

5. Probability Calculus

So far we have concentrated on descriptive statistics (deskriptiivinen eli kuvaileva tilastotiede), that is methods for organizing and summarizing data. As was already indicated in our discussion of sample versus population variance, another important aspect of statistics is to draw conclusions about a population based upon only a subset of the population, called a sample (otos). This branch of statistics is called inferential statistics (tilastollinen päättely eli inferenssi). Knowing some probability calculus (todennäköisyyslaskenta) will allow us to evaluate and control the likelihood that our statistical inference is correct and provide the mathematical basis for inferential statistics.

5.1. Combinatorics

When considering probabilities one is often concerned with calculating the number of events that might possibly occur in a certain situation. The branch of mathematics concerned with this is called combinatorics (kombinatoriikka).

The Addition Principle
(*Yhteenlaskuperiaate*)

Consider an experiment with n groups of *exclusionary* outcomes (in the sense that if event i occurs, then a different event $j \neq i$ cannot occur), such that

group 1 contains k_1 outcomes,
group 2 contains k_2 outcomes,
⋮
group n contains k_n outcomes.

Then:

The experiment has all in all $k_1 + k_2 + \dots + k_n$ possible outcomes.

We may restate the addition principle equivalently as: If an event P can happen in p different ways and a distinct event Q can happen in q different ways, then either P or Q can happen in $p + q$ different ways.

Example.

The number of roads leaving from Vaasa either north or south are the number of roads leading north plus the number of roads leading south.

Example.

There are 3 trains and 4 busses leaving from A to B. Then there are all in all $3+4=7$ possible ways of travelling from A to B.

Example.(football pool betting)

How many possibilities are there of betting the results of exactly 12 out of 13 football matches right?

We get exclusionary outcomes by considering the number of possibilities of betting the remaining match wrong. There are two ways of betting each of the 13 matches wrong. Therefore there are

$$\underbrace{2 + 2 + \dots + 2}_{13 \text{ times}} = 26 \text{ ways}$$

of betting exactly 12 out of 13 matches right (= betting 1 match wrong).

The Multiplication Principle
(Kertolasku-/ tuloperiaate)

Consider an experiment, that realizes in n separate steps. Assume that there are

k_1 possible outcomes of step 1,
 k_2 possible outcomes of step 2,
:
 k_n possible outcomes of step n ,

which are *independent* of the outcomes in the other steps. Then the total number of possible outcomes is

$$k_1 \cdot k_2 \cdot \dots \cdot k_n.$$

Equivalently we may state the multiplication principle as: If an event P can happen in p different ways and a distinct event can happen in q different ways, then P and Q can simultaneously happen in pq different ways.

Example.

If there are 3 ways from A to B and 4 ways from B to C then there are all in all $3 \cdot 4 = 12$ ways from A to C.

Example.

How many 3-digit numbers can be formed from the numbers 1 to 5 if each number may appear

	-not more than once?	-more than once?
digit 1:	5 possibilities	5 possibilities
digit 2:	4 possibilities	5 possibilities
digit 3:	3 possibilities	5 possibilities
numbers:	$5 \cdot 4 \cdot 3 = 60,$	$5 \cdot 5 \cdot 5 = 125.$

Note:

In the special case $k_1 = k_2 = \dots = k_n =: k$ the total number of possible outcomes is k^n .

Example.

In a football pool betting with 13 games there are $3^{13} = 1\,594\,323$ possible bets.

Counting Permutations

A permutation (permutaatio) of a set is any ordered selection of its elements.

Example:

The permutations of the set $A = \{a, b, c\}$ are: (a, b, c) , (a, c, b) , (b, a, c) , (b, c, a) , (c, a, b) , (c, b, a) .

The important point about permutations is the order of the elements. Two permutations are identical only if they contain the same elements and these appear in the same order.

It follows from the multiplication principle that the number of permutations of a set with n elements is given by the product

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

This is called n factorial (luvun n kertoma). Zero factorial ($0!$) is defined to be 1.

Example. There are $6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$ permutations of the letters A,B,C,D,E,F.

The multiplication principle may also be applied, when we use only $0 \leq k \leq n$ elements in each arrangement, which is called variation (variaatio) or permutation (k -permutaatio), of a set of n items choose k , denoted by " ${}_n P_k$ ", and given by:

$${}_n P_k = n \cdot (n - 1) \cdots (n - (k - 1)) = \frac{n!}{(n - k)!}.$$

Example.

How many 2-letter words can be formed from the word LAHTI?

$${}_5 P_2 = \frac{5!}{(5 - 2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 5 \cdot 4 = 20.$$

Counting Combinations

Combinations (kombinaatio) of a set are selections of its elements *without taking their order into account*.

Example: (a, b, c) and (b, a, c) are identical combinations of set $A = \{a, b, c\}$.

Note that taking the variation ${}_n P_k$ may be thought of selecting k out of n elements and then arranging them into order, which may be done in $k!$ possible ways. Therefore, *disregarding their order*, there are ${}_n C_k := {}_n P_k / k!$ combinations (kombinaatio) of k out of n elements. The expression

$${}_n C_k = \binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1},$$

read “ n choose k ” (n yli k), is called binomial coefficient (binomikerroin).

Example.

The number of possibilities to draw 7 out of 39 numbers (disregarding their order) in a lottery game is

$$\binom{39}{7} = \frac{39!}{7!32!} = \frac{39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 15\,380\,937.$$

Note.

There are just as many ways of dividing a set of n ordered elements into the categories of k selected elements and $n - k$ not selected elements as there are ways to put the elements back from their category “selected” or “not selected” into their original order.

${}_n C_k$ denotes therefore also the number of possibilities to arrange n elements, k of which are of category 1 (“selected”) and $n - k$ of which are of category 2 (“not selected”) back into order, if we disregard the order of elements of the same category. In other words, ${}_n C_k$ denotes also the number of permutations of n objects, k of which are of kind 1 but otherwise indistinguishable, and $n - k$ are of kind 2 but otherwise indistinguishable.

Permutations of Indistinguishable Objects

Suppose we have r sets with n_i ($i = 1, \dots, r$) *indistinguishable* elements, such that n_1 elements are of type 1, n_2 elements are of type 2, and so on. Suppose further that we wish to select

$$\begin{aligned} &k_1 \text{ elements of type 1,} \\ &k_2 \text{ elements of type 2,} \\ &\quad \vdots \\ &k_r \text{ elements of type } r \end{aligned}$$

with $0 \leq k_i \leq n_i$. Then, applying the multiplication principle, this can be done in

$$\binom{n_1}{k_1} \binom{n_2}{k_2} \cdots \binom{n_r}{k_r} \text{ ways.}$$

Example. An association has 10 male and 15 female members of which 2 men and 3 women are to be elected for board membership. This can be done in

$$\binom{10}{2} \binom{15}{3} = \frac{10!}{2!8!} \cdot \frac{15!}{3!12!} = 20\,475 \text{ ways.}$$

Assume next that the elements are distinguishable and we want to select n_1 elements into the first bin, n_2 elements into the second bin, and so on, without taking the order of the elements in any specific bin into account. Then:

$$\begin{aligned}
 & \binom{n}{n_1} \binom{n-n_1}{n_2} \cdots \binom{n-n_1-\dots-n_{r-1}}{n_r} \\
 &= \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \\
 & \quad \cdots \frac{(n-n_1-\dots-n_{r-1})!}{n_r!(n-n_1-\dots-n_{r-1}-n_r)!} \\
 &= \frac{n!}{n_1!n_2!\cdots n_r!} =: \binom{n}{n_1, n_2, \dots, n_r}.
 \end{aligned}$$

This is the so called multinomial coefficient (multinomikerroin). It may be alternatively interpreted as the number of distinct sequences of n elements that can be made from n_i elements of type i , where $n_1 + \dots + n_r = n$.

Example. How many words can be build from the letters of the word MISSISSIPPI?

$$\text{Answer: } \frac{11!}{1!4!4!2!} = 34\,650 \text{ words.}$$

5.2. The Definition of Probability

In probability calculus, everything that is somehow influenced by randomness, is called a random experiment (sattunaiskoe) or random event (sattunaisilmiö).

Notation:

Ω or S	<u>sample space/universal set=certain event</u> (perusjoukko = varma tapahtuma),
\emptyset	<u>empty set = impossible event</u> (tyhjä joukko = mahdoton tapahtuma),
A, B, C, \dots	(sub-) <u>sets</u> (of Ω) = <u>events</u> (Ω :n osajoukkoja = tapahtumat),
a, b, c, \dots	<u>elements = elementary events</u> (alkiota eli alkeistapaukset = yksittäiset tulostmahdollisuudet),
$A \cup B$	<u>union</u> of A and $B = A, B$, or both happen (yhdiste = A tai B tai molemmat),
$A \cap B$	<u>intersection</u> of A and B = A and B happen simultaneously (leikkaus = sekä A että B tapahtuvat),
A^C	<u>complement</u> = A does not happen (A :n komplementti = A ei tapahdu),
$A \cap B = \emptyset$	A and B are <u>disjoint</u> = A and B do not happen simultaneously (A ja B ovat erillisiä = A ja B eivät voi tapahtua yhtä aikaa).

Classical Probability (Klassinen todennäköisyys)

One considers some random experiment or random event with a limited number of outcomes (for example throwing a coin, taking balls out of an urn). One also assumes that the possible outcomes are symmetrical (tulomahdollisuudet ovat symmetrisiä) in the sense that elementary events are equally likely. If event A is made up of k of such equally likely elementary events, then the probability $P(A)$ is defined as:

$$P(A) := \frac{k}{n} = \frac{\# \text{ elementary events leading to } A}{\# \text{ all elementary events}}$$

Example. Casting a dice:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2\}, \quad B = \{1, 3, 5\}.$$

$$\Rightarrow \quad P(A) = \frac{1}{6}, \quad P(B) = \frac{3}{6} = \frac{1}{2}.$$

Note:

If the elementary outcomes of the experiment cannot be considered equally likely, then the symmetry principle cannot be applied.

The Relative Frequency Approach (*Empiirinen eli tilastollinen todennäköisyys*)

One assumes that the random experiment or random event under investigation can be repeated arbitrary many times with uncertain outcome. If the experiment is repeated n times, of which f_A times the event A occurs, then A gets assigned the relative frequency (suhteellinen frekvenssi):

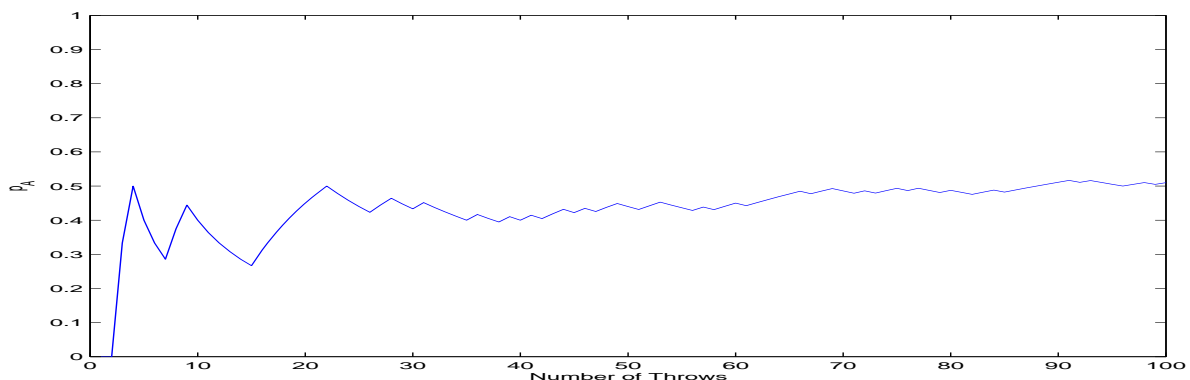
$$p_A := f_A/n.$$

Repeating the experiment many times yields then an approximation for the probability of event A , that is:

$$p_A \xrightarrow{n \rightarrow \infty} P(A).$$

Example. Throwing a coin 100 times yields 51 times head up, such that

$$P(\text{"head up"}) \approx 51/100 = 0.51.$$



Properties of Probabilities

Formally, probability is defined as a set function obeying the following three axioms due to Kolmogorov (1933):

1. $0 \leq P(A) \leq 1$ for all $A \subseteq \Omega$
2. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
(addition rule for disjoint events/
erillisten tapahtumien yhteenlaskusääntö)
3. $P(\Omega) = 1$

The following additional properties are implied by the axioms above:

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(general addition rule/ yhteenlaskusääntö)
5. $P(A^C) = 1 - P(A)$
6. $P(\emptyset) = 0$
7. $A \subseteq B \Rightarrow P(A) \leq P(B)$.

Example:

Drawing one card from a deck of 52 cards.
What is the probability of drawing an ace or a king?

$$A = \{\text{ace}\}, \quad B = \{\text{king}\} \quad \Rightarrow \quad A \cap B = \emptyset.$$

Therefore:

$$P(A \cup B) = P(A) + P(B) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}.$$

Example:(continued)

What is the probability of drawing an ace or a spade?

$$C = \{\text{spade}\} \quad \Rightarrow \quad A \cap C \neq \emptyset.$$

Therefore:

$$\begin{aligned} P(A \cup C) &= P(A) + P(C) - P(A \cap C) \\ &= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}. \end{aligned}$$

Example: (Coin-Tossing) What is the probability of getting tail up at least once when tossing a fair coin three times?

Hint: Expressions like “at least” or “at most” usually indicate that it pays to take a look at the complement of the event under investigation.

Let $A = \{\text{at least one tail up}\}$.

$\Rightarrow A^C = \{\text{no tail up}\} = \{\text{all head up}\}$.

There are $2^3 = 8$ elementary outcomes.

$$\Rightarrow P(A^C) = \frac{1}{8} \quad \Rightarrow \quad P(A) = 1 - \frac{1}{8} = \frac{7}{8}.$$

Example: (football pool betting continued.)

The probability of betting at least 12 out of 13 matches right is:

$$\begin{aligned} & P(12 \text{ or } 13 \text{ matches right}) \\ &= P(12 \text{ m. right}) + P(13 \text{ m. right}) \\ &= \frac{26}{313} + \frac{1}{313} \\ &= 0.000016953. \end{aligned}$$

Geometrical Probability

Consider a closed sample space Ω , such as a line in \mathbb{R} , an area in \mathbb{R}^2 or a volume in \mathbb{R}^3 , and let A be a subset of Ω ($A \subset \Omega$).

Then we may consider A as an event with geometrical probability (geometrisen todennäköisyys)

$$P(A) = \frac{m(A)}{m(\Omega)},$$

where m denotes the length, area, or volume of the corresponding set.

Note that this approach is no longer covered by the concept of classical probability, because there is an infinite number of possible outcomes, but the axioms by Kolmogorov do still apply.

Example. What is the probability of randomly throwing a dart such that it hits within the red area with a radius of $r = 3\text{cm}$, given that the dart will always land within the boundary of the target with a radius of $R = 15\text{cm}$?

$$P = \frac{\pi r^2}{\pi R^2} = \frac{9}{225} = 0.04.$$

5.3. Conditional Probability and Independence

In probability calculus statistical dependence (tilastollinen riippuvuus) between two events becomes visible by their probabilities depending upon whether the other event has happened or not. This leads to the concept of conditional probability (ehdollinen todennäköisyys).

Example. Consider two tosses of a fair coin, such that $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. Let:

$$\begin{aligned} A &= \{\text{identical tosses}\} = \{(H, H), (T, T)\}, \\ B &= \{\text{at least one head}\} = \{(H, H), (H, T), (T, H)\}. \\ \Rightarrow \quad P(A) &= \frac{1}{2}, \quad P(B) = \frac{3}{4}. \end{aligned}$$

The probability of A given that event B occurred is $\frac{1}{3}$, because (T, T) is no longer in the sample space, such that the favourable event has changed from A to $A \cap B = (H, H)$. This may be generalized as follows.

The conditional probability of A given B (A :n todennäköisyys ehdolla B) is defined as:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0.$$

This may be equivalently formulated as the so called general multiplication rules:

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

or:

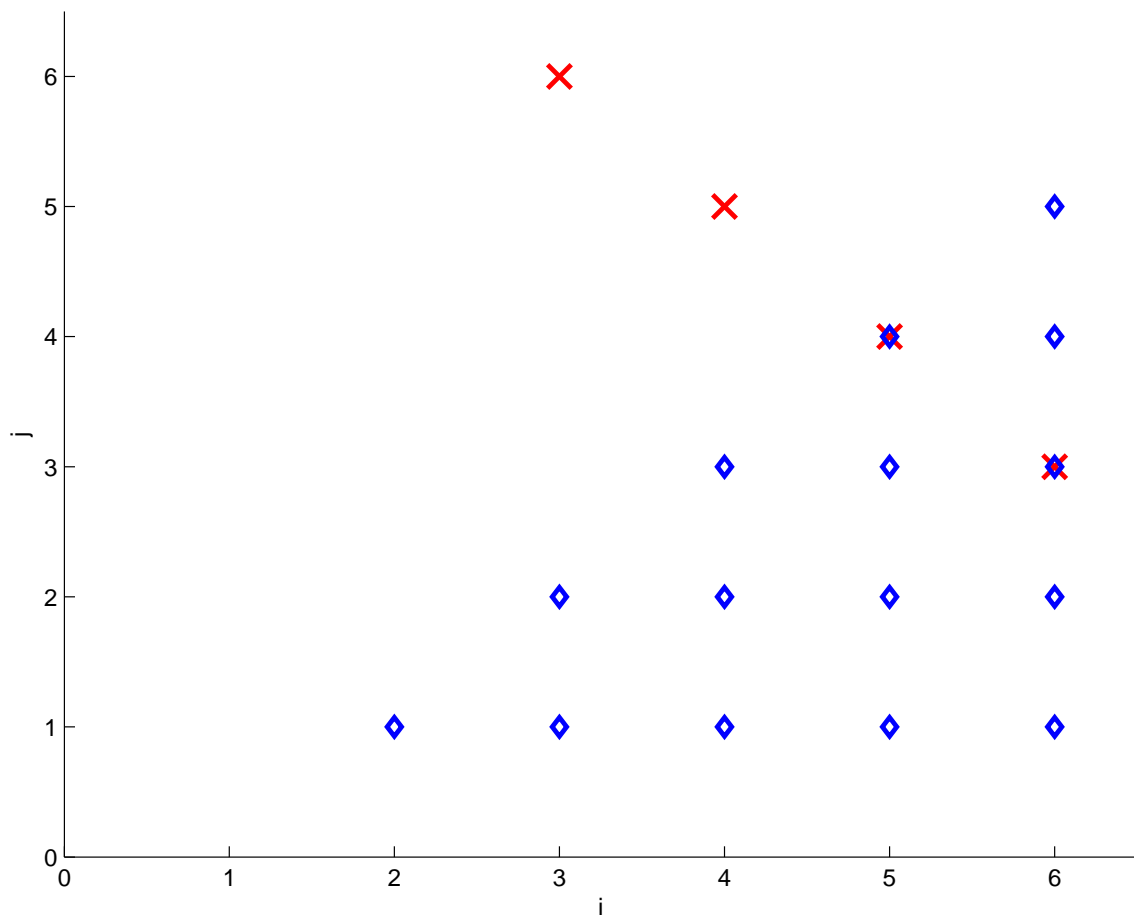
$$P(A \cap B) = P(A)P(B|A).$$

Example: What is the probability of first drawing an ace and then a red king from a deck of 52 cards?

Let $A = \{1\text{st card ace}\}$, $B = \{2\text{nd card red king}\}$.

$$P(A \cap B) = P(A)P(B|A) = \frac{4}{52} \cdot \frac{2}{51} = \frac{2}{663}.$$

Example. Two dice tosses, consider the events $A = \{(i, j) | i + j = 9\}$ and $B = \{(i, j) | i \geq j + 1\}$. What is $P(A|B)$?



$$P(B) = \frac{15}{36}, \quad P(A \cap B) = \frac{2}{36}$$

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{15/36} = \frac{2}{15}.$$

Example: Consider again two coin tosses. Let $A = \{1. \text{ toss head up}\}$ and $B = \{2. \text{ toss tail up}\}$. Then:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{1/2} = \frac{1}{2} = P(B).$$

In this case, the occurrence of the first event has no impact upon the probability of the second event.

Events A and B are called (statistically) independent (tilastollisesti riippumaton), if

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B).$$

This may be equivalently written as:

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

also known as the multiplication rule (kertolaskusääntö) for independent events.

Example. The Guests A, B, and C arrive independently of each other with probabilities 0.8, 0.6 and 0.9.

The probability that everybody arrives is:

$$P(A \cap B \cap C) = P(A)P(B)P(C) = 0.8 \cdot 0.6 \cdot 0.9 = 0.432.$$

The probability that nobody arrives is:

$$\begin{aligned} P(A^C \cap B^C \cap C^C) &= P(A^C)P(B^C)P(C^C) \\ &= 0.2 \cdot 0.4 \cdot 0.1 = 0.008. \end{aligned}$$

5.4. Total Probability and Bayes' Rule

The Law of Total Probability

Let B_1, B_2, \dots, B_n form a partition (ositus) of the sample space, that is:

$$B_1 \cup B_2 \cup \dots \cup B_n = \Omega$$

and

$$B_i \cap B_j = \emptyset \text{ for all } i \neq j.$$

$$\text{Then: } P(A) = \sum_{i=1}^n P(A \cap B_i).$$

Inserting $P(A \cap B_i) = P(A|B_i)P(B_i)$ from the definition of the conditional probability we get the law of total probability (kokonais-todennäköisyys):

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

The conditional probabilities $P(A|B_i)$ are sometimes referred to as transition probabilities (siirtymätodennäköisyyksiä) from state B_i to state A .

Example.

Concern drawing a ball from different urns depending upon the result of tossing a dice:

Dice no.	urn	white balls	black balls
1,2,3	1	1	2
4	2	2	1
5,6	3	3	3

Then, using the law of total probability,

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

with the events A =drawing a white ball, and B_i =drawing from urn i , the probability of drawing a white ball becomes:

$$P(A) = \frac{1}{3} \cdot \frac{3}{6} + \frac{2}{3} \cdot \frac{1}{6} + \frac{3}{6} \cdot \frac{2}{6} = \frac{4}{9}.$$

Bayes' Rule

Bayes' Rule (Bayesin Kaava) allows us to reverse the conditionality of events, that is, we can obtain $P(B|A)$ from $P(A|B)$. Recall for that purpose the definition of conditional probability:

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

which could be equivalently formulated as:

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

Inserting the latter equation into the former yields Bayes' Rule (Bayesin Kaava):

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

Example. (continued)

Dice no.	urn	white balls	black balls
1,2,3	1	1	2
4	2	2	1
5,6	3	3	3

Given that we drew a white ball, the probability that it was taken from urn 3 is using Bayes rule:

$$P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A)} = \frac{\frac{3}{6} \cdot \frac{2}{6}}{\frac{4}{9}} = \frac{3}{8},$$

where we have used our earlier result that the unconditional probability of drawing a white ball is $P(A) = \sum_{i=1}^3 P(A|B_i)P(B_i) = \frac{4}{9}$.

This result may be generalized to an alternative formulation of Bayes rule.

Bayes' Rule (extended)

Let B_1, B_2, \dots, B_n form a partition of Ω . Then:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}, \quad i = 1 \dots n.$$

Example.

Consider testing for a rare illness occurring with a probability of only $P(I) = 0.1\%$, so the unconditional probability of being healthy is $P(H) = 99.9\%$. The test, when administered to an ill person, reports a positive test result with probability $P(+|I) = 92\%$. However, if administered to a person who is not ill, it will erroneously report a positive test result with probability $P(+|H) = 4\%$. The probability of being ill, given a positive test result, is then:

$$\begin{aligned} P(I|+) &= \frac{P(+|I)P(I)}{P(+|I)P(I) + P(+|H)P(H)} \\ &= \frac{0.92 \cdot 0.001}{0.92 \cdot 0.001 + 0.04 \cdot 0.999} = 2.25\%. \end{aligned}$$