Week 2: Conditional probabilities, Bayes rules, Random variables

Yilin

Table of contents

Permutations

- A **permutation** of a set is an *ordered arrangement* of its elements.
- An **r-permutation** of S an *ordered selection* of r elements from S (with no repetitions allowed).
  - These are r-tuples $(a_1, \ldots, a_r)$ such that a_i's are pairwise distinct and $a_i \in S$ for all i.
  - If $|S| = n$, then an n-permutation is a permutation.
- The number of r-permutations of a set of size n is

$$P(n, r) = \frac{n!}{(n-r)!}.$$

The number of permutations of a set of size n is $P(n, n) = n!$

Combinations

order doesn't matter.

- An **r-combination** of a set S is a subset of size r of S.
- The number of r-combinations of a set of size n is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

$P(1000, 200)$
$C(1000, 200)$

Conditional probability

Events A and B with $P(B) > 0$. The **conditional probability of $\mathbf{A}$ given $\mathbf{B}$**, denoted $P(A|B)$, is

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

Independent events

- Two events A and B are **independent** if and only if

$$\boxed{P(A \cap B) = P(A)P(B).} \tag{1}$$

- If $P(A) > 0$ and $P(B) > 0$, (1) is equivalent to either

$$\begin{cases} P(A|B) = P(A) \text{ or} & (2) \\ P(B|A) = P(B). & (3) \end{cases}$$

- To prove the independence of A and B, we only need to prove one of the equations (1) or (2) or (3).

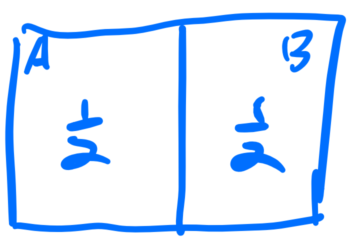
Explanation of Independent Events

- The independence of A and B means "the information that B occurs does not affect the probability that A occurs, and vice versa".
- **Remark**: Do not use any other definitions of independence such as "A and B have no influence on each other" or "A and B are disjoint". They are simply **incorrect**.

Question 1

Let A and B be disjoint events. Are A and B independent? If the answer is not, find a counterexample.

Solution.



$P(A) = P(B) = \frac{1}{2}$

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

$P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(A \cap B) \neq P(A)P(B)$.

A and B are dependent.

Example 1

A fair dice is rolled two times.

E_1: the 1st role gives 1.

E_2: the 2nd role gives 1.

Are E_1 and E_2 independent events?

Solution.

$P(E_1) = \frac{1}{6}$.

$P(E_2) = \frac{1}{6}$.

$P(E_1 \cap E_2) = \frac{1}{36}$.

$P(E_1 \cap E_2) = P(E_1)P(E_2) = \frac{1}{36}$

E_1 & E_2 are independent events.

Example 2

A number is chosen at random from $S = \{1, 2, \ldots, 9\}$.

A: the number is a prime. $A = \{2, 3, 5, 7\}$

B: the number is smaller than 5. $B = \{1, 2, 3, 4\}$

Are A and B independent? $A \cap B = \{2, 3\}$.

Solution. $P(A) = \frac{4}{9}$.

$P(B) = \frac{4}{9}$

$P(A \cap B) = \frac{2}{9}$

$P(A \cap B) \neq P(A)P(B)$

A and B are not independent.

Multiplication rule for conditional probability

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

$$P(A)P(B|A) = P(A\cap B)$$

Lemma 1. The following are multiplication rules for conditional probability.

a. $P(A \cap B) = P(A)P(B|A)$.

b. $P(A \cap B \cap C) = \underbrace{P(A)P(B|A)}_{P(A\cap B)}P(C|A \cap B)$.

Example 3 (Tutorial 1 Question 4)

You have a flight from Amsterdam to Sydney with a stopover in Dubai. The probabilities that a luggage is put on the wrong plane at the different airports are

0.95 0.97

$$\text{Amsterdam} : 0.05, \;\; \text{Dubai} : 0.03.$$

What is the probability that your luggage does not reach Sydney with you?

Example 3 solution

$$\underbrace{1 - \underbrace{0.95 \times 0.97}_{\text{have luggage in Sydney}}}_{\text{don't have luggage in Sydney}}$$

Exercise 2

You have a flight from Amsterdam to Sydney with stopovers in Dubai and Singapore. The probabilities that a luggage is put on the wrong plane at the different airports are

$$\text{Amsterdam} : 0.05, \;\; \text{Dubai} : 0.03, \;\; \text{Singapore} : 0.01.$$

0.95 0.97 0.99

What is the probability that your luggage does not reach Sydney with you?

Exercise 2 solution

$$1-(1-0.05)(1-0.03)(1-0.01)$$

$$=1-(0.95)(0.97)(0.99)$$

$$=1-0.9123\approx 8.77\%.$$

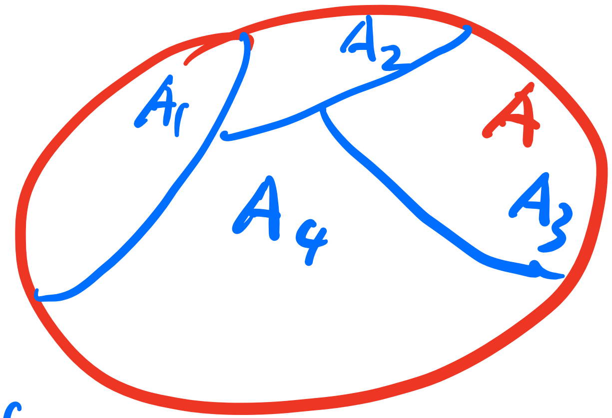
Partition of a set

Let A be a set. The sets $A_1, \ldots, A_n$ is called a **partition** of A if

(i) $A_i \subset A$ for any i,

(ii) A_i's are pairwise disjoint, and

(iii) $\cup_{i=1}^n A_i = A$.



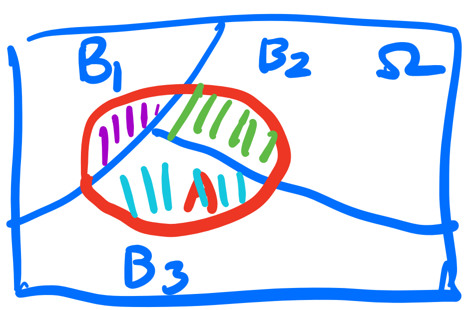
For examples:

- A and A^c is a partition of Ω. $A \cup A^c = \Omega$
- $\{1\}, \{2\}, \{3\}$ is a partition of $\{1, 2, 3\}$.

Law of total probability

Theorem 1. Let P be a probability measure on Ω. Assume that $B_1, \ldots, B_n$ is a partition of Ω. Then for any event A, we have

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{4}$$



$(A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$

$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$

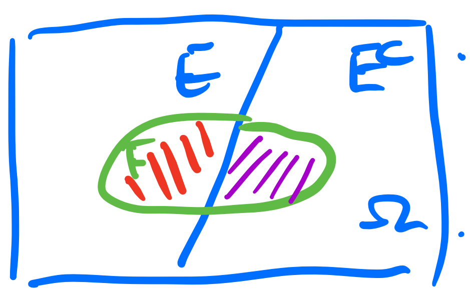
$+ P(A|B_3)P(B_3)$

Corollary of Theorem 1

Corollary 1. Let E and F be events in the sample space Ω. Then

$$P(F) = P(F|E)P(E) + P(F|E^c)P(E^c).$$

Proof.



Example 4

$P(A)$

One in $100,000$ people has a rare disease for which there is a fairly accurate diagnostic test. This test is correct 99.0% of the time when given to a person selected at random who has the disease, and correct 99.5% of the time when given to a person selected at random who does not have the disease. Find

$P(B|A) = 0.99$

$P(B^c|A^c) = 0.995$

(a) The probability that a person who tests positive actually has the disease? $P(A|B)$

information

(b) The probability that a person who tests negative does not have the disease? $P(A^c|B^c)$

information

Example 4 Solution

A=event that a randomly selected person has the disease.

B=event that a randomly selected person tests positive.

Need to compute $P(A|B)$ and $P(A^c|B^c)$.

A : has the disease A^c : doesn't have disease.

B : test positive B^c : test negative.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{\frac{1}{100{,}000} \times 0.99}{0.005} \approx 0.00198$$

$$P(B) = P(A)P(B|A) + \underbrace{P(A^c)}_{1-P(A)}\,\underbrace{P(B|A^c)}_{1-P(B^c|A^c)} = 0.005$$

Example 4 Continued

$$P(A^c|B^c) = \frac{P(B^c|A^c)\,P(A^c)}{P(B^c)}$$

$$P(B^c) = 1 - P(B).$$

$$= \frac{(0.995)\left(1-\frac{1}{100{,}000}\right)}{1-0.005} \approx 0.99.$$

Bayes' Rule (Simplified Version)

Theorem 2. Events E, F with $P(E) > 0$, $P(F) > 0$. Then

$$P(F|E) = \frac{\underline{P(E|F)}\,\underline{P(F)}}{P(E|F)P(F) + P(E|F^c)P(F^c)}. \leftarrow P(E)$$

Proof. Idea: Express $P(E \cap F)$ in two different ways

$$\underline{P(F|E)}\,\underline{P(E)} = P(E|F)P(F)$$

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

Bayes' Rule General

Theorem 3. $F_1, F_2, \ldots, F_n$ is a partition of Ω and E is an event. Assume $P(E) > 0$ and $P(F_i) > 0$ for $i = 1, \ldots, n$. Then for any $k \in \{1, \ldots, n\}$, we have

$$P(F_k|E) = \frac{P(E|F_k)P(F_k)}{\underbrace{\sum_{i=1}^{n} P(E|F_i)P(F_i)}_{P(E)}}.$$

Interpretation of Bayes' Rule

$$P(F_k|E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}.$$

- F_i's are possible cases for the occurrence of E.
- The Bayes' formula computes the probability that F_k caused E, given that E occurred.

Example 5

A factory uses 3 machines M_1, M_2, M_3 to produce certain items.

- M_1 produces 50% of the items, of which 3% are defective.
- M_2 produces 30% of the items, of which 4% are defective.
- M_3 produces 20% of the items, of which 5% are defective.

Suppose that a defective item is found. What is the probability that it came from M_2?

Example 5 Solution

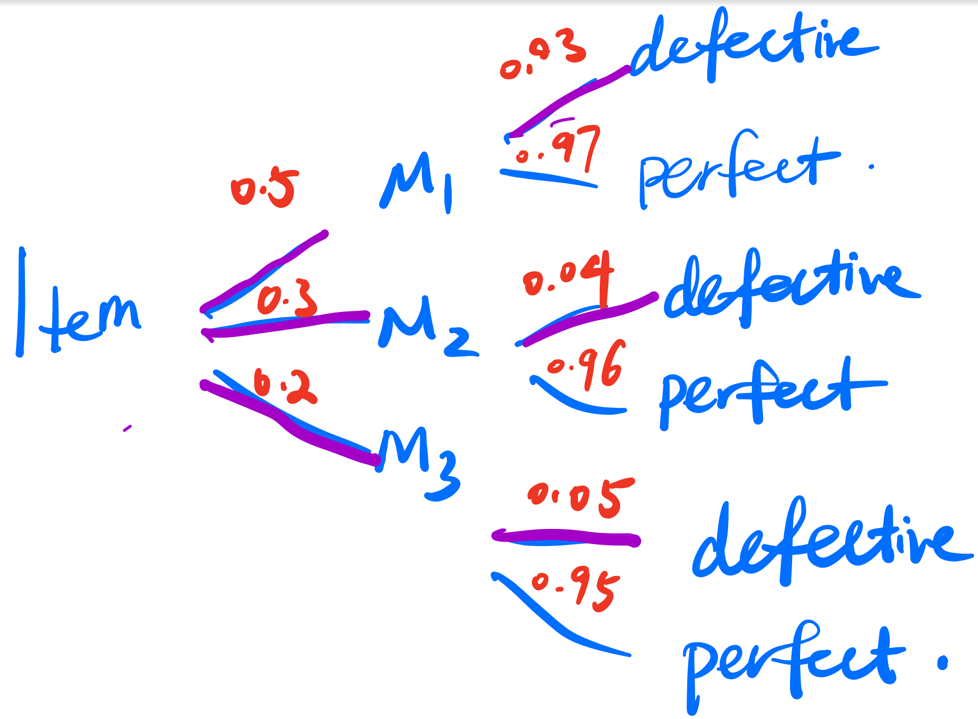
B_1, B_2, B_3 are events that a given item comes from M_1, M_2, M_3.
A is the event that a given item is defective. Compute $P(B_2|A)$.

$$P(B_2|A) = \frac{P(A|B_2)\,P(B_2)}{P(A)} = \frac{(0.3)(0.04)}{0.037} = \boxed{32\%}$$

$$P(A) = P(B_1)P(A|B_1) + \underbrace{P(B_2)P(A|B_2)}_{} + P(B_3)P(A|B_3)$$

$$= (0.5)(0.03) + (0.3)(0.04) + (0.2)(0.05) = 0.037.$$

Example 5 Continued



Example 6

- A coin is thrown three times

$$\Omega = \{\text{hhh, hht, htt, hth, ttt, tth, thh, tht}\}$$

- Usually we are not interested in the whole Ω (too complex), but only extract information of interests, for examples,
  - the total number of heads, or
  - the total number of tails, or
  - the number of heads minus the number of tails.
- Each of these quantities is a random variable.

Random Variables

- A **Random Variable** on the sample space Ω is a function

$$X : \Omega \to \mathbb{R},$$

that is, X assigns a real number to each possible outcome.

- The set of possible values of X is

$$X(\Omega) = \{X(w) : w \in \Omega\}.$$

- Capital letters $X, Y, Z, \ldots$ denote random variables. Small letters $x, y, z, \ldots$ denote possible values of random variables.

Example 7

- X= # heads in 3 coin tosses.

$$\Omega = \{\text{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}\}.$$

3 2 2 2 1 1 1 0

- X is a function $X : \Omega \to \mathbb{R}$ defined as follows

$$X(\text{HHH}) = 3,\ X(\text{HHT}) = X(\text{HTH}) = X(\text{THH}) = 2,$$

$$X(\text{HTT}) = X(\text{THT}) = X(\text{TTH}) = 1,\ X(\text{TTT}) = 0.$$

- The set of possible values of X is $\{0, 1, 2, 3\}$.

Exercise 3

Experiment: Roll a dice 5 times. Write out the sample space Ω and a few examples of random variables on Ω.

$$\Omega = \{(a_1, a_2, a_3, a_4, a_5),\ a_i \in \{1, 2, 3, 4, 5, 6\}\}.$$

$$6^5$$

Exercise 3 solution

Example : ① # of 6. $X = \{0, 1, 2, 3, 4, 5\}$

② Sum of all 5 numbers. "range"

$X = \{5, \cdots\cdots 30\}$

Probability measure - Recap

Probability measure P on Ω:

(i) $P(\Omega) = 1$

(ii) $P(A) \geq 0$ for any $A \subset \Omega$

(iii) $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for pairwise disjoint events $A_1, A_2, \ldots, A_n, \ldots$.

Common Probabilities of Interest

$S \subset \mathbb{R}$ and $x \in \mathbb{R}$. Several common probabilities of interest are

1. $P(X = x) = P\left(\{w \in \Omega : X(w) = x\}\right)$.
2. $P(X \in S) = P\left(\{w \in \Omega : X(w) \in S\}\right)$.
3. The **Cumulative Distribution Function (CDF)** F of X

$$F(x) = P(X \leq x).$$

Example 8

$X = \#$ number of heads in 3 consecutive fair-coin tosses. Find $P(X = 3), P(X \le 1)$ and $P(X \neq 2)$.

Solution.

$P(X=3) = \frac{1}{8}$

$P(X=2) = \frac{3}{8}$

$P(X=1) = \frac{3}{8}$

$P(X=0) = \frac{1}{8}$

$P(X \le 1) = P(X=0) + P(X=1)$

$= \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$

$P(X \neq 2) = 1 - P(X=2) = 1 - \frac{3}{8} = \frac{5}{8}$

Example 9

$P(H) = p$; $P(T) = 1-p$

$p \in [0,1]$. A calibrated coin has chance of landing head is p.

X = number of coin tosses until a head comes up. Given $n \in \mathbb{Z}^+$.

positive integers.

Find $P(X = n)$ and $P(X \le n)$.

geometric sequence with $r = 1-p$

Solution.

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

$$P(X \le n) = p + (1-p)p + (1-p)^2 p + \cdots + (1-p)^{n-1} p$$

$$= \frac{p(1-(1-p)^n)}{1-(1-p)} = 1-(1-p)^n$$

Countable Sets

$S \subset \mathbb{R}$ is **Countable** if there is an order to list all elements of S.

- Any finite subset $S = \{a_1, a_2, \ldots, a_n\}$ of $\mathbb{R}$ is countable.
- $S = \mathbb{N}$ is countable: $S = \{0, 1, 2, 3, \ldots\}$.
- $S = \mathbb{Q}^+$, the set of positive rational numbers, is countable. Its elements can be listed as $\frac{a}{b}$ with $a + b \in \{2, 3, 4, \ldots\}$:

$$\frac{1}{1}, \ \frac{1}{2}, \frac{2}{1}, \ \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \ \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \ldots$$

Discrete Random Variable

A random variable $X : \Omega \to \mathbb{R}$ is **discrete** if it takes on only countably many values, that is, the set of possible values of X, $X(\Omega) = \{X(w) : w \in \Omega\}$, is countable.

Examples of Discrete Random Variables

- X = number of heads in 3 coin tosses
  $X(\Omega) = \{0, 1, 2, 3\}$ is countable.
- X = number of coin tosses until a head comes up
  $X(\Omega) = \mathbb{N}$ is countable.
- **Remark**: $\mathbb{R}$ is not countable. The proof is beyond the scope of this course. You can *use this property without proof*.

Probability Mass Function (PMF)

The **Probability Mass Function (PMF)** of a discrete random variable X is a function $p : \mathbb{R} \to [0,1]$ defined by

$$p(x) = P(X = x).$$

$P(X=3) = \frac{1}{8}$.

$P(X=2) = \frac{3}{8}$

Properties of PMF

Lemma 2. If $X : \Omega \to R$ is a discrete random variable with PMF $p(x)$. Then

(a) $p(x) = 0$ for any $x \notin X(\Omega)$.

(b) $\sum_{x \in X(\Omega)} p(x) = 1$.

Example 10

- $X = \#$ heads in 3 independent fair-coin tosses. The set of possible values of X is $\{0, 1, 2, 3\}$ and

$$\begin{aligned} p(0) &= P(X=0) = P(\{\text{TTT}\}) = \frac{1}{8}, \\ p(1) &= P(X=1) = P(\{\text{HTT,THT,TTH}\}) = \frac{3}{8}, \\ p(2) &= P(X=2) = P(\{\text{HHT,HTH,THH}\}) = \frac{3}{8}, \\ p(3) &= P(X=3) = P(\{\text{HHH}\}) = \frac{1}{8}. \end{aligned}$$

- $p(0) + p(1) + p(2) + p(3) = 1.$

Example 11

$X = \#$ independent coin tosses until a head comes up. The set of all possible values for X is $\mathbb{Z}^+$. So

fair coin. $P(H)=P(T)=\frac{1}{2}$

- $p(x) = (\frac{1}{2})^x, x \in \mathbb{Z}^+$ and $p(x) = 0$ otherwise.

$(1-\frac{1}{2})^{x-1} \cdot \frac{1}{2}$.
$= (\frac{1}{2})^x$.

- We have

$$\sum_{x\in\mathbb{Z}^+} p(x) = \sum_{x=1}^{\infty} \frac{1}{2^x} = \lim_{n\to\infty} \sum_{x=1}^{n} \left(\frac{1}{2}\right)^n$$
$$= \lim_{n\to\infty} (1/2)\cdot\left(1 + (1/2) + \cdots + (1/2)^{n-1}\right)$$
$$= \lim_{n\to\infty} \frac{1}{2}\cdot\frac{1-(1/2)^n}{1-1/2}$$
$$= \frac{1}{2}\cdot\frac{1-0}{1/2} = 1.$$

geometric series.

$\frac{a}{1-r} = \frac{\frac{1}{2}}{1-\frac{1}{2}}$
$= 1$

Cumulative Distribution Function (CDF)

The **Cumulative Distribution Function (CDF)** of a random variable $X : \Omega \to \mathbb{R}$ is a function $F : \mathbb{R} \to [0, 1]$ defined by:

$$F(x) = P(X \leq x), \;\; x \in \mathbb{R}.$$

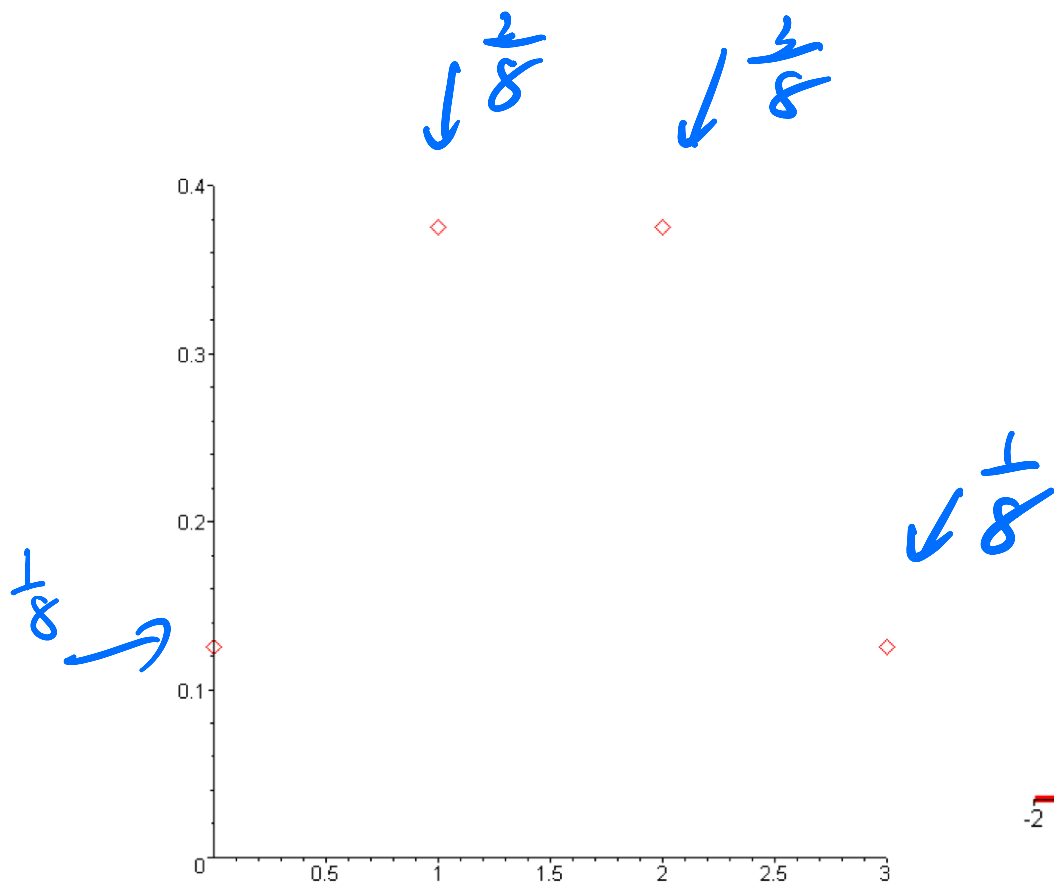
CDF is Nondecreasing

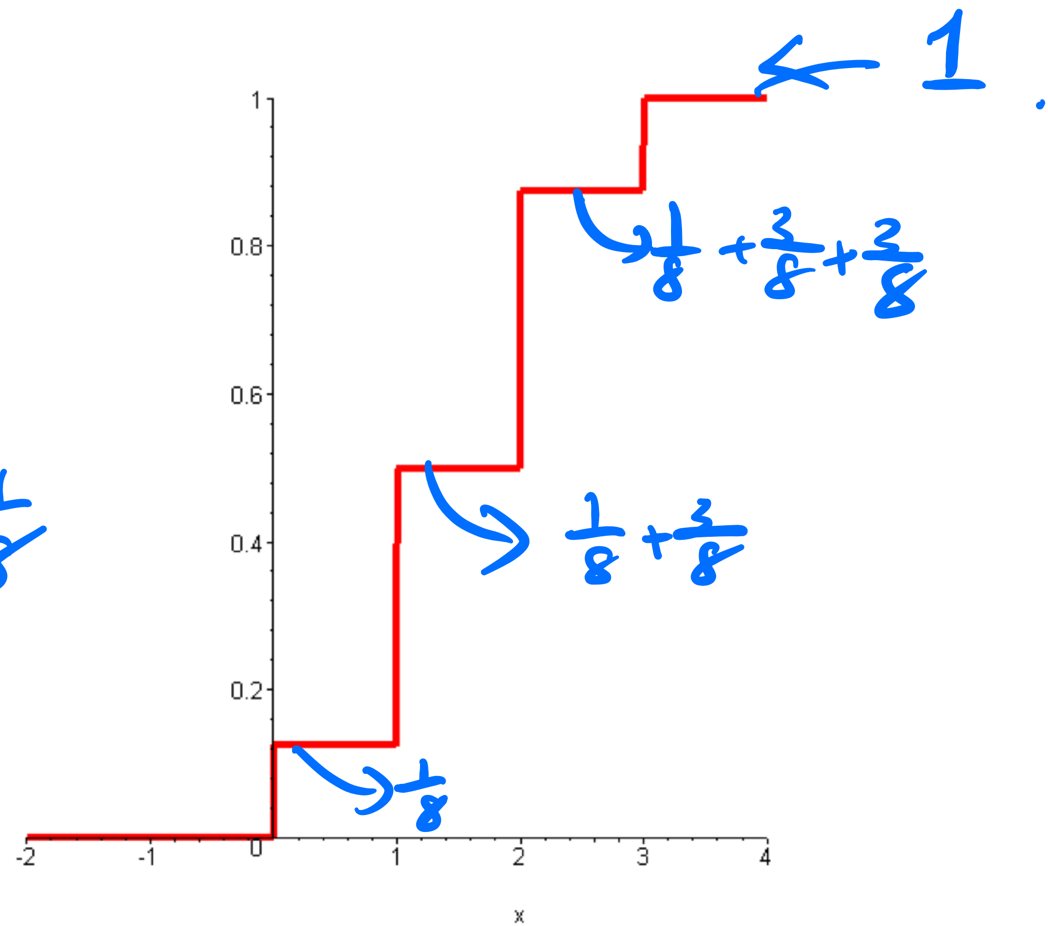
Lemma 3. F is a *nondecreasing* function, that is, $F(a) \leq F(b)$ whenever $a \leq b$.

Proof. For $a \leq b$ we have

$$\begin{aligned} F(b) &= P(X \leq b) \\ &= P(X \leq a) + P(a < X \leq b) \\ &\geq P(X \leq a) \\ &= F(a). \end{aligned}$$

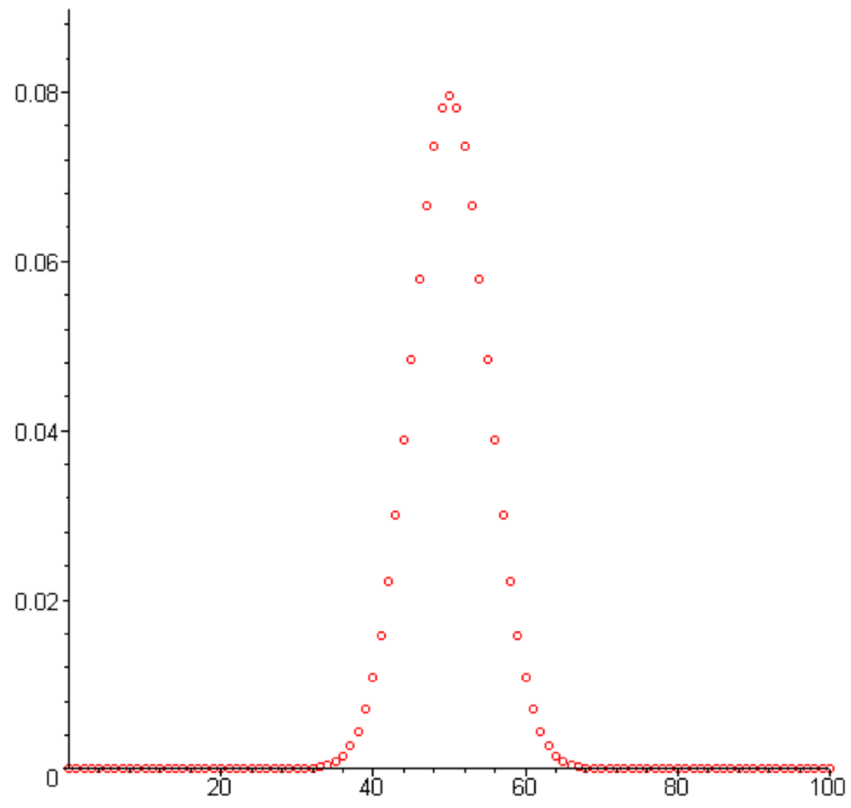
Graphs of CDF and PMF

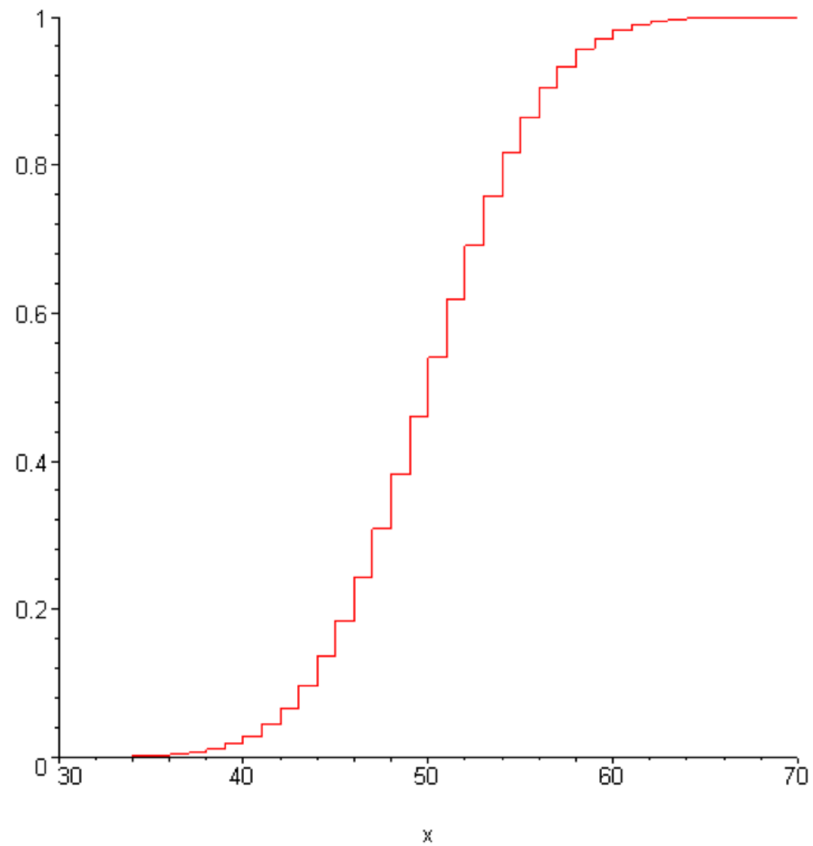




X = number of heads in 3 coin tosses

Graphs of CDF and PMF





X = number of heads in 100 coin tosses