e-value: Epistemic Value of Statistical Hypotheses

Julio Michael Stern [1]

in: Miodrag Lovric (2025, pp.863-867, en.188). *International Encyclopedia of Statistical Science*, 2nd.ed. Springer. doi: 10.1007/978-3-662-69359-9_188

The e-value, a.k.a. the *epistemic value* of hypothesis H given the observational data X or, the other way around, the *evidence value* rendered by the observational data X in support of hypothesis H, is a *significance measure* or *truth value* conceived for use in statistical modeling, see Pereira and Stern (1999).

The e-value and its associated statistical procedures, like the Full Bayesian Significance Test (FBST), comply with the best principles of Bayesian inference, including the likelihood principle, complete invariance, asymptotic consistency, etc. The e-value and the FBST also exhibit powerful logic or algebraic properties in situations involving the comparison or composition of distinct hypotheses that can be formulated either in the same or in different statistical models. Moreover, they effortlessly accommodate the case of sharp or precise hypotheses, a situation where alternative methods often require ad hoc and convoluted procedures. Furthermore, the e-value and the FBST exhibit excellent operational characteristics, like straightforward formulation, simple numerical implementation, robust and reliable behavior, etc. Finally, test procedures based on the e-value often outperform (in standard benchmark experiments) alternatives found in the literature for important applications in statistical modeling and operations research.

The e-value is defined within the standard Bayesian framework for parametric statistics, where observations, X, prior, $p_0(\theta)$, and posterior, $p_n(\theta)$, densities for the parameters are related by Bayesian learning steps, that is,

$$p_n(\theta \,|\, X) \;=\; (1/c_n) p_0(\theta) \prod_{i=1}^{n} p(x^{(i)} \,|\, \theta) \;=\; (1/c_n) p_0(\theta) p(X \,|\, \theta) \; . \tag{1}$$

A statistical hypothesis H states that the parameter θ^0 generating the observations X belongs to a region of the parameter space constrained by (vector) inequality and equality constraints, that is,

$$H = \{\theta \in \Theta \,|\, g(\theta) \le \mathbf{0} \wedge h(\theta) = \mathbf{0}\} \; . \tag{2}$$

The dimension of this hypothesis, h, is the dimension of its parameter space, t, minus the number of its equality constraints, q, that is,

$$h = \dim(H) = t - q \le \; t = \dim(\Theta) \; . \tag{3}$$

[1] Julio Michael Stern received his B.S. (1981) and M.Sc. (1983) in Physics from USP – the University of Sao Paulo; M.Eng. (1990) and Ph.D. (1991) in Operations Research from Cornell University; and Liv.Doc. (2002) in Computer Science from USP. He serves as Full Professor (2010) of Applied Mathematics at USP; Research Fellow (2006) at CNPq; President (2010-2011) of the Brazilian chapter of ISBA – the International Society for Bayesian Analysis; President (2017-2020) of IME-USP Research Council (CPq); Member (2017) of the Advisory Committee of ABJ – the Brasilian Jurimetrics Association; and Member (2019) of ABF – the Brazilian Academy of Philosophy. ORCID: 0000-0003-2720-3871; Google Scholar: 7spXyx8AAAAJ.

An hypothesis is called *sharp* or *precise* if the last inequality is strict, that is, if $h < t$; otherwise, if $h = t$, it is called a *slack* hypothesis. All major scientific theories in exact sciences are structured around precise natural laws formulated as mathematical equations that, in turn, are easily expressed in statistical modeling as sharp hypotheses. Nevertheless, many traditional significance measures or test procedures, specially in the Bayesian framework, face theoretical or methodological difficulties in the treatment of sharp hypotheses. The *e*-value and the FBST were developed to overcome these difficulties, giving a coherent and uniform treatment to either slack or sharp hypotheses. The successful accomplishment of these goals engender important philosophical, theoretical, methodological and practical consequences, see Stern (2017, 2020) and Stern et al. (2018, 2022).

The the *e*-value of H given X, $\text{ev}\,(H \,|\, X) \in [0, 1]$, and its complement, $\overline{\text{ev}}(H \,|\, X) = 1 - \text{ev}\,(H \,|\, X)$, are defined as follows; see Borges and Stern (2007) and Stern and Pereira (2014) for further details and explanations.

(i) $s(\theta)$, the *surprise function* in a statistical model is defined as the quotient between the posterior and the reference densities in the model,

$$s(\theta) = p_n(\theta)/r(\theta) \ . \tag{4}$$

The *reference density*, $r(\theta)$, can be interpreted as a representation of vague or weak information about θ, like the uniform distribution, $r(\theta) \propto 1$, an invariant prior, or a maximum entropy density, see Stern (2011). Alternatively, the reference density can be interpreted as a representation of the parameter space's underlying information metric, $r(\theta) = \sqrt{\det G(\theta)}$, given the metric $dl^2 = d\theta^t G(\theta) d\theta$, $\theta \in \Theta$.

(ii) s^*, the maximum (or supremum) of the surprise function constrained to the hypothesis H, is defined as

$$s^* = \sup_{\theta \in H} s(\theta) \ . \tag{5}$$

A maximizing argument, $\theta^* \,|\, s^* = s(\theta^*)$, is called a *tangential point*;

(iii) $T(v)$, the closed lower v-cut of the surprise function, and its complement, the open upper v-cut of the surprise function, $\overline{T}(v)$, are defined as

$$T(v) = \{\theta \in \Theta \,|\, s(\theta) \leq v\} \ , \quad \overline{T}(v) = \{\theta \in \Theta \,|\, s(\theta) > v\} \ . \tag{6}$$

The upper v-cut at level $v = s^*$, $\overline{T}(s^*)$, is called the *tangential set*, for its border corresponds to the contour line of the surprise function that is tangential to hypothesis H.

(iv) $W(v)$, the *truth function* or *Wahrheitsfunktion* at level v, is defined as the posterior probability mass inside the lower v-cut of the surprise function.

$$W(v) = \int_{T(v)} p_n\,(\theta)\, d\theta \ , \tag{7}$$

while its complement is defined as $\overline{W}(v) = 1 - W(v)$.

(v) $\text{ev}\,(H \mid X)$, the *epistemic value* of hypothesis H given the observed data X, is defined as the truth function $W(v)$ computed at level $v = s^*$, while its complement, $\overline{\text{ev}}(H \mid X)$, the evidence given by the observed data X against hypothesis H, has the complementary probability mass,

$$\text{ev}\,(H \mid X) = W(s^*)\ , \quad \overline{\text{ev}}(H \mid X) = \overline{W}(s^*) = 1 - \text{ev}\,(H)\ . \tag{8}$$

(vi) $\text{sev}(H \mid X)$, the *standardized e-value* of a hypothesis $H \subset \Theta$ of dimension $h = \dim(H) \leq t = \dim(\Theta)$, and its complement, $\overline{\text{sev}}(H \mid X)$, are defined as follows:

$$\text{sev}(H \mid X) = 1 - \overline{\text{sev}}(H \mid X)\ , \quad \overline{\text{sev}}(H \mid X) = \sigma(t, h, \overline{\text{ev}}(H \mid X))\ ; \tag{9}$$

where $\sigma(t, h, c)$, the *standardization function* on arguments $t, h \in \mathcal{N}_+$ and $c \in [0, 1]$, is defined in terms of the *chi-square* cumulative distribution with $d \in \mathcal{N}_+$ degrees of freedom, $Q(d, z)$, by the expression

$$\sigma(t, h, c) = Q\left(t - h, Q^{-1}(t, c)\right)\ . \tag{10}$$

Under appropriate regularity conditions, as the number of observations increases, that is, as $n \to \infty$, $\text{sev}(H \mid X)$ exhibits the following asymptotic behavior: If H is false, $\text{sev}(H \mid X) \to 0$; If H is true, $\text{sev}(H \mid X) \to U[0, 1]$. Concerning this behavior, $\overline{\text{sev}}(H)$ resembles the classical p-value and, accordingly, can replace (and outperform) it in many commonly used test procedures.

A logical formalism can be conceived as an algebra for obtaining truth-values of complex statements from its constituent or elementary parts. In this perspective, the e-value has strong logical properties captured by the algebraic formalism explained in the sequel. Let us consider alternative elementary hypotheses, $H^{(i,j)}$, $i = 1 \ldots q$, defined in $j = 1 \ldots k$ independent constituent models, $M^{(j)}$, and also a complex hypothesis, H, defined by *logical composition* in *homogeneous disjunctive normal form* (disjunction of conjunctions) of the aforementioned elementary hypotheses in the product model $M = M^{(1)} \times \ldots \times M^{(k)}$, that is:

$$H \;=\; \bigvee\nolimits_{i=1}^{q} \bigwedge\nolimits_{j=1}^{k} H^{(i,j)}\ , \quad M^{(i,j)} = \{\Theta^{(j)}, H^{(i,j)}, p_0^{(j)}, p_n^{(j)}, r^{(j)}\}\ , \tag{11}$$

$$M = \{\Theta, H, p_0, p_n, r\}\ , \quad \Theta = \prod\nolimits_{j=1}^{k} \Theta^{(j)}\ , \quad p_n = \prod\nolimits_{j=1}^{k} p_n^{(j)}\ , \quad r = \prod\nolimits_{j=1}^{k} r^{(j)}. \tag{12}$$

Then, $\text{ev}\,(H)$, the e-value supporting the complex hypothesis, is computed as:

$$\text{ev}\left(\bigvee\nolimits_{i=1}^{q} \bigwedge\nolimits_{j=1}^{k} H^{(i,j)}\right) \;=\; W\left(\max_{i=1}^{q} \prod\nolimits_{j=1}^{k} s^{*(i,j)}\right) \;=\; W\left(\max_{i=1}^{q} s^{*(i)}\right) \tag{13}$$

$$=\; \max_{i=1}^{q} W\left(s^{*(i)}\right) \;=\; \max_{i=1}^{q} \text{ev}\left(\bigwedge\nolimits_{j=1}^{k} H^{(i,j)}\right) \;=\; \max_{i=1}^{q} \text{ev}\left(H^{(i)}\right)\ ; \tag{14}$$

where the cumulative surprise distribution of the product model, $W(v)$, is given by the Mellin convolution operation, see Borges and Stern (2007), defined as

$$W = \bigotimes_{1 \leq j \leq k} W^{(j)}\ , \quad W^{(1)} \otimes W^{(2)}(v) \;=\; \int_0^\infty W^{(1)}(v/y) W^{(2)}(dy)\ . \tag{15}$$

The probability distribution of the product of two independent positive random variables is given by the Mellin convolution of their distributions. From this interpretation, we immediately see that $\otimes$ is a commutative and associative operator. Moreover, we observe that, in the extreme case of null-or-full support, that is, when, for $1 \leq i \leq q$ and $1 \leq j \leq k$, $s^{*(i,j)} = 0$ or $s^{*(i,j)} = \widehat{s}^{(j)}$, the e-values of the constituent elementary hypotheses are either 0 or 1, and the conjunction and disjunction composition rules of classical logic hold.

The *Generalized Full Bayesian Significance Test*, GFBST, *rejects* H if its e-value stays below an established threshold, c, that is, if $\text{ev}\,(H) < c$, and *accepts* H if it rejects its complement, that is, if $\text{ev}\,(\overline{H}) < c$, where $\overline{H} = \Theta - H$. In this context, the standard *modal logic* operators of *necessity*, $\Box$; *possibility*, $\Diamond$; *contingency*, ∇ ; and negation, $\neg$; are used to conveniently represent accepting H, $\Box H$, rejecting it (impossibility), $\neg\Diamond H$, or remaining agnostic (undecided), $\nabla H = \Diamond H \wedge \neg\Box H$.

Since the GFBST is directly engendered by the e-value, it inherits all its good statistical and compositional properties. Moreover, the GFBST obeys the following rules for consistent reasoning concerning the logical modalities of necessity, possibility and contingency or, alternatively, rules for consistent decision in accepting, rejecting, or remaining agnostic concerning interrelated statistical hypotheses:

(I) *Invertibility:* Applied to an hypothesis H and its complement, $\overline{H} = \Theta - H$.
(I.i) *Necessity inversion:* $\Box H \Leftrightarrow \neg\Diamond\overline{H}$;
(I.ii) *Possibility inversion:* $\Diamond H \Leftrightarrow \neg\Box\overline{H}$;
(I.iii) *Contingency inversion:* $\nabla H \Leftrightarrow \nabla\overline{H}$.

(M) *Monotonicity:* Applied to an hypothesis H and a superset, $H' \supset H$.
(M.i) *Monotonic necessity:* $\Box H \Rightarrow \Box H'$;
(M.ii) *Monotonic possibility:* $\Diamond H \Rightarrow \Diamond H'$.

(C) *Consonance:* Applied to an indexed set of hypotheses, $H^{(i)}$, for $i \in I$.
(C.i) *Union consonance*: $\Diamond(\cup_{i\in I} H^{(i)}) \Rightarrow \exists i \in I \,|\, \Diamond H^{(i)}$;
(C.ii) *Intersection consonance*: $\forall i \in I,\ \Box H^{(i)} \Rightarrow \Box(\cap_{i\in I} H^{(i)})$.

The aforementioned rules for consistent reasoning correspond to basic principles of rational argumentation that are natural and intuitive for human interpretation. Using inference or decision procedures that violate these rules of good reasoning brings the danger of miscommunication, misunderstanding, or misinformation. Moreover, using arguments of mathematical analysis, it is possible to use the same rules of consistent reasoning to give an (essentially) unique characterization of the GFBST, see Esteves et al. (2016) and Stern et al. (2018).

For detailed discussions of the GFBST, its statistical features, logical properties, and philosophical consequences, see Esteves et al. (2016), Madruga et al. (2001), Pereira et al. (2008), and Stern et al. (2018). The review articles by Pereira and Stern (2020) and Stern et al. (2022) include references for hundreds of applications of the e-value and the FBST in science and technology, several further theoretical, methodological, and computational developments, discussion of some consequences in epistemology and philosophy of science of this approach for evaluating and testing statistical hypotheses, and some directions for further research.

References

[1] Borges, Wagner de Souza; Stern, Julio Michael (2007). The Rules of Logic Composition for the Bayesian Epistemic E-Values. *Logic Journal of the IGPL*, 15, 5/6, 401-420. `doi:10.1093/jigpal/jzm032`

[2] Esteves, Luis Gustavo; Izbicki, Rafael; Stern, Julio Michael; Stern, Rafael Bassi (2016). The Logical Consistency of Simultaneous Agnostic Hypothesis Tests. *Entropy*, 18, 256. `doi:10.3390/e18070256`

[3] Madruga, Maria Regina; Esteves, Luis Gustavo; Wechsler, Sergio (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291-299. `doi:10.1007/BF02595698`

[4] Pereira, Carlos Alberto de Bragança; Stern, Julio Michael (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy*, 1, 99-110. `doi:1099-4300/1/4/99`

[5] Pereira, Carlos Alberto de Bragança; Stern, Julio Michael; Wechsler, Sergio (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis*, 3, 79-100. `doi:10.1214/08-BA303`

[6] Pereira, Carlos Alberto de Bragança; Stern, Julio Michael (2020). The e-value: A Fully Bayesian Significance Measure for Precise Statistical Hypotheses and its Research Program. *São Paulo Journal of Mathematical Sciences*, 1-19. `doi:10.1007/s40863-020-00171-7`

[7] Stern, Julio Michael (2011). Symmetry, Invariance and Ontology in Physics and Statistics. *Symmetry*, 3, 611–635. `doi:10.3390/sym3030611`

[8] Stern, Julio Michael; Pereira, Carlos Alberto de Bragança (2014). Bayesian Epistemic Values: Focus on Surprise, Measure Probability! *Logic Journal of the IGPL*, 22, 236-254. `doi:10.1093/jigpal/jzt023`

[9] Stern, Julio Michael (2017). Continuous Versions of Haack's Puzzles: Equilibria, Eigen-States and Ontologies. *Logic Journal of the IGPL*, 25, 4, 604-631. `doi:10.1093/jigpal/jzx017`

[10] Stern, Julio Michael (2018). Karl Pearson on Causes and Inverse Probabilities: Renouncing the Bride, Inverted Spinozism and Goodness-of-Fit. *South American Journal of Logic*, 4, 1, 219-252. `arXiv:1908.06346`

[11] Stern, Julio Michael; Izbicki, Rafael; Esteves, Luis Gustavo; Stern, Rafael Bassi (2018). Logically-Consistent Hypothesis Testing and the Hexagon of Oppositions. *Logic Journal of the IGPL*, 25, 741-757. `doi:10.1093/jigpal/jzx024`

[12] Stern, Julio Michael (2020). A Sharper Image: The Quest of Science and Recursive Production of Objective Realities. *Principia*, 24, 2, 255-297. `doi:10.5007/1808-1711.2020v24n2p255`

[13] Stern, Julio Michael; Pereira, Carlos Alberto de Bragança; Lauretto, Marcelo de Souza; Esteves, Luis Gustavo; Stern, Izbicki, Rafael; Rafael Bassi; Diniz, Marcio Alves (2022). The *e*-value and the Full Bayesian Significance Test: Logical Properties and Philosophical Consequences. `arXiv:2205.08010`

In the *International Encyclopedia of Statistical Science*, see the entries (pages, entry numb.):
• E-values (E for Expectation) in Hypothesis Testing; by Ruodu Wang, pp.867-872, en.189. • Full Bayesian Significant Test (FBST); by Carlos Alberto de Bragança Pereira, pp.996-1000, en.240. • Mixed-Paradigm Hypothesis Tests; by Mark Andrew Gannon, Luís Gustavo Esteves, Carlos Alberto de Bragança Pereira, pp.479-1484, en.365. • Bayes Factor; by Miodrag Lovric, pp.145-158, en.48. • P-Value Chronicles: The Unmasking Fourteen Widespread Misconceptions; by Rose R. Chamberlain, Matthew W. Ross, Calvin C. Long, and Miodrag Lovric, pp.2006-2024, en.705. • P-Values; by Raymond Hubbard, pp.2024-2025, en.706. • P-Values, Combining of; by Dinis Pestana, Maria de Fátima Brilhante, and Sandra Mendonça, pp.2026-2032, en.707.