
The Rules of Logic Composition for the Bayesian Epistemic e-Values

WAGNER BORGES, *CCH, Mackenzie Presbyterian University.*
E-mail: wborges@mackenzie.com.br

JULIO MICHAEL STERN, *IME, University of São Paulo, Brazil.*
E-mail: jstern@ime.usp.br

Abstract

In this paper, the relationship between the e -value of a complex hypothesis, H , and those of its constituent elementary hypotheses, H^j , $j=1\dots k$, is analyzed, in the independent setup. The e -value of a hypothesis H , $ev(H)$, is a Bayesian epistemic, credibility or truth value defined under the Full Bayesian Significance Testing (FBST) mathematical apparatus. The questions addressed concern the important issue of how the truth value of H , and the truth function of the corresponding FBST structure M , relate to the truth values of its elementary constituents, H^j , and to the truth functions of their corresponding FBST structures M^j , respectively.

Keywords: Bayesian models; Belief Calculi; Complex hypotheses; Compositionality; Convolution; Epistemic values; Possibilistic and Probabilistic reasoning; Significance tests; Truth functions.

1 Introduction

The e -value of an hypothesis H , $ev(H)$, is a Bayesian epistemic, credibility or truth value first defined in Pereira and Stern [39]. In this paper we analyze the relationship between the e -value, of a complex hypothesis, H , and those of its elementary constituents, H^j , $j=1\dots k$. This problem is known as the question of *Compositionality*, which plays a central role in analytical philosophy.

According to Wittgenstein [60], (2.0201, 5.0, 5.32):

- Every complex statement can be analyzed from its elementary constituents.
- Truth values of elementary statements are the results of those statements' truth-functions (Wahrheitsfunktionen).
- All truth-function are results of successive applications, to elementary constituents, of a finite number of truth-operations (Wahrheitsoperationen).

The compositionality question also plays a central role in far more concrete contexts, like that of reliability engineering, see Barlow and Prochan [4] and Birnbaum et al. [6], (1.4):

“One of the main purposes of a mathematical theory of reliability is to develop means by which one can evaluate the reliability of a structure when the reliability of its components are known. The present study will be concerned with this kind of mathematical development. It will be necessary for this purpose to rephrase our intuitive concepts of structure, component, reliability, etc. in more formal language, to restate carefully our assumptions, and to introduce an appropriate mathematical apparatus.”

When brought into a parametric statistical hypothesis testing context, a complex hypothetical scenario or complex hypothesis is a statement, H , concerning $\theta = (\theta^1, \dots, \theta^k) \in \Theta =$

$(\Theta^1 \times \dots \times \Theta^k)$ which is equivalent to a logical composition of statements, H^1, \dots, H^k , concerning the elementary components, $\theta^1 \in \Theta^1, \dots, \theta^k \in \Theta^k$, respectively.

Within this setting, means to evaluate the credibility of H , as well as that of each of its elementary components, H^1, \dots, H^k , is provided by the mathematical apparatus introduced in Pereira and Stern [40]. Further general references on the subject include Lauretto et al. [32] and Madruga et al. [36]. It is of interest, however, to know what can be said about the credibility of H , from the knowledge of the credibilities of its elementary components, H^1, \dots, H^k , and this is what the authors endeavor to explore in the present paper, in the case of independent corresponding FBST structures M^1, M^2, \dots, M^k .

In order to give the reader a clear and self contained view of the investigated questions and the answers provided, the paper has been organized as follows. In section 2, a panoramic review of several aspects of the FBST theory is presented. In section 3, the mathematical apparatus of the FBST is described. In section 4, the question of whether the truth value of a conjunctive hypothesis H , can be obtained from the truth values of its constituent elementary hypotheses, H^1, \dots, H^k , is explored. In sections 5 and 6 the question of whether the truth function of the FBST structure, M , corresponding to a conjunctive hypothesis H can be obtained from the truth functions of the structures, M^1, M^2, \dots, M^k , corresponding to H^1, \dots, H^k , respectively, is analyzed. Final remarks are presented in section 7.

2 FBST Review

The FBST was specially designed to give an epistemic value, or value of evidence, supporting a sharp hypothesis H . This support function is the e -value, $ev(H)$. Furthermore, the e -value has many necessary or desirable properties for a statistical support function, such as:

- (I) Provide an intuitive and simple measure of significance for the hypothesis in test, ideally, a *probability* defined directly in the original or *natural parameter space*.
- (II) Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the (manifold representing the) null hypothesis being tested, or the particular coordinate system chosen for the parameter space, i.e., be an *invariant* procedure.
- (III) Provide a measure of significance that is smooth, i.e. *continuous and differentiable*, on the hypothesis parameters and sample statistics, under appropriate regularity conditions of the model.
- (IV) Obey the *likelihood principle*, i.e., the information gathered from observations should be represented by, and only by, the likelihood function.
- (V) Require *no ad hoc artifice* like assigning a positive prior probability to zero measure sets, or setting an arbitrary initial belief ratio between hypotheses.
- (VI) Be a *possibilistic* support function, where the support of a logical disjunction is the maximum support among the support of the disjuncts.
- (VII) Be able to provide a *consistent* test for a given sharp hypothesis.
- (VIII) Be able to provide *compositionality* operations in complex models.
- (IX) Be an *exact* procedure, i.e., make no use of “large sample” asymptotic approximations when computing the e -value.
- (X) Allow the incorporation of previous experience or expert's opinion via (subjective) *prior distributions*.

The objective of this section is to provide a very short review of the FBST theoretical framework, summarizing the most important statistical properties of its support function, the e -value. It also summarizes the logical (algebraic) properties of the e -value, and its relations to other classical support calculi, including possibilistic calculus and paraconsistent or classical logic. Further details, demonstrations of theoretical properties, comparison with other statistical tests for sharp hypotheses, and an extensive list of references can be found in the author's previous papers.

2.1 The FBST Epistemic e -values

Let $\theta \in \Theta \subseteq R^p$ be a vector parameter of interest, and $L(\theta|x)$ be the likelihood associated to the observed data x , a standard statistical model. Under the Bayesian paradigm the posterior density, $p_n(\theta)$, is proportional to the product of the likelihood and a prior density,

$$p_n(\theta) \propto L(\theta|X)p_0(\theta).$$

The (null) hypothesis, H , states that the parameter lies in the null set, defined by inequality and equality constraints given by vector functions g and h in the parameter space.

$$\Theta_H = \{\theta \in \Theta | g(\theta) \leq \mathbf{0} \wedge h(\theta) = \mathbf{0}\}$$

From now on, we use a relaxed notation, writing H instead of Θ_H . We are particularly interested in sharp (precise) hypotheses, i.e., those in which $\dim(H) < \dim(\Theta)$, i.e. there is at least one equality constraint.

The FBST defines $ev(H)$, the e -value, the epistemic value or value of (presented or observed) evidence supporting (in favor of) the hypothesis H , and $\overline{ev}(H)$, the e -value against H , as

$$s(\theta) = \frac{p_n(\theta)}{r(\theta)}, \quad s^* = s(\theta^*) = \sup_{\theta \in H} s(\theta), \quad \widehat{s} = s(\widehat{\theta}) = \sup_{\theta \in \Theta} s(\theta),$$

$$T(v) = \{\theta \in \Theta | s(\theta) \leq v\}, \quad W(v) = \int_{T(v)} p_n(\theta) d\theta, \quad ev(H) = W(s^*),$$

$$\overline{T}(v) = \Theta - T(v), \quad \overline{W}(v) = 1 - W(v), \quad \overline{ev}(H) = \overline{W}(s^*) = 1 - ev(H).$$

The function $s(\theta)$ is known as the posterior surprise relative to a given reference density, $r(\theta)$. $W(v)$ is the cumulative surprise distribution. The surprise function was used, among other statisticians, by Good [23], Evans [16] and Royall [48]. Its role in the FBST is to make $ev(H)$ explicitly invariant under suitable transformations on the coordinate system of the parameter space, as shown in the next section.

The tangential (to the hypothesis) set $\overline{T} = \overline{T}(s^*)$, is a Highest Relative Surprise Set (HRSS). It contains the points of the parameter space with higher surprise, relative to the reference density, than any point in the null set H . When $r(\theta) \propto 1$, the possibly improper uniform density, \overline{T} is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set H . Small values of $\overline{ev}(H)$ indicate that the hypothesis traverses high density regions, thus favoring it.

The e -value, defined above, has a simple and intuitive geometric characterization. Figure 2.1 shows the null set H , the tangential HRSS \overline{T} , and the points of constrained

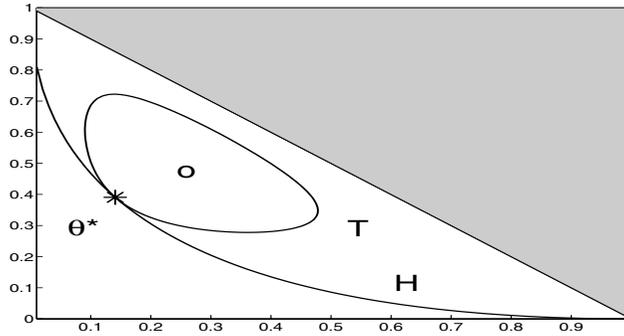


Figure 2.1. H-W Hypothesis and Tangential Set

and unconstrained maxima, θ^* and $\hat{\theta}$, for testing Hardy-Weinberg equilibrium law in a population genetics problem, as discussed in Pereira and Stern [40]. In this biological application: n is the sample size; x_1 and x_3 are the two homozygote sample counts and $x_2 = n - x_1 - x_3$ is heterozygote sample count. $\theta = [\theta_1, \theta_2, \theta_3]$ is the parameter vector. The posterior and maximum entropy reference densities for this trinomial model, the parameter space and the null set are:

$$p_n(\theta | x) \propto \theta_1^{x_1+y_1-1} \theta_2^{x_2+y_2-1} \theta_3^{x_3+y_3-1}, \quad r(\theta) \propto \theta_1^{y_1-1} \theta_2^{y_2-1} \theta_3^{y_3-1}, \quad y = [0, 0, 0],$$

$$\Theta = \{\theta \geq 0 | \theta_1 + \theta_2 + \theta_3 = 1\}, \quad H = \{\theta \in \Theta | \theta_3 = (1 - \sqrt{\theta_1})^2\}.$$

Finally, consider the situation where the hypothesis constraint, $H: h(\theta) = h(\delta) = 0, \theta = [\delta, \lambda]$ is not a function of some of the parameters, λ . This situation is described by Basu [5]:

“If the inference problem at hand relates only to δ , and if information gained on λ is of no direct relevance to the problem, then we classify λ as the Nuisance Parameter. The big question in statistics is: How can we eliminate the nuisance parameter from the argument?”

Basu goes on listing at least 10 categories of procedures to achieve this goal, like using max_λ or $\int d\lambda$, the maximization or integration operators, in order to obtain a projected profile or marginal posterior function, $f(\delta|x)$. The FBST does not follow the nuisance parameters elimination paradigm. In fact, staying in the original parameter space, in its full dimension, explains the “Intrinsic Regularization” property of the FBST, when it is used for model selection, see Pereira and Stern [42].

2.2 Reference, Invariance and Consistency

In the FBST the role of the reference density, $r(\theta)$ is to make $\bar{e}v(H)$ explicitly invariant under suitable transformations of the coordinate system. Invariance, as used in statistics, is a metric concept. The reference density can be interpreted as a compact and interpretable representation of the reference metric in the original parameter space. This metric is given by the geodesic distance on the density surface. The natural choice of the reference density is an uninformative prior, interpreted either as a representation of no information in the parameter space, or as the limit prior for no observations, or as the neutral ground state for the Bayesian operation.

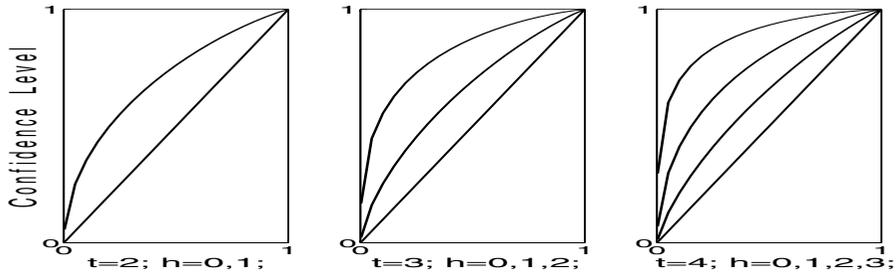


Figure 2.2. Test τ_c critical level vs. confidence level

Standard (possibly improper) uninformative priors include the uniform and maximum entropy densities, see Dugdale [15] and Kapur [29] for a detailed discussion. In the H-W example, using the notation above, the uniform density can be represented by $y=[1,1,1]$ observation counts, and the standard maximum entropy density can be represented by $y=[0,0,0]$ observation counts.

Let us consider the cumulative distribution of the e -value against the hypothesis, $\bar{V}(c) = \Pr(\bar{e}v \leq c)$, given θ^0 , the true value of the parameter. Under appropriate regularity conditions, for increasing sample size, $n \rightarrow \infty$, we can say the following:

- If H is false, $\theta^0 \notin H$, then $\bar{e}v$ converges (in probability) to 1, that is, $\bar{V}(0 \leq c < 1) \rightarrow 0$.
- If H is true, $\theta^0 \in H$, then $\bar{V}(c)$, the confidence level, is approximated by the function

$$QQ(t, h, c) = Q(t - h, Q^{-1}(t, c)) \quad , \quad \text{where}$$

$$Q(k, x) = \frac{\Gamma(k/2, x/2)}{\Gamma(k/2, \infty)} \quad , \quad \Gamma(k, x) = \int_0^x y^{k-1} e^{-y} dy \quad ,$$

$t = \dim(\Theta)$, $h = \dim(H)$ and $Q(k, x)$ is the cumulative chi-square distribution with k degrees of freedom. Figure 2.2 portrays $QQ(t, h, c) = Q(t - h, Q^{-1}(t, c))$ for $t=2 \dots 4$ and $h=0 \dots t-1$.

Under the same regularity conditions, an appropriate choice of threshold or critical level, $c(n)$, provides a consistent test, τ_c , that rejects the hypothesis if $\bar{e}v(H) > c$. The empirical power analysis developed in Stern and Zacks [57] and Lauretto [32], provides critical levels that are consistent and also effective for small samples.

Proof of invariance:

Consider a proper (bijective, integrable, and almost surely continuously differentiable) reparameterization $\omega = \phi(\theta)$. Under the reparameterization, the Jacobian, surprise, posterior and reference functions are:

$$J(\omega) = \left[\frac{\partial \theta}{\partial \omega} \right] = \left[\frac{\partial \phi^{-1}(\omega)}{\partial \omega} \right] = \begin{bmatrix} \frac{\partial \theta_1}{\partial \omega_1} & \cdots & \frac{\partial \theta_1}{\partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_n}{\partial \omega_1} & \cdots & \frac{\partial \theta_n}{\partial \omega_n} \end{bmatrix}$$

$$\tilde{s}(\omega) = \frac{\tilde{p}_n(\omega)}{\tilde{r}(\omega)} = \frac{p_n(\phi^{-1}(\omega)) |J(\omega)|}{r(\phi^{-1}(\omega)) |J(\omega)|}$$

Letting $\Omega_H = \phi(\Theta_H)$, it follows that

$$\tilde{s}^* = \sup_{\omega \in \Omega_H} \tilde{s}(\omega) = \sup_{\theta \in \Theta_H} s(\theta) = s^*.$$

Hence, the tangential set, $\bar{T} \mapsto \phi(\bar{T}) = \tilde{T}$, and

$$\tilde{\text{Ev}}(H) = \int_{\tilde{T}} \tilde{p}_n(\omega) d\omega = \int_{\bar{T}} p_n(\theta) d\theta = \bar{\text{ev}}(H).$$

Proof of consistency:

Let $\bar{V}(c) = \text{Pr}(\bar{\text{ev}} \leq c)$ be the cumulative distribution of the e -value against the hypothesis, given θ . We stated that, under appropriate regularity conditions, for increasing sample size, $n \rightarrow \infty$, if H is true, i.e. $\theta \in H$, then $\bar{V}(c)$ is approximated by the function

$$QQ(t-h, Q^{-1}(t, c)) .$$

Let $\theta^0, \hat{\theta}$ and θ^* be the true value, the unconstrained MAP (Maximum A Posteriori), and constrained (to H) MAP estimators of the parameter θ .

Since the FBST is invariant, we can chose a coordinate system where, the (likelihood function) Fisher information matrix at the true parameter value is the identity, i.e., $J(\theta^0) = I$. From the posterior Normal approximation theorem, see Gelman et al. [20], (appendix B), we know that the standarized total difference between $\hat{\theta}$ and θ^0 converges in distribution to a standard Normal distribution, i.e.

$$\sqrt{n}(\hat{\theta} - \theta^0) \rightarrow N(0, J(\theta^0)^{-1} J(\theta^0) J(\theta^0)^{-1}) = N(0, J(\theta^0)^{-1}) = N(0, I)$$

This standarized total difference can be decomposed into tangent (to the hypothesis manifold) and transversal orthogonal components, i.e.

$$d_t = d_n + d_{t-h} , \quad dt = \sqrt{n}(\hat{\theta} - \theta^0) , \quad d_h = \sqrt{n}(\theta^* - \theta^0) , \quad d_{t-h} = \sqrt{n}(\hat{\theta} - \theta^*) .$$

Hence, the total, tangent and transversal distances (L^2 norms), $\|d_t\|$, $\|d_h\|$ and $\|d_{t-h}\|$, converge in distribution to chi-square variates with, respectively, t , h and $t-h$ degrees of freedom.

Also from, the MAP consistency, we know that the MAP estimate of the Fisher information matrix, \hat{J} , converges in probability to true value, $J(\theta^0)$.

Now, if X_n converges in distribution to X , and Y_n converges in probability to Y , we know that the pair $[X_n, Y_n]$ converges in distribution to $[X, Y]$. Hence, the pair $[\|d_{t-h}\|, \hat{J}]$ converges in distribution to $[x, J(\theta^0)]$, where x is a chi-square variate with $t-h$ degrees of freedom. So, from the continuous mapping theorem, the e -value against H , $\bar{\text{ev}}(H)$, converges in distribution to $\bar{e} = Q(t, x)$, where x is a chi-square variate with $t-h$ degrees of freedom.

Since the cumulative chi-square distribution is an increasing function, we can invert the last formula, i.e., $\bar{e} = Q(t, x) \leq c \Leftrightarrow x \leq Q^{-1}(t, c)$. But, since x is a chi-square variate with $t-h$ degrees of freedom,

$$\text{Pr}(\bar{e} \leq c) = Q(t-h, Q^{-1}(t, c)) , \quad \text{Q.E.D.}$$

A similar argument, using a non-central chi-square distribution, proves the other asymptotic statement.

2.3 Loss Functions

In orthodox decision theoretic Bayesian statistics, a significance test is legitimate if and only if it can be characterized as an Acceptance (A) or Rejection (R) decision procedure defined by the minimization of the posterior expectation of a loss function, Λ . Madruga [36] gives the following family of loss functions characterizing the FBST. This loss function is based on indicator functions of θ being or not in the tangential set \overline{T} :

$$\Lambda(R, \theta) = aI(\theta \notin \overline{T}) \quad , \quad \Lambda(A, \theta) = b + dI(\theta \in \overline{T})$$

The interpretation of this loss function is as follows: If $\theta \in \overline{T}$ we want to reject H , for θ is more probable than anywhere in H ; If $\theta \in T$ we want to accept H , for θ is less probable than somewhere in H . The minimization of this loss function gives the optimal test:

$$\text{Accept } H \text{ iff } \text{ev}(H) \geq \varphi = (b + c)/(a + c) .$$

Note that this loss function is dependent on the observed sample (via the likelihood function), on the prior, and on the reference density, stressing the important point of non-separability of utility and probability, see Kadane and Winkler [27] and Rubin [49].

This type of loss function can be easily adapted in order to provide an asymptotic indicator checking if the true parameter belongs to the hypothesis set, $I(\theta^0 \in H)$. Consider the *tangential reference mass*,

$$\overline{m} = \left[\int_{\overline{T}(s^*)} r(\theta) d\theta \right]^\gamma$$

If $\gamma = 1$, \overline{m} is the reference density mass of the tangential set. If $\gamma = 1/t$, \overline{m} is a pseudo-distance from $\widehat{\theta}$ to θ^* . Consider also a threshold of the form $\varphi_1 = b\overline{m}$ or $\varphi_2 = b\overline{m}/(a + \overline{m})$, $a, b > 0$, in the expression of the optimal test above.

If $\theta^0 \notin H$, then $\widehat{\theta} \rightarrow \theta^0$ and $\theta^* \rightarrow \theta^{0*}$, where $\theta^{0*} \neq \theta^0$. Therefore $\|\widehat{\theta} - \theta^*\| \rightarrow c_1 > 0$. However, the standardized posterior, p_n , converges to a normal distribution centered on θ^0 . Hence, $\overline{m} \rightarrow c_2 > 0$ and $\varphi \rightarrow c_3 > 0$. Finally, since $\text{ev}(H) \rightarrow 0$, $\text{Pr}(\text{ev}(H) > \varphi) \rightarrow 0$.

If $\theta^0 \in H$, then $\widehat{\theta} \rightarrow \theta^0$ and $\theta^* \rightarrow \theta^0$. Therefore $\|\widehat{\theta} - \theta^*\| \rightarrow 0$, $\overline{m} \rightarrow 0$ and $\varphi \rightarrow 0$. Since $\text{ev}(H)$ converges to a proper distribution, see section 2.2, then, $\text{Pr}(\text{ev}(H) > \varphi) \rightarrow 1$.

2.4 Belief Calculi and Support Structures

Many standard Belief Calculi can be formalized in the context of Abstract Belief Calculus, ABC, see Darwiche and Ginsberg [13], Darwiche [14] and Stern [51]. In a Support Structure, $\langle \Phi, \oplus, \odot \rangle$, the first element is a Support Function, Φ , on a universe of statements, \mathcal{U} . Null and full support values are represented by $\mathbf{0}$ and $\mathbf{1}$. The second element is a support Summation operator, \oplus , and the third is a support Scaling or Conditionalization operator, \odot . A Partial Support Structure, $\langle \Phi, \oplus \rangle$, lacks the scaling operation.

The Support Summation operator, \oplus , gives the support value of the disjunction of any two logically disjoint statements from their individual support values, i.e.,

$$\neg(A \wedge B) \Rightarrow \Phi(A \vee B) = \Phi(A) \oplus \Phi(B) .$$

Table 2.1. Support structures for some belief calculi, $c = \Phi(C = A \wedge B)$.

$\Phi(\mathcal{U})$	$a \oplus b$	0	1	$a \leq b$	$c \oslash a$	Calculus
$\{0, 1\}$	$\max(a, b)$	0	1	$a \leq b$	$\min(c, a)$	Classical Logic
$[0, 1]$	$a + b$	0	1	$a \leq b$	c/a	Probability
$[0, 1]$	$\max(a, b)$	0	1	$a \leq b$	c/a	Possibility
$\{0 \dots \infty\}$	$\min(a, b)$	∞	0	$b \leq a$	$c - a$	Disbelief

The Support Scaling operator, \oslash , gives the conditional support value of B given A from the unconditional support values of A and the conjunction $C = A \wedge B$, i.e.,

$$\Phi_A(B) = \Phi(A \wedge B) \oslash \Phi(A) .$$

Support structures for some standard belief calculi are given in Table 2.1, where the support values of the two statements and their conjunction are given by $a = \Phi(A)$, $b = \Phi(B)$, $c = \Phi(C = A \wedge B)$.

In Table 2.1, the relation $a \leq b$ indicates that the value a represents a stronger support than the value b . Darwiche and Ginsberg [13] and Darwiche [14] also give a set of axioms defining the essential functional properties of a (partial) support function. Stern [51] shows that the support $\Phi(H) = \text{ev}(H)$ complies with all these axioms.

In the FBST, the support values, $\Phi(H) = \text{ev}(H)$, are computed using standard probability calculus on Θ which has an intrinsic conditionalization operator. The computed e -values, on the other hand, have a possibilistic summation, i.e., the e -values in favor of a composite hypothesis $H = A \vee B$, is the most favorable e -value of in favor of each of its terms, i.e., $\text{ev}(H) = \max\{\text{ev}(A), \text{ev}(B)\}$. It is impossible however to define a simple scaling operator for this possibilistic support that is compatible with the FBST's e -value, ev , as it is defined.

Hence, two belief calculi are in simultaneous use in the full Bayesian significance testing setup: ev constitutes a possibilistic partial support structure that coexists in harmony with the probabilistic support structure given by the posterior probability measure in the parameter space.

Requirements (V) and (VI), i.e. *no ad hoc artifice* and *possibilistic support*, find a rich interpretation in the juridical or legal context, where they correspond to the some of the most basic juridical principles, see Stern [51].

Onus Probandi is a basic principle of legal reasoning, also known as Burden of Proof, see Gaskins [19] and Kokott [31]. It also manifests itself in accounting through the Safe Harbor Liability Rule:

“There is no liability as long as there is a reasonable basis for belief, effectively placing the burden of proof (Onus Probandi) on the plaintiff, who, in a lawsuit, must prove false a defendant’s misstatement, without making any assumption not explicitly stated by the defendant, or tacitly implied by an existing law or regulatory requirement.”

The Most Favorable Interpretation principle, which, depending on the context, is also known as Benefit of the Doubt, *In Dubito Pro Reo*, or Presumption of Innocence, is a

consequence of the Onus Probandi principle, and requires the court to consider the evidence in the light of what is most favorable to the defendant:

“Moreover, the party against whom the motion is directed is entitled to have the trial court construe the evidence in support of its claim as truthful, giving it its most favorable interpretation, as well as having the benefit of all reasonable inferences drawn from that evidence.”

2.5 Sensitivity and Inconsistency

For a given prior, likelihood and reference density, let $\eta = \text{ev}(H; p_0, L_x, r)$ denote the e -value supporting H . Let $\eta', \eta'' \dots$ denote the e -value with respect to references $r', r'' \dots$. The degree of inconsistency of the e -value supporting H , induced by a set of references, $\{r, r', r'' \dots\}$ is defined by the index

$$I\{\eta, \eta', \eta'' \dots\} = \max\{\eta, \eta', \eta'' \dots\} - \min\{\eta, \eta', \eta'' \dots\}$$

The same index can be used to study the degree of inconsistency of the e -value induced by a set of priors, $\{p_0, p'_0, p''_0 \dots\}$. One could also study the sensitivity of the e -value to a set of virtual sample sizes, $\{1n, \gamma'n, \gamma''n \dots\}$, $\gamma \in [0, 1]$, corresponding to scaled likelihoods, $\{L, L', L'' \dots\}$.

This intuitive measure of inconsistency can be made rigorous in the context of paraconsistent logic and bilattice structures, see Abe et al. [1], Alcantara et al. [2], Arieli and Avron [3], Costa [7], Costa and Subrahmanian [8] and Costa et al. [9], [10], [11].

The bilattice $B(C, D) = \langle C \times D, \leq_k, \leq_t \rangle$, given two complete lattices, $\langle C, \leq_c \rangle$, and $\langle D, \leq_d \rangle$, has two orders, the knowledge order, \leq_k , and the truth order, \leq_t , given by:

$$\begin{aligned} \langle c_1, d_1 \rangle \leq_k \langle c_2, d_2 \rangle &\Leftrightarrow c_1 \leq_c c_2 \text{ and } d_1 \leq_d d_2 \\ \langle c_1, d_1 \rangle \leq_t \langle c_2, d_2 \rangle &\Leftrightarrow c_1 \leq_c c_2 \text{ and } d_2 \leq_d d_1 \end{aligned}$$

The standard interpretation is that C provides the “credibility” or value in favor of a hypothesis (or statement) H , and D provides the “doubt” or value against H . If $\langle c_1, d_1 \rangle \leq_k \langle c_2, d_2 \rangle$, then we have more information (even if inconsistent) about situation 2 than 1. Analogously, if $\langle c_1, d_1 \rangle \leq_t \langle c_2, d_2 \rangle$, then we have more reason to trust (or believe) situation 2 than 1 (even if with less information).

For each of the bilattice orders we define a join and a meet operator, based on the join and the meet operators of the single lattices orders. More precisely, \sqcup_k and \sqcap_k , for the knowledge order, and \sqcup_t and \sqcap_t , for the truth order, are defined by the following equations:

$$\begin{aligned} \langle c_1, d_1 \rangle \sqcup_k \langle c_2, d_2 \rangle &= \langle c_1 \sqcup_c c_2, d_1 \sqcup_d d_2 \rangle, \langle c_1, d_1 \rangle \sqcap_k \langle c_2, d_2 \rangle = \langle c_1 \sqcap_c c_2, d_1 \sqcap_d d_2 \rangle \\ \langle c_1, d_1 \rangle \sqcup_t \langle c_2, d_2 \rangle &= \langle c_1 \sqcup_c c_2, d_1 \sqcap_d d_2 \rangle, \langle c_1, d_1 \rangle \sqcap_t \langle c_2, d_2 \rangle = \langle c_1 \sqcap_c c_2, d_1 \sqcup_d d_2 \rangle \end{aligned}$$

The “unit square” bilattice, $\{[0, 1] \times [0, 1], \leq, \leq\}$ has been routinely used to represent fuzzy or rough pertinence relations, logical probabilistic annotations, etc. The lattice $\{[0, 1], \leq\}$ is the standard unit interval, where the join and meet coincide with the max and min operators, $\sqcup = \text{max}$ and $\sqcap = \text{min}$.

In the unit square bilattice the “truth”, “false”, “inconsistency” and “indetermination” extremes are t, f, \top, \perp , whose coordinates are given in Figure 2.3. As a simple example, let

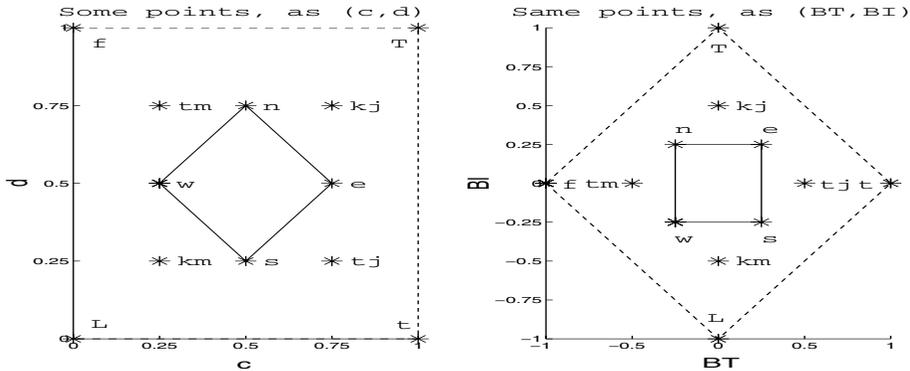


Figure 2.3. credibly-doubt and trust-inconsistency coordinates

region R be the convex hull of the four vertices n, s, e and w , given in Figure 2.3. Points kj, km, tj and tm are the knowledge and truth join and meet, over $r \in R$.

In the unit square bilattice, the degree of trust and degree of inconsistency for a point $x = \langle c, d \rangle$ are given by $BT(\langle c, d \rangle) = c - d$, and $BI(\langle c, d \rangle) = c + d - 1$, a convenient linear reparameterization of $[0, 1]^2$, to $[-1, +1]^2$. Figure 2.3 also compares the credibly-doubt and trust-inconsistency coordinates.

Let $\eta = ev(H)$, and $\bar{\eta} = \bar{ev}(H) = 1 - ev(H)$. The point $x = \langle \eta, \bar{\eta} \rangle$ in the unit square bilattice, represents herein a single e -value. Since $BI(x) = 0$, such a point is consistent. It is also easy to verify that for the multiple e -values, the definition of degree of inconsistency given above, is the degree of inconsistency of the knowledge join of all the single e -value points, i.e.,

$$I(\eta, \eta', \eta'' \dots) = BI(\langle \eta, \bar{\eta} \rangle \sqcup_k \langle \eta', \bar{\eta}' \rangle \sqcup_k \langle \eta'', \bar{\eta}'' \rangle \dots) .$$

Negation type operators are not an integral part of the bilattice structure but, in the unit square, one can define negation as $\neg \langle c, d \rangle = \langle d, c \rangle$, and conflation as $- \langle c, d \rangle = \langle 1 - c, 1 - d \rangle$, so that negation reverses trust, but preserves knowledge, and conflation reverses knowledge, but preserves trust.

As an example of sensitivity analysis we use the HW model with the standard uninformative references, the uniform and the maximum entropy densities, represented by $[1, 1, 1]$ and $[0, 0, 0]$ observation counts. Between these two uninformative references, we also consider perturbation references corresponding to $[0, 1, 1]$, $[1, 0, 1]$ and $[1, 1, 0]$ observation counts. Each of these references can be interpreted as the exclusion of a single observation of the corresponding type from the observed data set.

The e -values in the example are calculated using two sample proportions, $[x_1, x_2, x_3] = n[1/4, 1/4, 1/2]$ and $= n[1/4, 1/2, 1/4]$. The first exhibits the HW hypothesis symmetry, the second does not. The log 2 of sample size, $\log_2(n)$, ranged from 3 to 7. In Figure 2.4, the e -values corresponding to each choice of reference, are given by an interpolated dashed line. The interpretation of the vertical interval (solid bars) between the dashed lines is similar to that of the usual statistical error bars. However, the uncertainty represented by these bars does not have a probabilistic nature, being rather a possibilistic measure of inconsistency, defined in the partial support structure given by the FBST e -value, see Stern [52].

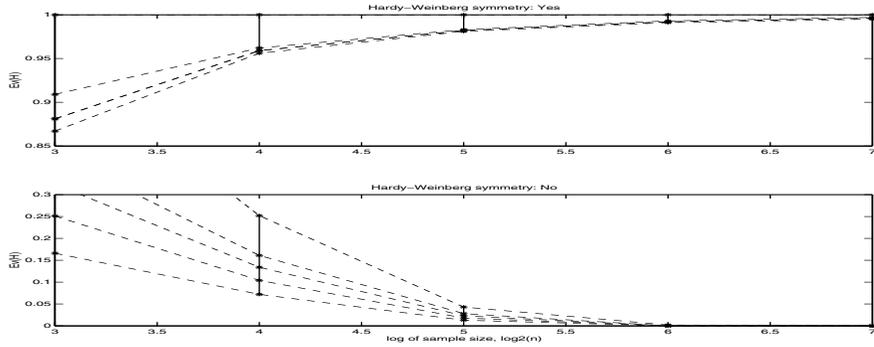


Figure 2.4. Sensitivity analysis

3 FBST Formal Structures

By a Full Bayesian Significance Test (FBST) *Structure*, we mean a quintuple

$$M = \{\Theta, H, p_0, p_n, r\}, \text{ where}$$

- Θ is the parameter space of an underlying statistical model $(S, \Sigma(S), P_\theta)$;
- $H: \theta \in \Theta_H \subseteq \Theta$ is a hypothesis on θ ; and
- p_0, p_n and r are, respectively, the prior, the posterior and the reference probability densities on Θ , all with respect to the same σ -finite measure μ on a measurable space $(\Theta, \Sigma(\Theta))$.

In the sequel we also use a relaxed notation, writing the hypothesis, H , instead of the set Θ_H defining it. Within a FBST structure, the following definitions are essential:

- The posterior *Surprise* function, $s(\theta)$, relative to the structure's reference density, $r(\theta)$, and its constrained and unconstrained suprema are defined as:

$$s(\theta) = p_n(\theta)/r(\theta),$$

$$s^* = s(\theta^*) = \sup_{\theta \in H} s(\theta), \quad \widehat{s} = s(\widehat{\theta}) = \sup_{\theta \in \Theta} s(\theta).$$

- The *Truth Function* or cumulative surprise distribution, $W: R_+ \mapsto [0, 1]$, and the *Untruth Function* of M , $\overline{W}(v)$, are defined as:

$$W(v) = \int_{\{s(\theta) \leq v\}} p_n(\theta) \mu(d\theta), \quad \overline{W}(v) = 1 - W(v).$$

- The *Truth Value*, $\text{ev}(H)$, or *e-value* supporting the hypothesis H in M , and the *Untruth Value*, $\overline{\text{ev}}(H)$ or *e-value* against H , are defined as:

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

As we will see in the next sections, it is not possible to obtain the truth value of a complex hypothesis only from the truth values of its elementary constituents. It is possible, however,

to obtain upper and lower bounds for the truth value of the complex hypothesis from the truth values of its elementary constituents.

We will also see that it is possible to obtain the truth function, W , of a complex structure, from the truth functions, W^j , of its elementary constituents, and the constrained supremum, s^* , of the complex structures surprise function from the elementary suprema, s^{*j} .

Since $\text{ev}(H) = W(s^*)$, the pair (W, s^*) will be referred to as the *Truth Summary* of the structure M .

Since we will be dealing in this paper, exclusively with complex hypotheses in an independent setup, we close this section by establishing the precise meaning of this framework. By an independent setup we mean that the FBST structures corresponding to the complex hypothesis H , $M = \{\Theta, H, p_0, p_n, r\}$, and to each of its elementary constituent hypotheses, H^j , $M^j = \{\Theta^j, H^j, p_0^j, p_n^j, r^j\}$, $j = 1, \dots, k$, bear the following relationships between their elements:

- the parameter space, Θ , of the underlying statistical model, $(S, \Sigma(S), P_\theta)$, is the product $\Theta^1 \times \Theta^2 \times \dots \times \Theta^k$;
- H , is equivalent to a logical composition (conjunctions and disjunctions) of H^1, H^2, \dots, H^k ;
- p_n and r , are probability densities with respect to the product measure $\mu = \mu^1 \times \mu^2 \times \dots \times \mu^k$ on $(\Theta, \Sigma(\Theta))$, where μ^j denote the σ -finite measure on $(\Theta^j, \Sigma(\Theta^j))$ with respect to which p_0^j, p_n^j and r^j are densities ; and
- the probability densities p_n and r are such that

$$p_n(\theta) = \prod_{j=1}^k p_n^j(\theta^j), \theta = (\theta^1, \dots, \theta^k) \in \Theta, \text{ and}$$

$$r(\theta) = \prod_{j=1}^k r^j(\theta^j), \theta = (\theta^1, \dots, \theta^k) \in \Theta.$$

4 Truth-Values Inequalities for Conjunctions

In this section we shall investigate, within the independent setup, the question of whether the truth value of a complex hypothesis, H , can be obtained from the truth values of its elementary constituents, H^1, H^2, \dots, H^k . We consider the case of a conjunctive composite hypothesis, that is, the case in which H is equivalent to $H^1 \wedge H^2 \wedge \dots \wedge H^j$.

In this case only bounds can be obtained for the truth and untruth values of H , from the corresponding truth and untruth values of the elementary constituents, H^j :

PROPOSITION 4.1

If H is equivalent to $H^1 \wedge H^2 \wedge \dots \wedge H^k$, then

$$\prod_{j=1}^k \text{ev}(H^j) \leq \text{ev}(H^1 \wedge H^2 \wedge \dots \wedge H^k), \text{ and}$$

$$\prod_{j=1}^k \bar{\text{ev}}(H^j) \leq \bar{\text{ev}}(H^1 \wedge H^2 \wedge \dots \wedge H^k).$$

In order to prove proposition 4.1, the following lemmas will be needed:

LEMMA 4.2

For any conjunctive composite hypothesis H with elementary constituents H^1, H^2, \dots, H^k ,

$$s^* = \sup_{\theta \in H} s(\theta) = \prod_{j=1}^k \sup_{\theta^j \in H^j} s^j(\theta^j) = \prod_{j=1}^k s^{*j}.$$

PROOF. Since for $\theta \in H$, $s^j(\theta^j) \leq s^{*j}$, for $1 \leq j \leq k$, $s(\theta) = \prod_{j=1}^k s^j(\theta^j) \leq \prod_{j=1}^k s^{*j}$ so that $s^* \leq \prod_{j=1}^k s^{*j}$. On the other hand, if for $\epsilon > 0$ and $s = \prod_{j=1}^k (s^{*j} - \epsilon)$, there must exist $\theta \in \bigwedge_{j=1}^k H^j$ such that $s(\theta) = \prod_{j=1}^k s^j(\theta^j) > \prod_{j=1}^k (s^{*j} - \epsilon)$. Consequently, $\sup_{\theta \in H} s(\theta) > \prod_{j=1}^k (s^{*j} - \epsilon)$, and the result follows by making $\epsilon \rightarrow 0$. ■

LEMMA 4.3

$$\prod_{j=1}^k W^j(v^j) \leq W(\prod_{j=1}^k v^j),$$

where W^j , $1 \leq j \leq k$, and W are the truth functions of M^j , $1 \leq j \leq k$, and M , respectively.

PROOF. Let $G: R_+^k \mapsto [0, 1]$ be defined as

$$G(v^1, \dots, v^k) = \int_{\{s^1(\theta^1) \leq v^1, \dots, s^k(\theta^k) \leq v^k\}} p_n(\theta) \mu(d\theta).$$

Since $s = \prod_{j=1}^k s^j$, $\mu = \prod_{j=1}^k \mu^j$, and

$$\begin{aligned} \{s^1(\theta^1) \leq v^1, \dots, s^k(\theta^k) \leq v^k\} &\subseteq \\ \left\{ \prod_{j=1}^k s^j(\theta^j) \leq \prod_{j=1}^k v^j \right\} &= \left\{ s(\theta) \leq \prod_{j=1}^k v^j \right\}, \end{aligned}$$

it follows that

$$\prod_{j=1}^k W^j(v^j) = G(v^1, \dots, v^k) \leq W(\prod_{j=1}^k v^j).$$

■

Proof of Proposition 4.1: In the inequality of Lemma 4.3, replacing each v^j by s^{*j} , $1 \leq j \leq k$, and then using Lemma 4.2, the first result in proposition 4.1 follows. The same argument proves the other assertion.

Consequently, if H is equivalent to $H^1 \wedge H^2 \wedge \dots \wedge H^k$, the truth values of the elementary constituent hypotheses give us lower and upper bounds for the truth value of the complex hypothesis. More precisely, ■

PROPOSITION 4.4

If H is equivalent to $H^1 \wedge H^2 \wedge \dots \wedge H^k$, then

$$\begin{aligned} \prod_{j=1}^k \text{ev}(H^j) &\leq \text{ev}(H^1 \wedge H^2 \wedge \dots \wedge H^k) \leq 1 - \prod_{j=1}^k (1 - \text{ev}(H^j)), \text{ and} \\ \prod_{j=1}^k \overline{\text{ev}}(H^j) &\leq \overline{\text{ev}}(H^1 \wedge H^2 \wedge \dots \wedge H^k) \leq 1 - \prod_{j=1}^k (1 - \overline{\text{ev}}(H^j)). \end{aligned}$$

In the null-or-full support case, that is, when, for $1 \leq j \leq k$, $s^{*j} = 0$ or $s^{*j} = \widehat{s}^j$, and the truth values of the simple constituent hypotheses are either 0 or 1, the bounds in proposition 4.4

are sharp. In fact, it is not hard to see that the composition rule of classical logic holds, that is,

$$\text{ev}(H^1 \wedge \dots \wedge H^k) = \begin{cases} 1 & , \quad \text{if } s^{*1} = \widehat{s}^1 \dots s^{*k} = \widehat{s}^k \\ 0 & , \quad \text{if, for some } j=1 \dots k, s^{*j} = 0 \end{cases}$$

In the example below, we show that the inequality in proposition 4.4 can, in fact be strict.

Example 4.5: In the third, first and second subplots of Figure 3, we have the graphs of truth functions corresponding, respectively, to the complex hypothesis $H^1 \wedge H^2$ and to its elementary constituents, H^1 and H^2 . Note that while $\text{ev}(H^1)=0.5$ and $\text{ev}(H^2)=0.7$, $\text{ev}(H^1 \wedge H^2)=0.64$, which is strictly greater than $\text{ev}(H^1)\text{ev}(H^2)=0.35$.

5 The Truth Operation for Conjunctions

In this section we shall investigate, also within the independent setup, the question of whether the truth function of the FBST structure corresponding to a complex hypothesis, H , can be obtained from the truth functions of the FBST structures corresponding to its elementary constituents, H^1, H^2, \dots, H^k . As in section 3, we consider the case of a conjunctive composite hypothesis, that is, the case in which H is equivalent to $H^1 \wedge H^2 \wedge \dots \wedge H^j$.

Definition 5.1: Given two probability distribution functions $G^1: R_+ \mapsto [0, 1]$ and $G^2: R_+ \mapsto [0, 1]$. Their Mellin convolution, $G^1 \otimes G^2$, is the distribution function defined by

$$G^1 \otimes G^2(v) = \int_0^\infty \int_0^{v/y} G^1(dx) G^2(dy) = \int_0^\infty G^1(v/y) G^2(dy) .$$

In probabilistic terms, the Mellin convolution $G^1 \otimes G^2$ gives us the distribution function of the product, of two independent random variables, X and Y , with distribution functions, G^1 and G^2 , respectively, see Kaplan and Lin [28], Springer [50] and Williamson [59]. From this interpretation, commutativity and associativity of Mellin convolution, \otimes , follows immediately. Furthermore, if G^1, G^2, \dots, G^k , are distribution functions defined on R_+ , their Mellin convolution is defined as

$$\bigotimes_{1 \leq j \leq k} G^j = G^1 \otimes G^2 \otimes \dots \otimes G^k(v) = G^1 \otimes (G^2 \otimes (\dots \otimes (G^{k-1} \otimes G^k(v)))$$

Lemma 5.2: For a conjunctive hypothesis H ,

$$H = \bigwedge_{j=1}^k H^j \quad , \quad W = \bigotimes_{1 \leq j \leq k} W^j .$$

Proof: 5.2 follows straight from the definition of W .

In view of the above result, we shall refer to the Mellin convolution, in the present context, as the *Truth Operation*.

In order to show that, together with the truth operation, truth summaries, (W^j, s^{*j}) , $1 \leq j \leq k$, efficiently synthesize the independent setup information, in the sense that the truth value of a complex hypothesis H can be obtained, the following lemma comes in handy.

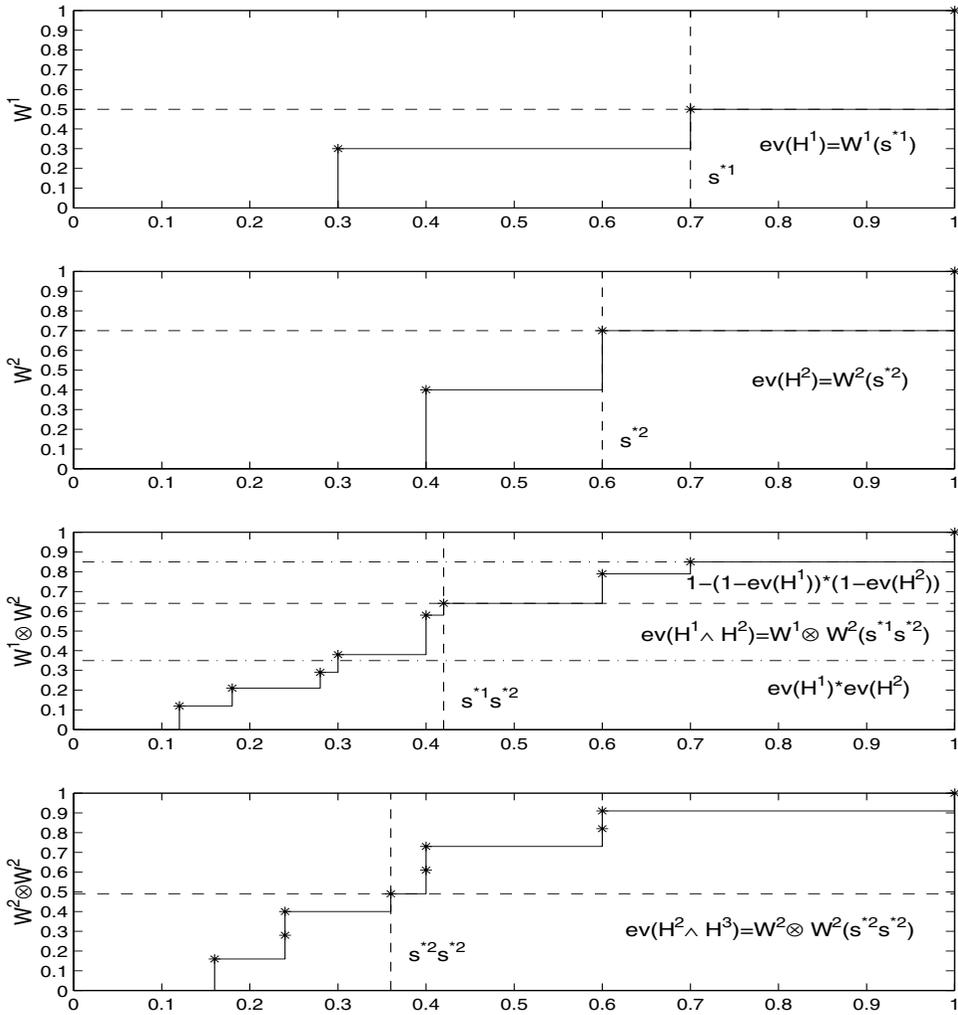


Figure 5.3. Subplots 1,2: W^j , s^{*j} , and $ev(H^j)$, for $j=1,2$;
 Subplot 3: $W^1 \otimes W^2$, $s^{*1} s^{*2}$, $ev(H^1 \wedge H^2)$ and bounds;
 Subplot 4: Structure M^3 is an independent replica of M^2 ,
 $ev(H^1) < ev(H^2)$, but $ev(H^1 \wedge H^3) > ev(H^2 \wedge H^3)$.

Proposition 5.3: If H is a complex hypothesis with elementary constituents H^1, H^2, \dots, H^k , and (W^j, s^{*j}) , $1 \leq j \leq k$, are their corresponding truth summaries, the truth value of H is given by

$$ev(H) = W(s^*) = \bigotimes_{1 \leq j \leq k} W^j \left(\prod_{j=1}^k s^{*j} \right).$$

PROOF. Immediate, from Lemmas 4.1 and 5.2. ■

6 Disjunctive Normal Form

Let us now consider the case where H is *Homogeneous* and expressed in *Disjunctive Normal Form*, that is:

$$H = \bigvee_{i=1}^q \bigwedge_{j=1}^k H^{(i,j)} ,$$

$$M^{(i,j)} = \{\Theta^j, H^{(i,j)}, p_0^j, p_n^j, r^j\} .$$

Let us also define $s^{*(i,j)}$ and $\widehat{s}^{(i,j)}$ as the respective constrained and unconstrained suprema of $s(\theta^{(i,j)})$ on the elementary hypotheses $H^{(i,j)}$.

Proposition 6.1:

$$\begin{aligned} \text{ev}(H) &= \text{ev}\left(\bigvee_{i=1}^q \bigwedge_{j=1}^k H^{(i,j)}\right) = W\left(\sup_{i=1}^q \prod_{j=1}^k s^{*(i,j)}\right) = \\ &= \max_{i=1}^q W\left(\prod_{j=1}^k s^{*(i,j)}\right) = \max_{i=1}^q \text{ev}\left(\bigwedge_{j=1}^k H^{(i,j)}\right) \end{aligned}$$

PROOF. Since the supremum of a function over the (finite) union of q sets, is the maximum of the suprema of the same function over each set, and W is monotonically increasing, the result follows. \blacksquare

Proposition 6.1 asserts the *Possibilistic* nature of the FBST truth value, that is, the e-value of a disjunction is the maximum e-value of the disjuncts, see Stern [51], Klir and Folger [30] and Zadeh [61].

It is worth mentioning that the present article does not abridge the most general composition cases of nested or heterogeneous structures, that is, the cases in which composite hypotheses are simultaneously assessed in heterogeneous sub-structures of (possibly) different dimensions. The following example indicates that this is not a trivial matter:

Example 6.2: Let $m = \text{argmax}_{j=1,2} \text{ev}(H^j)$ and H be equivalent to $(H^1 \vee H^2) \wedge H^3$. Is it true that $\text{ev}(H) =$

$$\max\{\text{ev}(H^1 \wedge H^3), \text{ev}(H^2 \wedge H^3)\} = \text{ev}(H^m \wedge H^3) ?$$

Interestingly the answer is in the negative. In the third and fourth subplots of Figure 3 we have the graphs of the Truth Functions corresponding, respectively, to the complex hypothesis $H^1 \wedge H^3$ and $H^2 \wedge H^3$, where the structure M^3 is an independent replica of M^2 . Observe that $\text{ev}(H^1) = 0.5 < \text{ev}(H^2) = 0.7$, but while $\text{ev}(H^1 \wedge H^3) = 0.64$, $\text{ev}(H^2 \wedge H^3) = 0.49$.

7 Final Remarks and Future Research

This paper gives a theoretical framework for the compositionality problem in the context of parametric statistical hypothesis testing, based on the FBST e -value, $\text{ev}(H)$. Forthcoming papers illustrate several applications, like simultaneous calibration of measurement procedures, and psychometric analysis on learning experiments, see Stern [53], [54], [55], [56], Foerster [18] and Piaget [46].

Forthcoming papers will also detail the implementation of computational procedures for estimating the truth function, $W(v), 0 \leq v \leq \widehat{s}$, by Markov Chain Monte Carlo (MCMC).

Such procedures require only minor adaptations, with small computational overhead, of the MCMC for estimating $\text{ev}(H) = W(s^*)$, see Gilks et al. [22], Evans and Swartz [17] and Stern and Zacks [58].

Forthcoming articles work on applications using the results obtained in this paper for testing complex hypotheses that are conditionally independent, in the context of Dirchlet-Multinomial models, Bayesian networks, and structural analysis, see Cozman [12] and Jensen [26], and MacDonald [34]. All these applications rely on conditional independence and decoupling, a pivotal concept in the discussion of causal relations in the context of mathematical statistics, see Pearl [37] and Stern [55], [56].

References

- [1] J.M.Abe, B.C.Avila, J.P.A.Prado (1998). *Multi-Agents and Inconsistence*. ICCIMA'98. 2nd International Conference on Computational Intelligence and Multimedia Applications. Traralgon, Australia.
- [2] J.Alcantara, C.V.Damasio, L.M.Pereira (2002). Paraconsistent Logic Programs. JELIA-02. 8th European Conference on Logics in Artificial Intelligence. *Lecture Notes in Computer Science*, 2424, 345–356.
- [3] O.Arieli, A.Avron (1996). Reasoning with Logical Bilattices. *Journal of Logic, Language and Information*, 5, 25–63.
- [4] R.E.Barlow, F.Prochan (1981). *Statistical Theory of Reliability and Life Testing Probability Models*. Silver Spring: To Begin With.
- [5] D.Basu, J.K.Ghosh (1988). Statistical Information and Likelihood. *Lecture Notes in Statistics*, 45.
- [6] Z.W.Birnbaum, J.D.Esary, S.C.Saunders (1961). Multicomponent Systems and Structures, and their Reliability. *Technometrics*, 3, 55-77.
- [7] N.C.A.Costa (1963). Calculs Propositionnels pour les Systemes Formales Incosistants. *Compte Rendu Acad. des Scienses*, 257, 3790–3792.
- [8] N.C.A.Costa, V.S.Subrahmanian (1989). Paraconsistent Logics as a Formalism for Reasoning about Inconsistent Knowledge Bases. *Artificial Inteligence in Medicine*, 1, 167–174.
- [9] N.C.A.Costa, C.A.Vago, V.S.Subrahmanian (1991). Paraconsistent Logics Ptt. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 37, 139-148.
- [10] N.C.A.Costa, J.M.Abe, V.S.Subrahmanian (1991). Remarks on Annotated Logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 37, 561–570.
- [11] N.C.A.Costa, J.M.Abe, A.C.Murolo, J.I.da Silva, C.F.S.Casemiro (1999). *Lógica Paraconsistente Aplicada*. São Paulo: Atlas.
- [12] F.G.Cozman (2000). Generalizing Variable Elimination in Bayesian Networks. Proceedings of the Workshop in Probabilistic Reasoning in Artificial Inteligence. Atibaia.
- [13] A.Y.Darwiche, M.L.Ginsberg (1992). A Symbolic Generalization of Probability Theory. AAAI-92. 10-th Conf. American Association for Artificial Intelligence.
- [14] A.Y.Darwiche (1993). *A Symbolic Generalization of Probability Theory*. Ph.D. Thesis, Stanford Univ.
- [15] J.S.Dugdale (1996). *Entropy and Its Physical Meaning*. Taylor & Francis.
- [16] M.Evans (1997). Bayesian Inference Procedures Derived via the Concept of Relative Surprise. *Communications in Statistics*, 26, 1125–1143.

- [17] M.Evans, T.Swartz (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford University Press.
- [18] H.von Foerster (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. NY: Springer Verlag. The following articles in this anthology are of special interest: (a) On Self-Organizing Systems and their Environments; pp 1–19. (b) On Constructing a Reality; pp 211–227. (c) Objects: Tokens for Eigen-Behaviors; pp 261–271. Also in: B.Inhelder, R.Gracia, and J.Voneche (1987). *Hommage a Jean Piaget: Epistémologie Génétique et Equilibration*. Paris: Delachaux et Niestlé. (d) For Niklas Luhmann: How Recursive is Communication? pp 305–323.
- [19] R.H.Gaskins (1992). *Burdens of Proof in Modern Discourse*. Yale Univ. Press.
- [20] A.Gelman, J.B.Carlin, H.S.Stern, D.B.Rubin, (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- [21] A.George, J.R.Gilbert, J.W.H.Liu (1993). *Graph Theory and Sparse Matrix Computations*. NY: Springer.
- [22] W.R.Gilks, S.Richardson, D.J.Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. NY: CRC Press.
- [23] I.J.Good (1983). *Good Thinking*. Univ. of Minnesota.
- [24] O.Häggström (2002). *Finite Markov Chains and Algorithmic Applications*. Cambridge Univ.
- [25] T.Z.Irony, M.Lauretto, C.A.B.Pereira, and J.M.Stern (2002). A Weibull Wearout Test: Full Bayesian Approach. In: Y.Hayakawa, T.Irony, M.Xie, edit. *Systems and Bayesian Reliability*, 287–300. *Quality, Reliability & Engineering Statistics*, 5, Singapore: World Scientific.
- [26] F.V.Jensen (2001). *Bayesian Networks and Decision Graphs*. NY: Springer.
- [27] J.B.Kadane, R.L.Winkler (1987). De Finetti's Methods of Elicitation. In Viertl (1987).
- [28] S.Kaplan, C.Lin (1987). An Improved Condensation Procedure in Discrete Probability Distribution Calculations. *Risk Analysis*, 7, 15-19.
- [29] J.N.Kapur (1989). *Maximum Entropy Models in Science and Engineering*. NY: Wiley.
- [30] G.J.Klir, T.A.Folger (1988). *Fuzzy Sets, Uncertainty and Information*. NY: Prentice Hall.
- [31] J.Kokott (1998). *The Burden of Proof in Comparative and International Human Rights Law*. Hague: Kluwer.
- [32] M.Lauretto, C.A.B.Pereira, J.M.Stern, S.Zacks (2003). Full Bayesian Significance Test Applied to Multivariate Normal Structure Models. *Brazilian Journal of Probability and Statistics*, 17, 147-168.
- [33] M.Lauretto, J.M.Stern (2005). FBST for Mixture Model Selection. MaxEnt 2005, 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. *American Institute of Physics Conference Proceedings*, 803, 121–128.
- [34] R.P.Macdonald (1985). *Factor Analysis and Related Methods*. London: LEA.
- [35] M.R.Madruga, L.G.Esteves, S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.
- [36] M.R.Madruga, C.A.B.Pereira, J.M.Stern (2003). Bayesian Evidence Test for Precise Hypotheses. *Journal of Statistical Planning and Inference*, 117, 185–198.
- [37] J.Pearl (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

- [38] C.A.B.Pereira, D.V.Lindley (1987). Examples Questioning the use of Partial Likelihood. *The Statistician*, 36, 15–20.
- [39] C.A.B.Pereira, J.M.Stern (1999). A Dynamic Software Certification and Verification Procedure. Proc. *ISAS-99, Int. Conf. on Systems Analysis and Synthesis*, 2, 426–435.
- [40] C.A.B.Pereira, J.M.Stern, (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy Journal*, 1, 69–80.
- [41] C.A.B.Pereira, J.M.Stern (2001). Full Bayesian Significance Tests for Coefficients of Variation. In: George, E.I. (Editor). *Bayesian Methods with Applications to Statistics*, 391-400. Monographs of Official Statistics, EUROSTAT.
- [42] C.A.B.Pereira, J.M.Stern, (2001). Model Selection: Full Bayesian Approach. *Environmetrics*, 12, 559–568.
- [43] C.A.B.Pereira, S.Wechsler (1993). On the Concept of p -value. *Brazilian Journal of Probability and Statistics*, 7, 159–177.
- [44] C.A.B.Pereira, S.Wechsler, J.M.Stern (2005). Can a Significance Test be Genuinely Bayesian? Submitted.
- [45] G.C.Pflug (1996). *Optimization of Stochastic Models: The Interface Between Simulation and Optimization*. Boston: Kluwer.
- [46] J.Piaget (1985). *Equilibration of Cognitive Structures: The Central Problem of Intellectual Development*. Univ. of Chicago.
- [47] H.Rouanet, J.M.Bernard, M.C.Bert, B.Lecoutre, M.P.Lecoutre, B.Le Roux (1998). *New Ways in Statistical Methodology. From Significance Tests to Bayesian Inference*. Berne: Peter Lang.
- [48] R.Royall (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.
- [49] H.Rubin (1987). A Weak System of Axioms for “Rational” Behaviour and the Non-Separability of Utility from Prior. *Statistics and Decisions*, 5, 47–58.
- [50] M.D.Springer (1979). *The Algebra of Random Variables*. NY: Wiley.
- [51] J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. 1st Bayesian Modeling Applications Workshop, UAI’03, and Laptec’03. 4th Cong. Logic Applied to Technology. *Frontiers in Artificial Intelligence and its Applications*, 101, 139–147.
- [52] J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA’04, *Lecture Notes Artificial Intelligence*, 3171, 134–143.
- [53] J.M.Stern (2005). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. Presented at FIS-2005, Third Conference on the Foundations of Information Science. *Cybernetics & Human Knowing*, 2007, 14, 1, 9–36.
- [54] J.M.Stern (2006). Language and the Self-Reference Paradox. Technical Report MAC-2006-02, Institute of Mathematics and Statistics, University of São Paulo. To appear in *Cybernetics & Human Knowing*, 2007, 14.
- [55] J.M.Stern (2006). Decoupling, Sparsity, Randomization and Objective Bayesian Inference. Technical Report MAC-2006-07, Institute of Mathematics and Statistics, University of São Paulo. To appear in *Cybernetics and Human Knowing*.
- [56] J.M.Stern (2006). Language, Metaphor and Metaphysics: The Subjective Side of Science. Technical Report MAC-2006-09, Institute of Mathematics and Statistics, University of São Paulo.
- [57] J.M.Stern, S.Zacks (2002). Testing Independence of Poisson Variates under the Holgate Bivariate Distribution. *Statistical and Probability Letters*, 60, 313–320.

- [58] J.M.Stern, S.Zacks (2003). Sequential Estimation of Ratios, with Applications to Bayesian Analysis. Technical Report MAC-2003-10, Institute of Mathematics and Statistics, University of São Paulo.
- [59] R.C.Williamson (1989). *Probabilistic Arithmetic*. Univ. of Queensland.
- [60] L.Wittgenstein (1921). *Tractatus Logico Philosophicus* (Logisch-Philosophische Abhandlung). (Ed.1999) NY: Dover.
- [61] L.A.Zadeh (1987). *Fuzzy Sets and Applications*. NY: Wiley.

Received June 21, 2006.