Three-Way Contingency Tables

Let X, Y and Z be random variables that have a, b and c states respectively. A *probability distribution* P for these random variables is an $a \times b \times c$-table of non-negative real numbers that sum to one.

The entries of the table P are the probabilities

$$P_{ijk} \;=\; \mathrm{Prob}(X = i, Y = j, Z = k).$$

The set of all distributions is a simplex Δ of dimension $abc - 1$.

A *statistical model* is a subset $\mathcal{M}$ of Δ which can be described by polynomial equations and inequalities in the coordinates P_{ijk}.

Typically, the model $\mathcal{M}$ is presented as the image of a polynomial map $P : \Theta \mapsto \Delta$ where Θ is a polynomially described subset of $\mathbb{R}^n$.

Independence

The distribution P is called *independent* if each probability is the product of the corresponding marginal probabilities:

$$P_{ijk} \;=\; P_{i++} \cdot P_{+j+} \cdot P_{++k}$$

Here, for instance,

$$P_{i++} \;= \mathrm{Prob}(X = i) \;=\; \sum_{j=1}^{b} \sum_{k=1}^{c} P_{ijk}$$

The independence model has the parametric representation

$$\begin{aligned} \Theta \;=\; \Delta_{a-1} \times \Delta_{b-1} \times \Delta_{c-1} \quad &\rightarrow \quad \Delta \;=\; \Delta_{abc-1} \\ (\alpha, \beta, \gamma) \quad &\mapsto \quad (P_{ijk}) \;=\; (\alpha_i \beta_j \gamma_k) \end{aligned}$$

The image is known as the Segre variety in algebraic geometry. Its points are the $a \times b \times c$-tables of tensor rank one.

Three Binary Variables

If $a = b = c = 2$ then the independence model (Segre variety) is the threefold in Δ_7 (or in $\mathbb{P}^7$) which has the parametrization:

$$\begin{aligned} P_{000} &= \alpha\beta\gamma & P_{001} &= \alpha\beta(1-\gamma) \\ P_{010} &= \alpha(1-\beta)\gamma & P_{011} &= \alpha(1-\beta)(1-\gamma) \\ P_{100} &= (1-\alpha)\beta\gamma & P_{101} &= (1-\alpha)\beta(1-\gamma) \\ P_{110} &= (1-\alpha)(1-\beta)\gamma & P_{111} &= (1-\alpha)(1-\beta)(1-\gamma) \end{aligned}$$

This threefold is cut out by the trivial constraint

$$P_{000} + P_{001} + P_{010} + P_{011} + P_{100} + P_{101} + P_{110} + P_{111} = 1$$

and the Markov basis which consists of nine *quadratic binomials*:

$$\begin{array}{lll} P_{100}P_{111} - P_{101}P_{110}, & P_{010}P_{111} - P_{011}P_{110}, & P_{010}P_{101} - P_{011}P_{100}, \\ P_{001}P_{111} - P_{011}P_{101}, & P_{001}P_{110} - P_{011}P_{100}, & P_{000}P_{111} - P_{011}P_{100}, \\ P_{000}P_{110} - P_{010}P_{100}, & P_{000}P_{101} - P_{001}P_{100}, & P_{000}P_{011} - P_{001}P_{010}. \end{array}$$