What is measure, what is probability.\(^1\)

The concept of measure does not appear in elementary calculus, but it is a fundamental and important concept. It is not very difficult to understand, since it is important. Besides, the introduction of the Lebesgue measure by Lebesgue is a good example of conceptual analysis, so let us look at its elementary part. A good introductory book for this topic is the already quoted Kolmogorov-Fomin. It is desirable that those who wish to study fundamental aspects of statistical mechanics and dynamical systems have proper understanding of the subject.

**What is the volume?**

For simplicity, let us confine ourselves to the two dimension. Thus, the question is: what is the area? Extension to higher dimensions should not be hard. If the shape of a figure is complicated, whether it has an area could be a problem,\(^2\) so let us begin with an apparently trivial case.

"The area of the rectangle \([0, a] \times [0, b]\) is \(ab\)."

Is this really so? If so, why is this true? Isn’t it strange that we can ask such a question before defining ‘area’? Then, if we wish to be logically conscientious, we must accept the following definition:

**Definition.** The area of a figure congruent to the rectangle \((0, a) \times (0, b)\) (here, ‘\('\)’ implies ‘[’ or ‘(’, ‘)’ is ‘]’ or ‘)’), that is, we do not care whether the boundary is included or not) is defined as \(ab\).\(^3\)

Notice that the area of an rectangle does not depend on whether its boundary is included or not. This is already included in the definition.

**The area of a fundamental set**

A figure made as the direct sum (that is, join without overlap except at edges and vertices) of a finite number of rectangles (whose edges are parallel to the coordinate axes and whose boundaries may or may not be included) is called a fundamental set (Fig. 0.4A.1). It should be obvious that the join and the product (common set) of two fundamental sets are both fundamental sets. The area of a fundamental set is defined as the total sum of the areas of the constituent rectangles.

![Figure 0.4A.1: Fundamental set](image)

**How to define the area of more complicated figures; a strategy**

For a more complicated figure, a good strategy must be to approximate it by a sequence of fundamental sets allowing increasingly smaller rectangles. Therefore, following Archimedes, we approximate the figure from inside and from outside (that is, the figure is approximated by a sequence of fundamental sets enclosed by

---

\(^1\)Appendix 2.4A of Oono *The Nonlinear World* (Springer, 2012)

\(^2\)(Under the usual axioms of mathematics) we encounter figures without areas.

\(^3\)Area connects the world of figures and that of numbers. Therefore, we need explicitly an ‘adaptor’ connecting these two disparate categories that must be free from ambiguous interpretation. The definition of the area of a rectangle given here is a typical example of the adaptor. Notice that it is operationally unambiguous. Recall how the units of various physical quantities are defined. A unit connects something extra-numerical to numbers, so it is a kind of adaptor, requiring an unambiguous specification. Thus, an explicit specification of an operational procedure is the core of the definition of a unit. The reason why physics appreciates operational definitions is that there is no room for interpretation.
the figure and by a sequence of fundamental sets enclosing the figure). If the areas of the inside and the outside approximate sequences agree in the limit, it is rational to define the area of the figure by the limit. Let us start from outside.

**Outer measure**
Let $A$ be a given bounded set (that is, a set that may be enclosed in a sufficiently large disk). Using a finite number of (or countably many) rectangles $P_k$ ($k = 1, 2, \cdots$), we cover $A$, where the boundaries of the rectangles may or may not be included, appropriately. If $P_i \cap P_j = \emptyset$ ($i \neq j$) and $\cup P_k \supset A$, $P = \{P_k\}$ is called a finite (or countable) cover of $A$ by rectangles (Fig. 0.4A.2O). Let the area of the rectangle $P_k$ be $m(P_k)$. We define the outer measure $m^\ast(A)$ of $A$ as follows:

$$m^\ast(A) \equiv \inf \sum_k m(P_k). \quad (0.4A.1)$$

Here, $\inf$ is taken over all the possible finite or countable covers by rectangles.

**Inner measure**
For simplicity, let us assume that $A$ is a bounded set. Take a sufficiently large rectangle $E$ that can enclose $A$. Of course, we know the area of $E$ is $m(E)$. The inner measure of $A$ is defined as

$$m_*(A) = m(E) - m^\ast(E \setminus A). \quad (0.4A.2)$$

It is easy to see that this is equivalent to the approximation from inside (Fig. 0.4A.2I). Clearly, for any bounded set $A$ $m^\ast(A) \geq m_*(A)$ holds.

**Area of figure, Lebesgue measure**
Let $A$ be a bounded set. If $m^\ast(A) = m_*(A)$, $A$ is said to be a measurable set (in the present case, a set for which its area is definable) and $\mu(A) = m^\ast(A)$ is called its area (2-dimensional Lebesgue measure).

At last the area is defined. The properties of a fundamental set we have used are the following two:

1. It is written as a (countable) direct sum of the sets whose areas are defined.
2. The family of fundamental sets is closed under $\cap$, $\cup$ and $\setminus$ (we say that the family of the fundamental

---

4The infimum of a set of numbers is the largest number among all the numbers that are not larger than any number in the set. For example, the infimum of positive numbers is 0. As is illustrated by this example, the infimum of a set need not be an element of the set. When the infimum is included in the set, it is called the minimum of the set. The above example tells us that the minimum need not exist for a given set.

5$A \setminus B$ in the following formula denotes the set of points in $A$ but not in $B$, that is, $A \cap B^c$. 
sets makes a set ring.\(^6\)

An important property of the area is its additivity: If \(P_i\) are mutually non-overlapping rectangles, \(\mu(\bigcup P_i) = \sum \mu(P_i)\). Furthermore, the \(\sigma\)-additivity for countably many summands also holds.\(^7\)

Notice that such a summary as that the area is a translationally symmetric \(\sigma\)-additive set-theoretical function which is normalized to give unity for a unit square does not work, because this does not tell us on what family of sets this set-theoretical function is defined. The above summary does not state the operational detail about how to measure the areas of various shapes, so no means to judge is explicitly given what figures can be measurable. Lebesgue’s definition of the area outlined above explicitly designates how to obtain the area of a given figure.

**Discussion 2.4A.1.** Read S. Wagon, *The Banach-Tarski Paradox* (Cambridge University Press, 1993). Is there any possibility for the Banach-Tarski theorem to be meaningful in understanding natural phenomena?\(^9\)

**General measure (abstract Lebesgue measure)**

The essence of characterization of the area is that there is a family of sets closed under certain ‘combination rules’ and that there is a \(\sigma\)-additive set-theoretical function on it. Therefore, we start with a \(\sigma\)-additive family \(\mathcal{M}\) consisting of subsets of a set \(X\): A family of sets satisfying the following conditions is called a \(\sigma\)-additive family:

1. \(X, \emptyset \in \mathcal{M}\),
2. If \(A \in \mathcal{M}\), then \(X \setminus A \in \mathcal{M}\),
3. If \(A_n \in \mathcal{M}\) \((n = 1, 2, \cdots)\), then \(\bigcup_{n=1}^\infty A_n \in \mathcal{M}\).

\((X, \mathcal{M})\) is called a measurable space. A nonnegative and \(\sigma\)-additive set-theoretical function \(m\) defined on a measurable set that assigns zero to an empty set is called a measure, and \((X, \mathcal{M}, m)\) is called a measure space. Starting with this measure \(m\), we can define the outer measure on a general set \(A \subset X\), mimicking the procedure already discussed above. The inner measure can also be constructed. When these two agree, the procedure already discussed above.

The completion is unique. In a complete measure space, if \(A \in \mathcal{M}\), then

\[ m(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty m(A_n). \]  

**The final answer to the question, “What is the area?” is:** the area is the completion of the Borel measure, where the Borel measure is the \(\sigma\)-additive translation-symmetric measure that gives unity for a unit square and is defined on the Borel family of sets which is the smallest \(\sigma\)-additive family of sets including all the rectangles. Generally speaking, a measure is something like a weighted volume. However, there is

---

\(^6\)More precisely, that a family \(\mathcal{S}\) of sets makes a ring implies the following two:

(i) \(\mathcal{S}\) includes \(\emptyset\),

(ii) if \(A, B \in \mathcal{S}\), then both \(A \cap B\) and \(A \cup B\) are included in \(\mathcal{S}\).

\(^7\)Indeed, if \(A = \bigcup_{n=1}^\infty A_n\) and \(A_n\) are mutually exclusive (i.e., for \(n \neq m\) \(A_n \cap A_m = \emptyset\)), for an arbitrary positive integer \(N\) \(A \supset \bigcup_{n=1}^N A_n\), so \(\mu(A) \geq \sum_{n=1}^N \mu(A_n)\). Taking the limit \(N \to \infty\), we obtain \(\mu(A) \geq \sum_{n=1}^\infty \mu(A_n)\). On the other hand, for the external measure \(m^*(A) \leq \sum_{n=1}^\infty m^*(A_n)\), so \(\mu(A) \leq \sum_{n=1}^\infty \mu(A_n)\).

\(^9\)If we assume that every set has an area, under the usual axiomatic system of mathematics, we are in trouble. See Banach-Tarski’s theorem (Discussion 2.4A.1).

\(^9\)Professor Lee Rubel (1927-1995) told the author in 1988 that any serious theorem will be needed by serious physics (according to the history). S. Banach (1892-1945), A. Tarski (1901-1983).

\(^10\)The completion is unique. In a complete measure space, if \(A\) is measure zero \((\mu(A) = 0)\), then its subsets are all measure zero. Generally, a measure with this property is called a complete measure. Completion of \((X, \mathcal{M}, m)\) may be understood as the extension of the definition of measure \(m\) on the \(\sigma\)-additive family generated by all the sets in \(\mathcal{M}\) + all the measure zero set with respect to \(m\).

\(^11\)E. Borel (1871-1956)
no guarantee that every set has a measure ($\mu$-measurable). It is instructive that a quite important part of the characterization of a concept is allocated to an ‘operationally’ explicit description (e.g., how to measure, how to compute). Recall that Riemann’s definition of the integral was based on this operational spirit, so it can immediately be used to compute integrals numerically.

**Jordan measure**

Before Lebesgue Jordan (1838-1922) defined his measure. His idea is to tessellate small elementary figures (e.g., squares of edge length $\epsilon$ in a given set $A$ as much as possible to estimate the area (from below). This estimate gives $a_\epsilon(A) = \text{the upper bound of the number of the elementary figures in } A \times \text{the area of the elementary figure}$.

Then, take the limit $\epsilon \to 0$ to define the inner measure $a(A)$ of $A$. The outer measure of $A$ may be analogously defined. Then, if $\overline{m}(A) = a(A)$, we say $A$ is (Jordan) measurable, and the agreed value is the area of $A$. Since the Lebesgue’s method of covering is more flexible than that by Jordan, we have the following inequalities:

$$a(A) \leq m(A) \leq \overline{m}(A) \leq a(A). \quad (0.4A.4)$$

Therefore, Jordan measurability implies Lebesgue measurability. However, the converse is not generally true. Since Jordan does not allow to use all the sizes at once, we cannot say, for example, that the outer measure of any countable set is zero within Jordan’s framework. $\sigma$-additivity cannot be asserted, either.

It is very interesting to note that the argument based on a certain unit is, even if the unit size is taken infinitesimal later, definitely weaker than the argument that allows the use of all the sizes at once. Since, eventually, we allow indefinitely small units, we might expect that the conclusions must be the same, but it is not. Is there any implication of this observation to the physical world or to physics?

Incidentally, we must clearly pay attention to the fact that humankind recovered the refined mathematical and logical level of Archimedes (287-212 BCE) of more than 2000 years ago only around or slightly before the time of Jordan. We must not forget the warning that culture can easily retrogress (medievalization occurs all too easily) with a sense of impending crisis.

**What is probability?**

Kolmogorov (1903-1987) defined probability as a measure whose total mass is normalized to unity.\footnote{A. N. Kolmogorov, *Foundations of the Theory of Probability* (2nd edition) (Chelsea, 1957); http://www.mathematik.com/Kolmogorov/index.html. Recommended. The third edition in Russian contains a reprint of this article with a nice outline of the history of probability theory by A. N. Siryaev.} Since long time ago ‘What is probability?’ has been a difficult problem.\footnote{A summary may be found in D. Gillis, *Philosophical theories of probability* (Routledge, 2000), but the argument given here is not described in this book.} An interpretation of the probability of an event is that it is a measure of our confidence in the occurrence of the event (subjective probability). It is something like a weight of the event, if we express the event as a set of elementary events compatible with the event. That is, if an event is interpreted as a set, its probability should be handled just as a measure of the set. Therefore, without further going into the problem of interpretation of probability, to specify only how to handle it clearly is the approach adopted by Kolmogorov. This approach may not squarely answer the question: “what is the probability?”. For example, there is no clear relation to relative frequencies. An important observation is that we can construct a theory of probability that is sufficiently rich and practical without answering any ‘philosophically (apparently) deep’ questions.

Avoiding the discussion of the meaning of probability and constructing only the algorithm for it may be admissible as mathematics, but the approach is incomplete if we wish to apply it to study Nature, the reality. If we wish to apply the concept of probability to the reality, we need its interpretation. Even Kolmogorov’s
definition is not aloof from the interpretation of probability, although he apparently avoids it. This approach contains the crucial idea that for the subjective probability (the extent of confidence) to be rational it must be interpreted as a measure.

For the cases of casting dice and tossing coins the numbers we call probability are based on our experience about frequencies and are consistent with the law of large numbers. It is not hard to accept intuitively that the empirical probabilities thus obtained obey the same logic as measures do. Such probabilities are understood as objective (and can be empirically confirmed with the aid of the law of large numbers). There is, however, a deep-rooted opinion that subjective probability is distinct from empirical probability (based on frequency). Such an opinion is, of course, due to the humanistic fallacy that our logic and language are unrelated to our empirical world.

Suppose there are two mutually exclusive events 1 and 2 with objective probabilities (relative frequencies) $p_1$ and $p_2$ ($> p_1$), respectively. If a subjective probability $p'$ of a gambler for these events become $p'_1 > p'_2$, then his gain on the average must be smaller than the gambler with $p = p'$ (i.e., whose subjective assessment is consistent with the objective reality). Thus, the agreement of subjective probability and empirical probability based on relative frequency is forced upon us (i.e., the subjects who choose), when we are subjected to natural selection. The probability based on relative frequency satisfies measure-theoretical axioms. Therefore, the subjective probability molded by natural selection follows, as long as it is useful for our survival, measure-theoretical axioms. Consequently, the assertion that subjective probability = extent of confidence behaves as volume or weight looks very natural. Or, we should say that our nervous system/emotion has been made to evolve so that this looks natural. The essence of probability is the amount of confidence backed by relative frequency, so even apparently subjective probabilities can be effective in empirical sciences.

There have been many philosophers who oppose the frequency theory of probability because the probability of a unique event is hard to think of. For example, Carnap thought that probability has a logical meaning independent of empirical facts, and tried to found probability on “the degree of confirmation” that is based on the logical relations among events. However, such attempts are typical humanistic fallacy totally forgetting about the fact that logical capacity has been formed under natural selection. When one tries to think about probability, even an apparently unique event is not considered as an event that occurs only once. It is embedded into the totality of experienced events (and logically inferred conclusions from them) during the evolution process that generated our nervous system. In other words, the probability of an event is gauged against what has been embodied by the phylogenetic learning.

---

14The reader might say what matters is not the subjective probability, but the probability of the person to take appropriate behaviors, because actual behavior is important, not the belief. However, it must be disadvantageous that the correspondence between thoughts and actions is not simple.


16Interestingly enough, we seem to have an ability to take statistics unconsciously from our experiences. Not only us but various animals have capacity for Bayesian learning under natural conditions as summarized in the following paper: T. J. Valone, “Are animals capable of Bayesian updating? An empirical review,” Oikos 112, 252 (2006). There is also an unconscious process of slow learning through trials and errors: P. J. Bayley, J. C. Frascino and L. R. Squire, “Robust habit learning in the absence of awareness and independent of the medial temporal lobe,” Nature 436, 550 (2005).
There have been numerous attempts to relate probability to randomness. The feeling of randomness comes from the experience when we choose items (events) that there is no way to reduce damage by any suitable bias. Consequently, if the world is ‘uniform,’ it is equivalent to equal probabilities for the events. However, as we will see later, it is hard to define randomness precisely, so it is not easy to found probability on randomness.\footnote{G. Shafer and V. Vovk, \textit{Probability and Finance: its only a game!} (Wiley-Interscience, 2001) is a notable attempt to develop probability theory not depending on the Kolmogorov axiomatic system. For example, it is possible to prove that if a gambler plays against Nature, and if there is no way to accumulate his wealth unboundedly, then the strategy of Nature (or the output of Nature) must satisfy the strong law of large numbers.}