The geometry of Iterative Proportional Scaling for log-linear models in 2D

Pedro N. Jorge

University of Porto

December 13, 2025

Outline

Part 1: Log-linear models

Part 2: Iterative Proportional Scaling

Outline

Part 1: Log-linear models

Part 2: Iterative Proportional Scaling

Probability Simplex and Models

- Let $\Omega = [n] = \{1, \dots, n\}$ be a discrete sample space. The *probability simplex* Δ_{n-1} is the set of all possible probability distributions on Ω:

$$\Delta_{n-1} = \left\{ (p_1, \dots, p_n) \mid p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \right\}.$$

- A *parametric statistical model* $\mathcal{M}_\Theta$ is a family of probability distributions parametrized by $\theta \in \Theta \subseteq \mathbb{R}^d$:

$$\mathcal{M}_\Theta = \{p_\theta \in \Delta_{n-1} : \theta \in \Theta\}.$$

- Therefore we have $\mathcal{M}_\Theta \subseteq \Delta_{n-1}$.

Example: Independence model in 2 binary r.v.

Simple non-trivial example: Two independent coin tosses.

Figure 1: Δ_3 (tetrahedron) with independence model surface

- A *log-linear model* is a special kind of subset inside the probability simplex (To be studied!)

Outline

Part 1: Log-linear models

Part 2: Iterative Proportional Scaling

Data and Maximum Likelihood Estimation

- We observe data $D = X^{(1)}, \dots X^{(n)}$ i.i.d from some unknown distribution.
- We choose a parametric model $\mathcal{M}_\Theta = \{p_\theta : \theta \in \Theta\}$ that we believe contains or approximates the true distribution.
- Objective: Find the best estimated distribution $p_{\hat{\theta}} \in \mathcal{M}_\Theta$.
- The *likelihood* of θ given data D is:

$$L(\theta \mid D) = p_\theta(D) = \prod_{i=1}^{n} p_\theta(X^{(i)})$$

- The *maximum likelihood estimate* (MLE) is:

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} L(\theta \mid D).$$

MLE calculation

In order to calculate the MLE, we can either:

- Analytically solve the *score equations*

$$\nabla_\theta \log L(\theta \mid D) = 0$$

- Use an algorithm like *Iterative Proportional Scaling* (IPS) to approximate a solution for log-linear models. (To be studied!)

IPS convergence

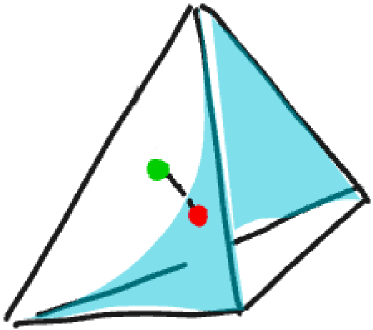

Figure 2: Green: true distribution. Red: MLE from IPS

- **Goal: Understand geometry and convergence properties of IPS.**