



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**Una introducción a métodos de estadística algebraica y  
sus aplicaciones en problemas**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**A C T U A R I O**

**P R E S E N T A:**

**FRANK PATRICK MURPHY HERNÁNDEZ**



**DIRECTOR DE TESIS:  
ACTUARIO FERNANDO DÍAZ LÓPEZ  
2015**

Ciudad Universitaria, D. F.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno	1. Datos del alumno
Apellido Paterno	Murphy
Apellido Materno	Hernández
Nombres	Frank Patrick
Teléfono	55-94-96-19
Universidad Nacional Autónoma de México	Universidad Nacional Autónoma de México
Facultad de Ciencias	Facultad de Ciencias
Número de Cuenta	405020508
2. Datos del tutor	2. Datos del tutor
Grado	Actuario
Apellido Paterno	Díaz
Apellido Materno	López
Nombre	Fernando
3. Datos del sinodal 1	3. Datos del sinodal 1
Grado	Doctora
Apellido Paterno	Alonso
Apellido Materno	Reyes
Nombres	María del Pilar
4. Datos del sinodal 2	4. Datos del sinodal 2
Grado	Doctora
Apellido Paterno	Fuentes
Apellido Materno	García
Nombres	Ruth Selene
5. Datos del sinodal 3	5. Datos del sinodal 3
Grado	Maestro en Ciencias
Apellido Paterno	Pérez
Apellido Materno	Carbajal
Nombre	Francisco
6. Datos del sinodal 4	6. Datos del sinodal 4
Grado	Maestro en Ciencias
Apellido Paterno	Pérez
Apellido Materno	Arriaga
Nombre	Fernando Daniel
7. Datos del trabajo escrito	Datos del trabajo escrito
Título	Una introducción a métodos de estadística algebraica y sus aplicaciones en problemas de tablas de contingencia
Número de páginas	67 páginas
Año	2015

## Agradecimientos

A mi mamá María Fé  
A mi papá Michael  
A mis hermanas Mitzy y Melanie  
A mi tía Bety  
A mi abuela Cuquita  
A mi tío Martín

---

# Índice

---

Introducción	6
Capítulo 1. Preliminares	8
Geometría Algebraica . . . . .	8
Bases de Gröbner. . . . .	11
Ideales Tóricos y Variedades Tóricas . . . . .	12
Tablas de Contingencia. . . . .	13
Complejos Simpliciales Abstractos . . . . .	15
Gráficas . . . . .	16
Capítulo 2. Una Introducción a la Estadística Algebraica	18
Un Ejemplo Motivacional. . . . .	19
Variedades y Modelos Estadísticos . . . . .	21
Modelos Mezclados. . . . .	24
Capítulo 3. Estadística Algebraica y Problemas en Tablas de Contingencia	27
Modelos Jerárquicos . . . . .	27
Bases de Markov . . . . .	32
Bases de Markov de Modelos Reducibles . . . . .	37
Un Ejemplo de Modelos Reducibles. . . . .	45
Capítulo 4. Aplicaciones	47
La prueba $\chi^2$ de Pearson . . . . .	47
Cadenas de Markov . . . . .	52
Tasa de convergencia . . . . .	54
No interacción en tres factores . . . . .	55
Modelos log-lineales . . . . .	57
Capítulo 5. Conclusiones	60

	ÍNDICE	5
ANEXOS-Códigos en R		61
Bibliografía		65

---

# Introducción

---

Esta tesis busca presentar dos cosas, primera ser un primer acercamiento a la estadística algebraica, y segundo y el tema principal de la tesis es ver sus aplicaciones a la teoría de tablas de contingencia. Es importante destacar que la estadística algebraica es de reciente creación. Aunque el álgebra ya era usada en la estadística en pruebas de hipótesis, estimación de parámetros, diseño de experimentos y en análisis multivariado, en específico en series de tiempo, hasta la creación de la estadística algebraica no se le había dado la importancia debida como herramienta de apoyo de la estadística. Cabe mencionar que la estadística algebraica no hace uso de todas las cuestiones algebraicas, las técnicas están enfocadas en el uso de la geometría algebraica y del álgebra conmutativa.

La estadística algebraica se puede decir que comenzó con el artículo Algebraic algorithms for sampling from conditional distributions de Sturmfels y Diaconis. De esta pareja de matemáticos, Diaconis es el estadista y Sturmfels es el geometra algebraico y algebrista conmutativo. Es importante destacar el papel de Sturmfels, ya que ha logrado integrar lo más abstracto de las matemáticas en un sin fin de aplicaciones que no sólo se limita a la estadística.

La idea de usar geometría en la estadística es original de Pearson en 1956 en su carta a la Real Sociedad Estadística, aunque se debe de destacar que Fisher (1921) hacen uso de una gran heurística geométrica para lograr sus resultados. El primero en usar geometría con tablas de contingencia fue Fienberg (1968).

En el primer capítulo se hace un repaso de todo lo que se necesita para poder entender esta tesis, también sirve para establecer una notación desde el principio. En el segundo capítulo se empieza con un ejemplo motivacional para aclarar de una manera introductoria que es la estadística algebraica, para continuar con una definición formal y resultados básicos de la materia. Al final

del capítulo se presentan un caso particular de modelos, los modelos mezclados, que después serán de interés.

En el capítulo tres se presenta como se aplica la teoría de la estadística algebraica a las tablas de contingencia, de hecho, a los modelos que cumplen este objetivo se les llama modelos jerárquicos. Uno de las principales virtudes de estos modelos es que se pueden descomponer en modelos más pequeños que se llaman modelos irreducibles. En el último capítulo se ven las aplicaciones de estos modelos, estas aplicaciones son desde un punto de vista histórico lo que motivo el nacimiento de la estadística algebraica, problemas donde las pruebas clásicas no se pueden efectuar por que las observaciones son pocas y lo que provoca que la estimación sea mala, se logra corregir con las técnicas presentadas en los capítulos anteriores.

# 1

---

## Preliminares

---

Aquí se presentarán los conceptos básicos de geometría algebraica y de estadística que se usarán a lo largo de la tesis, así como establecer la notación y otros aspectos.

Primero se considerará que los naturales incluirán al cero, y que para todo natural  $n$ ,  $[n]$  denota al conjunto  $\{1, \dots, n\}$ .

### 1. Geometría Algebraica

Para un anillo conmutativo  $R$  un ideal  $I$  es un subconjunto no vacío tal que es cerrado bajo la suma y  $RI \subseteq I$ , este hecho se denotará por  $I \leq R$ . Para  $S$  un subconjunto de  $R$ ,  $\langle S \rangle$  denota al conjunto de todas las  $R$ -combinaciones lineales de  $S$  y resulta ser el mínimo ideal que contiene a  $S$ . Un ideal se dice que es finitamente generado si existen  $x_1, \dots, x_n$  tal que  $\langle x_1, \dots, x_n \rangle = I$ . Un ideal  $I$  de  $R$  se llama primo si es un subconjunto propio de  $R$  y cada vez que  $ab \in I$  entonces  $a \in I$  o  $b \in I$ . Un ideal  $I$  de  $R$  se llama semiprimo si es un subconjunto propio de  $R$  y cada vez que  $a^2 \in I$  entonces  $a \in I$ . Un ideal  $I$  de  $R$  se llama radical, si cada vez que  $a^p \in I$  para algún natural  $p > 1$  entonces  $a \in I$ . Si un ideal es propio son equivalentes ser radical y ser semiprimo. La intersección de ideales radicales es un ideal radical, por lo que todo ideal está contenido en un ideal radical que es mínimo con respecto a esta propiedad, esto es, para  $I \leq R$  se define  $\sqrt{I} = \{a \in R \mid \text{existe } n > 0, a^n \in I\}$ .

PROPOSICIÓN 1.1. *Sea  $R$  un anillo conmutativo e  $I, J \leq R$ . Entonces:*

- $\sqrt{I} \leq R$ .
- $\sqrt{I}$  es un ideal radical.
- $I \subseteq \sqrt{I}$ .
- $I$  es un ideal radical si y sólo si  $\sqrt{I} = I$ .
- Si  $I \subseteq J$  y  $J$  es radical entonces  $\sqrt{I} \subseteq J$ .
- Si  $I \subseteq J$  entonces  $\sqrt{I} \subseteq \sqrt{J}$ .

Para  $K$  un campo,  $K[x_1, \dots, x_n]$  denota el anillo de polinomios en  $n$  indeterminadas con coeficientes en  $K$ , donde  $n$  es un natural. En esta tesis el campo  $K$  sólo podrá ser los racionales  $\mathbb{Q}$  (para cuestiones computacionales), los reales  $\mathbb{R}$  (para cuestiones puramente teóricas de las estadística algebraica) o los complejos  $\mathbb{C}$  (para cuestiones de geometría algebraica), aunque los resultados de este capítulo son válidos para cualquier campo. Un monomio  $f \in K[x_1, \dots, x_n]$  es un polinomio de la forma  $f(x_1, \dots, x_n) = \prod_{i=1}^n x_i^{\alpha(i)}$  con  $\alpha : [n] \rightarrow \mathbb{N}$  una función, una mejor notación para  $f$  es  $f = \bar{x}^\alpha$ . El anillo de polinomios en  $n$  indeterminadas con coeficientes en un campo  $K$  tiene estructura natural de  $K$  espacio vectorial y el conjunto de los monomios forma una base natural, por lo que todo polinomio en  $n$  indeterminadas se puede expresar de forma única como  $K$  combinación lineal de monomios, esto es, para todo  $f \in K[x_1, \dots, x_n]$ ,

$$f(x_1, \dots, x_n) = \sum_{\alpha: [n] \rightarrow \mathbb{N}} c_\alpha \bar{x}^\alpha$$

con  $c_\alpha \in K$  siendo todos cero excepto un número finito.

Ahora para  $S \subset K[x_1, \dots, x_n]$  se construye  $V(S) = \{(x_1, \dots, x_n) \in K^n \mid f(x_1, \dots, x_n) = 0 \text{ para todo } f \in S\}$ , a los subconjuntos de  $K^n$  de esta forma se les llaman variedades algebraicas afines.

En el caso de que  $S = \{f\}$  se escribirá  $V(f)$  en lugar de  $V(\{f\})$ .

**PROPOSICIÓN 1.2.** Sean  $K$  un campo,  $n$  un natural,  $S, T \subseteq K[x_1, \dots, x_n]$ ,  $I, J \leq K[x_1, \dots, x_n]$  e  $\{I_\lambda\}_{\lambda \in \Lambda}$  una familia de ideales de  $K[x_1, \dots, x_n]$ . Entonces:

- Si  $S \subseteq T$  entonces  $V(T) \subseteq V(S)$ .

- $V(K[x_1, \dots, x_n]) = \emptyset$ .
- $V(0) = K^n$ .
- $V(S) = V(\langle S \rangle) = V(\sqrt{\langle S \rangle})$ .
- $V(IJ) = V(I \cap J) = V(I) \cup V(J)$ .
- $V(\sum_{\lambda \in \Lambda} I_\lambda) = \bigcap_{\lambda \in \Lambda} V(I_\lambda)$ .

Ahora se da el proceso inverso, es decir, para  $T \subseteq K^n$  se le asigna  $I(T) = \{f \in K[x_1, \dots, x_n] \mid f(t) = 0 \text{ para todo } t \in T\}$ . Se tiene las siguientes propiedades:

**PROPOSICIÓN 1.3.** *Sean  $K$  un campo,  $n$  un natural,  $S, T \subseteq K^n$  y  $J \subseteq K[x_1, \dots, x_n]$ . Entonces:*

- $I(S) \subseteq K[x_1, \dots, x_n]$ .
- $I(S)$  es un ideal radical.
- Si  $S \subseteq T$  entonces  $I(T) \subseteq I(S)$ .
- $J \subseteq I(V(J))$ .
- $J = I(V(J))$  si y sólo si  $J$  es un ideal radical.
- $V(J) = V(I(V(J)))$ .
- $S \subseteq V(I(S))$ .
- $S = V(I(S))$  si y sólo si  $S$  es una variedad algebraica.
- $I(S) = I(V(I(S)))$ .

Un hecho importante a tomarse en cuenta es el teorema de Hilbert, pero para eso primero hay que recordar que un anillo conmutativo es neteriano, si todo ideal es finitamente generado.

**TEOREMA 1.1** (De las bases de Hilbert). *Sea  $R$  un anillo conmutativo. Si  $R$  es neteriano entonces  $R[x]$  es neteriano*

**COROLARIO 1.1.** *Sean  $K$  un campo y  $n$  un natural positivo. Entonces  $K[x_1, \dots, x_n]$  es un anillo noetheriano.*

Como referencias de álgebra conmutativa (los conceptos de ideal, anillo, radical, etc.) se recomienda Introduction To Commutative Algebra de Michael Atiyah, se resalta que el libro de es un clásico obligado. Con respecto a geometría algebraica se recomienda Elementary Algebraic Geometry de Hulek.

## 2. Bases de Gröbner

Un orden monomial  $\leq$  es un orden total en los monomios de  $K[x_1, \dots, x_n]$  de tal manera que el mínimo elemento sea 1 y para cualesquiera  $m_1, m_2, m_3$  monomios si  $m_1 \leq m_2$  entonces  $m_1 m_3 \leq m_2 m_3$ . Como ejemplo básico se tiene al orden lexicográfico donde  $\bar{x}^\alpha < \bar{x}^\beta$  si  $\alpha(1) < \beta(1)$  o existe  $i = 2, \dots, n$  tal que  $\alpha(i) < \beta(i)$  y  $\alpha(j) = \beta(j)$  para toda  $j = 1, \dots, i-1$ , en éste caso  $x_1^5 x_2 x_3^9 < x_1^7$ . En caso de que no se especifique el orden monomial siempre se supondrá que es el lexicográfico.

Para  $f \in K[x_1, \dots, x_n]$  y  $\leq$  un orden monomial se define el término principal de  $f$  según  $\leq$  como el monomio de mayor grado de  $f$  con respecto a  $\leq$  y se denota por  $LT_{\leq}(f)$ . Para  $I \leq K[x_1, \dots, x_n]$  se puede construir el ideal inicial de  $I$  con respecto a  $\leq$  que es  $\langle LT_{\leq}(f) \mid f \in I \rangle$  y éste se denota por  $LT_{\leq}(I)$ .

**DEFINICIÓN 2.1.** *Sea  $I$  un ideal de  $K[x_1, \dots, x_n]$  y  $\leq$  un orden monomial. Un conjunto finito  $G \subseteq I$  tal que  $LT_{\leq}(I) = \langle LT_{\leq}(g) \mid g \in G \rangle$  es una base de Gröbner de  $I$  con respecto a  $\leq$ .*

Las bases de Gröbner son útiles por la siguiente proposición.

**PROPOSICIÓN 2.1** (Teorema de la Eliminación). *Sea  $I$  un ideal de  $K[x_1, \dots, x_n]$  y  $\leq$  un orden monomial. Si  $G \subseteq I$  es una base de Gröbner entonces  $\langle G \rangle = I$ .*

Existe un algoritmo llamado algoritmo de Buchberger, que fue propuesto por Bruno Buchberger en 1965 en su tesis doctoral, el algoritmo toma como entradas un ideal y un orden monomial, y como salida calcula una base de Gröbner.

**DEFINICIÓN 2.2.** Sea  $I$  un ideal de  $K[x_1, \dots, x_n]$  y  $1 \leq m \leq n$ , el  $m$ -ésimo ideal de eliminación se define como  $I \cap K[x_m, \dots, x_n]$  y se denota por  $I_m$ .

Los ideales de eliminación son importantes por que se pueden computar usando bases de Gröbner lexicográficas.

**PROPOSICIÓN 2.2.** Sea  $I$  un ideal de  $K[x_1, \dots, x_n]$ ,  $\leq$  el orden lexicográfico y  $G$  una base de Gröbner para  $I$ . Entonces  $G \cap K[x_m, \dots, x_n]$  es una base de Gröbner respecto al orden lexicográfico para  $I_m$ .

Gracias a esta proposición se puede computar el ideal de la imagen de una función polinomial.

**PROPOSICIÓN 2.3.** Sea  $f : K^d \rightarrow K^n$  una función polinomial con  $f = (f_1, \dots, f_n)$  donde  $f_i \in K[x_1, \dots, x_d]$  con  $i = 1, \dots, n$  e  $I$  el ideal de  $K[x_1, \dots, x_d, y_1, \dots, y_n]$  generado por  $y_i - f_i(x_1, \dots, x_d)$  para  $i = 1, \dots, n$ . Entonces  $I(\text{im} f)$  es el ideal de eliminación  $I \cap K[y_1, \dots, y_n]$ .

Para los aspectos computacionales de la geometría algebraica se recomienda *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra* de Cox, Little y O'Shea. Se hace notar que el libro presenta todos los resultados de forma constructivista, no acepta el axioma de elección. Todas las pruebas son por medio de algoritmos y tiene una tendencia intuicionista, donde sólo se usa el tercer excluido para comprobar que los algoritmos se detienen.

### 3. Ideales Tóricos y Variedades Tóricas

Para  $R$  un anillo conmutativo  $M_{d \times n}(R)$  denota las matrices de  $d$  por  $n$  con coeficientes en  $R$ . Sea  $A \in M_{d \times n}(\mathbb{Z})$  tal que  $A_{ij}$  es un natural para toda  $1 \leq i \leq d$  y  $1 \leq j \leq n$ , se define  $f^A : K^d \rightarrow K^n$  como  $f^A = (f_1^A, \dots, f_n^A)$  donde

$$f_j^A(x_1, \dots, x_d) = \prod_{i=1}^d x_i^{A_{ij}}$$

Como ejemplo se considera la matriz  $A$ ,

$$A = \begin{pmatrix} 0 & 3 & 7 \\ 5 & 2 & 1 \end{pmatrix}$$

En este caso resulta que  $f^A(x_1, x_2) = (x_2^5, x_1^3 x_2^2, x_1^7 x_2)$ .

**DEFINICIÓN 3.1.** *La variedad tórica asociada a  $A$ , denotada por  $X_A$ , es  $V(I(\text{im}f^A))$ .*

**DEFINICIÓN 3.2.** *El ideal tórico asociado a  $A$ , denotado por  $I_A$ , es  $I(\text{im}f^A)$ .*

La siguiente proposición presenta una forma alternativa para caracterizar al ideal tórico en función de una familia infinita de generadores, aunque por el teorema de las bases de Hilbert ésta se puede reducir a una familia finita.

**PROPOSICIÓN 3.1.** *Para  $A$  una matriz de  $d$  por  $n$  con coeficientes naturales,  $I_A \leq K[\bar{x}] = K[x_1, \dots, x_n]$  tiene como generadores a  $\bar{x}^u - \bar{x}^v$  son tales que  $Au = Av$ .*

Por lo visto en la sección pasada usando el método de las bases de Gröbner se puede computar un conjunto finito de generadores de  $I_A$ .

## 4. Tablas de Contingencia

Una tabla de contingencia es una matriz que se usa en estadística para mostrar la frecuencia de distribuciones de un conjunto de variables aleatorias. El término tabla de contingencia fue usado por primera vez por Karl Pearson en "On the Theory of Contingency and Its Relation to Association and Normal Correlation" en 1904.

**4.1. Definición** Formalmente, se consideran  $X_1, \dots, X_k$  variables aleatorias que toma valores en  $[d_1], \dots, [d_k]$  respectivamente y se pone  $D = [d_1] \times \dots \times [d_k]$ ,  $\mathbb{R}^D$  denota el espacio vectorial de todas las funciones de  $D$  en los reales, este espacio tiene dimensión  $d_1 \dots d_k$ . La clasificación cruzada de  $n$  realizaciones independientes del vector aleatorio  $(X_1, \dots, X_k)$  produce un vector aleatorio con entradas naturales  $n \in \mathbb{R}^D$ , a esto se le llama una tabla de contingencia de  $k$ -formas donde  $n_{i_1, \dots, i_k}$  es el número de veces que la entrada  $(i_1, \dots, i_k)$  aparece en la muestra.

Un ejemplo de tabla de contingencia es:

	Diestro	Zurdo
Hombre	43	9
Mujer	44	4

En este caso la variable  $X_1$  es una variable Bernoulli que se interpreta como el sexo de una persona y la variable  $X_2$  también es una Bernoulli que se interpreta el lado con que una persona es más hábil.

De manera más practica una tabla contingencia (de 2-formas), es una forma de clasificar  $n$  objetos de acuerdo a dos criterios  $A$  y  $B$  donde el criterio  $A$  tiene  $r$  categorías  $A_1, \dots, A_r$  y el criterio  $B$  tiene  $s$  categorías  $B_1, \dots, B_s$ . En el ejemplo el criterio  $A$  es el sexo y el criterio  $B$  es el lado más hábil. También y comúnmente se calculan las sumas por columna y por renglón, ya que esto dice la clasificación sólo con respecto a un criterio. Es decir, una tabla de contingencia se presenta comúnmente como:

	Diestro	Zurdo	Total
Hombre	43	9	52
Mujer	44	4	48
Total	87	13	100

Esto dice que con respecto al sexo se tiene la siguiente clasificación:

Hombre	Mujer
52	48

Esto dice que con respecto al lado más hábil se tiene la siguiente clasificación:

Diestro	Zurdo
87	13

**4.2. Distribución Multinomial** Una hipótesis que se suele tener es que la probabilidad  $p_{ij}$  de uno de los  $n$  objetos de pertenecer a la categoría  $(A_i, B_j)$  es estocásticamente independiente y tiene distribución conjunta:

$$P(N = (n_{ij}), i = 1, \dots, r, j = 1, \dots, s) = \frac{n!}{\prod_{ij} n_{ij}!} \prod_{ij} p_{ij}^{n_{ij}}$$

Donde  $N$  es la variable que denota todas las posibles tablas de contingencias de  $r \times s$  que clasifican  $n$  objetos.

## 5. Complejos Simpliciales Abstractos

Los complejos simpliciales son construidos a partir de formar bloques llamados simplejos, los simplejos son puntos, segmentos de recta, triángulos rellenos, tetraedros sólidos y sus análogos de dimensiones mayores. Los complejos simpliciales tienen la ventaja de codificar información topológica en información puramente combinatoria. Sean  $x_1, \dots, x_n, x \in V$  un  $\mathbb{R}$ -espacio vectorial,  $x$  es una combinación convexa de  $x_1, \dots, x_n$  si existen  $t_1, \dots, t_n \in [0, 1]$  tal que  $x = \sum_{i=1}^n t_i x_i$  y  $\sum_{i=1}^n t_i = 1$ . Para  $x_0, \dots, x_n, x \in V$  se dirá que son afinmente independientes si  $x_1 - x_0, \dots, x_n - x_0$  son linealmente independientes. Para  $x_0, \dots, x_n \in V$  afinmente independiente el simplejo generado por  $x_0, x_1, \dots, x_n$  es el conjunto de todas las combinaciones convexas de  $x_0, \dots, x_n$  y se denota por  $[x_0, \dots, x_n]$ , a  $n$  se le llama la dimensión del simplejo y comunmente se llamará un  $n$ -simplejo para hacer referencia a su dimensión.

**DEFINICIÓN 5.1.** *Un complejo simplicial abstracto  $\mathcal{K}$  es un conjunto de conjuntos finitos no vacíos que cumple: Si  $S \in \mathcal{K}$  y  $T \subseteq S$  con  $T$  no vacío entonces  $T \in \mathcal{K}$ .*

Si  $\mathcal{K}$  es un complejo simplicial abstracto a sus elementos se les llama simplejos abstractos, cualquier elemento  $s$  de un simplejo abstracto se llama un vértice, mientras que cualquier subconjunto no vacío de un simplejo abstracto se llama una cara y una faceta es un simplejo abstracto que no esté contenido en ningún otro simplejo abstracto, se observa inmediatamente que un complejo simplicial está determinado por su conjunto de facetas. Se dice que  $\mathcal{K}$  es un complejo finito si  $\mathcal{K}$  como conjunto es finito y localmente finito si cualquier vértice pertenece sólo a un número finito de simplejos abstractos. La dimensión

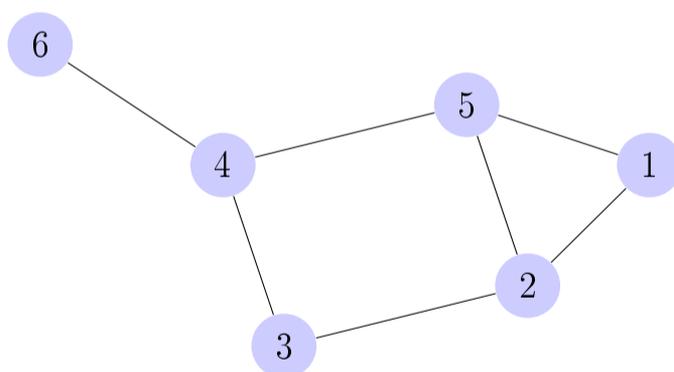
de un simplejo abstracto es su cardinalidad menos uno, si las dimensiones de un complejo simplicial abstracto están acotadas por arriba entonces se dice que el complejo es de dimensión finita y su dimensión es el supremo de las dimensiones de sus simplejos abstractos. Por último se define

$$|\mathcal{K}| := \bigcup \mathcal{K}$$

que resulta ser el conjunto de vértices. Los simplejos abstractos son una generalización de los simplejos usuales, y esta generalización guarda mucha de la información geométrica en un lenguaje combinatorio.

## 6. Gráficas

Una gráfica  $G$  es una pareja  $(V, E)$  de conjuntos al conjunto  $V$  se le llama el conjunto de vértices de la gráfica y sólo es un conjunto, en general en la teoría de gráficas se trabaja con  $V$  finito y en algunos casos numerable, mientras que  $E$  se le llama el conjunto de aristas de la gráfica y consta de subconjuntos de  $V$  de cardinalidad dos, naturalmente a los elementos de  $V$  se les llama vértices y a los de  $E$  aristas, notacionalmente para una arista se escribirá  $uv$  en lugar de  $\{u, v\}$ , se observa que bajo esta notación es lo mismo  $uv$  que  $vu$ , es decir, es simétrica la notación, así  $u \in G$  denotará que  $u \in V$  y  $uv \in G$  denotará que  $uv \in E$ . Las gráficas se pueden representar de manera geométrica, esto se hace representando a los vértices por círculos y a las aristas por segmentos de recta que unen a los círculos, por ejemplo, considérese la gráfica  $G$  cuyo conjunto de vértices es  $V = [6]$  y cuyo conjunto de aristas es  $E = \{12, 15, 23, 25, 34, 45, 46\}$ , representación geométrica resulta:



Una camarilla  $C$  de una gráfica  $G = (V, E)$  es un subconjunto del conjunto de vértices de tal modo que para cualquiera dos elementos distintos de  $C$  la arista que los une pertenece a la gráfica  $G$ , siguiendo con el ejemplo anterior  $C = \{1, 6\}$  no es una camarilla de  $G$  por que la arista 16 no pertenece a  $G$  mientras que  $\{1, 2, 5\}$  si es una camarilla de  $G$  de hecho es la camarilla de mayor cardinalidad. A una camarilla a la cual si se le agrega otro vértice deja de ser camarilla se le llama una camarilla máxima.

---

# Una Introducción a la Estadística Algebraica

---

En años recientes se ha hecho el uso del término estadística algebraica para hacer referencia al uso de técnicas de geometría algebraica y de álgebra conmutativa para estudiar problemas estadísticos relacionados con variables aleatorias finitas. De manera más formal, la estadística algebraica está dedicada al estudio de variedades algebraicas reales que surgen en el análisis multivariado discreto. El análisis multivariado discreto se interesa en modelos estadísticos para variables aleatorias que tienen un número finito de estados. El punto clave de la estadística algebraica es que la mayoría de los modelos estadísticos que se usan en la práctica para analizar variables aleatorias discretas son variedades algebraicas.

Históricamente el uso de herramientas algebraicas se había hecho presente en distintas áreas de la estadística como el uso de la teoría de grupos en el diseño experimental y el uso de formas cuadráticas en modelos fijos y aleatorios de efectos lineales. En la estadística algebraica se usan polinomios, aunque el uso de éstos no es nuevo en la estadística ya que desde el comienzo de ésta se usaron para modelos de regresión polinomial y en modelos multiplicativos derivados de modelos de independencia para tablas de contingencia o en terminología más moderna modelos para información categórica. La diferencia con los modelos de estadística algebraica es que en ésta se hace uso no sólo de los polinomios sino de la maquinaria de la geometría algebraica clásica, es decir, para analizar un polinomio se tienen varios objetos algebraicos alrededor de este como variedades algebraicas, ideales, el anillo de polinomios, entre otros. Los artículos donde se

empieza esta nueva rama de la estadística son (Diaconis and Sturmfels 1998) y (Pistone and Wynn 1996).

## 1. Un Ejemplo Motivacional

Se considera el modelo de regresión cuadrática en una variable dado por:

$$Y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \epsilon(x)$$

Si se observan cuatro puntos de diseño distintos sin réplica  $x_1, x_2, x_3$  y  $x_4$  se tiene la forma matricial usual de la regresión

$$\eta = E[Y] = X\theta$$

donde  $\eta = (\eta_1, \eta_2, \eta_3, \eta_4)^t$ ,  $\theta = (\theta_0, \theta_1, \theta_2)^t$ ,

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{pmatrix}$$

,  $Y$  es el vector de observación,  $\theta$  el vector de parámetros y los errores tienen esperanza cero. Por lo que se ve anteriormente se tiene el siguiente sistema de ecuaciones cuadráticas.

$$\begin{aligned} \theta_0 + \theta_1 x_1 + \theta_2 x_1^2 &= \eta_1 \\ \theta_0 + \theta_1 x_2 + \theta_2 x_2^2 &= \eta_2 \\ \theta_0 + \theta_1 x_3 + \theta_2 x_3^2 &= \eta_3 \\ \theta_0 + \theta_1 x_4 + \theta_2 x_4^2 &= \eta_4 \end{aligned}$$

Usando un poco de teoría de eliminación se puede ver que el sistema anterior es equivalente a:

$$\begin{aligned} & -(x_2 - x_3)(x_2 - x_4)(x_3 - x_4)\eta_1 + (x_1 - x_3)(x_1 - x_4)(x_3 - x_4)\eta_2 \\ & -(x_1 - x_2)(x_1 - x_4)(x_2 - x_4)\eta_3 + (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)\eta_4 = 0 \end{aligned}$$

Con la condición adicional de que si  $i \neq j$  entonces  $x_i \neq x_j$ . La versión del sistema de ecuaciones se podría llamar un modelo estadístico algebraico explícito y la última ecuación sería un modelo estadístico algebraico implícito. En lugar de un modelo de regresión lineal se puede considerar un modelo de regresión lineal generalizado donde  $Y_i$  se supone Poisson con parámetro  $\lambda_i$  para  $i = 1, 2, 3, 4$ , con log-vínculo:

$$\ln(\lambda_i) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2$$

para  $i=1, 2, 3, 4$ . En este caso se tiene:

$$\begin{aligned} & -(x_2 - x_3)(x_2 - x_4)(x_3 - x_4) \ln(\lambda_1) + (x_1 - x_3)(x_1 - x_4)(x_3 - x_4) \ln(\lambda_2) \\ & -(x_1 - x_2)(x_1 - x_4)(x_2 - x_4) \ln(\lambda_3) + (x_1 - x_2)(x_1 - x_3)(x_2 - x_3) \ln(\lambda_4) = 0 \end{aligned}$$

Esto implica, después de aplicar la función exponencial en la ecuación pasada, se tiene que:

$$\lambda_1^{(x_2-x_3)(x_2-x_4)(x_3-x_4)} \lambda_3^{(x_1-x_2)(x_1-x_4)(x_2-x_4)} - \lambda_2^{(x_1-x_3)(x_1-x_4)(x_3-x_4)} \lambda_4^{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = 0$$

Más aún se puede suponer que  $x_1, x_2, x_3$  y  $x_4$  son enteros, por ejemplo  $-1, 0, 1$  y  $2$  respectivamente. Después de exponenciar y haciendo una reducción se tiene

$$\lambda_1 \lambda_3^3 - \lambda_2^3 \lambda_4 = 0$$

Se hace notar que este modelo implícito define una variedad tórica. También se puede condicionar a que  $n = Y_1 + Y_2 + Y_3 + Y_4$ , en éste caso  $Y_1 + Y_2 + Y_3 + Y_4$  tiene distribución multinomial de parámetros  $p_i = \frac{\lambda_i}{n}$  para  $i = 1, 2, 3, 4$  y se tiene que éstos satisfacen  $p_1 p_3^3 - p_2^3 p_4 = 0$ . Como se ve en el ejemplo es posible llevar modelos estadísticos a planteamientos de naturaleza de la geometría algebraica. La idea de usar geometría algebraica en estadística es la siguiente: entender mejor los modelos ya conocidos, ayudar o innovar en la metodología

estadística e inferencia y definir una nueva clase de modelos en la cual se pueda explotar sus características algebraicas.

## 2. Variedades y Modelos Estadísticos

Considérese una variable aleatoria  $X$  con número finito de estados y que el conjunto de estados es  $[n]$  con  $n$  natural. La distribución de  $X$  se puede expresar como un vector  $(p_1, \dots, p_n) \in \mathbb{R}^n$  tal que  $\sum_{i=1}^n p_i = 1$  y  $p_i \geq 0$  para toda  $i = 1, \dots, n$ , donde  $p_i = P(X = i)$ . El conjunto de todos los vectores reales que cumplen lo anterior se denota por  $\Delta_n$  y se le llama el simplejo de probabilidad  $n - 1$ -dimensional, es decir,  $\Delta_n$  es el conjunto de todos los  $n$ -vectores de distribución. Un modelo estadístico para  $X$  es un subconjunto no vacío  $\mathcal{M}$  de  $\Delta_n$ .

De manera análoga al de la geometría algebraica para  $S \subseteq K[x_1, \dots, x_n]$  se define  $V_\Delta(S) = \{(x_1, \dots, x_n) \in \Delta_n \mid f(x_1, \dots, x_n) = 0 \text{ para todo } f \in S\}$ , a estos conjuntos se les llaman conjuntos semialgebraicos. Se observa inmediatamente que  $V_\Delta(S) = V(S) \cap \Delta_n$ . En el caso de que  $S = \{f\}$  se escribirá  $V_\Delta(f)$  en lugar de  $V_\Delta(\{f\})$ .

**DEFINICIÓN 2.1.** *Un modelo estadístico algebraico es un conjunto no vacío  $\mathcal{M}$  de la forma  $V_\Delta(S)$ .*

Se considera el siguiente ejemplo. Sea  $X = (X_1, X_2)$  una variable aleatoria 2-dimensional donde  $X_1$  y  $X_2$  toman valores en  $\{1, 2\}$ . Como se había planteado el modelo  $X$  toma valores en [4], para facilitar la notación en lugar de usar  $p_1, p_2, p_3$  y  $p_4$  se usará  $p_{11}, p_{12}, p_{21}$  y  $p_{22}$  donde  $p_{ij} = P(X_1 = i, X_2 = j)$ . Se pone  $p_{+i} = p_{1i} + p_{2i}$  y  $p_{i+} = p_{i1} + p_{i2}$  para  $i = 1, 2$ , es decir,  $p_{+i} = P(X_2 = i)$  y  $p_{i+} = P(X_1 = i)$ . A continuación una pequeña proposición que empieza a juntar las dos partes, la de geometría algebraica y la de estadística.

**PROPOSICIÓN 2.1.** *Sean  $X = (X_1, X_2)$  una variable aleatoria 2-dimensional donde  $X_1$  y  $X_2$  toman valores en  $\{1, 2\}$  y  $f \in K[x_{11}, x_{12}, x_{21}, x_{22}]$  con  $f(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11}x_{22} - x_{12}x_{21}$ . Entonces  $X_1$  y  $X_2$  son independientes si y sólo si  $(p_{11}, p_{12}, p_{21}, p_{22}) \in V_\Delta(f)$ .*

Demostración

Por la independencia se tiene que  $p_{ij} = p_{i+p+j}$  para  $i, j = 1, 2$ . Ahora calculando

$$\begin{aligned} f(p_{11}, p_{12}, p_{21}, p_{22}) &= p_{11}p_{22} - p_{12}p_{21} \\ &= p_{1+p+1}p_{2+p+2} - p_{1+p+2}p_{2+p+1} \\ &= 0 \end{aligned}$$

lo que significa que  $(p_{11}, p_{12}, p_{21}, p_{22}) \in V_{\Delta}(f)$ . Para el regreso se hace notar que despejando se obtiene

$$p_{22} = \frac{p_{12}p_{21}}{p_{11}}$$

De aquí se tiene que

$$\begin{aligned} p_{1+p+1} &= (p_{11} + p_{12})(p_{11} + p_{21}) \\ &= p_{11}^2 + p_{11}p_{12} + p_{11}p_{21} + p_{12}p_{21} \\ &= p_{11}\left(p_{11} + p_{12} + p_{21} + \frac{p_{21}p_{12}}{p_{11}}\right) \\ &= p_{11}(p_{11} + p_{12} + p_{21} + p_{22}) \\ &= p_{11} \end{aligned}$$

Análogamente se hace para  $p_{12}, p_{21}$  y  $p_{22}$  para obtener la independencia. ■

Lo que se acaba de demostrar es que  $V_{\Delta}(f)$  es un modelo estadístico algebraico para la independencia de dos variables binarias. Una generalización del resultado anterior es:

**PROPOSICIÓN 2.2.** Sean  $X_1$  y  $X_2$  variables aleatorias finitas que tienen  $d_1$  y  $d_2$  estados respectivamente. Se define  $f_{ijkl}(\bar{x}) = x_{ij}x_{kl} - x_{il}x_{kj}$  para  $i, k = 1, \dots, d_1$  y  $j, l = 1, \dots, d_2$ . Entonces  $V_{\Delta}(\{f_{ijkl} \mid 1 \leq i, k \leq d_1, 1 \leq j, l \leq d_2\})$  es un modelo estadístico algebraico para la independencia de  $X_1$  y  $X_2$ .

Aunque ya se tiene un modelo fuerte desde un punto de vista algebraico, éste presenta el problema de ser implícito (como en el ejemplo anterior) y no paramétrico como usualmente se presentan los modelos estadísticos, ahora se buscará dar una buena definición de lo que será un modelo estadístico paramétrico algebraico. Para esto primero hay que tener en cuenta la siguiente definición.

**DEFINICIÓN 2.2.** *Un modelo estadístico paramétrico es un conjunto  $f(\Theta)$ , donde  $\Theta \subseteq \mathbb{R}^d$  y  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  es una función continua tal que  $f(\Theta) \subseteq \Delta_n$ .*

A partir de esto se puede definir la versión algebraica como:

**DEFINICIÓN 2.3.** *Un modelo estadístico paramétrico algebraico es un conjunto  $f(\Theta)$ , donde  $\Theta \subseteq \mathbb{R}^d$  es un conjunto semialgebraico y  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  es una función polinomial tal que  $f(\Theta) \subseteq \Delta_n$ , con función polinomial se entiende que existen  $f_1, \dots, f_n \in \mathbb{R}[x_1, \dots, x_d]$  con  $f = (f_1, \dots, f_n)$ .*

Ahora considérese el siguiente ejemplo: sea  $\Theta = \Delta_2$  el espacio paramétrico y  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}$  dado por  $f(x, y) = (xy, x(1-y), (1-x)y, (1-x)(1-y))$ . Si se considera  $(\theta_1, \theta_2) \in \Theta$ , se tiene que  $\theta_1\theta_2 + \theta_1 - \theta_1\theta_2 + \theta_2 - \theta_1\theta_2 + 1 - \theta_1 - \theta_2 + \theta_1\theta_2 = 1$  y con el hecho de que  $0 \leq \theta_1\theta_2, \theta_1(1-\theta_2), (1-\theta_1)\theta_2, (1-\theta_1)(1-\theta_2) \leq 1$  se tiene que  $f(\theta_1, \theta_2) \in \Delta_{2 \times 2}$ , por lo que  $f(\Theta)$  es un modelo estadístico paramétrico algebraico. Como está dado  $f$  justo se contruye el modelo para la independencia que ya se había visto anteriormente, es decir, si se pone  $f(\theta_1, \theta_2) = (\theta_1\theta_2, \theta_1(1-\theta_2), (1-\theta_1)\theta_2, (1-\theta_1)(1-\theta_2)) = (p_{11}, p_{12}, p_{21}, p_{22})$  se tiene que  $f(\Theta) = V_{\Delta}(g)$  con  $g(x_{11}, x_{12}, x_{21}, x_{22}) = x_{11}x_{22} - x_{12}x_{21}$ .

En los modelos estadísticos paramétricos algebraicos también se puede pedir considera  $f(\Theta) \cap \Delta_n$  en vez de  $f(\Theta) \subseteq \Delta_n$  pero esto es el análogo a proyectivizar las variedades afines en geometría algebraica y corresponde a los modelos jerárquicos.

Ahora se plantea de nuevo el modelo de independencia paramétrico pero sin pedir que  $f(\Theta) \subseteq \Delta_n$ . Sean  $X_1$  y  $X_2$  dos variables aleatorias que tomen valores en  $[d_1]$  y  $[d_2]$  respectivamente,  $\Theta = [0, \infty)^{d_1+d_2}$  y  $f : \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_1 \times d_2}$  dada por  $f(\bar{x}) = (f_{ij}(\bar{x}))_{1 \leq i \leq d_1, 1 \leq j \leq d_2}$  donde  $f_{ij}(\bar{x}) = x_i x_{d_1+j}$  para  $1 \leq i \leq d_1, 1 \leq j \leq d_2$ , en éste caso  $\mathcal{M} = f(\Theta) \cap \Delta_n$ . Este modelo en geometría algebraica es la parte no negativa del encaje de Segre.

### 3. Modelos Mezclados

Siempre se tiene la contención  $f(\Theta) \subseteq V_{\Delta}(I(f(\Theta)))$  que no tiene porque ser una igualdad, es decir, la contención puede ser estricta, esta pregunta surge naturalmente en el contexto de la geometría algebraica, ya que la igualdad se da justo cuando el conjunto es un conjunto algebraicamente afín, pero desde un punto estadístico esto tiene mucha interpretación. Éste caso suele ocurrir en los modelos mezclados, o en modelos con una variable aleatoria oculta. Un modelo mezclado supone que se está observando propiedades de una población que es una mezcla de varias subpoblaciones independientes y que cuando se ve al individuo en la población no se sabe a que subpoblación pertenece.

**DEFINICIÓN 3.1.** *Sea  $\mathcal{M}$  un modelo estadístico para una variable aleatoria  $X$  en  $n$  estados y  $H$  un variable aleatoria en  $m$  estados. El modelo mezclado consiste en*

$$Sec^m(\mathcal{M}) = \left\{ \sum_{i=1}^m \pi_i p_i \mid \pi \in \Delta_m, p_i \in \mathcal{M} \text{ para todo } i = 1, \dots, m \right\}$$

Esto es, el conjunto de todas  $m$ -combinaciones convexas de  $\mathcal{M}$ . En el modelo mezclado la probabilidad de que un individuo pertenezca a una subpoblación  $i$  es  $p_i$  y no se observa a que subpoblación pertenece el individuo.

La notación  $Sec^m(\mathcal{M})$  viene de geometría algebraica y el modelo mezclado corresponde a la  $m$ -ésima variedad secante de  $\mathcal{M}$  pero aquí sólo se consideran las combinaciones convexas.

Sea  $\mathcal{M}$  el modelo de independencia de dos variables aleatorias  $X_1$  y  $X_2$  que toman valores en [4]. El ideal  $I(Sec^3(\mathcal{M}))$  es generado por el polinomio

$$f(\bar{x}) = \begin{vmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{vmatrix}$$

donde  $\bar{x} = (x_{11}, x_{12}, x_{13}, x_{14}, x_{21}, x_{22}, x_{23}, x_{24}, x_{31}, x_{32}, x_{33}, x_{34}, x_{41}, x_{42}, x_{43}, x_{44})$ .

El conjunto  $Sec^3(\mathcal{M})$  consiste en todas las matrices no negativas de rango 3 que se pueden escribir como combinación convexa de 3 matrices no negativas

de rango 1. Éste conjunto está contenido propiamente en  $V_{\Delta}(f)$  que consiste en la matrices no negativas de rango 3. Por ejemplo la matriz

$$\begin{pmatrix} \frac{1}{8} & \frac{1}{8} & 0 & 0 \\ 0 & \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & 0 & 0 & \frac{1}{8} \end{pmatrix}$$

no se puede escribir como combinación convexa de 3 matrices de rango 1.

A continuación se presenta el modelo Jukes-Cantor introducido en 1969 es el modelo más sencillo de secuencias de ADN, es otro modelo mezclado. Se llama  $T$  al octante positivo de  $\mathbb{R}^3$ , es decir,  $T = \{(t_1, t_2, t_3) \in \mathbb{R}^3 \mid t_1, t_2, t_3 \geq 0\}$  y se fija  $\alpha > 0$ . Para  $i = 1, 2, 3$  se considera la matriz de transición

$$M^i = \frac{1}{4} \begin{pmatrix} 1 + 3e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} \\ 1 - e^{-\alpha t_i} & 1 + 3e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} \\ 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 + 3e^{-\alpha t_i} & 1 - e^{-\alpha t_i} \\ 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 - e^{-\alpha t_i} & 1 + 3e^{-\alpha t_i} \end{pmatrix}$$

con  $i = 1, 2, 3$ .

Estas tres matrices se usan para construir un modelo de tres variables aleatorias observadas  $X_1, X_2$  y  $X_3$  y una variable aleatoria  $H$  oculta, cada una de éstas variables toman cuatro estados. Este modelo se parametriza con  $\phi : T \rightarrow \mathbb{R}^{64}$  con  $\phi(t_1, t_2, t_3) = (\phi_{jkl}(t_1, t_2, t_3))_{1 \leq j, k, l \leq 4}$  donde

$$\phi_{jkl}(t_1, t_2, t_3) = \frac{1}{4} \sum_{i=1}^4 M_{ij}^1 M_{ik}^2 M_{il}^3$$

El conjunto de distribuciones de probabilidad que surgen de esta para metrización se llama el modelo de Jukes-Cantor en el árbol de tres hojas-garra, como referencia del tema se sugiere revisar Lectures on Algebraic Statistics en el capítulo de Hidden Variables. Bajo un cambio de coordenadas se convierte este modelo en un modelo estadístico paramétrico algebraico, para esto se pone

$x_i = \frac{1}{4}(1 - e^{-\alpha t_i})$  para  $i = 1, 2, 3$ , esta transformación convierte al espacio paramétrico  $T$  en el espacio paramétrico  $\Theta = \{(x_1, x_2, x_3) \mid x_1, x_2, x_3 \in [0, \frac{1}{4}]\}$ . Por lo que las matrices de transición se convierten en

$$M^i = \frac{1}{4} \begin{pmatrix} 1 - 3x_i & x_i & x_i & x_i \\ x_i & 1 - 3x_i & x_i & x_i \\ x_i & x_i & 1 - 3x_i & x_i \\ x_i & x_i & x_i & 1 - 3x_i \end{pmatrix}$$

con  $i = 1, 2, 3$ .

### 3

---

# Estadística Algebraica y Problemas en Tablas de Contingencia

---

## 1. Modelos Jerárquicos

Sean  $X_1, \dots, X_n$  variables aleatorias que toman valores en  $[d_i]$  para  $i = 1, \dots, n$ , y en el caso que todas sean binarias, esto es que  $d_i = 2$  para  $i = 1, \dots, n$ , se supondrá que toman valores en  $\{0, 1\}$  en vez de de  $\{1, 2\}$ , también se pondrá  $D = \prod_{i=1}^n [d_i]$ . Un modelo jerárquico es un modelo en el cual se codifica la interacción entre estas variables mediante las llamadas funciones potenciales. Se considera  $\mathcal{K}$  un complejo simplicial abstracto tal que sus vértices pertenezcan a  $[n]$ , el mismo  $n$  que el número de variables aleatorias que se están considerando, ahora se pone como  $\Delta$  al conjunto de facetas de  $\mathcal{K}$  y se enumeran los elementos de ésta, es decir,  $\Delta = \{F_1, \dots, F_s\}$ .

El conjunto  $F \in \Delta$  intenta codificar la información de  $\{X_i \mid i \in F\}$ . Para cada  $F \in \Delta$  hay una función potencial  $\Phi_F$ , cuyo dominio es  $\prod_{i \in F} [d_i]$ . Para un vector  $\bar{x} = (x_1, \dots, x_n) \in D$  y  $F \in \Delta$ ,  $\bar{x}_F$  denota la imagen de  $\bar{x}$  bajo la proyección canónica  $\pi : D \rightarrow \prod_{i \in F} [d_i]$ , por ejemplo, en el caso binario con  $n = 5$  y  $F = \{1, 3, 5\}$ , si  $\bar{x} = (1, 0, 0, 1, 0)$  entonces  $\bar{x}_F = (1, 0, 0)$ . El modelo consiste en estudiar todas las posible distribuciones de  $(X_1, \dots, X_n)$  que cumplen

$$P(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{Z} \prod_{F \in \Delta} \Phi(\bar{x}_F)$$

donde  $Z$  es una constante normalizadora llamada la función de partición. Obviamente se tiene que

$$Z = \sum_{\bar{x} \in D} \left( \prod_{F \in \Delta} \Phi(\bar{x}_F) \right)$$

Un problema importante y no sencillo en modelos jerárquicos de varias variables es el de calcular eficientemente la función de partición.

Se define  $D_F = \prod_{i \in F} [d_i]$  para cada  $F \in \Delta$ , ahora si  $F = \{k_1, \dots, k_m\}$  se considera un conjunto de variable ( $\mathbb{R}$ -polinomialmente independientes)  $\{x_\phi^F\}_{\phi \in D_F}$

DEFINICIÓN 1.1. Sea  $f^\Delta : \prod_{F \in \Delta} \mathbb{R}^{D_F} \longrightarrow \mathbb{R}^D$  dado por

$$f_{\bar{x}}((x_\phi^F)_{\phi \in D_F})_{F \in \Delta} = \prod_{F \in \Delta} x_{\bar{x}_F}^F$$

con  $\bar{x} \in D$ , el modelo jerárquico de  $\mathcal{M}_\Delta$  está dado por  $\text{im} f \cap \Delta_D$ .

Como un primer ejemplo se considera dos variables binarias  $X_1$  y  $X_2$  y  $\Delta = \{\{0\}, \{1\}\}$ , en este caso  $D = \{0, 1\} \times \{0, 1\}$ ,  $D_{\{i\}} = \{0, 1\}$  para  $i = 1, 2$  y en particular se tiene que

$$\begin{aligned} ((x_\phi^F)_{\phi \in D_F})_{F \in \Delta} &= ((x_\phi^{\{0\}})_{\phi \in D_{\{0\}}}, (x_\phi^{\{1\}})_{\phi \in D_{\{1\}}}) \\ &= ((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) \end{aligned}$$

Por otro lado si  $\bar{x} \in D$  entonces  $\bar{x} = (i, j)$  con  $i, j = 0, 1$ , por lo que se tiene que

$$\begin{aligned} f_{00}((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) &= x_0^{\{0\}} x_0^{\{1\}} \\ f_{01}((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) &= x_0^{\{0\}} x_1^{\{1\}} \\ f_{10}((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) &= x_1^{\{0\}} x_0^{\{1\}} \\ f_{11}((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) &= x_1^{\{0\}} x_1^{\{1\}} \end{aligned}$$

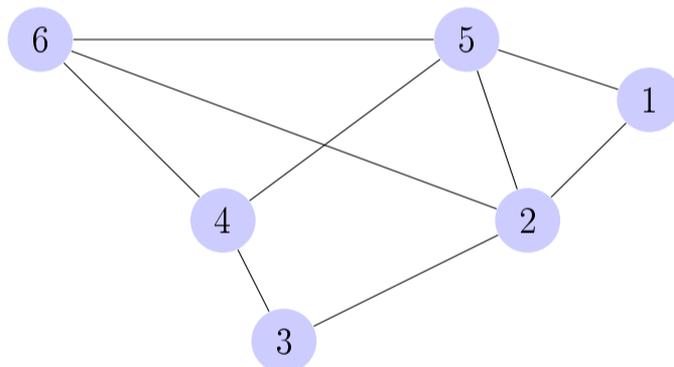
De aquí que la función polinomial  $f$  que se busca sea

$$f((x_0^{\{0\}}, x_1^{\{0\}}), (x_0^{\{1\}}, x_1^{\{1\}})) = (x_0^{\{0\}} x_0^{\{1\}}, x_0^{\{0\}} x_1^{\{1\}}, x_1^{\{0\}} x_0^{\{1\}}, x_1^{\{0\}} x_1^{\{1\}})$$

Si se renombran las variables es el modelo de independencia paramétrico que se observó anteriormente, el inducido por el encaje de Segre.

El modelo jerárquico de un simplejo simplicial abstracto  $\Delta$  se denotará por  $\mathcal{M}_\Delta$ , por ser un modelo estadístico paramétrico algebraico tiene asociado un ideal cuando se considera como variedad, el ideal de polinomios donde el modelo se anula. Como la parametrización es monomial el ideal asociado resulta ser un ideal tórico y se denotará por  $A_\Delta$  a la matriz que induce este ideal, siguiéndose la misma notación  $I_\Delta$  denotará el ideal tórico asociado a la variedad tórica.

Dentro de los modelos jerárquicos están los modelos de gráficas no dirigidas o mientras no se preste a confusión simplemente los modelos de gráficas, que son inducidos por un gráfica  $G = (E, V)$  y un vector de entero  $\mathbf{d}$ , estos modelos son bastante importantes. Para una gráfica  $G$  se puede considerar el complejo simplicial que induce  $\mathcal{K}(G)$  que es la familia de camarillas de  $G$  y como siempre basta considerar el conjunto de facetas  $\Delta(G)$  que en éste caso coincide con la familia de camarillas máximas. Por ejemplo se considera la siguiente gráfica  $G$ :



En este caso  $\Delta(G) = \{\{1, 2, 5\}, \{2, 5, 6\}, \{4, 5, 6\}\}$ . Dentro de los modelos gráficos se considerarán los descomponibles ya que tienen buenas propiedades como descripciones simples de sus ideales, así como formulas cerradas de estimados de sus estimadores máximos verosímiles. Los modelos descomponibles se definen recursivamente en términos de los modelos reducidos. En esencia lo que se busca es a un modelo grande dividirlo en varios más pequeños por lo

tanto más sencillos de trabajar a manera que los resultados de obtenidos en estos se puedan unir para obtener los resultados del modelo grande.

**DEFINICIÓN 1.2.** *Sea  $\mathcal{K}$  un complejo simplicial abstracto cuyos vértices pertenecen a  $[n]$ , se dice que  $\mathcal{K}$  es reducible si existen complejos simpliciales abstractos  $\mathcal{K}_1$  y  $\mathcal{K}_2$  tales que  $\mathcal{K}_1 \cup \mathcal{K}_2 = \mathcal{K}$  y  $\mathcal{K}_1 \cap \mathcal{K}_2 = 2^S$  para algún  $S$  subconjunto de  $[n]$  donde  $2^S$  se entiende por el conjunto potencia de  $S$  menos el vacío. A la pareja  $(\mathcal{K}_1, \mathcal{K}_2)$  se le llama una descomposición de  $\mathcal{K}$ .*

Lo primero que se observa es que esta definición se puede llevar a términos de los conjuntos de facetas. Por ejemplo sea  $\mathcal{K}$  el complejo simplicial abstracto dado por  $\Delta = \{\{1, 2\}, \{2, 3\}\}$  es reducible con  $\Delta_1 = \{\{1, 2\}\}$ ,  $\Delta_2 = \{\{2, 3\}\}$  y  $S = \{2\}$ . Por otro lado en el caso de que  $\Delta = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$  este complejo simplicial no es reducible ya que la única descomposición en esencia que se puede hacer es  $\Delta_1 = \{\{i, j\}\}$  y  $\Delta_2 = \{\{j, k\}, \{i, k\}\}$  con  $i, j, k \in \{1, 2, 3\}$  y  $i \neq j$ ,  $j \neq k$  e  $i \neq k$ , en éste caso  $\mathcal{K}_1 \cap \mathcal{K}_2 = \{\{i\}, \{j\}\}$  que no es el conjunto potencia de  $\{i, j\}$ . Se hace notar fuertemente que la intersección es sobre el complejo simplicial abstracto, pero sólo se mencionan los conjuntos de facetas dado que éstas determinan al complejo simplicial abstracto y se seguira abusando de la notación siempre y cuando no se preste a confusión.

**DEFINICIÓN 1.3.** *Sea  $\mathcal{K}$  un complejo simplicial abstracto reducible cuyos vértices pertenecen a  $[n]$ , se dice que  $\mathcal{K}$  es descomponible si para alguna descomposición  $(\mathcal{K}_1, \mathcal{K}_2)$  ambos como complejos simpliciales abstractos son reducibles o un conjunto unitario, es decir, ambos son reducibles, ambos son un conjunto unitario, o uno es reducible y el otro un conjunto unitario.*

La última definición puede ser un poco confusa, primero un conjunto unitario es un conjunto con un elemento, en el ejemplo dado por  $\Delta = \{\{1, 2\}, \{2, 3\}\}$  con  $\Delta_1 = \{\{1, 2\}\}$ ,  $\{\{2, 3\}\}$  resulta que también es descomponible por que en éste caso tanto  $\Delta_1$  como  $\Delta_2$  son conjuntos unitarios, lo que se busca es que estos de algún modo sean reducibles pero se manejarán por aparte y a veces como si lo fuesen. Ahora  $\Delta = \{\{1, 2\}, \{2, 3\}, \{4, 5\}, \{5, 6\}\}$  es reducible con  $\Delta_1 = \{\{1, 2\}, \{2, 3\}\}$ ,  $\Delta_2 = \{\{4, 5\}, \{5, 6\}\}$  y  $S = \emptyset$  y es descomponible por que  $\Delta_1$  y  $\Delta_2$  son reducibles. Por último  $\Delta = \{\{1, 2\}, \{2, 3\}, \{1, 3\}, \{3, 4\}\}$  es reducible con  $\Delta_1 = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ ,  $\Delta_2 = \{\{3, 4\}\}$  y  $S = \{3\}$ , en éste caso  $\Delta_1$  es reducible y  $\Delta_2$  es un conjunto unitario.

Como se ha definido un modelo jerárquico es fácil observar que el ideal que anula al modelo es un ideal tórico. Recordando lo mencionado sobre ideales tóricos en los preliminares si  $A_\Delta$  representa la matriz asociada al ideal tórico del modelo  $\Delta$  entonces  $I_\Delta = \{\bar{x}^u - \bar{x}^v \mid A_\Delta u = A_\Delta v, u, v \in \mathbb{N}^D\}$

Ahora se toman tres variables aleatorias binarias y el complejo simplicial abstracto con vértices en [3] inducido por el conjunto de facetas  $\Delta = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$ . Éste es el modelo de un tres ciclo. Entonces el ideal tórico está dado por:

$$I_\Delta = \langle x_{000}x_{011}x_{101}x_{110} - x_{001}x_{010}x_{100}x_{111} \rangle$$

Ahora se toman cuatro variables aleatorias binarias y el complejo simplicial abstracto con vértices en [4] inducido por el conjunto de facetas  $\Delta = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$ . Entonces el ideal tórico está dado por las veinticuatro cuádricas que se pueden obtener al calcular el determinante de los los menores de dos por dos de las siguientes matrices:

$$\begin{pmatrix} x_{0000} & x_{0001} & x_{0010} & x_{0011} \\ x_{1000} & x_{1001} & x_{1010} & x_{1011} \end{pmatrix}$$

$$\begin{pmatrix} x_{0100} & x_{0101} & x_{0110} & x_{0111} \\ x_{1100} & x_{1101} & x_{1110} & x_{1111} \end{pmatrix}$$

$$\begin{pmatrix} x_{0001} & x_{0101} & x_{1001} & x_{1101} \\ x_{0000} & x_{0100} & x_{1000} & x_{1100} \end{pmatrix}$$

$$\begin{pmatrix} x_{0011} & x_{0111} & x_{1101} & x_{1111} \\ x_{0010} & x_{0110} & x_{1100} & x_{1110} \end{pmatrix}$$

## 2. Bases de Markov

Los conjuntos de binomios generadores de ideales tóricos son conocidos como bases de Markov. El concepto de base de Markov de un modelo jerárquico es hoy en día uno de los más importantes de la estadística algebraica, de hecho, es éste el concepto con el que se pudo hacer la conexión con la geometría algebraica en el artículo de (Diaconis and Sturmfels 1998), estos resultados empezaron de manera práctica con la estadística algebraica.

Se consideran  $n$  variables aleatorias  $X_1, \dots, X_n$  que tomen valores en  $[d_1], \dots, [d_n]$  respectivamente,  $F = \{k_1, \dots, k_s\} \subseteq [n]$  ordenado en forma ascendente y  $\mathbf{u}$  una tabla con entradas enteras de  $d_1 \times \dots \times d_n$ , se denota por  $\mathbf{u}_F$  a la tabla  $F$ -marginal de  $\mathbf{u}$ . Esta nueva tabla  $\mathbf{u}_F$  es una tabla de  $d_{k_1} \times \dots \times d_{k_s}$  y dada por la regla:

$$(\mathbf{u}_F)_{i_{k_1} \dots i_{k_s}} = \sum_{j_{k_1}=i_{k_1}, \dots, j_{k_s}=i_{k_s}} \mathbf{u}_{j_1 \dots j_n}$$

Por ejemplo, si  $d_1 = 2$  y  $d_2, d_3 = 3$  las siguientes dos tablas dan todos los posibles valores de la tabla  $\mathbf{u}$ , esto se hace así por que  $\mathbf{u}$  es una tabla de  $2 \times 3 \times 3$

$$\mathbf{u}_1 = \begin{array}{|c|c|c|} \hline \mathbf{u}_{111} & \mathbf{u}_{112} & \mathbf{u}_{113} \\ \hline \mathbf{u}_{121} & \mathbf{u}_{122} & \mathbf{u}_{123} \\ \hline \mathbf{u}_{131} & \mathbf{u}_{132} & \mathbf{u}_{133} \\ \hline \end{array}$$

$$\mathbf{u}_2 = \begin{array}{|c|c|c|} \hline \mathbf{u}_{211} & \mathbf{u}_{212} & \mathbf{u}_{213} \\ \hline \mathbf{u}_{221} & \mathbf{u}_{222} & \mathbf{u}_{223} \\ \hline \mathbf{u}_{231} & \mathbf{u}_{232} & \mathbf{u}_{233} \\ \hline \end{array}$$

En el caso de que  $F = \{2, 3\}$  se tiene que:

$$\mathbf{u}_F = \begin{array}{|c|c|c|} \hline \mathbf{u}_{111} + \mathbf{u}_{211} & \mathbf{u}_{112} + \mathbf{u}_{212} & \mathbf{u}_{113} + \mathbf{u}_{213} \\ \hline \mathbf{u}_{121} + \mathbf{u}_{221} & \mathbf{u}_{122} + \mathbf{u}_{222} & \mathbf{u}_{123} + \mathbf{u}_{223} \\ \hline \mathbf{u}_{131} + \mathbf{u}_{231} & \mathbf{u}_{132} + \mathbf{u}_{232} & \mathbf{u}_{133} + \mathbf{u}_{233} \\ \hline \end{array}$$

En el caso de que  $F = \{1, 3\}$  se tiene que:

$$\mathbf{u}_F = \begin{array}{|c|c|c|} \hline \mathbf{u}_{111} + \mathbf{u}_{121} + \mathbf{u}_{131} & \mathbf{u}_{112} + \mathbf{u}_{122} + \mathbf{u}_{132} & \mathbf{u}_{113} + \mathbf{u}_{123} + \mathbf{u}_{133} \\ \hline \mathbf{u}_{211} + \mathbf{u}_{221} + \mathbf{u}_{231} & \mathbf{u}_{212} + \mathbf{u}_{222} + \mathbf{u}_{232} & \mathbf{u}_{213} + \mathbf{u}_{223} + \mathbf{u}_{233} \\ \hline \end{array}$$

En el caso de que  $F = \{1, 2\}$  se tiene que:

$$\mathbf{u}_F = \begin{array}{|c|c|c|} \hline \mathbf{u}_{111} + \mathbf{u}_{112} + \mathbf{u}_{113} & \mathbf{u}_{121} + \mathbf{u}_{122} + \mathbf{u}_{123} & \mathbf{u}_{131} + \mathbf{u}_{132} + \mathbf{u}_{133} \\ \hline \mathbf{u}_{211} + \mathbf{u}_{212} + \mathbf{u}_{213} & \mathbf{u}_{221} + \mathbf{u}_{222} + \mathbf{u}_{223} & \mathbf{u}_{231} + \mathbf{u}_{232} + \mathbf{u}_{233} \\ \hline \end{array}$$

En el caso en que cuando  $F = \{1\}$ :

$$\mathbf{u}_F = \begin{array}{|c|} \hline \mathbf{u}_{111} + \mathbf{u}_{121} + \mathbf{u}_{131} + \mathbf{u}_{112} + \mathbf{u}_{122} + \mathbf{u}_{132} + \mathbf{u}_{113} + \mathbf{u}_{123} + \mathbf{u}_{133} \\ + \mathbf{u}_{111} + \mathbf{u}_{112} + \mathbf{u}_{113} + \mathbf{u}_{121} + \mathbf{u}_{122} + \mathbf{u}_{123} + \mathbf{u}_{131} + \mathbf{u}_{132} + \mathbf{u}_{133} \\ \hline \mathbf{u}_{211} + \mathbf{u}_{221} + \mathbf{u}_{231} + \mathbf{u}_{212} + \mathbf{u}_{222} + \mathbf{u}_{232} + \mathbf{u}_{213} + \mathbf{u}_{223} + \mathbf{u}_{233} \\ + \mathbf{u}_{211} + \mathbf{u}_{212} + \mathbf{u}_{213} + \mathbf{u}_{221} + \mathbf{u}_{222} + \mathbf{u}_{223} + \mathbf{u}_{231} + \mathbf{u}_{232} + \mathbf{u}_{233} \\ \hline \end{array}$$

Ahora sea  $\Delta$  el conjunto de facetas de un complejo simplicial abstracto sobre  $[n]$ , si  $\Delta = \{F_1, \dots, F_r\}$  la matriz  $A_\Delta$  computa las tablas  $F_i$ -marginales con  $i = 1, \dots, r$  para cualquier tabla  $\mathbf{u}$  de  $d_1 \times \dots \times d_n$ , de la siguiente forma:

$$A_\Delta \mathbf{u} = (\mathbf{u}_{F_1}, \dots, \mathbf{u}_{F_r})$$

La siguiente es una definición alternativa a la dada de base de Markov. Un conjunto finito  $\mathcal{B}$  en el  $\mathbb{Z}$ -núcleo de  $A_\Delta$  será llamado base de Markov y se verá que las nociones coinciden.

**DEFINICIÓN 2.1.** *Sea  $A$  una matriz de enteros de  $d$  por  $n$ . Un conjunto finito  $\mathcal{B}$  en el  $\mathbb{Z}$ -núcleo de  $A_\Delta$  es llamado una base de Markov si para cualquier par de vectores de naturales  $u, v$  tales que  $Au = Av$ , existe una sucesión  $\{m_i\}_{i=1}^l$  en  $\mathcal{B}$  tal que  $v = u + \sum_{i=1}^l m_i$  y  $u + \sum_{i=1}^j m_i \geq 0$  para toda  $j = 1, \dots, l$ , donde que un vector sea positivo significa que todas sus entradas lo son.*

Los elementos de la base de Markov comúnmente se les llaman movimientos. La razón del nombre de base de Markov es por la siguiente aplicación; dada una tabla  $A_{\Delta \mathbf{u}}$  y una base de Markov  $\mathcal{B}$  para  $A_{\Delta}$  es posible generar dibujos aleatorios del conjunto de todas las tablas con entradas naturales que tengan las mismas marginales que  $A_{\Delta \mathbf{u}}$ . Esta caminata aleatoria se obtiene de acuerdo al siguiente algoritmo, se genera un movimiento aleatorio  $m \in \mathcal{B}$ , si  $\mathbf{u} + m \geq 0$  se reemplaza  $\mathbf{u}$  con  $\mathbf{u} + m$  con cierta probabilidad de acuerdo a una distribución estacionaria, este proceso se repite hasta que la estacionariedad se alcanza. Este procedimiento se puede usar para computar p-valores Monte Carlo para inferencia condicional en tablas de contingencia.

Ahora la siguiente proposición justifica el porque llamar una base de Markov a un conjunto de vectores con entradas naturales en vez de un conjunto de polinomios en varias variables. Para eso se hace saber que todo vector entero  $m$  se puede escribir de forma única como  $u^+ - u^-$  donde ambos vectores  $u^+$  y  $u^-$  tienen entradas naturales y soporte disjunto.

**PROPOSICIÓN 2.1.** *Sea  $A$  una matriz de enteros de  $d$  por  $n$  y  $\mathcal{B}$  un conjunto de vectores con entradas naturales. Entonces  $\{\bar{x}^{m^+} - \bar{x}^{m^-} \mid m \in \mathcal{B}\}$  es una base de Markov para el ideal tórico  $I_A$  si y sólo si  $\pm \mathcal{B}$  es una base de Markov para  $A$ .*

Demostración

Sean  $u$  y  $v$  vectores enteros tales que  $Au = Av$  entonces existen  $\{w_i\}_{i=1}^l$  vectores con entradas naturales y  $\{m_i\}_{i=1}^l$  movimientos tales que:

$$\bar{x}^u - \bar{x}^v = \sum_{i=1}^l \bar{x}^{w_i} (\bar{x}^{m_i^-} - \bar{x}^{m_i^+})$$

Despejando se obtiene que:

$$\bar{x} - \sum_{i=1}^l \bar{x}^{w_i} (\bar{x}^{m_i^-} - \bar{x}^{m_i^+}) = \bar{x}^v$$

De lo que se deduce por inducción sobre  $l$  que  $u + \sum_{i=1}^l m_i = v$ .

Para el regreso, se recuerda que un conjunto generador del ideal tórico es  $\{\bar{x}^u - \bar{x}^v \mid Au = Av\}$ , sean  $u$  un vector con entradas naturales y  $m$  un movimiento en  $\mathcal{B}$  tal que  $u + m \geq 0$  entonces:

$$\begin{aligned} \bar{x}^u - \bar{x}^{u-m^-}(\bar{x}^{m^-} - \bar{x}^{m^+}) &= \bar{x}^u - \bar{x}^{u-m^-+m^-} + \bar{x}^{u+m} \\ &= \bar{x}^{u+m} \end{aligned}$$

Esto da la idea de cómo se hace en general, es decir, sean  $u, v$  vectores enteros tales que  $Au = Av$ , entonces por hipótesis existe una familia  $\{m_i\}_{i=1}^l$  tal que  $v = u + \sum_{i=1}^l m_i$ , se pone:

$$\begin{aligned} w_1 &= u - m_1^- \\ w_2 &= u - m_2^- + m_1 \\ w_3 &= u - m_3^- + m_1 + m_2 \\ &\vdots \\ w_l &= u - m_l^- + \sum_{i=1}^{l-1} m_i \end{aligned}$$

Lo que implica que:

$$\bar{x} - \sum_{i=1}^l \bar{x}^{w_i}(\bar{x}^{m_i^-} - \bar{x}^{m_i^+}) = \bar{x}^v$$

Que despejando se tiene que:

$$\bar{x}^u - \bar{x}^v = \sum_{i=1}^l \bar{x}^{w_i}(\bar{x}^{m_i^-} - \bar{x}^{m_i^+})$$

De aquí  $\bar{x}^u - \bar{x}^v$  está en el ideal generado por  $\{\bar{x}^{m^+} - \bar{x}^{m^-} \mid m \in \mathcal{B}\}$ , lo que implica que genere el ideal tórico  $I_A$ . Es obvio que si  $m \in \mathbb{B}$  entonces  $\bar{x}^m \in \{\bar{x}^u - \bar{x}^v \mid Au = Av\}$  puesto que  $Am = 0 = A0$ . ■

Por el teorema de las bases de Hilbert todo conjunto generador tiene un subconjunto finito generador así que por la proposición anterior siempre existen las bases de Markov en el nuevo sentido.

Se dice que una base de Markov es mínima si es un conjunto generador mínimo de un ideal tórico. Una base de Markov mínima no necesariamente es única, notesé que no tiene porque ser única esta base de Markov. Para tratar este problema se considera:

**DEFINICIÓN 2.2.** *Sea  $A$  un matriz con entradas naturales de  $d$  por  $n$ , la base de Markov universal de  $A$  denotada por  $U(A)$  es la unión de todas las bases de Markov mínimas de  $A$ .*

Ahora una serie de definiciones de carácter técnico.

**DEFINICIÓN 2.3.** *Sean  $u, v$  y  $w$  vectores enteros con  $u = v + w$  se dice que ésta descomposición de  $u$  es semiconforme si  $v^+ \leq u^+$  y  $w^- \leq u^-$ . Una descomposición semiconforme se dice que es graduada si  $\|v\|_1 < \|u\|_1$  y  $\|w\|_1 < \|u\|_1$ . Sea  $A$  un matriz con entradas naturales de  $d$  por  $n$ , un vector en el  $\mathbb{Z}$ -núcleo de  $A$  se llama  $A$ -semiprimitivo si no existe una descomposición graduada semiconforme en elementos del núcleo.*

**DEFINICIÓN 2.4.** *Sea  $A$  un matriz con entradas naturales de  $d$  por  $n$  y  $A_i$  la  $i$ -ésima columna de  $A$  para  $i = 1, \dots, n$ . Si existe un vector  $w$  tal que  $w^t A_i = 1$  para toda  $i = 1, \dots, n$  se dice que  $A$  es una matriz homogénea.*

**OBSERVACIÓN 2.1.** *Se observa que cuando la matriz  $A$  es homogénea entonces el ideal tórico  $I_A$  es homogéneo, es decir, esta generado por polinomios cuyos monomios tienen el mismo grado.*

Gracias a éstas definiciones se puede postular la siguiente proposición:

**PROPOSICIÓN 2.2.** *Sea  $A$  un matriz homogénea con entradas naturales de  $d$  por  $n$ . Entonces la base universal de Markov de  $A$  está contenida en el conjunto de vectores semiprimitivos de  $A$ .*

Demostración

Sea  $u$  un vector  $A$ -semiprimitivo con descomposición graduada semiconforme  $u = v + w$ , entonces se tiene la siguiente identidad

$$\bar{x}^{u^+} - \bar{x}^{u^-} = \bar{x}^{u^+ - v^+} (\bar{x}^{v^+} - \bar{x}^{v^-}) + \bar{x}^{u^- - w^+} (\bar{x}^{w^+} - \bar{x}^{w^-})$$

Lo que significa que  $\bar{x}^{u^+} - \bar{x}^{u^-}$  se puede expresar como combinación de dos polinomios  $\bar{x}^{v^+} - \bar{x}^{v^-}$  y  $\bar{x}^{w^+} - \bar{x}^{w^-}$  de menor grado por lo que  $\bar{x}^{u^+} - \bar{x}^{u^-}$  no pertenece a una base de Markov mínima. ■

La siguiente es una proposición que intenta dar el recíproco de la anterior.

**PROPOSICIÓN 2.3.** *Sea  $A$  un matriz con entradas naturales de  $d$  por  $n$ . Entonces el conjunto de vectores en el  $\mathbb{Z}$ -núcleo de  $A$  que no tienen una descomposición semiconforme está contenido en cada base de Markov mínima de  $A$ .*

Demostración

Sea  $u$  un vector en el  $\mathbb{Z}$ -núcleo que no tenga una descomposición semiconforme y sea  $\mathcal{B}$  una base de Markov mínima de  $A$  tal que  $u$  no pertenezca. Entonces existe una sucesión de movimientos  $\{m_i\}_{i=1}^l$  en  $\mathcal{B}$  tales que conectan a  $u^+$  y a  $u^-$ , pero esto da  $u = m_1 + \sum_{i=2}^l m_i$  que es una descomposición semiconforme de  $u$ . ■

### 3. Bases de Markov de Modelos Reducibles

Como se mencionó anteriormente los modelos reducibles son de importancia por que se pueden llevar a modelos más sencillos, ahora se tomará un complejo simplicial abstracto  $\Delta$  y se verá como construir una base de Gröbner para cuando  $\Delta$  es reducible. Como se mencionó intuitivamente lo que se hará será resolver el problema para los modelos pequeños, es decir, si  $\Delta$  tiene una descomposición en  $\Delta_1$  y  $\Delta_2$  se encontrarán bases de Gröbner para  $\Delta_1$  y para  $\Delta_2$ . Esta técnica fue propuesta por Hosten y Sullivan en el 2002 en su artículo Gröbner bases and polyhedral geometry of reducible and cyclic models.

DEFINICIÓN 3.1. *Para un complejo simplicial abstracto  $\Delta$  en  $[n]$  reducible con  $\Delta_1$  y  $\Delta_2$ , al conjunto  $S \subseteq [n]$  tal que  $\Delta_1 \cap \Delta_2 = 2^S$  se le llama el separador de  $|\Delta_1| \setminus S$  y  $|\Delta_2| \setminus S$ , donde  $|\mathcal{K}|$  es el conjunto de vértices del complejo simplicial abstracto  $\mathcal{K}$ .*

Para  $f \in K[\bar{x}]$  un monomio, si es de la forma  $f = x_{K_1} \dots x_{K_t}$  con  $K_j = (k_{j1}, \dots, k_{jn})$  para  $j = 1, \dots, t$  se le puede asociar una tabla  $T$  de  $t$  por  $n$ , dada por  $T_{ij} = K_{ij}$  para todo  $i = 1, \dots, t$  y para todo  $j = 1, \dots, n$ . Si una variable esta elevada a la  $m$ -ésima potencia está en el monomio entonces aparecera  $m$  veces en la tabla  $T$ . Por ejemplo, para el monomio  $x_{000}x_{011}x_{101}x_{110}$  se tiene la tabla:

$$T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Como las variables conmutan entre sí, se tiene que dos tablas pueden representar al mismo monomio, esto es, el monomio pasado también se puede escribir como  $x_{011}x_{000}x_{110}x_{101}$ , en este caso se tendría una nueva tabla  $T'$  que sería:

$$T' = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Por esto se extiende el concepto de igualdad para tablas, en éste caso será: dos tablas  $T$  y  $T'$  de  $n$  por  $t$  son iguales si existe una permutación de los renglones de  $T$  que lleven a  $T'$ .

Un binomio  $f \in K[\bar{x}]$  se puede representar como la diferencia de tablas, por ejemplo,  $x_{000}x_{011}x_{101}x_{110} - x_{001}x_{010}x_{100}x_{111}$ , se puede ver como:

$$T - S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Esta diferencia solo se representa, no se puede operar, ya que no son matrices sino tablas. Se hace notar la diferencia ya que aunque se podría calcular la diferencia se perdería la información de que monomio es.

Si se tiene  $\mathbf{u}$  una  $d_1 \times \dots \times d_n$  tabla de contingencia y  $T$  su tabla asociada,  $F \subseteq [n]$  y  $\mathbf{u}_F$  la  $F$ -marginal de  $\mathbf{u}$ , entonces la tabla  $T|_F$  que representa  $\mathbf{u}_F$  se obtiene de eliminar las columnas de  $T$  indicadas por  $F$ .

**PROPOSICIÓN 3.1.** *Sea  $f$  un binomio representado por  $T - T'$  en la notación tabla. Entonces  $f$  está en  $I_\Delta$  si y sólo si  $T|_F = T'|_F$  para cada faceta  $F$  de  $\Delta$ .*

Demostración

Se llaman  $\mathbf{u}$  y  $\mathbf{v}$  a las tablas de contingencia de  $T$  y de  $T'$  respectivamente, por lo visto anteriormente  $f \in I_\Delta$  si y sólo si  $A_\Delta \mathbf{u} = A_\Delta \mathbf{v}$ , esto es, si y sólo si para cada faceta  $F$  de  $\Delta$  las  $F$ -marginales de  $\mathbf{u}$  y de  $\mathbf{v}$  coinciden y por lo observado anteriormente esto sucede si y sólo si  $T|_F = T'|_F$  para cada faceta  $F$  de  $\Delta$ . ■

**PROPOSICIÓN 3.2.** *Sea  $\Delta$  un complejo simplicial abstracto con  $\Delta = \{\{1\}, \{2\}\}$  que consiste de dos puntos aislados con  $\mathbf{d} = (d_1, d_2)$ . Entonces el conjunto de binomios de la forma*

$$\binom{i_1 \ j_2}{i_2 \ j_1} - \binom{i_1 \ j_1}{i_2 \ j_2}$$

con  $0 \leq i_1 < i_2 \leq d_1$  y  $0 \leq j_1 < j_2 \leq d_2$ , es una base de Gröbner del ideal  $I_\Delta$  con respecto al orden lexicográfico inverso.

Demostración

Véase Grobner Bases and Convex Polytopes de B. Sturmfels

■

DEFINICIÓN 3.2. Sea  $\Delta$  un complejo simplicial abstracto con exactamente dos facetas  $F_1$  y  $F_2$  y  $\mathbf{d} = (d_1, \dots, d_n)$ . Sea  $S = F_1 \cap F_2$  y se supone que  $F_1 \setminus S = \{1, \dots, a\}$ ,  $S = \{a+1, \dots, b\}$  y  $F_2 \setminus S = \{b+1, \dots, n\}$ . Se define  $Quad(F_1, F_2)$  como el conjunto de los binomios de la forma:

$$\binom{q_1 \ r \ s_1}{q_2 \ r \ s_2} - \binom{q_1 \ r \ s_2}{q_2 \ r \ s_1}$$

donde  $q_1, q_2 \in [d_1] \times \dots \times [d_a]$  con  $q_1 \neq q_2$ ,  $r \in [d_{a+1} - 1] \times \dots \times [d_b]$  y  $s_1, s_2 \in [d_{b+1}] \times \dots \times [d_n]$  con  $s_1 \neq s_2$ .

PROPOSICIÓN 3.3. Sea  $\Delta$  un complejo simplicial abstracto con exactamente dos facetas  $F_1$  y  $F_2$ . Entonces el conjunto  $Quad(F_1, F_2)$  es una base de Gröbner para  $I_\Delta$  con respecto al orden lexico gráfico.

Demostración

Si  $S = \emptyset$  el modelo es equivalente al modelo de dos puntos aislados de la proposición anterior con  $\Delta' = \{\{I\}, \{II\}\}$  con  $\mathbf{d}' = (d'_1, d'_2)$  donde  $d'_1 = d_1 \dots d_a$  y  $d'_2 = d_{a+1} \dots d_n$  por lo cual se tiene el resultado. Si  $S \neq \emptyset$  entonces se pueden reordenar sus columnas y renglones de la matriz  $A_\Delta$  para que se una matriz diagonal por bloques con bloques idénticos. Las columnas de cada bloque corresponden a las variables  $x_K$  con  $K = (k_1, \dots, k_n)$  con  $k_{a+1}, \dots, k_b$  fijos y este tipo de matrices es el del modelo de dos puntos aislados, por lo que se concluye que  $Quad(F_1, F_2)$  es la unión de bases Gröbner para modelos de dos puntos aislados que no comparten variables. ■

Ahora considérese el caso en el que  $\Delta$  es un complejo simplicial abstracto reducible con descomposición  $(\Delta_1, \Delta_2)$ ,  $\Delta_1 \cap \Delta_2 = 2^S$ ,  $\Delta_1 \setminus S = \{1, \dots, a\}$ ,  $S = \{a+1, \dots, b\}$  y  $\Delta_2 \setminus S = \{b+1, \dots, n\}$ . Se llamarán a  $\Delta_1$  y a  $\Delta_2$  los subcomplejos inducidos del complejo simplicial abstracto reducible  $\Delta$  y se denotan por  $\mathbf{d}_1$  y  $\mathbf{d}_2$  las partes correspondientes de  $\mathbf{d}$  con índices en  $|\Delta_1|$  y  $|\Delta_2|$ .

Se sabe que cualquier binomio en  $I_\Delta$  es homogéneo, en particular esto se puede seguir para los modelos en los que  $\Delta$  es un complejo simplicial abstracto descomponible en dos facetas por la proposición anterior, dado que se tiene una base de Gröbner de binomios se sigue que todo binomio es múltiplo de un binomio de la base y por lo tanto homogéneo. Interpretando esto en la notación

de tabla, todo binomio  $f$  se puede escribir como diferencia de dos tablas del mismo número de renglones, considerando que el binomio está en  $I_{\Delta_1}$ :

$$f = \begin{pmatrix} q_1 & r_1 \\ \vdots & \vdots \\ q_m & r_m \end{pmatrix} - \begin{pmatrix} q'_1 & r'_1 \\ \vdots & \vdots \\ q'_m & r'_m \end{pmatrix}$$

Como  $S$  es una cara de  $\Delta$ ,  $\Delta_1$  y  $\Delta_2$ , las tablas correspondientes deben tener las mismas  $S$ -marginales, por lo que se pueden permutar las columnas a modo de que  $r_i = r'_i$  para  $i = 1, \dots, m$ , por lo cual en notación tabla queda:

$$f = \begin{pmatrix} q_1 & r_1 \\ \vdots & \vdots \\ q_m & r_m \end{pmatrix} - \begin{pmatrix} q'_1 & r_1 \\ \vdots & \vdots \\ q'_m & r_m \end{pmatrix}$$

Ésto se puede hacer también para el ideal  $I_{\Delta_2}$ . Si se considera  $H$  un subconjunto de binomios de  $I_{\Delta_1}$ , cualquier elemento  $f$  se puede escribir como se describió anteriormente, se toma  $\bar{s} = (s_1, \dots, s_m) \in [d_{b+1}] \times \dots \times [d_n]$  y se crea un nuevo binomio  $f_{\bar{s}}$  dado por la tabla:

$$f_{\bar{s}} = \begin{pmatrix} q_1 & r_1 & s_1 \\ \vdots & \vdots & \vdots \\ q_m & r_m & s_m \end{pmatrix} - \begin{pmatrix} q'_1 & r_1 & s_m \\ \vdots & \vdots & \vdots \\ q'_m & r_m & s_m \end{pmatrix}$$

A continuación se define  $Ext(H, I_{\Delta}^{\Delta_1})$  como el conjunto de todos los  $f_{\bar{s}}$  donde  $f \in I_{\Delta_1}$  es un binomio y  $\bar{s} \in [d_{b+1}] \times \dots \times [d_n]$ . De manera análoga se define  $Ext(H, I_{\Delta}^{\Delta_2})$  para  $H$  un conjunto de binomios en  $I_{\Delta_2}$ , como el conjunto de todos los  $f_{\bar{s}}$  donde  $f \in I_{\Delta_2}$  es un binomio y  $\bar{s} \in [d_1] \times \dots \times [d_a]$ .

**PROPOSICIÓN 3.4.** *Sean  $\Delta$  un complejo simplicial abstracto reducible con descomposición  $(\Delta_1, S, \Delta_2)$ ,  $H$  un conjunto de binomios en  $I_{\Delta_1}$  y  $H'$  un conjunto de binomios en  $I_{\Delta_2}$ . Entonces  $Ext(H, I_{\Delta}^{\Delta_1}) \subseteq I_{\Delta}$  y  $Ext(H', I_{\Delta}^{\Delta_2}) \subseteq I_{\Delta}$ .*

*Demostración*

Sólo basta con demostrar un caso, el otro es análogo. Sea  $f \in Ext(H, I_{\Delta}^{\Delta_1})$  por lo que  $f = T - T'$ , para que  $f \in I_{\Delta}$  basta ver que las  $F$ -marginales de

$T$  y  $T'$  son iguales para cada faceta  $F$  del complejo simplicial abstracto  $\Delta$ , o lo que es equivalente  $f|_F = 0$  para cada faceta  $F$  de  $\Delta$ . Por como es  $\Delta$ , cada una de sus facetas ésta en  $\Delta_1$  ó en  $\Delta_2$  (no en ambos). Sea  $F$  una faceta del complejo simplicial abstracto  $\Delta$ . Si  $F \in \Delta_1$ , entonces  $f|_F = (f|_{\Delta_1})|_F$  y como  $f|_{\Delta_1} \in I_{\Delta_1}$ , se sigue que  $f|_F = 0$ . De la misma forma si  $F \in \Delta_2$ , entonces  $f|_F = 0$ . ■

Se hace ver que tanto  $I_{\Delta_1}$  como  $I_{\Delta_2}$  se pueden considerar encajados en anillos de polinomios más pequeños, de hecho, se considerará el anillo con la menor cantidad de indeterminadas, los cuales se denotarán por  $K[\bar{x}]_1$  y  $K[\bar{x}]_2$ .

**PROPOSICIÓN 3.5.** Sean  $\leq_1$  un orden monomial en  $K[\bar{x}]_1$ ,  $\leq_2$  un orden monomial en  $K[\bar{x}]_2$  y  $\leq$  un orden monomial en  $K[\bar{x}]$ . Entonces la relación  $\leq^*$  dada por

$$T \leq^* T' = \begin{cases} T|_{\Delta_1} <_1 T'|_{\Delta_1} \\ T|_{\Delta_1} = T'|_{\Delta_1} \text{ y } T|_{\Delta_2} <_2 T'|_{\Delta_2} \\ T|_{\Delta_1} = T'|_{\Delta_1} \text{ , } T|_{\Delta_2} = T'|_{\Delta_2} \text{ y } T < T' \end{cases}$$

es un orden monomial.

Demostración

Sean  $T$  y  $T'$  tablas, éstas caen en alguno de los tres casos, ya que lo peor que podría pasar sería que las dos restricciones fuesen iguales, por lo cual se compararían con respecto en  $\leq$  que es un orden total. Es fácil ver que el 1 es el menor elemento de  $\leq^*$  por que  $\leq_1$ ,  $\leq_2$  y  $\leq$  son ordenes monomiales, por la misma razón preservan el producto. Para la transitividad, sean  $T \leq^* T' \leq^* T''$  tablas. Se tienen los siguientes casos:

- Si  $T|_{\Delta_1} <_1 T'|_{\Delta_1}$  y  $T'|_{\Delta_1} <_1 T''|_{\Delta_1}$ . Como  $\leq_1$  es un orden monomial entonces  $T|_{\Delta_1} <_1 T''|_{\Delta_1}$ . Por lo que  $T <^* T''$ .
- Si  $T|_{\Delta_1} <_1 T'|_{\Delta_1}$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$  y  $T'|_{\Delta_2} <_2 T''|_{\Delta_2}$ . Entonces  $T|_{\Delta_1} <_1 T''|_{\Delta_1}$ , por lo que  $T <^* T''$ .

- Si  $T|_{\Delta_1} <_1 T'|_{\Delta_1}$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$ ,  $T'|_{\Delta_2} = T''|_{\Delta_2}$  y  $T' < T''$ . Entonces  $T'|_{\Delta_1} <_1 T''|_{\Delta_1}$ , por lo que  $T <^* T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} <_2 T'|_{\Delta_2}$  y  $T'|_{\Delta_1} <_1 T''|_{\Delta_1}$ . Entonces  $T|_{\Delta_1} <_1 T''|_{\Delta_1}$  de donde  $T <^* T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} <_2 T'|_{\Delta_2}$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$  y  $T'|_{\Delta_2} <_2 T''|_{\Delta_2}$ . Entonces  $T|_{\Delta_1} = T''|_{\Delta_1}$  y  $T|_{\Delta_2} <_2 T''|_{\Delta_2}$  de donde  $T <^* T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} <_2 T'|_{\Delta_2}$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$ ,  $T'|_{\Delta_2} = T''|_{\Delta_2}$  y  $T' < T''$ . Entonces  $T|_{\Delta_1} = T''|_{\Delta_1}$  y  $T|_{\Delta_2} <_2 T''|_{\Delta_2}$  de donde  $T <^* T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} = T'|_{\Delta_2}$ ,  $T < T'$  y  $T'|_{\Delta_1} <_1 T''|_{\Delta_1}$ . Entonces  $T|_{\Delta_1} <_1 T''|_{\Delta_1}$  de aquí  $T < T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} = T'|_{\Delta_2}$ ,  $T < T'$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$  y  $T'|_{\Delta_2} <_2 T''|_{\Delta_2}$ . Entonces  $T|_{\Delta_1} = T''|_{\Delta_1}$  y  $T|_{\Delta_2} <_2 T''|_{\Delta_2}$  de aquí  $T < T''$ .
- Si  $T|_{\Delta_1} = T'|_{\Delta_1}$ ,  $T|_{\Delta_2} = T'|_{\Delta_2}$ ,  $T < T'$ ,  $T'|_{\Delta_1} = T''|_{\Delta_1}$ ,  $T'|_{\Delta_2} = T''|_{\Delta_2}$  y  $T' < T''$ . Entonces  $T|_{\Delta_1} = T''|_{\Delta_1}$ ,  $T|_{\Delta_2} = T''|_{\Delta_2}$  y  $T < T''$  de aquí  $T <^* T''$ .

Por lo tanto el orden es transitivo. ■

**PROPOSICIÓN 3.6.** Sean  $G_1$  base de Gröbner de  $I_{\Delta_1}$  con respecto a un orden monomial  $\leq_1$  y  $G_2$  base de Gröbner de  $I_{\Delta_2}$  con respecto a un orden monomial  $\leq_2$  y sea  $\leq$  el orden lexicográfico en  $\text{Quad}(F_1, F_2)$ . Entonces el conjunto  $G = \text{Ext}(G_1, I_{\Delta_1}^{\Delta_1}) \cup \text{Ext}(G_2, I_{\Delta_2}^{\Delta_2}) \cup \text{Quad}(F_1, F_2)$  es una base de Gröbner para  $I_{\Delta}$  con respecto al orden monomial  $\leq^*$  descrito en la proposición pasada.

Demostración

El camino de la demostración es lograr ver que para cualquier monomio  $f = T - T' \in I_{\Delta}$  con término principal  $T$  con respecto a  $\leq^*$  (esto se hace sin pérdida de generalidad ya que si  $T$  no es el término principal lo es  $T'$  y la demostración se hace análoga), existe  $g \in G$  con término principal que divide a  $T$ . Se ve a  $f$  en notación tabla como:

$$f = T - T' = \begin{pmatrix} q_1 & r_1 & s_1 \\ \vdots & \vdots & \vdots \\ q_m & r_m & s_m \end{pmatrix} - \begin{pmatrix} q'_1 & r_1 & s'_1 \\ \vdots & \vdots & \vdots \\ q'_m & r_m & s'_m \end{pmatrix}$$

La demostración va por casos.

Primer caso si  $T|_{\Delta_1} \neq T'|_{\Delta_2}$ . Esto implica que el término principal de  $f|_{\Delta_1}$  con respecto a  $\leq_1$  es  $T|_{\Delta_1}$ , ya que en caso de no serlo lo sería  $T'|_{\Delta_1}$  lo que implicaría que  $T'$  es el término principal en  $f$  respecto a  $\leq^*$ . Como  $G_1$  es una base de Gröbner para  $I_{\Delta_1}$  con respecto a  $\leq_1$ , existe  $g \in G$  cuyo término principal respecto a  $\leq_1$  divide a  $T|_{\Delta_1}$ . Después de reordenar los renglones de  $f$  (esto se hace sólo para quitar los últimos  $m - i$  renglones y no andar quitando renglones intermedios), se puede poner a  $g$  de la siguiente forma en notación tabla:

$$g = U - U' = \begin{pmatrix} q_1 & r_1 \\ \vdots & \vdots \\ q_i & r_i \end{pmatrix} - \begin{pmatrix} q''_1 & r_1 \\ \vdots & \vdots \\ q''_i & r_i \end{pmatrix}$$

con  $U$  el término principal. Entonces el binomio:

$$g' = V - V' = \begin{pmatrix} q_1 & r_1 & s_1 \\ \vdots & \vdots & \vdots \\ q_i & r_i & s_i \end{pmatrix} - \begin{pmatrix} q''_1 & r_1 & s_1 \\ \vdots & \vdots & \vdots \\ q''_i & r_i & s_i \end{pmatrix}$$

pertenece a  $Ext(G_1, I_{\Delta_1}^{\Delta_1})$  y tiene término principal  $V$  que divide a  $T$ .

El segundo caso es cuando  $T|_{\Delta_1} = T'|_{\Delta_1}$  y  $T|_{\Delta_2} \neq T'|_{\Delta_2}$ , pero es análogo al que se acaba de mostrar. El tercer y último caso es cuando  $T|_{\Delta_1} = T'|_{\Delta_1}$  y  $T|_{\Delta_2} = T'|_{\Delta_2}$ . Entonces  $f$  no solo pertenece a  $I_{\Delta}$  si no también a  $I_{\Delta^*}$  donde  $\Delta^*$  es el complejo simplicial abstracto con exactamente dos facetas  $F_1$  y  $F_2$ . Pero  $Quad(F_1, F_2)$  es una base de Gröbner para  $I_{\Delta^*}$  por lo que  $T$  es divisible por el término principal de algún binomio de  $Quad(F_1, F_2)$ . Por lo tanto  $G$  es una base de Gröbner para  $I_{\Delta}$  con respecto a  $\leq^*$ . ■

**COROLARIO 3.1.** Sean  $H_1$  un conjunto de binomios que genera  $I_{\Delta_1}$  y  $H_2$  un conjunto de binomios que genera a  $I_{\Delta_2}$ . Entonces  $H = Ext(H_1, I_{\Delta_1}^{\Delta_1}) \cup Ext(H_2, I_{\Delta_2}^{\Delta_2}) \cup Quad(F_1, F_2)$ .

**COROLARIO 3.2.** Sean  $n_1$  el grado máximo de un generador mínimo en  $I_{\Delta_1}$  y  $n_2$  el grado máximo de un generador máximo en  $I_{\Delta_2}$ . Entonces el grado máximo  $n$  de un generador mínimo es  $n = \max\{2, n_1, n_2\}$ .

Demostración

La demostración se sigue de que  $Ext$  preserva el grado de los polinomios. ■

**COROLARIO 3.3.** Sea  $\Delta$  un complejo simplicial abstracto descomponible. Entonces  $I_{\Delta}$  tiene una base de Markov que consiste en binomios de grado dos.

Demostración

La demostración se hace por inducción en el número de facetas de  $\Delta$ . Si  $\Delta$  tiene una faceta, el ideal  $I_{\Delta}$  es el ideal cero. Si  $\Delta$  tiene dos facetas  $F_1$  y  $F_2$  el ideal  $I_{\Delta}$  está generado por  $Quad(F_1, F_2)$  que consiste de binomios de grado dos. Si no,  $\Delta$  es reducible y por hipótesis de inducción sobre  $I_{\Delta_1}$  e  $I_{\Delta_2}$  (ya que ambas tienen menos facetas) y aplicando el corolario anterior  $I_{\Delta}$  tiene una base de Markov de binomios de grado dos. ■

#### 4. Un Ejemplo de Modelos Reducibles

Se considera el simplejo simplicial abstracto dado por las facetas  $\Delta = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$  y  $\mathbf{d} = (2, 2, 2, 2)$ . El complejo es reducible con descomposición  $\Delta_1 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$  y  $\Delta_2 = \{\{2, 3\}, \{2, 4\}, \{3, 4\}\}$ . Ahora  $\Delta_1$  y  $\Delta_2$  son el modelo de un tres ciclos y se observó anteriormente su ideal tórico es un ideal cíclico. Que de hecho es generado por el polinomio que en notación tabla es:

$$T - T' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Si se le aplica el  $Ext$  al conjunto unitario  $\{T - T'\}$ , entonces  $Ext(\{T - T'\}, I_{\Delta}^{\Delta_1})$  consiste de dieciséis elementos cuya forma en notación tabla es:

$$\begin{pmatrix} 0 & 0 & 0 & r_1 \\ 0 & 1 & 1 & r_2 \\ 1 & 0 & 1 & r_3 \\ 1 & 1 & 0 & r_4 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 & r_1 \\ 0 & 1 & 0 & r_2 \\ 1 & 0 & 0 & r_3 \\ 1 & 1 & 1 & r_4 \end{pmatrix}$$

donde  $r_1, r_2, r_3, r_4 = 0, 1$ . De manera análoga  $Ext(\{T - T'\}, I_{\Delta}^{\Delta_2})$  también consta de dieciséis elementos que en notación tabla son:

$$T - T' = \begin{pmatrix} r_1 & 0 & 0 & 0 \\ r_2 & 0 & 1 & 1 \\ r_3 & 1 & 0 & 1 \\ r_4 & 1 & 1 & 0 \end{pmatrix} - \begin{pmatrix} r_1 & 0 & 0 & 1 \\ r_2 & 0 & 1 & 0 \\ r_3 & 1 & 0 & 0 \\ r_4 & 1 & 1 & 1 \end{pmatrix}$$

donde  $r_1, r_2, r_3, r_4 = 0, 1$ . Por último se observa que  $Quad(\{1, 2, 3\}, \{2, 3, 4\})$  consiste de cuatro binomios de la forma (en notación tabla):

$$\begin{pmatrix} 0 & r_1 & r_2 & 0 \\ 1 & r_1 & r_2 & 1 \end{pmatrix} - \begin{pmatrix} 0 & r_1 & r_2 & 1 \\ 1 & r_1 & r_2 & 0 \end{pmatrix}$$

con  $r_1, r_2 = 0, 1$ . La unión de estos tres conjuntos como es una base de Gröbner de  $I_{\Delta}$ , aunque no necesariamente es una base reducida.

## 4

---

# Aplicaciones

---

En este capítulo se presentan algunas aplicaciones. Se considera la siguiente información, que es el mes nacimiento con respecto al mes de muerte de 82 descendientes de la Reina Victoria, la tabla se puede encontrar en (Andrews and Herzberg, 1985).

Relación entre fecha de nacimiento y fecha de muerte													
Mes de nacimiento	Mes de muerte												Total
	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic	
Ene	1	0	0	0	1	2	0	0	1	0	1	0	6
Feb	1	0	0	1	0	0	0	0	0	1	0	2	5
Mar	1	0	0	0	2	1	0	0	0	0	0	1	5
Abr	3	0	2	0	0	0	1	0	1	3	1	1	12
May	2	1	1	1	1	1	1	1	1	1	1	0	12
Jun	2	0	0	0	1	0	0	0	0	0	0	0	3
Jul	2	0	2	1	0	0	0	0	1	1	1	2	10
Ago	0	0	0	3	0	0	1	0	0	1	0	2	7
Sep	0	0	0	1	1	0	0	0	0	0	1	0	3
Oct	1	1	0	2	0	0	1	0	0	1	1	0	7
Nov	0	1	1	1	2	0	0	2	0	1	1	0	9
Dic	0	1	1	0	0	0	1	0	0	0	0	0	3
Total	13	4	7	10	8	4	5	3	4	9	7	8	82

### 1. La prueba $\chi^2$ de Pearson

La prueba  $\chi^2$  de Pearson tiene dos usos principales, el primero es como prueba de bondad de ajuste y el segundo como prueba de independencia. En este caso el segundo uso es el que interesa.

Se empieza por considerar una tabla de contingencia de  $r \times c$ . La entrada  $ij$  que tiene la frecuencia observada se denota por  $n_{ij}$  con  $1 \leq i \leq r$  y  $1 \leq j \leq c$ . Para indicar la suma del  $i$ -ésimo renglón se usa  $n_{i+} := \sum_{j=1}^c n_{ij}$ . Para indicar la suma de la  $j$ -ésima columna se usa  $n_{+j} := \sum_{i=1}^r n_{ij}$ . Para indicar el total de observaciones se usa  $N = \sum_{i=1}^r n_{i+} = \sum_{j=1}^c n_{+j}$ .

La probabilidad de pertenecer a la clase  $ij$  se denota por  $p_{ij}$ . Obviamente, esto lleva a que la frecuencia esperada de la clase  $ij$ ,  $E_{ij}$ , sea  $Np_{ij}$ . Bajo el supuesto de que las frecuencias observadas siguen una distribución multinomial. Para hacer la notación intuitiva,  $p_{+j}$  es la probabilidad de pertenecer a la clase  $j$  y  $p_{i+}$  es la probabilidad de pertenecer a la clase  $i$ .

Que la variable aleatoria representada por las clases de las columnas sea independiente de la variable aleatoria representada por los renglones, quiere decir, que numéricamente se tiene  $p_{ij} = p_{i+}p_{+j}$  para toda  $1 \leq i \leq r$  y para toda  $1 \leq j \leq c$ . Aunque hay que recordar que estas probabilidades son desconocidas se pueden estimar a partir de sus frecuencias y sus mejores estimadores (estimadores máximos verosímiles) que están dados por  $\hat{p}_{ij} = \frac{n_{ij}}{N}$ ,  $\hat{p}_{i+} = \frac{n_{i+}}{N}$  y  $\hat{p}_{+j} = \frac{n_{+j}}{N}$ .

La estimación de  $p_{i+}$  y de  $p_{+j}$  permite evaluar la frecuencia esperada de la clase  $ij$ . Ésta está dada por  $E_{ij} = N\hat{p}_{i+}\hat{p}_{+j} = \frac{n_{i+}n_{+j}}{N}$ . Como es de esperarse, aunque las variables sean independientes no tiene por qué cumplirse que  $\hat{p}_{ij} = \hat{p}_{i+}\hat{p}_{+j}$ , pues son sólo estimaciones y no tienen que coincidir con las probabilidades reales. Por otro lado, en caso de ser independientes las variables la frecuencia estimada y la frecuencia observada sólo debe diferir por errores de aleatoriedad. Por lo mismo, en caso de no ser independientes, la diferencia antes mencionada se esperará que sea grande.

En 1904, Pearson propone como prueba de independencia de una tabla de contingencia el uso del estadístico:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

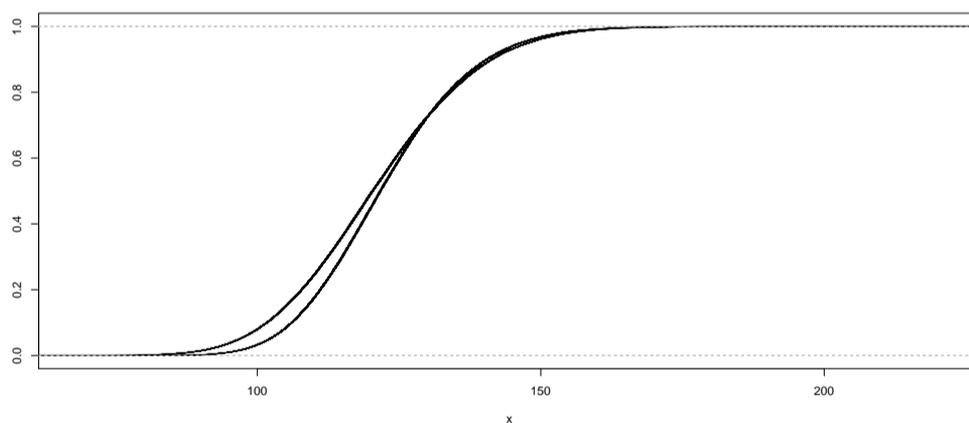
Como es de esperarse la hipótesis nula  $H_0$  es:

$$p_{ij} = p_{i+}p_{+j}$$

Se quiere un metodo para poder decidir para qué valores de  $\chi^2$  se debe aceptar  $H_0$  y para cuáles valores se debe rechazar  $H_0$ . El metodo se basa en deducir una distribución de probabilidad para  $\chi^2$  bajo el supuesto de que las variables son independientes. La aceptación o el rechazo de  $H_0$  está basado en la probabilidad de obtener el valor  $\chi^2$ , para valores con baja probabilidad se rechazará  $H_0$  y para los otros se aceptará. En general, baja probabilidad se toma como 0.05 o 0.01, y este número se le llama el valor de significancia de la prueba.

Bajo el supuesto de que las frecuencias observadas tienen distribución multinomial y que las frecuencias esperadas no son muy pequeñas, el estadístico  $\chi^2$  tiene distribución aproximada ji cuadrada. Así, es posible una prueba a la hipótesis de independencia al comparar los valores del estadístico  $\chi^2$  y los valores de la distribución ji cuadrada.

El supuesto de que las frecuencias esperadas no sean muy pequeñas se enuncia matemáticamente como que el valor mínimo en cada celda es 5. En el ejemplo presentado este supuesto es violado gravemente, no hay celda que tenga valor al menos de 5 o más. Si se hace una comparación entre la distribución por permutaciones del estadístico ji cuadrada contra la distribución de la ji cuadrada se puede ver que no es una buena aproximación como muestra la gráfica. La probabilidad muestral de que  $\chi^2 \leq 115.6$  es 0.3208 contra 0.3775. Por lo que el metodo no se puede utilizar, para esto hay que proponer un nuevo metodo.



Para desarrollar este nuevo metodo, se considera generar aleatoriamente una tabla de contingencia con suma de renglones fija y con suma de columnas fija. Para dos enteros positivos  $I$  y  $J$ , se toman dos vectores de enteros positivos

$\mathbf{r} = (r_1, \dots, r_I)$  y  $\mathbf{c} = (c_1, \dots, c_J)$ . Se pone  $\mathcal{F}(\mathbf{r}, \mathbf{c})$  como el conjunto de todas las tablas de contingencia de  $I$  por  $J$  con suma de renglones  $\mathbf{r}$  y con suma de columnas  $\mathbf{c}$ . Si se pone  $N = \sum_{i=1}^I c_i = \sum_{i=1}^J r_i$ , para  $X \in \mathcal{F}(\mathbf{r}, \mathbf{c})$ , se tiene:

$$H(X) = \frac{\prod_{j=1}^J \binom{c_j}{X_{1j} \dots X_{Ij}}}{\binom{N}{r_1 \dots r_I}}$$

la distribución hipergeométrica en  $\mathcal{F}(\mathbf{r}, \mathbf{c})$ . Ésta resulta ser la distribución condicional de la información dado el estadístico suficiente para el modelo clásico de independencia.

Se puede usar un método Monte Carlo para generar desde  $H$  como sigue: Si  $X$  es una tabla de contingencia que está en  $\mathcal{F}(\mathbf{r}, \mathbf{c})$ , se pueden elegir una pareja de columnas y una pareja de renglones. Éstas se intesectan en 4 entradas y se modifica  $X$  de forma aleatoria como:

$$\begin{array}{cc} + & - \\ - & + \end{array} \text{ o } \begin{array}{cc} - & + \\ + & - \end{array} \text{ con probabilidad } \frac{1}{2} \text{ cada una}$$

Sumando o restando 1 según sea el caso. Por la forma que se hace no se cambia la suma de las columnas ni la suma de los renglones. Si en un paso del proceso alguna entrada resulta negativa entonces se ignora y se repite. Esto resulta una cadena de Markov en  $\mathcal{F}(\mathbf{r}, \mathbf{c})$ . Esto es una forma de hacer el algoritmo de Metropolis, y como es usual la cadena resultante es conexa, aperiódica y reversible con distribución estacionaria  $H$ . En la simulación de esta cadena en la tabla de contingencia de nacimiento y muerte de los descendientes de la reina Victoria, se puede ejecutar la cadena con una longitud de 1,000,000 en 707.18 segundos. En general este método es muy eficiente a comparación de otros en 2-tablas de contingencia, pero para  $k$ -tablas de contingencia con  $k \geq 3$  parece ser el único.

Lo anterior puede hacerse de forma más general, sea  $\mathcal{H}$  un conjunto finito. Se considera la familia exponencial,

$$P_\theta(x) = Z(\theta)e^{\langle \theta, T(x) \rangle}$$

donde  $\theta \in \mathbb{R}^d$ ,  $Z(\theta)$  es una constante normalizante y  $T: \mathcal{H} \rightarrow \mathbb{N}^d \setminus \{(0, \dots, 0)\}$  una función, aquí los naturales empiezan desde cero.

Si  $X_1, \dots, X_N$  son variables aleatorias independientes e idénticamente distribuidas, de lo anterior el estadístico  $t = T(X_1) + \dots + T(X_N)$  es suficiente para  $\theta$ , esto se sigue del teorema de factorización de Fisher-Neyman.

Sea

$$\mathcal{Y}_t = \{(x_1, \dots, x_N) \in \mathcal{H}^d \mid T(x_1) + \dots + T(x_N) = t\}$$

Bajo la familia exponencial descrita anteriormente la ley de  $X_1, \dots, X_n$  dado  $t$  se distribuye uniforme sobre  $\mathcal{Y}_t$ . Evidentemente es un problema enumerar los elementos de  $\mathcal{Y}_t$  de manera efectiva o hacer un muestreo de la distribución uniforme de  $\mathcal{Y}_t$ .

Este problema, usualmente, se plantea en términos de la distribución hipergeométrica como sigue: se escribe

$$t = \sum_{i=1}^N T(X_i) = \sum_{i=1}^N G(x)T(x)$$

donde  $G(x) = |\{i = 1 \dots N \mid X_i = x\}|$ . La función  $G$  resulta ser un estadístico suficiente para cualquier información independiente e idénticamente distribuida. Se define el conjunto de todos los conjuntos de información con estadístico suficiente dado como:

$$\mathcal{F}_t = \{f: \mathcal{H} \rightarrow \mathbb{N} \mid \sum_x f(x)T(x) = t\}$$

Como  $T(x)$  no es cero para todo  $x$ , siempre se puede encontrar un conjunto de información con  $t$  suficiente. Claramente,  $\mathcal{F}_t$  es finito y no vacío. La imagen de la distribución uniforme en  $\mathcal{Y}_t$  bajo la función que va de  $\mathcal{Y}_t$  a  $\mathcal{F}_t$  es llamada la distribución hipergeométrica, dada por:

$$H_t(f) = \frac{N!}{|\mathcal{Y}_t|} \prod_x \frac{1}{f(x)!}$$

para toda  $f \in \mathcal{F}_t$ . Por lo que el problema se reduce a hacer muestreo de  $H_t$  en  $\mathcal{F}_t$ , dado  $t$ .

Las tablas de contingencia son el caso particular  $\mathcal{H} = \{(i, j) \mid 1 \leq i \leq I, 1 \leq j \leq J\}$ . El modelo usual de independencia para tablas de contingencia es  $T(i, j) \in \mathbb{N}^{I+J}$ , un vector de longitud  $I + J$  con dos entradas igual a 1 y el resto igual a 0. Las entradas que son 1, están en  $i$  de las primeras  $I$  entradas y en  $j$  de las últimas  $J$  entradas. El estadístico suficiente  $t$  contiene la suma de las columnas y la suma de los renglones de la tabla de contingencia asociada con  $N$  observaciones. El conjunto  $\mathcal{F}_t$  resulta ser todas las tablas de contingencia de  $I$  por  $J$  con esa suma de renglones y con esa suma de columnas.

## 2. Cadenas de Markov

Una base de Markov es un conjunto de funciones  $f_1, \dots, f_L: \mathcal{H} \rightarrow \mathbb{Z}$  tales que:

$$\sum_{x \in \mathcal{H}} f_i(x)T(x) = 0, \quad 1 \leq i \leq L$$

y para todo  $t$  y  $f, f' \in \mathcal{F}_t$  existen  $(\epsilon_1, f_{i_1}), \dots, (\epsilon_A, f_{i_A})$  con  $\epsilon_i = \pm 1$  tales que,

$$f' = f + \sum_{j=1}^A \epsilon_j f_{i_j}$$

y

$$f + \sum_{j=1}^a \epsilon_j f_{i_j} \geq 0, \quad 1 \leq a \leq A$$

Este nuevo concepto de base de Markov es el mismo que el que ya se había presentado. Las bases de Markov son importantes porque permiten construir cadenas de Markov.

Para  $f \in \mathcal{F}_t$ , se eligen  $I$  uniformemente de entre  $\{1, \dots, L\}$  y  $\epsilon$  de entre  $\{-1, 1\}$ . Si  $f + \epsilon f_I \geq 0$  la cadena se mueve ahí, en otro caso, permanece en  $f$ . La primera propiedad de ser base de Markov asegura que  $f + \epsilon f_I \in \mathcal{F}_t$ . La

segunda propiedad de ser base de Markov asegura que la cadena de Markov es conexa. La cadena se modifica para tener una distribución  $H_t$  con una muestra con distribución Bernoulli  $1/2$  para hacer otro paso en el proceso Metropolis.

LEMA 2.1. *Sea  $\sigma$  una función positiva en  $\mathcal{F}_t$  y una base de Markov  $f_1, \dots, f_L$ . Se genera la cadena de Markov eligiendo  $I$  uniformemente entre  $\{1, \dots, L\}$  y  $\epsilon = \pm$  con probabilidad  $1/2$ . Si la cadena esta en  $g \in \mathcal{F}_t$  y  $g + f_I\epsilon$  es no negativo, se mueve a  $g + f_I\epsilon$  con probabilidad*

$$\min\left\{\frac{\sigma(g + f_I\epsilon)}{\sigma(g)}, 1\right\}$$

*La cadena resulta ser conexa, reversible y aperiodica con distribución proporcional a  $\sigma$ .*

Un ejemplo muy útil de funciones en  $\mathcal{F}_t$  consiste en elegir una familia de funciones  $\{\omega_x: \mathbb{N} \rightarrow \mathbb{R}^+\}_{x \in \mathcal{H}}$ . Para  $g \in \mathcal{F}_t$ , se pone  $\sigma(g) = \prod_{x \in \mathcal{H}} \omega_x(g(x))$ . Por ejemplo, si  $\omega_x(a) = \frac{\theta_x^a}{a!}$  con  $0 < \theta_x \leq 1$ , de aquí  $\sigma$  resulta ser la distribución hipergeométrica múltiple. Cuando  $\theta_x = 1$  para toda  $x \in \mathcal{H}$ ,  $\sigma$  resulta ser la distribución hipergeométrica mencionada usual.

A continuación se toma el siguiente ejemplo de una tabla de contingencia de 4 por 4 (Snell 1974).

Relación entre color de cabello y color de ojos					
Color de Ojos	Color de Cabello				Total
	Negros	Castaño	Rojo	Rubio	
Café	68	119	26	7	220
Azul	20	84	17	94	215
Avellana	15	54	14	10	93
Verde	5	29	14	16	64
Total	108	286	71	127	592

El estadístico ji cuadrada para esta tabla es 138.29 con 9 grados de libertad para la distribución ji cuadrada. Diaconis y Efron (1985) se interesaron en la distribución de  $\chi^2$  bajo la hipótesis de que  $\mathcal{F}_t$  tiene distribución uniforme, el caso cuando  $\sigma = 1$ . Bajo un trabajo largo y arduo determinaron una estimación de aproximadamente 18.31 % de las tablas que tienen la misma suma de columnas y misma suma de renglones que la tabla anterior cumplen que  $\chi^2 \leq 138.29$ .

Esto se hizo usando métodos asintóticos y métodos Monte Carlo. No se usó el algoritmo Metropolis, solamente el procedimiento de sumar y restar descrito anteriormente. Para saber el número total de estas tablas, se usó el trabajo de Des Jardins donde se demostró que existen 1, 225, 914, 276, 276, 768, 514 con esta suma de renglones y esta suma de columnas.

**LEMA 2.2.** *Sea  $f_1, \dots, f_L$  una base de Markov en  $\mathcal{F}_t$ . Se genera una cadena de Markov eligiendo  $I$  uniformemente entre  $\{1, \dots, L\}$ . Si la cadena está en  $g \in \mathcal{F}_t$ , se determinan el conjunto de todos los enteros  $j$  tales que  $g + jf_I \in \mathcal{F}_t$ . Se elige  $j$  de este conjunto con probabilidad proporcional a*

$$\prod_{x \in C_I} \frac{1}{[g(x) + jf_I(x)]!}$$

con  $C_I = \{x \mid f_I(x) \neq 0\}$ . Entonces la cadena resulta ser conexa, reversible, aperiódica en  $\mathcal{F}_t$  con distribución estacionaria  $H_t$ .

El algoritmo de este lema se puede usar para medidas generales reemplazando el denominador por  $\prod_{x \in C_I} \frac{1}{\sigma(g(x) + jf_I(x))}$ . Este algoritmo toma pasos más grandes que los descritos en el algoritmo anterior.

### 3. Tasa de convergencia

Las cadenas de Markov descritas en la sección pasada pueden tardar de converger a su distribución estacionaria. Hay una teoría general para calcular la tasa de convergencia de estas cadenas. Para este tipo de cadenas, la teoría dice que se necesita un total  $\gamma^2$  pasos para garantizar la convergencia en variación total. Aquí  $\gamma$  es el diámetro de la gráfica inducida, que tiene por vértices los puntos del espacio de estados, en este caso,  $\mathcal{F}_t$  y vértices de  $f$  a  $g$  si  $g$  se puede alcanzar desde  $f$  en un paso.

**TEOREMA 3.1 (Diaconis).** *Para  $I$  y  $J$  enteros positivos,  $\mathbf{r} = (r_1, \dots, r_I)$  y  $\mathbf{c} = (c_1, \dots, c_J)$  con  $\sum_j c_j = \sum_i r_i = N$ . Sea  $\mathcal{F}(\mathbf{r}, \mathbf{c})$  el conjunto de todas las tablas de contingencia de  $I$  por  $J$  con suma de renglones  $\mathbf{r}$  y con suma de columnas  $\mathbf{c}$ . Sea  $U$  la distribución uniforme en  $\mathcal{F}(\mathbf{r}, \mathbf{c})$ . Sea  $K(x, y)$  la caminata inducida por  $\pm \mp$  los movimientos descritos anteriormente. Entonces:*

$$\|K_x^k - U\|_{TV} \leq A_1 e^{-A_2 c}$$

para  $k = c\gamma^2$  y  $c > 0$ . Aquí  $x \in \mathcal{F}(\mathbf{r}, \mathbf{c})$  es cualquier estado inicial y  $A_1, A_2$  son constantes explícitas que dependen de  $I$  y  $J$  y no de  $\mathbf{r}$  y  $\mathbf{c}$ . Más aún,  $\gamma$  el diametro de la gráfica cumple  $\gamma \leq \frac{N}{2}$ . Conversamente, existe  $x \in \mathcal{F}(\mathbf{r}, \mathbf{c})$  y existen constantes  $A_3, A_4$  como antes que satisfacen:

$$\|K_x^k - U\|_{TV} \leq A_3 e^{-A_4 c}$$

para  $k = c\gamma^2$ .

El teorema muestra que son suficientes y necesarios  $\gamma^2$  pasos para conseguir la estacionariedad. Las constantes  $A_i$  crecen de forma exponencial, poniendo  $D = IJ$ , las  $A_i$  son aproximadamente  $(\frac{D}{4})^{\frac{D}{4}}$ . Para tablas pequeñas, como la tabla de color de cabello y color de ojos, resulta una tasa razonable, para esta se necesitan alrededor de 100,000 pasos.

#### 4. No interacción en tres factores

Hasta ahora los ejemplos tenían solamente 2-tablas. Ahora se verán  $n$ -tablas para  $n$  mayor o igual a tres. Sean  $N$  objetos clasificados en tres categorías con  $I, J$  y  $K$  niveles, respectivamente. La probabilidad de un objeto de pertenecer a la categoría  $(i, j, k)$  es  $p_{ijk}$ . La no interacción en tres factores especifica log momios constante:

$$\frac{p_{111}p_{ij1}}{p_{i11}p_{1j1}} = \frac{p_{11k}p_{ijk}}{p_{i1k}p_{1jk}} \quad 2 \leq i \leq I, 2 \leq j \leq J, 2 \leq k \leq K.$$

La suma de las líneas son estadísticos suficientes para este modelo. Si se denota la entradas de la tabla como  $N_{ijk}$ , la suma de las líneas se denotarán como  $N_{+jk}, N_{i+k}$  y  $N_{ij+}$  donde  $N_{+jk} = \sum_{i=1}^I N_{ijk}$ ,  $N_{i+k} = \sum_{j=1}^J N_{ijk}$  y  $N_{ij+} = \sum_{k=1}^K N_{ijk}$ .

**4.1. Ejemplo** En 1978 Haberman tomó datos de 1972 del centro nacional de investigación de opinión con respecto a la postura entre cristianos blancos hacia el aborto. La información se puede ver como una tabla de  $3 \times 3 \times 3$ . La primer variable es el tipo de cristiano: protestante norteamericano, protestante sureño o católico. La segunda es la educación: baja (menos de 9 años), media (de 9 a 12 años) y alta (más de 12 años). La tercer variable es la actitud hacia el aborto no terapéutico: positivo, mezclada y negativa. La información es procesada como una muestra aleatoria de tamaño 1,055 de Estados Unidos de 1972.

		Protestantes Norteamericanos		
		Positivo	Mezclada	Negativa
Baja	9	16	41	
Media	85	52	105	
Alta	77	30	38	

		Protestantes Sureños		
		Positivo	Mezclada	Negativa
Baja	8	8	46	
Media	35	29	54	
Alta	37	15	22	

		Católicos		
		Positivo	Mezclada	Negativa
Baja	11	14	38	
Media	47	35	115	
Alta	25	21	42	

El estimador máximo verosímil de las entradas de las celdas según el modelo de no interacción en tres factores encontrado por adecuación proporcional iterativa es

		Protestantes Norteamericanos		
		Positivo	Mezclada	Negativa
Baja	12.01	14.43	39.58	
Media	85.75	52.51	103.8	
Alta	73.24	31.06	40.66	

Protestantes Sureños			
	Positivo	Mezclada	Negativa
Baja	9.44	12.25	40.27
Media	36.55	24.17	57.27
Alta	34.01	15.58	24.45

Católicos			
	Positivo	Mezclada	Negativa
Baja	6.55	11.32	45.13
Media	44.68	39.32	113.0
Alta	31.77	19.36	36.87

El estadístico ji cuadrada para bondad de ajuste es 13.37. Las usuales asintotas para esta distribución ji cuadrada le da 8 grados de libertad. Para calibrar las asintotas se corre una caminata aleatoria como la descrita anteriormente para obtener una muestra hipergeométrica con la suma de las líneas. La caminata tendrá 110 movimientos. Se concluye que el algoritmo corre fácil y bien, la aproximación de la ji cuadrada se ve bien y el modelos de no interacción en tres factores encaja con los datos.

## 5. Modelos log-lineales

Los modelos log-lineales son modelos para tablas de contingencia. El conjunto de índices es  $\mathcal{H} = \prod_{\gamma \in \Gamma} I_{\gamma}$  con  $\Gamma$  corriendo en las categorías e  $I_{\gamma}$  los valores de la categoría  $\gamma$ . Sea  $p(x)$  la probabilidad de caer en la celda  $x \in \mathcal{H}$ . El modelo log-lineal se puede definir poniendo:

$$\log(p(x)) = \sum_{a \subseteq \Gamma} \phi_a(x)$$

La suma corre sobre todos los subconjuntos de  $\Gamma$  y la notación  $\phi_a$  sólo depende de  $x$  por las coordenadas en  $a$ . Por lo que  $\phi_{\emptyset}$  es una constante. Determinando que clase de conjuntos  $\phi_a = 0$  se puede determinar el modelo.

Los modelos jerárquicos de Goodman (Goodman 1970, Haberman 1978, Darroch, Lauritzen y Speed 1980 ) empiezan con una clase  $\mathcal{C}$  de subconjuntos de  $\Gamma$  que cumpla que para todo  $a, b \subseteq \Gamma$ , si  $a \subseteq b$  o  $b \subseteq a$  entonces  $a = b$ . Un modelos jerárquico es definido poniendo  $\phi_a = 0$  al menos de que exista

$c \in \mathcal{C}$  tal que  $a \subseteq c$ . Por ejemplo si  $\Gamma = \{a, b, c\}$  y  $\mathcal{C} = \{(a, b), (b, c), (a, c)\}$  se obtiene el modelo no interacción en tres factores.

El estadístico suficiente para un modelo jerárquico es  $\{N(i_c)\}_{c \in \mathcal{C}}$  con  $i_c \in \mathcal{H}$  y  $N(i_c)$  la suma sobre todos los  $x$  que coinciden con  $i_c$  en las coordenadas determinadas por  $c$ , es decir, por la proyección en  $c$ .

Los modelos jerárquicos tienen un único estimador máximo verosímil que puede ser computado efectivamente usando el método de Newton-Raphson o el método de adecuación proporcional iterativa. Esto lleva a estimar  $\hat{p}_{\mathcal{C}}(x)$ . Si  $\mathcal{C} \subseteq \mathcal{D}$  son dos clases generadoras entonces una prueba exacta para suficiencia del modelo  $\mathcal{C}$  dentro de  $\mathcal{D}$  se puede basar en la distribución condicional bajo  $\mathcal{C}$  del estadístico ji cuadrada.

$$\sum_x \frac{(N\hat{p}_{\mathcal{C}}(x) - N\hat{p}_{\mathcal{D}}(x))^2}{N\hat{p}_{\mathcal{C}}(x)}$$

**5.1. Modelos gráficos** Los modelos gráficos son una subclase de los modelos jerárquicos que se obtienen de una gráfica con vértices  $\Gamma$  y aristas  $E$ . La clase generadora  $\mathcal{C}$  es el conjunto de camarillas de la gráfica. Estos modelos son caracterizados por propiedades de independencia condicional: para  $a, b, c \in \Gamma$ , las variables  $a$  y  $b$  son condicionalmente independientes si y sólo si cualquier camino de  $a$  a  $b$  tiene que pasar por  $c$ . Por ejemplo para tres vértices se tienen los siguientes modelos gráficos:

Completa Independencia

1

3

2

Modelo  $p_{ijk}$

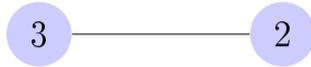
$$p_{i++}p_{+j+}p_{++k}$$

Estadístico Suficiente

$$N_{i++}, N_{+j+}, N_{++k}$$

Una Variable Independiente

1



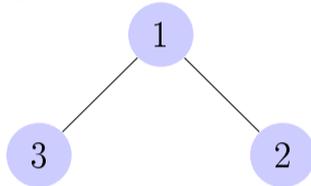
Modelo  $p_{ijk}$

$$p_{i++}p_{+jk}$$

Estadístico Suficiente

$$N_{i++}, N_{+jk}$$

Independencia Condicional



Modelo  $p_{ijk}$

$$\frac{p_{i+k}p_{+jk}}{p_{++k}}$$

Estadístico Suficiente

$$N_{i+k}, N_{+jk}$$

El modelo de no interacción en tres factores es el modelo jerárquico más sencillo que no es un modelo gráfico. Una subclase amigable de los modelos gráficos son los modelos descomponibles. Éstos surgen de gráficas para las cuales cualquier ciclo de longitud 4 tiene una cuerda. Los modelos descomponibles permiten una forma cerrada del estimador máximo verosímil y algoritmos simples para la generación de su distribución hipergeométrica.

---

## Conclusiones

---

La estadística algebraica es un gran avance en la estadística desde un punto de vista algebraico y computacional. Muchos de los modelos estadísticos se limitan a un punto de vista del álgebra lineal, esto es por razones computacionales ya que aunque existen muchos modelos que están basados en geometría diferencial o teoría de la medida cuando se intentan implementar tienen severos problemas computacionales. Estos problemas computacionales van desde la existencia de un algoritmo para la implementación hasta la eficiencia de dichos algoritmos. Por lo cual ha sido y continúan siendo lo más factible usar técnicas que dependan del álgebra lineal desde un punto de vista computacional.

La estadística algebraica trae una nueva geometría como herramienta, que tiene muchas bondades a nivel computacional. Aunque la definición de variedad algebraica depende de un sistema de ecuaciones polinomial esto se puede llevar a los algoritmos de álgebra lineal y conservan una complejidad computacional polinomial, aunque generalmente esta sea mayor.

En el caso de las tablas de contingencia la estadística algebraica demuestra ser muy original en sus métodos, ya que sirven para completar los métodos de los modelos gráficos que principalmente trabaja Lauritzen. En la tesis solo se muestran dos ejemplos de aplicaciones dado que aplicaciones más fuertes implicarían un estudio más profundo de un tercer tema como filogenética, que es donde la estadística algebraica se ha acomodado mejor a nivel de investigación aplicada. En el mundo los institutos que tienen investigación en biología computacional suelen tener un estadístico algebraico. Por otro lado, la importancia de la estadística algebraica es tal que ya hoy en día que ya existe una revista especializada solamente en el tema *Journal of Algebraic Statistics*.

---

## ANEXOS-Códigos en R

---

Aquí se presentan los códigos en R que se utilizarán.

### Prueba Chi Cuadrada de Pearson

```
> n=82
> g=12
> OBS=matrix(c(1,0,0,0,1,2,0,0,1,0,1,0,
+ 1,0,0,1,0,0,0,0,0,1,0,2,
+ 1,0,0,0,2,1,0,0,0,0,0,1,
+ 3,0,2,0,0,0,1,0,1,3,1,1,
+ 2,1,1,1,1,1,1,1,1,1,1,0,
+ 2,0,0,0,1,0,0,0,0,0,0,0,
+ 2,0,2,1,0,0,0,0,1,1,1,2,
+ 0,0,0,3,0,0,1,0,0,1,0,2,
+ 0,0,0,1,1,0,0,0,0,0,1,0,
+ 1,1,0,2,0,0,1,0,0,1,1,0,
+ 0,1,1,1,2,0,0,2,0,1,1,0,
+ 0,1,1,0,0,0,1,0,0,0,0,0), byrow=T,nrow=g)
> x.obs=rowSums(OBS)
> x.obs
 [1] 6 5 5 12 12 3 10 7 3 7 9 3
> y.obs=colSums(OBS)
> q.obs=0
> for(i in 1:12){
+     for(j in 1:12){
+         q.obs=q.obs+((OBS[i,j]-(x.obs[i]*y.obs[j])/n)^2)/
+         ((x.obs[i]*y.obs[j])/n)
+     }
+ }
```

```
+           }
> q.obs
[1] 115.5596
```

### Distribución muestral del estadístico ji cuadrada contra la la distribución ji cuadrada

```
> m=1000
> F=0
> n=200
> J=0
> for(a in 1:m){
+     S=sample(1:144,n,replace=TRUE)
+     X=matrix(1:144,ncol=12)*0
+     for(k in 1:144){
+         j=((S[k]-1)%12)+1
+         i=((S[k]-j)/12)+1
+         X[i,j]=1+X[i,j]
+     }
+
+ J[a]=sum(X)
+ y.obs=rowSums(X)
+ x.obs=colSums(X)
+ q.obs=0
+
+ for(r in 1:12){
+     for(s in 1:12){
+         q.obs=q.obs+((X[r,s]-(x.obs[r]*y.obs[s])/n)^2)/
+         ((x.obs[r]*y.obs[s])/n)
+     }
+ }
+ F[a]=q.obs
+ }
> plot(ecdf(F),main = "",ylab="", col="red",pch=22)
> H=rchisq(m, 121)
> lines(ecdf(H),col="blue")
```

**Cadena de Markov en tablas de contingencia** Este código corre la cadena de Markov descrita en la tabla de contingencia de los descendientes de la Reina Victoria.

```

> g=12
> n=82
> OBS=matrix(c(1,0,0,0,1,2,0,0,1,0,1,0,
+ 1,0,0,1,0,0,0,0,0,1,0,2,
+ 1,0,0,0,2,1,0,0,0,0,0,1,
+ 3,0,2,0,0,0,1,0,1,3,1,1,
+ 2,1,1,1,1,1,1,1,1,1,1,0,
+ 2,0,0,0,1,0,0,0,0,0,0,0,
+ 2,0,2,1,0,0,0,0,1,1,1,2,
+ 0,0,0,3,0,0,1,0,0,1,0,2,
+ 0,0,0,1,1,0,0,0,0,0,1,0,
+ 1,1,0,2,0,0,1,0,0,1,1,0,
+ 0,1,1,1,2,0,0,2,0,1,1,0,
+ 0,1,1,0,0,0,1,0,0,0,0,0), byrow=T,nrow=g)
> X=OBS
> m=1000
> x=0
> y=c(0,0)
> z=c(0,0)
> i=1
> ptm <- proc.time()
> while(i<= m){
+     x=rbinom(1,1,.5)
+     y=sample(1:12,2,FALSE)
+     z=sample(1:12,2,FALSE)
+
+     if(x==0 & X[y[1],z[2]]!=0 & X[y[2],z[1]]!=0 ){
+         X[y[1],z[1]]=X[y[1],z[1]]+1
+         X[y[1],z[2]]=X[y[1],z[2]]-1
+         X[y[2],z[1]]=X[y[2],z[1]]-1
+         X[y[2],z[2]]=X[y[2],z[2]]+1
+         i=i+1
+     }
+ }
+

```

```
+   if(x==1 & X[y[1],z[1]]!=0 & X[y[2],z[2]]!=0 ){
+       X[y[1],z[1]]=X[y[1],z[1]]-1
+       X[y[1],z[2]]=X[y[1],z[2]]+1
+       X[y[2],z[1]]=X[y[2],z[1]]+1
+       X[y[2],z[2]]=X[y[2],z[2]]-1
+       i=i+1
+   }
+ }
> proc.time() - ptm
  user system elapsed
 0.22   0.00   0.22
```

---

## Bibliografía

---

1. Andrews, D., Herzberg, A. Data, Springer, New York. 1985.
2. Atiyah, M., Introduction To Commutative Algebra, Addison-Wesley Series in Mathematics. 1994
3. Baldwin, E., Bérczi, G., Algebraic Geometry, 2013.
4. Bigatti, A., Robbiano, L., Toric ideals. *Matematica Contemporanea*, 21:1-25, 2001.
5. Bishop, Y. M. M., Fienberg, S. E., Discrete Multivariate Analysis: Theory and Practice. MIT Press. Holland, P. W. (1975).
6. Buchberger, B., An Algorithm for Finding the Basis Elements of the Residue Class Ring of a Zero Dimensional Polynomial Ideal. Ph.D. dissertation, University of Innsbruck. (1965).
7. Cox, D. A., Little, J., O'shea, D., Using Algebraic Geometry, Springer.
8. Cox, D. A., Little, J. and O'Shea, D., Ideals, Varieties, and Algorithms 3rd edn (New York, Springer-Verlag). (2008).
9. Darroch, J., Lauritzen, S., Speed, T., Markov fields a log-linear interactions models for contingency tables. *Ann. Statist.* 1980.
10. Diaconis, P., Efron, B. Testing for independence in a two-way table: new interpretations for the chi-square statistic. *Ann, Statist.* 1985
11. Diaconis, P. , Sturmfels, B., Algebraic algorithms for sampling from conditional distributions, *Annals of Statistics* 26(1), 363-97. (1998).

12. Drton, M., Sturmfels, B., Sullivan, S., Lectures on Algebraic Statistics. Birkhäuser Applied Probability and Statistics. 2009.
13. Eudey, T. T., Kerr, J. D., Trumbo, B. E., Using R to Simulate Permutation Distributions for Some Elementary Experimental Designs, Journal of Statistics Education, Volumen 18, Number 1, 2010
14. Fienberg, S. E., The geometry of an  $r \times c$  contingency table, Annals of Mathematical Statistics 39, 1186-90. 1968
15. Fisher, R. A. On the interpretation of  $\chi^2$  from contingency tables and the calculation of P, Journal of Royal Statistical Society 85,87-94. 1921.
16. Gibilisco, P., Riccomagno, E., Rogantin, M. P., Wynn, H. P., Algebraic and Geometric Methods in Statistics, Cambridge University Press 2010.
17. Goodman, L. The multivariate analysis of qualitative data: interactions among multiple calssifications. J. Amer. Statist. Assoc. 1970
18. Haberman, S. Analysis of Qualitative Data. Academic Press 1978.
19. Harris, J., Algebraic Geometry: A First Course, Springer.
20. Häjsgaard, S., Edwards, D., Lauritzen, S., Graphical Models with R, Springer, 2012.
21. Hosten, S., Sullivan, S. Gröbner Bases and Polyhedral Geometry of Reducible and Cyclic Models, Journal of Combinatorial Theory, Series A, Volume 100, Issue 2, November 2002, 277–301.
22. Hulek, K., Elementary Algebraic Geometry, AMS Student Mathematical Library, Vol. 20, 2003.
23. Lauritzen, S. Graphical Models. Oxford University Press, New York, 1996.
24. Lauritzen, S. L. (2002 electronic (1979, 1982, 1989)). Lectures on Contingency Tables (updated electronic version of the (University of Aalborg) 3rd (1989) ed.).
25. Lee, J. M., Introduction to Topological Manifolds, Springer 2011.

26. Li H., van Oystaeyen F., A primer of algebraic geometry: constructive computational methods. New York: Dekker, 2000.
27. Pearson, E. S., Some aspects of the geometry of statistics, Journal of Royal Statistical Society, Seires A 119 ,125-46, 1956.
28. Pistone, G. and Wynn, H. P. (1996). Generalised confounding with Gröbner bases, Biometrika 83(3), 653-66.
29. Slavković, A. B., Fienberg, S. E., Algebraic geometry of  $2 \times 2$  contingency tables.
30. Snee, Graphical display of two way contingency tables. Amer. Stats. 1974
31. Stark, P. B. <http://www.stat.berkeley.edu/~stark/SticiGui/index.htm>
32. Sturmfels, B. Grobner Bases and Convex Polytopes, American Mathematical Society, Univ. Lectures Series, No 8, Providence, Rhode Island, 1996.
33. Sturmfels, B. , "What is . . . a Gröbner Basis?", Notices of the American Mathematical Society 52 (10): 1199-1200, a brief introduction.