- 一般概念

设$\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N\}$为$N$个用于估计一未知参数$\boldsymbol{\theta}$的密度函数的样本，$\boldsymbol{x}^i$被一个接着一个逐次地给出。于是用贝叶斯定理，可以得到在给定了$\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^N$之后，$\boldsymbol{\theta}$的后验概率密度的迭代表示式为：

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^N) = \frac{p(\boldsymbol{x}^N \mid \boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\theta \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})}{p(\boldsymbol{x}^N \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})}$$

#注#

$$\begin{aligned} p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^N) &= \frac{p(\boldsymbol{x}^N \mid \boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})}{p(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^N)} \\ &= \frac{p(\boldsymbol{x}^N \mid \boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})}{p(\boldsymbol{x}^N \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})} \\ &= \frac{p(\boldsymbol{x}^N \mid \boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})\, p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})}{p(\boldsymbol{x}^N \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})} \end{aligned}$$

其中，对于$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^N)$而言，$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})$是它的先验概率，当加入新的样本$\boldsymbol{x}_N$后，得到经过修正的新的概率密度$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^N)$。如此一步步向前推，则$p(\boldsymbol{\theta})$应为最初始的先验概率密度，当读入第一个样本$\boldsymbol{x}^1$时，经过贝叶斯定理计算，可得到后验概率密度$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1)$。以此为新的一步，将$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1)$作为第二步计算的先验概率密度，读入样本$\boldsymbol{x}_2$，又得到第二步的后验概率密度$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \boldsymbol{x}^2)$，依此可以算出最后的后验概率密度$p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^N)$，从而得到最终的结果。

这里，需要先知道最初始的概率密度函数$p(\boldsymbol{\theta})$。至于全概率$p(\boldsymbol{x}^N \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1})$则可通过下式算出：

$$p(\boldsymbol{x}^N \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1}) = \int_x p(\boldsymbol{x}^N \mid \boldsymbol{\theta}, \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1}) p(\boldsymbol{\theta} \mid \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{N-1}) d\boldsymbol{\theta}$$

该值与未知量$\boldsymbol{\theta}$无关，可认为是一定值。