

# 第三堂课

## 信息的统计理论

### 摘要：

- ◆ 信息理论：概率论部分。严格逻辑和概率论逻辑的关系。凯恩斯对于概率论的理解。举例说明逻辑与严格的经典力学以及统计力学之间的关系。量子力学中的相应情况。
- ◆ 从数学角度看待严格逻辑向统计逻辑的转变。分析和组合论。
- ◆ 热力学的角度：信息和熵。
- ◆ 西拉德（Szilard）的理论。
- ◆ 香农的信息论。
- ◆ 叙述计算机的内部平衡的热力学性质。

### 一、自动机的鲁棒性

到此，关于信息的严格问题讨论已经告一段落，我们将继续从统计角度来讨论信息的本质。至少有两个原因可以说明统计和概率问题对于自动机和其功能实现是十分重要的：第一个原因可能显得有些任意且离题，虽然我并不这样认为。第二个原因则更加重要，下面我分别加以说明。

第一个原因是：实际上我们无法设想一台绝对可靠的自动机。假如你设计了一台自动机，并且严格地定义了它在任何情景下的全部行为。那么你一定忽略了问题的某些重要方面。如果你是一个新手，那么设计一台可以运行在完全确定环境下的自动机是一种很好的练习。但是只要稍具实际经验，我们就会知道这一步还仅是问题的最初阶段。

我们必须考虑统计因素的第二个原因是：如果你观察一台人造的，或者存在于自然界中的自动机，你会发现那些被严格程序所控制的仅仅是一些细节结构。大部分的控制是以一种允许错误，并且在错误发生时采取补救措施（多少有效）的方式来实现的。而且，说它们能够预防失误还有些夸大，因为这种机制其实根本就不可能消除所有错误，而是实现了一种发生个别的失误根本无关紧要的容错状态。在这种机制下，无论是错误还是失误带来的后果，都不能被彻底消除。我们可以努力去做到的事情，就是设计一台自动机，让它在遇到通常错误后仍然可以照常工作。这种设计的目的是减小错误的影响，而不是去消除错误。实际上，大多数常见自动机的构造和设计思想，都是属于这类容错型的。

为了允许错误作为一种独立的逻辑对象存在，我们不应该再以严格的方式表述公理，也就是说，公理不应该写成：“如果 A 和 B 发生，C 就会发生”这样的形式；而是“如果 A 和 B 发生，一定的概率下 C 会发生，也有一定的概率 D 会发生，等等”的形式。换句话说，每种给定情况下，都会有不同的结果，各自以不同的概率发生。从数学上说，我们可以简单地写出一个概率矩阵，说明各种状态之间发生转换的对应概率为何。你可以把问题写成这样“如果 A 和 B 已经发生，接下去发生 C 的概率有多大？”。这个概率矩阵就给出了一套以概率表示的逻辑系统。无论人工还是自然自动机，只要牵涉概率，都应该放到这个框架下研

究【原可参见冯纽曼的论文：《概然逻辑：用不可靠的元件组装可靠的系统(Probabilistic Logics and the Synthesis of Reliable Organs from Unreliable Components)》】。我接下来会谈及为何遇到复杂系统时候，我们就必须放弃严格逻辑而改用概然逻辑系统的原因【假如单个元件出故障的概率固定，那么自动机越是复杂系统崩溃的可能性也越大】

## 二、概率作为逻辑的扩展

上述原因使我们不得不把概率逻辑看作一般意义上的严格逻辑的一种扩展。这种把概率本身看作是逻辑的一种扩展的做法既非显而易见，也不为科学界所广泛接受，同时也远离了对概率的主流解释，但它却正是概率的经典解释之一。与此相对的是概率的频率解释，即认为逻辑本身是绝对严格的，但对一个我们不完全了解的现象，我们只能用出现频率大小来描述之。

我认为，这两种解释的区别，至少在拉普拉斯看来，是十分清楚的，他曾指出存在两种不同的方式来看待概率：频率和逻辑方式【《A philosophical Essay on Probabilities》】。在近代，经济学家凯恩斯也曾撰述概率方面的论文【《A Treatise on Probability》】，强调了两者的区别，并以此作为他的理论基础。凯恩斯相当详细地分析了概率问题，并说明了除了传统的频率解释以外，还可从逻辑的角度来解释概率。但他并没有试图把严格逻辑和概率区分开来，仅仅提到，如果你观察一个事件 A 和 B 的序列，这个序列可以用一个具体的量“B 紧跟着 A 发生的概率”来刻画。这里同严格逻辑的唯一联系之处就是如果此概率等于 1，你可以说 A 导致 B 的蕴涵关系，如果概率是 0，那么 A 的发生就排除了 B 的发生。但是，当这个概率接近 0 或者 1 的时候，你还是可以用一种模糊的方式进行推理<sup>1</sup>。

不可否认，逻辑分析的立场有其固有弊病，有时候，零概率事件明明在发生，否认其存在是很荒谬的<sup>2</sup>。同时，我们也不清楚在何种意义上，小概率事件表示我们可以认为这件事根本不会发生。尽管如此，凯恩斯还是提出了一种自洽的概率体系<sup>3</sup>。现代科学的其他研究，如量子力学中的发现，使得我们很倾向于接受这种自洽式的概率定义，尽管概率的本质现在远远还没有定论，而且可能长期得不到结论。无论如何，在量子力学中，我们倾向于改变传统的逻辑定义，而把概率看成逻辑的内在组成。

【在冯纽曼的著作《量子逻辑（严格的与概然的逻辑）》中，他做出结论：“概然逻辑是无法还原成严格逻辑的组成的，而应该被看成比后者本质上更广义的系统，量子力学中的概率形式  $P(a,b)=\Phi(0<\Phi<1)$  本身应该被看成自立自足的物理现实。

所以看起来概率逻辑应该是严格逻辑的重要延伸。这个观念，即所谓“逻辑概率论”构成了凯恩斯这方面的研究基础。”

可参阅冯纽曼和 Birkhoff 合著《The Logic of Quantum Mechanics》，以及冯纽曼同摩根斯坦合著的《博弈论和经济行为》3.3.3 节】

---

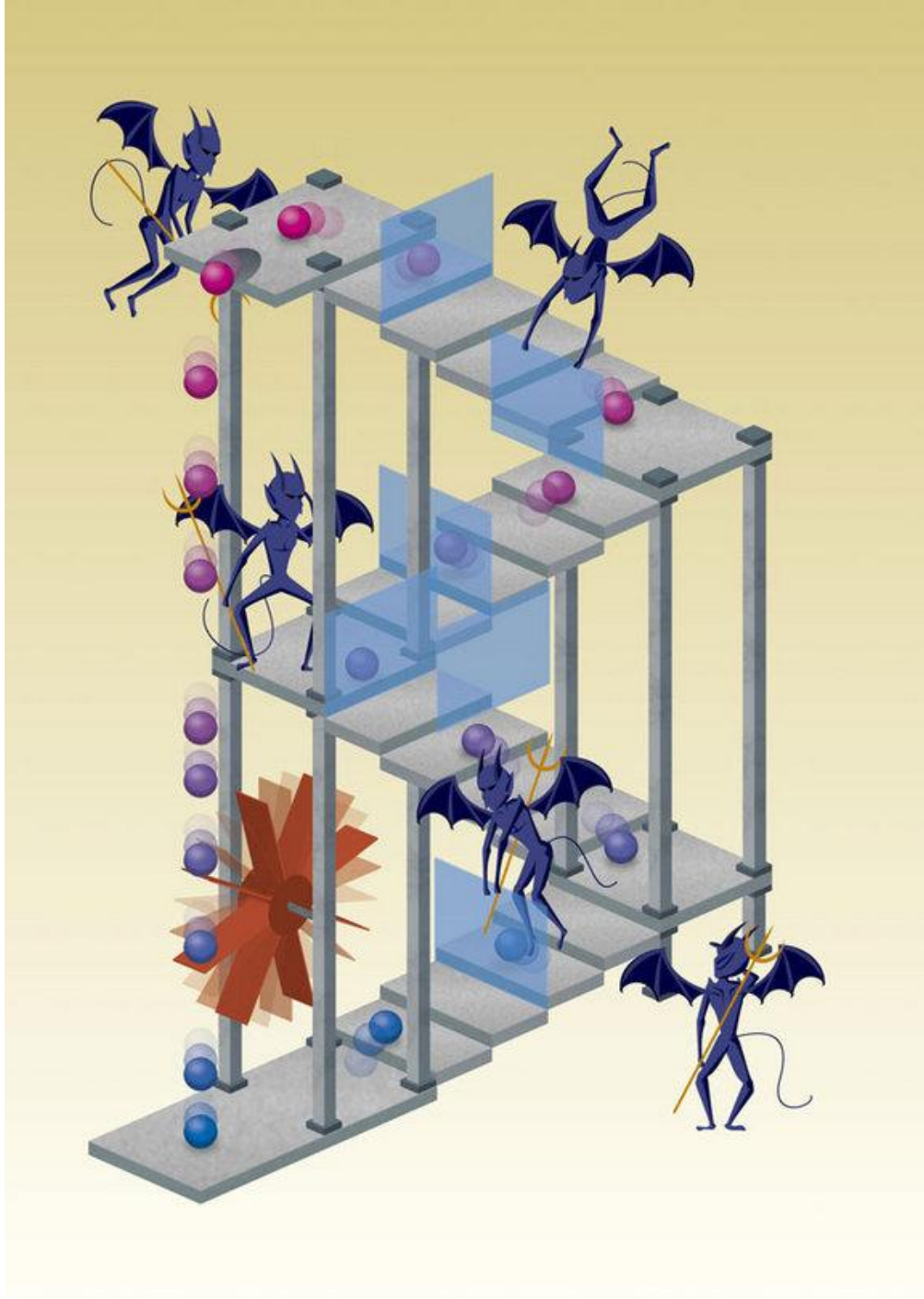
<sup>1</sup>译者按：当时模糊逻辑还没有被发明，要等到 1965 年俄裔数学家 Zadeh 提出模糊集之后，模糊数学和模糊控制才逐渐大行于世。冯纽曼在这里，已经明显暗示了这方面的学术机会。只是因为同主题不是太有关系，他没有深入分析。

<sup>2</sup>译者注：对于连续性随机变量，单个具体点的概率密度值为一有界常数，这个值可以是任意的，但因为点是没有长度的，所以该点的概率测度为 0，即该点所对应的事件发生的概率为 0，但这个事件仍然是可能发生的，因为这个事件在事件域内。也就是说，概率为 0 的事件并不一定不会发生。

<sup>3</sup>译者按：凯恩斯是主观概率学派的代表人物。他认为概率是无法客观定义的。这是因为从某种意义上说，概率的原因就是概率本身。人们对一件事情具有某种看法，这种看法本身就会导致事情如此发生。故概率是一种自洽的体系——以上引自凯恩斯《概率论》前言。

### 三、熵

【冯纽曼接着谈到两种不同的概率和“与它们相关的（不严格地说）”信息定义，一是热力学中的熵和信息；二是香农的信息论】



麦克斯韦妖（译者加）

图片来源：<http://www.livescience.com/8944-maxwell-demon-converts-information-energy.html>  
在熵和信息的关系这方面，冯纽曼在课上提到了玻尔兹曼，Hartley 和 Szilard。他详细

解释了麦克斯韦妖佯谬是怎么回事，Szilard 又是怎样用信息的角度来解释这个问题的<sup>4</sup>。

冯纽曼说，香农的信息论是一种用来测量信道传输信息能力的计量理论。他介绍了冗余度的概念，并指出通过冗余信息，我们可以实现纠错；这就像编辑校对原稿时候，不用询问作者，就可以改正拼写等常见错误的道理一样。冗余度是“使得我们可能书写较长文章（比如 10 页以上）的唯一原因”。换句话说，如果一种语言被压缩到了完全没有冗余的程度，那么它也不适用于传输较复杂的信息，因为你完全没有办法知道其中的文字的对错。这是一个原则问题，因此，我们使用的信息媒介的复杂度，和其传输信息所需要的冗余度之间，是有联系的。

在给维纳的《控制论》写的评论中，冯纽曼更广泛地阐述了熵和信息的关系，兹引用如下“对于物理学家来说，熵是属于热力学中间的一个概念，它牵涉到不同能量形式之间的转换。众所周知，封闭系统中的能量是守恒的，能量既不能被创生也不能被消灭，而只能在不同形式中间进行转换；这就是热力学第一定律（能量定律）。但是除此以外，还有关于热力学的第二定律（熵的定律）：能量存在不同的层次，机械能（包括动能和势能）的层次最高，接下来是以温度高低作为层次标志的热能，其他的能量也可以用类似的方式区分级别的高低。热二定律是说，能量的级别总是不断下降的，也即高级能量总是自发地转变成低级能量，或者，即使系统局部发生了能量的升级，系统其余的部分也必然要发生能量降级来抵消这种变化。如果我们要把这种变化记录下来，那么就可以用一个严格定义的物理量来表示能量的级别变化，这也就是熵。

“十九世纪中叶，物理学界就已经知道热力学和熵的关系，怎样用热力学方法来测量熵的变化等等。例如在玻尔兹曼早期的统计物理著作中，他就提到熵是和信息紧密联系在一起的。玻尔兹曼发现了，对于一个宏观信息（即人类可以直接观测的尺度）已经得到测量的系统来说，系统的熵正比于其微观状态数的对数。换句话说，熵正比于未知信息的对数。之后，许多科学家又在不同领域中间进一步发挥了这个思想，例如 Nyquist 和 Hartley 对于信道通信方面的研究，Szilard 在物理学和信息之间的关系的研究，以量子力学以及基本粒子物理学的评论者（(Mathematical Foundations of Quantum Mechanics, Berlin, 1932, Chapter V)）。

“我建议具有一定知识背景的读者参阅一些其他的文献，包括上面提到过的 Szilard 的著作，他对麦克斯韦妖做了相当清晰的分析。还应该研究一下香农最近的重要著作《信息论》，其中包括了信息的定义、人工语言、编码体系等等问题（Bell System Technical Journal, Vol.27,1948）。我们有理由认为，正像能量的降级伴随热力学熵的升高一样，信息熵也有类似的降级性质。因此，在热力学和我们正在讨论的严格逻辑的扩展之间，很可能存在某种深刻的联系。”<sup>5</sup>

---

<sup>4</sup>译者按：Szilard 认为，这个佯谬的关键在于：麦克斯韦妖判断一个分子的速度是快还是慢也需要能量。这个二进制的判定相当于一个比特的信息。他计算出，室温下存储一个比特的信息至少需要消耗三十万兆分之一焦耳的功。这个数目虽然不大，但足以保证热力学第二定律不会被打破。该理论中存在一个有趣的暗示，就是说信息本身就是一种能量，这意味着，如果小妖的确可以“免费”地获得信息的话，它也就获得了能量，因此是可以打破热二定律的。但直到 2010 年底，东京大学的研究人员在实验中首次成功将信息转化为自由能，“诱骗”粒子爬上电场制造的微小旋梯，这一貌似不可思议的想法才得到验证。实验中的粒子具有热能，在布朗运动下，原本可向任意方向移动，但研究人员用高速摄像机监视粒子的运动，当粒子沿旋梯向上移动时，允许它们自由移动，但当粒子试图向下移动时，就用电场创造的虚拟墙阻止它们，换言之，虚拟墙充当了“麦克斯韦妖”。只允许粒子朝一个方向移动，而粒子爬上阶梯完全由“自己的位置”这一信息所决定，无需任何外力。当粒子爬上旋梯时，就从移动到具有更高势能的地方获得了能量，研究人员便能够精确测量出有多少能量是由信息转换而来的。要说明的是，这个实验并非否定了热二定律，因为作为整体，并没有产生自由能，熵仍然是增加的。它只是说明了，通过传送信息而非能量，同样可以在局部改变熵的大小。这意味，信息、物质、能量三者是可以相互转换的。

<sup>5</sup>译者注：关于热力学和信息论、物理过程和逻辑演算之间究竟存在何种联系，到今天仍是一个众说纷纭的混乱领域。我认为关键之处有两点，一是热力学方程是以何种方式导致秩序的出现；二是这种秩序，用信息论的角度应该怎样刻画，它又如何反过来去控制热力学的结构？推荐普利高津所著《从存在到演化》，这本经典著作较好地阐述了相关知识。

冯纽曼接着提到了 Hamming 校验和纠错编码的概念。他说明数字系统（二进制、十进制等）是信息理论的一种应用，用数字化编码的漂亮方法，可以从低精度的组件得到非常高的精度。如果用 30 个元件代表 30 个二进制数字，其中每个元件都很粗糙，仅仅能分别有和无的两种状态（元件状态参数在 10% 范围内浮动），我们却可以用高达 10 亿分之一的精度来表达一个数字。除了数字化编码以外，我们不知道有任何其他的方法能做到这点。当然，从信息论角度说，这是理所当然的事，因为 30 个数字元件的熵是 30 单位，而 10 亿分之一的信息熵等于 10 亿以 2 为底的对数，也就是 30 个单位。

他还指出，生物传输信息用的规则类似于混合模拟电路，而从来没有用数字编码来表示信息（据我们所知），神经系统传输一个数字的时候，实质上是用一种频率调制的方式<sup>6</sup>，而不是用二进制编码表示大小。这可能是因为频率调制要比数字编码更加可靠的缘故。

## 四、未来的信息论

奇怪的是，虽然自动机的研究很需要这种新型的信息论，实际上却不存在这样的理论。只有在形式逻辑和热力学中间，才有一点点类似的蛛丝马迹。一种关于信息的理论跟逻辑有关是理所当然的，但是它同热力学也大有关系，却是令人惊讶，亟需解释的事了。

虽然这种新的信息理论在很多方面都同形式逻辑类似，它可能反而比较接近于一般的数学而非逻辑。原因是今天的形式逻辑与分析数学相差很远，逻辑所处理的是 1 或 0、有或无这样绝对的东西，这同数学最成熟的领域如数学分析的关系不大，但却与组合数学有些联系，但这是我们所知很少的数学领域。因此，我们认为相比于传统逻辑，这种新的概率逻辑和传统数学，尤其是和数学分析的关系要紧密的多，这是因为概率逻辑中间所有的公理都是用概率来表示的，而不是传统逻辑中的死板形式。在量子力学的概率基础之中，我们已经观察到了这种现象<sup>7</sup>。

热力学的概念也很可能会被吸纳到我们的新的信息理论中间来。种种迹象都说明，信息对应于熵，而熵增过程则对应于信息处理过程中的损耗。当然，两者之间也会存在区别，例如一台自动机的功能和效率不太可能用热力学中的那些统计量直接表达出来，自动机的统计参数一定会牵涉到一些温度以外的因素。但性质上，信息和热力学大致是类似的<sup>8</sup>。

此外，通过制造第一台计算机的实践，我们清楚地认识到计算机的关键性质牵涉到平衡。这包括不同组件之间的平衡、一个组件的速度和其他组件的大小之间的平衡，甚至是两个组件的速度之比和其他组件大小的平衡等等。在之前的第一堂课上，我在讲分级内存的时候提到过这点。这就像热力学中间，为了达到最高效率，系统必须平衡其中的各个组件的道理一样<sup>9</sup>。

如果自动机中间某一个部件太快了，其他部件跟不上；或者内存太小了；或者缓存和内存的速度比例设置的太大，缓存不够用；这些情况都会导致效率损失，就好象一台热机因为

---

<sup>6</sup>译者注：即以频率的高低，也即单位时间内神经脉冲的数量表示信号强度

<sup>7</sup>译者注：冯纽曼英年早逝，在创立概率逻辑，把概率和实际现象相结合的这项工作上，没有更多的时间继续下去。此后概率逻辑的发展很慢，直到最近，才逐渐发现，量子力学和量子概率也可以适用于现实世界。但相关理论还远没有建立，这也是我们的主要工作方向之一。

<sup>8</sup>译者注：自动机的功能和效率与它所工作的环境有关，我们必须通过环境的一些统计特征来刻画自动机的运行，就像我们在热力学中刻画系统所在环境的统计特征方法类似。

<sup>9</sup>译者注：热力学均衡和此处的“信息平衡”具有巨大的类似性。如理想卡诺机不损失能量，每一个过程都是可逆的。这就要求每一步都是准静态过程，亦即处于平衡状态的边缘。要之，平衡则熵不增，不平衡则熵增。这对应于信息处理的可逆性，如果用精巧的机器来实现信息的处理的话，信息的处理、运输和储存都是可逆的，只有在合并信息的时候，如把两个比特串进行异或运算，这时候之前的信息是无法取回的，就会导致不可逆的信息减少，这对应于热力学熵增过程。至于两者之间的关系究竟如何，这还是两门科学的前沿领域，尚无较清楚的理论体系。

部件之间的温差太大导致能量白白浪费的性质一样。这些我就不细讲了，总之，我相信热力学同自动机的联系是非常紧密的。

## Jake 点评

这一章虽然比较短，但是它涵盖的信息量却非常大，实则横跨了至少三个非常不同的学科。因此，阅读这一章，也许你会觉得冯纽曼故弄玄虚、言之无物。然而，这一章的意义非同小可，因为它几乎指明了复杂性科学未来几十年的发展方向。就拿我自己的探索经历来说，大概在 2004 年左右的时候，我就读过了这一章，然而，当时的感觉就是头晕眼花，不知所云。于是，把这本书撂倒了一边。有趣的是，大概 4 年的时间过去之后，偶然间读到了 E.T. Jaynes 对统计物理的新见解，使我对热力学、熵等概念发生了强烈的兴趣。之后，当我再次翻看了冯纽曼的这一章节的时候，我眼睛一亮，冯纽曼早就指出了主观概率、信息、热力学等等之间的联系。又过去 2 年的时间，当我认识到要理解概率的本质必须要了解量子力学的时候，我又一次看到冯纽曼在这一章里早就表达过此类观点了。

我已分不清 2004 年的阅读是否已经潜移默化地被冯纽曼影响了；还是我的探索轨迹恰巧与冯纽曼的思路一致？总而言之，我与这一章从某种层面上说的确发生了共振，以至于我对这短小精湛的一章越来越喜欢了。下面，我将分别从概率、热力学和信息论这三个层面综述一下相对重要的文献和进展。

### 一、概率

虽然概率论最早发源于 17 世纪，然而将近 400 年过去了，人们对于概率本质的争论却仍然没有停息。概率论从一开始就分成客观概率的频率学派以及主观概率的贝叶斯学派。主流观点是将概率解释为某事件发生的频率，它是事物发生的一种客观属性。然而，仍然有很多人坚持认为概率存在于人们的主观意识中。例如，当我们说明天将会下雨的概率是 50% 的时候，实际上我们表达了一种对明天是否会下雨的信念，而并不是说我们能够做 100 次重复试验，其中有 50 次明天下雨。那么我们是怎么得出明天下雨的概率是 50% 的呢？实际上，我们是根据今天的天气情况（例如是否阴天），以及以往的观测纪录（例如最近几天是否处于雨季）等等已有的信息，做出了明天下雨这一命题的不确定性推断。所以，明天下雨的概率 50% 并不是一个客观存在的概念，而是一种满足概率贝叶斯法则的推断。



Monty Hall 问题

也许你仍然认为这种认识只是看待概率的不同哲学观点，没有实质上的区别。那么，请看一个著名的决策问题的例子：Monty Hall 问题。Monty Hall 是美国一款著名娱乐节目的主持人。他曾提出了一个非常有趣的决策问题。假如舞台上三扇关闭的门，已知这三扇门后

面分别是羊、羊、汽车，主持人 Monty Hall 知道每扇门后面的情况，但是作为决策者的你并不知道。

现在，让你来选择一扇门，如果该门后面是汽车，那么你就赢得了大奖，把汽车开回家，如果是羊，那么你什么都得不到。问题有趣的地方在于，当你做出一个选择的时候，比如你选了 3 号门，Monty Hall 并不把 3 号门打开，而是打开了另一扇后面是羊的门给你看（因为有 2 只羊，因此，他总能找到一个不是你的选择，同时又是羊的门，比如 1 号门）。现在，主持人 Monty 问你，你还要不要坚持你的选择 3 号门呢？

这是一个典型的概率问题。按照常理思维，3 号门背后是汽车的概率是 1/3。尽管主持人 Monty 打开 1 号门给你看那是一只山羊，这仍然不会改变 3 号门背后是汽车的概率（因为概率是客观存在的，固定死的数字）。但是，与直觉相悖的是，真正最优的选择恰恰是改选 2 号门，你赢得汽车大奖的概率会更高。

只有当你认为概率并非一个事件所固有的客观属性，而是观测者判断某事情发生的主观信念的时候，概率分布才会因为主持人给你看了 1 号门背后的山羊这个事件而发生改变。事实上，根据贝叶斯原理，我们可以计算出新的概率分布。

我们记 A 事件为 3 号门背后是汽车，B 事件是 2 号门背后是汽车，C 表示 1 号门背后是汽车这个事件，X 表示主持人 Monty 打开了 1 号门的事件。这样，我们知道  $P(A)=P(B)=P(C)=1/3$ 。问题的难点在于 X 的概率计算，X 发生的前提有 3 种可能性：（1）如果 A 事件发生，即 3 号门背后是汽车，则主持人可能打开 1 号门，也可能打开 2 号门，这样 X 在 A 发生的条件下发生的概率是  $P(X|A)=1/2$ ；（2）如果 B 事件发生，X 发生的概率就是  $P(X|B)=1$ ，因为你选择了 3 号门，而 2 号门背后是汽车，主持人是了解这个情况的，所以他只能选择打开 1 号门而不是 2 号门。（3）如果 C 事件发生，X 发生的概率就是 0。因为主持人肯定会打开 2 号门，而不是 1 号。

这样，当你确认到 X 这个事件发生了之后，A 事件仍然发生的概率就需要按照贝叶斯法则计算：

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)} = \frac{P(X|A)P(A)}{P(A)P(X|A) + P(B)P(X|B) + P(C)P(X|C)}$$
$$= \frac{1/2 \times 1/3}{1/3 \times (1/2 + 1 + 0)} = 1/3$$

尽管它仍然是 1/3，但是，如果你改变选择 2 号门会有更多的机会赢得汽车，因为此时 B 事件（2 号门背后是汽车）发生的概率却变成：

$$P(B|X) = \frac{P(X|B)P(B)}{P(X)} = \frac{P(X|B)P(B)}{P(A)P(X|A) + P(B)P(X|B) + P(C)P(X|C)}$$
$$= \frac{1 \times 1/3}{1/3 \times (1/2 + 1 + 0)} = 2/3$$

所以，你的最优决策是改选 2 号门，因为它的概率会是你坚持选择 3 号门概率的 2 倍。从这个例子，我们可以清晰地看到后验的信息是如何改变事件概率的分布的。

冯纽曼在本章中提到的“概率作为一种逻辑的延伸”就是指概率的主观（贝叶斯）学派解释。在这个学派中，现代的一位集大成者是统计物理学家 E.T. Jaynes（他的[个人主页](#)），他从主观概率的角度出发，将统计物理看作是信息论的一个分支。此外，他也将概率视作逻辑的一种延伸。有关 E.T. Jaynes 的学术观点，大家可以参看他的著作：《概率论沉思录》（Probability theory – the Logic of Science）。



此外，随着人们对物理世界的探索，人们逐渐认识到了一种全新的概率论：量子概率。与经典概率最大的不同之处就在于量子概率定义了一种称为概率幅的量，它可以取复数值，而当你观测系统的时候，这个复数就会“塌缩”成经典概率。（关于量子概率，读者可参考我写的一篇科普文章：[《当概率成为复数——量子概率简介》](#)）。

人们起初认为：微观粒子世界不得被量子概率这套类似于魔法一样的数学去描述。然而，冯纽曼从数学和逻辑学的角度出发，将这套数学进行了大胆的抽象，从而奠定了量子逻辑这门学科（读者可以参考冯纽曼的经典著作：《*Mathematical Foundations of Quantum Mechanics*》Princeton University Press; Translation - from German edition (October 28, 1996)）。近年来，量子逻辑，特别是量子计算、量子通讯等领域的长足发展让人们看到了量子法则是如何突破经典计算和经典通信的（可参考：：Michael A. Nielsen, Isaac L. Chuang: *Quantum Computation and Quantum Information*, Cambridge University Press, 2000），这使得人们又重新注意到冯纽曼早在 20 年代就奠定的量子逻辑数学基础。

## 二、热力学

热力学是一门诞生于 18 世纪的学科，它起初是一门纯粹唯象的科学。人们首先建立了诸如温度、压强、热量等宏观物理量，并逐渐发现了热力学三大定律，从而引出了诸如熵、自由能等较为深刻的概念。到了 19 世纪之后，L. Boltzmann 的出现则使得热力学开始面貌一新。他将熵解释为系统宏观状态所对应的微观状态数，从而奠定了统计物理的基础。从此，统计力学成为了物理科学的一个庞大分支。

20 世纪的热力学和统计物理呈现出两种完全不同的走向。首先，人们迫切需要将热力学、统计物理扩展到非平衡态的复杂系统。然而，非平衡态下的热力学呈现出一种混乱的局面。早期，普利高津等人发展了近平衡态的热力学，从而把熵等概念进行了一定的扩充（可以参考普利高津的一系列著作，如《探索复杂性》、《从混沌到有序》等）；但后来更多的统计物理学家参与到了非平衡态复杂系统的研究中，他们更倾向于将各类物理方法和工具，诸如：相变、临界、幂律、稳定分布、福克-普朗克方程等用于各种各样的复杂系统中（请参考罗萨里奥·N·曼特尼亚，H·尤金·斯坦利所著：《经济物理学导论——金融中的相关性与复杂性》，2006，以及 D. Sornette 写的：《*Critical Phenomena in Natural Sciences*》，2000）。但是总体来说，这部分研究则呈现出群龙无首的局面。

近年来发展起来的另外一套关于流动的统计物理则另辟蹊径，尤其是最大熵产生原理有可能成为一种非平衡态系统的新的普适规律（可参考我写的一系列科普文章：[《生命之流》](#)以及[《流的探索》](#)）。

另外一条非常不一样的分支则是讨论信息与热力学之间的关系。在正文中冯纽曼反复提到的 Szilard 则是探索信息与热耗散之间联系的第一人。在此推荐 Harvey Leff, Andrew F. Rex 著：《*Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*》，2002）。

## 三、信息论

现代公理化概率论的创始人 Kolmogorov 曾经说过一句话：“Information theory must precede probability theory and not be based on it (信息论必将超越概率论而不是反过来成为概率论的推论)。

他的这句话表明了从信息论的角度，我们可以获得对概率以及不确定性的全新理解。这

就仿佛是基于微分方程的经典力学与基于最优化的变分原理的分析力学之间的关系,我们可以从概率论、马尔科夫链的角度来理解随机系统,也可以从最大熵、信息论的角度来理解它。

熵无疑是信息论的基础。它最早来源于热力学和统计物理,然而,Shannon 的奠基性工作却给我们展示了同样的数学表达形式也可以在编码、通信、压缩等信息论领域起到至关重要的作用(有关信息论,大家可以参考一本非常好的书:科弗.托马斯:《信息论基础》,机械工业出版社,2005)。

更有趣的是,近年来人们对 Shannon 熵的数学形式进行了一系列的扩展,其中最有趣的一套扩展就是 Renyi 熵(它被认为是多重分形的基础概念之一,请参考 D. Harte:《Multifractals》,Chapman&Hall/CRC,2001)以及后来 Tsallis 发展的非可延展统计物理:S. Abe, Y. Okamoto:《Nonextensive Statistical Mechanics and Its Applications》,Springer-Verlag Berlin,2001)。扩展的熵形式可推导出复杂系统中普遍存在的幂律分布。

最后,我想提一下 Fisher 信息。早在 Shannon 提出熵的数学形式之前,统计学家 Fisher 就提出了另外一种度量信息的方法,即 Fisher 信息。有趣的是,B.R. Frieden 对 Fisher 信息的数学形式进行了一定的扩展,并通过最小化 Fisher 信息,推导出了物理学中的电磁场的 Maxwell 方程、量子力学中的 Klein-Gordon 方程、广义相对论的爱因斯坦方程等。从此,人们第一次将物理定律建立在了 Fisher 信息的基础之上,这可以算是信息论的一个重大进展。

(参见: B.R. Frieden: Science From Fisher Information: A Unification, Cambridge University Press, 2004)。