MATH1231 Algebra, 2017
Chapter 9: Probability and Statistics

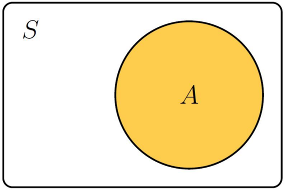
A/Prof. Daniel Chan

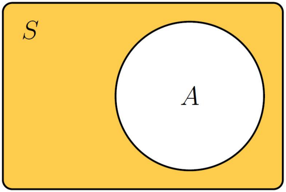
School of Mathematics and Statistics
University of New South Wales

danielc@unsw.edu.au

Review Venn diagrams

We need some basic set theory to study probability & statistics.
Let S be the universal set i.e. set containing all objects of interest in the question at hand & $\varnothing$ denote the empty set.

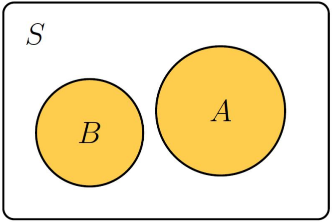




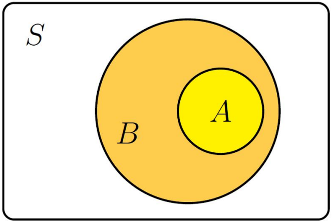
A^c

E.g. $\varnothing^c = S$.

Binary set operations & relations

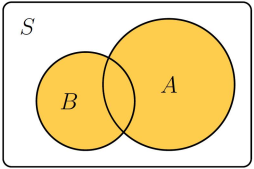


$A \cap B = \emptyset$

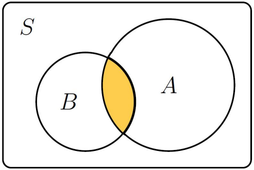


$A \subseteq B$

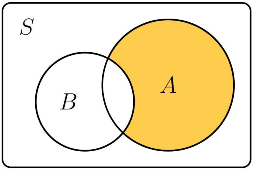
A, B are **disjoint** if $A \cap B = \varnothing$.



$A \cup B$



$A \cap B$



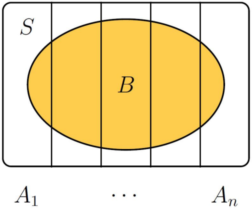
$A - B$

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. $A_i \cap A_j = \varnothing$ whenever $i \neq j$, and
- $A_1 \cup A_2 \cup \cdots \cup A_n = B$.



Proposition

Suppose $A_1, \ldots, A_n$ partition S & $B \subseteq S$ then $A_1 \cap B, \ldots, A_n \cap B$ partition B.

Laws of Set Algebra

Facts

- (Commutative law) $A \cup B = B \cup A, \quad A \cap B = B \cap A.$
- (Associative law)
  $A \cup (B \cup C) = (A \cup B) \cup C, \qquad A \cap (B \cap C) = (A \cap B) \cap C.$
- (Distributive laws) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$
- (De Morgan's Laws) $(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$

Why?

Size or Cardinality of a Set

Definition

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

Example

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

Obviously, if finite sets A and B are disjoint then

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

If not, we have

Theorem (The Inclusion-Exclusion Principle)

If A and B are finite sets then

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

Sample inclusion-exclusion principle question

Example

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

Solution

Let $T = \{3, 6, 9, \ldots, 99\}, F = \{5, 10, 15, \ldots, 100\}$.

Analysing 3 subsets

Example

In a class of 40 students: 2 have bulbasaur, charmander & squirtle, 7 have at least bulbasaur & charmander, 6 have at least charmander & squirtle, 5 have at least bulbasaur & squirtle, 17 have at least bulbasaur, 19 have at least charmander, & 14 have at least squirtle. How many have none of these pokemon?

Solution

Inclusion-Exclusion for 3 subsets

Proposition

Let A, B, and C be finite sets. Then

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

In eg of previous slide, let B, C, S be the sets of students who have bulbasaur, charmandeer & squirtle resp. Then

$$\begin{aligned} |B \cup C \cup S| &= |B| + |C| + |S| - |B \cap C| - |C \cap S| - |B \cap S| + |B \cap C \cap S| \\ &= 17 + 19 + 14 - 7 - 6 - 5 + 2 = 34 \end{aligned}$$

Hence $|(B \cup C \cup S)^c| = 40 - 34 = 6$.

Probability

In probability theory, we consider performing **experiments** e.g. rolling a die, polling voters etc and then observe or measure the **outcomes**.

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

The sample space is the set

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

Events

Definition (Event)

An **event** is a subset of the sample space i.e. a set of possible outcomes.

Eg The set $\{HHT, HTH, THH\}$ is the event "exactly two heads in 3 tosses". $\{HHH\}$ is event "all heads in 3 tosses".

Definition

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

Eg 2 events in previous example are mutually exclusive.

Definition of Probability

Definition

Suppose S is a sample space, then a **probability** on S is a real valued function P defined on the set of all events (subsets of S) such that

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

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

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

d) if A, B are mutually exclusive events then

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

E.g. $S = \{T, H\}$. Following define a probabilities on S

Fair coin: $P(\varnothing) = 0, P(\{T\}) = \frac{1}{2}, P(\{H\}) = \frac{1}{2}, P(\{T, H\}) = 1$

Biased coin: $P(\varnothing) = 0, P(\{T\}) = \frac{1}{3}, P(\{H\}) = \frac{2}{3}, P(\{T, H\}) = 1$

Rules for Probability

Theorem

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

Proof of (2).

□

Example

Suppose A and B are two mutually exclusive events in which $P(A) = 0.48, P(B) = 0.36$. 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

(Discrete) Uniform probability

Q Which probability function should you use?
A It depends on the experiment. **eg** For coin tosses, die rolls etc can invoke principle of symmetry & thm below.

Theorem

Let P be a probability on a finite sample space S.

- *Suppose all outcomes of S are equally likely, i.e. $P(\{a\})$ is constant*

Then $P(\{a\}) = \frac{1}{|S|}$ & $P(A) = \dfrac{|A|}{|S|}$.

The probability function in the thm is called the **discrete uniform probability**.

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 for simplicity.

Solution

If $n > 365$, by the pigeonhole principle, there will surely be at least two people born on the same day i.e. probability = 1.

Assume $n \leqslant 365$. Let Y be the set of the 365 days of the year. Consider sample space

$$S = \{(b_1, \ldots, b_n) \, : \, b_1, \ldots, b_n \in Y\} .$$

Let A = event "at least two of the n people share the same birthday".

Assume that it is equally likely for a person to be born on any date of a year, $\therefore$ outcomes of the sample space are equally likely.

Solution (Continued)

Kindergarten counting arguments give

$$|S| = 365^n \quad \textit{and} \quad |A^c| = 365 \times 364 \times \cdots \times (365 - n + 1).$$

Therefore,

$$P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|S|} = 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}.$$

That is, $P(A) = 1 - \dfrac{{}^{365}P_n}{365^n}$.

Gives some surprising probabilities:

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

Conditional Probability: simple example

Q What is the probability that there are two boys in a randomly chosen family of two children? (Assume boys & girls are equally likely).

		2nd Child	
		B	G
1st Child	B	BB	BG
	G	GB	GG

Probability of two boys $= \frac{1}{4}$.

Now, **suppose** we are told that there is at least one boy.

- The outcome "two girls" must be excluded from the sample space. New sample space $= \{BB, BG, GB\}$.
- The probability of two boys, *given there is at least one boy* will be $\frac{1}{3}$.

Conditional probability: definition

Let A and B be two events in a sample space S.

- If it is given that B has happened, then 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.

Definition (Conditional Probability)

The **conditional probability** of A **given** B is

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

Re-arranging terms gives
Multiplication Rule $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

Tree diagrams for successive experiments

Example

Urn 1 contains 2 red balls & 3 blue. Urn 2 contains 1 red ball & 2 blue. Suppose 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

Bayes' Rule computes conditional probabilities $P(A_i|B)$ from $P(B|A_i)$ as asked for in the previous example.

Theorem

If $A_1, \ldots, A_n$ partition the sample space S, and B is any event in S then
(Total probability rule)

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i), \qquad \textit{and}$$

(Bayes' rule)

$$P(A_k|B) = \frac{P(B|A_k)\,P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$$

Example: Bayes' rule

Example (Diagnostic Test)

A certain diagnostic test for a disease is

- 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

(A test is positive if it indicates the person has the disease.)

Q If 2 % of the population actually have the disease, what is the probability that a person does not have the disease when they test positive. (This is called a *false positive.*)

Solution

Statistical Independence

Informally, two events A and B are independent if the occurrence of one does not affect the probability of the other. i.e.

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

Now $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$ so we define

Definition

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

E.g. Below are the probabilities for being left-handed (L) or right-handed (R) and passing (P) or failing (F) MATH1231. Are L, P independent?

	P	F
L	.09	.01
R	.71	.19

Mutual independence of 3 or more events

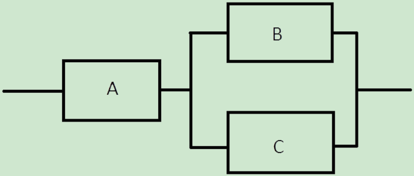
Definition

Events $A_1, \ldots, A_n$ are **mutually independent** if

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \times \cdots \times P(A_{i_k}) \quad \text{for any } i_1, \ldots, i_k.$$

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 (Continued)

Random Variables

Definition

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

E.g. Suppose your experiment is "randomly pick a student from this lecture theatre" (outcome is a student). $H =$ height of student is a random variable.

E.g. Consider a game where Jack and Jill toss two one-dollar coins simultaneously. After the toss, Jack gets all the heads & Jill the tails.

- Experiment = tossing four coins,
- Sample space $S = \{HHHH, HHHT, \dots\}$.

Consider random variables

$N =$ the number of heads.

$Y =$ the **payoff function** Jack's net gain.

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

Probabilities associated to random variables

Random variables often define events we are interested in **e.g.** the event A that Jack wins money is described by $Y > 0$ so
$A = \{HHHH, HHHT, HHTH, HTHH, THHH\}$,

Notation

If X is a random variable defined on the sample space S, we write

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

In above e.g. we have

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

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

Cumulative Distribution Function

Definition

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

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

E.g. In the Jack-Jill example we have

$$F_Y(0) = P(Y \leqslant 0) = 1 - P(Y > 0) = \frac{11}{16}.$$

- Usually drop the subscript X & write $F(x)$ for $F_X(x)$ if X understood.
- If $a \leqslant b$ then $F(b) - F(a) = P(a < X \leqslant b) \geqslant 0$.
- $\therefore F$ is non-decreasing i.e. $a \leqslant b \implies F(a) \leqslant F(b)$.
- $\lim_{x \to -\infty} F(x) = P(\varnothing) = 0$, and $\lim_{x \to \infty} F(x) = P(S) = 1$.

Discrete Random Variables

Let $X =$ random var on sample space S & $P =$ probability on S.

Definition

X is **discrete** if its image is countable i.e. we can list its values as $x_0, x_1, \ldots, x_n$ or $x_0, x_1, x_2, \ldots$.
The **probability distribution** of a discrete random variable X is the function $f(x) = P(X = x)$ on $\mathbb{R}$.
This function is encoded in the sequence $p_0, p_1, p_2, \ldots,$ where $p_k = P(X = x_k)$.

E.g. Consider the experiment of tossing 2 coins so $S = \{HH, HT, TH, TT\}$. Let $X =$ no. heads.

x_k	0	1	2
p_k			

Note any probability distribution satisfies: i) $p_k \geqslant 0$ for $k = 0, 1, 2, \ldots$, ii) $p_0 + p_1 + \cdots = 1$.

Example (Loaded Die)

X = toss of a loaded die with probability distribution $p_k = P(X = k) = \dfrac{\alpha}{k}$, for $k = 1, \ldots, 6$.

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

a) *Since* $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$, *where* $p_k = P(X = k)$, *we have*

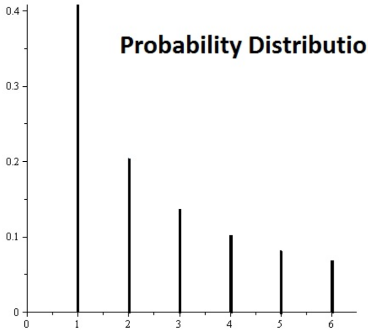
$$\alpha + \frac{\alpha}{2} + \frac{\alpha}{3} + \frac{\alpha}{4} + \frac{\alpha}{5} + \frac{\alpha}{6} = \frac{49\alpha}{20} = 1,$$

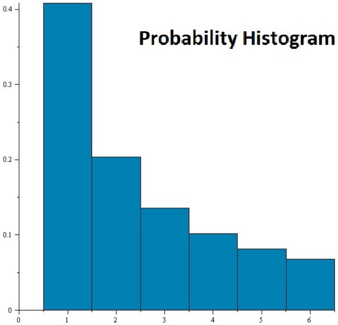
and hence $\alpha = \dfrac{20}{49}$.

Solution (Continued)

b) *The probability distribution for X:*

k	1	2	3	4	5	6
$P(X=k)$	$\frac{20}{49}$	$\frac{10}{49}$	$\frac{20}{147}$	$\frac{5}{49}$	$\frac{4}{49}$	$\frac{10}{147}$





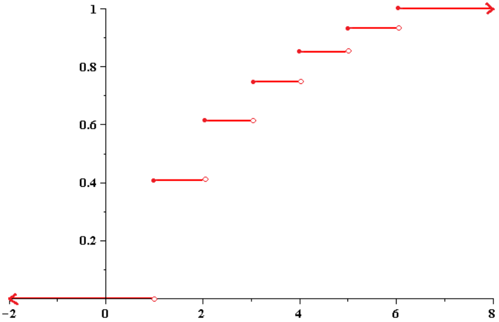
Solution (Continued)

c) *The cumulative distribution function of X: $F(x) = P(X \leqslant x)$.*

When $x < 1$, $F(x) = 0$; when $1 \leqslant x < 2$, $F(x) = P(X = 1) = \frac{20}{49}$;

when $2 \leqslant x < 3$, $F(x) = P(X = 1) + P(X = 2) = \frac{20}{49} + \frac{10}{49} = \frac{30}{49}$, etc.

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{20}{49} & \text{if } 1 \leqslant x < 2 \\ \frac{30}{49} & \text{if } 2 \leqslant x < 3 \\ \frac{110}{147} & \text{if } 3 \leqslant x < 4 \\ \frac{125}{147} & \text{if } 4 \leqslant x < 5 \\ \frac{137}{147} & \text{if } 5 \leqslant x < 6 \\ 1 & \text{if } x \geqslant 6 \end{cases}$$



Geometric Distribution

Definition

Fix the parameter $p \in (0,1)$. The **Geometric distribution** $G(p)$ is the function

$$G(p,k) := (1-p)^{k-1}p = p_k, \; k = 1,2,3,\dots .$$

If a discrete random variable X has probability distribution $P(X=k) = p_k$, then we write $X \sim G(p)$ & say X has **geometric distribution**.

Note The p_k can potentially define a probability distribution since all $p_k \geqslant 0$ &

$$\sum_{k\geqslant 1} p_k = p \sum_{j=k-1=0} (1-p)^j = p\frac{1}{1-(1-p)} = p\frac{1}{p} = 1.$$

This distribution occurs when you conduct an experiment with outcomes *success* with probability p, or *failure* and repeat until a success is reached. If X is the number of trials until the first success appears , then $X \sim G(p)$. See next slide for an example.

Example of a Geometric Distribution

Example

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

Solution

The possible values of X *are* 1, 2, 3, ...

$$\begin{aligned} P(X = k) &= P(\textit{not 6 for every toss in the first } (k-1) \textit{ times,} \\ &\qquad \textit{and 6 at the } k\textit{th toss}) \\ &= \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right) \\ &= \frac{5^{k-1}}{6^k} \end{aligned}$$

Hence $X \sim G(\frac{1}{6})$.

Sample statistics scenario

Example (Distribution of a Class Test Marks)

Marks of Algebra Test 1 of a Tutorial:

9	10	6	6	6	5	5	5	0	5
10	5	6	6	9	6	5	5	5	6

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											
Probability											

What is statistics?

Statistics is the science of handling data (e.g. test marks) with in-built variability or randomness. Statistics often organises and analyses data for various purposes such as the following.

- Present the data in more meaningful ways e.g. listing the frequencies or probabilities of marks as we did.
- Suppose in previous slide, you got one of the 9s, you might want to know how you fared compared to everyone else. One way is to compute the mean (here 6) or some other *measure of central tendency*.
- You might now ask, are you much better than average, or just a little? One way is to determine how spread out the marks are. In other words, we'd like some *measures of dispersion*.
- (Hypothesis testing) Suppose next year, the tutor changes his/her teaching method and the mean test score moves up to 6.5. Does this indicate effectiveness of the new teaching method, or can this be accounted for as a random fluctuation.

Mean or expected value

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

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

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

If X has n possible values $x_0, \ldots, x_{n-1}$ and has uniform distribution, then $p_k = \frac{1}{n}$ so $E(X) = \frac{1}{n}(x_0 + x_1 + \ldots + x_{n-1})$ is the usual mean or average.
E.g. In class test example p. 37, we find the mean is

$$\frac{1}{20} \times 0 + \frac{8}{20} \times 5 + \frac{7}{20} \times 6 + \frac{2}{20} \times 9 + \frac{2}{20} \times 10 = 6.$$

Variance & standard deviation

Let X = random var on sample space S & $g = \mathbb{R}$-valued function on $\mathbb{R}$.
We define the random variable $Y = g(X)$ to be the composite function $S \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$.

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

Definition (Variance and Standard Deviation)

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

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

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

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

These are both measures of dispersion or spread.

Alternate formula for variance

We normally use the following formula to calculate the variance.

Theorem

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

Proof.

Writing μ for $E(X)$, the definition of variance gives,

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_k (x_k - \mu)^2 p_k \\
&= \sum_k (x_k^2 - 2x_k\mu + \mu^2) p_k \\
&= \sum_k x_k^2 p_k - 2\mu \sum_k x_k p_k + \mu^2 \sum_k p_k \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - (E(X))^2 .
\end{aligned}$$

Calculating $\mu, \mathrm{Var}, \sigma$: Tossing 2 coins

Example

Let $X =$ the number of heads showing after 2 coins are tossed. Find mean, variance & standard deviation.

Solution

The probability distribution for X:

k	0	1	2
$P(X = k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Linear change of random variable

Theorem

Let X = discrete (or later continuous) random var with mean $E(X)$ & variance $\text{Var}(X)$. Consider linear change of variable $Y = aX + b$ for $a, b \in \mathbb{R}$. Then

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

Example (Tossing 2 coins from p. 42)

Jack gives 2 one-dollar coins to Jill. Jill will toss them and give back \$ 3 for every head showing. Let X = no. heads & $Y = 3X - 2$ be Jack's payoff function. Find $E(Y)$, $\text{Var}(Y)$, and $\text{SD}(Y)$.

The Binomial Distribution

A **Bernoulli trial** is an experiment with two outcomes labelled *Success* and *Failure*. Let $p =$ probability of success.

Consider random var $X =$ no. successes in n independent identical Bernoulli trials. The probability distribution of X is called a **binomial distribution**. Write $X \sim B(n, p)$.

Definition

The **Binomial distribution** $B(n, p)$ is the function

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

and 0 otherwise.

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

Mean & variance of binomial distribution

Theorem

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

Example

Hillary has a 17 poke balls. Each gives her a 20% chance of capturing any particular pokemon. She tries to capture as many as she can.

1. What's the expected number of pokemon she captures?
2. What is the probability that she captures 4?
3. What is the probability that she captures more than Donald's one pokemon.

Solution

Mean & variance of the geometric distribution

Theorem

If $X \sim G(p)$ *then*

$$E(X) = \frac{1}{p}, \quad Var(X) = \frac{1-p}{p^2}.$$

Proof for $E(X)$

We use the formula

$$\sum_{k \geqslant 1} kx^{k-1} = \frac{d}{dx} \sum_{k \geqslant 0} x^k = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}.$$

$E(X) =$

Example

A die is repeatedly rolled until a 6 appears.

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

b) What is the probability that the first 6 appears after the 5th roll?

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

Solution

Applications of the Binomial Distribution to Testing Claims

Example

The following experiment was conducted to test Mr Spock's alleged psychic abilities. A computer showed 4 cards to him and (randomly) picked one of them, and he guessed which card he thought the computer had picked. This process was repeated 288 times, and Spock managed to pick the right card 88 times (30.6 % of the trials).

Example

James T. Kirk did the same experiment. In 5 attempts, he picked the right card 2 times (40 % of the trials).

The expected number of correct guesses is only 25 % which may lead you to think that James & Spock are psychic.
We test the **Claim**: James & Spock are (not) psychic.

Test claim: Spock is not psychic

Q Can Spock's high guess rate be accounted for by chance?

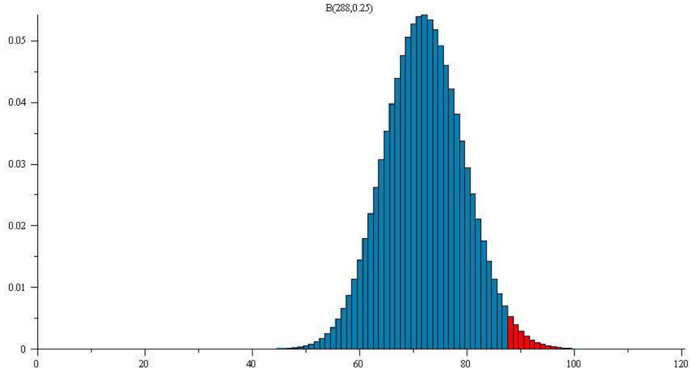
A Let $X =$ no. correct guesses in 288 trials.
We make the **hypothesis** that Spock is not psychic so $X \sim B(288, 0.25)$.
The tail probability that measures how unlikely it is that Spock guessed 88 times (or more) correctly in 288 trials is

$$P(X \geqslant 88) = \sum_{k=88}^{288} \binom{288}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{288-k} \approx 0.01902$$

which is less than 1 in 50.

This small probability suggests it's improbable the result is accounted for by random guessing as hypothesised. Instead we should *reject* the hypothesis & conclude the experiment gives evidence for Spock's psychic ability.

Tail probabilities



Is it clear why we call $P(X \geqslant 88)$ a tail probability?
In general, we call $P(X \geqslant t)$, $P(X > t)$ for $t > E(X)$, $P(X \leqslant t)$ for $t < E(X)$, and $P(|X - E(X)| > t)$ etc. **tail probabilities**.
In this course, if the tail probability is less than 5 % (1231 Algebra Notes p. 197) then we will regard this as significant.

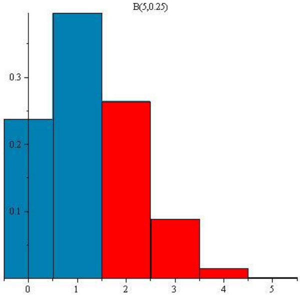
Test claim: James is not psychic

We now ask if James's high guess rate can be explained by chance.

Let X = no. correct guesses. Suppose James is not psychic so $X \sim B(5, 0.25)$.

Then $P(X \geqslant 2) =$

$$\binom{5}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3 + \binom{5}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + \binom{5}{4}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right) + \binom{5}{5}\left(\frac{1}{4}\right)^5 \approx 0.367.$$



This probability is no longer small. We see random guessing gives at least as good a strike rate, more than a third of the time.
James's good results can believably be accounted for by chance. The experiment does not give evidence supporting (or against) James's psychic ability.

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

- *We first replace each score with* $+$, 0, $-$, *when the reading is bigger than, the same, less than 96%, respectively.*
  This gives the following sequence with 12 plus signs

  $$+ \quad - \quad + \quad 0 \quad + \quad + \quad + \quad - \quad + \quad + \quad + \quad + \quad + \quad + \quad + .$$

Solution (Continued)

- *We* hypothesise *that there is no difference between the old system and the new system, i.e. + & - are equally likely, & calculate the tail probability of how improbable it is to get such readings under this hypothesis.*

  We ignore the zero (why?), so the number of plus signs, X, among the 14 measurements has a binomial distribution $B(14, 0.5)$. Hence the tail probability of getting as good as 12 plus signs is

$$\begin{aligned} & P(X \geqslant 12) \\ = \;& \binom{14}{12}\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^{2} + \binom{14}{13}\left(\frac{1}{2}\right)^{13}\left(\frac{1}{2}\right) + \binom{14}{14}\left(\frac{1}{2}\right)^{14} \\ \approx \;& 0.00647 \end{aligned}$$

Solution (Continued)

- *Conclusion.*

  Under the assumption that the new system is not better than the old one, the (tail) probability of getting 12 or more measurements higher than 96 % among 14 is approximately 0.65 % (less than 5 %). Hence there is evidence that the new system is better than the old one.

Continuous Random Variables

Unlike discrete random variables, probabilities of non-discrete random variables such as the height or weight of individuals cannot be determined from the probabilities $P(X = x)$, since these probabilities each equal 0. Instead, we will define probabilities for such random variables via 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** iff its cumulative distribution function $F_X(x)$ is continuous.

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

Example

The light rail goes up High St every 5 minutes. The Vice-Chancellor waits X minutes for the tram every afternoon. Suppose the cumulative distribution function for X is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{5} & \text{for } 0 \leqslant x \leqslant 5 \\ 1 & \text{for } x > 5 \end{cases}$$

What's the probability he waits more than 3 minutes?

Solution

Probability Density Function

For a discrete random variable X, the cumulative distribution function $F(x)$ is a sum over probability distribution values $p_k = P(X = x_k)$. For a continuous random variable, $F(x)$ is an integral of a continuous function analogue of a discrete probability distribution. The analogue is given in the following definition.

Definition

The **probability density function** $f(x)$ of a continuous random variable X is defined by

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

if $F(x)$ is differentiable, and $\lim_{x \to a^-} \frac{d}{dx}F(x)$ if $F(x)$ is not differentiable at $x = a$.

Example (Continued from the previous example)

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

Solution

Let the probability density function for X be $f(x)$. Besides $x = 0$, 5, we have

$$f(x) = F'(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{5} & \text{for } 0 < x < 5 \\ 0 & \text{for } x > 5 \end{cases}$$

By definition, $f(0) = \lim_{x \to 0^-} F'(x) = 0$.

Similarly, $f(5) = \lim_{x \to 5^-} F'(x) = \dfrac{1}{5}$.

Hence $f(x) = \begin{cases} 0 & \text{for } x \leqslant 0 \\ \frac{1}{5} & \text{for } 0 < x \leqslant 5 \\ 0 & \text{for } x > 5 \end{cases}$

Computing probabilities from probability density functions

Fundamental Theorem of Calculus $\implies F(x) = \int_{-\infty}^{x} f(t)dt.$

Hence for $a \leqslant b$,

Proposition

$$P(a \leqslant X \leqslant b) = P(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(x)dx$$

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

Suppose X is a random variable with probability density function

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

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

a) *Use the fact that* $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$, *and* $\displaystyle\int_{-\infty}^{\infty} f(x)dx = \int_1^{\infty} \frac{k}{x^4}dx$.

$$\lim_{R\to\infty}\left[-\frac{k}{3x^3}\right]_1^R = \frac{k}{3} = 1, \textit{ and hence } k = 3.$$

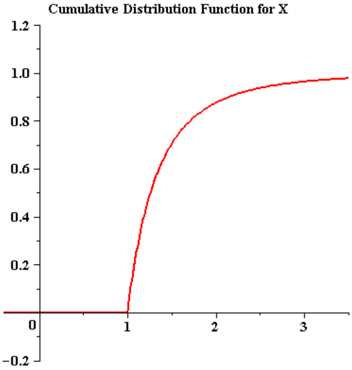
Solution (Continued)

b) $P(X \leqslant 3) = \int_{-\infty}^{3} f(x)dx = \int_{1}^{3} \frac{3}{x^4}dx = \left[-\frac{1}{x^3}\right]_1^3 = 1 - \frac{1}{27} = \frac{26}{27}$

$$P(2 \leqslant X \leqslant 3) = \int_{2}^{3} \frac{3}{x^4}dx = \left[-\frac{1}{x^3}\right]_2^3 = \frac{1}{8} - \frac{1}{27} = \frac{19}{216}$$

c)

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,dt \\ &= \begin{cases} 1 - \dfrac{1}{x^3} & \text{for } x > 1 \\ 0 & \text{for } x \leqslant 1 \end{cases} \end{aligned}$$



Mean and Variance of a Continuous Random Variable

The mean and variance of a continuous random variable can be obtained from those of a discrete random variable 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

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

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

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

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

Example

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

Solution

$$E(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{1}^{\infty} x\, \frac{3}{x^4} dx = \lim_{R\to\infty} \left[-\frac{3}{2x^2} \right]_1^R = \frac{3}{2}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{1}^{\infty} \frac{3}{x^2} dx = \lim_{R\to\infty} \left[-\frac{3}{x} \right]_1^R = 3$$

$$Var(X) = E(X^2) - (E(X))^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}, \textit{ hence } SD(X) = \frac{\sqrt{3}}{2}.$$

Standardised random variable

Theorem

If $E(X) = \mu$ *and* $SD(X) = \sigma$, *and* $Z = \dfrac{X - \mu}{\sigma}$, *then*

$$E(Z) = 0 \quad \textit{and} \quad Var(Z) = 1.$$

Proof.

$$E(Z) = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$\text{Var}(Z) = \text{Var}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1$$

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

The most important continuous distribution 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

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

Try the following applet.
`http://www.intmath.com/counting-probability/normal-distribution-graph-interactive.php`

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 $Z \sim N(0, 1)$.
- $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}$ does not have an elementary primitive function, so one cannot compute normal probabilities exactly using the fundamental theorem of calculus to integrate.

Computing normal probabilities from tables

To evaluate probabilities associated with a normal random variable $X \sim N(\mu, \sigma^2)$:
we standardise X to $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$. This gives

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

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.

Remember you can picture the value $P(Z \leqslant z)$ as the area to the left of z under the probability density curve of $Z \sim N(0, 1)$.

Calculating standard normal probabilities from tables

Example

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

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

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

Solution

Example: calculating non-standard normal probabilities

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 the probability that it is greater than 50.

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

Solution

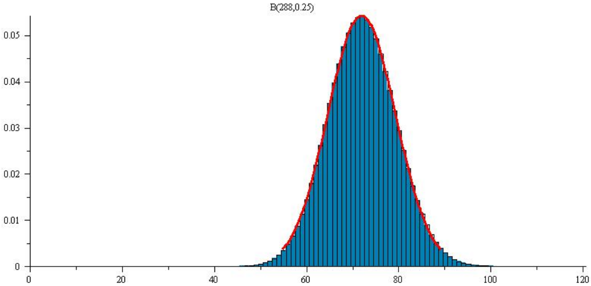
Solution (Continued)

Smoothing out the binomial distribution

Consider discrete random variable $X \sim B(288, .25)$. Probabilities

$$P(a \leqslant X \leqslant b) = \sum_{k=a}^{b} \binom{288}{k} .25^k .75^{288-k}$$

may involve large difficult sums. Alternately, we might view this sum as an area under the probability histogram (in blue) below.



Normal Approximation to Binomial

The histogram looks like the bell curve we see for the normal distribution. In fact, the red curve is the probability density for a normal variable $Y \sim N(72, 54)$.
Areas under the probability histogram in blue, can thus be approximated with areas under the red normal distribution curve.

To make this idea precise, let $X \sim B(n, p)$ be any discrete random variable with binomial distribution.

"Theorem"

Let Y be a continuous random variable with $Y \sim N(np, np(1-p))$ i.e. the same mean and standard deviation as X. Then for any integer x

- $P(X \leqslant x) \approx P(Y \leqslant x + .5)$,
- $P(X < x) \approx P(Y \leqslant x - .5)$.

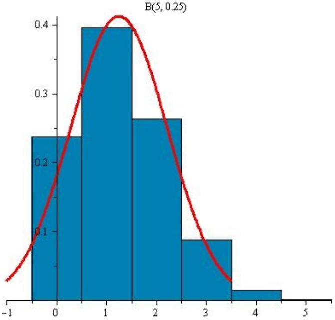
As a rule of thumb, the approximations are usually good enough if both $np > 10$ and $n(1-p) > 10$.

Why is there a "$\pm.5$" in the formulas?

When $n = 5$, $p = 0.25$,

$$\begin{aligned} np &= 1.25 \\ np(1-p) &= 0.9375 \end{aligned}$$

Formula approximates
$X \sim B(5, 0.25)$ with
$Y \sim N(1.25, 0.9375)$.
The red curve is $N(1.25, 0.9375)$.



To calculate a probability like $P(X \leqslant 1) = P(X = 0) + P(X = 1)$, need the area of the first two rectangles which goes up to $x = 1 + .5 = 1.5$. However, to calculate $P(X < 1)$, only need the area of the first rectangle which only goes up to $x = 1 - .5 = .5$.

Example of normal approximation to binomial

Example

In a poll, 53 % of the people in a random sample of 400 support Labor.

a) Suppose it is not true that more people support Labor i.e. the probability that a randomly chosen person supports Labor is 0.5. What is the probability that there are 53 % or more people in a sample of 400 supporting Labor?

b) Is there evidence that more people in Australia support Labor?

Solution

Solution (Continued)