Monty Hall problem

问题重述

[Wikipedia][https://en.wikipedia.org/wiki/Monty_Hall_problem]：
　Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

贝叶斯证明：换门好!

首先给出记号：

- X：选手选择的🚪；
- Y：🚗实际在的🚪；
- Z：主持人打开的🚪(必然是🐑)；按照题意，Z的取值同时受到X和Y的取值的限制，即Z的取值不能和X和Y的取值重复；

　　假设选手打开了No.1🚪，主持人打开了No.3🚪，选手该不该换门No.2🚪的问题，转化成概率问题，我们希望比较两个条件概率那个更大：

$$P(Y=1|X=1,Z=3)$$
$$P(Y=2|X=1,Z=3)$$

　　直觉上讲，主持人实际上在打开No.3🚪的时候帮我们排除了$Y=3$的可能，按理来说就只有剩下两个门会有🚗，换与不换🚗在后面的概率不应该是相等吗?

X=i	1	2	3
P	1/3	1/3	1/3

Y=i	1	2	3
P	1/3	1/3	1/3

由于Z同时受限与X和Y，所以实际上他的样本空间由27种可能缩减到12种可能；

	X	Y	Z
1	1	1	2
2	1	1	3
3	2	1	3
4	3	1	2
5	1	2	3
6	2	2	1
7	2	2	3
8	3	2	1
9	1	3	2
10	2	3	1
11	3	3	1
12	3	3	2

我们可以得到概率(后面计算要用的)

$$
\begin{aligned}
P(Z=3|X=1,Y=1)&=\frac{1}{2}\\
P(Z=3|X=1,Y=2)&=1\\
P(Z=3|X=1,Y=3)&=0
\end{aligned}
$$

同时我们要指明一点选手选择的🚪X是不受🚗实际在的🚪Y的影响，反过来也一样，就是说X和Y是独立的；那么由贝叶斯公式我们要计算

$$
\begin{aligned}
P(Y=k|X=1,Z=3)&=\frac{P(Y=k,X=1,Z=3)}{P(X=1,Z=3)}\\
&=\frac{P(Y=k,X=1,Z=3)}{\sum_{y=1}^{3}P(Y=y,X=1,Z=3)}
\end{aligned}
$$

也就是说我们只要计算$P(Y=y,X=1,Z=3)$，就万事大吉了！

$$
\begin{aligned}
P(Y=y,X=1,Z=3)&=P(Y=y)P(X=1,Z=3|Y=y)\\
&=P(Y=y)P(X=1|Y=y)P(Z=3|X=1,Y=y)\\
&=P(Y=y)P(X=1)P(Z=3|X=1,Y=y)
\end{aligned}
$$

（第三个等号由X和Y的独立性得到）这些概率我们前面都已经铺垫好啦，直接计算就可以了

$$
\begin{aligned}
P(Y=1,X=1,Z=3)&=P(Y=1)P(X=1)P(Z=3|X=1,Y=1)\\
&=\frac{1}{3}\times\frac{1}{3}\times\frac{1}{2}\\
P(Y=2,X=1,Z=3)&=P(Y=2)P(X=1)P(Z=3|X=1,Y=2)\\
&=\frac{1}{3}\times\frac{1}{3}\times 1\\
P(Y=3,X=1,Z=3)&=P(Y=3)P(X=1)P(Z=3|X=1,Y=3)\\
&=\frac{1}{3}\times\frac{1}{3}\times 0
\end{aligned}
$$

回到我们要计算的目标：

$$
\begin{aligned}
P(Y=1|X=1,Z=3)&=\frac{\frac{1}{3}\times\frac{1}{3}\times\frac{1}{2}}{\frac{1}{3}\times\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{3}\times 1+\frac{1}{3}\times\frac{1}{3}\times 0}=\frac{1}{3}\\
P(Y=2|X=1,Z=3)&=\frac{\frac{1}{3}\times\frac{1}{3}\times 1}{\frac{1}{3}\times\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{3}\times 1+\frac{1}{3}\times\frac{1}{3}\times 0}=\frac{2}{3}
\end{aligned}
$$

嗯哼！$P(Y=1|X=1,Z=3)<P(Y=2|X=1,Z=3)$，也就是说当选手选择了No.1🚪，主持人选择了No.3🚪之后，No.2🚪里有车的概率概率竟然会变大！！！真的好反直觉orz...

原因是啥?

　　我们为什么直觉上是主持人选择后换不换中奖的概率是相等的呢?

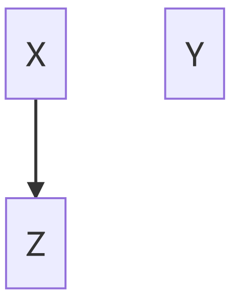
　　哈哈，其实是我们忽略了主持人对Y的认识，那么我们的样本空间只是从27种降到18种(仅限制X的取值不与Z的取值[大白话就是主持人不能打开选手选的门]，但是主持人不知道哪个🚪后面有🚗，就是说随机开剩下的🚪[Z的取值可以和Y的取值相等])，那么我们的样本空间如下：

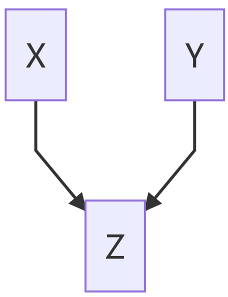
	X	Y	Z
1	1	1	2
2	1	1	3
3	2	1	1
4	2	1	3
5	3	1	1
6	3	1	2
7	1	2	2
8	1	2	3
9	2	2	1
10	2	2	3
11	3	2	1
12	3	2	2
13	1	3	2
14	1	3	3
15	2	3	1
16	2	3	3
17	3	3	1
18	3	3	2

$$P(Z=3|X=1,Y=1)=\frac{1}{2}$$
$$P(Z=3|X=1,Y=2)=\frac{1}{2}$$
$$P(Z=3|X=1,Y=3)=0$$

再根据上述步骤进行计算实际上就可以获得等概的结果。

实际上主持人选择打开哪扇门的时候，实际上是透露了真的有🚗的🚪的信息(即仅限制X的取值不与Z的取值，同时限制Y的取值不与Z的取值)，上面样本空间中序号为3/5/7/12/14/16实际上会因为主持人知道Y的信息而不会出现的情况；





如果用图模型来表示他们之间的关系，实际上是上图中靠下的这个情况，是一个对撞结构，Z受到其他两个变量的共同影响。

参考

[1] https://en.wikipedia.org/wiki/Monty_Hall_problem

[2] Jewell, Nicholas P , Glymour, et al. Causal Inference In Statistics.