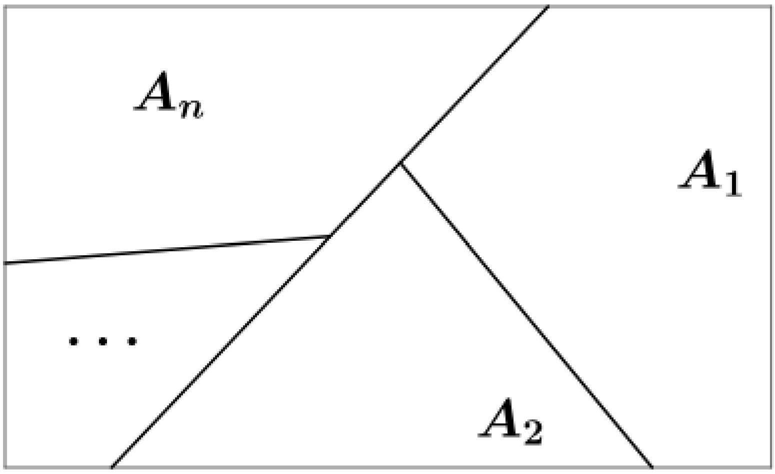
1.4.3 全概率

1.4.3.1 全概率公式

形式（1）

设A_1、A_2、$\cdots$、A_n为样本空间Ω的一个分割：



用代数表示就是：

$$A_i \cap A_j = \varnothing, (i \neq j) \quad 且 \quad \bigcup_{i=1}^{n} A_i = \Omega$$

若$P(A_i) > 0, i = 1, 2, \cdots, n$，则对任意事件$B$有：

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

形式（2）

设A_1、A_2、$\cdots$、A_n满足：

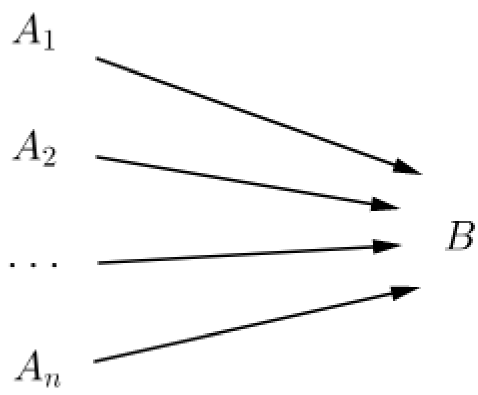
$$A_i \cap A_j = \varnothing, (i \neq j) \quad 且 \quad P(\bigcup_{i=1}^{n} A_i) = 1$$

若 $P(A_i) > 0, i = 1, 2, \cdots, n$，则对任意事件$B$有：

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B|A_i)$$

1.4.3.2 全概率公式通俗解释

以把分割A_1、A_2、$\cdots$、A_n看作导致B可能发生的各种各样的条件：



B发生的概率可以分解到各种条件上去，这就是全概率公式的通俗解释。

> 例题
>
> 袋中有50个乒乓球，其中20个是黄球，30个是白球，今有两人依次随机地从袋中各取一球，取后不放回，则第二个人取得黄球的概率是多少？
>
> 可以通过全概率公式来计算：
>
> $$P(C)=P(A)P(C|A)+P(\overline{A})P(C|\overline{A})$$
>
> $$=\frac{20}{50}\times\frac{19}{49}+\frac{30}{50}\times\frac{20}{49}$$
>
> $$=\frac{2}{5}$$
>
> 这个结论说明先抽、后抽对概率没有影响。
>
> 从全概率公式来看还是比较明显，全概率公式有两个分支：
>
> - 第一个分支是A事件的条件下，这个时候抽黄球的概率会减小
> - 第二个分支是B事件的条件下，这个时候抽黄球的概率会增大
>
> **两相抵消的结果就是顺序对概率没有影响。**
>
> **不过如果第一个人抽完之后公示了他的结果，那么就会对第二个人的概率产生影响，这相当于只走了全概率公式中的一个分支。**

1.4.4 贝叶斯与全概率

有了全概率公式后，可以得到贝叶斯定理真正的样子：

设A_1、A_2、$\cdots$、A_n为样本空间Ω的一个分割，则有：

$$
\begin{aligned}
P(A_i|B) &= \frac{P(BA_i)}{P(B)} \\
&= \frac{P(B|A_i)}{P(B)}P(A_i) \\
&= \frac{P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}P(A_i)
\end{aligned}
$$

也就是把$P(B)$分解到分割A_1、A_2、$\cdots$、A_n上去了。

之前已经是这么计算的了，当时没提是为了减小理解的难度。比如疾病检测问题，事件稍微重新定义下：

$$
A_1 = \text{“患病的人”}, \quad A_2 = \text{“没患病的人”}, \quad B = \text{“检测显示有病的人”}
$$

很显然，A_1、A_2就是一个分割，所以：

$$
\begin{aligned}
P(A_1|B) &= \frac{P(B|A_1)}{P(B)}P(A_1) \\
&= \frac{P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}P(A_1) \\
&= \frac{0.02}{0.02 \cdot 0.9 + 0.98 \cdot 0.1} \times 0.9 \\
&\approx 15.5\%
\end{aligned}
$$

> 例题
>
> 盒中有a个红球，b个黑球，今随机的从中取出一个，观察其颜色后放回
>
> 并加上同色球c个，再从盒中第二次取出一球，求第二次抽出黑球的概率与第一次抽出的概率是否相同？
>
> 设
>
> $$A = \{\text{第一次抽出的是黑球}\}, B = \{\text{第二次抽出的是黑球}\}$$
>
> 由题意得
>
> $$P(A) = \frac{b}{a+b}$$
>
> 则
>
> $$P(B|A) = \frac{b+c}{a+b+c}, P(B|\bar{A}) = \frac{b}{a+b+c}$$
>
> 因为$A, \bar{A}$是样本空间的一个划分，根据全概率公式可得

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

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

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

> 因此第二次抽出黑球的概率与第一次抽出的概率相同。

1.5 事件的独立性

1.5.1 严格定义

对于两个随机事件A、B，如果满足：

$$P(AB) = P(A)P(B)$$

则称A与B相互独立，或简称A与B独立，否则称A与B不独立或相依。

1.5.2 赌徒谬误

要是连续3次掷出正面，那下一次得到正面的概率有多高？

我们知道，连续出现4次正面的概率非常小：

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

有些赌徒据此认为，已经出现3次正面了，那么下一次也得到正面的概率就非常小了。这被称为赌徒谬误。

结合已经学习过的概率论知识，可以很好的反驳这个谬误。先画一下样本空间：

第一次	第二次	第三次	第四次
H	H	H	H
H	H	H	T
H	H	T	H
H	H	T	T
H	T	H	H
H	T	H	T
H	T	T	H
H	T	T	T
T	H	H	H
T	H	H	T
T	H	T	H
T	H	T	T
T	T	H	H
T	T	H	T
T	T	T	H
T	T	T	T

可以看到，总共有16种可能性，只有第一种可能性是四次都是正面，所以投掷出四次都是正面的概率确实是$\frac{1}{16}$。

投掷三次正面后，样本空间缩小了，变得只有两种可能性了（这其实是一个条件概率）：

第一次	第二次	第三次	第四次
H	H	H	H
H	H	H	T

所以第四次为正面的概率依然为$\frac{1}{2}$。

1.5.3 多个事件的独立性

三个事件

设A、B、C是三个事件，如果满足:
$$\begin{cases} P(AB)=P(A)P(B) \\ P(BC)=P(B)P(C) \\ P(AC)=P(A)P(C) \end{cases} \tag{1}$$

则称A、B、C是两两独立。若还有:

$$P(ABC)=P(A)P(B)P(C) \qquad (2)$$

则称A、B、C三者相互独立。

多个事件

设A_1、A_2、$\cdots$为有限个或者无限个事件，从中任取两个A_{i1}、A_{i2}，若满足：

$$P(A_{i1}A_{i2})=P(A_{i1})P(A_{i2})$$

则称A_1、A_2、$\cdots$是两两独立。

若从中任取有限个A_{j1}、A_{j2}、$\cdots$、A_{jm}，若满足：

$$P(A_{j1}A_{j2}\cdots A_{jm})=P(A_{j1})P(A_{j2})\cdots P(A_{jm})$$

则称A_1、A_2、$\cdots$是相互独立。

很显然：

$$\text{相互独立} \implies \text{两两独立}$$

而比较反直觉的是：

$$\text{相互独立} \not\Longleftarrow \text{两两独立}$$

> 例题
>
> 有两名选手比赛射击，轮流对同一目标进行射击，甲命中目标的概率为α,乙命中目标的概率为β。甲先射，谁先命中谁得胜。
>
> 问:甲获胜的概率是多少
>
> 记事件A_i为“第i次射击命中目标”,$i=1,2,\ldots$。因为甲先射，所以事件“甲获胜”可以表示为
>
> $$A_1\cup\bar{A}_1\bar{A}_2A_3\cup\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4A_5\cup\ldots$$
>
> 又因为各次射击是独立的，所以得
>
> $$P\{\text{甲获胜}\}=\alpha+(1-\alpha)(1-\beta)\alpha+(1-\alpha)^2(1-\beta)^2\alpha+\ldots$$
>
> $$=\alpha\sum_{0}^{\infty}(1-\alpha)^i(1-\beta)^i$$
>
> 根据等比数列求和,得
>
> $$\text{原式}=\alpha\lim_{n\to\infty}\frac{1-[(1-\alpha)(1-\beta)]^n}{1-(1-\alpha)(1-\beta)}$$
>
> 根据题意有$0<(1-\alpha)(1-\beta)<1$,所以
>
> $$\text{原式}=\frac{\alpha}{1-(1-\alpha)(1-\beta)}$$

2 随机变量及其分布

2.1 随机变量

2.1.1 随机变量的定义

定义在样本空间Ω上的实值函数:

$$X = X(\omega), \quad \omega \in \Omega$$

称为随机变量。

随机变量是一个函数，所以都用大写字母来表示，以示和自变量x的区别。

进一步规定，“随机变量X的取值为x”就是满足等式$X(\omega) = x$的一切ω组成的集合，简记为“$X = x$”，这是Ω的一个子集，即:

$$\text{“}X = x\text{”} = \{\omega : X(\omega) = x\} \subset \Omega$$

类似地，有:

$$\text{“}X \le x\text{”} = \{\omega : X(\omega) \le x\} \subset \Omega$$

2.1.2 引入随机变量的意义

数学最重要的特征就是抽象，引入随机变量就是《概率论与数理统计》这门课程进行抽象的重要步骤。

之后，本课程就主要研究随机变量以及相应的概率函数，并且只需要研究有限的类别，就可以覆盖生活、生产中大部分的需要。

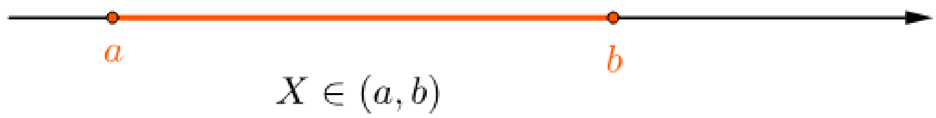
2.1.3 随机变量的离散与连续

对于随机变量:

$$X = X(\omega), \quad \omega \in \Omega$$

如果随机变量的函数值是实数轴上孤立的点（有限个或者无限个），则称为离散随机变量:

如果随机变量的函数值是实数轴上某个区间上所有的值（也可以是$(-\infty, +\infty)$区间），则称为连续随机变量:



2.2 二项分布

2.2.1 概率质量函数

有了随机变量之后，我们对概率的研究就变成了对一个个函数的研究。

2.2.1.1 概率质量函数定义

设X为离散型随机变量，其全部可能值为$x_1, x_2, \cdots$，则：

$$p_i = p(x_i) = P(X = x_i), \quad (i = 1, 2, \cdots)$$

称为X的概率质量函数（Probability Mass Function，缩写为PMF）。

上式还可写作列表的形式：

X	x_1	x_2	x_3	$\cdots$
P	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\cdots$

所以也称为X的概率分布列，或者简称为概率分布。

有时候也如下表示：

$$X \sim p(x)$$

读作X服从$p(x)$的概率分布。

2.2.1.2 概率质量函数严格定义

如果$p(x)$满足（$x \in \{x_i\}, i = 1, 2, \cdots$）：

- **非负性**：

$$p(x_i) \geq 0$$

- **规范性**：

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

则称其为概率质量函数（PMF）。

这两个性质决定了随机变量X不能随便选取。

2.2.2 伯努利分布

某样本空间只包含两个元素，$\Omega = \{\omega_1, \omega_2\}$，在其上定义随机变量$X$：

$$X = X(\omega) = \begin{cases} 1, & \omega = \omega_1 \\ 0, & \omega = \omega_2 \end{cases}$$

若$0 \leq p \leq 1$时，有：

$$p(1) = P(X = 1) = p$$

$$p(0) = P(X = 0) = 1 - p$$

或写作：

$$P(X = x) = p(x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

则此概率分布称作$0 - 1$分布，也称作伯努利分布，用以纪念数学家雅各布·伯努利.

2.2.3 n重伯努利试验

在数学中，类似于扔一次硬币这样的“是非题”称为一次伯努利试验，像上面这样独立地重复扔n次硬币（做同样的“是非题”n次），就称为n重伯努利试验。

注意，这里强调了要**独立**，也就是第n次扔硬币的概率不会受到前面的影响。

2.2.4 二项分布的定义

对于n重伯努利实验，如果每次得到“是”的概率为p，设随机变量：

$$X = \text{得到“是”的次数}$$

则称：

$$p(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \cdots, n$$

为随机变量X的二项分布，也可以记作：

$$X \sim b(n, p)$$

当$n = 1$的时候，对应的就是伯努利分布，所以伯努利分布也可以记作$b(1, p)$。

2.2.5 离散的累积分布函数

设X是一个随机变量，x是任意实数，函数：

$$F(x) = P(X \leq x) = \sum_{a \leq x} p(a)$$

因为是把概率分布函数累加起来，所以称为累积分布函数（Cumulative Distribution Function，或者缩写为CDF），也简称为分布函数。

在这里，定义累积分布函数主要是为了计算上的便利，以下常见计算都可以CDF来完成：

$P(X \leq a)$	$F(a)$
$P(X > a)$	$1-F(a)$
$P(a_1 < X \leq a_2)$	$F(a_2)-F(a_1)$

作为例子，我们来看看伯努利分布的累积分布函数：

已知伯努利分布的PMF为：

$$p(x)=\begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases}$$

那么根据定义容易得到伯努利分布的CDF为：

$$F(x)=\begin{cases} 0, & x<0 \\ 1-p, & 0 \leq x<1 \\ 1, & x \geq 1 \end{cases}$$

同样的，根据定义，二项分布的CDF为：

$$F(x)=\sum_{k \leq x}\binom{n}{k} p^k(1-p)^{n-k}$$

> 例题
>
> 预定餐厅座位而不来就餐的顾客比例为20%,
>
> 如今餐厅有3个座位，但预留给了5位顾客，问到时来到餐厅而没有座位的概率是多少?
>
> 设X表示到时来到餐厅的顾客人数，有X服从二项分布，根据题意有
>
> $$X \sim b(5，0.8)$$
>
> 故所求概率为
>
> $$P(X \geq 4)=P(X=4)+P(X=5)$$
>
> $$=\binom{5}{4} \times 0.8^4 \times 0.2+0.8^5=0.73728$$

2.3 离散的数学期望

2.3.1 期望的定义

设离散随机变量X的概率质量函数为：

$$p(x_i)=P(X=x_i), i=1,2, \cdots, n, \cdots$$

如果：

$$\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty$$

则称：

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

为随机变量X的数学期望（expected value，或，expectation），简称期望或均值（mean），也有很多文档会用μ_X来表示（如果不强调随机变量的话，也可以直接用μ来表示）：

$$\mu_X = \mu = \sum_{i=1}^{\infty} x_i p(x_i)$$

若级数$\sum_{i=i}^{\infty} |x_i| p(x_i)$不收敛，则称$X$的数学期望不存在。

2.3.2 矩

不管是"不确定性的计量"还是"加权平均"，都可以通过力矩来解释，所以数学期望也称作矩，更准确点说，由于数学期望：

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

中x_i是一次项，所以又称作一阶矩。这个称呼经常在统计的书上会遇到，特在此说明。

2.3.3 期望的性质

2.3.3.1 复合

假设$g(X)$为随机变量X的某一函数，则：

$$E\left[g(X)\right] = \sum_{i} g(x_i) p(x_i)$$

2.3.3.2 常数

若c为常数，则：

$$E(c) = c$$

2.3.3.3 线性组合

数学期望满足：

- 齐次性，对于任意常数a有：

$$E(aX) = aE(X)$$

- 可加性，对于随机变量的函数$g_1(X)$、$g_2(X)$有：

$$E\left[g_1(X) + g_2(X)\right] = E\left[g_1(X)\right] + E\left[g_2(X)\right]$$

2.3.4 二项分布的期望

伯努利分布和二项分布的期望分别如下：

	伯努利分布	二项分布
PMF	$p(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$	$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$
μ	p	np

收起详细

（1）伯努利分布的期望可以如下计算：

$$\mu = 1 \times p + 0 \times (1-p) = p$$

（2）二项分布的期望可以如下计算：

$$\begin{aligned} \mu &= \sum_{x=0}^{n} xp(x) & x = 0 \text{ 时 } xp(x) = 0 \\ &= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} & \binom{n}{x} = \frac{n}{x}\binom{n-1}{x-1} \\ &= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\ &= np(p+1-p)^{n-1} = np \end{aligned}$$

从这里也可以看出，伯努利分布是二项分布的特例。

2.3.5 辛普森悖论

下面是对美国某大学的两个学院的招生情况进行的一个统计，首先是法学院：

性别	录取	拒收	总数	录取比例
男生	8	45	53	15.1%
女生	51	101	152	33.6%
合计	59	146	205	28.8%

然后是商学院：

性别	录取	拒收	总数	录取比例
男生	201	50	251	80.1%
女生	92	9	101	91.1%
合计	293	59	352	83.2%

在这两个表格中，女生的录取比例都更高，这个时候我们往往会认为总体的录取比例就是女生更高。现在将两学院的数据汇总：

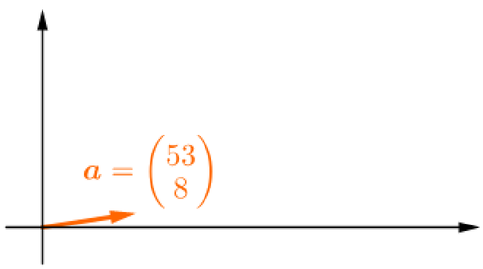
性别	录取	拒收	总数	录取比例
男生	209	95	304	68.8%
女生	143	110	253	56.5%
合计	352	205	557	63.2%

汇总之后，男生的录取比率反而更高，与我们的直觉相反。这就是辛普森悖论。

悖论的向量观点

这里再简单介绍一种理解辛普森悖论的向量观点，数学是抽象的，多一些视角总是有助于融会贯通的。

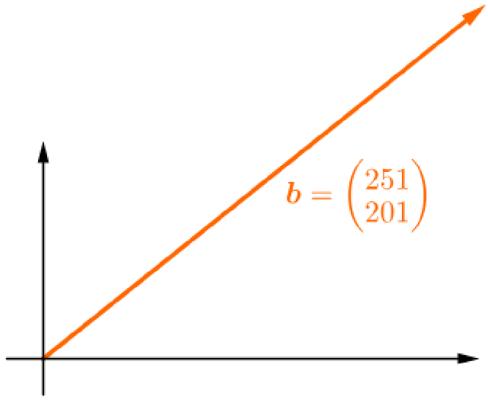
首先把法学院的男生招生情况组成一个向量：(招生总数，录取人数)。这个向量的斜率就是法学院男生的录取比例：



法学院

性别	录取	总数	录取比例
男生	8	53	15.1%
女生	51	152	33.6%
合计	59	205	28.8%

同理，商学院的男生招生情况可以组成另外一个向量，斜率就是商学院男生的录取比例：



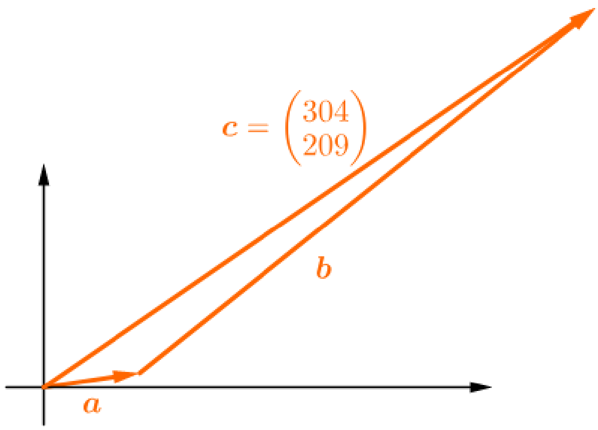
商学院

性别	录取	总数	录取比例
男生	201	251	80.1%
女生	92	101	91.1%
合计	293	352	83.2%

这两个向量相和构成一个新的向量：

$$\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b} = \begin{pmatrix} 304 \\ 209 \end{pmatrix}$$

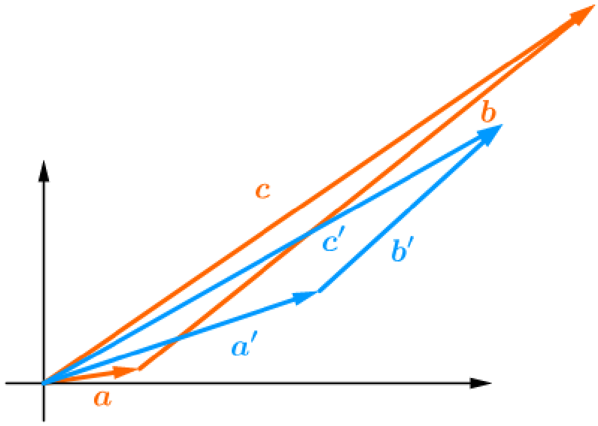
根据向量加法的三角形法则，$\boldsymbol{c}$正好与$\boldsymbol{a}$、$\boldsymbol{b}$组成一个三角形，斜率为汇总之后的男生录取率：



汇总表

性别	录取	总数	录取比例
男生	209	304	68.8%
女生	143	253	56.5%
合计	352	557	63.2%

女生的录取情况同样可以用一个三角形来表示（$\boldsymbol{a}'$表示法学院女生，$\boldsymbol{b}'$表示商学院女生，$\boldsymbol{c}'$表示汇总女生），把这两个三角形放在一起：



汇总表

性别	录取	总数	录取比例
男生	209	304	68.8%
女生	143	253	56.5%
合计	352	557	63.2%

可以看到，虽然$\boldsymbol{a}'$、$\boldsymbol{b}'$的斜率大于$\boldsymbol{a}$、$\boldsymbol{b}$，但是最终$\boldsymbol{c}$的斜率却大于$\boldsymbol{c}'$，这就是辛普森悖论的几何意义。