CHAPTER 1

A WEIBULL WEAROUT TEST: FULL BAYESIAN APPROACH

Telba Zalkind Irony Marcelo Lauretto
Carlos Alberto de Bragança Pereira Julio Michael Stern
tzi@cdrh.fda.gov , CDRH-FDA - Food and Drug Administration;
lauretto@supremum.com , Supremum Assessoria e Consultoria;
cpereira@ime.usp.br , IME-USP - University of Sao Paulo;
jstern@ime.usp.br , BIOINFO, NOPEF and IME-USP

The Full Bayesian Significance Test (FBST) for precise hypotheses is presented, with some applications relevant to reliability theory. The FBST is an alternative to significance tests or, equivalently, to *p-values*. In the FBST we compute the evidence of the precise hypothesis. This evidence is the probability of the complement of a credible set "tangent" to the sub-manifold (of the parameter space) that defines the null hypothesis. We use the FBST in an application requiring a quality control of used components, based on remaining life statistics.

Contents

1. Introduction

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 "Full" Bayesian and is based on the analysis of credible sets. By Full 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, that generates the Lindley's paradox effect. The FBST regards likelihoods as the proper means for representing statistical information, a principle stated by Royall (1997) to simplify and unify statistical analysis. Another important aspect of the FBST is its consistency with the "benefit of the doubt" juridical principle. These remarks will be understood in the sequel.

Significance tests are regarded as procedures for measuring the consistency of data with a null hypothesis, Cox (1977) and Kempthorne and Folks (1971). *p-values* are a tail area under the null hypothesis, calculated in the sample space, not in the parameter space where the hypothesis is formulated.

Bayesian significance tests defined in the literature, like Bayes Factor or the posterior probability of the null hypothesis, consider the *p-value* as a measure of evidence of the null hypothesis and present alternative Bayesian measures of evidence, Aitkin (1991), Berger and Delampady (1987), Berger *et al.* (1997), Irony and Pereira (1986, 1995), Pereira and Wechsler (1993), Sellke *et al.* (1999). As pointed out in Cox (1977), the first difficulty to define the *p-value* is the way the sample space is ordered under the null hypothesis. Pereira and Wechsler (1993) suggests a *p-value* that always regards the alternative hypothesis. One can find a great deal of objections agaist each of these measures of evidence. The most important argument against Bayesian tests for precise hypothesis is presented by Lindley (1957). The literature is full of objections to the classical *p-value*. The book by Royall (1997) and its review by Vieland *et al.* (1998) presents interesting and relevant arguments motivating statisticians to start thinking about new methods of measuring evidence. In a more philosophical terms, Carnap (1962), de Finetti (1989), Good (1983) and Popper (1989) discuss, in a great detail, the concept of evidence.

2. Motivation

In order to illustrate the FBST we discus a well known problem. Given a sample from a normal distribution with unknown parameters, we want to test if the standard deviation is equal to a constant. The hypothesis $\sigma = c$ is a straight line. We have a precise hypothesis since it is defined by a manifold (surface) of dimension (one) strictly smaller than the dimension of the parameter space (two).

It can be shown that the conjugate family for the Normal Distribution is a family of bivariate distributions, where the conditional distribution of the mean, μ, for a fixed precision, $\rho = 1/\sigma^2$, is normal, and the marginal distribution of the precision, ρ, is gamma, DeGroot (1970), Lindley (1978). We use the standard improper priors, uniform on $]-\infty, +\infty[$ for μ, and $1/\rho$ on $]0, +\infty[$ for ρ, in order to get a fair comparison with p-values, DeGroot (1970). Hence we have the parameter space, hypothesis and posterior joint distribution:

$$\Theta = \{(\mu, \rho) \in R \times R_+\} \quad , \quad \Theta_0 = \{(\mu, \rho) \in \Theta \,|\, \rho = c\}$$

$$f(\mu, \rho \,|\, x) \propto \sqrt{\rho}\; exp(-n\rho(\mu - m)^2/2) exp(-b\rho)\rho^{a-1}$$

$$x = [x_1 \ldots x_n] \ , \ a = \frac{n-1}{2} \ , \ m = \frac{1}{n}\sum_{i=1}^{n} x_i \ , \ b = \frac{n}{2}\sum_{i=1}^{n}(x_i - m)^2$$

Figure 1 shows the plot of 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, f^*, on the hypothesis. The interior of the tangent level curve, T^*, includes all points with posterior density greater than f^*, i.e. it is the highest probability density set tangent to the hypothesis.

The posterior probability of T^*, κ^*, gives an indication of inconsistency between the posterior and the hypothesis: Small values of κ^* indicate that the hypothesis traverses high density regions, favoring the hypothesis. Therefore we define $Ev(H) = 1 - \kappa^*$ as the measure of evidence (for the precise hypothesis).

In Figure 1 we test $c = 1$ with $n = 16$ observations of mean $m = 10$ and standard deviation $s = 1.02,\ 1.1,$ and 1.5. We present the FBST evidence, Ev, and the standard χ^2-test, $chi2$.

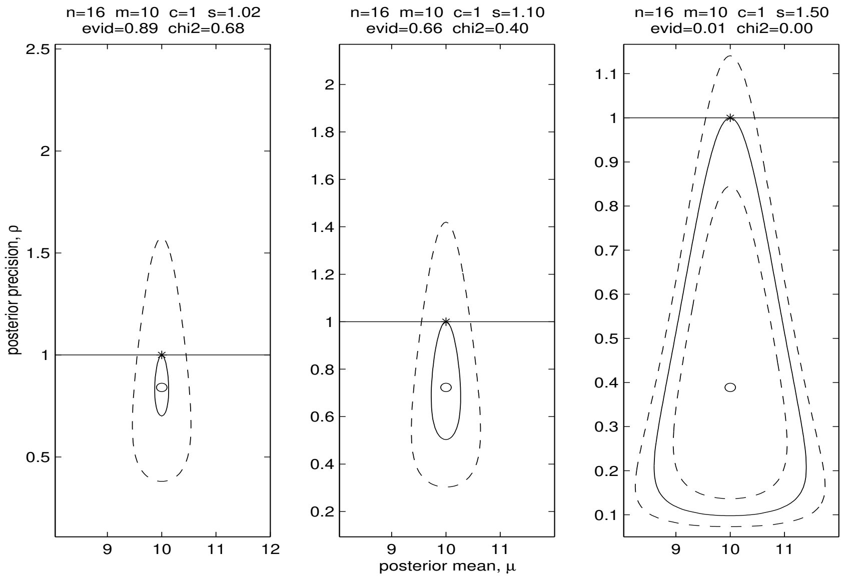

Fig. 1. Tangent and other Highest Probability Density Sets

It is clear that this example is only an illustration: there is no need of new methods to test the standard deviation of a normal distribution. However, efficient numerical optimization and integration computer programs, make it straightforward to extend the FBST to more complex structures. In sections 6 and 7 we present an important application involving the Weibull distribution, requiring a quality control test for used components, based on remaining life data. This problem appears in engineering as well as biological and pharmacological applications. The FBST is exact and performs well even for small samples and low frequencies. In the next section we give a more formal definition of the FBST.

3. The Evidence Calculus

Consider the random variable D that, when observed, produces the data d. The statistical space is represented by the triplet (Ξ, Δ, Θ) where Ξ is the sample space, the set of possible values of d, Δ is the family of measurable subsets of Ξ and Θ is the parameter space. We define now a prior model

(Θ, B, π_d), which is a probability space defined over Θ. Note that in this model $Pr\{A \mid \theta\}$ has to be Θ measurable. As usual, after observing data d, we obtain the posterior probability model (Θ, B, π_d), where π_d is the conditional probability measure on $\boldsymbol{B}$ given the observed sample point, d. In this paper we restrict ourselves to the case where the functions π_d has a probability density function f.

To define our procedure we should concentrate only on the posterior probability space (Θ, B, π_d). First, we define T_φ as the subset of the parameter space where the posterior density is greater than φ.

$$T_\varphi = \{\theta \in \Theta \mid f(\theta) \geq \varphi\}$$

The credibility of T_φ is its posterior probability,

$$\kappa = \pi_d(T_\varphi) = \int_{T_\varphi} f(\theta \mid d) d\theta = \int_\Theta f_\varphi(\theta \mid d) d\theta$$

where $f_\varphi(x) = f(x)$ if $f(x) \geq \varphi$ and zero otherwise.

Now, we define f^* as the maximum of the posterior density over the null hypothesis, attained at the argument θ^*,

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

and define $T^* = T_{f^*}$ as the set "tangent" to the null hypothesis, H, whose credibility is κ^*.

The measure of evidence we propose in this article is the complement of the probability of the set T^*. That is, the evidence of the null hypothesis is

$$Ev(H) = 1 - \kappa^* \quad \text{or} \quad 1 - \pi_d(T^*)$$

If the probability of the set T^* is "large", it means that the null set is in a region of low probability and the evidence in the data is 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 evidence in the data is in its favor. In the next section we give an operational construction of the FBST.

4. Numerical Optimization and Integration

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:

$$\Theta_0 = \{\theta \in \Theta \mid g(\theta) \le 0 \wedge h(\theta) = 0\}.$$

Since we are working with precise hypotheses, we have at least one equality constraint, hence $dim(\Theta_0) < dim(\Theta)$. Let $f(\theta)$ be the posterior probability density function, as defined in the last section.

The computation of the evidence measure defined in the last section is performed in two steps, a numerical optimization step, and a numerical integration step. The numerical optimization step consists of finding an argument θ^* that maximizes the posterior density $f(\theta)$ under the null hypothesis. The numerical integration step consists of integrating the posterior density over the region where it is greater than $f(\theta^*)$. That is,

- Numerical Optimization step:

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

- Numerical Integration step:

$$\kappa^* = \int_{\Theta} f_\varphi(\theta \mid d) d\theta$$

  where $f_\varphi(x) = f(x)$ if $f(x) \ge \varphi$ and zero otherwise.

Efficient computational algorithms are available, for local and global optimization as well as for numerical integration, Bazaraa *et al.* (1993), Horst *et al.* (1995), Luenberger (1984), Nocedal and Wright (1999), Pinter (1996), Krommer and Ueberhuber (1998), and Sloan and Joe (1994). Computer codes for several such algorithms can be found at software libraries as ACM, GSL and NAG, or at internet sites as *www.ornl.gov* and *www-rocq.inria.fr.*

We notice that the method used to obtain T^* and to calculate κ^* can be used under general conditions. Our purpose, however, is to discuss precise hypothesis testing, i.e. $dim(\Theta_0) < dim(\Theta)$, under absolute continuity of the posterior probability model, the case for which most solutions presented in the literature are controversial.

5. Weibull Distribution

The two parameter Weibull probability density, reliability (or survival probability) and hazard functions, for a failure time $t \ge 0$, given the shape, and characteristic life (or scale) parameters, $\beta > 0$, and $\gamma > 0$, are:

$$\begin{aligned}
w(t \,|\, \beta, \gamma) &= (\beta\, t^{\beta-1}/\gamma^{\beta})\, exp(-(t/\gamma)^{\beta}) \\
r(t \,|\, \beta, \gamma) &= exp(-(t/\gamma)^{\beta}) \\
z(t \,|\, \beta, \gamma) &\equiv w(\ )/r(\ ) = \beta\, t^{\beta-1}/\gamma^{\beta}
\end{aligned}$$

The mean and variance of a Weibull variate are given by:

$$\begin{aligned}
\mu &= \gamma\, \Gamma(1 + 1/\beta) \\
\sigma^2 &= \gamma^2(\Gamma(1 + 2/\beta) + \Gamma^2(1 + 1/\beta))
\end{aligned}$$

By altering the parameter, β, $W(t \,|\, \beta, \gamma)$ takes a variety of shapes, Dodson(1994). Some values of shape parameter are important special cases: for $\beta = 1$, W is the exponential distribution; for $\beta = 2$, W is the Rayleigh distribution; for $\beta = 2.5$, W approximates the lognormal distribution; for $\beta = 3.6$, W approximates the normal distribution; and for $\beta = 5.0$, W approximates the peaked normal distribution. The flexibility of the Weibull distribution makes it very useful for empirical modeling, specially in quality control and reliability. The regions $\beta < 1$, $\beta = 1$, and $\beta > 1$ correspond to decreasing, constant and increasing hazard rates. These three regions are also known as infant mortality, memoryless, and wearout failures. γ is approximately the 63rd percentile of the life time, regardless of the shape parameter.

The Weibull also has important theoretical properties. If n i.i.d. random variables have Weibull distribution, $X_i \sim w(t \,|\, \beta, \gamma)$, then the first failure is a Weibull variate with characteristic life $\gamma/n^{1/\beta}$, i.e. $X_{[1,n]} \sim w(t \,|\, \beta, \gamma/n^{1/\beta})$. This kind of property allows a characterization of the Weibull as a limiting life distribution in the context of extreme value theory, Barlow and Prochan (1975).

The affine transformation $t = t' + \alpha$ leads to the three parameter truncated Weibull distribution. A location (or threshold) parameter, $\alpha > 0$ represents beginning observation of a (truncated) Weibull variate at $t = 0$, after it has already survived the period $[-\alpha, 0[$. The three parameter truncated Weibull is given by:

$$\begin{aligned}
w(t \,|\, \alpha, \beta, \gamma) &= (\beta\, (t + \alpha)^{\beta-1}/\gamma^{\beta})\, exp(-((t + \alpha)/\gamma)^{\beta})/r(\alpha \,|\, \beta, \gamma) \\
r(t \,|\, \alpha, \beta, \gamma) &= exp(-((t + \alpha)/\gamma)^{\beta})/r(\alpha \,|\, \beta, \gamma)
\end{aligned}$$

6. Display Panels

We were faced with the problem of testing the wearout of a lot of used display panels. A panel displays 12 to 18 characters. Each character is displayed as a 5×8 matrix of pixels, and each pixel is made of 2 (RG) or 3 (RGB) individual color elements, (like a light emitting diode or gas plasma device). A panel fails when the first individual color element fails. The construction characteristics of a display panel makes the Weibull distribution specially well suited to model its life time. The color elements are "burned in" at the production process, so we assume they are not at the infant mortality region, i.e. we assume the Weibull's shape parameter to be greater than one, with wearout or increasing hazard rates.

The panels in question were purchased as used components, taken from surplus machines. The dealer informed the machines had been operated for a given time, and also informed the mean life of the panels at those machines. Only working panels were acquired. The acquired panels were installed as components on machines of a different type. The use intensity of the panels at each type of machine corresponds to a different time scale, so mean lifes are not directly comparable. The shape parameter however is an intrinsic characteristic of the panel. The used time over mean life ratio, $\rho = \alpha/\mu$, is adimensional, and can therefore be used as an intrinsic measure of wearout. We have recorded the time to failure, or times of withdrawal with no failure, of the panels at the new machines, and want to use this data to corroborate (or not) the wearout information provided by the surplus equipment dealer.

7. The Model

The problem described at the preceding sections can be tested using the FBST, with parameter space, hypothesis and posterior joint density:

$$\Theta = \{(\alpha,\beta,\gamma) \in \,]0,\infty] \times [1,\infty] \times [0,\infty[\ \}$$
$$\Theta_0 = \{(\alpha,\beta,\gamma) \in \Theta \,|\, \alpha = \rho\mu(\beta,\gamma)\ \}$$
$$f(\alpha,\beta,\gamma \,|\, D) \propto \prod_{i=1}^{n} w(t_i \,|\, \alpha,\beta,\gamma) \prod_{j=1}^{m} r(t_j \,|\, \alpha,\beta,\gamma)$$

where the data D are all the recorded failure times, $t_i > 0$, and the times of withdrawal with no failure, $t_j > 0$.

At the optimization step it is better, for numerical stability, to maximize the log-likelihood, $fl(\ )$. Given a sample with n recorded failures and m withdrawals,

$$\begin{aligned} wl_i &= \log(\beta) + (\beta - 1)\log(t_i + \alpha) - \beta\log(\gamma) - ((t_i + \alpha)/\gamma)^\beta + (\alpha/\gamma)^\beta \\ rl_j &= -((t_j + \alpha)/\gamma)^\beta + (\alpha/\gamma)^\beta \\ fl &= \sum_{i=1}^{n} wl_i + \sum_{j=1}^{m} rl_j \end{aligned}$$

the hypothesis being represented by the constraint

$$h(\alpha, \beta, \gamma) = \rho\,\gamma\,\Gamma(1 + 1/\beta) - \alpha = 0$$

The gradients of $fl(\ )$ and $h(\ )$ analytical expressions, to be given to the optimizer, are:

$$\begin{aligned} &dwl = \\ &\quad [\ (\beta - 1)/(t + \alpha) - ((t + \alpha)/\gamma)^\beta\beta/(t + \alpha) + (\alpha/\gamma)^\beta\beta/\alpha\,, \\ &\quad 1/\beta + \log(t + \alpha) - \log(\gamma) - ((t + \alpha)/\gamma)^\beta\log((t + \alpha)/\gamma) + (\alpha/\gamma)^\beta\log(\alpha/\gamma)\,, \\ &\quad -\beta/\gamma + ((t + \alpha)/\gamma)^\beta\beta/\gamma - (\alpha/\gamma)^\beta\beta/\gamma\ ] \\ &drl = \\ &\quad [\ -((t + \alpha)/\gamma)^\beta\beta/(t + \alpha) + (\alpha/\gamma)^\beta\beta/\alpha\,, \\ &\quad -((t + \alpha)/\gamma)^\beta\log((t + \alpha)/\gamma) + (\alpha/\gamma)^\beta\log(\alpha/\gamma)\,, \\ &\quad ((t + \alpha)/\gamma)^\beta\beta/\gamma\,,\ -(\alpha/\gamma)^\beta\beta/\gamma\ ] \\ &dh = \\ &\quad [\ -1\,,\ -\rho\,\gamma\,\Gamma'(1 + 1/\beta)\,\Gamma(1 + 1/\beta)/\beta^2\,,\ \rho\,\Gamma(1 + 1/\beta)\ ] \end{aligned}$$

For gamma and digamma functions efficient algorithms see Spanier and Oldham (1987).

8. Numerical Example

Table 1 displays 45 failure times (in years), plus 5 withdrawals, for a small lot of 50 panels, in a 3.5 years long experiment. The panels have supposedly been used, prior to acquisition, for 30% of its mean life, i.e. we want to test $\rho = 0.3$. In general, some prior distribution of the shape parameter is needed to stabilize the model. Knowing color elements' life time to be approximately normal, we consider $\beta \in [3.0, 4.0]$. Table 2 displays the evidence of some values of ρ. The maximum likelihood estimates of the Weilbull's parameters are $\alpha = 1.25$, $\beta = 3.28$ and $\gamma = 3.54$; so the estimates $\mu = 3.17$ and $\rho = 0.39$. The FBST corroborates the hypothesis $\rho = 0.3$ with an evidence of 98%.

Table 1. Failure times and withdrawals in years, $n = 45, m = 5$

0.01	0.19	0.51	0.57	0.70	0.73	0.75	0.75	1.11	1.16
1.21	1.22	1.24	1.48	1.54	1.59	1.61	1.61	1.62	1.62
1.71	1.75	1.77	1.79	1.88	1.90	1.93	2.01	2.16	2.18
2.30	2.30	2.41	2.44	2.57	2.61	2.62	2.72	2.76	2.84
2.96	2.98	3.19	3.25	3.31	+1.19	+3.50	+3.50	+3.50	+3.50

Table 2. Evidence for some values of ρ

ρ	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Evid	0.04	0.14	0.46	0.98	1.00	0.98	0.84	0.47	0.21	0.01

9. Final Remarks

The theory presented in this paper, grew out of the necessity of the authors' activities in the role of audit, control or certification agents, Pereira and Stern (1999a). These activities made the authors (sometimes painfully) aware of the benefit of the doubt juridical principle, or safe harbor liability rule. This kind of principle establishes that there is no liability as long as there is a reasonable basis for belief, effectively placing the burden of proof on the plaintiff, who, in a lawsuit, must prove false a defendant's misstatement. Such a rule also prevents the plaintiff from making any assumption not explicitly stated by the defendant, or tacitly implied by existing law or regulation. The use of an a priori point mass on the null hypothesis, as on standard Bayesian tests, can be regarded as such an ad hoc assumption.

As audit, control or certification agents, the authors had to check compliance with given requirements and specifications, formulated as precise hypotheses on contingency tables. In Pereira *et al.* (1999b) we describe several applications based on contingency tables, comparing the use of FBST with standard Bayesian and Classical tests. The applications presented in this paper are very similar in spirit, but we are not aware of any standard exact test in the literature. The implementation of FBST is immediate and trivial, as long as good numerical optimization and integration programs are at hand. In the applications in this paper, as well in those in Pereira *et al.* (1999b), it is desirable or necessary to use a test with the following characteristics:

- Be formulated directly in the original parameter space.
- Take into account the full geometry of the null hypothesis as a manifold (surface) imbedded in the whole parameter space.
- Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the (manifold representing the) null hypothesis being used.
- Be consistent with the benefit of the doubt juridical principle (or safe harbor liability rule), i.e. consider in the "most favorable way" the claim stated by the hypothesis.
- 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.
- Consider the alternative hypothesis in equal standing with the null hypothesis, in the sense that increasing sample size should make the test converge to the right (accept/reject) decision.
- 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 straightforward (computational) implementation. Moreover, as shown in Madruga *et al.* (2001), the FBST is also in perfect harmony with the Bayesian decision theory of Rubin (1987), in the sense that there are specific loss functions which render the FBST.

We remark that the evidence calculus defining the FBST takes place entirely in the parameter space where the prior was assessed by the scientist, Lindley (1983). We call it the "original" parameter space, although acknowledging that the parameterization choice for the statistical model semantics is somewhat arbitrary. We also acknowledge that the FBST is not invariant under general change of parameterization.

The FBST is in sharp contrast with the traditional schemes for dimensional reduction, like the elimination of so called "nuisance" parameters.

In these "reduced" models the hypothesis is projected into a single point, greatly simplifying several procedures. Problems with the traditional approach are presented in Pereira and Lindley (1987). The traditional reduction or projection schemes are also incompatible with the benefit of doubt principle, as stated earlier. In fact, preserving the original parameter space, in its full dimension, is the key for the intrinsic regularization mechanism of the FBST, when it is used in the context of model selection, Pereira and Stern (2000,2001).

Of course, there is a price to be paid for working with the original parameter space, in its full dimension: A considerable computational work load. But computational difficulties can be overcome with the used of efficient continuous optimization and numerical integration algorithms. Large problems can also benefit from program vectorization and parallelization techniques. Dedicated vectorized or parallel machines may be expensive and not always available, but most of the algorithms needed can benefit from asynchronous and coarse grain parallelism, a resource easily available, although rarely used, on any PC or workstation network through MPI, Portable Parallel Programming Message-Passing Interface, or similar distributed processing environments, Wilson and Lu (1996).

Finally, we notice that statements like "increase sample size to reject (accept) the hypothesis" made by many users of frequentist (standard Bayesian) tests, do not hold for the FBST. Increasing the sample size makes the FBST converge to the Boolean truth indicator of hypothesis being tested. In this sense, the FBST has good acceptance/rejection symmetry, even if the safe harbor rule prevents this symmetry from being perfect, introducing an offset for small samples. We believe that the existence of a precise hypothesis test with the FBST's symmetry properties has important consequences in knowledge theory, given the role played by the completely asymmetric standard statistical tests in some epistemological systems, Carnap (1962), Popper (1989).

References

1. Aitkin, M. (1991). Posterior Bayes Factors. *J R Statist Soc B* **53**(1), 111–142.
2. Barlow, R.E. and Prochan, F. (1988) *Statistical Theory of Reliability and Life Testing.* NY: Holt, Rinehart and Winston Inc.
3. Barlow, RE and Prochan, F (1988) Life Distributions Models and Incom-

plete Data. In: *Handbook of Statistics* **7***: Quality Control and Reliability.* (Krishnaiah, PR and Rao, CR eds.) Amsterdam: North-Holand pp 225-250.

4. Bazaraa, M.S., Sherali, H.D., and Shetty, C.M. (1993) *Nonliear Programming: Theory and Algorithms.* NY: Wiley.
5. Berger, J.O. and M Delampady, M. (1987). Testing precise hypothesis. *Statistical Science* **3**, 315–352.
6. Berger, J.O., Boukai, B. and Wang, Y. (1997). Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science* **3**, 315–352.
7. Bratley, P., Fox B.L. and Schrage, L. *A Guide to Simulation.* Springer-Verlag, 1987.
8. Dodson, B. (1994). *Weibull Analysis.* Milwaukee: ASQC Quality Press.
9. Carnap, R. (1962). *Logical Foundations of Probability.* Univ. of Chicago Press.
10. Cox, D.R. (1977). The role of significance tests. *Scand J Statist* **4**, 49–70
11. DeGroot, M.H. (1970). *Optimal Statistical Decisions.* NY: McGraw-Hill.
12. Evans, M. (1997). Bayesian Inference Procedures Derived via the Concept of Relative Surprise. *Communications in Statistics* **26**, 1125–1143.
13. Galassi, M., Davies J., Theiler, J., Gough, B., Priedhorsky, R., Jungman, G. and Booth, M. (1999). *GSL - GNU Scientific Library Reference Manual V-0.5.* WWW: lanl.gov.
14. Gomez, C. (1999). *Engineering and Scientific Computing with Scilab.* Berlin: Birkhäuser.
15. Good, I.J. (1983). *Good thinking: The foundations of probability and its applications.* University of Minnesota Press.
16. Finnetti, B. de (1989). *Decisão.* in Enciclopédia EINAUDI. Romano, R. (edtr.). O Porto: Imprensa Nacional.
17. Horst, R., Pardalos P.M. and Thoai N.V. (1995). *Introduction to Global Optimization.* Boston: Kluwer.
18. Irony, T.Z. and Pereira, C.A.B. (1986). Exact test for equality of two proportions: Fisher×Bayes. *J Statist Comp & Simulation* **25**, 93–114.
19. Irony, T.Z. and Pereira C.A.B. (1995). Bayesian Hypothesis Test: Using surface integrals to distribute prior information among hypotheses. *Resenhas* **2**, 27–46.
20. Kempthorne, O. and Folks, L. (1971). *Probability, Statistics, and Data Analysis.* Ames: Iowa State U.Press.
21. Knuth, D.E. *The Art of Computer Programming, vol 2 - Seminumerical Algorithms.* Addison Wesley, 1996.
23. Krommer, A.R. and Ueberhuber C.W. (1998). *Computational Integration.* Philadelphia: SIAM.
24. L'Ecuyer, P. Efficient and Portable Combined Pseudo-Random Number Generators. *Commun. ACM*, 1988.
25. Lepage, G.P (1978). VEGAS: An Adaptive Multidimensional Integration Program. *J. Comput. Phys.* **27**, 192–205.
26. Lehmann, E.L. (1986). *Testing Statistical Hypothesis.* NY: Wiley.

27. Lindley, D.V. (1957). A Statistical Paradox. *Biometrika* **44**, 187–192.
28. Lindley, D.V. (1978). The Bayesian Approach. *Scand J Statist* **5**, 1-26.
29. Luenberger, D.G. (1984). *Linear and Nonlinear Programming.* Reading: Addison-Wesley.
30. Madruga, M.R. Esteves, L.G. and Wechsler, S. (2001). On the Bayesianity of Pereira-Stern Tests. To appear in *TEST.*
31. Nocedal, J. and Wright, S. (1999). *Numerical Optimization.* NY: Springer.
32. Pereira,C.A.B. and Lindley,D.V. (1987). Examples Questioning the Use of Partial Likelihood. *The Statistician* **36**, 15–20.
33. Pereira,C.A.B. and Wechsler,S. (1993). On the Concept of *p-value. Braz J Prob Statist* **7**, 159–177.
34. Pereira,C.A.B. Stern,J.M. (1999a). A Dynamic Software Certification and Verification Procedure. *Proc. ISAS-99 - International Conference on Information Systems Analysis and Synthesis* II, 426–435.
35. Pereira,C.A.B. Stern,J.M. (1999b). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1**, 69–80.
36. Pereira,C.A.B. Stern,J.M. (2000). *Intrinsic Regularization in Model Selection using the Full Bayesian Significance Test.* Technical Report RT-MAC-2000-6, Dept. of Computer Science, University of Sao Paulo.
37. Pereira,C.A.B. Stern,J.M. (2001). Model Selection: Full Bayesian Approach. To appear in *Environmetrics.*
38. Pintér, J.D. (1996). *Global Optimization in Action. Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications.* Boston: Kluwer.
39. Popper, K.R. (1989). *Conjectures and Refutations: The Growth of Scientific Knowledge.* London: Routledge.
40. Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm.* London: Chapman & Hall.
41. Rubin, H. (1987). A Weak System of Axioms for "Rational" Behavior and the Non-Separability of Utility from Prior. *Statistics and Decisions* **5**, 47–58.
42. Sellke, T., Bayarri, M.J. and Berger, J. (1999). Calibration of p-values for Testing Precise Null Hypotheses. *ISDS Discussion Paper 99-13.*
43. Sloan, I.R. and Joe, S. (1994). *Lattice Methods for Multiple Integration.* Oxford: Oxford University Press.
44. Spanier, J. and Oldham, K.B. (1987). *An Atlas of Functions.* NY: Hemisphere Publishing.
45. Vieland, V.J. and Hodge, S.E. (1998). Book Reviews: *Statistical Evidence* by R Royall (1997). *Am J Hum Genet* **63**, 283–289.
46. Wichmann, B.A. and Hill, I.D. (1982). An Efficient and Portable Pseudo-Random Number Generator. *Appl. Stat.*, **31**, 188-190.
47. Wilson,G.V. and Lu,P. (1996). *Parallel Programming Using C++.* Cambridge: MIT Press.