516021910226 林耘丰

日出问题

在仅知道过去5000年太阳每日升起的条件下，求明天日出的概率。

解答

根据贝叶斯概率，可以将过去5000年的日出情况看作是已知信息E，明天日出的事件为假设H，而所求即为在已知信息下的后验概率(posterior probability)。由贝叶斯公式得后验概率为：

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

其中，$P(E|H)$ 为在给定明天日出情况的假设下，过去日出情况的概率，即为假设的可能性(likelihood)。$P(H)$ 为在不知道日出情况的条件下对明天日出的假设，即为先验概率(prior probability)。而$P(E)$是一个在不同假设下均不变的常数。

为了求出 $P(H|E)$，可以假设每天日出的概率为 p，且每次日出是独立事件。设过去的观察一共有 n 天，其中有 s 天为日出。则过去的日出概率可以看做，即：

$$P(E|H) = p^s(1-p)^{n-s}$$

而先验概率 $P(H)$ 是无任何信息下 p 的概率分布。此时遇到一个问题，太阳不升起到底能否发生，对这一信息的掌握与否将影响到 $P(H)$。因此，首先讨论已知太阳不升起能够发生这一信息的情况。此时，$P(H) = 1, 0 < p < 1$，即 p 均匀地取 $(0, 1)$ 上的任何值。

直接计算 $P(E)$ 比较困难，但是由于 $P(H|E)$ 是概率，因此在所有不同假设下，概率的和应为1。而假设的概率 $p \in (0, 1)$ ，于是转化为求$P(E|H)P(H)$对 p 的积分
$\int_0^1 p^s(1-p)^{n-s}dp = \mathrm{Beta}(s+1, n-s+1) = \frac{s!(n-s)!}{(n+1)!} = P(E)$。
于是 p 的概率密度分布函数为 $f(p) = \frac{(n+1)!}{s!(n-s)!}p^s(1-p)^{n-s}$
由于假设每天日出概率为 p ，因此明天日出概率也应该为 p 。只需求出 p 的数学期望即可。

$$E[p] = \int_0^1 pf(p)dp = \frac{s+1}{n+2}$$

设$s = n = 365 \times 5000$，则明天日出的概率为 $\frac{s+1}{n+2} \approx 99.9999452\%$

然后，讨论另一种条件下的先验概率，即我们不知道太阳能够升起或者不升起这一信息。此时，p 可取0和1，此时 p 的先验概率可以为Haldane prior，即 $P(H) = p^{-1}(1-p)^{-1}$ 。可以理解为，在仅观察到日出（或不日出）时，日出的概率只可能取1（或0）。其它类似的先验概率也可以使用，例如$\frac{1}{\sqrt{p(1-p)}}$ 等。注意这些不是概率密度函数，因为它的积分不收敛，而这并不影响后验概率的计算。重复前面计算过程，得到

$$E[p] = \frac{s}{n}$$

然而，在 $s = n$ 的条件下，计算得到 $P(E) = \frac{(s-1)!(n-s-1)!}{(n-1)!} = \infty$ 。此时公式将不再适用。解决的办法可以是采用 $\frac{1}{\sqrt{p(1-p)}}$ 作为先验概率，从而得到

$$E[p] = \frac{s+0.5}{n+1}$$

则明天日出概率为 $\frac{s+0.5}{n+1} \approx 99.9999726\%$

推广

事实上，上述期望值可以推广到一般情况。假设有 n 次互相独立的重复实验，其中有 s 次“成功”，则：

1. 假设我们仅仅知道实验会有成功和失败两种结果这一信息，则下一次实验成功的概率为

$$P = \frac{s+1}{n+2}$$

2. 假设我们对实验的结果的种类等毫不知情，则下一次实验成功的概率为

$$P = \frac{s}{n} \ \ or \ \ \frac{s+0.5}{n+1}$$

这一结果被称为拉普拉斯继承法则(Rule of succession)，是由拉普拉斯为了解决日出问题而给出的一个解法。它适用于在实验结果很少甚至不完整的情况下对实验概率的估计。

例如，两个人独立地抛同一枚未知的硬币，且不公开结果。假设第一个人结果为正面，第二个人为反面。那么，第一个人计算下一次抛硬币为正面的概率将为 $\frac{2}{3}$，而第二个人计算结果为 $\frac{1}{3}$ 。现在他们交流实验结果，他们将知道硬币分别有一次正面和反面朝上。于是他们的计算结果都将为 $\frac{1}{2}$ 。

补充

在实验结果不充分的时候，对实验的概率估计将很大程度上依赖于先验概率。而日出问题的前提正是无信息的先验概率。但实际上，对于了解日出日落背后的天体运行规律的我们来说，99.9999% 这一概率仍然是极低的。