1 Kullback-Leibler distance

Consider the uniform discrete probability distribution $p = (p_j)_{1\leq j\leq n}$ on the set $\Omega = \{\,\omega_1, \omega_2, \ldots, \omega_n\,\}$:

$$p_j = P(\{\omega_j\}) = 1/n.$$

If $x = (x_j)$ with $x_j > 0$ and $\sum x_j = 1$ is another probability distribution on Ω, then the *Kullback-Leibler* distance $d_{KL}(x, p)$ of x from p is defined as

$$d_{KL}(x,p) = \sum\nolimits_j p_j \log(p_j/x_j) = -log(n) - \frac{1}{n}\sum\nolimits_j \log(x_j). \tag{1}$$

This function is convex in the variable x and also symmetric in x. The symmetry uses the fact that the p_j are all equal and will be used for the analytic solution of the minimization problems below. Note that we have

$$\nabla d_{KL}(x,p) = -\tfrac{1}{n}(1/x_1, 1/x_2, \ldots, 1/x_n)' \quad \text{and} \tag{2}$$

$$\nabla^2 d_{KL}(x,p) = \tfrac{1}{n} diag(1/x_1^2, 1/x_2^2, \ldots, 1/x_n^2), \tag{3}$$

where $diag(\lambda)$ denotes the diagonal matrix with the vector λ on the diagonal as usual.

2 Minimization of d_{KL} under probability constraints

Now let $A_k \subseteq \Omega$, $k = 1, \ldots, m$ be *disjoint* events (subsets) and consider the convex minimization problem

$$x^* = argmin\{\, d_{KL}(x,p) \,:\, P^x(A_k) = q_k \,\}. \tag{4}$$

Here $P^x(A) = E^x[1_A]$ denotes the probability of the event A under the discrete probability distribution $x = (x_j)$ on the set Ω. Note that a constraint on the probabilities x of the form $P^x(A) = r$ has the form

$$r = P^x(A) = \sum\nolimits_j x_j 1_A(\omega_j)$$

and is therefore a linear constraint in the variable x. Moreover the right hand side is a symmetric function of the variables x_j. Consequently the solution x^* of (4) must be symmetric under all permutations of coordinates which leave the sets A_k invariant, in other words the probability function

$$x^* : \omega_j \mapsto x_j^* = P^*(\omega_j)$$

is constant on all the sets A_k as well as the complement $D = [\cup A_k]^c$. This uses the fact that the A_k are disjoint since this implies that points $\omega \in \Omega$ which are in the same set A_k or are in D cannot be distinguished by the conditions $\omega \in A_k$ (i.e. if it is only determined in which of the sets A_k they are).

More formally the system of constraints

$$r_k = P^x(A_k) = \sum\nolimits_j x_j 1_{A_k}(\omega_j) \tag{5}$$

is invariant under all permutations of the variables x_j which (when applied to the points ω_j) leave the sets A_k invariant. Thus the solution x^* has the form

$$x_j^* = \begin{cases} q_k & \text{if } j \in A_k \\ q_* & \text{if } j \in D \end{cases}$$

and the variables q_k, q_* can be computed from the following system of equations

$$\begin{aligned} r_k &= P^{x^*}(A_k) = q_k |A_k| \\ 1 - \sum\nolimits_k r_k &= P^{x^*}(D) = q_* |D| \end{aligned}$$

or explicitly

$$x_j^* = \begin{cases} r_k/|A_k| & \text{if } \omega_j \in A_k \\ \frac{1}{|D|}\left(1 - \sum_k r_k\right) & \text{if } \omega_j \in D. \end{cases} \tag{6}$$

Here $|D|$ denotes the cardinality of the set D as usual.