# Probability

Dr Wan Nor Arifin
Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.
wnarifin@usm.my

Last update: 19 September, 2018



## Contents

## Introduction

### Probability

- It is the chance that something will happen (Merriam-Webster, 2012).
- It is a branch of mathematics that studies the possible outcomes of given events together with the outcomes' relative likelihoods and distributions (Weisstein, 2018).
- Synonyms: Possibility, likelihood, chance.
- Expressed on a linear scale from 0 (Impossible) to 1 (Certain), or as a percentage (0% to 100%).
- Understanding of probability is important to understand statistical inference.
- Widely applicable to many different field of sciences, including medicine.
- Can be generally classified into:
    - Classical
    - Frequentist
    - Bayesian

### Classical

- 1654 Pierre de Fermat and Blaise Pascal – basic principles.
- 1812 Pierre-Simon Laplace – general theory.
- Game of chance e.g. flipping coin, rolling of dice, Russian roulette.
- Finite number of possible outcomes.
- If an event can occur in $N$ mutually exclusive and equally likely ways and if $N_A$ of these possess a characteristic $A$, the probability of occurrence of $A$ is defined by:

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

Example 1:

If a fair 6-sided die is rolled, probability of getting a 1 is:

$$P(1)=\frac{1}{6}$$

### Frequentist

- 1866 John Venn – first elaborate and systematic explanation.
- Views probability simply as a measure of the frequency of outcomes.
- A hypothesis is typically tested without being assigned a probability (posteriori).
- Probability of an event is estimated based on relative frequency after a number of repetition of random trials. As the number of trials is increased, it is expected the relative frequency to become a better estimate of the "true frequency" or probability.
- If $n_t$ is the total number of trials and $n_x$ is the number of trials where the event $x$ occurred, the probability $P(x)$ of the event occurring will be approximated by the relative frequency as:

$$P(x)\approx\frac{n_x}{n_t}$$

Example 2:

Let say on average, it rained 15 out of 30 days of September based on 200 years weather statistics data, with no pattern for which particular days it rained. So, the probability of rain on September 23 2012 is:

$$P(\text{rain on September 23}) = \frac{15}{30} = \frac{1}{2}$$

**Bayesian**

- 1763 Thomas Bayes – Bayes' Theorem
- Specifies some prior probability, which is then updated in the light of new, relevant data.
- Treats probability more subjectively as a statistical procedure that endeavors to estimate parameters of an underlying distribution based on the observed distribution.
- A probability is assigned to a hypothesis (priori).
- Bayes' theorem will be introduced later in this note.

Example 3:

General statement
The probability that a fair 20 cent coin would turn up Hibiscus is 50%.

Frequentist statement
If the experiment were repeated over and over again, in the long run the average frequency of Hibiscus would be close to 50%.

Bayesian statement
In my opinion, the probability that a fair 20 cent coin would turn up Hibiscus is 50%.

## Basic concepts

**Experiments**

- Situations for which the outcomes occur randomly.

**Sample Space**

- Set/list of all possible outcomes of an experiment.
- Denoted by Ω (OMEGA)
- Element of Ω denoted by ω (omega)
- $\Omega = \{\omega_1, \omega_1, \ldots, \omega_N)$

Example 4:

A fair 6-sided die is rolled and the number that come up is recorded. The sample space is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Example 5:

In a lecture room, there are 3 males and 2 females. If a lecturer want to choose a student to present an assignment on probability, the sampling space from which the lecturer can sample is:

$$\Omega = \{M_1, M_2, M_3, F_1, F_2\}$$

Example 6:

On the way to USM, a student passes through three intersections with traffic lights. At each intersection, he either stops (s) or continues (c). The sample space is:

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

where, for example ccs means he continues through the 1st intersection and 2nd intersection and stops at the 3rd intersection.

**Events**

- Subsets of sample space Ω.
- Looking at only a portion of sample space with some particular conditions/outcomes.

Example 7:

Using example 6, let say an event A is the event that the student stops at 2nd intersection:

$$A = \{css, csc, sss, ssc\}$$

Using example 5, let say event B is the event that the presenter is male:

$$B = \{M_1, M_2, M_3\}$$

**Union**

- Union of two events, let say A and B, is the event C that either A or B or both occur.
- Thus C is a set that consists all of the elements that belong to A and B.

$$C = A \cup B$$

Example 8:

If event A is the event that the student stops at 2nd intersection:

$$A = \{css, csc, sss, ssc\}$$

and event B is the event that the student stops at 3rd intersection:

$$B = \{sss, scs, ccs, css\}$$

then

$$C = A \cup B = \{css, csc, sss, ssc, \cancel{sss}, scs, ccs, \cancel{css}\} = \{css, csc, sss, ssc, scs, ccs\}$$

(remove repetitive common elements)

**Intersection**

- Intersection of two events, A and B is the the event C that both A and B occur.

$$C = A \cap B$$

Example 9:

Using previous example, thus

$$C = A \cap B = \{\cancel{css, csc, sss, ssc,}\ sss,\ \cancel{scs, ccs,}\ css\} = \{sss, css\}$$

(notice the elements that we strike through on *union* make up the intersection elements)

**Complement**

- Complement of an event, $A^c$, is the event that A does not occur, i.e. all elements in sample space that are NOT A.

Example 10:

Using example 6, let say an event A is the event that the student stops at 2nd intersection:

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

$$A = \{css, csc, sss, ssc\}$$

$$A^c = \{ccc, ccs, scc, scs\}$$

**Disjoint**

- Two events are disjoint when they have no elements in common/shared.

Example 11:

Continuing on previous example:

$$A = \{css, csc, sss, ssc\}$$

$$D = \{ccc\}$$

thus

$$A \cap D = \emptyset \quad \text{(empty set)}$$

Similarly,

$$A \cap A^c = \emptyset$$

## Laws of Set Theory

### Commutative Laws

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

### Associative Laws

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

### Distributive Laws

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

## Probability Measures

*Mutually exclusive* events are events that cannot occur at same time. If A occur, B cannot occur.

### Axioms

In mathematics, an axiom is “a proposition regarded as self-evidently true without proof” (Weisstein, 2018) that serves as a “starting point for further reasoning and arguments” (Wikipedia contributors, 2018).

For disjoint or mutually exclusive events, they satisfy these three axioms:

1. The sum of the probabilities of all mutually exclusive events is equal to 1.

$$P(\Omega) = 1$$

2. Probability of any event is non-negative.

$$\text{If } A \subset \Omega, \textit{then } P(A) \geqslant 0$$

3. If event $A_1$ and $A_2$ are disjoint, then probability of either $A_1$ or $A_2$ equals to the sum of their individual probabilities.

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

If A1, A2, …, An, … are mutually disjoint, then we have a more general form,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

**Properties**

These are derived properties based on the three axioms:

1. $P(A^c) = 1 - P(A)$
2. $P(\varnothing) = 0$
3. If $A \subset B$, then $P(A) \leqslant P(B)$
4. Addition Law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example 12:

Suppose a 20 cent coin is thrown twice. Let A denote the event of Hibiscus on 1st throw and B the event of Hibiscus on 2nd throw. What is the probability of $C = A \cup B$ ?

Since,

Ω = {hh, hc, ch, cc}, each outcome with P = 0.25

thus,

A = {hh, hc}, P(A) = 0.5
B = {hh, ch}, P(B) = 0.5
A∩B = {hh}, P(A∩B) = 0.25

Using addition law,

$$P(C) = P(A) + P(B) - A \cap B = 0.5 + 0.5 - 0.25 = 0.75$$

## Computing Probabilities: Counting Methods

**Multiplication Principle**

Basic multiplication principle
If one experiment has *m* outcomes and another experiment has *n* outcomes, then there are *mn* possible outcomes for the two experiments.

Example 13:

Playing cards have 13 face values and 4 suits. Thus there are 13 x 4 = 52 face values with suit combinations.

Example 14:

A lecturer selected 1 male student and 1 female student as class representatives in a class of 10 males and 10 females. Thus, there are 10 x 10 = 100 ways of class representative combinations.

Extended Multiplication principle
If there are *p* experiments and the 1st has $n_1$ possible outcomes, the 2nd $n_2$, …, and the *p*th $n_p$ possible outcomes, then there are a total of $n_1 \times n_2 \times \ldots \times n_p$ possible outcomes for the *p* experiments.

Example 15:

If a fair coin is thrown 4 times, how many different outcomes (hibiscus, congkak) are possible?

$$2 \times 2 \times 2 \times 2 = 2^4 = 16$$

**Permutation**

- It is an ordered arrangement of objects.
- Sampling *without* replacement – no duplication is allowed.
- Sampling *with* replacement – duplication is allowed.
- For a set of size *n* and a sample of size *r*, the number of different ordered samples (*without replacement*) is

$$n(n-1)(n-2)\ldots(n-r+1)=\frac{n!}{(n-r)!}={}^nP_r$$

- The number of orderings of *n* elements (*without replacement*) is

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

- For a set of size *n* and a sample of size *r*, there are

$$n^r$$

different ordered samples (*with replacement*).

Example 16:

How many different ways to arrange number 0 – 9 to form a 3 digits number sequence?

Sample size, $r = 3$
Number of elements, $n = 10$

$$n^r = 10^3 = 1000 \text{ ways}$$

(Sampling *with replacement* since same number can appear again and again)

Example 17:

If same number cannot appear twice, how many possible ways to arrange the sequence of number?

(Sampling *without replacement*)

$$n(n-1)(n-2)...(n-r+1) = 10(10 – 1)(10-3+1) = 10.9.8 = 720 \text{ ways}$$

Example 18:

How many ways 5 cups of different colors can be lined up?

$$5! = 5.4.3.2.1 = 120 \text{ ways.}$$

**Combination**

- It is the constituents of the samples regardless of the order (*unordered*) in which they were obtained.
- The number of *unordered* samples of *r* objects from *n* objects *without replacement* is

$$\frac{n(n-1)(n-2)\ldots(n-r+1)}{r!} = \frac{n!}{(n-r)!} \times \frac{1}{(r)!} = \frac{n!}{(n-r)!\,r!} = \binom{n}{r} = {}^{n}C_{r}$$

  since the number of ordered samples is $n(n-1)(n-2)\ldots(n-r+1)$ and a sample of size *r* can be ordered in *r*! ways.
- $\binom{n}{r}$ is known as binomial coefficients.
- The number of *unordered* samples of *r* objects from *n* objects *with* replacement is

$$\frac{(r+n-1)!}{r!(n-1)!}$$

Example 19:

Following an ANOVA comparing 5 groups, a student was asked to perform a number of independent *t*-tests. How many independent *t*-tests should he perform? (Remember *unordered* pairs)

*n* = 5, *r* = 2 (pair)

$$\binom{n}{r} = \frac{n!}{(n-r)!\,r!} = \frac{5!}{(5-2)!\,2!} = \frac{5.4.3!}{3!\,2!} = \frac{5.4}{2.1} = 10$$

Example 20:

How many combinations of 3 colors out of a sample of 7 colors are possible (with replacement)?

$$\frac{(r+n-1)!}{r!(n-1)!} = \frac{(3+7-1)!}{3!(7-1)!} = \frac{9!}{3!6!} = \frac{9.8.7}{3.2.1} = 84$$

- The number of ways that *n* objects can be grouped into *r* classes with $n_i$ in the *i*th class

$$i = 1, \ldots, r, \quad \text{and} \quad \sum_{i=1}^{r} n_i = n \quad \text{is}$$

$$\binom{n}{n_1 n_2 \ldots n_r} = \frac{n!}{n_1!\,n_2!\ldots n_r!}$$

- $\binom{n}{n_1 n_2 \ldots n_r}$ is known as multinomial coefficients.

Example 21:

A group of 7 eligible subjects in a clinical trial is allocated into 3 groups with size of 3, 2 and 2. How many ways the allocation could be done?

$$\binom{7}{3\,2\,2}=\frac{7!}{3!2!2!}=\frac{7.6.5.4.3!}{3!\,2!\,2!}=\frac{7.6.5.4}{2.1.2.1}=210$$

## Conditional Probability

**Definition**

- When probability is calculated with a subset of the sample space as denominator.
- When the occurrence an event A is dependent on the occurrence of event B.
- Probability of A restricted to the part of sample space containing event B only.
- Thus we say probability of A given B, P(A|B) → Conditional probability.

$$P(A|B)=\frac{P(A\cap B)}{P(B)},\ \text{ if }\ P(B)\neq 0$$

Example 22:

A researcher developed a new test kit to detect the presence of HIV infection using blood test. It was compared with the gold standard in detecting HIV infection ELISA test. The data was as follows:

| | | Detection by ELISA | | |
|---|---|---|---|---|
| | | D+ | D- | Total |
| Detection by new test | T+ | 30 | 15 | 45 |
| | T- | 5 | 50 | 55 |
| | Total | 35 | 65 | 100 |

T+ = Positive of HIV based on new test.
T– = Negative of HIV based on new test.
D+ = Positive of HIV based on ELISA.
D– = Negative of HIV based on ELISA.

What is the probability of HIV positive D+ when the test is positive T+ (focus on T+ row)?

P(D+ | T+) = 30/45 = .6667

What is the probability of test negative T– when HIV is negative D– (focus on D– column)?

P(T– | D–) = 50/65 = .7692

As you can see, conditional probability is easily calculated if we have the frequency of each row and column.

But if similar table is given in form of probability:

| | | Detection by ELISA | | |
|---|---|---|---|---|
| | | D+ | D- | Total |
| Detection by new test | T+ | .3000 | .1500 | .4500 |
| | T- | .0500 | .5000 | .5500 |
| | Total | .3500 | .6500 | 1.000 |

What is the probability of HIV positive D+ when the test is positive T+ (focus on T+ row)?

$$P(\text{D+} \mid \text{T+}) = \frac{P(\text{D+} \cap \text{T+})}{P(\text{T+})} = 0.3000/0.4500 = .6667$$

What is the probability of test negative T– when HIV is negative D– (focus on D– column)?

$$P(\text{T–} \mid \text{D–}) = \frac{P(\text{T-} \cap \text{D-})}{P(\text{D-})} = .5000/.65000 = .7692$$

**Multiplication Law**

- Manipulating the same formula for conditional probability, we have multiplication law to calculate intersection (or joint) probability

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

Example 23:

Suppose the probability of raining, R is 30% when it is cloudy, C. The probability that it is cloudy is 20%. Thus the probability that it is raining and cloudy

$$P(R \cap C) = P(R|C)P(C) = 0.3 \times 0.2 = 0.06$$

is 6%.

**Law of Total Probability**

- Also know as *partition theorem*.
- For events $B_{1,} B_{2,} \ldots, B_n$ that are disjoint events, whose union is $\bigcup_{i=1}^{n} B_i = \Omega$, and $B_i \cap Bj = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all *i*, then for any event A,

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

- Useful when direct information on P(A) is not available.

Example 24:

| | | Detection by ELISA | | |
|---|---|---|---|---|
| | | D+ | D- | Marginal *p* |
| Detection by new test | T+ | .6667 | .3333 | .4500 |
| | T- | .0909 | .9091 | .5500 |
| | * | * | * | * |

Let say from the same table, only conditional probabilities and marginal probabilities are known.

What is the probability of D+?

P(D+) = P(D+ | T+)P(T+) + P(D+ | T– )P(T–)
= .6667 x .4500 + .0909 x .5500
= 0.3500

What is the probability of D–?

P(D–) = P(D– | T+)P(T+) + P(D– | T–)P(T–)
= .3333 x .45 + .9091 x .5500
= 0.6500

*Compare the results with the full table data in Example 22.

**Bayes' Rule**

- For events $B_{1,} B_{2,} \ldots, B_n$ that are disjoint events, whose union is $\bigcup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all *i*, then

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

when linked to the Law of Total Probability and Multiplication Law, becomes

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

in an easier to understand format, is actually,

$$Posterior\ probability = \frac{Prior\ probability \times Likelihood}{Marginal\ probability}$$

Example 25:

| | | Detection by ELISA | | |
|---|---|---|---|---|
| | | D+ | D- | * |
| Detection by test T | T+ | .8571 | .2308 | * |
| | T- | .1429 | .7692 | * |
| | Marginal *p* | .3500 | .6500 | * |

From the table it is known the test T is positive T+ among 85.71% of HIV positive subjects, while it is falsely positive among 23.08% of HIV negative subjects. 35% of the sample is HIV positive while 65% is negative.

However, let say we do the new test of a drug addict, which we know based on our previous data or experience that most likely he is HIV positive with p = 0.85, and that he is unlikely to be HIV negative is p = 0.15.

Information known:

From **test T findings** (likelihood) , $P(A \mid B_i)$, for i = D+, D–:

P(T+ | D+) = .8571
P(T+ | D–) = .2308
P(T– | D+) = .1429
P(T– | D–) = .7692

From **previous knowledge** (prior probability), $P(B_i)$, for i = D+, D–:

P(D+) = .85
P(D–) = .15

What is the probability that the drug addict is HIV positive given that the test is positive?

$$P(\text{D+} \mid \text{T+}) = \frac{P(\text{T+}|\text{D+})\boldsymbol{P}(\textbf{D+})}{P(\text{T+}|\text{D+})\boldsymbol{P}(\textbf{D+})+P(\text{T+}|\text{D-})\boldsymbol{P}(\textbf{D-})}$$

$$= \frac{.8571\times\mathbf{.85}}{.8571\times\mathbf{.85}+.2308\times\mathbf{.15}}$$

= .9546 → Higher probability when expert opinion is considered.

What is the probability that the drug addict is HIV positive given that the test is negative?

$$P(\text{D+} \mid \text{T–}) = \frac{P(\text{T-}|\text{D+})\boldsymbol{P}(\textbf{D+})}{P(\text{T-}|\text{D+})\boldsymbol{P}(\textbf{D+})+P(\text{T-}|\text{D-})\boldsymbol{P}(\textbf{D-})}$$

$$= \frac{.1429\times\mathbf{.85}}{.1429\times\mathbf{.85}+.7692\times\mathbf{.15}}$$

= .5128 → Still high probability, in light of expert opinion.

## Independence

**Background**

- If probability of B is unaffected by prior occurrence of A and vice versa, we can say that the events are independent.
- Probability of B cannot be predicted based on probability of A.
- Formally,

$$P(A|B)=P(A) \text{ and } P(B|A)=P(B)$$

thus, modifying multiplication law

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

- A and B are said to be independent events if

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

Example 26:

| | | Detection by ELISA | | |
|---|---|---|---|---|
| | | D+ | D- | Total |
| Detection by new test | T+ | .3000 | .1500 | .4500 |
| | T- | .0500 | .5000 | .5500 |
| | Total | .3500 | .6500 | 1.000 |

The condition probability,

$$P(D+ \mid T+) = .3000/.45000 = .6667$$

If D+ and T+ are independent,

$$P(D+ \mid T+) = P(D+)$$

but P(D+) = .3500 ≠ .6667, thus they are not independent.

In other way,

$$P(D+ \cap T+) = .3000$$

not equal to,

$$P(D+)P(T+) = .3500 \times .4500 = .1575$$

thus they are not independent events.

Example 24:

Suppose a 20 cent coin is thrown twice. Let A denote the event of Hibiscus on 1st throw and B the event of Hibiscus on 2nd throw. Prove that A and B are independent events.

$$\Omega = \{hh, hc, ch, cc\}, \text{ each outcome with } P = 0.25$$

$$A = \{hh, hc\},\ P(A) = 0.5$$
$$B = \{hh, ch\},\ P(B) = 0.5$$
$$A \cap B = \{hh\},\ P(A \cap B) = 0.25$$

$$P(A \cap B) = P(A)P(B) = 0.5 \times 0.5 = 0.25 \rightarrow \text{Independent.}$$

## References


Daniel, W. W. (1995). *Biostatistics: A foundation for analysis in the health sciences* (6th ed.). USA: John Wiley & Sons.

Jones, P. W., & Smith, P. (2001). *Stochastic processes: An introduction*. New York, USA: Arnold.

Merriam-Webster (2012). *Merriam-Webster Online.* Retrieved from http://www.merriam-webster.com/

Pierce, R. (2014). Combinations and Permutations. In *Math Is Fun*. Retrieved from https://www.mathsisfun.com/combinatorics/combinations-permutations.html

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). USA: Duxbury Press.

Tijms, H. (2007). *Understanding probability: Chances rules in everyday life* (2nd ed.). New York, USA: Cambridge University Press.

Weisstein, E. W. (2012). *Wolfram MathWorld*. Retrieved from http://mathworld.wolfram.com/

Wikipedia contributors (2018). Axiom. In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Axiom&oldid=855955736