week 5 day 1

“Algebraic & Geometric Methods in Statistics”

Sonja Petrović
Created for Math/Stat 561

Feb 6, 2023.

Recap: Exponential families

- An {exponential family} is a *parametric statistical model* with probability distributions of a *certain form*.
- {General} enough to include many of the most common families of probability distributions:
  - multivariate normal
  - exponential
  - Poisson
  - binomial (with fixed number of trials)
- {Specific} enough to have nice properties:
  - likelihood function is strictly concave [next lecture]
  - have conjugate priors.

Objectives

- {What is} an exponential family?
- **How to find the polynomial ideal of an exponential family?**
  - Discrete exponential models: Hypothesis testing [future lecture]
    - Gaussian exponential submodels: Conditional independence implications [past lecture]

Recap: Discrete exponential families

Notation

- X a discrete random variable $X \in [r]$.
- $T(x) = a_x$, writing as a vector: $a_x = (a_{1x}, \dots, a_{kx})^t$. Assume $a_{jx} \in \mathbb{Z}$.
- $h(x) = h_x$, so $h = (h_1, \dots, h_r)$ is also a vector (of positive real numbers)
- $\eta = (\eta_1, \dots, \eta_k)^t$ and $\theta_i = \exp \eta_i$.

$$p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_i \theta_i^{a_{ix}}.$$

- The design matrix: $\mathcal{A} = (a_{jx})_{j \in [k], x \in [r]} \in \mathbb{Z}^{k \times r}$.
- For each value x of X, the monomial $\prod_j \theta_j^{a_{jx}} \leftrightarrow$ a column of $\mathcal{A}$.

Design matrix recipe

Columns of $\mathcal{A}$ are exponents of the parametrization of each given state.

Question from the previous lecture

- Consider the model $p_{ij} = \alpha_i \beta_j$ for $i \in [2]$ and $j \in [2]$.

Question from the previous lecture

- Consider the model $p_{ij} = \alpha_i \beta_j$ for $i \in [2]$ and $j \in [2]$. Binary independent random variables.
- The design matrix is

$$\mathcal{A} = \begin{array}{c} \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{array} \begin{array}{c} \begin{array}{cccc} p_{11} & p_{12} & p_{21} & p_{22} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

Finally, the vector $h = [1, 1, 1, 1]^t$.

$\rightarrow$ I *assume* that **each of you** has completed this example for not 2×2 but $r_1 \times r_2$ by hand, by now.

In this lecture

- log-affine models
- what to do with the h function in the parametrization of an exponential family model (nothing!)
- is there an "easy" way to compute the implicitization of all discrete exponential families?

Log-affine, log-linear discrete exponential families

- Let $\mathcal{A} = [a_{jx}]_{j \in [k], x \in [r]} \in \mathbb{Z}^{k \times r}$ be a design matrix.

$$p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_i \theta_i^{a_{ix}}.$$

- The **logarithm** of the exponential family model $p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_i \theta_i^{a_{ix}}$ is

Log-affine, log-linear discrete exponential families

- Let $\mathcal{A} = [a_{jx}]_{j \in [k], x \in [r]} \in \mathbb{Z}^{k \times r}$ be a design matrix.

$$p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_i \theta_i^{a_{ix}}.$$

- The **logarithm** of the exponential family model $p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_i \theta_i^{a_{ix}}$ is

$$\log p_\theta(x) = \log h_x + \sum_j a_{jx} \log \theta_j - \log Z(\theta).$$

- Assume $\mathcal{A}$ contains the vector $1 = (1, \ldots, 1)$ in the rowspan, then this is equivalent to requiring that $\log p$ belongs to the affine space $\log(h) + rowspan(A)$.

... "equivalent to requiring that log p belongs to the affine space $\log(h) + \mathit{rowspan}(A)$."

Definition

Definition 6.2.1. Let $A \in \mathbb{Z}^{k\times r}$ be a matrix of integers such that $\mathbf{1} \in \operatorname{rowspan}(A)$ and let $h \in \mathbb{R}^r_{>0}$. The *log-affine model* associated to these data is the set of probability distributions

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

If $h = \mathbf{1}$, then $\mathcal{M}_A = \mathcal{M}_{A,\mathbf{1}}$ is called a *log-linear model.*

Figure 1: Source: textbook

Ideal of a log-linear model

The *log-affine* model $\mathcal{M}_{\mathcal{A},h}$ given by design matrix $\mathcal{A}$ and vector h:

$$p_\theta(x) \propto h_x \prod_i \theta_i^{a_{ix}}.$$

- This is a model for the joint distribution for discrete random variables, whose states we may denote by $\{1, \ldots, r\}$. So the model is a parametric form of the joint probabilities $p_1, \ldots, p_r$.
- $\mathcal{M}_{\mathcal{A},h}$ is the set of all joint probability vectors $(p_1, \ldots, p_r)$ of the above form.
- The indeterminates p_i index the columns of the matrix $\mathcal{A}$.

Definition [Cf. 6.2.2. & 6.2.3. in the book]

The toric ideal of the model $\mathcal{M}_{\mathcal{A},h}$ is the ideal $I_{\mathcal{A},h}$ of the variety parametrized by $(p_1, \ldots, p_r)$. If $h = [1, \ldots, 1]$, we denote this as $I_{\mathcal{A}}$.

Proposition 6.2.4. *Let $A \in \mathbb{Z}^{k \times r}$ be a $k \times r$ matrix of integers. Then the toric ideal I_A is a binomial ideal and*

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

If $\mathbf{1} \in \operatorname{rowspan}(A)$, then I_A is homogeneous.

Figure 2: Proposition 6.2.4. from textbook

Class work: Before going into the proof, decipher:

- What is the definition of I_A, and what is really the claim in this proposition that needs to be proved?
- What is u? ($u \in \mathbb{N}^r \dots$)
- What is A? ($A \in \mathbb{Z}^{k \times r} \dots$)
- What is Au? example, meaning?
- What does p^u mean?

Proof.

On the board, draw out steps of "peeling terms" of any $f \in I_A$ one binomial at a time. Exactly as page 123 of the book

Remark on generality

We defined $I_{\mathcal{A},h}$ and I_A. The proposition only defines the binomial ideal I_A. * Why?? What happens to general h? * *Good news:* Generators for the toric ideal $I_{\mathcal{A},h}$ are easily obtained from generators of the toric ideal $I_{\mathcal{A}}$, by globally making the substitution $p_j \mapsto p_j/h_j$. Hence, it is sufficient to focus on the case of the toric ideal $I_{\mathcal{A}}$.

- All of these ideals I_A turn out to be *binomial* ideals; the proposition tells us which particular binomials to look for.
  - . . . and what is a "binomial ideal"? [Board as needed.]

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ?$

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2.$ There are 4 joint outcomes, and 2 parameters.
  - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.
  - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?
  $p_1 = \theta_2^3, p_2 = \theta_1\theta_2^2, p_3 = \theta_1^2\theta_2, p_4 = \theta_1^3$.
- What is an example of p^u?

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.
  - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?
  $p_1 = \theta_2^3, p_2 = \theta_1\theta_2^2, p_3 = \theta_1^2\theta_2, p_4 = \theta_1^3$.
- What is an example of p^u? $\rightarrow$ Say, $u = (0, 2, 0, 0)^t$. Then $p^u = p_2^2$.

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.
    - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?
  $p_1 = \theta_2^3, p_2 = \theta_1\theta_2^2, p_3 = \theta_1^2\theta_2, p_4 = \theta_1^3$.
- What is an example of p^u? $\rightarrow$ Say, $u = (0, 2, 0, 0)^t$. Then $p^u = p_2^2$.
    - What is $\mathcal{A}u$ in this case?

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.
    - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?
  $p_1 = \theta_2^3, p_2 = \theta_1\theta_2^2, p_3 = \theta_1^2\theta_2, p_4 = \theta_1^3$.
- What is an example of p^u? $\rightarrow$ Say, $u = (0, 2, 0, 0)^t$. Then $p^u = p_2^2$.
    - What is $\mathcal{A}u$ in this case? $\rightarrow$ $\mathcal{A}u$ is the value of the sufficient statistic in this exponential family. HW: verify that this is sufficient for the binomial model.
    - Can you come up with v such that $Au = Av$?

Example: Binomial with 3 trials [Ex. 6.2.5.]

Let $\mathcal{A} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{bmatrix}$.

- $k = ?, r = ? \rightarrow k = 4,\ r = 2$. There are 4 joint outcomes, and 2 parameters.
  - So the vector of joint probabilities is $p = [p_1, p_2, p_3, p_4]$.
- What do columns of $\mathcal{A}$ represent?
  $p_1 = \theta_2^3, p_2 = \theta_1\theta_2^2, p_3 = \theta_1^2\theta_2, p_4 = \theta_1^3$.
- What is an example of p^u? $\rightarrow$ Say, $u = (0, 2, 0, 0)^t$. Then $p^u = p_2^2$.
  - What is $\mathcal{A}u$ in this case? $\rightarrow$ $\mathcal{A}u$ is the value of the sufficient statistic in this exponential family. HW: verify that this is sufficient for the binomial model.
  - Can you come up with v such that $Au = Av$? \attn{$\rightarrow v = (1, 0, 1, 0)^t$.
- The corresponding binomial is $p_1 p_3 - p_2^2$.
  - VERIFY that this binomial evaluates to 0 *at all points in the model.*

Example 6.2.6. - class & board work

The model of independence of two discrete random variables. Say that $r_1 = 4$ and $r_2 = 3$.

- What is the parametrization of the model?
- What is the design matrix $\mathcal{A}$?
- From an observed table of counts u (which format is the table in, by the way??), what does Au compute?
- Find some generators of the toric ideal I_A by hand. Interpet them.
- How do you know when you have *all* binomials that suffice to capture (generate) the entire ideal of the model? (... That's the million dollar question.)

to do

I think I will leave 6.2.7 for self-study/reading. (Good for hw2.)

CODE for generating ideals. :)

Next up:

How to use implicit models for likelihood inference.

- next topic: likelihood inference from ch7,

course timeline update

Other resources

- TBD.

License

This document is created for Math/Stat 561, Spring 2023.