Lecture Notes for MA 611 Probability

Ionut Florescu

Chapter 1

Elements of Probability Measure

What is Probability? In essence:

> Mathematical modeling of random events and phenomena. It is fundamentally different from modeling deterministic events and functions, which constitutes the traditional study of Mathematics.

However, the study of probability uses concepts and notions straight from Mathematics; in fact Measure Theory and Potential theory are expressions of abstract mathematics generalizing the theory of Probability.

1.1 Probability. A Brief History

In the XVII-th century the first notions of Probability Theory appeared. More precisely in 1654 Antoine Gombaud, Chevalier de Méré, a French nobleman with an interest in gaming and gambling questions, was puzzled by an apparent contradiction concerning a popular dice game. The game consisted in throwing a pair of dice 24 times; the problem was to decide whether or not to bet even money on the occurrence of at least one "double six" during the 24 throws. A seemingly well-established gambling rule led de Méré to believe that betting on a double six in 24 throws would be profitable, but his own calculations based on many repetitions of the 24 throws indicated just the opposite. Using modern probability language de Méré was trying to establish if such an event has probability greater than 0.5. Puzzled by this and other similar gambling problems he called the attention of the famous mathematician Blaise Pascal. In turn this led to an exchange of letters between Pascal and another famous French mathematician Pierre de Fermat, the first documented evidence of the fundamental principles of the theory of probability. A few other particular problems on games of chance had been solved before in the XV-th and XVI-th centuries by Italian mathematicians; however, no general theory had been formulated before this famous correspondence.

In 1655 during his first visit to Paris the Dutch scientist Christian Huygens, learnt of the work on probability carried out in this correspondence. On his return to Holland in 1657, Huygens wrote a small work *De Ratiociniis in Ludo Aleae* the first printed work on the calculus of probabilities. It was a treatise on problems associated with gambling. Because of the inherent appeal of games of chance, probability theory soon became popular, and the subject developed rapidly during the XVIII-th century.

The major contributors during this period were Jacob Bernoulli (1654-1705) and Abraham de Moivre (1667-1754). Jacob (Jacques) Bernoulli was a Swiss mathematician who was the first to use the term integral. He was the first mathematician in the Bernoulli family, a family of famous scientists of the XVIII-th century. Jacob Bernoulli's most original work was *Ars Conjectandi* published in Basel in 1713, eight years after his death. The work was incomplete at the time of his death but it is still a work of the greatest significance in the theory of probability. De Moivre was a French mathematician who lived most of his life in England. A protestant, he was pushed to leave France after Louis XIV revoked the Edict of Nantes in 1685, leading to the expulsion of the Huguenots. De Moivre pioneered the modern approach to the theory of probability, when he published *The Doctrine of Chance: A method of calculating the probabilities of events in play* in 1718, although a Latin version had been presented to the Royal Society and published in the Philosophical Transactions in 1711. The definition of statistical independence appears in this book for the first time. *The Doctrine of Chance* appeared in new expanded editions in 1718, 1738 and 1756. The birthday problem (in a slightly different form) appeared in the 1738 edition, the gambler's ruin problem in the 1756 edition. The 1756 edition of *The Doctrine of Chance* contained what is probably de Moivre's most significant contribution to probability, namely the approximation of the binomial distribution by the normal distribution in the case of a large number of trials - which is honored by most probability textbooks as "The First Central Limit Theorem" (we will discuss this theorem in the course of this semester). He perceives as we will see the notion of standard deviation and is the first to write the normal integral. In *Miscellanea Analytica* (1730) he derives Stirling's formula (wrongly attributed to Stirling) which he uses in his proof of the central limit theorem. In the second edition of the book in 1738 de Moivre gives credit to Stirling for an improvement to the formula. De Moivre wrote:

> "I desisted in proceeding farther till my worthy and learned friend Mr James Stirling, who had applied after me to that inquiry, [discovered that $c = \sqrt{2}$]."

De Moivre also investigated mortality statistics and the foundation of the theory of annuities. In 1724 he publishes based on population data for the city of Breslau *Annuities on lives* one of the first statistical applications into finance. In fact in *A history of the mathematical theory of probability* (London, 1865), Todhunter says that probability:

> ... owes more to [de Moivre] than any other mathematician, with the single exception of Laplace.

Despite de Moivre's extraordinary scientific eminence his main income was as a private tutor of mathematics and he died in poverty. None of his influential friends: Leibnitz, Newton, Halley could help him find a university position.

De Moivre, like Cardan, is famed for predicting the day of his own death. He found that he was sleeping 15 minutes longer each night and summing the arithmetic progression, calculated that he would die on the day that he slept for 24 hours. He was right!

The XIX-th century saw the development and generalization of the early theory. Pierre-Simon de Laplace (1749-1827) publishes in 1812 *Théorie Analytique des Probabilités*. This is the first fundamental book in probability ever published (the second being Kolmogorov's monograph from 1933). Before Laplace, probability theory was solely concerned with developing a mathematical analysis of games of chance.The first edition was dedicated to Napoleon-le-Grand but, for obvious reason, the dedication was removed in later editions!

The work consisted of two books and a second edition two years later saw an increase in the material by about an extra 30 per cent. The work studies generating functions, Laplace's definition of probability, Bayes rule (so named by Poincaré many years later), the notion of mathematical expectation, probability approximations, a discussion of the method of least squares, Buffon's needle problem, and inverse Laplace transform. Later editions of the "Théorie Analytique des Probabilités" also contains supplements which consider applications of probability to determine errors in observations arising in astronomy, the other passion of Laplace.

Laplace had always changed his views with the changing political events of the time, modifying his opinions to fit in with the frequent political changes which were typical of this period. Laplace became Count of the Empire in 1806 and he was named a marquis in 1817 after the restoration of the Bourbons.

On the morning of Monday 5 March 1827 Laplace died. Few events would cause the Academy to cancel a meeting but they did so on that day as a mark of respect for one of the greatest scientists of all time.

Many workers have contributed to the theory since Laplace's time; among the most important are Chebyshev, Markov, von Mises, and Kolmogorov.

One of the difficulties in developing a mathematical theory of probability has been to arrive at a definition of probability that is precise enough for use in mathematics, yet comprehensive enough to be applicable to a wide range of phenomena. The search for a widely acceptable definition took nearly three centuries and was marked by much controversy. The matter was finally resolved in the 20th century by treating probability theory on an axiomatic basis. In 1933 a monograph by the Russian giant mathematician Andrey Nikolaevich Kolmogorov (1903-1987) outlined an axiomatic approach that forms the basis for the modern theory. In 1925 the year he started his doctoral studies, Kolmogorov published his first paper with Khinchin on the probability theory. The paper contains among other inequalities about partial series of random variables the three series theorem which provides important tools for stochastic calculus. In 1929 when he finished his doctorate he already had published 18 papers among them versions of the strong law of large numbers and the iterated logarithm.

In 1933, two years after his appointment as a professor at Moscow University, Kolmogorov publishes *Grundbegriffe der Wahrscheinlichkeitsrechnung* his most fundamental book. In it he builds up probability theory in a rigorous way from fundamental axioms in a way comparable with Euclid's treatment of geometry. He gives a rigorous definition of the conditional expectation which later becomes fundamental for the definition of Brownian motion, stochastic integration, and Mathematics of Finance. (Kolmogorov's monograph is available in English translation as *Foundations of Probability Theory*, Chelsea, New York, 1950). And he was not finished. In 1938 he publishes the paper *Analytic methods in probability theory* which lay the foundation work for the Markov processes, and toward a more rigurous approach to the Markov Chains.

Kolmogorov later extended his work to study the motion of the planets and the turbulent flow of air from a jet engine. In 1941 he published two papers on turbulence which are of fundamental importance. In 1953 and 1954 two papers by Kolmogorov, each of four pages in length, appeared. These are on the theory of dynamical systems with applications to Hamiltonian dynamics. These papers mark the beginning of KAM-theory, which is named after Kolmogorov, Arnold and Moser. Kolmogorov addressed the International Congress of Mathematicians in Amsterdam in 1954 on this topic with his important talk *General theory of dynamical systems and classical mechanics.* He thus demonstrated the vital role of probability theory in physics. His contribution in the topology theory is of outmost importance.

Kolmogorov had many interests outside mathematics, in particular he was interested in the form and structure of the poetry of Pushkin.

Like so many other branches of mathematics, the development of probability theory has been stimulated by the variety of its applications. In its turn, each advance in the theory has enlarged the scope of its influence. Mathematical statistics is one important branch of applied probability; other applications occur in such widely different fields as genetics, biology, psychology, economics, finance, engineering, mechanics, optics, thermodynamics, quantum mechanics, computer vision, etc.etc.etc.. In fact I compel the reader to find one area in today's science where no applications of the probability theory can be found.

For its immense success and wide variety of applications the Theory of Probability can be arguably viewed as the most important area of Mathematics.

1.2 Probability. A somewhat rigorous approach

The axiomatic approach of Kolmogorov is followed by most Probability Theory books. This is the approach of choice for most graduate level probability courses. However, the immediate applicability of the theory learned as such is questionable and many years of study are required to understand and unleash its full power.

On the other hand the Applied probability books completely disregard this approach and they go more or less directly into treating applications, thus leaving gaps into the reader's knowledge. At a cursory glance this approach appears to be very useful (the presented problems are all very real and most are difficult), however I question the utility of this approach when confronted with problems that are slightly different from the ones presented in such books.

Unfortunately, there is no middle ground between these two, hence the necessity of the present lecture notes. I will start with the axiomatic approach and present as much as I feel is going to be necessary for a complete understanding of the Theory of Probabilities. I will skip proofs which I will consider will not bring something new to the development of the student's understanding.

1.2.1 Probability Spaces

Let Ω be an abstract set. This is sometimes denoted with S and is called the sample space. It is a set containing all the possible outcomes or results of a random experiment or phenomenon. I called it abstract because it could contain anything. For example if the experiment consists in tossing a coin once the space Ω could be represented as $\{Head, Tail\}$. However, it could just as well be represented as $\{Cap, Pajura\}$, these being the romanian equivalents of $Head$ and $Tail$. The space Ω could as well contain an infinite number of elements. For example measuring the diameter of a doughnut could result in possible numbers inside a whole range. Furthermore, measuring in inches or in centimeters would produce different albeit equivalent spaces.

We will use $\omega \in \Omega$ to denote a generic outcome or a sample point.

We will use capital letters from the beginning of the alphabet A, B, C to denote events (any collection of outcomes).

So far so good. Now comes the first different notion. We have these events which are nothing more than subsets of Ω. Now we need to talk about collection of events. Think of the following possible situation: Poles of various sizes are painted in all the possible colors and nuances of colors. Suppose that in this model we have to calculate the probability of things like the next pole would be shorter than 15 inches and painted a nuance of red or blue. In order to deal with such examples we have to give a definition of probability that will be consistent and it will allow us to deal with such cases.

We will introduce the notion of σ-algebra (or σ-field) to deal with the problem of the proper domain of definition for the probability. Before we do that, we will introduce a special collection of events:

$$\mathscr{P}(\Omega) = \text{The collection of all possible subsets of } \Omega \tag{1.1}$$

Exercise 1. Roll a die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$. An example of a event is $A = \{$ Roll an even number$\} = \{2, 4, 6\}$. Find the cardinality (number of elements of $\mathscr{P}(\Omega)$ in this case.

Having defined sets we can now define operations with them: *union, intersection, complement* and slightly less important *difference and symmetric difference.*

$$\begin{cases} A \cup B & = \text{set of elements that are } \textbf{either} \text{ in} A \textbf{ or} \text{ in } B \\ A \cap B & = AB = \text{set of elements that are } \textbf{both} \text{ in } A \textbf{ and} \text{ in } B \\ A^c & = \bar{A} = \text{set of elements that are in } \Omega \text{ but } \textbf{not} \text{ in } A \end{cases} \tag{1.2}$$

$$\begin{cases} A \setminus B = & \text{set of elements that are in } A \text{ but } \textbf{not} \text{ in } B \\ A \triangle B = & (A \setminus B) \cup (B \setminus A) \end{cases}$$

We can of course express every operation in terms of union and intersection. There are important relations between these operations, I will stop here with the mention of the De Morgan laws:

$$\begin{cases} (A \cup B)^c & = A^c \cap B^c \\ (A \cap B)^c & = A^c \cup B^c \end{cases} \tag{1.3}$$

For more details read any textbook where set operations are defined.

Definition 1.2.1 (Algebra on Ω)**.** *A collection $\mathscr{F}$ of events in Ω is called an algebra (or field) on Ω iff:*

a) $\Omega \in \mathscr{F}$

b) Closed under complementarity: If $A \subseteq \mathscr{F}$ *then* $A^c \subseteq \mathscr{F}$

c) *Closed under finite union: If* $A, B \subseteq \mathscr{F}$ *then* $A \cup B \subseteq \mathscr{F}$

Remark 1.2.2. The first two properties imply that $\varnothing \in \mathscr{F}$. The third is equivalent by de Morgan laws (1.3) with $A \cap B \subseteq \mathscr{F}$

Definition 1.2.3 (σ-Algebra on Ω)**.** *If $\mathscr{F}$ is an algebra on Ω and in addition it is closed under countable unions then it is a σ-algebra (or σ-field) on Ω*

Note: Closed under countable unions means that the property c) in Definition 1.2.1 is replaced with: If $n \in \mathbb{N}$ is a natural number and $A_n \subseteq \mathscr{F}$ for all n then

$$\bigcup_{n\in\mathbb{N}} A_n \subseteq \mathscr{F}$$

Exercise 2 (An algebra which is not a σ-algebra). Let $\mathscr{B}_0$ be the collection of sets of the form: $(a_1, a_1'] \cup (a_2, a_2'] \cup \cdots \cup (a_m, a_m']$, for any $m \in \mathbb{N}^* = \{1, 2 \dots\}$ and all $a_1 < a_1' < a_2 < a_2' < \cdots < a_m < a_m'$ in $\Omega = (0, 1]$

Verify that $\mathscr{B}_0$ is an algebra. Show that $\mathscr{B}_0$ is not a σ-algebra.

Exercise 3. Let $\mathscr{F} = \{A \subseteq \Omega | \, A \text{ finite } \textbf{or } A^c \text{ is finite}\}$.

a) Show that $\mathscr{F}$ is an algebra

b) Show that if Ω is finite then $\mathscr{F}$ is a σ-algebra

c) Show that if Ω is infinite then $\mathscr{F}$ is **not** a σ-algebra

Exercise 4 (A σ-Algebra does not necessarily contain all the events in Ω). Let $\mathscr{F} = \{A \subseteq \Omega | \, A \text{ countable } \textbf{or } A^c \text{ is countable}\}$. Show that $\mathscr{F}$ is a σ-algebra.

Note that if Ω is uncountable implies that it contains a set A such that both A and A^c are uncountable thus $A \notin \mathscr{F}$.

The σ-algebra provides an appropriate domain of definition for the probability function. However, it is such an abstract thing that it will be hard to work with it. This is the reason for the next definition, it will be much easier to work on the generators of a *sigma*-algebra. This will be a recurring theme in probability, in order to show a property for a big class we show the property for a small generating set of the class and then use standard arguments to extend to the whole class.

Definition 1.2.4 (σ algebra generated by a class of Ω)**.** *Let $\mathscr{C}$ be a collection (class) of subsets of Ω. Then $\sigma(\mathscr{C})$ is the smallest σ-algebra on Ω that contains $\mathscr{C}$.*

Mathematically:

(a) $\mathscr{C} \subseteq \sigma(\mathscr{C})$

(b) $\sigma(\mathscr{C})$ *is a* σ*-field*

(c) *If* $\mathscr{C} \subseteq \mathscr{G}$ *and* $\mathscr{G}$ *is a* σ*-field then* $\sigma(\mathscr{C}) \subseteq \mathscr{G}$

Remark 1.2.5. Properties of σ-algebras:

- $\mathscr{P}(\Omega)$ is a σ-algebra, the largest possible σ-algebra on Ω
- If $\mathscr{F}$ is already a σ-algebra then $\sigma(\mathscr{F}) = \mathscr{F}$
- If $\mathscr{F} = \{\varnothing\}$ or $\mathscr{F} = \{\Omega\}$ then $\sigma(\mathscr{F}) = \{\varnothing, \Omega\}$, the smallest possible σ-algebra on Ω
- If $\mathscr{F} \subseteq \mathscr{F}'$ then $\sigma(\mathscr{F}) \subseteq \sigma(\mathscr{F}')$
- If $\mathscr{F} \subseteq \mathscr{F}' \subseteq \sigma(\mathscr{F})$ then $\sigma(\mathscr{F}') = \sigma(\mathscr{F})$

Example: Borel σ-algebra.

Let Ω be a topological space (think geometry exists in this space this assures us that the open subsets exist in this space). Then we define:

$$\begin{aligned} \mathscr{B}(\Omega) &= \text{The Borel } \sigma\text{-algebra} \\ &= \sigma\text{-algebra generated by the class of open subsets of } \Omega \end{aligned} \tag{1.4}$$

In the special case when $\Omega = \mathbb{R}$ we denote $\mathscr{B} = \mathscr{B}(\mathbb{R})$. $\mathscr{B}$ is the most important σ-algebra. The reason is that most experiments can be brought to equivalence with $\mathbb{R}$. Thus, if we define a probability measure on $\mathscr{B}$, we have a way to calculate probabilities for most experiments.

Most subsets of $\mathbb{R}$ are in $\mathscr{B}$. However, it is possible (though very difficult) to construct a subset of $\mathbb{R}$ explicitly which is not in $\mathscr{B}$. See Billingsley page 45 for such a construction in the case $\Omega = (0, 1]$.

Exercise 5. Show that $\mathscr{B} = \sigma\left(\{(-\infty, x] | \, x \in \mathbb{R}\}\right)$

We are finally in the position to define a space on which we can introduce the probability measure.

Definition 1.2.6 (Measurable Space.)**.** *A pair* $(\Omega, \mathscr{F})$*, where* Ω *is a set and* $\mathscr{F}$ *is a* σ*-algebra on* Ω *is called a Measurable Space.*

Definition 1.2.7 (Probability measure. Probability space)**.** *Given a measurable space $(\Omega, \mathscr{F})$, a probability measure is any function $\mathcal{P} : \mathscr{F} \to [0,1]$ with the following properties:*

i) $\mathcal{P}(\Omega) = 1$

ii) (countable additivity) For any sequence $\{A_n\}_{n\in\mathbb{N}}$ of disjoint events in $\mathscr{F}$ (i.e. $A_i \cap A_j = \varnothing$, for all $i \neq j$):

$$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

The triple $(\Omega, \mathscr{F}, \mathcal{P})$ is called a Probability Space.

Exercise 6 (Discrete Probability Space). Let Ω be a countable space. Let $\mathscr{F} = \mathscr{P}(\Omega)$. Let $p : \Omega \to [0, N)$ be a function on Ω such that $\sum_{\omega\in\Omega} p(\omega) = N < \infty$, where N is a finite constant. Define:

$$\mathcal{P}(A) = \frac{1}{N} \sum_{\omega\in A} p(\omega)$$

Show that $(\Omega, \mathscr{F}, \mathcal{P})$ is a Probability Space.

Remark 1.2.8. The previous exercise gives a way to construct discrete probability measures (distributions). For example take $\Omega = \mathbb{N}$, take $N = 1$. Then:

- $p(\omega) = \begin{cases} 1-p & \text{, if } \omega = 0 \\ p & \text{, if } \omega = 1 \\ 0 & \text{, otherwise} \end{cases}$, gives the Bernoulli(p) distribution.

- $p(\omega) = \begin{cases} \binom{n}{\omega} p^{\omega}(1-p)^{n-\omega} & \text{, if } \omega \le n \\ 0 & \text{, otherwise} \end{cases}$, gives the Binomial(n,p) distribution.

- $p(\omega) = \begin{cases} \binom{\omega-1}{r-1} p^{r}(1-p)^{\omega-r} & \text{, if } \omega \ge r \\ 0 & \text{, otherwise} \end{cases}$, gives the Negative Binomial(r,p) distribution.

- $p(\omega) = \frac{\lambda^n}{n!} e^{-\lambda}$, gives the Poisson ($\lambda$) distribution.

Proposition 1.2.9 (Elementary properties of Probability Measure)**.** *Let* $(\Omega, \mathscr{F}, \mathcal{P})$ *be a Probability Space. Then:*

(1) $\forall A, B \in \mathscr{F}$ *with* $A \subseteq B$ *then* $\mathcal{P}(A) \le \mathcal{P}(B)$

(2) $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$, $\forall A, B \in \mathscr{F}$

(3) (General Inclusion-Exclusion formula, also named Poincaré formula):

$$\mathcal{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathcal{P}(A_i) - \sum_{i<j\le n} \mathcal{P}(A_i \cap A_j) \tag{1.5}$$
$$+ \sum_{i<j<k\le n} \mathcal{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^n \mathcal{P}(A_1 \cap A_2 \cdots \cap A_n)$$

Successive partial sums are alternating between over-and-under estimating.

(4) (Finite subadditivity, sometimes called Boole's inequality):

$$\mathcal{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathcal{P}(A_i), \quad \forall A_1, A_2, \ldots, A_n \in \mathscr{F}$$

Exercise 7. Prove properties 1-4 above.

1.2.2 Null element of $\mathscr{F}$. Almost sure (a.s.) statements. Indicator of a set.

An event $N \in \mathscr{F}$ is called a null event if $P(N) = 0$.

Definition 1.2.10. *A statement* $\mathcal{S}$ *about points* $\omega \in \Omega$ *is said to be true* almost surely *(a.s.) or with probability 1 (w.p.1) if the set* N *defined as:*

$$N := \{\omega \in \Omega | \, \mathcal{S}(\omega) \text{ is true}\},$$

is in $\mathscr{F}$ *and* $\mathcal{P}(N) = 1$*, (or* N^c *is a null set).*

Definition 1.2.11. *We define the indicator function of an event* A *as the (simple) function* $\mathbf{1}_A : \Omega \to \{0, 1\}$,

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & , \quad \text{if } \omega \in A \\ 0 & , \quad \text{if } \omega \notin A \end{cases}$$

Remember this definition, it is one of the most important ones in probability.

1.3 Monotone Convergence properties of probability

Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a Probability Space.

Lemma 1.3.1. *The following are true:*

(i) If $A_n, A \in \mathscr{F}$ *and* $A_n \uparrow A$ *(i.e.,* $A_1 \subseteq A_2 \subseteq \dots A_n \subseteq \dots$ *and* $A = \bigcup_{n\geq 1} A_n$*), then:* $\mathcal{P}(A_n) \uparrow \mathcal{P}(A)$ *as a sequence of numbers.*

(ii) If $A_n, A \in \mathscr{F}$ *and* $A_n \downarrow A$ *(i.e.,* $A_1 \supseteq A_2 \supseteq \dots A_n \supseteq \dots$ *and* $A = \bigcap_{n\geq 1} A_n$*), then:* $\mathcal{P}(A_n) \downarrow \mathcal{P}(A)$ *as a sequence of numbers.*

(iii) (Countable subadditivity) If $A_1, A_2, \dots,$ *and* $\bigcup_{i=1}^{\infty} A_n \in \mathscr{F}$*, with* A_i*'s not necessarily disjoint then:*

$$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

Proof. (i) Let $B_1 = A_1, B_2 = A_2 \setminus A_1, \dots, B_n = A_n \setminus A_{n-1}$. Because the sequence is increasing we have that the B_i's are disjoint thus from Proposition 1.2.9 we obtain:

$$\mathcal{P}(A_n) = \mathcal{P}(B_1 \cup B_2 \cup \dots \cup B_n) = \sum_{i=1}^{n} \mathcal{P}(B_i).$$

Thus using countable additivity:

$$\mathcal{P}\left(\bigcup_{n\geq 1} A_n\right) = \mathcal{P}\left(\bigcup_{n\geq 1} B_n\right) = \sum_{i=1}^{\infty} \mathcal{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathcal{P}(B_i) = \lim_{n\to\infty} \mathcal{P}(A_n)$$

(ii) Note that $A_n \downarrow A \;\Leftrightarrow\; {A_n}^c \uparrow A^c$ which from part (i) implies that $1 - \mathcal{P}(A_n) \uparrow 1 - \mathcal{P}(A)$.

(iii) Let $B_1 = A_1, B_2 = A_1 \cup A_2, \dots, B_n = A_1 \cup \dots \cup A_n, \dots$. From the finite subadditivity we have that $\mathcal{P}(B_n) = \mathcal{P}(A_1 \cup \dots \cup A_n) \leq \mathcal{P}(A_1) + \dots + \mathcal{P}(A_n)$.

$\{B_n\}_{n\geq 1}$ is an increasing sequence of events, thus from (i) we get that $\mathcal{P}(\bigcup_{n=1}^{\infty} B_n) = \lim_{n\to\infty} \mathcal{P}(B_n)$. Combining the two relations above we obtain:

$$\mathcal{P}(\bigcup_{n=1}^{\infty} A_n) = \mathcal{P}(\bigcup_{n=1}^{\infty} B_n) \leq \lim_{n\to\infty} (\mathcal{P}(A_1) + \dots + \mathcal{P}(A_n)) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

□

Lemma 1.3.2. *The union of a countable number of $\mathcal{P}$-null sets is a $\mathcal{P}$-null set*

Exercise 8. Prove the above Lemma 1.3.2

1.4 Conditional Probability

Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a Probability Space. Then for $A, B \in \mathscr{F}$ we define the conditional probability of A given B as usual by:

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

We of course also have the *chain rule formulas*:

$$\begin{aligned}
&\mathcal{P}(A \cap B) = \mathcal{P}(A|\, B)\mathcal{P}(B), \\
&\mathcal{P}(A \cap B \cap C) = \mathcal{P}(A|\, B \cap C)\mathcal{P}(B|\, C)\mathcal{P}(C), \quad \text{etc.}
\end{aligned}$$

Total probability formula: Given $A_1, A_2, \ldots, A_n$ a partition of Ω (i.e. the sets A_i are disjoint and $\Omega = \bigcup_{i=1}^n A_i$), then:

$$\mathcal{P}(B) = \sum_{i=1}^{n} \mathcal{P}(B|\, A_i)\mathcal{P}(A_i), \quad \forall B \in \mathscr{F} \tag{1.6}$$

Bayes Formula: If $A_1, A_2, \ldots, A_n$ form a partition of Ω:

$$\mathcal{P}(A_j|\, B) = \frac{\mathcal{P}(B|\, A_j)\mathcal{P}(A_j)}{\sum_{i=1}^n \mathcal{P}(B|\, A_i)\mathcal{P}(A_i)}, \quad \forall B \in \mathscr{F}. \tag{1.7}$$

Exercise 9. Prove the *total probability formula* and the *Bayes Formula*