Contents

1 Introduction

The most common classification of probabilities distinguishes between objective, frequentist probability and subjective, Bayesian probability. As practical as this naive distinction has proven to be, it is worthwhile to take a closer look at the different facets of probability in order to better understand its role in statistics and across different schools of inference.

The historical development of probability theory[1] has been far from smooth. Compared with other mathematical and philosophical disciplines it seems as, for some reason, both mathematicians and philosophers had remarkable struggles to formalize and to deal with probability. In 1929 the British mathematician Bertrand Russell stated "*Probability is the most important concept in modern science, expecially as nobody has the slightest notion what it means.*" [5].

Though probability had practical and theoretical relevance for a long time, there is no calculus of probability before the seventeenth century. Until then, probability was handled qualitatively and mainly applied to propositions [5]. The classical definition of probability wasn't introduced until the early 18th century[2] by Jacob Bernoulli and Abraham De Moivre. At this stage, closely related to gamble settings, probability is seen as the fraction of the total number of possibilities in which a event of question occurs. In the following 200 years many attempts were made to extend the classical framework. In the early 19th century, attempts were made to develop a geometric foundation and with the invention of measure theory some mathematicians saw a strong connection to probability calculus. The mathematics of the 20th century was marked by a strong movement toward axiomatization, heavily influenced by David Hilbert and his 1921 proposal, now known as Hilbert's Program [7]. In line with this movement 1933 Andrei Kolmogorov published *Grundbegriffe der Wahrscheinlichkeitsrechnung* which set the foundation of modern probability calculus.

> "The purpose of this monograph is to give an axiomatic foundation for the theory of probability. The author set himself the task of putting in their natural place, among the general notions of modern mathematics, the basic concepts of probability theory—concepts which until recently were considered to be quite peculiar." [1]

Although his axiomatic approach aimed to be, as a mathematical theory, free from interpretation, he saw probability from a frequentistic perspective. In his chapter about elementary theory of probability, where he discussed probability in a finite setting, he added the section "The Relation to Experimental Data" where he briefly described how the theory of probability is applied to the actual world of experiments:

1) "There is assumed a complex of conditions, $\mathfrak{G}$, which allows of any number of repetitions."
2) "We study a definite set of events which could take place as a result of the establishment of the conditions $\mathfrak{G}$. In individual cases where the conditions are realized, the events occur, generally, in different ways. Let E be the set of all possible variants $\xi_1, \xi_2, ...$ of the outcome of the given events. Some of these variants might in general not occur, We include in the set E all the variants which we regard *a priori* as possible."
3) "If the variant of the events which has actually occurred upon realization of conditions $\mathfrak{G}$ belongs to the set A (defined in any way), then we say that the event A has taken place."

A different perspective on probability was offered by the German mathematician and philosopher Rudolf Carnap:

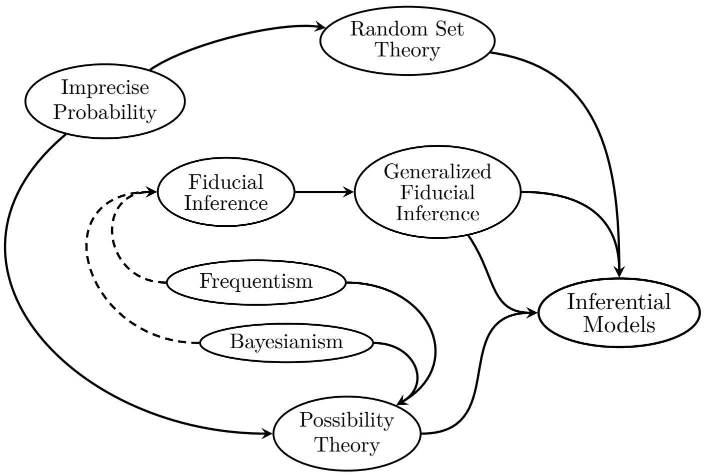
> "The various theories of probability are attempts at an explication of what is regarded as the prescientific concept of probability. In fact, however, there are two fundamentally different concepts for which the term 'probability' is in general use. The two concepts are as follows, here distinguished by subscripts.

[1] There is, both in English and German usage, some confusion about the terminology for probability calculus, probability theory, and stochastics. I use probability calculus when referring to the mathematical aspects, and the more general term probability theory when addressing probability from, for example, a philosophical perspective. I will avoid the term stochastics altogether.

[2]

(1) Probability$_1$ is the degree of confirmation of a hypothesis h with respect to an evidence statement e, e.g., an observational report. This is a logical, semantical concept. A sentence about this concept is based, not on observation of facts, but on logical analysis; if it is true, it is L-true (analytic).

(2) Probability$_2$ is the relative frequency (in the long run) of one property of events or things with respect to another. A sentence about is concept is factual, empirical." [3]



2 Different types of probability

2.1 Probability measure

Definition 2.1 (Sigma-Algebra)**.** Let Ω be a set and $\mathcal{P}(\Omega)$ denote the power set over Ω. Then $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called a σ-Algebra if

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$
3. $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F} \Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{F}$

holds. The tupel $(\Omega, \mathcal{F})$ is then called a measurable space.

Definition 2.2 (Measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A set-function $\mu : \mathcal{F} \to \mathbb{R}$ is called a measure if it fulfils

1. $\mu(\emptyset) = 0$
2. $\mu(A) \geq 0 \quad \forall\, A \in \mathcal{F} \quad$ (Non-negativity)
3. For any sequences of pairwise disjunct sets $A_i \in \mathcal{F}$, $i \geq 1$:
   $\mu\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mu(A_i) \quad$ (σ-additivity)

Measures that are normed to $\mu(\Omega) = 1$ (normalization property) are called *probability measure* and will be noted with $\mathbb{P}$. The *measure space* $(\Omega, \mathcal{F}, \mathbb{P})$ is then called a probability space where Ω is called a *sample space* and $\mathcal{F}$ a *event space*.

Definition 2.3 (Random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\Omega', \mathcal{F}')$ a measurable space. A function $X : \Omega \to \Omega'$ is a *random variable* if

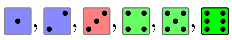
$$\{\omega \,|\, \omega \in \Omega \,\wedge\, X(w) \in A\} = X^{-1}(A) \in \mathcal{F} \quad \forall A \in \mathcal{F}' \quad (\mathcal{F} - \mathcal{F}' - \text{measurability})$$

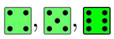
is satisfied. Using this definition the measurable space $(\Omega', \mathcal{F}')$ can be extended to a probability space through $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\Omega', \mathcal{F}', \mathbb{P}_X)$ where the according *push-forward-measure* $\mathbb{P}_X$ is defined through

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A) = \mathbb{P}(\{\omega | \omega \in \Omega \wedge X(\omega) \in A\}) = \mathbb{P}(\underbrace{X^{-1}(A)}_{\in \mathcal{F}}) \in [0, 1] \quad \forall A \in \mathcal{F}'.$$

Definition 2.4 (Information content)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. *Information* can be characterized as a subset $\mathcal{A} \subseteq \mathcal{F}$ of events we are capable of evaluating as having occurred or not.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\Omega', \mathcal{F}')$ a measurable space and X an *observable* $\mathcal{F} - \mathcal{F}' -$measurable random variable. Then $\mathcal{A}_X = \{X^{-1}(A) \mid A \in \mathcal{F}'\} \subseteq \mathcal{F}$ represents through X *observable* information.

Example 2.1. Consider a dice that is numbered and has coloured faces like this: $\Omega = \Big\{$  $\Big\}$.

If we are interested in the face values we can simply map the face values to the according numbers $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$. For each realization of X we can exactly know which events did occur or not. If we instead map to colours by $Y : \Omega \to \{blue, red, green\}$ some information is lost. If e.g. a green face was rolled, we can not know which of the green sides $\Big\{$  $\Big\}$ was rolled.

- $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$
$\mathcal{A}_X \subseteq \mathcal{A}$

- $Y : \Omega \to \{blue, red, green\}$
$\mathcal{A}_Y \subset \mathcal{A}$

3 False confidence and validity critereon

4 Imprecise probability theory

The term *imprecise probability* does not refer to a particular theory, but rather to a collection of different approaches that share a common feature of imprecision. In contrast to probability measures, imprecise probabilities are not additive. Some approaches are can be seen as a extension of the measure theoretic based probability theory, like *random sets*, while others might have a non-measure-theoretic foundation. In this chapter I will give deliver some motivation for imprecise probabilities in general and introduce some popular concepts of imprecise probabilities.

Consider a variation of the prison dilemma used in game theory, but from our perspective as a policeman. Assume there a two suspects of a crime, person A and B, where it is known, that exactly one of them must be guilty. If A is guilty B can not be guilty and otherwise. We, as a policeman, interrogate A without knowing anything about person B. After the interrogation a college asks about our opinion, how probable it is, that person A is guilty.

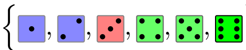
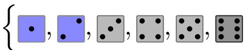
> College: *What do you think is the probability that person A is guilty?*
>
> Policeman: *I think the probability is around 20%.*
>
> College: *I see. So you assume the probability that person B is guilty must be around 80%?*

Although being intuitively clear how our college came to this conclusion, his response might seem puzzling. Our college implicitly chose $\Omega = \{\mathbf{A} \text{ is guilty}, \mathbf{B} \text{ is guilty}\}$ as sample space and $\mathcal{F} = \{\emptyset, \{\mathbf{A} \text{ is guilty}\}, \{\mathbf{B} \text{ is guilty}\}, \{\mathbf{A} \text{ is guilty}, \mathbf{B} \text{ is guilty}\}\}$ as event space. Simple probability calculus shows that $P(\{A \text{ is guilty}\}) = 0.2$ implies $P(\{B \text{ is guilty}\}) = 0.8$.

$$
\begin{aligned}
1 &= P(\Omega) && \text{(normalization)}\\
&= P(\{\textbf{A} \text{ is guilty}, \textbf{B} \text{ is guilty}\})\\
&= P(\{\textbf{A} \text{ is guilty}\} \,\dot\cup\, \{\textbf{B} \text{ is guilty}\})\\
&= P(\{\textbf{A} \text{ is guilty}\}) + P(\{\textbf{B} \text{ is guilty}\}) && \text{(additivity)}\\
&\Leftrightarrow\\
P(\{\textbf{B} \text{ is guilty}\}) &= 1 - P(\{\textbf{A} \text{ is guilty}\})
\end{aligned}
$$

Depending on the way someone interprets the statement "I think the probability is around 20%", this result might be more or less troubling. It is thinkable that a very experienced policeman had encountered many sufficiently similar situations to determine a relative frequency based solely on the information he gathered from interrogating A, without any knowledge of $B. Based on such a reading, the above conclusion seems quite reasonable - even though such a perspective might be rather uncommon in this context. Rather, one would interpret such an educated guess as a quantification of the strength of belief. Such an educated guess would more likely be meant as a quantification of the strength of belief.

4.1 Random sets

$\Omega = \{$  $\} \rightarrow \Omega_{Obs} = \{$  $\}$

Definition 4.1 (Random set)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\Omega', \mathcal{F}')$ a measurable space. A multi-valued mapping $\Gamma : \Omega \to \mathcal{P}(\Omega')$. For $A \in \mathcal{F}'$ the *upper inverse* is given by $\Gamma^*(A) = \{\omega \,|\, \omega \in \Omega,\, \Gamma(\omega) \cap A \neq \emptyset\}$ and the *lower inverse* is $\Gamma_*(A) = \{\omega \,|\, \in \Omega,\, \emptyset \neq \Gamma(\omega) \subseteq A\}$. When $\Gamma^*(A)$ and $\Gamma_*(A)$ belongs to $\mathcal{F}$ for all $A \in \mathcal{F}'$ the multi-valued mapping Γ is said to be strongly measurable and then called a *random set.*

Definition 4.2 (lower and upper probability)**.** Let $\Gamma : \Omega \to \mathcal{P}(\Omega')$ be a random set. For $A \in \mathcal{F}'$ the *upper probability* is defined by $P^*_\Gamma(A) = \dfrac{\mathbb{P}(\Gamma^*(A))}{\mathbb{P}(\Gamma^*(\Omega))}$ and the *lower probability* by $P_{*\Gamma}(A) = \frac{\mathbb{P}(\Gamma_*(A))}{\mathbb{P}(\Gamma^*(\Omega))}$.

"*Don't know*" probability $Pl(A) - Bel(A)$

4.2 Possibility measures

4.2.1 Boolean possibility theory

Boolean possibility theory is based on propositional logic where the Pinciple of Bivalence states that every proposition p is either true (1) or false (0). But instead of focusing directly on propositions the focus lies in modelling a rational agents belief about propositions. The current knowledge of an agent is represented by a *belief base* K which contains boolean formulas. K is required to be *consistent* i.e. it must be free of logical contradictions.

If a proposition p, based on K, is logically true, the agent must believe p to be true, written as $N(p) = 1$ and $N(p) = 0$ otherwise.

An agents state of belief is then represented by the pair $(N(p),\, N(\neg p))$ with 3 possible states:

- $(N(p),\, N(\neg p)) = (1,\, 0)$ agent believes p
- $(N(p),\, N(\neg p)) = (0,\, 1)$ agent believes $\neg p$
- $(N(p),\, N(\neg p)) = (0,\, 0)$ agent is completely ignorant about p

$(N(p),\, N(\neg p)) = (1,\, 1)$ is not a possible state since $p \wedge \neg p$ is a contradiction and can not be derived by a consistent belief base. It is important to notice that $N(p) = 0$ does not imply $N(\neg p) = 1$ since an agent is allowed to be fully ignorant. Therefore the question arises if a certain proposition is consistent with K. If p is consistent with K this relation is stated by $\Pi(p) = 1$. The relation between N and Π is given through $\Pi(p) = 1 - N(\neg p)$ and furthermore $N(p \wedge q) = \min\big(N(p), N(q)\big)$ and $\Pi(p \vee q) = \max\big(\Pi(p), \Pi(q)\big)$.

Possibility contour $\pi : \mathcal{X} \to [0, 1]$ with $\sup_x \pi(x) = 1$ Possibility measure $\Pi : \mathcal{P}(\mathcal{X}) \to [0, 1], \; A \mapsto \sup \pi(x), A \subseteq \mathcal{X}$

5 Fiducialism

5.1 Fisher's original fiducial argument

In 1930 Fisher introduced his idea the *fiducial principle* which turned out to be one of his most controversial ideas of his career. He criticised the concept of *inverse probability*, by which he meant Bayes's postulate fundamental to Bayesian inference [2].

> "Inverse probability has, I believe, survived so long in spite of its unsatisfactory basis, because its critics have until recent times put forward nothing to replace it as a rational theory of learning by experience." [4]

He disregarded the subjective nature behind the Bayesian approach and often inherently arbitrary choice of a particular *a priori* distribution for parameters. Therefore he tried to find a more objective alternative that is closer to frequentist probabilities. Although being convinced of the importance of his idea he failed to establish fiducialism. Fisher could not provide a coherent and comprehensive theory for fiducial inference and left behind a strongly limited theory, mostly built around exemplary examples and several changes of mind that lead to some confusion [6].

His proposed example takes a bivariate normal distribution with unknown, fixed correlation ϕ with a sample size of $n = 4$.

Let T be a statistic derived for observable sample correlations r with distribution function $F(r; \phi) = P(T \leq r \mid \Phi = \phi)$.

Fisher reasoned that, under repeated sampling, each possible value of $\phi \in [-1, 1]$ would be associated with a unique value of the $\gamma-$quantile of the sampling distribution.

Therefore by looking up the related e.g. 0.95-quantile to an observed sample correlation there is a corresponding *fiducial* 0.05-value.

Therefore by looking up the related e.g. 0.95-quantile to an observed sample correlation there is a corresponding *fiducial* 0.05-value. Fisher stated this as a relation of the form $P = F(T, \theta)$, where T is a statistic of continuous variation, P the probability, that T is less than any specified value and θ the fixed parameter of question. Therefore by looking up the fiducial 0.05-value for an observed θ we know

5.2 Generalized Fiducial Inference

Definition 5.1 (Data generating equation)**.** A data generating equation (DGE) has the form $\mathbb{X} = G(\theta, U)$ and contains

- observable data $\mathbb{X}$
- an association function G
- a random variable U with known distribution
- a parameter of interest θ.

After the observation this results in $X = G(\theta, U^*)$ with

- observed data X
- an unobserved realization U^*.

This can be reformulated through $X = G(\theta, U^*) \;\Leftrightarrow\; \theta = G^{-1}(X, U^*)$.

The basic idea behind DGEs can be shown with a motivational example.

Example 5.1. Lets have a look at a simple normal distribution $Y \sim \mathcal{N}(\theta,\, \sigma^2)$. We are interested in the unknown parameter θ. Now we can ask the question, which information would be sufficient to know the exact value of θ after observing e.g. X = (1.159) from $\mathcal{N}(2, 1.5^2)$

$$Y \sim \mathcal{N}(\theta,\, \sigma^2) = \theta + \mathcal{N}(0,\, \sigma^2), \quad U \sim \mathcal{N}(0,\, \sigma^2) \Rightarrow \; \theta \;=\; G^{-1}(X,\, U^*) \;=\; X - U^*$$

If the exact value of $U^* = -0.841$ was known, the true value of θ could directly be calculated by

$$\theta = X - U^* = 1.159 - (-0.841) = 2$$

So can we simply observe X and solve for $\theta \;=\; G^{-1}(X, U^*)$? The required information about U^* is typically not available and therefore the direct approach is not feasible.

Bibliography

[1] A. N. Kolmogorov. *Foundations of the Theory of Probability.* eng. 1950. URL: http://archive.org/details/kolmogorov_202112 (visited on 05/12/2025).

[2] John Aldrich. “R.A. Fisher and the making of maximum likelihood 1912-1922”. en. In: *Statistical Science* 12.3 (Sept. 1997). ISSN: 0883-4237. DOI: 10.1214/ss/1030037906. URL: https://projecteuclid.org/journals/statistical-science/volume-12/issue-3/RA-Fisher-and-the-making-of-maximum-likelihood-1912-1922/10.1214/ss/1030037906.full (visited on 05/09/2025).

[3] Rudolf Carnap. *Logical foundations of probability.* eng. Chicago : University of Chicago Press, 1950. URL: http://archive.org/details/logicalfoundatio00carn_0 (visited on 05/13/2025).

[4] R. A. Fisher. “Inverse Probability”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 26.4 (1930), pp. 528–535. DOI: 10.1017/S0305004100016297.

[5] Alan Hájek. “Interpretations of Probability”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta and Uri Nodelman. Winter 2023. Metaphysics Research Lab, Stanford University, 2023.

[6] S L Zabell. “R. A. Fish~r and the Fiducial Argument”. en. In: ().

[7] Richard Zach. “Hilbert’s Program”. In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta and Uri Nodelman. Winter 2023. Metaphysics Research Lab, Stanford University, 2023.