**<https://www.sohu.com/a/210850749_747155>**

**人工智能大数据背后的神秘公式：贝叶斯公式**

2017-12-16 05:00

大数据、人工智能、海难搜救、生物医学、邮件过滤，这些看起来彼此不相关的领域之间有什么联系？答案是，它们都会用到同一个数学公式——**贝叶斯公式**。它虽然看起来很简单、很不起眼，但却有着深刻的内涵。那么贝叶斯公式是如何从默默无闻到现在广泛应用、无所不能的呢？

一 什么是贝叶斯公式

18世纪英国业余数学家托马斯·贝叶斯（Thomas Bayes，1702～1761）提出过一种看上去似乎显而易见的观点：“用客观的新信息更新我们最初关于某个事物的信念后，我们就会得到一个新的、改进了的信念。” 这个研究成果，因为简单而显得平淡无奇，直到他死后的两年才于1763年由他的朋友理查德·普莱斯帮助发表。它的数学原理很容易理解，简单说就是，如果你看到一个人总是做一些好事，则会推断那个人多半会是一个好人。这就是说，当你不能准确知悉一个事物的本质时，你可以依靠与事物特定本质相关的事件出现的多少去判断其本质属性的概率。用数学语言表达就是：支持某项属性的事件发生得愈多，则该属性成立的可能性就愈大。**与其他统计学方法不同，贝叶斯方法建立在主观判断的基础上，你可以先估计一个值，然后根据客观事实不断修正。**

1774年，法国数学家皮埃尔-西蒙·拉普拉斯(Pierre-Simon Laplace，1749－1827)独立地再次发现了贝叶斯公式。拉普拉斯关心的问题是：当存在着大量数据，但数据又可能有各种各样的错误和遗漏的时候，我们如何才能从中找到真实的规律。拉普拉斯研究了男孩和女孩的生育比例。有人观察到，似乎男孩的出生数量比女孩更高。这一假说到底成立不成立呢？拉普拉斯不断地搜集新增的出生记录，并用之推断原有的概率是否准确。每一个新的记录都减少了不确定性的范围。拉普拉斯给出了我们现在所用的贝叶斯公式的表达：

P(A/B)=P(B/A)\*P(A)/P(B)，

该公式表示在B事件发生的条件下A事件发生的条件概率，等于A事件发生条件下B事件发生的条件概率乘以A事件的概率，再除以B事件发生的概率。公式中，P(A）也叫做先验概率，P(A/B)叫做后验概率。严格地讲，贝叶斯公式至少应被称为“贝叶斯-拉普拉斯公式”。

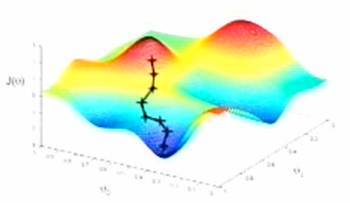
二 默默无闻200年

贝叶斯公式现在已经非常流行，甚至在热门美剧《生活大爆炸》中谢耳朵也秀了一下。但它真正得到重视和广泛应用却是最近二三十年的事，其间被埋没了200多年。这是为什么呢？原因在于我们有另外一种数学工具——经典统计学，或者叫频率主义统计学（我们在学校学的主要是这种统计学），它在200多年的时间里一直表现不错。从理论上讲，它可以揭示一切现象产生的原因，既不需要构建模型，也不需要默认条件，只要进行足够多次的测量，隐藏在数据背后的原因就会自动揭开面纱。

在经典统计学看来，科学是关于客观事实的研究，我们只要反复观察一个可重复的现象，直到积累了足够多的数据，就能从中推断出有意义的规律。而贝叶斯方法却要求科学家像算命先生一样，从主观猜测出发，这显然不符合科学精神。就连拉普拉斯后来也放弃了贝叶斯方法这一思路，转向经典统计学。因为他发现，如果数据量足够大，人们完全可以通过直接研究这些样本来推断总体的规律。

打个比方来帮助我们理解这两种统计学方法的区别。假如我们想知道某个区域里海拔最低的地方，经典统计学的方法是首先进行观测，取得区域内不同地方的海拔数据，然后从中找出最低点。这个数据量必须足够多，以反映区域内地形全貌的特征，这样我们才能相信找到的就是实际上的最低点。而贝叶斯方法是我不管哪里最低，就凭感觉在区域内随便选个地方开始走，每一步都往下走，虽然中间可能有一些曲折，但相信这样走早晚能够到达最低点。**可以看出，贝叶斯方法的关键问题是这个最终到达的低点可能不是真正的最低点，而是某个相对低点，它可能对该区域的地形（碗型、马鞍形等）和最初我们主观选择的出发点有依赖性。**如果问题域是碗型的，我们到达的就是最低点；但如果是马鞍形或者其他复杂曲面，那么我们到达的可能是多个相对低点（极点）中的一个，而不是真正的最低点。这是贝叶斯方法最受经典统计学方法诟病的原因，也是它在过去的200多年被雪藏的原因所在。

贝叶斯方法原理示意图：



三 初显威力

长期以来，贝叶斯方法虽然没有得到主流学界的认可，但其实我们经常会不自觉地应用它来进行决策，而且还非常有效。比如炮兵在射击时会使用贝叶斯方法进行瞄准。炮弹与子弹不同，它的飞行轨迹是抛物线，瞄准的难度更大，因此他们会先根据计算和经验把炮管调整到一个可能命中的瞄准角度（先验概率），然后再根据炮弹的实际落点进行调整（后验概率），这样在经过2-3次射击和调整后炮弹就能够命中目标了。

在日常生活中，我们也常使用贝叶斯方法进行决策。比如在一个陌生的地方找餐馆吃饭，因为之前不了解哪家餐馆好，似乎只能随机选择，但实际上并非如此，我们会根据贝叶斯方法，利用以往积累的经验来提供判断的线索。经验告诉我们，通常那些坐满了客人的餐馆的食物要更美味些，而那些客人寥寥的餐馆，食物可能不怎么样而且可能会被宰。这样，我们就往往通过观察餐厅的上座率来选择餐馆就餐。这就是我们根据先验知识进行的主观判断。在吃过以后我们对这个餐馆有了更多实际的了解，以后再选择时就更加容易了。**所以说，在我们认识事物不全面的情况下，贝叶斯方法是一种很好的利用经验帮助作出更合理判断的方法。**

而两个标志性的事件在让学术界开始重视贝叶斯方法上起到了重要作用。

1

联邦党人文集作者公案

1787年5月，美国各州（当时为13个）代表在费城召开制宪会议；1787年9月，美国的宪法草案被分发到各州进行讨论。一批反对派以“反联邦主义者”为笔名，发表了大量文章对该草案提出批评。宪法起草人之一亚历山大·汉密尔顿着急了，他找到曾任外交国务秘书（即后来的国务卿）的约翰·杰伊，以及纽约市国会议员麦迪逊，一同以普布利乌斯（Publius）的笔名发表文章，向公众解释为什么美国需要一部宪法。他们走笔如飞，通常在一周之内就会发表3-4篇新的评论。1788年，他们所写的85篇文章结集出版，这就是美国历史上著名的《联邦党人文集》。

《联邦党人文集》出版的时候，汉密尔顿坚持匿名发表，于是，这些文章到底出自谁人之手，成了一桩公案。1810年，汉密尔顿接受了一个政敌的决斗挑战，但出于基督徒的宗教信仰，他决意不向对方开枪。在决斗之前数日，汉密尔顿自知时日不多，他列出了一份《联邦党人文集》的作者名单。1818年，麦迪逊又提出了另一份作者名单。这两份名单并不一致。在85篇文章中，有73篇文章的作者身份较为明确，其余12篇存在争议。

1955年，哈佛大学统计学教授Fredrick Mosteller找到芝加哥大学的年轻统计学家David Wallance，建议他跟自己一起做一个小课题，他想用统计学的方法，鉴定出《联邦党人文集》的作者身份。

但这根本就不是一个小课题。汉密尔顿和麦迪逊都是文章高手，他们的文风非常接近。从已经确定作者身份的那部分文本来看，汉密尔顿写了9.4万字，麦迪逊写了11.4万字。汉密尔顿每个句子的平均长度是34.55字，而麦迪逊是34.59字。就写作风格而论，汉密尔顿和麦迪逊简直就是一对双胞胎。汉密尔顿和麦迪逊写这些文章，用了大约一年的时间，而Mosteller和Wallance甄别出作者的身份花了10多年的时间。

如何分辨两人写作风格的细微差别，并据此判断每篇文章的作者就是问题的关键。他们所采用的方法就是以贝叶斯公式为核心的包含两个类别的分类算法。先挑选一些能够反映作者写作风格的词汇，在已经确定了作者的文本中，对这些特征词汇的出现频率进行统计，然后再统计这些词汇在那些不确定作者的文本中的出现频率，从而根据词频的差别推断其作者归属。这其实和我们现在使用的垃圾邮件过滤器的原理是一样的。

他们是在没有计算机帮助的条件下用手工处理“大数据”，这一工程的耗时耗力是可想而知的。将近100个哈佛大学的学生帮助他们处理数据。学生们用最原始的方式，用打字机把《联邦党人文集》的文本打出来，然后把每个单词剪下来，按照字母表的顺序，把这些单词分门别类地汇集在一起。有个学生干得累了，伸了个懒腰，长长地呼了一口气。他这一口气用力太猛，一下子把刚刚归置好的单词条吹得如柳絮纷飞，一屋子学生瞬间石化，估计很多人连灭了他的心都有。而这只是手工大数据时代的日常。

Mosteller和Wallance这是要在干草垛里找绣花针。他们首先剔除掉用不上的词汇。比如，《联邦党人文集》里经常谈到“战争”、“立法权”、“行政权”等，这些词汇是因主题而出现，并不反映不同作者的写作风格。只有像“in”，“an”，“of”，“upon”这些介词、连词等才能显示出作者风格的微妙差异。一位历史学家好心地告诉他们，有一篇1916年的论文提到，汉密尔顿总是用“while”，而麦迪逊则总是用“whilst”。但仅仅有这一个线索是不够的。“while”和“whilst”在这12篇作者身份待定的文章里出现的次数不够多。况且，汉密尔顿和麦迪逊有时候会合写一篇文章，也保不齐他们会互相改文章，要是汉密尔顿把麦迪逊的“whilst”都改成了“while”呢？

当学生们把每个单词的小纸条归类、粘好之后，他们发现，汉密尔顿的文章里平均每一页纸会出现两次“upon”，而麦迪逊几乎一次也不用。汉密尔顿更喜欢用“enough”，麦迪逊则很少用。其它一些有用的词汇包括：“there”、“on”等等。1964年，Mosteller和Wallance发表了他们的研究成果。他们的结论是，这12篇文章的作者很可能都是麦迪逊。他们最拿不准的是第55篇，麦迪逊是作者的概率是240：1。

这个研究引起了极大的轰动，但最受震撼的不是宪法研究者，而是统计学家。Mosteller和Wallance的研究，把贝叶斯公式这个被统计学界禁锢了200年的幽灵从瓶子中释放了出来。

2

天蝎号核潜艇搜救

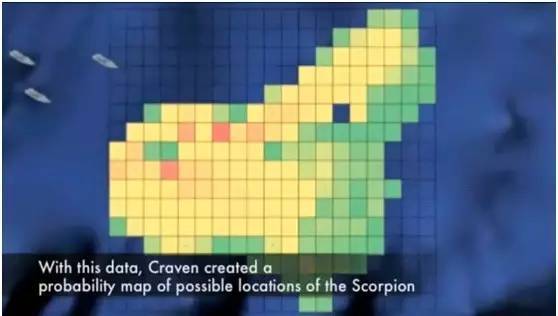
2014年初马航MH370航班失联，所有人都密切关注搜救的进展情况。那么我们是用什么方法在茫茫大海中寻找失联的飞机或者船只的呢？这要从天蝎号核潜艇说起。

1968年5月，美国海军的天蝎号核潜艇在大西洋亚速海海域突然失踪，潜艇和艇上的99名海军官兵全部杳无音信。按照事后调查报告的说法，罪魁祸首是这艘潜艇上的一枚奇怪的鱼雷，发射出去后竟然敌我不分，扭头射向自己，让潜艇中弹爆炸。

为了寻找天蝎号的位置，美国政府从国内调集了包括多位专家的搜索部队前往现场，其中包括一位名叫John Craven的数学家，他的头衔是“美国海军特别计划部首席科学家”。在搜寻潜艇的问题上，Craven提出的方案使用了上面提到的贝叶斯公式。他召集了数学家、潜艇专家、海事搜救等各个领域的专家。每个专家都有自己擅长的领域，但并非通才，没有专家能准确估计到在出事前后潜艇到底发生了什么。有趣的是，Craven并不是按照惯常的思路要求团队成员互相协商寻求一个共识，而是让各位专家编写了各种可能的“剧本”，让他们按照自己的知识和经验对于情况会向哪一个方向发展进行猜测，并评估每种情境出现的可能性。据说，为了给枯燥的工作增加一些趣味，Craven还准备了威士忌酒作为“投注”正确的奖品。

因为在Craven的方案中，结果很多是这些专家以猜测、投票甚至可以说赌博的形式得到的，不可能保证所有结果的准确性，他的这一做法受到了很多同行的质疑。可是因为搜索潜艇的任务紧迫，没有时间进行精确的实验、建立完整可靠的理论，Craven的办法不失为一个可行的办法。

由于失事时潜艇航行的速度快慢、行驶方向、爆炸冲击力的大小、爆炸时潜艇方向舵的指向都是未知量，即使知道潜艇在哪里爆炸，也很难确定潜艇残骸最后被海水冲到哪里。Craven粗略估计了一下，半径20英里的圆圈内的数千英尺深的海底，都是天蝎号核潜艇可能沉睡的地方，要在这么大的范围，这么深的海底找到潜艇几乎成了不可能完成的任务。



Craven把各位专家的意见综合到一起，得到了一张20英里海域的概率图。整个海域被划分成了很多个小格子，每个小格子有两个概率值p和q，p是潜艇躺在这个格子里的概率，q是如果潜艇在这个格子里，它被搜索到的概率。按照经验，第二个概率值主要跟海域的水深有关，在深海区域搜索失事潜艇的“漏网”可能性会更大。如果一个格子被搜索后，没有发现潜艇的踪迹，那么按照贝叶斯公式，这个格子潜艇存在的概率就会降低：

http://5b0988e595225.cdn.sohucs.com/images/20171216/3bbad5df2b91497c889ef765077d9565.png

由于所有格子概率的总和是1，这时其他格子潜艇存在的概率值就会上升：

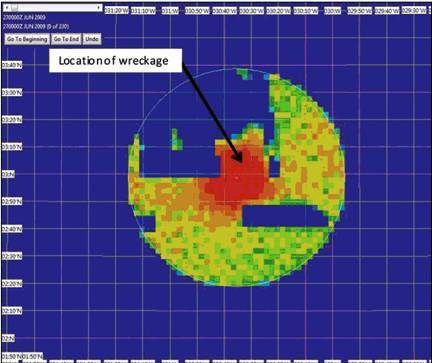
http://5b0988e595225.cdn.sohucs.com/images/20171216/e9755e91d1dc44eeafbcd4b8dadf158f.png

每次寻找时，先挑选整个区域内潜艇存在概率值最高的一个格子进行搜索，如果没有发现，概率分布图会被“洗牌”一次，搜寻船只就会驶向新的“最可疑格子”进行搜索，这样一直下去，直到找到天蝎号为止。

最初开始搜救时，海军人员对Craven和其团队的建议嗤之以鼻，他们凭经验估计潜艇是在爆炸点的东侧海底。但几个月的搜索一无所获，他们才不得不听从了Craven的建议，按照概率图在爆炸点的西侧寻找。经过几次搜索，潜艇果然在爆炸点西南方的海底被找到了。

由于这种基于贝叶斯公式的方法在后来多次搜救实践中被成功应用，现在已经成为海难空难搜救的通行做法。

2009年法航空难搜救的后验概率分布图：



2014马航MH370失联搜索区域：



与计算机的结合使贝叶斯公式巨大的实用价值进一步体现出来，它不但为我们提供了一条全新的问题解决路径，带来工具和理念的革命，而且甚至可能是人类大脑本身的认知和构建方式。

贝叶斯公式在联邦党人文集作者公案和天蝎号核潜艇搜救中大显身手后，开始引起学术界的注意和重视，而其上世纪八十年代在自然语言处理领域的成功，向我们展示了一条全新的问题解决路径。计算能力的不断提高和大数据的出现使它的威力日益显现，一场轰轰烈烈的“贝叶斯革命”正在发生。

四 真正的突破

自然语言处理就是让计算机代替人来翻译语言、识别语音、认识文字和进行海量文献的自动检索。但是人类的语言可以说是信息里最复杂最动态的一部分。人们最初想到的方法是语言学方法，让计算机学习人类的语法、分析语句等等。尤其是在乔姆斯基（有史以来最伟大的语言学家）提出 “形式语言” 以后，人们更坚定了利用语法规则的办法进行文字处理的信念。遗憾的是，几十年过去了，在计算机语言处理领域，基于这个语法规则的方法几乎毫无突破。

其实早在几十年前，数学家兼信息论的祖师爷香农 (Claude Shannon)就提出了用数学方法处理自然语言的想法。遗憾的是当时的计算机根本无法满足大量信息处理的需要，所以他的这一想法并没有引起重视。请加微信公众号：工业智能化(robotinfo) 马云都在关注

率先成功利用数学方法解决自然语言处理问题的是语音和语言处理大师贾里尼克 (Fred Jelinek)。**他引入一个全新的视角，认为语音识别就是根据接收到的一个信号序列推测说话人实际发出的信号序列（说的话）和要表达的意思。这就把语音识别问题转化为一个通信问题，而且进一步可以简化为用贝叶斯公式处理的数学问题。**

一般情况下，一个句子中的每个字符都跟它前面的所有字符相关，这样公式中的条件概率计算就非常复杂，难以实现。为了简化问题，他做了两个假设：

**1、 说话人说的句子是一个马尔科夫链，也就是说，句子中的每个字符都只由它前一个字符决定；**

**2、 独立输入假设，就是每个接受的字符信号只由对应的发送字符决定。**

这样的简化看起来有点简单粗暴，每个字符在语义上都是和文章的其他部分相关的，怎么可能只跟它前一个字符相关呢？很多人不相信用这么简单的数学模型能解决复杂的语音识别、机器翻译等问题。其实不光是一般人，就连很多语言学家都曾质疑过这种方法的有效性。但事实证明，这个基于贝叶斯公式的统计语言模型比任何当时已知的借助某种规则的解决方法都有效。贾里尼克和贝克夫妇在七十年代分别独立提出用这个模型进行语音识别，八十年代微软公司用这个模型成功开发出第一个大词汇量连续语音识别系统。现在我们手机上的语音识别和语音输入功能都已经非常成熟而且好用了。

更加可贵的是，这种语音识别系统不但能够识别静态的词库，而且对词汇的动态变化具有很好的适应性，即使是新出现的词汇，只要这个词已经被大家高频使用，用于训练的数据量足够多，系统就能正确地识别。**这反映出贝叶斯公式对现实变化的高度敏感，对增量信息有非常好的适应能力。**

自然语言处理方面的成功开辟了一条全新的问题解决路径：

**1、原来看起来非常复杂的问题可以用贝叶斯公式转化为简单的数学问题；**

**2、可以把贝叶斯公式和马尔科夫链结合以简化问题，使计算机能够方便求解；虽然我们不完全了解为什么这种看似粗暴的简化并不影响我们的研究过程，但从实践看来它非常有效；**

**3、将大量观测数据输入模型进行迭代——也就是对模型进行训练，我们就可以得到希望的结果。**

**随着计算能力的不断提高、大数据技术的发展，原来手工条件下看起来不可思议的进行模型训练的巨大工作量变得很容易实现，它们使贝叶斯公式巨大的实用价值体现出来。**

五 经典统计学的困难和贝叶斯革命

1

经典统计学的困难

当贝叶斯方法在实际应用中不断证明自己的同时，经典统计学却遇到了困境。经典统计学比较适合于解决小型的问题，同时该方法要求我们获得足够多的样本数据，而且要求这些样本能够代表数据的整体特征。在处理涉及几个参数的问题时，它可以得心应手。但如果相对于问题的复杂程度，我们只掌握少量的信息时，经典统计学就显得力不从心了，原因就是数据的稀疏性问题。

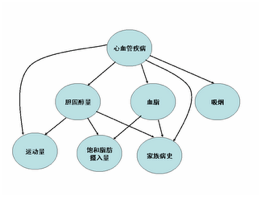
**都大数据时代了，还存在数据稀疏性问题吗？答案是肯定的。**具体来说，一个取决于n个参数，并且每个参数只有两种表现（0或者1）的系统，共有2的n次方种现象。如果某类癌症的产生过程中有100个基因参与（这其实很保守了，人类总共有几万个基因），那么它有2的100次方种可能的基因图谱；根据采样定理进行估算，采用经典统计学方法至少需要获得1%-10%的样本才能确定其病因，也就是需要制作出数万亿亿亿个患有该疾病的病人的基因图谱！这不具备可操作性。所以用经典统计学方法无法解释由相互联系、错综复杂的原因（相关参数）所导致的现象。

2

贝叶斯网络带来工具革命

而目前的情况是，相对简单的问题已经解决得差不多了，剩下的都非常复杂。龙卷风的形成、星系的起源、致病基因、大脑的运作机制等，要揭示隐藏在这些问题背后的规律，就必须理解它们的成因网络，把错综复杂的事件梳理清楚。由于经典统计学失效，科学家别无选择，他们必须从众多可能奏效的法则中选择一些可以信任的，并以此为基础建立理论模型。为了能做出这样的选择，为了能在众多可能性中确定他们认为最为匹配的，过去，科学家多少是依靠直觉来弥补数据上的缺失和空白。而贝叶斯公式正好以严谨的数学形式帮他们实现了这一点。科学家把所有假设与已有知识、观测数据一起代入贝叶斯公式，就能得到明确的概率值。而要破译某种现象的成因网络，只需将公式本身也结成网络，即贝叶斯网络，它是贝叶斯公式和图论结合的产物。

网络化想法的提出也不是一帆风顺的。直到上世纪80年代，美国数学家朱迪亚·珀尔才证明，使用贝叶斯网络应该可以揭示复杂现象背后的成因。操作原理是这样的：如果我们不清楚一个现象的成因，首先根据我们认为最有可能的原因来建立一个模型；然后把每个可能的原因作为网络中的节点连接起来，根据已有的知识、我们的预判或者专家意见给每个连接分配一个概率值。接下来只需要向这个模型代入观测数据，通过网络节点间的贝叶斯公式重新计算出概率值。为每个新数据、每个连接重复这种计算，直到形成一个网络图，任意两个原因之间的连接都得到精确的概率值为止，就大功告成了。即使实验数据存在空白或者充斥噪声和干扰信息，不懈追寻各种现象发生原因的贝叶斯网络依然能够构建出各种复杂现象的模型。贝叶斯公式的价值在于，当观测数据不充分时，它可以将专家意见和原始数据进行综合，以弥补测量中的不足。我们的认知缺陷越大，贝叶斯公式的价值就越大。



**心血管疾病成因的贝叶斯网络**

和前面提到的马尔可夫链类似，我们可以假设贝叶斯网络中每个节点的状态值取决于其前面的有限个状态。不同的是，贝叶斯网络比马尔可夫链灵活，它不受马尔可夫链的链状结构的约束，因此可以更准确地描述事件之间的相关性。可以说，马尔可夫链是贝叶斯网络的特例，而贝叶斯网络是马尔可夫链的推广，**它给复杂问题提供了一个普适性的解决框架**。

为了确定各个节点之间的相关性，需要用已知数据对贝叶斯网络进行迭代和训练。由于网络结构比较复杂，理论上，用现有的计算机是不可计算的（基于冯·诺依曼结构的计算机无法解决这种NP复杂度的问题，NP（Non-deterministic Polynomial）指用非确定机在多项式时间内可以解决的问题类）。**但对于一些具体的应用，可以根据实际情况对网络结构（采用网络拓扑的图同构技术）和训练过程进行简化，使它在计算上可行。**如果量子计算机开发成功，将能够完全解决其计算问题。这样，贝叶斯公式为科学家开辟的新路就完全打通了。

今天一场轰轰烈烈的“贝叶斯革命”正在发生：生物学家用贝叶斯公式研究基因的致病机制；基金经理用贝叶斯公式找到投资策略；互联网公司用贝叶斯公式改进搜索功能，帮助用户过滤垃圾邮件；大数据、人工智能和自然语言处理中都大量用到贝叶斯公式。既然在手工时代，我们无法预测到今天贝叶斯公式与计算机结合的威力，那么我们怎么能忽视贝叶斯网络与量子计算机结合可能蕴藏的巨大潜力呢？

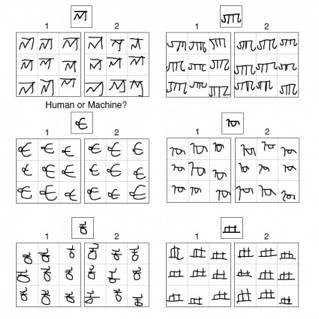
3

人类大脑的构建方式？

贝叶斯公式不仅在自然科学领域掀起革命，它的应用范围也延伸到了关于人类行为和人类大脑活动的研究领域。教育学家突然意识到，学生的学习过程其实就是贝叶斯公式的运用；心理学家证明贝叶斯方法是儿童运用的唯一思考方法，其他方法他们似乎完全不会。进一步，**心理学研究的成果使科学家思考人类的大脑结构是否就是一个贝叶斯网络。这个公式不仅是研究人类思维的工具，它可能就是大脑本身的构建方式。**这个观点十分大胆，但获得越来越广泛的认可。因为贝叶斯公式是我们在没有充分或准确信息时最优的推理结构，为了提高生存效率，进化会向这个模式演进。**贝叶斯公式突然渗透到一切科学领域，提供了通用的研究框架，这是十分罕见的事情。**

人工智能近年来取得了长足的进步，但目前的人工智能通常需要从大量的数据中进行学习，而人类具有“仅从少量案例就形成概念”的能力，两者之间存在巨大差距。比如，尽管你这辈子只见过一个菠萝，但你一眼就能看出菠萝的特征，很快就能从一堆水果中认出菠萝来，甚至还能在纸上画出菠萝的简笔画，而目前的人工智能算法得看成千上万张菠萝的图片才能做到。

不过，这种情况或许已经开始改变了。2015年底，一篇人工智能论文登上了《 科学 》杂志的封面，为人们带来了人工智能领域的一个重大突破： 三名分别来自麻省理工学院、纽约大学和多伦多大学的研究者开发了一个“只看一眼就会写字”的计算机系统。只需向这个系统展示一个来自陌生文字系统的字符，它就能很快学到精髓，像人一样写出来，甚至还能写出其他类似的文字——更有甚者，它还通过了图灵测试，我们很难区分下图中的字符是人类还是机器的作品。这个系统采用的方法就是贝叶斯程序学习（Bayesian Program Learning）——一种基于贝叶斯公式的方法。这不但是人工智能领域的重大突破，而且为我们认识人脑的学习机制提供了重要参考。



**人和机器作品对比图（图片来源：science）**

六 理念的革命

这不仅仅是一场科学的革命，同样也是一场理念的革命。当科学不断强调其对世界认识的客观性时，贝叶斯公式却融入了主观性因素：它并不向我们表述世界，而是表述我们所掌握的知识和经验。这些带有观察者个人因素的知识是脱离研究现象本身的；而它在向我们描述外部现实世界的同时，也描述了观察者对现实的认知的缺陷。更重要的，它迫使我们认识到，**科学理论和科学模型反映的是现实的心理意象，而不是现实本身。而现实为我们提供数据，以保证对现实的意象不会离现实本身太远。在寻找各种现象原因的同时，它也在规范着我们的思想。**

七 贝叶斯公式这么牛，与我何干？

我们经常需要在信息不充分或者不准确的情况下进行判断和决策，一条街上哪个饭馆最靠谱？在自习室惊鸿一瞥的女神有没有男朋友？老公的公文包里发现一只口红，他有没有出轨？新开发的App应该等做得尽善尽美再发布，还是应该尽早发布，用互联网的力量帮助它完善？我应该选择哪个工作offer或者还是考公务员才能使自己的收益最大化？

贝叶斯公式为我们提供了一些决策原则：

* 平时注意观察和思考，建立自己的思维框架，这样在面临选择时就容易形成一个接近实际情况的先验概率，这样经过少量的试错和纠错的迭代循环就可能得到理想的结果；在经过很多次选择和实践的历练后就能够形成自己的直觉，在面对陌生情况时，根据自己的经验和少量信息就能够快速地做出比较准确的判断。
* 大数据时代获得信息的成本越来越低，社会也变得更加开放和包容，初始状态（先验概率）的重要性下降了，即使最初选择不理想，只要根据新情况不断进行调整，仍然可以取得成功。所以如果当下觉得很难做出选择，那就倾听内心的声音，让直觉来选择，这有利于治疗选择恐惧症。

以开发App的例子来说，先按照自己的想法弄个可用的原型出来，然后充分利用互联网的力量，让活跃的用户社区帮助它快速迭代，逐渐使它的功能和体验越来越好。

* 对新鲜事物保持开放的心态，愿意根据新信息对自己的策略和行为进行调整。

“大胆假设，小心求证”，“不断试错，快速迭代”，这些都可以看成贝叶斯公式的不同表述。英国哲学家以赛亚·伯林（Isaish Berlin）曾经援引古希腊诗人的断简残片“狐狸多知而刺猬有一大知”，将人的策略分为狐狸和刺猬两类。刺猬用一个宏大的概念解释所有现象，而狐狸知道很多事情，用多元化的视角看待问题，它也愿意包容新的证据以使得自己的模型与之相适应。在这个快速变化的时代，固守一个不变的信条的刺猬很难适应环境的变化，而使用贝叶斯公式的灵活的狐狸才更容易生存。

参考文献

[1] 《新发现》杂志2013年2月：解密世界的方程式

[2] 吴军：《数学之美》

[3] 何帆：《先放一把火》

[4] 科学松鼠会：死理性派是怎么判断漂亮女孩是不是单身的？

[5] 统计之都创作小组：失联搜救中的统计数据分析

[6] 机器之心：《科学》封面重磅论文：人工智能终于能像人类一样学习

**Bayes**

如果说在数学中有什么方法即让人怀疑又让人迷恋，那么贝叶斯方法一定是一个典型。这个曾经被视为“民科”的方法竟然在今天的多个领域大行其道，不禁让人感慨它的神奇之处。一种方法竟然能写成一本书，必然有其牛逼之处。

我们先来读一句英语：

The girl saw the boy with a telescope.

这句话你怎么翻译呢？平常人肯定会说：那个女孩拿望远镜看见了那个男孩（即你对这个句子背后的实际语法结构的猜测是：The girl saw-with-a-telescope the boy ）。然而，仔细一想，你会发现这个句子完全可以解释成：那个女孩看见了那个拿着望远镜的男孩（即：The girl saw the-boy-with-a-telescope ）。那为什么平常生活中我们每个人都能够迅速地对这种二义性进行消解呢？这背后到底隐藏着什么样的思维法则？我们留到后面解释。

**贝叶斯思想的产生**

长久以来，人们对一件事情发生或不发生的概率，只有固定的0和1，即要么发生，要么不发生，从来不会去考虑某件事情发生的概率有多大，不发生的概率又是多大。而且概率虽然未知，但最起码是一个确定的值。

比如如果问那时的人们一个问题：“有一个袋子，里面装着若干个白球和黑球，请问从袋子中取得白球的概率是多少？”他们会想都不用想，会立马告诉你，取出白球的概率就是1/2，要么取到白球，要么取不到白球，即θ只能有一个值，而且不论你取了多少次，取得白球的概率θ始终都是1/2，即不随观察结果X的变化而变化。

这种频率派的观点长期统治着人们的观念，直到后来一个名叫Thomas Bayes的人物出现。



**Thomas Bayes**

贝叶斯(约1701-1761) Thomas Bayes，英国数学家。约1701年出生于伦敦，做过神甫。1742年成为英国皇家学会 会员。1761年4月7日逝世。贝叶斯在数学方面主要研究概率论。他首先将归纳推理 法用于概率论基础理论，并创立了贝叶斯统计 理论，对于统计决策函数、统计推断、统计的估算等做出了贡献。

贝叶斯所采用的许多术语被沿用至今。贝叶斯思想和方法对概率统计的发展产生了深远的影响。今天，贝叶斯思想和方法在许多领域都获得了广泛的应用。

贝叶斯在世时，并不为当时的人们所熟知，很少发表论文或出版著作，与当时学术界的人沟通交流也很少，用现在的话来说，贝叶斯就是活生生一民间学术“屌丝”，可这个“屌丝”最终发表了一篇名为“An essay towards solving a problem in the doctrine of chances”，翻译过来则是：机遇理论中一个问题的解。你可能觉得这篇论文的发表随机产生轰动效应，从而奠定贝叶斯在学术史上的地位。

事实上，上篇论文发表后，在当时并未产生多少影响，在20世纪后，这篇论文才逐渐被人们所重视。这种际遇，与梵高何其类似，画的画生前一文不值，死后价值连城，令人唏嘘。

“

频率派与贝叶斯派各自不同的思考方式：

频率派把需要推断的参数θ看做是固定的未知常数，即概率虽然是未知的，但最起码是确定的一个值，同时，样本X 是随机的，所以频率派重点研究样本空间，大部分的概率计算都是针对样本X 的分布；

而贝叶斯派的观点则截然相反，他们认为参数是随机变量，而样本X 是固定的，由于样本是固定的，所以他们重点研究的是参数的分布。

据此贝叶斯提出一种看上去似乎显而易见的观点：“用客观的新信息更新我们最初关于某个事物的信念后，我们就会得到一个新的、改进了的信念。”

这个研究成果，因为简单而显得平淡无奇，直到他死后的两年才于1763年由他的朋友理查德·普莱斯帮助发表。它的数学原理很容易理解，简单说就是，如果你看到一个人总是做一些好事，则会推断那个人多半会是一个好人。这就是说，当你不能准确知悉一个事物的本质时，你可以依靠与事物特定本质相关的事件出现的多少去判断其本质属性的概率。

用数学语言表达就是：支持某项属性的事件发生得愈多，则该属性成立的可能性就愈大。 与其他统计学方法不同，贝叶斯方法建立在主观判断的基础上，你可以先估计一个值，然后根据客观事实不断修正。

**拉普拉斯的助攻**

1774年，法国数学家拉普拉斯独立地再次发现了贝叶斯公式。拉普拉斯关心的问题是：当存在着大量数据，但数据又可能有各种各样的错误和遗漏的时候，我们如何才能从中找到真实的规律。

拉普拉斯研究了男孩和女孩的生育比例。有人观察到，似乎男孩的出生数量比女孩更高。这一假说到底成立不成立呢？拉普拉斯不断地搜集新增的出生记录，并用之推断原有的概率是否准确。每一个新的记录都减少了不确定性的范围。拉普拉斯给出了我们现在所用的贝叶斯公式的表达：

**P(A/B)=P(B/A)\*P(A)/P(B)**

**由条件概率到贝叶斯公式**

条件概率（又称后验概率）就是事件A在另外一个事件B已经发生条件下的发生概率。条件概率表示为P(A|B)，读作“在B条件下A的概率”。

在贝叶斯之前，人们已经能够计算“正向概率”，如“假设袋子里面有N个白球，M个黑球，你伸手进去摸一把，摸出黑球的概率是多大”。而一个自然而然的问题是反过来：“如果我们事先并不知道袋子里面黑白球的比例，而是闭着眼睛摸出一个（或好几个）球，观察这些取出来的球的颜色之后，那么我们可以就此对袋子里面的黑白球的比例作出什么样的推测”。这个问题，就是所谓的逆概问题。

我们用wiki上例子来说明：

一所学校里面有 60% 的男生，40% 的女生。男生总是穿长裤，女生则一半穿长裤一半穿裙子。有了这些信息之后我们可以容易地计算“随机选取一个学生，他（她）穿长裤的概率和穿裙子的概率是多大”，这个就是前面说的“正向概率”的计算。然而，假设你走在校园中，迎面走来一个穿长裤的学生（很不幸的是你高度近似，你只看得见他（她）穿的是否长裤，而无法确定他（她）的性别），你能够推断出他（她）是男生的概率是多大吗？

我们来算一算：

假设学校里面人的总数是 U 个。

60% 的男生都穿长裤，于是我们得到了 U \* P(Boy) \* P(Pants|Boy) 个穿长裤的（男生）（其中 P(Boy) 是男生的概率 = 60%，这里可以简单的理解为男生的比例；P(Pants|Boy) 是条件概率，即在 Boy 这个条件下穿长裤的概率是多大，这里是 100% ，因为所有男生都穿长裤）。

40% 的女生里面又有一半（50%）是穿长裤的，于是我们又得到了 U \* P(Girl) \* P(Pants|Girl) 个穿长裤的（女生）。加起来一共是 U \* P(Boy) \* P(Pants|Boy) + U \* P(Girl) \* P(Pants|Girl) 个穿长裤的，其中有 U \* P(Girl) \* P(Pants|Girl) 个女生。

两者一比就是你要求的答案。

下面我们把这个答案形式化一下：

我们要求的是 P(Girl|Pants) （穿长裤的人里面有多少女生），计算的结果是 ：

U \* P(Girl) \* P(Pants|Girl) / [U \* P(Boy) \* P(Pants|Boy) + U \* P(Girl) \* P(Pants|Girl)]

容易发现这里校园内人的总数是无关的，可以消去。于是得到：

P(Girl|Pants) = P(Girl) \* P(Pants|Girl) / [P(Boy) \* P(Pants|Boy) + P(Girl) \* P(Pants|Girl)]

注意，如果把上式收缩起来，分母其实就是 P(Pants) ，分子其实就是 P(Pants, Girl) 。而这个比例很自然地就读作：在穿长裤的人（ P(Pants) ）里面有多少（穿长裤）的女孩（ P(Pants, Girl) ）。

上式中的 Pants 和 Boy/Girl 可以指代一切东西，所以其一般形式就是：

P(B|A) = P(A|B) \* P(B) / [P(A|B) \* P(B) + P(A|~B) \* P(~B) ]

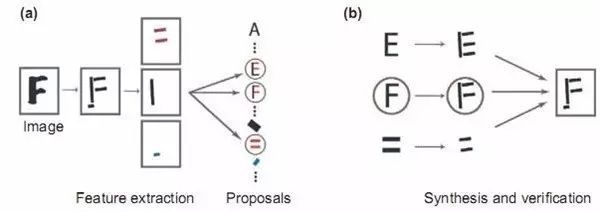
我们就得出了经典的贝叶斯公式：

P(B|A) = P(A|B) \* P(B) / P(A)

拉普拉斯说，概率论只是用数学公式把常识表现出来而已，大师的话确实让人深思。

**贝叶斯图像识别**

贝叶斯方法是一个非常 general 的推理框架。其核心理念可以描述成：Analysis by Synthesis （通过合成来分析）。认知科学新进展上有一篇 paper 就是讲用贝叶斯推理来解释视觉识别的，下图就是摘自这篇 paper ：

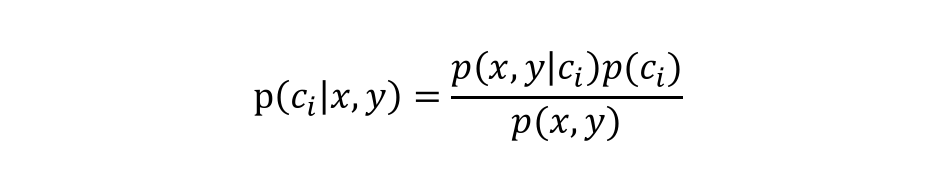


首先是视觉系统提取图形的边角特征，然后使用这些特征自底向上地激活高层的抽象概念（比如是 E 还是 F 还是等号），然后使用一个自顶向下的验证来比较到底哪个概念最佳地解释了观察到的图像。

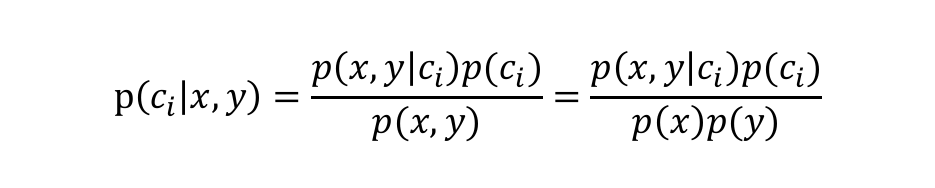
**朴素贝叶斯算法**

贝叶斯是从统计概率的角度来进行分类，确切来说是条件概率，例如要猜是哪一类动物，该动物具备的特征是：四条腿、高度超过x米、哺乳动物，那么在这些特征前提下，计算其为哪种动物的概率。

这里假设事物的属性只有两个x和y，类别用ci表示，可以通过条件概率或贝叶斯方法推测到：

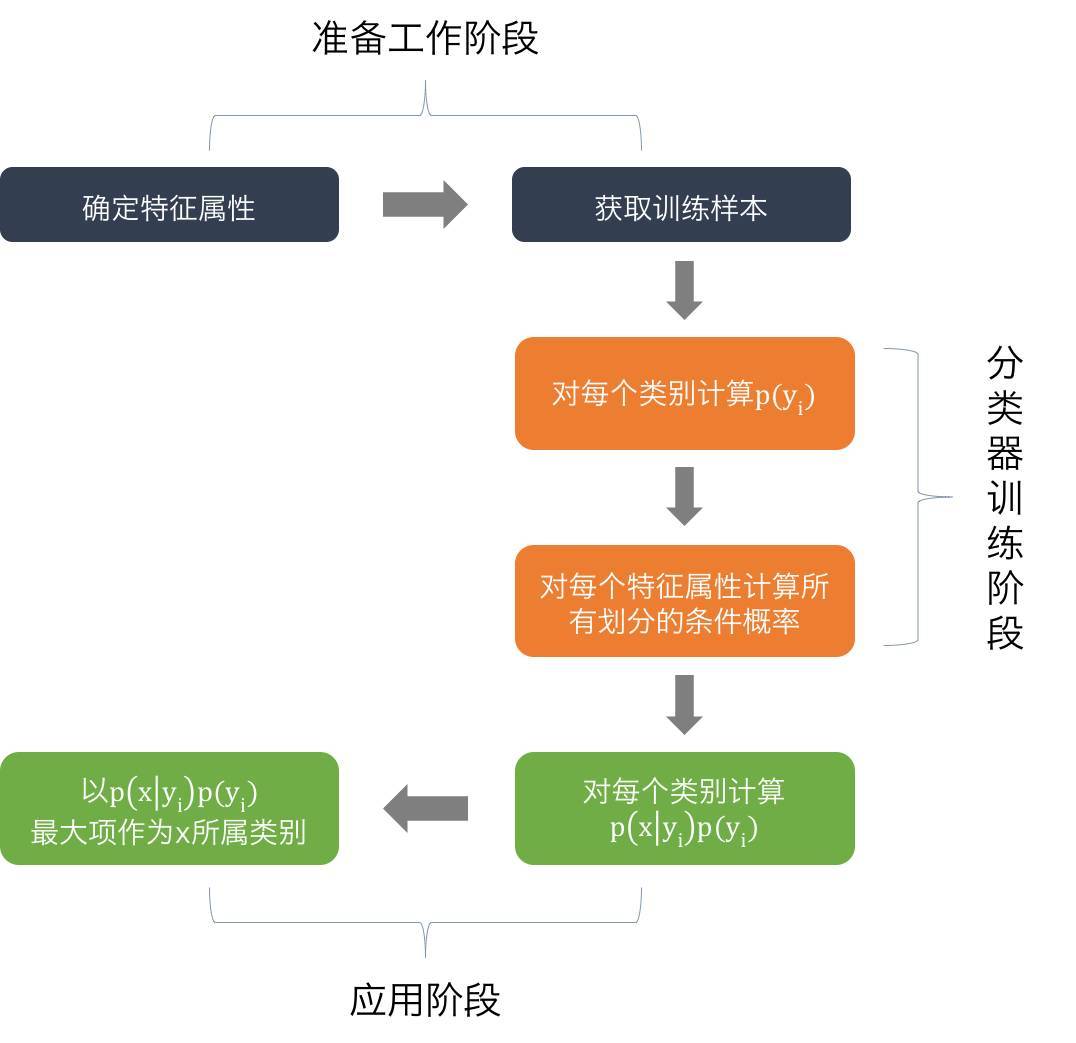


这样，就可以得出，在某个特定特征下属于某个类别的概率有多大，从而进行分类。所谓的朴素贝叶斯，假设的是：特征之间相互独立且是同等重要的，这样上面的公式就可以变为：



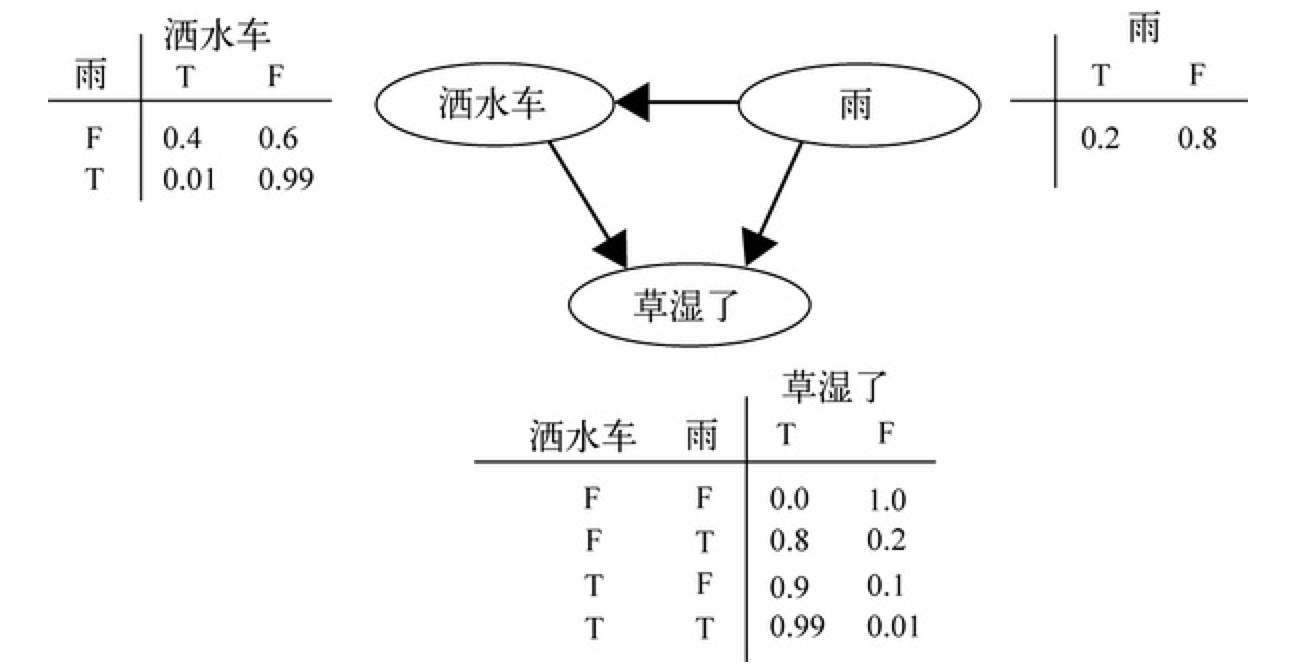
朴素贝叶斯的思想基础是这样的：对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。通俗来说，就好比这么个道理，你在街上看到一个黑人，我问你你猜这哥们哪里来的，你十有八九猜非洲。为什么呢？因为黑人中非洲人的比率最高，当然人家也可能是美洲人或亚洲人，但在没有其它可用信息下，我们会选择条件概率最大的类别，这就是朴素贝叶斯的思想基础。

朴素贝叶斯分类的流程可以由下图表示（暂时不考虑验证）：



**贝叶斯网络**

如果我们并不清楚一个现象的原因，就首先根据我们认为最有可能的原因来建立一个模型。然后将每个可能的原因当作网络中的节点连接起来，根据我们的预判或是专家的意见给每条连接分配一个先验概率值。换句话说，网络中的每个节点都通过贝叶斯公式和其他节点相连。接下来，只需向这个模型代入观察数据，通过网络节点间的贝叶斯公式重新计算出概率值即可。为每一个新数据、每一条连接重复这一计算，直到形成一个网络图，任意两个原因之间的连接都得出精确的后验概率值为止。这事便成了！



一个贝叶斯网络实例

**神奇的贝叶斯方法**

在经典统计学看来，科学是关于客观事实的研究，我们只要反复观察一个可重复的现象，直到积累了足够多的数据，就能从中推断出有意义的规律。而贝叶斯方法却要求科学家像算命先生一样，从主观猜测出发，这显然不符合科学精神。

实际上，统计学家和贝叶斯学家有一个有趣的争论，统计学家说：我们让数据自己说话。言下之意就是要摒弃先验概率。而贝叶斯支持者则说：数据会有各种各样的偏差，而一个靠谱的先验概率则可以对这些随机噪音做到健壮。

事实证明贝叶斯派胜利了，胜利的关键在于所谓先验概率其实也是经验统计的结果，譬如为什么我们会认为绝大多数硬币是基本公平的？为什么我们认为大多数人的肥胖适中？为什么我们认为肤色是种族相关的，而体重则与种族无关？先验概率里面的“先验”并不是指先于一切经验，而是仅指先于我们“当前”给出的观测数据而已，在硬币的例子中先验指的只是先于我们知道投掷的结果这个经验，而并非“先天”。

我们可以回到那个自然语言二义性的例子，并给出一个完美的解释了：如果语法结构是 The girl saw the-boy-with-a-telecope 的话，怎么那个男孩偏偏手里拿的就是望远镜——一个可以被用来 saw-with 的东东捏？这也忒小概率了吧。他咋就不会拿本书呢？拿什么都好。怎么偏偏就拿了望远镜？所以唯一的解释是，这个“巧合”背后肯定有它的必然性，这个必然性就是，如果我们将语法结构解释为 The girl saw-with-a-telescope the boy 的话，就跟数据完美吻合了——既然那个女孩是用某个东西去看这个男孩的，那么这个东西是一个望远镜就完全可以解释了（不再是小概率事件了）。

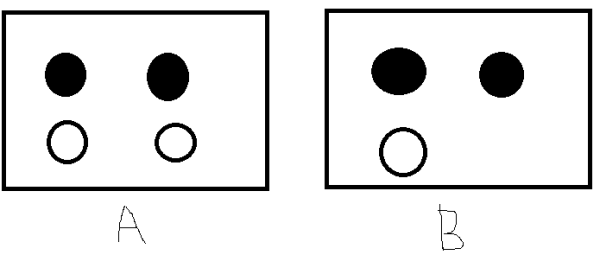
<https://www.cnblogs.com/zhoulujun/p/8893393.html>

# [贝叶斯公式由浅入深大讲解—AI基础算法入门](https://www.cnblogs.com/zhoulujun/p/8893393.html)

**1 贝叶斯方法**

    长久以来，人们对一件事情发生或不发生的概率，只有固定的0和1，即要么发生，要么不发生，从来不会去考虑某件事情发生的概率有多大，不发生的概率又是多大。而且概率虽然未知，但最起码是一个确定的值。比如如果问那时的人们一个问题：“有一个袋子，里面装着若干个白球和黑球，请问从袋子中取得白球的概率是多少？”他们会想都不用想，会立马告诉你，取出白球的概率就是1/2，要么取到白球，要么取不到白球，即θ只能有一个值，而且不论你取了多少次，取得白球的概率θ始终都是1/2，即不随观察结果X 的变化而变化。

    这种频率派的观点长期统治着人们的观念，但是：



假设我们有如下的7个球在A,B两个框中，如果我们随便取一个球，已知取到的球来自B框中，那么这个球是白球的概率是多少呢？或者问去除的球是白色，那么取自B框的概率是多少呢？这个问题不是很好解决，直到后来一个名叫**Thomas Bayes**的人物出现。

**1.1 贝叶斯方法的提出**

[托马斯·贝叶斯Thomas Bayes](https://en.wikipedia.org/wiki/Bayes%27_theorem)（1702-1763）在世时，并不为当时的人们所熟知，很少发表论文或出版著作，与当时学术界的人沟通交流也很少，用现在的话来说，贝叶斯就是活生生一民间学术“屌丝”，可这个“屌丝”最终发表了一篇名为“**An essay towards solving a problem in the doctrine of chances**”，翻译过来则是：机遇理论中一个问题的解。你可能觉得我要说：这篇论文的发表随机产生轰动效应，从而奠定贝叶斯在学术史上的地位。

事实上，上篇论文发表后，在当时并未产生多少影响，在20世纪后，这篇论文才逐渐被人们所重视。对此，与梵高何其类似，画的画生前一文不值，死后价值连城。

    回到上面的例子：“有一个袋子，里面装着若干个白球和黑球，请问从袋子中取得白球的概率θ是多少？”贝叶斯认为取得白球的概率是个不确定的值，因为其中含有机遇的成分。比如，一个朋友创业，你明明知道创业的结果就两种，即要么成功要么失败，但你依然会忍不住去估计他创业成功的几率有多大？你如果对他为人比较了解，而且有方法、思路清晰、有毅力、且能团结周围的人，你会不由自主的估计他创业成功的几率可能在80%以上。这种不同于最开始的“非黑即白、非0即1”的思考方式，便是贝叶斯式的思考方式。

    继续深入讲解贝叶斯方法之前，先简单总结下频率派与贝叶斯派各自不同的思考方式：

* 频率派把需要推断的参数θ看做是固定的未知常数，即概率θ虽然是未知的，但最起码是确定的一个值，同时，样本X 是随机的，所以频率派重点研究样本空间，大部分的概率计算都是针对样本X 的分布；
* 而贝叶斯派的观点则截然相反，他们认为参数θ是随机变量，而样本X 是固定的，由于样本是固定的，所以他们重点研究的是参数θ的分布。

    相对来说，频率派的观点容易理解，所以下文重点阐述贝叶斯派的观点。

    贝叶斯派既然把θ看做是一个随机变量，所以要计算θ的分布，便得事先知道θ的无条件分布，即在有样本之前（或观察到X之前），θ有着怎样的分布呢？

    比如往台球桌上扔一个球，这个球落会落在何处呢？如果是不偏不倚的把球抛出去，那么此球落在台球桌上的任一位置都有着相同的机会，即球落在台球桌上某一位置的概率服从均匀分布。这种在实验之前定下的属于基本前提性质的分布称为先验分布，或的无条件分布。

    至此，贝叶斯及贝叶斯派提出了一个思考问题的固定模式：

            先验分布 π(θ)+ 样本信息χ⇒  后验分布π(θ|x)

上述思考模式意味着，新观察到的样本信息将修正人们以前对事物的认知。换言之，在得到新的样本信息之前，人们对的认知是先验分布π(θ)，在得到新的样本信息后χ，人们对θ的认知为π(θ|x)。

而后验分布π(θ|x)一般也认为是在给定样本χ的情况下θ的条件分布，而使达到最大的值称为最大后θMD验估计，类似于经典统计学中的极大似然估计。

    综合起来看，则好比是人类刚开始时对大自然只有少得可怜的先验知识，但随着不断是观察、实验获得更多的样本、结果，使得人们对自然界的规律摸得越来越透彻。所以，贝叶斯方法既符合人们日常生活的思考方式，也符合人们认识自然的规律，经过不断的发展，最终占据统计学领域的半壁江山，与经典统计学分庭抗礼。

    此外，贝叶斯除了提出上述思考模式之外，还特别提出了举世闻名的贝叶斯定理。

**1.2 贝叶斯定理**

  在引出贝叶斯定理之前，先学习几个定义：

* **边缘概率**（又称**先验概率**）：某个事件发生的概率。边缘概率是这样得到的：在联合概率中，把最终结果中那些不需要的事件通过合并成它们的全概率，而消去它们（对离散随机变量用求和得全概率，对连续随机变量用积分得全概率），这称为边缘化（marginalization），比如A的边缘概率表示为P(A)，B的边缘概率表示为P(B)。
* **联合概率**表示两个事件共同发生的概率。A与B的联合概率表示为P(A∩B)或者P(A,B)。
* [**条件概率**](https://zh.wikipedia.org/wiki/%E6%9D%A1%E4%BB%B6%E6%A6%82%E7%8E%87)**（**又称**后验概率）**：事件A在另外一个事件B已经发生条件下的发生概率。条件概率表示为P(A|B)，读作“在B条件下A的概率”,。

  接着，考虑一个问题：P(A|B)是在B发生的情况下A发生的可能性。

* 首先，事件B发生之前，我们对事件A的发生有一个基本的概率判断，称为A的先验概率，用P(A)表示；
* 其次，事件B发生之后，我们对事件A的发生概率重新评估，称为A的后验概率，用P(A|B)表示；
* 类似的，事件A发生之前，我们对事件B的发生有一个基本的概率判断，称为B的先验概率，用P(B)表示；
* 同样，事件A发生之后，我们对事件B的发生概率重新评估，称为B的后验概率，用P(B|A)表示。

贝叶斯定理便是基于下述贝叶斯公式：

u=239894515,405307697&fm=58.jpeg

P(A|B)=P(B|A)P(A)/P(B)

上述公式的推导其实非常简单，就是从条件概率推出。

 根据条件概率的定义，在事件B发生的条件下事件A发生的概率是

P(A|B)=P(A∩B)/P(B)

同样地，在事件A发生的条件下事件B发生的概率

P(B|A)=P(A∩B)/P(A)

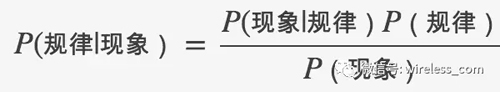
整理与合并上述两个方程式，便可以得到：

P(A|B)P(B)=P(A∩B)=P(B|A)P(A)

    接着，上式两边同除以P(B)，若P(B)是非零的，我们便可以得到贝叶斯定理的公式表达式：

P(A|B)=P(B|A)\*P(A)/P(B)

P(A|B) 和 P(B|A) 之类的经常让人混淆，陈老师给出了理解的一个关键点，区分出规律和现象，就是将A看成“规律”，B看成“现象”，那么贝叶斯公式看成：

[](http://www.zhoulujun.cn/zhoulujun/uploadfile/2017/0913/20170913094643189.jpg)

**贝叶斯推断的含义**

然后搜下，发现其实还有更好阐释，比如

对条件概率公式进行变形，可以得到如下形式：

我们把P(A)称为"先验概率"（Prior probability），即在B事件发生之前，我们对A事件概率的一个判断。P(A|B)称为"后验概率"（Posterior probability），即在B事件发生之后，我们对A事件概率的重新评估。P(B|A)/P(B)称为"可能性函数"（Likelyhood），这是一个调整因子，使得预估概率更接近真实概率。

所以，条件概率可以理解成下面的式子：

后验概率　＝　先验概率 ｘ 调整因子

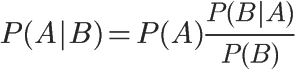
这就是贝叶斯推断的含义。我们先预估一个"先验概率"，然后加入实验结果，看这个实验到底是增强还是削弱了"先验概率"，由此得到更接近事实的"后验概率"。

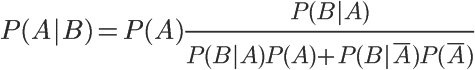
在这里，如果"可能性函数"P(B|A)/P(B)>1，意味着"先验概率"被增强，事件A的发生的可能性变大；如果"可能性函数"=1，意味着B事件无助于判断事件A的可能性；如果"可能性函数"<1，意味着"先验概率"被削弱，事件A的可能性变小。

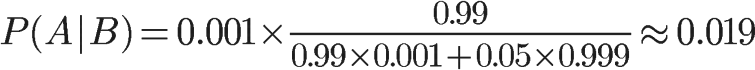
贝叶斯定理应用示例:

已知某种疾病的发病率是0.001，即1000人中会有1个人得病。现有一种试剂可以检验患者是否得病，它的准确率是0.99，即在患者确实得病的情况下，它有99%的可能呈现阳性。它的误报率是5%，即在患者没有得病的情况下，它有5%的可能呈现阳性。现有一个病人的检验结果为阳性，请问他确实得病的可能性有多大？

假定A事件表示得病，那么P(A)为0.001。这就是"先验概率"，即没有做试验之前，我们预计的发病率。再假定B事件表示阳性，那么要计算的就是P(A|B)。这就是"后验概率"，即做了试验以后，对发病率的估计。

根据条件概率公式，

用全概率公式改写分母，

将数字代入，

我们得到了一个惊人的结果，P(A|B)约等于0.019。也就是说，即使检验呈现阳性，病人得病的概率，也只是从0.1%增加到了2%左右。这就是所谓的"假阳性"，即阳性结果完全不足以说明病人得病。

或许换成这个公式 P(A|B)=P(A∩B)/B，看起来更加直白写：

阐释：

如果没有误报，那么得病率：.001\*.99

如果是误报，那么得病率为：.05\*(1-.0001)，

所以：

p(A|B)=.001\*.99/[.99\*.001+.05\*(1-.0001)]=.019

为什么会这样？为什么这种检验的准确率高达99%，但是可信度却不到2%？答案是与它的误报率太高有关。

（【习题】如果误报率从5%降为1%，请问病人得病的概率会变成多少？）

有兴趣的朋友，还可以算一下"假阴性"问题，即检验结果为阴性，但是病人确实得病的概率有多大。然后问自己，"假阳性"和"假阴性"，哪一个才是医学检验的主要风险？

网页上后面列了很多例子。