Chapter 1
Elements of Probability Measure

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 presenting 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 consider will not bring something new to the development of the student's understanding.

1.1 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 just as well contain an infinite number of elements. For example measuring the diameter of a doughnut could result in all 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.

Any collection of outcomes is called an event. That is, any subset of Ω is an event. We shall use capital letters from the beginning of the alphabet A, B, C to denote these events.

So far so good. The proper definition of Ω is one of the most important issues when treating a problem probabilistically. However, this is not enough. We have to make sure that we can calculate the probability of all the items of interest.

Think of the following possible situation: Poles of various sizes are painted in all the possible nuances of colors. In other words the poles have two characteristics of interest size and color. 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 answer such questions we have to define properly the sample space Ω and furthermore give a definition of probability that will be consistent. Specifically, we need to give a definition of the elements of Ω which **can be** measured.

To this end we have to group these events into some way that would allow us to say: yes we can calculate the probability of all the events in this group. In other words, we need to talk about the notion of collection of events.

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 introduce a special collection of events:

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

We could define probability on this very large set. However, this would mean that we would have to define probability for every single element of $\mathscr{P}(\Omega)$. This will prove impossible except on the case when Ω is finite. However, even in this case we have to do it consistently. For example if say the set $\{1,2,3\}$ is in Ω and has probability 0.2, how do we define the probability of $\{1,2\}$? How about probability of $\{1,2,5\}$? A much better approach would be to define probability only on the generators of the collection $\mathscr{P}(\Omega)$ or on the generators of a collection of sets as close as we can possibly make to $\mathscr{P}(\Omega)$.

How do we do this? Fortunately, algebra comes to the rescue. The elements of a collection of events are the events. So first we 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 to mention the De Morgan laws:

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

There is much more to be found out about set operations but for our purpose this is enough. Look at Billingsley (1995) or Chung (2000) for a wealth of more details.

Definition 1.1 (Algebra on Ω). A collection $\mathscr{F}$ of events in Ω is called an algebra (or field) on Ω iff:

1. $\Omega \in \mathscr{F}$
2. Closed under complementarity: If $A \subseteq \mathscr{F}$ then $A^c \subseteq \mathscr{F}$
3. Closed under finite union: If $A, B \subseteq \mathscr{F}$ then $A \cup B \subseteq \mathscr{F}$

Remark 1.1. The first two properties imply that $\varnothing \in \mathscr{F}$. The third is equivalent with $A \cap B \subseteq \mathscr{F}$ by the second property and the de Morgan laws (1.3).

Definition 1.2 (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 third property in Definition 1.1 is replaced with: If $n \in \mathbb{N}$ is a natural number and $A_n \subseteq \mathscr{F}$ for all n then

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

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 1.3 (σ algebra generated by a class $\mathscr{C}$ of sets in 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:

1. $\mathscr{C} \subseteq \sigma(\mathscr{C})$
2. $\sigma(\mathscr{C})$ is a σ-field
3. If $\mathscr{C} \subseteq \mathscr{G}$ and $\mathscr{G}$ is a σ-field then $\sigma(\mathscr{C}) \subseteq \mathscr{G}$

The idea of this definition is to verify a statement on the set $\mathscr{C}$. Then, due to the properties that would be presented later the same statement will be valid for all the sets in $\sigma(\mathscr{C})$.

Proposition 1.1. *Properties of σ-algebras:*

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

In general listing the elements of a sigma algebra explicitly is hard. It is only in simple cases that this is done.

Remark 1.2 (Finite space Ω). When the sample space is finite, we can and typically will take the sigma algebra to be $\mathscr{P}(\Omega)$. Indeed, any event of a finite space can be trivially expressed in terms of individual outcomes. In fact, if the finite space Ω contains M possible outcomes, then the number of possible events is finite and is equal with 2^M.

Example 1.1. Suppose a set $A \subset \Omega$. Let us calculate $\sigma(A)$. Clearly, by definition Ω is in $\sigma(A)$. Using the complementarity property we clearly see that A^c and $\emptyset$ are also in $\sigma(A)$. We only need to take unions of these sets and see that there are no more new sets. Thus:

$$\sigma(A) = \{\Omega, \emptyset, A, A^c\}.$$

□

Proposition 1.2 (Intersection and union of σ-algebras). *Suppose that $\mathscr{F}_1$ and $\mathscr{F}_2$ are two σ-algebras on Ω. Then:*

1. *$\mathscr{F}_1 \cap \mathscr{F}_2$ is a sigma algebra.*
2. *$\mathscr{F}_1 \cup \mathscr{F}_2$ is **not** a sigma algebra. The smallest σ algebra that contains both of them is: $\sigma(\mathscr{F}_1 \cup \mathscr{F}_2)$ and is denoted $\mathscr{F}_1 \vee \mathscr{F}_2$*

Proof. For part 2 there is nothing to show. Perhaps a counterexample. Take for instance two sets $A, B \subset \Omega$ such that $A \cap B \neq \emptyset$. Then take $\mathscr{F}_1 = \sigma(A)$ and $\mathscr{F}_2 = \sigma(B)$. Use the previous example and Exercise 1.2, part *c*.

For part 1 we just need to verify the definition of the sigma algebra. For example, take A in $\mathscr{F}_1 \cap \mathscr{F}_2$. So A belongs to both collections of sets. Since $\mathscr{F}_1$ is a sigma algebra by definition $A^c \in \mathscr{F}_1$. Similarly $A^c \in \mathscr{F}_2$. Therefore, $A^c \in \mathscr{F}_1 \cap \mathscr{F}_2$. The rest of the definition is verified in a similar manner. □

An example: Borel σ-algebra

Let Ω be a topological space (think geometry is defined in this space and this assures us that the open subsets exist in this space).

Definition 1.4. 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 $\mathscr{B} = \mathscr{B}(\mathbb{R})$, the Borel sets of $\mathbb{R}$. This $\mathscr{B}$ is the most important σ-algebra. The reason for this fact is that most experiments can be brought to equivalence with $\mathbb{R}$ (as we shall see when we will talk about random variables). 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 explicitly construct a subset of $\mathbb{R}$ which is not in $\mathscr{B}$. See (Billingsley, 1995, page 45) for such a construction in the case $\Omega = (0,1]$.

There is nothing special about the open sets, except for the fact that they can be defined in any topological space. In $\mathbb{R}$ we have alternate definitions which you will have to show are equivalent with the one given above in problem 1.7.

Probability measure

We are finally in the position to give the domain for the probability measure.

Definition 1.5 (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 1.6 (Probability measure. Probability space). Given a measurable space $(\Omega, \mathscr{F})$, a probability measure is any function $\mathbf{P} : \mathscr{F} \to [0,1]$ with the following properties:

i) $\mathbf{P}(\Omega) = 1$
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 $i \neq j$):

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

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

Note that the probability measure is a set function (i.e., a function defined on sets).

The next two definitions are given for completeness only. However, we will use them later in this class. They are both presenting more general notions than a probability measure and they will be used later in hypotheses of some theorems to show that the results apply to even more general measures than probability measures.

Definition 1.7 (Finite Measure). Given a measurable space $(\Omega, \mathscr{F})$, a finite measure is a set function $\mu : \mathscr{F} \to [0,1]$ with the same countable additivity property as

defined above and the measure of the space finite instead of one. More specifically the first property above is replaced with:

$$\mu(\Omega) < \infty$$

Definition 1.8 (σ-finite Measure). A measure μ defined on a measurable space $(\Omega, \mathscr{F})$ is called σ-finite if it is countably additive and there exist a partition[1] of the space Ω, $\{\Omega_i\}_{i\in I}$, and $\mu(\Omega_i) < \infty$ for all $i \in I$. Note that the index set I is allowed to be countable.

Example 1.2 (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:

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

We can show that $(\Omega, \mathscr{F}, \mathbf{P})$ is a Probability Space. Indeed, from the definition:

$$\mathbf{P}(\Omega) = \frac{1}{N} \sum_{\omega\in\Omega} p(\omega) = \frac{1}{N} N = 1.$$

To show the countable additivity property let A a set in Ω such that $A = \bigcup_{i=1}^{\infty} A_i$, with A_i disjoint sets in Ω. Since the space is countable we may write $A_i = \{\omega_1^i, \omega_2^i, \ldots\}$, where any of the sets may be finite, but $\omega_j^i \neq \omega_l^k$ for all i, j, k, l where either $i \neq k$ or $j \neq l$. Then using the definition we have:

$$\begin{aligned}\mathbf{P}(A) &= \frac{1}{N} \sum_{\omega\in\bigcup_{i=1}^{\infty} A_i} p(\omega) = \frac{1}{N} \sum_{i\geq 1, j\geq 1} p(\omega_j^i) \\ &= \frac{1}{N} \sum_{i\geq 1} \left(p(\omega_1^i) + p(\omega_2^i) + \ldots \right) = \sum_{i\geq 1} \mathbf{P}(A_i)\end{aligned}$$

□

This is a very simple example but it shows the basic probability reasoning.

Remark 1.3. The previous exercise gives a way to construct discrete probability measures (distributions). For example take $\Omega = \mathbb{N}$ the natural numbers and take $N = 1$ in the definition of probability of an event. 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.
- $p(\omega) = \begin{cases} \binom{n}{\omega} p^{\omega}(1-p)^{n-\omega} & \text{, if } \omega \leq n \\ 0 & \text{, otherwise} \end{cases}$, gives the Binomial(n,p) distribution.

[1] a partition of the set A is a collection of sets A_i, disjoint ($A_i \cap A_j = \emptyset$, if $i \neq j$) such that $\cup_i A_i = A$

- $p(\omega) = \begin{cases} \binom{\omega-1}{r-1} p^r (1-p)^{\omega-r} & \text{, if } \omega \geq r \\ 0 & \text{, otherwise} \end{cases}$, gives the Negative Binomial(r,p) distribution.
- $p(\omega) = \frac{\lambda^\omega}{\omega!} e^{-\lambda}$, gives the Poisson (λ) distribution.

Example 1.3 (Uniform Distribution on (0,1)). As another example let $\Omega = (0,1)$ and $\mathscr{F} = \mathscr{B}((0,1))$ the Borel sigma algebra. Define a probability measure U as follows: for any open interval $(a,b) \subseteq (0,1)$ let $U((a,b)) = b-a$ the length of the interval. For any other open interval O define $U(O) = U(O \cap (0,1))$.

Note that we did not specify $U(A)$ for all Borel sets A, rather only for the generators of the Borel σ-field. This illustrates the probabilistic concept presented above. In our specific situation, under very mild conditions on the generators of the σ-algebra any probability measure defined only on the generators can be uniquely extended to a probability measure on the whole σ-algebra (Carathèodory extension theorem). In particular when the generators are open sets these conditions are true and we can restrict the definition to the open sets alone. This example is going to be extended in Section 1.5.

Proposition 1.3 (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 $\mathbf{P}(A) \leq \mathbf{P}(B)$*
2. *$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B), \forall A, B \in \mathscr{F}$*
3. *(General Inclusion-Exclusion formula, also named Poincaré formula):*

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

   Note that successive partial sums are alternating between over-and-under estimating.
4. *(Finite subadditivity, sometimes called Boole's inequality):*

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

1.1.1 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 1.9. A statement $\mathcal{S}$ about points $\omega \in \Omega$ is said to be true *almost surely* (a.s.), almost everywhere (a.e.) or with probability 1 (w.p.1) if the set M defined as:

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

is in $\mathscr{F}$ and $\mathbf{P}(M) = 1$, (or, equivalently M^c is a null set).

We will use the notions a.s., a.e., and w.p.1. to denote the same thing – the definition above. For example we will say $X \geq 0$ a.s. and mean: $\mathbf{P}\{\omega | X(\omega) \geq 0\} = 1$ or equivalently $\mathbf{P}\{\omega | X(\omega) < 0\} = 0$. The notion of almost sure is a fundamental one in probability. Unlike in deterministic cases where something has to always be true no matter what, in probability we care about "the majority of the truth". In other words probability recognizes that some phenomena may have extreme outcomes, but if they are extremely improbable then we do not care about them. Fundamentally, it is mathematics applied to reality.

Definition 1.10. We define the indicator function of an event A as the (simple) function $\mathbf{1}_A : \Omega \to \{0,1\}$,

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

Sometimes this function is denoted with I_A.

Note that the indicator function is a regular function (not a set function). Indicator functions are very useful in probability theory. Here are some useful relationships:

$$\mathbf{1}_{\mathbf{A}\cap\mathbf{B}}(\cdot) = 1_A(\cdot)1_B(\cdot)$$

If $\{B_i\}$ form a partition of Ω (i.e. the sets A_i are disjoint and $\Omega = \bigcup_{i=1}^n A_i$):

$$\mathbf{1}_{\mathbf{A}}(\cdot) = \sum_i 1_{A\cap B_i}(\cdot)$$

1.2 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:

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

We can immediately rewrite the formula above to obtain the *multiplicative rule*:

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

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:

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

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

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

Example 1.4. A biker leaves the point O in the figure below. At each crossroad the biker chooses a road at random. What is the probability that he arrives at point A ?

Let B_k, $k = 1,2,3,4$ be the event that the biker passes through point B_k. These four events are mutually exclusive and they form a partition of the space. Moreover, they are equiprobable ($\mathbf{P}(B_k) = 1/4, \forall k \in \{1,2,3,4\}$). Let A denote the event "the biker reaches the destination point A". Conditioned on each of the possible points B_1-B_4 of passing we have:

$$\begin{aligned}
&\mathbf{P}(A|B_1) = 1/4 \\
&\mathbf{P}(A|B_2) = 1/2 \\
&\mathbf{P}(A|B_3) = 1
\end{aligned}$$

At B_4 is slightly more complex. We have to use the multiplicative rule:

$$\begin{aligned}
\mathbf{P}(A|B_4) &= 1/4 + \mathbf{P}(A \cap B_5|B_4) + \mathbf{P}(A \cap B_6 \cap B_5|B_4) \\
&= 1/4 + \mathbf{P}(A|B_5 \cap B_4)\mathbf{P}(B_5|B_4) + \mathbf{P}(A|B_6 \cap B_5 \cap B_4)\mathbf{P}(B_6|B_5 \cap B_4)\mathbf{P}(B_5|B_4) \\
&= 1/4 + 1/3(1/4) + 1(1/3)(1/4) = 3/12 + 2/12 = 5/12
\end{aligned}$$

Finally, by the law of total probability:

$$\begin{aligned}
\mathbf{P}(A) &= \mathbf{P}(A|B_1)\mathbf{P}(B_1) + \mathbf{P}(A|B_2)\mathbf{P}(B_2) + \mathbf{P}(A|B_3)\mathbf{P}(B_3) + \mathbf{P}(A|B_4)\mathbf{P}(B_4) \\
&= 1/4(1/4) + 1/2(1/4) + 1/4(1) + 5/12(1/4) = 13/24
\end{aligned}$$

□

Example 1.5 (De Méré's Paradox). As a result of extensive observation of dice games the French gambler Chevaliér De Méré noticed that the total number of spots

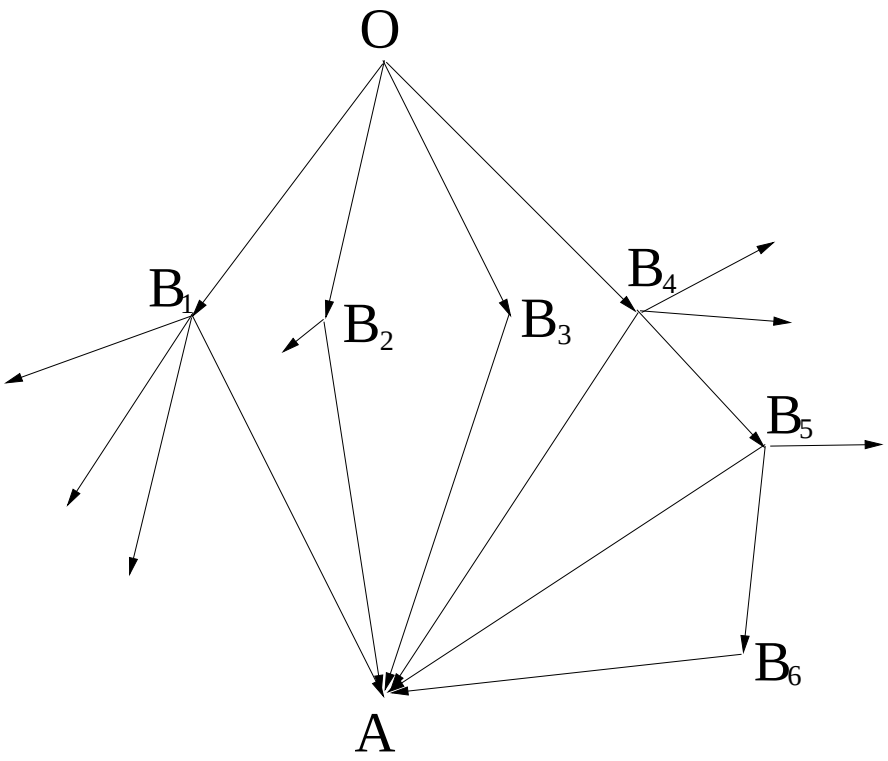

Fig. 1.1 The possible trajectories of the biker. O is the origin point and A is the arrival point. B_k's are intermediate points. Note that not all the ways lead to Rome, i.e. the probability of reaching Rome is less than 1.

showing on 3 dice thrown simultaneously turn out to be 11 more often than 12. However, from his point of view this is not possible since 11 occurs in six ways :

$(6:4:1);(6:3:2);(5:5:1);(5:4:2);(5:3:3);(4:4:3)$,

while 12 also in six ways:

$(6:5:1);(6:4:2);(6:3:3);(5:5:2);(5:4:3);(4:4:4)$

What is the fallacy in the argument?

Solution 1.1 (Solution due to Pascal). The argument would be correct if these "ways" would have the same probability. However this is not true. For example: (6:4:1) occurs in 3! ways, (5:5:1) occurs in 3 ways and (4:4:4) occurs in 1 way.

As a result we can easily calculate: $\mathbf{P}(11) = 27/216$; $\mathbf{P}(12) = 25/216$, and indeed his observation is correct and he should bet on 11 rather than on 12 if they have the same game payoff. □

Example 1.6 (Another De Mére's Paradox:). What is more probable?

1. Throw 4 dice and obtain at least one 6

2. Throw 2 dice 24 time and obtain at least once a double 6

Solution 1.2. For option 1: $1 - \mathbf{P}(\text{No 6}) = 1 - (5/6)^4 = 0.517747$.

For option 2: $1 - \mathbf{P}(\text{None of the 24 trials has a double 6}) = 1 - (35/36)^{24} = 0.491404$

Example 1.7 (Monty Hall problem). This is a problem named after the host of the American television show "Let's make a deal". Simply put at the end of a game you are left to chose between 3 closed doors. Two of them have nothing behind and one contains a prize. You chose one door but the door is not opened automatically. Instead, the presenter opens another door that contains nothing. He then gives you the choice of changing the door or sticking with the initial choice.

Most people would say that it does not matter what you do at this time, but that is not true. In fact everything depends on the host behavior. For example, if the host knows in advance where the prize is and always reveals at random some other door that does not contain anything then it is always better to switch.

Solution 1.3. This problem generated a lot of controversy since its publication (in 1970's) since the solution seems so counterintuitive. Articles talking about this problem in more detail Morgan et al. (1991), Mueser and Granberg (1991). We are presenting it here since it exemplifies the conditional probability reasoning. The key in any such problem is the sample space which has to be complete enough to be able to answer the questions asked.

Let D_i be the event that the price is behind door i. Let SW be the event that switching wins the price[2].

It does not matter which door we chose initially the reasoning is identical with all the three doors. So, we assume that initially we pick door 1.

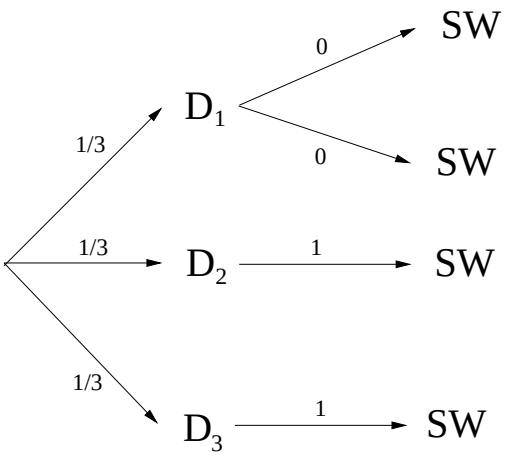

Fig. 1.2 The tree diagram of conditional probabilities. Note that the presenter has two choices in case D_1 neither of which results in winning if switching the door.

[2] As a side note this event is the same as the event "not switching loses"

Events D_i $i = 1,2,3$ are mutually exclusive and we can write:

$$\mathbf{P}(SW) = \mathbf{P}(SW|D_1)\mathbf{P}(D_1) + \mathbf{P}(SW|D_2)\mathbf{P}(D_2) + \mathbf{P}(SW|D_3)\mathbf{P}(D_3).$$

When the prize is behind door 1 since we chose door 1 the presenter has two choices for the door to show us. However, neither would contain the prize and in either case switching does not result in winning the prize, therefore $\mathbf{P}(SW|D_1) = 0$. If the car is behind door 2 since our choice is door 1 the presenter has no alternative but to show us the other door 3 which contains nothing. Thus switching in this case results in winning the price. The same reasoning works if the prize is behind door 3. Therefore:

$$\mathbf{P}(SW) = 1\frac{1}{3} + 1\frac{1}{3} + 0\frac{1}{3} = \frac{2}{3}$$

Thus switching has a higher probability of winning than not switching.

A generalization to n doors shows that it still is advantageous to switch but the advantage decreases as $n \to \infty$. Specifically, in this case $\mathbf{P}(D_i) = 1/n$; $\mathbf{P}(SW|D_1) = 0$ still, but $\mathbf{P}(SW|D_i) = 1/(n-2)$ if $i \neq 1$. Which gives:

$$\mathbf{P}(SW) = \sum_{i=2}^{n} \frac{1}{n}\frac{1}{n-2} = \frac{n-1}{n-2}\frac{1}{n} > \frac{1}{n}$$

Furthermore, different presenter strategies produce different answers. For example, if the presenter offers the option to switch only when the player chooses the right door then switching is always bad. If the presenter offers switching only when the player has chosen incorrectly then switching always wins. These and other cases can be analyzed in Rosenthal (2008).

Example 1.8 (Bertrand's box paradox). This problem was first formulated by Joseph Louis François Bertrand in his Calcul de Probabilités (Bertrand, 1889). In some sense this problem is related to the previous problem but it does not depend on any presenter strategy and the solution is much more clear. Solving this problem is an exercise in Bayes formula.

Suppose that we know that three boxes contain respectively: one box contains two gold coins, a second box with two silver coins, and a third box with one of each. We chose a box at random and from that box we chose a coin also at random. Then we look at the coin chosen. Given that the coin chosen was gold what is the probability that the other coin in the box chosen is also gold. At a first glance it may seem that this probability is $1/2$ but after calculation this probability turns out to be 2/3.

Solution 1.4. We plot the sample space in Figure 1.3. Using this tree we can calculate the probability:

$$\mathbf{P}(\text{Second coin is } G|\text{First coin is } G) = \frac{\mathbf{P}(\text{Second coin is } G \text{ and First coin is } G)}{\mathbf{P}(\text{First coin is } G)}.$$

Now, using the probabilities from the tree we continue:

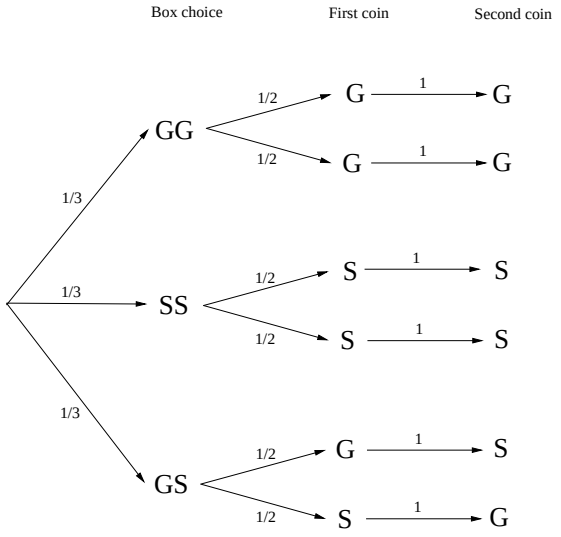

Fig. 1.3 The tree diagram of conditional probabilities.

$$= \frac{\frac{1}{3}\frac{1}{2}1 + \frac{1}{3}\frac{1}{2}1}{\frac{1}{3}\frac{1}{2}1 + \frac{1}{3}\frac{1}{2}1 + \frac{1}{3}\frac{1}{2}1} = \frac{2}{3}.$$

Now that we have seen the solution we can recognize a logical solution to the problem as well. Given that the coin seen is gold we can throw away the middle box. Then if this would be box 1 then we have two possibilities that the other coin is gold (depending on which we have chosen in the first place). If this is the box 2 then there is one possibility (the remaining coin is silver). Thus the probability should be 2/3 since we have two out of three chances. Of course this "logical" argument does not work if we do not choose the boxes with the same probability. □

Example 1.9. A blood test is 95% effective in detecting a certain disease when it is in fact present. However, the test yields also a false positive result for 1% of the people tested. If 0.5% of the population actually has the disease, what is the probability that the person is diseased given that the test is positive?

Solution 1.5. This problem illustrates once again the application of the Bayes rule. I do not like to use the rule literally instead work from first principles one will also obtain the Bayes rule without memorizing anything. We start by describing the sample space. Refer to the Figure 1.4 for this purpose.

So given that the test is positive means that we have to calculate a conditional probability. We may write:

$$\mathbf{P}(D|+) = \frac{\mathbf{P}(D \cap +)}{\mathbf{P}(+)} = \frac{\mathbf{P}(+|D)\mathbf{P}(D)}{\mathbf{P}(+)} = \frac{0.95(0.005)}{0.95(0.005) + 0.01(0.995)} = 0.323$$

How about if only 0.05% (i.e. 0.0005) of the population has the disease?

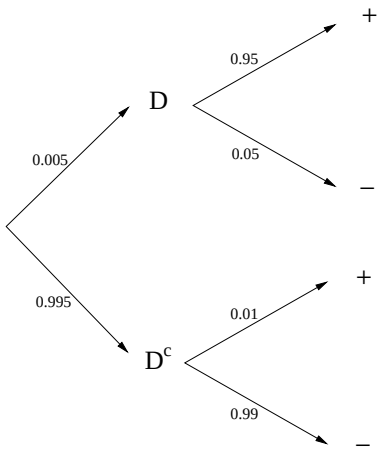

Fig. 1.4 Blood test probability diagram

$$\mathbf{P}(D|+) = \frac{0.95(0.0005)}{0.95(0.0005)+0.01(0.9995)} = 0.0454$$

This problem is an exercise in thinking. It is the same test device. In the first case the disease is relatively common and thus the test device is more or less reliable (though 32% right is very low). In the second case however the disease is very rare and thus the precision of the device goes way down. □

Example 1.10 (Gambler's Ruin Problem). We conclude this section with an example which we shall see many times throughout this book. I do not know who to credit with the invention of the problem since it is so mentioned so often in every probability treaties[3].

The formulation is simple. A game of heads or tails with a fair coin. Player wins 1 dollar if he successfully calls the side of the coin which lands upwards and loses \$1 otherwise. Suppose the initial capital is X dollars and he intends to play until he wins m dollars but no longer. What is the probability that the gambler will be ruined?

Solution 1.6. We will display what is called as a first step analysis.

Let $p(x)$ denote the probability that the player is going to be eventually ruined if he starts with x dollars.

If he wins the next game then he will have \$ $x+1$ and he is ruined from this position with prob $p(x+1)$.

If he loses the next game then he will have \$ $x-1$ so he is ruined from this position with prob $p(x-1)$.

Let R be the event he is eventually ruined. Let W be the event he wins the next trial. Let L be the event he loses this trial. Using the total prob. formula we get:

$$\mathbf{P}(R) = \mathbf{P}(R|W)\mathbf{P}(W) + \mathbf{P}(R|L)\mathbf{P}(L) \Rightarrow p(x) = p(x+1)(1/2) + p(x-1)(1/2)$$

[3] The formalization may be due to Huygens (1629-1695) in the XVII-th century

Is this true for all x? No. This is true for $x \geq 1$ and $x \leq w-1$. In the rest of cases we obviously have $p(0) = 1$ and $p(m) = 0$ which give the boundary conditions for the equation above.

This is a linear difference equation with constant coefficients. Please look at the general methodology in the following subsection on how to solve such equations.

Applying the method in our case gives the characteristic equation:

$$y = \frac{1}{2}y^2 + \frac{1}{2} \Rightarrow y^2 - 2y + 1 = 0 \Rightarrow (y-1)^2 = 0 \Rightarrow y_1 = y_2 = 1$$

In our case the two solutions are equal thus we seek a solution of the form $p(x) = (C + Dx)1^n = C + Dx$. Using the initial conditions we get: $p(0) = 1 \Rightarrow C = 1$ and $p(m) = 0 \Rightarrow C + Dm = 0 \Rightarrow D = -C/m = -1/m$, thus the general probability of ruin starting with wealth x is:

$$p(x) = 1 - x/m.$$

□

Solving difference equations with constant coefficients

This methodology is given for second order difference equations but higher order equations are solved in a very similar way. Suppose we are given an equation of the form:

$$a_n = Aa_{n-1} + Ba_{n-2},$$

with some boundary conditions.

The idea is to look for solutions of the form $a_n = cy^n$, with c some constant and y needs to be determined. Note that if we have two solutions of this form (say $c_1y_1^n$ and $c_2y_2^n$), then any linear combination of them is also a solution. We substitute this proposed form and obtain:

$$y^n = Ay^{n-1} + By^{n-2}.$$

Dividing by y^{n-2} we obtain the characteristic equation:

$$y^2 = Ay + B.$$

Next, we solve this equation and obtain real solutions y_1 and y_2 (if they exist). It may be possible that the characteristic equation does not have solutions in $\mathbb{R}$ in which case the difference equation does not have solutions either. Now we have two cases:

1. If y_1 and y_2 are distinct then the solution is $a_n = Cy_1^n + Dy_2^n$ where C, D are constants that are going to be determined from the initial conditions.

2. If $y_1 = y_2$ the solution is $a_n = Cy_1^n + Dny_1^n$. Again, C and D are determined from the initial conditions.

In the case when the difference equation contains p terms the procedure is identical even replicating the multiplicity issues. For more information one can consult any book on Ordinary Differential Equations such as Boyce and DiPrima (2004).

1.3 Independence

Definition 1.11. Two events A and B are called independent if and only if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$$

.

The events $A_1, A_2, A_3, \dots$ are called *mutually independent* (or sometimes simply independent) if for every subset J of $\{1, 2, 3, \dots\}$ we have:

$$\mathbf{P}\left(\bigcup_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$$

The events $A_1, A_2, A_3, \dots$ are called *pairwise independent* (sometimes jointly independent) if:

$$\mathbf{P}(A_i \cup A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j), \quad \forall i, j.$$

Note that jointly independent does not imply independence.

Two *sigma fields* $\mathscr{G}, \mathscr{H} \in \mathscr{F}$ *are* $\mathbf{P}$*–independent* if:

$$\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H), \quad \forall G \in \mathscr{G}, \forall H \in \mathscr{H}.$$

See Billingsley (1995) for the definition of independence of $k \geq 2$ sigma-algebras.

1.4 Monotone Convergence properties of probability

Let us take a step back for a minute and comment on what we have seen thus far. The σ-algebra differs from the regular algebra in that it allows us to deal with countable (not finite) number of sets. In fact this is a recurrent theme in probability, learning to deal with infinity. On finite spaces things are more or less simple. One has to define the probability of each individual outcome and everything proceeds from there. However, even in these simple cases imagine that one repeats an experiment over and over. Then again we are forced to cope with infinity. This section introduces a way to deal with this infinity problem.

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

Lemma 1.1. *The following are true:*

1. *If $A_n, A \in \mathscr{F}$ and $A_n \uparrow A$ (i.e., $A_1 \subseteq A_2 \subseteq \ldots A_n \subseteq \ldots$ and $A = \bigcup_{n\geq 1} A_n$), then: $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ as a sequence of numbers.*
2. *If $A_n, A \in \mathscr{F}$ and $A_n \downarrow A$ (i.e., $A_1 \supseteq A_2 \supseteq \ldots A_n \supseteq \ldots$ and $A = \bigcap_{n\geq 1} A_n$), then: $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ as a sequence of numbers.*
3. *(Countable subadditivity) If $A_1, A_2, \ldots,$ and $\bigcup_{i=1}^{\infty} A_n \in \mathscr{F}$, with A_i's not necessarily disjoint then:*

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

Proof. 1. Let $B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_n \setminus A_{n-1}$. Because the sequence is increasing we have that the B_i's are disjoint thus:

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

Thus using countable additivity:

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

2. Note that $A_n \downarrow A \Leftrightarrow {A_n}^c \uparrow A^c$ and from part 1 this means $1 - \mathbf{P}(A_n) \uparrow 1 - \mathbf{P}(A)$.

3. Let $B_1 = A_1, B_2 = A_1 \cup A_2, \ldots, B_n = A_1 \cup \cdots \cup A_n, \ldots$. From the finite subadditivity property in Proposition 1.3 we have that $\mathbf{P}(B_n) = \mathbf{P}(A_1 \cup \cdots \cup A_n) \leq \mathbf{P}(A_1) + \cdots + \mathbf{P}(A_n)$.

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

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

□

Lemma 1.2. *The union of a countable number of* **P***-null sets is a* **P***-null set*

This Lemma is a direct consequence of the countable subadditivity.

Recall from analysis: For a sequence of numbers $\{x_n\}_n$ lim sup and lim inf are defined:

$$\limsup x_n = \inf_m \{\sup_{n\geq m} x_n\} = \lim_{m\to\infty} (\sup_{n\geq m} x_n)$$
$$\liminf x_n = \sup_m \{\inf_{n\geq m} x_n\} = \lim_{m\to\infty} (\inf_{n\geq m} x_n),$$

and they represent the highest (respectively lowest) limiting point of a subsequence included in $\{x_n\}_n$.

Note that if z is a number such that $z > \limsup x_n$ then $x_n < z$ eventually[4].
Likewise, if $z < \limsup x_n$ then $x_n > z$ infinitely often[5].
These notions are translated to probability in the following way.

Definition 1.12. Let $A_1, A_2, \ldots$ be an infinite sequence of events, in some probability space $(\Omega, \mathscr{F}, \mathbf{P})$. We define the events:

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{m=n}^{\infty} A_m = \{\omega : \omega \in A_n \text{ for infinitely many n}\} = \{A_n \text{ i.o.}\}$$

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{m=n}^{\infty} A_m = \{\omega : \omega \in A_n \text{ for all n large enough}\} = \{A_n \text{ eventually}\}$$

Let us clarify the notions of "infinitely often" and "eventually" a bit more. We say that an outcome ω happens infinitely often for the sequence $A_1, A_2, \ldots, A_n, \ldots$ if ω is in the set $\bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m$. This means that for any n (no matter how big) there exist an $m \geq n$ and $\omega \in A_m$.

We say that an outcome ω happens eventually for the sequence $A_1, A_2, \ldots, A_n, \ldots$ if ω is in the set $\bigcup_{n=1}^{\infty} \bigcap_{m\geq n} A_m$. This means that there exist an n such that for any $m \geq n$, $\omega \in A_m$, so from this particular n and up ω is in all the sets.

Why so complicate definitions? The basic intuition is the following: say you roll a die infinitely many times, then it is obvious what it means for the outcome 1 to appear infinitely often. Also, we can say the average of the rolls will eventually be arbitrarily close to 3.5 (this will be shown later). It is not so clear cut in general. The framework above provides a generalization to these notions.

The Borel Cantelli lemmas

With this definitions we are now capable to give two important lemmas.

Lemma 1.3 (First Borel-Cantelli). *If $A_1, A_2, \ldots$ is any infinite sequence of events with the property $\sum_{n\geq 1} \mathbf{P}(A_n) < \infty$ then*

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m\right) = \mathbf{P}\left(A_n \textit{ events are true infinitely often}\right) = 0$$

This lemma essentially says that if the probabilities of events go to zero and the sum is convergent then necessarily A_n will stop occurring. However, the reverse of the statement is not true. To make it hold we need a very strong condition (independence).

[4] i.e., there is some n_0 very large so that $x_n < z$, for all $n \geq n_0$

[5] i.e., for any n there exists an $m \geq n$ such that $x_m > z$

Lemma 1.4 (Second Borel-Cantelli). *If $A_1, A_2, \ldots$ is an infinite sequence of **independent** events then:*

$$\sum_{n\geq 1} \mathbf{P}(A_n) = \infty \quad \Leftrightarrow \quad \mathbf{P}(A_n \text{ i.o.}) = 1.$$

Proof. First Borel-Cantelli.

$$\mathbf{P}(A_n \text{ i.o.}) = \mathbf{P}\left(\bigcap_{n\geq 1}\bigcup_{m=n}^{\infty} A_m\right) \leq \mathbf{P}\left(\bigcup_{n=m}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbf{P}(A_m), \forall n$$

where we used the definition and countable subadditivity. By the hypothesis the sum on the right is the tail end of a convergent series, therefore converges to zero as $n \to \infty$. Thus we are done. □

Proof. Second Borel-Cantelli:

"$\Rightarrow$" Clearly, showing that $\mathbf{P}(A_n \text{ i.o.}) = \mathbf{P}(\limsup A_n) = 1$ is the same as showing that $\mathbf{P}((\limsup A_n)^c) = 0$.

By the definition of lim sup and the DeMorgan's laws,

$$(\limsup A_n)^c = \left(\bigcap_{n\geq 1}\bigcup_{m=n}^{\infty} A_m\right)^c = \bigcup_{n\geq 1}\bigcap_{m=n}^{\infty} A_m^c.$$

Therefore, it is enough to show that $\mathbf{P}(\bigcap_{m=n}^{\infty} A_m^c) = 0$ for all n (recall that a countable union of null sets is a null set). However,

$$\begin{aligned}
\mathbf{P}\left(\bigcap_{m=n}^{\infty} A_m^c\right) &= \lim_{r\to\infty} \mathbf{P}\left(\bigcap_{m=n}^{r} A_m^c\right) = \underbrace{\lim_{r\to\infty} \prod_{m=n}^{\infty} \mathbf{P}(A_m^c)}_{\text{by independence}} \\
&= \lim_{r\to\infty} \prod_{m=n}^{r} (1 - \mathbf{P}(A_m)) \leq \underbrace{\lim_{r\to\infty} \prod_{m=n}^{r} e^{-\mathbf{P}(A_m)}}_{1-x\leq e^{-x} \text{ if } x\geq 0} \\
&= \lim_{r\to\infty} e^{-\sum_{m=n}^{r} \mathbf{P}(A_m)} = e^{-\sum_{m=n}^{\infty} \mathbf{P}(A_m)} = 0
\end{aligned}$$

The last equality follows since $\sum \mathbf{P}(A_n) = \infty$.
Note that we have used the following inequality: $1 - x \leq e^{-x}$ which is true if $x \in [0, \infty)$. One can prove this inequality with elementary analysis.

"$\Leftarrow$" This implication is the same as the first lemma. Indeed, assume by absurd that $\sum \mathbf{P}(A_n) < \infty$. By the First Borel-Cantelli Lemma this implies that $\mathbf{P}(A_n \text{ i.o.}) = 0$, a contradiction with the hypothesis. □

The Fatou lemmas

Again assume that $A_1, A_2, \ldots$ is a sequence of events.

Lemma 1.5 (Fatou lemma for sets). *Given any measure (not necessarily finite) μ we have:*

$$\mu(A_n \textit{ eventually}) = \mu(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mu(A_n)$$

Proof. Recall that $\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{m=n}^{\infty} A_m$, and denote this set with A. Let $B_n = \bigcap_{m=n}^{\infty} A_m$, which is an increasing sequence (less intersections as n increases) and $B_n \uparrow A =$. By the monotone convergence property of measure (Lemma 1.1) $\mu(B_n) \to \mu(A)$. However,

$$\mu(B_n) = \mu(\bigcap_{m=n}^{\infty} A_m) \leq \mu(A_m), \forall m \geq n,$$

thus $\mu(B_n) \leq \inf_{m\geq n} \mu(A_m)$. Therefore:

$$\mu(A) \leq lim_{n\to\infty} \inf_{m\geq n} \mu(A_m) = \liminf_{n\to\infty} \mu(A_n)$$

□

Lemma 1.6 (The reverse of the Fatou lemma). *If $\mathbf{P}$ is a finite measure (e.g., probability measure) then:*

$$\mathbf{P}(A_n \textit{ i.o.}) = \mathbf{P}(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} \mathbf{P}(A_n)$$

.

Proof. This proof is entirely similar. Recall that $\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{m=n}^{\infty} A_m$, and denote this set with A. Let $B_n = \bigcup_{m=n}^{\infty} A_m$. Then clearly B_n is a decreasing sequence and $B_n \downarrow A =$. By the monotone convergence property of measure (Lemma 1.1) and since the measure is finite $\mathbf{P}(B_1) < \infty$ so $\mathbf{P}(B_n) \to \mathbf{P}(A)$. However,

$$\mathbf{P}(B_n) = \mathbf{P}(\bigcup_{m=n}^{\infty} A_m) \geq \mathbf{P}(A_m), \forall m \geq n,$$

thus $\mathbf{P}(B_n) \geq \sup_{m\geq n} \mathbf{P}(A_m)$, again since the measure is finite . Therefore:

$$\mathbf{P}(A) \geq lim_{n\to\infty} \sup_{m\geq n} \mathbf{P}(A_m) = \limsup_{n\to\infty} \mathbf{P}(A_n)$$

□

Kolmogorov zero-one law

I like to present this theorem since it introduces the concept of *a sequence of* σ*-algebras*, a notion essential for stochastic processes.

For a sequence $A_1, A_2, \dots$ of events in the probability space $(\Omega, \mathscr{F}, \mathscr{P})$ consider the generated sigma algebras $\mathscr{T}_n = \sigma(A_n, A_{n+1}, \dots)$ and their intersection

$$\mathscr{T} = \bigcap_{n=1}^{\infty} \mathscr{T}_n = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \dots),$$

called the tail σ-field.

Theorem 1.1 (Kolmogorov's 0-1 Law). *If $A_1, A_2, \dots$ are independent then for every event A in the tail σ field ($A \in \mathscr{T}$) its probability $\mathbf{P}(A)$ is either* 0 *or* 1.

Proof. Skipped. The idea is to show that A is independent of itself thus $\mathbf{P}(A \cap A) = \mathbf{P}(A)\mathbf{P}(A) \Rightarrow \mathbf{P}(A) = \mathbf{P}(A)^2 \Rightarrow \mathbf{P}(A)$ is either 0 or 1. The steps of this proof are as follows:

1. First define $\mathscr{A}_n = \sigma(A_1, \dots, A_n)$ and show that is independent of $\mathscr{T}_{n+1}$ for all n.
2. Since $\mathscr{T} \subseteq \mathscr{T}_{n+1}$ and $\mathscr{A}_n$ is independent of $\mathscr{T}_{n+1}$, then $\mathscr{A}_n$ and $\mathscr{T}$ are independent for all n.
3. Define $\mathscr{A}_\infty = \sigma(A_1, A_2, \dots)$. Then from the previous step we deduce that $\mathscr{A}_\infty$ and $\mathscr{T}$ are independent.
4. Finally since $\mathscr{T} \subseteq \mathscr{A}_\infty$ by the previous step $\mathscr{T}$ is independent of itself and the result follows.

Note that $\limsup A_n$ and $\liminf A_n$ are tail events. However, it is only in the case when the original events are independent that we can apply Kolmogorov's theorem. Thus in that case $\mathbf{P}\{A_n \text{ i.o.}\}$ is either 0 or 1.

1.5 Lebesgue measure on the unit interval (0,1]

We conclude this chapter with the most important measure available. This is the unique measure that makes things behave in a normal way (e.g., the interval $(0.2, 0.5)$ has measure 0.3).

Let $\Omega = (0,1]$. Let $\mathscr{F}_0$=class of semiopen subintervals (a,b] of Ω. For an interval $I = (a,b] \in \mathscr{F}_0$ define $\lambda(I) = |I| = b - a$. Let $\varnothing \in \mathscr{F}_0$ the element of length 0. Let $\mathscr{B}_0$=the algebra of finite disjoint unions of intervals in (0,1]. Note that the problem 1.3 shows that this algebra is not a σ-algebra.

If $A = \sum_{i=1}^{n} I_n \in \mathscr{B}_0$ with I_n disjoint $\mathscr{F}_0$ sets; then

$$\lambda(A) = \sum_{i=1}^{n} \lambda(I_i) = \sum_{i=1}^{n} |I_i|$$

The goal is to show that λ is countably additive on the algebra $\mathscr{B}_0$. This will allow us to construct a measure (actually a prob. measure since we are working on (0,1]) using the next result (Caratheodory's theorem). The constructed measure is well defined and will be called the Lebesgue Measure.

Theorem 1.2 (Theorem for the length of intervals:). *Let $I = (a,b] \subseteq (0,1]$ and I_k of the form $(a_k,b_k]$ bounded but not necessarily in $(0,1]$.*

(i) If $\bigcup_k I_k \subseteq I$ and I_k are disjoint then $\sum_k |I_k| \leq |I|$
(ii) If $I \subseteq \bigcup_k I_k$ (with the I_k not necessarily disjoint) then $|I| \leq \sum_k |I_k|$.
(iii) If $I = \bigcup_k I_k$ and I_k disjoint then $|I| = \sum_k |I_k|$.

Proof. Exercise (*Hint:* use induction)

Note: Part (iii) shows that the function λ is well defined.

Theorem 1.3. *λ is a (countably additive) probability measure on the field $\mathscr{B}_0$. λ is called the Lebesgue measure restricted to the algebra $\mathscr{B}_0$*

Proof. Let $A = \bigcup_{k=1}^{\infty} A_k$, where A_k are disjoint $\mathscr{B}_0$ sets. By definition of $\mathscr{B}_0$,

$$A_k = \bigcup_{j=1}^{m_k} J_{k_j}, \quad A = \bigcup_{i=1}^{n} I_i,$$

where the J_{k_j} are disjoint. Then,

$$\lambda(A) = \sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} (\sum_{k=1}^{\infty} \sum_{j=1}^{m_k} |I_i \cap J_{k_j}|) = \sum_{k=1}^{\infty} \sum_{j=1}^{m_k} (\sum_{i=1}^{n} |I_i \cap J_{k_j}|)$$

and since $A \cap J_{k_j} = J_{k_j} \Rightarrow |A \cap J_{k_j}| = \sum_{i=1}^{n} |I_i \cap J_{k_j}| = |J_{k_j}|$, the above is continued:

$$= \sum_{k=1}^{\infty} \underbrace{\sum_{j=1}^{m_k} |J_{k_j}|}_{=|A_k|} = \sum_{k=1}^{\infty} \lambda(A_k)$$

□

The next theorem will extend the Lebesgue measure to the whole $(0,1]$, thus we define the probability space $((0,1], \mathscr{B}((0,1]), \lambda)$. The same construction with minor modifications works in $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ case.

Theorem 1.4 (Caratheodory's Extension Theorem). *A probability measure on an algebra has a unique extension to the generated σ-algebra.*

Note: The Caratheodory Theorem practically constructs all the interesting probability models. However, once we construct our models we have no further need of the theorem. It also reminds us of the central idea in the theory of probabilities: If one wants to prove something for a big set one needs to look first at the generators of that set.

Proof. (skipped), in the exercises.

Definition 1.13 (Monotone Class). A class $\mathscr{M}$ of subsets in Ω *is monotone* if it is closed under the formation of monotone unions and intersections, i.e.:

(i) $A_1, A_2, \cdots \in \mathscr{M}$ and $A_n \subset A_{n+1}, \bigcup_n A_n = A \Rightarrow A \in \mathscr{M}$
(ii) $A_1, A_2, \cdots \in \mathscr{M}$ and $A_n \supset A_{n+1} \Rightarrow \bigcap_n A_n \in \mathscr{M}$

The next theorem is only needed for the proof of the Caratheodory theorem. However, the proof is interesting and that is why is presented here.

Theorem 1.5. *If $\mathscr{F}_0$ is an algebra and $\mathscr{M}$ is a monotone class, then $\mathscr{F}_0 \subseteq \mathscr{M} \Rightarrow \sigma(\mathscr{F}_0) \subseteq \mathscr{M}$.*

Proof. Let $m(\mathscr{F}_0) =$ minimal monotone class over $\mathscr{F}_0 =$ the intersection of all monotone classes containing $\mathscr{F}_0$

We will prove that $\sigma(\mathscr{F}_0) \subseteq m(\mathscr{F}_0)$.

To show this it is enough to prove that $m(\mathscr{F}_0)$ is an algebra. Then exercise 1.11 will show that $m(\mathscr{F}_0)$ is a σ algebra. Since $\sigma(\mathscr{F}_0)$ is the smallest the conclusion follows.

To this end, let $\mathscr{G} = \{A : A^c \in m(\mathscr{F}_0)\}$.

(i) Since $m(\mathscr{F}_0)$ is a monotone class so is $\mathscr{G}$.
(ii) Since $\mathscr{F}_0$ is an algebra its elements are in $\mathscr{G} \Rightarrow \mathscr{F}_0 \subset \mathscr{G}$

(i) and (ii) $\Rightarrow m(\mathscr{F}_0) \subseteq \mathscr{G}$. Thus $m(\mathscr{F}_0)$ is closed under complementarity.

Now define $\mathscr{G}_1 = \{A : A \cup B \in m(\mathscr{F}_0), \forall B \in \mathscr{F}_0\}$.

We show that $\mathscr{G}_1$ is a monotone class:

Let $A_n \nearrow$ an increasing sequence of sets, $A_n \in \mathscr{G}_1$. By definition of $\mathscr{G}_1$, for all n $A_n \cup B \in m(\mathscr{F}_0), \forall B \in \mathscr{F}_0$.

But $A_n \cup B \supseteq A_{n-1} \cup B$ and thus the definition of $m(\mathscr{F}_0)$ implies:

$$\bigcup_n (A_n \cup B) \in m(\mathscr{F}_0), \forall B \in \mathscr{F}_0 \Rightarrow \left(\bigcup_n A_n\right) \cup B \in m(\mathscr{F}_0), \forall B,$$

and thus $\bigcup_n A_n \in \mathscr{G}_1$.

This shows that $\mathscr{G}_1$ is a monotone class. But since $\mathscr{F}_0$ is an algebra its elements (the contained sets) are in $\mathscr{G}_1$[6], thus $\mathscr{F}_0 \subset \mathscr{G}_1$. Since $m(\mathscr{F}_0)$ is the smallest monotone class containing $\mathscr{F}_0$ we immediately have $m(\mathscr{F}_0) \subseteq \mathscr{G}_1$.

Let $\mathscr{G}_2 = \{B : A \cup B \in m(\mathscr{F}_0), \forall A \in m(\mathscr{F}_0)\}$

$\mathscr{G}_2$ **is a monotone class.** (identical proof- see problem 1.10)

Let $B \in \mathscr{F}_0$. Since $m(\mathscr{F}_0) \subseteq \mathscr{G}_1$ for any set $A \in m(\mathscr{F}_0) \Rightarrow A \cup B \in m(\mathscr{F}_0)$. Thus, by the definition of $\mathscr{G}_2 \Rightarrow B \in \mathscr{G}_2 \Rightarrow \mathscr{F}_0 \subseteq \mathscr{G}_2$.

The previous implication and the fact that $\mathscr{G}_2$ is a monotone class implies that $m(\mathscr{F}_0) \subseteq \mathscr{G}_2$.

Therefore, $\forall A, B \in m(\mathscr{F}_0) \Rightarrow A \cup B \in m(\mathscr{F}_0) \Rightarrow m(\mathscr{F}_0)$ is an algebra. □

[6] one can just verify the definition of $\mathscr{G}_1$ for this.

Problems

1.1. 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.

1.2. Suppose two events A and B are in some space Ω. List the elements of the generated σ algebra $\sigma(A,B)$ in the following cases:
a) $A \cap B = \emptyset$
b) $A \subset B$
c) $A \cap B \neq \emptyset$; $A \setminus B \neq \emptyset$ and $B \setminus A \neq \emptyset$

1.3. 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 $m \in \mathbb{N}^* = \{1,2\ldots\}$ 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.

1.4. Let $\mathscr{F} = \{A \subseteq \Omega | A$ finite **or** A^c 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

1.5. A σ-Algebra does not necessarily contain all the events in Ω
Let $\mathscr{F} = \{A \subseteq \Omega | A$ countable **or** A^c 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 $A \notin \mathscr{F}$.

1.6. Show that the Borel sets of $\mathbb{R}$ $\mathscr{B} = \sigma(\{(-\infty, x] | x \in \mathbb{R}\})$.

Hint: show that the generating set is the same i.e., show that any set of the form $(-\infty, x]$ can be written as countable union (or intersection) of open intervals and viceversa that any open interval in $\mathbb{R}$ can be written as countable union (or intersection) of sets of the form $(-\infty, x]$.

1.7. Show that the following classes all generate the Borel σ-algebra, or put differently show the equality of the following collections of sets:

$$\begin{aligned}\sigma((a,b) : a < b \in \mathbb{R}) = \sigma([a,b] : a < b \in \mathbb{R}) = \sigma((-\infty, b) : b \in \mathbb{R}) \\ = \sigma((-\infty, b) : b \in \mathbb{Q}),\end{aligned}$$

where $\mathbb{Q}$ is the set of rational numbers.

1.8. Properties of probability measures
Prove properties 1-4 in the Proposition 1.3 on page 13.

Hint: You only have to use the definition of probability. The only thing non-trivial in the definition is the countable additivity property.

1.9. No mater how many zeros do not add to more than zero
Prove the Lemma 1.2 on page 23.
Hint: You may use countable subadditivity.

1.10. If $\mathscr{F}_0$ is an algebra, $m(\mathscr{F}_0)$ is the minimal monotone class over $\mathscr{F}_0$ and $\mathscr{G}_2$ is defined as:

$$\mathscr{G}_2 = \{B : A \cup B \in m(\mathscr{F}_0), \forall A \in m(\mathscr{F}_0)\}$$

Then show that $\mathscr{G}_2$ is a monotone class.
Hint: Look at the proof of theorem 1.5 on page 29, and repeat the arguments therein.

1.11. A monotone algebra is a σ-algebra
Let $\mathscr{F}$ be an algebra that is also a monotone class. Show that $\mathscr{F}$ is a σ-algebra.

1.12. Prove the *total probability formula* equation (1.6) and the *Bayes Formula* equation 1.7.

1.13. If two events are such $A \cap B = \emptyset$ are A and B independent? Justify.

1.14. Show that $\mathbf{P}(A|B) = \mathbf{P}(A)$ is the same as independence of the events A and B.

1.15. Prove that if two events A and B are independent then so are their complements.

1.16. Generalize the previous problem to n sets using induction.

1.17. One urn contains w_1 white balls and b_1 black balls. Another urn contains w_2 white balls and b_2 black balls. A ball is drawn at random from each urn, then one of the two such chose are selected at random.
a) What is the probability that the final ball selected is white?
b) Given that the final ball selected was white what is the probability that in fact it came from the first urn (with w_1 and b_1 balls).

1.18. At the end of a well known course the final grade is decided with the help of an oral examination. There are a total of m possible subjects listed on some pieces of paper. Of them n are generally considered "easy".

Each student enrolled in the class, one after another, draws a subject at random then presents it. Of the first two students who has the better chance of drawing a "favorable" subject?

1.19. Suppose an event A has probability 0.3. How many independent trials must be performed to assert with probability 0.9 that the relative frequency of A differs from 0.3 by no more than 0.1.

1.20. Show using the Cantelli lemma that when you roll a die the outcome $\{1\}$ will appear infinitely often. Also show that eventually the average of all rolls up to roll n will be within ε of 3.5 where $\varepsilon > 0$ is any arbitrary real number.

1.21. Andre Agassi and Pete Sampras decide to play a number of games together. They play non-stop and at the end it turns out that Sampras won n games while Agassi m where $n > m$. Assume that in fact any possible sequence of games was possible to reach this result. Let $P_{n,m}$ denote the probability that from the first game until the last Sampras is always in the lead. Find:

1. $P_{2,1}$; $P_{3,1}$; $P_{n,1}$
2. $P_{3,2}$; $P_{4,2}$; $P_{n,2}$
3. $P_{4,3}$; $P_{5,3}$; $P_{5,4}$
4. Make a conjecture about a formula for $P_{n,m}$.

1.22. My friend Andrei has designed a system to win at the roulette. He likes to bet on red, but he waits until there have been 6 previous black spins and only then he bets on red. He reasons that the chance of winning is quite large since the probability of 7 consecutive back spins is quite small. What do you think of his system. Calculate the probability the he wins using this strategy.

Actually, Andrei plays his strategy 4 times and he actually wins three times out of the 4 he played. Calculate what was the probability of the event that just occurred.

1.23. Ali Baba is caught by the sultan while stealing his daughter. The sultan is being gentle with him and he offers Ali Baba a chance to regain his liberty.
There are 2 urns and m white balls and n black balls. Ali Baba has to put the balls in the 2 urns however he likes with the only condition that no urn is empty. After that the sultan will chose an urn at random then pick a ball from that urn. If the chosen ball is white Ali Baba is free to go, otherwise Ali Baba's head will be at the same level as his legs.
How should Ali Baba divide the balls to maximize his chance of survival?

References

Bertrand, J. L. F. (1889). *Calcul des probabilités*. Paris: Gauthier-Villars et fils.

Billingsley, P. (1995). *Probability and measure* (3 ed.). Wiley.

Blæsild, P. and J. Granfeldt (2002). *Statistics with Applications in Biology and Geology*. CRC Press.

Boyce, W. E. and R. C. DiPrima (2004). *Elementary Differential Equations and Boundary Value Problems* (8 ed.). Wiley.

Cauchy, A. L. (1821). *Analyse algébrique*. Imprimerie Royale.

Chung, K. L. (2000). *A Course in Probability Theory Revised* (2nd ed.). Academic Press.

Dembo, A. (2008). Lecture notes in probability. available on http://www-stat.stanford.edu/∼adembo/.

Good, I. J. (1986). Some statistical applications of poisson's work. *Statistical Science 1*(2), 157–170.

Gross, D. and C. M. Harris (1998). *Fundamentals of Queueing Theory*. Wiley.

Jona-Lasinio, G. (1985). *Some recent applications of stochastic processes in quantum mechanics*, Volume 1159 of *Lecture Notes in Mathematics*, pp. 130–241. Springer Berlin / Heidelberg.

Karlin, S. and H. M. Taylor (1975). *A first course in stochastic processes* (2 ed.). Academic Press.

Kingman, J. F. C. (1993). *Poisson processes*. Oxford University Press.

Lu, T.-C., Y.-S. Hou, and R.-J. Chen (1996). A parallel poisson generator using parallel prefix. *Computers & Mathematics with Applications 31*(3), 33 – 42.

Morgan, J. P., N. R. Chaganty, R. C. Dahiya, and M. J. Doviak (1991). Let's make a deal: The player's dilemma. *American Statistician 45*, 284–287.

Mueser, P. R. and D. Granberg (1991). The monty hall dilemma revisited: Understanding the interaction of problem definition and decision making. working paper 99-06, University of Missouri.

Øksendal, B. (2003). *Stochastic Differential Equations* (5 ed.). Springer Verlag.

Rosenthal, J. (2008, September). Monty hall, monty fall, monty crawl. *Math Horizons*, 5–7.

Ross, S. (1995). *Stochastic Processes* (2nd ed.). Wiley.