Probability

- Probability theory is the mathematical study of randomness. A probability model of a random experiment is defined by assigning probabilities to all the different outcomes.
- Probability is a numerical measure of the likelihood that an event will occur. Thus, probabilities can be used as measures of degree of uncertainty associated with outcomes of an experiment. Probability values are always assigned on a scale from 0 to 1.
- A probability of 0 means that the event is impossible, while a probability near 0 means that it is highly unlikely to occur.
- Similarly an event with probability 1 is certain to occur, whereas an event with a probability near to 1 is very likely to occur.

Experiments and Outcomes

- In the study of probability any process of observation is referred to as an experiment.
- The results of an experiment (or other situation involving uncertainty) are called the outcomes of the experiment.
- An experiment is called a random experiment if the outcome can not be predicted.
- Typical examples of a random experiment are
  - a role of a die,
  - a toss of a coin,
  - drawing a card from a deck.

  If the experiment is yet to be performed we refer to possible outcomes or possibilities for short. If the experiment has been performed, we refer to realized outcomes or realizations.

Sample Spaces and Events

- The set of all possible outcomes of a probability experiment is called a ***sample space***, which is usually denoted by ***S***.
- The sample space is an exhaustive list of all the possible outcomes of an experiment. We call individual elements of this list ***sample points***.
- Each possible outcome is represented by one and only one sample point in the sample space.

Sample Spaces: Examples

For each of the following experiments, write out the sample space.

- Experiment: Rolling a die once
  - Sample space $S = \{1,2,3,4,5,6\}$
- Experiment: Tossing a coin
  - Sample space $S = \{Heads, Tails\}$
- Experiment: Measuring a randomly selected persons height (cms)
  - Sample space $S =$ The set of all possible real numbers.

Events

- An event is a specific outcome, or any collection of outcomes of an experiment.
- Formally, any subset of the sample space is an event.
- Any event which consists of a single outcome in the sample space is called an ***elementary*** or ***simple event***.
- Events which consist of more than one outcome are called ***compound events.***
- For example, an elementary event associated with the die example could be the "die shows 3".
- An compound event associated with the die example could be the "die shows an even number".

The Complement Event

- The complement of an event A is the set of all outcomes in the sample space that are not included in the outcomes of event A.
- We call the complement event of A as A^c.
- The complement event of a die throw resulting in an even number is the die throwing an odd number.
- Question: if there is a 40% chance of a randomly selected student being male, what is the probability of the selected student being female?

Set Theory : Union and Intersection

Set theory is used to represent relationships among events.

Union of two events:
The union of events A and B is the event containing all the sample points belonging to A or B or both. This is denoted $A \cup B$, (pronounce as “A union B”).

Intersection of two events:
The intersection of events A and B is the event containing all the sample points common to both A and B. This is denoted $A \cap B$, (pronounce as “A intersection B”).

More Set Theory

In general, if A and B are two events in the sample space S, then

- $A \subseteq B$ (A is a subset of B) = 'if A occurs, so does B
- $\varnothing$ (the empty set) = an impossible event
- S (the sample space) = an event that is certain to occur

Examples of Events

Consider the experiment of rolling a die once. From before, the sample space is given as $S = \{1,2,3,4,5,6\}$. The following are examples of possible events.

- A = score < 4 = $\{1,2,3\}$.
- B = 'score is even' = $\{2,4,6\}$.
- C = 'score is 7' = 0
- $A \cup B$ = 'the score is < 4 or even or both' = $\{1,2,3,4,6\}$
- $A \cap B$ = 'the score is < 4 and even = $\{2\}$
- A^c ='event A does not occur' = $\{4,5,6\}$

Sample Spaces

A complete list of all possible outcomes of a random experiment is called sample space or possibility space and is denoted by S.
A sample space is a set or collection of outcome of a particular random experiment.
For example, imagine a dart board. You are trying to find the probability of getting a bullseye. The dart board is the sample space. The probability of a dart hitting the dart board is 1.0. For another example, imagine rolling a six sided die. The sample space is 1, 2, 3, 4, 5, 6.

Sample Spaces

The following list consists of sample spaces of examples of random experiments and their respective outcomes.
The tossing of a coin, sample space is Heads, Tails
The roll of a die, sample space is 1, 2, 3, 4, 5, 6
The selection of a numbered ball (1-50) in an urn, sample space is 1, 2, 3, 4, 5, ..., 50

Sample Spaces

Percentage of calls dropped due to errors over a particular time period, sample space is $\{2\%, 14\%, 23\%, ...\}$
The time difference between two messages arriving at a message centre, sample space is 0, ..., infinity
The time difference between two different voice calls over a particular network, sample space is 0, ..., infinity

Probability

If there are n possible outcomes to an experiment, and m ways in which event A can happen, then the probability of event A (which we write as P(A)) is

$$P(A) = \frac{m}{n}$$

Probability

The probability of the event A may be interpreted as the proportion of times that event A will occur if we repeat the random experiment an infinite number of times.

Rules:

1. $0 \leq P(A) \leq 1$: the probability of any event lies between 0 and 1 inclusive.
2. $P(S) = 1$: the probability of the sample space is always equal to 1.
3. $P(A^c) = 1 - P(A)$: how to compute the probability of the complement.

Contingency Tables

Suppose there are 100 students in a first year college intake.

- 44 are male and are studying computer science,
- 18 are male and studying engineering
- 16 are female and studying computer science,
- 22 are female and studying engineering.

We assign the names M, F, C and E to the events that a student, randomly selected from this group, is male, female, studying computer science, and studying engineering respectively.

Contingency Tables

The most effective way to handle this data is to draw up a table. We call this a ***contingency table***. A contingency table is a table in which all possible events (or outcomes) for one variable are listed as row headings, all possible events for a second variable are listed as column headings, and the value entered in each cell of the table is the frequency of each joint occurrence.

	C	E	Total
M	44	18	62
F	16	22	38
Total	60	40	100

Contingency Tables

It is now easy to deduce the probabilities of the respective events, by looking at the totals for each row and column.

- P(C) = 60/100 = 0.60
- P(E) = 40/100 = 0.40
- P(M) = 62/100 = 0.62
- P(F) = 38/100 = 0.38

Remark:
The information we were originally given can also be expressed as:

- $P(C \cap M) = 44/100 = 0.44$
- $P(C \cap F) = 16/100 = 0.16$
- $P(E \cap M) = 18/100 = 0.18$
- $P(E \cap F) = 22/100 = 0.22$

Joint Probability Tables

A ***joint probability table*** is similar to a contingency table, but for that the value entered in each cell of the table is the probability of each joint occurrence. Often, the probabilities in such a table are based on observed frequencies of occurrence for the various joint events.

	C	E	Total
M	0.44	0.18	0.62
F	0.16	0.22	0.38
Total	0.60	0.40	1.00

Marginal Probabilities

- In the context of joint probability tables, a ***marginal probability*** is so named because it is a marginal total of a row or a column.
- Whereas the probability values in the cells of the table are probabilities of joint occurrence, the marginal probabilities are the simple (i.e. unconditional) probabilities of particular events.
- From the first year intake example, the marginal probabilities are $P(C)$, $P(E)$, $P(M)$ and $P(F)$ respectively.

Conditional Probabilities : Example 1

Recall the definition of conditional probability:

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

Using this formula, compute the following:

1. $P(C|M)$: Probability that a student is a computer science student, given that he is male.
2. $P(E|M)$: Probability that a student studies engineering, given that he is male.
3. $P(F|E)$: Probability that a student is female, given that she studies engineering.
4. $P(E|F)$: Probability that a student studies engineering, given that she is female.

Refer back to the contingency table to appraise your results.

Conditional Probabilities : Example 1

Part 1) Probability that a student is a computer science student, given that he is male.

$$P(C|M) = \frac{P(C \cap M)}{P(M)} = \frac{0.44}{0.62} = 0.71$$

Part 2) Probability that a student studies engineering, given that he is male.

$$P(E|M) = \frac{P(E \cap M)}{P(M)} = \frac{0.18}{0.62} = 0.29$$

Conditional Probabilities : Example 1

Part 3) Probability that a student is female, given that she studies engineering.

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{0.22}{0.40} = 0.55$$

Part 4) Probability that a student studies engineering, given that she is female.

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.22}{0.38} = 0.58$$

Remark: $P(E \cap F)$ is the same as $P(F \cap E)$.

Conditional Probability

Suppose B is an event in a sample space S with $P(B) > 0$. The probability that an event A occurs once Bhas occurred or, specifically, the conditional probability of A given B (written $P(A|B)$), is defined as follows:

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

- This can be expressed as a multiplication theorem

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

- The symbol $|$ is a vertical line and does not imply division.
- Also $P(A|B)$ is not the same as $P(B|A)$.

Remark: The Prosecutor's Fallacy , with reference to the O.J. Simpson trial.

Independent Events

Events A and B in a probability space S are said to be independent if the occurrence of one of them does not influence the occurrence of the other.

More specifically, B is independent of A if $P(B)$ is the same as $P(B|A)$. Now substituting $P(B)$ for $P(B|A)$ in the multiplication theorem from the previous slide yields.

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

We formally use the above equation as our definition of independence.

Mutually Exclusive Events

- Two events are mutually exclusive (or disjoint) if it is impossible for them to occur together.
- Formally, two events A and B are mutually exclusive if and only if $A \cap B = \varnothing$

Consider our die example

- Event A = 'observe an odd number' = $\{1, 3, 5\}$
- Event B = 'observe an even number' = $\{2, 4, 6\}$
- $A \cap B = \varnothing$ (i.e. the empty set), so A and B are mutually exclusive.

Addition Rule

The addition rule is a result used to determine the probability that event A or event B occurs or both occur. The result is often written as follows, using set notation:

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

- $P(A)$ = probability that event A occurs.
- $P(B)$ = probability that event B occurs.
- $P(A \cup B)$ = probability that either event A or event B occurs, or both occur.
- $P(A \cap B)$ = probability that event A and event B both occur.

Remark: $P(A \cap B)$ is subtracted to prevent the relevant outcomes being counted twice.

Addition Rule (Continued)

For mutually exclusive events, that is events which cannot occur together: $P(A \cap B) = 0$. The addition rule therefore reduces to

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

Addition Rule: Worked Example

Suppose we wish to find the probability of drawing either a Queen or a Heart in a single draw from a pack of 52 playing cards. We define the events Q = 'draw a queen' and H = 'draw a heart'.

- $P(Q)$ probability that a random selected card is a Queen
- $P(H)$ probability that a randomly selected card is a Heart.
- $P(Q \cap H)$ probability that a randomly selected card is the Queen of Hearts.
- $P(Q \cup H)$ probability that a randomly selected card is a Queen or a Heart.

Solution

- Since there are 4 Queens in the pack and 13 Hearts, so the $P(Q) = 4/52$ and $P(H) = 13/52$ respectively.
- The probability of selecting the Queen of Hearts is $P(Q \cap H) = 1/52$.
- We use the addition rule to find $P(Q \cup H)$:

$$P(Q \cup H) = (4/52) + (13/52) - (1/52) = 16/52$$

- So, the probability of drawing either a queen or a heart is $16/52 (= 4/13)$.

Multiplication Rule

The multiplication rule is a result used to determine the probability that two events, A and B, both occur. The multiplication rule follows from the definition of conditional probability.

The result is often written as follows, using set notation:

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

Recall that for independent events, that is events which have no influence on one another, the rule simplifies to:

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

Multiplication Rule

From the first year intake example, check that

$$P(E|F) \times P(F) = P(F|E) \times P(E)$$

- $P(E|F) \times P(F) = 0.58 \times 0.38 = 0.22$
- $P(F|E) \times P(E) = 0.55 \times 0.40 = 0.22$

Law of Total Probability

The law of total probability is a fundamental rule relating marginal probabilities to conditional probabilities. The result is often written as follows, using set notation:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

where $P(A \cap B^c)$ is probability that event A occurs and B does not.

Using the multiplication rule, this can be expressed as

$$P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c)$$

Law of Total Probability

From the first year intake example , check that

$$P(E) = P(E \cap M) + P(E \cap F)$$

with $P(E) = 0.40$, $P(E \cap M) = 0.18$ and $P(E \cap F) = 0.22$

$$0.40 = 0.18 + 0.22$$

Remark: M and F are complement events.

Bayes' Theorem

Bayes' Theorem is a result that allows new information to be used to update the conditional probability of an event.

Recall the definition of conditional probability:

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

Using the multiplication rule, gives Bayes' Theorem in its simplest form:

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

Probability: Worked Example

An electronics assembly subcontractor receives resistors from two suppliers: Deltatech provides 70% of the subcontractors's resistors while another company, Echelon, supplies the remainder.
1% of the resistors provided by Deltatech fail the quality control test, while 2% of the resistors from Echelon also fail the quality control test.

1. What is the probability that a resistor will fail the quality control test?
2. What is the probability that a resistor that fails the quality control test was supplied by Echelon?

Probability: Worked Example

Firstly, let's assign names to each event.

- D : a randomly chosen resistor comes from Deltatech.
- E : a randomly chosen resistor comes from Echelon.
- F : a randomly chosen resistor fails the quality control test.
- P : a randomly chosen resistor passes the quality control test.

We are given (or can deduce) the following probabilities:

- $P(D) = 0.70$,
- $P(E) = 0.30$.

Probability: Worked Example

We are given two more important pieces of information:

- The probability that a randomly chosen resistor fails the quality control test, given that it comes from Deltatech: $P(F|D) = 0.01$.
- The probability that a randomly chosen resistor fails the quality control test, given that it comes from Echelon: $P(F|E) = 0.02$.

Probability: Worked Example

The first question asks us to compute the probability that a randomly chosen resistor fails the quality control test. i.e. $P(F)$.

All resistors come from either Deltatech or Echelon. So, using the ***law of total probability***, we can express $P(F)$ as follows:

$$P(F) = P(F \cap D) + P(F \cap E)$$

Probability: Worked Example

Using the ***multiplication rule*** i.e. $P(A \cap B) = P(A|B) \times P(B)$, we can re-express the formula as follows

$$P(F) = P(F|D) \times P(D) + P(F|E) \times P(E)$$

We have all the necessary probabilities to solve this.

$$P(F) = 0.01 \times 0.70 + 0.02 \times 0.30 = 0.007 + 0.006 = 0.013$$

Probability: Worked Example

- The second question asks us to compute probability that a resistor that fails the quality control test was supplied by Echelon.
- In other words; of the resistors that did fail the quality test only, what is the probability that a randomly selected resistor was supplied by Echelon?
- We can express this mathematically as $P(E|F)$.
- We can use ***Bayes' theorem*** to compute the answer.

Probability: Worked Example

Recall Bayes' theorem

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

$$P(E|F) = \frac{P(F|E) \times P(E)}{P(F)} = \frac{0.02 \times 0.30}{0.013} = 0.46$$

More on probability

For this lecture and the next.

1. Contingency Tables
2. Conditional Probability: Worked Examples
3. Joint Probability Tables
4. The Multiplication Rule
5. Law of Total Probability
6. Bayes' Theorem
7. Exam standard Probability Question
8. Random Variables

1. Contingency Tables
2. Conditional Probability: Worked Examples
3. Joint Probability Tables
4. The Multiplication Rule
5. Law of Total Probability
6. Bayes' Theorem
7. Exam standard Probability Question
8. Random Variables

Contingency Tables

Suppose there are 100 students in a first year college intake.

- 44 are male and are studying computer science,
- 18 are male and studying statistics
- 16 are female and studying computer science,
- 22 are female and studying statistics.

We assign the names M, F, C and S to the events that a student, randomly selected from this group, is male, female, studying computer science, and studying statistics respectively.

Contingency Tables

The most effective way to handle this data is to draw up a table. We call this a ***contingency table***. A contingency table is a table in which all possible events (or outcomes) for one variable are listed as row headings, all possible events for a second variable are listed as column headings, and the value entered in each cell of the table is the frequency of each joint occurrence.

	C	S	Total
M	44	18	62
F	16	22	38
Total	60	40	100

Contingency Tables

It is now easy to deduce the probabilities of the respective events, by looking at the totals for each row and column.

- P(C) = 60/100 = 0.60
- P(S) = 40/100 = 0.40
- P(M) = 62/100 = 0.62
- P(F) = 38/100 = 0.38

Remark:

The information we were originally given can also be expressed as:

- $P(C \cap M) = 44/100 = 0.44$
- $P(C \cap F) = 16/100 = 0.16$
- $P(S \cap M) = 18/100 = 0.18$
- $P(S \cap F) = 22/100 = 0.22$

Joint Probability Tables

A ***joint probability table*** is similar to a contingency table, but for that the value entered in each cell of the table is the probability of each joint occurrence. Often, the probabilities in such a table are based on observed frequencies of occurrence for the various joint events.

	C	S	Total
M	0.44	0.18	0.62
F	0.16	0.22	0.38
Total	0.60	0.40	1.00

Marginal Probabilities

- In the context of joint probability tables, a ***marginal probability*** is so named because it is a marginal total of a row or a column.
- Whereas the probability values in the cells of the table are probabilities of joint occurrence, the marginal probabilities are the simple (i.e. unconditional) probabilities of particular events.
- From the first year intake example, the marginal probabilities are $P(C)$, $P(S)$, $P(M)$ and $P(F)$ respectively.

Conditional Probabilities : Example 1

Recall the definition of conditional probability:

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

Using this formula, compute the following:

1. $P(C|M)$: Probability that a student is a computer science student, given that he is male.
2. $P(S|M)$: Probability that a student studies statistics, given that he is male.
3. $P(F|S)$: Probability that a student is female, given that she studies statistics.
4. $P(S|F)$: Probability that a student studies statistics, given that she is female.

Refer back to the contingency table to appraise your results.

Conditional Probabilities : Example 1

Part 1) Probability that a student is a computer science student, given that he is male.

$$P(C|M) = \frac{P(C \cap M)}{P(M)} = \frac{0.44}{0.62} = 0.71$$

Part 2) Probability that a student studies statistics, given that he is male.

$$P(S|M) = \frac{P(S \cap M)}{P(M)} = \frac{0.18}{0.62} = 0.29$$

Conditional Probabilities : Example 1

Part 3) Probability that a student is female, given that she studies statistics.

$$P(F|S) = \frac{P(F \cap S)}{P(S)} = \frac{0.22}{0.40} = 0.55$$

Part 4) Probability that a student studies statistics, given that she is female.

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{0.22}{0.38} = 0.58$$

Remark: $P(S \cap F)$ is the same as $P(F \cap S)$.

Random Variables

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be ‘heads’ or ‘tails’.
- However, we often want to represent outcomes as numbers.
- A ***random variable*** is a function that associates a unique numerical value with every outcome of an experiment.
- The value of the random variable will vary from trial to trial as the experiment is repeated.
- Numeric values can be assigned to outcomes that are not usually considered numeric.
- For example, we could assign a ‘head’ a value of 0, and a ‘tail’ a value of 1, or vice versa.

Random Variables

There are two types of random variable - discrete and continuous. The distinction between both types will be important later on in the course.

Examples

- A coin is tossed ten times. The random variable X is the number of tails that are noted. X can only take the values $\{0, 1, ..., 10\}$, so X is a discrete random variable.
- A light bulb is burned until it burns out. The random variable Y is its lifetime in hours. Y can take any positive real value, so Y is a continuous random variable.

Discrete Random Variable

- A discrete random variable is one which may take on only a countable number of distinct values such as $\{0, 1, 2, 3, 4, ...\}$.
- Discrete random variables are usually (but not necessarily) counts.
- If a random variable can take only a finite number of distinct values, then it must be discrete.
- Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten.

Continuous Random Variable

- A continuous random variable is one which takes an infinite number of possible values.
- Continuous random variables are usually measurements.
- Examples include height, weight, the amount of sugar in an orange, the time required to run a computer simulation.

Random Variables

A pair of dice is thrown. Let X denote the minimum of the two numbers which occur. Find the distributions and expected value of X.

Random Variables

A fair coin is tossed four times. Let X denote the longest string of heads. Find the distribution and expectation of X.

Random Variables

A fair coin is tossed until a head or five tails occurs. Find the expected number E of tosses of the coin.

Random Variables

A coin is weighted so that P(H) = 0.75 and P(T) = 0.25
The coin is tossed three times. Let X denote the number of heads that appear.

- (a) Find the distribution f of X.
- (b) Find the expectation E(X).

Graphical Procedures for Statistics

- Bar-plots
- Histograms
- Boxplots

Histograms

- Consider an experiment in which each student in a class of 60 rolls a die 100 times.
- Each score is recorded, and a total score is calculated.
- As the expected value of rolled die is 3.5, the expected total is 350 for each student.
- At the end of the experiment the students reported their totals.
- The totals were put into ascending order, and tabulated as follows (next slide).

Outcomes of die-throw experiment

307	321	324	328	329	330	334	335	336	337
337	337	338	339	339	342	343	343	344	344
346	346	347	348	348	348	350	351	352	352
353	353	353	354	354	356	356	357	357	358
358	360	360	361	362	363	365	365	369	369
370	370	374	378	381	384	385	386	392	398

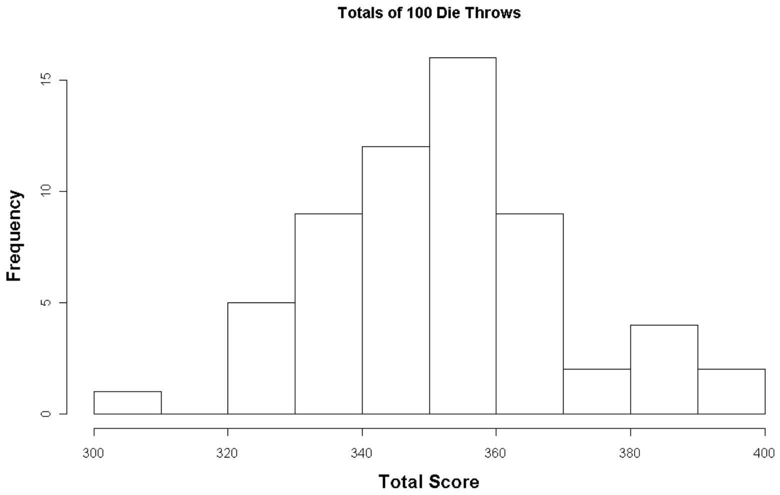
- What proportion of outcomes are less than or equal to 330? (Answer: 10%)
- What proportion of outcomes are greater than or equal to 370? (Answer: 16.66%)

What is a Histograms

TEXT HERE

Histograms

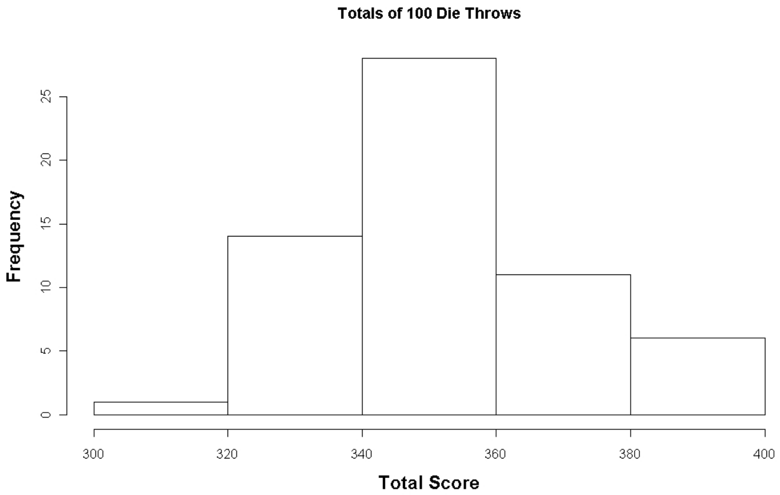
For the die-throw experiment;



Constructing Histograms

- Compute an appropriate number of class intervals.
- As a rule of thumb, the number of class intervals is usually approximately the square root of the number of observations.
- As there are 60 observations, we would normally use 7 or 8 class intervals.
- To save time, we will just use 5 class intervals.

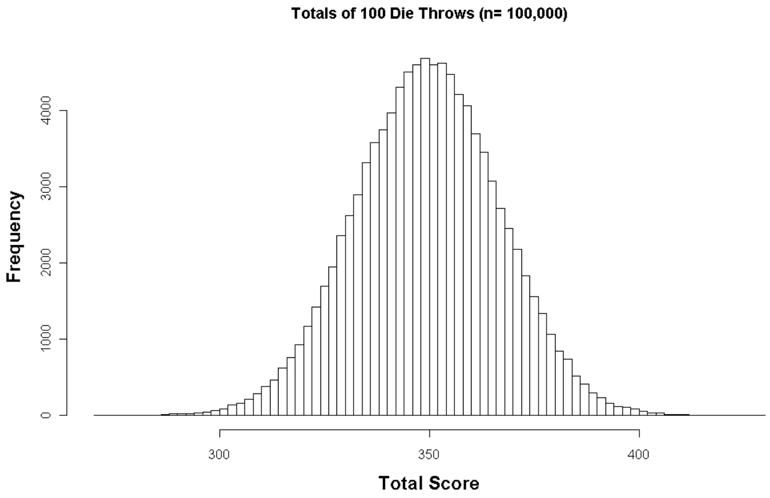
Histograms


Totals of 100 Die Throws
Frequency
0
5
10
15
20
25
300
320
340
360
380
400
Total Score

Histograms

- Suppose that the experiment of throwing a die 100 times and recording the total was repeated 100,000 times.
- (If implemented on a computer, we would call this a simulation study)
- The histogram of data (with a class interval width of 2) is shown on the next slide.
- How should the shape of the histogram be described?
- “Bell-shaped” would be a suitable description.

Histograms



Totals of 100 Die Throws (n= 100,000)
Frequency
0
1000
2000
3000
4000
300
350
400
Total Score

Simulation Study

A couple of remarks about the simulation study, some of which will be relevant later on.

- Approximately 68.7% of the values in the simulation study are between 332 and 367.
- Approximately 95% of the values are between 316 and 383.
- 2.5% of the values output are less than 316.
- 2.5% of the values study output are greater than 383.
- 175 values are greater than or equal to 400, whereas 198 values are less than or equal to 300.
- Results such as these are unusual, but they are not impossible.

Random Variables

A pair of dice is thrown. Let X denote the minimum of the two numbers which occur. Find the distributions and expected value of X.

Random Variables

A fair coin is tossed four times. Let X denote the longest string of heads. Find the distribution and expectation of X.

Random Variables

A fair coin is tossed until a head or five tails occurs. Find the expected number E of tosses of the coin.

Random Variables

A coin is weighted so that $P(H) = 0.75$ and $P(T) = 0.25$
The coin is tossed three times. Let X denote the number of heads that appear.

- (a) Find the distribution f of X.
- (b) Find the expectation $E(X)$.

- Now consider an experiment with only two outcomes. Independent repeated trials of such an experiment are called Bernoulli trials, named after the Swiss mathematician Jacob Bernoulli (16541705).
- The term ***independent trials*** means that the outcome of any trial does not depend on the previous outcomes (such as tossing a coin).
- We will call one of the outcomes the “success” and the other outcome the “failure”.

- Let p denote the probability of success in a Bernoulli trial, and so $q = 1 - p$ is the probability of failure. A binomial experiment consists of a fixed number of Bernoulli trials.
- A binomial experiment with n trials and probability p of success will be denoted by

$$B(n,p)$$

Counting Problems

- Sampling without replacement.
- Factorials
- Permutations
- Combinations

Sampling without replacement

- Sampling is said to be “without replacement” when a unit is selected at random from the population and it is not returned to the main lot.
- The first unit is selected out of a population of size N and the second unit is selected out of the remaining population of $N-1$ units and so on.
- For example, if you draw one card out of a deck of 52, there are only 51 cards left to draw from if you are selecting a second card.

Sampling without replacement

A lot of 100 semiconductor chips contains 20 that are defective. Two chips are selected at random, without replacement from the lot.

- What is the probability that the first one is defective?
  (Answer : 20/100 , i.e 0.20)
- What is the probability that the second one is defective given that the first one was defective?
  (Answer: 19/99)
- What is the probability that the second one is defective given that the first one was not defective?
  (Answer: 20/99)

Sampling With Replacement

Sampling is called "with replacement" when a unit selected at random from the population is returned to the population and then a second element is selected at random. Whenever a unit is selected, the population contains all the same units.

- What is the probability of guessing a PIN number for an ATM card at the first attempt.
- Importantly a digit can be used twice, or more, in PIN codes.
- For example 1337 is a valid pin number, where 3 appears twice.
- We have a one-in-ten chance of picking the first digit correctly, a one-in-ten chance of the guessing the second, and so on.
- All of these events are independent, so the probability of guess the correct PIN is $0.1 \times 0.1 \times 0.1 \times 0.1 = 0.0001$

Factorials Numbers

A factorial is a positive whole number, based on a number n , and which is written as "$n!$". The factorial $n!$ is defined as follows:

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$$

Remark $n! = n \times (n-1)!$

Example:

- $3! = 3 \times 2 \times 1 = 6$
- $4! = 4 \times 3! = 4 \times 3 \times 2 \times 1 = 24$

Remark $0! = 1$ not 0.

Permutations and Combinations

Often we are concerned with computing the number of ways of selecting and arranging groups of items.

- A ***combination*** describes the selection of items from a larger group of items.
- A ***permutation*** is a combination that is arranged in a particular way.

- Suppose we have items A,B,C and D to choose two items from.
- AB is one possible selection, BD is another. AB and BD are both combinations.
- More importantly, AB is one combination, for which there are two distinct permutations: AB and BA.

Combinations

Combinations: The number of ways of selecting k objects from n unique objects is:

$$^{n}C_{k} = \frac{n!}{k! \times (n-k)!}$$

In some texts, the notation for finding the number of possible combination is written

$$^{n}C_{k} = \binom{n}{k}$$

Example of Combinations

How many ways are there of selecting two items from possible 5?

$$ {}^{5}C_{2} \left(\text{also } \binom{5}{2} \right) = \frac{5!}{2! \times 3!} = \frac{5 \times 4 \times 3!}{2 \times 1 \times 3!} = 10 $$

Discuss how combinations can be used to compute the number of rugby matches for each group in the Rugby World Cup.

The Permutation Formula

The number of different permutations of r items from n unique items is written as ${}^{n}P_{k}$

$${}^{n}P_{k} = \frac{n!}{(n-k)!}$$

Permutations

Example: How many ways are there of arranging 3 different jobs, between 5 workers, where each worker can only do one job?

$$^{5}P_{3} = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60$$

Example of Combinations

A committee of 4 must be chosen from 3 females and 4 males.

- In how many ways can the committee be chosen.
- In how many cans 2 males and 2 females be chosen.
- Compute the probability of a committee of 2 males and 2 females are chosen.
- Compute the probability of at least two females.

Example of Combinations

Part 1

We need to choose 4 people from 7:
This can be done in

$$^7C_4 = \frac{7!}{4! \times 3!} = \frac{7 \times 6 \times 5 \times 4!}{4! \times 3!} = 35 \text{ ways.}$$

Part 2

With 4 men to choose from, 2 men can be selected in

$$^4C_2 = \frac{4!}{2! \times 2!} = \frac{4 \times 3 \times 2!}{2! \times 2!} = 6 \text{ ways.}$$

Similarly 2 women can be selected from 3 in

$$^3C_2 = \frac{3!}{2! \times 1!} = \frac{3 \times 2!}{2! \times 1!} = 3 \text{ ways.}$$

Example of Combinations

Part 2

Thus a committee of 2 men and 2 women can be selected in $6 \times 3 = 18$ ways.

Part 3

The probability of two men and two women on a committee is

$$\frac{\text{Number of ways of selecting 2 men and 2 women}}{\text{Number of ways of selecting 4 from 7}} = \frac{18}{35}$$

Example of Combinations

Part 4

- The probability of at least two females is the probability of 2 females or 3 females being selected.
- We can use the addition rule, noting that these are two mutually exclusive events.
- From before we know that probability of 2 females being selected is 18/35.

Example of Combinations

Part 4

- We have to compute the number of ways of selecting 1 male from 4 (4 ways) and the number of ways of selecting three females from 2 (only 1 way)
- The probability of selecting three females is therefore $\frac{4\times 1}{35} = 4/35$
- So using the addition rule

$$Pr(\text{ at least 2 females }) = Pr(\text{ 2 females }) + Pr(\text{ 3 females}$$

$$Pr(\text{ at least 2 females }) = 18/35 + 4/35 = 22/35$$

Factorials

- $5! = 5 \times 4 \times 3 \times 2 \times 1 (= 120)$
- $5! = 5 \times 4!$

The Choose Operator

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}$$

$$\binom{3}{1} = \frac{3!}{1! \times (3-1)!} \frac{3 \times 2!}{1! \times 2!} = \frac{3}{1} = 3$$

The Choose Operator

- $\binom{3}{0} = 1$
- (Remember 0! is always equal to 1)
- $\binom{3}{1} = 3$
- $\binom{3}{2} = 3$
- $\binom{3}{3} = 1$

Permutations

- The number of permutations of n objects is the number of ways in which the objects can be arranged in terms of order:
- Permutations of n objects :

$$n! = (n) \times (n-1) \times (n-2) \ldots \times 2 \times 1$$

- The symbol n! is read “n factorial”.
- In permutations and combinations problems, n is always positive. Also, note that by definition $0! = 1$ in mathematics.

Combinations

- In the case of permutations, the order in which the objects are arranged is important.
- In the case of combinations, we are concerned with the number of different groupings of objects that can occur without regard to their order.
- Therefore, an interest in combinations always concerns the number of different subgroups that can be taken from n objects. The number of combinations of n objects taken r at a time is

Permutations

Suppose a four letter code is made from the letters $\{a,b,c,d,e\}$, where repetitions are allowed and the order of the letters in the code is significant

For example a,a,e,c is a different code to a,c,e,a.

Permutations

- Let $\mathscr{U}$ be the set of all such codes.
- Let $\mathscr{V}$ be the set of all such codes beginning with a vowel.
- Let $\mathscr{P}$ be the set of all such codes which are palindromic.

(A palindromic code is a string of letters which read the same backwards as forwards, for example ***a,e,c,e,a*** is a 5 letter palindromic code.)

Permutations

How many elements are there in the set $\mathscr{U}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the set $\mathscr{V}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the set $\mathscr{P}$?

(i)	(ii)	(iii)	(iv)

Permutations

How many elements are there in the sets $\mathscr{V}$ and $\mathscr{P}$?

(i)	(ii)	(iii)	(iv)

Empty

Discrete Probability Distributions

- Poisson
- Binomial
- Geometric

What Is a Probability Distribution?

If you spend much time at all dealing with statistics, pretty soon you run into the phrase probability distribution. It is here that we really get to see how much the areas of probability and statistics overlap. Although this may sound like something technical, the phrase probability distribution is really just a way to talk about organizing a list of probabilities. A probability distribution is a function or rule that assigns probabilities to each value of a random variable. The distribution may in some cases be listed. In other cases it is presented as a graph.

Graph of a Probability Distribution

A probability distribution can be graphed, and sometimes this helps to show us features of the distribution that were not apparent from just reading the list of probabilities. The random variable is plotted along the x-axis, and the corresponding probability is plotted along the y - axis.

- For a discrete random variable, we will have a histogram
- For a continuous random variable, we will have the inside of a smooth curve

The rules of probability are still in effect, and they manifest themselves in a few ways. Since probabilities are greater than or equal to zero, the graph of a probability distribution must have y-coordinates that are nonnegative. Another feature of probabilities, namely that one is the maximum that the probability of an event can be, shows up in another way.

$$\text{Area} = \text{Probability}$$

Binomial Probability Distribution

The binomial distribution is a particular example of a probability distribution involving a discrete random variable. It is important that you can identify situations which can be modelled using the binomial distribution.

- There are n independent trials
- There are just two possible outcomes to each trial, success and failure, with fixed probabilities of p and q respectively, where q = 1 p.

The discrete random variable X is the number of successes in the n trials. X is modelled by the binomial distribution $B(n,p)$. You can write $X \sim B(n,p)$.

Poisson probability distribution

A discrete random variable that is often used is one which estimates the number of occurrences over a specified time period or space.
(remark : a specified space can be a specified length , a specified area, or a specified volume.)
If the following two properties are satisfied, the number of occurrences is a random variable described by the Poisson probability distribution

Properties

1) The probability of an occurrence is the same for any two intervals of equal length.
2) The occurrence or non-occurrence in any interval is independent of the occurrence or non-occurrence in any other interval.

The Poisson probability function is given by

- f(x) the probability of x occurrences in an interval.
- λ is the expected value of the mean number of occurrences in any interval. (We often call this the Poisson mean)
- e=2.71828284

Poisson Approximation of the Binomial Probability Distribution

The Poisson distribution can be used as an approximation of the binomial probability distribution when p, the probability of success is small and n, the number of trials is large. We set (other notation) and use the Poisson tables. As a rule of thumb, the approximation will be good wherever both and

Normal Probability Distribution

Bell Curve Bell curves show up throughout statistics. Diverse measurements such as diameters of seeds, lengths of fish fins, scores on the SAT and weights of individual sheets of a ream of paper all form bell curves when they are graphed. The general shape of all of these curves is the same. But all of these curves are different, because it is highly unlikely that any of them share the same mean or standard deviation. Bell curves with large standard deviations are wide, and bell curves with small standard deviations are skinny. Bell curves with larger means are shifted more to the right than those with smaller means.

Normal Probability Distribution

Characteristics of the Normal probability distribution

1. The highest point on the normal curve is at the mean, which is also the median and mode of the distribution.
2. The normal probability curve is bell-shaped and symmetric, with the shape of the curve to the left of the mean a mirror image of the shape of the curve to the right of the mean.
3. The standard deviation determines the width of the curve. Larger values of the the standard deviation result in wider flatter curves, showing more dispersion in data.
4. The total area under the curve for the normal probability distribution is 1.

The Binomial Probability Distribution

- The number of independent trials is denoted n.
- The probability of a ‘success’ is p
- The expected number of ‘successes’ from n trials is $E(X) = np$

Binomial Experiment

A binomial experiment (also known as a Bernoulli trial) is a statistical experiment that has the following properties:

- The experiment consists of n repeated trials.
- Each trial can result in just two possible outcomes. We call one of these outcomes a ***success*** and the other, a ***failure***.
- The probability of success, denoted by p, is the same on every trial.
- The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

Binomial Experiment

Consider the following statistical experiment. You flip a coin five times and count the number of times the coin lands on heads. This is a binomial experiment because:

- The experiment consists of repeated trials. We flip a coin five times.
- Each trial can result in just two possible outcomes : heads or tails.
- The probability of success is constant : 0.5 on every trial.
- The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials.

Binomial Probability

- A binomial experiment with n trials and probability p of success will be denoted by
$$B(n,p)$$
- Frequently, we are interested in the ***number of successes*** in a binomial experiment, not in the order in which they occur.
- Furthermore, we are interested in the probability of that number of successes.

Binomial Probability

The probability of exactly k successes in a binomial experiment B(n, p) is given by

$$P(X = k) = P(k \text{ successes }) = {}^{n}C_{k} \times p^{k} \times (1 - p)^{n-k}$$

- X: Discrete random variable for the number of successes (variable name)
- k : Number of successes (numeric value)
  - $P(X = k)$ “probability that the number of success is k”.
- n : number of independent trials
- p : probability of a success in any of the n trial.
- $1 - p$: probability of a failure in any of the n trial.

Binomial Example

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

Solution:

This is a binomial experiment in which

- a success is defined as an outcome of ‘4’.
- the number of trials is equal to $n = 5$,
- the number of successes is equal to $k = 2$,
- the number of failures is equal to 3,
- the probability of success on a single trial is 1/6,
- the probability of failure on a single trial is 5/6.

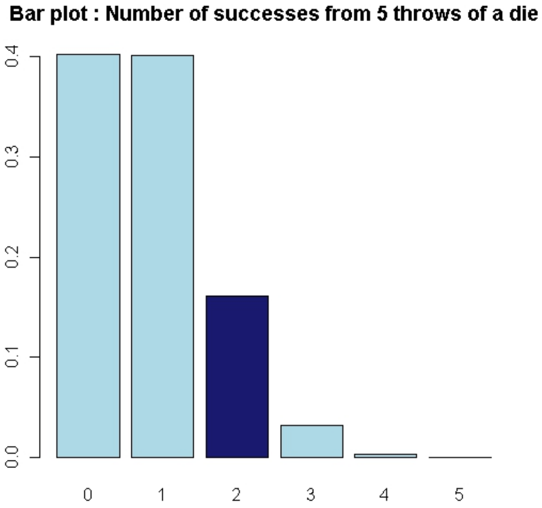
Binomial Example

Therefore, the probability of getting exactly 2 fours is:

$$P(X = 2) =^5 C_2 \times (1/6)^2 \times (5/6)^3 = 0.161$$

Remark: ${}^5C_2 = 10$

Binomial Example

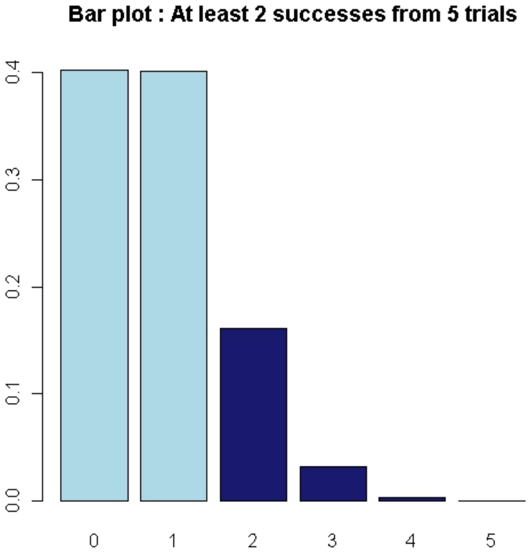


Binomial Probability

Remark : The sum of the probabilities of each of the possible outcomes (i.e. no fours, one four etc) is equal to one.

$$P(X=0)+P(X=1)+\ldots+P(X=5)=1$$

Binomial Example: At least two successes



Binomial Example: At least two successes

- Suppose we were asked to find the probability of ***at least*** 2 fours.
- Can you suggest the most efficient way of computing this?
- Suggestion: Compute $P(X=0)$ and $P(X=1)$.
- Together these probabilities are the complement probability of what we require.
- $P(X \geq 2) = 1 - (P(X=0) + P(X=1))$.
- (We will continue with this in future classes).

Cumulative Distribution Function

The cumulative distribution function (c.d.f.) of a discrete random variable X is the function $F(t)$ which tells you the probability that X is less than or equal to t.
So if X has p.d.f. P(X = x), we have:

$$F(t) = P(X \leq t) = \sum_{(i=0)}^{(i=t)} P(X = x)$$

In other words, for each value that X can be which is less than or equal to t, work out the probability that X is that value and add up all such results.

Binomial Example: Sample Problem

Suppose there are twelve multiple choice questions in an English class quiz. Each question has five possible answers, and only one of them is correct. Find the probability of having four or less correct answers if a student attempts to answer every question at random.

Binomial Example: Sample Problem

Solution: Since only one out of five possible answers is correct, the probability of answering a question correctly by random is $1/5 = 0.2$. We can find the probability of having exactly 4 correct answers by random attempts as follows.(Blackboard. Correct Answer is 13.29%)

The Poisson Probability Distribution

- A Poisson random variable is the number of successes that result from a Poisson experiment.
- The probability distribution of a Poisson random variable is called a Poisson distribution.
- Very Important: This distribution describes the number of occurrences in a ***unit period (or space)***
- Very Important: The expected number of occurrences is m

The Poisson Probability Distribution

We use the following notation.

$$X \sim Poisson(m)$$

Note the expected number of occurrences per unit time is conventionally denoted λ rather than m. As the Murdoch Barnes cumulative Poisson Tables

(Table 2) use m, so shall we. Recall that Tables 2 gives values of the probability $P(X \geq r)$, when X has a Poisson distribution with parameter m.

The Poisson Probability Distribution

Consider cars passing a point on a rarely used country road. Is this a Poisson Random Variable? Suppose

1. Arrivals occur at an average rate of m cars per unit time.
2. The probability of an arrival in an interval of length k is constant.
3. The number of arrivals in two non-overlapping intervals of time are independent.

This would be an appropriate use of the Poisson Distribution.

Changing the unit time.

- The number of arrivals, X, in an interval of length t has a Poisson distribution with parameter $\mu = mt$.
- m is the expected number of arrivals in a unit time period.
- μ is the expected number of arrivals in a time period t, that is different from the unit time period.
- Put simply : if we change the time period in question, we adjust the Poisson mean accordingly.
- If 10 occurrences are expected in 1 hour, then 5 are expected in 30 minutes. Likewise, 20 occurrences are expected in 2 hours, and so on.
- (Remark : we will not use μ in this context anymore).

Poisson Example

A motor dealership which specializes in agricultural machinery sells one vehicle every 2 days, on average. Answer the following questions.

1. What is the probability that the dealership sells at least one vehicle in one particular day?
2. What is the probability that the dealership will sell exactly one vehicle in one particular day?
3. What is the probability that the dealership will sell 4 vehicles or more in a six day working week?

Poisson Example

1. Expected Occurrences per Day: m = 0.5
2. Probability that the dealership sells at least one vehicle in one particular day?
$$P(X \geq 1) = 0.3935$$
3. Probability that the dealership will sell exactly one vehicle in one particular day?
$$P(X = 1) = P(X \geq 1) - P(X \geq 2) = 0.3935 - 0.0902 = 0.3031$$
4. Probability that the dealership will sell 4 vehicles or more in a six day working week?
   - For a 6 day week, m=3
   - $P(X \geq 4) = 0.3528$

Knowing which distribution to use

- For the end of semester examination, you will be required to know when it is appropriate to use the Poisson distribution, and when to use the binomial distribution.
- Recall the key parameters of each distribution.
- Binomial : number of ***successes*** in n ***independent trials***.
- Poisson : number of ***occurrences*** in a ***unit space***.

Characteristics of a Poisson Experiment

A Poisson experiment is a statistical experiment that has the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (m) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the region.
- The probability that a success will occur in an extremely small region is virtually zero.

Note that the specified region could take many forms. For instance, it could be a length, an area, a volume, a period of time, etc.

Poisson Distribution

A Poisson random variable is the number of successes that result from a Poisson experiment.
The probability distribution of a Poisson random variable is called a Poisson distribution.
Given the mean number of successes (m) that occur in a specified region, we can compute the Poisson probability based on the following formula:

The Poisson Probability Distribution

- The number of occurrences in a unit period (or space)
- The expected number of occurrences is m

Poisson Formulae

The probability that there will be k occurrences in a unit time period is denoted $P(X = k)$, and is computed as follows.

$$P(X = k) = \frac{m^k e^{-m}}{k!}$$

Poisson Formulae

Given that there is on average 2 occurrences per hour, what is the probability of no occurences in the next hour?

i.e. Compute $P(X=0)$ given that $m=2$

$$P(X=0) = \frac{2^0 e^{-2}}{0!}$$

- $2^0 = 1$
- $0! = 1$

The equation reduces to

$$P(X=0) = e^{-2} = 0.1353$$

Poisson Formulae

What is the probability of one occurrences in the next hour?

i.e. Compute $P(X = 1)$ given that $m = 2$

$$P(X = 1) = \frac{2^1 e^{-2}}{1!}$$

- $2^1 = 2$
- $1! = 1$

The equation reduces to

$$P(X = 1) = 2 \times e^{-2} = 0.2706$$

Poisson Expected Value and Variance

If the random variable X has a Poisson distribution with parameter m, we write

$$X \sim Poisson(m)$$

- Expected Value of X : E(X) = m
- Variance of X : $\text{Var}(X) = m$
- Standard Deviation of X : $SD(X) = \sqrt{m}$

Poisson Distribution : Example

- The number of faults in a fibre optic cable were recorded for each kilometre length of cable.
- The mean number of faults was found to be 4 faults per kilometre.
- The standard deviation of the number of faults was found to be 2 faults per kilometre.
- Is the Poisson Distribution is a useful technique for modelling the number of faults in fibre optic cable?
- (Looking at the last slide, the answer is yes, because the variance and mean are equal).

Poisson Approximation of the Binomial

- The Poisson distribution can sometimes be used to approximate the binomial distribution
- When the number of observations n is large, and the success probability p is small, the $B(n, p)$ distribution approaches the Poisson distribution with the parameter given by $m = np$.
- This is useful since the computations involved in calculating binomial probabilities are greatly reduced.
- As a rule of thumb, n should be greater than 50 with p very small, such that np should be less than 5.
- If the value of p is very high, the definition of what constitutes a "success" or "failure" can be switched.

Poisson Approximation: Example

- Suppose we sample 1000 items from a production line that is producing, on average, 0.1% defective components.
- Using the binomial distribution, the probability of exactly 3 defective items in our sample is

$$P(X = 3) = {}^{1000}C_3 \times 0.001^3 \times 0.999^{997}$$

Poisson Approximation: Example

Lets compute each of the component terms individually.

- ${}^{1000}C_3$

$$ {}^{1000}C_3 = \frac{1000 \times 999 \times 998}{3 \times 2 \times 1} = 166,167,000 $$

- 0.001^3

$$ 0.001^3 = 0.000000001 $$

- 0.999^{997}

$$ 0.999^{997} = 0.36880 $$

Multiply these three values to compute the binomial probability
$P(X = 3) = 0.06128$

Poisson Approximation: Example

- Lets use the Poisson distribution to approximate a solution.
- First check that $n \geq 50$ and $np < 5$ (Yes to both).
- We choose as our parameter value $m = np = 1000 \times 0.001 = 1$

$$P(X=3) = \frac{e^{-1} \times 1^3}{3!} = \frac{e^{-1}}{6} = \frac{0.36787}{6} = 0.06131$$

Compare this answer with the Binomial probability P(X = 3) = 0.06128. Very good approximation, with much less computation effort.

Poisson Approximation of the Binomial

- The Poisson distribution can sometimes be used to approximate the binomial distribution
- When the number of observations n is large, and the success probability p is small, the $Bin(n,p)$ distribution approaches the Poisson distribution with the parameter given by $m = np$.
- This is useful since the computations involved in calculating binomial probabilities are greatly reduced.
- As a rule of thumb, n should be greater than 50 with p very small, such that np should be less than 5.
- If the value of p is very high, the definition of what constitutes a "success" or "failure" can be switched.

Poisson Approximation: Example

- Suppose we sample 1000 items from a production line that is producing, on average, 0.1% defective components.
- Using the binomial distribution, the probability of exactly 3 defective items in our sample is

$$P(X=3) = {}^{1000}C_3 \times 0.001^3 \times 0.999^{997}$$

Poisson Approximation: Example

Lets compute each of the component terms individually.

- ${}^{1000}C_3$

$$^{1000}C_3 = \frac{1000 \times 999 \times 998}{3 \times 2 \times 1} = 166,167,000$$

- 0.001^3

$$0.001^3 = 0.000000001$$

- 0.999^{997}

$$0.999^{997} = 0.36880$$

Multiply these three values to compute the binomial probability $P(X = 3) = 0.06128$

Poisson Approximation: Example

- Lets use the Poisson distribution to approximate a solution.
- First check that $n \geq 50$ and $np < 5$ (Yes to both).
- We choose as our parameter value $m = np = 1000 \times 0.001 = 1$

$$P(X=3) = \frac{e^{-1} \times 1^3}{3!} = \frac{e^{-1}}{6} = \frac{0.36787}{6} = 0.06131$$

Compare this answer with the Binomial probability P(X = 3) = 0.06128. Very good approximation, with much less computation effort.

Geometric Distribution

Geometric Distribution

Geometric Distribution

Geometric distributions model (some) discrete random variables. Typically, a Geometric random variable is the number of trials required to obtain the first failure, for example, the number of tosses of a coin untill the first ’tail’ is obtained, or a process where components from a production line are tested, in turn, until the first defective item is found.

Geometric Distribution

A discrete random variable X is said to follow a Geometric distribution with parameter p, written $X \sim Ge(p)$, if it has probability distribution

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

where

- $x = 1, 2, 3, \ldots$
- p = success probability; $0 < p < 1$

Geometric Distribution

The trials must meet the following requirements:

(i) the total number of trials is potentially infinite; there are just two outcomes of each trial; success and failure;

(ii) the outcomes of all the trials are statistically independent;

(iii) all the trials have the same probability of success.

Geometric Distribution

The Geometric distribution has expected value and variance

$$E(X) = 1/(1-p)$$

$$V(X) = p/(1-p)^2$$

.

The Geometric distribution is related to the Binomial distribution in that both are based on independent trials in which the probability of success is constant and equal to p. However, a Geometric random variable is the number of trials until the first failure, whereas a Binomial random variable is the number of successes in n trials.

Current Status (Lecture 4B)

- Mid Term Examination next Monday (Week 5) at 4pm
- Currently covering : Continuous Probability Distributions
- Lecture notes are a bit out of synch with published class notes.
- The Exponential distribution will be examinable in Mid-Term 1
- Next Wednesday, we will start looking at the Normal Distribution.

Continuous Uniform Distribution

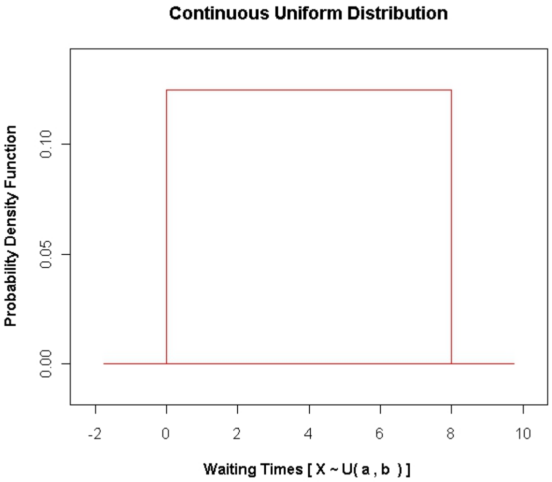
A random variable X is called a continuous uniform random variable over the interval (a,b) if it's probability density function is given by

$$f(x) = \frac{1}{b-a} \qquad \text{when } a \leq x \leq b \text{ (otherwise } f(x) = 0)$$

The corresponding cumulative density function is

$$F(x) = P(X \leq x) = \frac{x-a}{b-a} \qquad \text{when } a \leq x \leq b$$

The Continuous Uniform Distribution



Continuous Uniform Distribution

- The continuous uniform distribution is very simple to understand and implement, and is commonly used in computer applications (e.g. computer simulation).
- It is also known as the 'Rectangle Distribution' for obvious reasons.
- We specify the word "continuous" so as to distinguish it from it's discrete equivalent: the discrete uniform distribution.
- Remark; the dice distribution is a discrete uniform distribution with lower and upper limits 1 and 6 respectively.

Uniform Distribution Parameters

The continuous uniform distribution is characterized by the following parameters

- The lower limit a
- The upper limit b
- We denote a uniform random variable X as $X \sim U(a,b)$

It is not possible to have an outcome that is lower than a or larger than b.

$$P(X \leq a) = P(X \geq b) = 0$$

Interval Probability

- We wish to compute the probability of an outcome being within a range of values.
- We shall call this lower bound of this range L and the upper bound U.
- Necessarily L and U must be possible outcomes.
- The probability of X being between L and U is denoted $P(L \leq X \leq U)$.

$$P(L \leq X \leq U) = \frac{U - L}{b - a}$$

- (This equation is based on a definite integral).

Uniform Distribution: Cumulative Distribution

- For any value "c" between the minimum value a and the maximum value b, we can say
- $P(X \geq c)$
  $$P(X \geq c) = \frac{b-c}{b-a}$$
  here b is the upper bound while c is the lower bound
- $P(X \leq c)$
  $$P(X \leq c) = \frac{c-a}{b-a}$$
  here c is the upper bound while a is the lower bound.

Uniform Distribution: Mean and Variance

- The mean of the continuous uniform distribution, with parameters a and b is
$$E(X) = \frac{a+b}{2}$$
- The variance is computed as
$$V(X) = \frac{(b-a)^2}{12}$$

Uniform Distribution: Example

- Suppose there is a platform in a subway station in a large large city.
- Subway trains arrive **every three minutes** at this platform.
- What is the shortest possible time a passenger would have to wait for a train?
- What is the longest possible time a passenger will have to wait?

Uniform Distribution: Example

- What is the shortest possible time a passenger would have to wait for a train?
- If the passenger arrives just before the doors close, then the waiting time is zero.

$$a = 0 \text{ minutes } \ = 0 \text{ seconds}$$

Uniform Distribution: Example

- What is the longest possible time a passenger will have to wait?
- If the passenger arrives just after the doors close, and missing the train, then he or she will have to wait the full three minutes for the next one.

$$b = 3 \text{ minutes } \; = 180 \text{ seconds}$$

Uniform Distribution: Example

- What is the probability that he will have to wait longer than 2 minutes?

$$P(X \geq 2) = \frac{3-2}{3-0} = 1/3 = 0.33333$$

- See next slide (shaded area is 1/3 of rectangle)

The Continuous Uniform Distribution

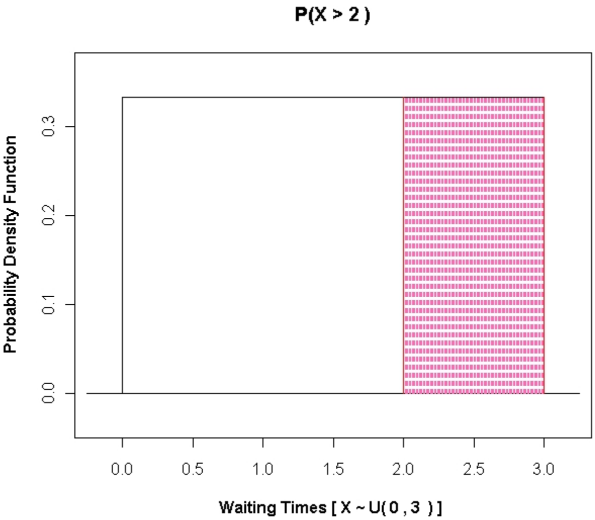


Figure:

Uniform Distribution: Example

- What is the probability that he will have to wait less than than 45 seconds (i.e. 0.75 minutes)?

$$P(X \leq 0.75) = \frac{0.75 - 0}{3 - 0} = 0.75/3 = 0.250$$

- See next slide (shaded area is 1/4 of rectangle)

The Continuous Uniform Distribution

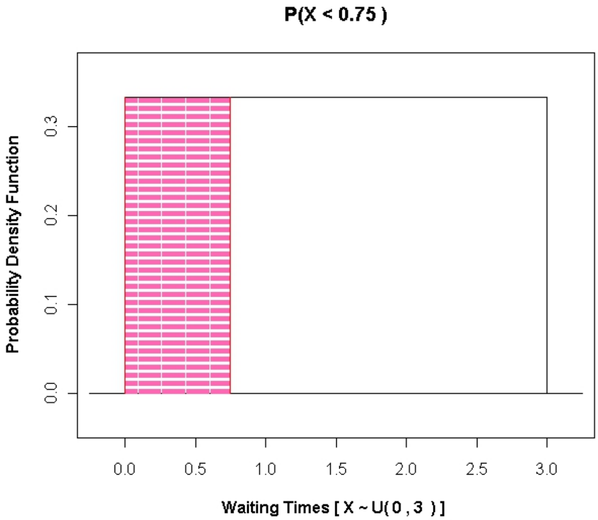


Figure:

Uniform Distribution: Expected Value

We are told that, for waiting times, the lower limit a is 0, and the upper limit b is 3 minutes.

The expected waiting time $\mathrm{E}[X]$ is computed as follows

$$\mathrm{E}[X] = \frac{b+a}{2} = \frac{3+0}{2} = 1.5 \text{ minutes}$$

Uniform Distribution: Variance

The variance of the continuous uniform distribution, denoted $\mathrm{V}[X]$, is computed using the following formula

$$\mathrm{V}[X] = \frac{(b-a)^2}{12}$$

For our previous example this is

$$\mathrm{V}[X] = \frac{(3-0)^2}{12} = \frac{3^2}{12} = \frac{9}{12} = 0.75$$

Exponential Distribution

The Exponential Distribution may be used to answer the following questions:

- How much time will elapse before an earthquake occurs in a given region?
- How long do we need to wait before a customer enters a shop?
- How long will it take before a call center receives the next phone call?
- How long will a piece of machinery work without breaking down?

Exponential Distribution

- All these questions concern the time we need to wait before a given event occurs.
- If this waiting time is unknown, it is often appropriate to think of it as a random variable having an **exponential distribution**.
- The time X we need to wait before an event occurs has an exponential distribution if the probability that the event occurs during a certain time interval is proportional to the length of that time interval.

Probability density function

The probability density function (PDF) of an exponential distribution is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The parameter λ is called ***rate*** parameter. It is the inverse of the expected duration (μ).

(If the expected duration is 5 (e.g. five minutes) then the rate parameter value is 0.2.)

Exponential Distribution: Cumulative density function

The cumulative distribution function (CDF) of an exponential distribution is

$$F(x;\lambda) = \begin{cases} 1-e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The CDF can be written as the probability of the lifetime being less than some value x.

$$P(X \leq x) = 1 - e^{-\lambda x}$$

Exponential Distribution: Expected Value and Variance

The expected value of an exponential random variable X is:

$$E[X] = \frac{1}{\lambda}$$

The variance of an exponential random variable X is:

$$V[X] = \frac{1}{\lambda^2}$$

Exponential Distribution: Example

Assume that the length of a phone call in minutes is an exponential random variable X with parameter $\lambda = 1/10$.

If someone arrives at a phone booth just before you arrive, find the probability that you will have to wait

(a) less than 5 minutes,

(b) greater than 5 minutes,

(c) between 5 and 10 minutes.

Also compute the expected value and variance.

Exponential Distribution: Example

Part a

Compute $P(X \leq 5)$ with $\lambda = 1/10$

$$P(X \leq x) = 1 - e^{-\lambda x}$$

Exponential Distribution: Example

Part a

Compute $P(X \leq 5)$ with $\lambda = 1/10$

- $P(X \leq x) = 1 - e^{-\lambda x}$
- $P(X \leq 5) = 1 - e^{-5/10}$
- $P(X \leq 5) = 1 - e^{-0.5}$
- $P(X \leq 5) = 1 - 0.6065$
- $P(X \leq 5) = 0.3934$

Exponential Distribution: Example

Part b

Compute $P(X \geq 10)$ with $\lambda = 1/10$

$$P(X \leq x) = 1 - e^{-\lambda x}$$

Complement rule

$$P(X \geq x) = 1 - P(X \leq x) = e^{-\lambda x}$$

Exponential Distribution: Example

Part b

Compute $P(X \geq 10)$ with $\lambda = 1/10$

$$P(X \geq x) = e^{-\lambda x}$$

- $P(X \geq x) = e^{-\lambda x}$
- $P(X \geq 10) = e^{-10/10}$
- $P(X \geq 10) = e^{-1}$
- $P(X \geq 10) = 0.3678$

Exponential Distribution: Example

Part c

Compute $P(5 \leq X \leq 10)$ with $\lambda = 1/10$

- Probability of being inside this interval is the complement of being outside the interval.
- The probability of being outside the interval is the composite event of being too low for the interval (i.e. $P(X \leq 5)$) and being too high for the interval (i.e. $P(X \leq 10)$).

$$P(5 \leq X \leq 10) = 1 - [P(X \leq 5) + P(X \geq 10)]$$

Exponential Distribution: Example

Part c

Compute $P(5 \leq X \leq 10)$ with $\lambda = 1/10$

$$P(5 \leq X \leq 10) = 1 - [P(X \leq 5) + P(X \geq 10)]$$

- **Too Low** $P(X \leq 5)$ = 0.3934
- **Too High** $P(X \geq 10)$ = 0.3678
- **Outside** $P(X \leq 5) + P(X \geq 10)$ = 0.7612
- **Inside** $P(5 \leq X \leq 10)$ = 1 - 0.7612 = 0.2388

Exponential Distribution

Expected Value and Variance

The expected value of an exponential random variable X is:

$$E[X] = \frac{1}{\lambda} = \frac{1}{1/10} = 10$$

The variance of an exponential random variable X is:

$$V[X] = \frac{1}{\lambda^2} = 100$$

Exponential Distribution

Exponential Distribution: Relationship to Poisson Mean

- The Exponential Rate parameter (λ) is related to the Poisson mean (m)
- If we expect 12 occurrences per hour - what is the rate of occurrences?
- We would expected to wait 1/12 of an hour (i.e. 5 minutes) between occurrences.
- Be mindful to keep your time units consistent, if working with both Poisson and Exponential.
- If working in minutes, our rate parameter values is $\lambda = 0.20$ (i.e. 1/5).
- (This could be the basis of an exam question).

Gamma Distribution

Applications The gamma distribution can be used a range of disciplines including queuing models, climatology, and financial services. Examples of events that may be modeled by gamma distribution include:

- The amount of rainfall accumulated in a reservoir
- The size of loan defaults or aggregate insurance claims
- The flow of items through manufacturing and distribution processes
- The load on web servers
- The many and varied forms of telecom exchange

The Pareto Distribution

Suppose the distribution of monthly salaries of full-time workers in the UK has a Pareto distribution with minimum monthly salary $x_m = 1000$ and concentration factor $\alpha = 3$.

The Pareto Distribution

1. Calculate the mean monthly salary of UK full-time workers.
2. Calculate the probability that a UK full-time worker earns more than 2000 per month.
3. Calculate the median monthly salary of UK full-time workers.

The Pareto Distribution

The expected value of a random variable following a Pareto distribution is

$$E(X) = \begin{cases} \infty & \text{if } \alpha \le 1, \\ \frac{\alpha x_\mathrm{m}}{\alpha - 1} & \text{if } \alpha > 1. \end{cases}$$

The Pareto Distribution

Because $\alpha = 3$, we will use this

$$E(X) = \frac{\alpha x_{\mathrm{m}}}{\alpha - 1}$$

Recall that $X_m = 1000$.

The Pareto Distribution

The cumulative distribution function of a Pareto random variable with parameters α and x_m is

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_\mathrm{m}}{x}\right)^\alpha & \text{for } x \geq x_\mathrm{m}, \\ 0 & \text{for } x < x_\mathrm{m}. \end{cases}$$

Using values for this example:

$$F_X(x) = \begin{cases} 1 - \left(\frac{1000}{x}\right)^3 & \text{for } x \geq 1000, \\ 0 & \text{for } x < 1000. \end{cases}$$

The Pareto Distribution

Calculate the probability that a UK full-time worker earns more than 2000 per month.

$$F_X(x) = \begin{cases} 1 - \left(\frac{1000}{x}\right)^3 & \text{for } x \geq 1000, \\ 0 & \text{for } x < 1000. \end{cases}$$

The Pareto Distribution

Calculate the probability that a UK full-time worker earns **more than** 2000 per month.

$$F_X(x) = \begin{cases} 1 - \left(\frac{1000}{x}\right)^3 & \text{for } x \geq 1000, \\ 0 & \text{for } x < 1000. \end{cases}$$

The Pareto Distribution

Calculate the median monthly salary of UK full-time workers.

$$\text{Median} : F_X(x) = 0.50$$

$$F_X(x) = \begin{cases} 1 - \left(\frac{1000}{x}\right)^3 & \text{for } x \geq 1000, \\ 0 & \text{for } x < 1000. \end{cases}$$

The Pareto Distribution

$$F_X(x) = 0.5 \qquad \rightarrow \qquad 1 - \left(\frac{1000}{x}\right)^3 = 0.50$$

$$\sqrt[3]{0.5} = 0.7937$$

The Pareto Distribution

$$\frac{1000}{0.7937} = 1259.92$$

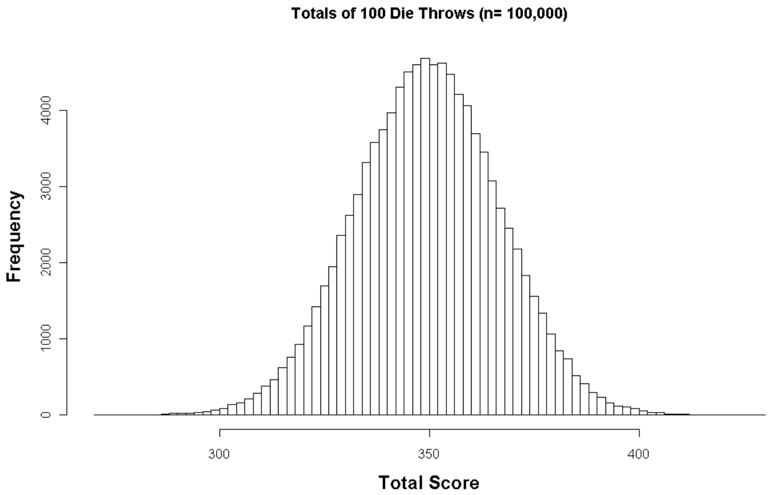
The Pareto Distribution

The Pareto Distribution

Introduction to the Normal Distribution

- Recall the experiment whereby a die was rolled 100 times, and the sum of the 100 values was recorded.
- This experiment was repeated a very large number of times (e.g. 100,000 times) in a simulation study.
- A histogram was drawn to depict the distribution of outcomes of this experiment.
- Recall that we agreed that “bell-shaped” was a good description of the histogram.

Normal Distribution



Normal Distribution

- The normal distribution is perhaps the most widely used type of probability distribution for a random variable.
- Normal distributions have the same general shape: the bell curve.
- The distributions are **symmetric** with values concentrated more in the middle than in the tails.
- Important The height of a normal distribution can be defined mathematically in terms of two fundamental parameters: the normal mean (μ) and the normal standard deviation (σ).
- A normally distributed random variable X is denoted $X \sim \mathrm{N}(\mu, \sigma^2)$ (note that we use the variance term here).
- The mean (μ) and standard deviation (σ) are vital for calculating probabilities.

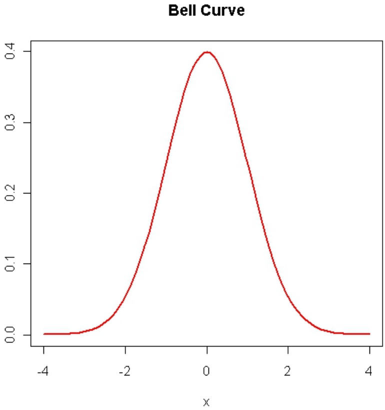
The Normal Distribution

The ***probability density function*** of the normal distribution is given as

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

Integrating this formula would allow us to compute probabilities.
However, it is not required to use this formula.

Normal Distribution



Characteristics of the Normal probability distribution

1. The highest point on the normal curve is at the mean, which is also the median of the distribution.
2. [VERY IMPORTANT] The normal probability curve is bell-shaped and symmetric, with the shape of the curve to the left of the mean a mirror image of the shape of the curve to the right of the mean. (This is the basis of an important rule, called the **Symmetry Rule**, that we shall meet later.)
3. The standard deviation determines the width of the curve. Larger values of the the standard deviation result in wider flatter curves, showing more dispersion in data.
4. As with all density curves, the total area under the curve for the normal probability distribution is 1.

Characteristics of the Normal probability distribution

Remark: It is useful to know the following statements as rules of thumb, but we will do all relevant calculations from first principles. However, in an exam situation, these rules of thumb may be invoked, and it is **NOT** required to show your workings.

- The interval defined by the mean $\pm 1\times$ standard deviation includes approximately 68% of the observations, leaving 16% (approx) in each tail.
- The interval defined by the mean $\pm 1.645\times$ standard deviation includes approximately 90% of the observations, leaving 5% (approx) in each tail.
- The interval defined by the mean $\pm 1.96\times$ standard deviation includes approximately 95% of the observations, leaving 2.5% (approx) in each tail.
- The interval defined by the mean $\pm 2.58\times$ standard deviation includes approximately 99% of the observations, leaving 0.5% (approx) in each tail.

The Standard Normal Distribution

- The standard normal distribution is a special case of the normal distribution with a mean $\mu = 0$ and a standard deviation $\sigma = 1$.
- We denote the standard normal random variable as Z rather than X.

$$Z \sim N(0, 1^2)$$

- The distribution is well described in statistical tables (i.e. Murdoch Barnes Table 3, aka MB3)
- Rather than computing probabilities from first principles, which is very difficult, probabilities from distributions other than the Z distribution (e.g. X $\sim$N($\mu = 100, \sigma = 15$)) can be computed using the Z distribution, a much easier approach. (We shall demonstrate how to do this shortly.)

Standardization formula

All normally distributed random variables have corresponding Z values, called ***Z-scores***.

For normally distributed random variables, the z-score can be found using the ***standardization formula***;

$$z_o = \frac{x_o - \mu}{\sigma}$$

where x_o is a score from the underlying normal ("X") distribution, μ is the mean of the original normal distribution, and σ is the standard deviation of original normal distribution.

Therefore z_o is the z-score that corresponds to x_o.

- Terms with subscripts mean particular values, and are not variable names.
- A computed Z-score is a normally distributed random variable only if the underlying distribution (X) is normally distributed. If the underlying distribution is not normal, then using Z-scores is not a valid approach.

The Standardized Value

- Suppose that mean $\mu = 80$ and that standard deviation $\sigma = 8$.
- What is the Z-score for $x_o = 100$?

$$z_{100} = \frac{x_0 - \mu}{\sigma} = \frac{100 - 80}{8} = \frac{20}{8} = 2.5$$

- Therefore the Z score is : $z_{100} = 2.5$

Z-scores

- A Z-score always reflects the number of standard deviations above or below the mean a particular score is.
- Suppose the scores of a test are normally distributed with a mean of 50 and a standard deviation of 9
- For instance, if a person scored a 68 on a test, then they scored 2 standard deviations above the mean.
- Converting the test scores to z scores, an X value of 68 would yield:

$$Z = \frac{68 - 50}{9} = 2$$

- So, a Z score of 2 means the original score was 2 standard deviations above the mean.

The Standard Normal (Z) Distribution Tables

- Importantly, probabilities relating to the z distribution are comprehensively tabulated in ***Murdoch Barnes Table 3***.
- This is available on sulis, in the "about this module" folder.
- Given a value of k (with k usually between 0 and 4), the probability of a standard normal "Z" random variable being greater than (or equal to) k $P(Z \geq k)$ is given in Murdoch Barnes table 3 .
- Other statistical tables can be used (e.g. the Dept. of Education Tables that many student would have used in school), but they may tabulate probabilities in a different way.

An Important Identity

If two values z_o and x_o are related in the following way, for some values μ and σ,

$$z_0 = \frac{x_0 - \mu}{\sigma}$$

Then we can can say

$$P(X \geq x_o) = P(Z \geq z_o)$$

or alternatively

$$P(X \leq x_o) = P(Z \leq z_o)$$

This is fundamental to solving problems involving normal distributions.

Using Murdoch Barnes Tables 3

- For some value z_o, between 0 and 4, the Murdoch Barnes tables set 3 tabulate $P(Z \geq z_o)$
- Ideally z_o would be specified to 2 decimal places. If it is not, round to the closest value.
- We call the third digit (i.e. the digit in the second decimal place) the "second precision".

Using Murdoch Barnes Tables 3

- To compute the relevant probability we express z_o as the sum of z_o without the second precision, and the second precision.(For example $1.28 = 1.2 + 0.08$.)
- Select the row that corresponds to z_o without the second precision (e.g. 1.2).
- Select the column that corresponds to the second precision(e.g. 0.08).
- The value that contained on the intersection is $P(Z \geq z_o)$

Find $P(Z \geq 1.28)$

	...	...	0.006	0.07	0.08	0.09
...	...	...	...	...	...	...
1.0	...	...	0.1446	0.1423	0.1401	0.1379
1.1	...	...	0.1230	0.1210	0.1190	0.1170
1.2	...	...	0.1038	0.1020	0.1003	0.0985
1.3	...	...	0.0869	0.0853	0.0838	0.0823
...	...	...	...	...	...	...

Using Murdoch Barnes tables 3

- Find $P(Z \geq 0.60)$
- Find $P(Z \geq 1.64)$
- Find $P(Z \geq 1.65)$
- Estimate $P(Z \geq 1.645)$

Find $P(Z \geq 0.60)$

	0.00	0.01	0.02	0.03	…	…
…	…	…	…	…	…	…
0.4	0.3446	0.3409	0.3372	0.3336	…	…
0.5	0.3085	0.3050	0.3015	0.2981	…	…
0.6	0.2743	0.2709	0.2676	0.2643	…	…
0.7	0.2420	0.2389	0.2358	0.2327	…	…
…	…	…	…	…	…	…

Find $P(Z \geq 1.64)$ and $P(Z \geq 1.65)$

	...	...	0.04	0.05	0.06	0.07
...	...	...	...	...	...	...
1.5	...	0.0630	0.0618	0.0606	0.0594	...
1.6	...	0.0516	0.0505	0.0495	0.0485	...
1.7	...	0.0418	0.0409	0.0401	0.0392	...
...	...	...	...	...	...	...

Using Murdoch Barnes tables 3

- $P(Z \geq 1.64) = 0.505$
- $P(Z \geq 1.65) = 0.495$

- $P(Z \geq 1.645)$ is approximately the average value of $P(Z \geq 1.64)$ and $P(Z \geq 1.65)$.
- $P(Z \geq 1.645)$ = (0.0495 + 0.0505)/2 = 0.0500. (i.e. 5%)

Exact Probability

Remarks: This is for continuous distributions only.

- The probability that a continuous random variable will take an exact value is infinitely small. We will usually treat it as if it was zero.
- When we write probabilities for continuous random variables in mathematical notation, we often retain the equality component (i.e. the "...or equal to..").
  For example, we would write expressions $P(X \leq 2)$ or $P(X \geq 5)$.
- Because the probability of an exact value is almost zero, these two expression are equivalent to $P(X < 2)$ or $P(X > 5)$.
- The complement of $P(X \geq k)$ can be written as $P(X \leq k)$.

Complement and Symmetry Rules

Any normal distribution problem can be solved with some combination of the following rules.

- **Complement rule**
- Common to all continuous random variables

$$P(Z \geq k) = 1 - P(Z \leq k)$$

Similarly

$$P(X \geq k) = 1 - P(X \leq k)$$

$$P(Z \leq 1.28) = 1 - P(Z \geq 1.28) = 1 - 0.1003 = 0.8997$$

Complement and Symmetry Rules

- **Symmetry rule**
- This rule is based on the property of symmetry mentioned previously.
- Only the probabilities corresponding to values between 0 and 4 are tabulated in Murdoch Barnes.
- If we have a negative value of k, we can use the symmetry rule.

$$P(Z \leq -k) = P(Z \geq k)$$

by extension, we can say

$$P(Z \geq -k) = P(Z \leq k)$$

Z Scores: Example 1

Find $P(Z \geq -1.28)$

Solution

- Using the symmetry rule

$$P(Z \geq -1.28) = P(Z \leq 1.28)$$

- Using the complement rule

$$P(Z \geq -1.28) = 1 - P(Z \geq 1.28)$$

$$P(Z \geq -1.28) = 1 - 0.1003 = 0.8997$$

Z Scores: Example 2

Find the probability of a Z random variable being between -1.8 and 1.96? i.e. Compute $P(-1.8 \leq Z \leq 1.96)$

Solution

- Consider the complement event of being in this interval: a combination of being too low or too high.
- The probability of being too low for this interval is $P(Z \leq -1.80) = 0.0359$ (check)
- The probability of being too high for this interval is $P(Z \geq 1.96) = 0.0250$ (check)
- Therefore the probability of being **outside** the interval is 0.0359 + 0.0250 = 0.0609.
- Therefore the probability of being **inside** the interval is 1- 0.0609 = 0.9391

$$P(-1.8 \leq Z \leq 1.96) = 0.9391$$

Application : Example

The mean time spent waiting by customers before their queries are dealt with at an information centre is 10 minutes.
The waiting time is normally distributed with a standard deviation of 3 minutes.

i) What percentage of customers will be waiting longer than 15 minutes
ii) 90% of customers will be dealt with in at most 12 minutes. Is this statement true or false? Justify your answer.
iii) What percentage of customers will wait between 7 and 13 minutes before their query is dealt with?

Solutions

Let x be the normal random variable describing waiting times
$P(X \geq 15) = ?$

First , we find the z-value that corresponds to x = 15 (remember $\mu = 10$ and $\sigma = 3$)

$$z_o = \frac{x_o - \mu}{\sigma} = \frac{15 - 10}{3} = 1.666$$

- We will use $z_o = 1.67$
- Therefore we can say $P(X \geq 15) = P(Z \geq 1.67)$
- The Murdoch Barnes tables are tabulated to give $P(Z \geq z_o)$ for some value z_o .
- We can evaluate $P(Z \geq 1.67)$ as 0.0475.
- Necessarily $P(X \geq 15) = 0.0475$.

Solutions

- ”90% of customers will be dealt with in at most 12 minutes.”
- To answer this question, we need to know $P(X \leq 12)$
- First , we find the z-value that corresponds to x = 12 (remember $\mu = 10$ and $\sigma = 3$)

$$z_o = \frac{x_o - \mu}{\sigma} = \frac{12 - 10}{3} = 0.666$$

Solutions

- We will use $z_o = 0.67$ (although 0.66 would be fine too)
- Therefore we can say $P(X \geq 12) = P(Z \geq 0.67) = 0.2514$
- Necessarily $P(X \leq 12) = P(Z \leq 0.67) = 0.7486$
- 74.86% of customers will be dealt with in at most 12 minutes.
- The statement that 90% will be dealt with in at most 12 minutes is false.

Solutions

What percentage will wait between 7 and 13 minutes ?
$P(7 \leq X \leq 13) = ?$
Solution:
Compute the probability of being too low, and the probability of being too high for the interval.
The probability of being inside the interval is the complement of the combination of these events.

Solutions

Too high:

$P(X \geq 13) = ?$

$$z_o = \frac{13-10}{3} = 1$$

From tables, $P(Z \geq 1) = 0.1587$. Therefore $P(X \geq 13) = 0.1587$

Too low:

$P(X \leq 7) = ?$

$$z_o = \frac{7-10}{3} = -1$$

By symmetry, and using tables, $P(X \leq 7) = P(Z \leq -1) = 0.1587$

Solutions

$$P(7 \leq X \leq 13) = 1 - [P(X \leq 7) + P(X \geq 13)]$$

$$P(7 \leq X \leq 13) = 1 - [0.1587 + 0.1587] = 0.6826$$

Normal Distribution : Solving problems

Recap:

- We must know the normal mean μ and the normal standard deviation σ.
- The normal random variable is $X \sim \mathrm{N}(\mu, \sigma^2)$.
- (If we don't, we usually have to determine them, given the information in the question.)
- The standard normal random variable is $Z \sim \mathrm{N}(0, 1^2)$.
- The standard normal distribution is well described in Murdoch Barnes Table 3, which tabulates $P(Z \geq z_o)$ for a range of Z values.

Normal Distribution : Solving problems

- For the given value x_o from the variable X, we compute the corresponding z-score z_o.

$$z_o = \frac{x_o - \mu}{\sigma}$$

- When z_o corresponds to x_o, the following identity applies:

$$P(X \geq x_o) = P(Z \geq z_o)$$

- Alternatively $P(X \leq x_o) = P(Z \leq z_o)$

Normal Distribution : Solving problems

- **Complement Rule**:

$$P(Z \leq k) = 1 - P(Z \geq k)$$

  for some value k

- Alternatively $P(Z \geq k) = 1 - P(Z \leq k)$
- **Symmetry Rule**:

$$P(Z \leq -k) = P(Z \geq k)$$

  for some value k

- Alternatively $P(Z \geq -k) = P(Z \leq k)$

Normal Distribution : Solving problems

- **Intervals**:

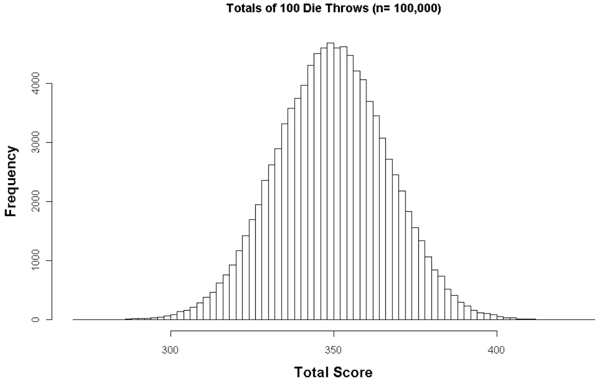
$$P(L \leq Z \leq U) = 1 - [P(Z \leq L) + P(Z \geq U)]$$

  where L and U are the lower and upper bounds of an interval.
- Probability of having a value too low for the interval : $P(Z \leq L)$
- Probability of having a value too high for the interval : $P(Z \geq U)$

Normal Distribution: Simulation Study

- Recall the experiment whereby a die was rolled 100 times, and the sum of the 100 values was recorded.
- This experiment was repeated a very large number of times (e.g. 100,000 times) in a simulation study.
- A histogram was drawn to depict the distribution of outcomes of this experiment.

Normal Distribution: Simulation Study



Normal Distribution: Simulation Study

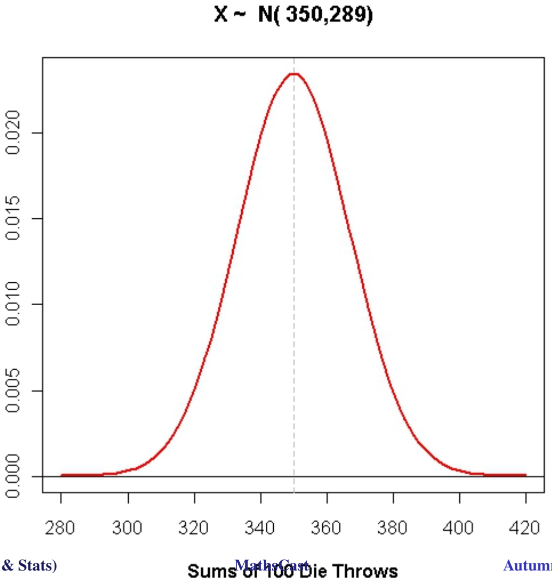
Recall some observations made about the results of the simulation study, made in a previous lecture.

- Approximately 68.7% of the values in the simulation study are between 332 and 367.
- Approximately 95% of the values are between 316 and 383.
- 2.5% of the values output are less than 316.
- 2.5% of the values study output are greater than 383.
- 175 values are greater than or equal to 400, whereas 198 values are less than or equal to 300.
- Results such as these are unusual, but they are not impossible.

Normal Distribution: Simulation Study

- Suppose we can ***approximate*** the summation of the die-throws using the normal distribution.
- The normal mean is necessarily $\mu = 350$.
- The normal standard deviation is approximately 17. (68% of values between 350 ± 17).
- Using the normal distribution, lets estimate the proportion of values greater than 383.

Normal Distribution: Simulation Study



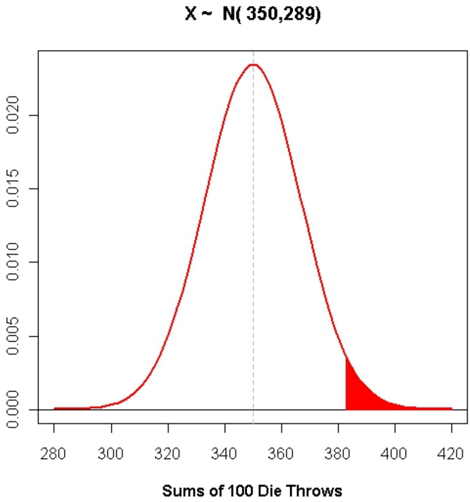
Normal Distribution: Simulation Study

- X is the normal random variable that approximates the sum of values from 100 throws of a die.
- Find $P(X \leq 383)$
- First use the standardization formula to find the Z-score.

$$z_o = \frac{383 - 350}{17} = \frac{33}{17} = 1.94$$

- Use the tables to compute $P(Z \geq 1.94)$ ($\mathit{Answer} : 0.0262$)
- Because $P(Z \geq 1.94) = 0.0262$, we can say $P(X \geq 383) = 0.0262$
- This is close to the proportion of observed values, which was 2.5%.
- Remark : The standard deviation of 17 was an estimate. The actual standard deviation should 17.12.

Normal Distribution: Simulation Study



Working Backwards

- Suppose we wish to find a value (lets call it A) from the normal distribution, such that a certain proportion of values is greater than A (e.g. 10%)
- Find A such that $P(X \geq A) = 0.10$. (with $\mu = 350$ and $\sigma = 17$)
- In general, our first step is to use the standardization equation to find the corresponding Z-score z_A.
- Because we don't know what value A has, we can't use this approach.
- However, we can say the following

$$P(X \geq A) = P(Z \geq z_A) = 0.10$$

- From the tables, we can approximate a value for z_A, by finding the closest probability value, and determining the corresponding Z-score.

Find z_A such that $P(Z \geq z_a) = 0.10$

- The closest probability value in the tables is 0.1003.
- The Z-score that corresponds to 0.1003 is 1.28.
- (Row : 1.2 , Column : 0.08)
- Therefore $z_A \approx 1.28$

	...	...	0.006	0.07	0.08	0.09
...	...	...	...	...	...	...
1.0	...	...	0.1446	0.1423	0.1401	0.1379
1.1	...	...	0.1230	0.1210	0.1190	0.1170
1.2	...	...	0.1038	0.1020	0.1003	0.0985
1.3	...	...	0.0869	0.0853	0.0838	0.0823
...	...	...	...	...	...	...

Working Backwards

- We can now use the standardization formula.
- We have only one unknown in the formula: A.

$$1.28 = \frac{A - 350}{17}$$

- Re-arranging (multiply both sides by 17):
  $21.76 = A - 350$
- Re-arranging (add 350 to both sides):
  $A = 371.76$
- $P(X \geq 371.76) \approx 0.10$
- (Remark: for sums of die-throws, round it to nearest value)

Working Backwards: Another Example

- Find B such that $P(X \geq B) = 0.90$. (with $\mu = 350$ and $\sigma = 17$)
- Necessarily $P(X \leq B) = 0.10$
- Find some value Z_B such that $P(Z \leq z_B) = 0.10$
- z_B could be negative.
- Use the symmetry rule $P(Z \leq z_B) = P(Z \geq -z_B)$
- $-z_B$ could be positive.
- Based on last example $-z_B = 1.28$. Therefore $z_B = -1.28$

Working Backwards

- Again ,we can now use the standardization formula
- We have only one unknown in the formula: B.

$$-1.28 = \frac{B - 350}{17}$$

- Re-arranging (multiply both sides by 17):
  $-21.76 = B - 350$
- Re-arranging (add 350 to both sides):
  $x_o = 350 - 21.76 = 328.24$
- $P(X \leq 328.24) \approx 0.10$

MA4413 Autumn 2008 paper

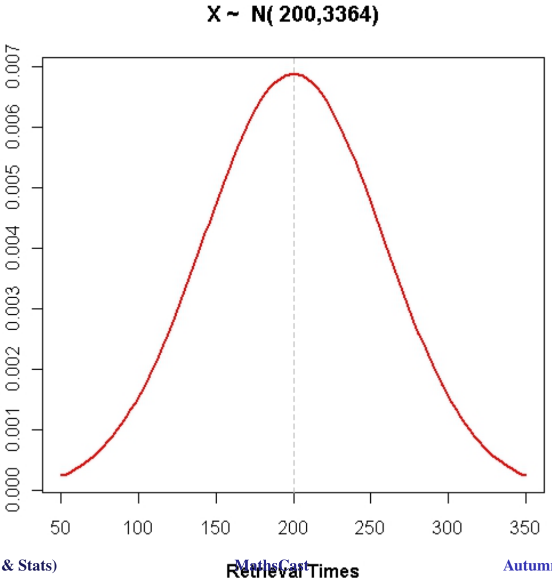
A model of an on-line computer system gives a mean times to retrieve a record from a direct access storage system device of 200 milliseconds, with a standard deviation of 58 milliseconds. If it can assumed that the retrieval times are normally distributed:

(i) What proportion of retrieval times will be greater than 75 milliseconds?

(ii) What proportion of retrieval times will be between 150 and 250 milliseconds?

(iii) What is the retrieval time below which 10% of retrieval times will be?

Normal Distribution



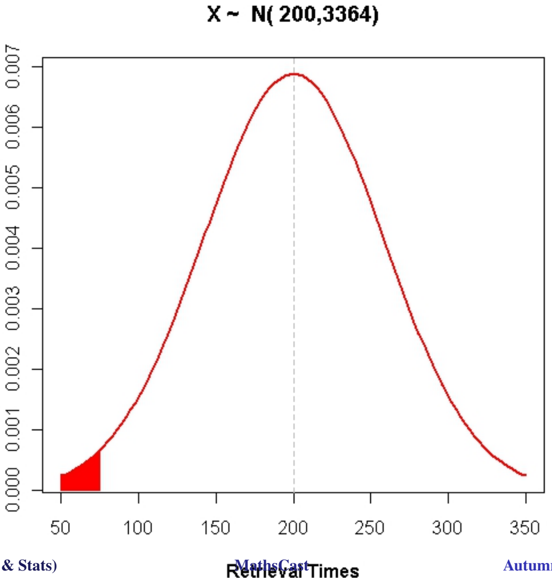
MA4413 Autumn 2008 paper (part 1)

What proportion of retrieval times will be greater than 75 milliseconds?

- Let X be the retrieval times, with $X \sim \mathrm{N}(200, 58^2)$.
- The first question asks us to find $P(X \geq 75)$.
- First compute the z score.

$$z_o = \frac{x_o - \mu}{\sigma} = \frac{75 - 200}{58} = -2.15$$

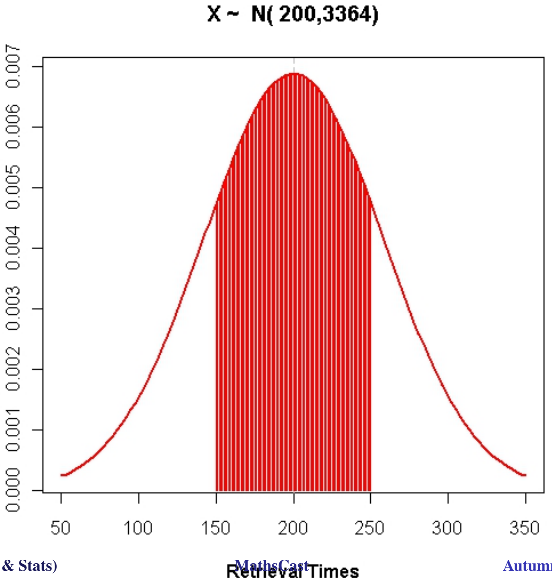
Normal Distribution



MA4413 Autumn 2008 paper (part 1)

- We can say
$$P(X \geq 75) = P(Z \geq -2.15)$$
- Using symmetry rule and complement rule
$$P(Z \geq -2.15) = P(Z \leq 2.15) = 1 - P(Z \geq 2.15)$$
- From tables $P(Z \geq 2.15) = 0.0158$
- Therefore $P(Z \leq 2.15) = 0.9842$
- Furthermore $P(X \geq 75) = \mathbf{0.9842}$ [Answer].

Normal Distribution



MA4413 Autumn 2008 paper (part 2)

- What proportion of retrieval times will be between 150 and 250 milliseconds?
- Find $P(150 \leq X \leq 250)$
- Use the 'Too Low / Too High ' approach.
- Too low $P(X \leq 150)$
- Too high $P(X \geq 250)$
- Find the z-scores for each.

$$z_{150} = \frac{150 - 200}{58} = -0.86$$

$$z_{250} = \frac{250 - 200}{58} = 0.86$$

MA4413 Autumn 2008 paper (part 2)

- We can now say

$$1.P(X \leq 150) = P(Z \leq -0.86)$$

$$2.P(X \geq 250) = P(Z \geq 0.86)$$

- By symmetry rule, $P(Z \leq -0.86) = P(Z \geq 0.86)$

$$P(X \leq 150) = P(X \geq 250)$$

- Let's compute $P(X \geq 250)$. Using tables

$$P(X \geq 250) = P(Z \geq 0.86) = 0.1949$$

MA4413 Autumn 2008 paper (part 2)

- Too high: $P(X \geq 250) = 0.1949$
- Too low: $P(X \leq 150) = 0.1949$
- Probability of being inside interval:

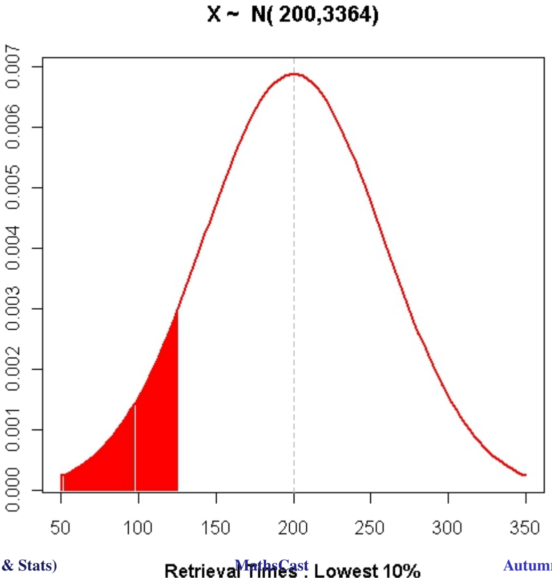
$$P(150 \leq X \leq 250) = 1 - [P(X \leq 150) + P(X \geq 250)]$$

- $P(150 \leq X \leq 250) = 1 - [0.1949 + 0.1949] = \mathbf{0.6102}$

MA4413 Autumn 2008 paper (part 3)

- What is the retrieval time below which 10% of retrieval times will be?
- Find A such that $P(X \leq A) = 0.10$.
- What z-score would correspond to A? Lets call it z_A.
- $P(Z \leq z_A) = 0.10$
- Remark: z_A could be negative.
- Using symmetry $P(Z \geq -z_A) = 0.10$
- Remark: $-z_A$ could be positive.

Normal Distribution



MA4413 Autumn 2008 paper (part 3)

- Use the Murdoch Barnes tables to get an approximate value for $-z_A$.
- The nearest value we can get is 1.28. ($P(Z \geq 1.28) = 0.1003$).
- If $-z_A = 1.28$, then $z_A = -1.28$
- We can now say
$$P(X \leq A) = P(Z \leq -1.28)$$

MA4413 Autumn 2008 paper (part 3)

- Necessarily A and Z_A are related by the standardization formula
- Recall that $\mu = 200$ and $\sigma = 58$.

$$-1.28 = \frac{A - 200}{58}$$

- Re-arranging (multiply both sides by 58)

$$-74.24 = A - 200$$

- Re-arranging again (Add 200 to both sides)

$$125.76 = A$$

MA4413 Autumn 2008 paper (part 3)

- Now we know the retrieval time below which 10% of retrieval times will be.
- $P(X \leq 125.76) = 0.10$ [Answer].