

On principles of inductive inference*

Ryszard Paweł Kostecki

Institute of Theoretical Physics, University of Warsaw, Hoża 69, 00-681 Warszawa, Poland

January 31, 2012

Abstract

We propose an *intersubjective epistemic* approach to foundations of probability theory and statistical inference, based on relative entropy and category theory, and aimed to bypass the mathematical and conceptual problems of existing foundational approaches.

1 Standard frameworks

The notion *inference* means ‘logical reasoning’. The *deductive inference* specifies premises by the valuations of sentences in *truth* values, and provides an inference procedure which is considered to lead to *certain* conclusion on the base of given premises. The *inductive inference* specifies premises by the valuations of sentences in *possible (plausible)* values and provides an inference procedure which is considered to lead to *most possible (most plausible)* conclusions on the base of given premises. From the mathematical perspective, the difference between deductive and inductive inference lays not in the form of logical valuations (these can be the same in both methods), but in the procedure of specifying conclusions on the base of premises. The conclusions of multiple application of deductive inference to the sequence of sets of premises depend *in principle* on all elements of all these sets, while the conclusions of the multiple application of inductive inference to the sequence of sets of premises depend *in principle* only on some elements of some of these sets. For this reason, the premises of inductive inference are also called *evidence*. An example of inductive inference procedure is any statistical reasoning based on probabilities. The evidence can be provided, for example, by particular quantities with units interpreted as ‘experimental data’ *together with* a particular choice of a method which incorporates these ‘data’ into statistical inference. Any choice of such method *defines* the evidential meaning of the ‘data’, and is a crucial element of the inference procedure. A standard example of such method is to ignore everything what is known about a sequence of numbers associated with a single abstract quality (such as a “position”), leaving only the value of arithmetic average and the value of a fluctuation around this average as a subject of comparison (e.g., by identification) with the mean and variance parameters of the gaussian probabilistic model.

According to frequentist interpretation (by Ellis [29], Venn [90], Fisher [31], von Mises [91, 92], Neyman [68], and others) probabilities can be given meaning only as relative frequencies of some experimental outcomes in some asymptotic limit. This interpretation was very influential in the last 160 years, *but* so far *none* mathematically strict and logically sound formulation of this interpretation exists (see, e.g., [48, 89, 39]). The separation of a formalism of inductive inference into ‘probability theory’ and ‘statistics’ is

*To be published in: Goyal P. (ed.), *Proceedings of the 31th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Waterloo, July 10-15, 2011*, AIP Conf. Proc., Springer, Berlin. [arXiv:1109.3142](https://arxiv.org/abs/1109.3142)

also a consequence of frequentist interpretation, which forbids consideration of probability (understood as relative frequency) as a subject of change based on change of evidence. *Thus, without frequentism, there is no reason for keeping the division of probabilistic and statistical part of the framework of inductive inference into two separated theories.* Moreover, the methods of statistical inference used within the frames of the frequentist approach are mainly based on *ad hoc* principles, which are justified by convention, and do not possess mathematically strict and logically sound justification (see e.g. [76, 46, 6, 81, 47]). This is a consequence of the lack of strict and sound foundations of frequentist interpretation of probability.

Beyond logically and mathematically unjustified frequentist approach (and even less successful [45, 26] propensity interpretation [71, 74, 75, 36]), probability theory and statistical inference theory can be considered as two parts (resp., kinematic and dynamic) of a single theory of quantitative inductive inference. The evidences used in this inference need not be restricted to frequencies. Nevertheless, the choice of particular methods of specification of evidences and drawing inferences requires some justification.

The ‘subjective’ bayesian approach (by Ramsey [77], de Finetti [20, 21, 22], Savage [79], and others) allows any kinematics and requires Bayes’ rule as dynamics, grounding both in requirement of *personal* consistency of betting. This is conceptually consistent, *but* by definition *lacks* any rules relating the probability assignments (theoretical model construction) with intersubjective knowledge (experimental setup construction, ‘experimental data’). Thus, it is often accused of arbitrariness. Such accusations *are* justified if they amount to saying that the methods of scientific inquiry *seem to be something more* than individual personal consistency of bets, *but are not* justified if they appeal to (operationally undefined!) notions of ‘objectivity’, ‘nature’, ‘reality’, etc., because *any* theoretical statement is after all an arbitrary mental construct.

The syntactic approach (by Johnson [52], Keynes [56], Carnap [9, 10], and others) amounts to construct probability theory as a predicative calculus in a formal language, *but* it does *not* provide neither any sound justification for the choice of language and calculus nor any definite methods of model construction, different from ‘subjective’ approach (see e.g. [43, 41, 95]). This makes syntactic approach foundationally irrelevant.

The ‘objective’ bayesian approach (by early Jeffreys [49], Cox [16, 17], Jaynes [47], Berger [5], and others) allows various mathematical rules of assignment of probabilities (see e.g. [55, 47]) and of inference (see e.g. [16, 17, 82, 83, 11, 12]). It attempts to select the preferred rules by an appeal to some notions of ‘rational consistency’ or ‘experimental reproducibility’, *but it fails* to provide sound conceptual justification for these rules which would be neither subjectively idealistic (personalist) nor ontologically idealistic (frequentist) [87, 88, 30, 54, 63]. (One of important problems is the lack of justification of the methods that associate particular constraints in entropic updating rule with some functions of ‘experimental evidence’, see e.g. [87, 88].) Yet, the idea to provide some rules of probabilistic model construction taking into account the role played by experimental evidence in *intersubjective consistency* of inductive inferences seems to be crucial.

The variety of above *conceptual frameworks* for probability theory and statistical inference corresponds to the variety of *mathematical frameworks*. There are four main approaches: by Bayes–Laplace [4, 23], Borel–Kolmogorov [7, 59], Whittle [93, 94], and Le Cam [64, 65]. But even more approaches is possible, because probability theory can be built upon two components: evaluational (kinematic) and relational (dynamic), and, apart from selection of one or two of these components, one can provide different

mathematical implementations thereof. For example, the evaluational component can be given either by an abstract measure theory on σ -algebras of subsets of some set, or by an integral theory on vector lattices, while the relational component might be given either by Bayes' rule, or by conditional expectations, or by entropic updating, etc.

The Borel–Kolmogorov framework [7, 59] is based on the notions of measure spaces $(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)$ and probabilistic models

$$\mathcal{M}(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu) \subseteq L_1(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)^+.$$

Building upon measure-theoretic integration theory, this framework is, from scratch, equipped with kinematic (evaluational) prescriptions, *but* has no *generic* notion of conditional updating of probabilities. (The reason of it is an associated, but by no means necessary, frequentist interpretation, which claims identification of probabilities with frequencies. This forbids ‘updating’ probabilities because it would mean updating the frequencies.) There are three facts to observe here. First, many different measure spaces $(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)$ lead to $L_1(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)$ spaces that are all isometrically isomorphic to the same abstract $L_1(\mathfrak{U})$ space, where \mathfrak{U} is camDcb-algebra (countably additive, Dedekind complete, boolean, allowing at least one strictly positive semi-finite measure), constructed by [35]

$$\mathfrak{U} := \mathfrak{U}(\mathcal{X}) / \{A \in \mathfrak{U}(\mathcal{X}) \mid \mu(A) = 0\}.$$

Thus, only $L_1(\mathfrak{U})$ is necessary for defining probabilistic models. But, given any camDcb-algebra \mathfrak{U} , the association of $L_1(\mathfrak{U})$ (and any other $L_p(\mathfrak{U})$) to \mathfrak{U} is functorial [35], and no appeal to representations in terms of measure space is ever required. Second, by the Loomis–Sikorski theorem [66, 84], each camDcb-algebra \mathfrak{U} can be represented as a measure space $(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)$, given the choice of measure μ on \mathfrak{U} . However, there are many different measure spaces that lead to the same algebra \mathfrak{U} [80]. *Thus, using the measure spaces $(\mathcal{X}, \mathfrak{U}(\mathcal{X}), \mu)$ instead of camDcb-algebras \mathfrak{U} is ambiguous.* Finally, as observed by Le Cam [64, 65] and Whittle [93, 94], probabilistic description in terms of measures μ on $(\mathcal{X}, \mathfrak{U}(\mathcal{X}))$ can be completely replaced by the description in terms of integrals ω on vector lattices $\mathfrak{A}(\mathfrak{U})$. (For every camDcb-algebra \mathfrak{U} there exists a canonically associated vector lattice $\mathfrak{A}(\mathfrak{U})$ of characteristic functions on the set of boolean homomorphisms $\mathfrak{U} \rightarrow \mathbb{Z}_2$.) *Thus, one can deal exclusively with positive finite integrals over commutative algebras \mathfrak{U} instead of measures on $\mathfrak{U}(\mathcal{X})$.* The normalised integrals are just expectation functionals, and probability of $a \in \mathfrak{U}$ is recovered by evaluation $p(a) := \omega(\chi_a)$ on characteristic function $\chi_a \in \mathfrak{A}(\mathfrak{U})$.

On the other hand, the Bayes–Laplace framework [4, 23] is based on finitely additive boolean algebras \mathcal{B} and conditional probabilities $p(x|\theta) : \mathcal{B} \times \mathcal{B} \rightarrow [0, 1]$. It is equipped from scratch with dynamical (relational) prescription, given by Bayes' rule

$$p(x|\theta) \mapsto p_{\text{new}}(x|\theta) := p(x|\theta) \frac{p(b|x \wedge \theta)}{p(b|\theta)},$$

but it provides no *generic* notion of probabilistic expectation over a continuous (countably additive) domain of infinite sets. Bayes' rule defines a concrete method of providing statistical inferences. Thus, *statistical inference can be understood as a dynamical component of probability theory.* As noticed by Jaynes [47], Bayes' rule and all Bayes–Laplace framework has precisely the same properties and the same range of validity if the conditional probabilities are evaluated into $[1, \infty]$ instead of $[0, 1]$. *Hence, the normalisation*

of probabilities is not a necessary feature of probabilistic/inferential framework (it is necessary only in the frequentist interpretation). Moreover, Bayes' rule is a special case of constrained relative entropic updating [13, 37]

$$p \mapsto p_{\text{new}} := \arg \inf_{q \in \mathcal{M}} \left\{ \int p \log(p/q) + F(q) \right\},$$

for p, q in a probabilistic model $\mathcal{M} = \mathcal{M}(\mathcal{X}, \mathcal{U}(\mathcal{X}), \mu)$, $\dim \mathcal{M} = n < \infty$, with parametrisation $\theta : \mathcal{M}(\mathcal{X}, \mathcal{U}(\mathcal{X}), \mu) \rightarrow \Theta \subseteq \mathbb{R}^n$, and constraints given by

$$F(q) = \lambda_1 \left(\int_{\mathcal{X}} \mu(x) \int_{\Theta} d\theta q(x|\theta) - 1 \right) + \lambda_2 \left(\int_{\Theta} d\theta q(x|\theta) - \delta(x - b) \right),$$

where λ_1 and λ_2 are Lagrange multipliers.

2 Towards new approach

In order to bypass the above problems, we need to take more careful look at the foundations of bayesian approach. Note that the Ramsey–de Finetti type [77, 20, 21] and Cox's type [16, 17] derivations of Bayes' theorem (or, equivalently, of the algebraic rules of 'probability calculus') *assume* that the conditional probabilities $p(A|I)$ are to be used in order to draw inferences on the base of premises (evidence) I . Hence, the function $p(A|I)$, or any other function used to derive it, already *assumes* that *some* rule of probability updating (inductive inference) has to be used, because only this assumption allows to speak of some elements of the algebra as 'evidence', or to speak of conditional probabilities as 'inferences'. Any particular algebraic rules of transformation of conditional probabilities arise only as a result of *additional* assumptions, which might not be relevant for general purposes and require anyway some additional justification. This observation allows us to consider spaces $\mathcal{M}(\mathcal{U}) \subseteq L_1(\mathcal{U})^+$ of unconditioned positive finite integrals (*information models*) as a candidate for kinematic component of inductive inference theory, and to consider *some* principle $\mathfrak{P} : \mathcal{M}(\mathcal{U}) \rightarrow \mathcal{M}(\mathcal{U})$ of updating of integrals as a candidate for its dynamic component (*information dynamics*). As opposed to approaches aimed at identification of algebraic and lattice theoretic relations underlying evaluations $p(A|I)$ and their transformations [57, 58], we do not require that inferences should be conditioned exclusively on the elements of the underlying algebra, and we allow infinite-dimensional algebras and information models in foundations.

The choice of any particular form of principle of inductive inference (information dynamics) is a delicate issue, because (for any particular form of information kinematics) it determines the range and form of allowed inferences. According to the arguments of [44, 15], there can be given no deductive logical premises for the claim that some inductive inference rule is absolute (universal), while inductive justification of induction is impossible due to circularity. However, if the chosen principle reduces in particular cases to a wide class of practically convenient and in some sense optimal techniques, then it can be considered as *appealing*. Moreover, if it allows characterisation by axioms possessing unambiguous interpretation, then it can be considered as *appealing* too.

Because Bayes' rule is a special case of constrained maximisation of relative entropy [13], we consider the latter as a candidate for a general principle \mathfrak{P} of quantitative inductive inference. For an evidence that it is appealing from the practical point of view,

let us note that: (i) the conditional expectations are characterised as maximisers of the expectations of Bregman relative entropy [3]; (ii) the maximum likelihood methods are just special case of application of Bayes' rule [47]; (iii) inference techniques based on Fisher–Rao information metric amount to using the second order Taylor expansion of Csiszár–Morimoto relative entropies [62, 27]; (iv) many standard frequentist techniques of statistical inference can be reexpressed in terms of relative entropy, see e.g. [62, 18, 97, 47, 28]. Regarding axiomatisation, Shore and Johnson [82, 83, 51], Paris and Vencovská [69, 70], and Csiszár [19] have provided characterisations of the principle of maximisation of constrained Kullback–Leibler relative entropy as a unique probability updating rule that satisfies some set of conditions (see also [85, 87, 88, 11, 12, 37]). Similar axiomatisation was provided by Jones and Byrne [53] and Csiszár [19] with respect to maximisation of Bregman relative entropies with linear constraints. More generally, the existence and uniqueness of constrained maximisation of any Bregman relative entropy is equivalent with the projection along ∇ -geodesic onto the non-empty set that is convex and closed in ∇^* -geodesics, where the torsion-free flat affine connections (∇, ∇^*) arise from the third order Taylor expansion of this relative entropy [1, 2]. If any of conditions of this type are accepted (what forms a particular *decision*), then the resulting updating rule is unique, what makes this rule axiomatically appealing. However, like in the case of derivation of Bayes' rule from Cox's type or the Ramsey–de Finetti type procedure, one might deny some of the premises of these derivations (such as normalisation), and *decide* to accept some other set of premises, leading to some other inductive inference rule. There also remains a problem how to relate the mathematical constraints of relative entropic updating with the operational descriptions specifying particular experiment.

The choice of a particular form of information dynamics is thus relativised to a particular set of decisions, which are *in principle* arbitrary. The same applies also to the choice of particular form of information kinematics (which includes model construction and model selection). But this arbitrariness is not necessarily unconstrained. According to the 'subjective' bayesian interpretation, it is constrained by consistency of decisions of a single person (individual). Thus, each person can in principle choose arbitrary method of kinematic model construction and arbitrary method of inductive inference, but he or she is required to maintain personal consistency of these choices in subsequent inferences. In our opinion, the *necessary* requirement for *scientific* inference (as opposed to *personal* inference) is to make these decisions consistent relatively to a particular community of users/agents. In other words, the decisions underlying construction and use of a particular form of information kinematics and information dynamics should be intersubjectively accepted and applied by all members of the given community. This way, within the range of intersubjective validity of these decisions, the notion of information model and its dynamics cannot be considered as 'subjective'.

This asks for the *sufficient* conditions that define the *scientific* inference. The crucial observation comes from Fleck [32, 33, 34] (see also [86] and [61]), who showed that the 'scientific facts' and 'experimental data' are always specified within the frames of some decisions (expressed in terms of particular assumptions and settings, including specification of the allowed response scales of measured outcomes, allowed experimental preparations and treatments, etc.), which are necessary to obtain intersubjective consistency with the preconceived notion of an 'experiment of a given type'. These decisions *define* the range of allowed variability of 'facts' and 'data', but are not determined by the theoretical model under consideration. Yet, *both* theoretical (inferential, informational) and experimental

(operational, knowledge-carrying) layers of scientific inquiry, as well as their mutual relationships, are determined by some intersubjectively shared *thought style* (which is an entity from meta-theoretic level). Hence, ‘scientific facts’ or ‘experimental data’ are relative to some particular intersubjectively shared decisions on construction and use of experimental setups. Everything that is individually (*personally*) experienced in a particular experimental situation, but does not fit into the frames rendered by the above decisions, is not considered as a valid ‘experimental data’ (‘scientific facts’) for an experiment *of a given type*. However, it *does not* mean that the ‘experimental data’ for a particular instance of experiment of a given type is completely *determined* by these frames. Taking under consideration all above restrictions and relativisations, there remains unexpectedness of a particular outcome that appears *within given frames* (of particular configuration and particular range of allowed outcomes). The aim of theoretical inquiry is to provide inductive inferences about these unexpected outcomes, dependently on the particular constraints that are taken into consideration as evidence.

We postulate that for any (stage of historical development of a) given thought style, the scientific character of inquiry requires to separate the theoretical abstract language used to intersubjectively define and communicate theories from the operational language used to intersubjectively define and communicate corresponding experiments. The specification of operational (‘experimental’) language for intersubjectively valid construction and use of experimental setups allows to define the range of validity of particular methods of inductive inference. More precisely, this allows to construct the *relational* justification of given rules of construction of information kinematics and information dynamics with respect to a particular description of experiments of a given type, provided in terms of a given operational language. In particular, this applies to justification of the choice of E and F in the constrained maximisation of relative entropy. They can be determined by the assumptions expressed in operational language *and* by the chosen relationship (correspondence) between operational and theoretical descriptions. The latter interprets terms of operational language in theoretical terms, and provides an operational (experimental) contextualisation of a given information (inductive inference) theory.

This means that the theoretical model is verified only *with respect to* certain context of intersubjectively shared decisions which construct the ‘experiment of a given type’. More precisely, the ‘experimental verification’ of a predictive theory means just an *intersubjective reproducibility (consistency)* of relationship between predictions (inferences) over a particular information model and results of use of an experimental setup of a given type, *with respect to* a given correspondence between the kinematics of this model and construction of experimental setup of a given type, as well as between the constraints of inductive inference and the particular use of this experimental setup. To simplify these relativisations, we can restrict the notion of *intersubjective validity*, as referring to fixed operational definitions *and* fixed operational contextualisation. This leads to *intersubjective interpretation*: the meaning of knowledge used to define particular theoretical model and its dynamics is provided by operational criteria that are sufficient and necessary in order to *intersubjectively reproduce* an ‘experiment of a given type’ that is considered to correspond to this theoretical model (which means that the inferences drawn from this model are interpreted as most plausible outcomes of corresponding experiment).

This interpretation does not define the absolute (passive, static) meaning of the notion of ‘knowledge’. It defines only the relational (active, dynamic) meaning of this notion, as a particular relationship between kinematics-and-dynamics of theoretical model and

construction-and-use of experimental setup. It differs from Bridgman’s operationalism [8] and various versions of logical positivism, because it postulates neither the *reduction* of theoretical layer to experimental layer, nor the absolute universality of their relationship. It differs from conventionalism of Duhem [24, 25], Poincaré [72, 73], and late Jeffreys [50] by an additional requirement of intersubjective consistency between experimental setup construction and theoretical model construction. This way it is capable of providing a solution to the problem that neither ‘subjective’ bayesianism nor ‘objective’ bayesianism can justify the particular use of ‘experimental data’ as evidence in inductive inference procedures. Personal betting behaviour and ontological postulates play no role in it. The intersubjective consistency (validity) of a particular relationship between theoretical model construction and operational construction of experimental setup, as well as a particular relationship between theoretical dynamics and operational use of experimental setup, is relative only to some community of users/agents which agree upon them. The constraint of intersubjective agreement is of meta-theoretical character and cannot be described in terms of inductive inference theory. Beyond any given community, the particular rules of models construction and inductive inference, as well as their relationship with the particular experimental setups and their use, are *irrelevant* (arbitrary, personalistic, ‘subjective’), but within this range they are *indispensable* (necessary, scientific, ‘objective’). This resolves the “subjective vs objective” bayesian debate, as well as it dissolves the bayesian version of the reference class problem [40], by removing them *beyond* inductive inference theory. The criteria for intersubjective consistency (coherence) of experimental verifiability establish the direct link of inductive inference with experimental data, which is independent of any assumptions of ontological or meta-theoretical character.

3 The principles of new foundations

On the level of *mathematical framework*, following earlier paper , we propose to unify kinematic (probabilistic, evaluational) and dynamic (statistic-inferential, relational) components, taking the best insights from analysis of the Borel–Kolmogorov and the Bayes–Laplace approaches. Thus, we replace measure spaces $(\mathcal{X}, \mathcal{U}(\mathcal{X}), \mu)$ by camDcb-algebras \mathcal{U} , and we consider integrals instead of measures. The failure of frequentism allows us to introduce statistical inference and lack of normalisation directly into foundations. We define: *information kinematics* as given by *information models* $\mathcal{M}(\mathcal{U}) \subseteq L_1(\mathcal{U})^+$ and their *information geometry* (specified by deviations, riemannian metrics, affine connections, etc., see e.g. [14, 1, 2, 38]); *information dynamics* as given by minimisation of information deviation $D : \mathcal{M}(\mathcal{U}) \times \mathcal{M}(\mathcal{U}) \rightarrow [0, \infty]$ such that $D(\omega, \phi) = 0 \iff \omega = \phi$, weighted by relative prior measure $E : \mathcal{M}(\mathcal{U}) \times \mathcal{M}(\mathcal{U}) \rightarrow [0, \infty]$, and subjected to constraints $F : \mathcal{M}(\mathcal{U}) \rightarrow]-\infty, \infty]$,

$$\mathfrak{P}_{D,E,F} : \mathcal{M}(\mathcal{U}) \ni \omega \mapsto \arg \inf_{\phi \in \mathcal{M}(\mathcal{U})} \left\{ \int_{\phi \in \mathcal{M}(\mathcal{U})} E(\varphi, \omega) D(\varphi, \phi) + F(\phi) \right\} \in \mathcal{M}(\mathcal{U}). \quad (3.1)$$

On the level of *information semantics*, the underlying algebra \mathcal{U} represents an *abstract qualitative language* subjected to quantitative evaluation, the space $\mathcal{M}(\mathcal{U})$ of finite integrals and its geometry represents *quantified knowledge*, while the entropic updating (3.1) represents *quantitative inductive inference procedure* determined by the triple (D, E, F) . The functions E and F specify the *evidence*, while the resulting projection $\mathfrak{P}_{D,E,F}$ is an

inductive inference: the most plausible state of knowledge subjected to given evidence. By *well-posed inference problem* we understand such triples (D, E, F) for which the solution $\mathfrak{P}_{D,E,F}$ exists and is unique. The lack of existence indicates *overdetermination* of an inference problem, while the lack of uniqueness indicates its *underdetermination*. For $E(\varphi, \omega) = d\varphi\delta(\varphi - \omega)$ the principle (3.1) says: given initial information state ω , choose such information state that is most close to ω in terms of distance defined by D , under constraints defined by F . Other relative prior measures E on $\mathcal{M}(\mathcal{U})$ allow more general selection of information state, which takes under consideration the relative distances to several different information states associated with the initial state. (This idea was first proposed by Zhu and Rohwer [96, 97], see also [78].)

To determine the particular operational and conceptual meaning attributed to the terms ‘knowledge’ and ‘change of knowledge’, the *information semantics* requires an additional *epistemic interpretation*. It should determine the choice of a particular information kinematics ($\mathcal{M}(\mathcal{U})$ and its information geometry) and a particular information dynamics (D, E, F) when applied to some particular experiments. According to *intersubjective epistemic interpretation*: (i) the construction of information kinematics should correspond to the intersubjective description of construction of experimental setup of a given type (provided in some operational terms); (ii) the construction of information dynamics should correspond to the intersubjective description of the use of a experimental setup of a given type. But what do ‘intersubjective description’ and ‘correspondence’ mean?

While it is impossible to formalise the intersubjective *decisions* within the framework of inductive *inference* theory, one can formalise the operational criteria that are necessary and sufficient in order to verify whether individual instances of experimentation (e.g., some mutually related preparations, actions and observations) can be considered as an *intersubjectively valid* instance of an ‘experiment of a given type’. These criteria *define* an intersubjective notion of an ‘experiment of a given type’, which is an abstraction, because one cannot step twice into the *same* stream. The intersubjective reproducibility (validation) of experiments can be understood as a preservation of their abstract structure in all individual (personal) instances of experimentation. We will encode this structure together with the corresponding structure-preservation criteria in terms of category theory. By an ‘experiment of a given type’ we understand an ‘experimental setup of a given type’ *and* a ‘given type of its use’. The former is defined by providing some particular categories: (i) **Subj** of abstract qualities (‘preparations’, ‘subjects’, ‘things’, ‘properties’, ‘questions’, ‘experimental units’) subjected to consideration in experiment; (ii) **Config** of active configurations of experimental inputs (‘treatments’, ‘interventions’, ‘acts’); (iii) **Scale** of registration scales of experimental outcomes (‘outputs’, ‘results’). The latter is defined by providing an association between inputs and outcomes conditioned on the choice of qualities. The particular association of outcome, input and quality in course of an individual (personal) instance of an experiment will be called ‘fact’ or ‘experimental data’. Following [42] we define an *observation* as a map from preparation (quality) to outcome (registration value), and define an *experimental design* as a map from preparation to treatment (act, configuration). This allows us to define *fact* as a map from experimental design to an outcome. We also require that facts can be decomposed into ‘active’ parts provided by treatments, and ‘passive’ parts provided by observations. Following [67] we assume that **Subj**, **Config** and **Scale** are equipped with forgetful functors $U. : \cdot \rightarrow \mathbf{Set}$ (allowing to deal with their objects as with sets), and we define: a *category*

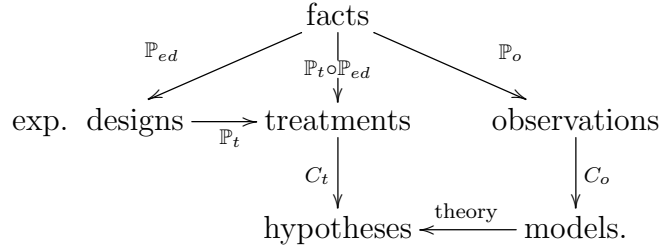
of observations as a product category $\mathbf{Scale} \times \mathbf{Subj}^{\text{op}}$; a category of *experimental designs* as a comma category $U_{\mathbf{Subj}} \downarrow U_{\mathbf{Config}}$; a *category of facts* as

$$\mathbf{Scale} \times (U_{\mathbf{Subj}} \downarrow U_{\mathbf{Config}})^{\text{op}}.$$

The decomposition of facts into experimental designs, treatments and observations is provided, respectively, by means of the projection functors

$$\begin{aligned} \mathbb{P}_{ed} &:= (\cdot)^{\text{op}} \circ \Pi_{(U_{\mathbf{Subj}} \downarrow U_{\mathbf{Config}})^{\text{op}}}, \\ \mathbb{P}_t &:= \Pi_{\mathbf{Config}}, \\ \mathbb{P}_o &:= \text{id}_{\mathbf{Scale}} \times \Pi_{\mathbf{Subj}^{\text{op}}}. \end{aligned}$$

Let theoretical layer be equipped not only with models $\mathcal{M}(\mathcal{U})$ and triples (D, E, F) , but also with spaces H of hypotheses. Define an *inductive inference theory* as a method that assigns to each model $\mathcal{M}(\mathcal{U})$ a (D, E, F) -dependent map into some H . If this method can be expressed as a covariant functor from the category of models $\mathcal{M}(\mathcal{U})$ equipped with information geometric structures and geometry-preserving morphisms as arrows to the category of spaces of hypotheses, then by *correspondence* we understand the functorial association of the respective categories:



This scheme is inspired and motivated by [42] and [67]. The functors C_t and C_o are correspondence rules that interpret active and passive components of experimental layer in terms of, respectively, active and passive components of theoretical layer. The theory will be called *experimentally verifiable* with respect to the given intersubjective context of facts and correspondence rules iff the above diagram commutes. This replaces the ‘absolute’ evaluation of the truth (verifiability, falsifiability) of a theory by the *mutual coherence* of experimental and theoretical layers of intersubjective scientific inquiry in face of the given correspondence rules. In analogy to [67] one can consider the algebras \mathcal{U} as objects of \mathbf{Subj} . In such case each \mathcal{U} can be understood as an abstract entity that is used for intersubjective communication of experimental preparations. The preparations represented by \mathcal{U} are subjected to observations valued into the objects of \mathbf{Scale} . The interpretation (correspondence rule) C_o maps collections of these valuations into the corresponding information models $\mathcal{M}(\mathcal{U})$ and their geometry. The models $\mathcal{M}(\mathcal{U})$ consists of theoretic valuations (integrals) into \mathbb{R}^+ , which are states of information. Due to C_o , the information states can be understood as the carriers of quantitative intersubjective knowledge describing the experimental setup of a given type. The choice of evidence E and F provides the description of active configurational settings (treatments, inputs). The choice of D and $\mathfrak{P}_{D,E,F}$ determines a range of allowed inferences. The experimental verifiability condition determines the relationship of these inferences with experimental facts.

References

- [1] Amari S.-i., 1985, *Differential–geometrical methods in statistics*, Lecture Notes in Statistics **28**, Springer, Berlin.
- [2] Amari S.-i., Nagaoka H., 1993, *Joho kika no hoho*, Iwanami Shoten, Tōkyō (engl. transl. rev. ed.: 2000, *Methods of information geometry*, American Mathematical Society, Providence).
- [3] Banerjee A., Guo X., Wang H., 2005, *On the optimality of conditional expectation as a Bregman predictor*, IEEE Trans. Inf. Theor. **51**, 2664.
- [4] Bayes T., 1763, *An essay towards solving a problem in the doctrine of chances*, Phil. Trans. Roy. Soc. London **53**, 370 (reprinted in: 1958, *Biometrika* **45**, 293).
- [5] Berger J.O., 2006, *The case for objective bayesian analysis*, Bayes. Anal. **1**, 385.
- [6] Berger J.O., Berry D.A., 1988, *Statistical analysis and the illusion of objectivity*, Amer. Scient. **76**, 159.
- [7] Borel É., 1909, *Les probabilités dénomérables et leurs applications arithmétiques*, Rend. Circ. Math. Palermo **27**, 247.
- [8] Bridgman P.W., 1927, *The logic of modern physics*, Macmillan, New York.
- [9] Carnap R., 1950, *Logical foundations of probability*, University of Chicago Press, Chicago.
- [10] Carnap R., 1952, *The continuum of inductive methods*, University of Chicago Press, Chicago.
- [11] Caticha A., 2004, *Relative entropy and inductive inference*, in: Erickson G., Zhai Y. (eds.) *Bayesian inference and maximum entropy methods in science and engineering*, AIP Conf. Proc. **707**, 75. [arXiv:physics/0311093](https://arxiv.org/abs/physics/0311093).
- [12] Caticha A., 2007, *Information and entropy*, in: Knuth K. et al. (eds.), *Bayesian inference and maximum entropy methods in science and engineering*, AIP Conf. Proc. **954**, 11. [arXiv:0710.1068](https://arxiv.org/abs/0710.1068).
- [13] Caticha A., Giffin A., 2006, *Updating probabilities*, in: Mohammad-Djafari A. (ed.), *Bayesian inference and maximum entropy methods in science and engineering*, AIP Conf. Proc. **872**, 31. [arXiv:physics/0608185](https://arxiv.org/abs/physics/0608185).
- [14] Chentsov N.N., 1972, *Statisticheskie reshayushchie pravila i optimal'nye vyvody*, Nauka, Moskva (engl. transl.: 1982, *Statistical decision rules and optimal inference*, American Mathematical Society, Providence).
- [15] Chwistek L., 1948, *The limits of science: outline of logic and of the methodology of the exact sciences*, Harcourt Brace, New York.
- [16] Cox R.T., 1946, *Probability, frequency and reasonable expectation*, Am. J. Phys. **14**, 1.
- [17] Cox R.T., 1961, *The algebra of probable inference*, John Hopkins University Press, Baltimore.
- [18] Cressie N., Read T.R.C., 1984, *Multinomial goodness-of-fit tests*, J. Roy. Statist. Soc. B **46**, 440.
- [19] Csiszár I., 1991, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Stat. **19**, 2032.
- [20] de Finetti B., 1931, *Sul significato suggestivo della probabilità*, Mathematicae **17**, 298.
- [21] de Finetti B., 1937, *Prevision: ses lois logiques, ses sources subjectives*, Ann. Inst. Henri Poincaré **7**, 1. (engl. transl. 1964, *Foresight: its logical laws, its subjective sources*, in: Kyburg H.E., Jr., Smokler H.E. (eds.), *Studies in subjective probability*, Wiley, New York, p.93).
- [22] de Finetti B., 1970, *Teoria delle probabilità*, Einaudi, Torino (engl. transl.: 1974, *Theory of probability*, Vol.1-2, Wiley, New York).
- [23] de Laplace P.-S., 1812, *Théorie analytique des probabilités*, Courcier, Paris.
- [24] Duhem P., 1894, *Quelques réflexions au sujet de la physique expérimentale*, Rev. Quest. Sci 2^{eme} ser. **36**, 179.

- [25] Duhem P., 1906, *La théorie physique, son objet, et sa structure*, Chevalier et Rivière, Paris (engl. transl. 1954, *The aim and structure of physical theory*, Princeton University Press, Princeton).
- [26] Eagle A., 2004, *Twenty-one arguments against propensity*, *Erkenntnis* **60**, 371.
- [27] Eguchi S., 1983, *Second order efficiency of minimum contrast estimators in a curved exponential family*, *Ann. Statist.* **11**, 793. [euclid:1176346246](#).
- [28] Eguchi S., Copas J., 2006, *Interpreting Kullback–Leibler divergence with the Neyman–Pearson lemma*, *J. Multivar. Anal.* **97**, 2034.
- [29] Ellis R.L., 1843, *On the foundations of the theory of probabilities*, *Trans. Cambridge Phil. Soc.* **8**, 1.
- [30] Fienberg S.E., 2006, *Does it make sense to be an “objective bayesian”? (Comment on articles by Berger and by Goldstein)*, *Bayes. Anal.* **1**, 429.
- [31] Fisher R.A., 1922, *On the mathematical foundation of theoretical statistics*, *Phil. Trans. Roy. Soc. London Ser. A* **222**, 309.
- [32] Fleck L., 1929, *Zur Krise der “Wirklichkeit”*, *Naturwiss.* **17**, 425 (engl. transl.: 1986, *On the crisis of “reality”*, in: Cohen R.S., Schnelle T. (eds.), *Cognition and fact: materials on Ludwik Fleck*, Reidel, Dordrecht, p.47).
- [33] Fleck L., 1935, *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*, Schwabe, Basel. (engl. transl. 1979, *Genesis and development of the scientific fact*, Chicago University Press, Chicago).
- [34] Fleck L., 1936, *Zagadnienie teorii poznawania*, *Przeegl. Filozof.* **39**, 3 (engl. transl.: 1986, *The problem of epistemology*, in: Cohen R.S., Schnelle T. (eds.), *Cognition and fact: materials on Ludwik Fleck*, Reidel, Dordrecht, p.79).
- [35] Fremlin D.H., 2000, 2001, 2002, 2003, *Measure theory*, Vol.1-4, Torres Fremlin, Colchester.
- [36] Giere R.N., 1973, *Objective single-case probabilities and the foundations of statistics*, in: Suppes P. et al. (eds.), *Logic, methodology and philosophy of science IV*, North-Holland, Amsterdam, p.467.
- [37] Giffin A., 2009, *Maximum entropy: the universal method for inference*, Ph.D. thesis, State University of New York, Albany. [arXiv:0901.2987](#).
- [38] Grasselli M.R., 2001, *Classical and quantum information geometry*, PhD thesis, King’s College, London.
- [39] Hájek A., 1997, *Mises redux’ – redux: fifteen arguments against finite frequentism*, *Erkenntnis* **45**, 209.
- [40] Hájek A., 2007, *The reference class problem is your problem too*, *Synthese* **156**, 563.
- [41] Hájek A., 2009, *Interpretations of probability*, in: Zalta E.N. (ed.), *Stanford encyclopedia of philosophy*, Stanford. Available at: plato.stanford.edu/entries/probability-interpret.
- [42] Hinkelmann K., Kempthorne O., 2005, *Design and analysis of experiments*, Vol.1-2, Wiley, Hoboken (2nd ed. of Vol.1, 2008).
- [43] Howson C., Urbach P., 1989, *Scientific reasoning: the bayesian approach*, Open Court, Chicago (2nd. ed., 1993).
- [44] Hume D., 1748, *Philosophical essays concerning human understanding*, Millar, London.
- [45] Humphreys P., 1985, *Why propensities cannot be probabilities*, *Phil. Rev.* **94**, 557.
- [46] Jaynes E.T., 1976, *Confidence intervals vs bayesian intervals*, in: Harper W.L., Hooker C.A. (eds.), *Foundations of probability theory, statistical inference, and statistical theories of science*, Reidel, Dordrecht, p.175.
- [47] Jaynes E.T., 2003, *Probability theory: the logic of science*, Cambridge University Press, Cambridge.
- [48] Jeffrey R.C., 1977, *Mises redux*, in: Butts R.E., Hintikka J. (eds.), *Basic problems in methodology and linguistics*, Reidel, Dordrecht, p.213.
- [49] Jeffreys H., 1931, *Scientific inference*, Cambridge University Press, London.

- [50] Jeffreys H., 1955, *The present position in probability theory*, Brit. J. Phil. Sci. **5**, 275.
- [51] Johnson R.W., Shore J.E., 1983, *Comments and correction to “Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy”*, IEEE Trans. Inf. Theory **29**, 942.
- [52] Johnson W.E., 1921, *Logic*, Cambridge University Press, Cambridge.
- [53] Jones L.K., Byrne C.L., 1990, *General entropy criteria for inverse problems, with applications to ada compression, pattern classification and cluster analysis*, IEEE Trans. Inf. Theory **36**, 23.
- [54] Kadane J.B., 2006, *Is “objective bayesian analysis” objective, bayesian, or wise? (comment on articles by Berger and by Goldstein)*, Bayes. Anal. **1**, 433.
- [55] Kass R.E., Wasserman L., 1996, *The selection of prior distributions by formal rules*, J. Amer. Stat. Assoc. **91**, 1343. (corr.: 1998, J. Amer. Stat. Assoc. **93**, 412).
- [56] Keynes J.M., 1921, *A treatise on probability*, Macmillan, London.
- [57] Knuth K.H., 2009, *Measuring on lattices*, [arXiv:0909.3684](https://arxiv.org/abs/0909.3684).
- [58] Knuth K.H., Skilling J., 2010, *Foundations of inference*, [arXiv:1008.4831](https://arxiv.org/abs/1008.4831).
- [59] Kolmogorov A.N., 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und Ihrer Grenzgebiete Bd. **2**, Springer, Berlin. (engl. transl. 1950, *Foundations of the theory of probability*, Chelsea, New York).
- [60] Kostecki R.P., 2010, *Quantum theory as inductive inference*, in: Mohammad-Djafari A., Bercher J., Bessière P. (eds.), *Proceedings of the 30th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIP Conf. Proc. **1305**, Springer, Berlin, p.24. [arXiv:1009.2423](https://arxiv.org/abs/1009.2423).
- [61] Kuhn T.S., 1962, *The structure of scientific revolutions*, Chicago University Press, Chicago.
- [62] Kullback S., 1959, *Information theory and statistics*, Wiley, New York.
- [63] Lad F., 2006, *Objective bayesian statistics... Do you buy it? Should we sell it? (comment on articles by Berger and by Goldstein)*, Bayes. Anal. **1**, 441.
- [64] Le Cam L., 1964, *Sufficiency and approximate sufficiency*, Ann. Math. Statist. **35**, 1419.
- [65] Le Cam L., 1986, *Asymptotic methods in statistical decision theory*, Springer, Berlin.
- [66] Loomis L.H., 1947, *On the representation of σ -complete boolean algebras*, Bull. Amer. Math. Soc. **53**, 757.
- [67] McCullagh P., 2002, *What is a statistical model?*, Ann. Stat. **30**, 1225. Available at: www.stat.uchicago.edu/~pmcc/pubs/AOS023.pdf.
- [68] Neyman J., 1937, *Outline of a theory of statistical estimation based on the classical theory of probability*, Phil. Trans. Roy. Soc. London A **236**, 333.
- [69] Paris J., Vecovská A., 1989, *Maximum entropy and inductive inference*, in: Skilling J. (ed.), *Maximum entropy and bayesian methods*, Kluwer, Dordrecht, p.397.
- [70] Paris J., Vecovská A., 1989, *A note on the inevitability of maximum entropy*, Int. J. Inexact Reas. **4**, 183.
- [71] Peirce C.S., 1910, *Notes on the doctrine of chances*, in: Hartshorne C., Weiss P. (eds.), 1931, *Collected papers of Charles Sanders Peirce*, vol.2, p.405.
- [72] Poincaré H., 1902, *La science et l’hypothese*, Flammarion, Paris.
- [73] Poincaré H., 1908, *Science et méthode*, Flammarion, Paris.
- [74] Popper K.R., 1957, *The propensity interpretation of the calculus of probability and the quantum theory*, in: Körner S. (ed.), *Observation and interpretation: a symposium of philosophers and physicists. Proceedings of the Ninth Symposium of the Colston Research Society held in the University of Bristol, April 1st–April 4th, 1957*, Butterworth Scientific, London, p.65.
- [75] Popper K.R., 1959, *The propensity interpretation of probability*, Brit. J. Phil. Sci. **10**, 25.
- [76] Pratt J.W., 1961, *Review of ‘Testing statistical hypothesis’ (Lehman, 1959)*, J. Am. Stat. Assoc. **56**, 163.

- [77] Ramsey F.P., 1931, *The foundations of mathematics and other logical essays*, Routledge and Kegan, London.
- [78] Rodríguez C.C., 2003, *A geometric theory of ignorance*, preprint, available at: omega.albany.edu:8008/ignorance/ignorance03.pdf.
- [79] Savage L.J., 1954, *The foundations of statistics*, Dover, New York. (1972, 2nd ed.).
- [80] Segal I.E., 1951, *Equivalences of measure spaces*, Amer. J. Math. **73**, 275.
- [81] Shirayayev A.N., Koshevnik Yu.A., 1989, *Review of: Asymptotic methods in statistical decision theory by Lucien Le Cam*, Bull. Amer. Math. Soc. **20**, 280 (engl. transl.).
- [82] Shore J.E., Johnson R.W., 1980, *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*, IEEE Trans. Inf. Theory **26**, 26.
- [83] Shore J.E., Johnson R.W., 1981, *Properties of cross-entropy minimization*, IEEE Trans. Inf. Theory **27**, 472.
- [84] Sikorski R., 1948, *On the representation of boolean algebras as fields of sets*, Fund. Math. **35**, 247.
- [85] Skilling J., 1988, *The axioms of maximum entropy*, in: Erickson G.J. and Smith C.R. (eds.), *Maximum-entropy and bayesian methods in science and engineering*, Kluwer, Dordrecht, p. 173.
- [86] Spengler O., 1918, 1923, *Der Untergang des Abendlandes. Umriße einer Morphologie der Weltgeschichte*, Vol.1-2, Braumüller, Wien. (2nd rev. ed. of Vol.1: 1922, Beck, München; engl. transl.: 1926, *The decline of the west*, Knopf, New York).
- [87] Uffink J.B.M., 1995, *Can the maximum entropy principle be explained as a consistency requirement?*, Stud. Hist. Phil. Mod. Phys. B **26**, 223.
- [88] Uffink J.B.M., 1996, *The constraint rule of the maximum entropy principle*, Stud. Hist. Phil. Mod. Phys. **27**, 47. Available at: www.phys.uu.nl/igg/jos/mep2def/mep2def.pdf.
- [89] van Lambalgen M., 1987, *Von Mises' notion of random sequence reconsidered*, J. Symb. Log. **52**, 725.
- [90] Venn J., 1866, *The logic of chance. An essay on the foundations and province of the theory of probability, with especial reference to its application to moral and social science*, MacMillan, London.
- [91] von Mises R., 1919, *Grundlagen der Wahrscheinlichkeitsrechnung*, Math. Z. **5**, 52.
- [92] von Mises R., 1928, *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit*, Springer, Berlin. (1972, 4th ed.; engl. transl. 1939, *Probability, statistics and truth*, Dover, New York).
- [93] Whittle P., 1970, *Probability*, Penguin, Harmondsworth.
- [94] Whittle P., 2000, *Probability via expectation*, (4th ed.), Springer, Berlin.
- [95] Zabell S.L., 2009, *Carnap and the logic of inductive inference*, in: Gabbay D.M., Hartmann S., Woods J. (eds.), *Handbook of the history of logic. Volume 10: Inductive logic*, Elsevier, Amsterdam, p.265.
- [96] Zhu H., Rohwer R., 1995, *Information geometric measurements of generalisation*, Tech. Rep. Neural Comp. Res. Group **4350**, Aston. Available at: eprints.aston.ac.uk/507/1/NCRG_95-005.pdf.
- [97] Zhu H., Rohwer R., 1998, *Information geometry, bayesian inference, ideal estimates and error decomposition*, Santa Fe Inst. Tech. Rep. **98-06-45**, Santa Fe. Available at: omega.albany.edu:8008/ignorance/zhu98.pdf.