Statistical Learning - Probability

How to Assess Uncertainty Using Probability?

Introduction

- Managers will have to cope with uncertainty in many decision situations. Concepts of probability will help you measure uncertainty and perform associated analysis that are essential in making effective business decisions.

Probability – Meaning and Concepts

greatlearning
Learning for Life

- Probability refers to chance or likelihood of a particular event – taking place
- An event is an outcome of an experiment.
- An experiment is a process that is performed to understand and observe possible outcomes.
- Set of all outcomes of an experiment is called the sample space.

Example

- In a manufacturing unit three parts from the assembly are selected. You are observing whether they are defective or non-defective. Determine

  a) The sample space.

  b) the event of getting at least two defective parts.

Solution

a) Let S = Sample space. It is pictured as under

GGG	GGD	GDG	DGG
GDD	DGD	DDG	DDD

D – Defective

G – Non-defective

b) Let E denote the event of getting at least two defective parts. This implies that E will contain two defective, and three defectives. Looking at the sample diagram above, E = { GDD, DGD, DDG, DDD}. It is easy to see that E is a part of S and commonly called as a subset of S. Hence an event is always a subset of the sample space

Definition of probability

greatlearning
Learning for Life

- Probability of an event A is defined as the ratio of two numbers m and n. In symbols

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

Where m = number of ways that are favourable to the occurrence of A and n = the total number of outcomes of the experiment (All possible outcomes)

Please note that P(A) is always >= 0 and always <=1.

P(A) is a pure number.

Portability values

- Probability 0 – Impossible event
- Probability 1: Certain event
- Probability always in the range of 0 to 1

Assessing Probability

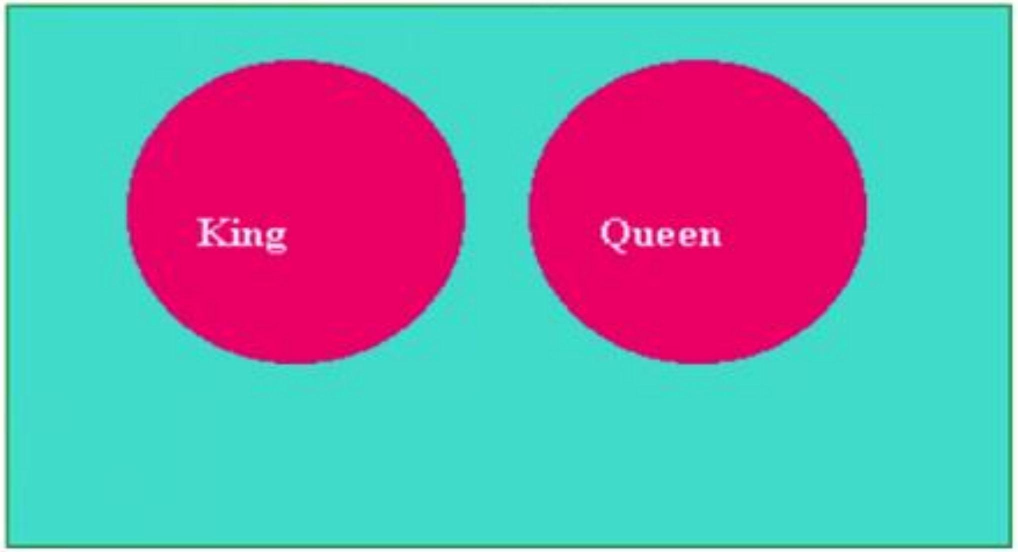
greatlearning
Learning for Life

- A Priori classical probability: Based on knowledge of the process
- Empirical probability: Based on data
- Subjective probability: Based on experience, analysis of situation and expert opinion

greatlearning
Learning for Life

Mutually exclusive events

- Two events A and B are said to be mutually exclusive if the occurrence of A precludes the occurrence of B. For example, from a well shuffled pack of cards, if you pick up one card at random and would like to know whether it is a King or a Queen. The selected card will be either a King or a Queen. It cannot be both a King and a Queen. If King occurs, Queen will not occur and Queen occurs, King will not occur.



Independents events

- Two events A and B are said to be independent if the occurrence of A is in no way influenced by the occurrence of B. Likewise occurrence of B is in no way influenced by the occurrence of A.

Rules for computing probability

greatlearning
Learning for Life

- **1) Addition Rule – Mutually exclusive events**

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

This rule says that the probability of the union of A and B is determined by adding the probability of the events A and B.

Here the symbol A U B is called A union B meaning A occurs, or B occurs or both A and B simultaneously occur. When A and B are mutually exclusive, A and B cannot simultaneously occur.

Rules for computing probability

- **1) Addition Rule – Events are not Mutually exclusive events**

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

This rule says that the probability of the union of A and B is determined by adding the probability of the event A and B and then subtracting the probability of the intersection of the events A and B.

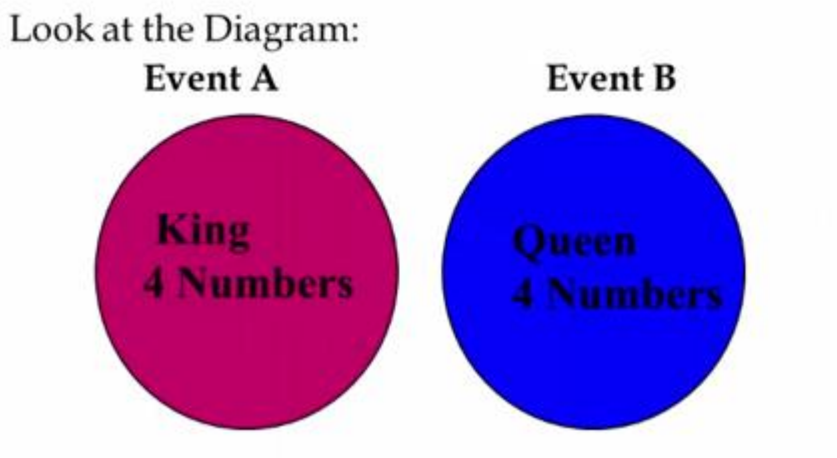
The symbol $A \cap B$ is called A intersection B meaning both A and B simultaneously occur.

Example of Addition Rules:

- From a pack of well-shuffled cards, a card is picked up at random.

1) What is the probability of selected card is a King or a Queen?
2) What is the probability that the selected card is a King or a Diamond

Solution to part 1)



- Let A = getting a King
- Let B = getting a Queen

There are 4 kings and there are 4 Queens. The events are clearly mutually exclusive. Apply the formula.

P(A U B) = P(A) + P(B) = 4/52 + 4/52 = 8/52 = 2/13

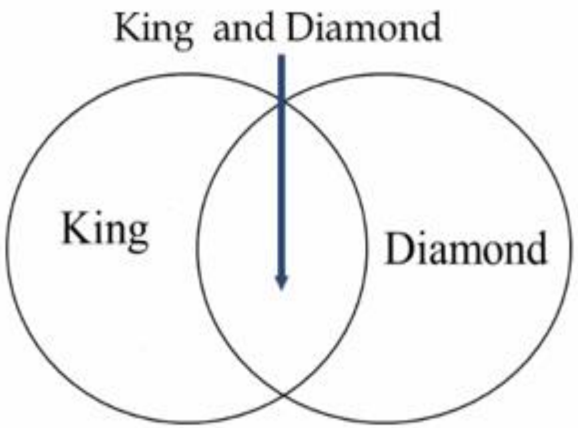
Solution to part 2

- Look at the diagram:

There are totally 52 cards in the pack out of which 4 are kings and 13 are diamonds. Let A = getting a King and B = getting a Diamond. The two events here are not mutually exclusive because you can have a card, which is both king and a Diamond called King Diamond.

$$P(K \cup D) = P(K) + P(D) - P(K \cap D)$$

$$= 4/52 + 13/52 - 1/52 = 16/52 = 4/13$$



Multiplication rule

- Independent events

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

This rule says when the two events A and B are independent the probability of the simultaneous occurrence of A and B (also known as probability of intersection A and B) equals the product of the probability of A and the probability of B.

Of course this rule can be extended to more than two events.

Multiplication rule

Independent Events-Example

- Example:

The probability that you will get an A grade in Quantitative Methods is 0.7. The probability that you will get an A grade in Marketing is 0.5. Assuming these two courses are independent, compute the probability that you will get an A grade in both these subjects.

- Solution:

Let A = getting A grade in Quantitative Methods

Let B =getting A grade in Marketing

It is given that A and B are independent.

$P(A \cap B) = P(A).P(B) = 0.7.0.5 = 0.35$

Multiplication Rule

- **Events are not independent**

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

- This rule says that the probability of the intersection of the events A and B equals the product of the probability of A and the probability of B given that A has happened or known to you. This is symbolized in the second term of the above expression as P (B/A). P(B/A) is called the conditional probability of B given the fact that A has happened.

- We can also write $P(A \cap B) = P(B).P(A/B)$ if B has already happened.

Multiplication rule

Events are not independent-Example

From a pack of cards, 2 cards are drawn in succession one after the other. After every draw, the selected card is not replaced. What is the probability that in both the draws you will get Spades?

Solution:

Let A = getting spade in the first draw

Let B = getting spade in the second draw.

The cards are not replaced.

This situation requires the use of conditional probability.

- P(A) = 13/52 (There are 13 Spades and 52 cards in a pack)
- P(B/A) = 12/51(There are 12 Spades and 51 cards because the first card selected
- $P(A \cap B) = P(B).P(A/B) = (13/52).(12/51)=156/2652= 1/17.$

Marginal Probability

- Contingency table consists of rows and columns of two attributes at different levels with frequencies or numbers in each of the cells. It is a matrix of frequencies assigned to rows and columns.

- The term marginal is used to indicate that the probabilities are calculated using contingency table (also called joint probability table).

Marginal Probability-Example

- A survey involving 200 families was conducted. Information regarding family income per year and whether the family buys a car are given in the following table.

Family	Income below Rs 10 Lakhs	Income of Rs. >=10 lakhs	Total
Buyer of Car	38	42	80
Non-Buyer	82	38	120
Total	120	80	200

- a) What is the probability that a randomly selected family is a buyer of the car?
- b) What is the probability that a randomly selected family is both a buyer of car and belonging to income of Rs.10 lakhs and above?
- c) A family selected at random is found to be belonging to income of Rs 10 lakhs and above. What is the probability that this family is buyer of car?

Solution

greatlearning
Learning for Life

- a) What is the probability that a randomly selected family is a buyer of the Car?

  80/200 = 0.40.

- b) What is the probability that a randomly selected family is both a buyer of car and belonging to income of Rs.10 lakhs and above?

  42/200 =0.21

- c) A family selected at random is found to be belonging to income of Rs 10 lakhs and above. What is the probability that this family is buyer of car?

  42/80 =0.525. Note this is a case of conditional probability of buyer given income is Rs.10 lakhs and above.

Some interesting problems for discussion

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD). 20% of the cars have both.

- What is the probability that a car has a CD player, given that it has AC?
  - –i.e., we want to find P(CD | AC)

Bayes' Theorem

greatlearning
Learning for Life

- Bayes' Theorem is used to revise previously calculated probabilities based on new information.
- Developed by Thomas Bayes in the 18th Century.
- It is an extension of conditional probability.

Bayes' Theorem

greatlearning
Learning for Life

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k)}$$

- Where:
  - Bi = i^{th} event of K mutually exclusive and collectively exhaustive events.
  - A = new event that might impact P(Bi)

Bayes' Theorem

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k)}$$

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|B')\,P(B')}$$

- Where:
  - B' = Complement of B
  - A = new event that might impact P(B)

Bayes Theorem Discussion Problem1

- ABCD Network has produced a new TV show. Of all the shows they produced, 40% hit shows and 60% flop shows.
- They are considering to employ a market research firm to conduct audience view pilot survey. The track record of the market research firm is – the shows that were actually hit, 90% of the times they had positive survey result. While the shows that were actually flop, they 70% of the times negative survey result.
- If the market research firm survey is positive, what is the probability that the show will be hit

Bayes Theorem Discussion Problem1

greatlearning
Learning for Life

	Hit	Flop
Prior Probability	40%	60%

	Hit	**Flop**
Survey Positive	P(Survey Pos | Hit) 90%	P(Survey Pos | Flop) 30%
Survey Negative	P(Survey Neg | Hit) 10%	P(Survey Neg | Flop) 70%

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|B')\,P(B')}$$

$$P(\text{Hit} \mid \text{Survey Positive}) = \frac{P(\text{Survey Positive}|\text{Hit}) * P(\text{Hit})}{P(\text{Survey Positive}|\text{Hit}) * P(\text{Hit}) + P(\text{Survey Positive} \mid \text{Flop}) * P(\text{Flop})}$$

$$= \frac{0.9 * 0.4}{(0.9 * 0.4) + (0.3 * 0.6)} = \frac{0.36}{0.36 + 0.18} = \frac{0.36}{0.54} = 0.67$$

- 67% probability that the show will be hit, given the survey is positive

Bayes Theorem Discussion Problem1

	Hit	Flop	
Shows Produced	40	60	100

	Hit	**Flop**
Survey Positive	P(Survey Pos | Hit) 0.90	P(Survey Pos | Flop) 0.30
Survey Negative	P(Survey Neg | Hit) 0.10	P(Survey Neg | Flop) 0.70

	Hit	**Flop**	
Survey Positive	= 40 * 0.9 **36**	= 60 * 0.3 **18**	**54**
Survey Negative	= 40 * 0.1 **4**	= 60 * .7 **42**	**46**
	40	**60**	**100**

P(Hit | Survey Positive) = 36 / 54 0.67