Exponential families

- Let X be a random variable taking values in a set $\mathscr{X}$.

- An exponential family is the set of probability distributions whose probability mass function or density function can be expressed as

$$f_\theta(x) = h(x)e^{\eta(\theta)^t T(x) - A(\theta)}$$

for a given statistic $T : \mathscr{X} \to \mathbb{R}^k$, natural parameter $\eta : \Theta \to \mathbb{R}^k$, and functions $h : \mathscr{X} \to \mathbb{R}_{>0}$ and $A : \Theta \to \mathbb{R}$.

Exponential families

- Three equivalent forms:
  - $f_\theta(x) = h(x)e^{\eta(\theta)^t T(x) - A(\theta)}$
  - $f_\theta(x) = h(x)g(\theta)e^{\eta(\theta)^t T(x)}$
  - $f_\theta(x) = e^{\eta(\theta)^t T(x) - A(\theta) + B(x)}$

Binomial distribution

$X \sim \text{Bin}(m, \theta)$, $\mathscr{X} = \{0,1,\ldots,m\}$

$$p(x) = \binom{m}{x} \theta^x (1-\theta)^{m-x} =$$

Binomial distribution

- $f_\theta(x) = h(x)e^{\eta(\theta)^t T(x) - A(\theta)}$
- Statistic $T : \mathscr{X} \to \mathbb{R}^k$, natural parameter $\eta : \Theta \to \mathbb{R}^k$, functions $h : \mathscr{X} \to \mathbb{R}_{>0}$ and $A : \Theta \to \mathbb{R}$
- Binomial distribution: $p(x) = \binom{m}{x} \theta^x (1-\theta)^{m-x} = \binom{m}{x} \exp\left[\left(\log \frac{\theta}{1-\theta}\right) x + m \log(1-\theta)\right]$
- Poll: What are k, T, η, h, A in this example?

1. $k = 1, T(x) = \log \frac{\theta}{1-\theta}, \eta = x, h = \binom{m}{x}, A = -m\log(1-\theta)$
2. $k = 1, T(x) = x, \eta = \log \frac{\theta}{1-\theta}, h = \binom{m}{x}, A = -m\log(1-\theta)$
3. $k = 2, T(x) = (x, m-x), \eta = (\theta, 1-\theta), h = \binom{m}{x}, A = 0$

Canonical form

- $f_\theta(x) = h(x)e^{\eta(\theta)^t T(x) - A(\theta)}$
- If $\eta(\theta) = \theta$, then the exponential family is said to be in canonical form.
- By defining a transformed parameter $\eta = \eta(\theta)$, it is always possible to convert an exponential family to canonical form.
- The function A is determined by the other functions: It makes the pdf (pmf) to integrate (sum) to one. Thus it can be written as a function of η.
- The canonical form is $f_\eta(x) = h(x)e^{\eta^t T(x) - A(\eta)}$.

Discrete exponential families

- Let X be a discrete random variable taking values in $\mathscr{X} = [r]$.
- Denote
  - $T(x) = a_x$ where $a_x = (a_{1x}, \ldots, a_{kx})^t$
  - $h(x) = h_x$, so $h = (h_1, \ldots, h_r) \in \mathbb{R}^r_{>0}$
  - $\eta = (\eta_1, \ldots, \eta_k)^t$ and $\theta_i = \exp(\eta_i)$
- Then $p_\eta(x) = h(x)e^{\eta^t T(x) - A(\eta)} =$

where $Z(\theta) = \sum_{x \in \mathscr{X}} h_x \prod_j \theta_j^{a_{jx}}$.

Discrete exponential families

$$p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_j \theta_j^{a_{jx}} \text{ where } Z(\theta) = \sum_{x \in \mathcal{X}} h_x \prod_j \theta_j^{a_{jx}}$$

- If a_{jx} are integers for all j and x, then the parametrizing functions are rational functions.
- The entries a_{jx} can be recorded in the matrix $A = (a_{jx})_{j \in [k], x \in [r]} \in \mathbb{Z}^{k \times r}$.
- For $x \in \mathcal{X} = [r]$, the monomials $\prod_j \theta_j^{a_{jx}}$ correspond to a column of the matrix A.

Example: Let $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}$ and $h = \mathbf{1}$. Then

$$p_\theta = \frac{1}{Z(\theta)} \left(\theta_2^3, \theta_1\theta_2^2, \theta_1^2\theta_2, \theta_1^3\right) \text{ where } Z(\theta) = \theta_2^3 + \theta_1\theta_2^2 + \theta_1^2\theta_2 + \theta_1^3.$$

Discrete exponential families

- Let $A = (a_{jx})_{j\in[k],x\in[r]} \in \mathbb{Z}^{k\times r}$.
- The logarithm of the exponential family representation $p_\theta(x) = \frac{1}{Z(\theta)} h_x \prod_j \theta_j^{a_{jx}}$ gives

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

- If we assume that the matrix A contains the vector $\mathbf{1} = (1,1,\ldots,1)$ in its row span, then this is equivalent to requiring that $\log p$ belongs belongs to the affine space $\log(h) + \text{rowspan}(A)$.

Log-affine model

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

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

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

Log-affine model

Def: Let $A \in \mathbb{Z}^{k\times r}$ be a matrix of integers such that $\mathbf{1} \in \mathrm{rowspan}(A)$ and let $h \in \mathbb{R}^r_{>0}$.
The monomial map associated to this data is the rational map

$$\phi^{A,h} : \mathbb{R}^k \to \mathbb{R}^r, \text{ where } \phi_j^{A,h} = h_j \prod_{i=1}^{k} \theta_i^{a_{ij}}.$$

NB! The normalizing constant $Z(\theta)$ is removed.

Example: Let $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}$. The monomial map is $\phi^A : \mathbb{R}^2 \to \mathbb{R}^4$ is given by

$$(\theta_1, \theta_2) \mapsto (\theta_2^3, \theta_1\theta_2^2, \theta_1^2\theta_2, \theta_1^3).$$

Discrete independent random variables

- Consider the parametrization

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

where $i \in [2], j \in [2]$ and α_i, β_j are independent parameters.

- This is the parametrization of two discrete independent random variables.

- Poll 1: What are the matrix A and vector h representing the above parametrization?

- Poll 2: What is the size of the matrix of A if $i \in [r_1]$ and $j \in [r_2]$?

Log-affine model

Def: Let $A \in \mathbb{Z}^{k\times r}$ and $h \in \mathbb{R}^r_{>0}$. The ideal

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

is called the toric ideal associated to the pair A and h.

- If $h = \mathbf{1}$, then we denote $I_A := I_{A,1}$.
- Generators for the ideal $I_{A,h}$ are obtained from generators of the ideal I_A.

Log-affine model

Prop: Let $A \in \mathbb{Z}^{k \times r}$ and $h \in \mathbb{R}^r_{>0}$. Then

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

Example: Let $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 \end{pmatrix}$. The monomial map is $\phi^A : \mathbb{R}^2 \to \mathbb{R}^4$ is given by

$$(\theta_1, \theta_2) \mapsto (\theta_2^3, \theta_1\theta_2^2, \theta_1^2, \theta_2, \theta_1^3).$$

The toric ideal is

$$I_A = \langle p_1p_3 - p_2^2, p_1p_4 - p_2p_3, p_2p_4 - p_3^2 \rangle. \text{ [Poll]}$$