MATH1241 Algebra, 2018
Group 2 — Tues 12 pm, Thurs 10 am

A/Prof. Catherine Greenhill

School of Mathematics and Statistics
University of New South Wales

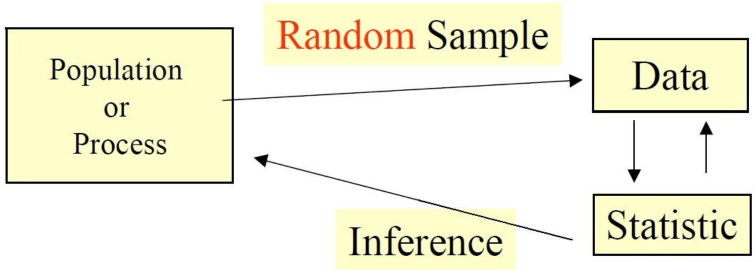
c.greenhill@unsw.edu.au

Acknowledgement: Lectures based on Dr. Chi Mak's notes.

Chapter 9 Probability and Statistics

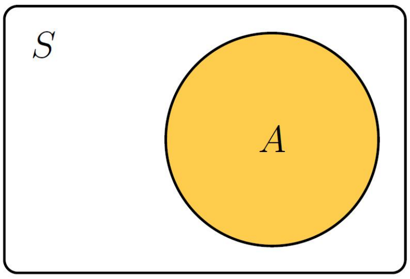
Statistics is the science of production, analysis and interpretation of numerical data for an objective.

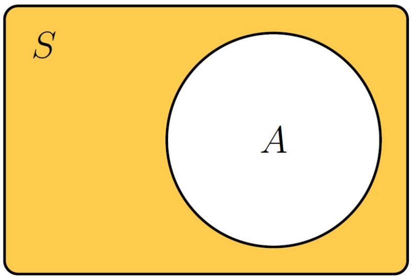
- **Data production**
  - What should be measured?
  - Size of the sample?
  - Ensure randomness.
- **Data analysis**
  - Organise the data in useful form for a specific purpose.
- **Statistical inference**
  - Draw valid conclusion from samples for the population.



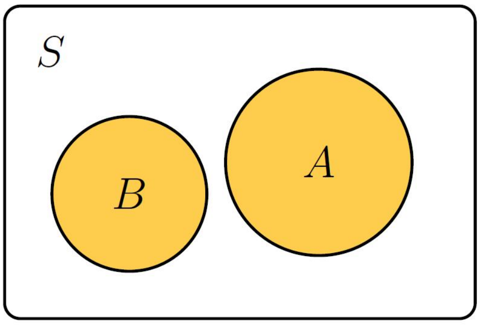
To draw valid conclusions for a population from random samples, we need to study probabilities. To study probability theory, we need the notion of sets. Assuming you are familiar with set notation, we start from Venn diagrams — illustrations of relationships among sets.

S is the universal set which contains all objects of given interest.

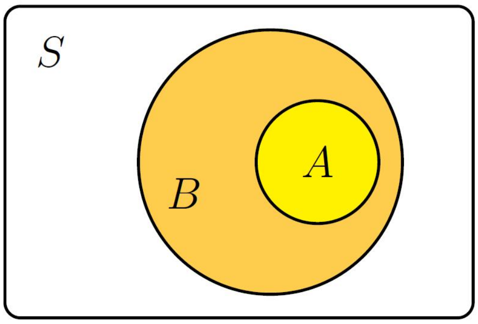




A^c

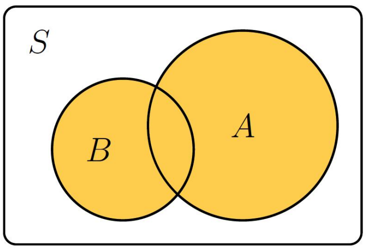


$A \cap B = \emptyset$

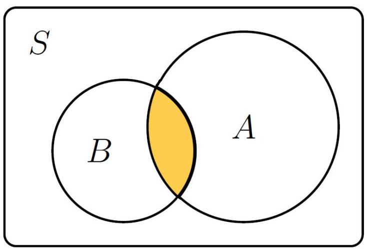


$A \subseteq B$

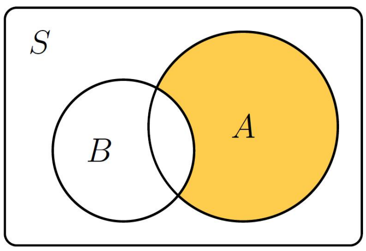
When the intersection of two sets is empty, the two sets are said to be **disjoint**.



$A \cup B$



$A \cap B$



$A - B$

Partition

Definition

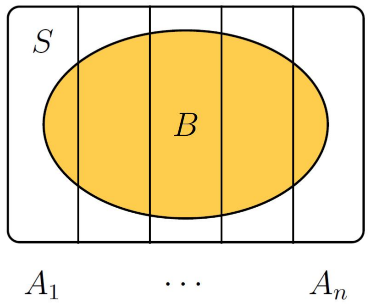
The sets $A_1, A_2, \ldots, A_n$ is said to **partition** a set B when

- $A_1, A_2, \ldots, A_n$ are pairwise disjoint, i.e. $A_i \cap A_j = \emptyset$ whenever $i \neq j$, and
- $A_1 \cup A_2 \cup \cdots \cup A_n = B$.

Proposition

If $A_1, \ldots, A_n$ partition S and B is a subset of S then $A_1 \cap B, \ldots, A_n \cap B$ partition B.

$A_1 \cap B, \ldots, A_n \cap B$ are pairwise disjoint, and the union of these sets is B.



Laws of set algebra

Union is a binary operation (with 2 operands), and so is intersection. Complement is a unary operation. Just like addition, multiplication and negative operations in numbers, these operations obey a number of laws. For example, union and intersection obey the commutative laws and associative laws, such as

$$A \cup B = B \cup A, \quad \text{and} \quad A \cap (B \cap C) = (A \cap B) \cap C.$$

Two less obvious ones:

Distributive laws

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

De Morgan's Laws

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

Example

Illustrate $(A \cup B)^c = A^c \cap B^c$ using Venn diagrams.

Proof.

$\square$

Size of a set

Definition

Let A be a finite set, i.e. it contains a finite number of elements. The **size** of A, denoted by $|A|$, is the number of elements in A.

Example

If $A = \{x : x \text{ is a lower case letter in the English alphabet}\}$, then $|A| = 26$.

Obviously, if finite sets A and B are disjoint then

$$|A \cup B| = |A| + |B|.$$

We can use this to prove the following theorem.

Theorem (The Inclusion-Exclusion Principle)

If A and B are finite sets then

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Example

How many integers between 1 and 100 inclusive are multiples of 3 or 5?

Solution

The inclusion-exclusion principle can be generalised to three or more sets.

Example

Let A, B, and C be finite sets. Prove that

$$\begin{aligned} |A \cup B \cup C| &= |A| + |B| + |C| - |A \cap B| - |A \cap C| \\ &\quad -|B \cap C| + |A \cap B \cap C|. \end{aligned}$$

The proof is omitted. To prove it, apply the inclusion-exclusion principle to the two sets A and $B \cup C$.

To solve enumeration problems involving more than two sets, it will be more efficient to use Venn diagrams.

Example

In a class of 40 people studying music: 2 play violin, piano and recorder, 7 play at least violin and piano, 6 play at least piano and recorder, 5 play at least recorder and violin, 17 play at least violin, 19 play at least piano, and 14 play at least recorder. How many play none of these instruments?

Solution

Infinite sets

Suppose that A is a finite set of n elements. We can always list out the elements as a sequence of n terms. However, for some infinite sets, their elements cannot be listed out as infinite sequences. For example $\mathbb{N}$ is a set which can be listed out as an infinite sequence $0, 1, 2, 3, 4, \ldots$, but we cannot do so for $\mathbb{R}$.

Definition

A set S is **countable** if its elements can be listed as a sequence.

Example

Prove that $\mathbb{Z}$ is countable.

Proof.

$\square$

Attempt Problems 9.1.

Probability

We consider probability as a theory which has been developed to analyse the **outcomes** of *repeated* **experiments**, where an experiment should usually be thought of as an observation or measurement made under precisely specified conditions.

Definition

The set of all possible **outcomes** of a given experiment is called the **sample space** for that experiment.

Example

Find the sample space for the experiment of tossing a coin three times.

Solution

In an experiment, we are often interested in a set of particular outcomes. We will refer to this as an **event**.

Definition (Event)

An **event** is a subset (possibly empty) of the sample space.

Example

The set $\{HHT, HTH, THH\}$ is the event of getting exactly two heads in 3 tosses.

Definition

Two events A and B are said to be **mutually exclusive** if $A \cap B = \emptyset$.

Note that mutually exclusive events are disjoint sets.

Now, we have gathered enough terms to define probability.

Definition of probability

Definition

Let S be a sample space. Then **probability** is a real-valued function P defined on the set of all events (subsets of S), such that

a) $0 \leq P(A) \leq 1$ for all $A \subseteq S$;

b) $P(\emptyset) = 0$;

c) $P(S) = 1$; and

d) if A, B are mutually exclusive events then

$$P(A \cup B) = P(A) + P(B).$$

Although the definition does **not** tell us how to find the probability of an event, there are some probability rules which follow immediately from it.

Rules for probability

Theorem

1. (Addition rule). *For events A, B*
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
2. $P(A^c) = 1 - P(A)$.
3. *If $A \subseteq B$ then $P(A) \leqslant P(B)$.*

Proof of (2).

$\square$

Example

Suppose that A and B are two mutually exclusive events with

$$P(A) = 0.48, \qquad P(B) = 0.36.$$

Write down the values of
a) $P(A \cup B)$, b) $P(A \cap B)$, c) $P(A^c)$, d) $P(A \cap B^c)$.

Solution

Probabilities are commonly allocated using either:

1. An analysis of the symmetries in the experiment.
2. The results of large data collections.
3. Use of expert opinion.

When a and b are two distinct outcomes, the events $\{a\}$ and $\{b\}$ are disjoint. Hence we have the following theorem.

Theorem

Let P be a probability on a countable sample space S and let A be an event in S. Then

1. $P(A) = \sum_{a \in A} P(\{a\})$. *In particular,* $\sum_{a \in S} P(\{a\}) = 1$.
2. *If all the outcomes of a finite sample space S are equally likely, i.e. $P(\{a\})$ is a constant for all $a \in S$, then* $P(A) = \dfrac{|A|}{|S|}$.

Example (Birthday Problem)

What is the probability that in a room of n people, at least two people were born on the same day of the year? We will exclude the possibility of leap years and take the number of days per year to be 365.

Solution

Solution (continued)

This probability is rather surprising, as the following table will reveal:

n	4	16	23	32	40	56
p_n	0.016	0.284	0.507	0.753	0.891	0.988

Conditional probability

Perform two coin tosses with a fair coin. What is the probability that the outcome is two tails? We can tabulate the equally likely outcomes in the following table.

		2nd throw H	2nd throw T
1st throw	H	HH	HT
	T	TH	TT

From the above table, the probability of two tails is $\frac{1}{4}$.

Now, we are told that there at least one coin came up tails.

- In this case, the outcome "two heads" has to be excluded from the sample space.
- If we know that there is at least one tail, the probability of two tails will be $\frac{1}{3}$.

Let A and B be certain events in a sample space S. If it is given that B has happened, the sample space is reduced from S to B. The probability of A under the assumption that B has happened is called the conditional probability of A given B. In the above example, the conditional probability of two tails, given that there is at least one tail, is $\frac{1}{3}$. Formally:

Definition (Conditional Probability)

The **conditional probability** of A **given** B is denoted and defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{provided} \quad P(B) \neq 0.$$

Example

The letters in the word MATHEMATICS are re-arranged at random to form a new string of letters. What is the probability that the new word ends in S if we know it begins with MT?

Solution

An immediate consequence of the definition of conditional probability is

Multiplication Rule

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

The conditional probabilities are implicitly used in tree diagrams.

Example

Urn 1 contains 2 red balls and 3 blue balls. Urn 2 contains 1 red ball and 2 blue balls. Suppose that a ball is drawn at random from Urn 1 and transferred to Urn 2. A ball is then drawn at random from the 4 balls in Urn 2.

a) What is the probability that the ball drawn from Urn 2 is red?

b) Given that the ball drawn from Urn 2 is red, what is the probability that the ball transferred from Urn 1 was blue?

Solution

Solution

Bayes' Rule

The previous example illustrated two important rules.

Theorem

Suppose that $A_1, \ldots, A_n$ partition the sample space S, and let B be any event in S. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\, P(A_i). \qquad \textbf{(Total probability rule)}$$

If A_k is one of the sets in the partition then

$$P(A_k|B) = \frac{P(B|A_k)\, P(A_k)}{\displaystyle\sum_{i=1}^{n} P(B|A_i)\, P(A_i)}. \qquad \textbf{(Bayes' rule)}$$

The rules can be easily explained by a tree diagram.

In applying Bayes' rule, it is a good idea to draw a tree diagram.

Example (Diagnostic Test)

A certain diagnostic test for a disease if 99 % sure of correctly indicating that a person has the disease when the person actually does have it and 98 % sure of correctly indicating that a person does not have the disease when they actually do not have it. In medical jargon the test is "positive" if it suggests that the person has the disease.

Suppose that 2 % of the population actually have the disease.

a) What is the probability that a person does not have the disease when they test positive. (This is called a *false positive.*)

b) Again suppose 2 % of the population actually have the disease. What is the probability of a false negative?

Solution

Solution (continued)

Statistical independence

Intuitively, two events A and B are independent if the occurrence of one does not affect the probability of the other. That is,

$$P(A|B) = P(A) \quad \textit{and} \quad P(B|A) = P(B).$$

Since $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$, we define independence as follows.

Definition

Two events A and B are **statistically independent** (or **independent** for short) if $P(A \cap B) = P(A) \cdot P(B)$.

For more than two events, pairwise independent is not sufficient to conclude that the events are independent.

Definition

Events $A_1, \ldots, A_n$ are **mutually independent** if and only if, for any $A_{i_1}, \ldots, A_{i_k}$ of these,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \times \cdots \times P(A_{i_k}).$$

Example

A green die and a red die are rolled, and we note the numbers shown on the uppermost faces. Let

A be the event 'the number on the green die is even',

B be the event 'the number on the red die is even',

C be the event 'the total of the 2 numbers is even'.

Are A, B, and C mutually independent?

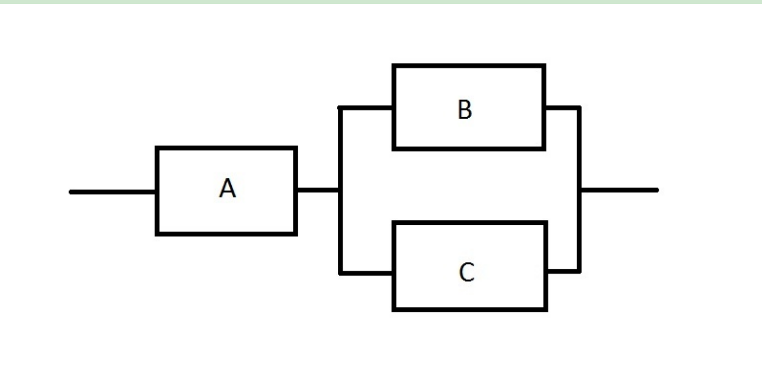
Solution

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Note: A, B and C are pairwise independent, but if A and B both occur then C definitely occurs.

Example

The following system consists of three independent components. It fails when either A fails or both B and C fail.



The probabilities that components A, B, and C fail in a year are 0.1, 0.3, and 0.35, respectively. Find the probability that the system fails in a year.

Solution

Attempt Problems 9.2.

Random variables

Suppose that in a random process, we want to measure some quantities which depend on the outcome.
For example, in each game, Jack and Jill toss two one-dollar coins simultaneously. After the toss, the coins showing heads belong to Jack and the coins showing tails belong to Jill.
Random process — tossing four coins.

N is the number of heads.

Y is the net gain of Jack in a game.

Another example:
Random process — randomly pick a student from this lecture theatre.

H is the height of the student.

We call N, Y, H random variables. Formally:

Definition

A **random variable** is a real-valued function defined on a sample space.

Jack and Jill toss two one-dollar coins simultaneously. After the toss, the coins showing heads belong to Jack and the coins showing tails belong to Jill.

Random process — tossing four coins.

> N is the number of heads.
>
> Y is the net gain of Jack in a game.

Example

$N(HHTT) = 2, \qquad Y(HHTT) = 0, \qquad Y(HHHT) = 1.$

Remarks

- The random variables N and Y are related by $Y = N - 2$.
- $P(Y = n - 2) = P(N = n)$. For instance, $P(Y = -1) = P(N = 1)$.

We are interested in events associated with the values of the random variable, and the probabilities of these events. For instance, the event A such that $Y < 0$ is

$$\{HTTT,\ THTT,\ TTHT,\ TTTH,\ TTTT\}.$$

We also use the notation $P(Y < 0)$ to denote the probability $P(A)$, so

$$P(Y < 0) = P(A) = \frac{5}{16}.$$

In general, if X is a random variable defined on the sample space S then we write

$$\begin{aligned} P(X = r) &= P(\{s \in S : X(s) = r\}), \\ P(X \leqslant r) &= P(\{s \in S : X(s) \leqslant r\}), \\ P(X > r) &= P(\{s \in S : X(s) > r\}), \end{aligned}$$

and so on.

Cumulative distribution function

Definition

The **cumulative distribution function** F_X of a random variable X is given by

$$F_X(x) = P(X \leqslant x) \quad \text{for all} \quad x \in \mathbb{R}.$$

We drop the subscript and say "let $F(x)$ be the cumulative distribution function of X" unless we are working on a few random variables simultaneously.

Facts:

- If $a \leqslant b$ then $F(a) \leqslant F(b)$.
- If $a \leqslant b$ then $P(a < X \leqslant b) = F(b) - F(a)$.
- $\lim_{x \to -\infty} F(x) = P(\emptyset) = 0$ and $\lim_{x \to \infty} F(x) = P(S) = 1$.

Discrete random variables

Definition

A random variable X is **discrete** if its image is countable. (The *image* of X is the range of X as a function)

To specify the probabilities of the events associated with a discrete random variable, we use:

Definition

The **probability distribution** of a discrete random variable X is the function f defined by

$$f(x) = P(X = x) \quad \text{for all} \quad x \in \mathbb{R}.$$

Since the image of X is countable, we can list the values of X as

$$x_0, x_1, x_2, \ldots,$$

which may be a finite or infinite sequence. Let $p_k = P(X = x_k)$ for each x_k in this sequence. Then the sequence

$$p_0, p_1, p_2, \ldots$$

specifies the probability distribution of X.

Example

Tabulate the values and probabilities distributions of the random variables N and Y in the motivating example.

Solution

To show that $p_0, p_1, p_2, \ldots$ is a probability distribution of a discrete random variable, we need to show

1. $p_k \geqslant 0$ for $k = 0, 1, 2, \ldots$
2. $p_0 + p_1 + \cdots = 1$.

Example

Show that the sequence defined by

$$p_k = (1-p)^{k-1}p \quad \text{for} \quad 1 \leqslant k < \infty, \text{ where } 0 < p < 1,$$

is the probability distribution of a random variable X such that $p_k = P(X = k)$.

Solution

To illustrate a probability distribution $P(X = x)$ of a discrete random variable X, we can sketch $P(X = x)$ as a function on $\mathbb{R}$. Alternatively, we can use a probability histogram, especially when the image of X is a set of consecutive integers.

Example (Loaded die)

Each face of a loaded die has a probability of occurring $P(X = k) = \dfrac{\alpha}{k}$, for $k = 1, 2, 3, 4, 5, 6$.

a) Find the value of α.

b) Sketch the probability distribution and the probability histogram of the random variable X.

c) Find and sketch the cumulative distribution function of X.

Solution

Solution (continued)

Solution (continued)

Example (of a geometric distribution)

In a game, we toss a die and I win if a six is thrown. Let X be the number of tosses until I win. Find and sketch the probability distribution for X.

Solution

Solution (continued)

Example (Distribution of class test marks)

Marks of Algebra Test 1 of a Tutorial:

9	10	8	6	8	5	4	9	0	7	10	5	3	0	9	1
10	9	6	10	9	6	5	10	7	10	9	5	7	9	6	8

Let X be the algebra test 1 mark of a randomly chosen student from the tutorial. Find the probability distribution of X.

Solution

Marks	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Tally											
Frequency	*2*	*1*	*0*	*1*	*1*	*4*	*4*	*3*	*3*	*7*	*6*
Probability	$\frac{1}{16}$	$\frac{1}{32}$	0	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{7}{32}$	$\frac{3}{16}$

Mean and variance

For a set of numbers $x_1, \ldots, x_n$,

1. the mean $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{k=1}^{n} x_k$,
2. the standard deviation is the root mean square deviation, that is

$$\sqrt{\frac{(x_1-\bar{x})^2 + \cdots + (x_1-\bar{x})^2}{n}} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k-\bar{x})^2}.$$

For large dataset with repetition, we can calculate the mean and standard deviation from the frequency table

x	x_1	x_2	$\cdots$	x_n
f	f_1	f_2	$\cdots$	f_n

by

$$\frac{1}{N}\sum_{k=1}^{n} f_k x_k \quad \text{and} \quad \sqrt{\frac{1}{N}\sum_{k=1}^{n} f_k(x_k-\bar{x})^2}, \quad \text{where } N = \sum_{k=1}^{n} f_k.$$

Let X be the value of a randomly chosen number from the dataset. We can rewrite the mean as

$$\sum_{k=1}^{n} x_k \frac{f_k}{N} = \sum_{k=1}^{n} x_k P(X = x_k).$$

The square of the standard deviation is known as the *variance*, which can be written as

$$\sum_{k=1}^{n} (x_k - \bar{x})^2 \frac{f_k}{N} = \sum_{k=1}^{n} (x_k - \bar{x})^2 P(X = x_k).$$

Definition (Mean)

If X is a discrete random variable with values $\{x_k\}_{k=0}^{\infty}$ and probability distribution $\{p_k\}_{k=0}^{\infty}$ then the **mean** or **expected value** of X is given by

$$E(X) = \sum_{k=0}^{\infty} x_k p_k,$$

provided the series converges. The mean is commonly denoted by μ or μ_X.

Theorem

If $Y = g(X)$ is the random variable whose values are related to those of X by $y = g(x)$, then

$$E(Y) = E(g(X)) = \sum_{k=0}^{\infty} g(x_k) p_k.$$

Definition (Variance and Standard Deviation)

The **variance** of a discrete random variable X is defined by

$$\text{Var}(X) = \sum_{k=0}^{\infty} (x_k - \mu_X)^2 p_k = E\left((X - E(X))^2\right).$$

The **standard deviation** of X, is then defined by

$$\text{SD}(X) = \sigma = \sigma_X = \sqrt{\text{Var}(X)}.$$

- The variance is the average square deviation from the mean.
- The standard deviation is the root mean square deviation.
- The variance and standard deviation measure the spread of the probability distribution of a random variable.
- We normally use the following formula to calculate the variance instead of using the definition.

Theorem

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2.$$

Proof.

Proof (continued).

$\square$

Example (loaded die, continuing from p. 44)

Find the mean, variance, and standard deviation of the random variable X in the loaded die example.

Solution

There are many ways to define a new random variable Y from an existing one X. Generally, there are no easily-described relationships between their means and between their variances. However, if $Y = aX + b$ then we have the following simple identities.

Theorem

Let X be a discrete random variable with mean $E(X)$ and variance $\mathrm{Var}(X)$. If we make the linear change of variable $Y = aX + b$, where a, b are real constants, then

1. $E(Y) = aE(X) + b$,
2. $\mathrm{Var}(Y) = a^2 \,\mathrm{Var}(X)$, *and*
3. $\mathrm{SD}(Y) = |a| \,\mathrm{SD}(X)$.

Example (Loaded Die from p. 44)

Jack plays a die game with the loaded die (p. 44) with a gambler. He has to pay \$5 to play a game. He then will receive \$$2x$ if he throws an x. Let Y be his net gain in a game. Find $E(Y)$, $\text{Var}(Y)$, and $\text{SD}(Y)$.

Solution

Will Jack win or lose money in long run?

Attempt Problems 9.3.

The binomial distribution

A **Bernoulli trial** is an experiment with two outcomes labelled *Success* and *Failure*. Suppose that the chance of success is p. Let X be the random variable counting the number of successes in n independent identical Bernoulli trials. The probability distribution of X is called a **binomial distribution**, and we write $X \sim B(n, p)$.

Definition

The **binomial distribution** $B(n, p)$ for $n \in \mathbb{N}$ is the function

$$B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where} \quad k = 0, 1, \ldots, n,$$

and 0 otherwise.

$B(n, p)$ **is** a probability distribution because

$B(n, p, k) \geqslant 0$ for all k, and $\displaystyle\sum_{k=0}^{n} B(n, p, k) = ((1-p) + p)^n = 1.$

Theorem

If $X \sim B(n, p)$ *then* $E(X) = np$ *and* $Var(X) = npq = np(1 - p)$.

We omit the proof of this theorem. A proof for $E(X) = np$ is provided in the Algebra Notes, Section 9.4. However, we need to remember the results of the theorem.

Example

On average, twenty percent of patients brought by ambulance to a hospital are admitted to the intensive care ward (ICW). On one particular night there were 7 spare beds in ICW and 17 people were brought to the hospital.

a) What is the probability that 4 spare beds in ICW were filled that night?

b) What is the probability that at least one was turned away that night because all the beds in ICW were full?

c) What would be the expected number of admissions to ICW on a night when 21 patients were brought in by ambulance?

Solution

Geometric distribution

Instead of performing n independent identical Bernoulli trials, we now perform independent identical Bernoulli trials until a success. If X is the number of trials until the first success appears, the probability distribution of X is known as a geometric distribution.

Definition

Fix $0 < p < 1$. The **geometric distribution** $G(p)$ is the function

$$G(p,k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \ldots$$

We proved earlier that the above sequence does define a probability distribution.

We write $X \sim G(p)$ to denote that X has this distribution.

Example

Prove that if $X \sim G(p)$ then $E(X) = \dfrac{1}{p}$.

Solution

If $X \sim G(p)$ then the variance of X is $\dfrac{1-p}{p^2}$.

Example

A die is repeatedly rolled until a 6 appears.

a) What is the probability that the first 6 appears on the 6th roll or later?

b) What is the probability that the first 6 appears within 5 rolls?

c) What is the expected number of rolls to get the first 6?

Solution

Applications of the binomial distribution to testing claims

Example

Olof Jonsson was tested on his psychic abilities in an experiment. In the experiment a computer showed 4 cards to him and (randomly) picked one of them, and he has to guess which card he thought the computer had picked. This process was repeated 288 times, and Olof managed to pick the right card 88 times.
Are you convinced that Olof was psychic?

He guessed correctly in 30.6 % of the trials.

Example

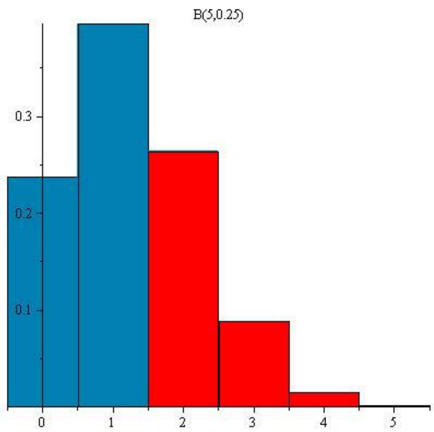
Jack did the same experiment.
In 5 attempts, and he picked the right card 2 times.
Is Jack better than Olof?

Jack guessed correctly in 40 % of the trials.

In Jack's case, he guessed 40 % correctly, which is better than the expected 25 %. Before we draw the conclusion that he is a psychic, we should ask the question "can such a good result (40 % or more) be explained by chance?".

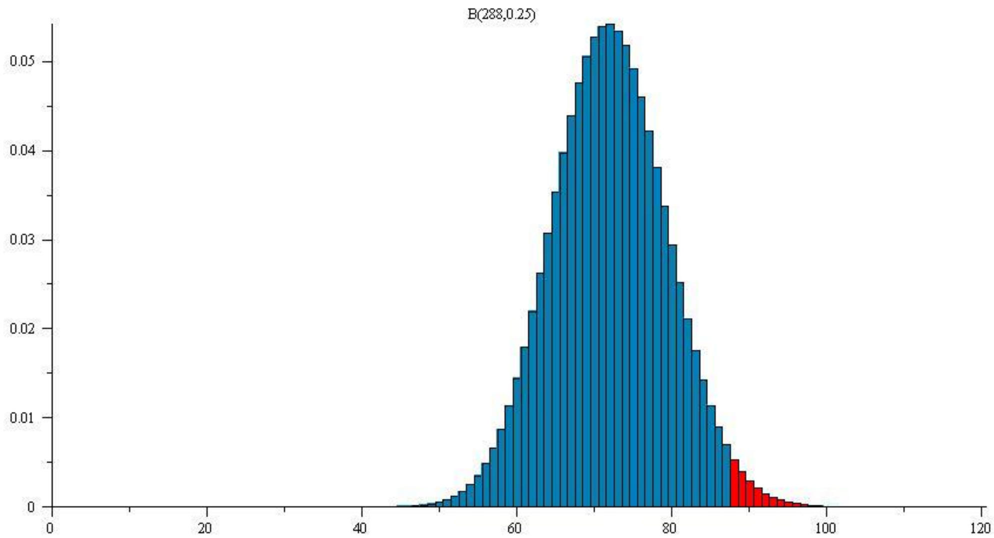
Assuming that Jack is not special, the distribution for the number of correct guesses, X, is the binomial distribution $B(5, 1/4)$.

We calculate $P(X \geqslant 2)$, which measures how unusual it is for Jack to guess 40 % or more correctly in 5 trials.



We call the probability $P(X \geqslant 2)$ a tail probability. (Can you see why?)
Jack's result cannot/can be explained by chance, so he is/is not a psychic.

For Olof,



Is it clear why we call $P(X \geqslant 88)$ a tail probability?

In general, we call expressions of the following type *tail probabilities*:

- $P(X \geqslant t)$ for $t > E(X)$,
- $P(X > t)$ for $t > E(X)$,
- $P(X \leqslant t)$ for $t < E(X)$,
- $P(|X - E(X)| > t)$.

The sign test

Example

The following data are 15 measurements of percentage moisture retention using a new sealing system.

97.5	95.2	97.3	96.0	96.8
99.8	97.4	95.3	98.2	99.1
96.1	97.6	98.2	98.5	99.4.

The previous system had a retention rate of 96 %. Can we claim that the new system is better?

Solution

- *First, replace each score with $+$, 0 or $-$, to indicate whether the reading is bigger than, the same as, or less than 96%, respectively:*
  $+$ - $+$ 0 $+$ $+$ $+$ - $+$ $+$ $+$ $+$ $+$ $+$ $+$

Solution (continued)

- *Assuming that there was no difference between the old system and the new system, calculate the tail probability of how unusual it is to get such readings.*

Solution (continued)

- *Conclusion.*

Attempt Problems 9.4.

Continuous random variables

In contrast to the discrete random variables, random variables such as the height or weight of individuals cannot be defined by probabilities $P(X = x)$ for single values x, since these probabilities each equal 0. Instead, we will define such random variables in terms of the cumulative distribution function

$$F_X(x) = F(x) = P(X \leqslant x) \qquad \text{for} \quad x \in \mathbb{R}\,.$$

Definition

A random variable X is **continuous** if and only if its cumulative distribution function $F_X(x)$ is continuous.

Strictly speaking, $F(x)$ must be differentiable except for at most countably many points.

Example

The director of a certain company is waiting to be picked up. Let the waiting time be X min. The director knows that their car will surely arrive within 5 minutes. The probability that the car will arrive within x minutes is proportion to x for $0 \leqslant x \leqslant 5$. That is,

$$P(X \leqslant x) = kx \quad \text{for } 0 \leqslant x \leqslant 5,$$

where k is a constant.
Find and sketch the cumulative distribution function for X.

Solution

Probability density function

For discrete random variables X, the cumulative distribution function $F(x)$ is a sum over the probability distribution values $p_k = P(X = x_k)$. For continuous random variables, $F(x)$ is an *integral* of the probability density function, which we now define:

Definition

The **probability density function** $f = f_X$ of a continuous random variable X is defined by

$$f(x) = f_X(x) = \frac{d}{dx}F(x)\,, \qquad x \in \mathbb{R}$$

if F is differentiable at x, and

$$f(x) = f_X(x) = \lim_{y \to x^-} \frac{d}{dx}F(y)$$

if F is not differentiable at x.

Example (continued from the previous example)

Find and sketch the probability density function for the random variable X defined in the previous example.

Solution

From the definition and the Fundamental Theorem of Calculus, we have

- $F(x) = \int_{-\infty}^{x} f(t)dt.$
- $P(a \leqslant X \leqslant b) = P(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(x)\,dx$ for $a \leqslant b$.
- Only functions $f : \mathbb{R} \to \mathbb{R}$ satisfying

  $$f(x) \geqslant 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

  can be probability density functions for some random variable.

Example

Let X be a random variable with probability density function

$$f(x) = \begin{cases} \dfrac{k}{x^4} & \text{if } x \geqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

a) Find the value of k.

b) Find $P(X \leqslant 3)$ and $P(2 \leqslant X \leqslant 3)$.

c) Find and sketch the cumulative distribution function for X.

Solution

Solution (continued)

Mean and variance of a continuous random variable

The definitions of mean and variance for continuous random variables are obtained from those of discrete random variables by replacing sums by integrals, and probability distributions by probability density functions.

Definition (Mean)

The **expected value** (or **mean**) of a continuous random variable X with probability density function $f(x)$ is defined to be

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx\,.$$

Theorem

If X is a continuous random variable with density function $f(x)$, and $g(x)$ is a real function, then the expected value of $Y = g(X)$ is

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx\,.$$

Definition (Variance)

The **variance** of a continuous random variable X is

$$\text{Var}\,(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2\,.$$

The standard deviation of X is $\sigma = \text{SD}(X) = \sqrt{\text{Var}\,(X)}$.

Example

Find the mean, the variance, and the standard deviation of the random variable defined in the previous example.

Solution

The mean and variance have the same properties under linear scaling as in the discrete case.

Theorem

If a and b are constants, then

$$\begin{aligned} E(aX+b) &= aE(X)+b, \\ Var(aX+b) &= a^2 Var(X), \\ SD(aX+b) &= |a|\, SD(X). \end{aligned}$$

Example

Let $Y = 3X - 2$, where X is the random variable in the previous example. Find $P(Y \leqslant 7)$. Find the mean and the standard deviation of Y.

Solution

Theorem

Let X be a random variable with $E(X) = \mu$ and $SD(X) = \sigma$. Define $Z = \dfrac{X-\mu}{\sigma}$. Then $E(Z) = 0$ and $Var(Z) = 1$.

Proof.

$\square$

The random variable $Z = \frac{X-\mu}{\sigma}$ is referred to as the **standardised** random variable obtained from X. Note that this theorem holds for discrete and continuous random variables.

Normal distribution

An important continuous distribution which is widely used in statistics is the **normal** distribution. It turns out that it is the limiting case of the binomial distribution.

Definition

A continuous random variable X is said to have **normal distribution** $N(\mu, \sigma^2)$ if it has probability density

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{where} \quad -\infty < x < \infty\,.$$

We write $X \sim N(\mu, \sigma^2)$ to denote that X has the normal distribution $N(\mu, \sigma^2)$.

Theorem (Mean and variance of a normal distribution)

If X is a continuous random variable and $X \sim N(\mu, \sigma^2)$, then

$$E(X) = \mu \quad \textit{and} \quad Var(X) = \sigma^2.$$

Remarks

- The probability density function of a normal distribution has a bell shape graph symmetrical about the mean.
- The distribution $N(0, 1)$ is called the **standard normal distribution**.
- If $X \sim N(\mu, \sigma^2)$ then the standardised random variable $Z = \dfrac{X - \mu}{\sigma}$ is normal and $Z \sim N(0, 1)$.

Let $X \sim N(\mu, \sigma^2)$. It is difficult to evaluate probabilities such as $P(X \leqslant x)$ by integration,

$$P(X \leqslant x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \, dt \, ,$$

because the probability density function does not have an elementary primitive. Instead, we standardise X to $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$. This gives

$$P(X \leqslant x) = P(Z \leqslant z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dt$$

where $z = \dfrac{x-\mu}{\sigma}$. The value of this integral for various z can be found either via a calculator or the table given in the Algebra Notes. This table gives the values of this integral for z in the range -3 to 3. For z less than -3, the value is essentially zero, while for z greater than 3, the value is essentially 1. The value $P(Z \leqslant z)$ is the area to the left of z under the probability density curve of $Z \sim N(0,1)$.

Example

Let $Z \sim N(0,1)$.

a) Find $P(-0.52 \leqslant Z < 1.23)$.

b) Find z such that $P(Z < z) = \dfrac{2}{3}$.

Solution

Example

In a certain examination, the marks are normally distributed with mean 65 and standard deviation 12. (Assume that a mark can be any real number.)

a) For a randomly chosen exam mark, what is probability that it is greater than 50?

b) Find c such that probability of getting a mark higher than c is 0.05.

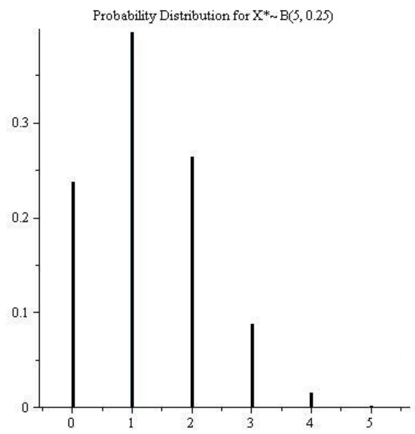
Solution

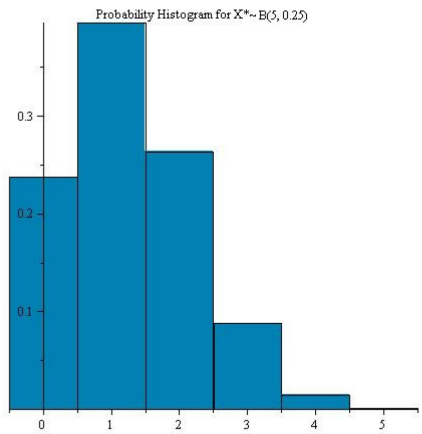
Solution (continued)

Normal approximation to binomial

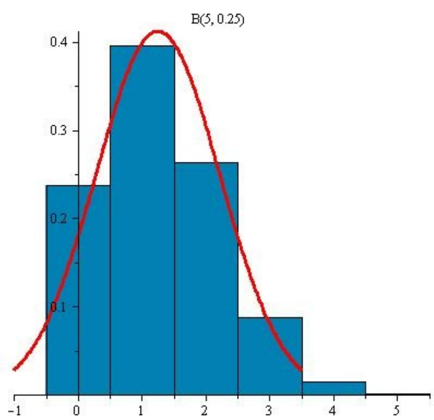
Consider a random variable X with binomial distribution $B(n, p)$. If we treat an integer x as the interval $[x - 0.5, x + 0.5]$, we can define a continuous random variable Y based on the probability histogram for X. Strictly speaking,

$$P(X = x) = P(x - 0.5 < Y \leqslant x + 0.5) \quad \text{for all } x = 0, 1, \ldots n.$$



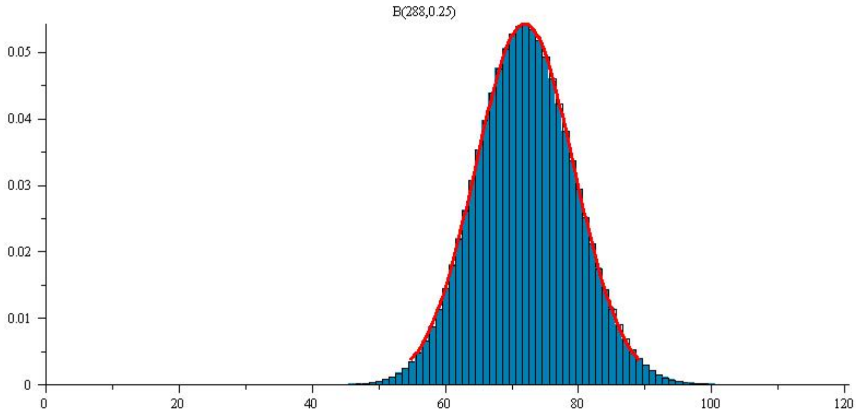


The mean and variance of the binomial distribution $B(n, p)$ are np and $np(1 - p)$, respectively. How good is the approximation if we approximate the binomial distribution with the continuous random variable Y with normal distribution $N(np, np(1 - p))$?



The approximation is not good when $n = 5$ and $p = 0.25$.

When $n = 288$ and $p = 0.25$, the approximation of $B(n,p)$ by $N(np, np(1-p)) = N(72, 54)$ is quite good.



A rule of thumb is that the normal distribution is a good approximation for $B(n,p)$ if both $np > 10$ and $n(1-p) > 10$.

Example

In a poll about ice-cream flavours, 53% of the people in a random sample of 200 prefer chocolate to strawberry.

a) Suppose that it is not true that more people prefer chocolate to strawberry. Then the probability that a randomly chosen person prefers chocolate is 0.5. Under this assumption, what is the probability that there are 53% or more people in a sample of 200 which prefer chocolate?

b) Is there evidence that more people prefer chocolate icecream to strawberry icecream?

c) Repeat part (a) and part (b) for a sample of 1400.

Solution

Solution (continued)

Solution (continued)

Attempt Problems 9.5.

Exponential distribution

Definition

Let $\lambda > 0$. A continuous random variable T has exponential distribution $\mathrm{Exp}(\lambda)$ if it has probability density

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geqslant 0, \\ 0 & \text{if } t < 0. \end{cases}$$

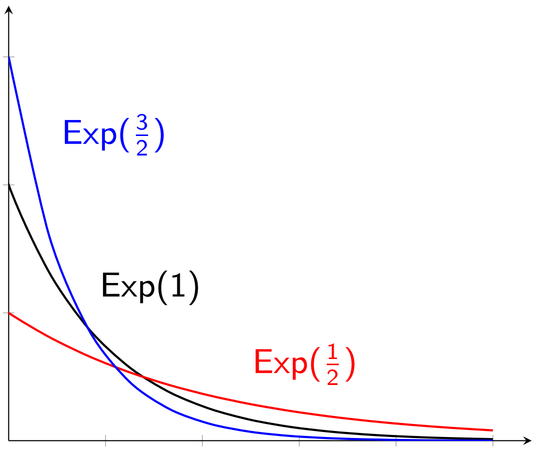
If T has distribution $\mathrm{Exp}(\lambda)$ then we write $T \sim \mathrm{Exp}(\lambda)$.

Exponential distribution

Definition

Let $\lambda > 0$. A continuous random variable T has exponential distribution $\mathrm{Exp}(\lambda)$ if it has probability density

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geqslant 0, \\ 0 & \text{if } t < 0. \end{cases}$$



Theorem (Mean and variance of the exponential distribution)

If T is a continuous random variable and $T \sim \mathrm{Exp}(\lambda)$ then

$$E(T) = \frac{1}{\lambda} \quad \textit{and} \quad Var(T) = \frac{1}{\lambda^2}.$$

Proof.

$\square$

Theorem (Cumulative distribution)

If T is a continuous random variable and $T \sim \mathrm{Exp}(\lambda)$ then

$$F_T(t) = \Pr(T \leqslant t) = \begin{cases} 1 - e^{-\lambda t} & \textit{if } t \geqslant 0, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof.

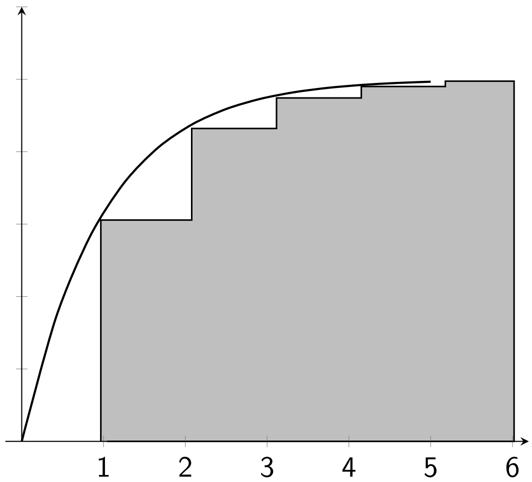
$\square$

Now compare $T \sim \text{Exp}(\lambda)$ and $X \sim G(p)$ where $p = 1 - e^{-\lambda}$. Observe that for all positive integers n,

$$F_T(n) = F_X(n) = 1 - (1-p)^n.$$

Therefore $\text{Exp}(\lambda)$ is approximated by the geometric distribution $G(p)$, with $p = 1 - e^{-\lambda}$.

Alternatively, $G(p)$ is interpolated by $\text{Exp}(\lambda)$, with $\lambda = \ln\left(\frac{1}{1-p}\right)$.



Example

An insurance company has found that $p = 5.02\%$ of its policies are claimed each year. For such a policy, find the

- probability that a claim occurs within the first 6 years;
- probability that a claim occurs within the first 6.5 years;
- probability that the first claim occurs within the 1st half of the 6th year;
- expected number of years until the first claim occurs.

Solution

Solution (continued)

End of Chapter 9. End of MATH1241!!!!