Algebraic Statistics

Benjamin C. W. Brown
B.Brown@ed.ac.uk

28th October 2020

Contents I

1. Independence Models

2. Classical Algebraic Geometry

3. Mixture Models & Secant Varieties

planetmath.org

A *statistical model* is usually parameterised by a function, called a *parametrisation*

$$\Theta \to \mathcal{P}, \quad \text{given by} \quad \theta \mapsto P_\theta, \quad \text{so that} \quad \mathcal{P} = \{P_\theta : \theta \in \Theta\},$$

where Θ is the *parameter space.* Θ is usually a subset of $\mathbb{R}^n$.

McCullagh, 2002

This should be defined using category theory.

A contingency table contains counts obtained by cross-classifying observed cases according to two or more discrete criteria.

Example

TODO: Figure (Florida death sentences)

We ask whether the sentences were made independently of the defendant's race.

Two-by-Two Contingency Tables

- Classify using two criteria with r and c levels, yields two random variables X and Y.
- Code outcomes as $[r] := \{1, \ldots, r\}$, and $[c] := \{1, \ldots, c\}$.

All information about X and Y is contained in the *joint probabilities*

$$p_{ij} = P(X = i; Y = j), \quad i \in [r], \ j \in [c].$$

- These in turn determine the *marginal probabilities*:

$$p_{i+} := \sum_{j=1}^{c} p_{ij} = P(X = i), \quad i \in [r],$$

$$p_{+j} := \sum_{i=1}^{r} p_{ij} = P(Y = j), \quad j \in [c].$$

Definition

Two random variables X and Y are *independent* if the joint probabilities factor as $p_{ij} = p_{i+} \cdot p_{+j}$, for all $i \in [r]$ and $j \in [c]$. Denote independence of X and Y by $X \perp\!\!\!\perp Y$.

Proposition

Two random variables X and Y are independent if and only if the $(r \times c)$-matrix, $p = (p_{ij})$, has rank one.

For a (2×2)-table, we thus have:

	$P(Y=1)$	$P(Y=2)$
$P(X=1)$	p_{11}	p_{12}
$P(X=2)$	p_{21}	p_{22}

$$\xrightarrow{X \perp\!\!\!\perp Y} p_{11}p_{22} = p_{12}p_{21}.$$

Suppose now we select n cases, giving rise to n independent pairs of discrete random variables:

$$\begin{pmatrix} X^{(1)} \\ Y^{(1)} \end{pmatrix}, \begin{pmatrix} X^{(2)} \\ Y^{(2)} \end{pmatrix}, \ldots, \begin{pmatrix} X^{(n)} \\ Y^{(n)} \end{pmatrix},$$

all drawn from the same distribution, i.e.:

$$P(X^{(k)} = i; Y^{(k)} = j) = p_{ij}, \quad \text{for all } i \in [r],\ j \in [c],\ k \in [n].$$

Joint probability matrix $p = (p_{ij})$ is an *unknown* element of the $(rc-1)$-dimensional *probability simplex*,

$$\Delta_{rc-1} = \left\{ q \in \mathbb{R}^{r\times c} :\ q_{ij} \geq 0,\ \text{for all } i, j,\ \text{and } \sum_{i=1}^{r} \sum_{j=1}^{c} q_{ij} = 1 \right\}.$$

Definitions

A *statistical model* $\mathcal{M}$ is a subset of Δ_{rc-1}. It represents the set of all candidates for the unknown distribution p.
The *independence model* for X and Y is the set

$$\mathcal{M}_{X \perp\!\!\!\perp Y} := \{p \in \Delta_{rc-1} : \mathrm{rank}(p) = 1\}.$$

$\mathcal{M}_{X \perp\!\!\!\perp Y}$ is the intersection of Δ_{rc-1} and the set of all matrices $p = (p_{ij})$ such that

$$p_{ij}p_{kl} - p_{il}p_{jk} = 0, \text{ for all } 1 \leq i < k \leq r, \text{ and } 1 \leq j < l \leq c.$$

These are examples of *Segre varieties* in algebraic geometry.

Projective Space

Playing field is *n-dimensional projective space*, $\mathbb{P}^n$:

$$\mathbb{P}^n := \{(z_0, \ldots, z_n) \in \mathbb{C}^n\}/(\mathbf{x} \sim \lambda \cdot \mathbf{y}), \quad \lambda \neq 0,$$

that is, its elements consists of *lines through the origin* in $\mathbb{C}^n$.

TODO: FIGURE

Varieties

Varieties are the geometric studied in algebraic geometry, and are the *vanishing sets*[1] for a system of polynomials.
TODO: FIGURES

[1] from '*Verschwindungsmenge*'

Segre varieties come from $\sigma : \mathbb{P}^n \times \mathbb{P}^m \to \mathbb{P}^{(n+1)(m+1)-1}$, that sends $([X],[Y])$ to the pairwise products of their components:

$$\sigma : ([X_1, \ldots, X_{n+1}], [Y_1, \ldots, Y_{m+1}]) \mapsto [\ldots, X_i Y_j, \ldots],$$

Example

$$\sigma : \mathbb{P}^1 \times \mathbb{P}^1 \to \mathbb{P}^3,\ ([X_1, X_2], [Y_1, Y_2]) \mapsto [X_1 Y_1, X_1 Y_2, X_2 Y_1, X_2 Y_2].$$

Set $[X_1 Y_1, X_1 Y_2, X_2 Y_1, X_2 Y_2] = [p_{11}, p_{12}, p_{21}, p_{22}]$,

$$\rightsquigarrow \det \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = 0 \iff \operatorname{rank} \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \leq 1.$$

Rulings

$\sigma(\mathbb{P}^1 \times \mathbb{P}^1) = \{[p_{11}, p_{12}; p_{21}, p_{22}] : \det(p_{ij}) = 0\}$ is an example of a *determinantal variety.*
This example has two families of lines inside of it; the images of $\sigma([p_{11}, p_{12}] \times \{Q\})$ and $\sigma(\{Q\} \times [p_{21}, p_{22}])$, which are called *rulings of the surface.*

- Let $\Delta \subset \mathbb{R}^4$ be the tetrahedron with vertices given by the four basis vectors, $A_i = e_i$, and let a general point $p = (p_{ij})$ inside of Δ be represented by

$$p_{ij} = (p_{11}, p_{12}, p_{21}, p_{22}) = \begin{array}{|c|c|} \hline p_{11} & p_{12} \\ \hline p_{21} & p_{22} \\ \hline \end{array}$$

- Fienberg and Gilbert have shown that the two rulings are given by

$$\begin{array}{|c|c|} \hline st & s(1-t) \\ \hline t(1-s) & (1-s)(1-t) \\ \hline \end{array} \qquad (0 \leq s, t \leq 1),$$

which is a hyperbolic paraboloid inside of Δ.

- They call this the *manifold of independence*; any point on this surface has independent row and column marginal totals.

TODO: FIGURE.

- Suppose $\mathcal{P} \subset \Delta_{r-1}$ is a model for a random variable X with state space $[r]$.
- Moreover, assume that there is a *hidden* or *latent* random variable Y with state space $[s]$, and for each $j \in [s]$, the conditional distribution of X given $Y = j$ is $p^{(j)} \in \mathcal{P}$.
- The hidden variable Y also has some probability distribution $\pi \in \Delta_{s-1}$.

So the joint distribution of Y and X is given by the formula

$$P(Y = j; X = i) = \pi_j \cdot p_i^{(j)}.$$

Mixture Models

- But as Y is hidden, we can only observe the marginal distribution of X, that is

$$P(X = i) = \sum_{j=1}^{s} \pi_j \cdot p_i^{(j)}.$$

- In other words, the marginal distribution of X is the convex combination of the s distributions $p^{(1)}, \ldots, p^{(s)}$, with weights given by π.

Definition

Let $\mathcal{P} \subset \Delta_{r-1}$ be a statistical model. The *s-th mixture model* is

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

Mixture Models

- Mixture models provide ways to build complex models out of simpler ones.
- Basic assumption is that the underlying population to be modelled can be split into s disjoint sub-populations.
- Restricted to each sub-population, the observable X follows a probability distribution from the simple model $\mathcal{P}$.
- After marginalisation though, the structure becomes significantly more complex as it is now a convex combination of these simple distributions.

LOOK AT 'ALGEBRAIC STATISTICS FOR COMPUTATIONAL BIOLOGY' - CHAPTER 14.