Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses

Carlos Alberto de Bragança Pereira
Julio Michael Stern

Entropy Journal 1999, I, 104-115
www.mdpi.org/entropy/

Bayesian Evidence Test for
Precise Hypotheses

R.Madruga, C.A.B.Prereira, J.M.Stern

Journal of Statistcal Planning and Inference
www.elsevier.com

Density Function $\quad f(x, \theta)$

Data $x \in S$, the Sample Space,
Parameter Space $\Theta = \{\theta \in R^k \,|\, g(\theta) \leq 0\}$

Given (after we have observed)
the data x, sample size n,

Likelihood Function $L(\theta \,|\, x)$
$c = \int_{\Theta} L(\theta \,|\, x) \, d\theta$

1- Likelihood, Plausibility, Favorable Chances.
2- Wahrscheinlichkeit , Verossimilhança,
Possibility it is Truth, Give Proof, Veracity.

1- Should we consider $(1/c) \, L(\theta \,|\, x)$
as a probability function of θ ?
Bayesians (yes) X Frequentists (no)

2- Posterior Function,
given a (Subjective) Prior $g(\theta)$,
$f(\theta \,|\, x) \propto L(\theta \,|\, x) g(\theta)$

Estimation:
What is (my, the) best guess for θ?

1- $\theta^* \in$ arg max $f(\theta \,|\, x)$

2- $\theta^* \in$ arg $\min_{\theta' \in \Theta} \int_\Theta f(\theta' \,|\, x)\, p(\theta', \theta)\, d\theta$

$p(\theta', \theta)$ Loss, penalty or distance function.

ex: $p(\theta', \theta) = ||\theta' - \theta||_p$
$\theta^* =$ Median, for $p = 1$,
$\theta^* =$ Mean, for $p = 2$,
$\theta^* =$ Mode, for $p = \infty$

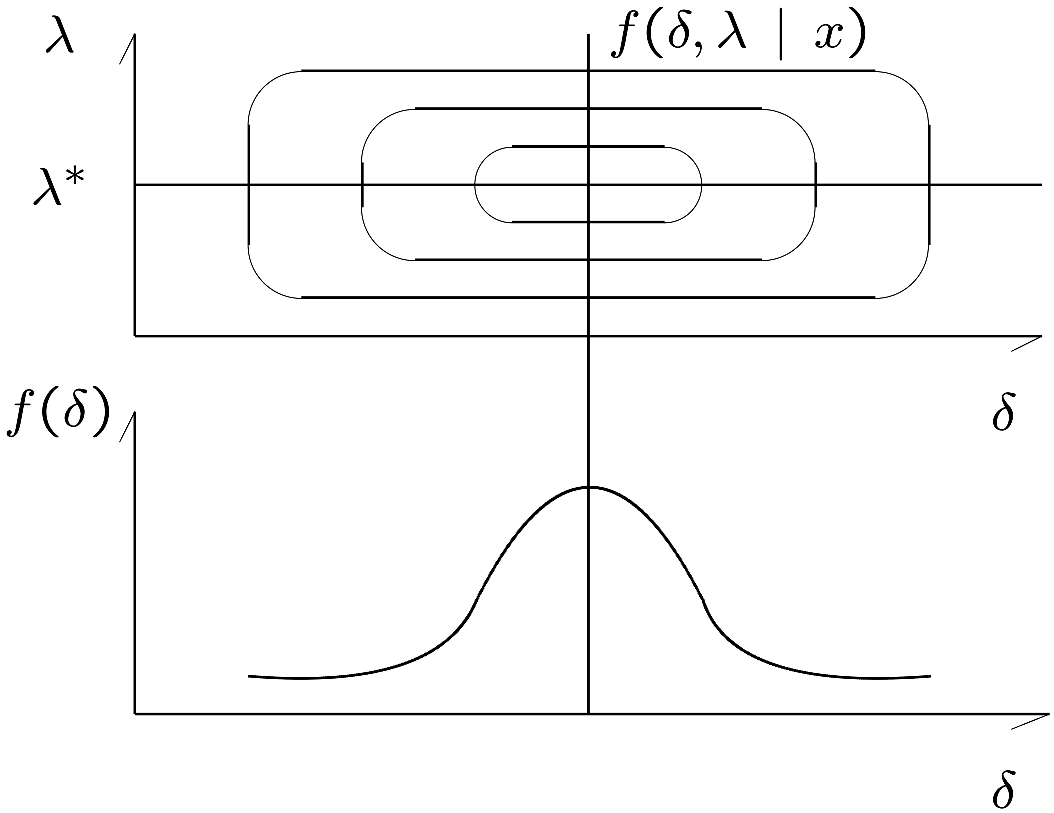
Operators: $\max_\Theta$, $\int_\Theta \; d\,\theta$

Parameter Space $\Theta = \Delta \times \Lambda$ best δ^* ?

λ are "nuisance" parameters
$\equiv$ we do not care about $\lambda^* \in \Lambda$

nonsense (meaningless, foolish) $<$ F. non sens
$<$ L. sentire (to feel), G. sinn (meaning)
annoyance $<$ anoye $<$ O.F. enoyer , SP. enojar
$<$ V.L. in odio (to hate)
nuisance (trouble) $<$ O.F. nuire
$<$ V.L. nocere (to harm), V.L. necare (to kill)

Eliminate λ
(get rid of the killer paramameters)
How? Std. Ans: Projection Π



Projection Operator, Π:

1- Profile Posterior:
$f(\delta \,|\, x) = f(\delta, \lambda^* \,|\, x)$

2- Marginal Posterior:
$f(\delta \,|\, x) = \int f(\delta, \lambda \,|\, x)\, d\lambda$

Hypothesis Testing:

$\Theta = \Theta_0 \; + \; \Theta_1$

Null, $H_0 : \theta \in \Theta_0$ Alternative, $H_1 : \theta \in \Theta_1$

$$LR = \quad \frac{\max_{\Theta_0} f(\theta \,|\, x)}{\max_{\Theta_1} f(\theta \,|\, x)} \quad = \quad \frac{f(\theta_0^* \,|\, x)}{f(\theta_1^* \,|\, x)}$$

Likelihood Ratio, Not a Probability

Neyman-Pearson Theorem $\Theta = \{\theta_0\} \cup \{\theta_1\}$

$LR \leq k$ criterion minimizes convex combination of errors: 1-reject $H_0 \,|\, \theta_0$, 2-accept $H_0 \,|\, \theta_1$

P-Value, Extreme events probability:

$T = \{x \in S \,|\, LR(x) \leq LR\}$, Tail points

$\alpha = Pr(T \,|\, H_0) \equiv Pr(T \,|\, \tilde{\theta})$, Nominal Level

$\tilde{\theta} = \text{arg}\ \max_{\Theta_0} Pr(T \,|\, \theta)$

1- α is a probability in the sample space

2- $\forall \theta$, $n \to \infty \;\; \Rightarrow \;\; \alpha \to 0$

"Increase sample size to reject"

If H_0 Posterior Probability

$\alpha = (1/c) \int_{\Theta_0} f(\theta \,|\, x)\, d\theta > 0$

1- Is a probability in Θ

2- If $\theta \in \dot{\Theta}_0$, $n \to \infty \;\Rightarrow\; \alpha \to 1$

3- If $\theta \in \dot{\Theta}_1$, $n \to \infty \;\Rightarrow\; \alpha \to 0$

Odds Ratio

$$OR(H_0\,,\,x) = \frac{\int_{\Theta_0} f(\theta \,|\, x)\, d\theta \;\; /c}{\int_{\Theta_1} f(\theta \,|\, x)\, d\theta \;\; /c}$$

$O(p) = p/(1-p)$

Precise Hypothesis:

$\Theta_0 = \{\theta \in \Theta \mid \; h(\theta) = 0\}$

$\dim(\Theta_0) < \dim(\Theta) \;\Leftrightarrow \dim(h) \geq 1$

we may use $\delta = h(\theta)$

$\int_{\Theta_0} 1 d\theta = 0 \Rightarrow OR = 0 \;,\; \forall x$!

Change measure over Θ_0,

$$OR = \quad \frac{\pi}{1-\pi} \quad \frac{\int_{\Theta_0} f(\delta{=}0,\lambda\,,x)g(\delta,\lambda)\, d\lambda}{\int_{\Theta_1} f(\theta\,,x)g(\theta)\, d\theta}$$

1- which mass π ?

2- which measure $d\lambda$ in Θ_0?

natural = easy parameterization of Θ_0 ?

3- $\forall\theta$, $\pi > 0$, $n \to \infty \;\Rightarrow\;$ OR $\to$ 1

"Increase sample size to accept"

Savage's allegory:
King Hiero suspects his goldsmith cheated, using impure gold for his new crown. Arquimedes must measure the crown density, $N(m, s)$, test $m = g$, and advise the king about the hanging.

Epistemological questions:

1- Do precise (sharp) hypothesis really exist?
Instrumental (model) versus
Essential (real) theory.
Galileo's "Epur si muove"

2- How can we "check" a precise hypothesis?

I.J.Good's positions on precise hypothesis:
0- We Do want to "Check" H
1- Denial of real sharp tests,
2- Favoring Jeffrey's scheme,
3- Reduction to dichotomies.

0- "Since I regard refutation and corroboration as both valid criteria for this demarcation it is convenient to use another term, Checkability, to embrace both processes. I regard checkability as a measure to which a theory is scientific, where checking is to be taken in both its positive and negative senses, confirming and disconfirming."

1- "Let us consider a null hypothesis that is a (sharp) simple statistical hypothesis H. This is often done is statistical practice, although it would usually be more realistic to lump in with H a small neighborhood of close hypotheses. ...

Of course if H is redefined to be a composite hypothesis by including within H a small neighborhood of the sharp null hypothesis, then it becomes possible to obtain much more evidence in favor of H, even without assuming a prior over the components of H.
Similarly, if by the truth of Newtonian mechanics we mean that it is approximately true in some appropriate well defined sense we could obtain strong evidence that it is true; but if we mean by its truth that it is exactly true then it has already been refuted.
Very often the statistician doesn't bother to make it quite clear whether his null hypothesis is intended to be sharp or only approximately sharp. ... It is hardly surprising then that many Fisherians (and Popperians) say that - you can't get (much) evidence in favor of the null hypothesis but can only refute it."

2- “My own view on induction is close to that of Jeffreys (1939) in that I think that the initial probability is positive for every self-consistent scientific theory with consequences verifiable in a probabilistic sense. No contradiction can be inferred from this assumption since the number of statable theories is at most countably infinite (enumerable).”

3- “It is very difficult to decide on numerical values for the probabilities, but it is not quite so difficult to judge the ratio of the subjective initial probabilities of two theories by comparing their complexities. This is one reason why the history of science is scientificaly important.”

Text book history is a post-mortem analysis.

When do we have a Precise Hypothesis?

A Scientific Theory, ex:

1- Classical Mechanics, errors in measurements

2- Statistical Physics, too many variables

3- Old (Einstein) Quantum Mechanics,
Unobservable (complex) wave function

4- Modern (Heisemberg) Quantum Mechanics,
No hidden states, von Neuman theorem

The Physical law is exact (precise) in Θ
Measurements in S are inexact, incomplete, unattainable, or essentially uncertain.

Software Controlled Randomized Process

Randomization may be:
- External, exogenous, or

- Internally induced, endogenous
1- Efficient Algorithms
   Simulated Annealing, Genetic, Tabu Search
2- Desirable unpredictability

Ex. of internal randomization applications:
1- Cryptography
2- Automatic Transactions,
   Exchange, Clearing, etc
3- Games and Lotteries

Precise hypothesis may be:
1- Software specifications
2- Functional specifications
3- Normative or Legal Requirements

Test $h(\theta) = 0$

No access to source code
Confidentiality, Outsourcing, Privacy, Security

Data from Black Box Simulation
or Historical Observations

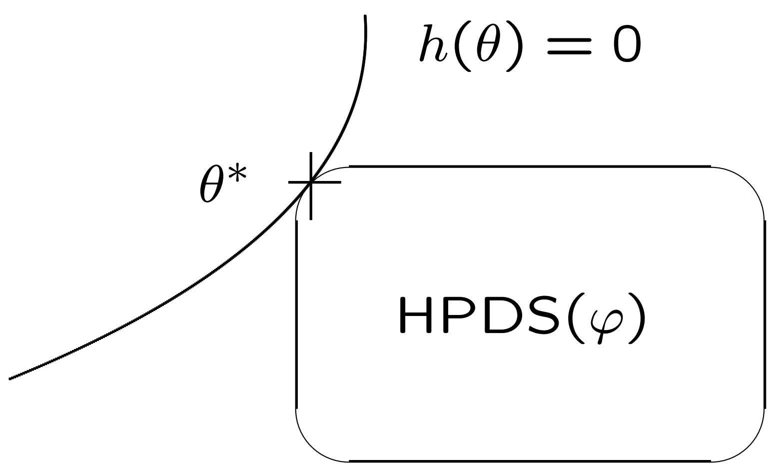
Benefit of the Doubt, Onus Probandi,
Presumption of Innocence, In Dubito Pro Reo
Safe Harbor Liability Rule:

This kind of principle establishes that there is no liability as long as there is a reasonable basis for belief, effectively placing the burden of proof on the plaintiff, who, in a lawsuit, must prove false a defendant's misstatement.

Such a rule also prevents the plaintiff of making any assumption not explicitly stated by the defendant, or tacitly implied by existing law or regulation.

Must consider the plaintif's claim, H_0,
in the most favorable way $\Rightarrow$
$\theta^* \in \arg\max_{\Theta_0} f(\theta \,|\, x)$, $\varphi = f(\theta^* \,|\, x)$
Can not integrate on Θ_0,
(it is a classic, not a quantic plaintif)

Can only consider as evidence againts H_0
the more likelly alternatives
$\text{HPDS}(\varphi) = \{\theta \in \Theta \,|\, f(\theta \,|\, x) \geq \varphi\}$
Tangent Highest Probability Density Set



FBST Operational Definition

$\Theta = \{\theta \in R^n \,|\, g(\theta) \leq 0\}$

$\Theta_0 = \{\theta \in \Theta \,|\, h(\theta) = 0\}$

1- Numerical Optimization step:

$\theta^* \in \arg\max_{\theta \in \Theta_0} f(\theta)$,

2- Numerical Integration step:

$Ev(H) = \int_\Theta f^*(\theta \,|\, d) d\theta$, where

$f^*(\theta) = 1(\theta \in T^*) f(\theta)$,

$T^* = \{\theta \in \Theta \,|\, f(\theta) \geq f(\theta^*)\}$

T^* HPDS is Tangent to H (manifold)
$Ev(H)$ Against H , $\overline{Ev}(H) = 1 - Ev(H)$

FBST properties:

Clear and Simple definition

Ev(H) is a Probability in Θ ,
nothing but Θ , the whole Θ

"Increase sample size
to get it right, accept / reject"

Testing $\sim$ "Checking" H,
as every scientist always wanted

... but also Compatible with the
Decision Theory framework, as shown by
Madruga, Esteves and Wechsler

Invariant on Hypothesis parameterization

Computationally Intensive:
Let Optimization, Num. Analysis
people earn a living :-)

FBST Explicitly Invariant Formulation

$\Theta = \{\theta \in R^n \,|\, g(\theta) \leq 0\}$

$\Theta_0 = \{\theta \in \Theta \,|\, h(\theta) = 0\}$

$r(\theta)$ a "Reference Density"

1- Numerical Optimization step:

$\theta^* \in \arg\max_{\theta \in \Theta_0} \frac{f(\theta)}{r(\theta)}$,

2- Numerical Integration step:

$Ev(H) = \int_{\Theta} f^*(\theta \,|\, d) d\theta$, where

$f^*(\theta) = 1(\theta \in T^*) f(\theta)$,

$T^* = \left\{ \theta \in \Theta \,|\, \frac{f(\theta)}{r(\theta)} \geq \frac{f(\theta^*)}{r(\theta^*)} \right\}$

$\dfrac{f(\theta)}{r(\theta)}$ is the Relative Surprise

If we change the parameter space a density changes by the Jacobian's determinant, and the ratio is unchanged

If $\theta' = \Phi(\theta)$, $f(\theta') = f(\Phi^{-1}(\theta'))\,\mathrm{det}(J)$

$$J = \frac{\partial\,\theta}{\partial\,\theta'} = \begin{bmatrix} \frac{\partial\,\theta_1}{\partial\,\theta'_1} & \cdots & \frac{\partial\,\theta_1}{\partial\,\theta'_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\,\theta_n}{\partial\,\theta'_1} & \cdots & \frac{\partial\,\theta_n}{\partial\,\theta'_n} \end{bmatrix}$$

If $r(\theta)$ is uniform (possibly improper) $Ev(H)$ is unchanged, we are just mapping T^* to the new coordinates

Jeffreys: The natural parameter space is the one where the non-informative prior is uniform in $]-\infty, +\infty[$

In this framework $r(\theta)$ is Jeffreys' non-informative prior

Ex. for $\sigma \in [0, \infty[$, $\phi = \log(\ )$
$\sigma' = \phi(\sigma)$ take us to $]-\infty, +\infty[$
$d\sigma' = d\sigma/\sigma$, and $r(\sigma) = 1/\sigma$
is the non-informative prior in $[0, \infty[$

$\phi = \log(\rho = \sigma^k) \Rightarrow r(\rho) = 1/\rho$

In $]-\infty, +\infty[$ the FBST Evidence and the Uniform Measure are both Invariant by Linear Transformations

Warning-1: $\theta' = \Phi(\theta)$ back to paradise
may be ambiguous, artificial and awkward

Transformations of $[0,1] \to \,]-\infty, +\infty[$
1, 2, 3, 4, ...

Correlations "are" Angles so why look
for a uniform in R^n and not in S^n ?

Warning-2: Invariance has different meanings

In Geometry (and therefore Physics)
we should talk about Invariance
by the Action of a Group

Noether Theorem: For every continuous symmetry of the laws of physics, there must exist a conservation law. For every conservation law, there must exist a continuous symmetry.