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

Marcelo Lauretto
lauretto@supremum.com
Carlos Alberto de Bragança Pereira
cpereira@ime.usp.br
Julio Michael Stern
jstern@ime.usp.br
BIOINFO and IME-USP,
University of Sao Paulo, Brazil

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) "tangent" 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 Multivariate Normal Structure (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.

In the application presented in this paper, as well as in those in Irony et al. (2002), Pereira and Stern (1999a,b 2001) or Stern (2001), 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 juridical principle (or safe harbor liability rule), i.e. consider in the "most favorable way" the claim stated by the hypothesis.

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

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

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

Moreover, as shown in Madruga et al. (2001), 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.

The FBST is based 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.

FBST: Invariant Procedure Definition

Parameter space, $\Theta \subset R^n$,
Hypothesis, $\Theta_0 \subset \Theta$,

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

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:

$$\begin{aligned} Ev(H) &= \int_{\Theta} f^*(\theta \mid d) d\theta \quad , \text{ where} \\ f^*(\theta) &= 1(\theta \in T^*)\, f(\theta) \\ T^* &= \left\{ \theta \in \Theta \mid \frac{f(\theta)}{r(\theta)} \geq \frac{f(\theta^*)}{r(\theta^*)} \right\} \end{aligned}$$

$f(\theta)$, the posterior p.d.f.
$r(\theta)$, the reference p.d.f.
$f(\theta)/r(\theta)$ is the *relative surprise*,
Good (1983) and Evans (1997).
T^* is the Highest Relative Surprise Set,
HRSS, “tangent” to the hypothesis Θ_0
$r(\theta) = 1 \Rightarrow$ HRSS $=$ HPDS

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.

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).

Ex: FBST for testing coefficients of variation, $CV(X) = \sigma(X)/E(X)$,

$$X \sim N(\beta, \sigma) \quad , \qquad H: \ \sigma/\beta = c$$

Using the standard improper priors, uniform on $]-\infty, +\infty[$ for β, and $1/\rho$ on $]0, +\infty[$ for ρ, we get the posterior joint distribution for β and ρ:

$$f(\beta, \rho \,|\, x) \propto \sqrt{\rho}\,\mathrm{exp}(\frac{-n\rho(\beta - \bar{x})^2}{2})\,\mathrm{exp}(-br)\rho^{a-1}$$

$$x = [x_1 \ldots x_n] \ , \ a = \frac{n-1}{2} \ ,$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ , \ b = \frac{n}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

FBST for H: CV=0.1

In Figure 1 we give the FBST evidence, Ev(H), when testing $CV = 0.1$ with a 3 samples of size $n = 16$, mean $\bar{x} = 10$ and standard deviations $s = 1.0$, $s = 1.1$ and $s = 1.5$.

Normal-Wishart Distribution

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 $\dot{n}R$, and given the statistics

$$\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{j=1}^{n} X_{\bullet,j} = \frac{1}{n}X\mathbf{1} \\ W &= (X-\beta)(X-\beta)' \end{aligned}$$

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 k-Normal with mean $\ddot{\beta}$ and precision $\ddot{n}R$.

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

$$
\begin{aligned}
\ddot{\beta} &= (n\bar{x} + \dot{n}\dot{\beta})/\ddot{n} \quad , \quad \ddot{n} = n + \dot{n} \\
\ddot{S} &= S + \dot{S} + \frac{n\dot{n}}{n+\dot{n}}(\dot{\beta} - \bar{x})(\dot{\beta} - \bar{x})'
\end{aligned}
$$

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, Box and Tiao (1973), DeGroot (1970), Zellner (1971).

Multivariate Normal Structure Models

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, $n = 4$.

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

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:

$$\begin{aligned} \Theta &= \{\theta = [\gamma', \beta', \delta]' \in R^{10+4+1} \ , \ V(\gamma) > 0\} \\ \Theta_0 &= \{\theta \in \Theta \,|\, h(\theta) = 0\} \\ h(\theta) &= \begin{bmatrix} \delta^2\gamma_1 - \gamma_3 \\ \delta^2\gamma_2 - \gamma_4 \\ \delta^2\gamma_5 - \gamma_6 \\ \delta\beta_1 - \beta_3 \\ \delta\beta_2 - \beta_4 \end{bmatrix} \end{aligned}$$

FBST for Minimum Total Error, $\alpha + \beta$

The minimum empirical total error, $\alpha + \beta$, as a function of the sample size, n, for the two experimental data available, are presented in Figure 2, showing interpolated values. As expected, Figure 2 indicates that the power of the test is an increasing function of n.

We are not aware of any competing test for this problem.