COMP9020 Lectures 9-11
Session 2, 2017
Counting, Probability and Expectation

- Textbook (R & W) - Ch. 5, Sec. 5.1–5.3; Ch. 9
- Problem sets 9–11
- Supplementary Exercises Ch. 5, 9 (R & W)

Announcements

Final Exam ...

- **Friday, 3 November, 8:45am**
- Multiple locations!

Final assignment ...

- Available Saturday
- Due Sunday October 22, 23:59

Lecture 8 recap

Big-O notation

- $O(f(n))$, $\Omega(f(n))$ and $\Theta(f(n))$
- Solving recurrence equations:
  - Unrolling
  - $T(n) = T(n-1) + b.n^k \quad \implies \quad T(n) = n^{k+1}$
  - $T(n) = c.T(n-1) + b.n^k \ (c > 1) \quad \implies \quad T(n) = c^n$.
  - Master theorem

Examples

Recall that $O(f(n))$ is the set of functions for which f is an *upper bound*.

So $3n \in O(n)$ but also $3n \in O(n^2)$, $3n \in O(n^3)$, etc.

In particular $6n^3 \in O(n^3)$ and $3n \in O(n^2)$ but

$$2n^2 = \frac{6n^3}{3n} \notin O(\frac{n^3}{n^2}) = O(n).$$

Note that if $f(n) \in O(h(n))$ and $g(n) \in O(k(n))$ then $f(n)g(n) \in O(h(n)k(n))$.

Examples

Recall that $O(f(n))$ is the set of functions for which f is an *upper bound*.

So $3n \in O(n)$ but also $3n \in O(n^2)$, $3n \in O(n^3)$, etc.

In particular $6n^3 \in O(n^3)$ and $3n \in O(n^2)$ but

$$2n^2 = \frac{6n^3}{3n} \notin O(\frac{n^3}{n^2}) = O(n).$$

Note that if $f(n) \in O(h(n))$ and $g(n) \in O(k(n))$ then $f(n)g(n) \in O(h(n)k(n))$.

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2/\log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2/\log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2 / \log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2/\log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2/\log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Examples

- $3n.\log(n) + 2n^2 \in O(n^2)$
- $\sqrt{7n^3 + 3n + 1} = (7n^3 + 3n + 1)^{\frac{1}{2}} \in O(n^{1.5})$
- $(2^{2.5})^{\log(n)} = (2^{\log(n)})^{2.5} = n^{2.5} \in O(n^{2.5})$
- $5n^{\log(\log(n))} \notin O(n^k)$ for any fixed k
- $n^2/\log(n) \in O(n^{2-\log(n)}) \subsetneq O(n^2)$

Properties of O and Θ

$(f, g) \in R$ if $f \in O(g)$:

- R is reflexive
- R is transitive
- R is *not* anti-symmetric: $n \in O(2n)$ and $2n \in O(n)$ but $n \neq 2n$.

$(f, g) \in S$ if $f \in \Theta(g)$:

- S is reflexive
- S is transitive
- S is symmetric

Master Theorem

Theorem

Suppose $T(n)$ is such that:

$$T(n) = d^{\alpha} \cdot T\left(\frac{n}{d}\right) + \Theta(n^{\beta})$$

- *(case 1)* $\alpha > \beta$: $T(n) = O(n^{\alpha})$
- *(case 2)* $\alpha = \beta$: $T(n) = O(n^{\alpha} \log n)$
- *(case 3)* $\alpha < \beta$: $T(n) = O(n^{\beta})$

Example

$$T(n) = 8T(n/2) + 2n^3$$

- $d = 2$, $\beta = 3$, $\alpha = 3$, so Case 2 applies.
- $T(n) \in O(n^3 \log n)$

Graphs revisited

Recall in a graph G with n vertices and m edges:

- $m \leq n^2$, so $|E| \in O(|V|^2)$
- If G is a tree then $n = m + 1$ so $|E| \in O(|V|)$

Overview

1. Counting techniques
2. Basic and conditional probability
3. Expectation
4. Probability distributions

NB

Combinatorics and probability arise in many areas of Computer Science, e.g.

- *Complexity of algorithms, data management*
- *Reliability, quality assurance*
- *Computer security*
- *Data mining, machine learning, robotics*

Counting Techniques

General idea: find methods, algorithms or precise formulae to count the number of elements in various sets or collections derived, in a structured way, from some basic sets.

Examples

Single base set $S = \{s_1, \ldots, s_n\}$, $|S| = n$; find the number of

- all subsets of S
- ordered selections of r different elements of S
- unordered selections of r different elements of S
- selections of r elements from S s.t. …
- functions $S \longrightarrow S$ (onto, 1-1)
- partitions of S into k equivalence classes
- graphs/trees with elements of S as labelled vertices/leaves

Basic Counting Rules (1)

Union rule — S and T *disjoint*

$$|S \cup T| = |S| + |T|$$

$S_1, S_2, \ldots, S_n$ pairwise disjoint ($S_i \cap S_j = \emptyset$ for $i \neq j$)

$$|S_1 \cup \ldots \cup S_n| = \sum |S_i|$$

Example

How many numbers in $A = [1, 2, \ldots, 999]$ are divisible by 31 or 41?

$\lfloor 999/31 \rfloor = 32$ divisible by 31
$\lfloor 999/41 \rfloor = 24$ divisible by 41
No number in A divisible by both
Hence, $32 + 24 = 56$ divisible by 31 or 41

Basic Counting Rules (1)

Union rule — S and T *disjoint*

$$|S \cup T| = |S| + |T|$$

$S_1, S_2, \ldots, S_n$ pairwise disjoint ($S_i \cap S_j = \emptyset$ for $i \neq j$)

$$|S_1 \cup \ldots \cup S_n| = \sum |S_i|$$

Example

How many numbers in $A = [1, 2, \ldots, 999]$ are divisible by 31 or 41?

$\lfloor 999/31 \rfloor = 32$ divisible by 31
$\lfloor 999/41 \rfloor = 24$ divisible by 41
No number in A divisible by both
Hence, $32 + 24 = 56$ divisible by 31 or 41

Basic Counting Rules (2)

Product rule

$$|S_1 \times \ldots \times S_k| = |S_1| \cdot |S_2| \cdots |S_k| = \prod_{i=1}^{k} |S_i|$$

If all $S_i = S$ (the same set) and $|S| = m$ then $|S^k| = m^k$

Example

Let $\Sigma = \{a, b, c, d, e, f, g\}$.
How many 5-letter words? How many with no letter repeated?

$$|\Sigma^5| = |\Sigma|^5 = 7^5 = 16,807$$

$$\prod_{i=0}^{4}(|\Sigma| - i) = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2,520$$

Exercises

S, T finite. How many functions $S \longrightarrow T$ are there?

$$|T|^{|S|}$$

$\boxed{5.1.19}$ Consider a *complete* graph on n vertices.

(a) No. of paths of length 3

Take any vertex to start, then every next vertex different from the preceding one. Hence $n \cdot (n-1)^3$

(b) paths of length 3 with all vertices distinct

$n(n-1)(n-2)(n-3)$

(c) paths of length 3 with all edges distinct

$n(n-1)(n-2)^2$

Exercises

S, T finite. How many functions $S \longrightarrow T$ are there?

$$|T|^{|S|}$$

$\boxed{5.1.19}$ Consider a *complete* graph on n vertices.

(a) No. of paths of length 3
Take any vertex to start, then every next vertex different from the preceding one. Hence $n \cdot (n-1)^3$

(b) paths of length 3 with all vertices distinct
$n(n-1)(n-2)(n-3)$

(c) paths of length 3 with all edges distinct
$n(n-1)(n-2)^2$

Basic Inferences

For arbitrary sets S, T, ...

$$
\begin{aligned}
|S \cup T| &= |S| + |T| - |S \cap T| \\
|T \setminus S| &= |T| - |S \cap T| \\
|S_1 \cup S_2 \cup S_3| &= |S_1| + |S_2| + |S_3| \\
&\quad - |S_1 \cap S_2| - |S_1 \cap S_3| - |S_2 \cap S_3| \\
&\quad + |S_1 \cap S_2 \cap S_3|
\end{aligned}
$$

Exercise

$\boxed{5.3.1}$ 200 people. 150 swim or jog, 85 swim and 60 do both.
How many jog?

S – (set of) people who swim, J – people who jog
$|S \cup J| = |S| + |J| - |S \cap J|$; thus $150 = 85 + |J| - 60$ hence
$|J| = 125$; answer *does not* depend on the number of people overall

$\boxed{5.6.38}$ (Supp) There are 100 problems, 75 of which are 'easy' and 40 'important'.
What's the smallest number of easy *and* important problems?

$$|E \cap I| = |E| + |I| - |E \cup I| = 75 + 40 - |E \cup I| \geq 75 + 40 - 100 = 15$$

Exercise

$\boxed{5.3.1}$ 200 people. 150 swim or jog, 85 swim and 60 do both.
How many jog?

S – (set of) people who swim, J – people who jog
$|S \cup J| = |S| + |J| - |S \cap J|$; thus $150 = 85 + |J| - 60$ hence
$|J| = 125$; answer *does not* depend on the number of people overall

$\boxed{5.6.38}$ (Supp) There are 100 problems, 75 of which are 'easy' and 40 'important'.
What's the smallest number of easy *and* important problems?

$$|E \cap I| = |E| + |I| - |E \cup I| = 75 + 40 - |E \cup I| \geq 75 + 40 - 100 = 15$$

Exercise

$\boxed{5.3.2}$ $S = [100 \ldots 999]$, thus $|S| = 900$.

(a) How many numbers have at least one digit that is a 3 or 7?

$A_3 = \{\text{at least one '3'}\}$
$A_7 = \{\text{at least one '7'}\}$

$(A_3 \cup A_7)^c = \{\ n \in [100, 999] : n \text{ digits } \in \{0, 1, 2, 4, 5, 6, 8, 9\}\ \}$

7 choices for the first digit and 8 choices for the later digits

$$|(A_3 \cup A_7)^c| = |\{1, 2, 4, 5, 6, 8, 9\}| \cdot |\{0, 1, 2, 4, 5, 6, 8, 9\}|^2$$

Therefore $|A_3 \cup A_7| = 900 - 448 = 452$

(b) How many numbers have a 3 *and* a 7?

$|A_3 \cap A_7| = |A_3| + |A_7| - |A_3 \cup A_7| =$
$(900 - 8 \cdot 9 \cdot 9) + (900 - 8 \cdot 9 \cdot 9) - 452 = 2 \cdot 252 - 452 = 52$

Exercise

$\boxed{5.3.2}$ $S = [100 \ldots 999]$, thus $|S| = 900$.

(a) How many numbers have at least one digit that is a 3 or 7?
$A_3 = \{\text{at least one '3'}\}$
$A_7 = \{\text{at least one '7'}\}$

$(A_3 \cup A_7)^c = \{\ n \in [100, 999] : n \text{ digits } \in \{0, 1, 2, 4, 5, 6, 8, 9\}\ \}$

7 choices for the first digit and 8 choices for the later digits

$$|(A_3 \cup A_7)^c| = |\{1, 2, 4, 5, 6, 8, 9\}| \cdot |\{0, 1, 2, 4, 5, 6, 8, 9\}|^2$$

Therefore $|A_3 \cup A_7| = 900 - 448 = 452$

(b) How many numbers have a 3 *and* a 7?
$|A_3 \cap A_7| = |A_3| + |A_7| - |A_3 \cup A_7| =$
$(900 - 8 \cdot 9 \cdot 9) + (900 - 8 \cdot 9 \cdot 9) - 452 = 2 \cdot 252 - 452 = 52$

Corollaries

- If $|S \cup T| = |S| + |T|$ then S and T are disjoint
- If $|\bigcup_{i=1}^{n} S_i| = \sum_{i=1}^{n} |S_i|$ then S_i are pairwise disjoint
- If $|T \setminus S| = |T| - |S|$ then $S \subseteq T$

These properties can serve to identify cases when sets are disjoint (resp. one is contained in the other).

Proof.

$|S| + |T| = |S \cup T|$ means $|S \cap T| = |S| + |T| - |S \cup T| = 0$

$|T \setminus S| = |T| - |S|$ means $|S \cap T| = |S|$ means $S \subseteq T$ $\square$

Combinatorial Objects: How Many?

permutations

Ordering of all objects from a set S; equivalently: Selecting all objects while *recognising* the order of selection.

The number of permutations of n elements is

$$n! = n \cdot (n-1) \cdots 1, \quad 0! = 1! = 1$$

r**-permutations**

Selecting any r objects from a set S of size n without repetition while *recognising* the order of selection.

Their number is

$$\Pi(n,r) = n \cdot (n-1) \cdots (n-r+1) = \frac{n!}{(n-r)!}$$

r-**selections** (or: r-**combinations**)
Collecting any r distinct objects without repetition;
equivalently: selecting r objects from a set S of size n and *not* recognising the order of selection.
Their number is

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} = \frac{n \cdot (n-1) \cdots (n-r+1)}{1 \cdot 2 \cdots r}$$

NB

These numbers are usually called binomial coefficients due to

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \ldots + b^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$$

Also defined for any $\alpha \in \mathbb{R}$ *as* $\displaystyle\binom{\alpha}{r} = \frac{\alpha(\alpha-1)\cdots(\alpha-r+1)}{r!}$

Simple Counting Problems

Example

$\boxed{5.1.2}$ Give an example of a counting problem whose answer is

(a) $\Pi(26, 10)$

(b) $\binom{26}{10}$

Draw 10 cards from a half deck (eg. black cards only)
(a) the cards are recorded in the order of appearance
(b) only the complete draw is recorded

Examples

- Number of edges in a complete graph K_n
- Number of diagonals in a convex polygon
- Number of poker hands
- Decisions in games, lotteries etc.

Simple Counting Problems

Example

$\boxed{5.1.2}$ Give an example of a counting problem whose answer is

(a) $\Pi(26, 10)$

(b) $\binom{26}{10}$

Draw 10 cards from a half deck (eg. black cards only)
(a) the cards are recorded in the order of appearance
(b) only the complete draw is recorded

Examples

- Number of edges in a complete graph K_n
- Number of diagonals in a convex polygon
- Number of poker hands
- Decisions in games, lotteries etc.

Exercise

$\boxed{5.1.6}$ From a group of 12 men and 16 women, how many committees can be chosen consisting of

(a) 7 members? $\binom{12+16}{7}$

(b) 3 men and 4 women? $\binom{12}{3}\binom{16}{4}$

(c) 7 women or 7 men? $\binom{12}{7}+\binom{16}{7}$

$\boxed{5.1.7}$ As above, but any 4 people (male or female) out of 9 and two, Alice and Bob, unwilling to serve on the same committee.

$\{$all committees$\}$ $-$ $\{$committees with both A and $B\}$
$= \binom{9}{4}-\binom{7}{2}=126-21=105$

equivalently, $\{$A in, B out$\}$ $+$ $\{$A out, B in$\}$ $+$ $\{$none in$\}$
$= \binom{7}{3}+\binom{7}{3}+\binom{7}{4}=35+35+35=105$

Exercise

$\boxed{5.1.6}$ From a group of 12 men and 16 women, how many committees can be chosen consisting of

(a) 7 members? $\binom{12+16}{7}$

(b) 3 men and 4 women? $\binom{12}{3}\binom{16}{4}$

(c) 7 women or 7 men? $\binom{12}{7}+\binom{16}{7}$

$\boxed{5.1.7}$ As above, but any 4 people (male or female) out of 9 and two, Alice and Bob, unwilling to serve on the same committee.

{all committees} − {committees with both A and B}
$= \binom{9}{4} - \binom{7}{2} = 126 - 21 = 105$

equivalently, {A in, B out} + {A out, B in} + {none in}
$$= \binom{7}{3} + \binom{7}{3} + \binom{7}{4} = 35 + 35 + 35 = 105$$

Counting Poker Hands

$\boxed{5.1.15}$ A poker hand consists of 5 cards drawn without replacement from a standard deck of 52 cards

$$\{A, 2\text{-}10, J, Q, K\} \times \{\text{club, spade, heart, diamond}\}$$

(a) Number of "4 of a kind" hands (e.g. 4 Jacks)

$|\text{rank of the 4-of-a-kind}| \cdot |\text{any other card}| = 13 \cdot (52 - 4)$

(b) Number of non-straight flushes, i.e. all cards of same suit but *not* consecutive (e.g. 8,9,10,J,K)

$|\text{all flush}| - |\text{straight flush}|$

$= |\text{suit}| \cdot |\text{5-hand in a given suit}| -$

$\quad |\text{suit}| \cdot |\text{rank of a straight flush in a given suit}|$

$= 4 \cdot \binom{13}{5} - 4 \cdot 10$

Counting Poker Hands

$\boxed{5.1.15}$ A poker hand consists of 5 cards drawn without replacement from a standard deck of 52 cards

$$\{A, 2\text{-}10, J, Q, K\} \times \{\text{club, spade, heart, diamond}\}$$

(a) Number of "4 of a kind" hands (e.g. 4 Jacks)
$|\text{rank of the 4-of-a-kind}| \cdot |\text{any other card}| = 13 \cdot (52 - 4)$

(b) Number of non-straight flushes, i.e. all cards of same suit but *not* consecutive (e.g. 8,9,10,J,K)
$|\text{all flush}| - |\text{straight flush}|$
$= |\text{suit}| \cdot |\text{5-hand in a given suit}| -$
$\quad |\text{suit}| \cdot |\text{rank of a straight flush in a given suit}|$
$= 4 \cdot \binom{13}{5} - 4 \cdot 10$

“Balls in boxes”

Example

Have n “distinguishable” boxes.
Have $k \leq n$ balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **at most one** ball per box?

NB

Case 2 is the same as the number of injections from K to N where $|K| = k$ and $|N| = n$.

"Balls in boxes"

Example

Have n "distinguishable" boxes.
Have $k \leq n$ balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **at most one** ball per box?

1. $\binom{n}{k}$
2. $\Pi(n, k)$

"Balls in boxes" continued

Example

Have n "distinguishable" boxes.
Have k balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **any number** of balls per box?

1. $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$
2. $\Pi(n+k-1, k)$

"Balls in boxes" continued

Example

Have n "distinguishable" boxes.
Have k balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **any number** of balls per box?

1. $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$
2. $\Pi(n+k-1, k)$

“Balls in boxes” continued

Example

Have n “distinguishable” boxes.
Have $k \geq n$ balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **at least** one ball per box?

NB

UPDATE (10/10) Case 2 is NOT the same as the number of surjections from K to N where $|K| = k$ and $|N| = n$

"Balls in boxes" continued

Example

Have n "distinguishable" boxes.
Have $k \geq n$ balls which are either:

1. Indistinguishable
2. Distinguishable

How many ways to place balls in boxes with **at least** one ball per box?
Place n balls in boxes. Distribute remaining $k-n$ balls however.

1. $\binom{n+(k-n)-1}{n-1} = \binom{k-1}{n-1}$
2. $\binom{k-1}{n-1}.k!$

“Balls in boxes” continued

Example

Have n “distinguishable” boxes.
Have k “distinguishable” and “replaceable” balls (i.e. many copies)
How many ways to place balls in boxes with **exactly one** ball per box?

NB

This is the same as the number of functions from K to N where $|K| = k$ and $|N| = n$.

“Balls in boxes” continued

Example

Have n “distinguishable” boxes.
Have k “distinguishable” and “replaceable” balls (i.e. many copies)
How many ways to place balls in boxes with **exactly one** ball per box?

n^k

Difficult Counting Problems

Example (Ramsay numbers)

An example of a *Ramsay number* is $R(3,3)=6$, meaning that

> *"K_6 is the smallest complete graph s.t. if all edges are painted using two colours, then there must be at least one monochromatic triangle"*

This serves as the basis of a game called S-I-M (invented by Simmons), where two adversaries connect six dots, respectively using blue and red lines. The objective is to *avoid* closing a triangle of one's own colour. The second player has a winning strategy, but the full analysis requires a computer program.

Using Programs to Count

Two dice, a red die and a black die, are rolled.
(Note: one *die*, two or more *dice*)

Write a program to list all the pairs $\{(R, B) : R > B\}$

Similarly, for three dice, list all triples $R > B > G$

Generally, for n dice, all of which are m-sided $(n \leq m)$, list all *decreasing* n-tuples

NB

In order to just find the number of such n-tuples, it is not necessary to list them all. One can write a recurrence relation for these numbers and compute (or try to solve) it.

Approximate Counting

NB

A Count may be a precise value or an **estimate**.

The latter should be asymptotically correct or at least give a good asymptotic bound, whether upper or lower. If S is the base set, $|S| = n$ its size, and we denote by $c(S)$ some collection of objects from S we are interested in, then we seek constants a, b s.t.

$$a \leq \lim_{n \to \infty} \frac{est(|c(S)|)}{|c(S)|} \leq b$$

Probability

Elementary Probability

Definition

Sample space:

$$\Omega = \{\omega_1, \ldots, \omega_n\}$$

Each point represents an outcome.

Event: a collection of outcomes = subset of Ω

Probability distribution: A function $P : \text{Pow}(\Omega) \to \mathbb{R}$ such that:

- $P(\Omega) = 1$
- E and F disjoint events then $P(E \cup F) = P(E) + P(F)$.

Fact

$$P(\emptyset) = 0, \quad P(E^c) = 1 - P(E)$$

Elementary Probability

Each outcome ω_i equally likely:

$$P(\omega_1) = P(\omega_2) = \ldots = P(\omega_n) = \frac{1}{n}$$

This a called a **uniform probability distribution** over Ω

Examples

Tossing a coin: $\Omega = \{H, T\}$

$$P(H) = P(T) = 0.5$$

Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

Exercises

$\boxed{5.2.7}$ Suppose an experiment leads to events A, B with probabilities $P(A) = 0.5, P(B) = 0.8, P(A \cap B) = 0.4$. Find

- $P(B^c) = 1 - P(B) = 0.2$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.9$
- $P(A^c \cup B^c) = 1 - P((A^c \cup B^c)^c) = 1 - P(A \cap B) = 0.6$

$\boxed{5.2.8}$ Given $P(A) = 0.6,\ P(B) = 0.7$, show $P(A \cap B) \geq 0.3$

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.6 + 0.7 - P(A \cup B) \\ &\geq 0.6 + 0.7 - 1 = 0.3 \end{aligned}$$

Exercises

$\boxed{5.2.7}$ Suppose an experiment leads to events A, B with probabilities $P(A) = 0.5, P(B) = 0.8, P(A \cap B) = 0.4.$
Find

- $P(B^c) = 1 - P(B) = 0.2$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.9$
- $P(A^c \cup B^c) = 1 - P((A^c \cup B^c)^c) = 1 - P(A \cap B) = 0.6$

$\boxed{5.2.8}$ Given $P(A) = 0.6,\ P(B) = 0.7$, show $P(A \cap B) \geq 0.3$

$$
\begin{aligned}
P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\
&= 0.6 + 0.7 - P(A \cup B) \\
&\geq 0.6 + 0.7 - 1 = 0.3
\end{aligned}
$$

Computing Probabilities by Counting

Computing probabilities with respect to a *uniform* distribution comes down to counting the size of the event.
If $E = \{e_1, \ldots, e_k\}$ then

$$P(E) = \sum_{i=1}^{k} P(e_i) = \sum_{i=1}^{k} \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}$$

Most of the counting rules carry over to probabilities wrt. a uniform distribution.

NB

The expression "selected at random", when not further qualified, means:
"subject to / according to / . . . a uniform *distribution."*

Examples

$\boxed{5.6.38}$ (Supp) Of 100 problems, 75 are 'easy' and 40 'important'.
(b) n problems chosen randomly. What is the probability that all n are important?

$$p = \frac{\binom{40}{n}}{\binom{100}{n}} = \frac{40 \cdot 39 \cdots (41-n)}{100 \cdot 99 \cdots (101-n)}$$

$\boxed{5.2.3}$ A 4-letter word is selected at random from Σ^4, where $\Sigma = \{a, b, c, d, e\}$. What is the probability that
(a) the letters in the word are all distinct?
(b) there are no vowels ("a", "e") in the word?
(c) the word begins with a vowel?

(a) $|E| = \Pi(5,4)$, $\quad P(E) = \frac{5 \cdot 4 \cdot 3 \cdot 2}{5^4} = \frac{120}{625} \approx 19\%$
(b) $|E| = 3^4$, $\quad P(E) = \frac{81}{625} \approx 13\%$
(c) $|E| = 2 \cdot 5^3$, $\quad P(E) = \frac{2}{5}$

Examples

$\boxed{5.6.38}$ (Supp) Of 100 problems, 75 are 'easy' and 40 'important'.
(b) n problems chosen randomly. What is the probability that all n are important?

$$p = \frac{\binom{40}{n}}{\binom{100}{n}} = \frac{40 \cdot 39 \cdots (41-n)}{100 \cdot 99 \cdots (101-n)}$$

$\boxed{5.2.3}$ A 4-letter word is selected at random from Σ^4, where $\Sigma = \{a, b, c, d, e\}$. What is the probability that
(a) the letters in the word are all distinct?
(b) there are no vowels ("a", "e") in the word?
(c) the word begins with a vowel?

(a) $|E| = \Pi(5,4)$, $\quad P(E) = \frac{5 \cdot 4 \cdot 3 \cdot 2}{5^4} = \frac{120}{625} \approx 19\%$
(b) $|E| = 3^4$, $\quad P(E) = \frac{81}{625} \approx 13\%$
(c) $|E| = 2 \cdot 5^3$, $\quad P(E) = \frac{2}{5}$

Exercise

$\boxed{5.2.11}$ Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?

$P(R + B \in \{2, 4, \ldots, 12\}) = \frac{18}{36} = \frac{1}{2}$

(b) the number on the red die is bigger than on the black die?

$P(R > B) = P(R < B)$; also $P(R = B) = \frac{1}{6}$
Therefore $P(R < B) = \frac{1}{2}(1 - P(R = B)) = \frac{5}{12}$

(c) the number on the black die is twice the one on the red die?

$P(R = 2 \cdot B) = P(\{(2,1), (4,2), (6,3)\}) = \frac{3}{36} = \frac{1}{12}$

$\boxed{5.2.12}$ (a) the maximum of the numbers is 4? $P(E_1) = \frac{7}{36}$
(b) their minimum is 4? $P(E_2) = \frac{5}{36}$

Check:

$P(E_1 \cup E_2) = \frac{7}{36} + \frac{5}{36} - P(E_1 \cap E_2) = \frac{7+5-1}{36} = \frac{11}{36}$

$P(\text{at least one '4'}) = 1 - P(\text{no '4'}) = 1 - \frac{5}{6} \cdot \frac{5}{6} = \frac{11}{36}$

Exercise

$\boxed{5.2.11}$ Two dice, a red die and a black die, are rolled.
What is the probability that
(a) the sum of the values is even?
$P\big(R + B \in \{2, 4, \ldots, 12\}\big) = \frac{18}{36} = \frac{1}{2}$

(b) the number on the red die is bigger than on the black die?
$P(R > B) = P(R < B)$; also $P(R = B) = \frac{1}{6}$
Therefore $P(R < B) = \frac{1}{2}(1 - P(R = B)) = \frac{5}{12}$

(c) the number on the black die is twice the one on the red die?
$P(R = 2 \cdot B) = P(\{(2,1), (4,2), (6,3)\}) = \frac{3}{36} = \frac{1}{12}$

$\boxed{5.2.12}$ (a) the maximum of the numbers is 4? $P(E_1) = \frac{7}{36}$
(b) their minimum is 4? $P(E_2) = \frac{5}{36}$

Check:

$P(E_1 \cup E_2) = \frac{7}{36} + \frac{5}{36} - P(E_1 \cap E_2) = \frac{7+5-1}{36} = \frac{11}{36}$
$P(\text{at least one `4'}) = 1 - P(\text{no `4'}) = 1 - \frac{5}{6} \cdot \frac{5}{6} = \frac{11}{36}$

Exercise

$\boxed{5.2.5}$ An urn contains 3 red and 4 black balls. 3 balls are removed without replacement. What are the probabilities that
(a) all 3 are red
(b) all 3 are black
(c) one is red, two are black

All probabilities are computed using the same sample space: all possible ways to draw three balls without replacement.

The size of the sample space is $\dfrac{7 \cdot 6 \cdot 5}{3!} = 35$

(a) $E =$ All balls are red: 1 combination
(b) $E =$ All balls are black: $\binom{4}{3} = 4$ combinations
(c) $E =$ One red and two black: $\binom{3}{1} \cdot \binom{4}{2} = 18$ combinations

Exercise

$\boxed{5.2.5}$ An urn contains 3 red and 4 black balls. 3 balls are removed without replacement. What are the probabilities that
(a) all 3 are red
(b) all 3 are black
(c) one is red, two are black

All probabilities are computed using the same sample space: all possible ways to draw three balls without replacement.

The size of the sample space is $\dfrac{7 \cdot 6 \cdot 5}{3!} = 35$

(a) $E =$ All balls are red: 1 combination
(b) $E =$ All balls are black: $\binom{4}{3} = 4$ combinations
(c) $E =$ One red and two black: $\binom{3}{1} \cdot \binom{4}{2} = 18$ combinations

Infinite sample spaces

Probability distributions generalize to infinite sample spaces **with some provisos**.

- In continuous spaces (e.g. $\mathbb{R}$):
  - Probability distributions are *measures*;
  - Sums are *integrals*;
  - Non-zero probabilities apply to *ranges*;
  - Probability of a single event is 0. Note: Probability 0 is not the same as impossible.
- In discrete spaces (e.g. $\mathbb{N}$):
  - Probability 0 is the same as impossible.
  - No uniform distribution!
  - Non-uniform distributions exist, e.g. $P(0) = 1$, $P(n) = 0$ for $n > 0$; or $P(0) = 0$, $P(n) = \frac{1}{2^n}$ for $n > 0$.

Asymptotic Estimate of Relative Probabilities

Example

Event $A \stackrel{\text{def}}{=}$ one die rolled n times and you obtain two 6's
Event $B \stackrel{\text{def}}{=}$ n dice rolled simultaneously and you obtain one 6

$$P(A) = \frac{\binom{n}{2} \cdot 5^{n-2}}{6^n} \qquad P(B) = \frac{\binom{n}{1} \cdot 5^{n-1}}{6^n}$$

Therefore $\frac{P(A)}{P(B)} = \frac{\binom{n}{2}}{\binom{n}{1}} \cdot \frac{1}{5} = \frac{n(n-1)}{2} \cdot \frac{1}{5n} = \frac{n-1}{10} \in \Theta(n)$

n	1	2	3	4	...	11	...	20	...
$P(A)$	0	$\frac{1}{36}$	$\frac{5}{72}$	$\frac{25}{216}$	...	0.296	...	0.198	...
$P(B)$	$\frac{1}{6}$	$\frac{10}{36}$	$\frac{25}{72}$	$\frac{125}{324}$	...	0.296	...	0.104	...

Inclusion-Exclusion

This is one of the most universal counting procedures. It allows you to compute the size of

$$A_1 \cup \ldots \cup A_n$$

from the sizes of all possible intersections

$$A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}, \;\; a_{i_1} < a_{i_2} < \ldots < a_{i_k}$$

Two sets $|A \cup B| = |A| + |B| - |A \cap B|$

Three sets $|A \cup B \cup C| = |A| + |B| + |C|$
$$-|A \cap B| - |A \cap C| - |B \cap C|$$
$$+|A \cap B \cap C|$$

NB

Inclusion-exclusion is often applied informally without making clear or explicit why certain quantities are subtracted or put back in.

Interpretation

Each A_i defined as the set of objects that satisfy some property P_i

$$A_i = \{\, x \in X : P_i(x) \,\}$$

Union $A_1 \cup \ldots \cup A_n$ is the set of objects that satisfy **at least one** property P_i

$$A_1 \cup \ldots \cup A_n = \{\, x \in X : P_1(x) \vee P_2(x) \vee \ldots \vee P_n(x) \,\}$$

Intersection $A_{i_1} \cap \ldots \cap A_{i_r}$ is the set of objects that satisfy **all** properties $P_{i_1}, \ldots, P_{i_r}$

$$A_{i_1} \cap \ldots \cap A_{i_r} = \{\, x \in X : P_{i_1}(x) \wedge P_{i_2}(x) \wedge \ldots \wedge P_{i_r}(x) \,\}$$

Special case $r = 1$: $A_{i_1} = \{x \in X : P_{i_1}(x)\}$

Inclusion-Exclusion is a very common method for deriving probabilities from other probabilities.

Two sets

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

Three sets

$$\begin{aligned}
P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\
&= P(A) + P(B) - P(A \cap B) + P(C) \\
&\qquad - P((A \cap C) \cup (B \cap C)) \\
&= P(A) + P(B) - P(A \cap B) + P(C) \\
&\qquad - \big(P(A \cap C) + P(B \cap C) - P(A \cap C \cap B \cap C)\big) \\
&= P(A) + P(B) + P(C) \\
&\qquad - P(A \cap C) - P(A \cap C) - P(B \cap C) \\
&\qquad + P(A \cap B \cap C)
\end{aligned}$$

Example

A four-digit number n is selected at random (i.e. randomly from $[1000 \ldots 9999]$). Find the probability p that n has each of 0, 1, 2 among its digits.

Let $q = 1 - p$ be the complementary probability and define

$$A_i = \{n : \text{no digit } i\}, A_{ij} = \{n : \text{no digits } i, j\}, A_{ijk} = \{n : \text{no } i, j, k\}$$

Then define
$T = A_0 \cup A_1 \cup A_2 = \{n : \text{ missing at least one of } 0, 1, 2\}$
$S = (A_0 \cup A_1 \cup A_2)^c = \{n : \text{ containing each of } 0, 1, 2\}$

Example (cont'd)

Once we find the cardinality of T, the solution is

$$q = \frac{|T|}{9000}, \; p = 1 - q$$

To find $|A_i|, |A_{ij}|, |A_{ijk}|$ we reflect on how many choices are available for the first digit, for the second etc. A special case is the leading digit, which must be $1, \ldots, 9$

Example (cont'd)

$$|A_0| = 9^4, \quad |A_1| = |A_2| = 8 \cdot 9^3$$
$$|A_{01}| = |A_{02}| = 8^4, \quad |A_{12}| = 7 \cdot 8^3$$
$$|A_{012}| = 7^4$$

$$\begin{aligned} |T| &= |A_0 \cup A_1 \cup A_2| \\ &= |A_0| + |A_1| + |A_2| - |A_0 \cap A_1| - |A_0 \cap A_2| - |A_1 \cap A_2| \\ &\qquad + |A_0 \cap A_1 \cap A_2| \\ &= 9^4 + 2 \cdot 8 \cdot 9^3 - 2 \cdot 8^4 - 7 \cdot 8^3 + 7^4 \\ &= 25 \cdot 9^3 - 23 \cdot 8^3 + 7^4 = 8850 \end{aligned}$$

$$q = \frac{8850}{9000}, \quad p = 1 - q \quad \approx 0.01667$$

Previous example generalised: Probability of an r-digit number having all of 0,1,2,3 among its digits.
We use the previous notation: A_i — set of numbers n *missing* digit i, and similarly for all $A_{ij\ldots}$
We aim to find the size of $T = A_0 \cup A_1 \cup A_2 \cup A_3$, and then to compute $|S| = 9 \cdot 10^{r-1} - |T|$.

$$
\begin{aligned}
|A_0 \cup A_1 \cup A_2 \cup A_3| &= \text{sum of } |A_i| \\
&- \text{sum of } |A_i \cap A_j| \\
&+ \text{sum of } |A_i \cap A_j \cap A_k| \\
&- \text{sum of } |A_i \cap A_j \cap A_k \cap A_l|
\end{aligned}
$$

Probability of Sequential Outcomes

Example

Team A has probability $p = 0.5$ of winning a game against B.
What is the probability P_p of A winning a best-of-seven match if
(a) A already won the first game?
(b) A already won the first two games?
(c) A already won two out of the first three games?

(a) Sample space S — 6-sequences, formed from wins (W) and losses (L)

$$|S| = 2^6 = 64$$

Favourable sequences F — those with three to six W

$$|F| = \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 20 + 15 + 6 + 1 = 42$$

Therefore $P_{0.5} = \frac{42}{64} \approx 66\%$

Probability of Sequential Outcomes

Example

Team A has probability $p = 0.5$ of winning a game against B.
What is the probability P_p of A winning a best-of-seven match if
(a) A already won the first game?
(b) A already won the first two games?
(c) A already won two out of the first three games?

(a) Sample space S — 6-sequences, formed from wins (W) and losses (L)

$$|S| = 2^6 = 64$$

Favourable sequences F — those with three to six W

$$|F| = \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 20 + 15 + 6 + 1 = 42$$

Therefore $P_{0.5} = \frac{42}{64} \approx 66\%$

Example (cont'd)

(b) Sample space S — 5-sequences of W and L

$$|S| = 2^5 = 32$$

Favourable sequences F — those with two to five W

$$|F| = \binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 10 + 10 + 5 + 1 = 26$$

Therefore $P_{0.5} = \frac{26}{32} \approx 81\%$

(c)

$$|S| = 2^4 = 16$$

$$|F| = \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 6 + 4 + 1 = 11$$

Therefore $P_{0.5} = \frac{11}{16} \approx 69\%$

Example (cont'd)

Redo for arbitrary p

(a)

$$P_p = \binom{6}{3} p^3(1-p)^3 + \binom{6}{4} p^4(1-p)^2 + \binom{6}{5} p^5(1-p) + \binom{6}{6} p^6$$

(b)

$$P_p = \binom{5}{2} p^2(1-p)^3 + \binom{5}{3} p^3(1-p)^2 + \binom{5}{4} p^5(1-p) + \binom{5}{5} p^5$$

(c)

$$P_p = \binom{4}{2} p^2(1-p)^2 + \binom{4}{3} p^3(1-p) + \binom{4}{4} p^4$$

Use of Recursion in Probability Computations

Question

Given n tosses of a coin, what is the probability of two HEADS *in a row? Compute for* $n = 5, 10, 20, \ldots$

Approaches:

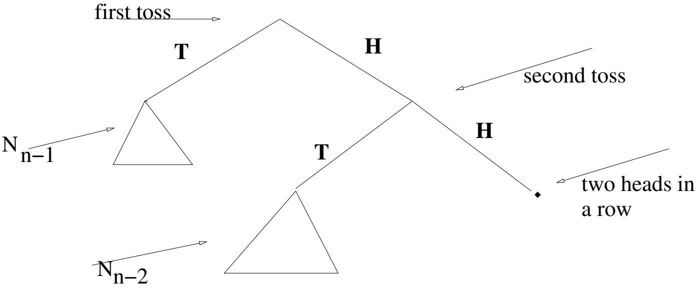
I. Write down all possibilities — 32 for $n = 5$, 1024 for $n = 10$, ...

II. Write a program; running time $O(2^n)$ — why?

III. Inter-relate the numbers of relevant possibilities
$N_n \stackrel{\text{def}}{=}$ No. of sequences of n tosses *without* ...HH... pattern
Initial values:
$N_0 = 1$, $N_1 = 2$, $N_2 = 3$ (all except "HH")
$N_3 = 5$ (why?) $N_4 = 8$ (why?)

Answer

We can summarise all possible outcomes in a **recursive tree**



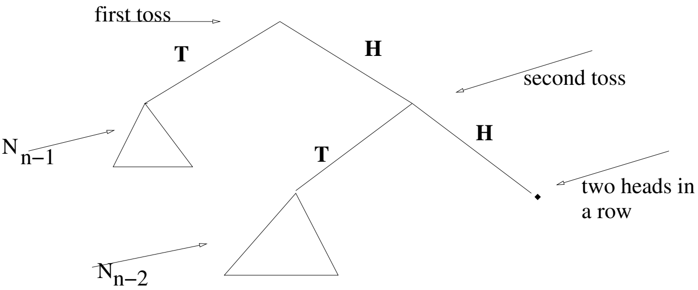
$N_n = N_{n-1} + N_{n-2}$ — Fibonacci recurrence: $N_n = \mathrm{FIB}(n+1)$

$N_n \approx \frac{1}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{2}\right)^{n+1} \approx 0.72 \cdot (1.6)^n$

$p_n = \frac{2^n - \mathrm{FIB}(n+1)}{2^n} \approx 1 - 0.72 \cdot (0.8)^n$

Answer

We can summarise all possible outcomes in a **recursive tree**



$N_n = N_{n-1} + N_{n-2}$ — Fibonacci recurrence: $N_n = \textsc{Fib}(n+1)$

$N_n \approx \frac{1}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{2}\right)^{n+1} \approx 0.72 \cdot (1.6)^n$

$p_n = \frac{2^n - \textsc{Fib}(n+1)}{2^n} \approx 1 - 0.72 \cdot (0.8)^n$

Example

Question

Given n tosses, what is the probability q_n of at least one HHH?

$q_0 = q_1 = q_2 = 0; q_3 = \frac{1}{8}$

Then recursive computation:

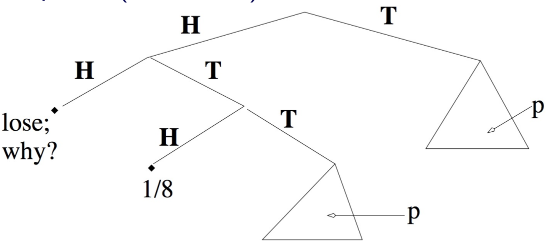
$$\begin{aligned} q_n = & \frac{1}{2} q_{n-1} && \text{(initial: T)} \\ & + \frac{1}{4} q_{n-2} && \text{(initial: HT)} \\ & + \frac{1}{8} q_{n-3} && \text{(initial: HHT)} \\ & + \frac{1}{8} && \text{(start with: HHH)} \end{aligned}$$

Example

Question

A coin is tossed 'indefinitely'. Which pattern is more likely (and by how much) to appear first, HTH or HHT?

let $p = P(HTH \text{ first})$



$$p = \frac{1}{8} + \frac{1}{8}p + \frac{1}{2}p \;\rightarrow\; \frac{3}{8}p = \frac{1}{8} \;\rightarrow\; p = \frac{1}{3}$$

NB

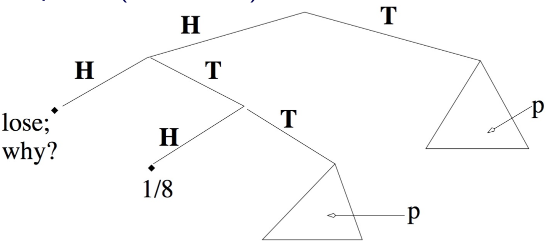
Probability that either pattern would appear at a given, prespecified point in the sequence of tosses is, obviously, the same.

Example

Question

A coin is tossed 'indefinitely'. Which pattern is more likely (and by how much) to appear first, HTH or HHT?

let $p = P(HTH \text{ first})$



$$p = \frac{1}{8} + \frac{1}{8}p + \frac{1}{2}p \;\rightarrow\; \frac{3}{8}p = \frac{1}{8} \;\rightarrow\; p = \frac{1}{3}$$

NB

Probability that either pattern would appear at a given, prespecified point in the sequence of tosses is, obviously, the same.

Example

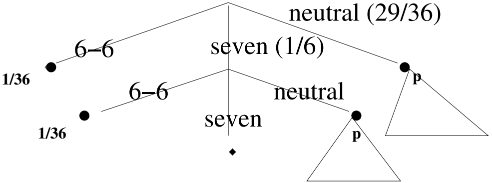
Question

Two dice are rolled repeatedly. What is the probability that '6–6' will occur before two consecutive (back-to-back) 'totals seven'?

NB

The probability of either occurring at a given roll is the same: $\frac{1}{36}$.

Let $p = P(\text{6–6 first})$



$$p = \frac{1}{36} + \frac{1}{6}\cdot\frac{1}{36} + \frac{1}{6}\cdot\frac{29}{36}p + \frac{29}{36}p \;\rightarrow\; 216p = 7 + 203p \;\rightarrow\; p = \frac{7}{13}$$

Example

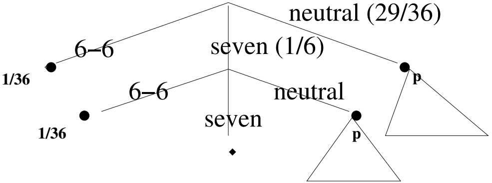
Question

Two dice are rolled repeatedly. What is the probability that '6–6' will occur before two consecutive (back-to-back) 'totals seven'?

NB

The probability of either occurring at a given roll is the same: $\frac{1}{36}$.

Let $p = P(6\text{–}6 \text{ first})$



$$p = \frac{1}{36} + \frac{1}{6}\cdot\frac{1}{36} + \frac{1}{6}\cdot\frac{29}{36}p + \frac{29}{36}p \;\rightarrow\; 216p = 7 + 203p \;\rightarrow\; p = \frac{7}{13}$$

NB

The majority of problems in probability and statistics do not have such elegant solutions. Hence the use of computers for either precise calculations or approximate simulations is mandatory. However, it is the use of recursion that simplifies such computing or, quite often, makes it possible in the first place.

Conditional Probability

Conditional Probability

Definition

Conditional probability of E **given** S:

$$P(E|S) = \frac{P(E \cap S)}{P(S)}, \quad E, S \subseteq \Omega$$

It is defined only when $P(S) \neq 0$

NB

$P(A|B)$ and $P(B|A)$ are, in general, not related — one of these values predicts, by itself, essentially nothing about the other. The only exception, applicable when $P(A), P(B) \neq 0$, is that $P(A|B) = 0$ iff $P(B|A) = 0$ iff $P(A \cap B) = 0$.

If P is the uniform distribution over a finite set Ω, then

$$P(E|S) = \frac{\frac{|E \cap S|}{|\Omega|}}{\frac{|S|}{|\Omega|}} = \frac{|E \cap S|}{|S|}$$

This observation can help in calculations...

Example

$\boxed{9.1.6}$ A coin is tossed four times. What is the probability of
(a) two consecutive HEADS
(b) two consecutive HEADS *given* that ≥ 2 tosses are HEADS

T	T	T	T	H	T	T	T
T	T	T	H	H	T	T	H
T	T	H	T	H	T	H	T
T	T	H	H	H	T	H	H
T	H	T	T	H	H	T	T
T	H	T	H	H	H	T	H
T	H	H	T	H	H	H	T
T	H	H	H	H	H	H	H

(a) $\frac{8}{16}$ (b) $\frac{8}{11}$

Some General Rules

Fact

- $A \subseteq B \to P(A|B) \geq P(A)$
- $A \subseteq B \to P(B|A) = 1$
- $P(A \cap B|B) = P(A|B)$
- $P(\emptyset|A) = 0$ *for* $A \neq \emptyset$
- $P(A|\Omega) = P(A)$
- $P(A^c|B) = 1 - P(A|B)$

NB

- $P(A|B)$ *and* $P(A|B^c)$ *are not related*
- $P(A|B), P(B|A), P(A^c|B^c), P(B^c|A^c)$ *are not related*

Example

Two dice are rolled and the outcomes recorded as b for the black die, r for the red die and $s = b + r$ for their total.
Define the events $B = \{b \geq 3\},\ R = \{r \geq 3\},\ S = \{s \geq 6\}$.

$$P(S|B) = \frac{4+5+6+6}{24} = \frac{21}{24} = \frac{7}{8} = 87.5\%$$

$$P(B|S) = \frac{4+5+6+6}{26} = \frac{21}{26} = 80.8\%$$

The (common) numerator $4 + 5 + 6 + 6 = 21$ represents the size of the $B \cap S$ — the common part of B and S, that is, the number of rolls where $b \geq 3$ and $s \geq 6$. It is obtained by considering the different cases: $b = 3$ and $s \geq 6$, then $b = 4$ and $s \geq 6$ etc.

The denominators are $|B| = 24$ and $|S| = 26$

Example (cont'd)

Recall: $B = \{b \geq 3\},\ R = \{r \geq 3\},\ S = \{s \geq 6\}$

$P(B) = P(R) = 2/3 = 66.7\%$

$P(S) = \frac{5+6+5+4+3+2+1}{36} = \frac{26}{36} = 72.22\%$

$P(S|B \cup R) = \frac{2+3+4+5+6+6}{32} = \frac{26}{32} = 81.25\%$

The set $B \cup R$ represents the event 'b or r'.
It comprises all the rolls except for those with *both* the red and the black die coming up either 1 or 2.

$P(S|B \cap R) = 1 = 100\%$ — because $S \supseteq B \cap R$

Exercise

$\boxed{9.1.9}$ Consider three red and eight black marbles; draw two without replacement. We write b_1 — Black on the first draw, b_2 — Black on the second draw, r_1 — Red on first draw, r_2 — Red on second draw

Find the probabilities

(a) both Red:

$$P(r_1 \wedge r_2) = P(r_1)P(r_2|r_1) = \frac{3}{11} \cdot \frac{2}{10} = \frac{3}{55}$$

Equivalently:

$|\text{two-samples}| = \binom{11}{2} = 55$; $|\text{Red two-samples}| = \binom{3}{2} = 3$

$P(\cdot) = \frac{\binom{3}{2}}{\binom{11}{2}} = \frac{3}{55}$

(b) both Black:

$$P(b_1 \wedge b_2) = P(b_1)P(b_2|b_1) = \frac{8}{11} \cdot \frac{7}{10} = \frac{28}{55} = \frac{\binom{8}{2}}{\binom{11}{2}}$$

Exercise

$\boxed{9.1.9}$ Consider three red and eight black marbles; draw two without replacement. We write b_1 — Black on the first draw, b_2 — Black on the second draw, r_1 — Red on first draw, r_2 — Red on second draw

Find the probabilities

(a) both Red:

$$P(r_1 \wedge r_2) = P(r_1)P(r_2|r_1) = \frac{3}{11} \cdot \frac{2}{10} = \frac{3}{55}$$

Equivalently:

$|\text{two-samples}| = \binom{11}{2} = 55$; $|\text{Red two-samples}| = \binom{3}{2} = 3$

$P(\cdot) = \frac{\binom{3}{2}}{\binom{11}{2}} = \frac{3}{55}$

(b) both Black:

$$P(b_1 \wedge b_2) = P(b_1)P(b_2|b_1) = \frac{8}{11} \cdot \frac{7}{10} = \frac{28}{55} = \frac{\binom{8}{2}}{\binom{11}{2}}$$

(c) one Red, one Black:

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3 \cdot 8}{\binom{11}{2}} \quad \text{— why?}$$

By textbook (the 'hard way')

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3}{11} \cdot \frac{8}{10} + \frac{8}{11} \cdot \frac{3}{10}$$

or

$$P(\cdot) = 1 - P(r_1 \wedge r_2) - P(b_1 \wedge b_2) = \frac{55 - 3 - 28}{55}$$

(c) one Red, one Black:

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3 \cdot 8}{\binom{11}{2}} \quad \text{— why?}$$

By textbook (the 'hard way')

$$P(r_1 \wedge b_2) + P(b_1 \wedge r_2) = \frac{3}{11} \cdot \frac{8}{10} + \frac{8}{11} \cdot \frac{3}{10}$$

or

$$P(\cdot) = 1 - P(r_1 \wedge r_2) - P(b_1 \wedge b_2) = \frac{55 - 3 - 28}{55}$$

Exercise

$\boxed{9.1.12}$ What is the probability of a flush given that all five cards in a Poker hand are red?

Red cards = $\diamondsuit$'s + $\heartsuit$'s
flush = all cards of the same suit

$$P(\text{flush} \mid \text{all five cards are Red}) = \frac{2 \cdot \binom{13}{5}}{\binom{26}{5}} = \frac{9}{230} \approx 4\%$$

Exercise

$\boxed{9.1.12}$ What is the probability of a flush given that all five cards in a Poker hand are red?

Red cards $= \diamondsuit$'s $+ \heartsuit$'s
flush $=$ all cards of the same suit

$$P(\text{flush} \,|\, \text{all five cards are Red}) = \frac{2 \cdot \binom{13}{5}}{\binom{26}{5}} = \frac{9}{230} \approx 4\%$$

Exercise

$\boxed{9.1.22}$ Prove the following:
If $P(A|B) > P(A)$ ("positive correlation") then $P(B|A) > P(B)$

$P(A|B) > P(A)$

$\rightarrow P(A \cap B) > P(A)P(B)$

$\rightarrow \frac{P(A \cap B)}{P(A)} > P(B)$

$\rightarrow P(B|A) > P(B)$

Exercise

$\boxed{9.1.22}$ Prove the following:
If $P(A|B) > P(A)$ ("positive correlation") then $P(B|A) > P(B)$

$$P(A|B) > P(A)$$

$$\to P(A \cap B) > P(A)P(B)$$

$$\to \frac{P(A \cap B)}{P(A)} > P(B)$$

$$\to P(B|A) > P(B)$$

Stochastic Independence

Definition

A and B are **stochastically independent** (notation: $A \perp B$) if $P(A \cap B) = P(A) \cdot P(B)$

If $P(A) \neq 0$ and $P(B) \neq 0$, all of the following are *equivalent* definitions:

- $P(A \cap B) = P(A)P(B)$
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$
- $P(A^c|B) = P(A^c)$ or $P(A|B^c) = P(A)$ or $P(A^c|B^c) = P(A^c)$

The last one claims that

$$A \perp B \leftrightarrow A^c \perp B \leftrightarrow A \perp B^c \leftrightarrow A^c \perp B^c$$

Basic non-independent sets of events

- $A \subseteq B$
- $A \cap B = \emptyset$
- Any pair of one-point events $\{x\}, \{y\}$:
  either $x = y$ and $P(x|y) = 1$
  or $x \neq y$ and $P(x|y) = 0$

Independence of $A_1, \ldots, A_n$

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

for all possible collections $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$.
This is often called (for emphasis) a *full* independence

Basic non-independent sets of events

- $A \subseteq B$
- $A \cap B = \emptyset$
- Any pair of one-point events $\{x\}, \{y\}$:
  either $x = y$ and $P(x|y) = 1$
  or $x \neq y$ and $P(x|y) = 0$

Independence of $A_1, \ldots, A_n$

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k})$$

for all possible collections $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$.
This is often called (for emphasis) a *full* independence

Pairwise independence is a *weaker* concept.

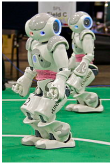
Example

Toss of two coins

$$\left.\begin{array}{l} A = \langle \text{first coin } H \rangle \\ B = \langle \text{second coin } H \rangle \\ C = \langle \text{exactly one } H \rangle \end{array}\right\} \begin{array}{l} P(A) = P(B) = P(C) = \frac{1}{2} \\ P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} \\ \text{However: } P(A \cap B \cap C) = 0 \end{array}$$

One can similarly construct a set of n events where any k of them are independent, while any $k+1$ are dependent (for $k < n$).

Independence of events, even just pairwise independence, can greatly simplify computations and reasoning in AI applications. It is common for many expert systems to make an approximating assumption of independence, even if it is not completely satisfied.



$$P(\text{sense}_t \mid \text{loc}_t, \text{sense}_{t-1}, \text{loc}_{t-1}, \ldots) = P(\text{sense}_t \mid \text{loc}_t)$$

Exercise

$\boxed{9.1.7}$ Suppose that an experiment leads to events A, B and C with $P(A) = 0.3$, $P(B) = 0.4$ and $P(A \cap B) = 0.1$

(a) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{4}$

(b) $P(A^c) = 1 - P(A) = 0.7$

(c) Is $A \perp B$? No. $P(A) \cdot P(B) = 0.12 \neq P(A \cap B)$

(d) Is $A^c \perp B$? No, as can be seen from (c).

Note: $P(A^c \cap B) = P(B) - P(A \cap B) = 0.4 - 0.1 = 0.3$
$P(A^c) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$

Exercise

$\boxed{9.1.7}$ Suppose that an experiment leads to events A, B and C with $P(A) = 0.3$, $P(B) = 0.4$ and $P(A \cap B) = 0.1$

(a) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{4}$

(b) $P(A^c) = 1 - P(A) = 0.7$

(c) Is $A \perp B$? No. $P(A) \cdot P(B) = 0.12 \neq P(A \cap B)$

(d) Is $A^c \perp B$? No, as can be seen from (c).

Note: $P(A^c \cap B) = P(B) - P(A \cap B) = 0.4 - 0.1 = 0.3$
$P(A^c) \cdot P(B) = 0.7 \cdot 0.4 = 0.28$

Exercise

$\boxed{9.1.8}$ Given $A \perp B, \; P(A) = 0.4, \; P(B) = 0.6$

$P(A|B) = P(A) = 0.4$

$P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.76$

$P(A^c \cap B) = P(A^c)P(B) = 0.36$

Exercise

$\boxed{9.1.8}$ Given $A \perp B, \;\; P(A) = 0.4, \;\; P(B) = 0.6$

$P(A|B) = P(A) = 0.4$

$P(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.76$

$P(A^c \cap B) = P(A^c)P(B) = 0.36$

Exercise

$\boxed{9.1.25}$ Does $A \perp B \perp C$ imply $(A \cap B) \perp (A \cap C)$?

No; this is almost never the case.
If somehow $(A \cap B) \perp (A \cap C)$ then it would give

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

As A is independent of B and of C it would suggest

$$P(A \cap B \cap C) \stackrel{?}{=} P(A) \cdot P(B) \cdot P(A) \cdot P(C)$$

instead of the correct

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

Exercise

$\boxed{9.1.25}$ Does $A \perp B \perp C$ imply $(A \cap B) \perp (A \cap C)$?

No; this is almost never the case.
If somehow $(A \cap B) \perp (A \cap C)$ then it would give

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

As A is independent of B and of C it would suggest

$$P(A \cap B \cap C) \stackrel{?}{=} P(A) \cdot P(B) \cdot P(A) \cdot P(C)$$

instead of the correct

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

Supplementary Exercise

$\boxed{9.5.5}$ (Supp) We are given two events with $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$. True, false or could be either?

(a) $P(A \cap B) = \frac{1}{12}$ — possible; it holds when $A \perp B$

(b) $P(A \cup B) = \frac{7}{12}$ — possible; it holds when A, B are disjoint

(c) $P(B|A) = \frac{P(B)}{P(A)}$ — false; correct is: $P(B|A) = \frac{P(B \cap A)}{P(A)}$

(d) $P(A|B) \geq P(A)$ — possible (it means that B "supports" A)

(e) $P(A^c) = \frac{3}{4}$ — true, since $P(A^c) = 1 - P(A)$

(f) $P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$ — true (also known as *total probability*)

Supplementary Exercise

$\boxed{9.5.5}$ (Supp) We are given two events with $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{3}$.
True, false or could be either?

(a) $P(A \cap B) = \frac{1}{12}$ — possible; it holds when $A \perp B$

(b) $P(A \cup B) = \frac{7}{12}$ — possible; it holds when A, B are disjoint

(c) $P(B|A) = \frac{P(B)}{P(A)}$ — false; correct is: $P(B|A) = \frac{P(B \cap A)}{P(A)}$

(d) $P(A|B) \geq P(A)$ — possible (it means that B "supports" A)

(e) $P(A^c) = \frac{3}{4}$ — true, since $P(A^c) = 1 - P(A)$

(f) $P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$ — true
(also known as *total probability*)

Expectation

Random Variables

Definition

An (integer) **random variable** is a function from Ω to $\mathbb{Z}$.
In other words, it associates a number value with every outcome.

Random variables are often denoted by $X, Y, Z, \ldots$

Example

Random variable $X_s \stackrel{\text{def}}{=}$ sum of rolling two dice

$\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$

$X_s((1,1)) = 2 \qquad X_s((1,2)) = 3 = X_s((2,1)) \quad \ldots$

$\boxed{9.3.3}$ Buy one lottery ticket for \$1. The only prize is \$1M.

$\Omega = \{win, lose\} \qquad X_L(win) = \$999,999 \qquad X_L(lose) = -\$1$

Expectation

Definition

The **expected value** (often called "expectation" or "average") of a random variable X is

$$E(X) = \sum_{k \in \mathbb{Z}} P(X = k) \cdot k$$

Example

The expected sum when rolling two dice is

$$E(X_s) = \frac{1}{36} \cdot 2 + \frac{2}{36} \cdot 3 + \ldots + \frac{6}{36} \cdot 7 + \ldots + \frac{1}{36} \cdot 12 = 7$$

$\boxed{9.3.3}$ Buy one lottery ticket for \$1. The only prize is \$1M. Each ticket has probability $6 \cdot 10^{-7}$ of winning.

$$E(X_L) = 6 \cdot 10^{-7} \cdot \$999,999 + (1 - 6 \cdot 10^{-7}) \cdot -\$1 = -\$0.4$$

NB

Expectation is a truly universal concept; it is the basis of all decision making, of estimating gains and losses, in all actions under risk. Historically, a rudimentary concept of expected value arose long before the notion of probability.

Theorem (linearity of expected value)

$E(X + Y) = E(X) + E(Y)$
$E(c \cdot X) = c \cdot E(X)$

Example

The expected sum when rolling two dice can be computed as

$$E(X_s) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$$

since $E(X_i) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \ldots + \frac{1}{6} \cdot 6$, for each die X_i

Example

$E(S_n)$, where $S_n \stackrel{\text{def}}{=} |\text{no. of HEADS in } n \text{ tosses}|$

- 'hard way'

$$E(S_n) = \sum_{k=0}^{n} P(S_n = k) \cdot k = \sum_{k=0}^{n} \frac{1}{2^n}\binom{n}{k} \cdot k$$

since there are $\binom{n}{k}$ sequences of n tosses with k HEADS, and each sequence has the probability $\frac{1}{2^n}$

$$= \frac{1}{2^n}\sum_{k=1}^{n} \frac{n}{k}\binom{n-1}{k-1} k = \frac{n}{2^n}\sum_{k=0}^{n-1}\binom{n-1}{k} = \frac{n}{2^n} \cdot 2^{n-1} = \frac{n}{2}$$

using the 'binomial identity' $\sum_{k=0}^{n}\binom{n}{k} = 2^n$

- 'easy way'

$$E(S_n) = E(S_1^1 + \ldots + S_1^n) = \sum_{i=1\ldots n} E(S_1^i) = nE(S_1) = n \cdot \frac{1}{2}$$

Note: $S_n \stackrel{\text{def}}{=} |\text{HEADS in } n \text{ tosses}|$ while each $S_1^i \stackrel{\text{def}}{=} |\text{HEADS in 1 toss}|$

Example

$E(S_n)$, where $S_n \stackrel{\text{def}}{=} |\text{no. of HEADS in } n \text{ tosses}|$

- 'hard way'

$$E(S_n) = \sum_{k=0}^{n} P(S_n = k) \cdot k = \sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} \cdot k$$

since there are $\binom{n}{k}$ sequences of n tosses with k HEADS, and each sequence has the probability $\frac{1}{2^n}$

$$= \frac{1}{2^n} \sum_{k=1}^{n} \frac{n}{k} \binom{n-1}{k-1} k = \frac{n}{2^n} \sum_{k=0}^{n-1} \binom{n-1}{k} = \frac{n}{2^n} \cdot 2^{n-1} = \frac{n}{2}$$

using the 'binomial identity' $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

- 'easy way'

$$E(S_n) = E(S_1^1 + \ldots + S_1^n) = \sum_{i=1\ldots n} E(S_1^i) = nE(S_1) = n \cdot \frac{1}{2}$$

Note: $S_n \stackrel{\text{def}}{=} |\text{HEADS in } n \text{ tosses}|$ while each $S_1^i \stackrel{\text{def}}{=} |\text{HEADS in 1 toss}|$

Example

$E(S_n)$, where $S_n \stackrel{\text{def}}{=} |\text{no. of HEADS in } n \text{ tosses}|$

- 'hard way'

$$E(S_n) = \sum_{k=0}^{n} P(S_n = k) \cdot k = \sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} \cdot k$$

since there are $\binom{n}{k}$ sequences of n tosses with k HEADS, and each sequence has the probability $\frac{1}{2^n}$

$$= \frac{1}{2^n} \sum_{k=1}^{n} \frac{n}{k} \binom{n-1}{k-1} k = \frac{n}{2^n} \sum_{k=0}^{n-1} \binom{n-1}{k} = \frac{n}{2^n} \cdot 2^{n-1} = \frac{n}{2}$$

using the 'binomial identity' $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

- 'easy way'

$$E(S_n) = E(S_1^1 + \ldots + S_1^n) = \sum_{i=1\ldots n} E(S_1^i) = nE(S_1) = n \cdot \frac{1}{2}$$

Note: $S_n \stackrel{\text{def}}{=} |\text{HEADS in } n \text{ tosses}|$ while each $S_1^i \stackrel{\text{def}}{=} |\text{HEADS in 1 toss}|$

NB

If $X_1, X_2, \ldots, X_n$ *are independent, identically distributed random variables, then* $E(X_1 + X_2 + \ldots + X_n)$ *happens to be the same as* $E(nX_1)$*, but these are very different random variables.*

Example

You face a quiz consisting of six true/false questions, and your plan is to guess the answer to each question (randomly, with probability 0.5 of being right). There are no negative marks, and answering four or more questions correctly suffices to pass. What is the probability of passing and what is the expected score?

To pass you would need four, five or six correct guesses. Therefore,

$$p(\text{pass}) = \frac{\binom{6}{4} + \binom{6}{5} + \binom{6}{6}}{64} = \frac{15+6+1}{64} \approx 34\%$$

The expected score from a single question is 0.5, as there is no penalty for errors. For six questions the expected value is $6 \cdot 0.5 = 3$

Example

You face a quiz consisting of six true/false questions, and your plan is to guess the answer to each question (randomly, with probability 0.5 of being right). There are no negative marks, and answering four or more questions correctly suffices to pass.
What is the probability of passing and what is the expected score?

To pass you would need four, five or six correct guesses. Therefore,

$$p(\text{pass}) = \frac{\binom{6}{4} + \binom{6}{5} + \binom{6}{6}}{64} = \frac{15 + 6 + 1}{64} \approx 34\%$$

The expected score from a single question is 0.5, as there is no penalty for errors. For six questions the expected value is $6 \cdot 0.5 = 3$

Exercise

$\boxed{9.3.7}$

An urn has $m + n = 10$ marbles, $m \geq 0$ red and $n \geq 0$ blue.
7 marbles selected at random without replacement.
What is the expected number of red marbles drawn?

$$\frac{\binom{m}{0}\binom{n}{7}}{\binom{10}{7}} \cdot 0 + \frac{\binom{m}{1}\binom{n}{6}}{\binom{10}{7}} \cdot 1 + \frac{\binom{m}{2}\binom{n}{5}}{\binom{10}{7}} \cdot 2 + \ldots + \frac{\binom{m}{7}\binom{n}{0}}{\binom{10}{7}} \cdot 7$$

e.g.

$$\frac{\binom{5}{2}\binom{5}{5}}{\binom{10}{7}} \cdot 2 + \frac{\binom{5}{3}\binom{5}{4}}{\binom{10}{7}} \cdot 3 + \frac{\binom{5}{4}\binom{5}{3}}{\binom{10}{7}} \cdot 4 + \frac{\binom{5}{5}\binom{5}{2}}{\binom{10}{7}} \cdot 5$$

$$= \frac{10}{120} \cdot 2 + \frac{50}{120} \cdot 3 + \frac{50}{120} \cdot 4 + \frac{10}{120} \cdot 5 = \frac{420}{120} = 3.5$$

Exercise

$\boxed{9.3.7}$

An urn has $m + n = 10$ marbles, $m \geq 0$ red and $n \geq 0$ blue.
7 marbles selected at random without replacement.
What is the expected number of red marbles drawn?

$$\frac{\binom{m}{0}\binom{n}{7}}{\binom{10}{7}} \cdot 0 + \frac{\binom{m}{1}\binom{n}{6}}{\binom{10}{7}} \cdot 1 + \frac{\binom{m}{2}\binom{n}{5}}{\binom{10}{7}} \cdot 2 + \ldots + \frac{\binom{m}{7}\binom{n}{0}}{\binom{10}{7}} \cdot 7$$

e.g.

$$\frac{\binom{5}{2}\binom{5}{5}}{\binom{10}{7}} \cdot 2 + \frac{\binom{5}{3}\binom{5}{4}}{\binom{10}{7}} \cdot 3 + \frac{\binom{5}{4}\binom{5}{3}}{\binom{10}{7}} \cdot 4 + \frac{\binom{5}{5}\binom{5}{2}}{\binom{10}{7}} \cdot 5$$

$$= \frac{10}{120} \cdot 2 + \frac{50}{120} \cdot 3 + \frac{50}{120} \cdot 4 + \frac{10}{120} \cdot 5 = \frac{420}{120} = 3.5$$

Example

Find the average waiting time for the first HEAD, with no upper bound on the 'duration' (one allows for all possible sequences of tosses, regardless of how many times TAILS occur initially).

$$A = E(X_w) = \sum_{k=1}^{\infty} k \cdot P(X_w = k) = \sum_{k=1}^{\infty} k \frac{1}{2^k}$$
$$= \frac{1}{2^1} + \frac{2}{2^2} + \frac{3}{2^3} + \dots$$

This can be evaluated by breaking the sum into a sequence of geometric progressions

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \dots$$

$$= \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots\right) + \left(\frac{1}{2^2} + \frac{1}{2^3} + \dots\right) + \left(\frac{1}{2^3} + \dots\right) + \dots$$

$$= 1 + \frac{1}{2} + \frac{1}{2^2} + \dots = 2$$

Example

Find the average waiting time for the first HEAD, with no upper bound on the 'duration' (one allows for all possible sequences of tosses, regardless of how many times TAILS occur initially).

$$A = E(X_w) = \sum_{k=1}^{\infty} k \cdot P(X_w = k) = \sum_{k=1}^{\infty} k \frac{1}{2^k}$$
$$= \frac{1}{2^1} + \frac{2}{2^2} + \frac{3}{2^3} + \dots$$

This can be evaluated by breaking the sum into a sequence of geometric progressions

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \dots$$

$$= \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots\right) + \left(\frac{1}{2^2} + \frac{1}{2^3} + \dots\right) + \left(\frac{1}{2^3} + \dots\right) + \dots$$

$$= 1 + \frac{1}{2} + \frac{1}{2^2} + \dots = 2$$

There is also a recursive 'trick' for solving the sum

$$A = \sum_{k=1}^{\infty} \frac{k}{2^k} = \sum_{k=1}^{\infty} \frac{k-1}{2^k} + \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{2} \sum_{k=1}^{\infty} \frac{k-1}{2^{k-1}} + 1 = \frac{1}{2} A + 1$$

Now $\quad A = \frac{A}{2} + 1$ and $A = 2$

NB

A much simpler but equally valid argument is that you expect 'half' a HEAD *in 1 toss, so you ought to get a 'whole'* HEAD *in 2 tosses.*

Theorem

The average number of trials needed to see an event with probability p is $\frac{1}{p}$.

Exercise

$\boxed{9.4.12}$ A die is rolled until the first 4 appears. What is the expected waiting time?

$P(\text{roll } 4) = \frac{1}{6}$ hence $E(\text{no. of rolls until first } 4) = 6$

Exercise

$\boxed{9.4.12}$ A die is rolled until the first 4 appears. What is the expected waiting time?

$P(\text{roll } 4) = \frac{1}{6}$ hence $E(\text{no. of rolls until first } 4) = 6$

Example

To find an object $\mathcal{X}$ in an unsorted list L of elements, one needs to search linearly through L. Let the probability of $\mathcal{X} \in L$ be p, hence there is $1-p$ likelihood of $\mathcal{X}$ being absent altogether. Find the expected number of comparison operations.

If the element is in the list, then the number of comparisons averages to $\frac{1}{n}(1+\ldots+n)$; if absent we need n comparisons. The first case has probability p, the second $1-p$. Combining these we find

$$E_n = p\frac{1+\ldots+n}{n} + (1-p)n = p\frac{n+1}{2} + (1-p)n = (1-\frac{p}{2})n + \frac{p}{2}$$

As one would expect, increasing p leads to a lower E_n.

Example

To find an object $\mathcal{X}$ in an unsorted list L of elements, one needs to search linearly through L. Let the probability of $\mathcal{X} \in L$ be p, hence there is $1 - p$ likelihood of $\mathcal{X}$ being absent altogether. Find the expected number of comparison operations.

If the element is in the list, then the number of comparisons averages to $\frac{1}{n}(1 + \ldots + n)$; if absent we need n comparisons. The first case has probability p, the second $1 - p$. Combining these we find

$$E_n = p\frac{1 + \ldots + n}{n} + (1-p)n = p\frac{n+1}{2} + (1-p)n = (1 - \frac{p}{2})n + \frac{p}{2}$$

As one would expect, increasing p leads to a lower E_n.

One may expect that this would indicate a practical rule — that high probability of success might lead to a high expected value. Unfortunately this is *not* the case in a great many practical situations.
Many lottery advertisements claim that buying more tickets leads to better expected results — and indeed, obviously you will have more potentially winning tickets. However, the expected value *decreases* when the number of tickets is increased.

As an example, let us consider a punter placing bets on a roulette (outcomes: $0, 1 \ldots 36$). Tired of losing, he decides to place \$1 on 24 'ordinary' numbers $a_1 < a_2 < \ldots < a_{24}$, selected from among 1 to 36.

His probability of winning is high indeed — $\frac{24}{37} \approx 65\%$; he scores on any of his choices, and loses only on the remaining thirteen numbers.

But what about his performance?

- If one of his numbers comes up, say a_i, he wins \$35 from the bet on that number and loses \$23 from the bets on the remaining numbers, thus collecting \$12.
  This happens with probability $p = \frac{24}{37}$.
- With probability $q = \frac{13}{37}$ none of his numbers appears, leading to loss of \$24.

The expected result

$$p \cdot \$12 - q \cdot \$24 = \$12\frac{24}{37} - \$24\frac{13}{37} = -\$\frac{24}{37} \approx -65¢$$

Many so-called 'winning systems' that purport to offer a winning strategy do something akin — they provide a scheme for frequent relatively moderate wins, but at the cost of an occasional very big loss.

It turns out (it is a formal theorem) that there can be *no system* that converts an 'unfair' game into a 'fair' one. In the language of decision theory, 'unfair' denotes a game whose individual bets have negative expectation.

It can be easily checked that any individual bets on roulette, on lottery tickets or on just about any commercially offered game have negative expected value.

Standard Deviation and Variance

Definition

For random variable X with expected value (or: **mean**) $\mu = E(X)$, the **standard deviation** of X is

$$\sigma = \sqrt{E((X - \mu)^2)}$$

and the **variance** of X is

$$\sigma^2$$

Standard deviation and variance measure how spread out the values of a random variable are. The smaller σ^2 the more confident we can be that $X(\omega)$ is close to $E(X)$, for a randomly selected ω.

NB

The variance can be calculated as $E((X - \mu)^2) = E(X^2) - \mu^2$

Example

Random variable $X_d \stackrel{\text{def}}{=}$ value of a rolled die

$$\mu = E(X_d) = 3.5$$

$$E(X_d^2) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 9 + \frac{1}{6} \cdot 16 + \frac{1}{6} \cdot 25 + \frac{1}{6} \cdot 36 = \frac{91}{6}$$

$$\text{Hence,} \quad \sigma^2 = E(X_d^2) - \mu^2 = \frac{35}{12} \quad \rightarrow \quad \sigma \approx 1.71$$

Exercise

$\boxed{9.5.10}$ (Supp) Two independent experiments are performed.
$P(\text{1st experiment succeeds}) = 0.7$
$P(\text{2nd experiment succeeds}) = 0.2$
Random variable X counts the number of successful experiments.

(a) Expected value of X? $E(X) = 0.7 + 0.2 = 0.9$

(b) Probability of exactly one success? $0.7 \cdot 0.8 + 0.3 \cdot 0.2 = 0.62$

(c) Probability of at most one success? (b)$+0.3 \cdot 0.8 = 0.86$

(e) Variance of X? $\sigma^2 = (0.62 \cdot 1 + 0.14 \cdot 4) - 0.9^2 = 0.37$

Exercise

$\boxed{9.5.10}$ (Supp) Two independent experiments are performed.
$P(\text{1st experiment succeeds}) = 0.7$
$P(\text{2nd experiment succeeds}) = 0.2$
Random variable X counts the number of successful experiments.

(a) Expected value of X? $E(X) = 0.7 + 0.2 = 0.9$

(b) Probability of exactly one success? $0.7 \cdot 0.8 + 0.3 \cdot 0.2 = 0.62$

(c) Probability of at most one success? (b)$+0.3 \cdot 0.8 = 0.86$

(e) Variance of X? $\sigma^2 = (0.62 \cdot 1 + 0.14 \cdot 4) - 0.9^2 = 0.37$

Cumulative Distribution Functions

Definition

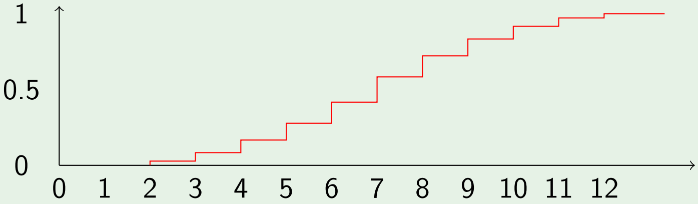
The **cumulative distribution function** $\mathrm{CDF}_X : \mathbb{Z} \longrightarrow \mathbb{R}$ of an integer random variable X is defined as

$$\mathrm{CDF}_X(y) \mapsto \sum_{k \leq y} P(X = k)$$

$\mathrm{CDF}_X(y)$ collects the probabilities $P(X)$ for all values up to y

Example

Cumulative distribution function for sum of 2 dice



Example: Binomial Distributions

Definition

Binomial random variables count the number of 'successes' in n independent experiments with probability p for each experiment.

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

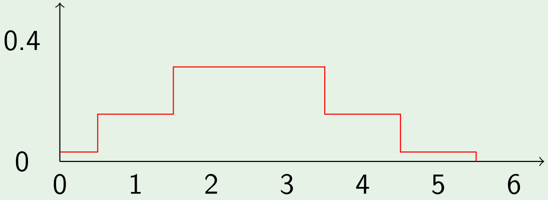
$$\mathrm{CDF}_B(y) \mapsto \sum_{k \le y} \binom{n}{k} p^k (1-p)^{n-k}$$

Theorem

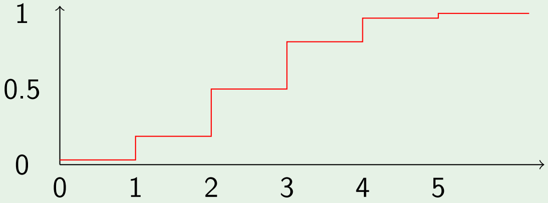
If X is a binomially distributed random variable based on n and p, then $E(X) = n \cdot p$ with variance $\sigma^2 = n \cdot p \cdot (1-p)$

Example (binomial distribution)

No. of HEADS in 5 coin tosses



CDF for no. of HEADS in 5 coin tosses



Exercise

$\boxed{9.4.10}$ An experiment is repeated 30,000 times with probability of success $\frac{1}{4}$ each time.

(a) Expected number of successes? $E(X) = 30,000 \cdot \frac{1}{4} = 7500$

(b) Standard deviation? $\sigma = \sqrt{30,000 \cdot \frac{1}{4} \cdot \frac{3}{4}} = 75$

Exercise

$\boxed{9.4.10}$ An experiment is repeated 30,000 times with probability of success $\frac{1}{4}$ each time.

(a) Expected number of successes? $E(X) = 30,000 \cdot \frac{1}{4} = 7500$

(b) Standard deviation? $\sigma = \sqrt{30,000 \cdot \frac{1}{4} \cdot \frac{3}{4}} = 75$

Normal Distribution

Fact

For large n, binomial distributions can be approximated by **normal distributions** *(a.k.a.* **Gaussian distributions***) with mean $\mu = n \cdot p$ and variance $\sigma^2 = n \cdot p \cdot (1 - p)$*

$$\frac{1}{\sqrt{2\sigma^2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Summary

- counting
  - union rule, product rule, $n!$, $\Pi(n, r)$, $\binom{n}{r}$
- events and their probability
- counting, inclusion-exclusion, recursion for probabilities
- conditional probability $P(A|B)$, independence $A \perp B$
- random variables X, expected value $E(X)$ $(= \text{mean } \mu)$
- CDF, standard deviation σ, variance σ^2