COMP9020 19T1
Week 9
Conditional Probability, Expectation

NB

1. Additional lecture (revision) on Wednesday week 10 (24 April)
2. Last tutorial on Wed, 24 April, 1–2pm in **Old Main 230**

- Textbook (R & W) - Ch. 9, Sec. 9.1–9.4
- Problem set week 9
- Supplementary Exercises Ch. 9 (R & W)
- Quiz week 9 (due Tuesday week 10)

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 $P(A \cap B) = 0$.

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) $\frac{8}{11}$

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

$\boxed{9.1.12}$ 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

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

Some General Rules

Fact

- $A \subseteq B \Rightarrow P(A|B) \geq P(A)$
- $A \subseteq B \Rightarrow P(B|A) = 1$
- $P(A \cap B|B) = P(A|B)$
- $P(\emptyset|A) = 0$ *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 $|S| = 26$

NB

Bayes' Formula: $P(S|B) \cdot P(B) = P(B|S) \cdot P(S)$

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
Using conditional probabilities, 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:

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

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
Using conditional probabilities, 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:

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

Exercise

(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

(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

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

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

$\Rightarrow P(A \cap B) > P(A) \cdot P(B)$

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

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

Exercise

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

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

$\Rightarrow P(A \cap B) > P(A) \cdot P(B)$

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

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

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 (if $P(A), P(B) > 0$)

- $A \subseteq B$
- $A \cap B = \emptyset$
- 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 (if $P(A), P(B) > 0$)

- $A \subseteq B$
- $A \cap B = \emptyset$
- 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 $k < n$).

NB

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.

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

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

$\boxed{9.5.5 \text{ (supp)}}$ 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) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ — true

NB

Total probability: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Exercise

$\boxed{\text{9.5.5 (supp)}}$ 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) $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ — true

NB

Total probability: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

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

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

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

Example

$\boxed{9.3.3}$ 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

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

Example

$\boxed{9.3.3}$ 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).

$$\begin{aligned} A = E(X_w) &= \textstyle\sum_{k=1}^{\infty} k \cdot P(X_w = k) = \sum_{k=1}^{\infty} k \frac{1}{2^k} \\ &= \tfrac{1}{2^1} + \tfrac{2}{2^2} + \tfrac{3}{2^3} + \ldots \end{aligned}$$

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

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

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

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

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).

$$\begin{aligned} A = E(X_w) &= \textstyle\sum_{k=1}^{\infty} k \cdot P(X_w = k) = \sum_{k=1}^{\infty} k \frac{1}{2^k} \\ &= \tfrac{1}{2^1} + \tfrac{2}{2^2} + \tfrac{3}{2^3} + \ldots \end{aligned}$$

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

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

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

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

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 expected number 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 expected number 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

$\boxed{\text{9.5.10 (supp)}}$ Two independent experiments are performed.
P(1st experiment succeeds) $= 0.7$
P(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

$\boxed{\text{9.5.10 (supp)}}$ Two independent experiments are performed.
P(1st experiment succeeds) $= 0.7$
P(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

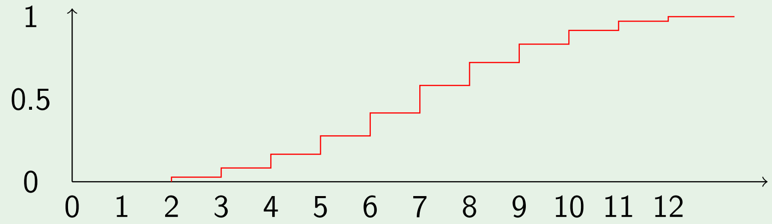
The **cumulative distribution function** $\text{CDF}_X : \mathbb{Z} \longrightarrow \mathbb{R}$ of an integer random variable X is defined as

$$\text{CDF}_X(y) \mapsto \sum_{k \le y} P(X = k)$$

$\text{CDF}_X(y)$ 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.

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

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

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 Distributions

Definition

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

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

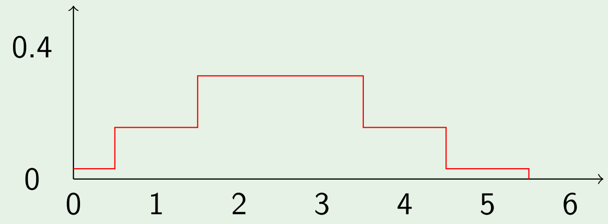
$$\text{CDF}_B(y) \mapsto \sum_{k \le y} \binom{n}{k} p^k (1 - p)^{n-k}$$

Theorem

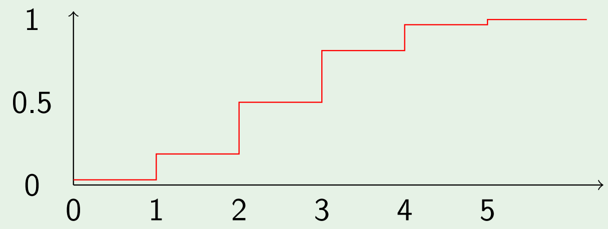
If X is a binomially distributed random variable based on n and p, then $E(X) = n \cdot p$ with variance $\sigma^2 = n \cdot p \cdot (1 - p)$

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

- Conditional probability $P(A|B)$, independence $A \perp B$
- Bayes' formula, total probability
- Random variables X
- Expected value $E(X)$
- Mean μ, CDF, standard deviation σ, variance σ^2