第二周　　条件概率和独立性

2.1 条件概率

学们好！本周我们学习条件概率、条件概率的计算方法以及独立性概念。

在实际问题中，对于随机事件 A，除了关心它本身的概率，有时还需要知道在某些附加条件下该事件发生的概率，这些附加条件通常以“某个事件已经发生”的形式给出。这就是已知某事件发生后，事件 A 的条件概率。

****

例 2.1.1　考虑恰有两个小孩的全部家庭，并且假定生男、生女是等可能的。若随机地选一个家庭，发现该家庭有一个女孩，问这一家另一个小孩是男孩的概率是多少？

解：　样本空间：{（男，男），（男，女），（女，男），（女，女）}，

设事件 A 为“其中一个是女孩”，事件 B 为“其中一个是男孩”

$A=\{(男，女),(女，男),(女，女)\}$

$B=\{(男，男),(男，女),(女，男)\}$

$AB=\{(男，女),(女，男)\}$

某家庭有一个女孩条件下，另一个小孩是男孩的概率为 $\dfrac{P(AB)}{P(A)}=\dfrac{2/4}{3/4}=\dfrac{2}{3}$

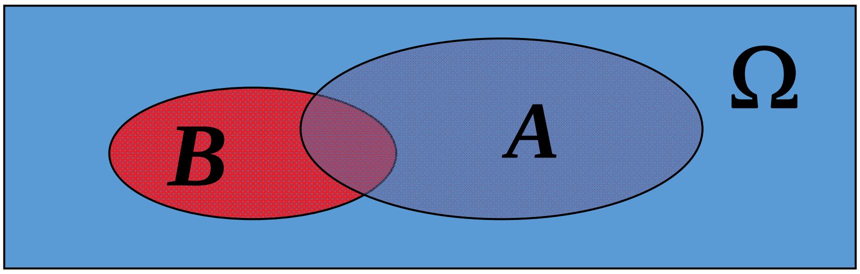
****

条件概率的定义

设 A、B 是两个事件，且 $P(B)>0$，则

$$P(A|B)=\frac{P(AB)}{P(B)}$$

称为“在事件 B 发生条件下，事件 A 发生的条件概率”，简称为条件概率.



****

例 2.1.2　某厂有甲、乙、丙三车间生产同一种产品，产量分别占总产量的 60%，30%和 10%。各车间的次品率分别是 2%，5%，6%。试用事件的语言表达如下概率

（1）各车间的次品率？

（2）若发现一件产品为次品，该次品来自甲车间的概率？

解：设产品是甲、乙、丙车间所生产分别为事件 A_1, A_2, A_3 ，产品是次品为事件 B，

（1） $P(B \mid A_1) = 0.02$ ， $P(B \mid A_2) = 0.05$， $P(B \mid A_3) = 0.06$

（2） $P(A_1 \mid B)$

若发现一件产品为次品，该次品来自甲车间的概率，可表示为以 B 为条件，A1 的条件概率表达式。这一概率的值就不是显然的了，要计算这一概率，需要进一步的计算工具。在学习条件概率计算方法之前，对例 2.1.1 我们还想做些引申。

****

对于例 2.1.1，有些同学直觉中的解答可能是 1/2,因为生男生女等可能，所以无论一个孩子是男是女，另一个孩子是男孩的概率都应该是 1/2。实际上 1/2 是另一个不同问题的正确解答。

****

例 2.1.3 考虑恰有两个小孩的全部家庭，并且假定生男、生女是等可能的。如果从这些家庭中随机地选择一个孩子，并发现她为女孩，问在她家里另一个孩子是男孩的概率是多少？

解：　样本空间：{ 男 g，男 b，女 g，女 b}，

设事件 A 为“这个孩子是女孩”，事件 B 为“这个孩子有一个兄弟”

A={ 女 g，女 b }， B={男 b，女 b }，

$AB=$ { 女 b }.

所求概率为 $P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{1/4}{2/4}=\dfrac{1}{2}$

这绝不是一个矫揉造作的例子，而是一个非常值得体会的例子，它说明正确理解概率统计学中“我们的抽样对象到底是什么”的重要性。这个例子也被著名概率学者钟开莱先生在他的《初等概率论》一书所采用。我们的分析也是沿着他书中的思路给出的。下一讲我们学习条件概率有关的几个重要计算公式。

****

2.2 条件概率有关条件概率的三个重要计算公式

上一讲中我们引入了条件概率，有了这一概念，我们对事件的表达就有了更丰富的工具。下面我们就希望能够有效地计算条件概率，得到我们想要的概率结果。对于条件概率而言呢，主要有三个计算公式，分别是乘法公式、全概率公式和贝叶斯公式。这三个计算公式的应用贯穿概率论的始终，是非常基本和重要的计算工具。下面我们看第一个乘法公式。

****

乘法公式

（1）设 A,B 是两个事件，$P(B)>0$，则 $P(AB)=P(B)P(A|B)$

证明： $P(A|B)=\dfrac{P(AB)}{P(B)} \Rightarrow P(AB)=P(B)P(A|B)$

（2）设 $A_1,A_2,\cdots,A_n$ 为 n 个事件，且 $P(A_1A_2\cdots A_{n-1})>0$，则

$$P(A_1A_2\cdots A_n)=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})。$$

证明： 数学归纳法， 设 $P(A_1\cdots A_k)=P(A_1)\cdot P(A_2\mid A_1)\cdots P(A_k\mid A_1\cdots A_{k-1})$，

$$P(A_1\cdots A_{k+1})=P(A_1A_2\cdots A_k)\cdot P(A_{k+1}\mid A_1A_2\cdots A_k)$$

$$=P(A_1)\cdot P(A_2\mid A_1)\cdots P(A_{k+1}\mid A_1A_2\cdots A_k).$$

直接验证：

$$P(A_1)\cdot P(A_2\mid A_1)\cdot P(A_3\mid A_1A_2)\cdots P(A_n\mid A_1A_2\cdots A_{n-1})$$

$$=P(A_1)\frac{P(A_1A_2)}{P(A_1)}\frac{P(A_1A_2A_3)}{P(A_1A_2)}\cdots\frac{P(A_1A_2\cdots A_n)}{P(A_1A_2\cdots A_{n-1})}$$

$$=P(A_1A_2\cdots A_n).$$

****

例 2.2.1 设箱子内有 a 个白球，b 个黑球，在其中不放回地连取 3 次，问前 2 次取到白球而第 3 次取到黑球的概率。

解： 设事件 A_i 表示第 i 次抽到白球，

$$P(A_1A_2\bar{A}_3)=P(A_1)\cdot P(A_2\mid A_1)\cdot P(\bar{A}_3\mid A_1A_2)$$

$$=\frac{a}{a+b}\cdot\frac{a-1}{a+b-1}\cdot\frac{b}{a+b-2}$$

思考：3 个均为白球，或抽到 2 黑 1 白，发生的概率分别是多少？

****

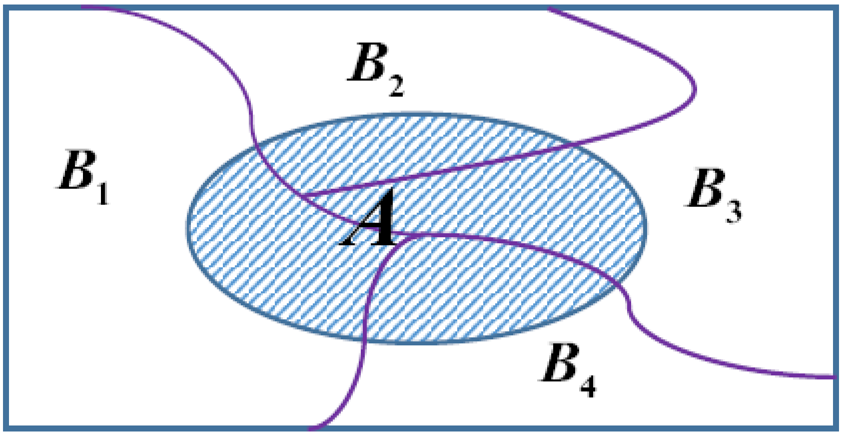
2. 全概率公式

设 $B_1,B_2,\cdots,B_n$ 是样本空间 Ω 的一个分割，即 $B_1,B_2,\cdots,B_n$ 互不相容，

且 $\bigcup_{i=1}^{n}B_i=\Omega$。如果 $P(B_i)>0\quad(i=1,2,\cdots,n)$，

则对任一事件 A，有 $P(A)=\sum_{i=1}^{n} P(B_i)P(A|B_i)$。

我们先用图示进一步明确对样本空间进行“分割”的含义。



$$P(A)=P(AB_1)+P(AB_2)+P(AB_3)+P(AB_4)$$

$$=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+P(B_3)P(A|B_3)+P(B_4)P(A|B_4)$$

****

2. 全概率公式

设 $B_1,B_2,\cdots,B_n$ 是样本空间 Ω 的一个分割，即 $B_1,B_2,\cdots,B_n$ 互不相容，

且 $\bigcup_{i=1}^{n} B_i=\Omega$。如果 $P(B_i)>0\quad(i=1,2,\cdots,n)$，

则对任一事件 A，有 $P(A)=\sum_{i=1}^{n} P(B_i)P(A|B_i)$。

证明： $P(A)=P(A\Omega)=P\left(A\cap(B_1\cup B_2\cup\cdots\cup B_n)\right)$

$$=P(AB_1\cup AB_2\cup\cdots\cup AB_n)$$

$$=P(AB_1)+P(AB_2)+\cdots+P(AB_n)$$

$$=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+\cdots+P(B_n)(A|B_n)$$

****

例 2.2.2 设甲箱中有a个白球，b个黑球，$a>0,b>0$；乙箱中有c个白球，d个黑球。自甲箱中任取一球放入乙箱，然后再从乙箱中任取一球。求最后由乙箱取出的是白球的概率。

解： 设事件A表示最后由乙箱取出的是白球，事件W表示从甲箱取出白球，

$$P\left(A\right)=P\left(W\right)P\left(A\,|\,W\right)+P\left(\overline{W}\right)P\left(A\,|\,\overline{W}\right)$$

$$=\frac{a}{a+b}\cdot\frac{c+1}{c+d+1}+\frac{b}{a+b}\cdot\frac{c}{c+d+1}$$

$$=\frac{ac+bc+1}{\left(a+b\right)\left(c+d+1\right)}.$$

****

例 2.2.3 买彩票，设n张彩票中有 1 张奖券，人们排成一队购买彩票，求第k个人购到奖券的概率。

对这一问题，通过一个直观的分析，即可得到结果。我们假想一种买彩票的过程，假设每个人买完彩票后都不离开，也不查看结果，而是 n 张彩票都卖完后，n 个人同时打开。这一假想过程，并不影响每个人的中奖可能。而 n 个人一起同时打开彩票时，奖券落在每个位置的机会均等，所以每个人的中奖概率都是相同的，均为 1/n。这样我们就得到了答案。但是现在我们不仅仅满足于得到得数，而是希望通过事件表达，运用标准的概率计算工具得到对这一问题的分析和理解，而这种分析和理解往往是更深刻的，其方法是<u>更有可能推广而适用于</u>更多问题的。

解 1：<u>首部分析法</u> 设 $A_k\left(n\right)$ 为事件“n个人买彩票，第k个人中奖”，则

$$P\left(A_k\left(n\right)\right)=P\left(A_1\left(n\right)\right)P\left(A_k\left(n\right)\,|\,A_1\left(n\right)\right)\ +\ P\left(\overline{A_1\left(n\right)}\right)P\left(A_k\left(n\right)\,|\,\overline{A_1\left(n\right)}\right)$$

$$=0+\frac{n-1}{n}P\left(A_{k-1}\left(n-1\right)\right)$$

$$=\frac{n-1}{n}\cdot\frac{n-2}{n-1}P\left(A_{k-2}\left(n-2\right)\right)$$

$$=\frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-(k-1)}{n-(k-2)}P\left(A_1\left(n-(k-1)\right)\right)$$

$$=\frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-(k-1)}{n-(k-2)}\cdot\frac{1}{n-(k-1)}=\frac{1}{n}.$$

****

例 2.2.3 买彩票，设n张彩票中有 1 张奖券，人们排成一队购买彩票，求第k个人购到奖券的概率。

解 2：设事件A_i表示“第i个人买到彩票”，则

$$P(A_k)=P\left(\overline{A_1}\,\overline{A_2}\cdots\overline{A_{k-1}}A_k\right)$$

$$=P\left(\overline{A_1}\right)P\left(\overline{A_2}\mid\overline{A_1}\right)\cdots P\left(\overline{A_{k-1}}\mid\overline{A_1}\,\overline{A_2}\cdots\overline{A_{k-2}}\right)P\left(A_k\mid\overline{A_1}\,\overline{A_2}\cdots\overline{A_{k-1}}\right)$$

$$=\frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-k+1}{n-k+2}\frac{1}{n-k+1}=\frac{1}{n}.$$

思考：n张彩票中有m张奖券，第k个人买到奖券的概率是多少？

****

3. 贝叶斯（Bayes）公式

设$B_1,B_2,\cdots,B_n$是样本空间Ω的一个分割，即$B_1,B_2,\cdots,B_n$互不相容，且$\bigcup_{i=1}^{n}B_i=\Omega$。

如果$P(A)>0$，$P(B_i)>0\ \ (i=1,2,\cdots,n)$，

则$P(B_i\mid A)=\dfrac{P(B_i)P(A\mid B_i)}{\sum_{j=1}^{n}P(B_j)P(A\mid B_j)}$.

证明：$$P(B_i\mid A)=\frac{P(AB_i)}{P(A)}$$

$$=\frac{P(B_i)P(A\mid B_i)}{P(B_1)P(A\mid B_1)+P(B_2)P(A\mid B_2)+\cdots+P(B_n)(A\mid B_n)}$$

****

例 2.2.4 某考生回答一道有 4 个选项的选择题，设会答该题的概率是 p，并且会答时一定能答对，若不会答时则在 4 个答案中任选 1 个。求该考生回答正确时他确实会答的概率。

解：设事件 A 表示“答对”，B 表示“会答”，则

$$P(B|A)=\frac{P(B)P(A|B)}{P(B)P(A|B)+P(\overline{B})P(A|\overline{B})}$$

$$=\frac{p\cdot 1}{p\cdot 1+(1-p)\cdot\frac{1}{4}}$$

$$=\frac{4p}{1+3p}.$$

****

例 2.2.5 一地区某疾病的发病率是 0.0004。现有一种化验方法，对真正患病的人，其化验结果 99%呈阳性，对未患病者，化验结果 99.9%呈阴性。求下列两事件的发生概率：

1. 检查结果呈阳性，是否真的患病？

2. 检查结果呈阴性，是否就可以放心地认为自己没有病？

解：设事件 A 表示“化验呈阳性”，B 表示“患病”，则

检查结果呈阳性，但实际上没有患病（虚惊一场的概率）

$$P(\overline{B}\mid A)=\frac{P(\overline{B})P(A\mid\overline{B})}{P(B)P(A\mid B)+P(\overline{B})P(A\mid\overline{B})}=\frac{0.9996\times 0.001}{0.0004\times 0.99+0.9996\times 0.001}=0.716$$

检查结果呈阴性，但事实上是患了病的概率（患病但没有查出来的概率）

$$P(B\mid\overline{A})=\frac{P(B)P(\overline{A}\mid B)}{P(B)P(\overline{A}\mid B)+P(\overline{B})P(\overline{A}\mid\overline{B})}=\frac{0.0004\times 0.01}{0.0004\times 0.01+0.9996\times 0.999}=4\times 10^{-6}.$$

****

正确地使用三个公式，要把握好乘法公式和贝叶斯公式中包含的时间因素。乘法

公式按照时间的顺序过程展开，A1 首先发生，然后依次是 A2，A3 到 An。贝叶斯公式是逆概率公式，它将结果为条件的概率转化为以原因为条件的概率计算。而全概率公式是分情况讨论，而且必须考虑到所有可能，不能有遗漏，所以要求对样本空间进行分割。

****

2.3 事件的独立性

事件的独立性

事件独立是指互不影响：$P(A|B)=P(A)$　　$P(B|A)=P(B)$

条件概率：$P(A|B)=\dfrac{P(AB)}{P(B)}$　$P(B|A)=\dfrac{P(AB)}{P(A)}$

定义. 对于事件 A,B，如果 $P(AB)=P(A)P(B)$，则称事件 A,B 相互独立，

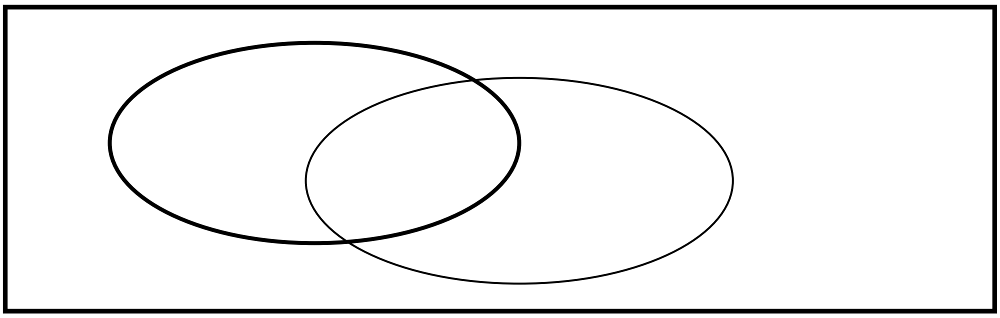
简称 A 与 B 独立，否则称 A 与 B 不独立或相关。

****

例 2.3.1 若事件 A 和 B 独立，则 A 与 $\overline{B}$ 独立，$\overline{A}$ 与 B 独立，$\overline{A}$ 与 $\overline{B}$ 独立。

$$P(A\overline{B})=P(A)-P(AB)=P(A)-P(A)P(B)$$

$$=P(A)(1-P(B))=P(A)P(\overline{B}).$$



利用“事件 A 和 B 独立，则 A 与 $\overline{B}$ 独立”的结果：

将 A 和 B 互换位置，则得到 $\overline{A}$ 与 B 独立

若 $\overline{A}$ 与 B 独立，则 $\overline{A}$ 与 $\overline{B}$ 独立。

****

例 2.3.2　3 人独立破译密码，他们单独能破译的概率分别为 $\frac{1}{5}$，$\frac{1}{3}$，$\frac{1}{4}$，试求此密码能被破译出的概率.

解：设事件 $B=\{$该密码被破译$\}$，$\bar{B}=\{$该密码未被破译$\}$.

设事件 $A_i=\{$第 i 个人能破译密码$\}$ $(i=1,2,3)$，则 $\bar{B}=\overline{A_1}\,\overline{A_2}\,\overline{A_3}$

$$P(B)=1-P(\bar{B})=1-P\left(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\right)=1-P\left(\overline{A_1}\right)P\left(\overline{A_2}\right)P\left(\overline{A_3}\right)$$

$$=1-\left(1-\frac{1}{5}\right)\times\left(1-\frac{1}{3}\right)\times\left(1-\frac{1}{4}\right)=\frac{3}{5}.$$

****

分赌本的例子

这是一个历史上曾经发生过的，一个很出名的问题叫做分赌本问题甲乙两个徒进行一场 9 局 5 胜制的赌博，先赢 5 局者获胜，假设每一局，都能分出胜负甲乙各压本金 100 元获胜方获得全部的 200 元本金那么这个规则非常明确　，应该没有任何问题但问题是呢如果当赌博进行到, 甲 3 比 1 领先的时候被迫中止了那么这 200 元本金该如何分配

****

多个事件相互独立，三个事件相互独立

$$P(A\mid BC)=P(A),\ P(C\mid A\cup B)=P(C),P(AB\mid C)=P(AB),\cdots\cdots$$

三个事件相互独立的定义对于 A,B,C 三个事件，如果它们之间两两独立，即：

$$P(AB)=P(A)P(B),\quad P(AC)=P(A)P(C),\quad P(BC)=P(B)P(C),$$

且 $P(ABC)=P(A)P(B)P(C)$，则事件 A,B,C 相互独立。

注：　A,B,C 两两独立不能保证 A,B,C 相互独立。

****

A,B,C 两两独立，但 A,B,C 不相互独立的例子

例 2.3.3 将一个正四面体，三个面分别涂红色、黄色和蓝色，剩下一个面涂上红、黄、蓝三色。

设事件 A,B,C 分别表示“将四面体投掷一次，底面含有红、黄、蓝色”　。

$P(A)=P(B)=P(C)=\frac{1}{2}$，$P(AB)=P(AC)=P(BC)=\frac{1}{4}\Rightarrow A,B,C$ 两两独立，

$P(A\mid BC)=1\neq P(A)$ 所以事件 A,B,C 不是相互独立的

$P(ABC)=\frac{1}{4}\neq P(A)P(B)P(C)=\frac{1}{8}$。

****

2.4 应用实例

应用实例一. 研究生招生是否有性别歧视?

1973 年,共有 8442 男生，4321 女生申请加州大学 Berkeley 分校的研究生院。最终男生录取比例大约 44%，女生录取比例大约 35%。

Science， Vol.187，398-404，7 February 1975，
Sex Bias in Graduate Admissions: Data from Berkeley
P. J. Bickel, E. A. Hammel, J. W. O'Connell

加州大学 Berkeley 分校 6 个最大专业的研究生入学资料

	男（1198/2691）		女（557/1835）	
专业	申请人数	录取百分比	申请人数	录取百分比
A	825	62	108	82
B	560	63	25	68
C	325	37	593	34
D	417	33	375	35
E	191	28	393	24
F	373	6	341	7

观察数据

1. A、B 两个专业容易考取。51.5%的男生申请，女生申请率只有 7.25%，

2. 其他四个专业较难考取，90%以上的女生申请这四个专业。

简单的看入学率是不合理的，简单的看各系的录取率同样不全面。更合理的考察应该是加权入学率，即综合考虑到各系的规模和录取率。

专业	A	B	C	D	E	F
申请人数	933	585	918	792	584	714
申请比例	0.21	0.13	0.20	0.18	0.13	0.16
男生录取率	62%	63%	37%	33%	28%	6%
女生录取率	82%	68%	34%	35%	24%	7%

男生的加权平均入学率:

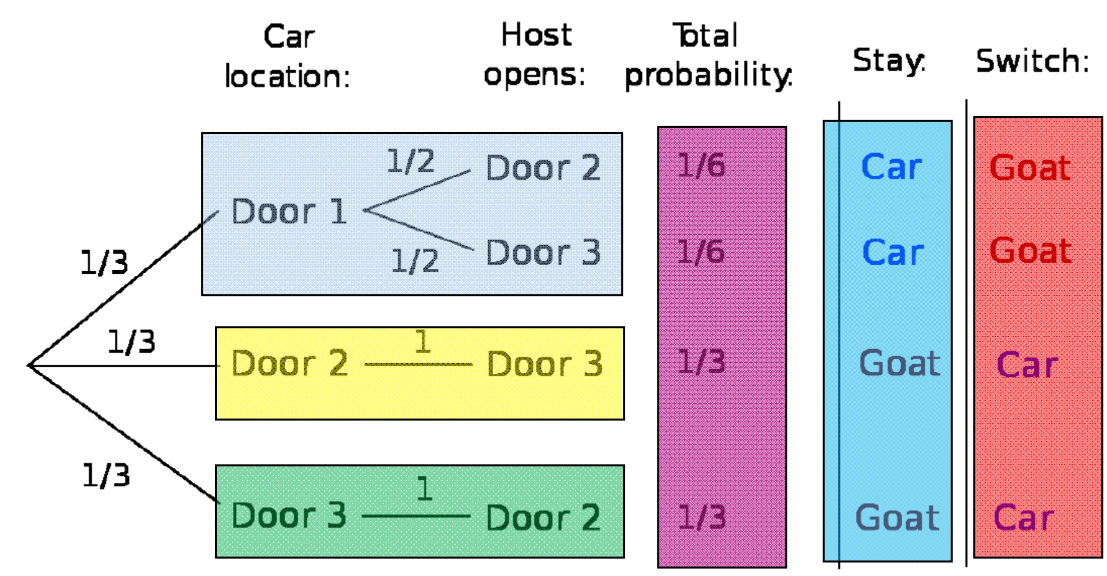
$$0.62\times0.21+0.63\times0.13+0.37\times0.20+0.33\times0.18+0.28\times0.13+0.06\times0.16\approx0.39$$

女生的加权平均入学率:

$$0.82\times0.21+0.68\times0.13+0.34\times0.20+0.35\times0.18+0.24\times0.13+0.07\times0.16\approx0.43$$

****

树形分叉图



Tree showing the probability of every possible outcome if the player initially picks door 1

应用实例二. Monty Hall 问题，要不要换门儿？

一个电视游戏节目，主持人在现场准备三扇门，分别编号为 1，2，3，并且事先随

机地在两扇门后各放一只羊，另一扇门后放汽车。节目开始后，主持人让参与互动的 1 名观众任选一门，然后在剩下的两扇门中打开一个有羊的，问此时尚未打开的两扇门中有汽车的概率分别是多少？

利用贝叶斯公式求解：

令 A_1, A_2, A_3 分别表示事件“汽车在 1，2，3 号门后”。

假设第一次选择了 1 号门，B 表示打开的是 2 号门。

要计算 $P(A_1 \mid B)$，$P(A_3 \mid B)$

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}, \quad P(B \mid A_1) = \frac{1}{2}, \quad P(B \mid A_2) = 0, \quad P(B \mid A_3) = 1,$$

根据全概率公式可得

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3) = \frac{1}{2},$$

所以 $P(A_1 \mid B) = \dfrac{P(A_1)P(B \mid A_1)}{P(B)} = \dfrac{1}{3}$，$P(A_3 \mid B) = \dfrac{P(A_3)P(B \mid A_3)}{P(B)} = \dfrac{2}{3}$。