20.19.2019.

Lecture 1:

Statistical Models
Exponential families
Toric varieties

We describe discrete random variables in terms of its probability distributions and continuos random variables in terms of its density functions.

X a random variale. discrete → probability distribution
continuous → density function

Example 1:

(1) Suppose $\theta \in (0,1)$. X is a binomial random variable

$$\underbrace{P(X=0)}_{\text{"success"}} = \theta \quad , \quad \underbrace{P(X=1)}_{\text{"failure"}} = (1-\theta)$$

(2) Suppose X is a univariate random variable with mean $= \mu$ and variance $= \sigma^2$
X has density function

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Def:

- A statistical model $\mathcal{M}$ is a collection of probability distributions or density functions over a given space.

- A parametric statistical model $\mathcal{M}_\Theta$ is the image of a map from a finite dimensional parameter space $\Theta \subset \mathbb{R}^d$ to a space of probability distributions or density functions. i.e. $p_\bullet : \Theta \longrightarrow \mathcal{M}_\Theta$, $\theta \longmapsto p_\theta$,

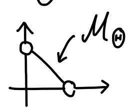
$\mathcal{M}_\Theta := \{ p_\theta : \theta \in \Theta \}$

- For each parameter value in the model, we want each p_θ to be uniquely determined by θ.
  $\rightarrow$ If this is the case, we say the model is identifiable.

Example 2:

(1) The binomial random variable model.
Take $\Theta = (0,1)$, $p_\bullet : \Theta \rightarrow \mathcal{M}_\Theta \subseteq \mathbb{R}^2$

$$\theta \longmapsto (\theta, 1-\theta)$$



(2) The normal random variable model
$$\Theta = \mathbb{R} \times \mathbb{R}_{>0}$$

$$(\mu, \sigma) \longmapsto f(x \mid \mu, \sigma)$$

(3) The probability simplex.
Consider a discrete random variable X with outcome space $\mathcal{X} = \{1, \ldots, r\}$, fix $p_i = P(X=i)$.
The set of all possible probability distributions for X is the probability simplex

$$\Delta_{r-1} = \{ (p_1, \ldots, p_r) \in \mathbb{R}^r : p_i \geq 0 , \ p_1 + p_2 + \cdots + p_r = 1 \}$$

$\rightarrow$ Discrete statistical models are subsets of the probability simplex

(4) The independence model. $X =$ discrete r.v. with outcome space $\{1, \ldots, r\}$
$Y =$ " ———————— " $\{1, \ldots, c\}$
Denote the joint distribution of X and Y by $p_{ij} = P(X = i, Y = j)$

$P_\bullet : \Theta = \Delta_{r-1} \times \Delta_{c-1} \to \Delta_{rc-1}$

$$(a_1, \dots, a_r) \times (b_1, \dots, b_r) \mapsto (a_1 b_1, \dots\dots a_r b_r)$$

$$\mathcal{M}_{X \perp\!\!\!\perp Y} = \{ P_\theta : \theta \in \Delta_{r-1} \times \Delta_{c-1} \}.$$

$r = 2$ $\quad (a_1 b_1,\ a_1 b_2,\ a_2 b_1,\ a_2 b_2)$

$c = 2$ $\quad p_{00} \quad p_{01} \quad p_{10} \quad p_{11}$

$$\mathcal{M}_{X \perp\!\!\!\perp Y} = V(p_{11} p_{00} - p_{01} p_{10}) \cap \Delta_3$$

→ The independence model is the Segre-embedding $\mathbb{P}^{c-1} \times \mathbb{P}^{r-1} \longrightarrow \mathbb{P}^{cr-1}$.

6.1. Regular exponential families

This definition is for discrete and continuous random variables.

$\mathcal{X}$ = sample space with σ-algebra $\mathcal{A}$.

ν = σ-finite measure on $\mathcal{A}$, $\quad \mathcal{X}$ is countable union of measurable sets with finite measure.

$T = \mathcal{X} \to \mathbb{R}^k$ is a statistic.

i.e a measurable map.

T^{-1}(meas. set) is meas. in $\mathcal{A}$.

We define the natural parameter space

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

For $\eta \in N$ we define a prob. density P_η on $\mathcal{X}$ as

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

where $\phi(\eta) = \log\left(\int_{\mathcal{X}} h(x)\, e^{\eta^t T(x)}\, d\nu(x) \right)$ → This is the normalizing constant

Let P_η be the probability measure on $(\mathcal{X}, \mathcal{A})$ that has ν-density p_η.

<u>Def:</u> Let K be a positive integer. The probability distributions $(P_\eta : \eta \in N)$ form a regular exponential family.

- To show a family is a regular exponential family, find $h(x)$, $T(x)$ and show the family has the desired form.

<u>Example:</u>

(1) Binomial random variable model

(2) Univariate Normal random variable.

More examples in Ch. 6.

Discrete Regular exponential families

Regular exponential families for discrete random variables.

- $\mathcal{X} = [r] = \{1, \dots, r\}$ → discrete outcome space.
- $T : \mathcal{X} \longrightarrow \mathbb{R}^K \Rightarrow T(x)$ is a vector for each $x \in [r]$.
- $h : \mathcal{X} \to \mathbb{R} \Rightarrow h$ is a vector $(h(1), \dots, h(r))$
- For $\eta \in \mathbb{R}^K$, the normalizing constant $\phi(\eta)$ is a sum

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

- If $x \in [r] \Rightarrow$ the exponential family is given by

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

Look at $\eta^t \cdot T(x)$. Write $T(x) = a_x = \begin{pmatrix} a_{1x} \\ \vdots \\ a_{Kx} \end{pmatrix}$, $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_K \end{pmatrix}$

$\theta_i = \exp(\eta_i)$

Set $h = (h_1, \dots, h_r) \in \mathbb{R}^r_{>0}$

$$\begin{aligned} P_\eta(x) &= h(x)\, e^{\eta^t T(x) - \phi(\eta)} \\ &= \frac{h_x\, e^{(\eta_1, \dots, \eta_K) \cdot (a_{1K}, \dots, a_{Kx})^t}}{\phi(\eta)} \\ &= \frac{h_x\, e^{\eta_1 a_{1K}} \cdots e^{\eta_K a_{Kx}}}{\phi(\eta)} \end{aligned}$$

$$= \frac{h_x\, \theta_1^{a_{1x}} \cdots \theta_K^{a_{Kx}}}{Z(\theta)}.$$

$$Z(\theta) = \sum_{x \in [r]} h_x\, \theta_1^{a_{1x}} \cdots \theta_K^{a_{Kx}}$$

If a_{jx} are integers $\Rightarrow$ $p_\theta(x)$ are rational functions on θ.

Another way to describe distributions in an exponential family.

$$\log(p_\theta(x)) = \log(h_x) + (\log(\theta_1), \ldots, \log(\theta_k))\begin{pmatrix} a_{1x} \\ \vdots \\ a_{kx} \end{pmatrix} - \log(Z(\theta))$$

$$\log(p_\theta) = \log(h) + (\log(\theta_1), \ldots, \log(\theta_k))A - \mathbf{1}\log(Z(\theta))$$

If $\mathbf{1} \in \text{rowspan}(A)$ $\Rightarrow$ $vA = \mathbf{1}$ for some $v \in \mathbb{R}^k$

$$\Rightarrow \log(p_\theta) = \log(h) + \left(\log(\theta)^t - \log(Z(\theta))v^t\right)A$$

$p_\theta \in \Delta_{r-1}$ is a member of the family if

$$\log(p_\theta) \in \text{Span}\{\log(h) + \text{rowspan}(A)\}.$$

<u>Def:</u> 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 these data is the set of probability distributions

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

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

These are toric varieties.

<u>Def:</u> h, A as before. The monomial map associated to this data is the rational map

$$\phi^{A,h} : \mathbb{R}^k \longrightarrow \mathbb{R}^r, \qquad \phi_j^{A,h}(\theta) = h_j \prod_{i=1}^{k} \theta_i^{a_{ij}}$$

we removed $Z(\theta)$ to homogenize the map.

Def: h, A as before. The ideal $I_{A,h} := I(\phi^{A,K}(\mathbb{R}^k)) \subseteq \mathbb{R}[p]$ is the toric ideal associated to the pair A, h.
If $h = 1$, we write $I_A = I_{A,1}$.

$$p_j \mapsto p_j / h_j$$

Prop: 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

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

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

Example: The independence model. Ex. 6.2.6.