Bayesian Possibilistic Support for Sharp Hypotheses

Julio Michael Stern

BIOINFO and

Computer Science Dept.,

University of São Paulo, Brazil

jstern@ime.usp.br

www.ime.usp.br/ ∼jstern

FBST:

The Full Bayesian Significance Test (FBST), first presented by Pereira and Stern in 1999, is a coherent Bayesian significance test for sharp hypotheses.

In several applications that motivated the FBST it was desirable or necessary to use a test of sharp (precise) hypotheses with the following characteristics:

• Give an intuitive and simple measure of significance for the (null) hypothesis, ideally, a **probability** defined directly in the original (natural) **parameter space**.

• Be a **consistent** indicator for the hypothesis being tested, in the sense that increasing sample size should make it converge to the right Boolean (0/1 false/true) indicator.

• Have an intrinsically geometric definition, independent of any non-geometric aspect, like:
- The hypothesis (manifold) parameterization,
- The coordinate system on the parameter space,
i.e., be an **invariant** procedure.

• Be an **exact** procedure, i.e., make no use of "large sample" asymptotic approximations.

• Comply with the **Onus Probandi** juridical principle (or *In Dubito Pro Reo* rule), i.e. consider in the "most favorable way" the claim stated by the hypothesis.

• Obey the **likelihood principle**, that is, the information gathered from observations should be represented by, and only by, the likelihood function.

• Require **no ad hoc artifice** that could lead to judicial contention, like assigning a positive prior probability to a zero measure set, or setting an arbitrary initial belief ratio between hypotheses.

• Allow, (only) if desired, the incorporation of previous experience or expert opinion via a **subjective prior** distribution.

Let $\theta \in \Theta \subseteq R^p$ be a vector parameter of interest, and $L_x = L(\theta \,|\, x)$ the likelihood associated to the observed data x.

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

$$p_x(\theta) \propto p(\theta) L(\theta \,|\, x).$$

The (null) hypothesis H states that the parameter lies in the null set Θ_H,

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

For sharp hypotheses, $\dim(\Theta_H) < \dim(\Theta)$.

The posterior surprise, $s(\theta)$, relative to a given reference density $r(\theta)$, is:

$$s(\theta) = p_x(\theta)/r(\theta).$$

The supremum of the relative surprise function over the hypothesis (manifold) is:

$$s^* \;=\; s^*(\Theta_H, p, L_x, r) \;=\; \sup_{\theta \,\in\, \Theta_H} s(\theta)$$

The contour or level sets and the Highest Relative Surprise Set (HRSS), at level φ, are:

$$C(\varphi, p, L_x, r) \;=\; \{\theta \in \Theta \,|\, s(\theta) = \varphi\}$$

$$D(\varphi, p, L_x, r) \;=\; \{\theta \in \Theta \,|\, s(\theta) > \varphi\}$$

The FBST value of evidence against a hypothesis H, Ev(H), is defined by:

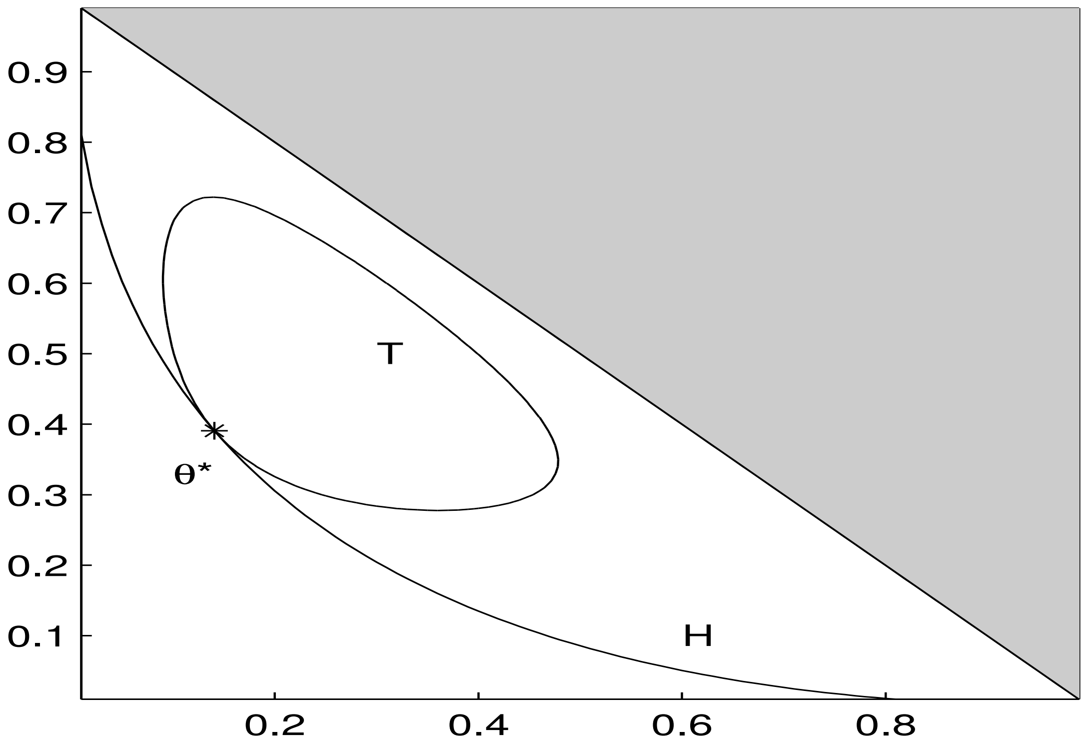
$$\begin{aligned}
\text{Ev}(H) &= \text{Ev}(\Theta_H, p, L_x, r) \\
&= \int_{T(H)} p_x(\theta)\, d\theta
\end{aligned}$$

$$\begin{aligned}
T(H) &= T(\Theta_H, p, L_x, r) \\
&= D(s^*, p, L_x, r)
\end{aligned}$$

The tangential HRSS, $T(H)$, contains the points in the parameter space whose surprise, relative to the reference density, is higher than that of any other point in the null set Θ_H. When the uniform reference density, $r(\theta) \propto 1$, is used, $T(H)$ is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set.

Interpretation: Small values of Ev(H) indicate that the hypothesis traverses high density regions, favoring the hypothesis.

$\overline{\text{Ev}}(H) = 1 - \text{Ev}(H)$ is the value of evidence supporting (or in favor of) hypothesis H.



Hardy-Weinberg genetic equilibrium:
n , sample size,
x_1, x_3 , homozygote counts,
$x_2 = n - x_1 - x_3$, heterozygote count,

$$
\begin{aligned}
p_x(\theta \mid x) &\propto \theta_1^{x_1-1}\theta_2^{x_2-1}\theta_3^{x_3-1} \\
r(\theta) &\propto \theta_1^{-1}\theta_2^{-1}\theta_3^{-1} \\
\Theta &= \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\} \\
\Theta_H &= \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1})^2\}
\end{aligned}
$$

The role of the reference density in the FBST is to make Ev(H) implicitly invariant under transformations of the coordinate system.

Invariance, as used in statistics, is a metric concept. The reference density is just a compact and interpretable representation for the reference metric in the original parameter space. This metric is given by the geodesic distance on the density surface.

The natural choice of reference density is an uninformative prior, 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 priors include the uniform and maximum entropy densities.

Possibilistic Support Structure:

Many standard Belief Calculi can be formalized in the context of Abstract Belief Calculus, ABC , of Darwiche and Ginsberg.

$\langle \Phi, \oplus, \oslash \rangle$ is a Support Structure , and
$\langle \Phi, \oplus \rangle$ is a Partial Support Structure.
Φ is the Support Function, on
$\mathcal{U}$, a universe of statements.
$\oplus$ is the support Summation operator,
$\oslash$ is the support Scaling or Conditionalization,
$\mathbf{0}$ and $\mathbf{1}$ indicate the minimal and maximal states of support.

$\oplus$ gives the support value of the disjunction $D = (A \vee B)$ of two logically disjoint statements from their individual support values, i.e.,

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

$\oslash$ 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 belief calculi, namely, classical logic (CL), probability calculus (PR), possibility calculus (PS), and disbelief calculus (DB), are given in the next table, where $C = A \wedge B$, $a = \Phi(A)$, $b = \Phi(B)$, $c = \Phi(C)$.

$\Phi(\mathcal{U})$	$a \oplus b$	$\mathbf{0}$	$\mathbf{1}$	$c \oslash a$	Calc.
$\{0, 1\}$	$\max(a, b)$	0	1	$\min(c, a)$	CL
$[0, 1]$	$a + b$	0	1	c/a	PR
$[0, 1]$	$\max(a, b)$	0	1	c/a	PS
$\{0 \ldots \infty\}$	$\min(a, b)$	∞	0	$c - a$	DB

In the FBST, the support values, $\Phi(H) = \overline{\text{Ev}}(H)$, are computed using standard probability calculus on Θ which has an intrinsic conditionalization operator.

The computed evidences, on the other hand, have a possibilistic summation, i.e., the value of evidence in favor of a composite hypothesis $H = A \vee B$, is the most favorable value of evidence in favor of each of its terms, i.e., $\overline{\text{Ev}}(H) = \max\{\overline{\text{Ev}}(A), \overline{\text{Ev}}(B)\}$.

It is impossible however to define a compatible scaling operator for this possibilistic support. Hence, two belief calculi are in simultaneous use in the FBST setup: $\overline{\text{Ev}}$ constitutes a possibilistic partial support structure coexisting in harmony with the probabilistic support structure given by the posterior probability measure in the parameter space.

Inconsistency aor Sensitivity Analysis:

For a given likelihood and prior density, let, $\eta = \mathsf{Ev}(\Theta_H, p, L_x, r)$ denote the value of evidence against a hypothesis H, with respect to reference r. Let $\eta', \eta'' \ldots$ denote the evidence against H with respect to references $r', r'' \ldots$.

The degree of inconsistency of the value of evidence against a hypothesis H, induced by a set of references, $\{r, r', r'' \ldots\}$ can be defined by the Inconsistency index

$$I \left\{ \eta, \eta', \eta'' \ldots \right\} =$$

$$\max \left\{ \eta, \eta', \eta'' \ldots \right\} - \min \left\{ \eta, \eta', \eta'' \ldots \right\}$$

This intuitive measure of inconsistency can be made rigorous in the context of paraconsistent logic and bilattice structures.

The degree of inconsistency for the evidence against H induced by multiple changes of the reference can be used as an index of imprecision or fuzziness of the value of evidence, Ev(H) , that can be interpreted within the possibilistic context of the partial support structure given by the evidence.

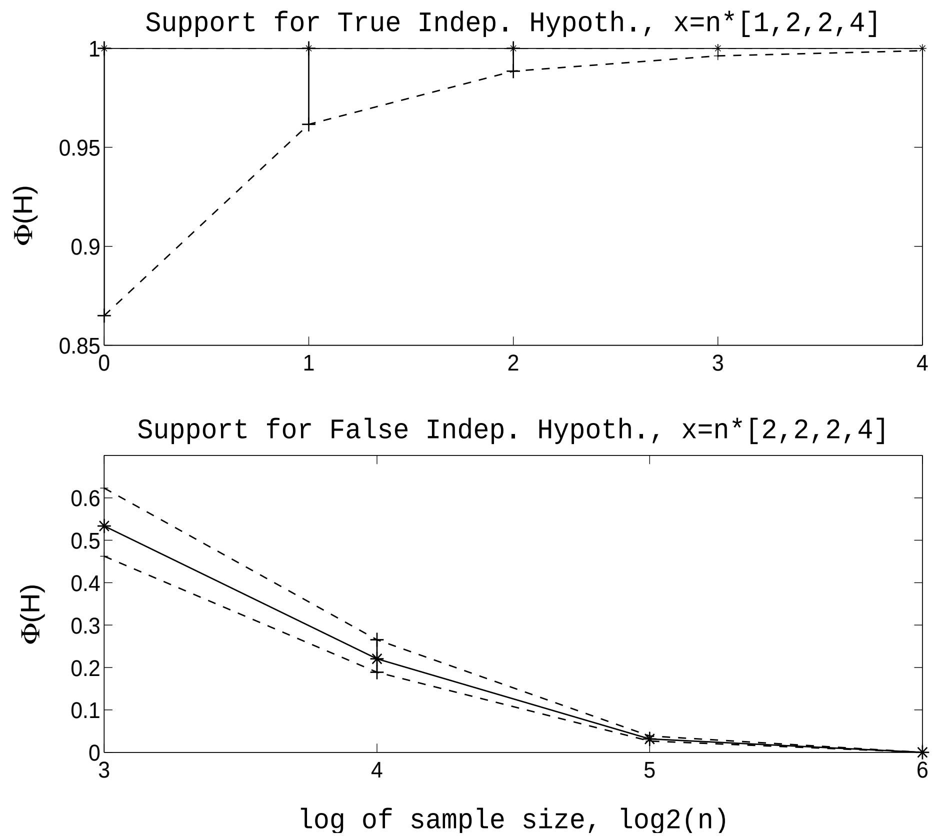
Some of the alternative ways of measuring the uncertainty of the value of evidence Ev(H), such as the *empirical power analysis*, have a dual possibilistic / probabilistic interpretation.

The degree of inconsistency also has the practical advantage of being inexpensive. When computing the evidence, only the integration limit, i.e. the threshold s^*, is changed, while the integrand, i.e. the posterior density, remains the same. Hence, when computing Ev(H), only a small computational overhead is required for the inconsistency analysis. In contrast, an empirical power analysis requires much more computational work than it is required to compute a single evidence.

Numerical Examples:
For the HW model we use as uniformative reference the standard maximum entropy density, that can be represented as $[-1, -1, -1]$ observation counts.

For the sensitivity analysis we also use the uniform reference, represented as $[0, 0, 0]$ observation counts, and intermediate “perturbation” references corresponding to $[-1, 0, 0]$, $[0, -1, 0]$ and $[0, 0, -1]$ observation counts.

The examples in Figure 2 are given by sample size factor and proportions,
$[x_1, x_2, x_3] = n * [1, 2, 1]$,
where the HW hypothesis is true, and
$[x_1, x_2, x_3] = n * [1, 1, 2]$,
where the HW hypothesis is false.



The induced degree of inconsistency is given by the vertical interval between the lines (solid bars), whose interpretation is similar to that of the usual statistical error bars.

Final Remarks:

The FBST evidence can be interpreted within the context of a possibilistic support structure. The semantic closure of such significance analysis relates to the notion of **Autopoiesis**, a term used by H.Maturana to express the concept of autonomy in Biology, Systems Theory and Epistemology.

The autopoietic nature of the FBST is in contrast with standard Bayesian significance analysis, using decision theoretic constructs based cost or utility functions.

If the original application problem context includes those economic concepts, we can still consider the decision theoretic analysis autopoietic. Therefore it is important to know that the FBST significance analysis is completely compatible with decision theory, this is shown by Madruga, Esteves and Wechsler in 2001.

However, there are many applications whose context does not necessarily include such economic constructs; as in problems found in science, law, politics, and other fields. In this case, an analysis based on exogenous concepts constitutes **Systemic Dedifferentiation** (Entdifferenzierung), a term by N.Luhmann.

In the specific situations where decision theoretic arguments based on economic concepts are not needed or recommended, the FBST allows for a statistical significance analysis that is immune to a dedifferentiation critique, like N.Luhmann:

"In this sense it is meaningless to speak of 'non-economic' costs. This is only a metaphorical way of speaking that transfers the specificity of the economic mode of thinking indiscriminately to other social systems".

Future Research:

Many traditional interpretations of Bayesian theory for hypothesis test rely on the epistemological tripod of empiricism, subjectivism and decision theory. The results of Madruga, Esteves and Wechsler imply that the FBST is fully compatible with this setting.

However, the autopoietic nature of the FBST analysis opens the possibility of providing Bayesian significance analysis that is friendly to users adopting other epistemological settings.

jstern@ime.usp.br

www.ime.usp.br/ ~jstern

Appendix: Bilattices

Given two complete lattices, $\langle C, \leq_c \rangle$, $\langle D, \leq_d \rangle$, $B(C, D)$ has Knowledge and Truth orders,

$$\begin{array}{rcl} B(C,D) & = & \langle C \times D, \leq_k, \leq_t \rangle \\ \langle c_1, d_1 \rangle \leq_k \langle c_2, d_2 \rangle & \Leftrightarrow & c_1 \leq_c c_2 \text{ and } d_1 \leq_d d_2 \\ \langle c_1, d_1 \rangle \leq_t \langle c_2, d_2 \rangle & \Leftrightarrow & c_1 \leq_c c_2 \text{ and } d_2 \leq_d d_1 \end{array}$$

Interpretation: C - credibility, D - doubt
If $\langle c_1, d_1 \rangle \leq_k \langle c_2, d_2 \rangle$, more information in 1 than 2 (even if inconsistent)
If $\langle c_1, d_1 \rangle \leq_t \langle c_2, d_2 \rangle$, more reason to trust 2 than 1 (even if with less information).

Join and a Meet operators, $\sqcup$ and $\sqcap$, for truth and knowledge orders:

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

Negation, $\neg$, and Conflation, $-$ properties, (if defined):

Ng1: $x \leq_k y \Rightarrow \neg x \leq_k \neg y$,

Ng2: $x \leq_t y \Rightarrow \neg y \leq_t \neg x$,

Cf1: $x \leq_k y \Rightarrow -y \leq_k -x$,

Cf2: $x \leq_t y \Rightarrow -x \leq_t -y$,

Ng3: $\neg\neg x = x$, Cf3: $--x = x$.

Ng: reverses trust, preserves knowledge,

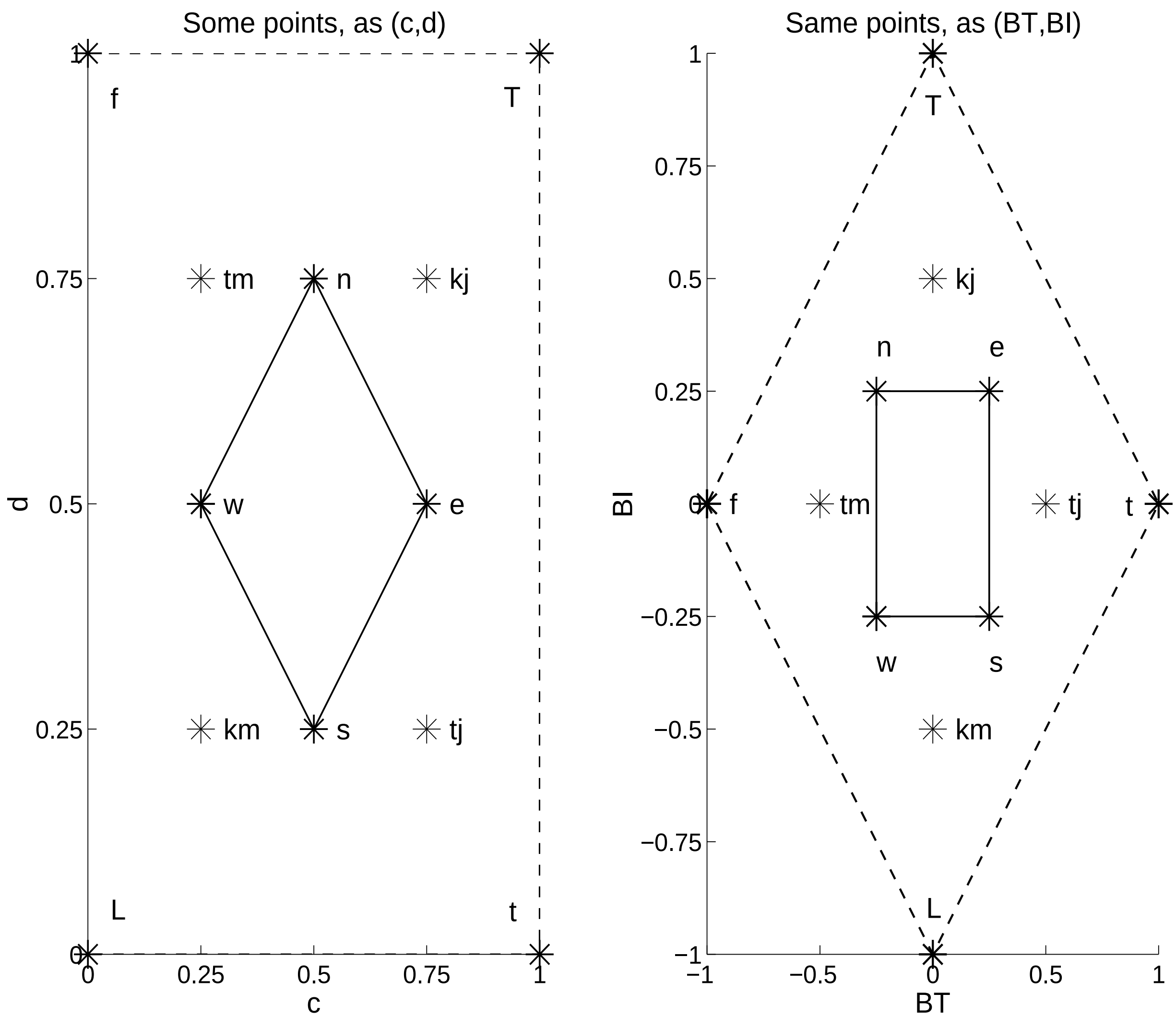
Cf: reverses knowledge, preserves trust.

Unit Square bilattice, over the standard Unit Interval lattice, $\langle [0,1], \leq \rangle$, where Join and Meet operators, $\sqcup$ and $\sqcap$, coincide with max and min operators. Negation and conflation operators are: $\neg \langle c,d \rangle = \langle d,c \rangle$, $- \langle c,d \rangle = \langle 1-c, 1-d \rangle$.

In Figure 2 we have the extremes points, t-truth, f-false, $\top$-inconsist., $\perp$-indeterm. Region R in the convex hull of points n-north, s-south, e-east and w-west. Points kj, km, tj and tm are knowledge and truth join and meet, over $r \in R$.

Degree of Trust and Inconsistency, for a point $x = \langle c,d \rangle$ in the Bilattice, are given by linear reparameterizations:

$$\mathrm{BT}(\langle c,d \rangle) = c - d \ , \quad \mathrm{BI}(\langle c,d \rangle) = c + d - 1 \ .$$


Some points, as (c,d)
Same points, as (BT,BI)
f
T
tm
n
kj
w
e
km
s
tj
L
t
d
c
0
0.25
0.5
0.75
1
BI
BT
−1
−0.5
−0.25
−0.75