1. Probability

1.1. Random experiments and sample spaces

- **Random experiment:** experiment, trial, or observation that can be repeated under the same conditions and whose outcomes are uncertain but observable.
- **Sample space**: The set Ω of possible outcomes of a random experiment.
- **Sample point (or "state of the nature")**: An outcome of a random experiment, $\omega \in \Omega$.
- **Example 1:** Random experiment: Roll a dice. $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$.
- **Example 2:** Random experiment: Flip a coin. $\Omega_2 = \{H, T\}$.
- **Example 3:** Random experiment: Roll a dice and flip a coin.

$$\Omega_3 = \Omega_1 \times \Omega_2 = \{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H), \\ (1, T), (2, T), (3, T), (4, T), (5, T), (6, T)\}.$$

- **Example 4:** Random experiment: You pick an individual from a large population and observe his/her wealth measured in cents.

$$\Omega_4 = \left\{ \begin{array}{l} \text{the set of all integers comprised between} \\ \text{the minimum and the maximum wealth} \end{array} \right\}.$$

- A set Ω is discrete or countable if there exists an injective (or one-to-one) function from Ω to the set of natural numbers $\mathbb{N}$.
- A set Ω is continuous if it is not discrete.
- Sample spaces can be either **discrete** (or countable) or **continuous**.
- Discrete sample spaces can be either **finite** or (countable) **infinite**.

1.2. Composition of experiments and combinatorics

- **Fundamental theorem of combinatorics.** Consider an experiment (or operation) that consists of K sub-experiments (or steps) and each step j can be made in n_j different ways, $j = 1, 2, ..., K$, then the whole operation can be made in $\prod_{j=1}^{K} n_j = n_1 \cdot n_2 \cdot ... \cdot n_K$ different ways.

- Sometimes the experiment consists of selecting permutations (or orders or arrangements) of n objects.

- The number of **permutations** P_n of n different objects is given by

$$P_n = n! \equiv n \cdot (n-1) \cdot (n-2) \cdot ... \cdot 3 \cdot 2 \cdot 1,$$

  where n is a natural number $(n = 0, 1, 2, ...)$ and the symbol "!" denotes the factorial.

- **Variations** (or permutations) of r objects taken from a set of n different objects ($n \geq r$): V_n^r.
- Here only r objects are taken from the set of n different objects (with $n \geq r$) and, thus, two variations may differ because their objects are different or because they are arranged in a different order.
- The number of variations of r objects taken from a set of n different objects is given by

$$V_n^r = \underbrace{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-r+1)}_{r \text{ elements}} = \frac{n!}{(n-r)!},$$

  where n and r are natural numbers with $n \geq r$.
- **Example:** A club has 24 members. Its executive committee is formed by 4 members of the club: the president, the vice-president, the treasurer and the secretary. Then, this executive committee can be formed in $24 \cdot 23 \cdot 22 \cdot 21 = 255\,024$ different ways.

- **Permutations with repeated objects.**

  The number of permutations of n objects with n_1 of type 1, n_2 of type 2,..., n_K of type K, with $\sum_{j=1}^{K} n_j = n$, is given by

  $$\frac{n!}{n_1!n_2!\ldots n_K!} = \frac{n!}{\prod_{j=1}^{K} n_j!}.$$

- **Example:** With the letters appearing in the word 'Tallahassee' we can form

  $$\frac{11!}{1!3!2!1!2!2!} = \frac{39\,916\,800}{48} = 831\,600$$

  different words.

- Number of **combinations** of r objects taken from a set of n different objects $(n \geq r)$, C_n^r:
  - number of ways we can select r objects from a set of n different objects (we do not care here about the order but only about the objects we pick); or
  - number of subsets of r elements from a set of n elements.
- The number of combinations is given by the following combinatorial number:

$$C_n^r = \binom{n}{r} \equiv \frac{n!}{r!(n-r)!} = \frac{V_n^r}{r!},$$

  where n and r are natural numbers with $n \geq r$.
- The combinatorial number $\binom{n}{r}$ also gives us the number of permutations of n objects with r of type 1 and $(n-r)$ of type 2.

- Number of ways of partitioning a set with n elements into K subsets of $n_{1,} n_2, ..., n_K$ elements, with $\sum_{j=1}^{K} n_j = n$:

$$\binom{n}{n_{1,} n_2, ..., n_K} \equiv \frac{n!}{n_1! n_2! ... n_K!},$$

which coincides with the formula of the number of permutations with repeated objects.

- **Properties of combinatorial numbers:**

1.

$$\binom{n}{r} = \binom{n}{n-r}.$$

Proof:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{n!}{(n-r)!r!} = \binom{n}{n-r}.$$

2.

$$0! = 1.$$

Proof:

$$\binom{n}{n} = 1 = \frac{n!}{n!0!} = \frac{1}{0!} \Longrightarrow 0! = 1.$$

- The combinatorial number $\binom{n}{r}$ is also a coefficient in the polynomial expansion of a binomial power (Newton's binomial theorem):

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r .$$

- Moreover,

$$\left(\sum_{j=1}^{K} x_j \right)^n =$$

$$\sum_{r_1=0}^{n} \sum_{r_2=0}^{n} \dots \sum_{r_K=0}^{n} \binom{n}{r_1, r_2, \dots, r_K} \left(x_1^{r_1} \cdot x_2^{r_2} \cdot \ldots \cdot x_K^{r_K} \right), \text{ with } \sum_{j=1}^{K} r_j = n.$$

- **Example.** Let $(x+y+z)^6$. Then, the value of the coefficient c in the term cx^3yz^2 of the expansion is

$$c = \binom{6}{3, 1, 2} = \frac{6!}{3!1!2!} = 60.$$

1.3. sigma-algebras, events, and measurable spaces.

- B^c denotes the complement of B relative to Ω, that is, $B^c = \{\omega \in \Omega \,|\omega \notin B\}$
- $A \cup B = \{\omega \in \Omega \,|\omega \in A, \text{ or } \omega \in B, \text{ or both}\}$.
- $A \cap B = \{\omega \in \Omega \,|\omega \in A \text{ and } \omega \in B\}$.
- $A \backslash B = \{\omega \in \Omega \,|\omega \in A \text{ and } \omega \notin B\}$. Therefore, $B^c = \Omega \backslash B$.

- **Definition.** A σ-algebra (or σ-field) on Ω is a collection $\mathcal{F}$ of subsets of Ω, such that

  1. $\emptyset \in \mathcal{F}$, $\Omega \in \mathcal{F}$.

  2. If $B \in \mathcal{F}$, then $B^c \in \mathcal{F}$.

  3. If $\{B_1, B_2, ...\}$ is a countable (finite or infinite) collection of elements of $\mathcal{F}$, then $\bigcup\limits_i B_i \in \mathcal{F}$.

- *Note:* If $\{B_1, B_2, \ldots\}$ is a countable (finite or infinite) collection of elements of the σ-algebra $\mathcal{F}$, then $\bigcap_i B_i \in \mathcal{F}$.

- **Proof.** Note that, if $\{B_1, B_2, \ldots\}$ is a countable collection of elements of the σ-algebra $\mathcal{F}$, then $\{B_1^c, B_2^c, \ldots\}$ is also a countable collection of elements of $\mathcal{F}$. Therefore,

$$\bigcup_i B_i^c \in \mathcal{F} \Longrightarrow \left(\bigcup_i B_i^c\right)^c \in \mathcal{F} \Longrightarrow \bigcap_i B_i \in \mathcal{F}. \quad Q.E.D. \qquad (\star)$$

- Moreover, part 3 of the definition of σ-algebra could be replaced by the following:

  3'. If $\{B_1, B_2, \ldots\}$ is a countable (finite or infinite) collection of elements of $\mathcal{F}$, then $\bigcap_i B_i \in \mathcal{F}$.

  This replacement is justified by just interchanging the set operations $\bigcup_i$ and $\bigcap_i$ in $(\star)$.

- In the definition of **algebra** on Ω we replace the word "countable" in (3) by "finite".
- Therefore, a σ-algebra is an algebra, but the converse is not true.
- The elements of a σ-algebra are called measurable sets, $B \in \mathcal{F}$.
- If Ω is a sample space and we define a σ-algebra $\mathcal{F}$ on it, the elements of $\mathcal{F}$ are called "events".
- B is an event if the question "Does $\omega \in B$?" has a definite "yes" or "not" answer after the experiment has been performed and delivered the outcome ω, for all outcomes $\omega \in \Omega$.

- The smallest σ-algebra on Ω is $\{\emptyset, \Omega\}$, whereas the largest is the collection of all subsets of Ω (the power set, denoted by 2^{Ω}).
- Let the number of elements (or cardinality) of Ω be $\#\Omega = n$. The number of elements of the power set 2^{Ω} is

$$\#2^{\Omega} = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = (1+1)^n = 2^n.$$

- **Example:** We roll a dice and flip a coin. This random experiment has 12 sample points (or outcomes):

$$\Omega = \{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H),$$
$$(1, T), (2, T), (3, T), (4, T), (5, T), (6, T)\}.$$

- Examples of events in the power set 2^{Ω}:

$A = \{(2, H), (4, H), (6, H)\}$ ←Even number of points and Head

$B = \{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H)\}$ ←Head

$C = \{(2, H), (4, H)\}$ ←Number of points even no larger than 4 and Head

- The power set 2^{Ω} contains $2^{12} = 4096$ different events.

- **Definition.** Let $S = \{A_1, A_2, ...\}$ be a collection of arbitrary subsets of Ω. The σ-algebra $\sigma(S)$ generated by S is the smallest σ-algebra containing S.

- **Note:** $\{\emptyset, \Omega\} \subset \sigma(S) \subset 2^{\Omega}$.

- **Example.** Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $S = \{\{1\}, \{1, 3, 4\}\}$. Then,

  $$\sigma(S) = \{\Omega, \emptyset, \{1\}, \{1, 3, 4\}, \{2, 3, 4, 5, 6\}, \{2, 5, 6\}, \{1, 2, 5, 6\}, \{3, 4\}\}.$$

- Usually, we consider σ-algebras generated by partitions of Ω.
- **Definition.** A partition of the set Ω is a collection of subsets of Ω, $\{B_1, B_2, ...\}$ such that

  (a) $B_i \neq \emptyset$, for all i,

  (b) $B_i \bigcap B_j = \emptyset$, for all pairs (i, j) with $i \neq j$,

  (c) $\bigcup\limits_i B_i = \Omega$.

- **Example.** Let $\Omega = \{1,2,3,4,5,6\}$ and $S = \{\{1,2\}, \{3,4\}, \{5,6\}\}$ is a partition of Ω. Then,

  $\sigma(S) = \{\Omega, \emptyset, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\}, \{3,4,5,6\}\}$.

- If $\mathcal{F}_1 \subset \mathcal{F}_2$, then we say that the σ-algebra $\mathcal{F}_2$ is finer than $\mathcal{F}_1$ (or that $\mathcal{F}_1$ is coarser than $\mathcal{F}_2$).

- $\{\emptyset, \Omega\}$ is the coarsest σ-algebra on Ω, whereas the power set 2^{Ω} is the finest σ-algebra on Ω.

- **Definition.** The pair $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra on Ω is called a measurable space.

1.4. Measure, probability, and probability spaces.

- **Definition.** A measure μ on the measurable space $(\Omega, \mathcal{F})$ is a set function on $\mathcal{F}$ taking values on the set of non-negative extended real numbers,

  $$\mu : \mathcal{F} \longrightarrow [0, \infty] \equiv \overline{\mathbb{R}}_+,$$

  such that

  1. If $\{B_1, B_2, ...\}$ is a countable (finite or infinite) collection of mutually disjoint measurable sets (elements of $\mathcal{F}$), then

  $$\mu\left(\bigcup_i B_i\right) = \sum_i \mu(B_i) \qquad \text{(countable additivity).}$$

  2. $\mu(\emptyset) = 0.$

- Property 2 is redundant if at least one of the measurable sets has finite measure.

- A signed measure on the measurable space $(\Omega, \mathcal{F})$ is a countably additive set function on $\mathcal{F}$ taking values on the set of extended real numbers $\overline{\mathbb{R}} \equiv [-\infty, \infty]$ such that $\mu(\emptyset) = 0$.
- Therefore, a measure is a signed measure that takes values on $\overline{\mathbb{R}}_{+}$.

- **Definition.** The triple $(\Omega, \mathcal{F}, \mu)$, where $(\Omega, \mathcal{F})$ is a measurable space and μ is a measure on it, is called a measure space.
- **Definition.** A probability measure (or "probability") P on $(\Omega, \mathcal{F})$, where Ω is a sample space and $\mathcal{F}$ is a σ-algebra of events, is a measure such that $P(\Omega) = 1$.
- **Definition.** The triple $(\Omega, \mathcal{F}, P)$, where P is a probability on $(\Omega, \mathcal{F})$ is called a probability space.
- *Note:* We only assign probability to events.

- Usually, if the sample space is discrete and the σ-algebra of events is the power set 2^{Ω}, we assign probabilities to all the sample points and, using the countable additivity property, we construct a probability on the measurable space $\left(\Omega, 2^{\Omega}\right)$.
- **Proposition.** Consider the probability space $\left(\Omega, 2^{\Omega}, P\right)$, where $\Omega = \{\omega_1, \omega_2, ...\}$ is a discrete sample space, then

$$P(B) = \sum_{\omega_i \in B} P\{\omega_i\}.$$

- **Proposition (Laplace).** Consider the probability space $\left(\Omega, 2^{\Omega}, P\right)$. If $\#\Omega = N$, $\#B = n$, and all the sample points are equally likely, then
$$P(B) = \frac{n}{N} \quad \text{(Laplace formula)}.$$

- **Proof:** Observe that $\Omega = \bigcup_{i=1}^{N} \{\omega_i\}$. Therefore,

  $P(\Omega) = \sum_{i=1}^{N} P\{\omega_i\}$. Since all the sample points are equally likely, $P\{\omega_i\} = p$ for all i. Then,
$$1 = P(\Omega) = \sum_{i=1}^{N} p = Np,$$
  which implies that $p = \frac{1}{N}$. Moreover,
$$P(B) = \sum_{\omega_i \in B} \underbrace{P\{\omega_i\}}_{=p} = \sum_{\omega_i \in B} \frac{1}{N} = \frac{n}{N}$$
  since the event B has n elements. $Q.E.D.$

Pierre Simon de Laplace (1749 - 1827)

- **Example:** We roll a dice. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega}$, $A = \{1, 2, 3\}$, $B = \{3, 4\}$, $C = \{4, 5\}$, and $D = \{4\}$.
- Then, $P(A) = 3/6 = 1/2$, $P(B) = 2/6 = 1/3$, $P(C) = 2/6 = 1/3$ and $P(D) = 1/6$.
- $A \cup C = \{1, 2, 3, 4, 5\}$ so that $P(A \cup C) = 5/6$.
- Note also that $A \cap C = \emptyset$ so that

$$P(A \cup C) = P(A) + P(C) = \frac{3}{6} + \frac{2}{6} = \frac{5}{6}.$$

- **Note on the relationship between probability and frequency:** If we randomly pick an object (or individual) from a population having a variable X distributed according to the relative frequency f_X, then

$$P\left\{X \in B\right\} = \sum_{x \in B} f_X\left(x\right),$$

where $\{X \in B\}$ is the event where the variable X of the object we pick takes a value belonging to the set B.

- **Notation:**

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

1.5. Properties of probability measures.

1. $P(A^c) = 1 - P(A)$.
2. If $A \subset B$, then $P(A) \leq P(B)$.
3. $0 \leq P(A) \leq 1$.
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
5. $$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

- Properties 2, 4, and 5 also hold for a general measure μ and Property 1 then becomes $\mu(A^c) = \mu(\Omega) - \mu(A)$, while Property 3 becomes $0 \leq \mu(A) \leq \mu(\Omega)$.

- All the previous rules of probability have to be satisfied both when the probability is objective (i.e., based on scientific predictions) and when the probability is subjective (i.e., based on beliefs).
- **Example:** We roll a dice. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = 2^{\Omega}$, $A = \{1, 2, 3\}$ and $B = \{3, 4\}$.
- Then, $P(A) = 3/6 = 1/2$ and $P(B) = 2/6 = 1/3$.
- $A \cup B = \{1, 2, 3, 4\}$ and $A \cap B = \{3\}$.
- Then, $P(A \cup B) = 4/6 = 2/3$ and $P\left(A \cap B\right) = 1/6$.
- Note that

$$P(A \cup B) = \frac{4}{6} = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6}.$$

1.6. Conditional probability.

- **Definition.** If A and B are events in the sample space Ω and $P(A) \neq (>) 0$, the conditional probability of B given A is

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

- Therefore, if A and B are events in the sample space Ω and $P(A) \neq (>) 0$, then

$$P(A \cap B) = P(A) \cdot P\left(B \,|A\right).$$

- **Proposition.** If $P\left(A_1 \cap A_2 \cap \ldots \cap A_{n-1}\right) > 0$, then

$$P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) =$$

$$P\left(A_1\right) \cdot P\left(A_2 \left| A_1 \right.\right) \cdot P\left(A_3 \left| A_1 \cap A_2 \right.\right) \cdot \ldots \cdot P\left(A_n \left| A_1 \cap A_2 \cap \ldots \cap A_{n-1} \right.\right).$$

- **Proof.**

$$P\left(A_1\right) \cdot P\left(A_2 \left| A_1 \right.\right) \cdot P\left(A_3 \left| A_1 \cap A_2 \right.\right) \cdot \ldots \cdot P\left(A_n \left| A_1 \cap A_2 \cap \ldots \cap A_{n-1} \right.\right)$$

$$= P\left(A_1\right) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdot \ldots \cdot \frac{P(A_1 \cap A_2 \cap \ldots \cap A_n)}{P(A_1 \cap A_2 \cap \ldots \cap A_{n-1})}$$

$$= P(A_1 \cap A_2 \cap \ldots \cap A_n). \qquad Q.E.D.$$

Note: If we randomly pick an object (or individual) from a population having the variables X and Y distributed according to the relative joint frequency $f_{X,Y}$ and, thus, with the relative marginal frequencies f_X and f_Y, and the conditional frequencies $f_{X|Y}$ and $f_{Y|X}$, then

$$P\{X \in B, Y \in C\} = \sum_{x \in B} \sum_{y \in C} f_{X,Y}(x,y) \Longrightarrow$$

$$P\{X = x, Y = y\} = f_{X,Y}(x,y),$$

$$P\{X \in B\} = \sum_{x \in B} f_X(x) \implies P\{X = x\} = f_X(x),$$

$$P\{Y \in C\} = \sum_{y \in C} f_Y(y) \implies P\{Y = y\} = f_Y(y),$$

$$P\{X \in B \,|Y \in C\} = \frac{P\{X \in B, Y \in C\}}{P\{Y \in C\}} = \frac{\sum\limits_{x \in B} \sum\limits_{y \in C} f_{X,Y}(x,y)}{\sum\limits_{y \in C} f_Y(y)}$$

and

$$P\{Y \in C \,|X \in B\} = \frac{P\{X \in B, Y \in C\}}{P\{X \in B\}} = \frac{\sum\limits_{x \in B} \sum\limits_{y \in C} f_{X,Y}(x,y)}{\sum\limits_{x \in B} f_X(x)}$$

$\Longrightarrow$

$$P\{X \in B \,|Y{=}y\} = \sum_{x \in B} f_{X|Y}(x\,|y)\,, P\{Y \in C \,|X{=}x\} = \sum_{y \in C} f_{Y|X}(y\,|x)$$

and

$$P\{X = x \,|Y = y\} = f_{X|Y}(x\,|y)\,, \quad P\{Y = y \,|X = x\} = f_{Y|X}(y\,|x)\,,$$

where $\{X \in B\}$ is the event where the variable X of the object we pick takes a value belonging to the set B and $\{Y \in C\}$ is the event where the variable Y of the object we pick takes a value belonging to the set C.

- **Example.** We extract two cards with "no replacement" from a deck of poker cards. The probability that the two cards will be aces is

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 \,|A_1\,) = \frac{4}{52} \cdot \frac{3}{51} = 0.0045.$$

- Assume now that the extractions are made "with replacement", that is, the cards are introduced back in the deck after each extraction. The probability that the two cards so extracted will be aces is

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 \,|A_1\,) = \frac{4}{52} \cdot \frac{4}{52} = 0.0059.$$

- **Example:** We extract three cards with "no replacement" from a deck of poker cards. The probability that all the three cards will be aces is

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2 \,|A_1\,) \cdot P(A_3 \,|A_1 \cap A_2\,) \\ &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} = 0.000181. \end{aligned}$$

1.7. Independent events.

- **Definition.** The events A and B are independent if

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

- **Proposition.** Let $P(A) > 0$. The events A and B are independent if and only if

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

- **Proof.** Since $P(A \cap B) = P(A) \cdot P(B | A)$ if $P(A) > 0$ and the events A and B are independent (i.e., $P(A \cap B) = P(A) \cdot P(B)$), we immediately obtain the desired result. $Q.E.D.$

Example. We toss 3 coins (or the same coin three times). This random experiment has 8 equally likely outcomes,

$$\Omega = \{(HHH), (HHT), (HTH), (THH), (TTH), (THT), (HTT), (TTT)\}$$

so that the probability of each outcome is 1/8.

- Consider the following events when $\mathcal{F} = 2^{\Omega}$:

  $A = \{(HHH), (HHT)\}, \leftarrow \{\text{Head in the first two coins}\}$

  $B = \{(HHT), (HTT), (THT), (TTT)\}, \leftarrow \{\text{Tail in the third coin}\}$

  $C = \{(HTT), (THT), (TTH)\} \leftarrow \{\text{Exactly two tails}\}$

- $A \cap B = \{(HHT)\}$ and $B \cap C = \{(HTT), (THT)\}$.

- $P(A) = 2/8 = 1/4$, $P(B) = 4/8 = 1/2$, $P(C) = 3/8$, $P(A \cap B) = 1/8$, and $P(B \cap C) = 2/8 = 1/4$.
- The events A and B are independent since

$$P(A \cap B) = \frac{1}{8} = P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}.$$

- The events B and C are not independent since

$$P(B \cap C) = \frac{1}{4} \neq P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16}.$$

- **Proposition.** If A and B are independent events, then
  (a) A^c and B are independent.
  (b) A^c and B^c are independent.

- **Proof.** **(a)** Observe that

$$B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B) .$$

Since $(A \cap B)$ and $(A^c \cap B)$ are disjoint events,

$$P(B) = P\left[(A \cap B) \cup (A^c \cap B)\right] = P(A \cap B) + P(A^c \cap B) ,$$

which is equivalent to

$$P(A^c \cap B) = P(B) - P(A \cap B) .$$

Therefore, since $P(A^c) = 1 - P(A)$ and the events A and B are independent $(\text{i.e., } P(A \cap B) = P(A) \cdot P(B))$, we have

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) = P(B) - P(A) \cdot P(B) \\ &= P(B)\left[1 - P(A)\right] = P(B) \cdot P(A^c), \end{aligned}$$

which proves the independency between A^c and B.

- **(b)** Obvious from (a). $Q.E.D.$

- **Definition.** The events in the collection $S = \{A_1, A_2, ...\}$ are independent if the probability of every finite intersection of events in S equals the product of their respective probabilities.

1.8. Theorem of total probability.

- **Theorem of Total Probability.** Let $\{B_1, B_2, ...\}$ be a countable (finite or infinite) collection of events that constitutes a partition of the sample space Ω and assume that $P(B_i) > 0$ for all i. Then,
$$P(A) = \sum_i P(B_i) \cdot P(A \,|B_i\,), \quad \text{for every event } A.$$

- **Proof.** Since
$$A = A \cap \Omega = A \cap \left(\bigcup_i B_i \right) = \bigcup_i (A \cap B_i)$$
and the events in the countable collection $\{A \cap B_1, A \cap B_2, ...\}$ are disjoint, we get
$$P(A) = P\left(\bigcup_i (A \cap B_i) \right) = \sum_i P(A \cap B_i) = \sum_i P(B_i) \cdot P(A \,|B_i\,),$$
where the last equality follows since $P(A \cap B_i) = P(B_i) \cdot P(A \,|B_i\,)$ if $P(B_i) > 0$. *Q.E.D.*

- **Example.** We extract two cards with "no replacement" from a deck of poker cards. The probability that the second card is an ace is

$$P\left(A_2\right) = P\left(A_1\right) \cdot P\left(A_2 \left| A_1\right.\right) + P\left(A_1^c\right) \cdot P\left(A_2 \left| A_1^c\right.\right)$$

$$= \left(\frac{4}{52} \cdot \frac{3}{51}\right) + \left(\frac{48}{52} \cdot \frac{4}{51}\right) = \frac{4}{52}.$$

- Assume now that the extractions are made "with replacement", that is, the cards are introduced back in the deck after each extraction. The probability that the second card is an ace is, obviously,

$$P\left(A_2\right) = P\left(A_1\right) \cdot P\left(A_2 \left| A_1\right.\right) + P\left(A_1^c\right) \cdot P\left(A_2 \left| A_1^c\right.\right)$$

$$= \left(\frac{4}{52} \cdot \frac{4}{52}\right) + \left(\frac{48}{52} \cdot \frac{4}{52}\right) = \frac{4}{52}.$$

- Note that, if the extractions are made with "no replacement", then the events A_1 and A_2 are not independent since

$$P\left(A_1 \cap A_2\right) = \frac{4}{52} \cdot \frac{3}{51} \neq P\left(A_1\right) \cdot P\left(A_2\right) = \frac{4}{52} \cdot \frac{4}{52}.$$

- If the extractions are made "with replacement", then the events A_1 and A_2 are independent since

$$P\left(A_1 \cap A_2\right) = \frac{4}{52} \cdot \frac{4}{52} = P\left(A_1\right) \cdot P\left(A_2\right) = \frac{4}{52} \cdot \frac{4}{52}.$$

- The theorem of total probability can be modified to apply for intersections of events as follows:
- **Proposition.** Let $\{B_1, B_2, ...\}$ be a countable (finite or infinite) collection of events that constitutes a partition of the event B and assume that $P(B_i) > 0$ for all i. Then,

$$P(A \cap B) = \sum_i P(B_i) \cdot P\left(A \,|B_i\right), \quad \text{for every event } A.$$

- **Proof.** Since $\bigcup_i B_i = B$, we have that $\bigcup_i (A \cap B_i) = A \cap B$, where the events in the countable collection $\{A \cap B_1, A \cap B_2, ...\}$ are disjoint. Thus,

$$\begin{aligned} P(A \cap B) &= P\left(\bigcup_i (A \cap B_i)\right) = \sum_i P\left(A \cap B_i\right) \\ &= \sum_i P(B_i) \cdot P\left(A \,|B_i\right). \qquad \textit{Q.E.D.} \end{aligned}$$

- If we make $B = \Omega$ in the previous proposition, then we recover the original theorem of total probability since $P(A \cap \Omega) = P(A)$.

1.9. Bayes' theorem.

Thomas Bayes (1702 - 1761)

- **Bayes' theorem.** Let $\{B_1, B_2, ...\}$ be a countable (finite or infinite) collection of events that constitutes a partition of the sample space Ω and assume that $P(B_i) > 0$ for all i. Then, for every event A such that $P(A) > 0$,

$$P\left(B_j \,|A\right) = \frac{P(B_j) \cdot P\left(A\,|B_j\right)}{\sum_i P(B_i) \cdot P\left(A\,|B_i\right)}, \text{ for all } j = 1, 2, ...$$

- **Proof.** Obvious since

$$P\left(B_j \,|A\right) = \frac{P(A \cap B_j)}{P(A)} \text{ when } P(A) > 0,$$

and $P(A \cap B_j) = P(B_j) \cdot P\left(A\,|B_j\right)$ if $P(B_j) > 0$, while $P(A) = \sum_i P(B_i) \cdot P\left(A\,|B_i\right)$ if $P(B_i) > 0$ for all i, as follows from the theorem of total probability. *Q.E.D.*

- **Note:** $P(B_j)$ is the prior probability of the event B_j, $P(A\,|B_j)$ is the conditional probability of the event A given B_j (also called the "likelihood" of A when B_j occurs), and $P(B_j\,|A)$ is the posterior probability of the event B_j given A, for $j = 1, 2, \ldots$

- **Example:** We want to select a candidate for a job and we observe the result of a test undertaken by the candidate.

- $\{g,\, b\} \equiv \{\text{candidate is good for a job, candidate is bad for a job}\}$, constitutes a partition of Ω. Thus, $b = g^c$.

- $\{p,\, f\} \equiv \{\text{candidate has passed a test, candidate has failed a test}\}$.

- *Prior probabilities:* $P(g) = 0.25$ so that $P(b) = 0.75$.

- *Likelihoods:* $P(p\,|g) = 0.99$ so that $P(f\,|g) = 1 - P(p\,|g) = 0.01$,

  and

  $P(f\,|b) = 0.83$ so that $P(p\,|b) = 1 - P(f\,|b) = 0.17$.

- Then,

$$P(g\,|p) = \frac{\overbrace{P(g)\cdot P(p\,|g\,)}^{P(p\cap g)}}{\underbrace{P(g)\cdot P(p\,|g\,) + P(b)\cdot P(p\,|b\,)}_{P(p)}}$$

$$= \frac{\overbrace{0.25\cdot 0.99}^{0.2475}}{\underbrace{(0.25\cdot 0.99) + (0.75\cdot 0.17)}_{=0.375}} = 0.66.$$

- Note that $P(b\,|p) = 1 - P(g\,|p) = 0.34$.
- Moreover, $P(p) = 0.375$ so that $P(f) = 0.625$.
- Finally, $P(p\cap g) = P\{\text{pass and good}\} = 0.2475$.

Exercises. Probability and Statistics. IDEA.
1. Probability

1. Prove that, for any positive integer (or natural number) n larger or equal than 2 and $r = 1, 2, ..., n-1$,
$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}.$$
Hint: Use the fact that
$$(1+y)^n = (1+y)(1+y)^{n-1} = (1+y)^{n-1} + y(1+y)^{n-1}.$$

2. Prove that
$$\sum_{r=0}^{k} \binom{m}{r}\binom{n}{k-r} = \binom{m+n}{k}.$$
Hint: Use the fact that
$$(1+y)^{m+n} = (1+y)^m (1+y)^n.$$

3. Suppose that we are concerned with the completion of a highway construction job, which may be delayed because of a strike. Suppose, furthermore, that the probabilities are 0.60 that there will be a strike, 0.85 that the job will be completed on time if there is no strike, and 0.35 that the job will be completed on time if there is a strike. What is the probability that the job will be completed on time?

4. A strictly positive integer I is selected, with $P\{I = n\} = \left(\frac{1}{2}\right)^n$, $n = 1, 2, ...$ If I takes the value n, a coin with probability e^{-n} of heads is tossed once.

   (a) Find the probability that the resulting toss is a head.

   (b) Find the conditional probability of $\{I = 5\}$ given that we know that the resulting toss has been a tail.

5. The members of a consulting firm rent cars from three rental agencies: 60% from agency 1, 30% from agency 2, and 10% from agency 3. If 9% of the cars from agency 1 need a tune-up, 20% of the cars from agency 2 need a tune-up, and 6% of the cars from agency 3 need a tune-up,

   (a) what is the probability that a rental car delivered to the firm will need a tune-up?

   (b) if a rental car delivered to the firm needs a tune-up, what is the probability that it came from the rental agency 2?

6. Let us define a generalized combinatorial number as

$$\binom{r}{x} = \frac{\prod_{i=0}^{x-1} (r-i)}{x!},$$

for any real number r and any natural number x. Moreover, by definition, we have

$$\binom{r}{0} = 1,$$

for any real number r.

Prove that,

(a) If n is a natural number such that $n \geq x$, then we obtain the traditional formula for a combinatorial number,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

(b) If n is a natural number such that $n < x$,

$$\binom{n}{x} = 0.$$

(c) For any real number r and any natural number x,

$$\binom{-r}{x} = (-1)^x \binom{r+x-1}{x}.$$

(d) If n is a strictly positive natural number,

$$\binom{-n}{x} = (-1)^x \binom{n+x-1}{n-1}.$$

(e) For any natural number x,

$$\binom{-1}{x} = (-1)^x \quad \text{and} \quad \binom{-2}{x} = (-1)^x(1+x).$$

(f) If r and z are two real numbers with $|z| < 1$,

$$(1+z)^r = \sum_{x=0}^{\infty} \binom{r}{x} z^x.$$

(g) For any real number z and any natural number n,

$$(1+z)^n = \sum_{x=0}^{n} \binom{n}{x} z^x.$$

7. In a scientific meeting of 100 people, 60 of them speak English only, 30 speak French only, and the remaining 10 speak both languages. Compute the probability that two randomly selected participants will be able to understand each other.

8. Ten individuals are randomly ordered and all possible orderings are equally likely. Find the probability that two given individuals be contiguous if the 10 individuals are ordered: a) in a row, b) in a circle.

9. We have 11 urns, which are numbered from 2 to 12. The composition of the urns is the following:

number of the urn	2	3	4	5	6	7	8	9	10	11	12
number of white balls	0	1	2	3	4	5	4	3	2	1	0
number of black balls	1	1	1	1	1	1	1	1	1	1	1

We roll two dices simultaneously. If the sum of points is k, we make an extraction with replacement of a ball from the urn k. If the extracted ball is black, the game is over. If the ball is white, we roll again the dices and repeat the whole process. Compute the probability of making at least 3 successive extractions.

10. There are n balls in a box. Each ball is either black or red and we assume that the $n+1$ different possible compositions of the box are equally likely.

    (a) We randomly pick one ball and it turns out to be red. What is the probability that k out of the n balls in the box were red?

    (b) Assume that $n = 6$ and that we randomly pick two balls with replacement (i.e., we put back in the box the picked ball after each extraction) and it turns out that both balls are red. What is the probability that 5 out of the 6 balls were red in the composition of the box?

    (c) Assume now that $n = 7$ and that we extract two balls with no replacement from the box. What is the probability that both balls will be black? What is the probability that the first extracted ball will be black and the second red?

11. We have two decks of Spanish cards (with 48 cards where 12 of them are figures). We randomly extract a card from the first deck and we insert it in the second deck. Then, we roll a dice. If we get one dot, we randomly extract a card from the first deck; if we get two dots, we randomly extract a card from the second deck; otherwise, we merge the two decks and we randomly extract a card from the resulting single deck. Find the probability of extracting a figure in this second extraction.

12. The Chicago Bulls and the Golden State Warriors are playing each other in the NBA (National Basketball Association) final playoffs. In this set of games, the first team to have won four games is declared to be the (world) champion.

(a) Assume that the Bulls are slightly better than the Warriors, such that

$$P\{\text{Bulls beat the Warriors in the first game}\} = 0.6$$

and that the outcome of a game affects slightly the probability of the outcome on the following game so that

$$P\{\text{Bulls win a game if they have won the previous game}\} = 0.7$$

but

$$P\{\text{Bulls win a game if they have lost the previous game}\} = 0.5.$$

(i) What is the probability that the Warriors will win a game if they have lost the previous game?

(ii) What is the probability that the Bulls will win the playoffs in only four games?

(iii) What is the probability that the Warriors will win the playoffs in only four games?

(iv) What is the probability that the Warriors have won the first game if we know that they have lost the second game?

(b) If the two teams were evenly matched, that is,

$$P\{\text{Bulls beat the Warriors in the first game}\} = 0.5$$

and the probability of winning a game did not depend upon the outcome of the previous game, would it be more likely or less likely that the series would end in just four games than under the conditions specified in (a)?

13. In a group of 20 students of the IDEA program, 12 of them have passed the Micro exam only, 6 have passed the Macro exam only, and the remaining two students have passed both the Micro and the Macro exams. We pick randomly three students from this group.

    (a) Compute the probability that all three students have passed the same exam.

    (b) If we know that the three students we have picked have passed the same exam, what is the probability that the three have passed the Micro exam.

14. Consider the events A, B, and C on the probability space $(\Omega, \mathcal{F}, P)$. Assume that $P(A) = 1/2$, $P(B) = 1/2$, $P(C) = 1/2$, $P(A \cap B) = 1/4$, $P(A \cap C) = 1/4$, $P(B \cap C) = 1/4$, and $P(A \cap B \cap C) = 1/4$.

    (a) Show that A and B are independent, A and C are independent, and B and C are independent, but A, B, and C are not independent. *Note:* this is an example of three events that are pairwise independent without being independent.

    (b) Prove that $P(A \cup B \cup C) = 1$.

(c) Find $P\left((A \cap B) \cap C^c\right)$, where C^c is the complement of the event C. Note that the event $(A \cap B) \cap C^c$ contains all the sample points belonging to both A and B but not belonging to C. That is, $(A \cap B) \cap C^c = (A \cap B) \backslash C$.

Consider now the events D, E, and F on the probability space $(\Omega, \mathcal{F}, P)$. Assume that $P(D) = 0.6$, $P(E) = 0.8$, $P(F) = 0.5$, $P(D \cap E) = 0.48$, $P(D \cap F) = 0.3$, $P(E \cap F) = 0.38$, and $P(D \cap E \cap F) = 0.24$.

(d) Are D and E are independent? Are D and F independent? Are E and F independent? Are D, E, and F independent? Show that $P(D \cap E \cap F) = P(D) \cdot P(E) \cdot P(F)$. *Note:* this is an example of three events whose probability of their intersection is equal to the product of their respective probabilities. However, they are not independent since they are not all pairwise independent.

(e) Find $P\left((D \cap E) \cap F^c\right)$, where F^c is the complement of the event F. Note that $(D \cap E) \cap F^c = (D \cap E) \backslash F$.

(f) Find $P\left((D \cup E \cup F)^c\right)$, where $(D \cup E \cup F)^c$ is the complement of the event $D \cup E \cup F$.

15. Prove the following generalization of the theorem of total probability:

    Theorem of total conditional probability. If A and C are events of the measurable space $(\Omega, \mathcal{F})$ and $\{B_1, B_2, ...\}$ is a discrete collection of events that constitutes a partition of the sample space Ω, with $P(B_i \cap C) > 0$ for $i = 1, 2, ...$, then

    $$P(A|C) = \sum_i P(B_i|C) \cdot P(A|B_i \cap C) \quad \text{for all } A \in \mathcal{F} \text{ and } C \in \mathcal{F}.$$

16. Consider a group of n individuals and a task that has to be done. Nobody wants to do this task. The "drawing straws" method is used to select the individual who is going to do it. Thus, individuals are randomly ordered and the last individual becomes the group leader. Then, the group leader takes n straws and ensures that one of them is shorter than the others. The leader then grabs all of the straws in her fist, such that all of them appear to be of the same length. The group leader offers the clenched fist to the group. Each member of the group draws sequentially a straw from the fist of the group leader. When an individual draws the shortest straw he must perform the task and the game finishes. Obviously, if nobody has picked the shortest straw, then this straw is the one left in the leader's fist and, then, the leader has to do the task. Does the order in which individuals pick the straw affect their probability of picking the shortest straw? *Hint:* Use the theorem of Exercise 15.