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

7th Conference on Multivariate Distributions with Applications
Aug. 8-13, 2010, Maresias, Brazil.

Julio Michael Stern
jstern@ime.usp.br
University of São Paulo,
Wagner Borges
wborges@mackenzie.com.br
Mackenzie Presbiterian University.

Logic J.of the IGPL, 2007, 15, 5-6, 401-420.

We analyze the relationship between the credibility, or truth value, of a complex hypothesis, H, and those of its elementary constituents, H^j, $j = 1 \ldots k$. This is the *Compositionality* question (ex. in analytical philosophy).

According to Wittgenstein,
(*Tractatus*, 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).

In reliability engineering, (Birnbaum, 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."

Composition operations:
Series and parallel connections;
Belief values and functions:
Survival probabilities and functions.

FBST - Full Bayesian Significance Test
Pereira and Stern (1999), Madruga (2003).

Bayesian paradigm: the posterior density, $p_n(\theta)$, is proportional to the product of the likelihood and a prior density,

$$p_n(\theta) \propto L(\theta \,|\, x)\, p_0(\theta).$$

Hypothesis: $H: \ \theta \in \Theta_H$,

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

Precise (sharp) hypothesis: $\dim(H) < \dim(\Theta)$, relaxed notation: H, instead of Θ_H.

Reference density, $r(\theta)$, interpreted as a representation of no information in the parameter space, or the limit prior for no observations, or the neutral ground state for the Bayesian operation. Standard (possibly improper) uninformative references include the uniform and maximum entropy(s) densities, $* * *$
see Dugdale (1996) and Kapur (1989)

FBST evidence value supporting and against the hypothesis H, $\mathsf{Ev}(H)$ and $\overline{\mathsf{Ev}}(H)$,

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

$$\widehat{s} = s(\widehat{\theta}) = \sup_{\theta\in\Theta} s(\theta) \ ,$$

$$s^* = s(\theta^*) = \sup_{\theta\in H} s(\theta) \ ,$$

$$T(v) = \{\theta \in \Theta \,|\, s(\theta) \le v\} \ , \quad \overline{T}(v) = \Theta - T(v) \ ,$$

$$W(v) = \int_{T(v)} p_n(\theta)\, d\theta \ , \quad \overline{W}(v) = 1 - W(v) \ ,$$

$$\mathsf{Ev}(H) = W(s^*) \ , \ \ \overline{\mathsf{Ev}}(H) = \overline{W}(s^*) = 1 - \mathsf{Ev}(H) \ .$$

$s(\theta)$ is the posterior surprise relative to $r(\theta)$. The tangential set $\overline{T}(v)$ is the HRSS. Highest Relative Surprise Set, above level v, $W(v)$ is the cumulative surprise distribution.

If $r \propto 1$ then $s(\theta) = p_n(\theta)$ and $\overline{T}$ is a HPDS. $r(\theta)$ implicitly gives the metric in Θ.

Hardy-Weinberg genetic equilibrium,
see (Pereira and Stern 1999).
n , sample size, x_1, x_3 , homozygote,
$x_2 = n - x_1 - x_3$, heterozygote count.

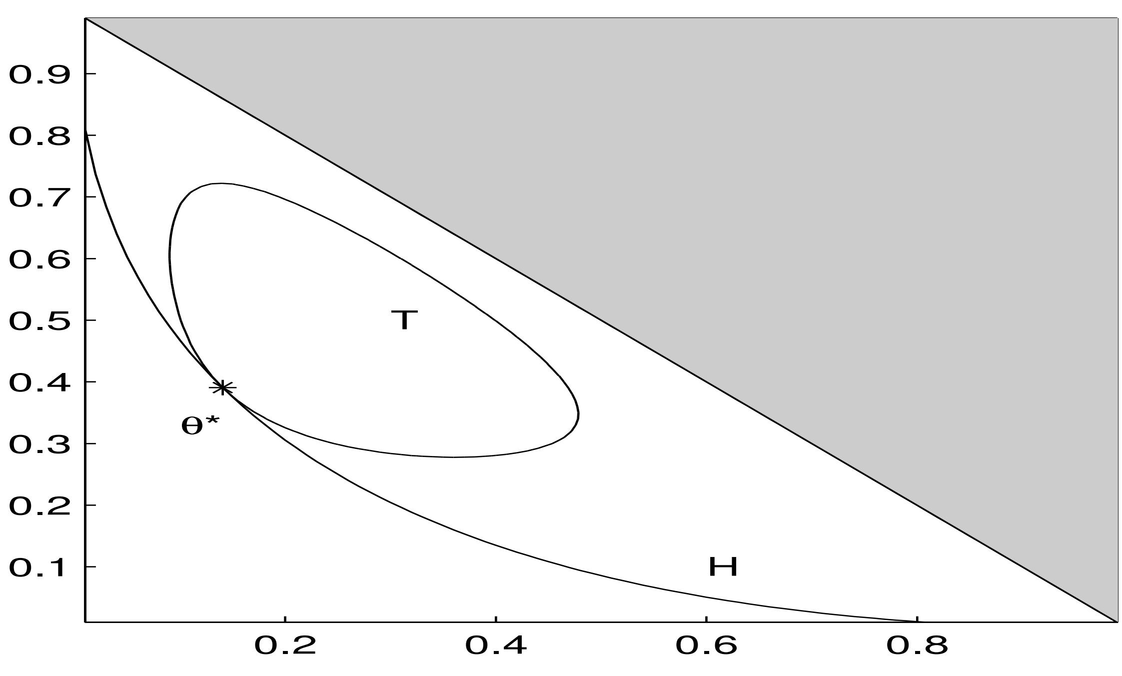
$$r(\theta) = p_0(\theta) \propto \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3} \ , \quad y =$$

$[0, 0, 0]$ (uniform) or $[-1, -1, -1]$ (max.ent.) ,

$$p_n(\theta \mid x) \propto \theta_1^{x_1 + y_1} \theta_2^{x_2 + y_2} \theta_3^{x_3 + y_3} \ ,$$

$$\Theta = \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\} \ ,$$

$$H = \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1}\ )^2\} \ .$$



Invariance:
Reparameterization of H (of $h(\theta)$): Trivial.
Reparameterization of Θ, (regularity cond.= bijective, integrable, a.s.cont.differentiable)

$$\omega = \phi(\theta) \quad , \quad \Omega_H = \phi(\Theta_H)$$

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

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

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

hence, $T(s^*) \mapsto \phi(T(s^*)) = \widetilde{T}(\widetilde{s}^*)$, and

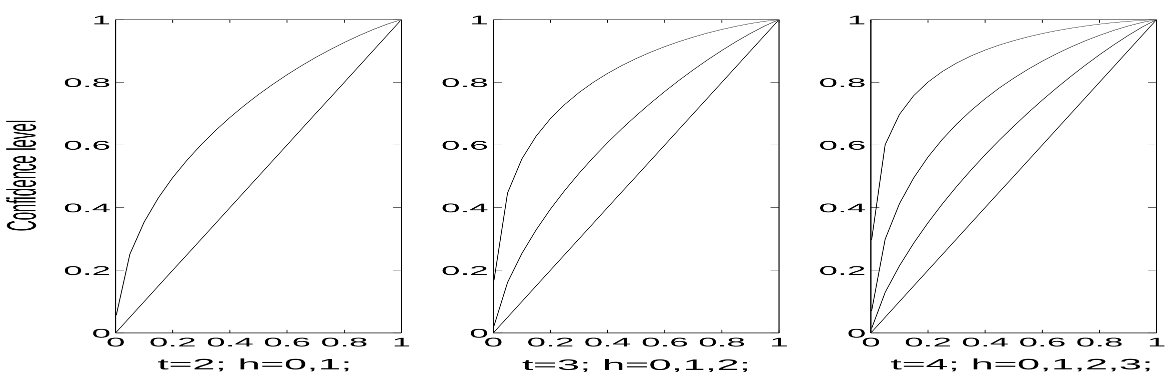
$$\widetilde{\mathsf{Ev}}(H) = \int_{\widetilde{T}(\widetilde{s}^*)} \widetilde{p}_n(\omega) d\omega =$$

$$\int_{T(s^*)} p_n(\theta) d\theta = \mathsf{Ev}(H) \quad , \quad \text{Q.E.D.}$$

Critical Level and Consistency:
$\overline{V}(c) = \Pr(\overline{\text{Ev}} \leq c)$, the cumulative distribution of $\overline{\text{Ev}}(H)$, given θ^0, the true parameter value.
Let $t = \dim(\Theta)$ and $h = \dim(H)$.
Under appropriate regularity conditions, for increasing sample size, $n \rightarrow \infty$,

- If H is false, $\theta^0 \notin H$, then $\overline{\text{Ev}}(H) \rightarrow 1$
- If H is true, $\theta^0 \in H$, then $\overline{V}(c)$, the confidence level, is approximated by the function

$$\text{Chi2}\left(t - h, \text{Chi2}^{-1}(t, c)\right) \; .$$



Test τ_c critical level vs. confidence level

Alternative approaches: Empirical power analysis Stern and Zacks (2002) and Lauretto (2004); Decision theory, Madruga (2001); Sensitivity analysis, Stern (2004). $* * *$

Comparative example:
Pereira, Stern, Wechsler (2005).
Independence in 2×2 contingency table.

$$H : \theta_{1,1} = (\theta_{1,1} + \theta_{1,2})(\theta_{1,1} + \theta_{2,1}) \ .$$

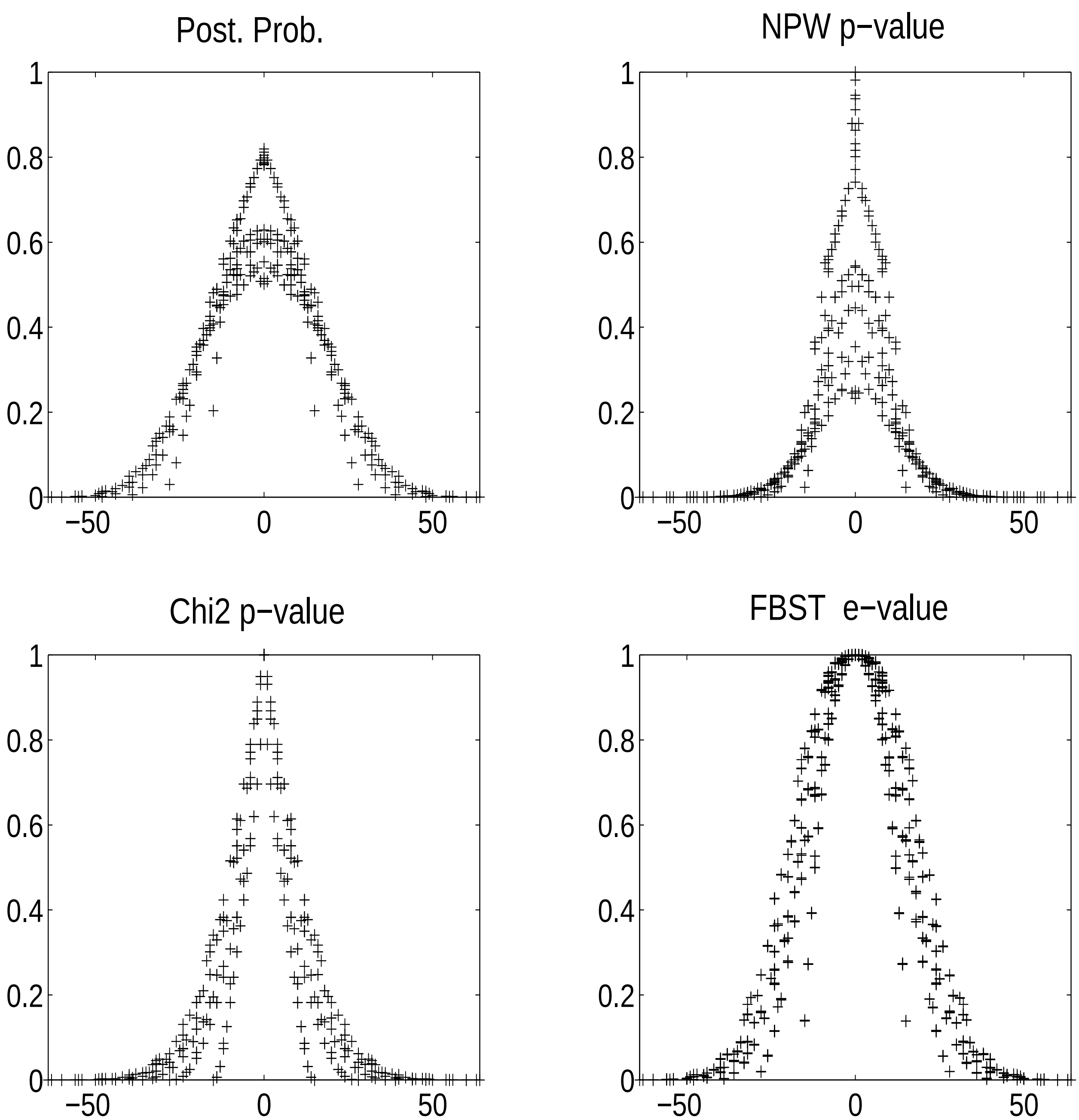
Figure 2 compares four statistics, namely,
-Bayes factor posterior probabilities (BF-PP),
-Neyman-Pearson-Wald (NPW) p-values,
-Chi-square approximate p-values, and the
-FBST evidence value in favor of H.

$$D = x_{1,1}x_{2,2} - x_{1,2}x_{2,1} \ ,$$

Horizontal axis: $D =$ diagonal asymmetry,
is the unormalized Pearson correlation,

$$\rho_{1,2} = \frac{\sigma_{1,2}}{\sigma_{1,1}\sigma_{2,2}} = \frac{\theta_{1,1}\theta_{2,2} - \theta_{1,2}\theta_{2,1}}{\sqrt{\theta_{1,1}\theta_{1,2}\theta_{2,1}\theta_{2,2}}} \ .$$

Wish list:
- Full symmetry gives H full support.
- Ev(H) in continuous and differentiable.



Independence Hypothesis, n=16

Numerical Computations: $\quad * * *$

- Integration Step, MCMC for $W(v)$:
(dominates computational time)
$g(\theta)$, importance sampling density,

$$W(v) = \frac{\int_\Theta Z_g^v(\theta) g(\theta) d\theta}{\int_\Theta Z_g(\theta) g(\theta) d\theta} \quad \text{where}$$

$$Z_g(\theta) = p_n(\theta)/g(\theta) \ , \quad Z_g^v(\theta) = I(v,\theta) Z_g(\theta) \quad ,$$

$$I(v,\theta) = 1(\theta \in T(v)) = 1(s(\theta) \le v) \ .$$

Precision analysis in Zacks and Stern (2003).

OBS: We can get $W : [0, \widehat{\theta}] \mapsto R$ at almost the same computational cost of $W(s^*) = \mathrm{Ev}(H)$.

- Optimization Step:
ALAG, Augmented Lagrangean Algorithm
(dominates program complexity)
Multimodality: SA, Simulated Annealing

Nuisance parameters and Model Selection:
see Basu (1988), and Pereira and Stern (2001).

Consider $H: \ h(\theta) = h(\delta) = 0$, $\theta = [\delta, \lambda]$ not a function of some of the parameters, λ.

"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? "

max_λ or $\int \ d\lambda$, the maximization or integration operators, are procedures to achieve this goal, in order to obtain a projected profile or marginal posterior function, $p_n(\delta)$.

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.

Abstract Belief Calculus, ABC,
see Darwiche, Ginsberg (1992).

$\langle \Phi, \oplus, \oslash \rangle$, Support Structure,
Φ , Support Function, for statements on $\mathcal{U}$.
Null and full support values are **0** and **1**.
$\oplus$, Support Summation operator,
$\oslash$, Support Scaling or Conditionalization,
$\langle \Phi, \oplus \rangle$, Partial Support Structure.

$\oplus$, gives the support value of the disjunction of any two logically disjoint statements from their individual support values,

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

$\oslash$, gives the conditional support value of B given A from the unconditional support values of A and the conjunction $C = A \wedge B$,

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

$\otimes$, unscaling: If Φ does not reject A,

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

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

$\Phi(\mathcal{U})$	$a \oplus b$	$\mathbf{0}$	$\mathbf{1}$	$a \preceq b$	$a \otimes b$	
$[0,1]$	$a + b$	0	1	$a \leq b$	$a \times b$	Pr
$[0,1]$	$\max(a,b)$	0	1	$a \leq b$	$a \times b$	Ps
$\{0,1\}$	$\max(a,b)$	0	1	$a \leq b$	$\min(a,b)$	CL
$\{0..\infty\}$	$\min(a,b)$	∞	0	$b \leq a$	$a + b$	DB

Pr= Probability, Ps= Possibility,
CL= Classical Logic, DB= Disbelief.

In the FBST setup, two belief calculi are in simultaneous use: Ev constitutes a possibilistic (partial) support structure coexisting in harmony with the probabilistic support structure given by the posterior probability measure in the parameter space, see also Zadeh (1987). See Klir (1988) for nesting prop. of $T(v)$.

FBST Compositionality:

Disjunction of (homogeneous) hypotheses
+ Possibilistic support structure $\Rightarrow$
Maximization as composition operation:
Stern (2003).

Structures: $\quad M^i = \{\Theta, H^i, p_0, p_n, r\}$.

$$\mathrm{Ev}\left(\bigvee_{i=1}^{q} H^i\right) = W\left(\max_{i=1}^{q} s^{*i}\right) = \max_{i=1}^{q}\left(\mathrm{Ev}(H^i)\right) \ ,$$

Onus Probandi, In Dubito Pro Reo, Presumption of Innocence, and Most Favorable Interpretation are basic principles of legal reasoning, see Gaskins (1992).

“The defendant 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.”

“The plaintiff has the burden of proof, and 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.”

A defendant describes a system (machine, software, etc.) by a parameter θ, and claims that θ has been set to a value in a legal or valid null set, H. Claiming that θ has been set at the most likely value must give the defendant’s claim full support, for being absolutely vague cannot put him in a better position.

$A : \theta \in \Theta$ and $\Rightarrow \mathrm{Ev}(A) = 1$, it is tautological.
$B : \theta \in \{\widehat{\theta}\} \Rightarrow \mathrm{Ev}(B) = 1$, for $\overline{T} = \emptyset$.

Conjunction of (homogeneous) hypotheses
+ Independent structures $\Rightarrow$
Mellin convolution as composition operation:
Borges and Stern (2005).

Structures: $\quad M^j = \{\Theta^j, H^j, p_0^j, p_n^j, r^j\}$.

$$\mathrm{Ev}\left(\bigwedge_{j=1}^{k} H^j\right) = \ W(s^*) = \bigotimes_{1\leq j\leq k} W^j \ \left(\prod_{j=1}^{k} s^{*j}\right) \ ,$$

Given two random variables, X and Y, with distributions $G^1, G^2 : R_+ \mapsto [0,1]$, the Mellin convolution, $G^1 \otimes G^2$, is the distribution of the product $Z = XY$, see Springer (1979),

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

$$\int_0^\infty G^1(z/y) G^2(dy) \ .$$

Ev(H), $W(v)$ and $\otimes$:
Truth value, function, operation.

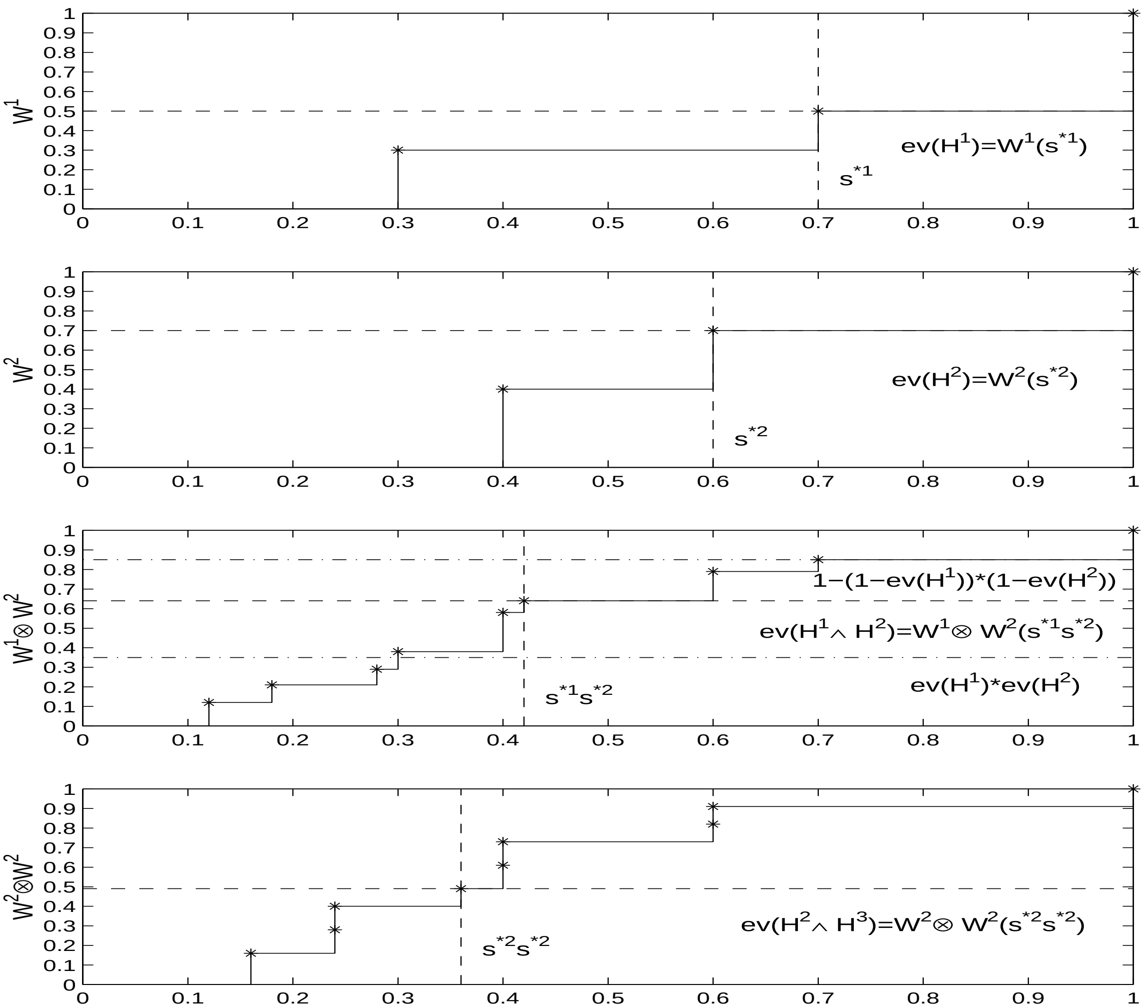


Fig.1,2: W^j, s^{*j}, and Ev(H^j), for $j = 1, 2$;
Fig.3: $W^1 \otimes W^2$, $s^{*1}s^{*2}$, Ev$(H^1 \wedge H^2)$ and bounds: Ev$(H^1) *$ Ev(H^2) and $1 - \overline{\text{Ev}}(H^1) * \overline{\text{Ev}}(H^2)$.
Fig.4: M^3 is an independent replica of M^2, Ev$(H^1) <$ Ev(H^2), but Ev$(H^1 \wedge H^3) >$ Ev$(H^2 \wedge H^3)$.

Compound H in Homogeneous Disjunctive Normal Form, (HDNF) + Independent (j) structures $\Rightarrow$

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

$$\mathrm{Ev}(H) \; = \; \mathrm{Ev}\left(\bigvee_{i=1}^{q} \bigwedge_{j=1}^{k} H^{(i,j)}\right) \; =$$

$$\max_{i=1}^{q} \mathrm{Ev}\left(\bigwedge_{j=1}^{k} H^{(i,j)}\right) \; =$$

$$W\left(\max_{i=1}^{q} \prod_{j=1}^{k} s^{*(i,j)}\right) \; ,$$

$$W = \bigotimes_{1 \leq j \leq k} W^j \; .$$

If all $s^* = 0 \vee \widehat{s}$, $\mathrm{Ev} = 0 \vee 1$, classical logic.

HDNF does not cover the most general composition cases of heterogeneous structures, dependent structures, etc. $* * *$

Constructivist Epistemology and Ontology: von Foerster (2003).

"Objects are tokens for eigen-behaviors."
(eigen-... = system's recurrent solution)
"Tokens stand for something else. In the cognitive realm, objects are the token names we give to our eigen-behavior. This is the constructivist's insight into what takes place when we talk about our experience with objects."
ex: ball, money (gold), wave (equation)...

"Eigenvalues have been found ontologically to be discrete (sharp), stable, separable and composable, while ontogenetically to arise as equilibria that determine themselves through circular processes. Ontologically, Eigenvalues and objects, and likewise, ontogenetically, stable behavior and the manifestation of a subject's 'grasp' of an object cannot be distinguished."

Scientific Production Diagram:
Maturana (1980), Krohn, Küppers (1990):

Experiment				Theory
Experiment design	$\Leftarrow$	Operatio- nalization	$\Leftarrow$	Hypotheses formulation
$\Downarrow$				$\Uparrow$
Effects observation		false/true eigensolution		Inter- pretation
$\Downarrow$				$\Uparrow$
Data acquisition	$\Rightarrow$	Explanation	$\Rightarrow$	Statistical analysis
Sample space				Parameter space

Scientific knowledge, structure and dynamics, as an autopoietic double feed-back system.

Statistical Inference:
Cognitive Constructivism or Idealism.
Rouanet (1998), Stern (2005).

- Predictive Probability Statements:
- Chance of observations in sample space.
- At the Experiment side of the diagram, the task of statistics is to make probabilistic statements about the occurrence of pertinent events, i.e. describe probabilistic distributions for what, where, when or which events can occur.

- Epistemic probability statements:
- Truth values in hypotheses space.
- At the Theory side of the diagram, the role of statistics is to measure the statistical support of (sharp) hypotheses, i.e. to measure, quantitatively, the hypothesis plausibility or possibility in the theoretical framework they were formulated, given the observed data.

OBS: Extravariability, measurement noise, and all other statistically significant factors ought to be incorporated into the model!

Noether theorems in physics, and
de Finetti type theorems in statistics: $***$

- NTs provide invariant physical quantities (conserv.laws) from symmetry transformation groups, and these are sharp hypotheses by excellence.
- dFTs provide invariant distributions from symmetry groups of the statistical model, generating prototypical sharp hypotheses in application areas, see Diaconis (1987,8), Eaton (1989), Feller (1968) and Ressel (1985,7,8).

Eigen-Solutions Composability:
Luhmann (1989), on the evolution of the scientific system. $***$

"This is something that idealization, mathematization, abstraction, etc. do not describe adequately. It concerns the increase in the capacity of decomposition and recombination, a new formulation of knowledge as the product of analysis and synthesis. ...uncovers an enormous potential for recombination."

-D.Basu, J.K.Ghosh (1988). Statistical Information and Likelihood. *Lecture Notes in Statistics*,45.
-W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J. IGPL*, 15, 5-6, 401-420.
-Z.W.Birnbaum, J.D.Esary, S.C.Saunders (1961). Multicomponent Systems and Structures Reliability. *Technometrics,* 3,55-77.
-A.Y.Darwiche, M.L.Ginsberg (1992). A Symbolic Generalization of Probability Theory. AAAI-92.
-J.S.Dugdale (1996). *Entropy and Its Physical Meaning.* London: Taylor & Francis.
-R.H.Gaskins (1992). *Burdens of Proof in Modern Discourse.* New Haven: Yale Univ. Press.
-J.N.Kapur (1989). *Maximum Entropy Models in Science and Engineering.* NY: Wiley.
-G.J.Klir, T.A.Folger (1988). *Fuzzy Sets, Uncertainty and Information.* NY: Prentice Hall.
-M.Lauretto, C.A.B.Pereira, J.M.Stern, S.Zacks (2003). Comparing Parameters of Two Bivariate Normal Distributions Using the Invariant Full Bayesian Significance Test. *Brazilian J. of Probability and Statistics,* 17, 147-168.
-M.R.Madruga, L.G.Esteves, S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.
-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.
-C.A.B.Pereira, S.Wechsler, J.M.Stern (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis,* 3, 1, 79-100.
-M.D.Springer (1979) *The Algebra of Random Variables.* NY:Wiley.
-J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. UAI'03 and Laptec'03, *Frontiers in Artificial Intell. and its Applications*, 101, 139–147.
-J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *LNAI*, 3171, 134–143.
-J.M.Stern (2005). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. MDPI, FIS2005, 61, 1–23.
-R.C.Williamson (1989) *Probabilistic Arithmetic.* Queensland Univ.
-L.Wittgenstein (1921). *Tractatus Logico Philosophicus* (Logisch-Philosophische Abhandlung). (Ed.1999) NY: Dover.