模式识别大作业

16231235 李谨杰

题目：一枚硬币，掷了一百次都是正面朝上，请问再掷一次正面朝上的概率是多少？

一、 问题分析

这个问题实际上是估计一个二项分布事件的发生的概率 p，是一个参数估计问题。

首先说一种错误的解答：我们小学初中都学过，对于理想的硬币模型，扔出正面和反面的概率均为 0.5。而每次掷硬币均为独立实验，即使前面都是正面朝上，那么再掷一次正面朝上的概率还是 0.5，因此很多人认为这个问题的答案是 0.5。犯这样的错误的原因在于：用理想模型估计问题概率，没有考虑实际情况。

事实上，扔一百次硬币全部朝上的概率为$\frac{1}{2^{100}} \approx 7.88 * 10^{-31}$,是一个几乎不可能发生的概率极小的小概率事件。因此，我们有充足的理由相信，硬币掷出正反两面的概率是不相等的。

因而这个问题被转换为，由 100 个实验样本估计掷一次硬币正面朝上的概率。我们不妨根据经验做出如下假设：每一次掷硬币事件相互之间都是独立的，且扔出正面的概率 p 保持不变。由概率论中所讲的二项分布概率模型可知，前一百次实验出现正面次数 X 发生 n 次的概率为：

$$P(X=n)=C_{100}^{n} p^{n}(1-p)^{100-n}$$

之后的问题即为根据样本数据估计该概率模型的参数 p。

参数估计分为点估计和区间估计，概率论上曾讲过区间估计与点估计中矩估计的内容，这里不再赘述。我们书上主要介绍的是点估计中的最大似然估计（MLE）与贝叶斯估计，老师还讲了最大后验估计（MAP），本文将采用这三种模型分别对参数 p 进行估计。

二、 最大似然估计（MLE）

2.1 基本思想

基于已有的样本 X，构造似然函数（likelihood function），找到一个 p 值，使得似然函数 L(p|X)达到最大，得到的 p 值即为最有可能的估计值。

2.2 具体求解

在本实验中，X=100 出现的概率最大，则 p 的选取应使得该实验结果出现的概率达到最大。据此构造似然函数为：

$$L(p|X) = P(X = 100) = p^{100}$$

由于 L(p)是一个单增函数，故当 p=1 时 L(p)最大。因此，由 MLE 出的正面朝上的概率为 1，即再掷一次正面朝上的概率为 1。

三、 最大后验估计（MAP）

3.1 基本思想

最大后验估计的思想也是构造后验概率函数，使得后验概率取极值。与最大似然估计相比，不同点在于 MAP 在估计未知量 θ 的函数中加入了先验概率$p(\theta)$，其中先验概率指的是

人们已经知道或普遍接受的规律。后验估计来自于贝叶斯公式：

$$p(\theta|X)=\frac{p(X|\theta)\cdot p(\theta)}{p(X)}$$

$$\text{posterior}=\frac{\text{likelihood}\cdot\text{prior}}{evidence}$$

因为公式中分母 p(X)与 θ 无关，该估计等价于选取 θ 使分子最大，即

$$\hat{\theta}_{MAP}=\text{argmax}_{\theta}\frac{p(X|\theta)p(\theta)}{p(X)}$$

$$=\text{argmax}_{\theta}\ p(X|\theta)p(\theta)$$

3.2 具体求解

在本实验中，对硬币的经验告诉我们，扔硬币正面朝上的次数满足二项分布，且满足先验概率在 θ=0.5 时取得最大值，因此可以选用二项分布的共轭先验分布函数 B 分布：

$$p(\theta)=p(p|\alpha,\beta)=\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}=\text{Beta}(p|\alpha,\beta)$$

其中 Beta 函数展开式为：

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

当 X 为正整数时，$\Gamma(n)=(n-1)!$。

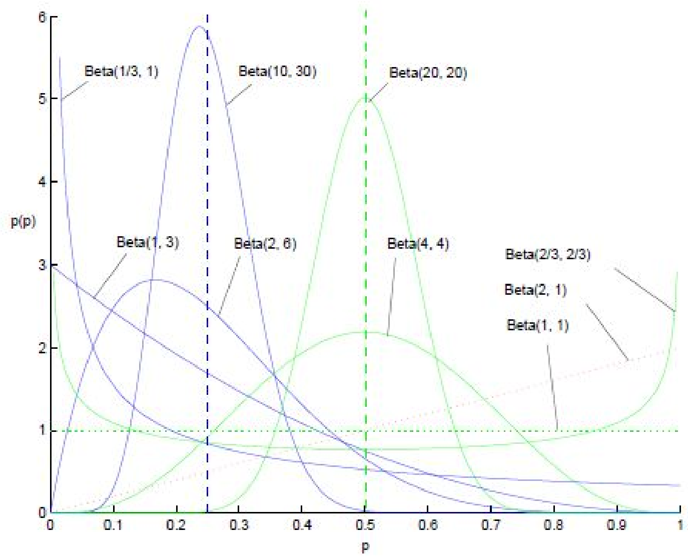


图 3.1 不同 α，β 值时 Beta 分布的概率密度函数

图 3.1 为不同 α，β 值时 Beta 分布的概率密度函数。如图可知，若在 θ=0.5 时取得最大值，需要$\alpha=\beta$。

由p(θ)与$\hat{\theta}_{MAP}$的定义，构造的 MAP 函数为

$$\hat{\theta}_{MAP} = \mathrm{p}^{100} \frac{1}{\mathrm{B}(\alpha，\beta)} \mathrm{p}^{\alpha-1}(1-\mathrm{p})^{\beta-1}$$

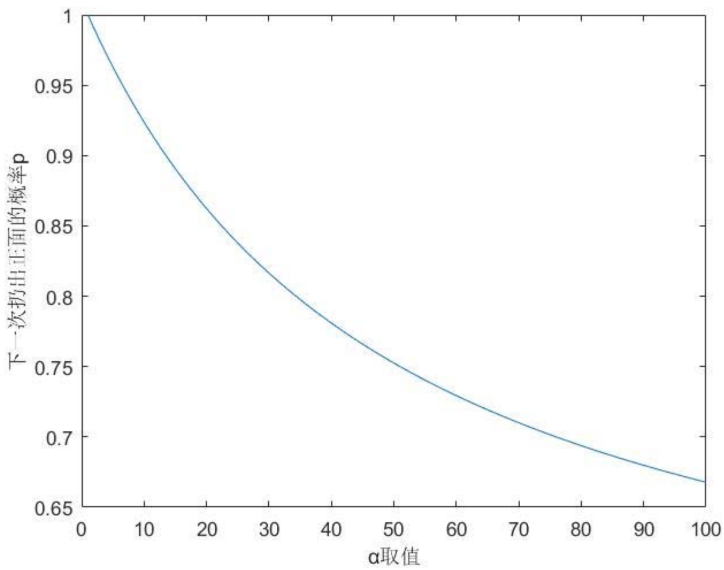
下面对该式求导：

$$\frac{\partial \hat{\theta}_{MAP}}{\partial p} = (\alpha+99)(1-\mathrm{p}) + \mathrm{p}(1-\beta) = 0$$

解得：

$$p = \frac{\alpha+99}{\alpha+\beta+98}$$

由上文可知，$\alpha = \beta$, α选值与得到正面概率 p 的关系作图如下：



从后验估计的思想可知，对参数的估计加入了先验概率。造币水平越高，我们认为 p 参数在 0.5 附近的可能性越大，因而最终对下一次掷出正面的估计概率就越小，符合上图的变化趋势。

若认为造币工艺较差，取$\alpha = \beta = 4$，则下一次掷出正面的概率为 p=0.9716；若认为造币工艺较好，取$\alpha = \beta = 20$，则下一次掷出正面的概率为 p=0.8623。可以看到α相当于将结果“拉”向先验概率的幅度，α越大，顶峰越尖锐，概率越集中。

四、 贝叶斯估计：

4.1 基本概念

贝叶斯估计和极大后验估计思想较为类似，均希望后验概率达到最大，但贝叶斯估计不具体估计参数的值，而是允许参数服从一定的概率分布，以整体的损失最小来估计参数。同时贝叶斯估计在估计后验概率时，不能忽略分母 p(X)。

回顾一下贝叶斯公式和全概率公式：

$$p(\theta \mid X) = \frac{p(X \mid \theta) \cdot p(\theta)}{p(X)}$$

$$p(X)=\int_{\theta\in\Theta}p(X\mid\theta)p(\theta)d\theta$$

由模式识别教材，贝叶斯估计的步骤为：

1. 根据对问题的认识确定θ的先验分布密度p(θ)
2. 由于样本独立同分布，而且已知样本密度函数的形式p(x|θ)，可以形式上求出样本集的联合分布如下，其中θ为变量。

$$p(\mathrm{x}\mid\theta)=\prod_{i=1}^{N}p(\mathrm{x}_i\mid\theta)$$

3. 利用贝叶斯公式求θ的后验概率分布：

$$p(\theta\mid x)=\frac{\mathrm{p}(\mathrm{x}\mid\theta)\,\mathrm{p}(\theta)}{\int_{\Theta}p(\mathrm{x}\mid\theta)\,\mathrm{p}(\theta)\,\mathrm{d}\,\theta}$$

4. 若采用平方误差损失函数，则θ的估计量为

$$\theta^{*}=\int_{\Theta}\theta p(\theta\mid\mathrm{x})\,\mathrm{d}\,\theta$$

4.2 具体求解

在这一小节中，为了避免混淆待估计参数 p 与概率符号 p，我将以θ代替带估计参数 p。

在本实验中，类似最大后验估计，我们假设先验分布为 B 分布。但构造贝叶斯估计时，不是采用使后验概率最大时的参数θ作为参数值，而是以满足 B 分布的θ的期望，即：

$$\theta^{*}=E[\theta\mid\mathrm{x}]=\int_{\Theta}\theta p(\theta\mid\mathrm{x})\,\mathrm{d}\,\theta$$

来作为 p 的参数估计值。

参数θ满足的概率分布为：

$$p(\theta\mid C,\alpha,\beta)=\frac{\prod_{i=1}^{N}p\left(C=c_i\mid\theta\right)p(\theta\mid\alpha,\beta)}{\int_0^1\prod_{i=1}^{N}p\left(C=c_i\mid\theta\right)p(\theta\mid\alpha,\beta)\mathrm{d}\theta}$$

$$=\frac{p^{n^{(1)}}(1-p)^{n^{(0)}}\dfrac{1}{\mathrm{B}(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}}{Z}$$

$$=\mathrm{Beta}\left(\theta\mid n^{(1)}+\alpha,n^{(0)}+\beta\right)$$

这里用到了公式

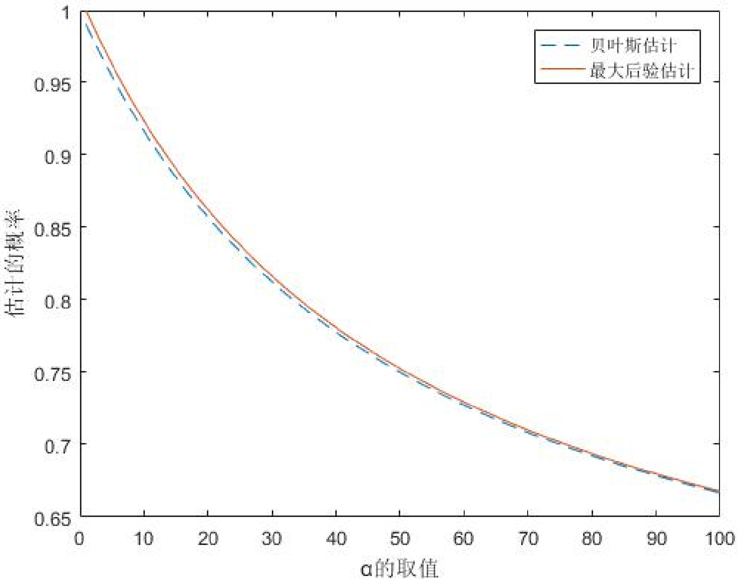
$$\int_p\prod_{t=1}^{|T|}P_t^{\alpha_t-1}=B(\alpha)$$

由该结果可知，根据贝叶斯估计，参数θ满足一个新的 B 分布，与先验分布形式相同，因此称二项分布与 Beta 分布是共轭分布。根据 B 分布的期望，我们有：

$$\theta^{*}=\frac{n^{(1)}+\alpha}{n^{(1)}+n^{(0)}+\alpha+\beta}=\frac{100+\alpha}{100+\alpha+\beta}$$

若认为造币工艺较差，取α = β = 4，则下一次掷出正面的概率为 p=0.9629；若认为造币工

艺较好，取$\alpha = \beta = 20$，则下一次掷出正面的概率为 p=0.8571。
贝叶斯估计与最大后验概率估计的 α 选值与得到正面概率 p 的关系作图如下：



可以看到相同先验概率分布的条件下，贝叶斯估计得出的概率稍低于最大后验概率，但整体上相差不大。这说明针对这个问题，最大后验分布与贝叶斯估计的结果类似。

五、　结论

本文通过最大似然估计、最大后验估计和贝叶斯估计这三种方法，对下一次掷硬币出现正面的概率进行预估，结果分别为 1, 0.9716 和 0.9629 ($\alpha = 4$)。因此，这是一枚严重有问题的硬币。

根据贝叶斯估计的原理，也可以随着样本数据实时调整估计量，即贝叶斯学习。当观测到新的数据时，可以通过这个数据调整后验概率，公式如下：

$$p(\theta \mid X^N) = \frac{\mathrm{p}(\mathrm{x}_N \mid \theta)\,\mathrm{p}(\theta \mid \mathrm{x}^{N-1})}{\int_p \mathrm{p}(\mathrm{x}_N \mid \theta)\,\mathrm{p}(\theta \mid \mathrm{x}^{N-1})\,\mathrm{d}\theta}$$

$$\theta^* = E[\theta \mid \mathrm{x}] = \int_\Theta \theta p(\theta \mid \mathrm{x})\,\mathrm{d}\theta$$

根据这一百个实验数据，按上述公式迭代一百次得到的概率与贝叶斯估计应该相同。但我用 matlab 编写上述程序时总是报错，可能是对连续变量的操作有些问题，目前还没有实现。

通过这个小例子，我还有两个感悟：1.　小概率事件发生意味着有什么地方出错了　2. 对生活中的概率问题一定要科学分析。

参考文献

1. 何书元. *概率论与数理统计*. 高等教育出版社，2006.
2. 张学工编著. *模式识别（第三版）*. 清华大学出版社，2010.
3. 三种参数估计方法（MLE，MAP，贝叶斯估计）
   https://blog.csdn.net/leo_xu06/article/details/51222215
4. 知乎问答 https://www.zhihu.com/question/29683794