Brazilian Journal of Probability and Statistics (2003), 17, pp. 147–168.

Full Bayesian significance test applied to multivariate normal structure models

Marcelo Lauretto[1], **Carlos Alberto de Bragança Pereira**[2]
Julio Michael Stern[2] and **Shelemyahu Zacks**[3]
[1] *Supremum Assessoria e Consultoria*
[2] *University of São Paulo*
[3] *State University of New York at Bighamton*

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) tangent to the sub-manifold (of the parameter space) that defines the null hypothesis. The MNS model we present appears when testing equivalence conditions for genetic expression measurements, using micro-array technology.

Key words: Credibility, evidence, full Bayesian significance test, relative surprise, structural models for multivariate normals.

1 Introduction

The Full Bayesian Significance Test (FBST) is presented in Pereira and Stern as a coherent Bayesian significance test, see Pereira and Stern (1999a,b, 2001a,b), Madruga et al. (2003), Stern (2003a,b). The FBST is intuitive and has a geometric characterization. It can be implemented using modern numerical optimization and integration techniques. Like all Bayesian analysis, the FBST regards likelihoods as the proper means for representing statistical information, a principle stated by Barnard (1947, 1949), Basu (1988), Birnbaum (1962, 1972), Finneti (1974, 1981, 1991, 1993), Good (1983), Kempthorne (1976, 1980), Kempthorne and Folks (1971), Royall (1997), and others, to simplify and unify statistical analysis. The method is "Fully" coherent with the Bayesian likelihood principle, meaning that the information gathered from observations is represented by (and only by) the likelihood function. The FBST needs no additional assumption, like a positive prior probability of the precise hypothesis, that leads to Lindley's paradox, Lindley (1957).

In this paper we study the dose-equivalence hypothesis.

The dose-equivalence hypothesis, H, asserts a proportional response of a pair of response measurements 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 test proved to be useful when comparing levels of genetic expression, as well as to calibrate micro array equipment at BIOINFO, the genetic research task force at University of Sao Paulo. 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 Anderson (1969), Bock and Bargnann (1966), Jiang and Sarkar (1998, 1999, 2000a,b), Jöreskog (1970), and McDonald (1962, 1974, 1975). We are not aware of any alternative test for the dose-equivalence hypothesis published in the literature.

In Section 2 we define the FBST, give an outline of its properties, and give a simple example of its application. In Section 3 we remember the basic facts about the Normal-Wishart distribution and its use in Bayesian statistics. In Section 4 we review Multivariate Structure Models. In Section 5 and 6 we describe the numerical optimization and integration algorithms used to implement the FBST. In Section 7 we describe the procedures used to establish the rejection level and the empirical power of the test. We also give the sufficient statistics of two datasets, and use them as case studies in the paper. In Section 8 we give some final remarks.

In this article we use the following matrix notation: The transpose of matrix M is M'. A family of matrices indexed by $h = 1, 2, \ldots$ is written $M^1, M^2 \ldots$. The i-th row, the j-th column, and the i,j-th element of matrix of the h-th matrix, M^h, are, respectively, $M^h_{i,\bullet}$, $M^h_{\bullet,j}$ and $M^h_{i,j}$. The vectors of zeros and ones, with appropriate dimension given by the context, are $\mathbf{0}$ and $\mathbf{1}$. In $(M + v)$, where v is a column (row) vector of compatible dimension, v is added to each column (row) of matrix M. The Hadamard or pointwise product, $\odot$, is defined by $M = A \odot B \Leftrightarrow M_{i,j} = A_{i,j}\, B_{i,j}$. The squared Frobenius norm of a matrix is defined by $\mathrm{frob2}(M) = \sum_{i,j} (M_{i,j})^2$.

2 FBST procedure definition

Let $X_1, \ldots, X_n$ be random variables having a joint density $f(x; \theta)$, with respect to a σ-finite measure μ. θ is a parameter vector in a parameter space $\Theta \subseteq R^p$ ($p \geq 1$). We are interested in a precise null hypothesis $H_0 : \theta \in \Theta_0$, $\Theta_0 \subset \Theta$, and $\dim(\Theta_0) < \dim(\Theta)$. Θ_0 is usually given by (vector) inequality and equality constraints. In the case of a sharp hypothesis, we have at least one equality constraint

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

Let $L(\theta; x)$ denote the likelihood function of θ on Θ. Let $p(\theta)$ be a prior density on Θ, and $r(\theta)$ a reference density on Θ. We denote by $p_n(\theta)$ the posterior density

of θ on Θ, i.e.,

$$p_n(\theta) \propto L(\theta; x)p(\theta) \ , \quad x = [x_1, \ldots, x_n] \ , \ \theta = [\theta_1, \ldots, \theta_p]$$

and define

$$\theta^* = \arg\max_{\theta \in \Theta_0} \left\{ \frac{p_n(\theta)}{r(\theta)} \right\} \ , \quad s_n^* = \max_{\theta \in \Theta_0} \left\{ \frac{p_n(\theta)}{r(\theta)} \right\} = \left\{ \frac{p_n(\theta^*)}{r(\theta^*)} \right\}.$$

The function $s_n(\theta) = p_n(\theta)/r(\theta)$ is called the "relative surprise", Good (1983). We define now, in the space Θ, the Highest Relative Surprise Set Θ_n^* of points $\theta \in \Theta$ with higher relative surprise $s(\theta)$ than any point in Θ_0, i.e.

$$\Theta_n^* = \left\{ \theta \in \Theta \,|\, \frac{p_n(\theta)}{r(\theta)} \geq s_n^* \right\}.$$

Notice that the set Θ_n^* is "tangential" to Θ_0 at θ^*. The evidence against H_0, given by the sample data x, is defined as the posterior probability of the tangential HRSS, i.e.,

$$Ev_n = \int_{\Theta_n^*} p_n(\theta) d\theta.$$

This definition of the evidence against H_0 is invariant with respect to a proper reparameterization. For instance, let $\omega = \phi(\theta)$, where ϕ is a measurable and integrable function. For the purpose of illustration, assume that ϕ is bijective (one-to-one) and continuously differentiable. Let $J(\omega)$ denote the Jacobian of the transformation, i.e.

$$J(\omega) = \left[\frac{\partial}{\partial \omega} \phi^{-1}(\omega) \right] = \left[\frac{\partial \theta}{\partial \omega} \right] = \begin{bmatrix} \frac{\partial \theta_1}{\partial \omega_1} & \cdots & \frac{\partial \theta_1}{\partial \omega_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_n}{\partial \omega_1} & \cdots & \frac{\partial \theta_n}{\partial \omega_n} \end{bmatrix}.$$

The posterior density of ω, given x, is

$$\tilde{p}_n(\omega) = p_n(\phi^{-1}(\omega)) \, |J(\omega)| \, .$$

Notice that the reference density under the reparameterization changes to

$$\tilde{r}(\omega) = r(\phi^{-1}(\omega)) \, |J(\omega)| \, .$$

Thus, the new surprise function is

$$\tilde{s}_n(\omega) = \tilde{p}_n(\omega)/\tilde{r}(\omega) = p_n(\phi^{-1}(\omega))/r(\phi^{-1}(\omega)).$$

Let $\Omega_0 = \phi(\Theta_0)$. It follows that

$$\tilde{s}_n^* = \sup_{\omega \in \Omega_0} \tilde{s}_n(\omega) = \sup_{\theta \in \Theta_0} s_n(\theta) = s_n^*.$$

Accordingly, $\Theta_n^* \mapsto \phi(\Theta_n^*) = \Omega_n^*$.

The evidence under reparameterization is

$$\tilde{Ev}_n = \int_{\Omega_n^*} \tilde{p}_n(\omega) d\omega = \int_{\Theta_n^*} p_n(\theta) d\theta = Ev_n.$$

This proves invariance by proper reparameterizations of Ev_n.

Remarks:

1. The original definition of the evidence against H_0, Pereira and Stern (1999a, b), did not employ the reference density $r(\theta)$. In the former definition, the "tangential" set Θ_n^* was the Highest Probability Density Set, HPDS (whose points have posterior density $p_n(\theta)$ greater than that of any point in Θ_0), instead of the Highest Relative Surprise Set, HRSS.

   The evidence in that former definition is the credibility of the set Θ_n^*. If this evidence is sufficiently high, it is customary to reject H_0. The former definition of evidence is not invariant under reparameterization, as can be shown by various examples.

   Taking the reference density as the (possibly improper) uniform density, $r(\theta) = U(\theta)$, the former and present definitions of evidence define the same tangent set, i.e. the HRSS and the HPDS coincide. In a proper reparameterization $\omega = \phi(\theta)$, using the present definition, we are just automatically mapping to the new coordinates the tangential set computed in the original coordinates, $\Omega_n^* = \phi(\Theta_n^*)$.

2. 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 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 $\omega = \phi(\theta)$ of the parameter space so that, in the new coordinate system, the "natural" uninformative prior is the uniform density. 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 uninformative prior is the uniform. We also observe that, in R^n, the uniform measure and the evidence computed by the former definition of the FBST are both invariant under proper linear transformations, Klein and Rota (1997) and Santalo (1976).

   Jeffreys suggests $\psi = \log(\sigma)$ as a suitable transformation for a parameter $\sigma \in]0, \infty[$. Using $d\psi = d\sigma/\sigma$, and assuming the uniform reference prior on ψ we obtain the reference prior for σ, $r(\sigma) \propto 1/\sigma$. This transformation also has the interesting property of being invariant under transformations of the form $\rho = \sigma^n$, i.e., $r(\rho) \propto 1/\rho$. We will use this prior in following examples.

In order to be consistent with the Onus Probandi juridical principle, see Stern (2003a), we will generally choose as reference density on Θ, the uniform density or a non-informative prior which yields a proper posterior density $p_n(\theta)$. For the examples in this and the next sections, we use the uniform reference density. Using the non-informative prior as reference density would have a minor impact (less than 5%) on all the numerical results presented for the examples analyzed in this paper. This situation is typical, unless the data sets are very small. It is possible to use other reference densities, although doing so may impair the adherence to the Onus Probandi principle, or change its interpretation, see Stern (2003a).

3. We notice that the FBST is used in the full dimensionality of the parameter space. In the way it is defined and to preserve all its properties, elimination of "nuisance" parameters is not recommended.

4. The tangential set Θ_n^* may have more than one connected component, a situation that can occur if the relative surprise has several local (or even global) maxima. This does not change the FBST definition, but may require additional computational effort, as briefly commented in the next remark.

5. The determination of the "tangential" set Θ_n^* might have to be done numerically, since analytic solutions might not be available, see Stern and Zacks (2002). Efficient numerical methods for optimization (finding θ^*) and integration are readily available. The reader is referred to Bertzekas and Tsitsiklis (1989), Censor and Zenios (1997), Evans and Swartz (2000), Iusem (1995). In case of several local maxima, global optimization techniques are required, see Pflug (1996), Spall (2003). The evidence can be estimated by stochastic integration techniques, like Monte Carlo, as shown in Stern and Zacks (2002), Zacks and Stern (2003). The final computer implementation makes use of user friendly, interactive and extensible environment, like Matlab, or open source software like R, Scilab, and Python.

6. As shown by Madruga et al. (2001), one can define a loss function with respect to which, the optimal Bayesian decision is to reject H_0 if Ev_n is greater than a critical level $0 < \lambda < 1$. The existence of such a cost function also ensures that the FBST is a coherent test from a decision theoretic perspective, a concept defined in Finetti (1972, 1974, 1981, 1991), see also Rubin (1987) and Loschi and Wechsler (2002).

2.1 Example: Testing coefficients of variation

As a simple example we present the FBST for testing coefficients of variation. Although this is a very simple example of FBST application, the alternative proposed in Lehmann (1959) gives only an asymptotic confidence interval. The Coefficient of Variation (CV) of a random variable X is defined as the ratio $CV(X) = \sigma(X)/E(X)$, i.e. the ratio of its standard deviation to its mean.

Let X be a normal random variable, with unknown mean and variance. We illustrate the FBST construction in a simple case, namely, testing the hypothesis that its coefficient of variation is equal to a given constant,

$$X \sim N(\beta, \sigma) \quad \text{and} \quad H: \ \sigma/\beta = c \ .$$

It can be shown that the conjugate family for this problem is the family of bivariate distributions, where the conditional distribution of the mean β, for a fixed precision $\rho = 1/\sigma^2$, is normal, and the marginal distribution of the precision ρ is gamma, DeGroot (1970). 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:

$$p_n(\beta, \rho \,|\, x) \propto \sqrt{\rho}\ exp(-n\rho(\beta - \bar{x})^2/2)\ \rho^{\frac{n-2}{2}} exp(-\rho s n/2)$$

$$x = [x_1 \ldots x_n]\ , \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad s = \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

In Figure 1 we plot some level curves of the posterior density function, including the level curve tangent to the hypothesis manifold. At the tangency point, θ^*, the posterior density attains its maximum, p_n^*, on the hypothesis. The interior of the

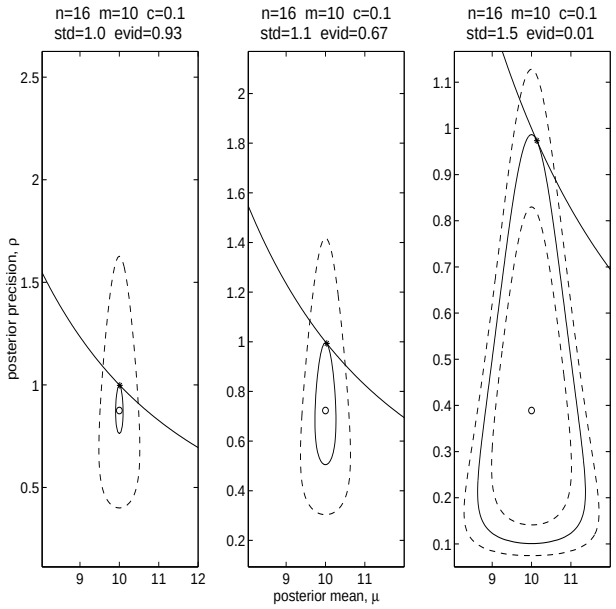

Figure 1 *FBST for H: CV=0.1*

tangent level curve, Θ_n^*, includes all points with posterior density greater than p_n^*, i.e., it is the Highest Density Probability Set tangent to the hypothesis. In Figure 1 we give the FBST evidence, $\mathrm{Ev}(H)$, when testing $CV = 0.1$ with a 3 samples of size $n = 16$, mean $\bar{x} = 10$ and standard deviations $std = 1.0$, $std = 1.1$ and $std = 1.5$. We can see the tangent set expanding as the sample standard deviation over mean ratio gets farther away from the coefficient of variation being tested, $CV(X) = \sigma(X)/E(X) = 0.1$. In this example we use the standard improper prior density and the uniform reference density. In the first plot, the sample standard deviation over mean ratio equals the coefficient of variation tested. Nevertheless, the evidence against the null hypothesis is not zero; this is because of the non uniform prior.

In order to test other hypotheses we only have to change the constraint(s) passed to the optimizer. Constraints for the hypothesis $\beta = c$ and $\sigma = c$ would be represented by, respectively, vertical and horizontal lines. All the details for these and other simple examples, as well as comparisons with standard frequentist and Bayesian tests, can be found in Irony et al. (2001), Pereira and Stern (1999b, 2000a,b) and Pereira and Wechsler (1993).

3 Normal-Wishart distribution

The conjugate family of priors for multivariate normal distributions is the Normal-Wishart family of distributions, DeGroot (1970). Consider the random matrix X with elements $X_{i,j}$, $i = 1 \ldots k$, $j = 1 \ldots n$, $n > k$ where each column, $X_{\bullet,j}$ contains a sample vector from a k-multivariate normal distribution with parameters β (mean vector) and V (covariance matrix), or $R = V^{-1}$ (precision matrix).

Let $\bar{x}$ and W denote, respectively, the statistics:

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

The random matrix W has Wishart distribution with n degrees of freedom and precision matrix R. The Normal and Wishart pdfs have the expressions:

$$\begin{aligned} f(\bar{x} \,|\, n, \beta, R) &= (\frac{n}{2\pi})^{k/2} |R|^{1/2} \exp(-\frac{n}{2}(\bar{x} - \beta)' R (\bar{x} - \beta)), \\ f(W \,|\, n, \beta, R) &= c\,|W|^{(n-k-1)/2} \exp(-\frac{1}{2}\mathrm{tr}(W\,R)), \\ c^{-1} &= |R|^{-n/2}\, 2^{nk/2}\, \pi^{k(k-1)/4} \prod_{j=1}^{k} \Gamma(\frac{n+1-j}{2}). \end{aligned}$$

Now consider the matrix X as above, with unknown mean β and unknown precision matrix R, and the statistic

$$S = \sum_{j=1}^{n} (X_{\bullet,j} - \bar{x})\,(X_{\bullet,j} - \bar{x})' = (X - \bar{x})(X - \bar{x})'.$$

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 multivariate normal with mean $\dot{\beta}$ and precision $\dot{n}R$, i.e.

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

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

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

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

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

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

As it is usual in the covariance structure literature, we will write $V(\gamma) = \sum \gamma_h G^h$, where the matrices G^h, $h = 1, \ldots k(k+1)/2$ form a basis for the space of $k \times k$ symmetric matrices; in our case, $k = 4$. Some matrix notation was presented at the end of Section 1.

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

$$G^h = \begin{bmatrix} \delta_{h,1} & \delta_{h,5} & \delta_{h,7} & \delta_{h,8} \\ \delta_{h,5} & \delta_{h,2} & \delta_{h,9} & \delta_{h,10} \\ \delta_{h,7} & \delta_{h,9} & \delta_{h,3} & \delta_{h,6} \\ \delta_{h,8} & \delta_{h,10} & \delta_{h,6} & \delta_{h,4} \end{bmatrix}$$

and the Kronecker-delta is $\delta_{h,i} = 1$ if $h = i$ and $\delta_{h,i} = 0$ if $h \neq i$.

The dose-equivalence hypothesis, H, asserts a proportional response of a pair of response measurements to two different stimuli. Each pair of response measurements is supposed to be a bivariate normal variate. H also asserts proportional standard deviations, and equivalent correlations for each pair of response 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 use in the numerical procedures an extended parameter space including the coefficient δ, and state the dose-equivalence optimization problem on the extended 15-dimentional space, with a 5-dimentional constraint:

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

In order to be able to compute some gradients needed in the next section, we recall some matrix derivative identities, see Anderson (1969), Harville (1997), McDonald and Swaminathan (1973), Rogers (1980). We use $V = V(\gamma)$, $R = V^{-1}$, and C for a constant matrix.

$$\frac{\partial V}{\partial \gamma_h} = G^h \ , \qquad \frac{\partial R}{\partial \gamma_h} = -R\, G^h\, R \ ,$$

$$\frac{\partial\, \beta' C\, \beta}{\partial\, \beta} = 2\, C\, \beta \ , \qquad \frac{\partial \log(|V|)}{\partial\, \gamma_h} = \mathrm{tr}(R\, G^h) \ ,$$

$$\frac{\partial\ \mathrm{frob2}(V-C)}{\partial\,\gamma_h} = 2\sum_{i,j}(V-C)\odot G^h\ .$$

We also define the auxiliary matrices:

$$P^h = R\,G^h \quad \text{and} \quad Q^h = P^h\,R\ .$$

5 Numerical optimization

To find θ^* we use an objective function, to be minimized on the extended parameter space, given by a centralization term minus the log-posterior kernel,

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

Large enough centralization factors, c, times the squared frobenius norm of $(V-C)$, where C are intermediate approximations of the constrained minimum, make the first points of the optimization sequence remain in the neighborhood of the empirical covariance (the initial C). As the optimization proceeds, we relax the centralization factor, i.e. make $c \to 0$, and maximize the pure posterior function. This is a standard optimization procedure following the regularization strategy of Proximal-Point algorithms, see Bertzekas and Tsitsiklis (1989), Iusem (1995), Censor and Zenios (1997). In practice this strategy let us avoid handling explicitly the difficult constraint $V(\gamma) > 0$.

Using the matrix derivatives given in the last section, we find the objective function's gradient, $\partial\,f/\partial\,\theta$,

$$\begin{aligned} \frac{\partial\,f}{\partial\,\gamma_h} &= \frac{a+n-k}{2}\mathrm{tr}(P^h)\ -\frac{1}{2}\mathrm{tr}(Q^h\,\ddot{S}) \\ &\quad -\frac{\ddot{n}}{2}(\beta-\ddot{\beta})'\,Q^h\,(\beta-\ddot{\beta}) \\ &\quad +2c\,n\ \sum_{i,j=1}^{n}(V-C)\odot G^h, \\ \frac{\partial\,f}{\partial\,\beta} &= \ddot{n}\,R\,(\beta-\ddot{\beta}). \end{aligned}$$

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 of the constraints, $\partial h / \partial \theta$, 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 Proximal-Point algorithms, working with the explicit analytical derivatives given above, 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". Optimization problems of small dimension, like above, allow us to use dense matrix representation without significant loss, Stern (1994).

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.

6 Numerical integration

The best approach to the numerical integration step required by the FBST is approximation by Monte Carlo (MC) simulation, see Evans and Swartz (2000) and Zacks and Stern (2003). We want an estimate of the ratio

$$Ev(n, \bar{x}, S) = \frac{\int_{\Theta_n^*} f(\theta; n, \bar{x}, S) d\theta}{\int_{\Theta} f(\theta; n, \bar{x}, S) d\theta}.$$

Since the space Θ is unbounded, we randomly chose the values of θ according to an "importance sampling" density $g(\theta)$, which is positive on Θ. The evidence function is equivalent to

$$Ev(n, \bar{x}, S) = \frac{\int_{\Theta} Z_g^*(\theta; n, \bar{x}, S) g(\theta) d\theta}{\int_{\Theta} Z_g(\theta; n, \bar{x}, S) g(\theta) d\theta},$$

where

$$\begin{aligned} Z_g(\theta; n, \bar{x}, S) &= \frac{f(\theta; n, \bar{x}, S)}{g(\theta)}, \\ Z_g^*(\theta; n, \bar{x}, S) &= I^*(\theta; n, \bar{x}, S) Z_g(\theta; n, \bar{x}, S), \\ I^*(\theta; n, \bar{x}, S) &= 1(\theta \in \Theta_n^*). \end{aligned}$$

Thus, a Monte Carlo estimate of the evidence is

$$\hat{Ev}_{g,m}(n, \bar{x}, S) = \frac{\sum_{i=1}^{m} Z_g^*(\theta_{i,\bullet}; n, \bar{x}, S)}{\sum_{i=1}^{m} Z_g(\theta_{i,\bullet}; n, \bar{x}, S)},$$

where $\theta_{i,\bullet}, i = 1 \ldots m$ are iid and independently chosen in Θ according to the importance sampling density $g(\theta)$. Thus,

$$\hat{Ev}_{g,m}(n, \bar{x}, S) \stackrel{m\to\infty}{\longrightarrow} Ev(n, \bar{x}, S) \text{ a.s.}[g].$$

The goodness of the MC estimation depends on the choice of g and m. Johnson (1980) describes a simple procedure to generate the cholesky factor of a Wishart variate $W = U'U$ with n degrees of freedom, from the cholesky factorization of the covariance parameter $V = R^{-1} = C'C$:

$$\begin{aligned} L_{i,j} &= N(0,1)\ ,\ i > j, \\ L_{i,i} &= \sqrt{\chi^2(n-i+1)}\ ;\ \ U = L'\,C\ . \end{aligned}$$

At the integration step it is important to perform all matrix computations directly from cholesky factors, Golub and van Loan (1989), Jones (1985). In this problem we can therefore use "exact sampling", what simplifies substatialy the integration step, i.e., $Z_g(\theta; n, \bar{x}, S) = 1$.

6.1 Precision of the MC simulation

In order to control the number of points, m, used at each MC simulation, we need an estimate of MC precision for evidence estimation. For a fixed large value m, the asymptotic distribution of $\hat{Ev}_{g,m}(n, \bar{x}, S)$ is normal with mean $Ev(n, \bar{x}, S)$ and asymptotic variance $V_g(n, \bar{x}, S)$. According to the delta method, Bickel and Doksum (2001), we obtain that

$$V_g(n, \bar{x}, S) = \frac{1}{m}\left(\frac{{\sigma_g^*}^2}{{\mu_g}^2} + \frac{{\sigma_g}^2{\mu_g^*}^2}{{\mu_g}^4} - 2\frac{\mu_g^*}{{\mu_g}^3}\gamma_g\right),$$

where

$$\begin{aligned} \mu_g &= \int_\Theta Z_g(\theta; n, \bar{x}, S) g(\theta) d(\theta), \\ \mu_g^* &= \int_\Theta Z_g^*(\theta; n, \bar{x}, S) g(\theta) d(\theta), \\ {\sigma_g}^2 &= \int_\Theta \left(Z_g(\theta; n, \bar{x}, S) - \mu_g\right)^2 g(\theta) d(\theta), \\ {\sigma_g^*}^2 &= \int_\Theta \left(Z_g^*(\theta; n, \bar{x}, S) - \mu_g^*\right)^2 g(\theta) d(\theta), \\ {\gamma_g}^2 &= \int_\Theta \left(Z_g(\theta; n, \bar{x}, S) - \mu_g\right)\left(Z_g^*(\theta; n, \bar{x}, S) - \mu_g^*\right) g(\theta) d(\theta) \end{aligned}$$

are the expected values, variances and covariance of $Z(\theta; n, \bar{x}, S)$ and $Z^*(\theta; n, \bar{x}, S)$ with respect to $g(\theta)$.

Define the coefficients

$$\xi_g = \frac{\sigma_g}{\mu_g} \text{ and } \xi_g^* = \frac{\sigma_g^*}{\mu_g}.$$

For abbreviation, let $\eta = Ev(n, \bar{x}, S)$. Also note that $\eta = \mu_g^*/\mu_g$. Then the asymptotic variance is

$$V_g(n, \bar{x}, S) = \frac{1}{m}\left(\xi_g^{*2} + \eta^2 {\xi_g}^2 - 2\,\frac{\eta\,\gamma_g}{{\mu_g}^2}\right).$$

Let us define the complementary variables

$$\begin{aligned} Z_g^c(\theta; n, \bar{x}, S) &= I^c(\theta; n, \bar{x}, S) Z_g(\theta; n, \bar{x}, S), \\ I^c(\theta; n, \bar{x}, S) &= 1 - I^*(\theta; n, \bar{x}, S), \\ \sigma_g^{c2} &= V_g\left(Z^c(\theta; n, \bar{x}, S)\right), \\ \xi_g^c &= \frac{\sigma_g^c}{\mu_g}. \end{aligned}$$

Some algebraic manipulation give us $V_g(n, \bar{x}, S)$ in terms of ξ_g^* and ξ_g^c, namely

$$V_g(n, \bar{x}, S) = \frac{1}{m}\left(\xi_g^{*2}(1-\eta)^2 + \xi_g^{c2}\eta^2 + 2\,\eta^2(1-\eta)^2\right).$$

For large values of m, the asymptotic $(1-\beta)$ level confidence level confidence interval for η is $\hat{Ev}_{g,m}(n, \bar{x}, S) \pm \Delta_{g,m,\beta}$, where

$$\Delta_{g,m,\beta}^2 = \frac{F_{1-\beta}(1,m)}{m}\left(\hat{\xi_g^*}^2(1-\hat{\eta})^2 + \hat{\xi_g^c}^2\hat{\eta}^2 + 2\,\hat{\eta}^2(1-\hat{\eta})^2\right),$$

where $F_{1-\beta}(1,m)$ is the $1-\beta$ quantile of the $F(1,m)$ distribution, and $\hat{\eta}$, $\hat{\xi_g^*}$ and $\hat{\xi_g^c}$ are consistent estimators of the respective quantities.

For large m, we can also use the approximation

$$\Delta_{g,m,\beta}^2 = \frac{\chi_{1-\beta}^2(1)}{m}\left(\hat{\xi_g^*}^2(1-\hat{\eta})^2 + \hat{\xi_g^c}^2\hat{\eta}^2 + 2\,\hat{\eta}^2(1-\hat{\eta})^2\right),$$

since $F(1,m)$ converges in distribution to the chi-square distribution with 1 degree of freedom, as $m \to \infty$.

If we wish to have $\Delta_{g,m,\beta} \leq \delta$, for a prescribed value of δ, then m should be such that

$$m \geq \frac{\chi_{1-\beta}^2(m)}{\delta^2}\left(\hat{\xi_g^*}^2(1-\hat{\eta})^2 + \hat{\xi_g^c}^2\hat{\eta}^2 + 2\,\hat{\eta}^2(1-\hat{\eta})^2\right).$$

The computation of the evidence, for a typical data set and 0.005 precision, takes less than a second on an Pentium microcomputer. Several other details for an efficient implementation of the evidence functions, as well as for refinement-estimation iterative procedures needed to estimate quantile and power functions can be found in Zacks and Stern (2003), and also in the program documentation, available from the authors, upon request.

7 Quantiles under the null hypothesis, empirical power and case studies

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, $\alpha + \beta$. (The origin of data-sets A and B is briefly explained at the end of this section). In order to accomplished the determination of this minimum total error we have to, after each estimation-refinement step, re-optimize the level α.

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 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 competing tests for this problem, so we can not compare the FBST power with any alternative.

As case studies for the FBST applied to the Multivariate Normal Structure problem presented in the last sections, we use sufficient statistics $\bar{x}$ and $\text{cov} = (1/n)S$, the experimental mean vector and covariance matrix, for samples A and B, both observations taken with sample size $n = 50$. These statistics come from calibration tests for micro-array measurement equipment, and its interpretation is very similar to classical pharmacological bio-equivalence studies. In this experiments, two dependent response measurements are taken in two different situations, supposed to be equivalent, up to a calibration factor.

In pharmacological bio-equivalence studies one usually measures maximum and total bio-availability (maximum and area under the plasma concentration curve) of a chemical compound delivered by two different formulations, in a crossover design, where the responses of the two formulations become correlated. For a detailed analysis and interpretation see Jiang and Sarkar (1998, 1999, 2000a,b). In the micro array calibration experiments red and green light intensities are measured in two experimental situations, supposed to be equivalent. The analysis of the micro array data is more complex, because the structure in the covariance matrix and in the mean vector as well as the calibration coefficient have to be assessed all at once. Perhaps this simultaneous assessment is also more realistic

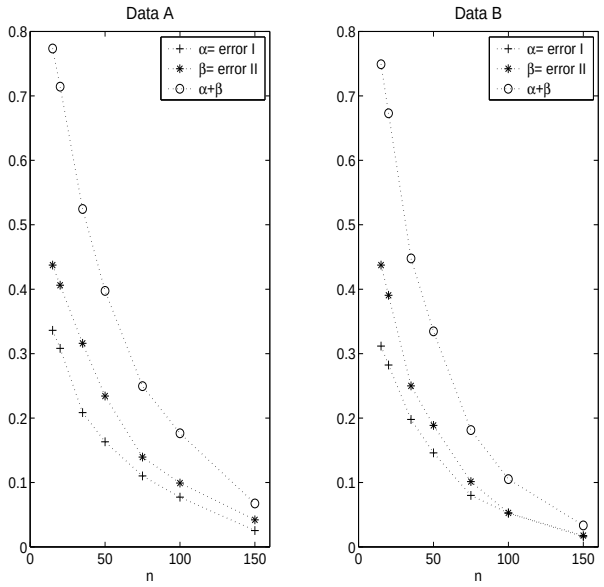

Figure 2 *FBST for Minimum Total Error,* $\alpha + \beta$

in several bio-equivalence experiments.

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

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

We perform the FBST, for samples A and B, where the hypothesis to be tested is described in Section 4. We also perform the test artificially varying the sample size value, n, but always using the same mean and covariance statistics, to give an idea of the test sensibility. Table 1 displays the evidence in favor of the null hypothesis according to the sample size value.

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

8 Final remarks

In the sequel we stress the fact that the FBST departs from two major statistical paradigms:

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, $f(\delta \,|\, x)$. 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 Lindley (1987), Pereira and Stern (2001a,b).

Neyman-Pearson:

The paradigm states hypothesis testing in a decision theoretic framework looking the test as an (optimal) choice between the (null) hypothesis, H_0, and an Alternative, H_1. The original Neyman-Pearson formulation deals with an unitary hypothesis and alternative: $H_0 = \{\theta_0\}$, $H_1 = \{\theta_1\}$. The extension of Neyman-Pearson paradigm, specially to sharp hypothesis, has been a source of endless controversy.

Kempthorne (1980) remarks the underlying semantic confusion:
"*This (Neyman-Pearson) led in 1933 to what I regard as a total alteration of the idea of quantifying evidence or degree of support for a model into a decision process.*"

The dominating asymmetry expressed by the maxima "Increase sample size to reject/accept", reveals the hazards of using this paradigm to test a sharp (zero Lebesgue measure) hypothesis, in either a frequentist or a Bayesian setting.

This situation is highly uncomfortable, inducing positions as extreme as denying the existence in science of real sharp hypothesis, as a justification for some statistical procedures. However this is a very fragile epistemological position, Harlow et al. (1997). Good (1983), confronts this dilemma with great intellectual

integrity. Sometimes he uses the precise hypothesis denial,

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

With the same line of argument Good also gives a justification for Jeffreys' tests:

"*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.*"

It is very interesting that a second argument used by Good to justify Jeffreys' tests, the holy grail of checkability, is naturally achieved by the FBST:

"*Since I regard refutation and corroboration as both valid criteria for this (science) 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.* "

"*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.*"

Finally, Good postulates that a real sharp hypothesis should have a proper alternative, not just everything else:

"*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 scientifically important.*"

This position is far more acceptable than the blunt denial of precise hypothesis, Kuhn (1996). Nevertheless, this position also runs into epistemological difficulties. As noticed by Feyerabend (1993), this position may be fine for "post mortem" historical analyses, but not for active living science, where statistical practice is most needed.

In contrast, to the Neyman-Pearson paradigm, the complement of the evidence computed by the FBST is a measure of significance for the Hypothesis itself, not a ratio against an alternative. Checkability, i.e. convergence to the Boolean indicator for the hypothesis, is just a natural characteristic of the FBST.

As noticed at the end of Section 5, the methodology presented in this article is very general, and several other structural hypotheses can be treated in a coherent, efficient, exact, robust, simple, and unified way. In order to do so we only have to replace the constraint and its Jacobian, passed to the optimizer. FBST computer programs for several related structural hypothesis, including also some other distributions, are now being implemented and will be made available to the scientific community as soon as possible.

Acknowledgements

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, and to Department of Mathematical Sciences at the State University of New York at Binghamton. The third author is also grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo, for his work was funded in part by grant FAPESP 2001-03484-1.

(Received May, 2003. Accepted October, 2003.)

References

Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. In Krishnaiah, P. *Multivariate Analysis II.* New York: Academic Press.

Barnard, G. A. (1947). The meaning of significance level. *Biometrika*, **34**, 179–182.

Barnard, G. A. (1949). Statistical inference. J. Roy. Statist. Soc. B, **11**, 115–149.

Basu, D. (1988). Statistical information and likelihood. Edited by J.K.Ghosh. *Lect. Notes in Statistics*, **45**.

Bertzekas, D. P. and Tsitsiklis, J. N. (1989). *Parallel and distributed computation, numerical methods.* Englewood Cliffs: Prentice Hall.

Bickel, P. J. Doksum, K. A. (2001). *Mathematical Statistics.* London: Prentice-Hall.

Birnbaum, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Assoc.,* **57**, 269–326.

Birnbaum, A. (1972). More on concepts of statistical evidence. *J. Amer. Statist. Assoc.,* **67**, 858–861.

Bock, R. D., Bargnann, R. E. (1966). Analysis of covariance structure. *Psycometrica*, **31**, 507–534.

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

Censor, Y. and Zenios, S. A. (1997). *Parallel Optimization: Theory, Algorithms, and Applications.* New York: Oxford.

Evans, M. (1997). Bayesian inference procedures derived via the concept of relative surprise. *Communications in Statistics*, **26**, 1125–1143.

Evans, M., Swartz, T. (2000). *Approximating Integrals Via Monte Carlo and Deterministic Methods.* London: Oxford Univ. Press.

Feyerabend, P. (1993). *Against Method.* Verso Books.

Finetti, B. de (1972). *Probability, Induction and Statistics.* New York: Wiley.

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

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

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

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

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

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

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

Harlow, L. L., Mulaik, S. A. and Steiger, J. H. (1997) *What If There Were No Significance Tests?* Lawrence Erlbaum Assoc.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* New York: Springer.

Irony, T. Z., Lauretto, M., Pereira, C. A. B. and Stern, J. M. (2001). A Weibull wearout test: Full Bayesian approach. In: Hayabawa, Y. Irony, T. and Xie, M. (Editors). *Systems and Bayesian Reliability,* 287–300. Singapore: World Scientific.

Iusem, A. N. (1995). *Proximal Point Methods in Optimization.* Rio de Janeiro: IMPA.

Jiang, G. and Sarkar, S. (1998). Some asymptotic tests for the equality of covariance matrices of two dependent bivariate normals. *Biometrical Journal*, **40**, 205–225.

Jiang, G., Sarkar, S. and Hsuan, F. (1999). A likelihood ratio test and its modifications for the homogeneity of the covariance matrices of dependent multivariate normals. *J. Stat. Plan. Infer.* , **81**, 95–111.

Jiang, G. and Sarkar, S. (2000a). Some combination tests for the equality of covariance matrices of two dependent bivariate normals. Proc. *ISAS-2000, Information Systems Analysis and Synthesis.*

Jiang, G. and Sarkar, S. (2000b). The likelihood ratio test for homogeneity of the variances in a covariance matrix with block compound symmetry. *Commun. Statist. Theory Meth.*, **29**, 1155–1178.

Johnson, M. E. (1980). *Multivariate Statistical Simulation.* New York: Wiley.

Jones, M. C. (1985). Generating inverse wishart matrices. *Commun. Statist. Simula. Computa.*, **14**, 511–514.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, **57**, 239–251.

Kempthorne, O. and Folks, L. (1971). *Probability, Statistics and Data Analysis.* Ames: Iowa State Univ. Press.

Kempthorne, O. (1976). Of what use are tests of significance and tests of hypothesis. *Comm. Statist. A,* **5**, 763–777.

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.* London: Cambridge Univ. Press.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions.* University of Chicago Press.

Lehmann, E. L. (1959). *Testing Statistical Hypothesis.* New York: Wiley.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Loschi, R. H. and Wechsler, S. (2002). Coherence, Bayes's theorem and posterior distributions. *Brazilian Journal of Probability and Statistics*, **16**, 169–185.

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

Madruga, M. R., Pereira, C. A. B. and Stern, J. M. (2003). Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference*, **117**, 185–198.

McDonald, R. P. (1962). A note on the derivation of the general latent class model. *Psychometrika*, **27**, 203–206.

McDonald, R. P. (1974). Testing pattern hypotheses for covariance matrices. *Psychometrika*, **39**, 189–201.

McDonald, R. P. (1975). Testing pattern hypotheses for correlation matrices. *Psychometrika*, **40**, 253–255.

McDonald, R. P. and Swaminathan. H. (1973). A simple matrix calculus with applications to multivariate analysis. *General Systems*, **18**, 37–54

Pereira, C. A. B. and Lindley, D. V. (1987). Examples questioning the use of partial likelihood. *The Statistician*, **36**, 15–20.

Pereira, C. A. B. and Stern, J. M. (1999a). A dynamic software certification and verification procedure. Proc. *ISAS-99, Information Systems Analysis and Synthesis,* **2**, 426–435.

Pereira, C. A. B. and Stern, J. M. (1999b). Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy*, **1**, 69–80.

Pereira, C. A. B. and 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. and Stern. J. M. (2001b). Model selection: Full Bayesian approach. *Environmetrics*, **12**, 559–568.

Pereira, C. A. B. and Wechsler, S. (1993). On the concept of p-value. *Brazilian Journal of Probability and Statistics*, **7**, 159–177.

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 and 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.

Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization.* New York: Wiley.

Stern, J. M. (1994). *Esparsidade, Estrutura, Estabilidade e Escalonamento em Álgebra Linear Computacional.* Recife: UFPE, IX Escola de Computação.

Stern, J. M. (2003a). Significance tests, belief calculi, and burden of proof in legal and scientific discourse. Laptec-2003, *Frontiers in Artificial Intelligence and its Applications,* **101**, 139–147.

Stern, J. M. (2003b). *Paraconsistent Sensitivity Analysis for Bayesian Significance Tests.* RT-MAC-IME-USP 2003-08. Submitted.

Stern, J. M. and Zacks, S. (2002). Testing the independence of Poisson variates under the Holgate bivariate distribution: The power of a new evidence test. *Statistical and Probability Letters*, **60**, 313–320.

Zacks, S. and Stern (2003). *Sequential Estimation of Ratios, with Application to Bayesian Analysis*. RT-MAC-IME-USP 2003-10. Submitted.

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

Marcelo Lauretto
Supremum Assessoria e Consultoria
Sao Paulo, Brazil
E-mail: lauretto@supremum.com

Carlos Alberto de Bragança Pereira
BIOINFO, NOPEF and IME-USP
University of São Paulo, São Paulo, Brazil
E-mail: cpereira@ime.usp.br

Julio Michael Stern
BIOINFO, NOPEF and IME-USP
University of São Paulo, São Paulo, Brazil
E-mail: jstern@ime.usp.br

Shelemyahu Zacks
State University of New York at Bighamton, USA
E-mail: shelly@math.binghamton.edu