最大信息熵

信息量跟概率成反比关系

信息量：$\log\frac{1}{p} = -\log p$

熵：$E_{p(x)}[-\log p] = \int -p(x)\log p(x)\,dx$

$= -\sum_x p(x)\cdot\log p(x)$

最大熵 $\Leftrightarrow$ 等可能　最大熵是对"等可能"的尽量描述

假设 x 是离散的：$\sum_{i=1}^{k} p_i = 1$

x	1	2	$\cdots$	k
p	p_1	p_2	$\cdots$	p_k

$$\begin{cases} \max H[p] = \max -\sum_{i=1}^{k} p_i \log p_i \\ s.t. \sum_{i=1}^{k} p_i = 1 \end{cases} \Leftrightarrow \begin{cases} \min \sum_{i=1}^{k} p_i \log p_i \\ s.t. \sum_{i=1}^{k} p_i = 1 \end{cases}$$

$p = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$

$\hat{p}_i = \arg\max H[p]$

$= \arg\min \sum_{i=1}^{k} p_i \log p_i$

$\mathcal{L}(p,\lambda) = \sum_{i=1}^{k} p_i \log p_i + \lambda(1-\sum_{i=1}^{k} p_i)$

$\frac{\partial \mathcal{L}}{\partial p_i} = \log p_i + \cancel{p_i}\cdot\frac{1}{\cancel{p_i}} - \lambda = 0$

$\log p_i + 1 - \lambda = 0 \Rightarrow \hat{p}_i = \exp(\lambda - 1)$ → constant

则 $\hat{p}_1 = \hat{p}_2 = \cdots = \hat{p}_k = \frac{1}{k}$　$p(x)$ 是均匀分布

最大熵 → 满足已知事实（约束）→ 最大熵原理

$Data = \{x_1, x_2, \cdots, x_N\}$

经验分布：$\hat{p}(X=x) = \hat{p}(x) = \frac{count(x)}{N}$

$E_{\hat{p}}[x]$，$Var_{\hat{p}}[x]$，$f(x)$ 是任意关于 x 的函数

$\boxed{E_{\hat{p}}[f(x)] = \Delta \rightarrow 已知}$