin{bmatrix} 5 & 3 & 2 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Pertenece a la familia dada

Vemos si puede aumentar su probabilidad por sustituciones de sus subtablas 2x2, en la forma indicada al final de la sección III.2.2 y según el teorema 3.1; no puede hacerlo, sus elementos verifican la condición (3.4). Por consiguiente todas sus subtablas 2x3 son de máxima probabilidad.

Vemos si puede aumentar su probabilidad por sustituciones por otra tabla 3x3 , en la forma indicada en el teorema 3.2; no puede hacerlo , sus elementos verifican la condición (3.6). Por consiguiente, la tabla T' es la tabla de máxima probabilidad para la familia dada.

Ejemplo 3.4. Determinar la tabla de máxima probabilidad para la familia de tablas 3x5 de marginales filas $R_1=14, R_2=12, R_3=4$, y $C_1=C_2=C_3=C_4=3, C_5=18$.

Las modas de las Distribuciones hipergeométricas multivariantes correspondientes a las filas de las tablas de la familia dada nos determinan la tabla

$$T = \begin{bmatrix} 2 & 1 & 1 & 1 & 9 \\ 1 & 1 & 1 & 1 & 8 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Cuyos marginales columnas $C'_1=3, C'_2=C'_3=C'_4=2, C'_5=21$, nos proporcionan el vector $\delta = (0, -1, -1, -1, +3)$, tendríamos que pasar ahora por diferentes familias isomarginales de marginales consecutivos

Familia con los C'_j anteriores $\xrightarrow{\text{tomando } \delta_2 \text{ y } \delta_5}$
 \mathfrak{T}_δ^0

Familia de marginales consecutivos $\xrightarrow{\text{tomando } \delta_3 \text{ y } \delta_5}$
 \mathfrak{T}_δ^1

Familia de marginales consecutivos $\xrightarrow{\text{tomando } \delta_4 \text{ y } \delta_5}$
 \mathfrak{T}_δ^2

Familia de marginales consecutivos \mathfrak{T}_δ^3
 coincidente con la familia isomarginal original

- Obtengamos un máximo de \mathfrak{T}_δ^1 .

Sustituimos la subtabla A_1 , 3×2 de T correspondiente a los marginales columnas C'_2 y C'_5

$$A_1 = \begin{bmatrix} 1 & 9 \\ 1 & 8 \\ 0 & 4 \end{bmatrix}$$

Por la subtabla 3×2 con los mismos marginales filas que la anterior, y marginales columnas C_2 y $C''_5 = C'_5 - 1 = C_5 + 2$, conectada a ella,

$$B_1 = \begin{bmatrix} 2 & 8 \\ 1 & 8 \\ 0 & 4 \end{bmatrix}$$

Para la tabla resultante de la sustitución

$$T_1 = \begin{bmatrix} 2 & 2 & 1 & 1 & 8 \\ 1 & 1 & 1 & 1 & 8 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Vemos si puede aumentar su probabilidad por sustituciones de sus subtablas 2×2 , en la forma indicada al final de la sección III.2.2 (probando solo las subtablas que hemos indicado como posibles para tal fin); No puede hacerlo, por consiguiente todas sus subtablas 2×5 son de máxima probabilidad.

Vemos si puede aumentar su probabilidad por sustituciones de sus subtablas 3×3 , en la forma indicada en el teorema 3.2; No puede hacerlo, por consiguiente T_1 es una tabla de máxima probabilidad para la familia de marginales filas R_1, R_2, R_3 , y marginales columnas $C''_1 = C_1 = 3, C''_2 = C_2 = 3, C''_3 = C_4 = 2, C''_5 = 20$.

- Obtengamos un máximo de \mathfrak{Z}_δ^2 .

Sustituimos la subtabla A_2 , 3×2 de T_1 correspondiente a los marginales columnas C''_3 y C''_5

$$A_2 = \begin{bmatrix} 1 & 8 \\ 1 & 8 \\ 0 & 4 \end{bmatrix}$$

Por la subtabla 3x2 con los mismos marginales filas que la anterior y marginales columnas C_3 y $C_5''' = C_5'' - 1 = C_5 + 1$, conectada a ella

$$B_2 = \begin{bmatrix} 2 & 7 \\ 1 & 8 \\ 0 & 4 \end{bmatrix}$$

Para la tabla resultante de la sustitución

$$T_2 = \begin{bmatrix} 2 & 2 & 2 & 1 & 7 \\ 1 & 1 & 1 & 1 & 8 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Vemos de la misma forma indicada para T_1 si puede aumentar su probabilidad; No puede hacerlo, por consiguiente T_2 es una tabla de máxima probabilidad para la familia \mathfrak{S}_δ^2 de marginales filas R_1, R_2, R_3 , y marginales columnas $C_1''' = C_1 = 3$, $C_2''' = C_2 = 3$, $C_3''' = C_3 = 3$, $C_4''' = C_4 - 1 = 2$, $C_5''' = C_5 + 1 = 19$.

- Obtengamos un máximo de $\mathfrak{S}_\delta^3 = \mathfrak{S}'$, o sea de la familia isomarginal original.

Sustituimos la subtabla A_3 , 3x2 de T_2 correspondiente a los marginales columnas C_4''' y C_5'''

$$A_3 = \begin{bmatrix} 1 & 7 \\ 1 & 8 \\ 0 & 4 \end{bmatrix}$$

Por la subtabla 3x2 con los mismos marginales filas que la anterior y marginales columnas C_4 y C_5

$$B_3 = \begin{bmatrix} 1 & 7 \\ 2 & 7 \\ 0 & 4 \end{bmatrix}$$

Para la tabla resultante de la sustitución

$$T_3 = \begin{bmatrix} 2 & 2 & 2 & 1 & 7 \\ 1 & 1 & 1 & 2 & 7 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Que ya pertenece a la familia original, vemos de la misma forma indicada para T_1 , y T_2 si puede aumentar su probabilidad; No puede hacerlo, por consiguiente T_3 es una tabla de máxima probabilidad de la familia original.

III.3: Algoritmo de Enumeración Incompleta para la determinación del p-valor de una tabla 3xc. Mejoras en el Algoritmo Reticulado de Mehta y Patel.

III.3.1: Introducción.

El Algoritmo reticulado de Mehta y Patel para la determinación del p-valor de una tabla rx_c , como ya ha sido señalado, necesita de la determinación de las trayectorias máximas correspondientes a los nodos de un estado, y caso de que el nodo no sea “excluido” del resto del proceso, necesita también, de la determinación de las trayectorias máximas correspondientes a los nodos hijos del estado siguiente. Como las trayectorias correspondientes a un nodo presentan un recorrido común hasta él, la trayectoria máxima queda determinada por la subtrayectoria máxima existente a partir del nodo.

En el caso $2xc$, aparte de desarrollar un método para la determinación de la subtrayectoria máxima a partir de un nodo, hemos relacionado la subtrayectoria máxima existente a partir de un nodo con la subtrayectoria máxima de un nodo hijo

determinado, y a su vez a través de esta subtrayectoria hemos determinado las subtrayectorias máximas correspondientes a los restantes nodos hijos, aportaciones que simplifican considerablemente el algoritmo; Trataremos a continuación de extender estas mejoras al caso de tablas 3xc. Hemos desarrollado en la sección anterior un método que nos permitirá determinar la subtrayectoria máxima correspondiente a un nodo, pero al igual que en el caso 2xc también aquí, se pueden tratar de encontrar relaciones que simplifiquen el proceso posterior de determinación de las subtrayectorias máximas para sus nodos hijos.

III.3.2: MMHD asociadas a los nodos del retículo de Mehta y Patel para tablas 3xc.

Todo nodo del retículo de M-P para tablas 3xc, tiene asociada en el siguiente sentido una MMHD.

Si $(k, R_{1k}, R_{2k}, R_{3k})$ es un nodo del estado k , con R_{ik} , $i=1,2,3$ determinados, las subtrayectorias posibles desde este nodo al nodo terminal vienen dadas por el conjunto de puntos de la

$$MMH(R_{1k}, R_{2k}, R_{3k}; C_1, \dots, C_k)$$

o, lo que es equivalente por las tablas de la familia

$$\mathfrak{S}_k = \{ Y / Y \text{ es } 3 \times k, \sum_{j=1}^k y_{ij} = R_{ik}, \sum_{i=1}^3 y_{ij} = C_j \}$$

Los nodos hijos del estado siguiente, se denotan en la forma

$(k-1, R_{1,k-1}, R_{2,k-1}, R_{3,k-1})$, y el rango de $R_{i,k-1}$, $i=1,2,3$ para estos nodos hijos está dado por

$$\max (0, R_{ik} - C_k + \sum_{l=1}^{i-1} (R_{lk} - R_{l, k-1})) \leq R_{i,k-1} \leq \min (R_{ik}, S_{k-1} - \sum_{l=1}^{i-1} R_{l, k-1})$$

donde $S_j = \sum_{l=1}^j C_l$, y se conviene que un sumatorio es nulo si su límite inferior excede

a su límite superior.

Las subtrayectorias posibles desde un nodo hijo $(k-1, R_{1,k-1}, R_{2,k-1}, R_{3,k-1})$ al nodo terminal vienen dadas por el conjunto de puntos de la

$$\text{MMH}(R_{1,k-1}, R_{2,k-1}, R_{3,k-1}; C_1, \dots, C_{k-1})$$

o, lo que es equivalente, por las tablas de la familia

$$\mathfrak{T}_{k-1} = \{Y / Y \text{ es } 3 \times (k-1), \sum_{j=1}^{k-1} y_{ij} = R_{i,k-1}, \sum_{i=1}^3 y_{ij} = C_j\}$$

Si denotamos por $\dot{x} = (\dot{x}_{ij})_{i=1,2,3, j=1, \dots, k}$ la tabla de máxima probabilidad de \mathfrak{T}_k , la subtabla $(\dot{x}_{ij})_{i=1,2,3, j=1, \dots, k-1}$ por el teorema 1.7 de H-J maximiza una de las familias \mathfrak{T}_{k-1} , aquella para la que sus marginales filas sean precisamente

$$R_{i,k-1} = \sum_{j=1}^{k-1} \dot{x}_{ij} \quad i=1,2,3$$

Y , al nodo hijo cuya familia asociada sea la anterior, le llamaremos nodo hijo principal.

El disponer de la tabla de máxima probabilidad para una de las familias \mathfrak{T}_{k-1} nos posibilita el acceder a las tablas de máxima probabilidad para las restantes familias \mathfrak{T}_{k-1} , pues partiendo de esa familia y a través de familias con dos marginales filas consecutivos se recorrerían todas las familias \mathfrak{T}_{k-1} , y se podrían determinar por la metodología desarrollada para tal fin sus máximos respectivos. La posibilidad señalada se justifica fácilmente,

Sean $R_{1,k-1}^*, R_{2,k-1}^*, R_{3,k-1}^*$, los valores concretos de $R_{1,k-1}, R_{2,k-1}, R_{3,k-1}$, correspondientes a los marginales filas de la familia asociada al nodo hijo principal. A todas las

familias con el marginal de la primera fila común con el dado se accede construyendo a partir de $R_{2,k-1}^*$ las dos sucesiones siguientes

$$R_{2,k-1}^*, R_{2,k-1}^* + 1, \dots (\text{valor máximo de } R_{2,k-1})$$

$$R_{2,k-1}^*, R_{2,k-1}^* - 1, \dots (\text{valor mínimo de } R_{2,k-1})$$

lo que nos determina una sucesión de familias de marginales consecutivos respecto a las filas segunda y tercera, con inicio en la familia considerada.

El resto de las familias corresponderían a aquellas que tienen el marginal de la primera fila diferente al de la familia considerada, pero también se podrían formar las sucesiones

$$R_{1,k-1}^*, R_{1,k-1}^* + 1, \dots (\text{valor máximo de } R_{1,k-1})$$

$$R_{1,k-1}^*, R_{1,k-1}^* - 1, \dots (\text{valor mínimo de } R_{1,k-1})$$

y, como si $R_{1,k-1}^*$ aumenta o disminuye en una unidad, $R_{2,k-1}^*$ disminuye o aumenta en una unidad, a partir de los marginales dados accederíamos a la familia de marginales consecutivos a través de las filas primera y segunda, y a partir de aquí se repetiría el primer proceso señalado.

Cualquiera que sea el nodo hijo $R_{1,k-1}$, $R_{2,k-1}$, $R_{3,k-1}$ siempre se puede llegar a él partiendo de $R_{1,k-1}^*$, $R_{2,k-1}^*$, $R_{3,k-1}^*$, dado que se cumple

$$\sum_{i=1}^3 R_{i,k-1} = \sum_{i=1}^3 R_{i,k-1}^* = C_1 + \dots + C_{k-1}, \text{ solo hay que dirigir de forma adecuada los}$$

procesos anteriormente señalados.

Por último, y en cuanto a extensión de los resultados conseguidos para tablas $2 \times c$ que pueden simplificar el algoritmo para computar el p-valor de una tabla $r \times c$,

señalar la extensión inmediata de la observación 2.7, que para tablas rxc quedaria en la siguiente forma.

Observación 3.1. Sea $(k; \mathbf{R}_k)$ un nodo madre, y $PAST_1, \dots, PAST_M$ las longitudes de los subcaminos recorridos hasta ese nodo, ordenados de menor a mayor. Para cada nodo hijo $(k-1; \mathbf{R}_{k-1})$ de dicho nodo madre,

$$PAST_j \cdot l \cdot LMA(k-1; \mathbf{R}_{k-1}) \leq PAST_{j'} \cdot l \cdot LMA(k-1; \mathbf{R}_{k-1}) \Leftrightarrow j \leq j'$$

y

$$PAST_j \cdot l \cdot LMI(k-1; \mathbf{R}_{k-1}) \leq PAST_{j'} \cdot l \cdot LMI(k-1; \mathbf{R}_{k-1}) \Leftrightarrow j \leq j'$$

siendo j y j' elementos del conjunto T , $T = \{1, \dots, M\}$.

III.4 :Discusión.

La metodología desarrollada para calcular una tabla 3xc de máxima probabilidad para unos marginales fijos, bajo la hipótesis de independencia entre filas y columnas o, equivalentemente para calcular la moda de una MMHD, y las relaciones encontradas entre los máximos de determinadas distribuciones condicionadas de una MMHD, pretenden ser solo el inicio de una generalización de las metodologías desarrolladas para tablas 2xc,y, no se pueden considerar cerradas sino abiertas en varios sentidos sobre los que es nuestra intención incidir en un futuro, cabe citar a título de ejemplo:

- Tratar de decidir, de forma inmediata, bajo qué condiciones la tabla resultante de sustituir en una tabla de máxima probabilidad una subtabla 2xk(o kx2) por su conectada sigue proporcionando tablas de máxima probabilidad (lo que experimentalmente , en los millones de casos inspeccionados, sucede en un porcentaje próximo al 80%).

- Investigar si el algoritmo desarrollado para determinar el máximo puede simplificarse cuando el proceso que se nos presenta necesita pasar por un elevado número de familias hasta acceder al máximo de la familia dada.
- Demostrar si una tabla $4 \times c$ con todas sus subtablas $h \times c$, $h < 4$, de máxima probabilidad, puede, y solo puede aumentar su probabilidad sumándole a sus subtabla 4×4 los elementos de un cuadrado latino 4×4 con los elementos $+1, 0, 0, -1$. Y generalizar esto para el caso $r \times c$.
- Demostrar que es condición necesaria y suficiente para que una tabla $r \times c$ ($r \leq c$) sea de máxima probabilidad, que todas sus subtablas $r \times r$ lo sean.

No obstante, y a pesar de las cuestiones que quedan abiertas, los procedimientos desarrollados para tablas $3 \times c$ suponen una serie de mejoras esenciales en el algoritmo para ejecutar el test de Fisher para estas tablas y que pueden quedar resumidas en los siguientes apartados:

- Hemos desarrollado un método, para una familia de tablas $3 \times c$ con marginales filas y columnas dados, que nos permite determinar, en algunos casos de forma inmediata y en los restantes casos de manera recursiva, una tabla de máxima probabilidad. Este método se ha basado sobre la aportación de una condición necesaria y suficiente para que una tabla tenga todas sus subtablas $2 \times c$ de máxima probabilidad, y también de una condición necesaria y suficiente para que una tabla $3 \times c$ con todas sus subtablas $2 \times c$ de máxima probabilidad tenga probabilidad máxima, cuando las metodologías existentes, salvo para el caso de marginales columnas coincidentes (Teorema 1.3 de Mehta y Patel), lo único que proporcionaban era una condición necesaria para tablas de máxima

probabilidad (Teorema 1.12 y Teorema 1.13 de Harry Joe), que en ocasiones resultaba demasiado laxa en el sentido de conducir a un conjunto bastante numeroso de tablas, donde ni tan siquiera todas tenían todas sus subtablas 2xc de máxima probabilidad y donde se hacia necesario determinar las probabilidades de las tablas del conjunto construido para decidir cual o cuales eran las que cumplían el requisito de tener máxima probabilidad.

- En el algoritmo reticulado de M-P, hay que utilizar el método de determinación de la moda de una MMHD para calcular las modas de todas las MMHD asociadas a nodos hasta los que se extiende el proceso, nosotros hemos encontrado un procedimiento que nos exige el utilizar el método solo para la MMHD asociada al nodo inicial de la red.
- En el algoritmo de M-P para un nodo madre $(k; \mathbf{R}_k)$ con una serie de valores PAST, $PAST_1, \dots, PAST_M$, hasta ese nodo , ordenados de menor a mayor, y con una serie de nodos hijos $(k-1; \mathbf{R}_{k-1})$, hay que discutir para cada nodo hijo combinado con todos los posibles valores PAST la verificación de $PAST_j. l. LMA(k-1; \mathbf{R}_{k-1}) \leq L(X)$ ó $PAST_j. l. LMI(k-1; \mathbf{R}_{k-1}) > L(X)$, $j=1, \dots, M$, nosotros podemos evitar esta discusión para numerosos nodos hijos haciendo uso de la observación 3.1, en la misma forma que lo hemos hecho para tablas 2xc.

BIBLIOGRAFÍA

AGRESTI, A., and WACKERLY, D., (1977) "Some exact conditional test of independence for $r \times c$ cross-classification tables", *Psychometrika*, Vol. 42, n° 1, pg. 111-125.

AGRESTI, A., (1992) "A survey of exact inference for contingency tables", *Statistical Science*, Vol. 7, n° 1, pg. 131-177.

BAGLIVO, J., OLIVIER, D., and PAGANO, M., (1985) " Computing Fisher and likelihood ratio exact tail probabilities for contingency tables", in *Proceedings of the Statistical Computing Section, American Statistical Association*, pg 70-77.

BAGLIVO, J., OLIVIER, D., and PAGANO, M., (1988) " Methods for the analysis of contingency tables with large and small cell counts", *Journal of the American Statistical Association, Computational Statistics*, 83, n° 404, pg 1006-1013.

BAGLIVO, J., OLIVIER, D., and PAGANO, M., (1992) "Methods for exact goodness-of-fit-test", *Journal of the American statistical association*, Vol. 87, n° 418, pg 464-469.

BAKER, R.J., (1977) "Exact distributions derived from two-way tables", *Algorithm AS 112*, pg 199-206.

BARNARD, G.A., (1945) " A new test for 2×2 tables", *Nature*, 156, 177 and 783-784.

BARNARD, G.A., (1947) " Significance test for 2×2 tables", *Biometrika*, 34, pg 123-138.

BARNARD, G. A. (1982), "Conditionality Versus Similarity in the Analysis of 2×2 Tables", *Statistics and Probability: Essays in Honor of C. R. Rao*, pg 59-65.

BENNETT, B.M., and NAKAMURA, E., (1963) "Tables for testing significance in a 2x3 contingency table", *Technometrics*, 5, pg. 501-511.

BERGER, R.L. and BOOS, D.D., (1994) "P values maximized over a confidence set for the nuisance parameter", *J. Amer. Statist. Ass.*, 89, pg 1012-1016.

BERKSON, J., (1978) "In dispraise of the exact test", *J. Statist. Plann Inference*, 2, pg 27-42. Discussion, 3, pg 181-213, (1979).

BOLAND, P.J. and PROSCHAN, F., (1987) "Schur Convexity of the Maximum Likelihood Function for the Multivariate Hypergeometric and Multinomial Distributions", *Statistics & Probability Letters*, 5, pg 317-322.

BOULTON, D.M., (1974) "Remark on algorithm 434", *Comm. ACM* 17, pg 326.

BOYETT, J. (1979) "Random RxC tables with given row and column totals", *J. Roy Statist. Soc. Ser. C* 28, pg 329-332.

CANTOR, A.B., (1979) "A computer algorithm for testing significance in MxK contingency tables", Fifth Proceeding of the Statistical Computing Section of the American Statistical Association.

CLARKSON, D.B., and FAN, Y. A. (1989) "Some improvements to the network algorithm for exact probabilities in contingency tables", *IMSL tech. Rep. 8903*. IMSL Inc., Houston, TX.

CLARKSON, D.B., and FAN, Y. A., and JOE, H. (1993) "A Remark on Algorithm 643: FEXACT: An Algorithm for performing Fisher's Exact Test in rxc Contingency tables", *ACM transactions on Mathematical Software*, Vol. 19, n° 4, pag 484-488.

CORMACK, R. S. and MANTEL, N. (1991), "Fisher's exact Test: The Marginal Totals as seen from two different angles", *The Statistician*, 40, pg 27-34.

- CRESSIE, N. and READ, T. R. C., (1989) "Pearson's X and the Loglikelihood Ratio Statistic G: A Comparative Review", *International Statistical Review*, 57, n° 1, pg 19-43.
- FISHER, R.A., (1934) "Statistical Methods for Research Workers", 5th Edition, Oliver and Boyd, Edinburgh.
- FISHER, R.A., (1935) "The Logic of inductive Inference", *J. Royal Statist. Soc.* 98.
- FREEMAN, G.H., and HALTON, J.H. (1951) "Note on an exact treatment of contingency, goodness of fit and other problems of significance", *Biometrika* 38, pg 141-149.
- FUNG, P.K. and LEE, J. (1989) "Extension of Fisher's exact test to 2-by-k contingency tables: a computer program in Basic", *Computer Methods and Programs in Biomedicine*, 28, pg 195-196.
- GAIL, M. and MANTEL, N., (1977) "Counting the Number of RxC Contingency Tables with Fixed Margins," *Journal of the American Statistical Association*, 72, n° 360, pg 859-862.
- GREENLAND, S., (1991), "On the Logical Justification of Conditional Tests for Two-By-Two Contingency Tables," *The American Statistician*, 45, n° 3, 248-251.
- HABER, M., (1989) "Do the marginal totals of a 2x2 contingency table contain information regarding the table proportions?," *Commun. Statist.- Theory Meth.*, 18, n° 1, 147-156.
- HABERMAN, S. J. (1988) "A warning on the use of chi-squared statistics with frequency tables with small expected cell counts," *Journal of the American Statistical Association*, " 83, 402, 555-560.
- HADLEY, G. (1962) "Linear Programming", Reading, Mass.: Addison-Wesley.
- HANCOCK, T.W. (1975) "Remark on Algorithm 434", *Commun ACM* 18, 117-119.

- HELLAND, I.S., (1995) "Simple counterexamples against the conditionality principle", *Amer. Statist.* 49 (4), pg 351-356.
- HIRJI, K.F. and JOHNSON, T.D. (1996) "A comparison of algorithms for exact analysis of unordered $2 \times k$ contingency tables" , *Computational Statistics & Data Analysis* 21, pg 419-429.
- HOWELL, D.C., and GORDON, L.R. (1976) " Computing the exact probability of an r by c contingency table with fixed marginal total" , *Behavior Research Method & Instrumentation*, Vol 8 (3), pg 317.
- IRWIN, j.d., (1935) " Test of significance for differences between percentages based on small numbers", *Metron*, 12(2), pg 84-94.
- JOE, H., (1985) " An ordering of dependence for contingency tables" , *Linear Algebra and its applications* 70, pg 89-103.
- 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, n° 11, pg 3677-3685.
- KEMPTHORNE, O.,(1979) " In dispraise of the exact test: reactions, " *Journal of Statistical Planning and Inference*, 3, pg 199-213.
- KOEHLER, K. and LARNZ, K. (1980) " An empirical investigation of goodness-of-fit statistics for sparse multinomials", *J. Amer. Statist. Assoc.* 75, pg 336-344.
- KOTZ, J., (1969) "Discrete Distributions", John Wiley and sons.
- KROONENBERG, P.M., and VERBEEK, A. (1987) " Comments on a generalization of Fisher's exact test in $p \times q$ contingency tables using more concordant relations" , *Commun. Statist. Simula*, 16(1), 301-306.

LARNTZ, K., (1978) " Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics, " Journal of the American Statistical Association, 73, n° 362, pg 253-263.

LEWOUTIN, R. C., and FELSENSTEIN, J, (1965) "The Robustness of homogeneity test in tables $2 \times N$ ", Biometrics, 21, pg 19-23.

LITTLE, R.J.A., (1989) " Testing the equality of two independent binomial proportions", Amer. Statist. 43 (4), pg 283-288.

LLOYD, C. J., (1988) " Some issues arising from the analysis of 2×2 contingency tables, " Austral. J. Statist., 30, n° 1, pg 35-46.

MARCH, D.L. (1972), "Exact Probabilities for $R \times C$ Contingency Tables," Communications of the Association of Computing Machinery, 15, pg 991-992.

MARTIN ANDRES, A. (1991) " A review of classic non-asymptotic methods for comparing two proportions by means of independent samples", Commun. Statist. Simulation and Comput., 20, pg 551-583.

MEHTA, C.R., and PATEL, N.R. (1980) "A network algorithm for the exact treatment of the $2 \times k$ contingency table" , Comm. Statist. Simulation Comput, B9, pg 649-664.

MEHTA, C. R. and PATEL, N. R., (1983) "A Network Algorithm for Performing Fisher's Exact Test in $R \times C$ Contingency Tables," Journal of the American Statistical Association, 78, n° 382, pg 427-434.

MEHTA, C.R., and PATEL, N.R. (1986) " Algorithm 643 FEXACT: a Fortran subroutine for Fisher's exact test on unordered $r \times c$ contingency tables" , ACM Transactions on mathematical software, vol 12 n°2, pg 154-161.

MEHTA , C. R., and PATEL, N. R., (1986 a) " A hybrid algorithm for Fisher`s exact test in unordered RxC contingency tables, " Commun. Statist.- Theor. Meth., 15, n° 2, pg 387-403.

MEHTA, C.R., and HILTON,J.F., (1993) "Exact power of conditional and unconditional test: going beyond the 2x2 contingency table" , The American Statistician, vol 47,n°2, pg 91-98.Comment by R.L. Berger and Reply, vol. 48, n°2, pg 175-176 (1994).

MEHTA,C.R. (1994) "The exact analysis of contingency tables in medical research" , Statistical Methods in Medical Research, 3, pg 135-156.

NGUYEN, T. T. (1985) "A Generalization of Fisher's Exact Test in PxQ Contingency Tables Using More Concordant Relations." Commun. Statist.- Simula. Computa; 14, n° 3, pg 633-645.

PAGANO, M., and HALVORSEN, K. T. (1981) " An algorithm for finding the exact significance levels of rxc contingency tables, " Journal of the American Statistical Association, 76, n° 376, pg 931-934.

PAGANO, M. and TRITCHLER, D., (1983) " On obtaining permutation distributions in polynomial time" , J. Amer. Statist. Assoc., 78, pg 435-440.

PATEFIELD, W.M. (1981) "An efficient method of generating random rxc tables with given row and column total" , J. Roy. Statist. Soc. Ser. C30, pg 91-97.

PEARSON, E.S., (1947) “ The choice of statistical tests illustrated on their interpretation of data classified in a 2x2 table “, Biometrika, 34, pg 139-167.

SAUNDERS, I.W (1984) "Algorithm as 205. Enumeration of rxc tables with repeated row totals", Royal Statistical Society, Applied Statistics, 340-352.

- SHAO, X. M, (1997) “ An efficient algorithm for the exact test on unordered $2 \times J$ contingency tables with equal column sums “. *Computational Statistics & Data Analysis* , 25, pg 273-285.
- SILVA MATO, A., (1993) “ Comparación de dos proporciones independientes por el método incondicionado”, Tesis Doctoral. Dpto. de Estadística e I.O.Universidad Complutense de Madrid.
- STATXACT, (1991) “Statxact: Statistical software for exact non-parametric inference”, versión 2, Cytel Software, Cambridge, Mass.
- UPTON, G.J.G. (1982) “ A comparison of alternative tests for the 2×2 comparative trial”, *J.R. Statist. Soc. A*, 145(1) pg 86-105.
- UPTON, G.J.G., (1992) “ Fisher’s exact test”, *J:R. Statist. Soc. A*, 155(3), pg 395-402.
- VERBEEK, A. and KROONENBERG, P.M. (1979) “Exact X^2 tests of independence in contingency tables with small numbers”, Preprint n°122. Dept. of Mathematics, University of Utrecht, The Netherlands, presented at 12th EMS, Varna, Bulgaria.
- VERBEEK, A. and KROONENBERG, P.M. (1985) " A survey of algorithms for exact distributions of test statistics in $r \times c$ contingency tables with fixed margins" , *Computacional Statistics & Data Analysis* 3, pg 159-185.
- YATES, F., (1934) “Contingency tables involving small numbers and the χ^2 test”, *J. R. Statist. Soc. Supp. 1*, pg 217-235.
- YATES , F. (1984) " Test of significance for 2×2 contingency tables, " *J.R. Statist. Soc. A*, 147, par 3, pg 426-463.