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. Or as the British mathematician Bertrand Russell stated in 1929 "*Probability is the most important concept in modern science, expecially as nobody has the slightest notion what it means.*" [10].

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 [10]. The classical definition of probability wasn't introduced until the early 18th century 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 [17]. In line with this movement 1933 Andrei Kolmogorov published *Grundbegriffe der Wahrscheinlichkeitsrechnung* which set the foundation of modern probability calculus where he wrote in the preface:

> "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." [2] [11, p.15].

Kolmogorov 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."[3]

Some interesting insights on probability and the controversies on different standpoints were 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] The following historical summary is primarily based on [15].
[2] The English translation is taken from [1, p. 15].
[3] Kolmogorov mentioned, that this section is not of interest for reader who are only interested in the purely mathematical develepment of his theory only. The axiomatic framework stands independently of this interpretative view.

(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[4] (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." [5, p.19].

It is worth noting that Carnap also treats probability$_1$ as an objective concept:

> "Deductive logic may be regarded as the theory of the relation of logical consequence, and inductive logic as the theory of another concept which is likewise objective and logical, viz., probability$_1$ or degree of confirmation. That probability$_1$ is an objective concept means this: if a certain probability$_1$ value holds for a certain hypothesis with respect to a certain evidence, then is value is entirely independent of what any person may happen to think about these sentences, just as the relation of logical consequence is independent in this respect." [5, p.43].

The multifaceted nature of probability still makes it hard to formalize and unify into a single coherent theory. Probability is expected to capture classical probability calculus while also being appropriate for modeling existing evidence and, more broadly, information based reasoning under uncertainty. In the context of statistical inference, probability serves dual roles: as a mathematical framework for modeling randomness (frequentist view) and as a measure of belief or uncertainty (Bayesian view). These interpretations lead to different methodologies and philosophical foundations, making unification challenging. Nonetheless, both approaches aim to derive meaningful conclusions from data in the presence of uncertainty. A probabilistic representation of the truth value of assumptions should align with the real world. If inference methods tended to assign high degrees of belief to false assumptions and low degrees of belief to true ones, they would be highly misleading and systematically deceptive. A requirement for "meaningful conclusions" must be that the probabilistic representation is *calibrated* in the sense that false assumptions are typically assigned low degrees of belief, and it is rare for them to receive high degrees of belief, and vice versa for true assumptions [12]. This requirement aligns well with the *Cournot's principle* which was firstly stated by Jacob Bernoulli in 1713 and later further developed by Antoine-Augustin Cournot and can be seen as an early attempt to bridge probability calculus with the real world:

> "An event with very small probability is morally impossible; it will not happen. Equivalently, an event with very high probability is morally certain;" [15].

It also resonates with the weak repeated sampling principle, as stated by Cox and Hinkley

> "The weak version of the repeated sampling principle requires that we should not follow procedures which for some possible parameter values would give, in hypothetical repetitions, misleading conclusions most of the time." [6, p. 45–46].

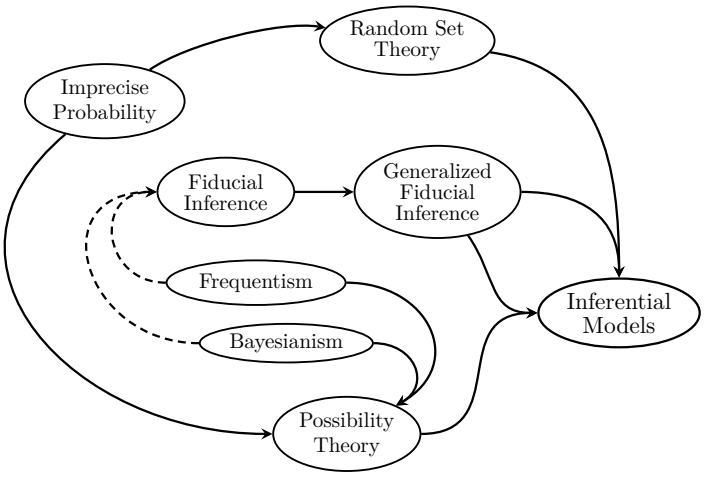
In this sense, *false confidence* should be avoided. Accordingly, the following statement from a paper on the false confidence phenomenon in satellite conjunction analysis, where reliable collision risk assessment is a serious concern, can only be seen as concerning:

> "Every real-world risk analysis problem involves a proposition of interest that is determined by the structure of the problem itself; e.g. 'Will these two satellites collide?'. Just as the practitioner will not seek out propositions strongly affected by false confidence, neither do practitioners have the option of avoiding such propositions when they arise." [3].

Inferential models, as a modern framework for "prior-free posterior probabilistic inference", introduced by Ryan Martin and Chuanhai Liu [13], are capable of offering, under certain conditions, calibration properties which can constrain false confidence.

In the following chapters, I will explain in more detail what is meant by false confidence, how and why it may lead to problems. To motivate this discussion, I will introduce key ideas from imprecise probability theory and fiducialism, and then explore their connection to inferential models.

[4] L-true refers here to "logically true" which means that the truthfullnes is depends on a (formal) logical representation instead of empirical facts.



2 Different types of probability

2.1 Probability measure

In this section, I will revisit some essential concepts from probability calculus in order to later highlight key differences from the established mathematical notion of probability. The underlying structure of the measure-theoretic construction of probability is based on specific types of set systems that exhibit useful structural properties — so-called σ-algebras. In measure theory, the symbol σ typically denotes (at most) countably infinite.

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. $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$
3. $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{F} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$

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

Therefore σ-algebras are non-empty collections of sets that contain the empty set and are closed under complementation and countable unions. Simplified, finite cases can often be seen as special cases of countably infinite ones by extending a finite index set $I = \{1, \dots, k\}$ to a countable one by defining $A_i = \emptyset \quad$ for $i > k$ so that a finite union $\bigcup_{i=1}^{k} A_i$ can be rewritten as a countable union $\bigcup_{i=1}^{\infty} A_i$, with $A_i = \emptyset$ for all $i > k$.

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)

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

Among the unfortunate terminological choices in statistics, random variable is perhaps one of the most misleading since they are neither random, nor variables by nature.

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

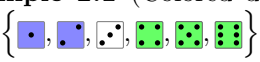
$$\{\omega \,|\, \omega \in \Omega \wedge X(w) \in A\} = X^{-1}(A) \in \mathcal{F} \quad \forall A \in \mathcal{F}' \quad (\mathcal{F} - \mathcal{F}' - \text{measurability})$$

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)**.** 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[9].

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 (Colored dice I)**.** 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 faces 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 $Y : \Omega \to \{blue, white, green\}$ 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 based on the piece of information "it was *green*".

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 generally not additive. Some approaches 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.

Example 4.1 (Prisoners dilemma I)**.** 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 $P(\{B \text{ is guilty}\}) = 0.8$.

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

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 quite uncommon in this context. Rather, one may interpret such an educated guess as a quantification of the strength of belief or a vague quantification of how the currently known evidence supports the assumption of A's guilt. However, it would seem highly unfair to hold a strong bias against B in this situation, or even to infer stronger evidence for B's guilt from weak evidence against A. I want to stress the fact that the additivity property has an important implication. By taking belief mass away from "A is guilty", it must be shifted to "B is guilty", even though this shift is not supported by any information concerning B. It therefore might seem reasonable to consider addressing certain problems using more weakly structured, non-additive alternatives.

4.1 Possibility measures

4.1.1 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 $\Pi(p \vee q) = \max\big(\Pi(p), \Pi(q)\big)$.

Example 4.2 (Prisoners dilemma II)**.** This framework enables us to model the initial situation in two particular cases. In the case where no suspect has yet been interrogated and nothing is known about either suspect (complete ignorance), and in the case where, after the interrogation of one (or possibly both) suspects, the guilt or innocence of a suspect is certain (complete knowledge).

The first case reveals a notable difference from a Bayesian flat prior approach. Instead of assuming a prior guilt (or innocence) of 0.5 for each suspect under ignorance, we can now adopt a more sophisticated perspective. The belief base contains our information that exactly one of the two suspects is guilty.

$N(A \text{ is guilty} \vee B \text{ is guilty}) = 1$. On the other hand there is no reason to believe in the guilt of A or B separately $N(A \text{ is guilty}) = N(B \text{ is guilty}) = 0$. On the other hand there is no reason to doubt that one of the subjects is innocent $\Pi(A \text{ is guilty} \vee B \textit{ is guilty}) = \Pi(A \text{ is guilty}) = \Pi(B \text{ is guilty}) = 1$.

In the second case e.g. for $N(A \text{ is guilty}) = 1$ it directly follows through $\Pi(A)$ is not guilty$) = 1 - N(A \text{ is guilty}) = 0$ and A is guilty $\Leftrightarrow$ B is not guilty that B must be believed to be innocent. The situation where an agent has incomplete knowledge between *complete ignorance* and *complete knowledge* can not be modelled so far.

A thought experiment suggests, without claiming to present a fully developed theory, how we can bridge this logical concept to the more familiar measure theory. Imagine a hypothetical universe consisting of many different worlds. Each world corresponds to a logical state and must be free of logical contradictions. By asking a question "How probable is p?" can thereby be understood as a question of "How probable is it to exist in a world where p is true?". Such probabilistic reasoning then follows roughly the following scheme. We must first consider which type of worlds qualify, so that only those sufficiently similar to our own are taken into account. This is very similar to narrowing down the population in statistical studies or considering theoretical assumptions and experimental conditions in physics. Then we narrow down the logical variables of interest. Only looking at these variables some worlds are indistinguishable and can be organized into sets ω representing such worlds with indistinguishable states. Let Ω denote the set that contains the resulting categories ω and $[p] \subseteq \Omega$ describe the conditions where p is true. We then refer to the conditions $[p]$ as the models of p. A proposition p is believed to be true if $[K] \subseteq [p]$ and believed to be false if $[K] \subseteq [p]^C$. If a proposition allows for at least one setting where it could be true it is considered to be *possible*. If a proposition is true in every setting it is *necessary* to believe in it.

4.1.2 Fuzzy set theory

The link to set-theoretic measure theory permits the incorporation of fuzzy set theory.

Definition 4.1 (Fuzzy set, membership function)**.** Let X be a set. A *fuzzy set* $\tilde{A}$ in X is a collection of ordered pairs of the form

$$\tilde{A} = \big\{(x, \mu_{\tilde{A}}(x)) \,|\, x \in X\big\}.$$

$\mu_{\tilde{A}} : X \to L, \sup(L) < \infty$ is called the membership function (or generalized characteristic function). For $\mu_{\tilde{A}}(x) = 1$ the fuzzy set is *normalized*. For $L = [0, 1]$. L is called a possibility scale.

Definition 4.2 (α-cut)**.** The set of elements that belong to $\tilde{A}$ at least to the degree α is called a *α-cut*:

$$A_\alpha = \big\{x \,|\, x \in X,\, \mu_{\tilde{A}} \geq \alpha\big\}$$

and $A'_\alpha = \big\{x \,|\, x \in X,\, \mu_{\tilde{A}} > \alpha\big\}$ defines a *strong α-cut*.

Here are a few useful basic operations on fuzzy sets $\tilde{A}$, $\tilde{B}$:

- Intersection (corresponding to a logical and): $\mu_{\tilde{A}\cap\tilde{B}} := \min\big(\mu_{\tilde{A}}(X), \mu_{\tilde{B}}(X)\big)\ \forall x \in X$
- Union (corresponding to an exclusive[5] or): $\mu_{\tilde{A}\cup\tilde{B}} := \max\big(\mu_{\tilde{A}}(X), \mu_{\tilde{B}}(X)\big)\ \forall x \in X$
- Complement (corresponding to a negation): $\mu_{\tilde{A}}(X) := 1 - \mu_{\tilde{A}}(X)\ \forall x \in X$

4.1.3 Numeric possibility theory

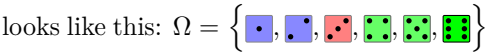
The key idea behind numeric possibility theory is that the current state of an agents incomplete knowledge can be captured by a *possibility distribution* and thus can be seen as an extension of boolean possibility theory.

$\pi : \mathcal{X} \to [0, 1]$ and the consistency assumption, that $\pi(x) = 1$ is fulfilled for at least one $x \in X$. $1 - \pi(x)$ might be interpreted as an agents surprise that X turns out to be x. A possibility distribution could be e.g. derived by intuition, an experts opinion or an assumed distribution. A *possibility measure* is defined as a set function that maps from $\mathcal{P}(\mathcal{X})$ to $[0, 1]$

[5] An inclusive or in $A \vee B$ describes a relation where A, B or A and B together are true. An exclusive or in $A \dot{\vee} B$ describes a relation where only A or B can be true, but not A and B together.

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

4.2 Random sets

Example 4.3 (Coloured dice II)**.** Consider a coloured dice that looks like this: $\Omega =$  and the random variable $Y \to \{blue, red, green\}$.

Given such a dice many people wouldn't struggle at all to precisely observe which colour was rolled while for some people it might be completely. Y is obviously only partially observable for people with red-green colour blindness since it could be look like this: $\Omega =$  $\to$ $\Omega_{Obs} =$.

So the question arises how to deal with such a situation? Assume, that some extra knowledge about that dice is provided: The faces can only be blue, red or green, but not all of those colours must be presented on the dice. Green is the only colour that can also come with a darker shade. The probability of being rolled is the same for all faces.

$$Y(\omega) = \begin{cases} blue & \text{for } \omega = \\ blue & \text{for } \omega = \\ red & \text{for } \omega = \\ green & \text{for } \omega = \\ green & \text{for } \omega = \\ green & \text{for } \omega = \end{cases} \qquad ? = \begin{cases} \{blue\} & \text{for } \omega = \\ \{blue\} & \text{for } \omega = \\ \{green, red\} & \text{for } \omega = \\ \{green, red\} & \text{for } \omega = \\ \{green, red\} & \text{for } \omega = \\ \{green\} & \text{for } \omega = \end{cases}$$

On the left we can see the *original* random variable Y_0 and on the right a failed attempt to write down a random variable.

Definition 4.3 (Strong measurability, random set, derived from [14])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\Omega', \mathcal{F}')$ a measurable space and $\Gamma : \Omega \to \mathcal{P}(\Omega')$ a multi-valued mapping. Given $A \in \mathcal{F}'$, its *upper inverse* is given by $\Gamma^*(A) = \{\omega \,|\, \omega \in \Omega,\ \Gamma(\omega) \cap A \neq \emptyset\}$ and the *lower inverse* is $\Gamma_*(A) = \{\omega \,|\, \omega \in \Omega,\ \emptyset \neq \Gamma(\omega) \subseteq A\}$. The multi-valued mapping Γ is said to be *strongly measurable* when $\Gamma^*(A)$ and $\Gamma_*(A)$ belong to $\mathcal{F}$ for all $A \in \mathcal{F}'$ and then be called a *random set*.

Definition 4.4 (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 $\overline{P}(A) := P^*_\Gamma(A) = \dfrac{\mathbb{P}(\Gamma^*(A))}{\mathbb{P}(\Gamma^*(\Omega))}$ and the *lower probability* by $\underline{P}(A) := P_{*\Gamma}(A) = \dfrac{\mathbb{P}(\Gamma_*(A))}{\mathbb{P}(\Gamma^*(\Omega))}$.

Example 4.4 (Coloured dice II, continued)**.** Here we face the situation that many random variables, $S(\Gamma) := \{Y : \Omega \to \Omega' \text{ is measurable} \,|\, Y(\omega) \in \Gamma(\omega)\, \forall \omega \in \Omega\}$, could potentially suit such an imprecise observation, we simply can not tell which one it exactly is. Consequently we can derive a set of suitable push-forward-measures $P(\Gamma) := \{P_Y \,|\, Y \in S(\Gamma)\}$ and a *credal set* $M(\overline{P}) := \{Q \text{ is a probability} \,|\, Q(A) \leq \overline{P}(A)\, \forall A \in \mathcal{F}'\}$ that defines a class of probabilities that are dominated by the derivable upper probabilities.

$$Y_1(\omega) = \begin{cases} blue & \text{for } \omega = \\ blue & \text{for } \omega = \\ red & \text{for } \omega = \\ red & \text{for } \omega = \\ red & \text{for } \omega = \\ green & \text{for } \omega = \end{cases} \quad , ..., \quad Y_n(\omega) = \begin{cases} blue & \text{for } \omega = \\ blue & \text{for } \omega = \\ green & \text{for } \omega = \\ green & \text{for } \omega = \\ green & \text{for } \omega = \\ green & \text{for } \omega = \end{cases}$$

On the left we can see the random variable Y_1 that describes the scenario where $\mathbb{P}_{Y_1}(\{green\}) = \underline{P}(\{green\}) = \frac{1}{6}$ has the lowest and $\mathbb{P}_{Y_1}(\{red\}) = \overline{P}(\{red\}) = \frac{3}{6}$ has the biggest suitable probability. On the right we can see the scenario where $\mathbb{P}_{Y_n}(\{red\}) = \underline{P}(\{red\}) = 0$ has the lowest and

$\mathbb{P}_{Y_n}(\{green\}) = \overline{P}(\{green\}) = \frac{4}{6}$ probability. In every scenario, since *blue* faces can be precisely observed, holds $\mathbb{P}_{Y \in S(\Gamma)}(\{blue\}) = \underline{P}(\{blue\}) = \overline{P}(\{blue\}) = \frac{2}{6}$.

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." [8]

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 [16].

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 $F(r; \phi) = P(T \leq r \,|\, \Phi = \phi)$.

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)[6] 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 P_U
- a parameter of interest θ.

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

- observed data X
- an unobserved realization U^* (of U).

[6] In reference to [4, Chapters 6.4.1.2 and 13.1]], with modified notation to maintain consistency in this exposition. Furthermore some simplifications were made to focus on the essential aspects.

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. But since the distribution of $U \sim P_U$ is known a copy of U, denoted as $u^* \sim P_U$, can be be repeatedly drawn. The new relation $X = G(\theta^*, u^*)$ contains observed (therefore known) data, simulated (therefore known) sample of u^* and a known relation G. Therefore θ^* can be derived as a *random estimator* (or *fiducial sample*) with an according fiducial distribution. This process is called a *stoachastic inversion process* and leads to an estimate, that looks familiar to an MLE, but instead of minimizing the distance between an assumed parametric model and data generating process, the distance between a DGE and the data gets minimized.

$$\theta^*_{FD} = \text{argmin}_{\theta^*} \| \underbrace{X}_{G(\theta, U^*)} - G(\theta, u^*) \|$$

6 Inferential models

Ryan Martin defines defines inferential models as follows:

Definition 6.1 (Inferential models, as given by [12])**.** Fix the sample space $\mathbb{Y}$, parameter space Θ and let $\mathscr{P} = \{P_{Y|\theta} : \theta \in \Theta\}$ be a statistical model. An inferential model is a map from $(\mathscr{P}, y, \dots)$ to a function $b_y : 2^{\Theta}$[7] $\to [0, 1]$, where $b_y(A)$ represents the analyst's belief in the assertion "$\theta \in A$".

This definition seems to have been purposefully left vague and requires further explanation. It formalizes how an assumed model $\mathscr{P}$ and the input data leads to an analysts degree of belief about θ. In the case of Bayesian inference the input would be $(\mathscr{P}, y, prior)$ and the output b_y would be the posterior distribution for θ and therefore an (additive) probability measure. But the IM framework offers more flexibility by allowing "$(\dots)$" to be something else, e.g. a random set, a DGE, and b_y is not required to be additive. The authors suggestions on proper use of this additional flexibility focus on avoiding false confidence by introducing criteria and descriptions how these can be fulfilled. A formalization of false confidence is given by the false confidence theorem.

Theorem 1 (False Confidence)**.** Consider probability measure $\mathbb{P}_y$ with a density function that is bounded and continuous for each y. Then for any $\theta \in \Theta$, any $\alpha \in (0, 1)$, and any $p \in (0, 1)$, there exists a set $A \subset \Theta$ such that

$$\theta \notin A \quad \text{and} \quad P_{Y|\theta}\{\mathbb{P}_y(A) \geq 1 - \alpha\} \geq p.$$

In this sense the *false confidence theorem* states that by definition, since additive belief functions are probability measures, all additive inferential models contain assertions that lead to false confidence. Such statements do not automatically cause complications. For example, when they are trivial, such as when the probability measure is based on the Lebesgue measure and therefore any assertion, or integration area, $\theta \in A \subseteq \mathbb{Q}$ must be $\mathbb{P}(A) = 0$ and $\mathbb{P}(A^C) = 1$ by construction since A is a Lebesgue-null set. Such trivial cases can be easily avoided, since it is known that relatively small subsets of Θ are tend to get assigned a small probability, and therefore the complement a large one. That are simply consequences of applying probability calculus. In a frequentist setting the consequences of applying probability calculus typically

[7] This is an alternative representation of the powerset over Θ.

can be easily understood and will not lead to problems[8], since they are typically explicitly intended. Problems may arise if small integration areas cannot be avoided and frequentist inference is not feasible — this was identified by Balch et al. as the reason for the aforementioned issue in satellite conjunction analysis, where the underlying phenomenon is known as probability dilution. Furthermore, they say:

> "False confidence is the inevitable result of treating epistemic uncertainty as though it were aleatory variability."[3].

Definition 6.2 (Validity)**.** An inferential model $(\mathscr{P}, y, \dots) \mapsto b_y$ is *valid* if

$$\sup_{\theta \notin A} P_{Y|\theta}\{b_y(A) > 1 - \alpha\} \leq \alpha, \quad \forall \alpha \in [0,1], \quad \forall A \subset \Theta.$$

That is, if $\theta \notin A$ and, hence, the hypothesis A is false, then the degree of belief, b_y (A), which is a random variable as a function of $Y \sim P_{Y|\theta}$, is stochastically no larger than $Unif(0,1)$.

Theorem 2. Let $(\mathscr{P}, y, \dots) \mapsto b_y$ be an inferential model that satisfies the validity condition (Definition 6.2). Fix $\alpha \in (0,1)$.

(a) Consider a testing problem with $H_0 : \theta \in A$ versus $H_1 : \theta \notin A$, where A is any subset of Θ. Then the test, T_y, that rejects H_0 if and only if $p_y(A) \leq \alpha$ has frequentist Type I error probability upper-bounded by α. That is,

$$\sup_{\theta \in A} P_{Y|\theta}(T_y \text{ rejects } H_0) \leq \alpha.$$

6.1 Construction of valid Inferential models

In the early stage of Inferential models the authors focused on IM construction based on random sets suggested by the following procedure:

> "*A-step.* Associate the unknown parameter θ to each possible (x, u) pair to obtain a collection of sets $\Theta_x(u)$ of candidate parameter values.
> *P-step.* Predict $\mu*$ with a valid predictive random set $\mathcal{S}$.
> *C-step.* Combine $X = x$, $\Theta_x(u)$, and $\mathcal{S}$ to obtain a random set $\Theta_x(\mathcal{S}) = \bigcup_{u \in \mathcal{S}} \Theta_x(u)$. Then, for any assertion $A \subseteq \Theta$, compute the probability that the random set $\Theta_x(\mathcal{S})$ is a subset of A as a measure of the available evidence in supporting A"[13].

The *a(ssociation)-step* is basically the same as the stating a data generating equation. The *p(rediction)-step* is conceptionally the same as simulating samples from copies of the unobserved auxiliary random variable U but with the major difference, that an extra layer of impression is added by the usage of a random set, characterized by as a subset of an *auxiliary space* $S(U) \subseteq \mathbb{U}$, the so-called *predictive random set*. In contrast to the GFI method, not simply just a single single random estimate and a single fiducial distribution are derived, but instead a set of random estimates and fiducial distributions, with one exception[9]. While GFI relies on a arbitrary choice of one single auxiliary variable U, which might be or not be close to the true (or optimal) U_0, a good choice of of the random set Γ makes it more probable that U_0 is captured in $S(\Gamma)$ which then makes it more probable, that the true (or optimal) value for θ is also captured. In particular, from $U_0 \in S(\Gamma)$, it follows that $\theta_0 \in \{\theta \mid X = G(U, \theta)\} := \Theta_x(U)$ and otherwise. The idea behind the *c(ombination)-step* is pretty simple. It suggests to construct, based on the results from the p-step, a new random set. Based on this random set the lower probability can be calculated to evaluate as an non-additive degrees of belief about any assertion of the form $A \subseteq \Theta$.

Example 6.1 (Based on example 2.1 in [7])**.** Suppose data X was sampled from $Y \sim \mathcal{N}(\mu, 1)$.

- a-step: Set $X = \mu + U, \quad U \sim \mathcal{N}(0, 1)$.
- p-step: Set the predictive random set through $S(u^*) = \{u' \,|\, |u'| \leq |u^*|\}$, where u^* is a copy of U and $u' \in \mathbb{U}$.

[8] Example 4.1 can be read in this way.

[9] The GFI method with an auxiliary variable U can be seen as a special case of such an additive IM where the random set Γ can be characterized by $S(\Gamma) = \{U\}$.

- c-step: Combine the results through $\Theta_X(\mathcal{S}) = \bigcup_{u' \in S(u*)} \{\mu | X = \mu + u'\} = [x - |u^*|,\, x + |u^*|]$.

If we are e.g. interested in the assertion $A = \{\mu \,|\, \mu > 0\}$ and observe $X = -1$, then $\Theta_{-1}(\mathcal{S}) = [-1 - |u^*|,\, -1 + |u^*|]$ is not a subset of A and the lower probability $\underline{P}(A)$ must be 0. The intersect is the set where $|u^*|$ is larger than 1 and $u^* \sim \mathcal{N}(0,\, 1)$ allows us to calculate the upper probability $\overline{P}(A) = \mathbb{P}(|\mu^*| > 1) = 2 \cdot (1 - \phi(1)) \approx 0.32$. For the negation A^C the lower probability can be calculated by $\underline{P}(A^C) = 1 - \overline{P}(A) = 0.68$ and the upper probability by $\overline{P}(A^C) = 1 - \underline{P}(A) = 1$.

In the early stage of IM the focus laid on the *hitting probability* of random sets defined as $\gamma(U^*) = P(U^* \in S(\Gamma))$. If $\gamma(U_0) \geq_{st}$[10] $Unif(0, 1)$, as a function of $U_0 \sim P_{U_0}$, holds, then the IM output is said to be valid, as stated by proposition 16.1 in [4]. This can be reformulated to the requirement that S must be such that $P_U\{\gamma_S \leq \alpha\} \leq \alpha \quad \forall \alpha \in [0,\, 1]$. In the validity theorem we can replace b_y, which is a pessimistic belief representation, though an optimistic belief representation p_y. Then

$$\sup_{\theta \in A} P_{Y|\theta}\{p_Y(A) \leq \alpha\} \;\leq\; \alpha \quad \forall \alpha \in [0,\, 1],\, \forall\, A \subseteq \Theta$$

shows the connection between an optimistic belief that is limited by α. This relation can be captured by a possibility distribution π by choosing the possibility measure $\Pi(A) := \sup_{u \in A} \pi(u)$ as the optimistic belief representation. Then the validity of an IM output does not rely on a random set, but on the chosen possibility distribution. This simplifies valid IM construction from "construct a random set which offers our desired properties to cover the uncertainty about U" to a much simpler problem, where the uncertainty can be directly covered by a possibility distribution. Every suitable possibility distribution directly leads to validity. A possibility distribution is suitable if for every assertion A and for every α, the resulting α-cuts A_α must cover at least as many possibilities as there is uncertainty.

[10] $\geq_{st}$ means "stochastically no smaller than" in the sense that the distribution function of $\gamma(U_0)$ is on or below that of $Unif(0, 1)$.

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] Michael Scott Balch, Ryan Martin, and Scott Ferson. "Satellite conjunction analysis and the false confidence theorem". en. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 475.2227 (July 2019), p. 20180565. ISSN: 1364-5021, 1471-2946. DOI: 10.1098/rspa.2018.0565. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.2018.0565 (visited on 05/19/2025).

[4] James O. Berger et al., eds. *Handbook of Bayesian, fiducial, and frequentist inference*. First edition. Chapman & Hall/CRC handbooks of modern statistical methods. Boca Raton, FL: CRC Press, 2024. ISBN: 978-0-429-34173-1.

[5] 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).

[6] D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. en. CRC Press, Sept. 1979. ISBN: 978-0-412-16160-5.

[7] Duncan Ermini Leaf and Chuanhai Liu. "Inference about constrained parameters using the elastic belief method". en. In: *International Journal of Approximate Reasoning* 53.5 (July 2012). Publisher: Elsevier BV, pp. 709–727. ISSN: 0888-613X. DOI: 10.1016/j.ijar.2012.02.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0888613X12000138 (visited on 05/23/2025).

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

[9] Robert Hable. "Data-Based Decisions under Complex Uncertainty". In: 2009. URL: https://api.semanticscholar.org/CorpusID:11222776.

[10] 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.

[11] *Kolmogoroff (1933) Grundbegriffe Der Wahrscheinlichkeitsrechnung*. ger. URL: http://archive.org/details/kolmogoroff-1933-grundbegriffe-der-wahrscheinlichkeitsrechnung (visited on 05/19/2025).

[12] Ryan Martin. "False confidence, non-additive beliefs, and valid statistical inference". en. In: *International Journal of Approximate Reasoning* 113 (Oct. 2019), pp. 39–73. ISSN: 0888613X. DOI: 10.1016/j.ijar.2019.06.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0888613X19300696 (visited on 05/17/2025).

[13] Ryan Martin and Chuanhai Liu. "Inferential Models: A Framework for Prior-Free Posterior Probabilistic Inference". In: *Journal of the American Statistical Association* 108 (Mar. 2013), pp. 301–313. DOI: 10.1080/01621459.2012.747960.

[14] Enrique Miranda, Inés Couso, and Pedro Gil. "Random sets as imprecise random variables". In: *Journal of Mathematical Analysis and Applications* 307.1 (2005), pp. 32–47. ISSN: 0022-247X. DOI: https://doi.org/10.1016/j.jmaa.2004.10.022. URL: https://www.sciencedirect.com/science/article/pii/S0022247X04008571.

[15] Glenn Shafer and Vladimir Vovk. *The origins and legacy of Kolmogorov's Grundbegriffe*. arXiv:1802.06071 [math]. Feb. 2018. DOI: 10.48550/arXiv.1802.06071. URL: http://arxiv.org/abs/1802.06071 (visited on 05/12/2025).
[16] S L Zabell. "R. A. Fish~r and the Fiducial Argument". en. In: ().
[17] 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.