随机、赌徒谬误、小数定律，概率论中最简单的智慧

2015-12-27 [算法与数学之美](javascript:void(0);)

**http://mmbiz.qpic.cn/mmbiz/dxyBLGFw5tsGTOzG57bY5UpDagnRkuDULOzQn9vyLjhBRibjEGGHiasu0uyUia1HnxichyuJcESWIOMk38xXOCP72g/0?wx_fmt=gif&tp=webp&wxfrom=5&wx_lazy=1**

*作者：万维钢*

*来源：选自《万万没想到——用理工科思维理解世界》，电子工业出版社出版。*

*转自：罗辑思维*

我认为人人都应该学一些概率知识，它现在是公民必备知识。

现在的世界比过去复杂得多，其中有大量不确定性，是否理解概率，直接决定一个人的开化程度。

1**随机：有些事情是无缘无故地发生的**

这个思想对我们的世界观有颠覆的意义。

古人没有这个思想，认为一切事物都是有因果的，甚至可能都是有目的的。人们曾经认为世界像一个钟表一样精确地运行。但真实世界不是钟表，它充满不可控的偶然。

更严格地说，有些事情的发生，跟他之前发生的任何事情，都可以没有因果关系。不论我们做什么都不能让它一定发生，也不能让它一定不发生。

一个人考了好大学，人们会说这是他努力的结果；一个人事业成功，人们会说这是他努力工作的结果。可是如果一个人买彩票中了大奖，这又是为什么呢？

答案是没有任何原因，这完全是一个随机事件。总会有人买彩票中奖，而这一期彩票中奖，跟他是不是好人，他在之前各期买过多少彩票，他是否关注中奖号码的走势，没有任何关系。

若一个人总是买彩票，他中奖的概率会比别人大点吧？的确，他一生之中中一次奖的概率比那些只是偶然买一次彩票的人大。但是当他跟上千万个人一起面对一次开奖的时候，他不具备任何优势。他之前所有的努力，对他在这次开奖中的运气没有任何帮助。一个此前没有买过任何彩票的人，完全有可能，而且有同样大的可能，在某一次开奖中把最高奖金拿走。

中奖，既不是他个人努力的结果，也不是“上天”对他有所“垂青”；不中，也不等于任何人与他做对。这就是“随机”，你没有任何办法左右结果。

理解随机性，我们就知道很多事情发生就发生了，没有太大可供解读的意义。我们不能从这件事获得什么教训，不值得较真，甚至不值得采取行动。

* 再完美的交通工具也不可能百分百安全，我们会因为极小的事故概率不坐飞机吗？我们只需要确定事故概率比其他旅行方式小就可以了。甚至连这都不需要，只需要确定这个小概率事件我们能够容忍就可以了。

2**赌徒谬误**

假如你在赌场玩老虎机，一上来运气不太好，连输好几把。这时候你是否有种强烈的感觉，你很快该赢了？

期货、彩票都是一样。连续好几把上来就亏损的情况下，是不是觉得下一把挣钱的概率很大？

这完全是一种错觉。赌博完全是独立的随机事件，这意味着下一把的结果和以前所有的结果都没有任何联系，已经发生了的事情不会影响将来。

“大数定律”说，如果进行足够多的抽奖，那么各种不同结果出现的频率就会等于他们的概率。

人们常常错误地理解为，随机就意味着均匀。如果过去一段时间内发生的事情不均匀，人们就错误的以为未来的事情会尽量往“抹平”的方向走。如果连输几把，那么下一把就应该会赢。

但大数定律的工作机制不是和过去搞平衡，它的真实意思是说如果未来进行非常多次的抽奖，你输非常多次、赢非常多次，以至于他们此前的一点点差异就会变得微不足道。

* 有个笑话说一个人乘坐飞机时总带着一颗炸弹，他认为这样就不会被恐怖分子炸飞机了，因为一架飞机上有两颗炸弹的可能性非常小。
* 战场上士兵有个说法，如果战斗中炸弹在你身边爆炸，你应该迅速跳进那个弹坑，因为两颗炸弹不大可能打到同一个地方。

这都是不理解独立随机事件导致的。

3**在没有规律的地方发现规律**

理解了随机性和独立随机事件，我们可以得到一个结论：独立随机事件的发生是没有规律和不可预测的，这是一个非常重要的智慧。

彩票分析师，相信中奖号码存在走势，相信其中的规律，所以近期多次出现的组合可能会继续出现，或者按照这个趋势可以预测下一个号码。

但这里根本没有规律，是完全随机的现象，即便存在缺陷，也需要大量的开奖后才能发现，而且缺陷的结果也很简单，无非是某个特定号码出现的可能性略大一些，完全谈不上什么复杂规律。

明明没有规律，这些彩票分析师是怎么看出规律来的呢？也许他们不是故意骗人，而很可能他们真的相信自己找到了彩票的规律。

发现规律是人的本能。

春天过后是夏天，乌云压顶常下雨，大自然中很多事情的确是有规律的。我们的本能工作得如此之好，以至于我们在明明没有规律的地方也能找出规律来。人脑很擅长理解规律，但是很不擅长理解随机性。

在没有规律的地方发现规律是很容易的事情，只要你愿意忽略所有不符合你这个规律的数据。而且如果数据够多，我们可以找到任何我们想要的规律。

* 有人拿圣经做字符串游戏，声称这是圣经对后世的预言。问题是，这些预言可以完美的解释已经发生的事情，但在预测未发生的事情时就不好使了。关键是圣经中有很多很多字符，如果仔细寻找，尤其是借助计算机的话，总能找到任何想要的东西。
* 把圣经换成毛选也一样，你会发现毛选也早就预言了中国后世发生的所有大事。

未来是不可被精确预测的，这个世界也并不像钟表那样运行。

4**小数定律**

现在我们知道，数据足够多的话，人们可以找到任何自己想要的重要规律，只要他不在乎这些规律的严格性和自洽性。那么在数据足够少的情况下又会如何？

如果数据足够少，有些规律会自己跳出来，你甚至不相信都不行。

人们抱着游戏或者认真的态度总结了世界杯足球赛的各种“定律”。比如——

* “巴西队的礼物”：只要巴西夺冠，下一届的冠军就将是主办大赛的东道主，除非巴西队自己将礼物收回。这一定律在2006年被破解。
* “1982轴心定律”：世界杯夺冠球队以1982年世界杯为中心呈对称分布，这个定律在2006年被破解。

还有一些未被破解的定律，比如——

* 凡是获得联合会杯或美洲杯，就别想在下一届世界杯夺冠。
* 中国队的“王治郅定律”：只要王治郅参加季后赛，八一队必然得总冠军，以及“0:2”落后无人翻盘定律。

如果仔细研究这些定律，会发现不易破解的定律其实都有一定的道理。王治郅和八一队都很强，0:2落后的确很难翻盘，而获得世界杯冠军是个非常不容易的事情，更别说同时获得联合会杯、美洲杯和世界杯。但不容易不等于不会发生，他们终究会被破解。

那些看似没有道理的神奇定律（正因为没道理，所以显得神奇），则大多数已经被破解。之所以神奇，是因为纯属巧合。世界杯总共才进行了80多年，20多届。只要数据足够少，我们总能发现一些没有破解的规律。

如果数据少，随机现象可以看上去很不随机。甚至非常整齐，感觉好像真有规律一样。

问题的关键是，随机分布不等于均匀分布。要想均匀分布，必须要样本总数非常大的时候才有效。一旦不均匀，人们就认为其中必有缘故（阴谋论），而事实却是这可能只是偶然事件。

* iPod最早推出“随机播放”功能的时候，用户发现有些歌曲会被重复播放，他们据此认为播放根本不随机。苹果公司只好放弃真正的随机算法，用乔布斯本人的话说，就是改进以后的算法使播放“更不随机以至于让人感觉更随机”。

如果统计数据很少，就很容易出现特别不均匀的情况。这个现象被诺奖得主丹尼尔·卡尼曼戏称为“小数定律”。卡尼曼说，如果我们不理解小数定律，就不能真正理解大数定律。

大数定律是我们从统计数字中推测真相的理论基础。大数定律说如果统计样本足够大，那么事物出现的频率就能无限接近他的理论概率——也就是他的“本性”。而小数定律说如果样本不够大，那么他就表现为各种极端情况，而这些情况可以跟他的本性一点关系都没有。

* 一个只有二十人的乡村中学某年突然有两人考上清华，跟一个有两千人的中学每年都有两百人考上清华，完全没有可比性。

如果统计样本不够大，就什么也说明不了。

正因为如此，我们才不能只凭自己的经验，哪怕加上家人和朋友的经验，去对事物做出判断。我们的经验非常有限。别看个例，看大规模统计。有的人听说两三个负面新闻就敢写文章把社会批得一文不值，这样的人非常无知。