离散变量的熵

- KL散度是度量两个分布之间差异的函数，在各种变分方法中，都有它的身影

KL散度的公式为：

$KL(p||q) = \sum p(x)log\frac{p(x)}{q(x)}$，这个是离散概率分布的公式，

$KL(p||q) = \int p(x)log\frac{p(x)}{q(x)}dx$，这个是连续概率分布的公式。

e.g.

- $x \in \{0, 1\}$ $\quad p \quad q$

  $p(0) = 1-r \quad p(1) = r$

  $q(0) = 1-s \quad q(1) = s$

  $KL(p||q) = (1-r)\log\frac{1-r}{1-s} + r\log\frac{r}{s}$

  $KL(q||p) = (1-s)\log\frac{1-s}{1-r} + s\log\frac{s}{r}$

  距离
  不对称.

定理

- $KL(p||q) \geq 0$ 当 $p=q$ 时为0.

两组正序列的KL距离

引理：若 $\sum a_i$ 与 $\sum b_i$ 是任意两组正序列.

(1) $\sum a_i \log \frac{b_i}{a_i} + \sum (a_i - b_i) \le 0$

$\sum a_i \log \frac{a_i}{b_i} + \sum (b_i - a_i) \ge 0$

(2) If $\sum a_i \ge \sum b_i$ then

$\sum a_i \log \frac{b_i}{a_i} \le 0 \iff \sum a_i \log \frac{a_i}{b_i} \ge 0$　　$a_i = b_i$ 取等号

证明:

For $x>0$　$\log x$　$x=1$ 处泰勒展开

$\because \log x = (x-1) - \frac{(x-1)^2}{2} \frac{1}{\theta^2}$　$\theta \in [1, x]$

$\therefore \log \frac{b_i}{a_i} = \left(\frac{b_i}{a_i} - 1\right) - \frac{1}{2\theta^2}\left(\frac{b_i}{a_i} - 1\right)^2$

$\Rightarrow a_i \log \frac{b_i}{a_i} = (b_i - a_i) - \frac{a_i^3}{2a_i^2\theta^2}\left(\frac{b_i}{a_i} - 1\right)^2$　　$\theta_i \in [1, \frac{b_i}{a_i}]$

$\sum$　　$\sum$　　$\sum$

$\Rightarrow \sum a_i \log \frac{b_i}{a_i} = \sum (b_i - a_i) - \sum \frac{a_i^3}{2a_i^2\theta_i^2}\left(\frac{b_i}{a_i} - 1\right)^2$　$a_i\theta_i \in [a_i, b_i)$

应用

- 数据压缩和密度估计之间有一种隐含关系，因为当我们知道真实分布后，就可以给出有效的压缩。

$$KL(p\|q) \simeq \frac{1}{N}\sum\{-\ln q(x_n|\theta) + \ln p(x_n)\}$$

通过有限观察的似然。

本式只与 θ 有关是第一项