Modelos Exponenciales y Variedades Tóricas

(por Nicolás Botbol)

Introducción

Nuestros objetos de estudio serán algunos modelos estadísticos para un espacio de estados finito que denotaremos χ.

Más precisamente, consideraremos familias de distribuciones exponenciales, que son de la forma:

$P_\theta(x) = Z(\theta).e^{<\theta,T(x)>}$,

donde $x \in \chi$ y $\theta \in [-\infty, \infty)^d$ Z es una constante que normaliza y T es un estadístico suficiente para el modelo, $T : \chi \to \mathbb{Z}^d_{\geq 0}$ (en la referencia [2] figura $\mathbb{Z}^d - \{0\}$, lo cual -creo- es incosistente con la conformación de la matriz A -ver más adelante).

De ahora en más olvidearemos la constante normalizadora Z. (Por ser positiva podemos agregársela a la exponencial).

Será conveniente reformular estos modelos. Identificaremos χ con el conjunto $\{1, 2, ...m\}$.

Una distribución de probabilidad sobre χ es un vector $P = (p_1, ..., p_m)$, donde $p_i = P(i)$ en la notación anterior, $P \in \mathbb{R}^m_{\geq 0}$ y $\sum_{i=1}^m p_i = 1$. Llamaremos $sop(P) = \{i \in \chi / p_i \neq 0\}$. Sea $A \in \mathbb{Z}^{d\times m}_{\geq 0}$, tal que si $A = (A_1|...|A_m)$ (A_i la i-ésima columna de A), entonces $A_i = T(i)$, es decir, A es la matriz de T.

Esta matriz induce la siguiente aplicación monomial:

$\phi_A : \mathbb{R}^d_{\geq 0} \to \mathbb{R}^m_{\geq 0}$

$(t_1, ..., t_d) \mapsto (\prod_{i=1}^m t_i^{A_{i1}}, ..., \prod_{i=1}^m t_i^{A_{im}})$.

Definición: Decimos que P se factoriza a través de A sii $P \in Im(\phi_A)$.

Si pensamos que P es una función de θ (i.e. $P = P_\theta = (p_1(\theta), ..., p_m(\theta))$), entonces la definición anterior es equivalente a que exista una flecha ψ que haga conmutar el diagrama siguiente:

$$\begin{tikzcd}[-\infty,\infty)^d \arrow[rr, "P_\theta"] \arrow[dr, dashed, "\psi"] & & \mathbb{R}^m_{\geq 0} \\ & \mathbb{R}^d_{\geq 0} \arrow[ur, "\phi_A"] & \end{tikzcd}$$

Los dos modelos anteriores son equivalentes tomando $t_i = e^{\theta_i}$, donde $e^{-\infty} = 0$. Sustituyendo obtenemos:
$(e^{\theta_1}, ..., e^{\theta_d}) \mapsto^{\phi_A} (e^{\sum_{i=1}^m \theta_i . A_{i1}}, ..., e^{\sum_{i=1}^m e^{\theta_i} . A_{im}})$.

Distribuciones que se factorizan

Estudiaremos a continuación la estructura algebraica de los modelos exponenciales. Daremos una caracterización de aquellos distribuciones que se factorizan de acuerdo a A y de aquellos modelos que son límite de distribuciones que se factorizan.

De acuerdo a la notación anterior, llamaremos A_i a la i-ésima columna de A. Entonces $sop(A_i) \subseteq \{1, 2, ..., d\}$.

Definición: Si $F \subseteq \{1, 2, ..., d\}$, decimos que F es nice si para cada $j \notin F$, $sop(A_j) \nsubseteq \bigcup_{i \in F} sop(A_i)$.

Lema 1: Si una distribución de probabilidad P se factoriza de acuerdo a A, entonces $sop(P)$ es nice.

Dem: Sea P una distribución de probabilidad que se factoriza de acuerdo a A. Queremos ver que $F = sop(P)$ es nice. Sean $(t_1, ..., t_d) \in \phi_A^{-1}(P)$. Entonces

$p_j = \prod_{i=1}^m t_i^{A_{ij}} > 0$ si $j \in F$ y
$p_j = \prod_{i=1}^m t_i^{A_{ij}} = 0$ si $j \notin F$.

Supongamos que F no es nice. Entonces $sop(A_k) \subseteq \bigcup_{i \in F} sop(A_i)$, para algún $k \notin F$. Entonces para cada $i \in sop(A_k), \exists l \in F$ tal que $A_{il} > 0$. Entonces si t_i se anulara, la coordenada l-ésima de ϕ_A sería nula, es decir $(\phi_A)_l = \prod_{i=1}^m t_i^{A_{il}} = 0$.
Luego $t_i > 0$ para todo $i \in sop(A_k)$.

Entonces $p_k = \prod_{i \in sop(A_k)} t_i^{A_{ik}} > 0$, luego $k \in F$, lo cual contradice lo supuesto. $\square$

Definiremos a continuación $X_A^{\geq 0}$, la variedad tórica asociada a A. Veamos cómo se relaciona la factorización de P con $X_A^{\geq 0}$.

Definición: Definimos $X_A^{\geq 0} = \{X \in \mathbb{R}^m_{\geq 0} / X^U = X^V\}$ donde $X = x_1, ..., x_m$. Donde $U, V \in \mathbb{Z}^m$ satisfacen que $U - V \in ker_{\mathbb{Z}}(A)$, es decir $A.(U - V) = 0$.

Observar que $X_A^{\geq 0} = V(I)$, $I =< X^U - X^V >_{U-V \in ker_{\mathbb{Z}}(A)} \subseteq \mathbb{R}^m[x_1, ..., x_m]$. Entonces $X_A^{\geq 0}$ es una variedad algebraica afín.

Supongamos ahora que permitimos soluciones $\tilde{U}, \tilde{V} \in \mathbb{R}^m$, es decir $\tilde{U} - \tilde{V} \in ker_{\mathbb{R}}(A)$. Y llamemos $\tilde{X}_A^{\geq 0} = \{X \in \mathbb{R}^m_{\geq 0} / X^{\tilde{U}} - X^{\tilde{V}} = 0\} \subseteq V(I) = X_A^{\geq 0}$.

Sabemos que existen escalares reales $\alpha_1, ..., \alpha_s$ tales que

$\tilde{U} - \tilde{V} = \sum_{i=1}^{s} \alpha_i.(U^{(i)} - V^{(i)})$, donde $(U^{(i)} - V^{(i)}) \in ker_{\mathbb{Z}}(A)$

Luego si $X = 0$, trivialmente se verifica la igualdad. Si $X \neq 0$ satisface la ecuación para U, V enteros (i.e. $X \in X_A^{\geq 0}$), $X^{U-V} = 1$ y $X^{\tilde{U}-\tilde{V}} = X^{\sum_{i=1}^{s} \alpha_i.(U^{(i)} - V^{(i)})} = \prod_{i=1}^{s} X^{\alpha_i.(U^{(i)} - V^{(i)})} = \prod_{i=1}^{s} (X^{(U^{(i)} - V^{(i)})})^{\alpha_i} = 1$ tal como queríamos ver (i.e. $X \in \tilde{X}_A^{\geq 0}$).

Entonces $\tilde{X}_A^{\geq 0} \supseteq X_A^{\geq 0}$.

Finalmente tenemos que $\tilde{X}_A^{\geq 0} = X_A^{\geq 0}$.

Recordemos que nuestro objetivo es caracterizar las distribuciones que se factorizan de acuerdo a A, y posteriormente caracterizar a aquellas que son límites de distribuciones que se factorizan.

Para ello probaremos el siguiente

Teorema 2: Una distribución de probabilidad P se factoriza de acuerdo a A si y solo si $P \in X_A^{\geq 0}$ y $sop(P)$ es nice.

Primero demostraremos antes un lema auxiliar:

Lema 3: Si una distribución de probabilidad P se factoriza de acuerdo a A entonces $P \in X_A^{\geq 0}$.

Dem(Lema 3): Queremos ver que si $P \in Im(\phi_A)$ entonces $P \in X_A^{\geq 0}$, (i.e. $Im(\phi_A) \subseteq X_(A)^{\geq 0}$). Para ello, tomemos $X \in Im(\phi_A)$ y $(t_1, ..., t_d) \in \mathbb{R}^d_{\geq 0}$ tal que $\phi_A(t_1, ..., t_d) = X$, o sea

$\phi_A(t_1, ..., t_d) = (\prod_{i=1}^d t_i^{a_{i1}}, ..., \prod_{i=1}^d t_i^{a_{im}})$.

Tenemos entonces que $x_j = \prod_{i=1}^d t_i^{a_{ij}}$ para $j = 1, ..., m$. Y queremos que $x_1^{U_1}...x_1^{U_m} = x_1^{V_1}...x_1^{V_m}$, donde $U = (U_1...U_m)$ y $V = (V_1...V_m)$.

Sustituyendo obtenemos que como $x_j = \prod_{i=1}^d t_i^{a_{ij}}$ entonces $x_1^{U_1}...x_m^{U_m} = (\prod_{i=1}^d t_i^{a_{i1}})^{U_1}...(\prod_{i=1}^d t_i^{a_{im}})^{U_m}$ y reagrupando obtenemos

$x_1^{\tilde{U_1}}...x_m^{\tilde{U_m}} = t_1^{\sum_{i=1}^d a_{i1}U_1}...t_d^{\sum_{i=1}^d a_{im}U_m} = t_1^{\sum_{i=1}^d a_{i1}V_1}...t_d^{\sum_{i=1}^d a_{im}V_m} = x_1^{\tilde{V_1}}...x_m^{\tilde{V_m}}$. Entonces $(x_1, ..., x_m) \in X_A^{\geq 0}$. $\square$

Demostraremos ahora el teorema.

Teorema 2: Una distribución de probabilidad P se factoriza de acuerdo a A si y solo si $P \in X_A^{\geq 0}$ y $sop(P)$ es nice.

Dem(Teo 2): Los lemas anteriores demuestran la parte "$\Rightarrow$".
Veamos entonces "$\Leftarrow$": Sea $P \in X_A^{\geq 0}$ tal que $sop(P) = F$ es nice.
Queremos ver que $P \in Im(\phi_A)$, es decir que el sistema

$p_j = \prod_{i=1}^m t_i^{A_{ij}} > 0$ si $j \in F$ y
$p_j = \prod_{i=1}^m t_i^{A_{ij}} = 0$ si $j \notin F$.

tiene solución no negativa $(t_1, ..., t_d)$.

Primero veamos que el sistema $p_j > 0$ si $j \in F$ tiene solución con todas sus coordenadas positivas. Entonces tomando logaritmo: $\tau_i = log(t_i)$. Obtenemos así un sistema equivalente:

$\sum_{i=1}^d a_{ij}.\tau_i = log(p_j)$ para $j \in F$.

Supongamos que el sistema anterior no tuviera solución. Sea $s = \#F$. Tenemos entonces un sistema de s ecuaciones con d variables que puede ser escrito como $B.y = c$, donde $B \in \mathbb{Z}^{d \times s}$ y $c_j = (log(p_j))$ para $j \in F$.

Por el supuesto anterior sabemos que c no es combinación lineal de las columnas de B. Entonces existe un vector q que es ortogonal a todas las colunas de B,pero no a c, es decir $qB = 0$ y $< q, c >\neq 0$, además podemos suponer que $< q, c >= 1$.

Tenemos que $\sum_{i\in F} q_i.log(p_i) =< q, c >= 1$ y $\sum_{i\in F} q_i.a_{ij} = q.B = 0$

Llamemos $U_j = max\{q_j, 0\}$ es aquel vector que se obtiene con los coeficiente no negativos de q y colocando ceros en los negativos y $V_j = max\{-q_j, 0\}$ es aquel vector que se obtiene con los coeficiente no positivos de q y colocando ceros en los positivos.

Luego, se obtiene $\sum_{i\in F} q_i.a_{ij} = \sum_{i\in F} (U_i - V_i).a_{ij} = \sum_{i=1}^{m} (U_i - V_i).a_{ij} = 0$ donde $U_i = V_i = 0$ si $i \notin F$.

Y como $\sum_{j\in F} q_j.log(p_j) = \sum_{j\in F} (U_j - V_j).log(p_j) = 1$, entonces:
$\sum_{j\in F} U_j.log(p_j) \neq \sum_{j\in F} V_j.log(p_j)$, entonces:
$\prod_{j\in F} t_j^{U_j} \neq \prod_{j\in F} t_j^{V_j}$.
Entonces $P \notin X_A^{\geq 0}$

Por lo tanto existe solución $(t_1, ..., t_d)$ con los $t_i > 0$, $t_i = e^{\tau_i}$ con $i \in \{1, ..., d\}$.

La solución $(t_1, ..., t_d)$ obtenida es independiente del índice i, para cualquier $i \notin I = \bigcap_{j\in F} sop(A_j)$,porque para cada $i \in I^c$, $a_{ij} = 0$ para todo $j \in F$.

Pongamos entonces $t_i = 0$ para todo $i \in I^c$.

Como F es nice, para cada índice $l \notin F$ existe un $i \in I^c$ tal que $a_{il} > 0$, luego $t_i^{a_{il}} = 0$, entonces $\prod_{i=1}^{d} t_i^{a_l} = 0$ para cada $l \notin F$. Entonces se verifica que este nuevo vector $(t_1, ..., t_d)$ satisface que $p_j = \prod_{i=1}^{m} t_i^{a_{ij}} > 0$ si $j \in F$ y $p_j = \prod_{i=1}^{m} t_i^{a_{ij}} = 0$ si $j \notin F$.

Entonces $P \in Im(\phi_A)$. $\square$

Distribuciones que son límite de distribuciones que se factorizan

Nos concentraremos ahora en el estudio de aquellas distribuciones que son límite de distribuciones que se factorizan de acuerdo al modelo A.

En general, $Im(\phi_A)$ no es un subconjunto cerrado de $\mathbb{R}_{\geq 0}^{m}$ (ver ejemplo * al final), esto es importante ya que hay distribuciones que no se factorizan de acuerdo a A, pero que son límites de distrubuciones que sí se factorizan.

Veremos ahora que aquellas distribuciones que caen en la variedad tórica, son aquellas se factorizan de acuerdo a A, o que son límites de distrubuciones que sí se factorizan, es decir, son aquellas que están en la clausura de $Im(\phi_A)$. En resumen, queremos ver que $X_A^{\geq 0} = cl(Im(\phi_A))$.

Una de las inclusiones es fácil ya que $X_A^{\geq 0}$ es cerrado (con la topología usual) por ser cerrado Zariski (por ser una variedad algebraica). Entonces como por el lema 3 sabemos que $Im(\phi_A) \subseteq X_A^{\geq 0}$ tenemos que $cl(Im(\phi_A)) \subseteq X_A^{\geq 0}$. Luego sólo resta ver que $X_A^{\geq 0} \subseteq cl(Im(\phi_A))$. Antes de ver esta inclusión, veremos un lema y algunas definiciones.

Definición: Sea A una matriz como hasta ahora, $A = (A_1|...|A_m)$ en columnas. Un subconjunto F se dice facial si existe un vector $c \in \mathbb{R}^d$ tal que: $< c, A_i >= 0$ para cada $i \in F$ y $< c, A_i \geq 1$ para cada $i \notin F$

Definición: Dado un subconjunto F definimos el vector característico de F como un vector $(v_1, ..., v_m)$ donde $v_i = 1$ si $i \in F$ y $v_i = 0$ si $i \notin F$.

Lema 4: Dado un subconjunto F y una matriz A, las siguientes afirmaciones son equivalentes:

(a) F es facial.

(b) El vector característico de F pertenece a la variedad tórica $X_A^{\geq 0}$.

(c) Existe un vector con soporte F en la variedad tórica $X_A^{\geq 0}$.

Dem: ver la referencia [2].

Ahora estamos en condiciones de enunciar y demostrar el último teorema, aquel que carecteriza a las distribuciones que caen en la variedad tórica $X_A^{\geq 0}$.

Teorema 5: Una distribución de probabilidad P se factoriza de acuerdo a A o es límite de distribuciones que se factorizan si y solo si $P \in X_A^{\geq 0}$.

Dem: Como decíamos al comienzo de esta sección, queremos ver que $X_A^{\geq 0} = cl(Im(\phi_A))$. La inclusión $cl(Im(\phi_A)) \subseteq X_A^{\geq 0}$ es clara. Tenemos que ver que $X_A^{\geq 0} \subseteq cl(Im(\phi_A))$.

Para ello tomemos un punto $P \in X_A^{\geq 0} - Im(\phi_A)$ y veamos que $P \in cl(Im(\phi_A))$.

Definiremos primero una sucesión de puntos $P(\epsilon)$, luego veremos que $P(\epsilon) \to P$ cuando $\epsilon \to 0$, y finalmente veremos que $P(\epsilon) \in Im(\phi_A)$ para todo $\epsilon > 0$

Sea $P \in X_A^{\geq 0} - Im(\phi_A)$ y sea $F = sop(P)$ (P es un vector con soporte F en la variedad tórica $X_A^{\geq 0}$, luego por el Lema 4 "(c)$\Rightarrow$ (a)" el conjunto $F = sop(P)$ es facial).

Consideremos el siguiente sistema de ecuaciones en las variables $t_1, ..., t_d$.

$\prod_{i=1}^d t_i^{a_{ij}} = p_j > 0$ si $j \in F$.

Vimos que este sistema tiene una solución $(t_1, ..., t_d)$ con todas las coordenadas reales positivas.

Como $F = sop(P)$ es facial. Entonces por definición podemos tomar un vector $c \in \mathbb{R}^d$ tal que

$< c, A_i >= 0$ para cada $i \in F$ y $< c, A_i \geq 0$ para cada $i \notin F$.

Sea $\epsilon > 0$ y sea $P(\epsilon) = (\epsilon^{<c,A_1>}. \prod_{i=1}^d t_i^{a_{i1}}, ..., \epsilon^{<c,A_m>}. \prod_{i=1}^d t_i^{a_{im}})$

Como $< c, A_i >= 0$ para cada $i \in F$ y $< c, A_i \geq 0$ para cada $i \notin F$, es claro que $\epsilon^{<c,A_i>} = 1$ para cada $i \in F$ y $\epsilon^{<c,A_i>} = \epsilon$ para cada $i \notin F$.

Se tiene:

$\epsilon^{<c,A_i>}. \prod_{j=1}^d t_j^{a_{ji}} = \prod_{j=1}^d t_j^{a_{ji}} \to \prod_{j=1}^d t_j^{a_{ji}}$ cuando $\epsilon \to 0$ para cada $i \in F$

y

$\epsilon^{<c,A_i>} \prod_{j=1}^d t_j^{a_{ji}} = \epsilon. \prod_{j=1}^d t_j^{a_{ji}} \to 0$ cuando $\epsilon \to 0$ para cada $i \notin F$.

Entonces hemos visto que $\lim_{\epsilon \to 0} P(\epsilon) = P$ tal como queríamos.

Sólo nos resta ver que $P(\epsilon) \in Im(\phi_A)$ para todo $\epsilon > 0$. Para ello observemos que:

$\epsilon^{<c,A_j>} = \epsilon^{c_1 a_{1j} + ... + c_d a_{dj}} = \epsilon^{c_1 a_{1j}} ... \epsilon^{c_d a_{dj}} = (\epsilon^{c_1})^{a_{1j}} ... (\epsilon^{c_d})^{a_{dj}} = \prod_{i=1}^d (\epsilon^{c_i})^{a_{ij}}$.

Esto muestra que podemos reescribir:

$P(\epsilon) = (\epsilon^{<c,A_1>}. \prod_{i=1}^d t_i^{a_{i1}}, ..., \epsilon^{<c,A_m>}. \prod_{i=1}^d t_i^{a_{im}}) =$

$\prod_{i=1}^d (t_i.\epsilon^{c_i})^{a_{i1}}, ..., \epsilon^{<c,A_m>}. \prod_{i=1}^d t_i^{a_{im}})$.

Luego $P(\epsilon) \in Im(\phi_A)$ para todo $\epsilon > 0$, ya que $P(\epsilon) = \phi_A(t_1.\epsilon^{c_1}, ..., t_d.\epsilon^{c_d})$. Lo cual concluye la demostración del teorema. $\square$

*Ejemplo **: Veamos ahora que en general, $Im(\phi_A)$ no es un subconjunto cerrado de $\mathbb{R}^m_{\geq 0}$. Sea

$A=\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$ Luego $\phi_A(x, y, z) = (z, xz, yx, y)$.

Notar que el elemento $(0, 1, 0, 0) \notin Im(\phi_A)$ ya que si $z = 0 \Rightarrow xz = 0$.

Consideremos el elemento $(t^{-1}, t, t^2) \in \mathbb{R}^m_{\geq 0}$ tomando límite con $t \to 0^+$obtenemos que $\phi_A(t^{-1}, t, t^2) = (t, 1, t, t^2) \to (0, 1, 0, 0)$.

Luego $(0, 1, 0, 0) \in cl(Im(\phi_A)) = X^{\geq 0}_A$ pero $(0, 1, 0, 0) \notin Im(\phi_A)$.

(Gracias a Martín Mereb por el ejemplo)

Referencias

[1] Las clases teóricas de la materia Variedades Tóricas - Alicia Dikenstein.

[2] El artículo: On The Toric Algebra of Grafical Models - Greiger, Meek, Sturmfels.

[3] Algebraic Geometry. A First Course. - Harris.

[4] Amigos con conocimiento de estadística - Andrés y Lucas.

[5] Martín M. (el ejemplo) y Alicia.