

# 第七章 概率基础

## 第 19 讲 概率论的基本概念

黄定江

DaSE @ ECNU

[djhuang@dase.ecnu.edu.cn](mailto:djhuang@dase.ecnu.edu.cn)

① 19.1 试验、样本空间和事件

② 19.2 概率

③ 19.3 独立事件、条件概率和贝叶斯理论

## 1 19.1 试验、样本空间和事件

## 2 19.2 概率

## 3 19.3 独立事件、条件概率和贝叶斯理论

## 引言：确定现象与不确定现象

人类在自然界的生产实践中，观察到的现象大致分为 2 类：

- 确定性现象：
  - 太阳肯定会东方升起
  - 标准气压下，水在  $100^{\circ}\text{C}$  会沸腾
- 不确定现象 (随机现象):
  - 感知的不确定性
  - 记忆的不确定性
  - 思维的不确定性

## 引言：数据科学中的不确定性

- 数据产生和分析过程中的不确定性：
  - 给定一个查询  $Q$  和文档集  $D$ ，从文档集中随机抽取一篇文档，查看是否与查询相关。
  - 观察股票  $A$  在某天中午 12 点时的股价。
  - 从一部电影的多条评论中随机挑选一条影评，观察是正类影评还是负类影评。
- 在数据科学中，数据通过采样得到的，具有一定的不确定性，它的结果是通过观测得到的，也具有一定的不确定性。数据科学中不确定性主要来自数据、模型和由模型产生的预测结果，因此使用概率模型对真实数据统计规律进行建模模拟。

## 引言：研究不确定性的学科——概率论

粗略的说，概率论是一门研究不确定现象的学科.

- 不确定现象在个别试验中呈现不确定结果，大量试验后呈现统计规律，概率论是研究不确定现象统计规律的一门学科。
- 概率可以被认为是一个事件发生的次数的分数，或者是对一个事件的信任程度。
- 我们用这个概率来衡量试验中发生某些事情的可能性。

下面我们从试验和样本空间入手，介绍概率的基本概念。

### 19.1.1 试验、样本空间和事件：试验

#### 随机试验

概率论是通过试验研究随机现象的统计规律，试验要满足 3 个条件：

- 试验在相同条件下可以重复进行。
- 试验的所有可能结果在试验前已知，结果不止一个。
- 试验前不知道哪种结果会出现。

具有上述特点的试验称为随机试验，简称试验。

## 例 1

给定一个查询  $Q$  和文档集  $D = \{d_1, d_2, d_3, \dots, d_n\}$ , 从  $D$  随机抽取一篇文档  $d_i$ , 查看  $d_i$  是否与查询相关。该试验满足上述三个条件:

- 将  $d_i$  放入文档集  $D$  中, 再次重复抽取。
- 已经知道试验的所有可能结果. 查询  $Q$  与抽取的文档  $d_i$  相关或者不相关。
- 随机抽取文档前, 我们并不知道抽取的文档是否与查询  $Q$  相关。



## 例 2

观察  $A$  股票在中午 12 点时的股价。这种试验也满足上述 3 个条件：多次在中午观察  $A$  股票价格；股票的所有可能价格属于非负实数集；在观测之前，并不知道本次试验结果。

## 例 3

从一部电影的众多影评中随机抽取一条影评，观察该影评是正类影评还是负类影评。这种试验显然也满足上述 3 个条件。

综上所述，可重复、结果多样、结果不可预测是随机试验的三个特点。

# 样本空间

## 样本空间

**样本空间**是随机试验所有可能结果的集合，记作  $\Omega$ 。每一种试验结果称为样本空间中的一个样本点。

- 例1中，给定一个查询  $Q$  和文档集  $D$ ，从文档集中的随机抽取文档  $d_j$  的试验的样本空间  $\Omega = \{\text{相关}, \text{不相关}\}$ ，即  $d_j$  要么与查询相关，要么不相关。
- 例2中中午 12 点观察  $A$  股票价格试验的样本空间  $\Omega = [0, \infty)$ 。某一天通过观测得知  $A$  股票单价是 100 元，则 100 是样本空间中的  $[0, \infty)$  一个样本点。
- 例3中随机从一堆影评中抽取一条影评，该试验的样本空间  $\Omega = \{\text{正类}, \text{负类}\}$ 。若抽得正类影评，则正类是样本空间  $\{\text{正类}, \text{负类}\}$  中的一个样本点。

# 事件

## 随机事件

满足某些条件的样本点组成样本空间的子集称为随机事件，简称事件. 例1中从文档集  $D$  中抽取与查询  $Q$  相关的文档是一随机事件. 例2中股票的价格大于 100 是一个随机事件. 例3中抽取的影评是正类影评是一个随机事件. 需要注意的是：

- 一个样本点也属于一个事件.
- 空集  $\emptyset$  是样本空间  $\Omega$  的子集，称为不可能事件.
- $\Omega$  是它自己的子集，称为必然事件.

### 19.1.2 事件的关系与运算：事件关系

事件是样本点的集合，事件之间的关系与运算可以按照集合之间的关系与集合运算来规定. 给定一个随机试验， $\Omega$  是试验的样本空间，事件  $A, B, C$  是  $\Omega$  的子集. 下列给出事件之间的 7 种关系.

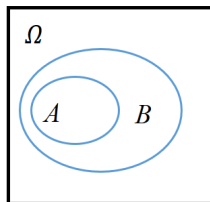


图 1: 包含关系

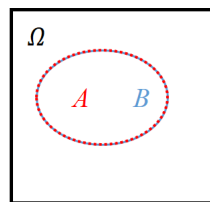


图 2: 相等关系

- **包含关系** 如果  $A \subset B$  或  $B \supset A$ ，称事件  $B$  包含事件  $A$ . 它的含义是：若事件  $A$  发生，则事件  $B$  必然发生.
- **相等关系** 如果  $A \subset B$  且  $A \supset B$ ，称事件  $B$  与事件  $A$  相等.

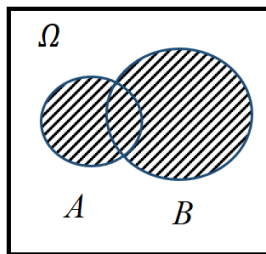


图 3: 事件和

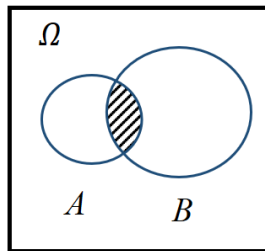


图 4: 事件积

- **事件和**  $A \cup B = \{\omega : \omega \in A \text{ 或 } \omega \in B\}$  称事件  $A$  与事件  $B$  的和事件. 它的含义是: 当且仅当事件  $A$  与事件  $B$  中至少一个发生时, 事件  $A \cup B$  发生.
- **事件积** 事件  $A \cap B = \{\omega : \omega \in A \text{ 且 } \omega \in B\}$  称事件  $A$  与事件  $B$  的积事件. 它的含义是: 当且仅当事件  $A$  与事件  $B$  中同时发生时, 事件  $A \cap B$  发生. 有时候我们将  $A \cap B$  记为  $AB$  或者  $(A, B)$ .

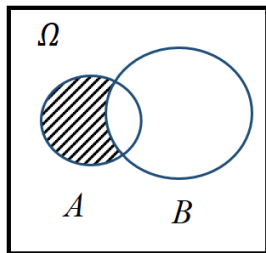


图 5: 事件差

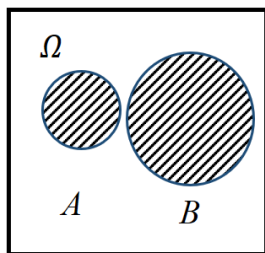


图 6: 互斥关系

- **事件差** 事件  $A - B = \{\omega : \omega \in A \text{ 且 } \omega \notin B\}$  称事件  $A$  与事件  $B$  的差事件. 它的含义是: 当且仅当事件  $A$  发生且事件  $B$  不发生时, 事件  $A - B$  发生.
- **互斥关系** 如果事件  $A \cap B = \emptyset$ , 称事件  $A$  与事件  $B$  互斥或不相容. 它的含义是: 在一次试验后, 事件  $A$  与事件  $B$  不会同时发生. 如果一组事件中任意两个事件互不相容, 这组事件两两不相容.

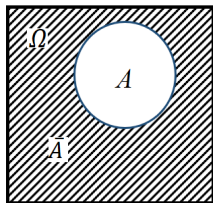


图 7: 逆事件

- **逆事件** 事件  $\Omega - A$  称为事件  $A$  的逆事件, 记作  $\bar{A} = \Omega - A$ . 它的含义是: 当且仅当事件  $A$  不发生时, 事件  $\bar{A}$  发生. 于是  $\bar{A} \cap A = \emptyset$ ,  $\bar{A} + A = \Omega$ . 由于  $A$  也是  $\bar{A}$  的对立事件, 因此称事件  $A$  与  $\bar{A}$  互逆.
- 有时也用  $A^c$  来表示  $A$  的逆事件。

## 事件运算

### 事件运算

与集合论中集合的运算一样，事件之间的运算满足下述定律：

- 交换律

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

- 结合律

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

- 分配律

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$



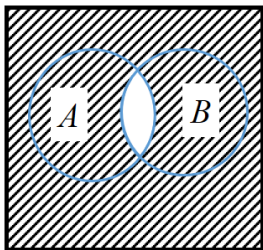


图 8: 法则 1

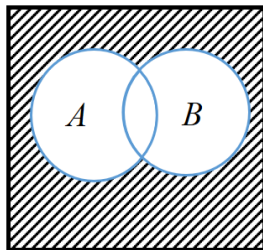


图 9: 法则 2

- 德·摩根法则

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

以上这些定律都可以扩展到任意多个事件.

## 定义 1

设  $E$  是随机试验,  $\Omega$  是它的样本空间。对于集合序列  $A_1, A_2, \dots, A_n$ , 若:

- $A_i \cap A_j = \emptyset, i \neq j, i, j = 1, 2, \dots, n$
- $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$

则称  $A_1, A_2, \dots, A_n$  为样本空间  $\Omega$  的一个划分。若  $A_1, A_2, \dots, A_n$  为样本空间  $\Omega$  的一个划分, 每次试验  $A_1, A_2, \dots, A_n$  中必有一个也仅有一个发生。

## 定义 2

给定事件  $A$ ，定义  $A$  的示性函数为

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

## 定义 3

如果集合序列  $A_1, A_2, \dots$  满足  $A_1 \subset A_2 \subset \dots$ ，则称该集合序列为单调递增序列，单调递增序列的极限定义为  $\lim_{n \rightarrow \infty} A_n = \bigcup_{i=1}^{\infty} A_i$

## 定义 4

如果集合序列  $A_1, A_2, \dots$  满足  $A_1 \supset A_2 \supset \dots$ ，则称该集合序列为单调递减序列，单调递减序列的极限定义为  $\lim_{n \rightarrow \infty} A_n = \bigcap_{i=1}^{\infty} A_i$

① 19.1 试验、样本空间和事件

② 19.2 概率

③ 19.3 独立事件、条件概率和贝叶斯理论

## 19.2 概率定义和公理

有了事件的定义后, 就可以在事件的基础上定义概率.

### 定义 5

设  $E$  是随机试验,  $\Omega$  是它的样本空间. 对于  $E$  的每一事件  $A$  赋予一个实数, 记为  $P(A)$ , 称为事件  $A$  的概率. 集合函数  $P(\cdot)$  称为概率分布或概率测度, 如果满足下列公理:

- 公理 1: 非负性 对每一个事件  $A$ ,  $P(A) \geq 0$ .
- 公理 2: 正则性 对必然事件  $\Omega$ ,  $P(\Omega) = 1$ .
- 公理 3: 可列可加性 设  $A_1, A_2, \dots$  是两两互补相容的事件. 即对于  $A_i \cap A_j = \emptyset, i \neq j, i, j = 1, 2, \dots$ , 有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

### $P$ 的两种常见解释

关于  $P$  有很多解释，最常见的有两种，频率和可信度：

- 频率  $P(A)$  就表示重复试验中事件  $A$  出现次数的最终比例。
- 可信度  $P(A)$  度量观测者对于  $A$  为真的信度。

无论哪一种解释，公理 1-3 都必须满足。两种不同解释在统计推断中会有很大的不同，派生了两个不同的学派：频率学派和贝叶斯学派。

## 推论

根据以上 3 个公理，可以推出如下 4 个推论：

- 推论 1 不可能事件概率为 0

$$P(\emptyset) = 0 \quad P(A \cap \bar{A}) = 0$$

- 推论 2 对任意两个事件  $A$  和  $B$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- 推论 3 事件  $A$  的补事件

$$P(\bar{A}) = 1 - P(A)$$

- 推论 4 对任意事件  $A$

$$0 \leq P(A) \leq 1$$

## 定理 1

(概率的连续性) 如果  $A_n \rightarrow A$ , 则当  $n \rightarrow \infty$  时,  $P(A_n) \rightarrow P(A)$ .

证明.

假定  $A_n$  是单调递增序列, 则  $A_1 \subset A_2 \subset \dots$ . 令  $A = \lim_{n \rightarrow \infty} A_n = \bigcup_{i=1}^{\infty} A_i$ . 定义  $B_1 = A_1, B_2 = \{\omega \in \Omega : \omega \in A_2, \omega \notin A_1\}, B_3 = \{\omega \in \Omega : \omega \in A_3, \omega \notin A_2, \omega \notin A_1\}, \dots$ . 容易证明  $B_1, B_2, \dots$  两两不相交, 对一切  $n$  有  $A_n = \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$  且  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ . 由公理 3 可得

$$P(A_n) = P\left(\bigcup_{i=1}^n B_i\right) = \sum_{i=1}^n P(B_i).$$

因此, 再利用公理 3 就可以得到

$$\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) = \sum_{i=1}^{\infty} P(B_i) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = P(A)$$





## 19.2 有限样本空间上的概率

假定样本空间  $\Omega = \{\omega_1, \dots, \omega_n\}$  有限。例如，连续将一颗骰子抛两次， $\Omega$  就有 36 个元素： $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ ，如果每一个结果都是等可能的，则有  $P(A) = |A|/36$  其中  $|A|$  表示集合  $A$  中的元素个数。因为只有两种情况可能满足骰子点数之和为 11，所以骰子点数之和为 11 的概率就是  $2/36$ 。

如果  $\Omega$  是有限的并且每种结果都是等可能的，那么

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

上式称为**均匀概率分布**。为求得概率，需要计算事件  $A$  包含的样本点，计算样本点个数的方法称为**组合法**。

① 19.1 试验、样本空间和事件

② 19.2 概率

③ 19.3 独立事件、条件概率和贝叶斯理论

### 19.3.1 独立事件

如果连续两次抛一枚均匀的硬币，则两次都出现正面的概率是  $\frac{1}{2} \times \frac{1}{2}$ ，之所以能将两者相乘是因为我们认为这两次抛硬币是独立的。

#### 定义 6

如果下式成立，则事件  $A$  和  $B$  是独立的：

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

记为  $A \perp B$ 。如果等式

$$P\left(\bigcap_{i \in \mathbb{J}} A_i\right) = \prod_{i \in \mathbb{J}} P(A_i)$$

对所有  $\mathbb{I}$  的子集  $\mathbb{J}$  都成立，则事件集  $\{A_i : i \in \mathbb{I}\}$  是独立的。

独立性可能以两种截然不同的方式出现。有时，直接假设两个事件是独立的。

#### 例 4

在连续两次抛一枚硬币的试验中，通常假设每次抛硬币是互相独立的，这也反映了硬币对抛第一次没有记忆性的事实。

而在另外一些时候，需要通过证明  $P(AB) = P(A)P(B)$  来推导两事件的独立性。

#### 例 5

抛一颗均匀骰子的试验中，令  $A = \{2, 4, 6\}$ ,  $B = \{1, 2, 3, 4\}$ ，则

$A \cap B = \{2, 4\}$ ,  $P(AB) = 2/6 = P(A)P(B) = (1/2) \times (2/3)$ ，所以说  $A$  与  $B$  是独立的。

在本例中，并没有假设  $A$  与  $B$  是独立的，而是证明它们是独立的。

假定  $A$  与  $B$  是互斥事件，并且每一个事件都有正的概率，它们有可能独立吗？答案是否定的，因为  $P(A)P(B) > 0$  而  $P(AB) = P(\emptyset) = 0$ 。除这种情况外，没有别的办法来判断维恩图中集合的独立性。

### 例 6

抛一枚均匀的硬币 10 次。令  $A =$  “至少出现一次正面”，令  $T_j$  表示反面出现在第  $j$  次的事件，从而

$$\begin{aligned} P(A) &= 1 - P(\bar{A}) \\ &= 1 - P(\text{全是反面}) \\ &= 1 - P(T_1 T_2 \dots T_{10}) \\ &= 1 - P(T_1)P(T_2) \dots P(T_{10}) \\ &= 1 - \left(\frac{1}{2}\right)^{10} \approx 0.9999 \end{aligned}$$

## 例 7

两个人轮流投篮，第 1 个人投进的概率为  $1/3$ ，第 2 个人投进的概率为  $1/4$ 。第 1 个人比第 2 个人先投进的概率是多少？令  $E$  表示所关心的事件，令  $A_j$  表示在第  $j$  轮由第 1 个人首次投进这一事件。注意到  $A_1, A_2, \dots$  是两两独立的，并且  $E = \bigcup_{j=1}^{\infty} A_j$ ，因此

$$P(E) = \sum_{j=1}^{\infty} P(A_j).$$

现在有  $P(A_1) = 1/3$ 。 $A_2$  表示第 1 轮两人都没投进，第 2 轮由第 1 个人首次投进，其概率为  $P(A_2) = (2/3)(3/4)(1/3) = (1/2)(1/3)$ 。以此类推可求得  $P(A_j) = (1/2)^{j-1}(1/3)$ ，从而

$$P(E) = \sum_{j=1}^{\infty} \frac{1}{3} \left(\frac{1}{2}\right)^{j-1} = \frac{2}{3}$$

这里用到公式，如果  $0 < r < 1$ ，那么  $\sum_{j=k}^{\infty} r^j = r^k / (1 - r)$ 。

### 独立性小结

- $A$  和  $B$  是独立的当且仅当  $P(AB) = P(A)P(B)$
- 独立有时用于假设而有时需要推导。
- 正概率的互斥事件不可能是独立的。

## 19.3.2 条件概率

### 定义 7

假设  $P(B) > 0$ , 定义在  $B$  发生的情况下  $A$  的条件概率为

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

$P(A|B)$  可认为是  $A, B$  同时发生的次数占  $B$  发生次数的比例。对任意固定  $B$  只要  $P(B) > 0$ ,  $P(\cdot|B)$  就是一个概率测度 (即它满足概率的 3 条公理), 也即  $P(A|B) \geq 0$ ,  $P(\Omega|B) = 1$ 。如果  $A_1, A_2, \dots$  互斥, 则  $P(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$ 。但是:

- 一般  $P(A|B \cup C) = P(A|B) + P(A|C)$  是不成立的。有关概率的法则只适用于竖杠左边的事件。
- 一般  $P(A|B) = P(B|A)$  也是不成立的。

### 例 8

得麻疹时身上有斑点的概率是 1, 但身上有斑点时得麻疹的概率并不是 1。

在这个例子里,  $P(A|B)$  和  $P(B|A)$  的差异是很显然的, 但是在有些情况下, 却未必能这么显而易见。这一错误在法律案件中经常发生, 有时将其称为检察官谬论。



## 例 9

疾病  $D$  的医学检查结果可能为  $+$  和  $-$ ，它们的概率如下

	$D$	$\bar{D}$
$+$	0.009	0.099
$-$	0.001	0.891

由条件概率的定义可得

$$P(+|D) = \frac{P(+ \cap D)}{P(D)} = \frac{0.009}{0.009 + 0.001} = 0.9$$

$$P(-|D) = \frac{P(- \cap \bar{D})}{P(\bar{D})} = \frac{0.891}{0.891 + 0.099} = 0.9$$

显然，该检验是相当精确的，对患者的检验结果有 90% 呈阳性，而对于健康者的检验结果有 90% 呈阴性，假定去作检查的结果是阳性，患这种病的概率会是多大呢？很多人认为是 0.90，而正确的结果是

$$P(D|+) = \frac{P(+ \cap D)}{P(+)} = \frac{0.009}{0.009 + 0.099} \approx 0.08$$

## 定理 2

如果  $A$  与  $B$  是相互独立的事件则  $P(A|B) = P(A)$ 。对任意两事件  $A, B$  有

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

根据该定理，可以知道独立性的另一个解释为在知道  $B$  的情况下不会改变  $A$  的概率，即

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

公式  $P(AB) = P(A)P(B|A)$  有些时候对计算概率很有帮助。

## 例 10

从一副扑克中不重复抽两张牌，令  $A$  表示第一次抽取的牌是梅花  $A$ ，令  $B$  表示第二次抽取的牌是红桃  $K$ ，则

$$P(AB) = P(A)P(B|A) = (1/52)(1/51)$$

## 条件概率小结

- 如果  $P(B) > 0$  则

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

- 对固定的  $B$ ,  $P(\cdot|B)$  满足概率公理, 但一般地, 对固定的  $A, P(A|\cdot)$  不满足概率公理。
- 一般地,  $P(A|B) \neq P(B|A)$
- $A$  和  $B$  独立当且仅当  $P(A|B) = P(A)$

### 19.3.3 贝叶斯理论

贝叶斯理论是机器学习算法的基础，在信息检索、邮件过滤、文本分类等诸多方面有广泛的应用。

#### 定理 3

(全概率法则) 令  $A_1, A_2, \dots, A_k$  是  $\Omega$  的一个划分，则对任意事件  $B$  有

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

证明.

定义  $C_j = BA_j$  并注意到  $C_1, \dots, C_k$  是互斥的， $B = \bigcup_{j=1}^k C_j$ ，由条件概率定义知  $P(BA_j) = P(B|A_j)P(A_j)$ ，因此

$$P(B) = \sum_j P(C_j) = \sum_j P(BA_j) = \sum_j P(B|A_j)P(A_j)$$



# 贝叶斯定理

## 定理 4

(贝叶斯定理) 令  $A_1, \dots, A_k$  是  $\Omega$  的一个划分, 对每一个  $i$  有  $P(A_i) > 0$ , 如果  $P(B) > 0$ , 则对  $i = 1, \dots, k$  有

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

注通常称  $P(A_i)$  为  $A$  的先验概率, 称  $P(A_i|B)$  为  $A$  的后验概率。

证明.

由条件概率的定义以及全概率法则可得

$$P(A_i|B) = \frac{P(A_i B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$



## 例 11

我将自己的邮件分为三类： $A_1 =$ “垃圾邮件”， $A_2 =$ “低优先级邮件”， $A_3 =$ “高优先级邮件”，由以前的经验发现  $P(A_1) = 0.7, P(A_2) = 0.2, P(A_3) = 0.1$ 。令  $B$  表示邮件中含有单词“free”这一事件，由以前的经验有  $P(B|A_1) = 0.9, P(B|A_2) = 0.01, P(B|A_3) = 0.01$ 。我收到一封邮件其中含有单词“free”，这封邮件是垃圾邮件的概率为多少？

## 解

由贝叶斯理论可求得

$$P(A_1|B) = \frac{0.9 \times 0.7}{(0.9 \times 0.7) + (0.01 \times 0.2) + (0.01 \times 0.1)} = 0.995$$

## 本讲小结

### 基本概念

- 随机试验
- 样本空间
- 随机事件
- 概率和概率公理

### 基本概念

- 独立性
- 条件概率
- 全概率法则
- 贝叶斯定理：先验概率和后验概率

概率的两种解释产生统计推断的两个学派：频率学派和贝叶斯学派！