Statistics and Innovation - Technological and Theoretical

Julio Michael Stern*
IME-USP, The Institute of Mathematics and Statistics
of the University of São Paulo, * *jstern@ime.usp.br*



ABE - SINAPE-2014, July 20-25, Natal.

http://www.ime.usp.br/∼jstern/miscellanea/jmsslide/sinape14.pdf

This Presentation

- I - This Presentation and Presenter
- II - Some Technological Projects in(?) Statistics
  - with IME-USP++ Bayesian Group*
- III - A Theoretical Project in(?) Statistics
  - with IME-USP++ Bayesian Group*
- IV - Innovation - Some Characteristics and Advice
- V - Getting it Done - Analysis: Divide and Conquer
- VI - Getting it Used - Synthesis: Unite and Conquer
- VII - Some References
- VIII - Frequently Asked Questions

IME-USP++ Bayesian Group = Turma do Carlinhos

This Presenter

- 1981, 1983 - Bachelor, M.Sci., Mathematical Physics (Semi-Riemannian Geometry, General Relativity), IF-USP
- 1989, 1991 - M.Eng., Ph.D., Operations Research (Algorithms for Sparse Matrices / Optimization), Cornell University
- 2001 - Liv.Doc. (Priv.Doz.), Computer Science, IME-USP
- 2010 - Full Prof. (Titular), Applied Mathematics, IME-USP
- 2003-2014 - CNPq Research Fellow (Pesquisa Operacional)

– Operations Research = Optimization (Linear & Non-Linear Programming) + Stochastic proc. + Statistics + Computer Sci.

- at IME-USP*, working for:

• 2008 - 28th MaxEnt - Bayes. Inference & Max. Entropy Meth.

• 2012 - ISBrA - Brazil. Chapt. Int. Soc. for Bayesian Analysis

*Only Bayesian group to host both?

-H.Jeffreys (1939, 46) Theory of probability; An invariant form for the prior probability in estimation problems.

-E.T.Jaynes (1957, 68) Information th.& statistical mechanics; Prior probabilities.

1st: 1981 Laramie, WY; 1979 València.

Some Technological Projects in(?) Statistics

- with IME-USP++ Bayesian Group

- Media Insertion Mean-Variance Optimizer
- Client: IPSOS, São Paulo and France (ARF-2005)
- Innovation: Integrated reach-penetration (hit/individual) analysis and optimization for advertising campaigns.

- Algorithmic Analysis of Paternity Tests.
- Client: Genomic, São Paulo (Genet. & Molec. Biol. 09)
- Innovations: Relax Statistical hypotheses concerning:
(1) Mutations, (2) Independence (no consanguinity),
(3) Homogeneity (Hardy-Weinberg population equilibrium);
(4) Development of new algorithms for Bayesian Networks.

- Hierarchical Forecasting with Polynomial Networks
- Client: Editora Abril, São Paulo (KES-2009)
- Innovation: Integration of Time Series econometrics with tools of AI (artificial intelligence) for qualitative (subjective) factors.

Some Technological Projects in(?) Statistics

- Token-ring Clearing Heuristic for Currency Circulation.
- Client: FinanTech / Politec / Banco do Brasil / CIP-SPB Câmara Interbancária, Sistema de Pagamentos Brasileiro INPI: 00042036; XI EBEB, AIP Proc. 1490, 179-188 (2012);
- Innovation: Real time Interbank Payments System (2001); + Alternative Currency Payments System - TORC3 (2012)

- E.C.Colla, J.M.Stern (2008). Sparse Factorization Methods for Inference in Bayesian Networks. (MaxEnt-08, KES-09).
- Nice hammer looking for some nails...

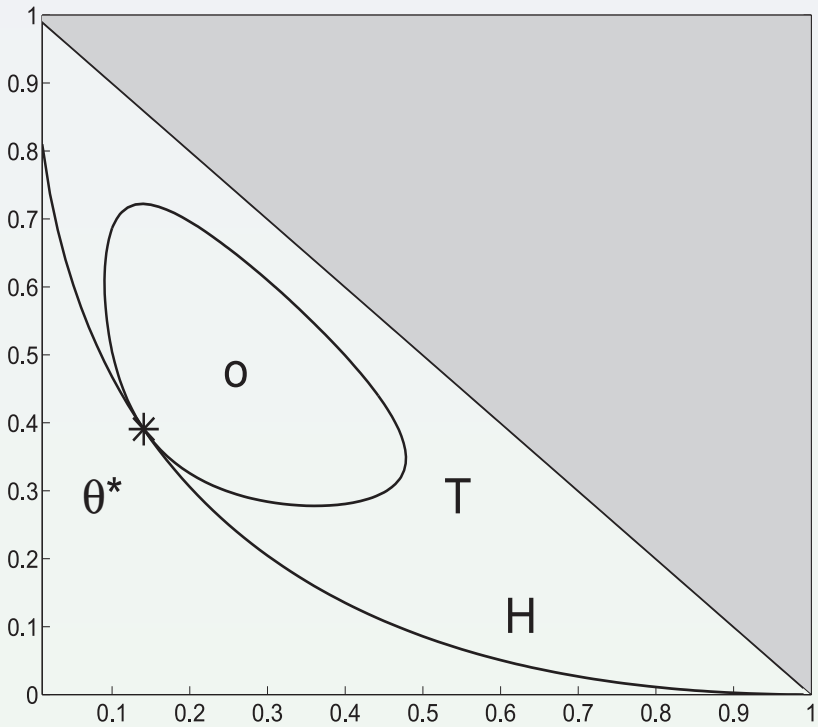
- Several ongoing tecnological projects...

- M.Diniz et al. (2011 / 2012). Unit Roots / Cointegration: Bayesian Significance Test. *Communications in Statistics - Theory and Methods*, 40, 23, 4200-4213. / 41, 19, 3562-3574.
- Innovative solution for a large class of econometric problems.

A Theoretical Project in(?) Statistics - FBST

- $\mathrm{ev}(H \mid X)$ - The *Epistemic Value* of statistical hypothesis H, given the observed data X, or the *Evidence Value* of data X supporting hypothesis H - Full Bayesian Significance Test.
- H is a Sharp or Precise hypothesis:

H states that the model's parameter lies in a zero volume set, $\Theta_H : \theta \in \Theta \mid \mid g(\theta) \leq 0 \wedge h(\theta) = 0$.



Ex - Hardy-Weinberg H: homo/heterozygote freqs.

$\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1$;

$h(\theta) = (1 - \sqrt{\theta_1}\,)^2 - \theta_3$.

$p_n(\theta \mid X) \propto p_0(\theta) L(\theta \mid X)$

$\propto \theta_1^{x_1+y_1} \theta_2^{x_2+y_2} \theta_3^{x_3+y_3}$;

$y_j = 0, \frac{-1}{2}, -1$; Const, Invar, MaxEnt.

Reference Density and Surprise Function

- $r(\theta)$, the reference density, is a representation of no, minimal or vague information about the parameter θ. If $r \propto 1$ then $s(\theta) = p_n(\theta)$ and $\overline{T}$ is a HPDS.
- $r(\theta)$ defines the reference metric in Θ, $dl^2 = d\theta' J(\theta) d\theta$, directly from the Fisher Information Matrix,

$$J(\theta) \equiv -\mathrm{E}_{\mathcal{X}} \frac{\partial^2 \log p(x \mid \theta)}{\partial \theta^2} = \mathrm{E}_{\mathcal{X}} \left(\frac{\partial \log p(x \mid \theta)}{\partial \theta} \frac{\partial \log p(x \mid \theta)}{\partial \theta} \right).$$

- The *surprise function*, $s(\theta) = p_n(\theta)/r(\theta)$, measures changes in the posterior relative to the reference density.
- The 'hat' and 'star' superscripts indicate unconstrained and constrained (to the hypothesis H) maximal arguments and supremal surprise values, as follows:

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

Model's Truth Function and Epistemic Value of H

- The surprise function's *v-cut*, $T(v)$, and its complement, the *highest surprise function set* (HSFS) above level v, $\overline{T}(v)$, are

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

- If the reference density is the uniform (possibly improper) density, $r(\theta) \propto 1$, then $s(\theta) \propto p_n(\theta)$ and the HSFS are standard *highest probability density sets* (HPDS)

- The statistical model's *truth function*, $W(v)$, is the cumulative probability function up to surprise level v, $0 \leq v \leq \widehat{s}$.

$$W(v) = \int_{T(v)} p_n(\theta)\, d\theta\,.$$

- Finally, the *e*-value (epistemic value) of hypothesis H, is

$$\text{ev}(H) = W(v^*)\,.$$

e-values have a Logic! (compositionality rules)

- H in Homogeneous Disjunctive Normal Logical Form:

Independent statistical Models, $j = 1, 2, \ldots,$ each model with stated Hypotheses $H^{(i,j)}$, $i = 1, 2 \ldots,$ defining the Structures $M^{(i,j)} = \{\Theta^j, H^{(i,j)}, p_0^j, p_n^j, r^j\}$.

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

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

- Composition operators: max and $\bigotimes$ (Mellin convolution);
- Invariant Possibilistic Belief Calculus defined over the statistical model's Posterior *Probability* (invariant) Measure;
- Classical logic limit: If all $\mathrm{ev}(H^{i,j}) \simeq 0 \vee 1$, then $\mathrm{ev}(H) \simeq 0 \vee 1$.

e-values deserve an Epistemological Framework

- *p*-value - Ronald Fisher, Jerzy Neyman, Egon Pearson;
- Epistemology: *Falsificationism* - Karl Popper;
- Metaphor: “The Scientific Tribunal”.

- Bayes Factor - Bruno deFinetti, L.J. Savage, I.J. Good;
- Epistem: *Utility Th.* - John v.Neumann, Oskar Morgenstern;
- Metaphor: “The Scientific Casino”, “Betting odds”.

- *e*-values - Carlos A.B. Pereira, J.M. Stern, S. Wechsler...
- Epistem: *Cognitive Constructivism* - Humberto Maturana, Francisco Varela, Heinz von Foerster; +*Invariance* - Felix Klein, A. Einstein, Emmy Noether, Eugene Wigner, Hermann Weyl;
- Metaphors: “Objects are Tokens for Eigen-Solutions”;
+ “Objectivity means invariance by a group of automorphisms”.
- Essential propert.: Precise, Stable, Separable, Composable.

Innovation - Some Characteristics and Advice

- Fire uphill, Water downhill and True Innovation, nobody can stop, contain, fence, lock inside a box...

- Innovation likes to "jump fences", inspiring / stimulating trans-disciplinarity, systemic thinking, holistic approaches.
- Who invented: Fourier and Wavelet analysis? Kalman filter? MCMC / Particle filters? Rank reducing matrix factorizations?
- ...Simplex? ParTan method? (G. Dantzig, O. Kempthorne)

- Do your work, publish it and, most importantly, ...enjoy it!
- Publish! Fast! ...where more relevant for you(r) readers.
- Doing so establishes intellectual rights and responsibilities, tracks ownerships and accountabilities of all involved parties.
- Patent, Copy-Right; or Copy-Left, GPLicence it; but do it!
- Avoid outsourced jobs in "assembly line" research programs.
- Disregard narrow-minded scope / publication control policies.
- No *vira-latas** complex!

*Nelson Rodrigues - Brazilian mongrel, stray dog.

Getting it Done - Analysis: Divide and Conquer

– Strategies and Tactics for Software & Technological Projects –

- Make it easy for potential users to actually use your stuff!
- Almost invariably this means: Implement software package(s).

- Plane for a cascade unfolding of manageable tasks,
from High-level prototypes to Low-level production software.
- Critical tasks: Parallel teams, alternative/competing solutions.

- May use (but do not be manipulated by) Integration Agents:
- Volunteer only for the tasks you want (+time & competence);
- Have well defined “borders” (tasks, environ., interfaces, I/O).

- Use only reliable, stable & well -developed / -documented,
programming environments (ANSI C, Octave, R, Python, etc.);
- GPL: *GNU is good for you!* - Even in commercial projects!
- Avoid uncontrollable dependencies, draconian conditions, etc.
(have your development tools’ source code)

Getting it Used - Synthesis: Unite and Conquer

– Finding Interesting (my personal taste) Theoretical Projects –

- New algorithms that make methods more efficient.
- New methods that meet real demands / applications.
- Breakthrough interpretations based on new methods.
- Ex.: Finding new causal links / relational pathways
- Theoretical foundations for new methods.
- Theoretical repercussion of new developments in related areas, ex: Statistics to/from Probability, Numerical analysis, Optimization, Logic, Epistemology, etc.
- All of the above have a tendency to integrate distinct (sub)areas, and disciplines, make new connections, etc.

Some References, see www.ime.usp/∼jstern

- M.S.Lauretto, C.A.B.Pereira, J.M.Stern (2008). 28th MaxEnt, Int. Work. on Bayesian Inference & Maximum Entropy Methods in Science and Engineering. *AIP proc.* 1073.
- J.M.Stern, M.S.Lauretto, A.Polpo, M.A.Diniz (2012). EBEB-2012, XI Brazilian Meeting on Bayesian Statistics. *American Institute of Physics Conference Proceedings*, v.1490.
- P.J.Fernandes, J.M.Stern, M.S.Lauretto (2007). A New Media Optimizer Based on the Mean-Variance Model. ARF-05 / *Pesquisa Operacional*, 27, 427-456.
- M.Lauretto et al. (2009). A Straightforward Multiallelic Significance Test for the Hardy-Weinberg Equilibrium Law. *Genetics and Molecular Biology*, 32, 3, 619-625.
- M.Lauretto, F.Nakano, C.A.B.Pereira, J.M.Stern (2009). Hierarchical Forecasting with Polynomial Nets. *Studies in Computational Intelligence*, 199, 305-315.
- C.A.B.Pereira, J.M.Stern, S.Wechsler (2008). Can a Significance Test Be Genuinely Bayesian? *Bayesian Analysis*, 3, 1, 79-100.
- W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J. of the IGPL*, 15, 5-6, 401-420.
- J.M.Stern, C.A.B.Pereira (2014). Bayesian Epistemic Values: Focus on Surprise, Measure Probability! *Logic J. of the IGPL*, 22, 2, 236-254.
- J.M.Stern (2014) Cognitive-Constructivism, Quine, Dogmas of Empiricism, and Münchhausen's Trilemma. To appear in the proceedings of EBEB-2014.
- F.V.Cerezetti, J.M.Stern (2012). Non-Arbitrage in Financial Markets: A Bayesian Approach for Verification. *AIP Conf. Proc.*, 1490, 87-96.
- C.Humes, M.S.Lauretto, F.Nakano, C.A.B.Pereira, G.Rafare, J.M.Stern (2012). TORC3 Token-Ring Clearing Heuristic for Currency Circulation. *AIP* 1490, 179-188.
- V.Fossaluza, M.S.Lauretto, C.A.B.Pereira, J.M.Stern (2014). Combining Optimization and Randomization Approaches for the Design of Clinical Trials. EBEB-2014.