Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses

Julio Michael Stern
jstern@ime.usp.br
University of São Paulo,
Brazil
and (several) Coauthors

www.ime.usp.br/∼jstern/slide/maxent08.pdf

$***$ = current research topics

FBST:

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

- Better performance (more power, etc.);

- Simpler formulation (more/new applications);

- Theoretical Statistical properties;

- Logical (Compositionality) properties;

- Epistemological / Ontological consequences.

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

1- Give an intuitive and simple measure of significance for sharp hypotheses, ideally, a **probability** defined directly in the original (natural) **parameter space**.

2- 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.

3- Give a smooth measure of significance, i.e. **continuous and differentiable**, on the hypothesis parameters and sample statistics, under appropriate regularity conditions.

4- **Likelihood principle**, i.e., the information gathered from observations should be represented (only) by the likelihood function.

5- Be able to provide an **exact** procedure, making no use of "large sample" asymptotic approximations.

6- Require **no ad hoc prior** information that could lead to judicial contention, like a positive probability mass on a zero measure set, or a belief ratio between hypotheses, etc.

7- Be able to provide a **consistent** test for a given sharp hypothesis, in the sense that increasing sample size should make it converge to the right accept/reject decision.

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

9- Provide a **possibilistic** support function.

10- Provide **compositionality** operations in complex models.

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, $\mathrm{Ev}(H)$ and $\overline{\mathrm{Ev}}(H)$,

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

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

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

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

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

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

$s(\theta)$ 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.

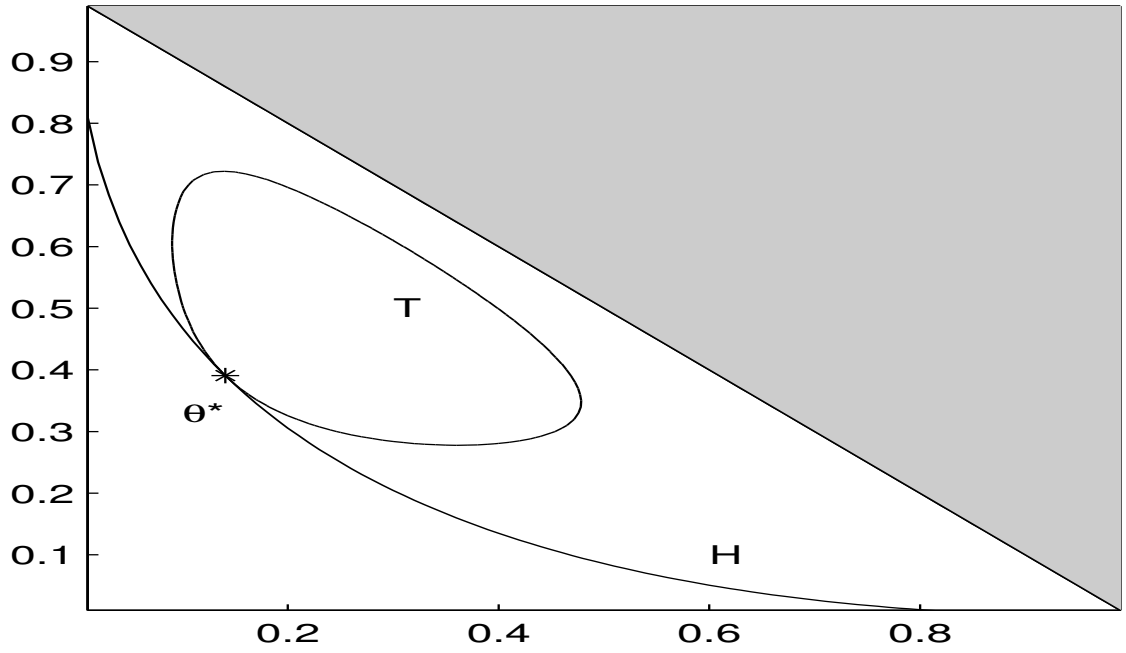
$$r(\theta) = p_0(\theta) \propto \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3} \ , \quad y =$$

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

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

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

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



Numerical Computations: $***$

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

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

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

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

Precision analysis in Lauretto (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 Lagrangian Algorithm
(dominates program complexity)
Multimodality: SA, Simulated Annealing

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

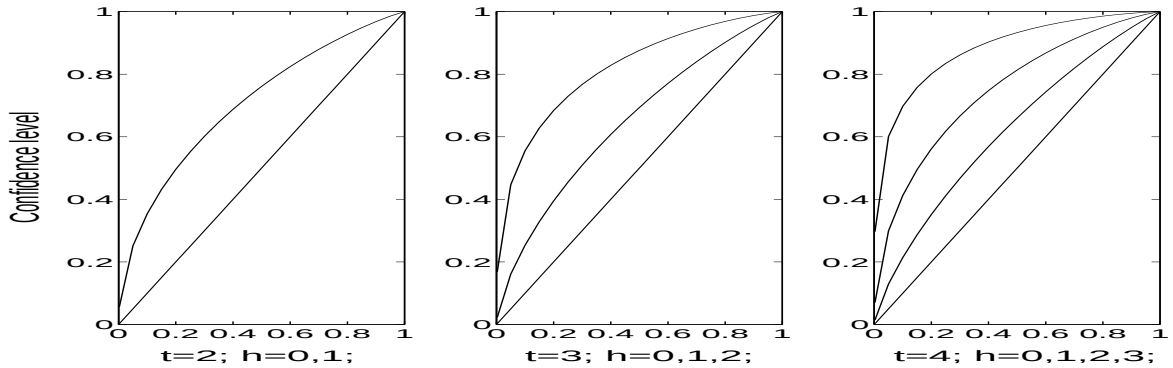
$$\widetilde{\mathsf{Ev}}(H) = \int_{\widetilde{T}(\widetilde{s}^*)} \widetilde{p}_n(\omega)d\omega =$$

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

Confidence 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, $n \to \infty$,

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

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



Test τ_c critical level vs. confidence level

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

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

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

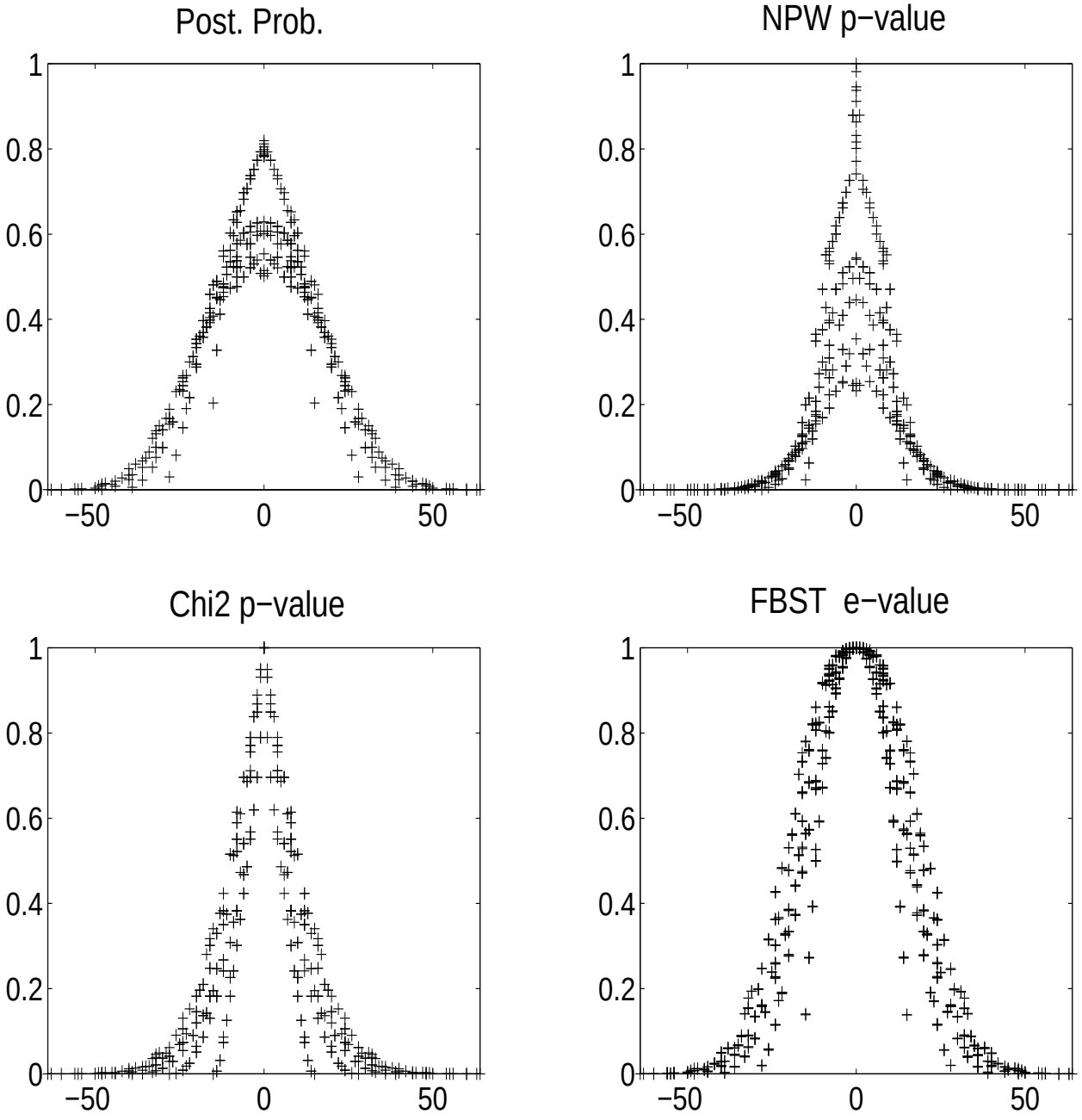
Figure 2 compares four statistics, namely,
-Bayes factor posterior probabilities (BF-PP),
-Neyman-Pearson-Wald (NPW) p-values,
-Chi-square approximate p-values, and the
-FBST evidence value in favor of H.

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

Horizontal axis: $D =$ diagonal asymmetry,
is the unnormalized 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

Samples "compatible with the hypothesis", having no asymmetry, are near the center, incompatible samples are to the sides.

The envelope curve for the FBST e-values, is smooth (differentiable) and therefore level at its maximum, where it reaches the value 1.

The envelope curves for the p-values take the form of a cusp, i.e. a pointed or broken curve. NPW p-values have, at the top of the cusp, a "spike" with symmetric samples, but having different outcome probabilities, "competing" for the higher p-value.

Collateral effect of the artifice that converts an epistemic question about H (in the parameter space), into a predictive question about X (in the sample space, conditional on H).

"increase sample size to reject".

Decision-theoretic, Orthodox Bayesian view:

"Gambling problems embrace the whole of (decision) theoretical statistics."

Epistemic questions about H are questions on How to Gamble on an H_0 against an H_1.

Standard procedure is Jeffrey's test:
- Gives a positive (ad hoc) mass to (sharp) H;
- Lindley's Paradox, unavoidable consequence: *"increase sample size to accept"*.

Dubins and Savage (1965):
The unacceptability of extreme (sharp) null hypotheses is perfectly well known; it is closely related to the often heard maxim that science disproves, but never proves, hypotheses...
The role of extreme (sharp) hypotheses in science and other statistical activities seems to be important but obscure.

In (a less) orthodox decision theoretic Bayesian approach, 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.

FBST loss function, Madruga (2001), based on indicator functions of θ being or not in the tangential set $\overline{T}$:

$$\Lambda(R,\theta) = a\, I(\theta \notin \overline{T}) \ , \quad \Lambda(A,\theta) = b + d\, I(\theta \in \overline{T}) \ .$$

Note that Λ depends 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 (1987).

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 several compositionality properties of the FBST.

Let us 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."

Goal: An analogy between the reliability of series / parallel structures and the likelihood of composite hypotheses in HDNF, Homogeneous Disjunctive Normal Form.

Abstract Belief Calculus, ABC,
see Darwiche, Ginsberg (1992),
and Stern (2003).

$\langle \Phi, \oplus, \oslash \rangle$, Support Structure,
Φ , Support Function, for statements on $\mathcal{U}$.
Null and full support values are $\mathbf{0}$ and $\mathbf{1}$.
$\oplus$, 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 $C = A \wedge B$,

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

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

$\Phi(\mathcal{U})$	$a \oplus b$	$\mathbf{0}$	$\mathbf{1}$	$a \preceq b$	$c \oslash a$	
$[0,1]$	$a+b$	0	1	$a \leq b$	c/a	Pr
$[0,1]$	$\max(a,b)$	0	1	$a \leq b$	c/a	Ps
$\{0,1\}$	$\max(a,b)$	0	1	$a \leq b$	$\min(c,a)$	CL
$\{0..\infty\}$	$\min(a,b)$	∞	0	$b \leq a$	$c-a$	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: $M^i = \{\Theta, H^i, p_0, p_n, r\}$.

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

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

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

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

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

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.

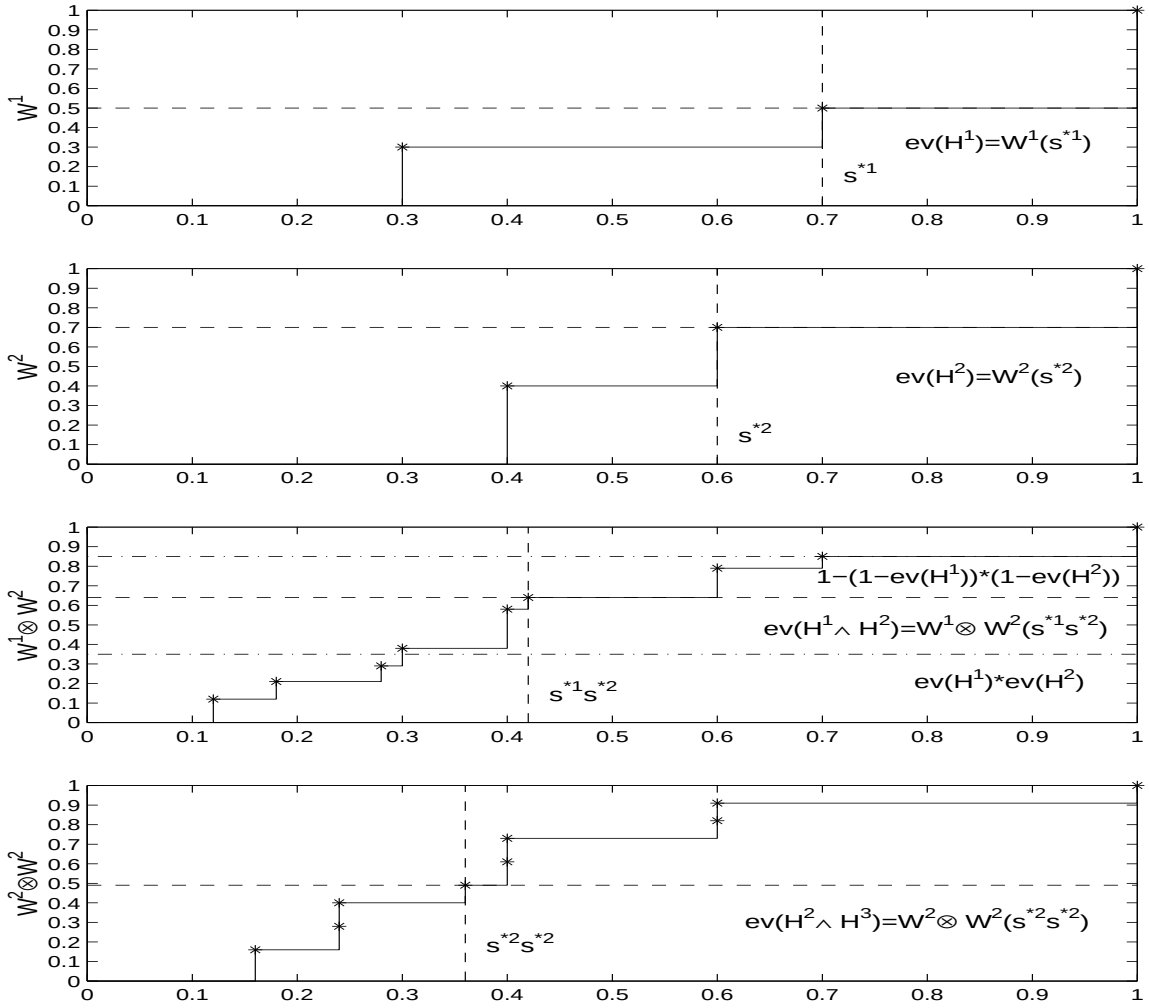


Fig.1,2: W^j, s^{*j}, and $\mathrm{Ev}(H^j)$, for $j = 1, 2$;
Fig.3: $W^1 \otimes W^2$, $s^{*1}s^{*2}$, $\mathrm{Ev}(H^1 \wedge H^2)$ and bounds: $\mathrm{Ev}(H^1) * \mathrm{Ev}(H^2)$ and $1 - \overline{\mathrm{Ev}}(H^1) * \overline{\mathrm{Ev}}(H^2)$.
Fig.4: M^3 is an independent replica of M^2, $\mathrm{Ev}(H^1)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

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

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

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

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

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

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

HDNF does (does not) cover the cases:
- No: General heterogeneous structures.
- Yes: Conditionally independent models,
  (Nested Dirichlet, Bayes Networks).

Inconsistency or Sensitivity Analysis:

For a given likelihood and prior density, let, $\eta = \text{Ev}(\Theta_H, p_0, L_x, r)$ denote the value of evidence against a hypothesis H.

Let $\eta, \eta', \eta'' \ldots$ denote the evidence against H with respect to a set of references, $r, r', r'' \ldots$, or priors, $p_0, p_0', p_0'' \ldots$, or scaled posteriors, $\{p_n^1, p_n^{\gamma'}, p_n^{\gamma''} \ldots\}$, $1 > \gamma' > \gamma'' > \ldots 0$, corresponding to virtual sample sizes $\{1n, \gamma' n, \gamma'' n \ldots\}$.

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.

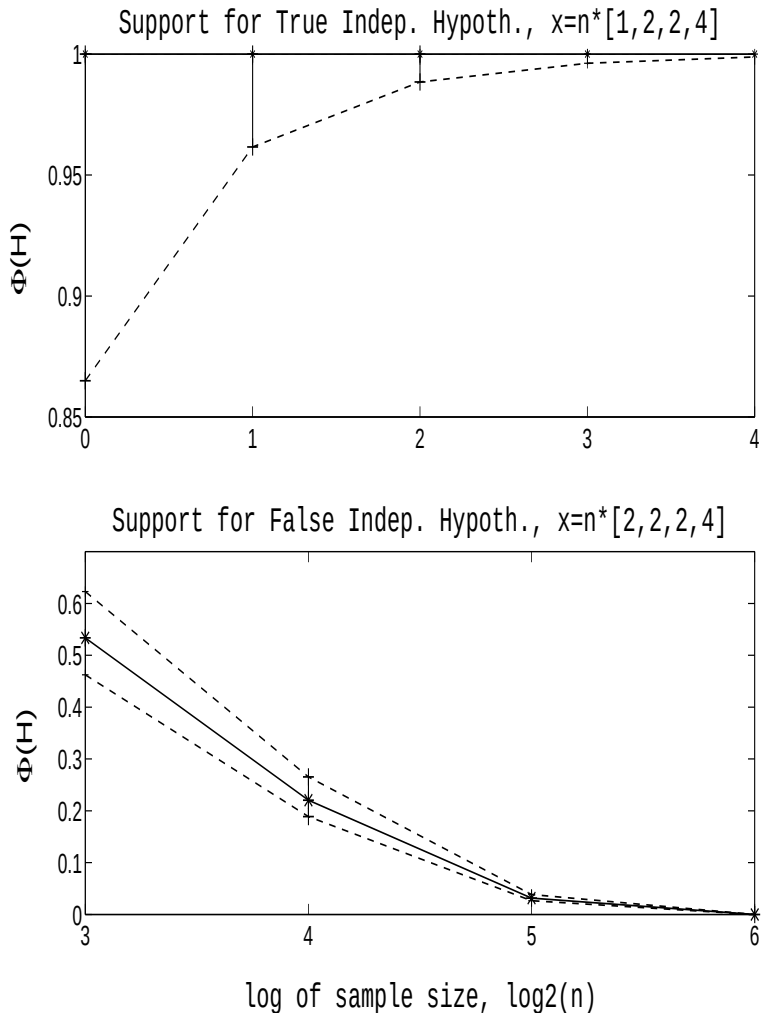
Some of the alternative ways of measuring the uncertainty of the value of evidence Ev(H), such as the *empirical power analysis*, have a dual possibilistic / probabilistic interpretation.

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 uninformative 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.

Bilattices:

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

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

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:

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

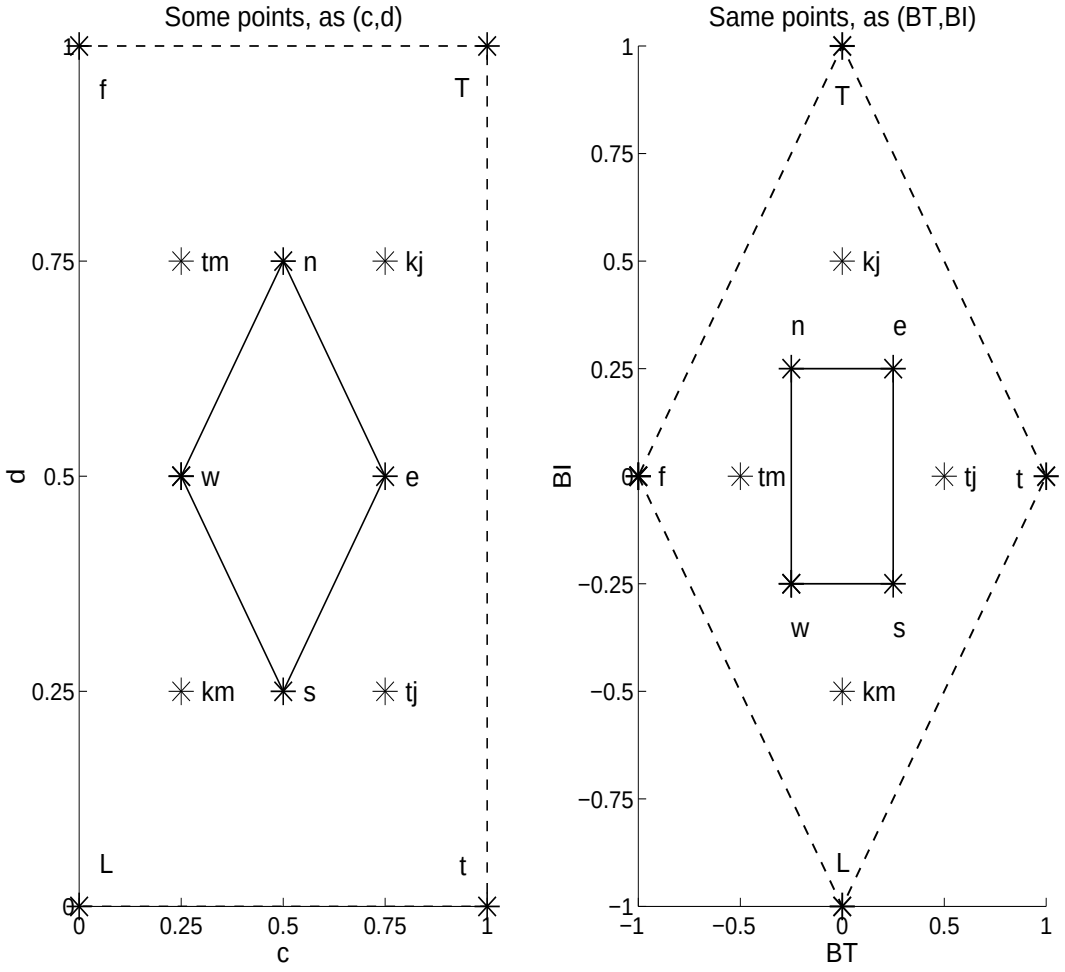
Negation, $\neg$, and Conflation, $-$
properties, (if defined):
Ng1: $x \leq_k y \Rightarrow \neg x \leq_k \neg y$,
Ng2: $x \leq_t y \Rightarrow \neg y \leq_t \neg x$,
Cf1: $x \leq_k y \Rightarrow -y \leq_k -x$,
Cf2: $x \leq_t y \Rightarrow -x \leq_t -y$,
Ng3: $\neg\neg x = x$, Cf3: $--x = x$.
Ng: reverses 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., $\bot$-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:

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



(c, d) and (BT, BI) coordinates

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, 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!

Constructivist Epistemology and Ontology:
von Foerster (2001, 2003):

"Objects are tokens for eigen-behaviors. 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."

"The meaning of recursion is to run through one's own path again. One of its results is that under certain conditions there exist indeed solutions which, when reentered into the formalism, produce again the same solution. These are called "eigen-values", "eigen- functions", "eigen-behaviors", etc., depending on which domain this formation is applied - in the domain of numbers, in functions, in behaviors, etc."

“ Out of an infinite continuum of possibilities, recursive operations carve out a precise set of discrete solutions. Eigen-behavior generates discrete, identifiable entities. Producing discreteness out of infinite variety has incredibly important consequences. It permits us to begin naming things. Language is the possibility of carving out of an infinite number of possible experiences those experiences which allow stable interactions...”

“Eigenvalues have been found ontologically to be discrete (lower-dimensional, precise, sharp or singular), stable (limit or fixed point), 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.”

Sharp objects => Identifiable entities =>
can be Named => Language (composition)

Theorems of Noether (physics), Haar (continuous Groups), de Finetti (statistics), etc.

- NTs and HT provide invariant physical quantities (conserv.laws) and invariant measures (integrals) from symmetry transformation groups. These become sharp hypotheses by excellence.
- dFTs provide invariant distributions from symmetry groups of the statistical model, generating prototypical sharp hypotheses in application areas, see Eaton (1989), Feller (1968), Nachbin (1965) and Deitmar (2002).

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."

Decoupling, Randomization, Sparsity, and Objective Inference.

General strategy:
Coupled equations in $x_1, \ldots, x_n$.
define new variables $y_1, \ldots, y_n$, s.t.
evolution of y_i depends only of y_i
(and not on y_j , $j \neq i$).
Solve $y_i(t)$, $i = 1 \ldots n$,
convert to old coordinates, $x_i(t)$.

Discrete chord system's dynamics:

$$\ddot{x} + Kx = 0 \ , \quad w_0^2 = \frac{h}{ms} \ ,$$

$$K = w_0^2 \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} .$$

Decoupling operator:
Orthogonal matrix Q, diagonalizing K, i.e.

$$Q^{-1} = Q' \ , \text{ and } \ Q'KQ = D = \text{diag}(d) \ .$$

$$Q'(Q\ddot{y}) + Q'K(Qy) = I\ddot{y} + Dy = 0 \ ,$$

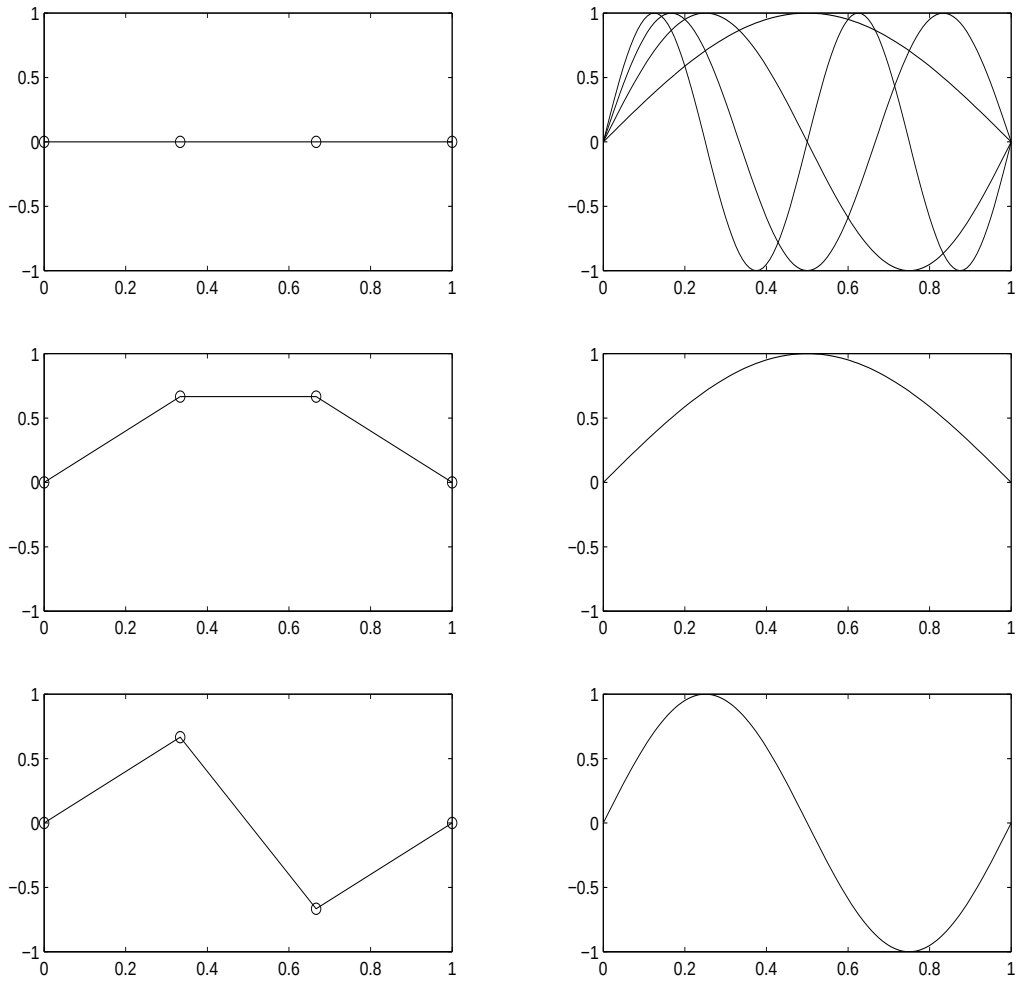
equivalent to the n decoupled scalar equations for harmonic oscillators, $\ddot{y}_k + d_k y_k = 0$.

Solution in normal coordinates:

$$y_k(t) = \sin(\varphi_k + w_k t) \ ,$$

Columns of Q are the eigenvectors of matrix K, multiples of the un-normalized vectors z^k. Their corresponding eigenvalues, $d_k = w_k^2$, for $j, k = 1 \ldots n$,

$$z_j^k = \sin\left(\frac{jk\pi}{n+1}\right) \ , \ w_k = 2w_0 \sin\left(\frac{k\pi}{2(n+1)}\right) \ .$$



Eigen-Solutions, Discrete and Continuous Chord.

Given a (vector) random variable, x, its expected (mean) vector, β, and covariance matrix, V, are defined as:

$$\beta = E(x) \ , \ \ V = \text{Cov}(x) = E((x-\beta)\otimes(x-\beta)') \ .$$

Since the expectation operator is linear,

$$E(Ax + b) = AE(x) + b \ \text{ and}$$

$$\text{Cov}(Ax + b) = A\text{Cov}(x)A' \ .$$

For numerical and structural model estimation we write $V(\gamma) = \sum \gamma_t G^t$, where G^t is a basis for the space of symmetric matrices of dimension $n \times n$, see Lauretto (2002), For example, for dimension $n = 4$,

$$V(\gamma) = \sum_{t=1}^{10} \gamma_t \, G^t \ = \ \begin{bmatrix} \gamma_1 & \gamma_5 & \gamma_7 & \gamma_8 \\ \gamma_5 & \gamma_2 & \gamma_9 & \gamma_{10} \\ \gamma_7 & \gamma_9 & \gamma_3 & \gamma_6 \\ \gamma_8 & \gamma_{10} & \gamma_6 & \gamma_4 \end{bmatrix} \ .$$

How can we decouple the estimated model?

A possible decoupling operator is the lower triangular Cholesky factor, $L \mid V = LL'$.

Let us consider $y = L^{-1}x$, or $x = Ly$.

In the new variables the model is decoupled, i.e., has uncorrelated random components,

$$\text{Cov}(y) = L^{-1}VL^{-t} = L^{-1}LL'L^{-t} = I \ .$$

Let us consider a simple example:

$$V = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 8 \end{bmatrix} \ , \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 2 \end{bmatrix} \ .$$

This example has two peculiarities:

- The matrix V is *sparse*, has several zero elements, and also *structured*, the zeros are arranged in nice a (block) pattern.

- This is also an example of *perfect* factorization (or elimination), i.e., no position with a zero in V is *filled* with a non-zero in L.

Perfect eliminations are rare, however, there are several techniques that can be used to obtain sparse (and structured) Cholesky factorizations in which the fill in is minimized, that is, the sparsity of the Cholesky factor is maximized. In practice, large models can only be computed with the help of these techniques.

Bayesian Networks, rely on sparse factorization techniques that, from an abstract graph theoretical perspective, are almost identical to sparse Cholesky factorization, see for example Lauritzen (2006), Stern (2006a, sec.9-11) and Colla (2007).

Simpson's Paradox and the
Control of Confounding Variables:

Table 1: Simpson's Paradox.

Sex	T	R	NR	Tot	R%
All	T	20	20	40	50%
All	NT	16	24	40	40%
Male	T	18	12	30	60%
Male	NT	7	3	10	70%
Fem	T	2	8	10	20%
Fem	NT	9	21	30	30%

Simpson's Paradox (Lindley's example):
The association between two variables, Treatment and Recovery, is reversed if the data is aggregated / disaggregated over a *confounding* variable, Sex.

How can we design a statistical experiment in order to avoid (separate, decouple) spurious associations?

1) Control possible confounding variables, imposing some form of invariance, constancy or equality.

2) Measure possible confounding variables and include the relevant ones in the statistical model.

Keeping everything under control in a statistical experiment (or in life in general) constitutes, in the words of Fisher:

"a totally impossible requirement in our example, and equally in all other forms of experimentation."

Solution: Box et al. (1978, p.102-103):

"Control what you can,
and randomize what you can not."

Pearl (2000, p. 340,348. Epilogue:
The Art and Science of Cause and Effect):

*"...Fisher's 'randomized experiment'...
consists of two parts, 'randomization' and intervention'."*

"Intervention means that we change the natural behavior of the individual: we separate subjects into two groups, called treatment and control, and we convince the subjects to obey the experimental policy. We assign treatment to some patients who, under normal circumstances, will not seek treatment, and give placebo to patients who otherwise would receive treatment. That, in our new vocabulary, means 'surgery' - we are severing one functional link and replacing it with another. Fisher's great insight was that connecting the new link to a random coin flip 'guarantees' that the link we wish to break is actually broken. The reason is that a random coin is assumed to be unaffected by anything we can measure on macroscopic level..."

"Statistics is Prediction";
but is that all there is?
Cognitive Constructivism: No!

Abelson (1995, p.xiii): *"The purpose of statistics is to organize a useful argument from quantitative evidence, using a form of principled rhetoric."*

Einstein (1950): *"There exists a passion for comprehension, just as there exists a passion for music."*

"I believe that every true theorist is a kind of tamed metaphysicist... The metaphysicist believes that the logically simple is also the real. The tamed metaphysicist believes that not all that is logically simple is embodied in experienced reality, but that the totality of all sensory experience can be 'comprehended' on the basis of a conceptual system built on premises of great simplicity."

Are scientific hypotheses or theories supposed to be static (dogmaic) ?

No. Science is an Evolving System!

Evolution of complex systems =>

Stable modular structures =>

Quantization =>

Objective probability!

(at least in Bohr complementarity theory)

- J.Alcantara, C.V.Damasio, L.M.Pereira (2002). Paraconsistent Logic Programs. JELIA-02. *LNCS,* 2424, 345-356.

- O.Arieli, A.Avron (1996). Reasoning with Logical Bilattices. *J.Logic, Language & Information,* 5,25-63.

- D.Basu, J.K.Ghosh (1988). Statistical Information and Likelihood. *Lecture Notes in Statistics*,45.

- Z.W.Birnbaum, J.D.Esary, S.C.Saunders (1961). Multicomponent Systems and Structures Reliability. *Technometrics,* 3,55-77.

- W.Borges, J.M.Stern (2005). On the Truth Value of Complex Hypothesis.*CIMCA-05,MaxEnt-06,SMPS-06.*

- N.C.A.Costa (1963). Calculs Propositionnels pour les Systemes Formales Incosistants. Compte Rendu Acad. des Scienes, 257, 3790–3792.

- N.C.A.Costa, J.M.Abe, A.C.Murolo, J.I.da Silva, C.F.S. Casemiro (1999). *Lógica Paraconsistente Aplicada.* Atlas.

- A.Y.Darwiche, M.L.Ginsberg (1992). A Symbolic Generalization of Probability Theory. AAAI-92.

- A.Deitmar (2002). *A First Course in Harmonic Analysis.* Springer.

- J.S.Dugdale (1996). *Entropy and Its Physical Meaning.* Taylor & Francis.

- M.L.Eaton (1989). *Group Invariance Applications in Statistics.* Hayward: IMA.

- W.Feller (1971). *An Introduction to Probability Theory and Its Applications.* Wiley.

- M.Fitting (1989). Bilattices and the Theory of Truth. *J. Philosophical Logic*, 18, 225–256.

- R.H.Gaskins (1992). *Burdens of Proof in Modern Discourse.* Yale Univ. Press.

- J.B.Kadane, R.L.Winkler (1987). De Finetti's Methods of Elicitation. In: Viertl (1987). *Probability and Bayesian Statistics.* Plenum.

- J.N.Kapur (1989). *Maximum Entropy Models in Science and Engineering.* Wiley.

- G.J.Klir, T.A.Folger (1988). *Fuzzy Sets, Uncertainty and Information.* 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.Probability and Statistics,*17,147-168.

- M.Lauretto, J.M.Stern (2005). FBST for Mixture Model Selection. MaxEnt-05, *American Institute of Physics Conference Proceedings,* 803, 121–128.

- M.Madruga, L.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.

- L.Nachbin (1965). *The Haar integral.* Van Nostrand.

- C.A.B.Pereira, J.M.Stern (1999). Evidence and Credibility: Full Bayesian Significance Test Precise Hypotheses. *Entropy*,1,69–80.

- C.A.B.Pereira, J.M.Stern, (2001). Model Selection: Full Bayesian Approach. *Environmetrics*, 12, 559–568.

- C.A.B.Pereira, S.Wechsler, J.M.Stern (2005). Can a Significance Test be Genuinely Bayesian? Submitted.

- L.S.Pontriaguin (1982). *Grupos Continuos.* MIR.

- 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 Intelligence 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. FIS2005, 61, 1–23.

- J.M.Stern, S.Zacks (2002). Testing Independence of Poisson Variates under the Holgate Bivariate Distribution. The Power of a New Evidence Test. *Statistical and Probability Letters*, 60, 313–320.

- R.C.Williamson (1989) *Probabilistic Arithmetic.* Ph.D. Thesis, Queensland University.

- L.Wittgenstein (1921). *Tractatus Logico Philosophicus* (Logisch-Philosophische Abhandlung). Dover.