Basic Statistics and Probability

Chapter 3: Probability

- Events, Sample Spaces and Probability
- Unions and Intersections
- Complementary Events
- Additive Rule. Mutually Exclusive Events
- Conditional Probability
- Multiplicative Rule. Independent Events
- Counting Rules
- Bayes Rule

Probability is the science that studies randomness or chance. It plays an important role in Inferential Statistics, since it provides the link between sample and population.

The core idea in the start of probability theory was that certain kinds of phenomenon occur in an unpredictable, random manner, but some kind of regularity is observed after a large number of occurrences. The aim is to look for rules governing these random events.

The practical problems leading to an interest in the topic of probability were games of chance and errors in experimental science.

Events, Sample Spaces and Probability

A (random) experiment is an act or process of observation leading to a single outcome (or event) that cannot be predicted with certainty.

Example 3.1: Coin toss

We consider the experiment of flipping a coin and recording the result (head or tails).

Example 3.2: Die toss

We consider the experiment of tossing a die and recording the result (1, 2, 3, 4, 5 or 6).

Statistical experiments are also:

- measuring the percentage of bodyfat and the diameter of neck, chest, abdomen,... and recording the age of a man;
- choosing diamonds for sale in internet ads and recording their price, Carat weight, cut and other characteristics;
- measuring BMI, glucose, insulin, and other variables and recording age and presence/absence of breast tumor in a woman;
- counting the number of goals of Real Madrid team in a match;
- recording how long a student can speak with its colleague beside before the instructor tells them off.

A sample point or a simple event is each of the possible basic outcomes of an experiment (a basic outcome cannot be decomposed into more basic ones).

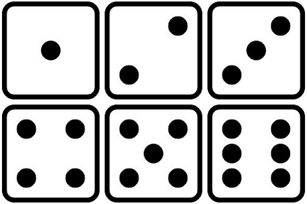
Example 3.1: Coin toss

All the possible simple events are two:

1. heads;
2. tails.

Example 3.2: Die toss

The possible simple events are six:



Example 3.3: Tossing two coins

If we toss two coins, all the possible outcomes are:

HH HT TH TT

where H in the first position means "Heads on coin 1".

The collection of all the possible simple events (outcomes) of an experiment is the sample space S.

Example 3.1: Coin toss

$$S : \{\text{Head}, \text{Tail}\}$$

Example 3.2: Die toss

$$S : \{1, 2, 3, 4, 5, 6\}$$

Example 3.3: Tossing two coins

$$S :$$

The probability P of a random event is a number between 0 and 1 that measures the likelihood that the outcome will occur when the experiment is performed. This number is usually taken as the relative frequency of the event when the random experiment is repeated a large number of times.

Example 3.1: Coin toss

If the coin is fair, then

$$P(\text{Head}) = 1/2 \quad P(\text{Tail}) = 1/2$$

Example 3.2: Die toss

For a fair die

$$P(1) = 1/6 \quad P(2) = 1/6 \quad P(3) = P(4) = P(5) = P(6) = 1/6$$

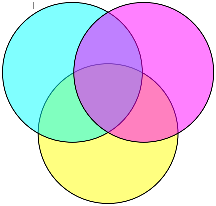
Connection with set theory

Sample space $\leftrightarrow$ Universe

Event $\leftrightarrow$ Subset

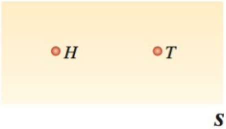
Sample point $\leftrightarrow$ Element

Like in set theory, we can use Venn diagrams to represent events.

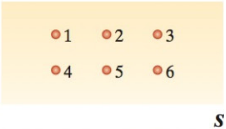


A Venn diagram is a graphical representation of a closed figure (like a circle) containing some elements.

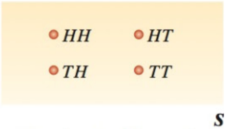
Venn diagrams for examples 3.1, 3.2 and 3.3:



a. Experiment: Observe the up face on a coin



b. Experiment: Observe the up face on a die



c. Experiment: Observe the up faces on two coins

Represent the following events with Venn diagrams as a subset of the sample space S:

Example 3.2: Die toss

- Obtain an even number of dots in the die.

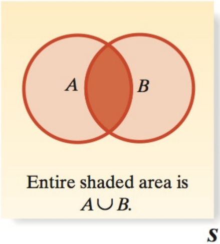
- The number of dots is less than or equal to 3.

Example 3.3: Tossing two coins

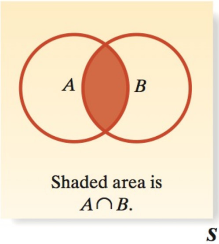
- Obtain at least one head.

The union of two events A and B, $A \cup B$, is the event that either A **or** B occurs in a single performance of the experiment. $A \cup B$ consists of all the sample points belonging to A, to B or to both.

The intersection of two events A and B, $A \cap B$, is the event that both A **and** B occur on a single performance of the experiment. $A \cap B$ consists of all the sample points belonging to both A and B.

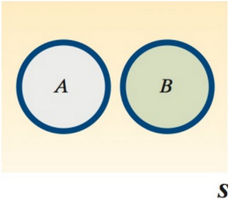


a. Union



b. Intersection

Two events A and B are disjoint or mutually exclusive when their intersection is empty: $A \cap B = \varnothing$.



Compound events are those which are not simple, but they can be decomposed into a union of simple events:

Example 3.2: Die toss

- Obtain an even number of dots in the die.

- The number of dots is less than or equal to 3.

Example 3.3: Tossing two coins

- Obtain at least one head.

Strictly speaking, probability is a function P associating, to each event A, a number $P(A)$ between 0 and 1 satisfying:

- $P(S) = 1$
- If $A_1, A_2, \ldots, A_n, \ldots$ are disjoint events, then $P(A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots) = P(A_1) + P(A_2) + \ldots + P(A_n) + \ldots$

To compute the probability of an event the idea is to decompose it into a union of simple events and sum up the probabilities of these.

Example 3.2: Die toss

Consider the event of obtaining an odd number of dots:

The probability of this event is:

Example 3.2: Die toss

Consider the following events

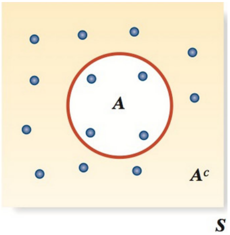
A : Toss an even number B : Toss a number less than or equal to 3

- Describe $A \cup B$

- Describe $A \cap B$

- Calculate $P(A \cup B)$ and $P(A \cap B)$, assuming the die is fair.

The complement, A^c, of an event A is the event that A does not occur. A^c comprises all the sample points in the sample space S but not in A.



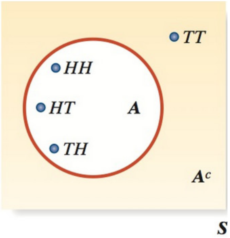
$$A \cup A^c = S \qquad\qquad A \cap A^c = \varnothing$$

$$P(S) = P(A \cup A^c) = P(A) + P(A^c) = 1 \Rightarrow \boxed{P(A^c) = 1 - P(A)}$$

$$\boxed{P(\varnothing) = 0}$$

Example 3.3: Tossing two coins

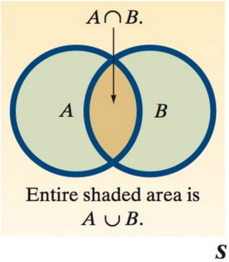
Define the event A : "Observing at least one head". Find $P(A)$.



Additive Rule. Mutually Exclusive Events

Additive Rule of Probability

Let A and B be two events, not necessarily disjoint.



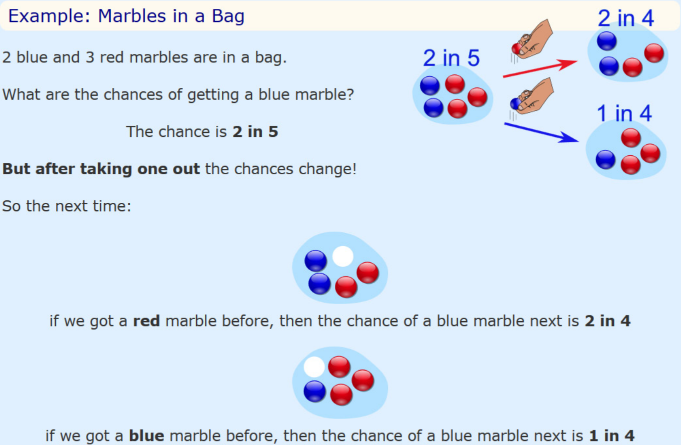
Then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

An event A is a subset of B, $A \subseteq B$, if every sample point in A is also in B. If $A \subseteq B$, then $P(A) \leq P(B)$.

Example 3.4: Hospital admissions

Hospital records show that 12% of all patients are admitted for surgical treatment, 16% are admitted for obstetrics and 2% receive both obstetrics and surgical treatment. If a new patient is admitted to the hospital, what is the probability that the patient will be admitted for surgery, for obstetrics, or for both?

Sometimes we have additional knowledge on the possible outcome of a random experiment. Then the probability of an event will be derived conditional on this additional knowledge.



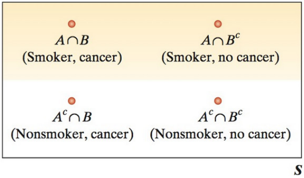
The probability of event A conditional on event B (or given that B has occurred) is

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

This formula adjusts the probability of $A \cap B$ from its original value in the complete sample space S to a conditional probability in the reduced sample space B.

Example 3.5: Smoking and Cancer

Consider an individual randomly selected from the adult male population in the US. Let A represent the event that the individual smokes, and let A^c denote the event that the individual does not smoke. Similarly, let B represent the event that the individual develops cancer.



Probabilities

	Develops Cancer	
Smoker	Yes, B	No, B^c
Yes, A	.05	.20
No, A^c	.03	.72

What is the probability that a smoker develops cancer? What are the chances that a non-smoker develops cancer?

Example 3.6: Product Complaints

A manufacturer of an electromechanical kitchen utensil conducted an analysis of a large number of consumer complaints and found that they fell into the six categories shown in the table. If a consumer complaint is received, what is the probability that the cause of the complaint was the appearance of the product, given that the complaint originated during the guarantee period?

	Reason for Complaint			
	Electrical	Mechanical	Appearance	Totals
During Guarantee Period	18%	13%	32%	63%
After Guarantee Period	12%	22%	3%	37%
Totals	30%	35%	35%	100%

Multiplicative Rule of Probability

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

Example 3.7: Light-colored Eyes

A population is formed by three ethnic groups: A (30%), B (10%) and C (60%). The relative frequencies of the character "light-colored eyes" are 20%, 40% and 5% respectively in each group.

$$P(\text{light}|A) = 0.2 \quad P(\text{light}|B) = 0.4 \quad P(\text{light}|C) = 0.05$$

What is the probability that a person from that population has light eyes and comes from the group A?

Events A and B are independent if the occurrence of B does not alter the probability that A occurs, that is, if

$$P(A|B) = P(A).$$

This is equivalent to

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

and also equivalent to

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

Example 3.2: Die toss

Define the events

$$A: \text{ Obtain an even number} \quad B: \text{ Obtain a number} \leq 4$$

Are A and B independent?

Example 3.6: Product Complaints

Define the events

$$\begin{array}{ll} A: & \text{Cause of complaint is product appearance} \\ B: & \text{Complaint occurred during guarantee term} \end{array}$$

Are A and B independent?

Events that are not independent are said to be dependent.

Sometimes experiments have so many sample points that it is impractical or unfeasible to list them all. But we can try to apply certain counting rules to count the number of simple events in the random experiment.

Combinations Rule

We want to draw a sample of n elements without replacement from a set of N elements. Then the number of possible different samples (not taking into account the ordering of the individuals in the sample) is

$$\binom{N}{n} = \frac{N!}{n!(N-n)!},$$

where $n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1$.

Example 3.8: Three balls

I have three balls with different colors • • •. How many different samples of two balls taken without replacement can I obtain?

• •   • •   • •

Observe that the sample • • is considered equal to the sample • •, because order is not taken into account.

Example 3.9: Movie Reviews

Suppose a movie reviewer for a blog reviews 5 movies each month. This month the reviewer has 20 movies from which to make the selection. How many different samples of 5 movies can be selected from the 20?

Permutations Rule

Given a set of N different elements, we want to select n elements from the total N and arrange them within n positions (here we take into account the ordering). The number of different permutations (or, better, variations) of the N elements taken n at a time is

$$P_n^N = N(N-1)(N-2)\cdots(N-n+1) = \frac{N!}{(N-n)!}$$

Example 3.8: Three balls

The possible permutations of the $N = 3$ balls taken $n = 2$ at a time are

Partitions Rule

We want to partition a set of N elements into k subsets, the first one containing n_1 elements, the second containing n_2 elements, ... and the kth one containing n_k elements, with $n_1 + n_2 + \ldots + n_k = N$. Then the number of different possible partitions is

$$\frac{N!}{n_1!n_2!\ldots n_k!}.$$

Example 3.10: Construction workers

Suppose you have 12 construction workers and you wish to assign 3 to jobsite 1, 4 to jobsite 2, and 5 to jobsite 3. In how many different ways can you make this assignment?

Total Probability Rule: Let $A_1, \ldots, A_m$ be events such that $\bigcup_{i=1}^m A_i = S$ and $A_i \cap A_j = \varnothing$ for all $i \neq j$. Then, for any event B,

$$P(B) = \sum_{i=1}^{m} P(B|A_i)P(A_i).$$

Bayes Rule: Let $A_1, \ldots, A_m$ be events such that $\bigcup_{i=1}^m A_i = S$ and $A_i \cap A_j = \varnothing$ for all $i \neq j$. Then, for any event B,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^m P(B|A_i)P(A_i)} = \frac{P(B|A_j)P(A_j)}{P(B)}.$$

Example 3.7: Light-colored Eyes

(a) Calculate the probability that a person chosen at random has light-colored eyes.

(b) Calculate the probability that an individual with dark eyes is from group A.

(c) If an individual chosen at random has light eyes, from which ethnic group is it more likely that he comes?

Example 3.11: Medical Test

Suppose there is a certain disease randomly found in 2 out of every 1,000 persons of the general population (prevalence in the population). A certain clinical blood test is 99% effective in detecting the presence of this disease (sensitivity of the test), that is, it will yield an accurate positive result in 99 out of every 100 individuals with the disease. But it also yields false-positive results in 5% of the cases where the disease is not present (specificity of the test is then 95%).

Compute the probability that a person who has a positive result in the test actually suffers the disease.