EMS702P Statistical Thinking and Applied Machine Learning

Week 1.2 – Sets and Probability

Jun Chen

Introduction to Statistics and Machine Learning

Edition: v1.1

Table of Contents

1 Sets

If we can identify a property which is common to several objects, it is often useful to group them together. Such a grouping is called a **set**.

1.1 Terminologies

Set: A collection of objects. For Example,

A = {the resistors produced in a factory on a particular day}

B = {all odd numbers}

D = {on, off}

E = {0, 1, 2, 3, 4, 5, 6}

F = {3, 1, 2, 0, 4, 5, 6}

G = {1, 2, 3}

The elements of a set can be described in words if they cannot be listed or listed easily, such as *A* and *B*. Two sets are equal if they contain exactly the same elements, such as *E* and *F*.

Finite Set: A set with a finite number of elements, such as *A*, *D*, *E, F*.

Infinite Set: Otherwise, such as *B*.

Subset: A set contained completely within another set, such as $G \subseteq E = F$.

$\in$ ($\notin$)**:** An element is (is not) a member of a set, e.g. $4 \in E$ but $4 \notin G$.

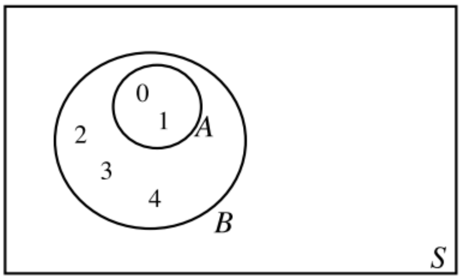
Universal Set S: the set containing all the objects of interest in a particular situation, e.g. $S = B$ if we are only interested in odd numbers.

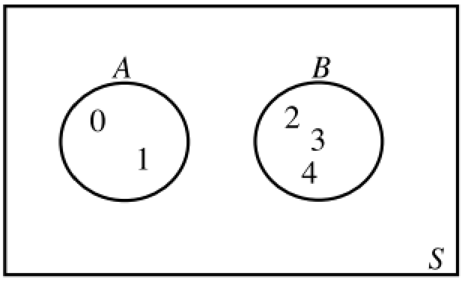
Complement: The complement of *G*, denoted by $G^{'}$ contains all the elements of the universal set that are not in *G*, e.g. if $S = F$, then $G^{'} = \{0, 4, 5, 6\}$.

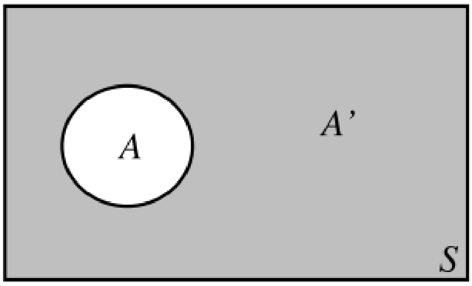
Empty Set ϕ: A set that contains no members, e.g. if $S = F$, then $F^{'} = \phi$.

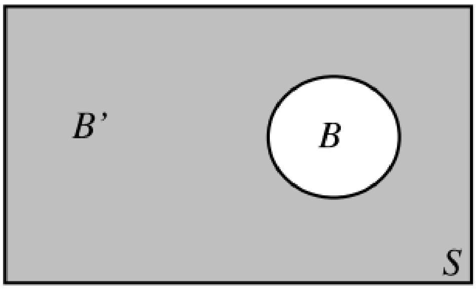
1.2 Venn Diagrams

Sets are often represented graphically by **Venn Diagrams**.



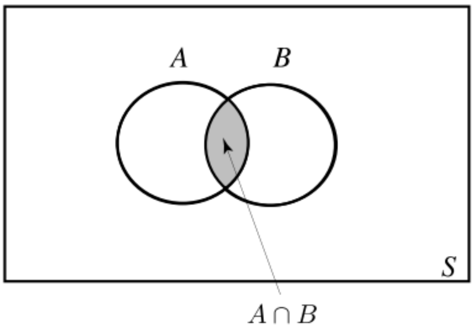






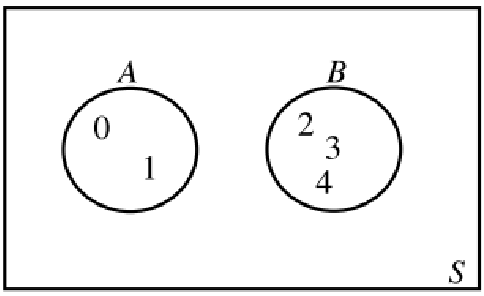
1.3 The Intersection and Union of Sets

Intersection of Sets: $A \cap B = B \cap A = \{x: x \in A$ **and** $x \in B\}$.



Example 1

From the Venn diagram below, state $A \cap B$



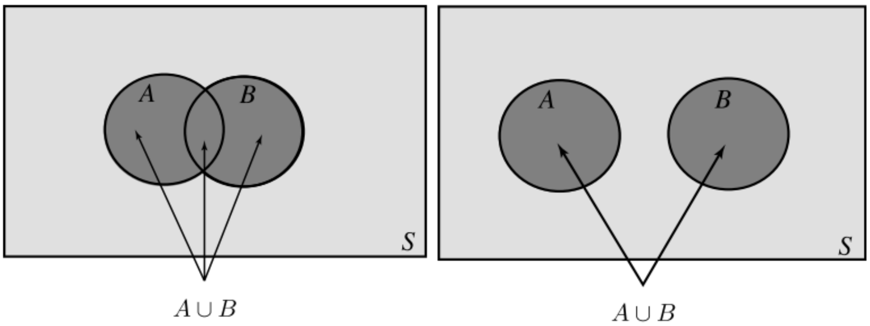
Solution

A and *B* are **disjoint** sets, which are represented by separate area regions in the Venn diagram.

The rule of the intersection for multiple (>3) sets

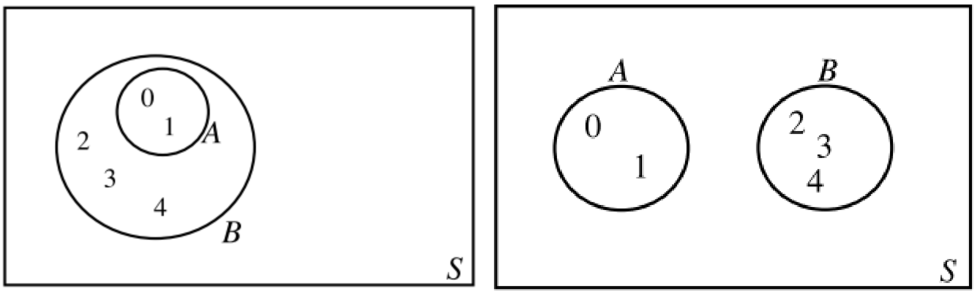
$$(A \cap B) \cap C = A \cap (B \cap C)$$

The Union of Sets: $A \cup B = B \cup A = \{x: x \in A$ **or** $x \in B\}$.



Example 2

From the Venn diagram below, state $A \cup B$



Solution

There is **no** need to repeat elements in a set.

2 Elementary Probability

Probability is about the study of uncertainty and provides us the tools to set up mathematical models of systems and processes which are affected by random occurrences or 'chance'.

There are many sources of randomness in applied machine learning. Randomness is used as a tool to help the learning algorithms be more robust and ultimately result in better predictions and more accurate models. Sources of randomness in machine learning include

- Randomness in Data

  There is a random element to the sample of data.

  Data may have mistakes or errors.

  Data contain noise.
- Randomness in Algorithms

  Randomness allows the algorithm to achieve a better performing mapping of the data than if randomness was not used.

  Randomness allows an algorithm to attempt to avoid overfitting the small training set and generalize to the broader problem.
- Randomness in Evaluation

  We work with only a small sample of the data.

  We fit and evaluate the model on different subsets of the available dataset to infer how the model works on average rather than on a specific set of data.

Examples of randomness that affects machine learning algorithms and has been exploited by machine learning algorithms include

- The shuffling of training data prior to each training epoch in stochastic gradient descent (introduced in W5.3).
- The random subset of input features chosen for split points in a random forest algorithm (introduced in W11.3).
- The random initial weights in an artificial neural network (introduced in W5.3).

Therefore, we must find a way to keep randomness under control, so as to limit its impact of e.g. noise in the data, to evaluate model skills, and to improve algorithms. Probability theory lay down a foundation for machine learning practitioners to do this.

Before we define probability, let's recap those terminologies in W1.1 through examples.

Example 3

Obtain the **sample space** of the experiment throwing a single coin.

Solution

In this particular case, the order of the **outcomes** is unimportant like for sets in general. Heads or Tails are equally likely to occur. Therefore, for a fair coin, H and T are **equally likely outcomes**.

The sample space obtained in Example 3 is an **equi-probable space** as the chance that any one sample point in it occurs is equal to the chance that any other sample points occurs.

Discrete Sample Space: The sample space in the form of a list (possibly infinite) like in Example 3.

Continuous Sample Space: The sample space that is not possible to list.

Example 4

Obtain the **sample space** of the experiment measuring the tensile strength of small gauge steel wire.

Solution

Example 5

Two fair coins are thrown. List the ordered outcomes for the event when just one Tail is obtained.

Solution

Here the order does matter unlike for sets in general. The event is a **subset** of the sample space. Unlike sets, events can be ordered when it is explicitly stated or within a certain context.

Complementary Event: The set of outcomes which are not members of the event from the sample space.

Example 6

Obtain the complement of the event in Example 5.

Solution

2.1 The Definition of Probability

The **Principle** of Equally Likely Outcomes enables us to deduce the probabilities that **simple** events (and hence more complicated events which are combinations of simple events) occur.

If a sample space *S* consists of *n* simple outcomes which are equally likely and an event *A* consists of *m* of those simple outcomes, then

$$P(A) = \frac{m}{n}, 0 \leq P(A) \leq 1$$

- $P(A) = 1$: The event A is certain because A is identical to S.
- $P(A) = 0$: The event A is impossible because A is empty, i.e. $A = \phi$.

The Complement Rule: The probability of the complement of A occurring is equal to 1 minus the probability of A occurring.

$$P(A') = 1 - P(A)$$

Example 7

Find the probability of obtaining a total score of at least five when three dice are thrown.

Solution

The use of the event A' can sometimes simplify the calculation of the probability $P(A)$.

2.2 Laws of Elementary Probability

When we require the probability of two events occurring simultaneously or the probability of one or the other or both of two events occurring then we need probability laws to carry out the calculations.

Events, like sets, can be combined to produce new events.

- $A \cup B$ – event A or event B (or both) occur when the experiment is performed.
- $A \cap B$ – both A and B occur together.

There are four types of events. We will define the following two types as the other two types are simply the opposite of the following two event.

Mutually exclusive events are events that by definition cannot happen together. The probability of mutually exclusive events occurring together is

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

If the occurrence of one event A does not affect, nor is affected by, the occurrence of another event B then A and B are **independent events**.

Example 8

Two box contains 20 nuts each, some have a metric thread, some have a British Standard Fine (BSF) threads and some have a British Standard Whitworth (BSW) thread. A nut is picked out of each box. Decide if A and B are mutually exclusive.

$A = \{nut\ picked\ out\ of\ the\ first\ box\ is\ BSF\}$

$B = \{nut\ picked\ out\ of\ the\ second\ box\ is\ metric\ \}$

Solution

Example 9

A box contains 20 nuts, some have a metric thread, some have a British Standard Fine (BSF) threads and some have British Standard Whitworth (BSW) thread. Two nuts are picked out of the box. Decide if A and B are independent.

$A = \{first\ nut\ picked\ out\ of\ the\ box\ is\ BSF\}$

$B = \{second\ nut\ picked\ out\ of\ the\ box\ is\ metric\ \}$

Solution

The Addition Law of Probability:

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

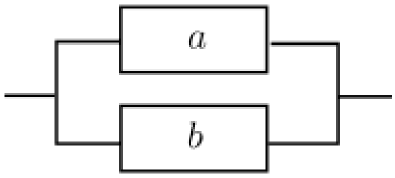
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

For a sample space S consist of the n **simple distinct** events $E1, E2 \ldots En$, we have

$$P(E1) + P(E2) + \cdots + P(En) = 1$$

Example 10

The diagram shows a simplified circuit in which two independent components a and b are connected in parallel.



The circuit functions if either or both of the components are operational. It is known that if A is the event 'component a is operating' and B is the event 'component b is operating' then $P(A) = 0.99$, $P(B) = 0.98$ and $P(A \cap B) = 0.9702$. Find the probability that the circuit is functioning.

Solution

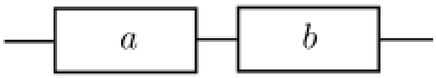
In this case, the probability that the circuit functions is greater than the probability that either of the individual components functions.

The Multiplication Law of Two Independent Events:

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

Example 11

Two components a and b are connected in series as shown below.



Define two events

- A is the event 'component a is operating'
- B is the event 'component b is operating'

Previous testing has indicated that $P(A) = 0.99$, and $P(B) = 0.98$. The circuit functions only if a and b are both operating simultaneously. What is the probability that the circuit is operating?

Solution

What if the Event A and B are not independent? Let's revisit Example 9.

Example 12

A box contains 20 nuts, 5 have a metric thread, 5 have a British Standard Fine (BSF) threads and 10 have British Standard Whitworth (BSW) thread. Two nuts are picked out of the box.

$A = \{first\ nut\ picked\ out\ of\ the\ box\ is\ BSF\}$

$B = \{second\ nut\ picked\ out\ of\ the\ box\ is\ metric\ \}$

What is the probability of $P(A \cap B)$?

Solution

In general, the multiplication law of probability needs to be revised to

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

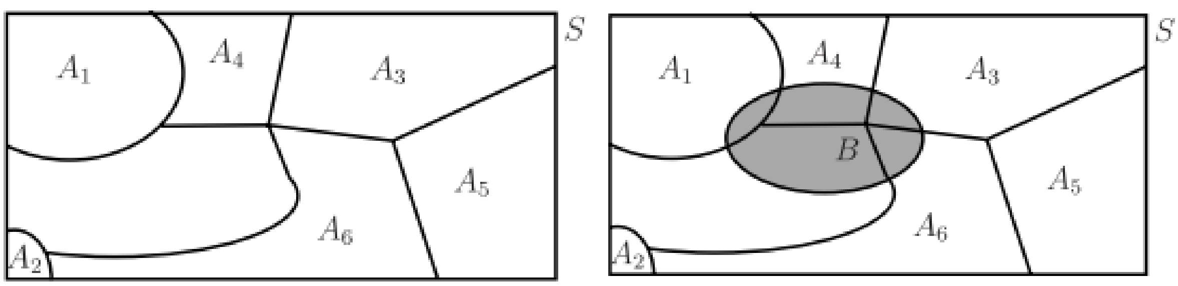
Where $P(B|A)$ is the **conditional probability** of an event B occurring given that event A has occurred, and can be obtained by rearranging the above equation into

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{number\ of\ outcomes\ in\ A\ \cap\ B}{number\ of\ outcomes\ in\ A}$$

In a sense we have a new sample space which is firstly defined by the event A. For B to occur some of its members must also be members of event A. So, for example, in an equi-probable space, $P(B|A)$ must be the number of outcomes in $A\ \cap\ B$ divided by the number of outcomes in A.

2.3 Total Probability and Bayes' Theorem

The Total Probability theorem enables us to calculate the probability of the event B, i.e. $P(B)$.



If B is any event within S, then B can be expressed as the union of subsets:

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$$

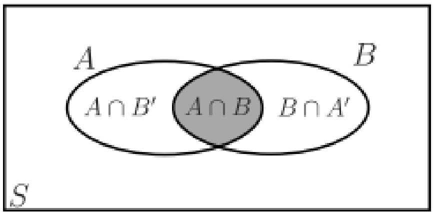
$(B \cap A_1), (B \cap A_2) \ldots (B \cap A_n)$ are mutually exclusive. Using the addition law of probability for mutually exclusive events, we have

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$$

Expressing the above equation in terms of conditional probabilities gives

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)$$

A special case of the Total Probability theorem considers A to be an event in a sample space S (the sample space is partitioned by A and A') as shown below.



$$P(B) = P(A \cap B) + P(B \cap A')$$

With the conditional probability, $P(B)$ can be rewritten as

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

It is easy to see that the probability of $P(B)$ in this special case is a **simplified** version of the theorem of Total Probability theorem.

The Bayes' theorem enables us to calculate the probability of the event $A_i, i = 1 \ldots n$, given the event B.

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

$$= \frac{P(A \cap B)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}$$

$$= \frac{P(B|A)\, P(A)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}$$

This is true for any event A and so, replacing A by A_1 gives the Bayes' theorem.

$$P(A_i|B) = \frac{P(B|A_i)\,P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P\,(B|\,A_n)P\,(A_n)}$$

Example 13

A factory production line is manufacturing bolts using three machines, A, B and C. Of the total output, machine A is responsible for 25%, machine B for 35% and machine C for the rest. It is known from previous experience with the machines that 5% of the output from machine A is defective, 4% from machine B and 2% from machine C. A bolt is chosen at random from the production line and found to be defective. What is the probability that it came from machine A?

Solution

3 Probability Distributions

A **random variable** X is a quantity whose value cannot be predicted with certainty. Variables may be:

Quantitative: i.e. numerical, e.g. numbers of vehicles.

Qualitative: usually categorical, e.g. make of a car, voting intention of a voter, cause of failure in a machine.

Quantitative variables may be

Continuous: i.e. measured on a continuous scale. E.g. height of a student, breaking strength of a cord.

Discrete: i.e. only certain separate values, e.g. integers, are possible. E.g. number of right-turning vehicles, goal difference of a football team.

Distributions are often described in terms of their **mass** or **density functions**. Mass or Density functions are functions that describe how the proportion of data or likelihood of the proportion of observations change over an **ordered list** (discrete) or the range (continuous) of the distribution.

In practice, we are interested in discrete random variables and their distributions, and continuous random variables and their distributions. Discrete distributions arise from experiments involving counting, for example, road deaths, car production and aircraft sales, while continuous distributions arise from experiments involving measurement, for example, voltage, corrosion and oil pressure.

3.1 Discrete vs. Continuous Probability Distributions

A random variable X and its distribution are said to be **discrete** if the values of X can be presented as an **ordered list** say $x_1, x_2, x_3, \ldots$ with probability values

$p_1, p_2, p_3, \ldots$. For example, the number of times a particular machine fails during the course of one calendar year is a discrete random variable.

Probability Mass Function (PMF) for a discrete random variable is defined as

$$P(X = x_i) = p_i$$

Two necessary conditions for a valid PMF are below

- $P(X = x_i) \geq 0$
- $\sum_{i=1}^{n} P(X = x_i) = 1, where\ n\ can\ be\ infinite$

Cumulative Distribution Function (CDF) for a discrete random variable is defined as the summation of p_i for which x_i is less than or equal to x.

$$\sum_{x_i \leq x} P(X = x_i)$$

Example 14

A traffic engineer is interested in the number of vehicles reaching a particular crossroads during periods of relatively low traffic flow. The engineer finds that the number of vehicles X reaching the crossroads per minute is governed by the probability distribution:

x	**0**	**1**	**2**	**3**	**4**
$P(X = x)$	0.37	0.39	0.19	0.04	0.01

Graph the PMF and the corresponding CDF.

Solution

The expectation of a Discrete Random Variable: The mean or expected value or expectation of X, which is written $E(X)$ is defined as

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) \triangleq \mu$$

The expectation $E(X)$ of X is the value of X which we expect on average. We can think of $P(X = x_i)$ as the weight for x_i to appear in the sample space, i.e. the population.

The variance of a Discrete Random Variable: The variance of X, which is written $V(X)$ is defined by

$$V(X) = \sum_{i=1}^{n} p_i (X - \mu)^2 \triangleq \sigma^2$$

Where, $\sigma = \sqrt{V(X)}$ is the **standard deviation** of X.

Example 15

Prove $V(X) = E(X^2) - [E(X)]^2$

Solution

Example 16

Calculate the expected value, the variance and the standard deviation of the random variable X in Example 14.

Solution

A random variable X is said to be **continuous** if it can assume any value in a given **interval**. This contrasts with the definition of a discrete random variable which can only assume discrete values.

For a continuous random variable, we can never determine the probability if the random variable assume a particular value as there are always small variations so as to deviate from that particular value. Instead, we can only calculate the probability that a continuous random variable lies within a given range of values. A continuous random variable is characterised, not by probabilities of the type $P(X = c)$ (as was the case with a discrete random variable), but by a function $f(x)$ called the **Probability Density Function** (PDF).

Similar to discrete random variables, two necessary conditions for a valid PDF are below

- $f(x) \geq 0 \qquad for\ all\ x$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

Analogous to the formula for the CDF for discrete random variables, **CDF for continuous random variables** is defined by replacing the summation with the integral.

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

The probability of a continuous variable within an interval is calculated by

$$P(a < X < b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$$

The expectation of a Continuous Random Variable is defined as

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

The variance of a Continuous Random Variable is defined as

$$V(X) = \int_{-\infty}^{\infty} f(x)(X - \mu)^2 dx \triangleq \sigma^2$$

Example 17

For the variable X with PDF

$$f(x) = \begin{cases} \frac{1}{2}x, & 0 \leq x \leq 2 \\ 0, & elsewhere \end{cases}$$

find $E(X)$ and then $V(X)$.

Solution

3.2 Binomial Distribution

When an experiment (or trial) is repeated a fixed number of times, under certain assumptions, it can be modelled by the binomial distribution. Within each trial we focus attention on a particular outcome. If the outcome occurs, we label this as **a success**. The binomial distribution allows us to calculate the probability of observing a certain number of successes in a given number of trials. The Binomial Distribution is one of the commonly used discrete distributions.

For the binomial model to be applied the following **four criteria** must be satisfied:

- the trial is carried out a fixed number of times n
- the outcomes of each trial can be classified into two 'types' conventionally named 'success' or 'failure'
- the probability p of success remains constant for each trial, likewise for $q = 1 - p$ for failure
- the individual trials are independent of each other.

The name of Binomial Distribution is given since the probabilities of observing a certain number of successes in a given number of trials are in fact the terms which arise in binomial expansion of $(p + q)^n$. E.g. if $n = 2$, i.e. 2 trials, we have the following

$$P(2\ successes\ in\ 2\ trials) = p^2$$

$$P(1\ successes\ in\ 2\ trials) = 2pq$$

$$P(0\ successes\ in\ 2\ trials) = q^2$$

Let X be a discrete random variable, being the number of successes occurring in n independent trials of an experiment. If X is to be described by the binomial model, the probability of exactly r successes in n trials is given by

$$P(X = r) = {}^{n}C_r p^r q^{n-r}$$

Here there are r successes (each with probability p), $n - r$ failures (each with probability q) and

$$^{n}C_r = \frac{n!}{r!\,(n-r)!}$$

is the number of ways of taking r objects among the n trials.

If a random variable X follows a binomial distribution in which an experiment is repeated n times each with probability p of success then we write $X \sim B(n, p)$. Without proof, the **expectation** and **variance** of the Binomial distribution are given by the formulae

$$E(X) = np$$

$$V(X) = np(1 - p) = npq$$

Example 18

In a box of switches it is known 10% of the switches are faulty. A technician is wiring 30 circuits, each of which needs one switch. What is the probability that (a) all 30 work, (b) at most 2 of the circuits do not work? (c) Find $E(X)$ and $V(X)$ where the switches work?

Solution

3.3 The Poisson Distribution

The **Poisson** distribution is a probability model which can be used to find the probability of a single event occurring a given number of times in an interval of (usually) time. It is worth noting that only the occurrence of an event can be counted; the non-occurrence of an event cannot be counted. This contrasts with the Binomial distribution where we know the number of trials, the number of events occurring and therefore the number of events not occurring.

If X is the random variable 'number of occurrences in a given interval' for which the average rate of occurrence is λ then, according to the Poisson model, the probability of r occurrences in that interval is given by

$$P(X = r) = e^{-\lambda}\frac{\lambda^r}{r!} \qquad r = 0,1,2,3,\ldots$$

We also have

$$P(X = r) = \frac{\lambda}{r}P(X = r - 1) \qquad for\, r \geq 1$$

Example 19

A Council is considering whether to base a recovery vehicle on a stretch of road to help clear incidents as quickly as possible. The road concerned carries over 5000 vehicles during the peak rush hour period. Records show that, on average, the number of incidents during the morning rush hour is 5. The Council won't base a vehicle on the road if the probability of having more than 5 incidents in any one morning is less than 30%. Based on this information should the Council provide a vehicle?

Solution

One application of the Poisson distribution is to approximate the probability of the outcome $X = r$ of the Binomial distribution subject to the following conditions.

1) n is large
2) p is small
3) $np = \lambda$

Example 20

A worn machine which is known to produce 1% defective components is used for a production run of 40 components. Use both the Binomial distribution and its Poisson approximation to calculate the probability that two defective items are produced for comparison.

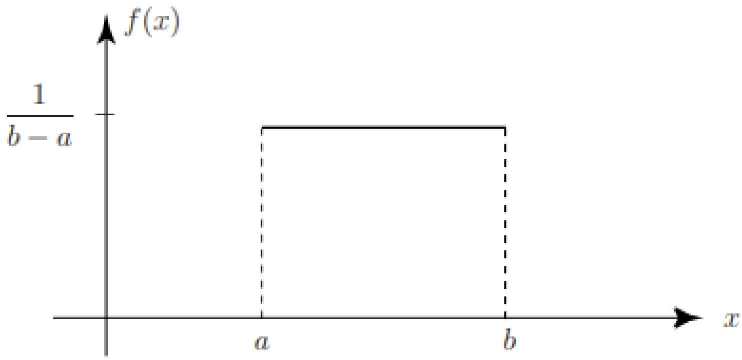
Solution

Recalling that the expectation and variance of the Binomial distribution are given by the results $E(X) = np$ and $V(X) = np(1-p) = npq$. When n is large and p is small and $np = \lambda$, we have the **expectation** and **variance** for the Poisson distribution as below.

$$E(X) = np = \lambda$$
$$V(X) = np(1-p) \approx np = \lambda$$

3.4 The Uniform Distribution

The **Uniform** or **Rectangular** distribution has random variable X restricted to a finite interval $[a, b]$ and has $PDF = f(x)$ a constant over the interval. An illustration is shown below.



The function $f(x)$ has the following form

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & elsewhere \end{cases}$$

The **expectation** and **variance** of a Uniform distribution can be found using the formula defined earlier in Section 3.1.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{b} x\frac{1}{b-a}dx = \frac{1}{2(b-a)}[x^2]_a^b = \frac{b+a}{2}$$

$$V(X) = E(X^2) - [E(X)]^2 = \int_{a}^{b} x^2\frac{1}{b-a}dx - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}$$

Example 21

The current (in mA) measured in a piece of copper wire is known to follow a uniform distribution over the interval $[0, 25]$. Write down the formula for the probability density function $f(x)$ of the random variable X representing the current. Calculate the mean and variance of the distribution and find the cumulative distribution function $F(x)$.

Solution

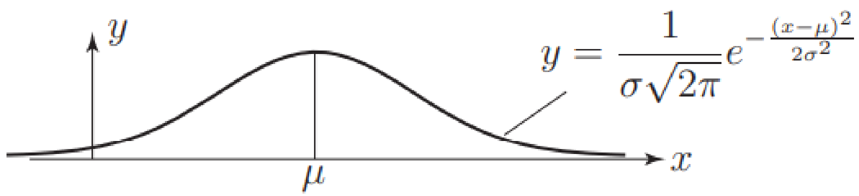
3.5 The Normal Distribution

The **Gaussian** distribution, named for Carl Friedrich Gauss, is the focus of much of the field of statistics. Data from many fields of study surprisingly can be described using a Gaussian distribution, so much so that the distribution is often called the **Normal** distribution because it is so common.

A Normal distribution has mean μ and variance σ^2. A random variable X following this distribution is usually denoted by $N(\mu, \sigma^2)$ and we often write $X \sim N(\mu, \sigma^2)$. The PDF of a Normal distribution with mean μ and variance σ^2 is given by the formula

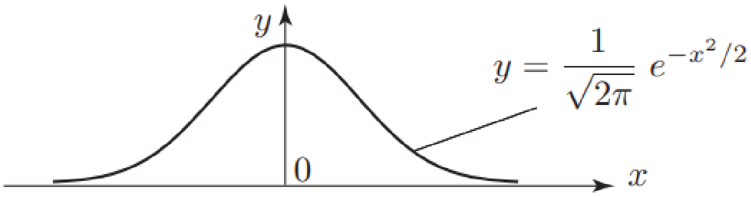
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The height of the bell is controlled by the value of σ. As with all normal distribution curves it is symmetrical about the centre and decays as $x \to \pm\infty$. As with any probability density function the area under the curve is equal to 1.



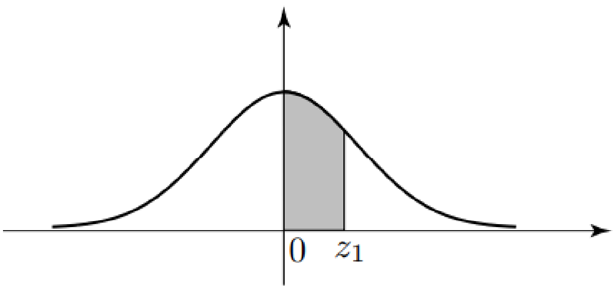
Due to the integral used in the PDF of the Normal distribution, calculating the probability of an interval (i.e. areas under the normal curve) is intractable. We can instead tabulate the distribution to provide information of the probability of an interested interval. Clearly, since μ and σ^2 can both vary, there are infinitely many Normal distributions and it is impossible to give tabulated information concerning them all.

We can however, tabulate only on – the **Standard Normal Distribution** – and convert all problems involving the normal distribution into problems involving the Standard Normal Distribution. The Standard Normal Distribution has a mean of **zero** and a variance of **one**.



If the behaviour of a continuous random variable X is described by the distribution $N(\mu, \sigma^2)$ then the behaviour of the random variable $Z = \frac{(X-\mu)}{\sigma}$ is described by the standard normal distribution $N(0, 1)$. We call Z the **standardised normal variable** and we write $Z \sim N(0, 1)$.

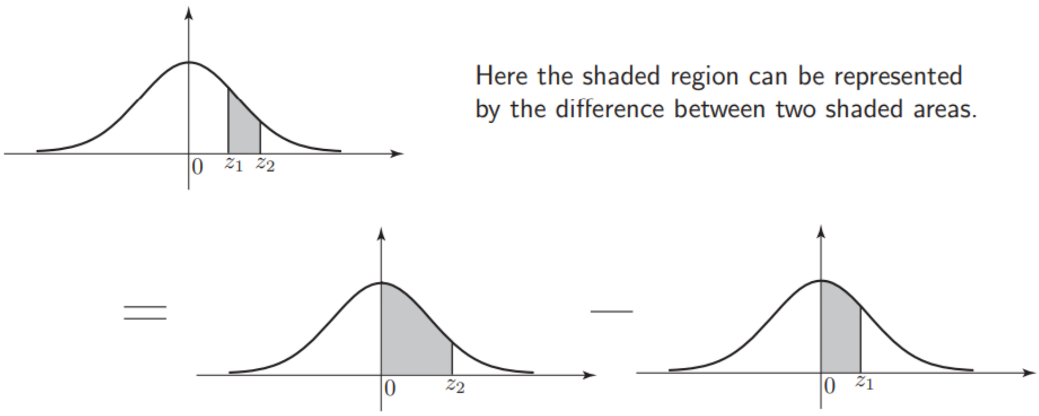
Since the Standard Normal distribution is used so frequently a table of values has been produced to help us calculate probabilities (See **Appendix**). It is based upon the following diagram.



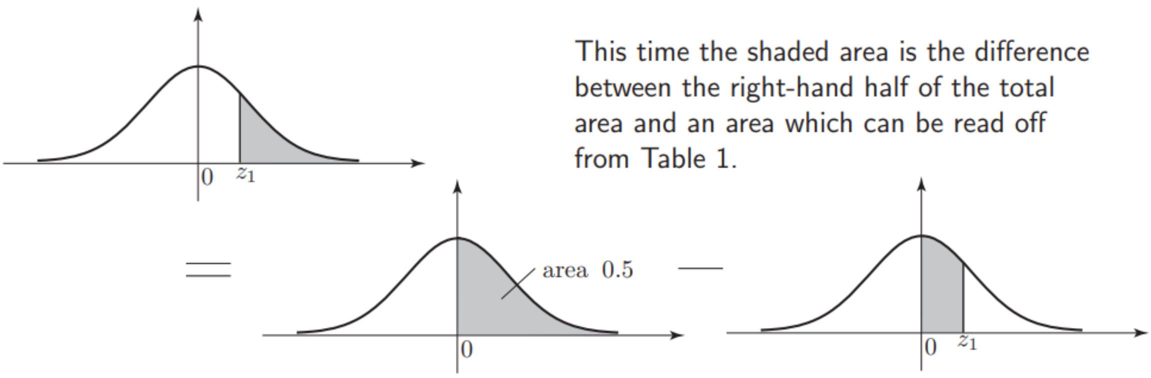
Since the total area under the curve is equal to 1 it follows from the symmetry in the curve that the area under the curve in the region $Z > 0$ is equal to 0.5. The shaded area is the probability that Z takes values between 0 and z_1. When we 'look-up' a value in the table we obtain the value of the shaded area.

The table enables us to calculate the probabilities of **any** given intervals of **any** normal distributions. Before we do that, let's look at 5 cases that we may encounter when calculating the probabilities of any given intervals.

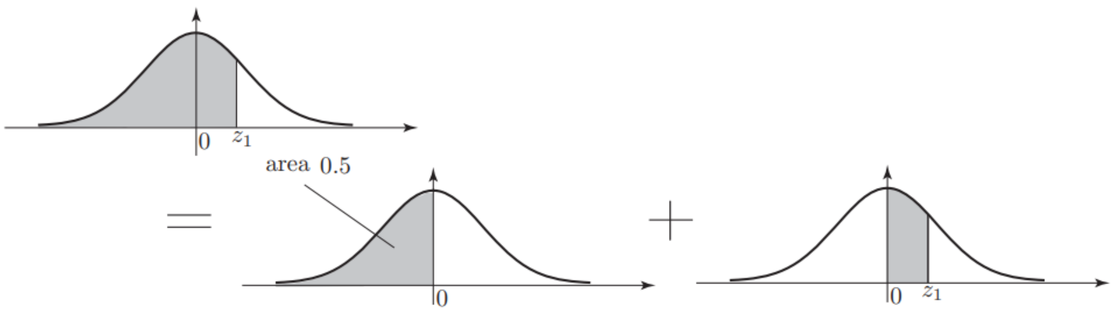
Case 1



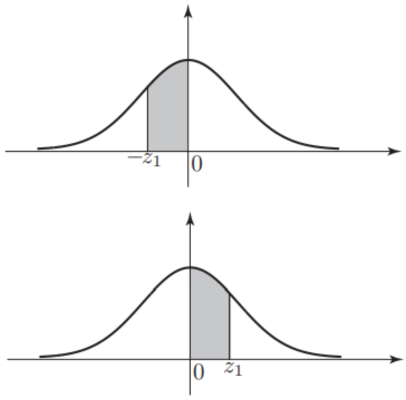
Case 2



Case 3

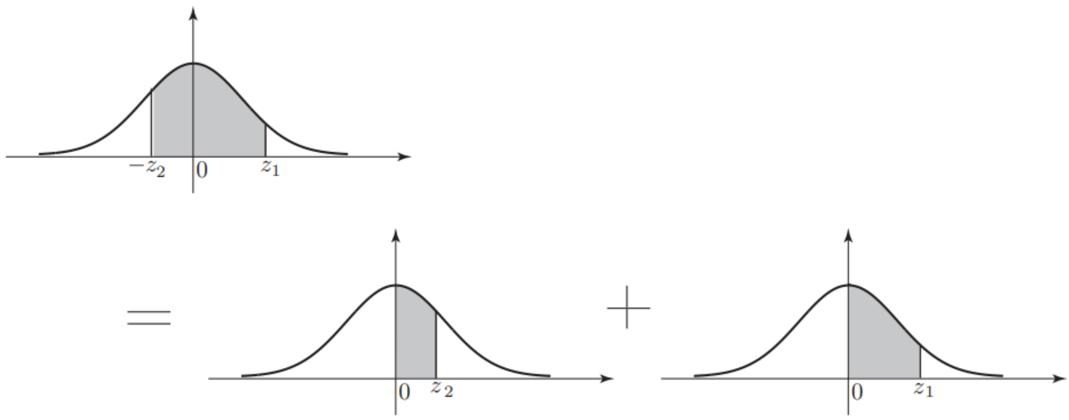


Case 4



By symmetry this shaded area is equal in value to the one above.

Case 5



For any continuous random variable X, we can only calculate the probability that

- X lies between two given values;
- X is greater than a given value;
- X is less than a given value.

We cannot calculate the probability for individual values.

Example 22

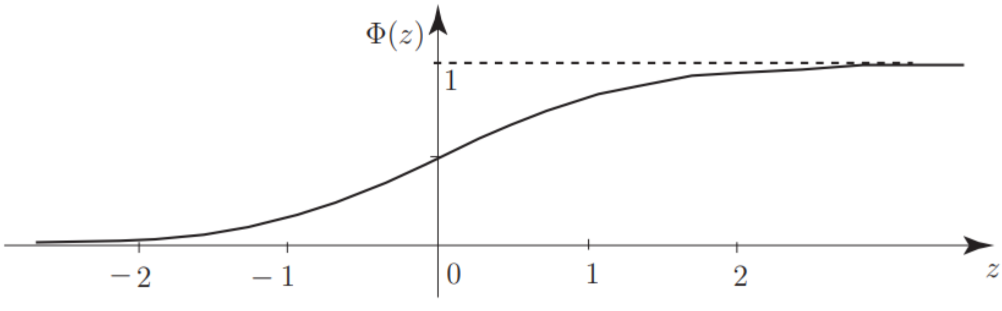
Piston rings are mass-produced. The target internal diameter is 45 mm but records show that the diameters are normally distributed with mean 45 mm and standard deviation 0.05 mm. An acceptable diameter is one within the range

44.95 mm to 45.05 mm. What proportion of the output is unacceptable? If the standard deviation is halved by improved production practices what is now the proportion of unacceptable items?

Solution

In the case of the cumulative distribution for the standard normal curve, we have the following for the CDF.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-u^2/2} du$$



$$P(a < X < b) = F(b) - F(a) = \Phi(\frac{b - \mu}{\sigma}) - \Phi(\frac{a - \mu}{\sigma})$$

Where, $F(x)$ is the CDF of any given Normal distribution.

We have already seen that the Poisson distribution can be used to approximate the Binomial distribution for large values of n and small values of p provided that the correct conditions exist. The approximation is only of practical use if just a few terms of the Poisson distribution need be calculated. This can be avoided by using the Normal distribution to approximate the Binomial distribution.

In order to do this, the following **conditions** need to be satisfied.

- $np > 5$
- $n(1 - p) > 5$

Example 23

A particular production process used to manufacture ferrite magnets used to operate reed switches in electronic meters is known to give 10% defective magnets on average. If 200 magnets are randomly selected, what is the probability that the number of defective magnets is between 24 and 30?

Solution

In general, **adding/subtracting** n independent Normal random variables with different means and variances result in a new Normal random variable with its being

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

$$V(X_1 \pm X_2 \pm \cdots \pm X_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

Example 24

In a certain mass-produced assembly, a 3 cm shaft must slide into a cylindrical sleeve. Shafts are manufactured whose diameter S follows a Normal distribution $S \sim N(3, 0.0042)$ and cylindrical sleeves are manufactured whose internal diameter C follows a Normal distribution $C \sim N(3.010, 0.0032)$. Assembly is performed by selecting a shaft and a cylindrical sleeve at random. In what proportion of cases will it be impossible to fit the selected shaft and cylindrical sleeve together?

Solution

4 Further Readings

[1] HELM Workbook 35 Sets and Probability.

https://nucinkis-lab.cc.ic.ac.uk/HELM/HELM_Workbooks_31-35/WB35-all.pdf

[2] HELM Workbook 37 Discrete Probability Distributions.

https://nucinkis-lab.cc.ic.ac.uk/HELM/HELM_Workbooks_36-40/WB37-all.pdf

[3] HELM Workbook 38 Continuous Probability Distributions

https://nucinkis-lab.cc.ic.ac.uk/HELM/HELM_Workbooks_36-40/WB38-all.pdf

[4] HELM Workbook 39 The Normal Distribution

https://nucinkis-lab.cc.ic.ac.uk/HELM/HELM_Workbooks_36-40/WB39-all.pdf

5 Appendix

The Standard Normal Probability Integral

$Z = \frac{x-\mu}{\sigma}$	0	1	2	3	4	5	6	7	8	9
0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
.1	0398	0438	0478	0517	0577	0596	0636	0675	0714	0753
.2	0793	0832	0871	0909	0948	0987	1026	1064	1103	1141
.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
.4	1555	1591	1628	1664	1700	1736	1772	1808	1844	1879
.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
.7	2580	2611	2642	2673	2703	2734	2764	2794	2822	2852
.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4946	4947	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
	4987	4990	4993	4995	4997	4998	4998	4999	4999	4999