

FBST for Generalized Poisson Distribution

Paulo C. Hubert Jr.^{*,†}, Marcelo S. Lauretto^{*,**} and Julio M. Stern^{*,‡}

^{*}University of São Paulo, Brazil.

[†]paulo.hubert@gmail.com

^{**}marcelolauretto@usp.br

[‡]jmstern@hotmail.com

Abstract. The Generalized Poisson Distribution (GPD) adds an extra parameter to the usual Poisson distribution. This parameter induces a loss of homogeneity in the stochastic processes modeled by the distribution. Thus, the generalized distribution becomes an useful model for counting processes where the occurrence of events is not homogeneous. This model creates the need for an inferential procedure, to test for the value of this extra parameter. The FBST (*Full Bayesian Significance Test*) is a Bayesian hypotheses test procedure, capable of providing an evidence measure on sharp hypotheses (where the dimension of the parametric space under the null hypotheses is smaller than that of the full parametric space). The goal of this work is study the empirical properties of the FBST for testing the nullity of extra parameter of the generalized Poisson distribution. Numerical experiments show a better performance of FBST with respect to the classical likelihood ratio test, and suggest that FBST is an efficient and robust tool for this application.

Keywords: Hypotheses testing, Inference, Bayesian inference, Counting processes, Poisson distribution

PACS: 01.30.Cc, 02.50.-r, 02.50.Ng, 02.50.Tt

INTRODUCTION

The Poisson distribution is applied to models resulting in counting processes. The distribution depends on a single parameter, α , which represents the occurrence rate for the events being counted. The parameter α is also both the mean and the variance of this distribution.

The model implicitly assumes an homogeneous occurrence rate for the process. There are, however, a variety of situations where such assumption fails to be fulfilled; one example is the distribution of the number of organisms in a colony of bacteria, where the birth rates for any site in space or time are dependent upon the previous number of organisms near the site in question (due to the availability of food, temperature, social behavior, etc).

A variety of Generalized Poisson Distributions (GPDs) have been proposed; in particular, the works of Consul and Jain [6,7,8] brought forth a distribution with two parameters, α and λ , that generalizes the usual Poisson distribution in a flexible manner. Changing λ , it is possible to induce an increase or decrease in the occurrence rate being modeled.

GPDs arise in several discrete models where the probabilities depend either on the previous number of occurrences or on the average number of events in any given range. Some urn models, where the probability distributions depend on the total number of balls in several urns, also lead to a GPD. There is also the Generalized Poisson Process,

a point process where the occurrence probabilities in any small range depend on the previous state of the process, as well as on the particular time and length of the range. Other models leading to a Generalized Poisson Distribution are: The Negative Binomial distribution, that converges in probability to a GPD when the number of trials goes to infinity in such a way that $n \times p = \theta$; The Quasi-binomial distribution, in similar conditions, when $n \rightarrow \infty$, $\rho \rightarrow 0$ and $\phi \rightarrow 0$; see [4, 5].

Since the flexibility of the generalized distribution depends heavily upon the value of λ , the need for an inference procedure on the hypothesis $H_0 : \lambda = 0$ is clear: ultimately, it is a test of homogeneity for the stochastic processes generating the data under analysis.

The *Full Bayesian Significance Test*, FBST, is an inference procedure well-suited for testing sharp hypotheses. Since the hypothesis $H_0 : \lambda = 0$ defined above is sharp, in the sense of Stern [20], FBST is a natural candidate to be used as inference tool.

The goal of this work is study the empirical properties of the FBST for testing the nullity of extra parameter of the generalized Poisson distribution. Firstly, we introduce the generalized Poisson distribution and the Full Bayesian Significance Test in the present context. Then, we discuss the numerical experiments performed in order to compare performances of FBST and Likelihood Ratio test.

THE GENERALIZED POISSON DISTRIBUTION (GPD)

Let X be a discrete random variable, taking only non-negative values. X follows a generalized Poisson distribution if its probability distribution function is given by:

$$p(x|\alpha, \lambda) = \begin{cases} (1/x!) \alpha (\alpha + x\lambda)^{x-1} e^{-\alpha - x\lambda} & x = 0, 1, 2, \dots \\ 0 & \text{if } x > m, \text{ when } \lambda < 0 \end{cases} \quad (1)$$

where $\alpha > 0$, $\max(-1, -\alpha/m) < \lambda \leq 1$, and $m(\geq 4)$ is the largest integer such that $\alpha + m\lambda > 0$ when $\lambda < 0$. This lower bound on λ is imposed in order to assure at least five points in the sample space with positive probabilities when λ is negative. It can be easily seen that the distribution reduces to the usual Poisson distribution when $\lambda = 0$. As a consequence of the definition, when $\lambda < 0$, the distribution does not sum to unity in $[0, 1, \dots]$, and must be normalized. For an analysis of this truncated distribution, and an account for estimating its parameter using Bayesian methods, see Scollnik [14].

The mean and variance of X are given by

$$E(X) = \frac{\alpha}{1 - \lambda} \quad \text{Var}(X) = \frac{\alpha}{(1 - \lambda)^3} \quad (2)$$

and when $\lambda = 0$, $E(X) = \text{Var}(x) = \alpha$.

The probability distribution thus defined has some interesting properties, such as convolution (analog to the regular Poisson distribution) and unimodality [4]. Furthermore, there exist a variety of probability models which lead to this distribution; a few examples can be found in Consul [4].

Many applications for the generalized Poisson distribution can be found in the literature: Scollnik [14] mentions some applications in the analysis of actuarial data; Famoye and Singh [10] build a regression model with the generalized Poisson distribution and

apply it to domestic violence data; and Sheth [16] uses it as the limit distribution on a clustering model with Poisson initial conditions.

FULL BAYESIAN SIGNIFICANCE TEST

The Full Bayesian Significance Test (FBST) is presented by Pereira and Stern [12] as a coherent Bayesian significance test. FBST is intuitive and has a geometric characterization. In this article the parameter space, Θ , is a subset of R^n , and the hypothesis is defined as a restricted subset defined by vector valued inequality and equality constraints: $H : \theta \in \Theta_H$ where $\Theta_H = \{\theta \in \Theta | g(\theta) \leq 0 \wedge h(\theta) = 0\}$. For simplicity, we often use H for Θ_H . We are interested in precise hypotheses, with $\dim(H) < \dim(\Theta)$. $f_x(\theta)$ is the posterior probability density function. In the present context of Generalized Poisson Distribution, $\theta = [\alpha, \lambda]$

The computation of the evidence measure used on the FBST is performed in two steps:

- The *optimization step* consists of finding the maximum (supremum) of the posterior under the null hypothesis, $\theta^* = \arg \sup_H f_x(\theta)$, $f^* = f_x(\theta^*)$.
- The *integration step* consists of integrating the posterior density over the Tangential Set, \bar{T} , where the posterior is higher than anywhere in the hypothesis, i.e.,

$$\begin{aligned} \bar{T} &= \{\theta \in \Theta : f_x(\theta) > f^*\} \\ \bar{\text{Ev}}(H) &= \Pr(\theta \in \bar{T} | x) = \int_{\bar{T}} f_x(\theta) d\theta \end{aligned}$$

$\bar{\text{Ev}}(H)$ is the evidence against H , and $\text{Ev}(H) = 1 - \bar{\text{Ev}}(H)$ is the evidence supporting (or in favor of) H . For a better understanding of this evidence measure, Figure 1 illustrates two examples of FBST application to the hypothesis $H : \lambda = 0$. The level curves represent the posterior distributions for two simulated samples of size $n = 50$. The first sample (left) was generated with parameters $\alpha = 3, \lambda = -0.02$ and the second sample (right) was generated with parameters $\alpha = 3, \lambda = -0.20$. The vertical (dotted) line crossing $\lambda = 0$ represents the space under hypothesis, Θ_H . Marker “*” represents the point θ^* of maximum posterior density in Θ_H , and the black level curve tangent to θ^* corresponds to \bar{T} frontier. Intuitively, if the hypothesis set is in a region of “high” posterior density (as in example 1), then \bar{T} is a “small” set, and therefore $\bar{\text{Ev}}(H)$ is “small” ($\simeq 0.131$), meaning “weak” evidence against H . On the other hand, if the hypothesis set is in a region of “low” posterior density (as in example 2), then \bar{T} is “heavy” and therefore $\bar{\text{Ev}}(H)$ is “large” ($\simeq 0.946$), meaning “strong” evidence against H .

For acceptance/rejection of H , critical levels for the evidence measure can be derived as follows. Let us consider the cumulative distribution of the evidence value against the hypothesis, $\bar{V}(\tau) = \Pr(\bar{\text{Ev}} \leq \tau)$, given θ^0 , the true value of the parameter. Under appropriate regularity conditions, for increasing sample size, $n \rightarrow \infty$, we can state the following [13]:

- If H is false, $\theta^0 \notin H$, then $\bar{\text{Ev}}$ converges (in probability) to one, that is, $\bar{V}(\tau) \rightarrow \delta(1)$.

- If H is true, $\theta^0 \in H$, then $\bar{V}(\tau)$, the confidence level, is approximated by the function $\bar{W}(t, h, \tau) = \text{Chi2}(t - h, \text{Chi2}^{-1}(t, c))$, where $t = \dim(\Theta)$, $h = \dim(H)$ and $\text{Chi2}(k, x)$ is the cumulative chi-square distribution with k degrees of freedom.

Hence, to reject H with a significance level δ , one can set an asymptotic threshold $\tau = \bar{W}^{-1}(t, h, 1 - \delta)$, i.e. set τ such that $\bar{W}(t, h, \tau) = 1 - \delta$. For the numerical experiments performed in this work, we take the traditional significance level $\delta = 0.05$.

Several FBST applications and examples, efficient computational implementation, interpretations, and comparisons with other techniques for testing sharp hypotheses, can be found in the authors' papers in the reference list. For a FBST review see Pereira et al. [13].

In this work, we consider an (improper) uniform distribution for $\theta = [\alpha, \lambda]$, and hence the posterior distribution is proportional to the likelihood function:

$$f_x(\theta) = f_x(\alpha, \lambda) \propto \prod_{i=1}^n p(x_i | \alpha, \lambda) = \prod_{i=1}^n \frac{\alpha(\alpha + x_i \lambda)^{x_i - 1} e^{-\alpha - x_i \lambda}}{x_i!}$$

The evidence against H is computed as:

$$\begin{aligned} \bar{T} &= \{\theta \in \Theta : f_x(\theta) > f_x(\theta^*, 0)\} \\ \bar{\text{Ev}}(H) &= \frac{\int_{\bar{T}} f_x(\theta) d\theta}{\int_{\Theta} f_x(\theta) d\theta} \end{aligned}$$

The point θ^* of maximum posteriori over H is easily obtained, since when $\lambda = 0$, $f_x(\theta)$ equals the Poisson density, and therefore $\theta^* = (\bar{x}, 0)$, where \bar{x} denotes the sample mean.

The numerical integration over the posterior measure is performed via Monte Carlo method, by generating values for parameters α, λ of GPD. Since there are no conjugate distributions for these parameters, we used the adaptive sampler HITRO [11].

HITRO combines the multivariate Ratio-of-Uniforms method with the Hit-and-Run sampler. The Ratio-of-Uniforms transformation maps the region below the p.d.f f , i.e. $G(f) = \{(x, y) : 0 < y < f(x)\}$ into the region

$$A(f) = A_{r,m}(f) = \left\{ (u, v) : 0 < v < f\left(\frac{u}{v^r} + m\right)^{1/(r+1)} \right\}$$

by means of the transformation

$$(u, v) \mapsto (x, y) = \left(\frac{u}{v^r} + m, v^{r+1}\right).$$

The vector m must be a point near the mode (in our implementation, we set m as the mode). The method relies on the theorem that, if (u, v) is uniformly distributed over $A(f)$, then $x = u/v^r + m$ has probability density function $f(x)/\int f(z)dz$. The Hit-and-run sampler is used for generating points (u, v) uniformly over $A(f)$.

The task of finding the unconstrained mode of GPD is performed numerically, via the Algencan-Tango solver, which source code and detailed description are freely distributed [1, 2] (see internet link at the reference).

RESULTS AND DISCUSSION

We run some numerical experiments based on simulated data in order to evaluate the FBST performance on the present context.

As a benchmark, we used the Likelihood Ratio Test (LRT), which is vastly known in statistics literature; it is based on the frequentist inference theory [3]. The test statistic is defined by

$$q = -2 \ln \frac{\sup_{(\alpha, \lambda) \in \Theta_H} L(\alpha, \lambda | X)}{\sup_{(\alpha, \lambda) \in \Theta} L(\alpha, \lambda | X)}$$

and q asymptotically follows a Chi-square distribution, with degrees of freedom given by the difference between the dimension of the unconstrained parameter space and the dimension of parameter space under H_0 ; in the present context, q follows a Chi-square distribution with 1 degree of freedom.

The parameter's range in our empirical study is motivated by two applications: Car accidents in non-life insurance, and default risk modeling in financial banking operations. In both applications the range of interest is on parameters $\theta = [\alpha, \lambda]$ that generate countings around 2 or 3. Hence, we choose to simulate samples with fixed means $E(X) \in \{2, 3\}$ and $\lambda \in \{0.0, \pm 0.01, \pm 0.02, \pm 0.05, \pm 0.1, \pm 0.15, \pm 0.2, \pm 0.25, \pm 0.3\}$. Each value of $E(X)$ and λ determines the corresponding value of α , according to the expectancy equation presented previously: $\alpha = E(X)(1 - \lambda)$. Figure 2 presents the values of α computed from the values of $E(x), \lambda$.

For each pair $[\lambda, E(X)]$, we generated a total of 300 samples of size $n = 15$. Our interest is to investigate the frequency of cases where FBST and LRT rejected the null hypothesis $\lambda = 0$, using for both tests asymptotic significance $\delta = 0.05$. Figure 3 shows the rejection rates obtained by FBST and LRT, for samples generated with mean 2 (left) and 3 (right). The vertical (dotted) line corresponds to samples generated under the hypothesis.

The graphics show that FBST is more capable than LRT to detect deviances from hypothesis when $\lambda < 0$. When $\lambda > 0$, both methods present very similar results. For $\lambda = 0$, the rejection rates are: for $E(X) = 2$, both FBST and LRT provide a rejection rate of 6.4% (slightly above the theoretical significance of 5%). For $E(X) = 3$, FBST and LRT provide a rejection rates of 5.3% and 4.7%, respectively, which are close to the theoretical significance.

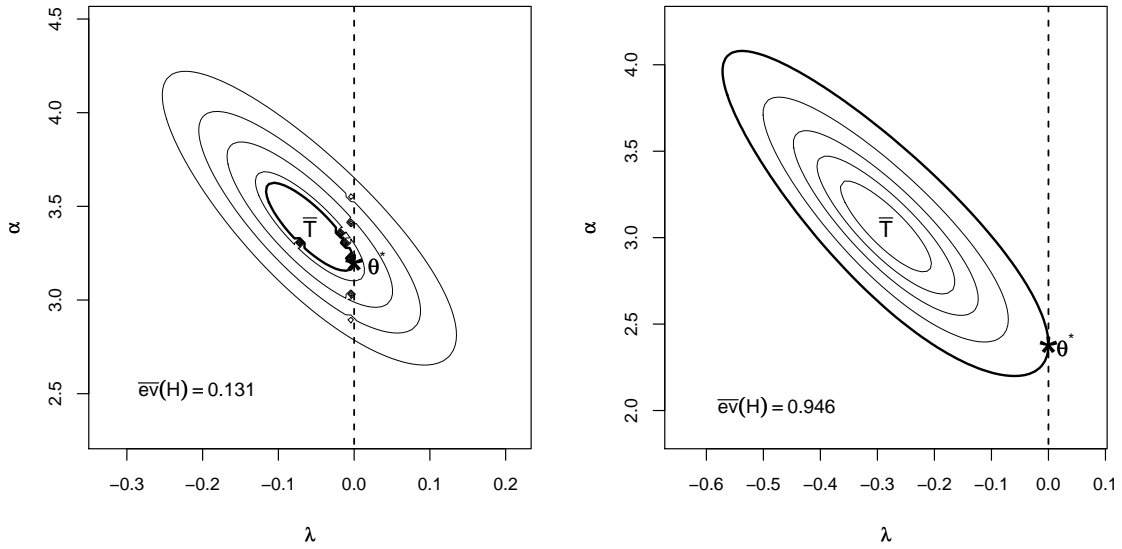


Figure 1: Contour plots of posterior densities and evidence against $H : \lambda = 0$ for two simulated samples of size $n = 100$ and parameters $\alpha = 3, \lambda \in \{-0.02, -0.2\}$.

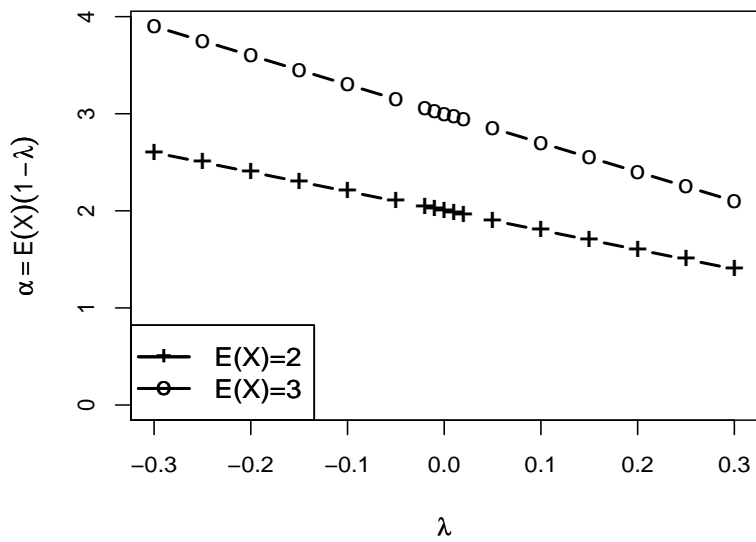


Figure 2: Values of α computed from fixed values for $E(X), \lambda$.

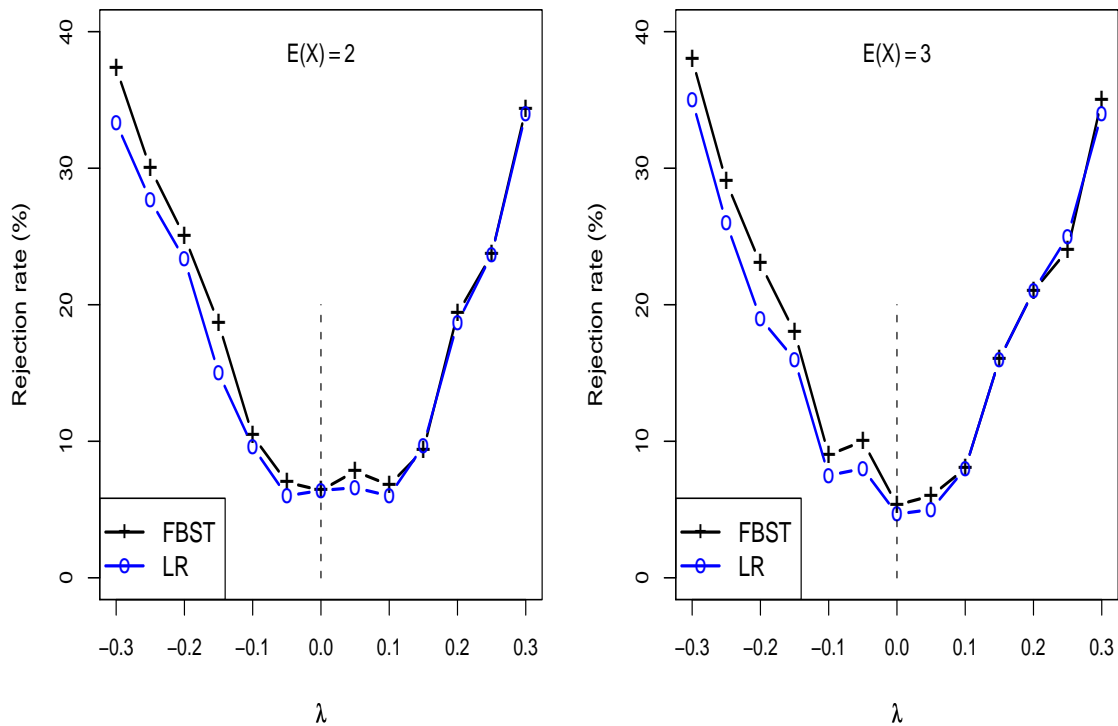


Figure 3: Rejection rates (%) obtained by FBST and LRT, with samples of size $n = 15$, mean 2 (left) and 3 (right), and λ taking several values in the interval $[-0.3, 0.3]$.

CONCLUSION

The goal of this paper was to study the empirical properties of the FBST for testing the nullity of extra parameter of the generalized Poisson distribution. Numerical experiments have shown a better performance of FBST, compared to the classical likelihood ratio test, and suggest that the FBST is an efficient and robust inference tool for this application.

When comparing the performances of the two tests, FBST was more responsive than that of the competing likelihood-ratio test: The FBST test rejection rate attains its minimum when H_0 is true, and increases rapidly when $\lambda \neq 0$. The rejection rate when the hypothesis is false increases faster than that of the competing test, specially when λ decreases, that is, when $\lambda \leq 0$.

Bayesian methods for model selection, like [15,17], could be used as an alternative approach to the problem at hand, and will be the subject of future research. We must stress, however, that the correct interpretation of the answers generated by alternative Bayesian methods may be quite different, as studied in [18,19,20].

The authors are grateful for the support of COPPE-UFJR, EACH-USP, IME-USP, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP).

REFERENCES

1. E.G.Birgin , J.M.Martinez (2002). Large-scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients. *Computational Optimization and Applications*, 23, 101-125.
2. E.G.Birgin, J.M.Martinez and M.Raydan (2000). Nonmonotone Spectral Projected Gradient Methods on Convex Sets. *SIAM Journal on Optimization*, 10, 1196-1211. Software and documentation available at www.ime.usp.br/~egbirgin/tango/.
3. G.Casella, R.L.Berger (2002). *Statistical Inference*, 2nd Edition. Pacific Grove, Thomson Learning.
4. P.C.Consul (1988). *Generalized Poisson Distributions*. CRC, Florida, USA.
5. P.C.Consul, F.Famoye (2006). *Lagrangian probability distribution*. Birkhauser, Boston, USA.
6. P.C.Consul, G.C.Jain (1970). On the generalization of Poisson distribution. *Ann. Math. Statist.*, 41, 1387.
7. P.C.Consul, G.C.Jain (1973a). A generalization of the Poisson distribution. *Technometrics*, 15, 4, 791–799.
8. P.C.Consul, G.C.Jain (1973b). On some interesting properties of the gene realized Poisson distribution. *Biometrische*, 15, 495–500.
9. P. C.Consul, L.R.Shenton (1973). Some interesting properties of Lagrangian distributions. *Comm. Stat. Theor. Meth.*, 2, 3, 263–272.
10. F.Famoye, K.P.Singh (2006). Zero-inflated generalized Poisson regression model with an application to domestic violence data. *Journal of Data Science*, 4, 117–130.
11. R.Karawatzki, J.Leydold, K.Pötzelberger (2005). Automatic Markov Chain Monte Carlo Procedures for Sampling from Multivariate Distributions. Department of Statistics and Mathematics Wirtschaftsuniversität Wien Research Report Series. Report 27, December 2005. Software available at <http://statistik.wu-wien.ac.at/arvag/software.html>.
12. C.A.B.Pereira, J.M.Stern (1999). Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, 1, 99–110.
13. C.A.P.Pereira, J.M.Stern, S.Wechsler (2008). Can a significance test be genuinely Bayesian? *Bayesian Analysis*, 3, 1, 79–100.
14. D.P.M.Scollnik (1998). On the analysis of the truncated generalized Poisson distribution using a Bayesian method. *ASTIN Bulletin*, 28, 135–152.
15. G.E.Schwarz (1978). Estimating the Dimension of a Model. *Ann. Statistics*, 6, 2, 461-464.
16. R.K.Sheth (1998). The generalized Poisson distributions and a model of clustering from Poisson initial conditions. *Monthly Notices Royal Astronomical Society*, 299, 1, 207–217.
17. D.S.Sivia, C.J.Carlie (1992). Molecular Spectroscopy and Bayesian Spectral Analysis: How many lines are there? *J.of Chemical Physics*, 96, 1, 170-178.
18. J.M.Stern (2003). 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.
19. J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *Lecture Notes Artificial Intelligence*, 3171, 134–143.
20. J.M.Stern (2007). Cognitive constructivism, Eigen-solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing*, 14, 1, 9–36.