

文章编号:1002—1566(2006)02 - 0240 —04

概率与直觉

林志兴

(福州大学数学与计算机学院,福建,350002;莆田学院,福建,351100)

摘要:通过“免费抽奖问题”、“生日问题”、“肇事车认定问题”一个例子说明:在求某个事件的概率时,如果不经认真分析,常常会被直觉误导。并分析了导致错觉的原因是错误地选择了样本空间,进而说明正确选择样本空间的重要性。

关键词:直觉;概率;样本空间

中图分类号:O211

文献标识码:A

Probability and Intuition

L IN Zhi-xing

(The college of Maths and Computer Sciences, Fuzhou University, Fujian, 350002;
Putian University, Fujian, 351100 China)

Abstract: Through the question of drawing a lottery free, the birthday question, and the trouble-causing car cognition question, the paper explain the fact that if we don't think it over, we will often be misled by the intuition while reckoning the probability of a certain incident. And find the reason that cause illusion is to choose the sample space wrongly, then explain the importance of choosing the sample space correctly.

Key words: intuition; probability, sample space

0 引言

几乎没有一个数学教师不知道概率对生活的重要性,按约瑟夫·巴特勒的说法,概率是“生活的真正指南”。在现实生活中,人们常常对某一事件发生的概率作出估计,这种估计往往是依靠直觉进行的。如果没有真正领会概率论的原理,很有可能作出错误的判断。下面的这些例子说明,直觉在很多情况下是不可靠的。

1 例子

1.1 免费抽奖问题

某个经营洗涤用品的公司推出一个非常有吸引力的促销活动:

“本公司为了答谢广大顾客长期以来对本公司产品的支持与厚爱,特推出免费抽奖活动。抽奖方式:箱子里有 20 个球,10 个 10 分和 10 个 5 分。从箱子里摸出 10 个球,把各个球上的分数加总,按总分设置奖项如下:

收稿日期:2004 年 11 月 19 日

一等奖:100分,29寸彩电一台,价值3000元	七等奖:85分,香皂一块,价值4元
二等奖:50分,DVD一台,价值700元	八等奖:65分,牙膏一盒,价值3元
三等奖:95分,榨汁机一台,价值200元	九等奖:80分,牙刷一把,价值2元
四等奖:55分,电饭煲一个,价值100元	十等奖:70分,洗衣粉一小袋,价值0.5元
五等奖:90分,XX牌洗发露2瓶,价值50元	十一等奖:75分,以成本价25元购买XX版洗发露1瓶
六等奖:60分,XX牌洗发露1瓶,价值25元	

“抽奖结果共有11个,即50、55、……、100分。其中有10个分数可免费得到奖品,中奖率超过90%。”

我们的直觉似乎和商家宣传的一致,中奖率为 $10/11 \approx 91\%$ 。但如果你站在一旁观察抽奖情况,你会发现得十一等奖的摸奖者很多,而且就算中其他免费奖项,也大多是一些价值较低的小奖品。难道是箱子里暗藏机关?

实际上,箱子里没有任何“机关”,而且关键的问题在于十一个奖项出现的概率不一样。随机摸出10个球,设其中10分球的个数为 X ,则5分球的个数为 $10 - X$ 。 X 服从超几何分布,即

$$P(X = i) = \frac{C_{10}^i C_{10}^{10-i}}{C_{20}^{10}}, (i = 0, 1, \dots, 10)$$

根据上式,可求得各事件发生的概率如下:

表1

10分球的个数 X	0	1	2	3	4	5
概率 $P(X = i)$	0.000005	0.000541	0.01096	0.077941	0.238693	0.343718
分数	50	55	60	65	70	75
奖项	—	三	五	七	九	十一
10分球的个数 X	6	7	8	9	10	
概率 $P(X = i)$	0.238693	0.077941	0.01096	0.000541	0.000005	
分数	80	85	90	95	100	
奖项	十	八	六	四	二	

摸奖者需要购买产品的概率超过 $1/3$,而且价值越高的奖项,得到的概率就越低。特别是中两个大奖的概率只有十万分之一左右。再来研究一下一次摸奖中商家的期望收益。经调查,各奖项奖品的成本如下表所示:

表2

奖项	—	二	三	四	五	六
概率 $P(X = i)$	0.000005	0.000005	0.000541	0.000541	0.01096	0.01096
奖品成本	2500	500	150	80	20	10
商家收益	- 2500	- 500	- 150	- 80	- 20	- 10
奖项	七	八	九	十	十一	
概率 $P(X = i)$	0.077941	0.077941	0.238693	0.238693	0.343718	
奖品成本	2.9	2.1	1.2	0.4	10	
商家收益	- 2.9	- 2.1	- 1.2	- 0.4	15	

一次摸奖中商家的期望收益 = $\sum$ 商家收益 $\times$ 相应概率,计算结果为3.91元。看似“免费

抽奖”,实为商家推销产品、获取利润的手段。

1.2 生日问题

某个班级有 55 名学生,假定每个人的生日在一年中的任何一天的概率为 $1/365$,问该班级至少有两人生日在同一天概率有多大?

一年的天数远大于学生的人数,似乎至少有两人的生日在同一天概率不会很大,但下面的计算结果可能会让你感到意外。上面的生日问题可转化为分房问题:“有 n 个人,每人都以同样的概率 $1/N (n \leq N)$ 被分配到 N 间房的任一间中,求至少有两个人在同一房间的概率。”设 A = “至少有两个人在同一房间”,则 $\bar{A}$ = “恰有 n 间房中各有一人”。把某个人分配到 N 间房中之一去,又 N 种可能,把 n 个人分配到 N 房中去共有 N^n 种分配法,即基本事件总数为 N^n 。恰好有 n 间房可从 N 间房中任意选出,有 C_N^n 种选法,对每一种这样的选法, n 个人又有 $n!$ 种不同的分配法,所以 $\bar{A}$ 包含的基本事件总数为 $C_N^n \cdot n!$,所以

$$P(\bar{A}) = \frac{C_N^n \cdot n!}{N^n}, P(A) = 1 - P(\bar{A})$$

再回到生日问题。一年的 365 天相当于 365 间房,55 个学生相当于 55 个客人,即 $N = 365, n = 55$ 。设 B = “至少有两人生日在同一天”,则

$$P(B) = 1 - P(\bar{B}) = 1 - \frac{C_{365}^{55} \cdot 55!}{365^{55}} \approx 0.99$$

几乎必有两人生日在同一天,与直觉有很大的差异。

1.3 肇事车认定问题

一辆出租车涉及一起夜间肇事逃逸事故。在这个城市里,有绿色和蓝色两种出租车。其中绿色的占 85%,蓝色的占 15%。有一位目击者认定肇事的出租车是蓝色的。法庭在与出事当夜相同环境下测试目击者的可信度,并得出结论,在 80%时间里,目击者能正确识别这两种颜色,20%的时间里不能(对两种颜色的识别率一致)。问肇事车是蓝色的概率大还是绿色的概率大?

有些人可能会觉得这个问题是多些一举,目击者认这肇事车是蓝色的,而他能正确识别颜色的概率为 80%,这不明摆着肇事车是蓝色的概率就是 80%吗?真的就是这样吗?且看下面分析。

设 A = “肇事车是蓝色的”, $\bar{A}$ = “肇事车是绿色的”。在没有目击者的情况下, $P(A) = 0.15, P(\bar{A}) = 0.85$ 。另设 B = “目击者认为肇事车是蓝色的”, $\bar{B}$ = “目击者认为肇事车是绿色的”。由已知条件有 $P(B/A) = 0.80, P(B/\bar{A}) = 0.20$ 。在目击者认为肇事车是蓝色的情况下肇事车确实是蓝色的概率为

$$\begin{aligned} P(A/B) &= \frac{P(A) P(B/A)}{P(A) P(B/A) + P(\bar{A}) P(B/\bar{A})} \\ &= \frac{0.15 \times 0.80}{0.15 \times 0.80 + 0.85 \times 0.20} \approx 0.41 \end{aligned}$$

在目击者认为肇事车是蓝色的情况下而肇事车却是绿色的概率为

$$\begin{aligned} P(\bar{A}/B) &= 1 - P(A/B) \\ &= 1 - 0.41 = 0.59 \end{aligned}$$

$P(A/B) < P(\bar{A}/B)$,即在目击者认为肇事车是蓝色的情况下,肇事车是绿色的概率大于蓝色的概率,与直观判断刚好相反。我们从另外一个角度来计算,也许能更直观地说明这个问

题:一辆车被目击者认为是蓝色的概率

$$\begin{aligned}P(B) &= P(A) P(B | A) + P(\bar{A}) P(B | \bar{A}) \\&= 0.15 \times 0.80 + 0.85 \times 0.20 \\&= 0.12 + 0.17 = 0.29\end{aligned}$$

即每 100 辆被指认的出租车中,平均有 29 辆被目击者认为是蓝色的。其中有 $100 \cdot P(A) \cdot P(B | A) = 12$ 辆是蓝色的,而有 $100 \cdot P(\bar{A}) \cdot P(B | \bar{A}) = 17$ 辆是绿色的,即绿色的更多。

2 小结

上面通过三个例子说明,在求某个事件的概率时,如果不经认真分析,容易被直觉误导。在日常生活中,由于条件的限制,我们不可能在判断某个事件的概率时都拿出纸和笔来计算一番。但为了减少判断错误,我们应当仔细领会概率论的原理。我们平时遇到的很多概率问题是古典概率问题。在古典概率问题中,样本空间的选择是非常关键的。在这三个例子中,产生错误的原因均是由样本空间的选择不当引起的。在“免费抽奖问题”中,样本空间应当是由 C_{20}^{10} 个不同的抽样结果组成,而不是 11 个不同的总分;在“生日问题”中,样本空间应当是 365⁵⁵ 种不同的生日分布情况,而不是一年中的 365 天;在“肇事车认定问题”中,产生错误的原因表面上看是混淆了 $P(A | B)$ 和 $P(B | A)$ 两个概率,其实也是由于错误地选择了样本空间引起的,即错误地把“所有蓝色的车”当作样本空间,正确的样本空间应该是“所有被指认为蓝色的车。”

[参考文献]

- [1] 缪铨生,概率与数理统计[M].上海:华东师范大学出版社.1997.
- [2] 杨镜华,高额奖金后面有陷阱——一项抽奖活动里的数学[J].数理统计与管理.2003(3),54~55.