# 计算机问题求解 – 论题2-7 - 离散概率基础

2016年04月7日

# 预习检查

To compute probabilities, we assign a **probability weight** $P(x)$ to each element of the sample space so that the weight represents what we believe to be the relative likelihood of that outcome. There are two rules in assigning weights. First, the weights must be nonnegative numbers, and second, the sum of the weights of all the elements in a sample space must be 1. We define the **probability** $P(E)$ of the event $E$ to be the sum of the weights of the elements of $E$. Algebraically, we write

$$P(E) = \sum_{x:x\in E} P(x). \tag{5.1}$$

## 问题1:

你能否给我们讲讲书上那个关于邮购商店的“故事”，那里什么地方体现了“碰运气”？

If we have a table with 100 buckets and 50 keys to put in those buckets, it is possible that all 50 of those keys could be assigned (hashed) to the same bucket in the table. However, someone who is experienced with using hash functions will tell you that you'd never see this in a million years. But that same person might also tell you that neither would you ever see, in a million years, all the keys hash into different locations. In fact, it is far less likely that all 50 keys would hash into one place than that all 50 keys would hash into different places, but both events are quite unlikely. Being able to understand just how likely or unlikely such events are is a major reason for taking up the study of probability.

## 问题2：

## 这里的you'd never…和neither would you ever…有什么意义？

# 离散概率模型

## 问题3：

## 你能否以下面的过程为例解释离散概率模型中的主要概念？

Sample space
Element
Event
Probability

**A process**: 掷两个色子

# Axioms for a probability space

满足下列性质的*P*称为一个probability distribution 或者一个probability measure。

1. $P(A) \geq 0$ for any $A \subseteq S$.
2. $P(S) = 1$.
3. $P(A \cup B) = P(A) + P(B)$ for any two disjoint events $A$ and $B$.

记住：*P*是一个函数。

**问题4：**

**为什么任何事件的概率值不会大于1？**

**如果A和B相交，P(AUB)是什么？**

# 问题5：

# 有限样本空间中的离散概率与计数有什么关系？

$$P(E) = \sum_{x:x\in E} P(x).$$

# 有限样本空间

- There are only finite outcomes.
- Each outcome individually consists an elementary event.
  - For one coin toss, there are two outcomes – head and tail. "Head" is an elementary event.
- The probability of an elementary event  corresponds a specific outcome.
- If all outcomes are equally likely, then the probability of an event *E* can be computed as:

$$p(E) = \frac{|E|}{|A|} = \frac{\text{total number of outcomes in } E}{\text{total number of outcomes}}$$

问题6：

你能举一个不满足equally likely分布的例子吗？

这种情况下，概率公式应该如何确定？

# 交集非空的事件

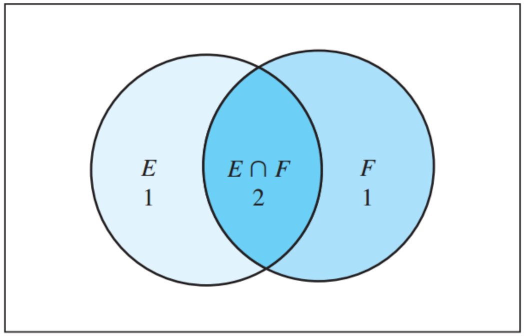
- 掷均匀的色子，掷3次。出现事件“或者3次均相等，或者没有一次是4”的概率是多少？
- 合理假设：每个outcome出现的可能性是一样的。
- 样本空间大小是 $6^3=216$。
- 用$F$表示事件“3次结果一样”， … )
- 用$G$表示事件“没有一次结果是4”，则$|G|=5^3=125$ (G是从集合{1,2,3,5,6}中任选3个数（可重复）的方案数)
- 要求的事件为$F$和$G$的并集:

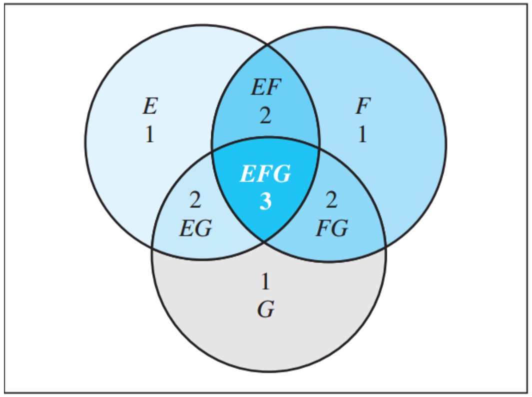
$$|F\cup G|=|F|+|G|-|F\cap G|=6+125-5=126$$

- 因此，最终结果是：126/216 = 7/12

This is a special case of so-called **inclusion-exclusion principle**

# 容斥原理




$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

# 容斥定律: 否定形式

## 包含在若干个子集的并集中的元素个数:

$$N(A_1 \cup A_2 \cup \ldots \cup A_n) = S_1 - S_2 + S_3 - \ldots + (-1)^{k+1} S_k + \ldots + (-1)^{n+1} S_n$$

where, $S_k = \sum_{1 \le i_1 \le i_2 \le \ldots \le i_k \le n} | A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k} | \quad k = 1,2,\ldots,n$

For an example：the formula for 4 subsets

$N - (|A_1| + |A_2| + |A_3| + |A_4|)$

$+ (|A_1 \cap A_2| + |A_1 \cap A_2| + |A_1 \cap A_4| + |A_2 \cap A_3| + |A_2 \cap A_4| + |A_3 \cap A_4|)$

$- (|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4|)$

$+ |A_1 \cap A_2 \cap A_3 \cap A_4|$

# 容斥定律: 否定形式

没有被包含在若干个子集的并集中的元素个数:

$$N(\overline{A_1}\,\overline{A_2}\ldots\overline{A_n}) = N - S_1 + S_2 - S_3 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

$$\text{where,}\quad S_k = \sum_{1 \le i_1 \le i_2 \le \ldots \le i_k \le n} |A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}| \qquad k = 1,2,\ldots,n$$

For an example：the formula for 4 subsets

$N - (|A_1| + |A_2| + |A_3| + |A_4|)$

$+ (|A_1 \cap A_2| + |A_1 \cap A_2| + |A_1 \cap A_4| + |A_2 \cap A_3| + |A_2 \cap A_4| + |A_3 \cap A_4|)$

$- (|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| + |A_2 \cap A_3 \cap A_4|)$

$+ |A_1 \cap A_2 \cap A_3 \cap A_4|$

# 容斥定律

**(Principle of Inclusion and Exclusion for Probability)** The probability of the union $E_1 \cup E_2 \cup \cdots \cup E_n$ of events in a sample space $S$ is given by

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{i_1, i_2, \ldots, i_k: \\ 1 \le i_1 < i_2 < \cdots < i_k \le n}} P\left(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}\right). \tag{5.6}$$

$$P\left(\prod_{i=1}^{n} \overline{E}_i\right) = P(S) - P\left(\bigcup_{i=1}^{n} E_i\right)$$

# Hatcheck Problem

- 大剧院衣帽间的员工太粗心，将$n$个客人的帽子上的标签搞乱了。他将$n$顶帽子随意地递交给每个客人。
  - 问题：“每个客人都拿错了帽子”的概率是多少？
- 数学模型：随机地排列自然数 1,2,3,…,$n$，生成一个序列：$i_1, i_2, i_3,\ldots,i_n$。出现下述情况的概率是多少： 对任意的 $k(1\le k\le n)$, $i_k\ne k$？
- 这样的序列称为 ***derangement***.

# Number of Derangement

- Define $i_k=k$ as Property $A_k$ , and $A_k$ is used for the subset of all permutations satisfying property $A_k$.

The number of derangement is :

$$N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = N - S_1 + S_2 - S_3 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

where $N = n!$

where $S_k$ is $\sum_{1 \le i_1 \le i_2 \ldots \le i_k \le n} | A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k} |$

Note : $S_k$ is the number of permutations (possible with multi - count) keeping exactly k elements in their original positions, and the other $n - k$ elements as any possible permutation. So :

$$S_1 = \binom{n}{1}(n-1)!\,;\, S_2 = \binom{n}{2}(n-2)!\,;\ldots,\, S_k = \binom{n}{k}(n-k)! = \frac{n!}{k!}$$

# The Probability of Derangement

We have known that the number of derangement is

$$N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = N - S_1 + S_2 - S_3 + \ldots + (-1)^k S_k + \ldots + (-1)^n S_n$$

where $N = n!$, and $S_k = \binom{n}{k}(n-k)! = \dfrac{n!}{k!}\ (k = 1,2,3,\ldots,n)$

$$\therefore N(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\ldots\overline{A_n}) = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!};$$ and the probability: $\displaystyle\sum_{k=0}^{n}\frac{(-1)^k}{k!}$

Since $\displaystyle\sum_{k=0}^{\infty}\frac{(-1)^k}{k!} = e^{-1}$, the difference between the probability

and $e^{-1} \approx 0.367879\ldots$ is less than $\dfrac{1}{n!}$, which means that the

probability is about 0.368, independent of $n$, except for very small $n$.

# 条件概率

The **conditional probability** of $E$ given $F$, denoted by $P(E|F)$ and read as "the probability of $E$ given $F$," is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}. \tag{5.13}$$

书上的例子：掷两个特殊的色子：出现三角形、圆、正方形的面数分别是1，2，3。事件E: 至少一个有圆的面朝上；事件F: 朝上的两个面图案相同。如果知道事件F已经发生，那么事件E的概率是多少？

- 按照常识来分析：可以理解为样本空间改变了。
- 利用上面的定义式来计算。

| TT | TC | TS | CT | CC | CS | ST | SC | SS |
|---|---|---|---|---|---|---|---|---|
| $\frac{1}{36}$ | $\frac{1}{18}$ | $\frac{1}{12}$ | $\frac{1}{18}$ | $\frac{1}{9}$ | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{4}$ |

原来的样本空间大小是36，如果只考虑两面相同，样本空间大小缩小为14，其中出现圆图案的是4。
关键假设：三种图案间的比例不变!

问题7：

条件概率那个式子是定义还是定理？

# 相互独立的事件

$E$ is **independent** of $F$ if $P(E|F) = P(E)$

**(Product Principle for Independent Probabilities)** Suppose $E$ and $F$ are events in a sample space. Then $E$ is independent of $F$ if and only if $P(E \cap F) = P(E)P(F)$.

## 问题8：

如何理解P(E|F)和P(E∩F)？

事件之间的相互“independent”与“disjoint”有什么不同？

# 独立试验过程

一个有5个步骤的独立试验，其样本空间有多大？

flip a coin $n$ times, $S_i = \{\text{H}, \text{T}\}$ for each $i$, $1 \leq i \leq n$.) A process that occurs in stages is called an **independent trials process** if

$$P(x_i = a_i | x_1 = a_1, \ldots, x_{i-1} = a_{i-1}) = P(x_i = a_i) \qquad (5.17)$$

In an independent trials process, the probability of a sequence $a_1, a_2, \ldots, a_n$ of outcomes is $P(\{a_1\})P(\{a_2\}) \cdots P(\{a_n\})$.

# 例：

- 掷了100次硬币，有人告诉你，他看到了99个正面，另外一个是正还是反？
- 掷了99次硬币，都是正面，你赌下一次是正还是反？
- 掷100次硬币，你敢赌都是正面吗？

这三个例子都是反映掷骰子是独立的！

In an independent trials process, the probability of a sequence $a_1, a_2, \ldots, a_n$ of outcomes is $P(\{a_1\})P(\{a_2\})\cdots P(\{a_n\})$.

# 又例：Hashing相关的“事件”

What sample space and probabilities have we been using when discussing hashing? Using these, show that the event “key $i$ hashes to position $r$” and the event “key $j$ hashes to position $q$” are independent when $i \neq j$. Are they independent if $i = j$?

如果将$n$个keys “哈希”到大小为$k$的表中，则样本空间包含长度为$n$，元素为$\{1, 2, \cdots, k\}$中任意元素的序列。

The event that key $i$ hashes to some number $r$ consists of all $n$-tuples with $r$ in the $i$th position, so its probability is $k^{n-1}/k^n = 1/k$. The probability that key $j$ hashes to some number $q$ is also $1/k$. If $i \neq j$, then the event that key $i$ hashes to $r$ and key $j$ hashes to $q$ has probability $k^{n-2}/k^n = 1/k^2$, which is the product of the probabilities that key $i$ hashes to $r$ and key $j$ hashes to $q$. Therefore, these two events are independent. If $i = j$, the probability of key $i$ hashing to $r$ and key $j$ hashing to $q$ is 0, unless $r = q$, in which case it is 1. Thus, if $i = j$, these events are not independent.

# 条件概率应用与概率分析 – 考试成绩

If a student knows 80% of the material in a course, what do you expect her grade to be on a (well-balanced) 100-question short-answer test about the course? What is the probability that she answers a question correctly on a 100-question true-false test if she guesses at each question for which she does not know the answer? (We assume she knows what she knows—that is, if she thinks she knows the answer, then she really does.) What do you expect her grade to be on a 100-question true-false test?

**问题8:**

**你能否从“直观”上判断该期望多少分?**

$$
\begin{aligned}
P(R) &= P(R \cap K) + P(R \cap \overline{K}) \\
&= P(R|K)P(K) + P(R|\overline{K})P(\overline{K}) \\
&= 1 * 0.8 + 0.5 * 0.2 = 0.9
\end{aligned}
$$

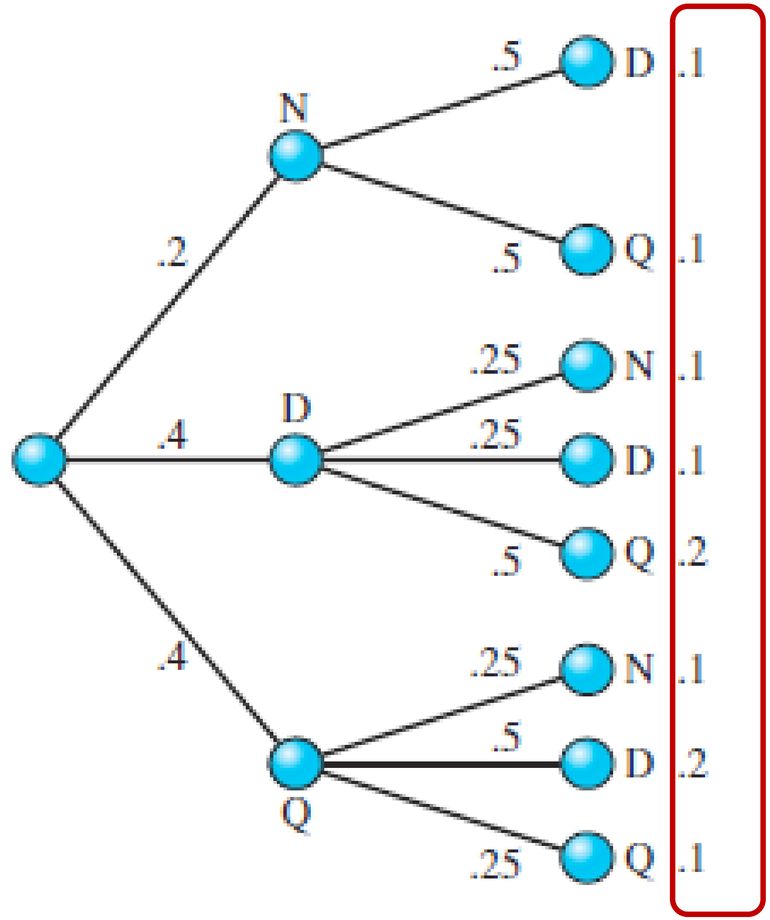
附注：R：回答正确；K：知道正确答案

# Tree Diagrams: 结合计数与概率

**过程:**

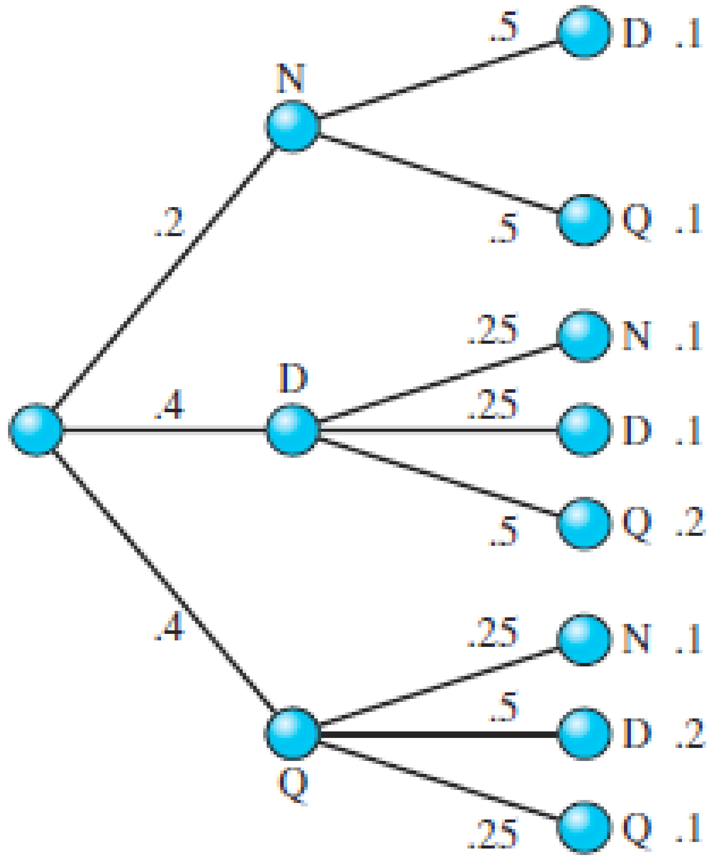
从下列硬币中依次取两枚:

nickel: 1
dime: 2
quarter: 2




为什么这个值等于该路径中前面结点上的值的乘积?

# 讨论：




为什么这个值等于该路径中前面结点上的值的乘积？

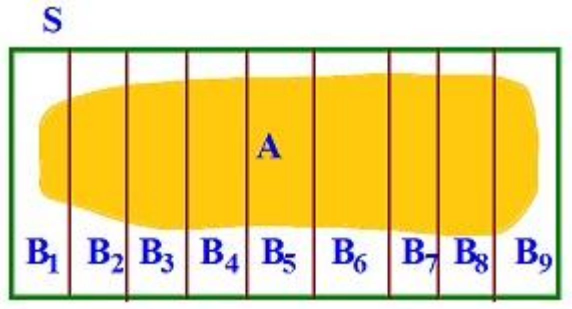
A tree for an independent trials process has the property that at each level, for each node at that level, the labeled tree consisting of that node and all its children is identical to each labeled tree consisting of another node at that level and all its children. If we have such a tree, then it automatically satisfies the definition of an independent trials process.

# law (or formula) of total probability

Suppose A is any event defined in the sample space S. If the sample is partitioned by the disjoint events. If $\{B_n: n = 1,2,3 \dots\}$ is a finite or countably infinite partition of a sample space S.

$$\Pr(A) = \sum_n \Pr(A \cap B_n)$$

$$\Pr(A) = \sum_n \Pr(A \mid B_n) \Pr(B_n),$$




$$P(A) = P(A \cap B) + P(A \cap B')$$

$$P(A) = P(A|B)P(B) + P(A|B')P(B')$$

注：本定义并不完整

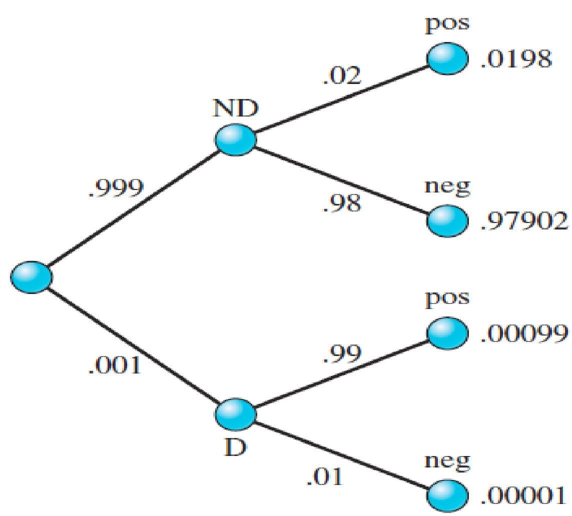
# 问题9：

# Bayes定理是什么内容？它为什么成立？它有什么意义？

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

# 不那么“直观”的概率分析

A test for a disease that affects 0.1% of the population is 99% effective on people with the disease (that is, the test says they have the disease with probability .99). The test gives a false reading (saying that a person who does not have the disease is affected with it) for 2% of the population without the disease. We can think of choosing someone and testing them for the disease as a two-stage process. In Stage 1, we either choose someone with the disease or we don't. In Stage 2, the test is either positive or it isn't. Give a tree diagram for this process. What is the probability that someone selected at random and given a test for the disease tests positive? What is the probability that someone who tests positive in fact has the disease?

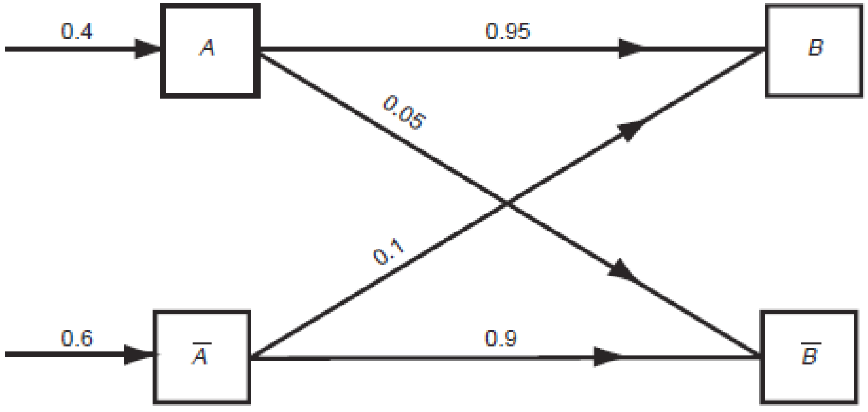



$$P(\mathrm{D}|\mathrm{pos}) = \frac{P(\mathrm{D} \cap \mathrm{pos})}{P(\mathrm{pos})}$$

$$P(\mathrm{D} \cap \mathrm{pos})/P(\mathrm{pos}) = .00099/.02097 = \underline{.0472}$$

# 网络通信的例子

Problem: a simple binary communication channel carries messages by using only two signals, say 0 and 1. We assume that, for a given binary channel, 40% of the time a 1 is transmitted; the probability that a transmitted 0 is correctly received is 0.90, and the probability that a transmitted 1 is correctly received is 0.95. Determine (a) the probability of a 1 being received, and (b) given a 1 is received, the probability that 1 was transmitted.




$P(A) = 0.4, \quad P(\bar{A}) = 0.6;$

$P(B|A) = 0.95, \quad P(\bar{B}|A) = 0.05;$

$P(\bar{B}|\bar{A}) = 0.90, \quad P(B|\bar{A}) = 0.10.$

**b** $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)} = \dfrac{0.95(0.4)}{0.44} = 0.863.$

**a** $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.95(0.4) + 0.1(0.6) = 0.44$

# 问题10：

## 举例说明什么是随机变量？解释为什么它本质上是函数？

**A process**: 掷两个色子

# 贝努利试验 – 成败之间

The probability of having exactly $k$ successes in a sequence of $n$ independent trials with two outcomes and probability $p$ of success on each trial is given by

$$P(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

又是二项系数！

A student takes a 10-question objective test.[6] Suppose that a student who knows 80% of the course material has probability .8 of success on any question, independent of how he did on any other problem. What is the probability that he earns a grade of 80 or better (out of 100)?

$$P(80 \text{ or better}) = \binom{10}{8}(.8)^8(.2)^2 + \binom{10}{9}(.8)^9(.2)^1 + \binom{10}{10}(.8)^{10}(.2)^0 \approx 0.678$$

# 问题11：

$$E(X) = \sum_{i=1}^{k} x_i P(X = x_i)$$

# 什么是随机变量的期望值，它与平均值有什么不同和关联？

If a random variable $X$ is defined on a (finite) sample space $S$, then its expected value is given by

$$E(X) = \sum_{s: s \in S} X(s) P(s). \tag{5.26}$$

大数定律（**law of large numbers (LLN)**）

a theorem that describes the result of performing the same experiment a large number of times. According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

# 期望的线性可加性

**Theorem 5.10**

Suppose $X$ and $Y$ are random variables on the (finite) sample space $S$. Then

$$E(X+Y)=E(X)+E(Y).$$

**Theorem 5.11**

Suppose $X$ is a random variable on a sample space $S$. Then for any number $c$, we have

$$E(cX)=cE(X).$$

Theorems 5.10 and 5.11 are very useful in proving facts about random variables. Taken together, they are typically called **linearity of expectation**. (The idea that the expectation of a sum is the same as the sum of expectations is called the **additivity of expectation**.) The idea of linearity will often allow us

For example, on one flip of a coin, our expected number of heads is .5. Suppose we flip a coin $n$ times and let $X_i$ be the number of heads we see on flip $i$, so that $X_i$ is either 0 or 1. (For example, in five flips of a coin, $X_2(\text{HTHHT}) = 0$ and $X_3(\text{HTHHT}) = 1$.) Then $X$, the total number of heads in $n$ flips, is given by

$$X = X_1 + X_2 + \cdots + X_n, \tag{5.27}$$

$$\begin{aligned} E(X) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= .5 + .5 + \cdots + .5 \\ &= .5n. \end{aligned}$$

# Indicator Random Variable

A random variable that is 1 if a certain event happens and 0 otherwise is called an **indicator random variable**

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs ,} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

Indicator random variables provide a convenient method for converting between probabilities and expectations.

问题12：

这句话是什么意思？

$$E(X_i) = P(X_i = 1) = P(\text{the event occurs}). \qquad (5.28)$$

# 期望值与算法分析

**问题13：**
**这是什么?它为什么在算法分析中很有用？**

```
FindMin(A,n)
// Finds the smallest element in Array A, where n = |A|
(1) min = A[1]
(2) for i = 2 to n
(3)     if (A[i] < min)
(4)         min = A[i]
(5) return min
```

随机变量$X$:
赋值语句执行次数

We solve this problem by letting $X$ be the number of times that min is assigned a value and $X_i$ be the indicator random variable for the event that $A[i]$ is assigned to min. Then $X = X_1 + X_2 + \cdots + X_n$, and $E(X_i)$ is the probability that $A[i]$ is the smallest element in the set $\{A[1], A[2], \ldots, A[i]\}$. Because $(i-1)!$ of the $i!$ permutations of these elements have $A[i]$ as the smallest element, $E(X_i) = 1/i$. Thus,

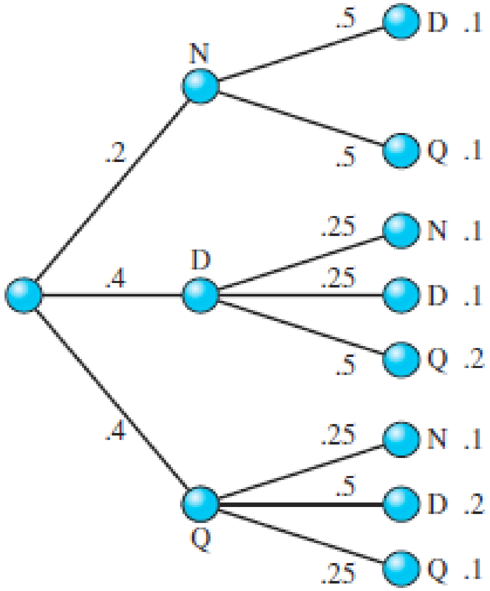
$$E(X) = \sum_{i=1}^{n} \frac{1}{i}.$$

# 讨论

1. Proof and applications of **Principle of Inclusion and Exclusion**

2. A tree for an independent trials process has the property that at each level, for each node at that level, the labeled tree consisting of that node and all its children is identical to each labeled tree consisting of another node at that level and all its children. If we have such a tree, then it automatically satisfies the definition of an independent trials process.

How to prove it?



# 家庭作业

- CS pp.260-: 6, 10-13
- CS pp.274-: 2, 9, 10, 14, 15
- CS pp.290-: 3-4, 8, 11-13
- CS pp.307-: 5, 6, 8, 10, 17, 20, 21