Gregory Kordas
Last update: March 21, 2023

LECTURE 2
PROBABILITY MEASURES

Abstract: We discuss the assignment of probability first to events belonging to finite, then to countably infinite, and finally to uncountably infinite sample spaces.

1. Introduction.

Although the term *probability* is used frequently in everyday speech, it is difficult to pin down exactly what we mean by it. For example, it is generally agreed that a probability of $\frac{1}{2}$ should be assigned to the event of obtaining a head in the toss of a fair coin, but it is natural to seek an explanation as to why this should be so.

One line of reasoning might be that if there are only two possible outcomes and it is agreed that neither seems favored over the other, they should be assigned equal probabilities, i.e., each should receive a probability of $\frac{1}{2}$.

Laplace's Principle of Insufficient Reason : If there are n mutually exclusive outcomes and there is no reason (or there is insufficient reason) to believe that any of them are more or less likely to occur than the others, the probability assigned to each one should be $1/n$.

A second way of assigning probabilities to the possible outcomes advocated by Savage (1954), would be to argue that probabilities are subjective assessments of the likelihood of particular events. There is no reason why we should all agree on the number assigned to the probability of a head, as long as we all obey some natural "constraints", like the "reasonable" requirement that probabilities should be positive and add to 1.

Savage's Principle of Subjective Probability: As long as we all agree on the sample space, and we all obey the rule that our subjective probabilities add to 1, we may assign any number between 0 and 1 we wish to the probability of an individual event.

This is the so-called "Bayesian" definition of probability. Von Mises (1957) couldn't disagree more. He argued that a purely subjective view of probability would not be enough to take Statistics into the realm of the Sciences: If we cannot agree on an objective way of assigning probabilities to particular events, then we are allowing the possibility that different scientists

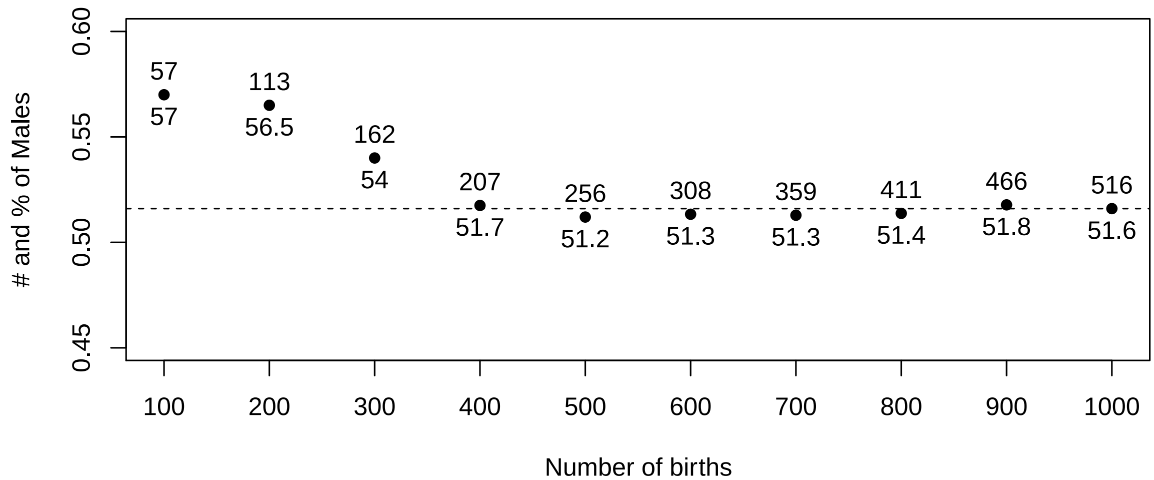

FIGURE 1. Number and percent of males at every 100 births up to a 1,000 births. (Data: 1,000 births at *Marika Iliadi Hospital*, Athens, Greece, starting at Jan. 1, 1983.)

working on the same problem might reach different conclusions without the means of deciding which answer is the "correct" one!

Von Mises's Principle of Objective Probability: The probability of an event is its long run frequency as the number of trials becomes infinitely large.

This is the so-called "classical" definition of probability. Figure 1 presents an empirical determination of the human sex ratio at birth, which is internationally accepted to be $51 - 52\%$ males and $48 - 49\%$ females. In our sample of 1,000 births at *Marika Iliadi Hospital* in Athens, Greece, the sex ratio converges to a value of 51.6% boys and 48.4% girls after 1,000 births.This is a good example for both Laplace's definition as well as the classical definition: The reason why boys are born with higher probability than girls is yet unknown and it is the subject of a great deal of research currently in biology and medicine. Laplace's theory of insufficient reason would assign a 50-50 chance, but the empirical outlook of the objective probability viewpoint leads us to a 52-48 chances in favor of males.

The classical definition of probability has several draw-backs, not least of which is that it rests on the unrealistic concept of experiments that can be repeated *ad infinitum*. Von Mises' criticism of Savage's concept is not that convincing either, especially if one thinks of an economic situation where two investors considering buying a financial asset might very well assign different probabilities to future movements of the asset's price, and therefore arrive at different investment decisions. After all, if all investors had the same expectations about the future movements of asset prices, there would be no trade in these assets!

These issues have been the subject of a fierce debate over the years[1]. In what follows, we will bypass these philosophical (meta-mathematical) issues by adopting Kolmogorov's agnostic (mathematical) position:

Kolmogorov's Axiomatic Definition of Probability: Like all mathematical objects, probability can (and should!) be defined on an axiomatic basis. Exactly like the concepts of "point" and "line" in the axiomatization of Euclidean geometry, the origin or meaning of the concept of probability and the justification of the axioms on which our probability calculus will be based is *not* part of mathematics. Statements appearing in this context may be considered as axioms connecting undefined terms.

Obviously, all the interpretations of probability discussed above are consistent with Kolmogorov's agnostic axiomatics, so one is free to interpret the axioms (which we will discuss presently) either from the frequentist/classical or the Bayesian/subjective viewpoint.

2. Probability Measures on Finite Sample Spaces.

A process such as tossing a coin, rolling a die, or selecting a playing card from a pack is called an *experiment*. We start by considering experiments that admit only a finite number of *elementary outcomes* (ω's): in the coin-tossing experiments there are 2 possible outcomes (head and tails); in the second and third there are 6 and 52, respectively. An *event* is a collection of elementary outcomes. The collection of all the elementary outcomes is an event called the *sample space*, and is denoted by Ω.

One might, of course, be interested in the occurrence of an odd-numbered face rather than an even one in the die experiment, so it is not only the elementary events themselves that are of interest, but also all sorts of other events that can be constructed from them. For example, in the coin tossing experiment there are

$$\mathrm{C}_0^6 + \mathrm{C}_1^6 + \cdots + \mathrm{C}_6^6 = 2^6$$

potentially interesting events altogether, where C_x^n denotes the $n!/[(n-x)!x!]$ possible combinations of n objects in groups of x objects.

[1]For a recent account of these historical debates from a Bayesian (i.e., the defending) viewpoint see Gelman and Robert (2013), as well as, the comments by other authors and the rejoinder in *The American Statistician*, 2013, 67(1), pp. 1-17.

A finite class $\mathscr{F}$ of subsets of Ω is a *field* if

(i) $\Omega \in \mathscr{F}$;
(ii) (closed under complementation) $A \in \mathscr{F}$ implies that $A^c \in \mathscr{F}$;
(iii) (closed under unions) $A, B \in \mathscr{F}$ implies that $A \cup B \in \mathscr{F}$.

Given a class $\mathscr{F}$ it is easy to check if it is a field or not. For example, it is easy to see that $\mathscr{F} = \{\emptyset, \Omega\}$ is the smallest possible field. In the die-rolling experiment, $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the class $\mathscr{F} = \{\emptyset, 1 \cup 3 \cup 5, 2 \cup 4 \cup 6, \Omega\}$ of subsets of Ω, is indeed a field. Note however, that this field contains none of the individual ω's and omits many subsets of Ω such as $5 \cup 6$. This field is actually the smallest that contains the odd and even outcomes and is known as the field *generated* by these outcomes. In general, the smallest field generated by some event A is given by $\{\emptyset, A, A^c, \Omega\}$.

It is instructive to see precisely how a *class* of events generates a field. Let $\mathscr{A} = \{A_1, A_2\}$ be a class of events ω. At the outset Ω is placed in the field together with its complement $\emptyset$. Then all possible sets and their complements are added to give $\Omega, \emptyset, A_1, A_1^c, A_2, A_2^c$. These are supplemented by all possible unions that are not already present: $A_1 \cup A_2, A_1 \cup A_2^c, A_1^c \cup A_2$ and $A_1^c \cup A_2^c$. Finally, these new unions must have their complements, like $(A_1 \cup A_2)^c$, added to generate the field.

Our next task is that of assigning probabilities to events in any given field $\mathscr{F}$.

A *probability measure* is a set function $P(\cdot)$, mapping elements of $\mathscr{F}$ into $[0, 1]$, with the following properties:

(i) For all $A \in \mathscr{F}$, $0 \leq P(A) \leq 1$;
(ii) $P(\emptyset) = 0$ and $P(\Omega) = 1$;
(iii) For $A_1, A_2 \in \mathscr{F}$ and $A_1 \cap A_2 = \emptyset$, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Example 1.

(1) Consider the experiment of tossing a coin once. Then $\Omega = \{H, T\}$, $\mathscr{F} = \{\emptyset, H, T, \Omega\}$ is a field, and it seems reasonable to assign $P(H) = P(T) = \frac{1}{2}$, so that $P(\Omega) = 1$, $P(\emptyset) = 0$, and $P(\cdot)$ is a probability measure on $\mathscr{F}$.

(2) A variation of the previous experiment is the case where a pair of identical coins are tossed. There are at least two sample spaces that one may consider here: $\Omega = \{HH, HT, TT\}$, or $\Omega = \{HH, HT, TH, TT\}$. The first sample space corresponds to an experiment in which the order of the appearance of heads or tails is of no consequence, while the later sample space takes the order into account. If the coins are fair and order is inconsequential, a reasonable assignment of probabilities would be, $P(HH) = P(TT) = \frac{1}{4}$ and $P(HT) = \frac{1}{2}$. If the order of

the appearance of H and T is taken into account, we should instead assign $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$. This example illustrates the importance of being clear at the outset about the nature of what events are interesting, so that suitable sample spaces, fields and probability measures can be defined.

(3) In the experiment of rolling a die once, $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $\mathscr{F}$ might consist of all 2^6 possible outcomes of Ω, i.e. the $\emptyset$, the C_1^6 elementary events, the C_2^6 events like $1 \cup 2$, the C_3^6 events like $2 \cup 4 \cup 6$,..., the $C_6^6 = 1$ event $1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6$, and Ω. As before the elementary events belong to $\mathscr{F}$ and it is reasonable to assign $P(1) = P(2) = ... = P(6) = \frac{1}{6}$. The probability of any other member of $\mathscr{F}$ can now be calculated via property (iii) of the definition of a probability measure.

(4) A modification of the above example would be the experiment in which $\Omega = \{1, 2, 3, 4, 5, 6\}$ as before, but $\mathscr{F} = \{\emptyset, 1 \cup 3 \cup 5, 2 \cup 4 \cup 6, \Omega\}$, so that we only consider even and odd outcomes. Note that the elementary outcomes are not members of $\mathscr{F}$, so our previous definition of $P(\cdot)$ will not do here. We can however, still define an external ancillary function $p(\cdot)$ and assign $p(1) = p(2) = ... = p(6) = \frac{1}{6}$. Clearly $p(\cdot)$ is very similar to a probability measure but it is not defined on a field and it has no defined additivity properties. Nonetheless, $p(\cdot)$ may be used to define a valid probability measure $P(\cdot)$ on $\mathscr{F}$. Set $P(1 \cup 3 \cup 5) = p(1) + p(3) + p(5)$, $P(2 \cup 4 \cup 6) = p(2) + p(4) + p(6)$, $P(\emptyset) = 0$, $P(\Omega) = 1$. Then $P(\cdot)$ satisfies all the requirements in the definition of a probability measure, and is indeed a probability measure on $\mathscr{F}$.

■

The elementary events in Example 4, such as "6", are not members of $\mathscr{F}$, so they are said to be *not measurable* with respect to (this particular) $\mathscr{F}$. The triplet $(\Omega, \mathscr{F}, P)$ is called a *probability space*.

3. Probability Measures on Countably Infinite Sample Spaces.

It is now time to extend the notions of a field and a probability measure to cases where Ω is composed of an infinite number of outcomes. The sample space of many experiments can be modeled by setting $\Omega = \{1, 2, ...\}$, the set of all positive integers. In such cases, the sample space is said to be *countably infinite*. Examples of experiments with such sample space might be the number of particles emitted by a radioactive substance, or the number of messages arriving at a switching board in a given period of time. The sample space is permitted to be infinite because there is no reason to fix an upper bound to the counts that could be observed in advance of the experiment.

The starting point, as before, is to define the set of interesting events about which probabilistic statements are to be made for the countable infinite sample space $\Omega = \{1, 2, ...\}$. It is pertinent to ask whether a field constructed by the definition of the previous section is sufficient to generate all potentially interesting events when Ω is countably infinite. As usual Ω must belong to the field, but our previous definition is not adequate to guarantee that it does. To see this, note that Ω can certainly be constructed from the union of its odd and its even members, that is $\Omega = \{1\} \cup \{3\} \cup \cdots \cup \{2\} \cup \{4\} \cup \cdots$, but the event $\{1\} \cup \{3\} \cup \cdots$ is not necessarily a member of any field, since it is not constructed as a finite union of events, being itself an infinite union. Hence, using members of a field constructed as in the previous section only, it is not possible to write Ω as $\Omega = \{1\} \cup \{3\} \cup \cdots \cup \{2\} \cup \{4\} \cup \cdots$.

A similar problem arises if an attempt is made to use a probability measure as defined in the previous section. Since Ω has to be a member of any field, and $P(\Omega) = 1$, it follows that $P(\{1\} \cup \{2\} \cup \cdots) = 1$. But we cannot write this as $P(\Omega) = P(1) + P(2) + \cdots$, since the $P(\cdot)$ as defined above is only finitely additive and the event in question is an infinite union.

It should by now be clear that we will need to extent the definitions of fields and probability measures to accommodate countably infinite union and countable additivity, respectively.

To distinguish a field from one which also has the additional property that countable (possibly infinite) unions of sets in the field are also in the field, the terminology *sigma-field*, or σ-field is often employed, and is denoted by σ-$\mathscr{F}$.

We say that σ-$\mathscr{F}$ is a σ-field if

(i) $\Omega \in \sigma$-$\mathscr{F}$;
(ii) (closed under complementation) $A \in \sigma$-F implies that $A^c \in \sigma$-$\mathscr{F}$;
(iii) (closed under countable unions) $A_i \in \sigma$-$\mathscr{F}$, $i = 1, 2, ...$ implies that $\cup_i A_i \in \sigma$-$\mathscr{F}$.

Likewise, it is possible to refine our definition of a probability measure so that it can assign probabilities to (potentially) countably infinite unions of events.

A *probability measure* is a set function, $P(\cdot)$, mapping elements of σ-$\mathscr{F}$ into $[0, 1]$, with the following properties:

(i) For all $A \in \sigma$-$\mathscr{F}$, $0 \leq P(A) \leq 1$;
(ii) $P(\emptyset) = 0$ and $P(\Omega) = 1$;
(iii) (countable additivity) For disjoint events $A_i \in \sigma$-$\mathscr{F}$, $i = 1, 2, ...$, $P(\cup_i A_i) = \sum_i P(A_i)$.

Example 2. Let $\Omega = \{1, 2, ...\}$ and consider the function

$$p(i) = \frac{1}{(e-1)i!}, \quad i = 1, 2, ...$$

where is e the base of natural logarithms. It is easy to verify that $\sum_i p(i)$ over the positive integers is unity. Define $A_i = \{2i\}$ so that $p(A_i)$ attaches weights to even elementary events in the sample space. Let the σ-filed of interest, σ-$\mathscr{F}$, be the set of all possible subsets of Ω, the power set. Then $P(\cup_i A_i)$ is the probability assigned to an even outcome of the experiment. Using the function $p(\cdot)$, $P(\cup_i A_i) = \sum_i p(2i)$ and $P(\cdot)$ constructed in this way is a valid probability measure on σ-$\mathscr{F}$. The even numbers form a countably infinite set and the probability of obtaining an even value actually turns out to be about 0.316 (try it). ■

It is worth noting here that it is not possible to have a countably infinite sample space that is composed of events that are equally likely, for no function $p(\cdot)$ can exist which sums to unity across such Ω when $p(\omega)$ is constant across all outcomes. That is, one cannot find a countably additive measure such that $P(\Omega) = \sum_i P(\{i\}) = 1$ when $P(\{i\})$ is constant. An important consequence of this is that **a countably infinite space cannot be found to serve as a model for tossing a fair coin an infinite number of times**. This realization foreshadows one aspect of the difficulties that arise when Ω is permitted to be the real line, rather than the positive integers, to which we turn presently.

4. Probability Measures on Uncountably Infinite Sample Spaces.

Consider a manufacturing process that produces candy bars. No such process is perfect and no two bars will be identical to each other. The manufacturer, however, would presumably wish to have the weight W of any given bar lay within some tolerance interval with high probability. That is to say, it will be of interest to construct a framework within which a statement like $P(W \in [a, b]) = 0.95$, relating to an interval $[a, b]$, can be made. A natural space to consider for developing such mechanisms is then the real line $\mathbb{R}$ and intervals thereof. Accordingly, the sample space $\Omega = \mathbb{R}$ is used and interesting events are constructed using intervals in that space.

The contrast between this Ω and the ones previously discussed is that the real line has a complicated mathematical structure in comparison to the set of positive integers. To be sure, $\mathbb{R}$ consists of a countable set points, the rational numbers such as $\frac{1}{2}$ and $\frac{1}{3}$ etc., and also an uncountable set, the irrationals such as $\sqrt{2}$, π and e. Of course, the sets of rational numbers and the of irrational numbers are complements of each other and their union is, by definition, the whole real line.

In order to devise an appropriate field and probability measure in the uncountably infinite case, one might think of extending the conditions in the previous section to cover arbitrary, rather than countable, unions. This would emerge by analogy with the transition from the finite to the countable infinite case and would constitute an attempt to define a probability measure on all possible subsets of $\Omega = \mathbb{R}$. The set of all possible subsets is sometimes called the *power set*. However, for various reasons, it proves infeasible to construct a probability measure on the class of *all* subsets of $\mathbb{R}$. The main reason for this failure is because the power set of $\mathbb{R}$ is just too large to permit the construction of a sensible measure on it. For one thing, we would be forced to include *uncountable unions* of sets in any σ-field constructed out of the power set of $\mathbb{R}$, and uncountable unions are nor permitted in set theory! We will return to this later.

It is thus necessary to restrict the class of sets to be used but not to exclude any probabilistically useful ones. When, as here, the sample is $\Omega = \mathbb{R}$ it is quite natural to think of an interval like $(a, b]$ as an elementary event. Define $\mathscr{I}$ to be the class of intervals of the type $(a, b]$ on $\mathbb{R}$. The subsets of $\Omega = \mathbb{R}$ in which we shall be interested are unions of intervals.

Let us begin by considering the class of intervals, $\mathscr{I}_1$, of $\Omega = (0, 1]$, and for the sake of concreteness, we may define the probability (measure) of such intervals as $P(a, b] = b - a$, which is known as *Lebesgue measure*. An immediate difficulty arises with $\mathscr{I}_1$ (and $\mathscr{I}$) in that it is not closed under complementarity of unions. For example, if $I = (0.5, 0.75]$, then $I^c = (0, 0.5] \cup (0.75, 1]$, so the complement of an interval may not itself be an interval but rather a union of disjoint intervals. Provided we assume that $\emptyset \in \mathscr{I}_1$, this class of intervals is closed under intersections, but it is not, of course, a field because of the problem with the complements.

Building on previous discussion, it can be seen that the class $\mathscr{I}_u$ of *finite unions of intervals* is a field, since it is closed under unions and complements. We shall employ the notation $\mathscr{I}_u$ to mean the class of finite unions of intervals either on $(0, 1]$ or on $\mathbb{R}$ where this causes no ambiguity. The Lebesgue probability measure can now be used to give

$$P_u\left(\bigcup_{i=1}^{n}(a_i, b_i]\right) = \sum_{i=1}^{n} P(a_i, b_i] = \sum_{i=1}^{n}(b_i - a_i)$$

whenever the n intervals are non-overlapping and all belong to $(0, 1]$. It should also be clear that P_u is finitely additive over the field $\mathscr{I}_u$, so our discussion above regarding finite measures applies.

However, the restriction to finite unions turns out to be inadequate, and countably infinite unions must also be considered. The simple case of a singleton $\omega = x$ can be used to show this: since x is the (infinite) intersection $\cap A_n$ where $A_n = (x - \frac{1}{n}, x]$, it follows from De Morgan's Law [which states that, for two events A_i and A_j, $(A_i \cup A_j)^c = A_i^c \cap A_j^c$] that this is a countable, but not finite, union of events. More complicated sets may require both countability and complementarity. For instance, each rational number, being a singleton, is a countable intersection of intervals as above and, since the set of rationals is countable, it follows that the set of rationals may be formed by countable unions. The set of irrationals, which is uncountable, is then formed by complementarity.

The construction of a suitable model for the infinite coin-tossing experiment mentioned above, can now be used to provide some probabilistic motivation for considering countably infinite unions. Let $\Omega = (0, 1]$ and consider the *binary* representation of any point $\omega \in \Omega$. Thus $\frac{1}{2}$ may be represented by 0.1, $\frac{1}{4} = 0.01$, $\frac{1}{8} = 0.001$, and so on. Some points, such as these, have terminating binary expansions, but many require infinite ones. Indeed, even those numbers with a terminating expansion have an equivalent non-terminating representation, e.g. $\frac{3}{4} = 1 - \frac{1}{4} = 1 - 0.01 = 0.10111...$, $\frac{1}{2} = 0.01111...$, and it is these infinite expansions that will be used here. Each point $\omega \in (0, 1]$ is therefore represented by an infinite sequence of zeros and ones which may be used to label the outcomes from tossing a coin an infinite number of times, using 1 and 0 to represent heads and tails.

It is clear that it is impossible to find a countable mechanism that would generate a σ-filed that contains *all* intervals in $(0, 1]$, so a more roundabout way of finding the elusive σ-field must be attempted. Now $\mathscr{I}_u$ is a field, even though is does not admit of the countably infinite unions permitted with σ-fields. However, it is a member of various σ-fields, for example the power set. The intersection of all σ-fields of which $\mathscr{I}_u$ is a member is also a σ-field and it does contain $\mathscr{I}$ as well as $\mathscr{I}_u$. This intersection is the smallest σ-field that contains $\mathscr{I}$ and is known as the σ-filed generated by $\mathscr{I}$. This important σ-field is denoted by $\mathscr{B}$ and is called the *Borel σ-field*, and its elements are called *Borel sets.*

The class of countable unions of intervals is a proper subset of the Borel field, which is, in turn, a subset of the power set, i.e., $\mathscr{I} \subset \mathscr{B} \subset 2^{(0,1]}$. From this, it becomes obvious that $\mathscr{B}$ does not contain all possible subsets, or, in other words, there are sets in $(0, 1]$ (or $\mathbb{R}$ if we consider the entire real line) that are not measurable with respect to $\mathscr{B}$. An example of such a set due to Vitali, is given in Billingsley (1995), p.45. The crucial importance of $\mathscr{B}$ is that sets that are nonmeasurable with respect to it are really very exotic and rarely, if ever, would be expected to arise in applications. Also, as long as we confine attention to constructing sets based on countable unions of intervals, we can never generate a set outside $\mathscr{B}$. A drawback is that $\mathscr{B}$ is not constructively defined because, as asserted above, it cannot be generated by any countable mechanism (the arbitrary intersections by means of which the $\mathscr{B}$ is constructed, are uncountably many).

Once we have $\mathscr{B}$, we can construct a probability measure $P(\cdot)$ on it, assigning measure on intervals of $\mathbb{R}$. A difficulty that arises is that any candidate probability measure must first be shown to be countably additive on the class of intervals $\mathscr{I}$. That is, if $(a_i, b_i]$ are disjoint intervals in $\mathscr{I}$ and if $(a, b] = \cup_i (a_i, b_i]$, then $P((a, b]) = \sum_i P((a_i, b_i])$. Of course, this would be trivial if the union of intervals were finite, but as countable unions are not sufficient to encompass $\mathscr{B}$, another step is necessary. This is done by first defining $P(\cdot)$ on $\mathscr{I}$ where countable operations are enough, and then *extending* $P(\cdot)$ on $\mathscr{I}$ to another measure $P^*(\cdot)$ on $\mathscr{B}$, by ensuring that for any set $A \subset \mathscr{I} \subset \mathscr{B}$, $P(A) = P^*(A)$. The famous *Carathéodory extension theorem* asserts that a probability measure on a field has a unique extension to the generated σ-field, which means that since $\mathscr{I}$ is a field and $\mathscr{B}$ is the σ-field generated by it, the extension from $P(\cdot)$ to $P^*(\cdot)$ is indeed unique (see Billingsley, 1995, Theorem 3.1, and Theorem 11.3). The functions P^* and P are different in the sense that they are defined on different domains but, since they are equal on $\mathscr{I}$, we generally use P unless there is a reason to stress the difference.

Example 3. (Lebesgue measure). Let $\Omega = (0, 1]$ and $\mathscr{I}_1$ be the class of intervals on $(0, 1]$. For intervals in $\mathscr{I}_1$, define $P((a, b]) = b - a$ as a candidate probability measure, so that the probability is just the length of the interval. Obviously $P(\cdot)$ is nonnegative and $P(\Omega) = 1$. Furthermore, the probability (measure) of any singleton can be seen to be zero by setting $a = b$. It is easy to see that P is finitely additive on $\mathscr{I}_1$, and it is true, thought much harder to prove, that P is countably additive on $\mathscr{I}_1$, and that this is also the case with the Borel σ-field $\mathscr{B}_1$. Thus P is a probability measure defined on a σ-field and it is known as *Lebesgue measure* on $(0, 1]$. Note that this is the same as a uniform distribution on $(0, 1]$. ■

The theory of measure is a powerful analytical tool that, as the following example shows, can even be used to shed light on the structure of $\mathbb{R}$ itself!

Example 4. Consider how Lebesgue measure might be attached to the set of rational numbers in $(0, 1]$. The rationals are members of $\mathscr{B}_1$, being formed by countable unions. But although there is an infinity of them, they are countable and thus their Lebesgue measure is zero! This means that the irrationals, which are uncountable, have measure one, although they don't form an interval as the rationals are dense in $\mathbb{R}$. Put in plain language, if one draws a number at random from $\mathbb{R}$, it is almost certain that the number would be irrational. Similarly, the algebraic numbers have also measure zero, while the transcendentals (the complement of the algebraics) have measure one. Thus our randomly drawn number would almost certainly also be transcendental, as well as, irrational. An intriguing thought is that, if Nature at the time of the Big Bang drew the universal constants randomly from $\mathbb{R}$, then they should all be irrational and transcendental. And they (probably) are! – all known universal constants, like π, e, and Euler's γ, have been shown to be irrational, and both π and e have also been shown to be transcendental. There is yet no proof that Euler's γ is also transcendental, although it certaintly appears to be so. ■

Although very important theoretically, the Lebesgue measure is, of course, not the only valid probability measure on $\mathbb{R}$. Lebesque measure on $[0, 1]$ correspents to, what we would call in Statistics, a *Uniform distribution* on $[0, 1]$. As the following example shows, all other distributions are defined by their c.d.f. $F(\cdot)$ "with respect to Lebesque measure", as we say (see the "*probability transformation*" discussed in Lecture 3).

Example 5. Let $\Omega = \mathbb{R}^+$, where $\mathbb{R}^+ = [0, \infty)$ is the non-negative real line, and let $\mathscr{I}^+$ be the class of intervals on $\mathbb{R}^+$. Define the probability measure P on $\mathscr{I}^+$ by $P(a, b] = e^{-a} - e^{-b}$. The set function P is clearly nonnegative and $P(\mathbb{R}^+) = e^{-0} - e^{-\infty} = 1$. It can also be shown that P is countably additive on $\mathscr{I}^+$ and, therefore, has a unique extension to the corresponding Borel σ-field $\mathscr{B}^+$. Thus, this is a valid probability measure. Note that this measure corresponds to an *Exponential distribution* on $\mathbb{R}^+$. In general, if $F(\cdot)$ is the c.d.f. of a continuous random variable, $P((a, b]) = F(b) - F(a)$ is a valid measure. ■

References

Billingsley, P. (1995). *Probability and Measure*, Wiley Series in Prob. & Mathematical Statistics.

Gelman, A., and Robert, C.P. (2013) – "Not Only Defended But Also Applied" – The Perceived Absurdity of Bayesian Inference", with comments and rejoinder, *The American Statistician*, 67(1), 1–17.

Kolmogorov, A.N. (1956). *The Foundations of Probability Theory.*

Laplace, P.S. (1820). *A Philosophical Essay on Probability*, available from Dover editions.

Savage, L.J. (1954). *The Foundations of Statistics*, available from Dover editions.

von Mises, R. (1957). *Probability and Statistics*, available from Dover editions.

> [In school] not a word was said to us about he meaning or utility of mathematics: we were simply asked to explain how an equilateral triangle could be constructed by the intersection of two circles, and to do sums in a, b and x instead of in pence and shillings, leaving me so ignorant that I concluded that a and b must mean eggs and cheese and x nothing, with the result that I rejected algebra as nonsense, and never changed that opinion until in my advanced twenties Graham Walls and Karl Pearson convinced me that instead of being taught mathematics I had been made a fool of.
>
> — *George Bernard Shaw*