Comparing Parameters of Two Bivariate Normal Distributions using the Invariant Full Bayesian Significance Test

M. Lauretto, *lauretto@supremum.com*
C.A.B. Pereira, *cpereira@ime.usp.br*
J.M. Stern, *jstern@ime.usp.br*
BIOINFO and IME-USP, University of Sao Paulo, Brazil

December 22, 2002

Abstract

The Full Bayesian Significance Test (FBST) for precise hypotheses is applied, to a Multivariate Normal Structure (MNS) model. In the FBST we compute the evidence against the precise hypothesis. This evidence is the probability of the Highest Relative Surprise Set (HRSS) "tangential" to the sub-manifold (of the parameter space) that defines the null hypothesis. The FBST formulation presented in this paper provides an invariant procedure under general coordinate transformations of the parameter space, provided a reference density has been established.

The MNS model we present appears when testing equivalence conditions for genetic expression, using micro-array technology. FBST departs from major statistical paradigms, like nuisance parameters elimination. We discuss some of the statistical and epistemological consequences of this departure.

Key Words. Credibility, Epistemology, Evidence, Full Bayesian Significance Test, Invariant procedure, Nuisance parameters elimination, Numerical integration and optimization, Onus Probandi, Reference prior, Relative surprise, Structural models for multivariate normals.

1 Introduction

In this paper we present the dose-equivalence hypothesis. The dose-equivalence hypothesis, H, asserts a proportional response of a pair of expression levels to two different stimuli. The hypothesis also asserts proportional standard deviations, and equivalent correlations for each response pair. The proportionality coefficient, δ, is interpreted as the second stimulus dose equivalent to one unit of the first. This can be seen as a simultaneous generalization of the linear mean structure, the linear covariance structure, and the Behrens-Fisher problems. The application of the dose-equivalence model is similar to the much simpler bio-equivalence model used in pharmacology, and closely related by several other classic covariance structure models used in biology, psychology, and social sciences, as described in Jiang and Sarkar (2000) and McDonald (1974, 1975). We are not aware of any alternative test for the dose-equivalence hypothesis published in the literature.

The Full Bayesian Significance Test (FBST) is presented in Pereira and Stern (1999b) as a coherent Bayesian significance test. The FBST is intuitive and has a geometric characterization. It can be easily implemented using modern numerical optimization and integration techniques. The method is "Fully" Bayesian and consists in the analysis of credible sets. By Fully we mean that we need only the knowledge of the parameter space represented by its posterior distribution. The FBST needs no additional assumption, like a positive probability for the precise hypothesis. The FBST regards likelihoods as the proper means for representing statistical information, a principle stated by Basu (1988), Birnbaum (1972), Finetti (1974,1981,1992,1993), Good (1983), Kempthorne (1980), Pereira and Lindley (1987), Royall(1997), and others, to simplify and unify statistical analysis. Besides the information in the data, the FBST can also incorporate the scientist's prior knowledge. Another important aspect of the FBST is its consistency with the "Onus Probandi" juridical principle, Gaskins (1992), Kokott (1998), Stern(2001).

The FBST formulation presented in this paper provides an invariant procedure under general coordinate transformations of the parameter space, provided a reference density (usually the uniform or the uninformative prior) has been established. The definition of the invariant procedure is based on the concept of relative surprise, Good (1983), Evans (1997).

In the application presented in this paper, as well as in those in Irony et al. (2002), Pereira and Stern (1999a,b 2001a,b) or Stern and Zacks (2002), it is desirable or necessary to use a test with the following characteristics:

1- Be formulated directly in the original (natural) parameter space.

2- Take into account the full geometry of the null hypothesis as a manifold (surface) imbedded in the whole parameter space.

3- Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the (manifold representing the) null hypothesis being tested.

4- Be an invariant procedure under general bijective and smooth transformations of the parameter space coordinate system.

5- Be consistent with the Onus Probandi (burden of proof) juridical principle, i.e. consider in the “most favorable way” the claim stated by the hypothesis.

6- Consider only the observed sample, needing no artifice that could lead to judicial contention, like a positive prior probability distribution on the precise hypothesis.

7- Be consistent, in the sense that increasing sample size should make the evidence converge to the right 0/1 indicator for the hypothesis being tested (accept/reject decision).

8- Give an intuitive and simple measure of significance for the (null) hypothesis, ideally, a probability in the parameter space.

FBST has all these theoretical characteristics and can be efficiently implemented with the appropriate computational tools. Moreover, the FBST is also in perfect harmony with Bayesian decision theory of Rubin (1987), in the sense that there are specific loss functions which render the FBST. Although we do can cast the FBST in a decision theoretic framework, it was originally defined in a pure operational form, based only on the Onus Probandi juridical principle, Pereira and Stern (1999b). Compliance with this juridical principle, also known as Benefit of the Doubt, Presumption of Innocence or (in accounting) Safe Harbor Liability Rule, was imperative in some of our consulting projects, Pereira and Stern (1999a).

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 (Onus Probandi) on the plaintiff, who, in a lawsuit, must prove false a defendant’s misstatement, without making any assumption not explicitly stated by the defendant, or tacitly implied by existing law or regulation.

Interesting connections of some of the characteristics stated above, with ethics, epistemology, law, psychology and statistics can be found in extensive references given at the authors previous articles.

2 FBST: Invariant Procedure Definition

We restrict the parameter space, Θ, to be always a subset of R^n, and the hypothesis is defined as a further restricted subset $\Theta_0 \subset \Theta \subseteq R^n$. Usually, Θ_0 is defined by vector valued inequality and equality constraints:

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

We are interested in precise hypotheses, so we have at least one equality constraint, hence $dim(\Theta_0) < dim(\Theta)$. $f(\theta)$ is the posterior probability density function.

The computation of the evidence measure against the null hypothesis, $Ev(H)$, used on the FBST is performed in two steps, a numerical optimization step, and a numerical integration step. In order to provide an explicitly invariant formulation for the evidence, under general non-degenerate smooth transformations of the parameter space coordinate system, we use an extra factor, $r(\theta)$, a reference density.

The FBST procedure is defined by:

- Numerical Optimization step:

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

- Numerical Integration step:

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

$$f^*(\theta) = 1(\theta \in T^*)\, f(\theta) \quad , \quad T^* = \left\{ \theta \in \Theta \,|\, \frac{f(\theta)}{r(\theta)} \geq \frac{f(\theta^*)}{r(\theta^*)} \right\}$$

The coefficient $f(\theta)/r(\theta)$ is called the *relative surprise*, and its use is the key to obtain an invariant procedure. This concept was used by Good (1983), Evans (1997) and others.

The "tangential" set T^* is the Highest Relative Surprise Set (HRSS) having the points of the parameter space with higher surprise, relative to the reference density, than any point on the hypothesis. When the reference density is the (possibly improper) uniform, $r(\theta) = 1$, then T^* is the Highest Density Probability Set (HDPS) tangential to H.

If the probability of the set T^* is "large", it means that the null set is in a region of low probability, and the data gives strong evidence, $Ev(H)$ large, against the null hypothesis. On the other hand, if the probability of T^* is "small", then the null hypothesis is in a region of high probability and the data does not provide strong evidence against the hypothesis. $\overline{Ev}(H) = 1 - Ev(H)$ may be interpreted as the evidence in the data in favor of the hypothesis.

After any bijective C^1 (continuos and differentiable) transformation of the parameter space coordinate system, $f(\ )$ and $r(\ )$ are multiplied by the determinant of Jacobian of the transformation, leaving the ratio unchanged. So we are just mapping to the new coordinates the tangential set T^* computed in the original coordinates.

The strict interpretation of the Onus Probandi principle is to take the reference density as the (possibly improper) uniform density, $r(\theta) = 1$. We can generalize the procedure using other reference densities. For example, we may use as reference density the uninformative prior (also known as neutral or reference prior), if one is available. This possibility is suggested by the paper of Evans (1997), in conjunction with Jeffreys' rules to obtain uninformative priors, Zellner (1971, appendix to chapter 2).

One of Jeffreys' rules to obtain an uninformative prior is to define a transformation $\theta' = \Phi(\theta)$ of the parameter space so that in the new coordinate system the uniform uninformative prior in R^n is "natural". According to this perspective, using the uninformative prior as reference density is equivalent to specify a transformation Φ of the parameter space, so that, in the transformed parameter space, the reference density (or uninformative prior) is uniform. We also observe that, in R^n, the uniform measure and the FBST are both invariant under non-degenerate linear transformations, Klein (1997), Santalo (1976).

Finally let us remark that it is possible to use a reference density other than the uniform or uninformative priors. By doing so however, one may impair the adherence of the FBST to the Onus Probandi principle, or change its interpretation. In the following application, for simplicity, we use as reference the (improper) uniform density.

3 Normal-Wishart Distribution

The conjugate family of priors for multivariate normal distributions is the Normal-Wishart family of distributions, DeGroot (1970). Taking as prior distribution for the precision matrix R the wishart distribution with $a > k-1$ degrees of freedom and precision matrix $\dot{S}$ and, given R, taking as prior for β a multivariante normal with mean $\dot{\beta}$ and precision $\dot{n}R$, i.e.

$$
\begin{aligned}
f(\beta, R) &= f(R)\, f(\beta \,|\, R) \\
f(R) &\propto |R|^{(a-k-1)/2} \exp(-\frac{1}{2}\mathrm{tr}(R\,\dot{S})) \\
f(\beta \,|\, R) &\propto |R|^{1/2} \exp(-\frac{\dot{n}}{2}(\beta - \dot{\beta})' R (\beta - \dot{\beta}))
\end{aligned}
$$

The posterior distribution for the parameters β and R has the form:

$$
\begin{aligned}
f(\beta, R \,|\, n, \bar{x}, S) &= f(R \,|\, n, \bar{x}, S)\, f(\beta \,|\, R, n, \bar{x}, S) \\
f(R \,|\, n, \bar{x}, S) &\propto |R|^{(a+n-k-1)/2} \exp(-\frac{1}{2}\mathrm{tr}(R\,\ddot{S})) \\
f(\beta \,|\, R, n, \bar{x}, S) &\propto |R|^{1/2} \exp(-\frac{\ddot{n}}{2}(\beta - \ddot{\beta})' R (\beta - \ddot{\beta})) \\
\ddot{\beta} &= (n\bar{x} + \dot{n}\dot{\beta})/\ddot{n} \quad , \quad \ddot{n} = n + \dot{n} \\
\ddot{S} &= S + \dot{S} + \frac{n\dot{n}}{n+\dot{n}}(\dot{\beta} - \bar{x})(\dot{\beta} - \bar{x})'
\end{aligned}
$$

Hence, the posterior distribution for R is a Wishart distribution with $a + n$ degrees of freedom and precision $\ddot{S}$, and the conditional distribution for β, given R, is Normal with mean $\ddot{\beta}$ and precision $\ddot{n}R$. All covariance and precision matrices are supposed to be positive definite, $n > k$, $a > k-1$, and $\dot{n} > 0$.

Non-informative improper priors are given by $\dot{n} = 0$, $\dot{\beta} = 0$, $a = 0$, $\dot{S} = 0$, i.e. we take a Wishart with 0 degrees of freedom as prior for R, and a constant prior for β, DeGroot (1970), Zellner (1971). We use these priors for simplicity. Then, the posterior for R is a Wishart with n degrees of freedom and precision S, and the posterior for β, given R, is k-Normal with mean $\bar{x}$ and precision nR.

We can now write the simplified log-posterior kernels:

$$
fl(\beta, R \,|\, n, \bar{x}, S) = fl(R \,|\, n, \bar{x}, S) + fl(\beta \,|\, R, n, \bar{x}, S)
$$

$$\begin{aligned} fl(R \,|\, n, \bar{x}, S) = flr &= \frac{a+n-k-1}{2}\log(|R|) - \frac{1}{2}\mathrm{tr}(R\,\ddot{S}) \\ fl(\beta \,|\, R, n, \bar{x}, S) = flb &= \frac{1}{2}\log(|R|) - \frac{\ddot{n}}{2}(\beta-\ddot{\beta})' R (\beta-\ddot{\beta}) \end{aligned}$$

For the surprise kernel, relative to the uninformative prior, we only have to replace the factor $(a+n-k-1)/2$ by $(a+n)/2$.

4 Multivariate Normal Structure Models

We will use the following notation: A family of matrices numbered $1, 2 \ldots$ is written $M^1, M^2 \ldots$. The i-th row, the j-th column, and the i,j-th element of matrix M^t are, respectively, $M^t_{i,\bullet}$, $M^t_{\bullet,j}$ and $M^t_{i,j}$. The Hadamard or pointwise product, $\odot$, is defined by $M = A \odot B \Leftrightarrow M_{i,j} = A_{i,j}\, B_{i,j}$. The Frobenius norm of a matrix is defined by $\mathrm{frob2}(M) = \sum_{i,j=1}^{n} (M_{i,j})^2$.

As it is usual in the covariance structure literature, we will write $V(\gamma) = \sum \gamma_t G^t$, where the matrices G^t form a basis for the space of $n \times n$ symmetric matrices; in our case, $n = 4$.

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

We recall some matrix derivative identities, Anderson (1969), McDonald (1973), Rogers (1980):

$$\frac{\partial V}{\partial \gamma_t} = G^t \ , \qquad \frac{\partial R}{\partial \gamma_t} = -R\, G^t\, R \ ,$$

$$\frac{\partial\, \beta' M \beta}{\partial \beta} = 2\, M\, \beta \ , \qquad \frac{\partial\, \log(|V|)}{\partial \gamma_t} = \mathrm{tr}(R\, G^t) \ ,$$

$$\frac{\partial\ \mathrm{frob2}(V - M)}{\partial \gamma_t} = 2 \sum_{i,j=1}^{n} (V - M) \odot G^t \ ,$$

We also define the auxiliary matrices: $P^t = R\, G^t$ and $Q^t = P^t\, R$.

The dose-equivalence hypothesis, H, asserts a response, mean of a second bivariate normal, proportional to a reference, first bivariate normal. H also asserts proportional standard deviations, and equivalent correlations for each response pair of measurements. The proportionality coefficient, δ, is interpreted as the dose, calibration or proportionality coefficient.

In order to get simpler expressions for the log-likelihood, the constraints and its gradients, we extend the parameter space including the coefficient δ, and state the dose-equivalence optimization problem on the extended 15-dimentional space, with a 5-dimentional constraint:

$$\Theta = \{\theta = [\gamma', \beta', \delta]' \in R^{10+4+1} \ , \ V(\gamma) > 0\}$$

$$\Theta_0 = \{\theta \in \Theta \,|\, h(\theta) = 0\} \ , \ h(\theta) = \begin{bmatrix} \delta^2\gamma_1 - \gamma_3 \\ \delta^2\gamma_2 - \gamma_4 \\ \delta^2\gamma_5 - \gamma_6 \\ \delta\beta_1 - \beta_3 \\ \delta\beta_2 - \beta_4 \end{bmatrix}$$

5 Numerical Optimization and Integration

At the optimization step we minimize a centralization term minus the log-posterior kernel, on the extended parameter space,

$$\begin{aligned} f(\theta) &= c\,n\,\mathrm{frob2}(V - C) \, - flr \, - flb \\ &= c\,n\,\mathrm{frob2}(V - C) \, - \frac{a+n-k}{2}\log(|R|) \\ &\quad + \frac{1}{2}\mathrm{tr}(R\,\ddot{S}) \, + \frac{\ddot{n}}{2}(\beta - \ddot{\beta})'R(\beta - \ddot{\beta}) \end{aligned}$$

The objective function's gradient, $\partial f / \partial \theta$, is:

$$\begin{aligned} \frac{\partial f}{\partial \gamma_t} &= \frac{a+n-k}{2}\mathrm{tr}(P^t) \, - \frac{1}{2}\mathrm{tr}(Q^t\,\ddot{S}) \\ &\quad - \frac{\ddot{n}}{2}(\beta - \ddot{\beta})'\,Q^t\,(\beta - \ddot{\beta}) + 2c\,n \sum_{i,j=1}^{n} (V - C) \odot G^t \\ \frac{\partial f}{\partial \beta} &= \ddot{n}\,R\,(\beta - \ddot{\beta}) \end{aligned}$$

For the surprise kernel and its gradient, relative to the uninformative prior, we only have to replace the factor $(a+n-k)/2$ by $(a+n+1)/2$.

The Jacobian matrix, $\partial h / \partial \theta$, is:

$$\begin{bmatrix} \delta^2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\delta\gamma_1 \\ 0 & \delta^2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\delta\gamma_2 \\ 0 & 0 & 0 & 0 & \delta^2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2\delta\gamma_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & -1 & 0 & \beta_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta & 0 & -1 & \beta_2 \end{bmatrix}$$

At the optimization step, Variable-Metric Trust-Region algorithms, working with explicit analytical derivatives, proved to be very stable, in contrast with the often unpredictable behavior of some methods found in most statistical software, like Newton-Raphson or “Scoring”, Luenberger (1984), Nocedal (1999), Pflug (1998).

Optimization problems of small dimension, like above, allow us to use dense matrix representation, without significant loss, Golub (1989). Large enough centralization factors, c, times the squared frobenius norm of $(V - C)$, for intermediate approximations of the constrained minimum $C \approx V(\gamma^*)$, make the first points of the optimization sequence remain in the neighborhood of the empirical covariance. As the optimization proceeds, we relax the centralization factor, i.e. make $c \rightarrow 0$, and maximize the pure posterior function. In practice this strategy let us avoid handling explicitly the difficult constraint $V(\gamma) > 0$.

The best approach to the numerical integration step required by the FBST is approximation by Monte Carlo (MC) simulation. Several other details for efficient implementations of the evidence functions, as well as for refinement-estimation iterative procedures needed to estimate quantile and power functions can be found in Jones (1985), Lange (1999), Ökten (1997), Spanier and Oldham (1987), Stern and Zacks (2002), Tezuka (1995), and also in the program documentation, available from the authors, upon request. The final computer implementation makes use of user friendly, interactive and extensible environments, like Matlab, or similar open source software like Scilab and Python, Beazley (2001), Gomez (1999). The computation of the evidence, for a typical data set and 1% precision, takes less than a second on an Pentium microcomputer.

6 Numerical Examples

The FBST proved to be useful for testing calibration of micro array equipment, as well as to compare levels of genetic expression, at BIOINFO, the genetic research task force at University of São Paulo. Table 1 displays the evidence in favor of the hypothesis for two typical experimental mean and covariance statistics, A and B, obtained with $n = 50$ observations. We also test some variations on the sample size to give an idea of the test sensibility.

$$\bar{x}^A = \begin{bmatrix} 0.9909 \\ 0.7631 \\ 1.8485 \\ 1.7373 \end{bmatrix} \qquad \mathrm{cov}^A = \begin{bmatrix} 1.1271 & 0.5075 & 0.4891 & 0.4373 \\ 0.5075 & 1.2392 & 0.5356 & 0.5400 \\ 0.4891 & 0.5356 & 2.3241 & 1.1486 \\ 0.4373 & 0.5400 & 1.1486 & 2.1694 \end{bmatrix}$$

$$\bar{x}^B = \begin{bmatrix} 0.9496 \\ 0.7352 \\ 1.4163 \\ 1.6411 \end{bmatrix} \qquad \mathrm{cov}^B = \begin{bmatrix} 1.3820 & 0.7482 & 0.2157 & -0.0085 \\ 0.7482 & 1.4807 & 0.5087 & 0.3908 \\ 0.2157 & 0.5087 & 1.5811 & 0.5275 \\ -0.0085 & 0.3908 & 0.5275 & 2.3600 \end{bmatrix}$$

Table 1: Evidence in Favor of δ-equivalence Hypothesis

Sample	Sample Size						
	100	75	60	50	40	30	25
A	0.47	0.77	0.90	0.96	0.99	1.00	1.00
B	0.24	0.55	0.76	0.88	0.96	0.99	1.00

We want to estimate the empirical power of the FBST for a given experimental data set. Given the sufficient statistic, $n, \bar{x}, S$, we consider $\hat{\theta}$ and θ^*, the unconstrained and constrained posteriori maxima.

Given the constrained maximum θ^* we simulate $s = 10^4$ independent samples $\bar{x}_i, S_i$ of size n. For each of the s simulated samples generated around θ^*, we then estimate the evidence $\eta_i^* = Ev(n, \bar{x}_i, S_i)$ according to the last section. Finally we establish the rejection level estimating the $(1-\alpha)$ quantile $\lambda = q_{\alpha,n}(\theta^*)$ from the $\eta_i^*,\ i = 1 : s$.

Next we consider the unconstrained maximum, $\hat{\theta}$. We simulate $t = 10^4$ samples around $\hat{\theta}$, and estimate the test empirical power as the fraction of these t samples around $\hat{\theta}$ whose evidence against H, $\hat{\eta}_j,\ j = 1 : t$, is above the rejection level λ at θ^*. The power is $1 - \beta$, where β is the probability of accepting H when H is not valid, the type II error. As in the quantile estimation, a careful estimation-refinement procedure is necessary to obtain the desired accuracy in reasonable computation time.

In the following example we choose α in order to minimize the total error, $\alpha + \beta$. In order to accomplished the determination of this minimum total error we have to, after each estimation-refinement step, re-optimize the level α.

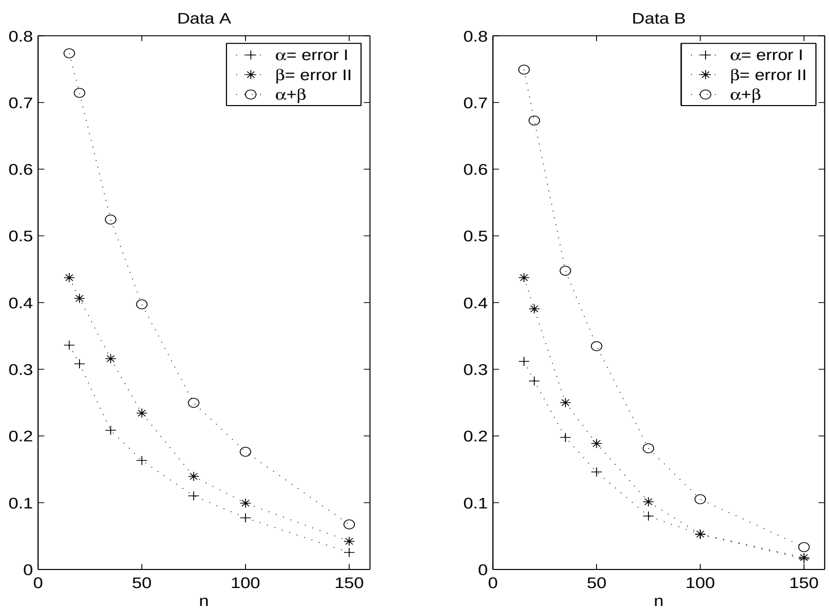

Figure 1: Figure-1 FBST for Minimum Total Error, $\alpha + \beta$

The minimum empirical total error estimate, $\alpha + \beta$, as a function of the sample size, n, for the two experimental data available, are presented in Figure 1, showing interpolated values. As expected, Figure 1 indicates that the power of the test is an increasing function of n. We are not aware of any competing test for this problem, so we can not compare the FBST power with alternative tests.

7 Final Remarks and Acknowledgements

We must stress that the FBST departs from a major statistical paradigm, nuisance parameters elimination: Consider the situation where the hypothesis constraint, $H:\ h(\theta) = h(\delta) = 0\ , \theta = [\delta, \lambda]$ is not a function of some of the parameters, λ.

This situation is described by Basu (1988): "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?"

Basu goes on listing at least 10 categories of procedures to achieve this goal, like using the max_λ or $\int\ d\lambda$ operators, in order to obtain a projected profile or marginal posterior function, $f(\delta \mid x)$.

The very word Nuisance, used by Fisher, reveals how strong is this paradigm in statistical thinking. Compare the etymological derivation of the radical Nuisance, related to (An)noyance, in contrast to Nonsense, far more natural in the context:
Nuisance $<$ O.F. Nuire $<$ V.L. Nocere (to harm), Necare (to kill).
Annoy $<$ O.F. Enoier $\approx$ Sp. Enojar $<$ V.L. in-Odiare (to hate).
Nonsense $<$ F. non Sens (meaningless) $<$ L. Sentire (to feel) $\approx$ Ger. Sinn (meaning).

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, Pereira and Stern (2001a,b).

In order to handle several other structural hypotheses, we only have to replace the constraint, and its Jacobian, passed to the optimizer. Hence, many different hypothesis about the mean and covariance or correlation structure can be treated in a coherent, efficient, exact, robust, simple, and unified way. FBST computer programs for several related structural hypothesis, including also other distributions, are now being implemented, and will be made available to the scientific community as soon as possible.

The authors are grateful to the Prof. Shelemyahu Zacks for several important suggestions. The authors are grateful for the support of the Departments of Computer Science and Statistics of the Institute of Mathematics and Statistics of the University of São Paulo, The last author is also grateful to Department of Mathematical Sciences at the State University of New York at Binghamton; his work was funded in part by grant FAPESP 2001-03484-1.

References

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

Beazley,D.M. (2001). *Python Essential Reference.* Indianapolis: New Riders.

Birnbaum,A. (1972). More on Concepts of Statistical Evidence. *J. Amer. Statist. Assoc.* 67, 858–861.

Box, G.E.P., Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis.*. NY: Wiley.

Evans,M. (1997). Bayesian Inference Procedures Derived via the Concept of Relative Surprise. *Communications in Statistics* 26, 1125–1143.

DeGroot,M.H. (1970). *Optimal Statistical Decisions.* NY: McGraw-Hill.

Finetti,B.de (1974). *Theory of Probability,* V1 and V2. London: Wiley.

Finetti,B.de (1981,1991). *Scritti.* V1: 1926-1930, V2: 1931-1936. Padova: CEDAM

Finetti,B.de (1993). *Probabilitá e Induzione.* Bologna: CLUEB.

Gaskins,R.H. (1992). *Burdens of Proof in Modern Discourse.* New Haven: Yale Univ. Press.

Golub,G.H., van Loan,C.F. (1989). *Matrix Computations.* Baltimore: Johns Hopkins.

Gomez,C. (1999). *Engineering and Scientific Computing with Scilab.* Berlin: Birkhäuser.

Good,I.J. (1983). *Good thinking: The foundations of probability and its applications.* University of Minnesota Press.

Irony,T.Z., Lauretto,M., Pereira,C.A.B., Stern,J.M. (2001). A Weibull Wearout Test: Full Bayesian Approach. In: Hayabawa,Y., Irony,T., and Xie,M., Edts. *Systems and Bayesian Reliability.* pp 287-300. World Scientific, Singapore.

Jiang,G., Sarkar,S. (2000). The Likelihood Ratio Test for Homogeneity of the Variances in a Covariance Matrix with Block Compound Symmetry. *Commun. Statist. Theory Meth.* 29, 1155-1178.

Jones,M.C. (1985). Generating Inverse Wishart Matrices. *Commun. Statist. Simula. Computa.* 14, 511–514.

Kempthorne,O. (1980). Foundations of Statistical Thinking and Reasoning. *Australian CSIRO-DMS Newsletter.* 68, 1–5; 69, 3–7.

Klein,D.A. and Rota,G.C. (1997). *Introduction to Geometric Probability.* Cambridge Univ. Press.

Kokott,J. (1998). *The Burden of Proof in Comparative and International Human Rights Law.* The Hague: Kluwer.

Lange,K. (1999). *Numerical Analysis for Statisticians.* NY: Springer.

Luenberger,D.G. (1984). *Linear and Nonlinear Programming.* Reading: Addison-Wesley.

McDonald,R.P., Swaminathan.H. (1973). A Simple Matrix Calculus with Applications to Multivariate Analysis. *General Systems*, 18, 37–54

McDonald,R.P. (1974). Testing Pattern Hypotheses for Covariance Matrices. *Psychometrica*, 39, 189–201.

McDonald,R.P. (1975). Testing Pattern Hypotheses for Correlation Matrices. *Psychometrica*, 40, 253–255.

Madruga,M.R., Esteves,L.G., Wechsler,S. (2001). On the Bayesianity of Pereira-Stern Tests. *Test* 10, 291-299.

Madruga,M.R., Pereira,C.A.B., Stern,J.M. (2002). Explicitly Invariant Formulation for the Full Bayesian Significance Test. RT-MAE-IME-USP 2002-06, Tech. Rep. Univ. of São Paulo. Submitted.

Nocedal,J., Wright,S.J. (1999). *Numerical Optimization.* NY: Springer.

Ökten,G. (1995). *Contributions to the Theory of Monte Carlo and Quasi Monte Carlo Methods.* Dissertation.com.

Pereira,C.A.B., Lindley,D.V. (1987). Examples Questioning the use of Partial Likelihood. *The Statistician*, 36, 15-20

Pereira,C.A.B., Stern,J.M. (1999a). A Dynamic Software Certification and Verification Procedure. *Proc. ISAS/SCI-99*, 2, 426–435.

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

Pereira,C.A.B., Stern,J.M. (2001a). Full Bayesian Significance Tests for Coefficients of Variation. In: George, E.I. (Editor). Bayesian Methods with Applications to Statistics, 391-400. Monographs of Official Statistics, EUROSTAT.

Pereira,C.A.B., Stern.J.M. (2001b). Model Selection: Full Bayesian Approach. *Environmetrics* 12, (6), 559-568.

Pflug,G.C. (1996). *Optimization of Stochastic Models: The Interface Between Simulation and Optimization.* Boston: Kluwer.

Rogers,G.S. (1980). *Matrix Derivatives.* Lecture Notes in Statistics Vol. 2.

Basel: Marcel Dekker.

Royall,R. (1997). *Statistical Evidence: A Likelihood Paradigm.* London: Chapman & Hall.

Rubin,H. (1987). A Weak System of Axioms for "Rational" Behavior and the Non-Separability of Utility from Prior. *Stat. and Decisions* 5, 47-58.

Santalo,L.A. (1976). *Integral Geometry and Geometric Probability.* Encyclopedia of Mathematics and its Applications, 1. London: Addison-Wesley.

Spanier,J., Oldham,K.B. (1987). *An Atlas of Functions.* NY: Hemisphere Publishing.

Stern,J.S. and Zacks,S. (2002). *Testing the Independence of Poisson Variates under the Holgate Bivariate Distribution The Power of a new Evidence Test.* RT-MAC-IME-USP 2002-03, Tech. Rep. Univ. of São Paulo. Submitted.

Tezuka,S. (1995). *Uniform Random Numbers: Theory and Practice.* Dordrecht: Kluwer.

Zellner,A. (1971). *An Introduction to Bayesian Inference in Econometrics.* NY: Wiley.