$$\Omega \Rightarrow (\Omega, \mathbb{F}) \Rightarrow (\Omega, \mathbb{F}, \mathrm{P})$$

条件概率 $(\boldsymbol{A}, \mathbb{F}_A, \mathrm{P}_A)$

第二章 条件概率与统计独立性

第一节 条件概率，全概率公式，贝叶斯公式

一 条件概率

我们知道, 对任何概率问题的讨论，都必须建立起一个概率空间$(\Omega,\mathbb{F},\mathrm{P})$, 并且是在$\boldsymbol{P}(\Omega)=\mathbf{1}$的前提下进行讨论. 但有时除了这个总前提之外, 还会出现附加前提. 例如，抛掷一枚均匀的骰子, 已知掷出的点数为奇数(记为事件$\boldsymbol{B}$)，要求出点数大于$\mathbf{1}$（记为事件$\boldsymbol{A}$)的概率. 显然，这个附加前提“掷出的点数为奇数”会影响事件$\boldsymbol{A}$发生的概率.

下面求“在事件B发生的前提下，事件A发生的概率”——以$P(A|B)$表示，称之为条件概率.

$$\Omega=\{1,2,3,4,5,6\},\quad B=\{1,3,5\},\quad A=\{2,3,4,5,6\}.$$

由于B已经发生，因此掷出的所有可能结果只有3种情况：点数为1, 点数为3, 点数为5. 若事件A再发生，则事件$AB=\{3,5\}$发生. 这相当于在样本空间$\Omega'=B=\{1,3,5\}$中, 求事件AB发生的概率. 这也是古典概型问题，因此

$$P(A|B)=\frac{|AB|}{|B|}=\frac{2}{3}$$

那么在原概率空间Ω中，如何求$P(A|B)$呢？有

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

这个例子启发我们：可以以$P(AB)$与$P(B)$之比作为条件概率$P(A|B)$的定义.

定义2.1.1 设$(\Omega,\mathbb{F},\mathrm{P})$是一个概率空间，$B\in\mathbb{F}$，而且$P(B)>0$，则对任意$A\in\mathbb{F}$，记

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

并称$P(A|B)$为在事件B发生的前提下，事件A发生的条件概率.

在具体计算$P(A|B)$时, 可以用公式右端来求, 也可以像刚才的例子那样，直接从缩小了的样本空间来求，后一种求法有时更方便、实用.

可验证上述定义的$P(A|B)$满足非负性、规范性、可列可加性.

（1）非负性　对任何事件A，有 $P(A|B)\geq 0;$

（2）规范性　$P(\Omega|B)=1;$

（3）可列可加性　若$A_1,A_2,\cdots$两两互不相容，则

$$P\left(\sum_{n=1}^{+\infty}A_n\Big|B\right)=\sum_{n=1}^{+\infty}P(A_n|B).$$

因此$\left(\Omega,\mathbb{F},P(\cdot|B)\right)$也是概率空间，对条件概率$P(\cdot|B)$也可由三条基本性质导出概率的其他性质.

【例1】某班级共有100名同学，其中男生60名，女生40名．在期末的高等数学考试中有8人考试不及格，其中有6名男生，2名女生. 若A表示同学是男生这一事件, B表示考试不及格这一事件. 试求

(1) 条件概率$P(B|A)$，$P(A|B)$，$P(\overline{A}|B)$.

(2) 条件概率$P(B|A)$，$P(B)$的含义, 并比较大小.

解：(1) $P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{6/100}{60/100}=10\%$,

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{6/100}{8/100}=75\%.$$

$$P(\bar{A}|B)=1-P(A|B)=25\%.$$

(2)条件概率$P(B|A)$表示男生的不及格率，$P(B)$表示班级的不及格率．由于

$$P(B|A)=10\%,\ \ P(B)=8\%.$$

因而$P(B|A)>P(B)$，这表明男生的不及格率高于班级的不及格率.

下面给出条件概率特有的三个非常实用的公式：乘法公式、全概率公式、贝叶斯公式．这些公式可以帮助我们计算一些复杂事件的概率.

二　乘法公式

根据条件概率的定义，不难得下列乘法公式：

定理2.1.1 设A,B与$A_1,A_2,\cdots,A_n$都是概率空间$(\Omega,\mathbb{F},\mathrm{P})$中的事件．则

（1）$P(AB)=P(B)P(A|B)$（这里要求$P(B)>0$）.

$P(AB)=P(A)P(B|A)$（这里要求$P(A)>0$）.

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

（这里要求$P(A_n|A_1A_2\cdots A_{n-1})>0$）

证明根据条件概率的定义，即得（1）
由于

$$P(A_1) \geq P(A_1A_2) \geq \cdots \geq P(A_1A_2\cdots A_{n-1}) > 0.$$

因此(2)中的各条件概率均有意义，根据条件概率的定义，(2)式的右边等于

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

【例2】（坛子模型）设坛子中有a个白球及b个黑球, 每次随机取出一个球，取出后将原球放回, 并加入c个同色球和d个异色球. 再摸第二次，这样下去共摸了n次. 记A_i表示第 i 次取出的是白球，B_j表示第j次取出的黑球，试求.

（1）$P(A_1B_2B_3)$，$P(B_1A_2B_3)$，$P(B_1B_2A_3)$,

（2）$P(A_1A_2\cdots A_{n_1}B_{n_1+1}B_{n_1+2}\cdots B_n)$.

解　(1) $P(A_1B_2B_3)=P(A_1)P(B_2|A_1)P(B_3|A_1B_2)$

$$=\frac{a}{a+b}\cdot\frac{b+d}{a+b+c+d}\cdot\frac{b+d+c}{a+b+2c+2d}.$$

$$P(B_1A_2B_3)=P(B_1)P(A_2|B_1)P(B_3|B_1A_2)$$
$$=\frac{b}{a+b}\cdot\frac{a+d}{a+b+c+d}\cdot\frac{b+d+c}{a+b+2c+2d},$$
$$P(B_1B_2A_3)=P(B_1)P(B_2|B_1)P(A_3|B_1B_2)$$
$$=\frac{b}{a+b}\cdot\frac{b+c}{a+b+c+d}\cdot\frac{a+2d}{a+b+2c+2d}.$$

以上概率与白球在第几次被取出有关.

（2） $P(A_1)P(A_2|A_1)\cdots P(A_{n_1}|A_1A_2\cdots A_{n_1-1})$

$\cdot P(B_{n_1+1}|A_1A_2\cdots A_{n_1})\cdots\cdot P(B_n|A_1A_2\cdots A_{n_1}B_{n_1+1}\cdots B_{n-1})$

$$=\prod_{k=0}^{n_1-1}\left(\frac{a+kc}{a+b+k(c+d)}\right)\cdot\prod_{k=n_1}^{n-1}\left(\frac{b+n_1d+(k-n_1)c}{a+b+k(c+d)}\right).$$

坛子模型也称为波利亚模型，这个模型有如下各种变化：

（1） 当$c=-1, d=0$时，即为不返回抽样. 此时前次抽取结果会影响后次的抽样结果. 但只要抽取的黑球与红球个数一样，则概率与抽出球的次序无关, 它们都是一样的，本例中有

$$P(A_1B_2B_3)=P(B_1A_2B_3)=P(B_1B_2A_3)$$
$$=\frac{ab(b-1)}{(a+b)(a+b-1)(a+b-2)}.$$

（2） 当$c=0, d=0$时，即为返回抽样. 此时前次抽取结果会影响后次的抽样结果. 但只要抽取的黑

球与红球个数一样，则概率与抽出球的次序无关，它们都是一样的，本例中有

$$P(A_1B_2B_3) = P(B_1A_2B_3) = P(B_1B_2A_3)$$

$$= \frac{ab^2}{(a+b)^3}.$$

(3) 当$c > 0, d = 0$时，即为传染病模型. 此时每次取出球后就会增加下一次取到同色球的概率，或换句话说，每次发现一个传染病患者，以后都会增加再传染的概率. 与(1)、(2)一样，以上三个概率相等.

$$P(A_1B_2B_3) = P(B_1A_2B_3) = P(B_1B_2A_3)$$

$$= \frac{ab(b+c)}{(a+b)(a+b+c)(a+b+2c)}.$$

（4）当$c=0, d>0$时，称为安全模型．此模型可解释为：每当事故发生了(红球被取出)，安全工作就抓紧一些，下次再发生的概率就会减少；而当事故没有发生时(黑球被取出)，安全工作就放松一些，下次再发生事故的概率就会增大．在这种场合，上述三个概率分别为

$$P(A_1B_2B_3)=\frac{a}{a+b}\cdot\frac{b+d}{a+b+d}\cdot\frac{b+d}{a+b+2d},$$

$$P(B_1A_2B_3)=\frac{b}{a+b}\cdot\frac{a+d}{a+b+d}\cdot\frac{b+d}{a+b+2d},$$

$$P(B_1B_2A_3)=\frac{b}{a+b}\cdot\frac{b}{a+b+d}\cdot\frac{a+2d}{a+b+2d}.$$

$$\Omega \Longrightarrow (\Omega,\ \mathbb{F}) \Longrightarrow (\Omega,\mathbb{F},\mathrm{P})$$

条件概率 $\left(\boldsymbol{A},\mathbb{F}_{A},\mathrm{P}_{A}\right)$

【前例】某班级共有100名同学，其中男生60名，女生40名.在期末的高等数学考试中有8人考试不及格，其中有6名男生，2名女生. 若A表示同学是男生这一事件，B表示考试不及格这一事件.

(无条件概率)$P(B)$表示班级的不及格率，

条件概率$P(B|A)$表示男生的不及格率，

条件概率$P(B|\bar{A})$表示女生的不及格率.

(1) 三种概率之间的关系?——全概率公式

(2) 由$P(B|A) \Rightarrow\Rightarrow P(A|B)$?——贝叶斯公式

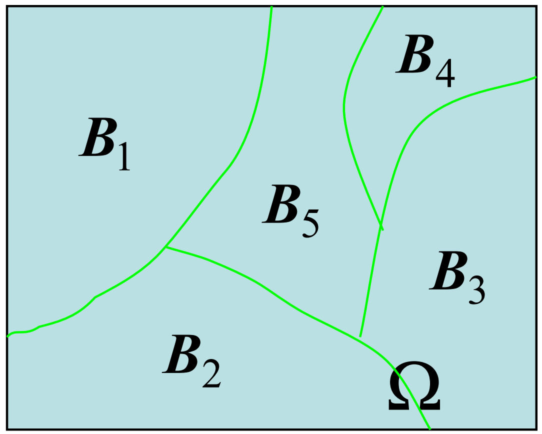
三　全概率公式　（加法公式+乘法公式）

全概率公式是概率论的一个重要公式，它提供了计算复杂事件概率的一条有效途径，使一个复杂事件的概率计算问题化繁为简.

定义2.1.1　设$B_1, B_2, \cdots, B_n, \cdots$是概率空间$(\Omega, \mathbb{F}, \mathrm{P})$中的一组事件，若它们两两互不相容，而且

$$\sum_{k=1}^{\infty} B_k = \Omega.$$

则称它们是样本空间Ω的一个分割，亦称完备事件组.

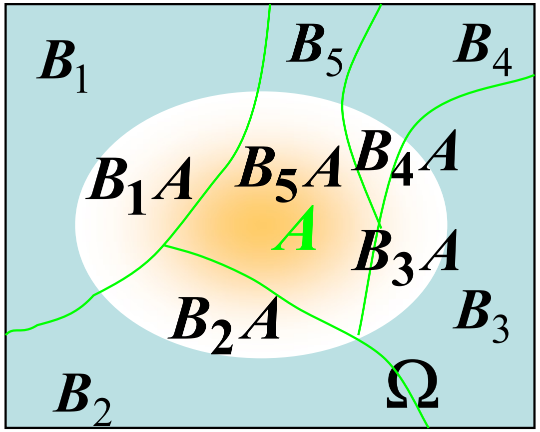


定理2.1.2 全概率公式 设$(\Omega,\mathbb{F},\mathrm{P})$是概率空间，$B_1,B_2,\cdots,B_n,\cdots$是$\Omega$的一个分割，则对$\mathbb{F}$中的任何事件$A$，有

$$P(A)=\sum_{k=1}^{\infty}P(B_k)P(A|B_k).$$

全概率公式

证明 由于



$$A=A\Omega=A\sum_{k=1}^{\infty}B_k=\sum_{k=1}^{\infty}AB_k$$

利用加法公式和乘法公式得

$$P(A)=\sum_{k=1}^{\infty}P(AB_k)=\sum_{k=1}^{\infty}P(B_k)P(A|B_k)$$

【例3】盒中装有5个乒乓球，其中3个是新球2个是旧球，比赛时从中任取一球，用后不放回，求第二次取出新球的概率.

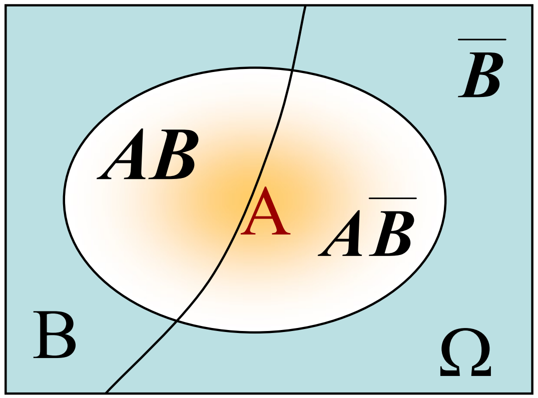
解：设A表示事件“第二次取出新球”，

B表示事件“第一次取出新球”.

则 $B,\bar{B}$是样本空间Ω的一个分割. 因而

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

$$=\frac{3}{5}\times\frac{2}{4}+\frac{2}{5}\times\frac{3}{4}=\frac{3}{5}.$$



无条件概率是条件概率的加权平均.

【例4】某工厂的第一、第二、第三号车间生产同一种产品，产量各占总产量的50%,30%,20%. 次品率分别为1%,1.5%,2%. 现从该厂产品中随机地抽取一件，试求该产品是次品的概率.

解：以B_1, B_2, B_3分别表示 “产品是由第一、第二、第三号车间生产”,则它们是样本空间的一个分割，以A表示“产品是次品”，则

$$P(A)=\sum_{k=1}^{3}P(B_k)P(A|B_k)$$

$$=0.5\times 0.01+0.3\times 0.015+0.2\times 0.02$$

$$=0.0135$$

【例5】(坛子模型)设坛子中有a个白球及b个黑球，每次随机取出一个球，取出后将原球放回，并加入c个同色球，如此反复进行. 试证明：在第n次取球时取出白球的概率为$\dfrac{a}{a+b}$.

证明：以A_k表示在第k次取球时取出白球的事件，则$\overline{A}_k$表示在第k次取球时取出黑球的事件.

我们对n作归纳，显然有

$$P(A_1)=\frac{a}{a+b}$$

假设$n=k-1$, $P(A_{k-1})=\dfrac{a}{a+b}$

$k>1$时结论成立，要证$n=k$结论成立.

以A_1与$\overline{A}_1$作为对样本空间Ω的分割.

此时$P(A_k|A_1)$可看作是从放有$a+c$个白球和b个黑球的坛子中按规则取球, 并且在第$k-1$次取球时取出白球的概率.

因此由归纳假设知

$$P(A_k|A_1)=\frac{a+c}{a+b+c}, \quad P(A_k|\bar{A}_1)=\frac{a}{a+b+c}.$$

于是根据全概率公式得

$$P(A_k)=P(A_{k-1})P(A_k|A_{k-1})+P(\bar{A}_{k-1})P(A_k|\bar{A}_{k-1})$$

$$=\frac{a}{a+b}\cdot\frac{a+c}{a+b+c}+\frac{b}{a+b}\cdot\frac{a}{a+b+c}=\frac{a}{a+b}.$$

因此结论对一切n成立.

思考：

上述解答中如果用A_{k-1}与$\overline{A}_{k-1}$对Ω进行分割是否可行？为什么？

练习题：

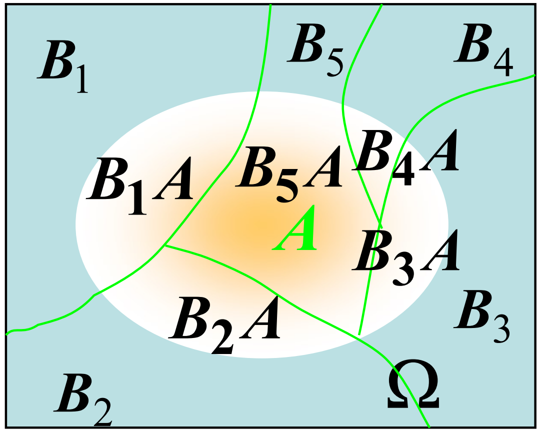
（摸彩模型）设在n张彩票中有一张奖券．求第二个人摸到奖券的概率．

四　贝叶斯公式

定理2.1.3　贝叶斯公式　设$(\Omega,\mathbb{F},\mathrm{P})$是概率空间，$B_1,B_2,\cdots,B_n,\cdots$是$\Omega$的一个分割，$P(B_i)>0$，$i=1,2,\cdots$，则对$\mathbb{F}$中的任何事件$A$，有

Bayes公式

$$P(B_k|A)=\frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{\infty}P(B_i)P(A|B_i)}.\qquad k=1,2,\cdots$$



证明　由条件概率的定义知

$$P(B_k|A)=\frac{P(AB_k)}{P(A)},$$

上式的分子用乘法公式，分母用全概率公式即得证.

贝叶斯公式在概率论与数理统计中有着多方面的应用，假定$B_1, B_2, \cdots$是导致试验结果的“原因”，称$P(B_i)$先验概率，它反映了各种原因发生的可能性的大小，一般是以往经验的总结，在这次试验前.已经知道, 现在若试验产生了事件A，这个信息将有助于探讨事件发生的“原因”，条件概率$P(B_j \mid A)$称为后验概率，它反映了试验之后对各种“原因”发生的可能性大小的新知识，例如：在医疗诊断中，有人为了诊断病人到底是患了$B_1, B_2, \cdots$中的那一种病, 对病人进行观察与检查，确定了某个指标(譬如体温、

脉搏、转氨酶含量等)想用这类指标来帮助诊断, 这时可以用贝叶斯公式来计算有关概率，首先必须确定先验概率$P(B_i)$，这实际上是确定患各种疾病的大小, 以往资料可以给出一些初步数据(称为发病率)，其次要确定$P(A|B_i)$. 这当然要依靠医学知识，在实际工作中检查指标A一般有多个，综合所有的后验概率，会对诊断有很大的帮助，在实现计算机自动诊断或辅助诊断中，这种方法有是实用价值的.

【例6】 某地区居民的肝癌发病率为0.0004，现用甲胎蛋白法进行普查.医学表明，化验结果是存有错误的. 已知患有肝癌的人其化验结果99%呈现阳性(有病)，而没有患肝癌的人其化验结果99.9%呈现阴性(无病. 现某人的化验结果呈现阳性，问他真的患肝癌的概率是多少？

解 记A为事件“被检查者患有肝癌”,

B为事件“化验结果呈现阳性”.

由题设知

$$P(A)=0.0004,\quad P(\overline{A})=0.9996.$$

$$P(B|A)=0.99,\quad P(B|\overline{A})=1-0.999=0.001.$$

由贝叶斯公式得

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

$$=\frac{0.0004\times0.99}{0.0004\times0.99+0.9996\times0.001}=0.284.$$

这表明，在化验结果呈现阳性的人中，真患有肝癌的人不到30%.

在实际应用中，常采用复查的方法来减少错误率，比如上例中，可对首次检查得阳性的人群再进行复查，此时

$$P(A)=0.284.$$

若复查仍然显阳性，这时再用贝叶斯公式得

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

$$=\frac{0.284\times 0.99}{0.284\times 0.99+0.716\times 0.001}=0.997.$$

这就大大提高了甲胎蛋白法的准确率了. 复查结果呈现阳性的人中，真患有肝癌的人达到了99.7%.

思考题:

（1）第一次检查得阴性，复查结果仍是阴性的人中，真正患有肝癌的概率？

（2）第一次阳性，第二次阴性与第一次阴性、第二次阳性的相比，谁患病的概率大？

【例7】(遗传风险)在人类遗传学中,某中坏的基因会引起夭折.设a是这样的一个基因，基因型aa将不能长大成人，基因型Aa的人称为带菌者(a具有隐性性状).假定在一般总体中(不论性别如何)带菌者的概率为p. 现考察下述问题:

（1）已知某成年男子有一个哥哥或姐姐在童年夭折(原于aa). 求该男子为带菌者的概率.

（2）若该男子跟一个女人结婚(不知她的家庭遗传情况)，问其子代的基因型为AA, Aa, aa的概率各是多少?

解：首先，由于他有一个哥哥或姐姐在童年夭折，他的双亲都是带菌者(为什么?)，因此他们孩子的基因型的分布为

基因型	AA	Aa	aa
p	1/4	1/2	1/4

于是，所求概率为

$$P(\text{带菌者}\mid\text{成人})=\frac{P(Aa)}{P(AA\cup Aa)}=\frac{1/2}{1/4+1/2}=\frac{2}{3},$$

$$P(\text{不带菌者}\mid\text{成人})=1-\frac{2}{3}=\frac{1}{3}.$$

（2）若该人跟一个女人结婚(不知她的家庭遗传情况)，则基因结合的类型见下表

父本	母本	结合的概率	产生AA的概率	Aa 的概率	aa的概率
AA	AA	$\frac{1}{3}(1-p)$	1	0	0
AA	Aa	$\frac{1}{3}p$	$\frac{1}{2}$	$\frac{1}{2}$	0
Aa	AA	$\frac{2}{3}(1-p)$	$\frac{1}{2}$	$\frac{1}{2}$	0
Aa	Aa	$\frac{2}{3}p$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

由全概率公式：

$$P(子AA)=\frac{1}{3}(1-p)+\frac{p}{6}+\frac{1}{3}(1-p)+\frac{p}{6}=\frac{2}{3}-\frac{p}{3},$$

$$P(子Aa)=\frac{1}{3}(1-p)+\frac{p}{6}+\frac{p}{3}=\frac{1}{3}+\frac{p}{6}.$$

$$P(子aa)=\frac{p}{6}.$$

由此可知，在这种背景下，就子一代而言，一个成人是带菌者的概率为

$$P(子Aa\,|\,子AA\cup子Aa)=\frac{\frac{1}{3}+\frac{p}{6}}{\frac{2}{3}-\frac{p}{3}+\frac{1}{3}+\frac{p}{6}}=\frac{2+p}{6-p}\approx\frac{1}{3}.$$

补充练习

1 假设有三张形状完全相同的卡片，第一张两面全是红色，第二张两面全是黑色，第三张一面红色一面黑色，将这三张卡片随机地选出一张，并抛在桌面上，发现这张卡片朝上的一面为红色，求其另一面是黑色的概率.

2 伊索寓言“孩子与狼”讲的是一个小孩每天到山上放羊,山里有狼出没. 第一天, 他在山上喊: “狼来了, 狼来了”山下的村民闻声便去打狼, 可到山上, 发现狼并没来, 第二天，仍是如此；第三天狼真的来了, 可无论小孩怎么喊叫, 也没有人来救他. 现在请你用**Bayes**公式来分析此寓言中村民对小孩的可信程度是如何下降的.


1
狼來了！
2
農人們趕來救羊，卻發現被騙了。
3
哈哈！騙你們的
4
狼來了！

A ={小孩说谎}，　B ={小孩可信}

原来　$P(B) = 0.8 \qquad P(\overline{B}) = 0.2$

假设　$P(A|B) = 0.1 \qquad P(A|\overline{B}) = 0.6$

第一次说谎后

$$P(B|A) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})} = \frac{0.8\times 01}{0.8\times 0.1 + 0.2\times 0.6} = 0.4$$

$$P'(B) = P(B|A) = 0.4, \quad P'(\overline{B}) = P(\overline{B}|A) = 0.6$$

第二次说谎后

$$P''(B) = P'(B|A) = 0.1 \quad P''(\overline{B}) = P'(\overline{B}|A) = 0.9$$

这表明村民们经过两次上当，对这个小孩的可信程度已经从原来的0.8下降到了0.1.如此低的可信度，村民们听到第三次呼叫时就不会再上山打狼了.

贝叶斯公式虽然很简单，但是它却很有哲理意义. 这个公式是以**18**世纪英国哲学家贝叶斯命名的，贝叶斯本人并不专门研究概率统计，只不过对统计问题感兴趣而已. 他生前并没有发表这个公式，而是在他去世后两年，由他的一个朋友整理遗物时从他的笔记中发现后发表的.

我们可以这样来理解这个公式：假设某个过程具有n个可能的前提（原因）$B_1, B_2, \ldots, B_n$. 而$P(B_1)$, $P(B_2), \cdots, P(B_n)$是人们对这n个可能的前提（原因）的概率的事前估计，称之为先验概率. 当这个过程有了一个结果之后，人们便会通过条件概率.

$$P(B_1|A), P(B_2|A), \cdots, P(B_n|A)$$

来对这个可能前提的可能性大小的做出一种新的认识，因此将这些条件概率称之为后验概率. 而贝叶斯公式恰恰提供了一种计算后验概率的工具. 后来从这种先验概率和后验概率的理念中发展出了一整套统计理论和方法，并形成了概率统计的一个很大的学派——贝叶斯学派.

贝叶斯学派对概率统计问题有自己的独特理解，在处理许多问题时有自己的独到之处，给出了许多的统计方法，后来发展出所谓经验贝叶斯方法，在计算机广为应用的今天，贝叶斯方法和经验贝叶斯方法的使用价值大大提高.