

Lectures 1–2. Axioms of probability. Facts belonging to the “set-theoretic introduction” to both probability theory and measure theory.

It seems clear that probability theory must have to do, first of all, with random events and their probabilities. Axioms of probability are taught in every middle-level course of probability theory. Let me write these axioms.

First of all, at the beginning of every problem in probability theory is a *sample space*: speaking formally, an arbitrary set with elements of arbitrary nature. We’ll be using the letter Ω (capital “omega”) to denote it (in some books, different letters are used for the sample space, in particular, S).

The elements $\omega \in \Omega$ are called *sample points*.

Of course, we can study probability theory only in its purely mathematical aspect, without caring about possible applications. But if we do care about applications, at least in principle, we better say how is the sample space chosen in these possible applications.

First of all, probability theory is applied, in principle, to situations where there is an *experiment*, which can be repeated, under the same conditions, many times, but its outcome is not completely determined by these conditions: the outcome of the experiment may be different at its different repetitions.

As a *sample point* ω we choose something that describes the outcome of our experiment: I can say, a *complete account* of how this experiment went. As the sample space Ω we take the set of *all possible outcomes* of our experiments, the set of all possible such complete accounts.

An *event* (the epithet *random* is usually dropped: it goes without saying that the events that we consider in probability theory are random ones) is a subset of the sample space: say, $A \subseteq \Omega$. There are two signs to use to express the fact that one set is a part of another: \subset and \subseteq (similar to the signs for inequalities: $<$ and \leq). The first one is used when we want to stress that the case of the two sets being in fact the same is excluded; while the second sign means that the equality is allowed. So, $A \subseteq B$ means that A is a part of B , including the case of A being the same as B : just that *every element of A is an element of B* ; while $A \subset B$ means that every element of A is an element of B , *and A cannot be the same as B : there exists at least one element of B that does not belong to A* .

In particular, we can consider Ω itself as an event (as its own subset).

It is supposed that to every event A is associated a number $P(A)$ called *the probability of A* .

So probability is a number-valued function of events.

This function must satisfy certain axioms.

This axioms can be formulated differently; I am choosing the following formulation:

Axiom 1. $0 \leq P(A) \leq 1$, $P(\Omega) = 1$.

Axiom 2. If we have two disjoint events A and B (*disjoint* means that A and B have no common points: $A \cap B = \emptyset$), then $P(A \cup B) = P(A) + P(B)$.

Axiom 3. Suppose we have an infinite sequence of disjoint events: $A_1, A_2, \dots, A_n, \dots$ (the word *disjoint*, applied to more than two sets means that no two of them have common elements: $A_i \cap A_j = \emptyset$ for $i \neq j$). Then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

My axioms are not independent from one another: so, **Axiom 2** can be deduced from **3** if we take $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \dots = \emptyset$ (it can be proved very easily from **Axiom 3** that $P(\emptyset) = 0$). But I wanted to formulate the simple thing with two events first.

Example 1–2.1: Ω is a countable (i. e., finite or countably infinite: such that its elements can be numbered with integers) space, $p(\omega)$, $\omega \in \Omega$, are nonnegative numbers that add up to 1: $p(\omega) \geq 0$, $\sum_{\omega} p(\omega) = 1$. (If a sum, an integral, or a product is used without any mention of the set over which the index of summation or multiplication (the integration variable) changes, it will mean that *all possible* values are taken.) We take $\mathcal{F} = \mathcal{P}(\Omega)$ (the class of all subsets of Ω),

$$P(A) = \sum_{\omega \in A} p(\omega). \quad (1-2.1)$$

Then P is a probability, i. e. satisfies **Axioms 1–3**.

That **Axioms 2, 3** are satisfied follows from the fact that a finite sum or the sum of an infinite series with all terms of the same sign (all nonnegative, in our case) does not depend on the order of summation.

Example 1–2.2: Ω is the square $[0, 1] \times [0, 1] = [0, 1]^2 = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ (make a picture of the square, and some subset in it; $P(A)$, the area of the set A).

The first axiom is clearly satisfied: the area of the square is equal to 1, and the area of any subset of Ω is between 0 and the area of the square. The second axiom: this is that the area of the union of two or several non-overlapping figures is equal to the sum of the areas of these figures. But this is the main tool that the ancient Greeks (Euclid) used to find areas of figures in the plane: say, they represented a pentagon as the union of non-overlapping triangles.

As for **Axiom 3**, Euclid probably did not know about it; but Archimedes did know, and used it to find, say, the area of the figure between a parabola and a straight line.

So, based on what the ancient Greeks knew, we see that all probability axioms are satisfied in this case.

However the state of mathematics at those times is not what we expect to have in our days; and so the concept of area (also of volume, and of length, in the one-dimensional case) came under scrutiny around the year 1900, mainly in the work of Henri Lebesgue and Emile Borel. They discovered that the area cannot be defined for *all* subsets of the square: there are sets that have an area (*measurable* sets), and those that have no area (*non-measurable*). A new branch of mathematics, called measure theory, emerged. It turned out that area, volume of three-dimensional sets, length of sets in the real line, are particular cases of what we call now the *Lebesgue measure*, in different numbers of dimensions; and this Lebesgue measure does satisfy **Axioms 2 and 3** (clearly, it satisfies **Axiom 1** if the Lebesgue measure of the largest set we are considering is equal to 1 – which is the case for $\Omega = [0, 1]^2$).

So, what sets (subsets of our square) have no area (no Lebesgue measure)? It turned out that we cannot really construct an example of such a set: every set that we are able to describe turns out to be measurable. But non-measurable sets do exist, only the proof

of this fact (which is not very simple, by the way) is not of the kind that allows us to construct an example of such a set. (A similar situation: 1000 people enter a room where we have 100 coins, and some of them take coins; when they go out, we may be sure that there are people without coins – because $100 < 1000$, but we cannot point at a concrete person without a coin.)

Measure theory studies not only the Lebesgue measure in different numbers of dimensions, but also other mathematical objects called *measures*.

Let a set X be given, and a class \mathcal{X} of its subsets (that is, a set \mathcal{X} of its subsets; only saying “a set consisting of sets” we might get our tongue tangled – so we are using the words *class* of sets; not to be confused in what is just a set and what a set – or class – of such, we will be using *script* letters $\mathcal{A}, \mathcal{B}, \mathcal{C}$, etc. for classes of sets, and just the usual capital letters for all other sets). We say that a function $m: \mathcal{X} \mapsto [0, \infty]$ (i. e., a function on \mathcal{X} taking nonnegative real values and, possibly, the value $+\infty$) is a *measure* if

$$\emptyset \in \mathcal{X}, \quad m(\emptyset) = 0; \tag{1-2.2}$$

for disjoint $A_1, A_2, \dots, A_n, \dots$ such that all $A_i \in \mathcal{X}$, and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{X}$, have

$$m\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(A_i) \tag{1-2.3}$$

(we can add up an infinite series of nonnegative numbers, and the series can either converge, with its sum a number, or diverge, in which case we take ∞ as its sum; and if one of $m(A_i)$ is equal to ∞ , we take the sum $\sum_{i=1}^{\infty} m(A_i)$ to be equal to ∞).

In Kolmogorov and Fomin’s book “Introductory Real Analysis”, the terminology is different: a set function m is called a measure if for finitely many disjoint A_i ’s we have $m(\bigcup_i A_i) = \sum_i m(A_i)$; and what is called a measure here, is called a *countably additive* measure.

In the 1920s and 1930s it became clear to some specialists in probability theory that this theory should have measure theory as its foundation. In the 1930s Kolmogorov formulated the axioms of probability, based on the concept of measure.

So measure theory lies in the foundations of modern probability theory.

We tell our students that *an event is a subset of the sample space Ω* , and we don’t tell them anything that is wrong; but we do not tell the whole truth: in our theory, every event is a subset of Ω , but not every subset is necessarily an event.

So in probability theory we should say something about the class \mathcal{F} of subsets of Ω whose elements are proclaimed *events*. This is done again in the language of measure theory.

To be precise, not exactly in the language of measure theory, because it is not about the measure, but rather about sets and classes thereof. So I would formulate it like this: in the language of the *set-theoretic introduction* to measure theory. Or, still better: *the set-theoretic introduction to both measure theory and probability theory* – because it is common to both of them.

Let \mathcal{A} be a class of subsets of a set X . This class is called an *algebra* of subsets of X (or: an algebra in X) if

1) $X \in \mathcal{A}$;

2) $A \in \mathcal{A} \Rightarrow$ its complement $A^c (= X \setminus A)$ also belongs to \mathcal{A} (... \Rightarrow ... is the notation for “from ... follows ...”).

3) $A_1, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$.

Example: Finite unions of intervals (finite or infinite ones) form an algebra in the real line \mathbb{R}^1 .

A class \mathcal{A} of subsets of X is a σ -*algebra* if **1)**, **2)** are satisfied, and

3 σ) $A_1, \dots, A_n, \dots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

(The letter σ is used in terms having to do with countable unions.)

It is easy to see that every σ -algebra is an algebra: but not every algebra is a σ -algebra (e. g., not that of our example above).

The axioms of probability theory can be formulated like this: there is a σ -algebra \mathcal{F} of subsets in the sample space Ω , whose elements are proclaimed *events*; and probability P is a measure on the σ -algebra \mathcal{F} whose value at Ω is equal to 1.

A triple (Ω, \mathcal{F}, P) of an arbitrary set Ω consisting of elements ω of an arbitrary nature, a σ -algebra \mathcal{F} of its subsets, and a measure P on \mathcal{F} such that $P(\Omega) = 1$ is called a *probability space*.

We can develop probability theory without any regard to its possible applications; but it is not usually done so. I have already spoken about how the sample space Ω is chosen. The σ -algebra \mathcal{F} is chosen usually in a standard way: if Ω is countable, we take $\mathcal{F} = \mathcal{P}(\Omega)$, the class of *all* subsets of Ω ; and if it is uncountable, say, a region on a multidimensional space, the choice of \mathcal{F} is also a standard one – but what this standard is I’ll tell you later. The main thing is: how do we choose, in applications, the *probability* P ; and if this P has been chosen, what practical (extra-mathematical) implications does this have?

There are some principles of application of probability theory.

One of them is as follows: *If the probability $P(A)$ is small enough, we can act as if the event A were impossible.*

In fact, all of Inferential Statistics is based on this principle.

This principle is formulated not in our usual precise and frank mathematical way, but somewhat vaguely. But mathematical precision is in its place in mathematics, and is quite out of place here, which is the borderline between mathematics and the outside world.

What does it mean “small enough”? Is it “not greater than 0.05”? or “ ≤ 0.0001 ”?

But this is not a mathematical question: it is a practical one, and the answer to it depends on what the consequences assuming that the event A will not occur and being wrong are. If the only consequence is that I’ll be late by five minutes to the beginning of a lecture, probably I would risk the bound of 5%; but perhaps I would refrain from boarding a plane that will crash with probability 0.05.

There is exactly one nonnegative number that is *small enough* for all practical purposes we can think of: namely, 0. So if $P(A) = 0$, in all applications of probability theory we can assume that the event A will not happen.

This is why probability theory pays attention to events with probability equal to 0: to systematically disregard them.

This is its common trait with *measure theory*; but the terms used for this in these branches of mathematics are a little different.

In measure theory, if some property is satisfied except for elements x of a set A with $m(A) = 0$, we say that this property is satisfied *almost everywhere* (with respect to the measure m).

In probability theory, if some property is satisfied for sample points ω except for $\omega \in A$, and $P(A) = 0$, we say that this property is satisfied *almost surely*.

I should have mentioned this in the lecture – but I am including it in the lecture note anyway: Since probability theory originated in the seventeenth century, much earlier than measure theory or set theory, there is a traditional probabilistic terminology and ways of expression that do not mention any *sets*. This traditional language of probability theory is rather convenient, and appealing in some degree to our intuition; so it is still in use. I'll give some expressions in this traditional language, and their translations in the language of sets (I would say “set-theoretic language”, only you don't need any *theory* to speak in terms of *sets*).

Instead of saying: “the sample point ω under consideration belongs to the event A ”: $\omega \in A$, we use any of the following synonymous expressions: *A has occurred*, or *occurs*, or *will occur*, or *happened*, etc.

The *event sure to occur* is such an event that always occurs, i. e. such that always $\omega \in it$ – for every sample point $\omega \in \Omega$. In other words, this is the event Ω (we remember that any set can be considered as its own subset).

The *impossible event* is such that never happens, i. e., $\omega \in it$ for no $\omega \in \Omega$; – that is, the empty set \emptyset .

Mutually exclusive events A_1, \dots, A_n, \dots are such that cannot happen together, i. e., it cannot be that $\omega \in A_i, \omega \in A_j$ for the same ω if $i \neq j$. In the language of sets it means that each pair of the sets has an empty intersection; – that is, the events are *disjoint*.

The *opposite event* to an event A is such that occurs if and only if A does not take place. It is another verbal expression for the complement A^c of the set (event) A .

“If the event A occurs, then B occurs too” means, in the language and the notations of sets, that $A \subseteq B$.

The words: *the event consisting in that ...*, or simply *the event that ...* mean: the set of sample points ω such that ...; for this set, as we know, the following notation is used:

$$\{\omega: \dots\}. \tag{1-2.4}$$

For example, you can read such an expression in some books on probability theory: *the event consisting in that the event A or B will happen*. To the language of sets and set-theoretic notations this is translated so: $\{\omega: \omega \in A \text{ or } \omega \in B\}$. This is simply, in the set-theoretic notation, $A \cup B$.

Further: *the probability that ...* means: the probability of the event (consisting in) that ...; which is denoted as

$$P(\{\omega: \dots\}). \quad (1-2.5)$$

Since we know that only events, i.e., sets consisting of sample points ω can stand under the probability sign, – the mention of ω here is usually dropped out. Also for shortness two kinds of parantheses inside each other is not written. I leave the more characteristic braces:

$$P\{\dots\}. \quad (1-2.6)$$

Instead of “ $P(A) =$ say, 0.82” such expressions are used: *A has probability 0.82*, or *the event A occurs with probability 0.82*. The expression: *such and such thing with probability so-and-so* means:

$$P(\{\omega: \text{such and such thing}\}) = \text{so-and-so}, \quad (1-2.7)$$

or

$$P\{\text{such and such thing}\} = \text{so-and-so}. \quad (1-2.8)$$

End of the piece not mentioned in the lecture.

I’ll mention very little about *probability* before getting to *random variables*.

The first thing I want to prove here is that the probability of the limit of a sequence of events is equal to the limit of their probabilities.

Except that we don’t know what a limit of a sequence of sets is.

The limit of a sequence of numbers is not defined always; so it is not surprising that the limit of a sequence of sets also will be defined only in some cases. I’ll define it in the following two cases.

If $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots \subseteq B_n \subseteq \dots$ is a non-decreasing sequence of sets, I take

$$\lim_{n \rightarrow \infty} B_n = \bigcup_{i=1}^{\infty} B_i \quad (1-2.9)$$

(please draw a picture illustrating this); and if $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots \supseteq B_n \supseteq \dots$ is a non-increasing sequence of sets, I take

$$\lim_{n \rightarrow \infty} B_n = \bigcap_{i=1}^{\infty} B_i. \quad (1-2.10)$$

Theorem 1 – 2.1. *For both a non-decreasing sequence of events B_n and for a non-increasing,*

$$P\left(\lim_{n \rightarrow \infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n). \quad (1-2.11)$$

Proof. It is clear that we need to prove the theorem only for, say, non-decreasing sequences (if it is proved, the non-increasing ones are handled by introducing $C_i = B_i^c$: *this* sequence is non-decreasing).

We have:

$$\lim_{n \rightarrow \infty} B_n = B_1 \cup (B_2 \setminus B_1) \cup (B_3 \setminus B_2) \cup \dots \cup (B_n \setminus B_{n-1}) \cup \dots ; \quad (1-2.12)$$

these sets are mutually disjoint (make a picture). So by Axiom 3

$$P\left(\lim_{n \rightarrow \infty} B_n\right) = P(B_1) + P(B_2 \setminus B_1) + P(B_3 \setminus B_2) + \dots + P(B_n \setminus B_{n-1}) + \dots \quad (1-2.13)$$

(the sum of an infinite series). By definition it is equal to

$$\lim_{n \rightarrow \infty} [P(B_1) + P(B_2 \setminus B_1) + P(B_3 \setminus B_2) + \dots + P(B_n \setminus B_{n-1})]. \quad (1-2.14)$$

Again we use the fact that the events $B_1, B_2 \setminus B_1, \dots, B_n \setminus B_{n-1}$ are disjoint, and their union is B_n ; so by Axiom 2 (we even don't need countable additivity, Axiom 3, here)

$$P(B_1) + P(B_2 \setminus B_1) + P(B_3 \setminus B_2) + \dots + P(B_n \setminus B_{n-1}) = P(B_n), \quad (1-2.15)$$

and we come to formula (1-2.11).

We know that a function $f: \mathbb{R}^1 \mapsto \mathbb{R}^1$ is continuous if and only if it follows from $\lim_{n \rightarrow \infty} x_n = x_0$ that $f(\lim_{n \rightarrow \infty} x_n) = \lim_{n \rightarrow \infty} f(x_n)$. So Theorem 1-2.1 means, in fact, that the probability function P is *continuous*.

Theorem 1-2.1 is true also for arbitrary measures m if the m -measures of all sets considered are *finite*.

Theorem 1-2.2. *Let \mathcal{A} be an algebra of sets in a space X ; let m be a finite nonnegative finitely additive function on \mathcal{A} (i.e. $0 \leq m(A) < \infty$, and $m(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n m(A_i)$ for disjoint $A_1, A_2, \dots, A_n \in \mathcal{A}$ such that $\bigcup_{i=1}^n A_i \in \mathcal{A}$. Then the function m is countably additive (i.e. satisfies (1-2.3) for disjoint $A_1, A_2, \dots, A_n, \dots \in \mathcal{A}$ such that $\bigcup_{i=1}^{infy} A_i \in \mathcal{A}$) if and only if it is "continuous at zero":*

$$B_1 \supseteq B_2 \supseteq \dots \supseteq B_n \supseteq B_{n+1} \supseteq \dots, \bigcap_{n=1}^{\infty} B_n = \emptyset \Rightarrow \lim_{n \rightarrow \infty} m(B_n) = 0. \quad (1-2.16)$$

The **proof** was not given in the lecture, but here it is. That (1-2.16) follows from countable additivity is, in fact, Theorem 1-2.1. Let us prove that countable additivity follows from (1-2.16).

Let $A_1, A_2, \dots, A_n, \dots \in \mathcal{A}$, and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. We have to prove that $m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$.

We have for every natural n :

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots \cup A_n \cup \bigcup_{i=n+1}^{\infty} A_i, \quad (1-2.17)$$

where the “summands” are disjoint. We have:

$$\bigcup_{i=n+1}^{\infty} A_i = \left[A_1 \cup A_2 \cup \dots \cup A_n \cup \left(\bigcup_{i=1}^{\infty} A_i \right)^c \right]^c \quad (1-2.18)$$

(make a picture – or prove, without one, that a point $x \in X$ belongs to the set in the left-hand side if and only if it belongs to the right-hand side of (1-2.18)). Since A_1, A_2, \dots, A_n , and $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, and \mathcal{A} is an algebra (not necessarily a σ -algebra), we have also $\bigcup_{i=n+1}^{\infty} A_i \in \mathcal{A}$.

Let us take $B_n = \bigcup_{i=n}^{\infty} A_i$. Clearly $B_1, B_2, \dots, B_n, \dots$ is a non-increasing sequence; and we have $\bigcap_{n=1}^{\infty} B_n = \emptyset$ (no point $x \in X$ belongs to all of B_n – no matter whether it belongs to one – and only one! – of A_i , or not). So $\lim_{n \rightarrow \infty} m\left(\bigcup_{i=n+1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} m(B_{n+1}) = \lim_{n \rightarrow \infty} m(B_n) = 0$, and it follows from (1-2.16) that $m\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} (m(A_1) + \dots + m(A_n)) = \sum_{i=1}^{\infty} m(A_i)$.

Now back to probabilities (though the results that follow are true for all measures).

A couple of theorems are so obvious that I give them without proofs:

Theorem 1–2.3. $P(A^c) = 1 - P(A)$.

Theorem 1–2.4. If $A \subseteq B$ for events A and B , then $P(A) \leq P(B)$.

The next one is a little more complicated:

Theorem 1–2.5. For any events $A_1, A_2, \dots, A_n, \dots$ (disjoint or not) we have:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i). \quad (1-2.19)$$

Proof. Again we represent the non-disjoint union $\bigcup_{i=1}^{\infty} A_i$ as one of disjoint sets; but since A_i do not form a non-decreasing sequence as in Theorem 1–2.1, it will be a little more complicated:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \dots \cup (A_n \setminus (A_1 \cup \dots \cup A_{n-1})) \cup \dots \quad (1-2.20)$$

So

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P(A_1) + P(A_2 \setminus A_1) + P(A_3 \setminus (A_1 \cup A_2)) + \dots \\ &\quad + P(A_n \setminus (A_1 \cup \dots \cup A_{n-1})) + \dots \end{aligned} \quad (1-2.21)$$

By $A_i \setminus (A_1 \cup \dots \cup A_{i-1}) \subseteq A_i$ and Theorem 1–2.4, we have $P(A_i \setminus (A_1 \cup \dots \cup A_{i-1})) \leq P(A_i)$; from which (1-2.19) follows.

Theorem 1–2.6 (Borel–Cantelli’s Lemma). If $A_1, A_2, \dots, A_n, \dots$ are events such that $\sum_{i=1}^{\infty} P(A_i) < \infty$, then almost surely only finitely many of the events A_i occur.

The statement of the theorem requires a translation. According to our meaning of the expression “almost surely”, this statement means that

$$P\{\text{infinitely many of } A_i \text{ occur}\} = 0. \quad (1-2.22)$$

Again this requires a translation into the language of sets and operations on them. “Infinitely many” means that however large a natural number n we take, for at least one $i > n$ the event A_i occurs. So here is the translation:

$$\{\text{infinitely many of } A_i \text{ occur}\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n+1}^{\infty} A_i. \quad (1-2.23)$$

By Theorem 1-2.5 we have: $P(\bigcup_{i=n+1}^{\infty} A_i) \leq \sum_{i=n+1}^{\infty} P(A_i)$; the events $B_n = \bigcup_{i=n+1}^{\infty} A_i$ clearly form a non-increasing sequence, so we have:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n+1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{i=n+1}^{\infty} A_i\right) \leq \lim_{n \rightarrow \infty} \sum_{i=n+1}^{\infty} P(A_i). \quad (1-2.24)$$

But we know that for every convergent series its “tail” after the n -th summand goes to 0 as $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \sum_{i=n+1}^{\infty} P(A_i) = \lim_{n \rightarrow \infty} \left[\sum_{i=1}^{\infty} P(A_i) - \sum_{i=1}^n P(A_i) \right] = 0; \quad (1-2.25)$$

which proves (1-2.22).

Still we have a little to do before we start on random variables; but what follows is not probability or measure theory, but something in the common set-theoretic introduction to them.

Theorem 2.7. *For any class \mathcal{A} of subsets of a set X there exists the smallest σ -algebra \mathcal{C} in X that contains \mathcal{A} .*

That is:

- \mathcal{C} is a σ -algebra, $\mathcal{C} \supseteq \mathcal{A}$;
- for any σ -algebra $\mathcal{D} \supseteq \mathcal{A}$ we have:

$$\mathcal{D} \supseteq \mathcal{C}. \quad (1-2.26)$$

It is clear at once that if such a σ -algebra \mathcal{C} exists, it is *unique*; but how to prove its existence?

The simplest **proof** seems to be this: Let us consider all σ -algebras \mathcal{D} that contain the class \mathcal{A} . At least one such σ -algebra exists: namely, the σ -algebra $\mathcal{P}(X)$ of *all* subsets of X (clearly, it contains \mathcal{A} , and clearly it is a σ -algebra: all σ -algebra requirements are

of the form: (if this or that, then) a certain subset of X belongs to $\mathcal{P}(X)$; but *every* subset belongs to it). Then we take

$$\mathcal{C} = \bigcap_{\mathcal{D}: \mathcal{D} \text{ is a } \sigma\text{-algebra, and } \mathcal{D} \supseteq \mathcal{A}} \mathcal{D}. \quad (1-2.27)$$

This class of sets clearly contains \mathcal{A} (because all “intersecants” contain it); let us prove that it is a σ -algebra.

We have to prove:

$$X \in \mathcal{C}; \quad (1-2.28)$$

$$A \in \mathcal{C} \Rightarrow A^c \in \mathcal{C}; \quad (1-2.29)$$

$$A_1, A_2, \dots, A_n, \dots \in \mathcal{C} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{C}. \quad (1-2.30)$$

The first one is obvious: X belongs to each of intersecants \mathcal{D} , all of them being σ -algebras. Let us prove (1-2.29): Suppose a set A belongs to the class \mathcal{C} ; then A belongs to every class \mathcal{D} mentioned in (1-2.27). Since these classes are σ -algebras, we have also for each of them $A^c \in \mathcal{D}$; and a thing (A^c) belonging to each of the intersecants belongs to the intersection.

In the same way (1-2.30) is proved: $\bigcup_{i=1}^{\infty} A_i$ belongs to \mathcal{C} because it belongs to every of the intersecants \mathcal{D} – which is because \mathcal{D} are σ -algebras.

The smallest σ -algebra containing a class of sets \mathcal{A} is denoted $\sigma(\mathcal{A})$. And the following term is used for it: the σ -algebra *generated by* \mathcal{A} .

In absolutely the same way the following theorem is proved:

Theorem 1–2.8. *For any class \mathcal{A} of subsets of a set X there exists the smallest algebra \mathcal{C} in X that contains \mathcal{A} .*

The notation for this smallest *algebra* will be $\alpha(\mathcal{A})$.

While we can describe the algebra $\alpha(\mathcal{A})$ generated by a class of sets \mathcal{A} easily enough (e. g., for \mathcal{A} being the class of all intervals, finite or infinite, in the real line \mathbb{R}^1 the algebra $\alpha(\mathcal{A})$ is the class of all finite unions of intervals – this class was mentioned before), not so with the σ -algebra $\sigma(\mathcal{A})$. Say, for the same example of the class of all intervals, is $\sigma(\mathcal{A})$ the class of all *countable* unions of intervals? – No, e. g., it is easy to check that the set Ir of all irrational numbers belongs to the σ -algebra generated by all intervals, but it is not a countable union of intervals. – Is $\sigma(\mathcal{A})$ the class of all countable unions of intervals, and complements of such unions? – Again no: we can take irrational numbers in one half of the real line, and rational numbers in the other half. – Is it the class of all countable unions of intervals, and their complements, and countable intersections of such unions? – Again no – but it is a little harder to invent a counterexample. – Is it the class of all countable unions, and countable intersections thereof, and all countable unions of countable intersections of countable unions, and all countable intersections of countable unions of countable intersections of countable unions, *etc.*? – Still it is “no” – even if we are able to represent much more sets belonging to $\sigma(\mathcal{A})$ in this form. – We can go in this

direction specifying in a more precise and more elaborate way what “etc.” should mean – but we see it is pretty complicated.

So how can we work with σ -algebras generated by classes of sets? We can do so in some indirect ways.

But about this, after the hurricane.

I am including in this lecture note some problems for you to solve (they will be also in a special file for problems). Only I cannot work it out, what deadline I should put on these problems.

1 Specialists in probability theory have endeavored to apply it to number theory; e. g. Chebyshev.

For some subsets A of the set N of natural numbers what may be called *asymptotic density* exists:

$$\nu(A) = \lim_{n \rightarrow \infty} \frac{\text{number of elements in } A \cap \{1, 2, \dots, n\}}{n}. \quad (1-2.31)$$

For example, for the set D_3 of natural numbers divisible by 3 we have $\nu(D_3) = 1/3$; for the set Pr of all prime numbers $\nu(Pr) = 0$ (even if it is not so easy to prove as in the case of D_3).

Let \mathcal{N} be the class of sets for which the limit (1-2.31) exists.

Is the class \mathcal{N} a σ -algebra in N ?

Is it an algebra?

2 Is the set function ν of the previous problem finitely additive (i. e., does, for disjoint $A_1, \dots, A_n \in \mathcal{N}$ such that $\bigcup_{i=1}^n A_i \in \mathcal{N}$, the equality $\nu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \nu(A_i)$ necessarily hold?)?

3 Is the set function ν of problem **1** countably additive?

4 Can you produce an example of a probability space and a sequence of events A_i in it, such that $\sum_{i=1}^{\infty} P(A_i) = \infty$, $P\{\text{infinitely many of } A_i \text{ occur}\} = 1$?

5 Can you produce an example of a probability space and a sequence of events A_i in it, such that $\sum_{i=1}^{\infty} P(A_i) = \infty$, $P\{\text{infinitely many of } A_i \text{ occur}\} \in (0, 1)$?

6 Can you produce an example of a probability space and a sequence of events A_i in it, such that $\sum_{i=1}^{\infty} P(A_i) = \infty$, $P\{\text{infinitely many of } A_i \text{ occur}\} = 0$?