

忙总谈数学：瞎扯贝叶斯理论的基本思想

January 22, 2022

Abstract

现在贝叶斯的思想已经成为归纳逻辑的核心,并且逐步发展为一套一般性的科学推理理论和方法。贝叶斯思想现在是机器学习的核心方法之一。

简单总结:

(1)、当假定世界本质是随机的,那么我们认识世界的结论也是不确定的,只能通过不断累积的经验去逼近;

(2)、所以我们认识世界本质基于经验积累;

(3)、我们判断事物是什么的准确概率,往往基于经验积累程度多少。或者说,我们经验积累,能够使我们逐步减少判断事件错误的概率。

所以贝叶斯对先验概率的指定既是主观的,又是理性的,而且随经验积累逐渐优化。

从哲学角度看,贝叶斯思想是一种逻辑+历史的方法,是归纳推理方法的一次革命(归纳推理就是根据过去的经验预测未来的推理),把经验数据量化,并直接带入预测判断。解决了休谟对归纳推理的合理性提出的质疑。

豆瓣小组“管理实践与学习”

<https://www.douban.com/group/542139/?ref=sidebar>

瞎扯贝叶斯理论的基本思想

来自: wxmang 2015-12-25 13:28:14

<https://www.douban.com/group/topic/82509566/>

既然是瞎扯，就不是很严谨的，因为要简单明了，可能有的地方细节删除太多了，导致说不通。所以只能大概齐。真的要弄懂，还是请看教科书为好。写这种文章，华罗庚先生写得最好。

1 倒向问题

自从人类有自我意识，可能就在讨论一个至今没有结论的问题：机遇（或者运气，或者机会）到底是什么？怎么把握，怎么预测，怎么估计或计算大小等等。这是人类的一个核心问题。其实这也是一个典型的倒向问题。

人类思考问题有两个方向，一个是正向，也即知道结果找原因（例如现在我们经常讨论的明朝灭亡的原因）；一个是倒向，也即根据一些现象判断结果（例如如果我们事先并不知道黑箱里面黑白球的比例，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，就此对黑箱里面的黑白球的比例进行推测。现实需要大量的倒向计算，例如现在某些现象出现，企业会不会破产，现在应该怎么办等等）。

用正向思维方式研究问题的，我们叫他们事后诸葛亮，历史学家便是。这种研究只能用于经验总结和知识储备上。

用倒向思维方式研究问题的，我们叫他们预测大师，见微知著的真人，管蠡窥测或以蠡测海的超人等等。实际上有价值的问题多数都是倒向问题，例如：股市上，通过那几点征兆就能判断是一次多或空的机会；医院中，通过那几个症状就能判断是什么病；科学研究上，通过几个实验数据，就能构造什么理论解释模型等等。一般说来，数学家，物理学家等等都是研究倒向问题的，或者说，他们不能通过很少征兆或现象来预测或判断结果，就没存在价值（顺便说一句，不知道倒向问题思维方式的人，没法再金融市场或股市搏击，目前投机市场最前沿的研究几乎就是倒向随机过程和鞅论为主，中国的彭实戈院士是倒向随机过程领域的其中一个领军人物）。

如何用倒向思维方式研究机遇？也即如何从一些征兆或现象，判断机遇，或者把问题进行等价推广：如何用一些已知的信息或经验，判断或预测未知。

1763年，英国的长老会牧师贝叶斯发表了一篇论文“论有关机遇问题的求解”，提出了解决的框架：那就是用不断增加的信息和经验，可以逐步逼近未知的真相或理解未知。并给出了算法（其实贝叶斯由于是一个牧师，他关心的原始问题本来的表述是：人能不能根据凡人世界的经验和现实世界的证据，证明上帝的存在，因为宗教人士的逻辑是机遇就是上帝存在的主要证据，能够认识机遇的规律，几乎等同于证明上帝存在）。

后来经拉格朗日等数学家进一步努力，获得了大突破，贝叶斯理论成为现代统计学两大支柱之一。

由于我们不讨论数学，所以不进一步讨论贝叶斯思想的各种复杂数学表达，我们只讨论其基本思想。

2 贝叶斯基本思想

下面我们用一个例子来介绍贝叶斯思想。

假定甲乙两人做一个游戏：甲蒙上眼睛，乙随手在一张纸上画一条线段，两个端点分别A点和B点，再随机在线段上画一点C，现在游戏是：通过一些信

息，让甲判断C点位置。

信息1：乙随机在线段上划点，并告诉每一个点是离A端点近还是离B端点近；

信息2：乙必须告诉每一个点是位于AC之间还是CB之间。

显然只需要有限次划点，甲就能基本确定C的大致位置。

为简单说明，我们假设乙划点总是在中位点，也即第一个点是A+B的中点，根据上面游戏规则，必须告诉甲的两个信息，甲这样就能判断C靠A点近还是B点近；如果C靠A点近，那么第二个点划在A+C的中点，同样重复上述步骤，甲就能判断C点是靠A点近还是 $(A+C)/2$ 近，继续循环。我们知道中位点划分会形成一个 $1/(2N)$ 的收敛级数，可以在有限的N次内，就能获得C点位置范围，误差不大于 $1/(2N)$ 次方）。

这就是用逐步增加的信息确定未知的例子。这也就是贝叶斯思想的基本模型。

当然贝叶斯考虑的逐步增加信息是人的经验和知识构成的先验信息，而先验信息会因为个人的经验偏见，视野局限，测量误差，外部环境变化，数据丢失等等导致一定的不准确，也即只有概率意义上的正确。所以他能够得到的对未知的判断也只能是概率意义上的。

简单总结一下贝叶斯的基本观点是：

(1)、由于经验或知识是否正确带有不确定性因素，所以基于以前经验和知识（也即先验），根据一些随机出现或观察到的现象来判断事物真相，原因或未知，就有一定的不确定性。（其实这个观点有一个强大的前提假设：未知世界本质是随机的，所以任何未知都具有不确定性，这是现代量子力学和复杂系统证明了的假设）。

(2)、由于我们依靠已知的经验或知识来分析观察到的现象，以此推测或断未知，所以任何未知的判断或推测都是是不确定的，能用一个概率分布去描述，也即未知的不确定性程度由先验概率分布和现象出现的概率分布决定。

上面这句话的本质意思是：是不是能够通过一些现象正确判断或推测未知，取决于我们经验多少和掌握的现象多少。例如，医生能否通过症状判断疾病，取决于医生知识（这也是一种经验，是从前人身上学习的）和经验积累，也取决于掌握的症状多少（掌握症状一般通过各种检查实现，例如为了掌握症状而进行的量体温，照X光，核磁共振，CT检查，超声波检查等等）。

但是光掌握症状并不能完全判断病症（否则就不需要医生，只需要检测工程师了），还需要医生知识和经验。

因为任何检查都不可能完备，再加上任何医生的经验和知识也是不完备的（例如对一种新疾病原有的经验和知识就无能为力），所以任何判断都有一定的对错概率。

(3)、由于基于N个现象（或症状）和个人主观经验来判断未知带有一定不确定，我们往往需要做多次检测和由具有不同个人主观经验的人来判断，然后按照极大似然原则选择结果（例如医院诊断重大疾病往往要请不同医生会诊，也要进行不同系列，不同类型的医疗检查）。

(4)、先验信息可以通过的收集、挖掘和加工而数量化，形成先验分布（也即所谓专家知识库可以提高判断精度）。

上面这些观点，综合起来就是：

现实世界本身是不确定的，人类的观察能力是有局限性的（例如如果人类能够直接观察到电子的运行，还需要假设什么模型），人类所观察到的只是事物

表面上的结果（若干症状或现象），例如往往只能知道从黑箱里面取出来的球是什么颜色，而并不能直接看到黑箱里面实际的情况。

贝叶斯思想给我们提供了一个猜测黑箱里面情况的方法。当然这种方法得到的结果是不确定的（因为世界本质就是不确定的，而且这种方法依赖于人的主观经验，本身是否正确也是不确定的）。

贝叶斯的方法其实是一个算法：第一步，算出各种不同猜测的可能性大小；第二步算出最可能的猜测是什么。

第一步就是计算特定猜测的后验概率（对于连续的猜测空间就是计算猜测的概率密度函数），第二步则是极大似然方法。

极大似然的定义是：事件A与参数 $\theta \in \Theta$ 有关， θ 取值不同，则 $P(A)$ 也不同，若A发生了，则认为此时的 θ 值就是 θ 的估计值。这就是极大似然。

例如两人一起打猎，只响一枪，就打中一猎物，那么按正常逻辑，一枪命中，肯定是枪法好的那个，这个推断就体现了极大似然法的基本思想。

再例如，袋中装有许多黑、白球，不同颜色球的数量比为3:1，试设计一种方法，估计任取一球为黑球的概率P。

显然P的值无非是1/4或3/4，需要通过抽样来决定分布中参数究竟是1/4还是3/4。现从袋中有放回地任取3只球，显然白球出现次数多的话，P就是1/4，黑球出现次数多的话，P就是3/4）。

也即贝叶斯思想就是：对于给定观测数据，一个猜测是否正确，取决于这个猜测本身独立的可能性大小（先验概率）和这个猜测生成我们观测到的数据的可能性大小（似然）的乘积。（也即最可能的猜测=先验概率*似然最大的这个值）。

3 贝叶斯公式

先用一个wikipedia上的例子介绍贝叶斯公式的原理。

假设一所学校里面有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。问题：随机在校园乱走，看到在校园里穿长裤的人里面有多少女生的概率？

假设学校里面人的总数是N个。60%的男生都穿长裤，于是我们得到了 $N * P(\text{Boy}) * P(\text{Pants}|\text{Boy})$ 个穿长裤的（男生）（其中 $P(\text{Boy})$ 是男生的概率=60%，这里可以简单的理解为男生的比例； $P(\text{Pants}|\text{Boy})$ 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 $N * P(\text{Girl}) * P(\text{Pants}|\text{Girl})$ 个穿长裤的（女生）。加起来一共是 $N * P(\text{Boy}) * P(\text{Pants}|\text{Boy}) + N * P(\text{Girl}) * P(\text{Pants}|\text{Girl})$ 个穿长裤的。

整理一下： $P(\text{Girl}|\text{Pants})$ （穿长裤的人里面有多少女生的概率）= $N * P(\text{Girl}) * P(\text{Pants}|\text{Girl}) / [N * P(\text{Boy}) * P(\text{Pants}|\text{Boy}) + N * P(\text{Girl}) * P(\text{Pants}|\text{Girl})]$ 。很容易发现公式与校园内人的总数是无关的，可以消去。于是得到

$$P(\text{Girl}|\text{Pants}) = P(\text{Girl}) * P(\text{Pants}|\text{Girl}) / [P(\text{Boy}) * P(\text{Pants}|\text{Boy}) + P(\text{Girl}) * P(\text{Pants}|\text{Girl})]$$

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：

$$P(B|A) = P(A-B) * P(B) / [P(A|B) * P(B) + P(A|\sim B) * P(\sim B)]$$

即已知 $P(A|B)$ 、 $P(A)$ 和 $P(B)$ ，可以计算出 $P(B|A)$ 。

这就是著名的贝叶斯公式。所以拉普拉斯说概率论只是把常识用数学公式表达了出来。

4 贝叶斯思想在决策中

当然贝叶斯思想用的最多的领域是决策。贝叶斯决策就是在不完全情报下，对部分未知的状态用主观概率估计，然后用贝叶斯公式对发生概率进行修正，最后再利用期望值和修正概率做出最优决策。

贝叶斯决策属于风险型决策，决策者虽不能控制客观因素的变化，但却掌握其变化的可能状况及各状况的分布概率，并利用期望值即未来可能出现的平均状况作为决策准则。

贝叶斯思想能够用于决策的原因是：除样本（例如决策调查获得的一些事件或资料）提供的基础信息外，人类的经验的先验信息也是决策判断的重要依据。

例如以对神童出现的概率 p 的估计为例。按经典统计的做法，完全由样本提供的信息(即基础抽样信息)来估计，认为参数 p 是一个“值”。而人类经验其实对 p 已经有了一个了解，如 p 可能取 p_1 与 p_2 ，且取 p_1 的机会很大，取 p_2 机会很小。先验信息关于参数 p 的信息是一个“分布”，如 $P(p = p_1) = 0.9$ ， $P(p = p_2) = 0.1$ ，即在抽样之前已知道(先验的) p 取 p_1 的可能性为0.9。若不去抽样便要作出推断，自然会取 $p = p_1$ 。但若抽样后，除非采样信息(即样本提供的信息)包含十分有利于“ $p = p_2$ ”的支持论据，否则采纳先验的看法“ $p = p_1$ ”。

下面用一个例子解释这种决策。

假设患者有肺炎为事件 B ， $P(B)$ 是医生在检测前，基于经验对当时当地的肺炎患病率的概率判断，例如1%，也即如果没有任何检查，医生基于经验，会判断得肺炎概率很小， $P(\sim B)$ 是不是肺炎的概率，例如是99%；

$P(A)$ 是全体人群（包括肺炎和不是肺炎人群）检查出现各种指标异常的概率，这可以是一个函数曲线，自变量是各种检测指标，总和等于1（可以经过规范整理得到）；

$P(A|B)$ 是当病人是肺炎时，各种检测指标出现异常的概率，是医院经过大量临床试验和检测数据分析得到的；

$P(A|\sim B)$ 是当病人不是肺炎时，各种检测指标出现异常的概率，是医院经过大量临床试验和检测数据分析得到的；

那么根据贝叶斯公式，就可以计算出 $P(B|A)$ ，也即当某一个异常指标出现时，确诊是肺炎的概率。

例如假设我们知道检测肺炎有12个指标，根据多年来的经验和大量临床数据，如果是肺炎，可以得到这12个检测指标的出现异常的概率分布情况，也即知道 $P(A_i|B)$ ；对不是肺炎的，这12个指标的异常概率分布情况我们也知道，也即 $P(A_i|\sim B)$ 。

在诊疗过程中，医生要根据临床经验对各种病症状 A_i 进行权衡。当然由于经验数据误差和知识不完备，存在误诊概率。

举个例子来说明，假设有一台肺炎诊断仪，通过对它以往的诊断记录的分析，如果患者确实患有肺炎它的确诊率为90%，若果患者没有患肺炎，被诊断成肺炎的概率为10

那么如果一个人被这台诊断仪确诊成肺炎（这是现象），这个人患有肺炎的概率是多少（这是判断）？

设 A ：肺炎诊断仪给出肺炎诊断。 B_1 ：病人是肺炎患者。 B_2 ：病人不是肺炎患者。

根据贝叶斯公式： $P(A|B_1) = 90\%$ ； $P(A) = 90\% * P(B_1) + 10\% * P(B_2)$ ；

则 $P(B_1|A)=P(B_1)*90\%/(90\%*P(B_1)+10\%*P(B_2))$;

我们知道当时当地人群中肺炎患者的比重很小，假设为1%（这就是经验，有一定的主观性），则 $P(B_1)=1\%$ ； $P(B_2)=99\%$ ；

可以算出： $P(B_1|A)=8\%$ ，也即一个人被这台诊断仪确诊成肺炎，而这个人真正患有肺炎的概率只有8%。显然这是一个不能饶恕的结果。

但是实际上医生经常这么做，只是他们不是用肺炎诊断仪，而是用一些简单诊断手段就随便下结论了。

所以医院需要加入更多的检查项目，增加更多的检测设备，加入更多医生判断的判断，通过极大似然原则，筛选最可能的结果，逐步逼近真相，解决掉判断的错误风险。

5 先验和后验

先验就是指在抽样前就有的经验信息的概率表述，先验不必有客观的依据，它可以部分地或完全地基于主观信念。

再以肺炎为例子，某人认为自己得肺炎，去看病，医生在没有检测之前，基于经验，认为本地在此时此地，得肺炎情况很小，认为他极可能只是感冒，这就是先验。

当然医生不会完全相信自己先验，必然会给某人进行完整的检查，例如测体温，量血压，甚至照X光等等，这些检测结果数据就构成检测样本X。

医生诊断时，必然不会只使用X（检测数据）提供的信息进行诊断，而必须先验（否则就不需要医生，只需要工程师就能治病了）。

假定 $\theta=0$ 是没病， $\theta=1$ 是有病，贝叶斯理论认为X的分布取决于 θ 是0还是1，同时根据贝叶斯公式，我们知道了X，就有助于推断 θ 是否为1。

通过使用贝叶斯公式计算后验分布，也即根据抽样样本（检测数据）X的分布 $p(x)$ 及 θ 的先验分布 $p(\theta)$ （当时当地某种疾病的分布的先验概率，由医生经验和知识积累决定），用贝叶斯公式就可计算出在已知 $X=x$ 的条件下， θ 的条件分布 $p(\theta|x)$ （这就是后验概率，也即经取样和先验概率计算后的出来的）。

显然这个分布综合了样本X（检测数据）及先验分布 $p(\theta)$ （医生经验）所提供的有关的信息。假定设 $p(\theta=1)=0.001$ ，经计算， $p(\theta=1|x)=0.86$ ，则含义为：在某人的检测指标量出之前，根据医生经验（ $p(\theta=1)$ ），他患病的可能性定为0.001，而在得到X后，认识发生了变化：其患病的可能性提高为0.86。这一判断既与X有关，但这也离不开先验分布。也即若当时当地肺炎的发病率很小，医生将倾向于只有在样本X显示出很强的证据时，才诊断某人有肺炎，这样就可以避免误诊。

6 奥卡姆剃刀对决策的筛选

贝叶斯思想本质上是一个经验归纳推理的计算公式，解决了逐步逼近真实的过程，但是显然，这种逼近有不确定性，因为即便一个判断与经验数据非常符合，也并不代表这个判断就是正确的判断，因为经验数据总是会有各种各样的误差，比如观测误差，记忆误差等等，再加上有时无法分别背景噪音，把背景噪音纳入经验数据，来加以判断（背景噪音可以看成与事件完全无关的因素）。

这时，我们就必须应用奥卡姆剃刀：如果两个理论具有相似的解释力度，那么优先选择那个更简单的（往往也正是更平凡的，更少繁复的，更常见的）。

所以现实中，我们建立解释问题的模型往往只提取出几个与结果相关度很高，很重要的因素，而不会面面俱到，太复杂的模型不但成本高，而且会因为无法辨别背景噪声因素，导致失真。

例如我们知道现在各种计算健康体重的模型都是用身高和体重近似于一个二阶多项式的关系（对人群随机抽取了N个样本，用最小二乘法拟合出一个二阶多项式），但大家都知道并不是只有身高才会对体重产生影响。但是总体上来说，绝大多数人的身高与体重有关，成正态分布，这个分布就保证了身高体重相关模型能够在大多数情况下做出靠谱的预测。当然人有胖瘦，密度也有大小，所以完美符合身高体重的二阶多项式关系的人是不存在的，只是近似。这就是人类对事物理解的模式。

7 简单总结

根据上述描述，我们可以得出几个结论：

（1）、当假定世界本质是随机的，那么我们认识世界的结论也是不确定的，只能通过不断累积的经验去逼近；

（2）、所以我们认识世界本质基于经验积累；

（3）、我们判断事物是什么的准确概率，往往基于经验积累程度多少。或者说，我们经验积累，能够使我们逐步减少判断事件错误的概率。

所以贝叶斯对先验概率的指定既是主观的，又是理性的，而且随经验积累逐渐优化。

从哲学角度看，贝叶斯思想是一种逻辑+历史的方法，是归纳推理方法的一次革命（归纳推理就是根据过去的经验预测未来的推理），把经验数据量化，并直接带入预测判断。解决了休谟对归纳推理的合理性提出的质疑。

现在贝叶斯的思想已经成为归纳逻辑的核心，并且逐步发展为一套一般性的科学推理理论和方法。贝叶斯思想现在是机器学习的核心方法之一。

不用数学公式介绍数学思想，真的很费力，也不知道讲清楚没有。

问答部分（注意：排列不一定是按照时间顺序）

jianduanle：忙总从想写这篇文章到发表，历时2个半小时。全文6312字。播音员播报的语速是每分钟500字，中等水平的打字员速度是120字/分钟。确实如忙总所说，是累翻两个打字员的节奏。

卡冈诺维奇：忙总写文章的捷才真是让人佩服，我憋个论文，一天下来能有一千字就不错了，多少天一个字也写不出来，真是囧！

一休哥们儿：请教忙总，如果基于样本和经验得到了一定的概率分布，但由于不确定性，仍不能确保决策正确，这时做决策时如何对风险和不确定性进行管理，有没有相关的框架可遵循。

wxmang：请教忙总，如果基于样本和经验得到了一定的概率分布，但由于不确定性，仍不能确保决策正确，这时?...?一休哥们儿

所以贝叶斯方法引入了似然·极大似然方法。其实在前面介绍中有：多检测，会诊。

猾心：请教芒总可能不相关的问题，研究广义相对论中有个叫因果的律理论，从贝叶斯理论能否推导因果律的正确否？

wxmang：?2015-12-26 17:40:19 量子力学其实推翻了严格的因果关系。所以基于世界本质是随机假设的贝叶斯理论，也推翻了严格因果关系的存在，认为事物之间因果关系只存在于概率分布意义下，与量子力学差不多。

xienan：忙总，大数据分析关注事件之间的相关性，而不关注因果性；量子力学中用概率解释代替因果解释。这二者之间是否有某种联系？

wxmang：量子力学其实不完全是推翻微观世界因果性这么简单，因为波恩的概率解释、海森堡的不确定性原理和玻尔的互补原理是完整理解量子力学的基础，简单来讲，前两者摧毁了经典世界的（严格）因果性，互补原理和不确定性原理又合力捣毁了世界的（绝对）客观性。建议看看《量子论的哥本哈根解释》。

xienan：我本人是赞同哥本哈根学派的。个人感觉，因果性和时间的连续性，是人类的某种错觉。或者说是在人类认知尺度上，对真实世界的一种近似。建立保罗万象的大一统理论是件遥不可及的事。

wxmang：这个不懂，尤其是弦论。不过我的同班同学，既有弦论大师，也有反对弦论的，认为他们走到邪路上去了，都是一流大学的一流教授，所以不知道谁对。

张小涛：文科生完全看不懂.....?所以直接看结论。经验累积足以让一些反应成为本能，这些本能将有助于躲避危险和把握机遇。年轻人阅历少没有获取直接经验的机会，可行办法就是读历史读案例做情景模拟模拟代入前人角色汲取间接经验。不知道这个思路对不对？

wxmang：不是这样的，贝叶斯是一种归纳推理算法：也即推理的步骤。不是一些单纯概念。现在垃圾邮件拦截，网络搜索（例如Google），机器翻译等等都是贝叶斯方法。贝叶斯方法是硬技术，相当于车床，冲床，不是随便说说的东西。

wxmang：马尔科夫过程在豆瓣写不了，因为豆瓣不支持符号。想介绍马尔科夫过程的核心思想，必须介绍布朗运动，维纳分布，郎之万方程和时间对称等等概念，这些玩意没数学公式，光凭口水话，我觉得写的人都看不懂。

wxmang：概率讨论稍微深入一点，就得讨论勒贝格测度，不然进行不下去。我没勇气在这里讨论，这玩意需要的预备知识太多了。当年科大上了一学期课，讲了26个引理，才算引导到可以开始讨论阶段。

wxmang：不是特别喜欢数学，不建议自学数学，浪费时间，因为歧路太多，容易迷路。

明月当空：忙总，请教您一个问题，不知道您是否了解现在的一种叫做“全脑教育（全脑思维课程）”，下午参加了一次这样的课程，里面的小孩子展现的蒙眼识字、快速作诗，快速记忆等一系列神奇能力，这种培训可信吗？

wxmang：不了解，我觉得现在骗子太多，傻子都不够用了。智力开发不是这种以培养巫婆神汉为目标的模式。

柯贤达：大赞芒总，写代码应用过贝叶斯。期待芒总有空时能再来一篇普及

下马尔可夫过程

wxmang: 马尔可夫过程解释得最好的一本书是钟开来的名著《马尔可夫过程, 布朗运动和时间对称》, 建议有兴趣看看, 豆瓣没法科普, 因为没有办法展示数学符号。没数学符号, 解释不了转移矩阵这些最简单概念。

Cousteau: 公式符号一般是用“公式编辑器”做好再截图发到网上的。不清楚网友们怎么直接过度到马尔可夫过程了, 忙总如果有暇有意, 可以简单介绍一下lebesgue测度和kolmogorov给概率论做的公理化基础。

wxmang: 显示不了符号, 柯尔莫哥洛夫定义是用集合符号定义的。勒贝格测度也是集合符号, 没法显示。

wxmang: 柯尔莫哥洛夫概率论公理假设我们有一个基础集 Ω , 其子集 X 为 σ 代数, 和一个给 X 的要素指定一个实数的函数 P 。 X 的要素是 Ω 的子集, 称为“事件”。设 Ω 为非空集合。满足以下条件的集合系 X 称为 Ω 上的一个 σ 代数: Ω 在 X 中; 如果一个集合 A 在 X 中, 那么它的补集 $-A$ 也在 X 中; 如果有若干个(甚至可数个)集合 A_1, A_2, \dots, A_n 都在 X 中, 那么它们的并集也在 X 中。也即设 Ω 为非空集合。则 Ω 上的一个 σ 代数是指一个关于可数并集运算封闭的集合代数。记号 (Ω, X) 称为一个可测空间。所以测度论和勒贝格积分的所有定理都能自然使用。对于确定的 σ 代数 X , Ω 的子集中属于 X 的称为可测集合。而在概率论中, 这些集合被称为随机事件。某个集合 X 上的 σ 代数又叫 σ 域, 是 X 的所有子集的集合(也就是幂集)的一个子集。这个子集满足对于可数个集合的并集运算和补集运算的封闭性(因此对于交集运算也是封闭的)。 σ 代数可以用来严格地定义所谓的“可测集合”, 是测度论的基础概念之一。第一公理对于任意一个集合 $E \in X$, 即对于任意的事件 $P(E) \in [0, 1]$ 。即, 任一事件的概率都可以用0到1区间上的一个实数来表示。第二公理 $P(\Omega)=1$ 。即, 整体样本集合中的某个基本事件发生的概率为1。也即在样本集合之外已经不存在基本事件了。第三公理任意两两不相交事件 E_1, E_2, \dots 的可数序列满足 $P(E_1 \cup E_2 \cup \dots) = \sum P(E_i)$ 。即, 不相交子集的并的事件集合的概率为那些子集的概率的和。这也被称为是 σ 可加性。如果存在子集间的重叠, 这一关系不成立。

wxmang: 某个事件 E 的概率 $P(E)$ 定义在所有可能基础事件的样本空间 Ω 时, 概率 P 必须满足柯尔莫哥洛夫公理。也即概率就是定义在样本空间的子集的 σ 代数上的一个测度, 那些子集为事件, 使得所有集的测度为1。

欧拉陀螺: 请教, 本忙总文中写的现代统计学两大支柱, 是不是指: 贝叶斯理论和显著性检验。分别对应贝叶斯派和频率派, 两派一直纠缠不清。我见到的多数大学教材都是采用频率派的理论框架, 夹杂者贝叶斯派的知识。我的理解两派争论的原因是决策者和决策对象, 主观和客观的矛盾。

wxmang: 是的, 这个问题实际上在柯尔莫哥洛夫用公理化定义概率就出现了, 概念其实只在实践中有用, 因为他基本只能处理离散事件, 连续性的问题, 频率派处理不了。而最优控制里面的随机问题, 都是连续性的。现在大学教科书大多用频率派的框架, 是因为简单, 不需要太多数学, 例如不需要懂实变函数, 泛函分析。其实高级的概率论问题, 频率派无法解决, 甚至根本看不懂。至于高等数理统计, 例如陈希孺先生写的, 就与频率派无关了。其实钟开来写的数理统计, 也与频率派无关。

wxmang: 更具体点来说, 由于采用的数学基础不同, 频率学派把分布参数当成一个固定值的数据, 而贝叶斯认为分布参数是随机变量。也即频率学派认为世界的本质是确定的, 而贝叶斯学派认为世界的本质是不确定的。其实就是频率派认为分布参数是客观存在, 不会改变, 虽然未知, 但却是固定值, 所以

频率派核心概念是似然函数。贝叶斯派则认为分布参数是随机值，因为人不是上帝，无法确定没有观察到的东西，所以是一个随机数，因此参数也可以有分布，所以贝叶斯派核心概念是后验分布。但是单纯从公式上来看，后验分布其实就是似然函数乘以先验分布再normalize一下使其积分到1。频率学派出现这种对分布参数的假设，是因为他们掌握的数学工具无法理解参数空间的所有细节，所以只能用最大似然和置信区间这样的方法来处理，争取尽可能确定唯一的真实参数。贝叶斯学派认为世界本质是随机的，所以是无法确定参数空间里的具体数值的，所以参数空间里的每个值都有可能是真实模型使用的值，区别只是概率不同而已。所以用先验分布和后验分布来处理参数空间上的每个值的概率。结果就是：如果后验分布是双峰的，频率学派就会选这两个峰当中较高的那一个对应的值作为他们的猜测，而贝叶斯学派则会同时报告这两个值，并给出对应的概率。贝叶斯学派上世纪刚开始时发展不快的原因是当时人类的计算能力不足，例如如何确定后验分布是需要大量计算的。而频率学派主要使用确定方法，所以在上个世纪大部分时间占据统计领域核心地位。但是计算机发展和抽样算法的进步，使得对于任何模型任何先验分布都可以有效地求出后验分布。因为贝叶斯派认为所有的参数都是随机变量，都有分布，因此可以使用一些基于采样的方法使得我们能够构建复杂模型，解决复杂问题。而频率派基于黑箱和猜测的做法，导致出现很多问题，例如Regina Nuzzo 2014年2月份发在Nature杂志Volume 506, Issue 7487上批判p值滥用的文章试图证明Fisher的显著性检验的内在逻辑特征及其局限性。认为真实参数除上帝，凡人是不知而知的；凡人对真实参数做出的所有零假设都是错误的，都是应该且能够被拒绝的。在样本量足够大的条件下，可以拒绝所有的零假设，总是可以得到 $p < 0.05$ 的结论；p值依赖于一个人为确定的、假想的、错误的但又难以证实难以证伪的零假设下某个统计量的概率分布；若不能拒绝零假设，也不能称“接受零假设”，因为零假设总是错误的，在增加样本量后总是可以拒绝的； $p < 0.05$ ，拒绝零假设，有统计学意义并不代表有实际意义。零假设和真实参数差距再小也是可以拒绝的；一次试验得到一个 $p < 0.05$ 的结果，拒绝了零假设，并不意味这样的结果在以后的试验中能够重复。贝叶斯学派现在扬眉吐气了。Fisher本人批判贝叶斯方法，但是他也离不开贝叶斯理论。

jianduanle：如果应用到工作中，就是说面对一个问题的时候，掌握的信息越多，越逼近事情的真相和问题的根源。我在工作中，遇到复杂问题时，一般的工作方法是尽可能多的了解相关的情况，先摸清体系，然后再梳理问题并考虑解决方案。一些年轻人习惯犯的方法有两种：一是上来就按照自己的直觉去干，生怕问题处理的不够快；二是只关注问题本身，只关注问题是否解决，不愿意或没想到去了解体系。

wxmang：贝叶斯给出的计算方法比概念重要，不但告诉你干什么，也告诉你你怎么干。这点与传统社会科学不同，社会科学不会告诉你怎么干。

猾心：这就是社会科学最软肋的地方，不知道怎么干。我个人理解是社会科学无法定量分析，没法精确计算给出路径。很多社会科学家以为弄条粗糙的公式就是定量分析。芒总，您觉得社会学、法学这些什么领域能够用到比较好的数学理论？博弈论？还是控制论？

wxmang：社会科学未来的前途我觉得不是使用定量工具。使用定量工具，无论如何比不过自然科学，他们需要的是思想的穿透力，洞察力，判断力，例如《易经》就没有什么数量工具概念，《黄帝内经》也没有，但是能够屹立几

千年不倒。现在这些社会科学家自己无能，思想没有穿透力，洞察力，判断力，只好往数量工具上靠，试图换点饭吃，就像走江湖的用《易经》的术数算命混饭吃一样，是走到邪路上去了。

落日：忙总的提法很有启发性！还让我想到了军事学里的兵棋。战争是一场非常复杂不可预测的活动，兵棋却提炼了很多战争经验的总结，比如每个算子包含攻击力、防御力、机动力、士气和补给等众多属性，战争中的偶然因素用扔骰子查裁决规则表来表示，加上基于历史背景和训练经验的众多行动规则，竟然也能模拟变化莫测的战场形势，军人可以利用兵棋推演来改善战术计划、训练战术素养。值得一提的是，它的模拟效果比单纯的军事运筹学和系统分析要强得多。我觉得这和易经挺类似的，都是提取本质经验建立推演模型，筮法和扔骰子也差不多，给用户一个训练学习的模型，并在学习训练中不断改进规则模型，类似贝叶斯推理一样。我觉得这才是社会科学未来发展的趋势。

必须勤力：最近在处理实际问题的时候，对社会科学有一点提炼。在社会科学研究中，我的观点始终是：定量分析适合去认识一个较为普遍的现象，但这种数据不能用来对应到个体身上。比如对某个地区内的群众的大致饮食习惯、偏好的调查，可以获得一个大致的数据，但不能因此推断出该地区内的某个人的特定的饮食习惯、偏好。某个教育理念、教育方法的有效性可以从一个群体内调研出来，但不能确定是否对某一个具体的教育对象有什么样的效果。所以，社会科学要落实到应用层面，是需要做定性分析的，而且更有研究价值，这需要研究者的综合智力非常高（洞察力、敏感性以及表达能力、策略的选择等等）。但现在社会科学的一个普遍特点是非常不自信，努力让自己变成“科学”，拼命套穿自然科学的研究工具的外衣，干什么都得编一些数字才显得有水平。最明显的学科是心理学，它自己看不上社会科学，选择站在自然科学的队伍里去，但关键人家自然科学也根本看不上它，而实际上它是不折不扣的社会科学，用文科生思维就可以搞定它（在我校号称国内第一的心理学，本科录取的是理科生，学位是理学学位，其实是1979年从我们教育学专业分出去的子学科，它的硕博士入学考试，我们教育学的文科生随便考考都能靠第一，心理学具体应用方面，文科生出身的并不比理科生背景的要弱，事实上还更好。而我们教育学的文科生大部分理科不好：记得非常简单的高数课，小一半人要重考，统计学什么的其实真的很基础，工具会用即可，但挂科的也真是不少，人体解剖生理学课及其实实验，他们基本上是当文科来背诵的）。

卖小号：有种说法，内经和伤寒论使用的是“定性”的方法。

wxmang：建议看我以前讨论《黄帝内经》的帖子，很详细。

wxmang：我体会陈先生在贝叶斯和频率派上采取中立态度（他其实是Fisher和Neyman的粉丝，他在波兰学习过），是因为他以前的世界观是确定性的，不能接受贝叶斯学派的世界是随机的观点。可是随着量子力学和复杂系统的发展，似乎世界的确是不确定性的，甚至复杂非线性系统里面严格的因果关系是不成立的，所以他又觉得贝叶斯不是一无可取。在他的书里，仍然是确定性的粉丝，不过比较中立和客观了。

阳小样：很像，波粒二象性和测不准原理

wxmang：与那个还真无关，就是一个对世界看法的问题。

欧拉陀螺：忙总，这么描述不确定性可以吗：量子力学和系统学所掌握的关于对象系统的规律是一种统计规律。有两种可能的解释：1它们的参数和物理量本来是确定的，只是目前人们只掌握到统计规律的层次；2它们的参数和物理量

根本没有确切值，按统计规律来取值是其根本特点。虽然不清楚不确定性的来源，但其统计规律本身是确定无疑的。

wxmang: 随机事件和不确定性事件是根本不同的。随机事件可以通过无限抽样逐步逼近了解其分布（大数定理），也即局部信息包含整体信息。而不确定性事件不能通过抽样确定分布，甚至可能根本没有稳定的分布，也即不确定性事件是没法通过抽样把握规律的，也即局部信息并不包含整体信息。概率论只研究随机事件。在复杂系统中，相当多类型的湍流，混沌等等都是不确定性事件，无法用概率分布描述。

欧拉陀螺: 忙总，那段是参考喀兴林先生的表达。他的思想是和稀泥。虽然深一层次的规律争论不休，但是避开这些问题，但是把建立一个唯象公理化的理论体系是可能的和恰当的。以后不论那种意见被证明是正确的，其公理化结构不会重大改变。当然我也蒙圈，工作中只是用到一些相关结论，目前涉及不到原理层次。上述表达是个人理解，肯定与喀先生的有误差。

wxmang: 这其实是一个哲学问题。我们知道哥德尔不完备定理，稍微推广一下，我们知道就知道一般公理系统都是不完备的，也即总有在公理系统以外的概念或事物，而公理系统定义的原则，就是我们认识的边界。根据这个原则，柯尔莫哥洛夫定义的概率公理系统可能也是不完备的，所以根据他的定义得到的结论，可能也是不完备的。根据柯尔莫哥洛夫公理系统对概率的定义，我们知道实际上随机事件组成的集合是一个勒贝格可测集，概率实际上是一个勒贝格可积函数。由于存在勒贝格不可测集合和勒贝格不可积函数，也即从逻辑上讲，存在不可通过局部信息了解全局信息的事件。所以严格讲，有限的知识，总是有边界的，可是边界外面是什么？目前我们从逻辑上讲就无法知道。

伯威: 原来学数学的也有这种解释现实的争论，真是没想到。原以为数学是抽象学科，是不是自然科学都得另说，怎么会有这种解释现实问题？只有像物理这种才有哥本哈根学派和爱因斯坦之争。话说当初刚看到玻尔那个氢原子模型也太粗糙了，这也弄得出一个学派？另外概率和统计要算应用数学，不算所谓的纯数学吧？

wxmang: 科大数理统计现在不在数学系，在管理学院。概率论算应用数学。

wxmang: 这可争论是不是如何解释现实，而是争论知识是否有边界和世界是否可认识这样核心的问题。

Cousteau: 贝叶斯理论的个人理解：研究对象的样本数不确定是有限还是无限的，而没有无限概念，就没有“连续性”和分析概念。您说的确定性，是否一般是指集合能够被准确定义，至少要求对任意一个元素都能够确切判断其是否属于此集合？对于大样本有限集合，其计算结果和无限样本的结果肯定不同，虽然集合乍看起来差不多。针对一个不确定是否有限的集合的定义，既不能一一列出，也不能用公式表述，只能描述其部分特性。贝叶斯的“理性”一部分就是它不预先假设研究对象是确切的，无限的，也就不假设其分布符合分析结果，但是依然保留“部分特性”。在结果预测时就会在确定频率的基础上考虑具体的因素，得出更具有针对性的结果。请指正

wxmang: 不是你理解的，这个问题讲清楚，需要很多篇幅，或者你去看陈希孺先生的《数理统计简史》，或者等我有时间，慢慢整理。因为在数理统计诸多分支中，情况看不完全相同。1950年前，数理统计学派是英国学派占主流，也即Fisher为代表这一派，他们特点是：坚定反对贝叶斯（因为他们是确定论世界观），反对Neyman为代表的数学化（当年Neyman就是因为看不惯英国学派的数学基础太差，民科水平才去的伯克利，结果通过严格数学

化数理统计的基础，把伯克利建立成为世界数理统计重镇和核心），他们强调实用，应用。，初等化，平民化，工具化，总之，就是数理统计是傻瓜工具，没有什么逻辑基础（不考虑，当年Neyman写的一篇置信区间严格定义的论文，被Fisher拒绝发表，但是这篇论文奠定了现代数理统计的逻辑基础）。所以我觉得讨论数理统计的基础，得从Neyman的工作及其以后数理统计的发展开始，Fisher的东西在我看来太不严谨，基本是民科的水平（能够用，但是严格追究概念和逻辑，就要抓瞎，因为他经常武断的给出一个东西，至于为什么，有什么逻辑基础，不知道）。这一部分工作，得看钟开来的书。需要很多时间。我暂时还没大块时间来整理，当年看过钟开来的书，也忘得差不多了。最糟糕是豆瓣不能展示数学公式。

wxmang：钟开莱的名言：我最讨厌统计学家。我想他的理解是：统计学就是民科吧。

cfms：大师钟开莱http://www.qstheory.cn/wz/renw/201112/t20111209_129030.htm (没忍住，还是粘一下)

Cousteau：谢谢，我没认真看过数理统计，有空一定补上。当初看的是王竹溪的热力学与统计力学的几本书，当时还是被王先生本身叙述的优美吸引的。不过我的一个主要问题就是您后来说的世界观上面的，就是您如何区分无限集合和较大有限集合呢？

wxmang：可穷举的就是有限的，不可穷举的，就是无限的，例如对任意 N ，总存在 $N+1$ ，就是无限的。

Cousteau?2016-01-22 18:23:05 可是现实中不能按穷举法判别...比如海里有多少条鱼，对于不同组织来说就不一样

wxmang：数学理解的无限与常识是不一样的。

Cousteau：是这样的。数学书本上的能够理解，自己帮忙写过的讲义（考研辅导兼职）也被认可。可是无限和连续统概念和现实中对对应起来总是隔着一层：以前看过庞加莱写的《最后的沉思》，不知道是翻译问题还是怎么回事看着也隔着一层。后来看毛主席同李政道对话录（有演义成分），感觉反倒清楚些。

wxmang：看来你的数学基础不够，主要是实变函数没学好。这个是初等数学和高等数学之间的门槛，你还没有跨过去。你说这些，在实变函数中，都能解决。

Cousteau：忙总批评。我不是科班出身，实变函数是课能讲、题能做、论文能看（大言不惭呵，纯粹为谋食性质在换工作间隙干过小半年），可是自问懂不懂我打怀疑。

wxmang：丘成桐就把理不理解在集合上定义抽象算法来判断是不是适合学数学。

wxmang：继续瞎扯。数理统计频率学派的数学基础是概率论的大数定理和中心极限定理，如果没有这个基础，直接就是民科裸奔（这可能就是钟开莱先生讨厌统计学家的原因，因为他们学科的基础太脆弱了）。这两类定理在柯尔莫哥洛夫公理化定义的概率前提下，是可以严格证明的，也即逻辑上没有漏洞。不过建立这个基础的工作不是坚定的反贝叶斯学派的Fisher完成的，而是Neyman。不过算是这样，数理统计的数学基础仍然很不完善，经常还是要加入人为因素才能处理问题，就不说贝叶斯学派的先验概率逻辑漏洞了：因为无法进行精确定义和度量，就算是频率学派的极大似然函数也要考人的经验参与，甚至数理统计的数学建模，都需要经验，这时经验其实跟算命的作用差不多-瞎蒙而已，因为经验并不能对从未见过或发生过的事情起作用，而数理

统计建模，却经常要这样做。Neyman曾师从于波兰谢尔品斯基等数学家，是假设检验的统计理论的创始人之一。他与K·皮尔逊的儿子E·S·皮尔逊合著《统计假设试验理论》，发展了假设检验的数学理论，把假设检验问题简化为一个最优化问题来处理，引进了检验功效函数的概念，以此作为判断检验程序好坏的标准，这种思想使统计推断理论变得非常明确。当时英国数理统计学派（以Fisher为主要代表）主流观点是：统计应该是为人民大众服务的通俗易懂的工具。而Neyman则认为纯数学应作为统计学理论基础，强调数学上的严格和美。Neyman的一篇长篇论文《基于经典概率论的统计估计理论纲要》，寄小皮尔森主编的《生物计量》，小皮尔森拒绝采用，寄给Fisher控制的刊物，也被拒绝。这时世界数理统计学界的主要刊物都被小皮尔森（接替其父亲老皮尔森）和Fisher控制。Neyman最后只能依靠Fisher的敌人，当时的贝叶斯学派领袖推荐，才得以在《皇家学会哲学会报》发表。所以当时就爆发了Neyman不满Fisher和皮尔森父子数学上的无能，认为搞的统计学没有数学基础支撑，纯属民科级别的冲突。所以当Neyman接到伯克利数学系主任Evans邀请后，就去了美国，成立了统计学界著名的伯克利统计实验室（我国的许宝禄先生也出身于这个实验室）。Evans当时想法是“找一个个性极强的中心人物来建立一个理论统计学学派”。二十年后，伯克利彻底取代了伦敦大学独霸世界统计学界龙头几十年的地位。从学派的分歧来看，英国数理统计学派的传统观点是统计学要面向实际通俗易操作，能为实际工作者直接使用，所以Fisher相信直觉（Fisher认为当 $N=1, 2, 3$ 如果都成立的话应该就没问题了）。而Neyman一直想为数理统计建立严格的数学基础，在Neyman领导下，统计学主流成为注重数学上的严密性，形式简洁但理论深奥。而且Neyman认为当时Fisher面临的问题是：从实际问题出发研究通用方法已经走投无路。

伯克利的J. W. Tukey 于1962年发表了《数据分析的未来》，基本代表了美国现代数理统计学派的主要思想：

- 1、任何学科想发展，都应当寻求全新的问题来研究；
- 2、应当在更现实的框架下去研究老问题；
- 3、不要为了因为数学能力不足而采用过于简化的模型或者以及与现实不符的假定；
- 4、应当寻求新的处理观测数据的方法，弄清楚其有用的性质，以备选用（不要老是均值方差这一套）；
- 5、不要给定参数的概率模型后去做参数估计，数据分析应该在确定概率模型前（也即不要先确定结论，再去找原因。频率学派的数理统计方法实际上能够做到：对任何给定的结果，找到对应的原因。基本属于马屁专业学科，只是批一个数学马甲）；
- 6、数据分析必须容许适当程度的错误，以使所获的不周全的证据可以经常启示一个正确的解法；
- 7、勿本末倒置，数据分析使用数学论证和结果，是以此作为一种根据，而非用于证明或方法合法性的印记；
- 8、如果只是在新增因素的基础上进行解的优化，就不必了；
- 9、多用简单的、易行易懂的方法；
- 10、稳健性重于有效性；
- 11、数理统计的重点是数据分析，是data investigation, 而不是传统的数据处理。

数理统计必须强调人的判断的重要性，对分析结果的解释采取不盲从而是有批判性的立场。也即Tukey把数理统计学工作分为两类，一类是对数据分析有贡

献的（而英国传统频率学派，大多数时候仅仅是数据处理而不是数据分析），另一类是纯数学工作。任何一件数理统计学工作必须从实用或纯数学中选其一，否则就毫无价值。陈希孺先生认为：统计学不只是一种方法或技术，还含有世界观的成分——它是看待世界上万事万物的一种方法，我们常讲某事从统计观点看如何如何，指的就是这个意思。同时，陈先生还认为，自从柯尔莫哥洛夫定义了概率论的公理体系，人们就可以抛开“概率究竟是什么”的问题，专注于概率在推理运算中该如何使用。

乌金沙：这个系列帖子很费精力和心力，我想这是您对研究生阶段及之前所学的一个回顾，当然加入了工作近30年的经验。机器学习对未来几十年的各类决策辅助，应该影响很大。归纳推理逻辑将走入一个前所未有的顶峰。

deepVar：现在伯克利也是计算机尤其是机器学习的顶级强校之一，有机器学习界的领袖级科学家迈克尔·乔丹。其实我觉得机器学习和统计几乎合流。开句玩笑，忙如果是这十年的学统计出身的话，弄不好也去美帝搞机器学习了。对现在的学生吸引力太大。估计忙总同学也有不少人在这个圈子。

wxmang：的确，现在机器学习基本就是贝叶斯学派的天下。不过我的同学搞这玩意不多，因为他们基本是反贝叶斯学派的，还是搞传统的假设检验，方差分析，回归分析的比较多，只有很少在搞参数估计（因为这是贝叶斯方法的强项，他们玩不过）。我如果继续搞数理统计，我会搞回归分析，这是当时热门：预测学的基础。

winternight39：请教忙总如何评价翁文波院士的国家周期表？感觉好神秘。

wxmang：不了解。

wxmang：目前机器学习核心就是贝叶斯决断。基本独霸天下了。google的搜索算法，在线翻译，都是贝叶斯决断。

Leon温陵：不完全是，google发家的搜索排序算法是pagerank算法，后面的数学是矩阵、线性代数和马尔科夫过程。当然现在是什么就不是很清楚了。

wxmang：我看的是前几年介绍资料，说他们搜索基础算法是马尔科夫链+贝叶斯决断。

乌金沙：好奇地问一句，是因为师门的缘故么？老师少搞或不搞贝叶斯？

wxmang：是的，当时国内没人搞贝叶斯，因为在文革时，贝叶斯被批判为资产阶级唯心论（同时倍批判的还有量子力学和相对论，如果能够找到当年工人批判组的文章，才会笑掉大牙）。所以大学里面禁止搞。

乌金沙：近日科大讯飞玩了一把语音现场转换为文字，效率远超速记员。不过，现在单兵突破已经到了顶点，需要多学科协同。假设有一个库，库里有近三十年全国各地级市（州、公署）政府工作报告全文，另外还有不是特别完全的统计数据库，需要摸索提出问题的话，需要建立不少假设，其中一些核心假设与贝叶斯方法可以完全无关。比如一个核心假设：工作报告是决策层中各种势力的一个妥协，一些低频率关键词代表决策层的某种意图。在这一连串流程中，贝叶斯方法仅仅只能用于搜索后端。也就是说通过人工对低频关键词进行筛选，但这些词是由地级市水平的决策者先筛选过的，这基本杜绝了极低水平的决策干扰（劣决策剔除，显而易见，县级政府极多劣决策）；然后通过各种方法包括回归分析来找一些关联，同时利用贝叶斯方法进行一些搜索动作（主要目的是机器替代人工，节省时间）。

猾心：继续延伸乱推论一下，以双盲实验为依据的西药学……属于这个范畴否？

wxmang：是的，这个双盲实验的科学性争论极大，我有一个同学在美国大医药公司负责设计新药的双盲实验方案，去年大半年，我们同班同学争论他

的工作是否科学还是蒙人，很多争论的都是数理统计教授，还有的是国际数理统计界的大腕。争论半天，公说公有理，我也没看出谁对谁错，不过倒是知道了双盲试验可能存在逻辑漏洞：因为未了实验可行和降低成本，必须对实验模型的因素和变量进行删减，而删谁不删谁，全看经验和已经知道的知识，而新药往往没有经验，也没有知识储备，这时删除，就是靠蒙或赌，也即可能删除的恰好是最关键的因素。所以说双盲实验就绝对科学，我觉得纯属扯淡。

猾心：这个推论万一有了定论，西药新药制度就很呵呵了。。。

[已注销]：双盲不是绝对科学，药物试验更多的是有目的的应用。方法自身的局限性都是清楚的，只是商人需要强调罢了。双盲的理论依据是因果推断理论，基于虚拟事实假设的一套完整的西方逻辑体系，忙总说的争吵其实是对现实结论和数字计算结果之间差距的争吵，而这个差距在非量子级别的体系水平上，以目前人类的智慧来说无法超越。

[已注销]：至于删谁不删谁，那也只是表象。美国生物统计学和因果推断是两个方向，就和数学和理论物理关系有点类似，关系极其密切，最终目的不完全相同。双盲实验的起点是频率学派，但是总有前导知识，总能正式地或者非正式地使用贝叶斯方法。但是贝叶斯法也是双刃剑，就像一把枪，看你怎么用了。

wxmang：我想说的是：双盲试验本质就是江湖蒙古大夫批一个科学马甲，在赌经验。当年葛兰素史克吃死人的文迪雅也是通过双盲试验的。在被美国FDA禁用之前，我也在吃，我运气好，没死。吃死多少人？这就是赌输了的例子。葛兰素史克在这个药上亏损20多亿美元，再加上赔偿死者的上百亿美元，双盲试验其实风险极大。

[已注销]：真是同一个意思。我说的也正是您在另一个回帖里说的认识论的问题。不过目前没有比双盲更好的办法，这是人类目前的认识水平决定的。物理学遇到过这个问题，因果性在量子层面引发无数讨论，但并没有最终结论。而在宏观层面，至少目前没有看到绕过去的方法。关于GSK文迪雅的问题，忙总过于武断了。逻辑上吃死人是必然的，只是需要衡量成本收益罢了。双盲试验提高揭示治疗效果的把握度的同时，也是降低了发现不良反应的能力，这是逻辑上的必然，也就是双刃剑的意思。现实的做法是知道这种方法的问题所在，而不是把大夫都说成是蒙古大夫。

wxmang：我其实想说的意思是：葛兰素史克把双盲试验当成程序正义了，认为只要做了双盲试验，就可以免死。当年复方丹参滴丸（天津天士力）去美国申请上市，葛兰素史克的人就说：这玩意没经过双盲试验，是假冒伪劣货。双盲试验不能避免病人死亡，同样，没经过双盲试验的复方丹参滴丸也未必就是坏蛋，我吃了15年，连续吃，每天吃，也没见什么问题。在法律上我们都知道，程序正义不能代替事实正义。葛兰素史克偏要这么做，其实算，双盲试验是西方医药集团打压中药的基本工具和武器。可是这个武器本身也是不靠谱的。一切争论都是利益在刀光剑影。每年中国几万亿的药品市场，让多少美国医药公司老板半夜梦醒，咬牙切齿。

[已注销]：这没办法。西方药厂之间用这个程序正义相互厮杀，而文迪雅恰恰就是他们互撕的好例子。当然天天吃也不证明什么，比如公鸡天天在太阳升起之前打鸣，但是却不是太阳升起的原因。

wxmang：你精神上太自虐了，你的说法相当于：中国就算成为世界第一，因为采用的是非民主体制，所以不能证明是对的；洋大人就算死翘翘了，但是因为采用的民主体制，所以不能证明是错的。这太自虐了。不好意思再说了，也不忍心再说了。

伊洛：忙总，文迪雅被重罚，主要是故意隐瞒不利的临床数据，而非主要因为副作用。一个药物上市后，还会经历四期临床试验、药物警戒等，确保安全、有效、质量可控。比如2000年上市、2010年退市的减肥药西布曲明（国内销售最好的是太极的曲美），退市的主要原因也是上市后双盲试验提示心血管疾病风险提升，并未罚款：“SCOUT（Sibutramine Cardiovascular Outcome Trial）名为“西布曲明心血管终点试验”，是由欧盟人用药品专家委员会批准，雅培公司开展的上市后研究。这项随机、双盲、安慰剂对照研究始于2002年，长达约6年，注册了约1万名55岁及以上、超重或肥胖、并已有心血管疾病史或II型糖尿病史、或还存在其他心血管疾病风险的受试者，用于评估使用西布曲明后体重下降对心血管的影响。SCOUT研究结果显示西布曲明3年平均减重仅为约2.7公斤，而由此引发的死亡和心血管事件（包括非致死性心梗，非致死性卒中，可复苏的心脏骤停，心血管死亡等）上升了约16%，减肥治疗的风险大于效益——这也直接促成了欧美随后对西布曲明的退市限令。”如何证明一个药物是安全、有效、质量可控？目前为止，双盲试验可能是现阶段最好的方法了，上市后还需要对不良反应持续监测，药监局有专门部门负责，国内2002年好像才开始建立。药品研发历史其实很短暂，监管其实是在一次次重大药品事故中逐步完善的。美国1938年“磺胺事件”推动了FDA监管法案的批准，要求①将所有药品成分公布②药品需做上市前审评，证明其安全性。1950年代末，沙利度胺上市时用于孕妇晨吐，最终发现对胎儿有致畸性，全世界一万名新生儿受影响，大多数成为没有四肢的“海豹儿”。FDA从此声名鹊起；严格临床试验及立法就此铺开。这款药物是德国公司Grünenthal研发，2010年赔偿5亿欧元给受害者。也正是这一事件，欧美60年代、70年代清理了一大批历史原因造成的无效药物批文。国内的药品批文数量是天量，主要形成于大跃进、文革时期，中央大幅向地方放权，省乃至地级市均可以批准药品。新药审批权直到1985年才收归中央。90年代大放权时代，又一大批药品获批。1998年全国食药监局才建立，郑筱萸任局长。02年“地标”转“国标”，本想清理这些垃圾批文，最终虎头蛇尾。新药审批也一直松动的很，直到05年郑筱萸被抓，07年被枪毙。现在市面上，很多药物如果追根溯源，其实都不太清白。现任药监局局长是毕井泉，曾担任国务院副秘书长配合岐山工作（后来好像配合汪），去年上任以来一系列举措还是深受业界欢迎，现在正在清理存量的申请上市批文，未来肯定要转向存量，拭目以待。

wxmang：我稍微多说一点。争夺评价标准这个不是政治，是国家地位和实力的标志，是中国国力发展一个必然的结果，不可避免。中国以前弱，与西方在标准上没冲突，因为我们一个是没资格抗争，第二个是没利益动力抗争，所以我们国家是大国中最没有独立自主标准的国家（不如印度、南非、巴西、俄罗斯等等），我们几乎所有质量标准，产业标准，甚至经济评价标准，都全盘接受西方的标准（这跟286实用主义的白猫黑猫指导理论有关）。可是当我们开始发展后，立即面临标准的利益冲突。产业标准上第一个大冲突是通信领域的无线通信标准，最后两败俱伤，不过中国最终在付出巨大代价后，还是建立了自己的3G和4G标准，虽然从经济上讲其实是得不偿失，但是从此在安全上不再受制于人，而且获得了产业进一步发展的空间，从长远来说是值得的。紧跟着的冲突就是产品质量标准。中国本来是完全融入ISO9000系列标准体系中的国家，可能比欧盟的国家（例如德国、法国等等）更彻底，因为我们国标=ISO9000+美国+德国+日本质量指标弱化一个等级。可是美欧日为了阻挡中国企业扩张市场，提出了很多自称为非关税壁垒的质量标准，例如欧盟的绿色标准，日本的节能标准，美国的劳工保护标准，这些标准都是为中国量身定做

的，所以出现了著名的轮胎反倾销，不锈钢反倾销，聚酯切片反倾销等等，中国受伤很重（例如山东广饶做出口轮胎生产的成片企业倒闭，中国是世界轮胎第一生产大国，广饶是中国轮胎第一生产大县，占全国生产能力20%，几乎全部出口）。这也教导我们，我们必须有自己的标准，不然就是单项被宰。所以开始制定自己的产业标准和质量标准，所以这几年这种被宰的情况少了，因为我们也有宰人的标准了，美欧日再想宰，就会出现对宰现象，互相放血，看谁顶不住。例如荷兰飞利浦曾经发起节能灯反倾销，结果被宰一次（商业贿赂和价格垄断），就老实了。医药标准也是这一系列标准中的一个。我们的西药，基本都是仿制药（自己独创的一只手就能数过来），所以西方标准，就是我们西药的标准，但是中国巨大的中成药市场（目前占国内药品市场规模接近50%，超过一万亿人民币）却是西方医药企业一直想用西药标准来打倒并占领的地盘，如果我们放弃自己独立的标准，中药将逐渐不复存在。但是中药如果按照西方标准，的确都将被淘汰。可是我们用几千年的人命测试出来的药，难道真的不如靠几个统计学家人为设计双盲试验测试出来的安全？再说双盲试验还是其他实验，逻辑上也是难以保证万无一失的。不能用双盲试验当成免死金牌，只要双盲试验过关，即使药品出问题，也是上帝的事情；而中国人的东西，即使不出问题，也是必须淘汰的，因为你的程序不正义。这是典型的双重标准。我们必须争夺这种标准。其实自从我们调查葛兰素史克贿赂案开始（四个中国区高管判刑，20人被拘留），就一直在争夺，最近葛兰素史克又把五个中国人以偷窃知识机密告上美国法庭为反击（中国也许会对其他西方大的医药企业一起动手例如阿斯利康、罗氏、辉瑞、拜耳、赛诺菲和礼来等等都进入了调查程序）。当然另外一个更大的标准，那就是大宗商品定价标准，这个我们还在努力，估计很快上海就要推出石油期货交易所了。铜的定价权目前正在争夺中。

王金沙：我国铜产量占全球占比太低，铜定价权要抢过来不容易啊。中期内打成个均势，让铜价尽量低位运行，这倒是很符合国家利益。

wxmang：现在利用实际用量+美元储备两个工具，可以逼空。其实正在在做。有的玩家被逼交现货，已经认赔出场。

王金沙：有点像公司法里面的否决权，34%则可。如果这么说，乐观点，十年之内，大宗商品主导权基本可以收入囊中。

wxmang：全部大宗不太可能，我们稀缺的一定要控制。

弈理指归：嘉能可之前就是一贯玩霸盘哄抬价格的，这次相当够呛，破产都有可能，还有来宝，农业部门要卖给中粮了。

wxmang：这次不倒下几个大家伙，这些投机商真的都不知道自己姓什么了。

然后203：高盛从15年年中转向唱多，似乎是看到这个苗头了。2015年美国多个著名对冲基金也是颜面扫地。

wxmang：高盛跟中国政府的关系非外界所能想象。

土豆烧牛肉：铁矿石这块有计划么？前些年可是被老外狠狠的宰了刀，当然，国内这边也不争气。说到正义标准，似乎就普通民众而言更看重结果，估计也是经历得多了，口号也听得多了。其实西方国家现在投票率低也有这个因素。

wxmang：铁矿石主要是内奸，某个机构想自己盈利，当然最后是自己赚一块钱，让洋人赚10块钱。我想纪检委应该知道怎么回事，就看有没有胆量动这块。

deepVar: 贝叶斯理论非常核心。个人觉得机器学习的两个主要理论基础是PAC学习理论和概率图模型。其中概率图包括贝叶斯网和马尔科夫网，即有向图和无向图。实际中，使用频率观点求参，还是贝叶斯对参数积分，往往由问题的复杂度决定。现在即使由于计算机发展，采样变得容易，但很多情况下仍然困难。另外，曾经问过一个大统计的学生，感觉其学科体系没有包括统计学习的内容，也许有些老化。但科大的学生在机器学习领域还是很厉害的。

wxmang: 我讲的情况仅限于78级数学系数理统计专业。现在情况不是这样，现在科大数理统计不在数学系，而在管理学院，现在他们主流是贝叶斯了。

罗阿宝: 忙总，建议您单独发一个主题帖，让更多的人看见上面的内容，这里面水太深了，难怪很多人都看不清，各论坛上中医西医的辩论，基本上都是秀逻辑秀智商，很少有人从争夺评价标准这个层面去思考。

wxmang: 其实西方与我们争论什么民主，自由，平等，科学，其实最后一脱干净，都是争的钱。如果中国还是民国初年军阀割据时代，绝对没西方人跟我们讨论什么民主自由平等博爱科学，他们唯恐你不乱，乱才能多卖军火，多赚钱。我最近看到一些当年西方公司卖军火给贵州军阀的原始文件，那时不跟我们讲什么民主博爱了，教会的牧师神父当担保，抵押是贵州汞矿40年开采权（现在是世界储量第三的汞矿，中国第一）。

弈理指归: 民主博爱这些玩意可以称作洋儒学，也是很有忽悠能力的，我看现在各年龄层的不少人都吃这套。

wxmang: 其实还是不自信+无知导致的。读书太少。

欧拉陀螺: 多谢忙总点播，收益良多。频率派的文章经常被批判使用贝叶斯理论。没有接触过高等数理统计，所以对两者的区别理解的不透彻。看到他们吵架都吵到nature级的杂志了<http://www.nature.com/news/scientific-method-statistical-errors-1.14700>

wxmang: 实话实说，我其实也有点蒙圈，也是一家之言，陈希孺先生其实也是蒙圈的，当年在科大上课时，对这个问题也是说不清道不明的（可以参考他写的：《数理统计学简史》，2002年，湖南教育出版社），陈先生算是中国数理统计学界的泰斗了。所以说这个问题各有各的说法，数学家说数学家的观点，统计学家说统计学家的观点，谁对谁错其实无所谓，只要解决问题就行。就现在企业质量控制、风险管理和设计等等实用场合，主要还是贝叶斯方法多，fisher的方法在保险或医药行业用的多。我觉得该用什么用什么吧。

欧拉陀螺: 该用什么用什么，很赞同。春节假期看看陈先生的讲解。

吃狼的绵羊: 感觉陈希孺也是看低贝叶斯派的，虽然没有明说，但是他在这本书中，是这样描述的：他们“自称”为贝叶斯统计。”所谓的”贝叶斯统计派“号称”xxx。。。用自称，所谓，号称这几个词感觉就是得不到他的认可

wxmang: 陈希孺的世界观不能接受世界是不确定的，所以不能接受贝叶斯学派。

longshen: 个人认为，时间序列比贝叶斯用起来更方便些。

wxmang: 时间序列只包括系统的历史信息，不包括观察者的先验信息，所以只能预测系统惯性，对系统突变无能为力，但是贝叶斯方法由于包括了观测者的先验信息，能够处理突变问题。这种例子很多。在一个非线性系统处于不

稳定环境中时，贝叶斯方法的效果好于时间序列。股市就可以证明这点。

kdzn: 股市里的江恩是时间序列，缠论是贝叶斯技术，垃圾邮件的判别基本上是贝叶斯加正向判断:关键字，邮件指纹

wxmang: 机器学习本质都是贝叶斯的。这次这个下围棋的狗，也是基于经验积累和分析算法的。

snoweagle: 这次狗的表现可以说是震惊了世人，没有人能想到在围棋这种复杂度的项目上面，机器学习进展这么快，联想到3个月前的那场比赛，可以说是指数级的提高速度

伯威: 第4局李世石中盘翻盘了，刚才乐视的直播室好欢乐啊，柯洁（暂居世界排名第一）哈哈大笑。网络直播比电视直播喜感多了。AI还是有点不靠谱啊，不过应该是工程问题，本质上AI还是胜了。

nt4587: 不一定，如果它不能处理这么复杂的局面的话，那么对人的胜率就会很低了。要不是小李没拿到过go以前的棋谱，它的弱点可能早就暴露出来了。我觉得他们的算法是有缺陷的。

伯威: 仅仅是猜测，猜测。AI的算法显然是有缺陷的，不然今天就不会输了。可能这个缺陷是可修复的，可以通过打补丁解决。不属于像大众的DSG变速箱那样，是天生的设计缺陷（大众这个评论俺是听说的，其实不懂这个）。

wxmang: 算法没问题，只是学习背景需要增加而已。机器学习是渐进过程，效率比人学习效率还高。

daze: 机器能用算法出感情吗，搞出所谓的情感心理医生机器人？

wxmang: 人工智能定义不是这样定义的，我专门写一个帖子介绍一下。

napolean: 那忙总您怎么看ai未来的发展，会在哪些领域最有可能对人形成替代

wxmang: 这个无法判断，原则上不需要创造概念，范式和算法的领域，机器都可以替代。