Chapter 1

Introduction to Probability

Conceptual approaches

- Historically, three different conceptual approaches have been developed for defining probability and for determining probability values: the classical, relative frequency, and subjective approaches.

- If N(A) possible elementary outcomes are favorable to event A, N(S) possible outcomes are included in the sample space, and all the elementary outcomes are equally likely and mutually exclusive, then the probability that event A will occur is

$$P(A) = \frac{N(A)}{N(S)}$$

Examples

When a fair dice is thrown, what are the possible outcomes? There are 6 possible outcomes. The dice can role any number between one and six. Each outcome is equally likely. The probability of each outcome is 1/6.

In a well-shuffled deck of cards which contains 4 aces and 48 other cards, the probability of an ace (A) being obtained on a single draw is;

$$P(A) = N(A)/N(S) = 4/52 = 1/13$$

1.0.1 Histograms

A histogram is constructed from a frequency table. The intervals are shown on the X-axis and the number of scores in each interval is represented by the height of a rectangle located above the interval. A histogram of the response times from the dataset Target RT is shown below.

1.0.2 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:

$$F(t) = P(X \leq 1)$$

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.

Example

In the above example where the die is thrown repeatedly, lets work out $P(X \leq t)$ for some values of t.

$P(X \leq 1)$ is the probability that the number of throws until we get a 6 is less than or equal to 1. So it is either 0 or 1.

- $P(X = 0) = 0$

- $P(X = 1) = 1/6$.

- Hence $P(X \leq 1) = 1/6$

Similarly, $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
$= 0 + 1/6 + 5/36 = 11/36$

1.1 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:

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

From the first year intake example, check that

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

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

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

1.2 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

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

1.3   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.

Examples

When a fair dice is thrown, what are the possible outcomes? There are 6 possible outcomes. The dice can role any number between one and six. Each outcome is equally likely. The probability of each outcome is 1/6.

   In a well-shuffled deck of cards which contains 4 aces and 48 other cards, the probability of an ace (A) being obtained on a single draw is;

$$P(A) = N(A)/N(S) = 4/52 = 1/13$$

1.4   How to Compute Probability: Equally Likely Outcomes

- Sometimes, a statistical experiment can have n possible outcomes, each of which is equally likely. Suppose a subset of r outcomes are classified as "successful" outcomes.

- The probability that the experiment results in a successful outcome (S) is:

$$P(S) = \frac{\text{Number of successful outcomes}}{\text{Total number of equally likely outcomes}} = r/n$$

- Consider the following experiment. An urn has 10 marbles. Two marbles are red, three are green, and five are blue. If an experimenter randomly selects 1 marble from the urn, what is the probability that it will be green?

- In this experiment, there are 10 equally likely outcomes, three of which are green marbles. Therefore, the probability of choosing a green marble is 3/10 or 0.30.

1.5   Basic of Compound Events

- pairwise disjoint sets

- The addition principle

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

1.6 Probability Functions

In general there are four types of functions related to probability distributions:

1. Random Number Generation (which we have covered as far as we are going to with MATLAB already).

2. Probability Density Funtion,

3. Cumulative Distribution Function,

4. Inverse Cumulative Distribution Function (also known as the Quantile Function).

Probability Density Funtion
The probability density function (or "density" for short, and denoted f) is a measure of the likelihood of a particular value. For Discrete Probability distribution (where the set of all possible outcomes are integers), it is the probability of a random variable taking a specific value.

$$f(x) = P(X = x)$$

Cumulative Distribution Function
The cumulative distribution function (or "distribution" for short, and denoted F) is the probability that a random variable X will not exceed a certain specified value x.

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

Inverse Cumulative Distribution Function
The Inverse Cumulative Distribution Function, also known as the quantile function specifies, for a given probability in the probability distribution of a random variable, the value at which the probability of the random variable will be less than or equal to that probability.

For specified probability level k where $0 \leq k \leq 1$, the quantile function yields the value x such that

$$P(X \leq x) = k$$

In MATLAB, the density commands have `inv` as the second part of the name.
Some Examples

When using the normal distribution, the relevant commands have `norm` as the first part of their name. If no additional arguments are specified (i.e. for mean and standard deviation), then the default values are used. These default settings are $\mu = 0$ and $\sigma = 1$, the specifications for the Standard Normal (Z) distribution.

For the Poisson distribution, the functions have `poiss` at the start of their names (i.e. `poisscdf`).

The additional parameter for the Poisson distribution is the `lambda` (rate) parameter.

```
% Z distribution.
norminv(0.95)
norminv(0.975)


poisspd(4,lambda=5)
```

Exercises

1. Compute the following value : $P(Z \leq 1.50)$ where Z is the standard normal distribution.

2. Determine the value for k such that $P(Z \leq k) = 0.90$.

3. Find an interval that contains 95% of the values from a standard normal distribution.

4. Where X is a Poisson Random Variable with rate parameter `lambda` $= 3$, compute the probaility of $P(X = 1)$, $P(X = 3)$ and $P(X = 5)$.

5. Use the help command to determine the additional arguments for functions related to the Binomial distribution.

- An example of the use of probability is in decision making. Decision making usually involves uncertainty. For example, should we invest in a company if there is a chance it will fail?

- Should we start production of a product even though there is a likelihood that the raw materials will arrive on time in poor? Having a number which measures the chances of these events occurring helps us to make a decision.

- Why are we interested in probability in this module? Many statistical methods use the idea of a probability distribution for this data.

- We have already looked at relative frequency distribution in Section 2. Probability distributions are based on the same concepts as relative frequency distributions. They are used to calculate probabilities of different values occurring in the data collected.

- We will examine probability distributions in more detail in Section 4. First we need to learn about the basic concepts of probability.

- Permutations where repetition is allowed:
$$n!$$

- Permutations where repetition is not allowed

$$\frac{n!}{(n - k)!}$$

Worked Example

Two people look at the letters in the word discovery. Independently of each other, each person writes down two of the letters from the word discovery. What is the probability that

(i) One person writes down two vowels and the other person

(ii)

Three cards are drawn, one after the other, without replacement, from a pack of 52 playing cards. Find the probability that the
Discrete probability distributions
The discrete probability distributions that described in this course are

- the binomial distribution,

- the geometric distribution,

- the hypergeometric distribution,

- the Poisson distributions.

Given that event B has already occurred , what is the probability of event A
P(A — B) probability of A given B
P(A and B) probability of A and B
$\widehat{C}$omplement event

What is the probability of A not happening What is the probability of outcomes not included in A.
Section 4: Probability Distributions
Expected value of the outcome
E[X]
By definition the expected value is the value that 50

$$P(X \geq E[X]) = 0.50$$

$$P(X \leq E[X]) = 0.50$$

1.7   Section 2 : Descriptive Statistics

Descriptive Statistics

1. Mean

2. Variance and standard deviation

3. Median and IQR

4. Histograms and Barcharts

5. Quantile statistics

- Define quota sampling. In what circumstances would you use it?

- In what circumstances would you use stratified random sampling?

- Give two ways in which stratified random sampling differs from quota sampling.

1.8   Continuous Joint Probability Distribution

Continuous Joint Probability Distributions arise from groups of continuous random variables. Continuous joint probability distributions are characterized by the Joint Density Function, which is similar to that of a single variable case, except that this is in two dimensions. The joint density function $f(x, y)$ is characterized by the following:

Section 3 : Probability

How to Compute Probability: Equally Likely Outcomes Sometimes, a statistical experiment can have n possible outcomes, each of which is equally likely. Suppose a subset of r outcomes are classified as "successful" outcomes.

The probability that the experiment results in a successful outcome (S) is:

$$P(S) = \frac{(Number of successful outcomes)}{(Total number of equally likely outcomes)} = r/n$$

Consider the following experiment. An urn has 10 marbles. Two marbles are red, three are green, and five are blue. If an experimenter randomly selects 1 marble from the urn, what is the probability that it will be green?

In this experiment, there are 10 equally likely outcomes, three of which are green marbles. Therefore, the probability of choosing a green marble is 3/10 or 0.30.

1.9 Rules of 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

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

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.

part 2 : Discrete Probability distributions

1.10 Section 6 : Bivariate Analysis

Pearson correlation coefficient Scatterplots Simple linear regression

Correlation Coefficient

Strength of a linear relationship between X and Y

```
M=1000
CorrData=numeric(M)
for (i in 1:M)
{
CorrData[i] = cor(rnorm(10),rnorm(10))
}
```

Exponential Distribution

This distribution is useful in modelling lifetimes.

Lecture 8A

In the last class, we looked at how to compute the mean, variance and standard deviation.

As these are key outcomes of this part of the course, we shall briefly go over this material again

The mean

The mean (i.e. average) value is denoted with a bar over the set name i.e. " ".

(pronounced x bar) is the sample mean.

x=c(15, 34, 7, 12, 18, 9, 1, 42, 56, 28, 13, 24, 35) boxplot(x, horizontal =TRUE)

Reminder: Mid Term Exam is tomorrow (Wednesday 16th March) at 9 a.m. 15 Multiple Choice Questions Answer all 15 Worth 15No Calculators Allowed Duration : 45 minutes

Last Week Section 1 - Introduction to Statistics and Probability ImportantDiscrete variables V Continuous variables Samples and Populations Section 2 Graphical Methods Frequency Distributions Bar Charts , Pie Charts and Histograms Shape of Histograms (Symmetric / Left Skewed / Right Skewed) This Class Completion of graphical methods Ogives Numerical Methods Measures of Centrality Measures of Variability The Ogive A graph of the cumulative frequency distribution or the cumulative relative frequency distribution is called an Ogive.

Audit Data ClassCumul. FreqCum. Rel. Freq X¡1540.2 X¡20120.6 X¡25170.85 X¡30190.95 X¡35201.00 [OVERHEAD] We have already used graphical methods 1)To compare the numbers in different categories (e.g. the number of employees with different levels of education working in a company. 2)To look at the frequency distribution of measurements in different class intervals e.g. the frequency of audit time for an accounting firm.

We can also use graphical methods 1)To express the change of some quantity of over a period of time e.g. profits of a company over the past decade.(Trendlines)

2)To express the relationship between two measurements where they occur in pairs. E.g. profits of a company with the amount spent on research and development. (Scatterplots)

[Pgs 25]

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.

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 $\boldsymbol{S}$, then

- $A \subseteq B$(A is a subset of B) = 'if A occurs, so does B'.

- $\oslash$ (the empty set) = an impossible event.

- S (the sample space) = an event that is certain to occur.

1.11   Independent and Mutually Exclusive Events

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

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

- There are many situations in everyday life where the outcome is not known with certainty. For example; applying for a job or sitting an examination.

- We use words like "Chance", "the odds", "likelihood" etc but the most effective way of dealing with uncertainty is based on the concept of probability.

- Probability can be thought of as a number which measures the chance or likelihood that a particular event will occur.

- An example of the use of probability is in decision making. Decision making usually involves uncertainty. For example, should we invest in a company if there is a chance it will fail?

- Should we start production of a product even though there is a likelihood that the raw materials will arrive on time in poor? Having a number which measures the chances of these events occurring helps us to make a decision.

- Why are we interested in probability in this module? Many statistical methods use the idea of a probability distribution for this data.

- We have already looked at relative frequency distribution in Section 2. Probability distributions are based on the same concepts as relative frequency distributions. They are used to calculate probabilities of different values occurring in the data collected.

- We will examine probability distributions in more detail in Section 4. First we need to learn about the basic concepts of probability.

- There are many situations in everyday life where the outcome is not known with certainty. For example; applying for a job or sitting an examination. We use words like "Chance", "the odds", "likelihood" etc but the most effective way of dealing with uncertainty is based on the concept of probability. Probability can be thought of as a number which measures the chance or likelihood that a particular event will occur.

- An example of the use of probability is in decision making. Decision making usually involves uncertainty. For example, should we invest in a company if there is a chance it will fail?

- Should we start production of a product even though there is a likelihood that the raw materials will arrive on time in poor? Having a number which measures the chances of these events occurring helps us to make a decision.

- Why are we interested in probability in this module? Many statistical methods use the idea of a probability distribution for this data.

- We have already looked at relative frequency distribution in Section 2. Probability distributions are based on the same concepts as relative frequency distributions. They are used to calculate probabilities of different values occurring in the data collected.

- We will examine probability distributions in more detail in Section 4. First we need to learn about the basic concepts of probability.

- Probability concerns itself with random phenomena or probability experiments. These experiments are all different in nature, and can concern things as diverse as rolling dice or flipping coins. The common thread that runs throughout these probability experiments is that there are observable outcomes. If we collect all of the possible outcomes together, then this forms a set that is known as the sample space.

- In this set theory formulation of probability the sample space for a problem corresponds to an important set. Since the sample space contains every outcome that is possible, it forms a setting of everything that we can consider. So the sample space becomes the universal set in use for a particular probability experiment.

1.12   Basic definitions of probability

- The symbol P is used to designate the probability of an event. Thus P(A) denotes the probability that event A will occur in a single observation or experiment.

- The smallest value that a probability statement can have is 0 (indicating the event is impossible) and the largest value it can have is 1 (indicating the event is certain to occur). Thus, in general: $0 \leq P(A) \leq 1$

- The smallest value that a probability statement can have is 0 (indicating the event is impossible) and the largest value it can have is 1 (indicating the event is certain to occur).

- Thus, in general: $0 \leq P(A) \leq 1$

- In a given observation or experiment, an event must either occur or not occur. Therefore, the sum of the probability of occurrence plus the probability of nonoccurrence always equals 1. Thus, where A' indicates the nonoccurrence of event A, we have $P(A) + P(A') = 1$

1.13 Basic of Compound Events

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

1.14 The Complement Rule

- The probability of an event not occurring is one minus the probability of it occurring.

$$P(E^C) = 1 - P(E)$$

1.15 Complement Rule

The complement rule in Probability

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

If the probability of C is 70% then the probability of C' is 30%

1.16 Combining Probabilities

Events rarely occur in isolation. Usually we are interested in a combination or compound of events; for example

- The probability that two sections of a factory will be understaffed on the same day

- The probability of having a car accident today, given that you have had a car accident in the last five years.

We will look at two laws of probability for combining events

- The Addition Law

- The multiplication Law

1.16.1 Probability trees

The setting out of solutions to problems requiring the manipulation of the probabilities of mutually exclusive and independent events can sometimes be helped by the use of probability tree diagrams. These have useful applications in decision theory.

The best choice of probability tree structure often depends upon the question and the natural order in which events like A and B above occur.

Example

In the above example where the die is thrown repeatedly, lets work out $P(X \leq t)$ for some values of t.

P(X $\leq$ 1) is the probability that the number of throws until we get a 6 is less than or equal to 1. So it is either 0 or 1.

- P(X = 0) = 0

11

- $P(X = 1) = 1/6$.

- Hence $P(X \leq 1) = 1/6$

Similarly, $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$
$= 0 + 1/6 + 5/36 = 11/36$

- If two events are mutually exclusive, then the probability of either occurring is the sum of the probabilities of each occurring.

- Specific Addition Rule: Only valid when the events are mutually exclusive.

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

Probability Distribution

A probability distribution is a table of values showing the probabilities of various outcomes of an experiment.

For example, if a coin is tossed three times, the number of heads obtained can be 0, 1, 2 or 3. The probabilities of each of these possibilities can be tabulated as shown:

Number of Heads	0	1	2	3
Probability	1/8	3/8	3/8	1/8

A discrete variable is a variable which can only take a countable number of values. In this example, the number of heads can only take 4 values (0, 1, 2, 3) and so the variable is discrete. The variable is said to be random if the sum of the probabilities is one.

1.16.2 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:

$$F(t) = P(X \leq 1)$$

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.

1.17 Sample Space

A probability distribution is a table of values showing the probabilities of various outcomes of an experiment.

For example, if a coin is tossed three times, the number of heads obtained can be 0, 1, 2 or 3. The probabilities of each of these possibilities can be tabulated as shown:

Number of Heads	0	1	2	3
Probability	1/8	3/8	3/8	1/8

1.18   What is a contingency table?

A contingency table is essentially a display format used to analyse and record the relationship between two or more categorical variables. It is the categorical equivalent of the scatterplot used to analyse the relationship between two continuous variables.

1.19   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:

$$F(t) = P(X \leq 1)$$

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.

1.20   Basic Probability

1. Axioms of Probability

2. Basic Rules of Probability

3. Conditional Probability

1.20.1   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.

1.20.2   Using Confidence Limits

- Alternatively, we can use the confidence interval to make a decision on whether or not we should reject or fail to reject the null hypothesis.

- If the null value is within the range of the confidence limits, we fail to reject the null hypothesis.

- If the null value is outside the range of the confidence limits, we reject the null hypothesis.

- Occasionally a conclusion based on this approach may differ from a conclusion based on the p-value. In such a case, remark upon this discrepancy.

Kolmogorov Smirnov Tests Non-parametric procedures Compute the test statistic Same distribution

1.20.3   Normal Probability Plot (QQ plot)

- The normal probability plot is a graphical technique for assessing whether or not a data set is approximately normally distributed.

- The data are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line.

- Departures from this straight line indicate departures from normality.

- The relevant R functions are `qqnorm()` and `qqline()`.