

<https://helda.helsinki.fi>

Approaching probabilistic laws

Niiniluoto, Ilkka

2021-12-30

Niiniluoto , I 2021 , ' Approaching probabilistic laws ' , Synthese , vol. 199 , no. 3-4 , pp. 10499-10519 . <https://doi.org/10.1007/s11229-021-03256-8>

<http://hdl.handle.net/10138/341377>

<https://doi.org/10.1007/s11229-021-03256-8>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.



Approaching probabilistic laws

Ilkka Niiniluoto¹

Received: 27 August 2020 / Accepted: 8 June 2021 / Published online: 9 July 2021
© The Author(s) 2021

Abstract

In the general problem of verisimilitude, we try to define the distance of a statement from a target, which is an informative truth about some domain of investigation. For example, the target can be a state description, a structure description, or a constituent of a first-order language (Sect. 1). In the problem of legisimilitude, the target is a deterministic or universal law, which can be expressed by a nomic constituent or a quantitative function involving the operators of physical necessity and possibility (Sect. 2). The special case of legisimilitude, where the target is a probabilistic law (Sect. 3), has been discussed by Roger Rosenkrantz (*Synthese*, 1980) and Ilkka Niiniluoto (*Truthlikeness*, 1987, Ch. 11.5). Their basic proposal is to measure the distance between two probabilistic laws by the Kullback–Leibler notion of divergence, which is a semimetric on the space of probability measures. This idea can be applied to probabilistic laws of coexistence and laws of succession, and the examples may involve discrete or continuous state spaces (Sect. 3). In this paper, these earlier studies are elaborated in four directions (Sect. 4). First, even though deterministic laws are limiting cases of probabilistic laws, the target-sensitivity of truthlikeness measures implies that the legisimilitude of probabilistic laws is not easily reducible to the deterministic case. Secondly, the Jensen–Shannon divergence is applied to mixed probabilistic laws which entail some universal laws. Thirdly, a new class of distance measures between probability distributions is proposed, so that their horizontal differences are taken into account in addition to vertical ones (Sect. 5). Fourthly, a solution is given for the epistemic problem of estimating degrees of probabilistic legisimilitude on the basis of empirical evidence (Sect. 6).

Keywords Divergence · Legisimilitude · Nomic constituent · Probabilistic law · Truthlikeness · Verisimilitude

This article belongs to the topical collection "Approaching Probabilistic Truths", edited by Theo Kuipers, Ilkka Niiniluoto, and Gustavo Cevolani.

✉ Ilkka Niiniluoto
ilkka.niiniluoto@helsinki.fi

¹ Department of Philosophy, History, and Art Studies, University of Helsinki, Unioninkatu 40 A, P. O. BOX 24, 00014 Helsinki, Finland

1 The similarity approach to truthlikeness

Karl Popper's (1963) original work on truthlikeness was based on the concepts of truth value (true or false) and logical deduction (entailment). Theories were represented as deductively closed sets of sentences in some language L , and the comparative notion "more truthlike" was characterized by set-theoretical comparisons of the truth content and falsity content of rival theories. The main lesson from the dramatic failure of Popper's definition in 1974 was the need to add the notion of similarity or resemblance to the logical toolbox. In the first formulations of the *similarity approach*, Risto Hilpinen (1976) represented theories as classes of possible worlds and employed spheres of similarity from David Lewis' approach to counterfactuals, while Pavel Tichý defined theories as disjunctions of propositional constituents and Ilkka Niiniluoto as disjunctions of monadic constituents, adding a function to measure the distance between constituents. Soon these notions were extended to full first-order languages by Tichý, Niiniluoto, Raimo Tuomela and Graham Oddie, with systematic summaries in Oddie's *Likeness to Truth* (1986), Niiniluoto's *Truthlikeness* (1987), and Theo Kuipers' edition *What Is Closer-to-the-Truth?* (1987).

In Hilpinen's treatment, the truthlikeness of a theory depends on its maximum and minimum distances from the actual world. Tichý and Oddie favor the average distance, while Niiniluoto combines the minimum distance with the normalized sum of all distances. In the linguistic formulations, the *degree of truthlikeness* $\text{Tr}(H, C^*)$ of a theory H in language L depends on the similarity of the disjuncts of H with the true constituent C^* of L . Here the target C^* is the most informative truth expressible in the conceptual framework L , and $\text{Tr}(H, C^*)$ is maximal when H is identical with this complete truth C^* .¹

A successful theory H should give a full and correct description of a domain of investigation by its conceptual resources in language L . In other words, H should specify the L -structure of the actual world with respect to L , and strongest theories are able to do this up to isomorphism. Here L may include qualitative or quantitative concepts. But the choice of the logical complexity of language L allows a finer discrimination: the *target* can be chosen for the purposes of the relevant cognitive problem, so that it may be a propositional constituent, a state description, a structure description, a monadic constituent, a polyadic constituent of depth- d , or a complete first-order theory.² For each of these choices, the task is to define the distances of statements in L from the given target.

Following Popper, one should distinguish here two problems. In the *logical problem* of truthlikeness, we are given the true target C^* , and we ask what it means to say that a theory is close to C^* or closer to C^* than another theory. In the *epistemic problem* of truthlikeness, the true target C^* is unknown, and we ask how we can rationally claim or estimate on available evidence E that one theory is close to C^* or closer to C^* than another theory.

¹ For surveys, see Niiniluoto (1998), Oddie (2016).

² For details, see Niiniluoto (1987), pp. 204–208. Most of these logical concepts were developed for the purposes of inductive logic by Rudolf Carnap and Jaakko Hintikka (cf. Niiniluoto, 2011).

To illustrate the similarity approach to these two problems, let L be a monadic first-order language with k one-place predicates, and let $\mathbf{Q} = \{Q_1, \dots, Q_K\}$ be the Q -predicates of L ($K = 2^k$). The Q -predicates can be defined by the conjunction of negated or unnegated primitive predicates of L , so that there is a natural distance $\rho_{uv} = d(Q_u, Q_v)$ between them.³ The Q -predicates are the strongest predicates expressible in L , and they constitute a classification system of individuals in the domain of L . A *state description* in L locates each individual in one and only one “cell” defined by a Q -predicate, while a *structure description* specifies the proportions of individuals in these cells. A *monadic constituent* C_i of L specifies which Q -predicates are empty and which are non-empty:

$$C_i = (+/-)(Ex)Q_1(x) \& \dots \& (+/-)(Ex)Q_K(x) \tag{1}$$

where $(+ / -)$ is replaced by negation or nothing. As an empty universe is excluded, the number of constituents is $q = 2^K - 1$. If CT_i is the class of occupied cells by C_i , then (1) can be rewritten in the following form:

$$C_i = \prod_{j \in CT_i} (Ex)Q_j(x) \& (x) [\bigvee_{j \in CT_i} Q_j(x)]. \tag{2}$$

If for the true constituent C^* there are no empty cells, so that $CT^* = \mathbf{Q}$, the world is atomistic in the sense that there are no true universal generalizations. For example, the truth of the generalization $(x)(Fx \rightarrow Gx)$ means that the cell $F \& \sim G$ is empty. A simple distance between monadic constituents is the *Clifford-measure*:

$$\Delta_C(C_i, C_j) = |CT_i \Delta CT_j| / K = \text{the number of disagreements of } C_i \text{ and } C_j, \tag{3}$$

where Δ is the symmetric difference (see Fig. 1). Variants of (3), which take into account distances between Q -predicates, have been considered by Tichý, Oddie, and Niiniluoto.⁴ Then the degree of truthlikeness $\text{Tr}(H, C^*)$ of a generalization H in L depends on the Clifford-distances (or their variants) of the disjuncts of H from the true constituent C^* . A comparative notion “ H_1 is closer to the truth than H_2 ” is explicated by the condition $\text{Tr}(H_1, C^*) > \text{Tr}(H_2, C^*)$.

If C^* is unknown, but a rational epistemic probability measure P is defined over the class of constituents of L , then the unknown degree $\text{Tr}(H, C^*)$ can be estimated by its expected value on the basis of evidence E :

$$\text{ver}(H/E) = \sum P(C_i/E) \text{Tr}(H, C_i), \tag{4}$$

³ This is the Manhattan or Hamming distance between Q_u and Q_v (i.e., the number of disagreements in their definition), but normalized to take values between 0 and 1. Q -predicates can also be defined by Carnapian families of predicates, and their distance by the Euclidean metric. See Niiniluoto (1987), pp. 44–47.

⁴ See Tichý (1976). For discussion, see Oddie (1986), pp. 99–105, Niiniluoto (1987), pp. 310–335. Cf. Section 5 below.

where the sum goes over $i = 1, \dots, q$.⁵ For monadic languages, the relevant posterior probabilities $P(C_i/E)$ of constituents (i.e., degrees of belief on the truth of C_i given E) are given by Jaakko Hintikka's system of inductive logic.⁶

2 Verisimilitude vs. Legisimilitude

Following the difference between accidental and lawlike generalizations, a distinction between verisimilitude and legisimilitude has been proposed by L. J. Cohen. In the logical problem of *legisimilitude*, the target is not just the strongest true statement about the world (in a given language), but a genuine law of nature. A solution of this problem for universal or deterministic laws can be based on S. Uchii's notion of nomic constituent.⁷ Let $L(\Box)$ be a modal monadic language with the operators of nomic necessity \Box and nomic possibility \Diamond , satisfying the system S5. Then a *nomic constituent* tells which Q-predicates are possible and which are impossible:

$$B_i = \prod_{j \in CT_i} \Diamond(Ex)Q_j(x) \& \Box(x) \left[\bigvee_{j \in CT_i} Q_j(x) \right]. \quad (5)$$

The number of nomic constituents in $L(\Box)$ is again $q = 2^K - 1$. As actuality implies possibility, and impossibility implies non-actuality, nomic constituent (5) is partly weaker and partly stronger than constituent (3). Laws of nature are disjunctions of nomic constituents. For example, the law $\Box(x)(Fx \rightarrow Gx)$ is equivalent to the disjunction of all nomic constituents which state that the cell $F \& \sim G$ is physically impossible. The distance $\Delta(B_1, B_2)$ between nomic constituents B_1 and B_2 can be defined by the Clifford-measure $|CT_1 \Delta CT_2|/K$ or its variants. The *degree of legisimilitude* of a law of nature H depends on its distance to the true nomic constituent B^* :

$$\text{leg}(H, B^*) = 1 - \Delta(H, B^*).$$

Alternatively, if the cognitive aim is to combine verisimilitude and legisimilitude, the target could be the conjunction $B^* \& C^*$.⁸ Estimation of legisimilitude can again employ expected values based on inductive probabilities.⁹

Nomic constituents (5) represent *laws of coexistence*, i.e., lawlike connections between attributes or properties. To define *laws of succession*, introduce a discrete temporal index t to Q-predicates:

⁵ This solution to the epistemic problem was proposed by Niiniluoto in 1977 (cf. Niiniluoto, 1987, Ch. 7).

⁶ For a survey, see Niiniluoto (2011).

⁷ See Niiniluoto (1987), pp. 91–98. Nomic constituents correspond to what Kuipers (1982) called "theoretical truth" (as opposed to "descriptive truth") and later "nomic truth" (see Kuipers, 2019).

⁸ See Niiniluoto (1987), p. 377.

⁹ For monadic nomic constituents, the relevant posterior probabilities are again obtained from Hintikka's inductive logic (see Niiniluoto, 1987, pp 98–102).

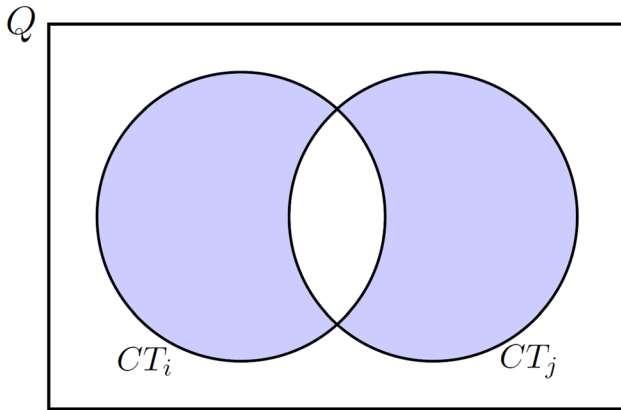


Fig. 1 Clifford-distance between monadic constituents

$$\prod_{\langle i,j \rangle \in T} \diamond(Ex)(Q'_i(x) \& Q'_j(x)) \& \square(x) [\vee_{\langle i,j \rangle \in T} (Q'_i(x) \& Q'_j(x))]. \tag{6}$$

Here T lists all possible transitions between successive states. For deterministic laws, for each i there is only one j such that $\langle i,j \rangle \in T$. Again the Clifford-measure can be applied to measure the distance between laws of succession: $|T_1 \Delta T_2| / K^2$.¹⁰

The class of Q-predicates of a monadic language can be generalized to a quantitative *state space* $Q \subseteq R^k$ generated by real-valued quantities h_1, \dots, h_k .¹¹ In the simplest case, Q is the real line or its part with the geometrical distance between points $\rho(x, x') = |x - x'|$. More generally, Q is a k-dimensional metric space with the Euclidean metric. Here laws of coexistence specify regions of nomically possible states

$$F(x) = \{x \in Q \mid f(h_1(x), \dots, h_k(x)) = 0 \}.$$

The Clifford-distance between two such laws F_1 and F_2 is defined by

$$\int_{F_1 \Delta F_2} dQ.$$

An alternative approach to quantitative laws expresses how a function h_k necessarily depends on h_1, \dots, h_{k-1} :

$$h_k(x) = g(h_1(x), \dots, h_{k-1}(x)).$$

The distances between two such real-valued functions can be defined by the Minkowski or L^p -metrics for functions:

¹⁰ See Niiniluoto (1987), p. 379.

¹¹ See Niiniluoto (1987), Ch. 3.

$$\Delta_p(f, g) = (\int |f(x) - g(x)|^p)^{1/p}. \quad (7)$$

Here $p=1$ is the Manhattan metric, $p=2$ the Euclidean metric, and $p=\infty$ the Tchebycheff metric $\sup |f(x) - g(x)|$.¹² The degree of legisimilitude of the law f then depends on its distance to the true law f^* :

$$\text{leg}(f, f^*) = \frac{1}{1 + \Delta(f, f^*)}.$$

Further, f is closer to the truth than g if and only if $\text{leg}(f, f^*) > \text{leg}(g, f^*)$.

Quantitative laws of succession can be formulated by relativizing the state x with a time: $h(x, t)$ = state of x at time t . Then deterministic *dynamical laws* tell how the state depends on time t and some initial state at time t_0 :

$$(h_1(x, t), \dots, h_k(x, t)) = F(t, h_1(x, t_0), \dots, h_k(x, t_0)).$$

A law of succession specifies nomically possible trajectories $F : \mathbf{R} \times \mathbf{Q} \rightarrow \mathbf{Q}$. The distance between such laws can be defined by taking for each $Q \in \mathbf{Q}$ the Minkowski distance between the trajectories $F_1(t, Q)$ and $F_2(t, Q)$, for $t \in \mathbf{R}$, and then summing over all possible initial states $Q \in \mathbf{Q}$.¹³

3 Probabilistic laws

The notion of a universal or deterministic law introduced in Sect. 2 can be generalized to *probabilistic laws*, if an objective physical probability measure is available.¹⁴ Following Leibniz, such physical probabilities express “degrees of possibility”. This can be understood in terms of single-case propensities: $P(G/F) = r$ means that a physical set-up has a numerical disposition of strength r to produce an outcome of type G in each trial of type F . Thus, probability statements involve a dispositional modal operator, so that they differ from extensional statistical statements about actual relative frequencies of attributes in reference classes (i.e., structure descriptions). Universal laws of coexistence and deterministic laws of succession are limiting special cases of probabilistic laws (with propensities 0 and 1).¹⁵ Genuine probabilistic laws presuppose that the world is indeterministic, but in statistical modelling one may assign in some sense objective probabilities to random phenomena (e.g., coin tossing, roulette) even when the underlying reality is deterministic. For the task

¹² See Niiniluoto (1987), p. 385. The metric (7) is based on the differences between the values of two functions, but it does not reflect the similarity of their mathematical form (see Niiniluoto, 2019, p. 131). For a proposal to measure the distance between quantitative laws as a combination of accuracy and nomicity, see Garcia Lapeña (2021).

¹³ See Niiniluoto (1987), p. 393.

¹⁴ See Niiniluoto (1987), pp. 118–121.

¹⁵ For probabilistic laws with single-case propensities, see Fetzer (1981).

of defining approximation to such probabilities, the philosophical issue of indeterminism and determinism can be left open.

To define *probabilistic constituents*, replace in a nomic constituent (5) the operator of physical possibility \diamond with a probability measure P over the discrete state space \mathbf{Q} of Q -predicates, now understood as the “sample space” or the class of outcomes of a trial x :

$$\prod_{Q_i \in \mathbf{Q}} [P(Q_i(x)) = p_i]. \tag{8}$$

Probabilities (8) over \mathbf{Q} define a multinomial context. Then $p_i = P(Q_i(x)) > 0$ if and only if Q_i is physically possible, for all $i = 1, \dots, K$, so that here P is applied to the open formula $Q_i(x)$ instead of the existential statement in (5). Now a probabilistic constituent (8) is compatible with the nomic constituent B_i with CT_i if and only if it assigns a positive probability to the Q -predicates in CT_i and zero probability to other Q -predicates. This means that typically a nomic constituent is an infinite disjunction of probabilistic constituents.

In many statistical applications, the trial x counts the number of successes in a repeated experiment (e.g., binomial and Poisson distributions), so that the state space \mathbf{Q} is a subclass of the set \mathbf{N} of natural numbers. Then the distance $\rho(x, x')$ between points in \mathbf{Q} is their normalized arithmetical difference.

An important distinction can be made between pure and mixed probabilistic laws. A probabilistic constituent, where CT_i is a proper subset of \mathbf{Q} , is a *mixed* law in the sense that it entails a universal law (cells in $\mathbf{Q}-CT_i$ are necessarily empty). *Pure* probabilistic laws have no such entailments: the world is atomistic in the sense that all Q -predicates in \mathbf{Q} are nomically possible (so that no universal laws hold), and a positive probability is assigned to all Q -predicates.

To define probabilistic laws of succession, for a discrete space \mathbf{Q} the set T of possible transitions between states is replaced by a matrix of transition probabilities

$$p_{j/i} = P(Q_j^{t+1}(x)/Q_i^t(x)), \tag{9}$$

where $p_{1/i} + \dots + p_{K/i} = 1$ for each i . This definition involves the Markov condition, i.e., the next state depends only on the present state. If transition probabilities are 0 or 1, this law reduces to the deterministic law (6).¹⁶ Equation (9) determines the n -step transition probabilities

$$p_{j/i}(n) = P(Q_j^{t+n}(x)/Q_i^t(x)),$$

and for an irreducible stationary Markov chain the limits of $p_{j/i}(n)$, for $n \rightarrow \infty$, give a long-run probability distribution. These notions can be generalized to Markov

¹⁶ Discrete state systems, both deterministic and probabilistic, with a Markov condition have been studied by Rescher (1970).

processes with a continuous time.¹⁷ Special cases of probabilistic laws of succession can be formulated by quantitative dynamic laws like the law of radioactive decay

$$P(Q(x, t)) = 1 - e^{-\lambda t}, \quad (10)$$

where $Q(x, t)$ states that atom x decays within the time-interval $[0, t]$ and λ is a constant. Finally, for a quantitative state space \mathbf{Q} a probability measure on \mathbf{Q}^∞ (i.e., infinite sequences of successive states) assigns a physical probability to possible trajectories of a time-continuous stochastic process.

4 Distance between probabilities

The problem of legisimilitude for probabilistic laws has not yet received much attention. The main focus in the literature has been on cases, where the target is a universal or deterministic law, either qualitative or quantitative. The only detailed proposals have been given by Rosenkrantz (1980) and Niiniluoto (1987), 403–405, who apply the Kullback–Leibler notion of divergence as a measure of distance from a probabilistic truth.

Mathematicians have suggested a great number of measures for distances between probability distributions. In a comprehensive survey, Cha (2007) lists 45 different measures,¹⁸ which have been used for various purposes. For example, the central limit theorem (the sum of n independent random variables approximates in the limit the normal distribution) and laws of large numbers (observed relative frequencies and predictive probabilities approach almost surely objective probabilities in a multinomial Bernoulli process) express distances between epistemic probabilities q and objective probabilities p by their geometrical distance $|q-p|$.¹⁹ This amounts to the Manhattan metric

$$\Delta_1(p, q) = \sum |p_i - q_i|.$$

For discrete probabilities, the squared Euclidean or quadratic metric

$$\Delta_2(p, q) = \sum (p_i - q_i)^2,$$

or its variant χ^2 , is a standard way of measuring the fit between two distributions or structural descriptions.²⁰ In the special case of *scoring*, where q_i are probabilistic estimates of the truth values p_i of n rival exclusive hypotheses ($p_j = 1$, otherwise 0), Glenn Brier's 1950 measure of inaccuracy is quadratic, i.e., $d(1, q) = (1-q)^2$

¹⁷ See Parzen (1962), pp. 248, 277.

¹⁸ Cf. Niiniluoto (1987), pp. 7–8.

¹⁹ See Festa (1993).

²⁰ See Niiniluoto (1987), pp. 15–16, pp. 302–303, pp. 321–322.

and $d(0, q) = q^2$,²¹ while I. J. Good in 1952 favored the logarithmic measure $d(1, q) = -\ln q$, $d(0, q) = -\ln(1 - q)$.²² From these local measures the total scoring measure is obtained by summing the inaccuracies of all q_i .

Some measures are based on the inner products $p_i q_i$. Hellinger’s 1909 proposal was modified in 1946 in Bhattacharyya’s dissimilarity coefficient:

$$-\log \sum \sqrt{p_i q_i}.$$

The *directed divergence* of a discrete random variable p from another q was defined by Solomon Kullback and Richard Leibler in 1951 as the expected logarithmic difference between p and q with respect to p :

$$\text{div}(p, q) = \sum p_i \log(p_i/q_i). \tag{11}$$

Here \log can be taken to have the binary base 2, so that $\log 2 = 1$. Formula (11) is also called the relative entropy of p with respect to q . This measure is only a semimetric: non-negative, $\text{div}(p, p) = 0$, $\text{div}(p, q) = 0$ if and only if $p = q$, but non-symmetric (usually $\text{div}(p, q) \neq \text{div}(q, p)$), and the triangle equation is not satisfied. For continuous probability densities f and g on \mathbb{R} , Eq. (11) is replaced by

$$\text{div}(f(x), g(x)) = \int_{-\infty}^{+\infty} f(x) \log \left[\frac{f(x)}{g(x)} \right] dx.$$

The *symmetric divergence* between p and q is defined by

$$\text{div}_s(p, q) = \text{div}(p, q) + \text{div}(q, p) = \sum (p_i - q_i) \log (p_i/q_i).$$

Other variants include the λ -divergence

$$\text{div}_\lambda(p, q) = \lambda \text{div}(p, \lambda p + (1 - \lambda)q) + (1 - \lambda) \text{div}(q, \lambda p + (1 - \lambda)q)$$

For $\lambda = 1/2$, it gives the *Jensen-Shannon divergence*

$$\begin{aligned} \text{div}_{JS}(p, q) &= \frac{1}{2} \text{div}(p, \frac{p+q}{2}) + \frac{1}{2} \text{div}(q, \frac{p+q}{2}) = \frac{1}{2} H(\frac{p+q}{2}) - \frac{H(p) + H(q)}{2} \\ &= \frac{1}{2} \sum p_i \log p_i + \frac{1}{2} \sum q_i \log q_i - \frac{1}{2} \sum (p_i + q_i) \log [(p_i + q_i)/2] \\ &= \frac{1}{2} \sum p_i \log [2p_i/(p_i + q_i)] + \frac{1}{2} \sum q_i \log [2q_i/(p_i + q_i)]. \end{aligned} \tag{12}$$

where $H(p) = -\sum p_i \log p_i$ is the Shannon entropy of p . Formula (12) is non-negative and symmetric, and its square root is a metric. Renyi divergence is defined by

²¹ See Pettigrew (2015). Oddie (2019) accepts the quadratic scoring rule, but argues that it is incompatible with the principles of truthlikeness.
²² See McCutcheon (2019).

$$D_{\alpha}(p, q) = \frac{1}{\alpha - 1} \log \sum (p_i^{\alpha} / q_i^{\alpha-1}).$$

For $\alpha=1$, it gives in the limit the Kullback–Leibler divergence, and for $\alpha=1/2$ twice the Bhattacharyya distance.

Divergence was originally intended as a tool in information theory.²³ In Bayesian statistics it has been used to measure the difference between prior and posterior distributions. It can be also used for assessing the similarity of a probabilistic model with some aspect of reality.²⁴ The first connection to the studies in truthlikeness was developed by Roger Rosenkrantz (1980). Inspired by I. J. Good's notion of the weight of evidence, Rosenkrantz suggested that for a random experiment x and truth h^* , hypothesis h is more truthlike than hypothesis h' if

$$\text{div}(P(x/h), P(x, h^*)) < \text{div}(P(x/h'), P(x/h^*)).$$

This idea was connected to the similarity approach to truthlikeness by Niiniluoto (1987): the distance of a probabilistic hypothesis h from the probabilistic target h^* is measured by $\text{div}(h^*, h)$. Thus, hypothesis h is more truthlike than h' if and only if $\text{div}(h^*, h) < \text{div}(h^*, h')$. When the relevant hypotheses h , h' , and h^* are specified by probabilities p_i , q_i , and p_i^* , this comparative condition holds if and only if $\sum p_i^* \log(q_i/p_i) < 0$. Generalization to probability density functions is immediate.

More generally, Niiniluoto (1987) recommends divergence div as a solution to the problem of probabilistic legisimilitude:

- for probabilistic laws of coexistence (8) in qualitative conceptual spaces, the distance to the true probabilistic constituent
- for probabilistic laws of succession (9) in qualitative languages, the distance to the matrix of true probability transitions
- for probabilistic laws of succession in the quantitative space \mathbf{Q}^{∞} , the distance to the true probability on \mathbf{Q}^{∞} .

Alternative solutions could replace div by some other distance measure, e.g., Manhattan, Euclidean, or Bhattacharyya.

The Kullback–Leibler divergence $\text{div}(p, q)$ has a limitation which is not noted in Niiniluoto (1987). Its definition (11) presupposes that p is absolutely continuous with respect to q , i.e., if $q_i=0$, then $p_i=0$. Further, when $p_i=0$, the factor $0 \log 0$ in the sum vanishes. The same condition is required for probability densities: if $g(x)=0$, then $f(x)=0$. This means that the KL-divergence can be applied only to pure probabilistic laws, since for mixed probabilistic constituents mistakes in the empty cells (or zero points in \mathbf{Q}) in the target and hypothesis would not count at all. The same problem is faced by the Bhattacharyya distance, whose factors vanish as soon as p_i or q_i is 0, but not by the Minkowski metrics.

²³ See Kullback (1959).

²⁴ For example, Sober (2002) follows the statistician H. Akaike in measuring the distance between a fitted model (with fixed parameter values) and the truth by the Kullback–Leibler distance.

This problem with divergence is observed by Rosenkrantz (1980), who suggests that in the evaluation of $\text{div}(P(x, h), P(x, h^*))$ zero probabilities are replaced by a slightly positive possibility of misclassification, but this ad hoc move is unsatisfactory. As a better solution one can recommend the use of the Kullback–Leibler directed divergence div for pure probabilistic laws, and the Jensen-Shannon divergence div_{JS} (instead of div) to measure the distance between mixed probability distributions over cells \mathbf{Q} or in the transition matrix. The JS-divergence shares the good properties of the KL-divergence, but it has a finite value in all cases even when some of the probabilities are zero.²⁵

The following examples illustrate various possibilities in analyzing approach to probabilistic laws.

Example 1. Let L be a monadic language with two primitive predicates $Fx = x$ is a swan and $Gx = x$ is white. Then there are four Q -predicates in L :

$$Q_1x = Fx \ \& \ Gx$$

$$Q_2x = Fx \ \& \ \sim Gx$$

$$Q_3x = \sim Fx \ \& \ Gx$$

$$Q_4x = \sim Fx \ \& \ \sim Gx.$$

Then the true constituent C^* in L states that all Q -predicates are instantiated. Let H be the false universal generalization “All swans are white”. H states that Q_2 is empty and leaves other cells as question marks. Applying the min-sum definition with weights γ and γ' for the min and sum factors, respectively, the degree of truthlikeness of H is²⁶

$$\text{Tr}(H, C^*) = 1 - \gamma/4 - 5\gamma'/8.$$

Choosing $\gamma = 2/3$ and $\gamma' = 1/3$, this is equal to $5/8$. The degree of truthlikeness of the false constituent C_1 with $CT_1 = \{Q_1, Q_3, Q_4\}$ is²⁷

$$\text{Tr}(C_1, C^*) = 1 - \gamma/4 - \gamma'/32 = 79/96 > 5/8.$$

The same numerical results hold for the nomic versions of C^* , H , and C_1 . In the probabilistic framework, H corresponds to the law $P(Gx/Fx) = 1$, but now the target is the true probability distribution P^* over the cells Q_1, \dots, Q_4 , and H is a disjunction of probabilistic constituents with probability 0 for Q_2 . As the number of black swans is small in comparison to white swans, the true probabilistic law is something like $P(Gx/Fx) = 0.95$. It follows, for any reasonable distance measure, that H has a

²⁵ div_{JS} is absolutely continuous, since in (12) $(p_i + q_i)/2 = 0$ implies $p_i = 0$ and $q_i = 0$.

²⁶ See formula (9.21) with $b = b' = 1$ and $q = 4$ in Niiniluoto (1987), p. 338.

²⁷ See formula (6.88) with $||l = 24$ and $\text{av}(*, B) = 1/2$ in Niiniluoto (1987), p. 229.

relatively high degree of truthlikeness, and in any case higher than that of the law $P(Gx/Fx) = 0.5$. But if the cognitive interest of the investigator is to know both the nomic and actual features of birds, so that the target is the conjunction P^* & C^* , then H 's overall truthlikeness is reduced, since it mistakenly excludes the cell Q_2 , while laws of the form $P(Gx/Fx) = r < 1$ allow for the actual existence of black swans.

Example 2. Already Example 1 illustrates the fact that the comparison of ordinary, nomic and probabilistic constituents is a complicated matter, as they involve different targets. For example, a probabilistic constituent P is equivalent to a single nomic constituent B only in the special case where just one cell Q_i is physically possible – and, hence, has probability 1. In other cases, the true nomic constituent B^* is an infinite disjunction of probabilistic constituents, and the target-sensitivity does not allow a direct comparison of the degrees of truthlikeness of these different types of hypotheses. In particular, the atomistic nomic constituent, which states the possibility of all Q -predicates, is the disjunction of all pure probabilistic laws. This means that there is no connection between divergence and Clifford-distance for pure probabilistic laws. To see this, assume that P_1 and P_2 are two different laws, and B_1 and B_2 are the nomic constituent entailed P_1 and P_2 . If P_1 and P_2 are pure laws, then $CT_1 = CT_2 = Q$ and $CT_1 \Delta CT_2 = \emptyset$, so that $\text{div}(P_1, P_2) > 0$ but the Clifford-distance $\Delta_C(B_1, B_2) = 0$.²⁸ But some simple comparisons can be made for the special case of uniform mixed laws. Thus, suppose C_1 and C_2 are monadic nomic constituents in a language with K Q -predicates with $|CT_1 - CT_2| = A$, $|CT_2 - CT_1| = B$, and $|CT_1 \cap CT_2| = D$, so that the Clifford distance Δ_C between C_1 and C_2 is $(A + B)/K$ (see (3)). Let P_1 and P_2 be probabilistic constituents which allocate probability uniformly to CT_1 and CT_2 ($1/c$ and $1/c'$, respectively), where $c' \geq c$. Now $D = c - A = c' - B$. Then the Manhattan distance satisfies

$$\Delta_1(P_1, P_2) = \frac{A}{c} + \frac{B}{c'} + \left(\frac{1}{c} - \frac{1}{c'}\right)D = \frac{1}{c'}(A + B) - \frac{c}{c'} + 1 = \frac{K}{c'} \Delta_C(C_1, C_2) + \frac{c' - c}{c'}.$$

If $c = c'$, this value equals $K \Delta_C(C_1, C_2)/c$. For the Euclidean distance with $c = c'$ we have

$$\Delta_2(P_1, P_2) = (A + B)/c^2 = K \Delta_C(C_1, C_2)/c^2.$$

A similar connection to the Clifford measure holds for the Jensen-Shannon divergence:

²⁸ Fig. 2 in Sect. 5 shows that symmetric difference still has an interesting connection to the distance between probability densities.

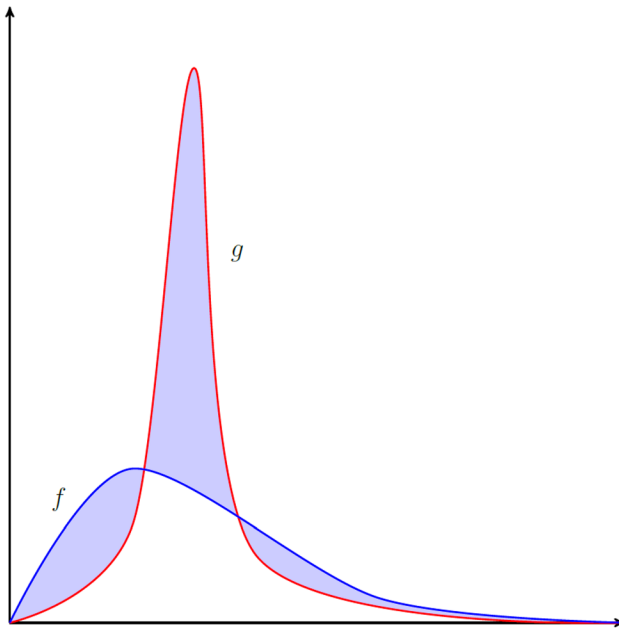


Fig. 2 Distance between probability densities f and g

$$\begin{aligned}
 \text{div}_{\text{JS}}(P_1, P_2) &= \frac{\log 2c}{2c}(A + B) + \frac{\log c}{c} D - \log c \\
 &= \frac{\log 2c}{2c}(A + B) - \left(1 - \frac{D}{c}\right) \log c \\
 &= \frac{\log 2c}{2c}(A + B) - \frac{\log c}{c} A \\
 &= \frac{\log 2}{c} A \\
 &= \frac{K}{2c} \Delta_C(C_1, C_2).
 \end{aligned}$$

However, such connections fail for non-uniform laws. For example, if nomic constituents B_1 and B_2 are otherwise almost equal, but B_1 makes correct possibility claims about cells Q_i with high true probability p_i^* while B_2 makes such claims about cells Q_j with low probability p_j^* , then it may happen that truthlikeness ordering is reversed when the target changes from B^* to P^* : $\Delta_C(B_1, B^*) > \Delta_C(B_2, B^*)$, but $d(B_1, P^*) < d(B_2, P^*)$.²⁹

²⁹ Festa (2007) has proposed a way of measuring the distance between a monadic generalization and “the statistical truth” (i.e., true probabilistic constituent). The idea, roughly speaking, is to divide Q-predicates into “statistically common” and “rare” ones, and then demand that a truthlike generalization should make true existential claims about common predicates and false exclusion claims only about rare predicates.

Example 3 If p and q are disjoint mixed probabilistic laws (i.e., $CT_p \cap CT_q = \emptyset$), then

$$\Delta_1(p, q) = \sum p_i + \sum q_i = 1 + 1 = 2$$

$$\text{div}_{JS}(p, q) = - \sum \frac{p_i}{2} \log \frac{p_i}{2} - \sum \frac{q_i}{2} \log \frac{q_i}{2} - \frac{1}{2} H(p) - \frac{1}{2} H(q) = 1.$$

Example 4 The Poisson distribution for a randomly occurring rare event $p(i)$, $i = 1, 2, \dots$, with a constant mean λ is defined by

$$p(i) = \frac{\lambda^i}{i!} e^{-\lambda}.$$

The KL-divergence between two Poisson distributions with rates λ and λ' (where $\lambda' > \lambda$) is

$$\lambda' - \lambda - \lambda \log \frac{\lambda'}{\lambda}.$$

The proof uses the Taylor series

$$e^\lambda = \sum_0^\infty \frac{\lambda^i}{i!}.$$

Example 5 The Manhattan difference between two exponential laws (10) with decay rates λ and λ' (where $\lambda' > \lambda$) is

$$\Delta_1(1 - e^{-\lambda t}, 1 - e^{-\lambda' t}) = \int_0^\infty (e^{-\lambda' t} - e^{-\lambda t}) dt = \frac{\lambda' - \lambda}{\lambda \lambda'}.$$

Example 6 Let p and q be deterministic laws of succession such that $p_{2/1} = 1$, $p_{1/1} = 0$ and $q_{2/1} = 0$, $q_{1/1} = 1$. Then

$$\text{div}_{JS}(p, q) = -\left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2}\right) - (0 + 0)/2 = -\log \frac{1}{2} = \log 2 = 1.$$

For indeterministic r with $r_{2/1} = r_{1/1} = 1/2$,

$$\text{div}_{JS}(p, r) = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} - \frac{1}{2} \left(-\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2}\right) = 2 \log 2 - \frac{3}{4} \log 3 = 0.811.$$

These examples illustrate that several alternative distance measures give fairly similar comparative results. For uniform nomic constituents their results are related to the Clifford-measure between ordinary constituents, but this relation is not straightforward for non-uniform constituents and disappears for pure probabilistic laws. When it comes to measure the distance between particular probability values, geometrical and quadratic differences seem simple and useful, but for the distance between whole probability distributions or densities divergence is a convenient

choice. The applicability of the Kullback–Leibler divergence is restricted to pure probabilistic laws, so that the Jensen–Shannon divergence turns out to be valuable complement which can be applied to mixed laws which assign zero probabilities to some Q-predicates or sample points.

5 Vertical versus horizontal distance measures

An important debate about the explication of truthlikeness for monadic languages concerned the question, whether the distance between constituents should reflect distances between Q-predicates. The Clifford-measure $\Delta_C(C_1, C^*)$ counts all errors of C_1 about the Q-predicates equally: mistaken existence claims in $CT_1 - CT^*$ and mistaken non-existence claims in $CT^* - CT_1$ have the same weight $1/K$ in (3). It is natural to consider also situations where the cognitive seriousness of errors in a false constituent are treated differently, so that the distance from the truth is not simply the cardinality of the symmetric difference. Niiniluoto proposed in 1976 two modifications of the Clifford-measure on the basis of the ρ -measure between Q-predicates.³⁰ In the Jyväskylä measure d_j false existence claims are weighted by their distance to the nearest non-empty cell, and false non-existence claims by their distance to the nearest empty cell, while in the weighted symmetric difference d_w the first condition holds, but false non-existence claims are weighted by the minimum distance to a really non-empty cell. Then Δ_C and d_w (unlike d_j) are symmetric, and Δ_C and d_j (unlike d_w) are specular, where a specular distance (in the sense of Festa, 1993) satisfies the condition that the maximally distant constituent from C_i is its photographic negative (i.e., all positive claims are replaced by negative ones and vice versa).³¹ If the ρ -measure reflects resemblances between predicates in a family (e.g. colors), then for the Jyväskylä measure the generalization “All ravens are grey” is closer to the truth than “All ravens are white”.³²

Tichý’s (1976) general definition of truthlikeness implies for the monadic case a distance measure between constituents which differs from the Clifford-measure Δ_C and its modifications d_j and d_w .³³ A *linkage* η between sets CT_i and CT_j is a surjective mapping from the larger of the sets to the smaller one. The cardinality $\text{card}(\eta)$ of η is then $\max\{|CT_i|, |CT_j|\}$, and the breadth of η is the average distance between the linked predicates:

$$B(\eta) = \frac{1}{\text{card}(\eta)} \sum_{\langle Q_u, Q_v \rangle \in \eta} \rho(Q_u, Q_v). \quad (13)$$

The distance $d_T(C_i, C_j)$ between constituents C_i and C_j is then defined as the breadth of the narrowest linkage between CT_i and CT_j .

³⁰ See Niiniluoto (1978). See also the refined treatment of truth approximation by Kuipers (2019).

³¹ See Niiniluoto (1987), p. 319.

³² See Niiniluoto (1987), p. 340.

³³ See also Oddie (1986), pp. 91–99, who applies this method to depth-d constituents.

Niiniluoto (1987) rejects Tichý's proposal for several reasons. The use of average in (13) leads to unintuitive examples, and constituents should not be treated as if they consisted only of existence claims. Indeed, d_T is not specular and does not reflect the cognitive goal of finding true universal generalizations. The most fundamental objection is that $d_T(C_i, C^*)$ can be derived as the minimum distance between two state descriptions s and s' , where s entails the uniformly distributed infinite structure description entailing C_i and s' entails C^* . Thus, Tichý is not defining the distance between C_i and C^* in terms of the counted or weighted differences in claims about the Q -predicates (and thereby the ability of C_i to express true generalizations), but rather in terms of putting an infinite number of individuals in their right places in a classification system.³⁴ The latter problem should be solved by choosing the target as the true state description and by replacing constituents (in a non ad hoc way) as disjunctions of state descriptions.

In spite of this criticism, Tichý's basic idea is interesting, since the notion of a linkage resembles metrics defined for trees in terms of the number of transformations needed to change one tree to another.³⁵ A linkage takes seriously (but perhaps in a wrong way) the demand that "horizontal" distances between Q -predicates are relevant. The goal of distributing an infinite number of individuals to their right places could be viewed as analogous to the task of distributing a probability mass (of measure 1) to its right place. Indeed, a discrete probabilistic constituent (8) allocates the probabilities to a finite number of points in the space Q of Q -predicates, and a continuous probability density f on a state space $Q \subset \mathbb{R}^n$ does the corresponding assignment to an infinite number of points. This can be illustrated by the simple case where Q is a subset of the real line \mathbb{R} and $f: Q \rightarrow \mathbb{R}^+$. If we denote by D_f the region between the curve $f(x) \geq 0$ and the real axis, i.e.,

$$D_f = \{ \langle x, y \rangle \mid x \in Q, 0 \leq y \leq f(x) \},$$

then the density f gives the probability measure 1 to D_f . For two probability densities f and g , the symmetric difference $D_f \Delta D_g$ covers the region between the functions $f(x)$ and $g(x)$ (see Fig. 2). The Manhattan distance is simply the area of this region:

$$\Delta_1(f, g) = \left| D_f \Delta D_g \right|,$$

which is a direct analogue of the Clifford-distance (3) between constituents.

The Manhattan measure, as well as its Euclidean and divergence alternatives, are in an obvious sense *vertical*, since they measure the distance between probability distributions by the absolute, quadratic, or logarithmic differences between the values of $f(x)$ and $g(x)$, without consideration of horizontal distances between points in the sample space Q . This verticality is dramatically seen in Example 3, where the distances between the values of two disjoint discrete probabilistic constituents

³⁴ See Niiniluoto (1987), pp. 328–330.

³⁵ See the Boorman-Olivier and Fu metrics for trees in Niiniluoto (1987), pp. 11–14.

p and q are maximal, and the distances $\Delta_1(p,q)$ and $\text{div}_{JS}(p,q)$ have their maximal values quite independently of the location of p and q with respect to the space \mathbf{Q} . A counterpart of this result for probability densities is the following observation: if f_1 and f_2 are geometrical distributions with the same shape but disjoint domains, then $\Delta_1(f_1,f_2)$, $\Delta_2(f_1,f_2)$, and $\text{div}_{JS}(f_1,f_2)$ have their maximal values quite independently of the geometrical distance a between these densities (see Fig. 3). In fact, all distance measures surveyed by Cha (2007), which are applicable to mixed probabilistic laws, share this feature of verticality.

The observations above motivate the idea that one could try to find measures which in some way take into account the *horizontal* distances between probability distributions (in addition to their vertical ones). Then a modification of Tichý’s linkages might be fruitful. The detailed development of this suggestion has to be left for another occasion, but a simple illustration of the idea can be given here. Consider again real-valued probability densities f which define regions D_f in a subspace S of \mathbb{R}^2 . Let $\beta : S \rightarrow S$ be an area-preserving function, so that $\beta[A]$ has the same area as A for all subregions A of S . Thus, β maps D_f onto D_g by moving the whole probability mass from D_f to D_g .³⁶ The length of the vector $\langle x,y \rangle, \beta(x,y) \rangle$ is defined by the metric of S , and the breadth of β is defined as the sum (integral) of all these lengths for points $\langle x,y \rangle$ in D_f . Then the distance between probabilities f and g is the breadth of the narrowest transformation β between D_f and D_g . For example, in Fig. 3 the mapping $\beta(x,y) = (x - a, y)$, i.e., linear shift to the left,³⁷ gives a linkage between f_2 and f_1 whose breadth is a , since

$$\int_{D_{f_2}} | \langle x,y \rangle, \langle x - a, y \rangle | dx dy = \int_{D_{f_1}} a dx dy = a.$$

For probabilities on the discrete sample space \mathbf{Q} , which in effect define columns on the points of \mathbf{Q} with the total length one, the corresponding idea is to measure the distance between p and q by looking for the shortest length-preserving transformation between p and q . Such a transformation divides the columns of q into pieces and moves them in order to reach a fit with p . If a part of a column q_i is moved to Q_j , then the length of this part is multiplied with the distance $\rho(Q_i,Q_j)$. For example, let $\mathbf{Q} = \{0, 1, 2\}$, $\rho(0, 1) = \rho(1, 2) = 1/2$, $\rho(0, 2) = 1$. Then p and q have the maximal distance 1, if p gives all probability to 0 and q to 2. If

³⁶ A transformation $\beta(x,y) = \langle f(x,y), g(x,y) \rangle$, where f and g are linear functions, is area-preserving, if the absolute value of its Jacobian determinant is one at every point. The Jacobian is composed of the partial derivatives of f and g :
 $|\partial f(x,y)/\partial x \ \partial f(x,y)/\partial y|$
 $|\partial g(x,y)/\partial x \ \partial g(x,y)/\partial y|$

³⁷ Note that the Jacobian determinant of this transformation is.
 $|1 \ 0|$
 $|0 \ 1|$
 so that its value is $1 \times 1 + 0 \times 0 = 1$.

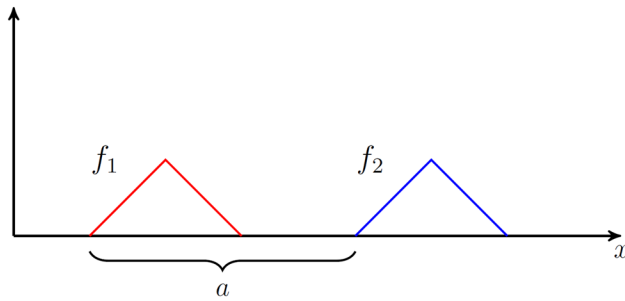


Fig. 3 Disjoint geometrical distributions

$$p_1 = p_2 = p_3 = 1/3$$

$$q_1 = 1/6, q_2 = 2/3, q_3 = 1/6.$$

then the distance between p and q is

$$\frac{1}{6} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{6}.$$

But if

$$q_1 = 1/6, q_2 = 1/6, q_3 = 2/3,$$

then the distance between p and q is $1/4$. These measures, which combine vertical and horizontal aspects, are applicable to both pure and mixed probabilistic laws.

6 Estimating distance from probabilistic truth

According to the similarity approach to the epistemic problem of truthlikeness, unknown degrees of truthlikeness can be estimated by their expected value (4) using a posterior probability distribution over constituents. The same idea can be applied for the estimation of unknown degrees of divergence, which measure distance from the true probabilistic law.

Example 7³⁸ If p is the true probability of success in a binomial model

$$B(p, s) = \binom{n}{s} p^s (1-p)^{n-s}$$

and q is our guessed value, then the divergence of q from p in a single trial is

³⁸ Cf. Niiniluoto (1987), p. 404.

$$\text{div}(q, p) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}.$$

As divergence is additive for independent distributions, this divergence in n trials is

$$n \left[p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q} \right].$$

If the prior distribution $g(p)$ of p is the uniform Beta(1,1), i.e., $g(p)=1$ for $0 \leq p \leq 1$, by Bayes’ theorem the posterior distribution $g(p/s)$ of p with s successes and $n-s$ failures is Beta($s + 1, n - s + 1$), i.e.,

$$g(p/s) = \frac{\Gamma(n + 2)}{\Gamma(s + 1)\Gamma(n - s + 1)} p^s (1 - p)^{n-s},$$

whose mean is $(s + 1)/(n + 2)$.³⁹ Then the estimated distance of q from p can be calculated by

$$\int_0^1 g(p/s) \text{div}(q, p) dp.$$

It follows that, given s successes in n trials, guess q' is estimated to be closer to the truth than q if and only if

$$\begin{aligned} \log \frac{q'}{q} \int_0^1 g(p/s) p dp &> \log \frac{1 - q'}{1 - q} \int_0^1 g(p/s) (1 - p) dp \\ \text{iff } \log \frac{q'}{q} \cdot \frac{s + 1}{n + 2} &> \log \frac{1 - q'}{1 - q} \cdot \left[1 - \frac{s + 1}{n + 2} \right] \\ \text{iff } \log \frac{q'}{q} / \log \frac{1 - q}{1 - q'} &> \frac{n - s + 1}{s + 1}. \end{aligned}$$

Note that for the deterministic hypothesis $q=1$ the value of $\text{div}(q,p)$ is not well defined, but for the Jensen-Shannon distance

$$\text{div}_{JS}(1, p) = \frac{1}{2} p \log \frac{p}{2} - \frac{1}{2} (p + 1) \log \frac{p + 1}{2} + \frac{1}{2} \log 2.$$

Hence,

$$\text{div}_{JS}(1, 0) = \frac{1}{2} \log 2 - \frac{1}{2} \log \frac{1}{2} = \log 2 = 1.$$

³⁹ See Festa (1993, pp. 60–61. Here $\Gamma(n)=(n-1)!$ is the gamma function.

Example 8 ⁴⁰ Let x_1, \dots, x_n be independent measurements of an unknown real-valued quantity θ with a normal distribution $N(\theta, \sigma^2)$:

$$f(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\theta)^2/2\sigma^2}.$$

Then their mean value $y = (x_1 + \dots + x_n)/n$ is normally distributed $N(\theta, \sigma^2/n)$. If the prior probability of θ is sufficiently flat normal, then the posterior distribution $g(\theta/y)$ of θ is approximately $N(y, \sigma^2/n)$, where y is the observed mean. If $f(x/\theta)$ is the true distribution, and $f(x/\theta_0)$ is our guess, then their estimated directed divergence is

$$\begin{aligned} & \int g(\theta/y) \int f(x/\theta) \log[f(x/\theta)/f(x/\theta_0)] dx d\theta \\ &= \int g(\theta/y) [(\theta-\theta_0)/2\sigma^2] d\theta \\ &= \frac{1}{2\sigma^2} \int \theta g(\theta/y) d\theta - \frac{\theta_0}{2\sigma^2} \\ &= (y-\theta_0)/2\sigma^2. \end{aligned}$$

Here the mean y as the best estimate agrees with the result of the Bayes-rule of minimizing expected quadratic loss.

7 Conclusion

We have seen in this paper that the basic idea of the similarity approach to truthlikeness can be extended from qualitative and quantitative first-order languages to cases where probabilistic statements (and their disjunctions) are compared with probabilistic targets. Sections 2 and 3 show how one can naturally proceed from universal and deterministic laws to probabilistic laws. Section 4 argues that the Kullback–Leibler divergence has to be supplemented by the Jensen–Shannon divergence as a measure between mixed probabilistic laws, i.e., laws which assign zero probabilities to some sample points and thereby entail some universal laws. Section 5 formulates a research program for studying a new class of measures which account for the horizontal differences between probability densities, based on distances between sample points. In this way the theory of probabilistic truth approximation does not only lend tools from probability calculus but may suggest novel kinds of problems for mathematicians. Finally, Sect. 6 gives examples to show that the method of estimating degrees of legisimilitude by their expected value can be generalized from the case of deterministic laws to probabilistic laws.

Acknowledgements I am grateful to Gustavo Cevolani and Theo Kuipers for useful comments on probabilistic truth approximation in general and my paper in particular.

⁴⁰ Cf. Niiniluoto (1987), pp. 281–283, 428.

Funding Open access funding provided by University of Helsinki including Helsinki University Central Hospital.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, *1*, 300–307.
- Festa, R. (1993). *Optimum inductive methods*. Kluwer.
- Festa, R. (2007). Verisimilitude, qualitative theories, and statistical inferences. In S. Pihlström, P. Raatikainen, & M. Sintonen (Eds.), *Approaching truth: Essays in honour of Ilkka Niiniluoto* (pp. 143–178). College Publications.
- Fetzer, J. (1981). *Scientific knowledge: Causation, explanation, and corroboration*. D. Reidel.
- García Lapeña, A. (2021). Truthlikeness for quantitative deterministic laws. *The British Journal for the Philosophy of Science*.
- Hilpinen, R. (1976). Approximate truth and truthlikeness. In M. Przelecki, K. Szaniawski, & R. Wojcicki (Eds.), *Formal methods in the methodology of empirical sciences* (pp. 19–42). D. Reidel.
- Kuipers, T. (1982). Approaching descriptive and theoretical truth. *Erkenntnis*, *18*, 343–387.
- Kuipers, T. (Ed.). (1987). *What is closer-to-the-truth?* Rodopi.
- Kuipers, T. (2019). *Nomic truth approximation revisited*. Springer.
- Kullback, S. (1959). *Information theory and statistics*. Wiley.
- McCutcheon, R. (2019). In favor of logarithmic scoring. *Philosophy of Science*, *86*, 286–303.
- Niiniluoto, I. (1978). Truthlikeness in first-order languages. In J. Hintikka, I. Niiniluoto, & E. Saarinen (Eds.), *Essays on mathematical and philosophical logic* (pp. 437–458). D. Reidel.
- Niiniluoto, I. (1987). *Truthlikeness*. D. Reidel.
- Niiniluoto, I. (1998). Verisimilitude: The third period. *The British Journal for the Philosophy of Science*, *49*, 1–29.
- Niiniluoto, I. (2011). The development of the Hintikka program. In D. Gabbay, S. Hartmann, & J. Woods (Eds.), *Handbook of the history of logic* (Vol. 10, pp. 311–356). North-Holland.
- Niiniluoto, I. (2019). *Truth-seeking by abduction*. Springer.
- Oddie, G. (1986). *Likeness to truth*. D. Reidel.
- Oddie, G. (2016). Truthlikeness. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*, <https://plato.stanford.edu/archives/win2016/entries/truthlikeness/>. Read Oct. 20, 2020.
- Oddie, G. (2019). What accuracy could not be. *The British Journal for the Philosophy of Science*, *70*, 551–580.
- Parzen, E. (1962). *Stochastic processes*. Holden-Day.
- Pettigrew, R. (2015). *Accuracy and laws of credence*. Oxford University Press.
- Popper, K. (1963). *Conjectures and refutations*. Routledge and Kegan Paul.
- Rescher, N. (1970). *Scientific explanation*. The Free Press.
- Rosenkrantz, R. (1980). Measuring truthlikeness. *Synthese*, *45*, 463–488.
- Sober, E. (2002). Instrumentalism, parsimony, and the Akaike framework. *Philosophy of Science*, *69*(S3), S112–S123.
- Tichý, P. (1976). Verisimilitude redefined. *The British Journal for the Philosophy of Science*, *27*, 25–42.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.