

贝叶斯估计、最大似然估计、最大后验概率估计



SnailTyan (/u/7731e83f3a4e) [+ 关注](#)

2.5 2018.05.31 18:22* 字数 4804 阅读 10617 评论 3 喜欢 32

(/u/7731e83f3a4e)

Bayes' Theorem

文章作者：Tyan

博客：noahsnail.com (<https://links.jianshu.com/go?to=http%3A%2F%2Fnoahsnail.com>)

| CSDN ([https://links.jianshu.com/go?](https://links.jianshu.com/go?to=http%3A%2F%2Fblog.csdn.net%2Fquincuntial)

[to=http%3A%2F%2Fblog.csdn.net%2Fquincuntial](https://links.jianshu.com/go?to=http%3A%2F%2Fblog.csdn.net%2Fquincuntial)) | 简书

(https://www.jianshu.com/users/7731e83f3a4e/latest_articles)

1. 引言

贝叶斯估计、最大似然估计(MLE)、最大后验概率估计(MAP)这几个概念在机器学习和深度学习中经常碰到，读文章的时候还感觉挺明白，但独立思考时经常会傻傻分不清楚(ಥ_ಥ)，因此希望通过本文对其进行总结。

2. 背景知识



注：由于概率与数理统计需要了解的背景知识很多，因此这里只列出了部分内容，且写的较简略，许多概念的学习需要根据标题自己查找答案。

(/apps/
utm_sc
banner

2.1 概率与统计

概率统计是很多人都学过的内容，但概率论与统计学的关系是什么？先看一下概率论与统计学在维基百科中的定义：

概率论是集中研究概率及随机现象的数学分支，是研究随机性或不确定性等现象的数学。
统计学是在数据分析的基础上，研究如何测定、收集、整理、归纳和分析反映数据数据，以便给出

下面的一段话引自LarrB Wasserman的《All of Statistics》，对概率和统计推断的研究内容进行了描述：

The basic problem that we studB in probabilitB is:
Given a data generating process, what are the properities of the outcomes?

The basic problem of statistical inference is the inverse of probabilitB:
Given the outcomes, what can we saB about the process that generated the data?

概率论是在给定条件（已知模型和参数）下，对要发生的事件（新输入数据）的预测。
统计推断是在给定数据（训练数据）下，对数据生成方式（模型和参数）的归纳总结。
概率论是统计学的数学基础，统计学是对概率论的应用。

2.2 描述统计和推断统计

统计学分为描述统计学和推断统计学。描述统计，是统计学中描绘或总结观察量基本情况的统计总称。推断统计指统计学中研究如何根据样本数据去推断总体数量特征的方法。

描述统计是对数据的一种概括。描述统计是罗列所有数据，然后选择一些特征量（例如均值、方差、中位数、四分中位数等）对总体数据进行描述。推断统计是一种对数据的推测。推断统计无法获取所有数据，只能得到部分数据，然后根据得到的数据推测总体数据的情况。

2.3 联合概率和边缘概率

假设有随机变量 A 和 B ，此时 $P(A = a, B = b)$ 用于表示 $A = a$ 且 $B = b$ 同时发生的概率。这类包含多个条件且所有条件同时成立的概率称为**联合概率**。请注意，联合概率并不是其中某个条件成立的概率，而是所有条件同时成立的概率。与之对应地， $P(A = a)$ 或 $P(B = b)$ 这类仅与单个随机变量有关的概率称为**边缘概率**。

联合概率与边缘概率的关系如下：



$$P(A = a) = \sum_b P(A = a, B = b)$$

$$P(B = b) = \sum_a P(A = a, B = b)$$

(/apps/
utm_sc
banner

2.4 条件概率

条件概率表示在条件 $B = b$ 成立的情况下， $A = a$ 的概率，记作 $P(A = a|B = b)$ ，或者说条件概率是指事件 $A = a$ 在另外一个事件 $B = b$ 已经发生条件下的发生概率。为了简洁表示，后面省略 a, b 。

联合概率、边缘概率、条件概率的关系如下：

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

转换为乘法形式：

$$P(A, B) = P(B) * P(A|B) = P(A) * P(B|A)$$

2.5 全概率公式

如果事件 $A_1, A_2, A_3, \dots, A_n$ 构成一个完备事件组，即它们两两互不相容（互斥），其和为全集；并且 $P(A_i)$ 大于0，则对任意事件 B 有

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \dots + P(B|A_n)P(A_n) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

上面的公式称为全概率公式。全概率公式是对复杂事件 A 的概率求解问题转化为了在不同情况下发生的简单事件的概率的求和问题。

2.6 贝叶斯公式

由条件概率的乘法形式可得：

$$P(A|B) = \frac{P(B|A)}{P(B)} * P(A)$$

上面的式子称为贝叶斯公式，也叫做贝叶斯定理或贝叶斯法则。在贝叶斯定理中，每个名词都有约定俗成的名称：

- $P(A|B)$ 是已知 B 发生后 A 的条件概率，也由于得自 B 的取值而被称作 A 的后验概率，表示事件 B 发生后，事件 A 发生的置信度。
- $P(A)$ 是 A 的先验概率或边缘概率，表示事件 A 发生的置信度。
- $P(B|A)$ 是已知 A 发生后 B 的条件概率，也由于得自 A 的取值而被称作 B 的后验概率，也被称作似然函数。



- $P(B)$ 是 B 的先验概率或边缘概率，称为标准化常量。
- $\frac{P(B|A)}{P(B)}$ 称为标准似然比(这个叫法很多，没找到标准统一的叫法)，表示事件 B 为事件 A 发生提供的支持程度。

因此贝叶斯公式可表示为：后验概率=似然函数先验概率/标准化常量=标准似然比先验概率。根据标准似然比的大小，可分为下面三种情况：

- 如果标准似然比 > 1 ，则先验概率 $P(A)$ 得到增强，事件 B 的发生会增大事件 A 发生的可能性；
- 如果标准似然比 $= 1$ ，则先验概率 $P(A)$ 保持不变，事件 B 的发生不影响事件 A 发生的可能性；
- 如果标准似然比 < 1 ，则先验概率 $P(A)$ 得到削弱，事件 B 的发生会降低事件 A 发生的可能性。

由全概率公式、贝叶斯法则可得：

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

2.7 似然与概率

在英文中，似然 (likelihood) 和概率 (probability) 是同义词，都指事件发生的可能性。但在统计中，似然与概率是不同的东西。概率是已知参数，对结果可能性的预测。似然是已知结果，对参数是某个值的可能性预测。

2.8 似然函数与概率函数

对于函数 $P(x|\theta)$ ，从不同的观测角度来看可以分为以下两种情况：

- 如果 θ 已知且保持不变， x 是变量，则 $P(x|\theta)$ 称为概率函数，表示不同 x 出现的概率。
- 如果 x 已知且保持不变， θ 是变量，则 $P(x|\theta)$ 称为似然函数，表示不同 θ 下， x 出现的概率，也记作 $L(\theta|x)$ 或 $L(x;\theta)$ 或 $f(x;\theta)$ 。

注：注意似然函数的不同写法。

2.9 推断统计中需要了解的一些概念

- 假设实际观测值与真实分布相关，试图根据观测值来推测真实分布
- 由于观测值取值随机，因此由它们计算得到的估计值也是随机值
- 估计方式多种多样，且不同估计方式得到的估计值也有所不同

样本、样本容量、参数统计、非参数统计、估计量、真实分布、经验分布。

2.10 频率学派与贝叶斯学派



注：频率学派与贝叶斯学派只是解决问题的角度不同。

频率学派与贝叶斯学派探讨「不确定性」这件事时的出发点与立足点不同。频率学派从「自然」角度出发，试图直接为「事件」本身建模，即事件 A 在独立重复试验中发生的频率趋于极限 p ，那么这个极限就是该事件的概率。

(/apps/
utm_sc
banner

贝叶斯学派并不从试图刻画「事件」本身，而从「观察者」角度出发。贝叶斯学派并不试图说「事件本身是随机的」，或者「世界的本体带有某种随机性」，这套理论根本不言说关于「世界本体」的东西，而只是从「观察者知识不完备」这一出发点开始，构造一套在贝叶斯概率论的框架下可以对不确定知识做出推断的方法。

频率学派的代表是最大似然估计；贝叶斯学派的代表是最大后验概率估计。

2.11 共轭先验

在贝叶斯统计中，如果后验分布与先验分布属于同类，则先验分布与后验分布被称为共轭分布，而先验分布被称为似然函数的共轭先验。

2.12 Beta分布

在概率论中，Beta分布也称B分布，是指一组定义在 $(0, 1)$ 区间的连续概率分布，有两个参数 $\alpha, \beta > 0$ 。Beta分布的概率密度为：

$$\begin{aligned} f(x; \alpha, \beta) &= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 \mu^{\alpha-1}(1-\mu)^{\beta-1} d\mu} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\ &= \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} \end{aligned}$$

其中， $\Gamma(z)$ 是 Γ 函数。随机变量 X 服从Beta分布写作 $X \sim \text{Beta}(\alpha, \beta)$ 。

3. 问题定义

以抛硬币为例，假设我们有一枚硬币，现在要估计其正面朝上的概率 θ 。为了对 θ 进行估计，我们进行了10次实验（独立同分布，i.i.d.），这组实验记为

$X = x_1, x_2, \dots, x_{10}$ ，其中正面朝上的次数为6次，反面朝上的次数为4次，结果为 $(1, 0, 1, 1, 0, 0, 1, 1, 1)$ 。

4. 最大似然估计(MLE)

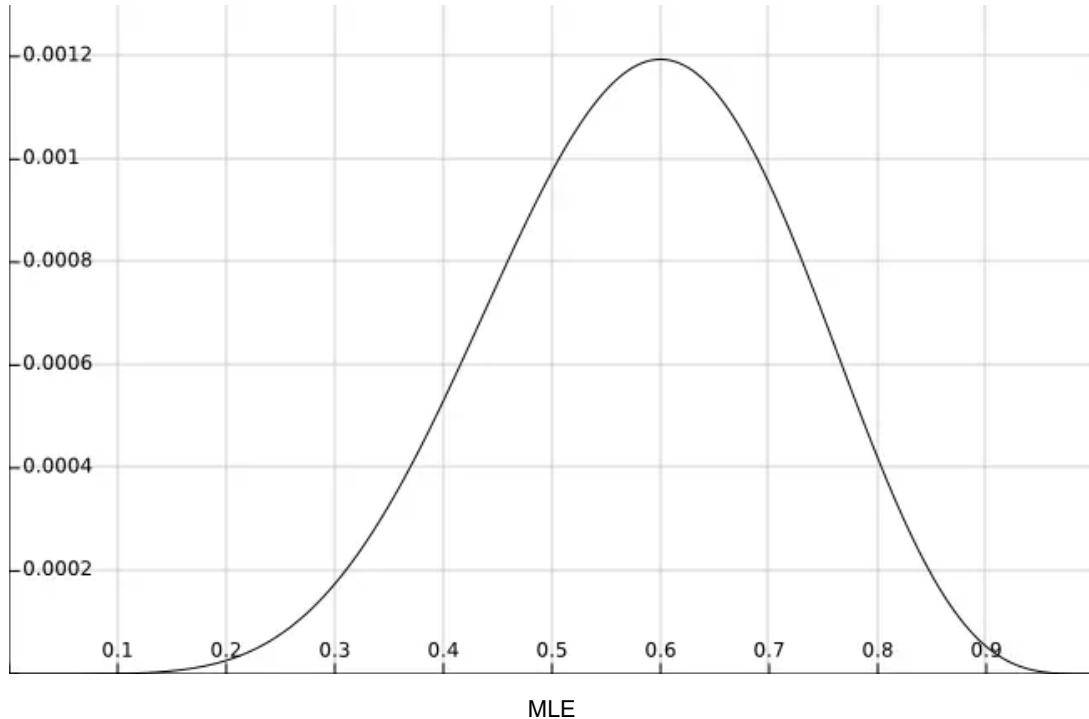
最大似然估计，英文为Maximum Likelihood Estimation，简称为MLE，也叫极大似然估计，是用来估计概率模型参数的一种方法。最大似然估计的思想是使得观测数据（样本）发生概率最大的参数就是最好的参数。



对一个独立同分布的样本集来说，总体的似然就是每个样本似然的乘积。针对抛硬币的问题，似然函数可写作：

$$L(X; \theta) = \prod_{i=0}^n P(x_i | \theta) = \theta^6 (1 - \theta)^4$$

根据最大似然估计，使 $L(X; \theta)$ 取得最大值的 θ 即为估计结果，令 $L(X; \theta)' = 0$ 可得 $\hat{\theta} = 0.6$ 。似然函数图如下：



由于总体的似然就是每个样本似然的乘积，为了求解方便，我们通常会将似然函数转成对数似然函数，然后再求解。可以转成对数似然函数的主要原因是对数函数并不影响函数的凹凸性。因此上式可变为：

$$\ln L(X; \theta) = \ln \prod_{i=0}^n P(x_i | \theta) = \sum_{i=0}^n \ln(P(x_i | \theta)) = 6 \ln(\theta) + 4 \ln(1 - \theta)$$

令 $\ln(L(X; \theta))' = 0$ 可得 $\hat{\theta} = 0.6$ 。

正态分布的最大似然估计

假设样本服从正态分布 $N \sim (\mu, \sigma^2)$ ，则其似然函数为

$$L(\mu, \sigma^2) = \prod_{i=0}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

对其取对数得：

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^n (x_i - \mu)^2$$

分别对 μ ， σ^2 求偏导，并令偏导数为0，得：

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=0}^n (x_i - \mu) = 0 \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=0}^n (x_i - \mu)^2 = 0 \end{cases}$$

解得：

$$\begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=0}^n x_i = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^n (x_i - \bar{x})^2 \end{cases}$$

$\hat{\mu}$, $\hat{\sigma}^2$ 就是正态分布中 μ , σ^2 的最大似然估计。

最大似然估计的求解步骤：

- 确定似然函数
- 将似然函数转换为对数似然函数
- 求对数似然函数的最大值（求导，解似然方程）

5. 最大后验概率估计(MAP)

最大后验概率估计，英文为Maximum A Posteriori Estimation，简称为MAP。回到抛硬币的问题，最大似然估计认为使似然函数 $P(X|\theta)$ 最大的参数 θ 即为最好的 θ ，此时最大似然估计是将 θ 看作固定的值，只是其值未知；最大后验概率分布认为 θ 是一个随机变量，即 θ 具有某种概率分布，称为先验分布，求解时除了要考虑似然函数 $P(X|\theta)$ 之外，还要考虑 θ 的先验分布 $P(\theta)$ ，因此其认为使 $P(X|\theta)P(\theta)$ 取最大值的 θ 就是最好的 θ 。此时要最大化的函数变为 $P(X|\theta)P(\theta)$ ，由于 X 的先验分布 $P(X)$ 是固定的（可通过分析数据获得，其实我们也不关心 X 的分布，我们关心的是 θ ），因此最大化函数可变为 $\frac{P(X|\theta)P(\theta)}{P(X)}$ ，根据贝叶斯法则，要最大化的函数 $\frac{P(X|\theta)P(\theta)}{P(X)} = P(\theta|X)$ ，因

此要最大化的函数是 $P(\theta|X)$ ，而 $P(\theta|X)$ 是 θ 的后验概率。最大后验概率估计可以看作是正则化的最大似然估计，当然机器学习或深度学习中的正则项通常是加法，而在最大后验概率估计中采用的是乘法， $P(\theta)$ 是正则项。在最大似然估计中，由于认为 θ 是固定的，因此 $P(\theta) = 1$ 。

最大后验概率估计的公式表示：

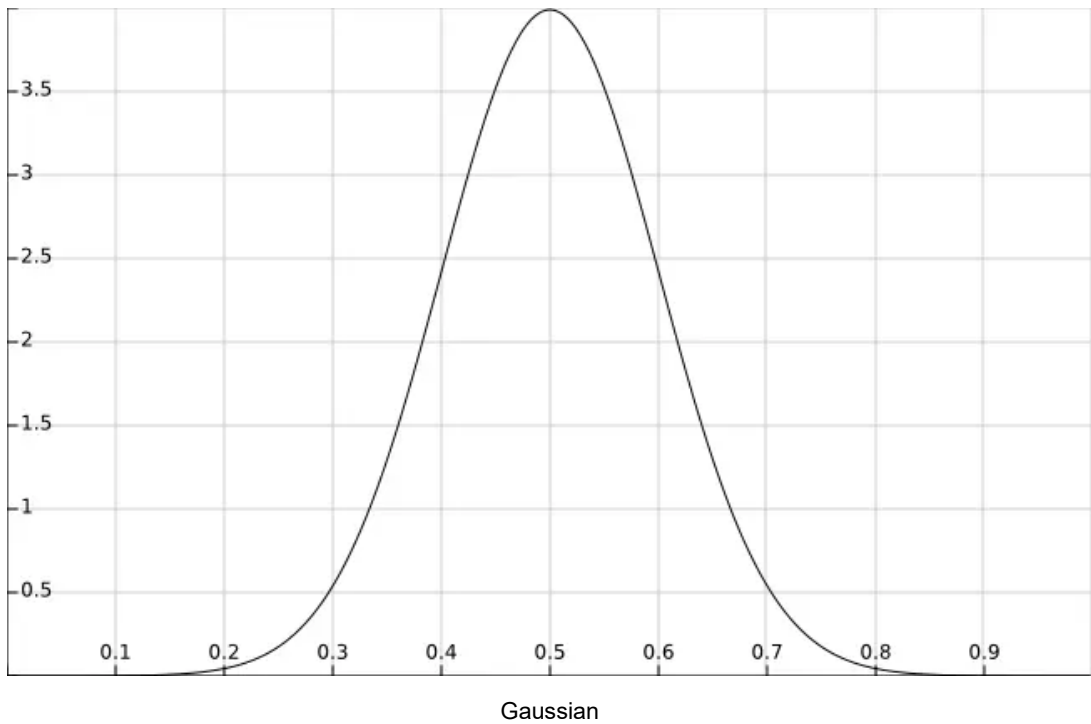
$$\operatorname{argmax}_{\theta} P(\theta|X) = \operatorname{argmax}_{\theta} \frac{P(X|\theta)P(\theta)}{P(X)} \propto \operatorname{argmax}_{\theta} P(X|\theta)P(\theta)$$

在抛硬币的例子中，通常认为 $\theta = 0.5$ 的可能性最大，因此我们用均值为 0.5 ，方差为 0.1 的高斯分布来描述 θ 的先验分布，当然也可以使用其它的分布来描述 θ 的先验分布。 θ 的先验分布为：

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} = \frac{1}{10\sqrt{2\pi}} e^{-50(\theta-0.5)^2}$$

先验分布的函数图如下：





(/apps/
utm_sc
banner

在最大似然估计中，已知似然函数为 $P(X|\theta) = \theta^6(1-\theta)^4$ ，因此：

$$P(X|\theta)P(\theta) = \theta^6 \times (1-\theta)^4 \times \frac{1}{10\sqrt{2\pi}} \times e^{-50(\theta-0.5)^2}$$

转换为对数函数：

$$\ln(P(X|\theta)P(\theta)) = \ln(\theta^6 \times (1-\theta)^4 \times \frac{1}{10\sqrt{2\pi}} \times e^{-50(\theta-0.5)^2}) = 6\ln(\theta) + 4\ln(1-\theta) + \ln\left(\frac{1}{10\sqrt{2\pi}}\right) - 50(\theta-0.5)^2$$

令 $\ln(P(X|\theta)P(\theta))' = 0$ ，可得：

$$100\theta^3 - 150\theta^2 + 40\theta + 6 = 0$$

由于 $0 \leq \theta \leq 1$ ，解得： $\hat{\theta} \approx 0.529$ 。 $P(X|\theta)P(\theta)$ 的函数图像如下，基本符合 θ 的估计值 $\hat{\theta}$ ：

如果我们用均值为0.6，方差为0.1的高斯分布来描述 θ 的先验分布，则 $\hat{\theta} = 0.6$ 。由此可见，在最大后验概率估计中， θ 的估计值与 θ 的先验分布有很大的关系。这也说明一个合理的先验概率假设是非常重要的。如果先验分布假设错误，则会导致估计的参数值偏离实际的参数值。

先验分布为Beta分布

如果用 $\alpha = 3, \beta = 3$ 的Beta分布来描述 θ 的先验分布，则

$$P(X|\theta)P(\theta) = \theta^6 \times (1 - \theta)^4 \times \frac{1}{B(\alpha, \beta)} \times \theta^{\alpha-1}(1 - \theta)^{\beta-1}$$

令 $P(X|\theta)P(\theta)'$ = 0求解可得：

$$\hat{\theta} = \frac{\alpha + 5}{\alpha + \beta + 8} = \frac{8}{3 + 3 + 8} \approx 0.57$$

$Beta(3, 3)$ 的概率密度图像如下图：

Beta(3,3)

最大后验概率估计的求解步骤：

- 确定参数的先验分布以及似然函数
- 确定参数的后验分布函数
- 将后验分布函数转换为对数函数
- 求对数函数的最大值（求导，解方程）

6. 贝叶斯估计



贝叶斯估计是最大后验估计的进一步扩展，贝叶斯估计同样假定 θ 是一个随机变量，但贝叶斯估计并不是直接估计出 θ 的某个特定值，而是估计 θ 的分布，这是贝叶斯估计与最大后验概率估计不同的地方。在贝叶斯估计中，先验分布 $P(X)$ 是不可忽略的。回到抛硬币的例子中，在已知 X 的情况下，描述 θ 的分布即描述 $P(\theta|X)$ ， $P(\theta|X)$ 是一种后验分布。如果后验分布的范围较窄，则估计值的准确度相对较高，反之，如果后验分布的范围较广，则估计值的准确度就较低。

贝叶斯公式：

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

在连续型随机变量中，由于 $P(X) = \int_{\Theta} P(X|\theta)P(\theta)d\theta$ ，因此贝叶斯公式变为：

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_{\Theta} P(X|\theta)P(\theta)d\theta}$$

从上面的公式中可以看出，贝叶斯估计的求解非常复杂，因此选择合适的先验分布就非常重要。一般来说，计算积分 $\int_{\Theta} P(X|\theta)P(\theta)d\theta$ 是不可能的。对于这个抛硬币的例子

来说，如果使用共轭先验分布，就可以更好的解决这个问题。二项分布参数的共轭先验是Beta分布，由于 θ 的似然函数服从二项分布，因此在贝叶斯估计中，假设 θ 的先验分布服从 $P(\theta) \sim \text{Beta}(\alpha, \beta)$ ，Beta分布的概率密度公式为：

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

因此，贝叶斯公式可写作：

$$\begin{aligned} P(\theta|X) &= \frac{P(X|\theta)P(\theta)}{\int_{\Theta} P(X|\theta)P(\theta)d\theta} \\ &= \frac{\theta^6(1-\theta)^4 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}}{\int_{\Theta} \theta^6(1-\theta)^4 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} d\theta} \\ &= \frac{\theta^{\alpha+6-1}(1-\theta)^{\beta+4-1}}{\int_{\Theta} \theta^{\alpha+6-1}(1-\theta)^{\beta+4-1} d\theta} \\ &= \frac{\theta^{\alpha+6-1}(1-\theta)^{\beta+4-1}}{B(\alpha+6-1, \beta+4-1)} \\ &= \text{Beta}(\theta|\alpha+6-1, \beta+4-1) \\ &= \text{Beta}(\theta|\alpha+6, \beta+4) \end{aligned}$$

从上面的公式可以看出， $P(\theta|X) \sim \text{Beta}(\theta|\alpha+6, \beta+4)$ 。其中 B 函数，也称Beta函数，是一个标准化常量，用来使整个概率的积分为1。 $\text{Beta}(\theta|\alpha+6, \beta+4)$ 就是贝叶斯估计的结果。

如果使用贝叶斯估计得到的 θ 分布存在一个有限均值，则可以用后验分布的期望作为 θ 的估计值。假设 $\alpha = 3, \beta = 3$ ，在这种情况下，先验分布会在0.5处取得最大值，则 $P(\theta|X) \sim \text{Beta}(\theta|9, 7)$ ， $\text{Beta}(\theta|9, 7)$ 的曲线如下图：

Beta(9,7)

从上图可以看出，在 $\alpha = 3, \beta = 3$ 的情况下， θ 的估计值 $\hat{\theta}$ 应该在0.6附近。根据Beta分布的数学期望公式 $E(\theta) = \frac{\alpha}{\alpha + \beta}$ 可得：

$$\hat{\theta} = \int_{\Theta} \theta P(\theta|X) d\theta = E(\theta) = \frac{\alpha}{\alpha + \beta} = \frac{9}{9 + 7} = 0.5625$$

注：二项分布参数的共轭先验是Beta分布，多项式分布参数的共轭先验是Dirichlet分布，指数分布参数的共轭先验是Gamma分布，高斯分布均值的共轭先验是另一个高斯分布，泊松分布的共轭先验是Gamma分布。

贝叶斯估计要解决的不是如何估计参数，而是用来估计新测量数据出现的概率，对于新出现的数据 \tilde{x} ：

$$P(\tilde{x}|X) = \int_{\Theta} P(\tilde{x}|\theta) P(\theta|X) d\theta = \int_{\Theta} P(\tilde{x}|\theta) \frac{P(X|\theta) P(\theta)}{P(X)} d\theta$$

贝叶斯估计的求解步骤：

- 确定参数的似然函数
- 确定参数的先验分布，应是后验分布的共轭先验
- 确定参数的后验分布函数
- 根据贝叶斯公式求解参数的后验分布



7. 总结

从最大似然估计、最大后验概率估计到贝叶斯估计，从下表可以看出 θ 的估计值 $\hat{\theta}$ 是逐渐接近0.5的。从公式的变化可以看出，使用的信息是逐渐增多的。最大似然估计、最大后验概率估计中都是假设 θ 未知，但是确定的值，都将使函数取得最大值的 θ 作为估计值，区别在于最大化的函数不同，最大后验概率估计使用了 θ 的先验概率。而在贝叶斯估计中，假设参数 θ 是未知的随机变量，不是确定值，求解的是参数 θ 在样本 X 上的后验分布。

注：最大后验概率估计和贝叶斯估计都采用Beta分布作为先验分布。

Type	MLE	MAP	BE
$\hat{\theta}$	0.6	0.57	0.5625
f	$P(X \theta)$	$P(X \theta)P(\theta)$	$\frac{P(X \theta)P(\theta)}{P(X)}$

参考资料

- 书籍：程序员的数学2——概率统计
- 概率论与统计学的关系是什么？ (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F20269390>)
- 贝叶斯学派与频率学派有何不同？ (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F20587681>)
- 概率论 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%25A6%2582%25E7%258E%2587%25E8%25AE%25BA>)
- 推论统计学 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%258E%25A8%25E8%25AB%2596%25E7%25B5%25B1%25E8%25A8%2588%25E5%25AD%25B8>)
- 描述统计学 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%258F%258F%25E8%25BF%25B0%25E7%25BB%259F%25E8%25AE%25A1%25E5%25AD%25A6>)
- 统计学 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E7%25BB%259F%25E8%25AE%25A1%25E5%25AD%25A6>)
- 详解最大似然估计（MLE）、最大后验概率估计（MAP），以及贝叶斯公式的理解 (<https://links.jianshu.com/go?to=https%3A%2F%2Fblog.csdn.net%2Fu011508640%2Farticle%2Fdetails%2F72815981>)
- 如何理解条件概率？ (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F27462939>)

(/apps/
utm_sc
banner



10. 贝叶斯定理 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E8%25B4%259D%25E5%258F%25B6%25E6%2596%25AF%25E5%25AE%259A%25E7%2590%2586>)
11. 贝叶斯推断及其互联网应用（一）：定理简介 (https://links.jianshu.com/go?to=http%3A%2F%2Fwww.ruanyifeng.com%2Fblog%2F2011%2F08%2Fbayesian_inference_part_one.html)
12. 全概率公式 (<https://links.jianshu.com/go?to=https%3A%2F%2Fbaike.baidu.com%2Fitem%2F%25E5%2585%25A8%25E6%25A6%2582%25E7%258E%2587%25E5%2585%25AC%25E5%25BC%258F>)
13. 怎样用非数学语言讲解贝叶斯定理 (Bayes's theorem) ? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F19725590>)
14. 似然 (likelihood) 与概率 (probability) 的区别 (<https://links.jianshu.com/go?to=http%3A%2F%2Fyangfangs.github.io%2F2018%2F04%2F06%2Fthe-different-of-likelihood-and-probability%2F>)
15. 如何通俗地理解概率论中的「极大似然估计法」? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F24124998>)
16. 如何通俗地理解“最大似然估计法”? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.matongxue.com%2Fmadocs%2F447.html>)
17. 概率论与数理统计 (<https://links.jianshu.com/go?to=http%3A%2F%2F218.194.248.2%2F%7Ewuxiaogang%2Fcpcourse%2Fother%2Fstatic.pdf>)
18. All of Statistics: A Concise Course in Statistical Inference (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.ic.unicamp.br%2F%7Ewainer%2Fcursos%2F1s2013%2Fml%2Flivro.pdf>)
19. MLE, MAP, EM 和 point estimation 之间的关系是怎样的? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F19894595>)
20. 最大后验概率 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%259C%2580%25E5%25A4%25A7%25E5%2590%258E%25E9%25AA%258C%25E6%25A6%2582%25E7%258E%2587>)
21. 从最大似然估计开始，你需要打下的机器学习基石 (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.jiqizhixin.com%2Farticles%2F2018-01-09-6>)
22. 如何理解似然函数? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F54082000>)
23. 共轭先验 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E5%2585%25B1%25E8%25BD%25AD%25E5%2585%2588%25E9%25AA%258C>)

(/apps/
utm_sc
banner



24. 参数估计: 最大似然估计 (MLE) , 最大后验估计(MAP), 贝叶斯估计, 经验贝叶斯 (Empirical Bayes)与全贝叶斯(Full Bayes) (<https://links.jianshu.com/go?to=https%3A%2F%2Fblog.csdn.net%2Flin360580306%2Farticle%2Fdetails%2F51289543>)
25. 什么是最大似然估计、最大后验估计以及贝叶斯参数估计 (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.jiqizhixin.com%2Farticles%2F041505>)
26. 先验概率、后验概率以及共轭先验 (<https://links.jianshu.com/go?to=https%3A%2F%2Fblog.csdn.net%2Fbaimafujinji%2Farticle%2Fdetails%2F51374202>)
27. 认识Beta/Dirichlet分布 (<https://links.jianshu.com/go?to=https%3A%2F%2Fcosx.org%2F2013%2F01%2Flda-math-beta-dirichlet>)
28. B分布 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25CE%2592%25E5%2588%2586%25E5%25B8%2583>)
29. B函数 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25CE%2592%25E5%2587%25BD%25E6%2595%25B0>)
30. Beta distribution (https://links.jianshu.com/go?to=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FBeta_distribution)
31. Beta function (https://links.jianshu.com/go?to=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FBeta_function)
32. Beta Distribution PDF Grapher (<https://links.jianshu.com/go?to=http%3A%2F%2Feurekastatistics.com%2Fbeta-distribution-pdf-grapher%2F>)
33. 文本语言模型的参数估计-最大似然估计、MAP及贝叶斯估计 (<https://links.jianshu.com/go?to=https%3A%2F%2Fblog.csdn.net%2Fyangliuy%2Farticle%2Fdetails%2F8296481>)
34. Γ 函数 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25CE%2593%25E5%2587%25BD%25E6%2595%25B0>)
35. 使用的绘图工具 (<https://links.jianshu.com/go?to=http%3A%2F%2Ffooplot.com%2F>)
36. 求解一元三次方程的工具 (<https://links.jianshu.com/go?to=http%3A%2F%2Fwww.yunsuan.org%2Fapp%2Fs2084>)
37. 你对贝叶斯统计都有怎样的理解? (<https://links.jianshu.com/go?to=https%3A%2F%2Fwww.zhihu.com%2Fquestion%2F21134457>)
38. Bayesian inference (https://links.jianshu.com/go?to=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FBayesian_inference%23Formal_description_of_Bayesian_inference)
39. 概率密度函数 (<https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%25A9%259F%25E7%2>)

58E%2587%25E5%25AF%2586%25E5%25BA%25A6%25E5%2587%25BD%25E6%2595%25B8)

40. 累积分布函数 ([https://links.jianshu.com/go?](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E7%25B4%25AF%25E7%25A7%25AF%25E5%2588%2586%25E5%25B8%2583%25E5%2587%25BD%25E6%2595%25B0)

[to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E7%25B4%25AF%25E7%25A7%25AF%25E5%2588%2586%25E5%25B8%2583%25E5%2587%25BD%25E6%2595%25B0](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E7%25B4%25AF%25E7%25A7%25AF%25E5%2588%2586%25E5%25B8%2583%25E5%2587%25BD%25E6%2595%25B0))

41. 似然函数 ([https://links.jianshu.com/go?](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E4%25BC%25BC%25E7%2584%25B6%25E5%2587%25BD%25E6%2595%25B0)

[to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E4%25BC%25BC%25E7%2584%25B6%25E5%2587%25BD%25E6%2595%25B0](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E4%25BC%25BC%25E7%2584%25B6%25E5%2587%25BD%25E6%2595%25B0))

42. 概率质量函数 ([https://links.jianshu.com/go?](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%25A6%2582%25E7%258E%2587%25E8%25B4%25A8%25E9%2587%258F%25E5%2587%25BD%25E6%2595%25B0)

[to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%25A6%2582%25E7%258E%2587%25E8%25B4%25A8%25E9%2587%258F%25E5%2587%25BD%25E6%2595%25B0](https://links.jianshu.com/go?to=https%3A%2F%2Fzh.wikipedia.org%2Fwiki%2F%25E6%25A6%2582%25E7%258E%2587%25E8%25B4%25A8%25E9%2587%258F%25E5%2587%25BD%25E6%2595%25B0))

43. Introduction to Bayesian Inference ([https://links.jianshu.com/go?](https://links.jianshu.com/go?to=https%3A%2F%2Fwww.datascience.com%2Fblog%2Fintroduction-to-bayesian-inference-learn-data-science-tutorials)

[to=https%3A%2F%2Fwww.datascience.com%2Fblog%2Fintroduction-to-bayesian-inference-learn-data-science-tutorials](https://links.jianshu.com/go?to=https%3A%2F%2Fwww.datascience.com%2Fblog%2Fintroduction-to-bayesian-inference-learn-data-science-tutorials))

(/apps/
utm_sc
banner

小礼物走一走，来简书关注我

赞赏支持

📖 数学之美 (/nb/25590714)

举报文章 © 著作权归作者所有



SnailTyan (/u/7731e83f3a4e)

写了 761532 字，被 1049 人关注，获得了 918 个喜欢

(/u/7731e83f3a4e)

+ 关注

纸上得来终觉浅，绝知此事要躬行。 个人博客: <http://noahsnail.com> CSDN博客: <http://blog.csdn.net/quinn>

喜欢 | 32



更多分享

被以下专题收入，发现更多相似内容



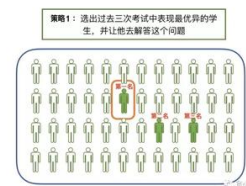
数学之美 (/c/e9654414bd33?utm_source=desktop&utm_medium=notes-included-collection)





看场电影 (/c/d174a9693692?utm_source=desktop&utm_medium=notes-included-collection)

(/p/7b9b979f1bff?)



(/apps/
utm_sc
banner

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommend
最大似然估计、最大后验概率与贝叶斯估计 (/p/7b9b979f1bff?utm_campa...

Intro 众所周知, 机器学习的过程一般分两个要点, 第一是“模型”, 第二是“优化”。“模型”这一点很好理解, 就是要找到适合用于解决手头上预测问题的模型。机器学习训练过程的核心在于“优化”。而训练 (training) ...



uglybachelor (/u/9705bed6f110?)

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommend

NLP面试-最大似然估计与贝叶斯估计的区别 (/p/ead99acd6437?utm_cam...

1 相关理论 1.1 全概率公式 全概率公式为概率论中的重要公式, 它将对一复杂事件A的概率求解问题转化为在不同情况下发生的简单事件的概率的求和问题。如果事件B1、B2、B3...Bn 构成一个完备事件组, 即...



致Great (/u/261e23a40f71?)

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommend

(/p/8d9790c13475?)



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommend
竟有这种操作?贝叶斯方法的简单学习 (/p/8d9790c13475?utm_campaign=...

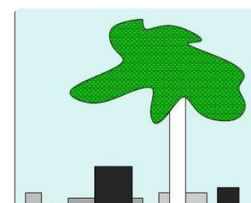
各位小伙伴们大家好,前些日子,我看了一些关于贝叶斯方法的文章,其中以今天这一篇文章觉得最好,不仅讲的简单通俗易懂并且很多的方法都有概括到这是一篇关于贝叶斯方法的科普文,我会尽量少用公式,多用平...



云时之间 (/u/7ebb78d8c38a?)

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommend

(/p/9d3a91cb2117?)



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommend
数学之美番外篇: 平凡而又神奇的贝叶斯方法 (/p/9d3a91cb2117?utm_ca...

本文转自 刘未鹏的博客! 概率论只不过是把常识用数学公式表达了出来。记得读本科的时候, 最喜欢到城里的计算机书店里面去闲逛, 一逛就是好几个小时; 有一次, 在书店看到一本书, 名叫贝叶斯方法。当时...

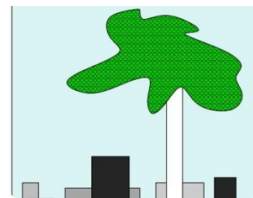


Bioquan (/u/4f87e59ba7e4?)

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommend



(/p/15ec5bc5a3a6?



(/apps/
utm_sc
banner

utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommenc
为何选最小二乘 兼看贝叶斯方法 (/p/15ec5bc5a3a6?utm_campaign=male...

http://mindhacks.cn/2008/09/21/the-magical-bayesian-method/ 数学之美番外篇：平凡而又神奇的贝叶斯方法
概率论只不过是把常识用数学公式表达了出来。——拉普拉斯 目录 0. 前言 1. 历史 1.1 一个例子：自然...



麒麟楚庄王 (/u/0a24b5397da3?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommenc

(/p/1568e1a053cc?



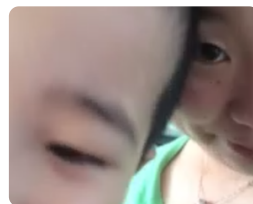
utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommenc
亲们，请不要再投稿这样的照片了，好吗？ 1 (/p/1568e1a053cc?utm_cam...



x九叔x (/u/99e15dd28677?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommenc

(/p/ae931130ed96?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommenc
维铭父母正面管教打卡第18期12号刘葵&姜熳麒93/90 (/p/ae931130ed96?u...

上班日... 今天一户也没有，所以有很多时间做自己的事。下班去舞蹈学校开例会，学期结束了，总结，总结，新学期要求... 晚上回家，和嗖嗖嗖交谈，因为临时的原因，不能带着他去看大恐龙了，请求嗖嗖嗖的...



s_o_s_o_爸爸 (/u/1588cee7f0e7?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommenc

(/p/4f38d689aaf1?



utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommenc
老妈不给我买书钱 (/p/4f38d689aaf1?utm_campaign=maleskine&utm_co...

我有一个老妈。这句话显然是句废话，大伙儿都有老妈。而且大部分老妈一定很爱儿女，我妈也是，她挺爱我的。但是！因为爱而产生的包容也不是没底线的吧？！我计划考研要买参考书复习资料，没有经济来源...




Ayesha23 (/u/256972c9e8bd?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommenc



我养你啊！不用，我养你！ (/p/68944ada083f?utm_campaign=maleskine...

近日成都小甜甜一句“养活我，带我吃饭就够了”让她红遍了全抖音，很多人去成都找她，抖音上全都是用各种交通工具去成都的。这件事其实很让人费解，现在甚至还演变成了男女两大阵营，战斗场面一度十分激烈...

 清雅静兮 (/u/97de884ad659?

utm_campaign=maleskine&utm_content=user&utm_medium=seo_notes&utm_source=recommenc
(/apps/
utm_sc
bannet

