UNA INTRODUCCIÓN AMIGABLE A LA ESTADÍSTICA ALGEBRAICA

ERIC ZEPEDA

Resumen. En este artículo tiene por objetivo introducir amigablemente al lector al área de Estadística Algebraica sin dejar de lado los tópicos claves para quien no tenga experiencia previa en el área, más que teoría de Probabilidades y un poco de Algebra.

Índice

> ”Statisticians, like artists, have the bad habit of falling in love with their models.”
>
> *George Box*

1. Introducción

Historicamente, El tópico principal de Estadística Algebraica inicia con el trabajo de Diaconis y Sturmfels (1998) con su paper “Algebraic algorithms for sampling from conditional distributions”, pero no fue atribuido el concepto de Estadística Algebraica hasta que Pistone, Riccomagno, y Wynn con su monografía “ algebraic Statics, Monographs on Statistics and Applied Probability ”, el 2001, donde exploraron las conexiones entre Geometría Algebraica y el diseño de experimentos. Desde entonces, ha existido un gran desarrollo del área debido al crecimiento que esta obtuvo a través del impacto que tuvo la monografía creando lo que hoy conocemos como Estadística Algebraica.
Como se mencionó anteriormente, la Estadística Algebraica es un área relativamente nueva, donde se emplean herramientas provenientes de Geometría Algebraica, Algebra Conmutativa, Geometría Combinacional y otras relacionadas con computación simbolica para abarcar problemas de Teoría de probabilidades, estadisticas y sus respectivas aplicaciones.
Nuestro objetivo principal es introducir las nociones básicas para poder estudiar esta área y ver un poco de Familia Exponencial Discreta en nuestra *Introducción amigable a la Estadística Algebraica.*

2. Preliminares: Cadenas de Markov.

En esta Sección se entregarán todos los resultados a utilizar en este articulo por lo que se sugiere su lectura para quienes se aventuren a estudiar Estadistica Algebraica por primera vez sin conocimiento previo de algunos resultados de Geometría Algebraica o de Inferencia Estadística.
Para esto, empezaremos por ver un poco de Cadenas de Markov para introducir conceptos como espacio de estados y así seguir con lo que usaremos de Geometría Algebraica para entrar de a poco a hablar de Inferencia Estadística y finalizar con un poco de Estadística Algebraica.
Sea $X_1, X_2, ..., X_m$ una secuencia de variables aleatorias del mismo espacio de estados Σ, un alfabeto finito. Como son variables aleatorias, existe una distribución conjunta asociada a ellas. Es decir, para cada $(x_1, ..., x_m) \in \Sigma$ existe un número real

$$P(X_1 = x_1, X_2 = x_2, ..., X_m = x_m)$$

La cual es ea probabilidad de $X_1 = x_1, X_2 = x_2, ..., X_m = x_m$ a ser una distribución de probabilidad. Estos valores están dados entre 0 y 1 y su suma $(x_1, ..., x_m) \in \sum^m$ es uno.

Definición 2.1. La secuencia $X_1, X_2, X_3, ..., X_m$ es llamada Cadena de Markov sí para todo $i = 3, ..., m$ y para todo $x_1, ..., x_n \in \Sigma$

$$P(X_i = x_i | X_1 = x_1, X_2 = x_2, ..., X_{i-1} = x_{i-1}) = P(X_i = x_i | X_{i-1} = x_{i-1})$$

Además definimos el espacio de estados como el espacio de todos los posibles estados de la Cadena de Markov.

Ejemplo 2.2. (Cadena de Markov con 3 pasos). Sea $m = 3$ y $\Sigma = \{0, 1\}$ en un modelo de cadena de amrkov. Naturalmente podemos asociar una distribucion de probabilidad con un punto en $\mathbb{R}^8$. De hecho, la distribución conjunta es determinado por las 8 combinaciones posibles de $P(X_1 = i, X_2 = j, X_3 = k)$ para $i, j, k \in \{0, 1\}$. Usaremos la siguiente abreviación para las distribuciones de probabilidad, $p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$. De este modo, la distribución conjunta de tres variables aleatorias binarias es el punto

$$(p_{000}), (p_{001}), (p_{010}), (p_{011}), (p_{100}), (p_{101}), (p_{110}), (p_{111}) \in \mathbb{R}^8$$

Calculando las distribuciones condicionales tenemos una expresión racional en termino de las probabilidades conjuntas:

$$P(X_3 = k | X_1 = i, \ X_2 = j) = \frac{p_{ijk}}{p_{ij+}}$$

Donde + denota

$$p_{ij+} = \sum_{k \in \{0,1\}} p_{ijk}$$

Luego, Por la condición que implica ser una cadena de markov (la de la definición), una distribución que provenga de una cadena de Markov, se traduce en la expresión racional

$$\frac{p_{ijk}}{p_{ij+}} = \frac{p_{+jk}}{+j+} \quad \forall i, j, k \in \{0, 1\}$$

Análogamente, se puede representar esto como

$$\frac{p_{ijk}}{p_{ij+}} = \frac{p_{i'jk}}{i'j+} \quad \forall i, i', j, k \in \{0, 1\}$$

Definición 2.3. Un vector $(p_{000}), (p_{001}), (p_{010}), (p_{011}), (p_{100}), (p_{101}), (p_{110}), (p_{111}) \in \mathbb{R}^8$ es una distribución de probabilidades de una Cadena de Markov si y solo si satisface

1. $p_{ijk} \geq 0 \quad \forall i, j, k \in \{0, 1\}$
2. $\Sigma_{i,j,k \in \{0,1\}} p_{ijk} = 1$
3. $p_{000}p_{101} - p_{001}p_{100} = 0$
4. $p_{010}p_{111} - p_{011}p_{110} = 0$

3. Preliminares: Geometría Algebraica

3.1. Variedades. Sea $\mathbb{K}$ un cuerpo. Para propositos de este artículo, existen tres cuerpos que son de nuestro interes: El cuerpo de los racionales $\mathbb{Q}$, el cual será de mucha útilidad para realizar cálculos, el de los números reales $\mathbb{R}$, el cual es fundamental, pues la imagen de una función de probabilidad es real , y el cuerpo de los complejos $\mathbb{C}$ que será empleado en algunas circunstancias para probar algunos teoremas en esta sección.
Sean $p_1, p_2, ..., p_r$ variables polinomiales. Un monomio de estas variables polinomiales, es una expresión de la forma

$$p^u := p_1^{u_1} p_2^{u_2} \dots p_r^{u_r}$$

Donde $u = (u_1, u_2, ..., u_r) \in \mathbb{N}^r$ es un vector de enteros positivos y p^u en nuestra expresión es llamado vector monomial.
Un polinomio en variables polinomiales $p_1, p_2, ..., p_r$ en $\mathbb{K}$ es una combinación lineal finita de monomios con coeficientes en $\mathbb{K}$. De esta manera, los polinomios admiten una representación como sigue:

$$\sum_{u\in A} c_u p^u$$

Donde A es un subconjunto finito de $\mathbb{N}^r$, cada coeficiente $c_u \in \mathbb{K}$.

Ejemplo 3.1. A continuación enlistaremos un par de ejemplos de lo que **si** puede ser un monomio polinomial

- La expresión $p_1^2 p_3^4$ es un monomio en variables polinomiales p_1, p_2, p_3 con $u = (2, 0, 4)$ satisface todas las condiciones para ser un monomio polinomial.
- $f = p_1^2 p_3^4 + 6p_1^2 p_2 - \sqrt{2} p_1^3$ con $\mathbb{K} = \mathbb{R}$.

Ejemplo 3.2. Por su parte, también se muestran ejemplos donde la expresión **no** es un monomio polinomial.

- $p_1^2 p_3^{-4}$ **no** es monomial, porque posee un exponente negativo en la variable polinomial p_3.
- Cualquier serie de potencias no puede generar un monomio polinomial. Esto sucede por el hecho de que A debe ser un conjunto finito, cosa que no se cumple en una serie de potencias cualquiera. Un ejemplo de ello puede ser

$$\sum_{i=0}^{\infty} p_1^j$$

Sea

$$\mathbb{K}[p] := \mathbb{K}[p_1, p_1, ..., p_r]$$

El conjunto de todos los polinomios con variables polinomiales $p_1, p_2, ..., p_r$ cuyos coeficientes pertenecen a $\mathbb{K}$. Este conjunto de polinomios forma un anillo, pues la suma de dos polinomios es un polinomio y el producto también. Naturalmente cada polinomio es considerado como una función multivariada de $\mathbb{K} \to \mathbb{K}$ evaluandola. Por esta razón, nos referiremos a $\mathbb{K}[p]$ como el anillo de funciones de polinomiales.

Definición 3.3. Sea $S \subseteq K[p]$ un conjunto de polinomios. La *variedad* definida por S es el conjunto

$$V(S) = \{a \in \mathbb{K}^r : f(a) = 0, \quad \forall f \in S\}$$

La variedad $V(S)$ es también llamada conjunto de ceros de S (ó locus cero).

Observación 3.4. El conjunto de puntos en una variedad va a depender principialmente de $\mathbb{K}$. Es por esto que cuando sea necesario denotaremos la variedad como $V_{\mathbb{K}}(S)$.

Ejemplo 3.5. (Algunas variedades en $\mathbb{R}^2$). La variedad

- $V(\{p_1 - p_2^2\}) \subset \mathbb{R}^2$ es la familia de curvas parabolicas que se abren por la derecha en el plano.
- $V(\{p_1^2 + p_2^2 - 1\}) \subset \mathbb{R}^2$ es la cirscunsferncia centrada en el orígen.
- La variedad $V(\{p_1 - p_2^2, p_1^2 + p_2^2 - 1\})$ es el conjunto de puntos que satisfacen ambas ecuaciones

$$p_1 - p_2^2 = 0 \quad p_1^2 + p_2^2 - 1 = 0$$

Así, es la intersección $V(\{p_1 - p_2^2\}) \cap V(\{p_1^2 + p_2^2 - 1\})$. De este modo, es posible concluir que la variedad $V(\{p_1 - p_2^2, p_1^2 + p_2^2 - 1\})$ es una cantidad finita de puntos

$$\begin{aligned}
&V_{\mathbb{Q}}(\{p_1 - p_2^2, p_1^2 + p_2^2 - 1\}) = \emptyset \\
&V_{\mathbb{R}}(\{p_1 - p_2^2, p_1^2 + p_2^2 - 1\}) = \left\{ \left(\frac{-1+\sqrt{5}}{2}, \pm\sqrt{\frac{-1+\sqrt{5}}{2}} \right) \right\} \\
&V_{\mathbb{C}}(\{p_1 - p_2^2, p_1^2 + p_2^2 - 1\}) = \left\{ \left(\frac{-1\pm\sqrt{5}}{2}, \pm\sqrt{\frac{-1\pm\sqrt{5}}{2}} \right) \right\}
\end{aligned}$$

En algunas situaciones, no es raro considerar polinomios de la forma $p_1^2 + 4p_2 = -17$. Es decir, polinomios que no tienen una forma directa $f(p) = 0$, pero se puede llegar a ella a través de las operaciones algebraicas antes descritas (suma, resta o multiplicación de polinomios).

Por otra parte, en el ambito de la estadística algebraica. Muchas de las variedades son presentadas como conjuntos parámetricos. En partícular, existe un mapeo polinomial $\phi : \mathbb{K}^d \to \mathbb{K}^r$ tal que $V = \{\phi(t) : t \in \mathbb{K}^d\}$ es la imagen del mapeo. También denotaremos a la imagen $\phi(\mathbb{K}^d)$ para simplificar la notación.

Ejemplo 3.6. (Cubo Torcido). Considere el siguiente mapeo polinomial

$$\phi : \mathbb{C} \to \mathbb{C}^3, \quad \phi(t) = (t, t^2, t^3)$$

La imagen inversa de este mapeo es la *curva del cubo torcido* en $\mathbb{C}^3$. Cada punto del cubo torcido satisface las ecuaciones

$$p_1^2 = p_2 \quad p_1^3 = p_3$$

Reciprocamente, cada punto tal que satisfaga ambas condiciones pertenece a la imagen de ϕ.

El siguiente ejemplo consiste en una variedad paramétrica que parece de manera natural en el estudio de modelos estadísticos.

Ejemplo 3.7. (Variabels aleatorias binomiales). Considere el mapeo polinomial $\phi : \mathbb{C} \to \mathbb{C}^{r+1}$ cuya función en la cordenada i es

$$\phi_i(t) = \binom{r}{i} t^i (1-t)^{r-i}, \quad i = 0, 1, ..., r.$$

Para un parámetro real $\theta \in [0, 1]$, el valor $\phi_i(\theta)$ es la probabilidad $P(X = i)$ donde X es una variable binomial aleatoria con probabilidad de exito θ. Esto resulta en que si por ejemplo, θ corresponde a la probabilidad de obtener cara en una moneda, $\phi_i(\theta)$ corresponde a la probabilidad de ver i caras en r lanzamientos independientes de esta. En consecuencia, el vector $\phi(\theta)$ denota la distribución de probabilidad de una variable aleatoria binomial.
La imagen del intervalo real, [0,1], es una curva en el simplice de probabilidad definido como

$$\Delta_r = \left\{ p \in \mathbb{R}^{r+1} : \sum_{i=0}^{r} p_i = 1, p_i \geq 0, \ \forall i \right\}$$

El cual consiste en todas las distribuciones posibles de una variable binomial aleatorias. Por ejemplo, si consideramos el caso en que $r = 2$, la curva en $\phi([0, 1])$ es la curva dentro del simplice mostrado a continuación:

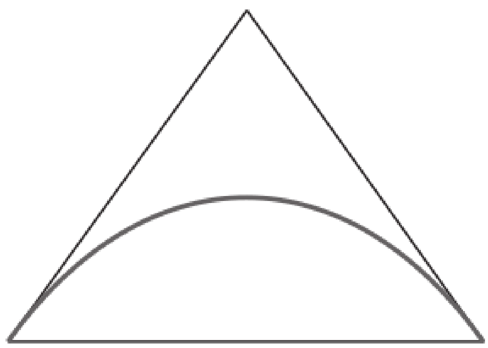

FIGURA 1. El conjunto de todas las distribuciones binomiales para 2 intentos ($r = 2$).

La imagen de una parametrización compleja $\phi(\mathbb{C})$ es una curva en $\mathbb{C}^{r+1}$, que también es una variedad algebraica. En partícular, es el conjunto de puntos que satisfacen la ecuación trivial $\sum_{i=0}^{r} p_i = 1$ junto con todos los subdeterminantes de 2×2, que se anulan en la matriz de $2 \times r$:

$$\begin{pmatrix} p_0 \Big/ \binom{r}{0} & p_1 \Big/ \binom{r}{1} & p_2 \Big/ \binom{r}{2} & \cdots & p_{r-1} \Big/ \binom{r}{r-1} \\ p_1 \Big/ \binom{r}{1} & p_2 \Big/ \binom{r}{2} & p_3 \Big/ \binom{r}{3} & \cdots & p_r \Big/ \binom{r}{r} \end{pmatrix}$$

En aplicaciones de estadística algebraica, es común considerar no solo mapeos cuyas funciones cordenadas son polinomios, también puede suceder el caso en que sean funciones racionales. Estas últimas son de la forma $\frac{f}{g}$, donde f y g son polinomios. De este modo, un mapeo racional es de la forma $\phi : \mathbb{C}^d \dashrightarrow \mathbb{C}^r$, donde cada función cordenada $\phi_i = f_i/g_i$, donde cada $f_i, g_i \in \mathbb{K}[t] := \mathbb{K}[t_1, ..., t_d]$ son polinomios.

Observación 3.8. El mapeo ϕ no es definido en todo $\mathbb{C}^d$, sino fuera de la hipersuperficie $V(\{\prod_i g_i\})$. Es por esto, que la imagen del mapeo antes mencionado pertence al conjunto $\mathbb{C}^d \backslash V(\{\prod_i g_i\})$, donde el mapeo está bien definido. El simbolo $\dashrightarrow$ es usado en la definición notacional de unn mapeo racional para indicar que el mapeo no necesariamente esta definido en todo su dominio.

3.2. Ideales. En la parte anterior, mostramos como tomar una colección de polinomios (un objeto algebraico) para producir su conjunto de ceros comunes (un objeto geométrico). Esta construcción puede hacerce al revés, desde un objeto geométrico a uno algebraico. Los objetos de este estilo vienen de un radical ideal. Esta interacción entre variedades e ideales es el corazón de la geometría algebraica, que es lo que veremos a continuación.

Definición 3.9. Sea $W \subseteq \mathbb{K}^r$. El conjunto de polinomios

$$I(W) = \{f \in \mathbb{K}[p] : f(a) = 0, \ \forall a \in W\}$$

Es el ideal que se anula de W o el ideal definido de W.
Recuerde que un subconjunto I de un anillo R es un *ideal* si $f + g \in I$ para todo $f, g \in I$ y $hf \in I$ para todo $f \in I$ y $h \in R$. Como el nombre indica, $I(W)$ es un ideal.
Se dice que un polinomio $\mathcal{F}$ genera al ideal I, si está el caso en que cada polinomio $f \in I$ puede ser escrito como combinación lineal finita de la forma $f = \sum_{i=1}^{k} h_i f_i$, donde $f_i \in \mathcal{F}$ y los h_i polinomios en $\mathbb{K}[p]$. El número k puede dependerde f. En este caso, denotraremos $\langle \mathcal{F} \rangle$ a que I es generado por $\mathcal{F}$.

Teorema 3.10. *(Teorema de la base de hilbert). Para cada ideal I en un anillo polinomial $K[p_1, ..., p_r]$ existe un conjunto finito de polinomios $\mathcal{F} \subset I$ tal que $I = \langle \mathcal{F} \rangle$. Esto es que cada ideal en $K[p]$ es generado por un conjunto finito.*

Observación 3.11. Note que el teorema de la base de hilbert solo se cumple para anillos polinomiales considerando finitas variables. En general, un anillo que satisfaga la propiedad de que cada ideal sea generado finitamente por un conjunto, es llamado un anillo Noetheriano.

Cuando es presentado con un conjunto $W \subseteq \mathbb{K}^r$, es común tener por tarea encontrar un conjunto de polinomios en $I(W)$ tal que genere un ideal. El caso especial donde el conjunto $W = \phi(\mathbb{K}^d)$ para algún mapeo racional ϕ es llamado problema implicito y muchos problemas en estadística algebraica pueden ser interpretados de esta forma.

Ejemplo 3.12. (Variables binomiales aleatorias). Sea $\phi : \mathbb{C} \to \mathbb{C}^{r+1}$ el mismo mapeo del ejemplo 2,9. Entonces el ideal $I(\phi(\mathbb{C})) \subseteq \mathbb{C}[p_0, ..., p_r]$ es generado por los polinomios $\sum_{i=0}^{r} p_1 - 1$ más los subdeterminantes de 2×2 de la matriz de ese ejemplo.
En el caso especial que vimos, $r = 2$, tenemos

$$I(\phi(\mathbb{C})) = \langle p_0 + p_1 + p_2 - 1, 4p_0p_2 - p_1^2 \rangle$$

Así, los polinomios $p_0 + p_1 + p_2 - 1$ y $4p_0p_2 - p_1^2$ resuelven el problema implicito en esta instancia.

Si dado un ideal, podemos calcular iterativamente la variedad definida por los polinomios que se anulan. Siempre se tiene $I \subseteq I(V(I))$. Sin embargo, esta contención no necesariamente es de igualdad. Los ideales que vienen de ideales que se anulan, son forzados a tener una estructura adicional.

Definición 3.13. Un ideal I es llamado radical si $f^k \in I$ para algún polinomio f y un entero positivo k se tiene que $f \in I$. El radical de I, denotado $\sqrt{(I)}$ es el ideal radical más pequeño que contenga a I y consiste en todos los polinomios

$$\sqrt{I} = \left\{ f \in \mathbb{K}[p] : f^k \in I, \ k \in \mathbb{N} \right\}$$

La siguiente proposición explica la importancia de los ideales radicales en geometría algebraica.

Proposición 3.14. *Para cualquier cuerpo $\mathbb{K}$ y cualquier conjunto $W \subset \mathbb{K}^r$, el ideal $I(W)$ es un radical ideal.*

El resultado que sigue, es mucho más significativo y nos muestra el proceso de tomar ideales que se anulan de alguna variedad en el caso de un cuerpo algebraicamente cerrado[1].

Teorema 3.15. *(Nullstellensatz).Sea I un ideal de $\mathbb{K}[p_1, ..., p_n]$, donde $\mathbb{K}$ es algebraicamente cerrado[2]. Supongamos que $f \in \mathbb{K}[p_1, ..., p_n]$ se anula en I. Entonces existe un $m \in \mathbb{N}$ tal que $f^m \in I$. Alternativamente,*

$$I(V(I)) = \sqrt{I}$$

En lo que sigue, recordemos que en la topologia de Zariski, es una topología generada a partir de sus cerrados. Estos son los ceros comunes de los polinomios que pueden darse. Formalmente, estos son como sigue:

Definición 3.16. Dado un conjunto $W \subseteq \mathbb{K}^r$ es un cerrado de Zariski si es una variedad de la forma $W = V(S)$ para algún $S \subseteq \mathbb{K}[p]$

Con esta pequeña definición es posible hablar de *Clausuras de Zariski* de algún conjunto.

Proposición 3.17. *El conjunto $V(I(W))$ es llamado clausura de zariski de W. Esta es la variedad algebraica más pequeña que contiene a W*

Ejemplo 3.18. (Una claurusa de Zariski simple). Como un ejemplo de clausura de Zariski, sea $W \subset \mathbb{R}^2$ que consiste en el eje p_1 sin el orígen (0,0). Esto por el simple hecho de que si un polinomio se anula W, entonces también lo hará en en orígen. Así, la clausura de Zariski contiene el eje p_1. Por otra parte, el eje p_1 es una variedad definida por el ideal $\langle p_2 \rangle$, en consecia, la clausura de Zariski en de W es un cerrado de Zariski.

Ejemplo 3.19. (Variable Aleatoria Binomial). Sea $\phi : \mathbb{C} \to \mathbb{C}^{r+1}$ el mimsmo mapeo que usamos para los ejemplos de variable aleatoria binomial anteriores. La imagen del intervalo $[0, 1] \in \Delta_r$ no es un cerrado de Zariski. En efecto, la clausura de $\phi([0, 1])$ es igual a $\phi(\mathbb{C})$. Esto es porque $\phi(\mathbb{C})$ es cerrado y la clausura de Zariski de $[0, 1] \subseteq \mathbb{C}$ en el simplice es todo $\mathbb{C}$.

Observación 3.20. La clausura de Zariski **no** es $\mathbb{R} \subseteq \mathbb{C}$. Además, $\mathbb{R}$ no es una variedad si es considerado como subconjunto de $\mathbb{C}$, desde que $\mathbb{R} = \{z \in \mathbb{C} : z = \overline{z}\}$ y la conjugación no es una operación operación en el anillo de polinomios.
Si solo consideramos subconjuntos de Δ_r, entonces $\phi([0, 1])$ es igual a la intersección de la clausura de Zariski con el simplice de probabilidad. En este caso, la variedad entrega una buena aproximación para el modelo estadístico de variables aleatorias binomiales.
Ahora, recordemos que en cualquier espacio topologico, el complemento de un conjunto cerrado es uno abierto y en el caso de la Topología de Zariski[3], el complemento de un cerrado de Zariski es un *Abierto de Zariski*.

Ejemplo 3.21. (Mezcla de Variables Aleatorias Binomiales). Sea $\mathcal{P} \subseteq \Delta_{r-1}$ una familia de distribuciones probabilidades (por ejemplo, un modelos estadístico) y sea $s \in \mathbb{N}$. El modelo mezclado i, $Mixt^s(\mathcal{P})$ consiste en todas las distribuciones de probabilidad que vienen de combinaciones convexas de s distribuciones en $\mathcal{P}$. Esto es

$$Mixt^s(\mathcal{P}) = \left\{ \sum_{j=1}^{s} \pi_j p^{(j)} : \pi \in \Delta_{s-1} \ y \ p^{(j)} \in \mathcal{P}, \ j \in \{1, ..., s\} \right\}$$

Los modelos de mezcla permiten construir modelos más complejos a partir de un modelo simple $\mathcal{P}$.
Note que En el caso especial en que $r = 2$ y $s = 2$. $Mixt^2(\mathcal{P})$ es la región bajo la curva en la figura antes mostrada en el primer ejemplo de distribución aleatoria binomial. Por el hecho de que $Mixt^2(\mathcal{P})$ es un subconjunto bidimensional de Δ_2. Veremos que la clausura de Zariski $Mixt^2(\mathcal{P})$ es todo el plano $V(p_0 + p_1 + p_2 - 1)$.
De esta menera, la intersección de la clausura de Zariski y el simplice de probabilidad Δ_2 no es un modelo de mezcla.

Este ejemplo ilustra el hecho que sucede en algunas situaciones en estadística algebraica, que no es suficiente considerar solo variedades. De hecho, el modelo de mezcla es el conjunto

$$\Delta_2 \cap \left\{ p : det \begin{pmatrix} 2p_0 & p_1 \\ p_1 & 2p_2 \end{pmatrix} \geq 0 \right\}$$

[1]Recuerde que un cuerpo es algebraicamente cerrado si cada polinomio con coeficientes en ese cuerpo posee todas sus raices en ese cuerpo.
[2]Por ejemplo $\mathbb{C}$
[3]Para aclarar, la topologría de Zariski es la topología inducida por los cerrados de Zariski.

Observación 3.22. En el caso especial de mezcla de variables aleatorias binomiales es muy empleada en estadística bayesiana por su conexión a intercambiabilidad via el teorema de Finetti.

Habiendo construido las nociones de topología de Zariski y postulado el teorema Nullstelensatz, es posible hacer una correspondencia natural entre ideales radicales y variedades como se muestra en el siguiente resultado.

Teorema 3.23. *(Correspondencia Ideal-Variedad). Sea $\mathbb{K}$ un cuerpo algebraicamente cerrado. Entonces los mapeos*

$$V : \left\{ I \subseteq \mathbb{K}[p] : I = \sqrt{I} \right\} \to \{ W \subseteq \mathbb{K}^r : W = V(J) \text{ para algún } J \subseteq \mathbb{K}[p] \}$$
$$I : \{ W \subseteq \mathbb{K}^r : W = V(J) \text{ para algún } J \subseteq \mathbb{K}[p] \} \to \left\{ I \subseteq \mathbb{K}[p] : I = \sqrt{I} \right\}$$

Cada una es una inclusión-inversa biyectiva entre el conjunto de ideales radicales y el conjunto de variedades.
Demostración:
Por la definición de variedad, dada una variedad V, $V(I(V)) = V$. Por otra parte, por Nullstellensatz nos indica que $I(V(I)) = I$ si I es un ideal radical y K es algebraicamente cerrado. Esto indica que V e I son biyecciones. Para ver que estos dos sean inclusión- inversa, note que si $I \subseteq J$ entonces $V(J) \subseteq V(I)$, desde que en la definición de variedad, $V(J)$ tiene más restricciones que $V(I)$ y así menos puntos que satisfagan esas condiciones.

El resultado anterior, tiene como consecuencia, nos otorga un vinculo directo entre geometría y algebra.

Proposición 3.24. *Sean I, J dos ideales de $K[p]$. Entonces*

$$V(I+J) = V(I) \cap V(J)$$

Donde $I+J = \{f+g : f \in I, g \in J\}$. Además, si $\mathbb{K}$ es algebraicamente cerrado, y v, W subconjuntos arbitrarios de $\mathbb{K}^r$, entonces

$$I(V \cap W) = \sqrt{I(V) + I(W)}$$

Demostración:
Para esta parte, basta con demostrar solo la primera ecuación, pues es consecuencia de la correspondencia ideal-variedad.
Sea $a \in V(I+J)$.

$$\begin{aligned} &\Rightarrow f(a) = 0 \text{ para todo } f \in I \text{ y todo } f \in J. \\ &\Rightarrow a \in V(I) \cap V(J). \\ &\therefore a \in V(I) \end{aligned}$$

Por otra parte, si $a \in V(I) \cap V(J)$ todos los polinomios satisfacen esa condición en I y J por lo que también $I+J$.

Un resultado análogo conectando intersecciones de ideales y uniones de variedades.

Proposición 3.25. *Sean I y J ideales de $K[p]$. Entonces,*

$$V(I \cap J) = V(I) \cup V(J)$$

.
Si V y W son subconjuntos arbitrarios de $\mathbb{K}^r$, entonces

$$I(V \cup W) = I(V) \cap I(W)$$

Demostración:
La demostración es trivial y se deja como ejercicio para el lector.

3.3. Bases de Gröbner. Las bases de Gröbner son una herramienta algebraica para estudiar los ideales computacionalmente. El uso más básico de las bases Gröbner para responder al *problema de la membresía ideal*. Es decir, para ver si un polinomio cualquiera f pertenece o no a un ideal particular I. Iniciaremos esta parte como sigue:

Definición 3.26. Una relación de orden $\prec$ establece un orden total en el conjunto de monomios en el anillo polinomial $\mathbb{K}[p]$ tal que satisface estas dos condiciones:

1. $1 = p^0 \preceq p^u$ para todo $u \in \mathbb{N}^r$.
2. $p^u \prec p^v$ implica $p^w \cdot p^u \prec p^w \cdot p^v$ para todo $w \in \mathbb{N}^r$

Los ejemplos mas comunes donde se aplica dicho orden, está en el orden lexicografico y orden lexicografico al revés. Empezaremos revisando uno por uno.

Ejemplo 3.27. (Orden Lexicografico). Este tipo de orden es el mismo que tienen las guias telefonicas, haciendo distinción entre los caracteres como por ejemplo, aa con ab asignando el segundo despues del primero porque se sobrentiende que primero esta el a antes que el b. En el caso polinomial, en vez de considerar un alfabeto como en el ejemplo dado, se busca ordenar según los indices que se condiere en nuestro conjunto de polinomios haciendo la distinción como $p^u \prec_{lex} p^v$ si el término de más a la izquierda de $v - u$ es positivo. Un ejemplo de ello es la cadena

$$p_3^3 \prec_{lex} p_2 \prec_{lex} p_1^2 p_3^3 \prec_{lex} p_1^2 p_3^3 \prec_{lex} p_1^3$$

Ejemplo 3.28. (Orden Lexicografico reverso). Esto es al revés del ejemplo anterior, pues considera que $p^u \prec_{rev} p^v$ si y solo si

1. $\sum_{i=1}^r u_i < \sum_{i=1}^r v_i$
2. $\sum_{i=1}^r u_i = \sum_{i=1}^r v_i$ y el elemento de más a la derecha es negativo.

En principio no podría ser clara la aseveración que se esta tomando, asi que tomemos la cadena siguiente

$$p_1 \prec_{rev} p_2 p_3^3 \prec_{rev} p_2^3 p_3^1 \prec_{rev} p_2^3 p_3^2 \prec_{rev} p_1^5$$

Donde lo más pequeño es el p_1^5 bajo esta relación de orden.

Definición 3.29. Dada una relación de orden $\prec$ y un polinomio $f \in \mathbb{K}[p]$, entonces el *monomio inicial*, *relación inicial* o *relación principal* de f con respecto a $\prec$ es el monomio más grande p^u con respecto al orden $\prec$ sobre todos los monomios en f con coeficientes cero. Denotaremos el monomio inicial de f como $in_{\prec}(f)$.

Definición 3.30. El *ideal inicial* $in_{\prec}(I)$ de un ideal $I \subseteq \mathbb{K}[p]$ con respecto a la relación de orden $\prec$ es el ideal generado por los monomios iniciales de todos los polinomios en I:

$$in_{\prec}(I) = \langle in_{\prec}(f) : f \in I \rangle$$

.

Definición 3.31. (Base de Gröbner). Una Base de Gröbner de un ideal I con respecto a la relación de orden $\prec$ es una colección finita de polinomios $\mathcal{G} \subset I$ tal que los monomios iniciales de los elementos de $\mathcal{G}$ generan el ideal inicial:

$$\langle in_{\prec}(g) : g \in \mathcal{G} \rangle = in_{\prec}(I)$$

Una definición equivalente de una base de Gröbner es como un conjunto finito de polinomios $\mathcal{G} \subseteq I$ talque para todo $f \in I$ no nulo, existe un $g \in \mathcal{G}$ tal que $in_{\prec}(g)$ divide $in_{\prec}(f)$.

Observación 3.32. Observe que siempre las bases finitas de Gröbner siempre existen para cualquier ideal y cualquier relación de orden como consecuencia del teorema de la base de hilbert por el simple hecho de que $in_{\prec}$ es un ideal con generadores monomiales finitos. Esto permite escoger un conjunto $\mathcal{G} \subseteq I$ cuyos terminos iniciales general $in_{\prec}$.

A continuación, describiremos como calcular las bases de Gröbner e ideales iniciales de ideales.

Algoritmo 1: División Multivariada

Entrada: $\mathcal{G} = \{g_1, ..., g_k\}$ un conjunto finito de polinomios, f otro polinomio, y $\prec$ una relación de orden antes establecida.

Salida : Una representación de la forma $\sum_1^k h_i g_i + r$ para f tal que ningún termino de r es divisible por un termino inicial de cualquier polinomio en $\mathcal{G}$

Inicio Algoritmo de División Multivariada;

Fijar $h_i = 0$ para todo i y $r = f$;
while *(r posee un termino $c_a p^a$ divisible por el termino principal de algún g_i)* **do**
  $h_i \to h_i + c_a p^a / in_{\prec}(g_i)$;
  $r \to r - c_a p^a / in \prec (g_i) \cdot g_i$;
end

Observación 3.33. Este algoritmo está bién definido por el hecho de que para nuestra relación de orden no puede existir una secuencia infinita decreciente de monomios con respecto a esta.

Cabe mencionar, que el resto r resultante de dividir f po $\mathcal{G}$ es llamado forma normal de f y se denota $Nf_{\mathcal{G}}(f)$. Note que por el algoritmo de división multivariada, esta representación no necesariamente es única al depender del orden en que se presenten los polinomios. Es por ello que la unicidad dependerá solo en caso de fijarse un orden que sea invariante para el mismo conjunto de polinomios $\mathcal{G}$. Este puede ser empezando del monomio más pequeño al más grande y viceversa.

Ahora, consideraremos el caso especial cuando $Nf_{\mathcal{G}}(f) = 0$ para introducir una caracterización de las bases de Gröbner.

Teorema 3.34. *Sea I un ideal, $\prec$ una relación de orden y $\mathcal{G}$ es un subconjunto finito de I. Entonces, $NF_{\mathcal{G}}(f) = 0$ para todo $f \in I$ si y solo si, $\mathcal{G}$ es una base de Gröbner de I con respecto a $\prec$.*
Demostración:
Supongamos que $NF_{\mathcal{G}}(f) = 0$ para todo $f \in I$. Entonces, si $f \in I$, existe un $g \in \mathcal{G}$ tal que $in_{\prec}(g)$ divide $in_{\prec}(f)$ por lo que es eventualmente cancelado en el proceso de división. Así, $\mathcal{G}$ es una base de Gröbner para I.

Por otra parte, sea $\mathcal{G}$ una base de Gröbner para I, $f \in I$ y $r = NF_{\mathcal{G}}(f)$. Luego, por el algoritmo de división, existen $h_1, ..., h_k$ tales que $f = \sum_{i=1}^k h_i g_i + r$. Reordenando la expresión se tiene

$$r = f - \sum_{i=0}^{k} h_i g_i$$

De este modo, $r = NF_{\mathcal{G}}(f) \in I$. No obstante, por el algoritmo de división, $in_{\prec}(r)$ no es divisible por cualquier término principal de $\mathcal{G}$ (Contradicción). Luego, como $\mathcal{G}$ es una base de Gröbner, $r = 0$.

Una aplicación del teorema anterior es la de verificar ideales a partir de una base de Gröbner. En partícular, si $\mathcal{G}$ es una base de Gröbner para I, entonces $f \in I$ si y solo si $NF_{\mathcal{G}}(f) = 0$. Otra consecuencia de la definición de la base de Gröbner es que toda base de Gröbner de un ideal es siempre un conjunto generador para dicho conjunto.

Corolario 3.35. *Sea I un ideal, $\prec$ una relación de orden y $\mathcal{G}$ una base de Gröbner para I con respecto a $\prec$. Entonces $\mathcal{G}$ genera I.*
Demostración:
Sea $f \in I$ y $\mathcal{G} = \{g_1, ..., g_k\}$. El teorema anterior nos indica que $NF_{\mathcal{G}}(f) = 0$. Esto implica que existen $h_1, ..., h_k \in \mathbb{K}[p]$ tal que f puede ser expresado como $f = \sum_{i=0}^k h_i g_i$. Es decir, $\mathcal{G}$ es un conjunto generador para I.

Mientras que vimos que las bases de Gröbner siempre existen, pero su calculo puede resultar no ser obvio a partir de un conjunto generador. Una manera de lograrlo, es a través de los $S-$polinomios que son de la forma:

$$S(f,g) = \frac{p^v}{gcd(p^u,p^v)}f - \frac{p^u}{gcd(p^u,p^v}g$$

Donde f,g son polinomios y p^u, p^v son términos principales. Estos $S-$polinomios nos serán de gran utilidad para el siguiente resultado:

Teorema 3.36. *(Criterio de Buchberger). Sea I un ideal, $\prec$ una relación de orden y $\mathcal{G}$ un conjunto que genere I. Entonces, $\mathcal{G}$ es una base de Gröbner de I con respecto a $\prec$ si y solo si $NF_{\mathcal{G}}(S(f,g)) = 0$ para todo $f,g \in \mathcal{G}$.*
La demostración se deja como ejercicio para el lector.

Algoritmo 2: Algoritmo de Buchberger

Entrada: Un conjunto $\mathcal{F}$ generadora de I y una relación de orden
Salida : Una base de Gröbner para I con respecto a $\prec$

Inicio Algoritmo de Buchberger;

Fijar $\mathcal{G} := \mathcal{F}$;
while *(Mientras $\mathcal{G}$ no satisfaga el criterio de Buchberger)* **do**
  Existe un par $f,g \in \mathcal{G}$ tal que $NF_{\mathcal{G}}(S(f,g)) = h \neq 0$.
  $\mathcal{G} \to \mathcal{G} \cup \{h\}$;
end

La importancia de este algoritmo en lo que veremos de estadística algebraica será en pocas palabras el calculo de bases de Gröbner a través de Macaulay2.

4. Preliminares: Inferencia Estadística.

En principio, la teoría de probabilidades se enfoca en variables aleatoria y lo que le suceden de acuerdo a varias transformaciones o medidas son realizadas. Implicitamente en el estudio de esto está el hecho de que las distribuciones de estas probabilidades son conocidas y estan fijas. Es aquí donde la estadística entra en juego para inferir las propiedades que tienen estas distribuciones.

La Estadistica en simples palabras es esta idea. Tomamos una colección de datos, los cuales asumimos que son aleatorios de algun conjunto de muestra cuya distribución desconocemos. En esto nos gustaría desarrollar los metodos para *inferir* estas propiedades a partir de la información obtenida. Más aún es como si estas proveniesen de una generalización de distribuciones a las cuales denominaremos *Familias de distribuciones* y medir cuan seguros estamos en esta primicia.

Estadística es una rama muy rica y nosotros solo tocaremos los tópicos que nos serán relevantes para nuestro tabajo con Geometría Algebraica por lo que a continuación definiremos la nociones más básicas de esta.

4.1. Modelos Estadísticos.

Definición 4.1. Un modelo estadístico $\mathcal{M}$ es una colección de distribuciones probabilisticas o de funciones de densidad que pueden ser catalogados como

- Paramétrico: Cuando $\mathcal{M}_\Theta$ es un mapeo de un espacio de parámetros de dimensión finita $\Theta \subseteq \mathbb{R}^d$ a un espacio de distribuciones de probabilidad o de funciones de densidad. Es decir,

$$p_\bullet : \Theta \to \mathcal{M}_\theta, \quad \theta \mapsto p_\theta$$

  De esta forma, el modelo es la imagen del mapeo $p_\bullet$, $\mathcal{M}_\theta = \{p_\theta : \theta \in \Theta\}$
- No paramétrico: Cuando Θ es de dimensión infinita o no euclidiana.
- Modelo Semi-paramétrico: Cuando $\Theta \approx \Theta_1 \times \Theta_2$, donde Θ_1 es de dimensión finita y Θ_2 es de dimensión infinita.

Un estadístico es *identificable* si el mapeo $p_\bullet$ es inyectivo (uno a uno).

Ejemplo 4.2. (Variable aleatoria binomial). Sea X una variable aleatoria discreta con $r+1$ estados, que denotaremos por $0, 1, ..., r$. Sea $\Theta = [0,1]$ y para $\theta \in \Theta$ considere la distribución

$$P_\theta(X=i)=\binom{r}{i}\theta^i(1-\theta)^{r-i}$$

La cual es la distribucion de una variable aleatoria binomial con r ejemplos y un parámetro θ. El modelo de variables aleatorias binomiales $\mathcal{M}$ consiste en todas las distribuciones probabilisticas se plantean de esta manera.

Ejemplo 4.3. (Vector aleatorio normal para mupltiples varibales). Sea $X \in \mathbb{R}^m$ un vector real n-dimensional. Sea $\Theta = \mathbb{R}^m \times PD_m$, donde PD_m es un cono de matrices simétricas definidas potivas $m\times m$. Para $\theta = (\mu, \Sigma) \in \Theta$ sea

$$p_\theta(x)\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

La función de densidad de una variable aleatoria conjuntamente normal con esperanza μ y matriz de covarianza Σ. El modelo $\mathcal{M}_\theta$ consiste en todas las funciones de densidad para un vector real de dimensión m.

Definición 4.4. Definimos un estadístico T como una función, que va del espacio de estados de una variable aleatoria a otro conjunto. Para modelos estadisticos paramétricos $\mathcal{M}_\theta$ un estadistico T es suficiente para un modelo si

$$P(X=x|T(X)=t,\theta)=P(X=x|T(X)=t)$$

Equivalentemente, la distribución conjunta $p_\theta(x)$ se puede factorizar como

$$p_\theta(x)=f(x)g(T(x),\theta)$$

Donde f es una función que **no** depende de θ. Un estadistico T es minimal suficiente si todo estadístico distinto de T es una función de T ($T=g(S)$).

4.2. Estimación de parámetros. Dado un modelo estadístico paramétrico y un conjunto de datos. Un problema clásico en la estadística es estimar uno o todos los parámetros del modelo basado en los datos. Idealmente, nos gustaría hallar un procedimiento en el cual, a medida que aumenta el tamaño de nuestro conjunto de datos, si la distribución de los iniciales proviene del modelo, el parámetro estimado converja al verdadero parámetro subyacente. Tal estimador lo llamaremos consistente.

Definición 4.5. Sea $\mathcal{M}_\theta$ un modelo estadístico paramétrico con un espacio de parámetros Θ. Un parámetro de un modelo estadístico es una función $s:\Theta\to\mathbb{R}$. Un estimador de s es una función del espacio de datos D a $\mathbb{R}$, $\hat{s}:D\to\mathbb{R}$. El estimador $\hat{s}$ es consistente si $\hat{s}\to_p s$ a la medida que el conjunto de datos tiende al infinito.

Definición 4.6. Sea T un estadístico y $X=(X^{(1)},...,X^{(n)})$ los datos. T se dice completo si dada una transformación h tal que

$$\int h(T)dP_{T,\theta}=0, \quad \forall\theta\in\Theta$$

Entonces, $h(T)=0$

Teorema 4.7. *Todo Estadístico suficiente y completo es minimal*

Definición 4.8. Sea D Un conjunto de datos de algún modelo con espacio de parámetros Θ. La función de verocimilitud

$$L(\theta|D):=p_\theta(D) \quad (Caso\ Discreto)$$

$$L(\theta|D:=f_\theta(D)) \quad (Caso\ Continuo)$$

Aquí $p_\theta(D)$ es la probabilidad de observar el conjunto de datos D dado el parámetro θ en el caso discreto y $f_\theta(D)$ es la función de densidad evaluada en D en el caso continuo. El estimador Máximo verosimil (abreviado MLE en inglés) $\hat{\theta}$ es el parámetro que maximiza la función de verosimilitud. Es decir,

$$\hat{\theta}=\operatorname*{argmax}_{\theta\in\Theta} L(\theta|D)$$

Observe, que consideramos la función de verosimilitud como una funcion respecto a θ con el conjunto de datos D fijo. Esto contrasta la interpretación de distribución de probabilidad, donde el parámetro es considerado fijo y la variable aleatoria es una incognita.
En el caso de tener observaciones i.i.d. $D = (X^{(1)}, ..., X^{(n)})$ la función de verosimilitud se puede factorizar como

$$L(\theta|D) = L(\theta|X^{(1)}, ..., X^{(n)}) = \prod_{i=1}^{n} L(\theta|X^{(i)})$$

En el caso de datos discretos, esta función verosimilitud es así una función vectorial que cuenta $u = (u_0, ..., u_r)$, donde $u_i = \#\{j : X^{(j)} = i\}$ y se tiene

$$L(\theta|X^{(1)}, ..., X^{(n)}) = \prod_{j} p_\theta(j)^{u_j}$$

5. Una introducción a la estadística Algebraica desde la familia Exponencial

5.1. Familia exponencial. Luego de pasar por las secciones anteriores, al fin tenemos herramientas para poder estudiar la Familia exponencial. La cual juega un rol importante en la estadistica moderna, porque entrega un extenso marco de referencia para describir modelos estadísticos. Es por esto que la Familia exponencial es una de las más estudiadas (incluso tambien las probabilidades conjuntas proveniente de distribuciones normales).

En lo que sigue, veremos una gran cantidad de modelos estadisticos que son intepretados naturalmente como submodelos en 2 modos: primero que el parámetro del modelo pertenece al espacio de parámetros, y que los estadisticos suficientes de la familia exponencial mapea los datos a un cono de estadisticos suficientes del largo de la Familia exponencial misma. Este último hecho permite el analisis matemático de muchos modelos estadisticos interesantes y complejos a través de la Familia exponencial regular.

La literatura sobre Estadística por su parte ha sido enfocada po mucho tiempo en el caso en que los submodelos de la familia exponencial vienen de variedades suaves del espacio de parámetros. Estos son conocidos como Familias exponenciales curvadas. Es por ello, que nos enfoco será el caso donde es un subconjunto semi-algebraico del espacio de parámetros natural de la familia exponencial a la cual llamaremos *Familia exponencial Algebraica.*

5.2. Familia exponencial Regular. Sea X un espacio muestral con $\sigma-$algebra $\mathcal{A}$ en el cual es definida una medida $\sigma-$finita ν. Sea un estadístico $T : X \to \mathbb{R}^k$ (función medible), y $h(x) : X \to \mathbb{R}_{>0}$ una función medible. Definimos el espacio natural de parámetros como

$$N = \left\{ \eta \in \mathbb{R}^k : \int_X h(x) \cdot e^{\eta^t T(x)} d\nu(x) < \infty \right\}$$

Para $\eta \in N$, podemos definir la función de probabilidad p_η en X como

$$p_\eta(x) = h(x) e^{\eta^t T(x) - \phi(\eta)}$$

Donde

$$\phi(\eta) = log \int_X h(x) e^{\eta^t T(x)} d\nu(x)$$

Sea P_η la medida de probabilidad en $(X, \mathcal{A})$ que tiene ν-densidad p_η. Definimos $\nu^T = \nu \circ T^{-1}$ como la medida inducida por el estadístico T en la $\sigma - algebra$ de Borel en $\mathbb{R}^k$. El soporte de ν^T es la intersección de todos los conjuntos cerrados $A \subset \mathbb{R}^k$, que cumplen $\nu^T(\mathbb{R}^k \backslash A) = 0$. La función $\phi(\eta)$ puede ser interpretado como el logaritmo de la transformación de Laplace de ν^T.Recuerde que la dimensión afín de $A \subseteq \mathbb{R}^k$ es la dimensión del espacio lineal generado por todas las diferencias $x - y$ de dos vectores $x, y \in A$.

Definición 5.1. Sea k un entero positivo. Las distribuciones de probabilidad $(P_\eta : \eta \in N)$ forma una Familia exponencial regular de orden k si N es un conjunto abierto de $\mathbb{R}^k$ y la dimensión afín del soporte de ν^T es k. El estadístico $T(x)$, que induce la Familia exponencial Regular es llamado *estadístico suficiente canónico.*

El orden de una Familia exponencial Regular es único y si la misma familia es representada usando dos estadísticos suficientes canonicos, entonces estos son transformaciones no singulares afines entre sí.

Ejemplo 5.2. (Variables aleatorias discretas). Sea X un espacio muestral y $[r]$ un conjunto de enteros. Sea ν la medida de conteo en X, es decir, la medida $\nu(A)$ de $A \subset X$ es igual a la cardinalidad de A. $T : X \to \mathbb{R}^{r-1}$

$$T(x) = (1_{\{1\}}, ..., 1_{\{r-1\}(x)})^t$$

Cuyas indicatrices 1_im $i \in \{1, ..., r-1\}$ nos van a indicar cual valor en X al que el argumento x es igual. En partícular, cuando $x = r$, $T(x)$ es el vector nulo. Sea h(x) = 1 para todo $x \in [r]$. La medida inducida ν^T es la medida en la $\sigma - algebra$ (de Borel) de $\mathbb{R}^{r-1}$ con soporte igual a los r vectores en $\{0,1\}^{r-1}$ tales que tienen al menos una componente distinta de cero. Las diferencias entre estos r vectores incluyen toda la base canonica de vectores de $\mathbb{R}^{r-1}$. Así la dimensión afín del soporte de ν^T es igual a $r-1$. Además satisface $\forall \eta \in \mathbb{R}^{r-1}$

$$\phi(\eta) = log \int_X h(x) e^{\eta^t T(x)} d\nu(x) = log \left(1 + \sum_{x=1}^{r-1} e^{\eta x}\right) < \infty$$

Así el Espacio natural de parámetros N es igual a $\mathbb{R}^{r-1}$ y en partícular es abierto. La $\nu - densidad$ p_η es un vector de probabilidad en $\mathbb{R}^r$. Los componentes de $P_\eta(x)$ para $1 \geq x \geq r-1$ son positivos y dados por

$$p_\eta(x) = \frac{e^{\eta x}}{1 + \sum_{x=1}^{r-1} e^{\eta x}}$$

El último componente de p_η es también positivo e igual

$$p_\eta = 1 - \sum_{x=1}^{r-1} p_\eta(x) = \frac{1}{1 + \sum_{x=1}^{r-1} e^{\eta x}}$$

La Familia de probabilidades inducidas en distribución $(P_{eta} : \eta \in \mathbb{R}^{r-1})$ es una Familia exponencial Regular de orden $r-1$. Los parámetros naturales η_x pueden ser interpretados como la razón logarítmica de probabilidades, pues p_η es igual a un vector positivo de probabilidad $(p_1, ..., p_r)$ si y solo si $\eta_x = log(p_x/p_r)$ para $x = 1, ..., r-1$. Esto establece una correspondencia entre el espacio natural de parámetros $N = \mathbb{R}^{r-1}$ y el interior del simplex de probabilidades de dimensión $r-1$, Δ_{r-1}.

Ejemplo 5.3. (Variables aleatorias normales multivariadas). Sea X un espacio muestral euclidiano de dimensión n, dotado con su $\sigma - algebra$ de Borel y medida de *Lebesgue* ν. Consideremos el estadístico $T : X \to \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2}$ dado por

$$T(x) = (x_1, ..., x_m, -x_1^2/2, ..., -x_m^2/2, -x_1 x_2, ..., -x_{m-1} x_m)^t$$

Las funciones polinomiales que forman los componentes de $T(x)$ son linealmente independientes y así el soporte de ν^T tiene toda la dimensión afín $m + \frac{m(m+1)}{2}$. Sea $h(x) = 1$ para todo $x \in \mathbb{R}^m$.
Si $\eta \in \mathbb{R}^m \times \mathbb{R}^{m(m+1)/2}$, entonces denotamos $\eta_{[m]} \in \mathbb{R}^m$ al vector de los primeros p componentes η_i, $1 \geq i \geq m$. Similarmente, escribimos $\eta[m \times m]$ a la matriz simétrica formada por los últimos $m(m+1)/2$ componentes η_{ij} $1 \geq i \geq j \geq m$. La función $x \mapsto e^{\eta^t T}$ es $\nu - integrable$ si y solo si $\eta_{[m \times m]}$ es definida positiva. Así, el espacio natural de parámetros N es igual al producto cartesiano de $\mathbb{R}^m$ y el cono de matrices definidas positivas de $m \times m$. Si η es en el espacio abierto N, entonces

$$\phi(\eta) = -\frac{1}{2}\left(\log det(\eta_{[m \times m]}) - \eta_{[m]}^t \eta_{[m \times m]_{\eta_{[m]}}} - m \log(2\pi)\right)$$

Su densidad de Lebesgue p_η puede ser escrito como

$$p_\eta(x) = \frac{exp\left\{\eta_{[m]}^t x - tr(\eta_{[m \times m]} x x^t)/2 - \eta_{[m]}^t \eta_{[m \times m]} \eta_{[m]}/2\right\}}{\sqrt{(2\pi)^m det(\eta_{[m \times m]}^{-1})}}$$

En esta expresión, podemos usar el truco de la traza para reescribirla como $tr(\eta_{[m \times m]} x x^t)/2 = x^t \eta_{[m \times m]} x$. Esto nos permite escribir la forma cuadratica en x como un producto punto entre $\eta_{[m \times m]}$ y el vector del estadisctico suficiente. Tomando $\Sigma = \eta_{[m \times m]}^{-1}$ y $\mu = \eta_{[m \times m]} \eta_{[m]}$, se tiene que

$$\frac{1}{\sqrt{(2\pi)^m det(\Sigma)}} exp\left\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right\}$$

Es la función de densidad de la distribución normal multivariada $\mathcal{N}_m(\mu, \Sigma)$. Así la familia de distribuciones normales multivaradas en $\mathbb{R}^m$ con matriz de covarianza definida positiva es una Familia exponencial Regular de orden $m\binom{m+1}{2}$.

Observación 5.4. La inversa de la matriz de covarianza Σ, juega un rol fundamental en los modelos Gaussianos como un parámetro natural de la Familia a través de su aparición en restricciones de independencia condicionales saturadas. Por esta razón, la matriz inversa recibe un nombre especial: *Matriz de concentración* o *Matriz de presición*. Esta es denotada como $K = \Sigma^{-1}$.

En algunas circunstancias es util considerar la clausura de una Familia exponencial en una representación adecuada. Esto ocurre usualmente en el caso de Familias exponenciales Discretas, donde vimos que la representación del modelo multinomial entrega el interior de un simplex de probabilidad. Por otra parte, el conjunto de todas las distribuciones que pertenecen en la clausura de una Familia exponencial Regular es llamada *Familia exponencial Extendida*. Esta última involucra distribuciones de probabilidad que no poseen funciones de densidad. Por ejemplo en el caso Gausseano, el operador clausura posee matrices de covarianza que son semi-definidas positivas, pero singulares. Estos poseen una distribución de probabilidad que son soportados en espacios vectoriales con menor dimensionalidad, y de este modo, no tienen funciones de densidad.

Todo lo anterior fue para poder describir modelos estadístico que pueden ser intepretados naturalmente como sub-modelos de Familias exponenciales a través de su espacio natural de parámetros o versiones transformadas de estos espacios de parámetros.

5.3. Familia exponencial Regular Discreta (FERD).

En esta parte, exploraremos la FERD para variables aleatorias discretas con un número finito de estados, y como estan relacionados a los objetos en algebra. Del hecho de que X se asumirá Discreto con espacio de estados $\Omega = [r]$, existe un número de simplificaciones que pueden ser hechas a la representación de una Familia exponencial Regular de la presentada en la parte anterior. Primero, un estadístico $T(x)$ da lugar a un mapeo medible de $[r]$ a $\mathbb{R}^k$. Esto es asociar un vector $T(x) \in \mathbb{R}^k$ a cada $x \in [r]$. De manera análoga, la función medible $h : [r] \to \mathbb{R}_{>0}$ simplemente asocia un valor real positivo a cada $x \in [r]$.

Observación 5.5. Cualquier función no negativa es medible en este contexto y cualquier lista de vectores $T(1), ..., T(r)$ puede ser usada. La transformada de laplace $\int_\Omega h(x)e^{\eta^t T(x)} d\nu$ es la sumatoria

$$Z(\eta) = \sum_{x\in[r]} h(x)e^{\eta^t T(x)}$$

Ahora veamos los elementos individuales del producto. Llamaremos nosotros al vector $T(x) = a_x$, escrito explícitamente como $a_x = (a_{1x}, ..., a_{kx})^t$. De manera similar, $\eta = (\eta_1, ..., \eta_k)^t$ y $\theta_i = exp(\eta_i)$. Por simplicidad, denotamos $h(x) = h_x$ de modo que $h = (h_1, ..., h_r)$ es un vector en $\mathbb{R}^r_{>0}$. Esto quiere decir, que

$$p_\eta(x) = h(x)e^{\eta^t T(x) - \phi(\eta)}$$

Puede ser reescrito como

$$p_\theta(x) = \frac{1}{Z(\theta))} h_x \prod_j \theta_j^{a_{jx}}$$

Donde

$$Z(\theta) = \sum_{x\in[r]} h_x \prod_j \theta_j^{a_{jx}}$$

En partícular, si a_{jx} son enteros para todo j y x, entonces la Familia exponencial Discreta es una familia de distribuciones de probabilidad parametrizada, donde las funciones paramétricas son funciones racionales.
Una Tercera representación equivalente del modelo en su contexto discreto viene de determinar si dada una distribución de probabilidad $p \in int(\Delta_{r-1})$, esta pertenece a la familia exponencial Discreta asociada a $A =$

$(a_{jx})_{j\in[k],x\in[r]} \in \mathbb{Z}^{k\times r}$ y $h \in \mathbb{R}^r_{>0}$. De hecho, tomando el logaritmo de la representación de la familia exponencial

$$\log p_\eta(x) = \log h(x) + \eta^t T(x) - \phi(\eta)$$

O equivalentemente,

$$\log p_\theta(x) = \log h_x + \log(\theta)^t A - \log Z(\theta)$$

Si asumimos, que la matriz A contiene un vector $\mathbf{1}$= $(1, 1, ..., 1)$ en su espacio generado , entonces esto es equivalente a requerir que $\log p + rowspan(A)$ donde $rowspan(A)$ corresponde al espacio generado por los vectores filas de la matriz A. Por esta razón, las Familias exponenciales Discretas son llamados modelos log $-$afín. Esta conexión también explica la interpretación de estos modelos como provenientes de tomar sub-espacios vectoriales del espacio natural de parámetros del FERD Completo.

Definición 5.6. Sea $A \in \mathbb{Z}^{k\times r}$ una matríz de enteros tal que $\mathbf{1}\in$ $rowspan(A)$ y sea $h \in \mathbb{R}^r_{>0}$. El modelo log-afín asociado a esta información es el conjunto de distribuciones probabilisticas

$$\mathcal{M}_{A,h} := \{p \in int(\Delta_{r-1}) : \log p \in \log h + rowspan(A)\}$$

Si $h = 1$, entonces $\mathcal{M}_A = \mathcal{M}_{A,\mathbf{1}}$ es llamado modelo $log - lineal$

En adición a los nombres de las Familias exponenciales y modelos log-linal, estos modelos son también conocidos en la comunidad de Estadística Algebraica como Modelos Tóricos.

Definición 5.7. Sea $h \in \mathbb{R}^r_{>0}$ un vector de números positivos y $A = (a_{ij}) \in \mathbb{Z}^{k\times r}$ una matriz de enteros y suponemos que $\mathbf{1} \in rowspan(A)$. El mapeo asociado a estos datos es el mapeo racional

$$\phi^{A,h} : \mathbb{R}^k \to \mathbb{R}^r,\ Donde \quad \phi^{A,h}_j = h_j \prod_{i=1}^k \theta_i^{a_{ij}}$$

Note que la definición del mapeo racional $\phi^{A,h}$ hemos removido la constante normalizadora $Z(\theta)$. Esto tiene un efecto homogeneizar la parametrización por lo que podemos cálcular el ideal homogeneo que se cancela del modelo.

Definición 5.8. Sea $h \in \mathbb{R}^r_{>0}$ un vector con valores positivos y $A = (a_{ij}) \in \mathbb{Z}^{k\times r}$ una matriz de $k \times r$. El ideal

$$I_{A,h} := I(\phi^{A,h}(\mathbb{R}^k)) \subset \mathbb{R}[p]$$

Es llamado Ideal Tórico asociado al par A y h. En el caso especial en que $h = 1$, denotamos esto como $I_A := I_{A,\mathbf{1}}$

Note que los generadores del Ideal Tórico $I_{A,h}$ se puede obtener fácilmente de loss generadores del Ideal I_A, haciendo sustituciones $p_j \to \frac{p_j}{h_h}$. Así esto es suficiente para enfocarnos en el caso del ideal tórico I_A. Estos ideales resultan ser ideales binomiales.

Proposición 5.9. *Sea $A \in \mathbb{Z}^{k\times r}$ una matriz de enteros. Entonces el ideal tórico I_A es un ideal binomial y*

$$I_A = \langle p^u - p^v : u, v \in \mathbb{N}^r,\ y\ Au = Av\rangle$$

Si $\mathbf{1}\in rowspan(A)$, entonces I_A es homogéneo.
Demostración:
Es claro que cualquier binomio $p^u - p^v$ tal que $Au = Av$ es en el ideal tórico I_A. Para mostrar que I_A es generado por estos binomios, usaremos algo más fuerte: El hecho de que estos binomios generan el ideal I_A como un espacio vectorial. De hecho, sea $f(p) \in I_A$ y $c_u p^u$ un monomio cualquiera que "aparezca" en f con un coeficiente c_u no nulo. Para cualquier θ, se debiese satisfacer $f(\phi^A(\theta)) = 0$. Esto implica, que $c_u p^u \Rightarrow c_u \theta^{Au}$. Para cancelar esto, tomemos un mismo monomio (que no necesariamente satisfaga $c_u = c_v$) $c_v p^v$ Tal que cumpla $f(\phi^A(\theta)) = 0$.

$$\Rightarrow Au = Av$$

De modo que podemos formar otro polinomio con menos terminos

$$f - c_u(p^u - p^v)$$

Luego, de acuerdo a los números de términos que posea nuestro f, el resultado se puede alcanzar mediante inducción.
Finalmente, para ver que I_A es homogeneo, observe que $\mathbf{1} \in rowspan(A)$ fuerza a que $Au = Av$ implique $\mathbf{1}u = \mathbf{1}v$. Así todos los binomios generados son homogeneos.

Observación 5.10. Los ideales tóricos son también ideales reticulados.
De hecho, $I_A = I_{\mathcal{L}}$ donde $\mathcal{L}$ es el reticulado $\mathcal{L} = ker_{\mathbb{Z}}A$. Además, si $h = \mathbf{1}$, entonces los generadores de los ideales pueden ser elegidos a ser binomios cuyos coeficientes son±1, por lo que el cuerpo de coeficientes del anillo no importan realmente.

Ejemplo 5.11. (Cubo torcido). Sea $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}$. El ideal tórico I_A es el ideal que se anula con la parametrización:

$$p_1 = \theta_2^3, \quad p_2 = \theta_1\theta_2^2, \quad p_3 = \theta_1^2\theta_2, \quad p_4 = \theta_1^3$$

El ideal tórico es generado por tres binomios cuadraticos

$$I_A = \langle p_1p_3 - p_2^2, p_1p_4 - p_2p_3, p_2p_4 - p_3^2 \rangle$$

Por ejemplo el primer binomio $p_1p_3 - p_2^2$ es de la forma $p^u - p^v$ donde $u = (1,0,1,0)^T$ y $v = (0,2,0,0)^T$. Además, es fácil verificar que $Au = Av$ en este caso. El estadístico suficiente de este modelo, es obtenido del vector de conteo de u, es ell vector Au.

Observación 5.12. La variedad $V(I_A) \subset \Phi^3$ el cubo torcido suave es el que vimos en la parte de geometría algebraica en su forma afín. Además note que por su parte, si tomamos $h = (1,3,3,1)$, entonces $I_{A,h}$ entrega el ideal que se anula del modelo de una variable binomial aleatoria con tres intentos.

Ejemplo 5.13. (Variables Aleatorias discretas e independientes). Siempre que podemos tener una variedad parametrizada por funciones monomiales de parámetros, el ideal que se anula es un ideal tórico. Por ejemplo, considere la parametrización

$$p_{ij} = \alpha_i\beta_j$$

Donde, $i \in [r_1]$, $j \in [r_2]$ y además, α_i y β_j son parámetros independientes. Esto es, la parametrización (homogenea) de dos variables aleatorias discretas e independientes. La matriz A que representa el ideal tórico es una matríz $(r_1 + r_2) \times (r_1r_2)$ con una fila para cada parámetro y una columna para cada estado de cada par de variables aleatorias. La columna ij de A, da el vector exponente para $p_{ij} = \alpha_i\beta_j$ es el vector $e_i \oplus e_j^{'}$ donde e_i denota el vector de la base canónica i en $\mathbb{R}^{r_1}$ y análogamente, $e_j^{'}$el vector de la base canónica j en $\mathbb{R}^{r_2}$. Tomemos el ejemplo con $r_1 = 3$ y $r_2 = 4$. La matriz A es

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

Observe que la matriz A actúa como un operador lineal que calcula las sumas de filas y columnas del arreglo $r_1 \times r_2$ u. El ideal tórico I_A es generado por todos los binomios $p^u - p^v$ tal que u,v tienen los mismos estadísticos suficientes en el modelo. Esto indíca que los arreglos u y v tienen las mismas sumas de columnas y filas. Note que $I_A = I_{1 \perp\!\!\!\perp 2}$ por lo que I_A es generado por los subdeterminantes de 2×2 de la matriz generica

$$I_A = \langle p_{i_1j_1}p_{i_2j_2} - p_{i_1j_2}p_{i_2j_1} : i_1, i_2 \in [r_1], j_1, j_2 \in [r_2] \rangle$$

Ejemplo 5.14. (Independencia Completa). Sea A una matriz de 6×8

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

El siguiente código singular calcula una base de Gröbner para el ideal tórico I_A

```
ring r = 0, (p1,p2,p3,p4,p5,p6,p7,p8),dp;
intmat A[6][8] = 1,1,1,1,0,0,0,0,
0,0,0,0,1,1,1,1,
1,1,0,0,1,1,0,0,
0,0,1,1,0,0,1,1,
1,0,1,0,1,0,1,0,
0,1,0,1,0,1,0,1;
intvec v = 1,1,1,1,1,1,1,1;
LIB "toric.lib";
ideal I = toric_ideal(A,"hs",v);
std(I);
```

Usando los resultados obtenidos a través del código, se puede afirmar que el ideal I_A tiene nueve generadores cuadraticos homogeneos

$$p_6p_7 - p_5p_8, ..., p_2p_3 - p_1p_4$$

Renombrando las variables como $p_{111}, ..., p_{222}$, podemos ver que I_A es igual al ideal de independencia condicional $I_{\mathcal{C}}$ de completa independencia de tres variables aleatorias binarias. Esto es

$$\mathcal{C} = \{1 \perp\!\!\!\perp \{2,3\}, 2 \perp\!\!\!\perp \{1,3\}, 3 \perp\!\!\!\perp \{1,2\}\}$$

Observación 5.15. El singular tiene una libreria construida para cálculo con ideales tóricos. Macaulay2 también tiene esa funcionalidad con la librería $4ti2$.

5.4. Geometría Algebraica Real.

Definición 5.16. Sea $\mathcal{F}$,$\mathcal{G} \subset \mathbb{R}[p_1, ..., p_r]$ conjuntos finitos de polinomios. La base semialgebraica definida por $\mathcal{F}$ y $\mathcal{G}$ es el conjunto

$$\{a \in \mathbb{R}^r : f(a) = 0, \quad \forall f \in \mathcal{F} \ y \ g(a) > 0, \forall g \in \mathcal{G}\}$$

Un conjunto semialgebraico es cualquier conjunto que pueda ser escrito como una union finita de semialgebras básicos. En el caso en que $\mathcal{G} = \emptyset$ el conjunto semialgebraico resultante es llamado *variedad real algebraica.*

Ejemplo 5.17. (Simplice de probabilidad). El interior de un simplice de probabilidad es un conjunto semialgebraico con $\mathcal{F} = \{p_1 + \ldots + p_r - 1\}$ y $\mathcal{G} = \{p_i : i \in [r]\}$. El simplice de probabilidad Δ_{r-1} es un conjunto semialgebraico, pues se puede obtener al unir $2^r - 1$ de semialgebras básicas las que consisten en el interior relativo de las caras de Δ_{r-1}.
En general, si $\mathcal{F}$ y $\mathcal{G}$ son conjuntos finitos de polinomios, un conjunto de la forma

$$\{a \in \mathbb{R}^r : f(a) = 0, \quad \forall f \in \mathcal{F} \ y \ g(a) > 0, \forall g \in \mathcal{G}\}$$

es comunmente llamado *conjunto semialgebraico básico cerrado.* Un importante ejemplo de este tipo de semialgebra es el conjunto de poliedros, donde ambos conjuntos, $\mathcal{F}$, $\mathcal{G}$ consisten en polinomios lineales.

Ejemplo 5.18. (Conjunto semialgebraico no básico). Sea $\Theta = D \cup R$ la union del disco $D = \{p \in \mathbb{R}^2 : p_1^2 + p_2^2 \leq 1\}$ y el rectangulo $R = \{p \in \mathbb{R}^2 : 0 \leq p_1 \leq 1, -1 \leq p_2 \leq 1\}$. Este conjunto semialgebraico no puede ser escrito como un conjunto de la forma $\{p \in \mathbb{R}^2 : g(p) \geq 0, \forall g \in \mathcal{G}\}$ para una unión finita de polinomios $\mathcal{G}$. De hecho, desde que la mitad de la circunsferencia $p_1^2 + p_2^2 = 1$ forma parte de la frontera de Θ, el polinomio $g = 1 - p_1^2 - p_2^2$ podría requerir pertenecer a $\mathcal{G}$ o como factor de algun polinomio en $\mathcal{G}$. No todo punto en Θ satisface $g \geq 0$, y de manera similar, los de la forma fg tampoco.

La mayor parte de los modelos de estadística algebraica vienen de tomar un espacio de parámetros Θ, donde este es un poliedro y el modelo será la imagen de Θ bajo un mapeo racional. Como podemos ver, estos conjuntos son siempre conjuntos semialgebraicos.

Teorema 5.19. *(Tarski-Seindenberg). Sea $\Theta \subseteq \mathbb{R}^{r+1}$ un conjunto semialgebraico y sea $\pi : \mathbb{R}^{r+1} \to \mathbb{R}^r$ una proyección de cordenadas. Entonces $\pi(\Theta) \subseteq \mathbb{R}^r$ es también un conjunto semialgebraico.*

El teorema de Tarski$-$Seidenberg es la base de muchos resultados útiles en geometría algebraica real. En partícular, si se construye un grafo de un mapeo racional y se aplica varias veces este teorema se deduce:

Teorema 5.20. *Sea $\phi : \mathbb{R}^{r_1} \to \mathbb{R}^{r_2}$ un mapeo racional y $\Theta \subseteq \mathbb{R}^{r_1}$ un subconjunto semialgebraico. Entonces la imagen de Θ bajo ϕ*

$$\phi(\Theta) := \{\phi(\theta) : \theta \in \Theta, y\ \phi(\theta)\ Es\ bien\ definido\}$$

es un subconjunto semialgebraico de $\mathbb{R}^{r_2}$.
Demostración:
Sea $\Gamma_\phi : \{(\theta, \phi(\theta)) : \theta \in \Theta\}$ el grafo de ϕ. Desde que ϕ es dada por funciones racionales, $\phi_i = \frac{f_i}{g_i}$ este es un conjunto semialgebraico (el cual viene de tomar las representaciones de Θ como conjunto semialgebraico y agregando las restricciones pertinentes para que $y_i g_i(\theta) - f_i(\theta) = 0$, y entonces intersectando los conjuntos de semialgebras, $\prod g_i(\theta) \neq 0$). La imagen $\phi(\Theta)$ es la proyección del grafo Γ_ϕ.

Definición 5.21. El cono tangente $TC_{\theta_0}(\Theta)$ de un conjunto semialgebraico $\Theta \subseteq \mathbb{R}^k$ en el punto $\theta_0 \in \Theta$ es el conjunto de limites de las sucesiones $\alpha_n(\theta_n - \theta_0)$ donde α_n es una secuencia de valores reales positivos y $\theta_n \in \Theta$ converge a θ_0. Los elementos del cono tangente son llamados *vectores tangentes*. Diremos que θ_0 es un punto suave de Θ si $TC_{\theta_0}(\Theta)$ es un espacio vectorial.

Observación 5.22. El cono tangente es un conjunto cerrado, y es literalmente un cono. Esto por el simple hecho de que al tomar cualquier vector en el cono gangente y multiplicarlo por un escalar positivo, este nuevo vector aún pertenece al cono tangente.

Definición 5.23. Sea $\mathbb{K}$ un cuerpo algebraicamente cerrado, e $I \subseteq \mathbb{K}[p]$ un ideal. Supongamos que el origen esta en la variedad afín $V(I)$. Entonces, el *cono tangente algebraico* es definido como $TC_0(V(I)) = V(min(I))$ donde $min(I) := \langle min(f) : f \in I\rangle$

El cono tangente algebraico puede ser definido en cada punto de la variedad. Simplemente basta con tomar la serie de taylor de cada polinomio perteneciente al ideal en torno a dicho punto (o equivalentemente, aplicando la transformación afin para llevar el punto de interes al origen). Además, el cono tangente sobre los números complejos en nuestro punto, puede ser calculado directamente usando las bases de Gröbner.
La idea general es generalizar la siguiente observación:
Sea $f \in I(\Theta)$ y $\theta_0 \in \Theta$. Extender f como una serie de taylor sobre el punto θ_0 y de este modo, el termino más pequeño de la serie se anula en el cono tangente.

Ejemplo 5.24. (Cono Tangente en un Nodo Simple). Considere el polinomio $f = p_2^2 - p_1^2(p_1 + 1)$. La variedad $V(f)$ es el cubo nodal. El polinomio $min(f) = p_2^2 - p_1^2$, donde podemos observar que el cono tangente en el origen de $V(f)$ es la unión de dos lineas.
Para calcular el ideal $min(I)$ y tambien el cono tangente algebraico, se necesita del cálculo de las bases de Gröbner.

Proposición 5.25. *Sea $I \subseteq \mathbb{K}[p]$ un ideal. Sea $I^h \subset \mathbb{K}[p_0, p_1]$ una homogeneización de I con respecto al polinomio lineal p_0. Sea $\prec$ un orden de eliminación talque anula p_0, y sea $\mathcal{G}$ una base de Gröbner para I^h con respecto a $\prec$. Entonces,*

$$min(I) = \langle min(g(1,p)) : g \in \mathcal{G}\rangle$$

Ejemplo 5.26. (Un punto singular en una variedad tórica). Considere la parametrización monomial ϕ^A dada por la matriz

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 3 & 4 \end{pmatrix}$$

El punto (1,0,0,0) pertenece a la variedad V_A y es un punto singular. El siguiente código en Macaulay2 calcula el cono tangente en este punto.

```
S = QQ[t1,t2];
R = QQ[p1,p2,p3,p4];
f = map(S,R,{t1, t1*t2^2, t1*t2^3,t1*t2^4});
g = map(R,R,{p1+1, p2,p3,p4});
h = map(R,R,{p1-1, p2,p3,p4});
I = kernel f
Ishift = g(I)
T = tangentCone Ishift
Itangetcone = h(T)
```

Los mapeos g y h estaban involucradas en la traslación de $(1,0,0,0)$ al origen. El ideal resultante es $\langle p_3^2, p_4\rangle$. Por esto, el cono tangente es un plano bidimensional, pero con multiplicidad 2.

Observación 5.27. El cono tangente algebraico sobre números complejos solo proporciona un poco de información sobre el cono tangente real. De hecho, si $\Theta \subset R^k$ es un conjunto semialgebraico, $I(\Theta)$ el ideal que se anula, y $0 \in \Theta$, entonces la variedad algebraica real $V(min(I(\Theta)))$ contiene el cono tangente $TC_0(\Theta)$, pero no necesariamente son iguales en general.

Ejemplo 5.28. (Cono tangente en una cúspide). Sea $f = p_2^2 - p_1^3$ y $V(f)$ el cubo cuspidal en un plano. Entonces $min(f) = p_2^2$ y así el cono tangente complejo es el eje p_1. Por otra parte, cada punto en $V_{\mathbb{R}}(f)$ satisface $p_1 \geq 0$ y así el cono real tangente satisface esta restricción. De esta manera, el cono tangente del cubo cuspidal en el orígen es la parte positiva del eje p_1.

Generalizando el ejemplo anterior, es posible obtener información del cono tangente de conjuntos semialgebraicos en puntos más generales mirando los polinomios que son positivos en los puntos dados.

Proposición 5.29. *Sea Θ un conjunto semialgebraico. Supongamos $0 \in \Theta$ y sea $f \in \mathbb{R}[p]$ un polinomio tal que $f(\theta) \geq 0$ para todo $\theta \in \Theta$. Entonces, $min(f)(\theta) \geq 0, \quad \forall\theta \in TC_0(\Theta)$.*
Demostración:
Si $f(\theta) \geq 0, \forall\theta \in \Theta$, entonces haciendo $\theta \to 0$, se tiene $min(f)(\theta) > 0$. Esto sigue del hecho de que en $min(f)(\theta)$ dominan los terminos de mayor orden de f cuando θ es cercano al origen.

Luego de esto, hemos definido las nociones básicas de geometría algebraica real por lo que finalmente podemos establecer las bases de la *Familia exponencial Algebraica.*

5.5. Familia Exponencial Algebraica.

Definición 5.30. Sea $\{P_\eta : \eta \in N\}$ una familia exponencial regular de orden k. La sub-familia inducida por el conjunto $M \subseteq N$ es una *familia exponencial algebraica* si existe un conjunto $\overline{N} \subseteq \mathbb{R}^k$ tal que para el difeomorfismo $g : N \to \overline{N}$ y el conjunto semi-algebraico $A \subseteq \mathbb{R}^k$ se tiene $M = g^{-1}(A \cap \overline{N})$.

Observación 5.31. Por convención, nos referiremos al conjunto M, el conjunto $A \cap \overline{N}$ o la familia de distribuciones de probabilidad resultante, $\{P_\eta : \eta \in M\}$ como la familia exponencial algebraica.

Ejemplo 5.32. (Variables aleatorias discretas). Sea $\overline{N} = int(\Delta_{k-1})$ y considere el difeomorfismo

$$\eta \mapsto (exp(\eta_1), ..., exp(\eta_k))/\sum_{i=1}^{k} exp(\eta_i)$$

Donde $\eta_k = 1 - \sum_{j=1}^{k-1} \eta_j$. Esto muestra que cada conjunto semialgebraico del interior de un simplice de probabilidad en una familia exponencial algebraica.

Este ejemplo, nos mnuestra que no podemos tomar solo funciones racionales g en la definición. Es posible que g sea un difeomorfismo más complicado al momento de introducir una familia exponencial algebraica. Observe que puede ser util y provechoso tomar la clausura del conjunto $\overline{N}$ y el modelo $M \subseteq \overline{N}$. La razón para considerar conjuntos abiertos en este contexto es para aplicaciones al test de razón de verosimilitud.

Ejemplo 5.33. (Variables Aleatorias Gausseanas). Sea $N = \overline{N} = \mathbb{R}^m \times PD_m$ el cono de matrices definidas positivas y considere el difeomorfismo

$$\mathbb{R}^m \times PD_m \to \mathbb{R}^m \times PD_m, (\eta, K) \mapsto (K^{-1}\eta, K^{-1})$$

Esto muestra que cualquier subconjunto algebraico de $\mathbb{R}^m \times PD_m$ de esperanza y matrices de covarianza produce una familia exponencial algebraica. Observe que como el mapeo $K \mapsto K^{-1}$ es racional, podemos tener subconjuntos algebraicos de promedios y matrices de concentración: El teorema de Tarski-Seidenberg garantiza que los conjuntos semialgebraicos en un régimen también son semialgebraicos en el otro.
La necesidad de considerar modelos estadísticos como conjuntos semialgebraicos es claro, desde que podemos estudiar los modelos de variables ocultas.

Ejemplo 5.34. (Análisis Factorial). El modelo de Análisis factorial $\mathcal{F}_{m,s}$ es la familia miltivariada de distribuciones normales $\mathcal{N}_m(\mu, \Sigma)$ en $\mathbb{R}^m$, cuyo vector esperanza μ es un vector arbitrario en $\mathbb{R}^m$ y su matriz de covarianza pertenece al conjunto

$$F_{m,s} = \{\Omega + \Lambda\Lambda^T \in PD_m : \Omega \succ 0 \ diagonal, \Lambda \in \mathbb{R}^{m \times s}\}$$

Por notación, $\Omega \succ 0$ se refiere a que la matriz Ω es definida positiva. Note que $F_{m,s}$ es un conjunto semialgebraico por el hecho de que es la imagen de un conjunto semialgebraico $\mathbb{R}^m_{>0} \times \mathbb{R}^{m \times s}$ bajo un mapeo racional. De ahí, el modelo de analisis factorial es un modelo estadístico algebraico. Sin embargo, en general no es de la forma $PD_m \cap V(I)$ para cualquier ideal I. En partícular, la estructura semialgebraica realmente importa en este modelo.
Como ejemplo, considere el caso con $m = 3$, $s = 1$, un modelo 1-factor con tres covariables. Es decir, el modelo consiste en todas las matrices de covarianza de 3×3 de la fomra

$$\Sigma = \Omega + \Lambda\Lambda^T = \begin{pmatrix} \omega_{11} + \lambda_1^2 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ \lambda_1\lambda_2 & \omega_{22} + \lambda_2^2 & \lambda_2\lambda_3 \\ \lambda_1\lambda_3 & \lambda_2\lambda_3 & \omega_{33} + \lambda_3^2 \end{pmatrix}$$

Donde $\omega_{11}, \omega_{22}, \omega_{33} \geq 0$ y $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$.

Referencias

[EH16] David Eisenbud and Joe Harris, *3264 and All That: A Second Course in Algebraic Geometry*, Cambridge University Press, 2016.
[Har13] Joe Harris, *Algebraic Geometry: A First Course*, Springer Science and Business Media. **133** (2013).
[Vak17] Ravi Vakil, *The Rising Sea: Foundations of Algebraic Geometry*, Preprint (2017).
[Sul18] Sullivant, Seth. *Algebraic Statistics.* Vol. 194. American Mathematical Soc., 2018.
[DP98] Diaconis, Persi, and Bernd Sturmfels. *Algebraic Algorithms For Sampling From Conditional Distributions.* The Annals of statistics 26.1 (1998): 363-397.
[PGE01] Pistone, Giovanni, Eva Riccomagno, and Henry P. Wynn. *Algebraic Statistics*, volume 89 of Monographs on Statistics and Applied Probability. (2001).

Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile.
Email address: `eric.zepeda.14@sansano.usm.cl`