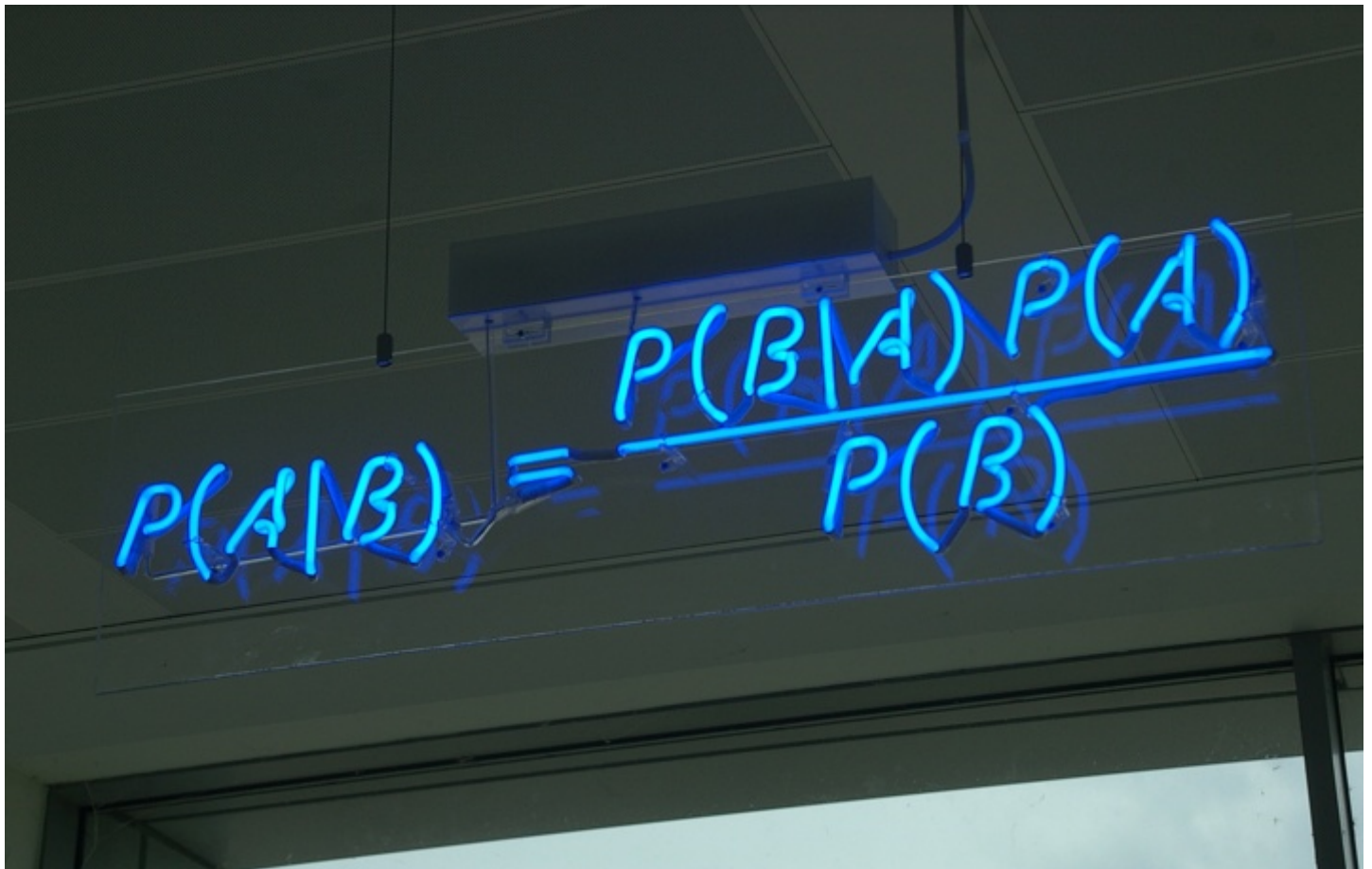


知乎

首发于
川流不息

贝叶斯统计



石川

量化交易 话题的优秀回答者

已关注

84 人赞同了该文章

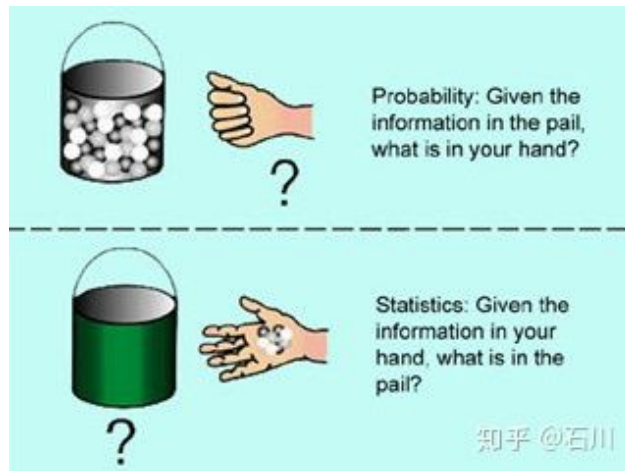
1 概率论和统计学的区别

简单来说，**概率论和统计学解决的问题是互逆的**。假设有一个具有不确定性的过程（process），然后这个过程可以随机的产生不同的结果（outcomes）。则概率论和统计学的区别可以描述为：

- 在**概率论（probability theory）**中，我们已知该过程的概率模型，该模型的不确定性由相应的概率分布来描述；概率论要回答的问题是该过程产生某个结果的可能性有多大这类问题。
- 在**统计学（statistics）**中，该过程的概率模型对我们来说是未知的，但是我们有一系列该过程产生的结果的观测值；我们希望通过这些观测值来推断出这个过程中的不确定性是什么样的。

总结来说就是：**通过已知的概率模型来精确的计算各种结果的可能性就是概率论；根据观测的结果来推断模型的不确定性就是统计学。**

如果上面的描述依然晦涩，请看下面这个例子。假设桶里面有 100 个小球，小球分为白色和黑色。如果已知桶里面一共有 30 个白球和 70 个黑球，想回答随机从桶中摸出一个白球（或者黑球）的概率是多少这样的问题，这就属于概率论的范畴。而如果已知通过有放回的采样抽出了 10



对于概率论来说，每一个问题都有唯一的答案。通过相关计算，总可以计算出我们关心的结果发生的概率。反观统计学，它更像是一门艺术。因为要推断的模型是未知的，因此很难说哪种推断方法就优于另一种方法，或者哪种推断结果就比其他结果更加正确。就拿上面的例子来说，虽然观测到的 10 个球中有 4 个白球和 6 个黑球，但我们仍不能断言桶里白球占 40% 的推断就一定比桶里白球占 50% 或者 30% 的推断更加准确。

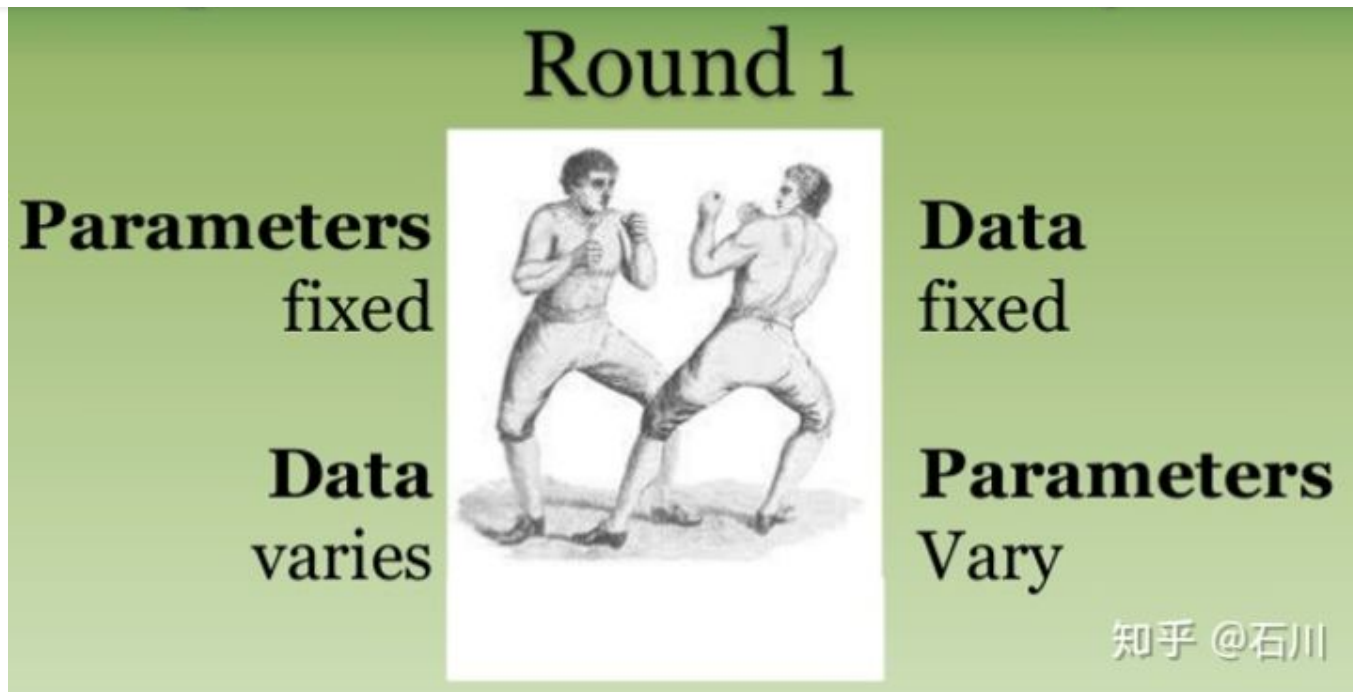
2 古典统计学和贝叶斯统计学

统计学领域中有两大学派：古典统计学（classical）和贝叶斯统计学（Bayesian，以英国数学家托马斯·贝叶斯命名）。古典统计学又称为频率论（frequentist）。

关于这两大学派孰优孰劣已有一个世纪的争论。它们的本质区别在于对待未知模型或者参的方法是不同的：

- 古典统计学认为，未知的模型或者参数是**确定的**，只不过我们**不知道**它确切的形式或者取值。
- 贝叶斯统计学认为，未知的模型或者参数变量是**不确定的**，但是这种不确定性可以由一个**概率分布**来描述。





古典统计学通过进行大量重复实验并统计某个特定结果出现的频率作为对未知参数的估计。以猜桶中白球的比例为例，频率论者会进行大量的带放回的独立抽取实验（实验可以做到天荒地老海枯石烂），然后计算所有结果中白球出现的频率，以此作为对小球中白球比例的推断。古典统计学的核心在于通过大量的实验来消除模型或者参数估计中的不确定性（因为它假设未知模型或者参数是确定的）。

贝叶斯统计学则截然不同。贝叶斯统计学“使用概率的方法来解决统计学问题”。如前所述，贝叶斯统计学认为未知的模型或者参数是不确定的、符合某个概率分布。特别的，我们会首先根据主观判断或者过去的经验，对这个概率分布有一个猜测，称为**先验分布 (prior distribution)**；然后根据越来越多的**观测值 (new data 或者 new evidence)**来修正对该概率分布的猜测，最后得到的概率分布称为**后验分布 (posterior distribution)**。贝叶斯统计学中的“概率”的概念可以被解释为我们对未知变量不同取值的信心程度的测度 (measure of confidence)。贝叶斯统计不消除未知变量的不确定性，而是通过越来越多的新的观测点来持续更新我们对于该未知变量不确定性的认知，提高我们对不确定性的判断的信心。

对于上面这个例子，假设在观测值出现之前，我们猜测桶中有 50% 的白球和 50% 的黑球。因此 50% 是我们对白球比例的先验信仰 (prior belief)。随着不断进行抽取实验，我们会根据得到的观测值更新我们的信仰。假设 10 次抽取后得到 4 个白球和 6 个黑球，那么此时我们对白球比例的信仰就会从最初的 50% 减少一些，这是因为我们结合新的证据（即观测的 10 个球中仅有 40% 是白球）更新了猜测。假设 100 次抽取后得到了 35 个白球和 65 个黑球，那么此时我们对白球比例的信仰又会继续更新。随着越来越多的观测值，我们会持续更新猜测，**并且对该猜测的信心程度也会越来越高，即未知变量（在这里是白球比例）后验分布的标准差会越来越小**（后面会通过一个扔硬币的例子说明）。

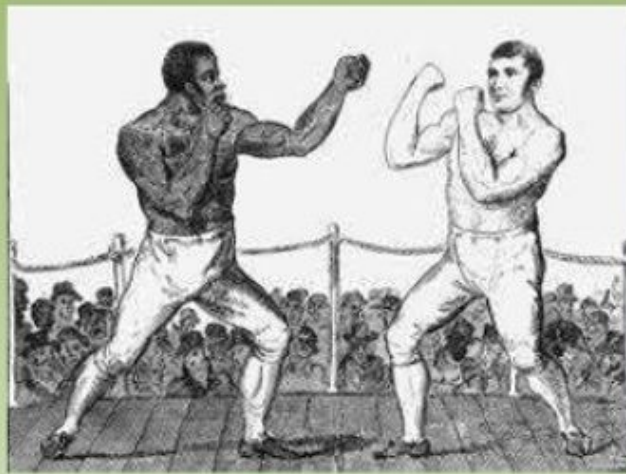


知乎

首发于
川流不息

会同意 50% 是一个合理的先验猜测。但是，极端的人也许会使用 0% 或者 100% 白球作为他的先验猜测。不过，尽管不同人可以有不同的先验分布，但是随着他们结合新的观测点来更新自己的信仰，我们会发现他们最终得到的后验分布是会逐渐收敛的。此外，对很多生活中的实际问题，使用一个合理的猜测 (educated guess) 作为先验是很有好处的。

Frequentist vs. Bayesian Round 2



“The Strength of the Prior” 知乎 @石川

3 为什么要学习贝叶斯统计

贝叶斯统计在生活以及量化投资中有着广泛的应用。从下面两个意义上说，相对古典统计，贝叶斯统计有明显的优势：

1. 虽然在上面对抽小球的例子中我们进行大量重复性的实验并计算白球的频率（古典统计学手段），但对于是在生活中的很多实际问题，大量重复实验是不现实的。比如我们想推断川普当选美国总统的概率。显然，我们没法让美国人进行成千上万次不同的投票选举，然后计算川普获胜的频率。即便是通过民意调查的方式，进行成千上万次也是不切实际的（简单从成本的角度考虑就不可能）。因此，对于这个问题我们只能有非常有限的几次民意调查结果。我们当然可以只通过这些有限的结果利用古典统计学对川普获胜的概率做出估计，但是可以想象的是这个估计的误差会非常大。而贝叶斯统计则提供了新的视角。

2. 合理的先验分布对未知量的估计是非常有益的。对生活中很多实际问题的判断都和人们的学识、经验、见识有关。在这种情况下，如果我们把有限和观测数据和根据知识和经验得到的先验

知乎

首发于
川流不息

产收益率作为先验，将基金经理的主观判断作为观测值，通过把它们两者结合来得到后验判断。它的本质也是贝叶斯统计。

可见，掌握贝叶斯统计并且使用它做推断，即**贝叶斯推断 (Bayesian inference)**，十分重要。贝叶斯统计框架的核心无疑就是**贝叶斯定理 (Bayes' rule)**。

4 贝叶斯定理

本节简要介绍贝叶斯定理，它是贝叶斯推断的核心。（对数学不感兴趣的读者可以跳过本节，这么做不会影响对后文的理解。）

贝叶斯定理的推导始于条件概率。**条件概率可以定义为：在事件 B 发生的前提下，事件 A 发生的概率。数学上用 $P(A|B)$ 来表示这个条件概率。**

生活中条件概率屡见不鲜。比如在没有赶上 8 点这趟地铁，上班迟到的概率是多少？

条件概率 $P(A|B)$ 的数学定义为：

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

这个公式的白话解释为：“当 B 发生前提下 A 发生的概率”等于“**A 和 B 同时发生的概率**”除以“**B 发生的概率**”。用我们的例子来说，那就是“在没有赶上 8 点这趟地铁的前提下，上班迟到的概率”等于“没赶上 8 点这趟地铁且上班迟到的概率”除以“没赶上 8 点这趟地铁的概率”。将这个式子左右两边同时乘以 $P(B)$ 得到 $P(B)P(A|B) = P(A \cap B)$ 。

类似的，我们也可以求出 $P(B|A)$ ，即在 A 发生的前提下，B 发生的概率是多少。在上面例子中，这对应着“在上班迟到的前提下，没有赶上 8 点这趟地铁的概率是多少”？（上班迟到的原因可能很多，比如没赶上这趟地铁是一个，又比如赶上地铁了但是下地铁后去办公楼咖啡馆里耽搁了 10 分钟也是一个，或者因为早上发烧先去医院了等等。）根据定义：

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

同样，两边同时乘以 $P(A)$ （并且由 $P(A \cap B) = P(B \cap A)$ ）得到

$P(A)P(B|A) = P(A \cap B)$ 。由此可知 $P(B)P(A|B) = P(A)P(B|A)$ 。这个结果也可以写作如下形式，即大名鼎鼎的**贝叶斯定理**：





5 贝叶斯推断

由贝叶斯定理可以顺其自然得到贝叶斯推断。前文提到，贝叶斯统计的核心是通过新的观测数据（或者新的证据）来不断的更新我们对未知量的认知。

考虑一个假想的例子。假设我们的先验认知是明天太阳不会升起（即明天太阳不会升起的概率为1）。然而，实际观测到的证据是每天太阳都照常升起。由此，我们会不断的修正之前那个先验，由此得到的后验认知是下一天太阳不会升起的概率越来越低。通过**新证据或者数据来更新认知的过程就是贝叶斯推断**。下面我们来正式的描述它。

假设我们有一个需要估计的未知量 θ ，并且针对该变量有一个先验分布 $P(\theta)$ 。令 D 为一系列观测值或者证据。我们希望通过 D 来修正对 θ 的分布的认知，即 $P(\theta|D)$ 是我们感兴趣的。由贝叶斯定理可得：

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

在贝叶斯推断的框架下，上面公式中的这些概率量都有约定俗成的名字：

- $P(\theta)$ ： θ 的**先验分布 (prior)**。它表示在没有任何观测值序列 D 时我们对于 θ 的不确定性的认知。
- $P(\theta|D)$ ： θ 的**后验分布 (posterior)**。它表示在考虑了观测值序列 D 后，我们对于 θ 的不确定性的改进的认知。
- $P(D|\theta)$ ：**可能性、似然度 (likelihood)**。它是当未知变量服从 θ 的前提下，我们观察到序列 D 的条件概率。
- $P(D)$ ：**观测值或证据 (evidence)**。这是在考虑所有可能的 θ 的分布下，所能观测到序列 D 的非条件概率。

可见，通过使用贝叶斯推断，我们可以合理的将先验认知和实际证据结合在一起，得到一个更新的后验认知。

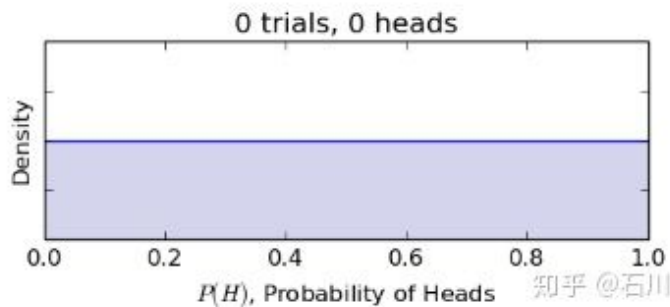
此外，贝叶斯推断框架的强大之处在于我们可以迭代的看问题，即在每次有新观测数据后我们可以得到一个新的后验分布，然后把它作为下个新数据出现前的（新的）先验分布。换句话说，在这个过程中我们通过反复迭代使用贝叶斯定理，持续更新对未知量的分布的认知。

6 一个扔硬币的例子



币便可以得到一个由正面和（或）反面结果组成的观测序列，这就是观测序列 D 。

假设在开始扔硬币前，我们对 θ 的分布 $P(\theta)$ 有如下先验猜想： θ 可以是 0 到 1 范围内的任何取值，符合连续均匀分布（uniform distribution，比如 θ 等于 0 说明该硬币两面都不是头像； θ 等于 1 说明该硬币两面都是头像； θ 等于 0.5 意味着该硬币一面头像一面非头像，且质地均匀等）。在这个假设下， θ 的先验概率密度函数为 0 到 1 之间的一条水平线（下图）。



下面我们就来说说如何通过贝叶斯定理、利用新的实验结果来更新这个先验分布。为此，引入一个非常有用的概念——**共轭先验 (conjugate priors)**。有点晕？别着急往下看。为了解释它，我们先来介绍另一个应用非常广泛的分布——**Beta 分布 (Beta distribution)**。

Beta 分布是一组定义在 0 到 1 区间上的连续概率分布，其具体形态由两个参数 α 和 β 决定，其概率密度函数为：

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

上式中 $B(\alpha, \beta)$ 是一个由 α 和 β 决定的系数，以满足 f 在 0 到 1 上的积分为 1。我们将上述概率密度函数简写为： $\text{Be}(\alpha, \beta)$ 。

前面说过，我们对于 θ 的先验分布猜测是 uniform distribution，而它是一种特殊的 Beta 分布，其对应 Beta 分布的参数为 $\alpha = 1$ 以及 $\beta = 1$ ，因此有 $\theta \sim \text{Be}(1, 1)$ 。

当我们抛掷概率为 θ 的硬币时，得到正面的概率为 θ ，反面的概率为 $1 - \theta$ 。因此，假如我们抛掷 n 次，得到 $m (\leq n)$ 次正面的概率实际上是一个**二项分布 (binomial distribution)**，且满足（以下 D 代表抛掷 n 次中得到 m 次正面这件事）：

$$P(D|\theta) = C \cdot \theta^m (1-\theta)^{n-m}$$

上式中 $C = n!/(m!(n-m)!)$ 是一个系数。

一般的，当先验满足参数为 α 和 β 的 Beta 分布时，由**贝叶斯定理**可知，后验概率满足：

知乎

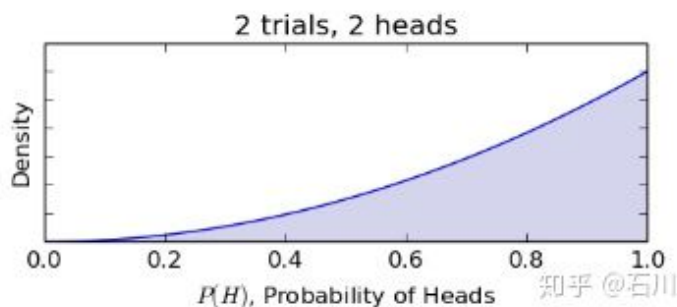
首发于
川流不息

$$\propto \theta^{\alpha+m-1} (1-\theta)^{\beta+n-m-1}$$

可见此时后验满足参数为 $\alpha + m$ 和 $\beta + n - m$ 的 Beta 分布。在贝叶斯统计中，如果先验和后验属于同类分布，则它们称作共轭分布，而先验称作是似然函数（本例中是二项分布）的共轭先验。

好了，数学基础已经打好，现在我们可以扔硬币了。别忘了我们的先验 uniform distribution 恰好是 $\text{Be}(1, 1)$ 。

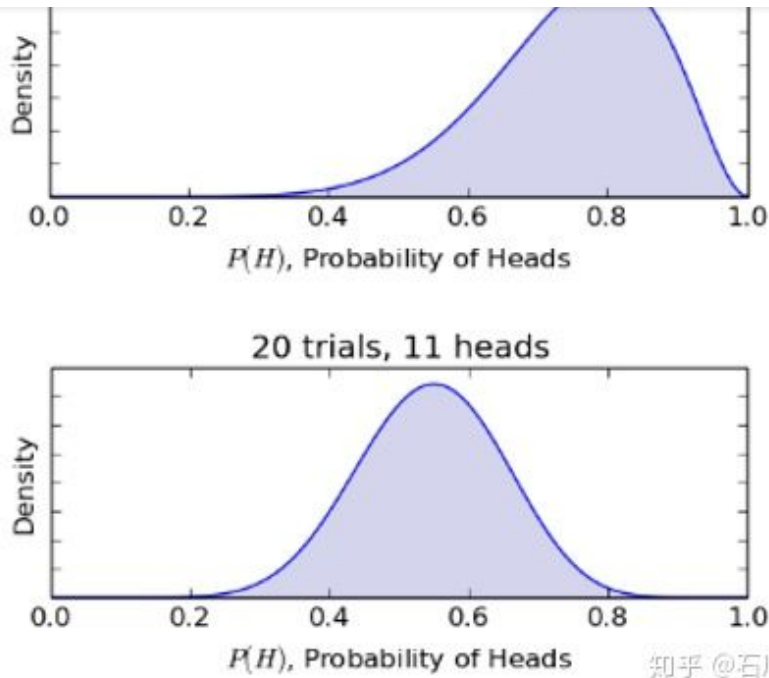
下面我们开始扔硬币。假设扔了两次后，得到了两次头像（ $n = m = 2$ ）。根据贝叶斯推断，我们得到关于 θ 的更新后的后验概率密度函数为 $\text{Be}(3, 1)$ ，如下图所示。可见由于连续看到两次头像面的结果，我们开始倾向于认为 θ 的取值是越接近 1 越有可能。



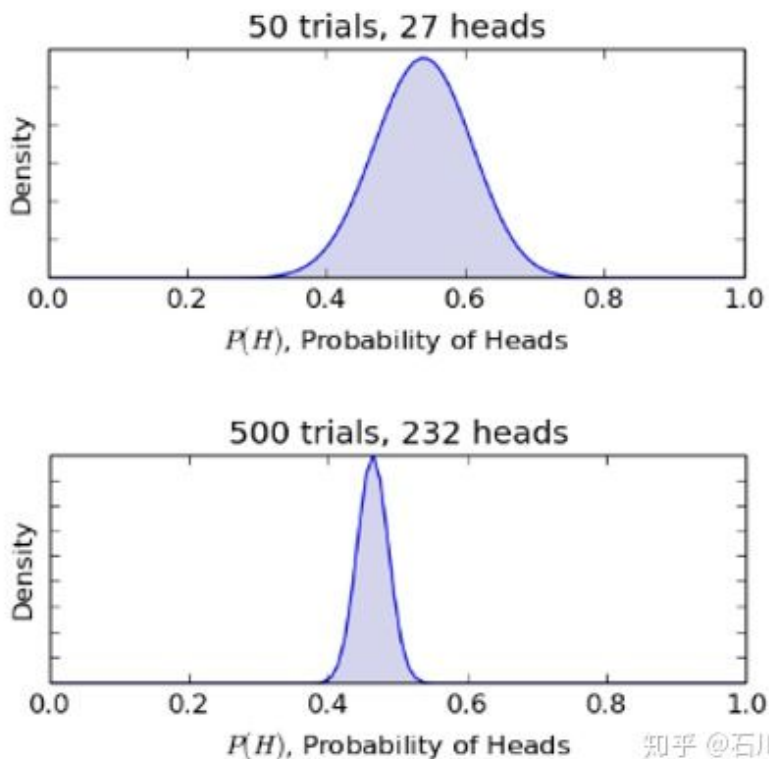
让我们继续实验。假如我们扔了 10 次后得到 8 次正面，而扔了 20 次后得到了 11 次正面。根据这些结果，我们不断更新 θ 的后验分布（下图）。当 10 次中有 8 次正面时，我们会认为这个硬币很有可能是不公平的，即正面和反面出现的概率不同（ $\text{Be}(9, 3)$ ）。而当 20 次中出现 11 次正面时，我们的认知会再次根据新的结果得到修正，我们开始认为这个硬币可能是公平的了（ $\text{Be}(12, 10)$ ）。



知乎

首发于
川流不息

最后，下面两张图是经过了 50 次（27 次正面）和 500 次（232 次正面）实验后的 θ 的后验分布（分别为 $\text{Be}(28, 24)$ 和 $\text{Be}(233, 269)$ ）。



随着越来越多的新结果的出现，我们对于 θ 的不确定性的认知越来越清晰；对于 θ 的不同取值的信心越来越高。特别的，我们越来越有把握的说 θ 最有可能的取值是 0.5 附近。这体现在 500 次实验后， θ 的后验分布 $P(\theta|D)$ 已经非常狭窄（换句话说， θ 的取值的标准差越来越小且集中在 0.46 附近。假如这枚硬币确实是一枚公平的硬币，那么如果再进行 500 此实验，会发现 $P(\theta|D)$ 会更加狭窄且 θ 的取值一定会集中在 0.5 附近。

知乎

首发于
川流不息

用贝叶斯定理，逐步细化、完善我们对 θ 的不确定性的认知，最终得到了关于 θ 的不确定性的非常自信的后验分布（即 θ 的分布以 0.5 为中心，标准差非常小，它最有可能的取值就是 0.5）。

贝叶斯统计是一个强大的工具；不熟悉它的人却对其敬而远之。下面是网上关于贝叶斯统计的一个笑话。它可能代表着很多吃瓜群众对贝叶斯统计的看法，以及贝叶斯统计学派的自嘲：

A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.

译：一个贝叶斯学派的学者是这样的：他模糊的期待着一匹马（先验），然而却看到了一头驴（证据），于是便自信的认为那是一头骡子（后验）。

（全文完）

免责声明：文章内容不可视为投资意见。市场有风险，入市需谨慎。

原创不易，请保护版权。如需转载，请联系获得授权，并注明出处，谢谢。已委托“维权骑士”（维权骑士_领先的原创内容监测、保护及快速授权平台）为进行维权行动。

编辑于 2019-07-02

统计 贝叶斯统计 贝叶斯理论

▲ 赞同 84 ▼ 10 条评论 分享 ★ 收藏 ...

文章被以下专栏收录



川流不息

北京量信投资管理有限公司是一家在中国基金业协会备案登记的专业私募基金管理人...

关注专栏

推荐阅读



10 条评论

切换为时间排序

写下你的评论...



MixViper

1 年前

未来是Bayesian的!

1 赞



SongLe

1 年前

lol strongly believe

2 赞



阿七啊

1 年前

一打开知乎就看见 是为了提醒我，我贝叶斯作业还没做吗

1 赞



偶尔

1 年前

如果能把那个硬币的案例计算也写进去会好些。贝叶斯定理。

赞



石川 (作者) 回复 偶尔

1 年前

感谢建议。我考虑一下，争取做个更新。

赞



石川 (作者) 回复 偶尔

1 年前

Done。再次感谢建议。

赞



唐子明

11 个月前


深刻!


赞





小. 张. 华. 11

10 个月前

 赞

 回复

 踩

 举报



sky new

8 个月前

学习了，感谢！

 赞



小沧海

1 个月前

写的很清晰

 赞

