On Distributions, Probability, and Numbers: a Structuralist Point of View

Uwe Saint-Mont, Nordhausen University of Applied Sciences

August 24, 2026

Abstract. Much has been written about the notion of probability, and how this concept may be interpreted. In contrast, philosophers of science seem to have neglected the concept of (stochastic) distribution, although distributions are at least as fundamental. One the one hand, they generalise the basic mathematical notion of number, on the other, from a structuralist perspective, distributions precede probabilities and their interpretation.

Keywords

Interpretation of Probability; Stochastic Distributions; Philosophy of Mathematics

1 Introduction

The nature of probability has been discussed much. Yet considerably less attention has been devoted to distributions. This is surprising, since from a structuralist point of view, distributions come first (section 2). Also from an applied perspective, distributions and their properties (see, e.g., Johnson et al. (1994, 1995, 2005)) are much more important than specific probabilities (section 3). Statistics further strengthens this point of view, and sheds more light on the various interpretations of the concept of probability (section 4). In a nutshell, the most appropriate interpretation of a specific probability depends on the situation at hand.

2 The structuralist point of view

In the philosophical literature, much has been written about randomness and the concept of probability, or, more precisely, about the latter's concept various meanings (classical/logical, frequentist, subjectivist, evidential, as a propensity ,...). Mathematicians may be rather puzzled by these distinctions and the amount of ink that has been spilled, as ever since Hilbert's formalism, syntax has been (much) more important than semantics. If asked about the concept of probability, the straightforward reference therefore is to its axiomatic treatment, in particular Kolmogorov (1933).

In other words, the latter formal treatment clarifies the concept of probability independently of any interpretation: Given a set S (of all possible events), any function $p(\cdot)$ satisfying Kolmogorov's classical axioms,

(i) non-negativity: for all $A \subseteq S$ we have $p(A) \geq 0$,

(ii) unit measure: $p(S) = 1$,

(iii) additivity: If $A, B \subseteq S$, and $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$,

is named a 'probability (measure)' on the 'probability space' S.

The formal treatment of the concept of probability is no outlier, rather, it is a textbook example, since for more than one hundred years, the axiomatic approach has become the hallmark of mathematics. Instead of endless and ultimately rather fruitless discussions on the 'nature' or 'essence' of some concept, object or theory, a set of axioms is able to specify and thus clarify the formal properties of an idea, and relegate interpretations to 'philosophy', i.e., some realm outside that of strict logic. In other words, this ansatz puts syntax in first place and assigns a subordinate role to the interpretation or meaning of a term.

Consistently, the twin concept of probability, information $I(p) = -\log(p)$, also does not refer to any specific meaning. This list could be extended: any mathematical object that adheres to Peano's axioms is called a natural number system (or a version of the natural numbers $\mathbb{N}$), any set that satisfies the axioms of associativity, has an identity element and possesses an inverse element is called a group, and every formal system that adheres to the axioms of Euklid (or Hilbert) is a (Euklidean) geometry. In general, consequentially, the core concept of contemporary mathematics is that of a so-called 'space', i.e., a set endowed with some kind of structure. No matter whether the latter is called geometric, topological, algebraic, or functional, it is always defined by an ensemble of axioms.

The received idea that mathematics is the study of quantity — numbers — has thus been generalised to the notion that it is concerned with quantity and structure, i.e. numbers, sets, and all kinds of formal relations. Following this train of thought, Hilbert's time-honoured philosophy of mathematics, formalism, has become structuralism, the prevailing contemporary philosophy of mathematics (Resnik 1997, Shapiro 1997).

This point of view has two major consequences. For one, a certain formal structure may be applied wherever its seems (to) fit. Therefore, such logically consistent structures have been named 'free-floating', since they can be linked to any kind of empirical phenomenon, which has a similar structure, or 'free-standing' — since they have no 'real' or definite foundation (Burke 2015).

Second, a specific mathematical object is just (or, at least, first and foremost) a position in a certain structure. For instance, instead of pondering what the number 3 really means or how it may be interpreted, structuralism starts with the Peano axioms, which define $\mathbb{N}$, for instance $\{\emptyset\}, \{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}, \ldots$, and says that the number three is tantamount to the third position in this series. More generally, any natural number is some element of the set $\mathbb{N}$, every element of a linear space is called a vector, and an open set is defined on a topological space.

In the same vein, a probability space is just a set S with a stochastic structure on it. In probabilistic jargon, every set $A \subseteq S$ is named an event, and the probability $p(A)$ is a function from the set of all subsets $\mathcal{P}(S)$ of S to the interval $[0, 1]$ that adheres to Kolmogorov's axioms. (In total generality, that turns out to be tricky, therefore one restricts attention to a reasonable subset of $\mathcal{P}(S)$ if necessary.) In other words, every set A is endowed with a number $p(A)$, its probability — mass, size, area, volume — whatever term seems the be the most suitable for a certain application.

Note that these verbal explanations are still rather abstract and have nothing to do with physical randomness, lack of knowledge, some kind of fuzzyness or observed variability. The axioms just state reasonable properties that may or not be fitting in a practical application:

The first axiom says that $p(A)$ is non-negative, thus it is not suitable for charges, which are positive and negative. Owing to $p(S) = 1$, the total mass of all objects combined is limited to one. However, in a certain field, the total mass could be larger than any finite number. Finally, the third axiom can be interpreted as a conservation law; upon combining two objects, mass neither appears out of nowhere nor does it vanish into thin air. In the realm of politics and coalitions this is a quite dubious assumption, since the power of two parties A, B, considered in isolation, is typically less than or equal to both parties combined, that is, $p(A) + p(B) \leq p(A \cup B)$. Quite the opposite may be the case in chemical reactions, when mass disappears. That is, a reaction product may weigh less than the reactants combined, $p(A \cup B) \leq p(A) + p(B)$.

If the calculus of probabilities is not appropriate for a certain situation, it could be advisable to use a more general or even a totally different formal framework, such as imprecise probabilities (Walley 1991), the theory of belief functions (Shafer 1990), or some uncertainty calculus (Dubois and Prade 2001).

Despite the last sentence, much can be said to support the eminent importance of probability calculus on the theoretical side. First, Cox (1946) demonstrated that the very notion of rationality may be expressed in terms of probability. Second, in a strong sense, the reverse is also true: If one places bets and does not want to be cheated, it is inevitable to abide by Kolmogorov's axioms. Otherwise, inconsistencies arise that may be exploited by an opponent with the help of so-called 'Dutch books' (Ramsey 1926, de Finetti 1937, Earman 1992, Greenland 1998).

Third, probability calculus may be interpreted as a significant extension of classical logic: deductive logic has two truth values; true (1), and false (0). It is a $\{0, 1\}$-logic. Its 'inverse', inductive logic, does not just distinguish between 'right' and 'wrong', day and night so-to-speak, but admits that there are many shades of grey in-between. If you know that all swans are white, you know that every observed swan has to be of that color. However, if you observe (many) white swans your conviction might grow gradually that they are indeed all white. This is a kind of personal belief, which seemingly follows a $[0, 1]$-logic, i.e., you are more or less certain. The Cox-Jaynes argument (Cox

1946, Jaynes 2003) states that *every* allowed extension of classical logic to plausibility theory is isomorphic to probability theory. In other words, a rational agent has no choice, if he abides by classical logic, he also has to be a probabilist (for the derivation see, e.g., Beck (2010), section 2). Hájek (2023) extends this point of view further:

> Utilities (desirabilities) of outcomes, their probabilities, and rational preferences are all intimately linked. The Port Royal Logic (Arnauld 1662) showed how utilities and probabilities together determine rational preferences; de Finetti's betting interpretation derives probabilities from utilities and rational preferences; von Neumann and Morgenstern (1944) derive utilities from probabilities and rational preferences. And most remarkably, Ramsey (1926) (and later, Savage (1954) and Jeffrey (1966) derives both probabilities and utilities from rational preferences alone.

In a sense, that is all there is to probability, numbers, vectors, etc. Instead of asking the rather fruitless question as to what these objects or concepts 'really mean', it pays to simply *apply* them whenever appropriate. That is, to measure, to calculate, to model and to infer with their assistance. On the other hand, any kind of *theory* should have 'beautiful' properties, such as those just mentioned. In other words, and perfectly in line with the structuralist point of view, basic concepts and their connections define abstract formal structures, which are crucial; their parts and the elements they are made of come second, and finally one may think about appropriate interpretations.

3 Distributions

In this vein, specific numbers are of some interest, and so are particular probabilities. For example, numbers such as π and e, or objects such as circles and triangles are important and are thus thoroughly studied. It may also be important to know how many people are living in India, the size of Canada, or the age of the oldest mammal. The same with specific probabilities. What is the probability that a patient survives a certain diagnosis (medical statistics), that we detect extraterrestrial life in the next decade (www.seti.org), or of winning a game of chance that was interrupted before its close (Sheffer 2019)? The information that is conveyed be all those numbers is surely helpful.

However, sets of numbers, such as $\mathbb{N}, \mathbb{R}$, $\mathbb{C}$ (i.e.,the set of all natural, real and complex numbers, respectively), and the structures defined on them are actually even more important than any specific number. The same with probability theory and its applications: the significance of a concrete numerical value of some 'random' event pales in comparison to random variables and their distributions.

Distributions are a very natural and useful probabilistic structure, namely a set of events (which are often just numbers) plus a set of corresponding weights. In the simplest case, one may just list a finite number of elementary events (outcomes) and their corresponding probabilities. For instance, a model for rolling a dice one time is the discrete uniform distribution on the numbers $\{1, \ldots, 6\}$, that is,

x_k	1	2	3	4	5	6
p_k	1/6	1/6	1/6	1/6	1/6	1/6

Equivalently, $p(X = x_k) = p(X = k) = 1/6$ for $k = 1, \ldots, 6$, describes the (uncertain) behaviour of the dice. In other words, X is a 'random variable' that models the dice, and the collection of all events (numbers that may occur) plus their probabilities, $(x_k, p_k)_{k=1,\ldots,6}$, say, may denote its distribution in a compact way. Typically, important distributions have a name, which will be capitalized in what follows, with the parameters of the distribution added, if necessary. In the case of the dice, we have a discrete uniform distribution on the set $\{1, \ldots, 6\}$, i.e., a Discrete Uniform$(\{1, \ldots, 6\})$.

It is important to note that one may also proceed from the bottom up. In this perspective, a (single) number is the most elementary mathematical object and the notion of a distribution is a natural generalisation. In particular, it is straightforward to identify a specific number a with the 'one-point distribution' $(a, 1) = \delta_a$. In other words, the number a is endowed with the weight $w = 1$. In general, a distribution is just a set of numbers a_k plus associated weights w_k, and a probability distribution obeys Kolmogorov's axioms, which means that the weights have specific properties.

Countable collections $(x_k, p_k)_{k=1,2,\ldots}$ are called discrete (probability) distributions, and the mapping $k \mapsto p_k$ is called a probability mass function (pmf). If the underlying set S is uncountable, one typically uses a probability density function (pdf), which has analogous properties, i.e., $f(x) \geq 0$, $\int_S f(x)\,\mathrm{dx} = 1$. Here, some regularity condition on f is needed that guarantees that the latter integral exists. (In many cases of practical importance, f is continuous.)

The advantage of the latter derivation is twofold. On the one hand, it shows that the concept of distribution is rooted in mathematics' most basic notion (i.e., number). On the other hand, weights w_k do not need to have anything to do with 'probability', 'chance' or 'randomness' (whatever these terms mean). Given this point of view, 'distribution' and 'probability' are just technical terms (e.g. Jaynes (2003), pp. 41-42) which may possess but need not necessarily have a classic probabilistic meaning: a distribution is just a weighted set of numbers, and if the weights are non-negative and add up to one, they are called probabilities.

3.1 Elementary distributions

An elementary route towards defining distributions with 'nice' properties starts with important (regular) mathematical objects and translates them into probabilistic terms.

Perhaps the most elementary object is the unit square with area $1 \cdot 1 = 1$, which defines the pdf $f(x) = 1$ of a Standard Uniform $U(0, 1)$ on the unit interval. The circle is associated with the number π. Moreover, many non-negative smooth functions have an integral that is related to or even equal to π, for instance

$$\int_{-\infty}^{\infty} \frac{1}{1+t^2}\,\mathrm{dt} = \int_{-\infty}^{\infty} \frac{2}{e^t + e^{-t}}\,\mathrm{dt} = \int_0^1 \frac{1}{\sqrt{t(1-t)}}\,\mathrm{dt} = \int_{-\pi/2}^{\pi/2} (1 + \cos(2t))\,\mathrm{dt} = \pi$$

Therefore, in order to get a probability distribution, it suffices to divide by the right hand side, which gives the pdfs of the Cauchy, the Secanshypberbolicus (on $\mathbb{R}$), the Arcussinus (on the unit interval), and the Raisedcosine distribution (on $(-\pi/2, \pi/2)$), respectively. Most importantly,

$$\int_{-\infty}^{\infty} e^{-t^2/2}\,\mathrm{dt} = \sqrt{2\pi} \quad \text{and} \quad \int_0^{\infty} e^{-t}\,\mathrm{dt} = 1$$

yield the pdfs of a Normal and an Exponential distribution.

An analogous remark holds for series. Suppose the elements of a series are non-negative and sum up to a finite number, such as

$$1 + t + t^2 + t^3 + \ldots = \frac{1}{1-t}$$

if $|t| < 1$. Dividing by the number on the right-hand side gives the geometric distribution $G(t)$, defined by

x_k	0	1	2	3	4	...	$\sum$
p_k	$1-t$	$(1-t)t$	$(1-t)t^2$	$(1-t)t^3$	$(1-t)t^4$	...	1

A more abstract and less ad hoc way is to think of a probability distribution as a certain *additive* decomposition of 1, that is, one always has $\sum p_i = 1$. However, the number 1 may also be decomposed in an interesting *multiplicative* way, in particular $1 = e^{\lambda-\lambda} = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \lambda^k/k!$, which yields the Poisson (λ), i.e.,

$$p_k = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

for $k = 0, 1, 2, \ldots$, where $\lambda > 0$.

Finally, back to geometry. Given a set of points (vectors) $x_1, \ldots, x_n$ in $\mathbb{R}^p$, their convex hull is the set of all points m that can be written as

$$m = w_1 x_1 + \ldots + w_n x_n \tag{1}$$

where all $w_i \geq 0$ and $\sum_{i=1}^{n} w_i = 1$. Although this general geometric construction has nothing to do with probability, formally, if one keeps a set of feasible coefficients $\{w_1, \ldots, w_n\}$ fixed, (x_i, w_i) is a well-defined discrete probability distribution, and m its centre of gravity.

3.2 Derived distributions

Distributions such as the above can be used as starting points for the derivation of other distributions. In general, there are two major strategies to determine a distribution:

- Bottom-up (via construction): Start with elementary building blocks, and combine them into more complex structures, often iteratively.
- Top-down (axiomatic approach): Impose boundary conditions that determine a distribution. Typically, the domain is subscribed, and there are criteria that have to be met or optimized.

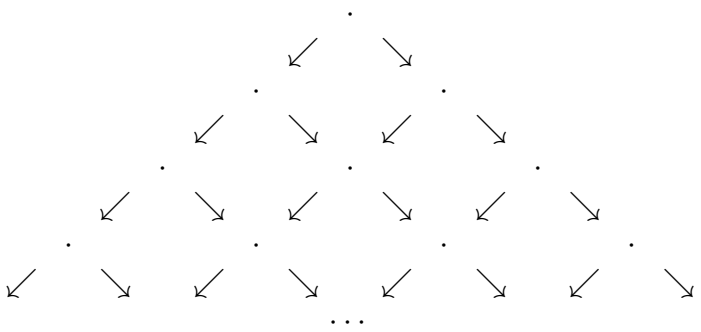
Constructive approach

The first ansatz establishes a strong link between combinatorics and the theory of distributions. To this end, the most important building block is the 'binary bifurcation'

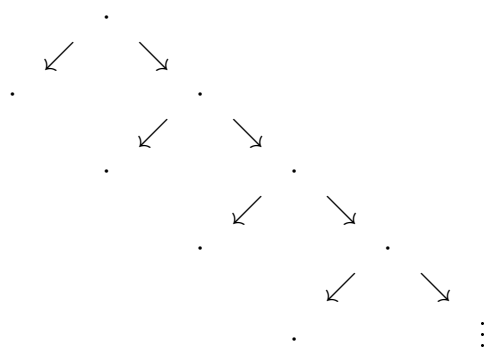
$$\begin{array}{ccc} & \cdot & \\ (1-p) \swarrow & & \searrow p \\ b & & a \end{array}$$

That is, the mass at the top is distributed to points a and b at the bottom. Given unit mass, the bifurcation forwards mass $1-p$ to b and mass p to a, i.e., it creates a two-point distribution. If $a=1$ and $b=0$, we have the so-called Bernoulli distribution $B(p)$.

Iterating this procedure straightforwardly yields Pascal's triangle,



and the binomial distribution, or just the 'Binomial' $B(n,p)$, where $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ for $k=0,\ldots,n$. Completely analogously, the structure



defines the Geometric $G(p)$, where the corresponding random variable is given by $p(X=k)=p^k(1-p)$ for $k=0,1,2,\ldots$

The general idea is easy to grasp: Take an interesting combinatorial structure, which, in total generality, defines a directed a-cyclical graph, and distribute the unit mass at the top along its edges. The corresponding probability distribution is then determined by the masses $p(X=k)$ that arrive at the final edges, for instance the numbers $\{0,\ldots,n\}$ in the case of the Binomial, and $\{0,1,2,\ldots\}$ in the case of the Geometric. Since the number of combinatorial schemes is vast, so is the number of corresponding distributions. The most basic model in this respect seem to be urns that are filled with balls of different colours; the composition of such an urn plus the way the balls are withdrawn then defines a specific distribution.

Essentially, these approaches are constructive and also iterative. For instance, the Panjer (1981) family of distributions is straightforwardly defined with the help of the algorithm

$$p_{j+1}=p_j\cdot\left(c+\frac{d}{j+1}\right),$$

where $c+d\geq 0$, and a suitable probability p_0. In a sense, *internal* forces or a mechanism are shaping the resulting distribution. In applications, scientists working in the field may thus try to model an empirical mechanism rather directly, i.e., with the help of a suitable probabilistic law.

Boundary conditions imply distributions

Maximum Entropy (MaxEnt) distributions are typical examples for the second approach. The n-th moment of a r.v. with pdf $g(x)$ is $\int_S x^n g(x)\,\mathrm{dx}$. Given a certain domain S and some conditions on the moments, these assumptions determine a functional form, i.e., a specific distribution. For instance, if one fixes the domain $(0, \infty)$ and the first moment $E(X) = \int_0^\infty x g(x)\,\mathrm{dx} = \mu$, the exponential distribution $\mathrm{Exp}(\lambda)$ with pdf $f_\lambda(x) = \lambda e^{-\lambda x}$, which has mean $\mu = 1/\lambda$ has maximum entropy. If the domain is the real line and one prescribes the expected value and the variance, i.e. $\int_{-\infty}^\infty x g(x)\,\mathrm{dx} = \mu$, and $\int_{-\infty}^\infty x^2 g(x)\,\mathrm{dx} - \mu^2 = \sigma^2$, the normal distribution with expected value μ and standard deviation $\sigma = \sqrt{\sigma^2}$, named $N(\mu, \sigma)$ is MaxEnt. See Kapur (1993) for an exhaustive collection and systematic derivation of MaxEnt distributions.

The Poisson distribution may also be defined in such an axiomatic way: If events are independent from one another, occur just one at a time, and at a certain constant rate, then the number of occurrences in a fixed time interval follows a Poisson distribution. Such characterizations exist for many distributions (Johnson et al. 1994, 1995, 2005).

In general, sufficient boundary conditions fix a concrete object. For instance, a four-sided figure with equal angles and equal side lengths has to be a square; and a triangle with distinct vertices A, B, C on a circle, where $\overline{AC}$ is a diameter, has to be right-angled. Classic boundary conditions in physics (and beyond) are differential equations, which together with initial conditions determine a function. With respect to probability distributions, one typically has an idea about the domain of the random variable, and about further conditions, such as the moments, some formal relationships or specific characteristics of a field under study.

Graphically speaking, the top-down definition of a distribution is like working with silly putty that assumes a certain shape, given *external* forces. The nature of these forces then determines the concrete form of the distribution in question. For example, the Exponential can be derived as a compromise between the opposing forces of gravity and diffusion, see Mandelbrot (1997), p. 227.

3.3 Connecting distributions

Iteratively constructed discrete distributions almost immediately lead to the idea of summing (and multiplying) random variables. For instance, the Binomial $B(n, p)$ is just the distribution of the sum of n independent Bernoulli(p) random variables. Summing k independent geometrically distributed random variables $X_p \sim G(p)$ gives a Negative Binomial distribution $Y \sim \mathrm{NB}(k, p)$, where $p(Y = x) = \binom{k-1+x}{k-1}(1-p)^k p^x$ for $x = 0, 1, 2, \ldots$

Moreover, just as finite sums $\sum_{i=1}^n x_i$ of numbers have been extended to series $\sum_{i=1}^\infty x_i$, so have sums of random variables, yielding so-called asymptotic distributions. The Poisson, the Exponential and many other distributions may be constructed that way. Perhaps the most important classic example is the De Moivre–Laplace theorem. If $X_n \sim B(n, p)$ then the associated expected value is $\mu_n = E(X_n) = np$, and the corresponding variance $\sigma_n^2 = \sigma^2(X_n) = np(1-p)$. In the limit, the random variable

$$Z_n = \frac{X_n - \mu_n}{\sigma_n}$$

tends toward a Standard Normal $Z \sim N(0,1)$, which has pdf $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$. In other words, $p(a \leq Z \leq b) = \int_a^b \frac{\exp(-t^2/2)}{\sqrt{2\pi}}\,\mathrm{dt}$. This list could be extended tremendously.

New distributions can also be created quite directly, i.e. by means of transformations of known distributions. For example, the tangent of the Uniform is the Cauchy, the latter distribution's logarithm is a Sech, and the inverse of the Uniform $U \sim U(0,1)$, as well as $U/(1-U)$, are Pareto distributions. $\ln U - \ln(1-U)$ has a Logistic distribution, and the sine of a Uniform is an Arcsin. The ratio of two independent standard Normal r.v.'s is a standard Cauchy C r.v. with pdf $1/(\pi(1+x^2))$; and if A has an Arcsin distribution on the unit interval, C^2 and $A/(1-A)$ have the same distribution, etc., etc.

Moreover, there are strong links between discrete and continuous distributions, such as various analytical connections and limit relationships. Discretisation straightforwardly turns a continuous into a discrete distribution. For example, the floor function maps all probability mass on the interval $[j, j+1)$ to the number j. Therefore, if $X \sim$ Exponential(λ), then $Y = floor(X) \sim G(\theta)$, where $\theta = e^{-\lambda}$, since for $j = 0, 1, 2, \ldots$,

$$\begin{aligned} p(Y=j) &= p(floor(X)=j) = p(j \leq X < j+1) = \int_j^{j+1} \lambda e^{-\lambda t}\,\mathrm{dt} \\ &= [1-e^{-\lambda t}]_j^{j+1} = e^{-\lambda j} - e^{-\lambda j}e^{-\lambda} = (e^{-\lambda})^j(1-e^{-\lambda}) = \theta^j(1-\theta) \end{aligned}$$

In a nutshell, without much mathematical effort, an enormously large number of distributions appears. Each of them has individual characteristics, but they are all connected and seem to be organised in a large network-like structure. Just as in the case of physics' 'particle zoo' or chemistry's periodic table, it is not trivial to arrange this vast collection of distributions in a natural way. Saint-Mont (2026) proposes stochastic versions of the Askey scheme, which categorises orthogonal polynomial systems.

Whatever the details, the crucial point is that probability theorists and applied statisticians are working with distributions: they are mainly interested in their properties and relations. By contrast, the associated probabilities are of limited concern, and it is an exception that the latter's meaning needs to be discussed.

3.4 Generalising distributions and numbers

A basic standard density is just an elementary function with corresponding integral, for instance e^{-t} and $\int_0^\infty e^{-t}\,\mathrm{dt} = 1$. Adding parameters gives larger families of distributions, in the case of the Exponential, the most important generalisation is the Gamma(k, λ), which has a shape parameter k and a rate parameter $\lambda > 0$. With the gamma function $\Gamma(k)$, its pdf on the non-negative reals is

$$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}.$$

Another example could be the pdf of a Normal $N(\mu, \sigma)$ with location μ and standard deviation σ which generalises $\exp(-x^2/2)$ and is

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

This elementary way to generalise works for continuous distributions, since a shift (i.e., a change of location) or a scale transformation (i.e., the variance becoming larger or smaller) either does not affect the domain, or affects the domain in a transparent way.

In the case of discrete random variables, which are typically defined on the natural numbers or a subset thereof, the range of transformations that keep the domain fixed (or at least manageable) is rather limited. Although one cannot therefore just add a parameter to a Poisson, say, there is a general technique to introduce further parameters: mixing. That is, some parameter of a discrete distribution - a fixed number - is replaced by a random variable. For instance, if the parameter of the Poisson is replaced by a Gamma(k, ν)-distributed r.v., the result, a 'Poisson-Gamma' distribution is also defined on $\mathbb{N}$, and has two parameters (namely k, and ν). More precisely, the latter distribution is a Negative Binomial$(k, \nu/(\nu+1))$, and is also the sum of k geometrically-distributed independent random variables.

Actually, the basic idea that any number can be treated as a trivial (one-point) distribution, and that - vice versa - probability distributions generalise the concept of number can be applied in (at least) three ways:

(i) If the number λ is a parameter, mixing is tantamount to replacing λ by a random variable. Particular important examples are Poisson-X distributions, when X is some other distribution. However, mixing may also be applied to continuous random variables. For instance, if $Z \sim N(\mu, 1/\sqrt{R})$ and $R \sim \text{Gamma}(a, 1/b)$, then the mixed distribution is a three-parameter Student $t \sim (\mu, \sqrt{b/a}, 2a)$. Since $E(R) = a/b$ and thus $E(1/\sqrt{R}) = \sqrt{b/a}$, one could also say that one starts with $N(\mu, \sqrt{b/a})$, and that the additional variability introduced by the random second parameter (i.e., standard deviation) results in a Student t distribution. (Example to be continued at the end of section 4.3.)

(ii) A single number μ may be replaced by a random variable, in particular a random variable X with expected value $\mu = E(X)$. In error theory, μ represents the true value of some constant, and X might be a process measuring μ. In other words, we observe μ plus/minus some 'noise'.

(iii) A probability p is also just a number. Endowing it with a distribution leads to the vast field of imprecise probability. Once again, numbers give way to distributions (Augustin et al. 2014, Walley 1991).

Even more general, almost any mathematical object can be 'randomized', i.e., may be replaced by a weighted set of similar objects. Prominent examples are random distribution functions (Dubins and Freedman 1963), or the Dirichlet process (Ferguson 1973) that can be interpreted as a distribution over distributions.

4 Statistics: kinds of distribution and probability

Statistics applies distributions (i.e., weighted sets of numbers), probabilities and related concepts to real-world data. Thus, it is no coincidence that the issue of interpretation shows up prominently at this point, and actually statistics distinguishes thoroughly between two kinds of distribution:

- empirical distributions, i.e., the weights are observed quantities (frequencies in particular),
- theoretical distributions, i.e., the weights are probabilities.

For example, consider equation (1). A sample consists of n observations (real numbers) $x_1, \ldots, x_n$, and if $w_i = h_i = n_i/n$ is the observed relative frequency of some observation x_i, the collection (x_i, h_i) is the empirical (or sample) distribution. Thus, the number m becomes the latter's weighted arithmetic mean $\bar{x}$, and the ordinary mean $\sum_{i=1}^{n} x_i/n$ if $n_i = 1$ for all i. If we had to represent the distribution by a single number, m would be a straightforward choice. Philosophically speaking, if the weights are relative frequencies, $w_i = h_i$, we are in the domain of Frequentist statistics, and some frequentists claim that all probabilities have to be interpreted in this way.

Another interpretation / application of equation (1) are n groups (or sub-samples) of size n_i and means $\bar{x}_i$. Then $\sum_{i=1}^{n} h_i \bar{x}_i$ is the overall (total) mean of the sample. Here, although h_i is the relative frequency of sample i with respect to all observations, h_i is rather determined by the experiment(er), and not a (random, fluctuating) number that could be different for the next sample. Philosophically, that might be a difference, in practice, the mathematical formalism is the same and therefore causes no headaches.

Essentially, classical statistics assumes that there is a 'true' theoretical distribution (x_i, p_i), say, and that this distribution produces a sample of n (independent) observations. In other words, the sample's empirical distribution (x_i, h_i) is known to us. Statistics works since its basic theorems (such as Bernoulli (1713), Glivenko (1933), Cantelli (1933)) guarantee that observed quantities converge towards ('estimate') their theoretical counterparts - in particular and most basically $h_i \to p_i$ if $n \to \infty$, but also $m = \sum h_i x_i \to \sum p_i x_i = \mu$. In general, given mild assumptions, the empirical distribution approximates the theoretical distribution on which it is based.

4.1 Distributions and probabilities: a finer distinction

The basic distinction above may be refined, for there are three major interpretations of probability, depending on how the latter are determined:

(i) empirical: the values x_i are observed and the weights are frequencies, $w_i = h_i$.

(ii) logical: the distribution and its corresponding probabilities are determined by a well-defined theoretical setting or structure (such as a 'perfect' coin or dice).

(iii) subjective: potential values and their probabilities are arbitrarily chosen numbers.

The first option is perhaps the most unequivocal, since observed numbers are just as they are — they are 'objective' if you like. Without an explicit reference to empirical experience, there are several kinds of theoretical distribution: one may either look for a well-structured theoretical object such as an urn scheme that defines an associated distribution, which is the second option, or concede that without any boundary conditions other than Kolmogorov's axioms imposed, the choice of distribution is (completely) subjective, which is the third option. Typical (applied) cases lie in-between, i.e., a distribution often stems from rather 'objective' boundary conditions or mechanisms

on the one hand, 'augmented by' softer criteria such as mathematical convenience on the other, and eventually personal judgement.

The second option has served as the blueprint for defining *classical probability*: given an urn with k white balls and m black balls, choosing a ball at random (or just fairly, i.e., without any preference) gives the probability $p = m/(m+k)$ that it will be black. In other words, the proportion $k : m$, which is a property of the urn, translates into a property of an unknown single ball, namely the latter's tendency of being colored in a particular way. One could also argue that the idea of a discrete *uniform distribution* comes first (n possible outcomes, which are on a par), and that the classical definition of probability is a corollary.

The last option (resort), probability as a subjective degree of belief, has sparked much discussion, in particular between the schools of Frequentist and Bayesian statistics. However, in our view, this rather concerns the choice of distribution in a specific setting than the theory of probability as a mode of reasoning.

4.2 Bayesian statistics

Let us make the latter point more precise: In practice, statisticians are rather agnostic with respect to the 'correct' interpretation of probability, and many would also say that they are eclectic when it comes to philosophical schools. However, theorists of all quarters emphasise that probability has nice *formal* properties, in particular that $p(A)$ and $p(B|A) = p(A \cap B)/p(A)$ can be inverted without much ado. That is, if $\bar{A} = S \backslash A$ is the complement of A, then $p(\bar{A}) = 1 - p(A)$, and

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B|A)p(A)}{p(B)} \quad \text{if} \quad p(B) \neq 0. \tag{2}$$

A more subtle look finds that the inversion of conditional probabilities works because of the basic symmetry (Bai et al. 2025)

$$p(A|B)p(B) = p(A \cap B) = p(B \cap A) = p(B|A)p(A)$$

Bayesian statistical inference applies these elementary formulas: If A is a prior hypothetical distribution (hypothesis) and B is the data or evidence, $p(A \cap B) = p(B|A)p(A)$ gives the posterior probability of the hypothesis and the data combined, and $p(A|B)$ is the 'updated' probability of the hypothesis given the data.

Again, single numbers are not really important; instead, it is the collection of all possible outcomes and their corresponding probabilities, the prior and the posterior *distributions*, that really matter. Therefore, the Bayesian inferential mechanism (2) is often written as:

$$\textit{posterior} = \frac{\textit{likelihood}}{\textit{evidence}} \cdot \textit{prior} \tag{3}$$

Traditionally, so-called 'conjugate priors' have the nice property that the posterior belongs to the same distributional family as the prior, and thus the mathematical mechanism becomes easily tractable.

4.3 Precision

In contemporary practice, the prior (distribution) encodes what is known about some parameter prior to the present observations, and the posterior collects all the information after the observations are in. In a nutshell, the more restricted these distributions are, the more we know about a parameter. Fig. 1 depicts this situation: the smaller the variance, the better, and therefore the inverse of the variance has been named precision. In the extreme, i.e., in the best case, we know the parameter, and the variance vanishes.

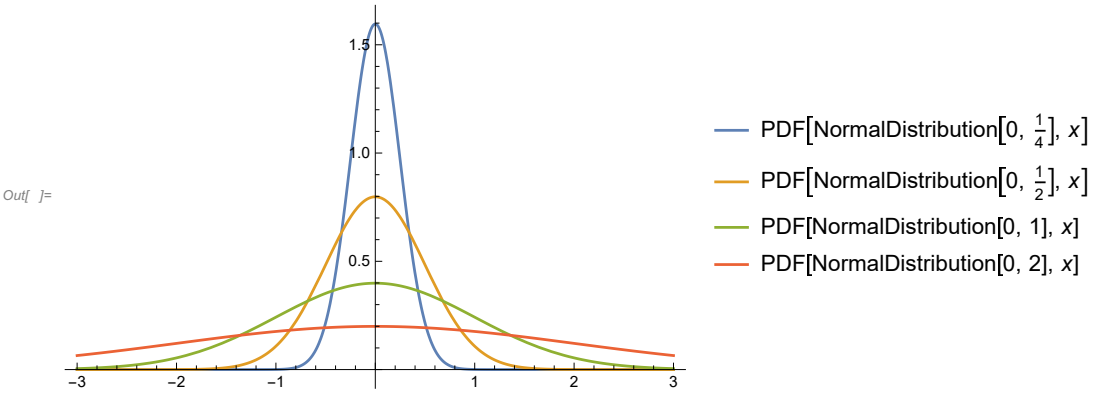


Figure 1: Gaussian (Normal) distributions with standard deviations $1/4, 1/2, 1, 2$

Goodfellow et al. (2016), pp. 334, 336, elaborate:

> Priors can be considered weak or strong depending on how concentrated the probability density in the prior is. A weak prior is a prior distribution with high entropy, such as a Gaussian distribution with high variance. Such a prior allows the data to move the parameters more or less freely. A strong prior has very low entropy, such as a Gaussian distribution with low variance. Such a prior plays a more active role in determining where the parameters end up.
>
> An infinitely strong prior places zero probability on some parameters and says that these parameter values are completely forbidden, regardless of how much support the data give to those values.

A strong prior should be based on facts, i.e., knowledge that has been created before the current study. It is the hallmark of a good theory that it restricts observable values to a small (allowed) area. If such a theory is not available, empirical Bayes methods use the data at hand to estimate the prior (Carlin und Louis 2000). Without empirical data, one may still ask experts to get an informative prior (Mikkola et al. 2024).

Following this train of thought, we should also consider the worst case, i.e., if there is no prior information. In modern parlance, the weakest possible prior is a distribution with maximum entropy (Jaynes 1957a,b). Its predecessor is the 'law of insufficient reason' (Laplace 1812), since in the case of a finite interval $[a, b]$, the 'MaxEnt distribution' there is the continuous uniform distribution. Laplace chose the Uniform "[because of] a lack of sufficient reason for assuming nonuniform priors", and thus the name.

As "the laws of probability are laws of consistency, an extension to partial beliefs of formal logic, the logic of consistency" (Ramsey (1926), p. 182, see section 2), and

since Bayes' formula (2) follows immediately from Kolmogorov's axioms, it is hard to escape the impression that Bayesian inference (3) is almost inevitable. Essentially, it is just 'common sense reduced to calculation' (Laplace 1812), and if combined with empirical data may be used as an 'evidential calculus' (Goodman 1999). At the very least, Bayesian reasoning is straightforward and very general, even if one thinks that the Cox-Jaynes argument goes to far.

Fig. 1 can also be interpreted in a non-Bayesian way that focusses on information. To know a number c is tantamount to knowing the one-point distribution δ_c. Uncertainty, imprecision or a lack of information can thus be modelled straightforwardly with the help of a random variable X whose distribution is centered at c, i.e. $EX = c$, and the spread or variance of X measures the information deficit. The more X deviates from the 'truth', the less we know. Actually, the specific distribution of X even tells us *how* our state of knowledge departs from the exact value.

The classic statistical example is provided by the Normal and Student's t distribution, which we introduced towards the end of section 3.4 (in the following, $n = 2a = 2b$). Suppose $X, X_i \sim N(\mu, \sigma)$ are independent random variables. Then the default estimator of the mean is $\hat{\mu} = \bar{X} = \sum_{i=1}^{n} X_i/n$ and that of the variance is $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \bar{X})/(n-1)$. Now, if both parameters have to be determined from the data, this is less convenient than if only μ has to be estimated. In mathematical terms, the 'price' one has to pay for estimating σ too, should be a less favourable distribution. Actually,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{and} \quad \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Since the variance of the latter distribution is $n/(n-2) > 1$ if $n > 2$, Student's t is in a strict sense worse than the Standard Normal. Consistently, a well-known rule of thumb says that in order to replace t_n by a Normal, at least 30 observations are necessary.

5 Finding common ground in varied interpretations

Just as was to be expected, due to structuralism, the preceding sections have shown that distributions are more important than probabilities. Alternatively, one may start from the prominent idea that statistics is much concerned with the study of variation (Fisher (1925a), p. 1), and that distributions are excellently suited for this purpose. Given this point of view, the various interpretations of probability differ with respect to the *source* of variation:

(i) *Empirical frequencies* are just observed proportions, which obey Kolmogorov's axioms. It is the aim of descriptive statistics to record the variability of the values in a sample; for instance, one may count the number of girls a and boys b in a classroom, and calculate the relative frequency of the girls $a/(a+b)$ and the boys $b/(a+b)$. Formally, this is the same as the classical definition of probability, and putting $p = a/(a+b)$, the distribution is equivalent to that of a coin toss, $B(p)$.

(ii) If the 'classroom' is small, statistics speaks of a sample and sample variability, if it is large, of a population. In the latter case, empirical frequencies describe some population's *natural variability* with respect to certain properties (sex, age,

health, income, wealth, etc.), which is also a main topic of official statistics. The difference between sample and population, i.e., the effective treatment of the additional amount of variability due to the (much) smaller sample size has propelled the evolution of statistics. Succinctly: given a sample, what can be said about a larger population?

(iii) In the case of *logical probability* the variability is produced by an object that admits several outcomes. Thus we are uncertain what is going to happen; however, the composition of the object (the situation if you like), only admits certain possibilities and it is also possible to assign a weight to each outcome. Paradigmatic examples are a dice and a coin that admits heads and tails with probability p and $1-p$. Formally, this is a Bernoulli distribution $B(p)$ with success probability p. Without loss of generality, $p(X=1)=p$ and $p(X=0)=1-p$.

(iv) On the microscopic level, the variability is due to quantum *objects* that are inherently indeterministic. Thus Popper (1957) proposed that these objects have a certain (ontic) *propensity* to exhibit certain kinds of behaviour. For instance, within a second, a radioactive atom may decay or not, and p is the probability that a decay occurs. Again, the model for the quantum object is $B(p)$, and Schrödinger's famous cat enlarges this distribution to the macroscopic world. (As is to be expected, the distribution is preserved as long as there is no interaction of the quantum object with its environment. Otherwise, we either observe 0 (dead) or 1 (alive).)

(v) More generally speaking, (some) laws of nature may be probabilistic and imply the *chance* of seeing a particular event. *Best-systems interpretations* following Lewis (1986, 1994) posit that 'the chances are the probabilities that are determined by these probabilistic laws' (Hájek (2023). Therefore, this kind of chance has been named 'fundamental *dynamical* chance' (Schwarz (2016), my emphasis). Loewer (2004) extends this point of view to initial conditions, with an initial condition thus turning into a stochastic distribution. He gives two concrete and important examples:

> By adding the micro-canonical distribution to Newtonian laws the resulting system (and the proposition that the entropy in the distant past was much lower than currently) entails all of statistical mechanics. By adding the quantum equilibrium distribution ($P(x)=\Psi^2$) to the Bohmian dynamical laws the resulting system entails standard quantum mechanics.

In the classic Dürr et al. (1992), the authors elaborated the latter example and introduced the fitting term *absolute uncertainty*.

(vi) The classical way to make a *subjective probability* precise is to offer a bet. If you are willing to bet 4 dollars against 1 dollar from a bookmaker, you assume that your team has a 4:1 chance of winning, i.e. your probability is 4/5 that your team will win. Although it is difficult to pinpoint the object that produces the distribution in this case, it is obvious that future possible events are determined by the situation (a game may end with a success, a loss, or a draw), and that the associated weights are the result of a thought process, combining former experience with up-to-data sports news and personal intuition. This decision

process might be complex and in large parts implicit, nevertheless, its result is an explicit probability distribution.

(vii) *Algorithmic probability*, largely neglected by philosophers,[1] formalizes Occam's idea (razor) that simple objects should be more probable than complex objects (Li und Vitányi (2008), p. 272). To this end, suppose that $\mathbf{x} = x_1, x_2, \ldots, x_n$ is a series of zeros and ones, which is interpreted as the output of a computer (typically named 'universal Turing machine' U in this context). $K(\mathbf{x})$, its so-called Kolmogorov complexity, is the length of the smallest computer program $\mathbf{y}$, that if fed into U, produces $\mathbf{x}$. U reads programs from left to right and the crucial condition is that no program $\mathbf{y}$ is allowed to occur as a 'prefix' of any other program $\mathbf{y}^*$. Then the algorithmic probability of $\mathbf{x}$ is $p(\mathbf{x}) = 2^{-K(\mathbf{x})}$, which is the probability that a fair coin produces a specific sequence of length $K(\mathbf{x})$. (Note that because of the last sentence algorithmic probability may be interpreted as a specific logical probability.)

Any kind of regularity (aka lack of variability) may be used to compress a given sequence. Consistently, sequences that are virtually incompressible have no intrinsic regularity, and may thus be called random. A major result of the theory of Kolmogorov complexity is that there are two kinds of sequence: most sequences are random and only a few can be described by a short program, i.e., they follow a simple rule (ibid., chaps. 3, 4, and p. 224 in particular).

(viii) *Universal probability* is closely related (ibid., pp. 272f). That is, if $|\mathbf{y}|$ denotes the length of some program $\mathbf{y}$ and $U(\mathbf{y}) = \mathbf{x}$ says that the latter implies the output (data) $\mathbf{x}$, then $q(\mathbf{x}) = \sum_{U(\mathbf{y})=\mathbf{x}} 2^{-|y|}$ is the universal probability of $\mathbf{x}$. Note that $q(\mathbf{x})$ is large, if there is at least one short program $\mathbf{y}$ that produces $\mathbf{x}$. Basically, algorithmic and universal probability translate the variability in a string $\mathbf{x}$ into its description length (Rissanen 1983).

More often, one speaks of the universal *distribution*, since thus each sequence $\mathbf{x}$ can be assigned a (prior) probability. (This idea goes back to Carnap, see section 4.3 of the article on 'information' in the Standard Encyclopedia of Philosophy (summer 2024 ed.).)

In a nutshell, the 'correct' interpretation of a probability statement depends on the situation at hand. This is why the semantics of the concept of probability are discussed in philosophy *and* statistics. At times, a probability may exist in the real world, however, given other circumstances, it might be nothing more than a personal belief (de Finetti 1974). In a sense, the former interpretation is 'stronger' (and thus 'better') than the latter, and therefore propensities and frequencies are in general preferred to more subjective notions of probability.

However, and much more importantly, just as in the classical case of numbers this kind of diversity does not come as a surprise, and despite the long-standing controversy between Frequentist and Bayesian statistics actually is an advantage, since distributions and probabilities can therefore be applied to many fields and settings. As already mentioned at the beginning, the same holds for the formal definition of information, and this list may be extended to almost any concept in all branches of mathematics, such as algebra, calculus or geometry.

[1]For instance, this kind of probability is not mentioned in the article on the 'interpretation of probability' in the Stanford Encyclopedia of Philosophy (summer 2024 ed.).

Consistently, this article bears on the philosophy of mathematics. Hilbert pointed out that formal relationships define mathematical objects and structures, i.e. that syntax is crucial. On the other hand, semantics is not just secondary, but also depends on the field of application or even the specific situation. Thus it is to be expected that there are numerous and divergent interpretations of the concept of probability, which turns out to be correct, and these likewise favor some kind of *algebraic* structuralism in Hilbert's tradition. That is, 'mathematics does not have a single well-defined content, and that, to the extent that it contains assertions at all, these are always relative to a particular theory, framework, language, or structure'. [Moreover,] 'if mathematics is said to have content at all, it is only in a local sense' (see Burke (2015), p. 181).

Numbers quantify some object, process or phenomenon. In the case of probabilities, there is always some kind of variability, imprecision or uncertainty, a certain lack of information if you like. The latter may be of a principled nature or due to our imperfect state of knowledge — aleatory or epistemic in philosophical terms. In particular, there might be a 'random' mechanism in the wings, or there may just not be an obvious pattern in the data.

The 'pseudo-random' numbers produced by so-called 'random number generators' are actually periodic, but what about the decimal places of π (and other mathematical constants, see Bailey et al. (1997), Podnieks (2014), Roba and Podnieks (2025), and Li und Vitányi (2008), pp. 224-226, in particular), or the location of the non-trivial zeros of $\zeta(s)$? (See Katz and Sarnak (1999), Iwaniec and Kowalski (2004), chap. 25, and Conrey (2019), fig. 2.)

The received view that a random sequence is generated by a random experiment is too coarse to be of much help here. However, in contemporary terms, a random sequence is very complex, hardly compressible, and not computable — and these terms may be used almost interchangeably. Moreover, since there are several grades of 'computability', it is possible to distinguish between several degrees of randomness (see Li und Vitányi (2008), pp. 232-237).[2]

Amazingly, it turns out that the lowest grade of randomness is tantamount to 'lower semicomputability'. (ibid., p. 236; one might also consider complexity oscillations, pp. 232ff.) This major result of the modern formal treatment coincides with the observation that 'randomness can be considered an extreme form of chaos' (Elston und Glasbey (1990), p. 340). Since random and algorithmically created sequences are connected, it is virtually impossible to discriminate between 'genuine randomness' on the one hand, and fully-developed 'deterministic chaos' on the other.

In classical terms, if the events A, B are stochastically independent, there is no flow of information between them; knowing A does not tell anything about B, and vice versa. This corresponds to the view of Werndl (2009), pp. 213f, 217, who explains why, in the limit, chaos coincides with random behavior: 'chaos can be defined in terms of mixing ..., mixing goes along with loss of information, [such that] all sufficiently past events are approximately probabilistically irrelevant.' In other words, mixing weakens the link between past and present events, and chaos finally cuts the connection.

[2]In the finite case this is rather obvious, 'because it would be clearly unreasonable to say that a sequence x of length n is random and to say that a sequence y obtained by flipping the first bit 1 in x is nonrandom.' (ibid., p. 166)

6 Conclusion

Questions such as those just considered might be named fundamental and are worth studying. But perhaps we are also putting the cart before the horse. That is, as long as the theories of random variables and their distributions, probability, information and complexity, are able to deal with data in a productive way, it may be rather irrelevant if a chaotic or a stochastic process produced the data, if a certain probability actually exists (or not), or why we always seem to encounter a mixture of signal and noise. In this vein, Li und Vitányi (2008), p. 166, write pragmatically:

> What we can do is to express the degree of incompressibility of a finite sequence in the form of its Kolmogorov complexity, and then analyze the statistical properties of the sequence—for example, the numbers of 0's and 1's in it.

The search for 'true' randomness might be just as pointless (or even misleading), as trying to draw a sharp line between chaos and chance. But no matter as to what the source of the uncertainty is, as long as it is possible to say what may happen, i.e., to specify all potential outcomes (events) and to attach a weight (i.e., a non-negative number) to each of them, the result is a (probability) distribution. There may be some debate with respect to the events and the weights, i.e., to the exact shape of the distribution. However, from there on, Kolomogorov's axioms and their implications allow to reduce rational reasoning to calculation.

The theory of probability, and distributions in particular, are tools that capture uncertainty in all its guises — be it a certain lack of information, 'natural' variability in a population, sampling or measurement error, the inherent propensity of an object, etc., etc. Thus, it becomes possible to argue with degrees of indeterminacy in a precise mathematical way. This turns out to be much more useful than to ask fundamental questions such as what randomness 'really' is, or which of the many interpretations of probability is the best (or correct). At times, there may be a (random, chaotic, simple) mechanism that generated the data at hand, and in some situation a certain interpretation of probability might stick out.

However, be this as it may, typically a researcher hardly considers such 'fundamental' issues. Quite the opposite: what he really cares about are distributional properties, such as the functional form (do we have a Normal, approximately?), the location, skewness or variance of the data, and if his model is adequate, i.e., if it captures crucial properties of a phenomenon under study. In other words, practical work is also, typically, pragmatic; since the calculus of probabilities is very useful in quantifying and handling volatility - wherever the latter may come from - it has been elaborated, extended, and applied extensively.

Finally, it may be stressed once again that there is a contrast between the importance of distributions and probabilities. Whereas the former are pivotal, the latter are rather auxiliary. For example, in statistics, probabilities are often just a(nother) parameter that may be estimated from the data: there is no substantial difference between estimating the rate parameter of a Poisson(λ) and the success probability of a Bernoulli(p), say. Upon focussing on the latter concept and its interpretation, philosophers seem to have been missing a large part of the story, and the leading role played by distributions, in particular. One might even say that, probably, they have been barking up the wrong tree.

References

Arnauld, A. (1662). Logic, or, The Art of Thinking ("The Port Royal Logic"), translatedy by J. Dickoff and P. James, Indianapolis: Bobbs-Merrill, 1964

Augustin, T.; Coolen, F.P.A.; De Cooman, G.; and M.C.M. and Troffaes (2014). Introduction to imprecise probabilities. *John Wiley & Sons.*

Bai, G.; Buscemi, F.; and V. Scarani (2025). Quantum Bayes' Rule and Petz Transpose Map from the Minimum Change Principle. *Physical Review Letters* **135**, 090203.

Bailey, D.H.; Borwein, P.B.; and S. Plouffe (1997). On the Rapid Computation of Various Polylogarithmic Constants. Mathematics of Computation. 66(218), 903–913.

Beck, J.L. (2010). Bayesian system identification based on probability logic. Struct. Control Health Monit. **17**, 825–847.

Bernoulli, J. (1713). Ars conjectandi.

Burke, M. (2015). Frege, Hilbert, and Structuralism. PhD thesis, Univ. of Ottawa. http://dx.doi.org/10.20381/ruor-2700

Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. Giorn. Ist. Ital. Attuari 4: 92–99.

Carlin, B.P.; and Lois, T.A. (2000). Bayes and Empirical Bayes Methods for Data Analysis. (2nd ed.) *Chapman & Hall/CRC, Boca Raton, FL.*

Conrey, B. (2019). Riemann's hypothesis. https://aimath.org/ kaur/publications/90.pdf

Cox, R.T. (1946). Probability, Frequency, and Reasonable Expectation. *American J. of Physics* **14**, 1-13.

Dubins, L.E.; and D.A. Freedman (1963). Random distribution functions. *Bull. American Math. Soc.* **69**, 548-551.

Dubois, D.; and H. Prade (2001). Possibility Theory, Probability Theory and Multiple-Valued Logics: A Clarification. Ann. Math. Artif. Intell. 32, 35-66.

Dürr, D., Goldstein, S., and N. Zanghí (1992). Quantum equilibrium and the origin of absolute uncertainty. J Stat Phys 67, 843–907.

Earman, J. (1992). Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. *The MIT Press, Cambridge, Mass.*

Elston, D.A.; and Glasbey, C.A. (1990). Comment on Bartlett, M.S. (1990). Chance or Chaos? *J. of the Royal Statistical Society, Ser. A* **153(3)**, 321-347.

Ferguson, T. (1973). Ferguson, T. (1973). Bayesian analysis of some nonparametric problems. Annals of Statistics **1(2)**, 209–230.

de Finetti, B. (1937). La Prévision: ses Lois Logiques, ses Sources Sujectives. *Ann. Inst. H. Poincaré* **7**, 1-68. English translation: H. E. Kyburg Jr., reprinted in Kotz und Johnson (1993: Vol. I, 134-174).

de Finetti, B. (1974), Theory of Probability, Vol. 1. *John Wiley and Sons, New York.*

Fisher, R.A. (1925). Statistical Methods for Research Workers (14th ed. 1973). *Macmillan, New York.*

Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probabilità. Giorn. Ist. Ital. Attuari 4: 421–424.

Goodfellow, I.; Bengio, Y.; and A. Courville (2016). Deep Learning. *The MIT Press, Cambridge, MA, and London.*

Goodman, S.N. (1999). Toward Evidence-Based Medical Statistics. 2: The Bayes Factor. *Annals Intern Med.* **130**, 1005-1013.

Greenland, S. (1998). Probability Logic and Probabilistic Induction. *Epidemiology* **9(3)**, 322-332.

Hájek, A. (2023). Interpretations of Probability. The Stanford Encyclopedia of Philosophy (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/win2023/entries/probability-interpret/>.

Iwaniec, H.; and E. Kowalski (2004). Analytic Number Theory. Colloquium Publ. 53. *American Mathematical Society.*

Jaynes, E.T. (1957a). Information theory and statistical mechanics (part I) *Physical Review of Physics* **106(4)**, 620-630.

Jaynes, E.T. (1957b). Information theory and statistical mechanics (part II) *Physical Review of Physics* **108(2)**, 171-190.

Jaynes, E.T. (2003). Probability Theory. The Logic of Science.

Jeffrey, R. (1966). The Logic of Decision, Chicago: University of Chicago Press.

Johnson, N.L.; Kemp, A.W.; and Kotz, S. (2005). Univariate Discrete Distributions. (3rd ed.) *Wiley.*

Johnson, N.L.; Kotz, S.; and Balakrishnan, N. (1994). Continuous Univariate Distributions, vol. 1. (2nd ed.) *Wiley.*

Johnson, N.L.; Kotz, S.; and Balakrishnan, N. (1995). Continuous Univariate Distributions, vol. 2. (2nd ed.) *Wiley.*

Kapur, J. N. (1993). Maximum-Entropy Models in Science and Engineering (revised edition) *Wiley Eastern Limited*, New Delhi, India.

Katz, N.M.; and P. Sarnak (1999). Zeroes of zeta functions and symmetry. Bull. Amer. Math. Soc. 36, 1-26.

Kolmogorov, A.N. (1933). Grundbegriffe der Wahrscheinlichkeitsrechnung. *Springer, Berlin.*

Kotz, S.; and Johnson, N.L. (1993). Breakthroughs in Statistics. Bd. I: Foundations and Basic Theory. Bd. II: Methodology and Distribution. *Spinger, New York.*

Laplace, P.-S. (1812). Théorie Analytique des Probabilités. *Courcier Imprimeur, Paris.*

Lewis, D. (1986). Philosophical papers, vol. II. Oxford Univ. Press, Oxford.

Lewis, D. (1994). Humean Supervenience Debugged. Mind **103**, 473–490.

Li, M.; and Vitányi, P. (2008). An Introduction to Kolmogorov Complexity and its Applications. (3rd ed.) *Springer, New York.*

Loewer, B. (2004). David Lewis's Humean Theory of Objective Chance. Philosophy of Science **71 (5)**, 1115–1125.

Mandelbrot, B.B. (1997). Fractals and Scaling in Finance. *Springer.*

Mikkola, P.; Osvaldo A. Martin, Suyog Chandramouli, Marcelo Hartmann, Oriol Abril Pla, Owen Thomas, Henri Pesonen, Jukka Corander, Aki Vehtari, Samuel Kaski, Paul-Christian Bürkner, Arto Klami (2024). Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Anal.* **19(4)**, 1129-1161.

Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions. *Astin Bulletin* **12**, 22-26.

Podnieks, K. (2014). Digits of pi: limits to the seeming randomness. arXiv:1411.3911

Popper, K. R., (1957). The Propensity Interpretation of the Calculus of Probability and the Quantum Theory, in S. Körner (ed.), The Colston Papers, 9: 65–70.

Ramsey, F.P. (1926). Truth and Probability. In: Ramsey (1931), The Foundations of Mathematics and other Logical Essays, ch. VII (pp. 156-198), edited by Braithwaithe, R.B. *Kegan, Paul, Trench, Trubner & Co., London.*

Resnik, M.D. (1997). Mathematics as a Science of Patterns. *Clarendon Press, Oxford.*

Rissanen, J. (1983). A Universal Prior for Integers and Estimation by Minimum Desription Length. *Annals of Statistics* **11(2)**, 416-431.

Roba, P.N.; and K. Podnieks (2025). Digits of pi: limits to the seeming randomness II. arXiv:2504.10394

Saint-Mont, U. (2026). Information is Key. *ISBN 979-8-249695-18-7.*

Savage, L. J. (1954). The Foundations of Statistics, John Wiley.

Schwarz, W. (2016). Best system approaches to chance. The Oxford Handbook of Proabability and Philosophy, chap. 20 (pp. 423-439), edited by Hájek, A., and C. Hitchcock. *Oxford University Press, Oxford.*

Shafer, G. (1990). Shafer, G. (1990). Perspectives on the theory and practice of belief functions. International Journal of Approximate Reasoning 3, 1-40.

Shapiro, S. (1997). Philosophy of Mathematics. Structure and Ontology. *Oxford University Press, Oxford.*

Sheffer, G. (2019). Pascal's and Huygens's game-theoretic foundations for probability. Sartoniana 32, 117-145.

von Neumann, J., and O. Morgenstern (1944). Theory of Games and Economic Behavior, Princeton: Princeton University Press.

Walley, P. (1991). Statistical Reasoning with Imprecise Probabilities. *Chapman and Hall, London.*

Werndl, C. (2009). What are the New Implications of Chaos for Unpredictability? *Brit. J. Phil. Sci.* **60**, 195-220.