The Full Bayesian Significance Test for Symmetry in Contingency Tables

Gustavo G. Bernardo*, Marcelo S. Lauretto† and Julio M. Stern*,**

*University of Sao Paulo, Institute of Mathematics and Statistics
†University of Sao Paulo, School of Arts, Sciences and Humanities
**jstern@ime.usp.br

Abstract. The test for symmetry in contingency tables constitutes a broad and important subarea in Statistics, and several methods have been devised for this problem. In this work we propose the Full Bayesian Significance Test (FBST) for the problems of symmetry and point-symmetry in contingency tables. FBST is an intuitive Bayesian approach which avoids to assign positive probabilities to zero measure sets when testing sharp hypotheses. Numerical experiments comparing FBST performance to power-divergence statistics suggest that FBST is a good alternative for problems concerning tests for symmetry in contingency tables.

Keywords: Contingency tables, FBST, Power divergence, Significance Tests, Symmetry.
PACS: 01.30.Cc, 02.50.Ng, 02.50.-r

1. INTRODUCTION

The problem of symmetry hypothesis is fundamental in statistics analysis, where the researcher must assess the existence of a certain symmetry condition, see [1,2]. In several applications, the state of compliance, normality or health is characterized by the existence of symmetries - for example, the cylindrical symmetry of an equipment part or a cornea of a human eye, or the specular symmetry of a pendulous movement or the human walk. In these situations, the lack of symmetry is an indicator of non-compliance, abnormality or illness. The early detection of the lack of symmetry can frequently allow the repair, maintenance or simplified treatment, thus avoiding much more expensive and complex late procedures. This kind of early detection may be helpful in avoiding severe consequences, e.g. the breaking of an important part in a machine during its operation.

The test for symmetry in contingency tables constitutes a broad and important subarea in Statistics. Symmetry is usually referred to the main diagonal. When symmetry in a contingency table is defined with respect to a point, namely the center of the table, we have a point-symmetry model, see [3].

Several methods have been devised for symmetry in contingency tables, e.g. classical Chi-square test [4], power-divergence statistic [1,5] and entropy methods [2].

In this work we propose the FBST, Full Bayesian Significance Test [6], for the problems of symmetry and point-symmetry in contingency tables. FBST is an intuitive Bayesian approach which avoids to assign positive probabilities to zero measure sets when testing sharp hypotheses.

In the next sections we introduce the Full Bayesian Significance Test and the problems of symmetry. Then we discuss some numerical results comparing FBST performance to power-divergence statistics.

2. FULL BAYESIAN SIGNIFICANCE TEST

The Full Bayesian Significance Test (FBST) is presented by Pereira and Stern [6] as a coherent Bayesian significance test. FBST is suitable for cases where 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 $\dim(H) < \dim(\Theta)$. In this work, $f_x(\theta)$ denotes the posterior probability density function.

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, $\overline{T}$, where the posterior is higher than anywhere in the hypothesis, i.e.,

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

$\overline{\mathrm{Ev}}(H)$ is the evidence against H, and $\mathrm{Ev}(H) = 1 - \overline{\mathrm{Ev}}(H)$ is the evidence supporting (or in favor of) H. Intuitively, if the hypothesis set is in a region of "righ" posterior density, then $\overline{T}$ is a "small" set, and therefore $\overline{\mathrm{Ev}}(H)$ is "small", meaning "weak" evidence against H. On the other hand, if the hypothesis set is in a region of "low" posterior density, then $\overline{T}$ is "heavy" and therefore $\overline{\mathrm{Ev}}(H)$ is "large", meaning "strong" evidence against H.

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 detailed FBST review see [7].

3. TESTS FOR SYMMETRY

In this section we present the formulation of test for diagonal and point symmetry in contingency tables, and the FBST and power-divergence strategies for these problems.

For simplicity, we introduce the formulation considering two-dimensional contingency tables, observing that the extension to the multi-dimensional case is straightforward. A two-dimensional contingency table represents the observed frequencies of cross-classified cases, according to two variables A and B. The contingency table will have r rows representing the categories $A_1, A_2, \ldots, A_r$ of variable A, and c columns representing the categories $B_1, B_2, \ldots, B_c$ of variable B. We consider a sample of n individuals, each one classified in unique categories in A and B. The contingency table is represented by an array $X = (x_{i,j}), i = 1...r, j = 1...c$, where $x_{i,j}$ is the number of sample cases belonging to categories A_i, B_j.

The parameter of interest is the joint distribution of probability for the categories in A and B. This distribution is denoted by the array $\theta = (\theta_{i,j}), i = 1...r, j = 1...c$, where $\theta_{i,j}$ is the probability that an individual drawn at random from the population belongs to categories A_i and B_j. Hence, the parameter space, Θ, is a simplex

$$\Theta = \{\theta \geq \mathbf{0} \,|\, \theta'\mathbf{1} = 1\},$$

where $\mathbf{0}$ and $\mathbf{1}$ denote vectors of zeros and ones with same size of θ, and θ' is the transpose of θ.

Using this notation, the diagonal symmetry and point symmetry hypotheses are defined as constrained subspaces of Θ:

- Diagonal Symmetry: $H = \{\theta \in \Theta \,|\, \theta_{i,j} = \theta_{j,i},\;\; i = 1...r, j = 1...c\};\;\; r = c$.
- Point Symmetry: $H = \{\theta \in \Theta \,|\, \theta_{i,j} = \theta_{r-i+1,c-j+1},\;\; i = 1...r, j = 1...c\}$.

3.1. FBST formulation on tests for symmetry

We assume that the frequencies X follow a Multinomial distribution, and consider that θ follows, a priori, a Dirichlet distribution with parameters (in matrix form) $\dot{X} = (\dot{x}_{i,j}), i = 1...r, j = 1...c$:

$$\begin{aligned} M(X|n,\theta) &= n!/\prod_{i,j} x_{i,j}! \prod_{i,j} \theta_{i,j}{}^{x_{i,j}} \,, \\ D(\theta|\dot{X}) &= \Gamma\left(\sum_{i,j} \dot{x}_{i,j}\right) / \prod_{i,j} \Gamma(\dot{x}_{i,j}) \prod_{i,j} \theta_{i,j}{}^{\dot{x}_{i,j}-1} \end{aligned}$$

Thus, the posterior distribution of θ is a Dirichlet with parameters $(\ddot{x}_{i,j})$, where $\ddot{x}_{i,j} = x_{i,j} + \dot{x}_{i,j}, i = 1...r, j = 1...c$:

$$f_x(\theta) = \Gamma\left(\sum_{i,j} \ddot{x}_{i,j}\right) / \prod_{i,j} \Gamma(\ddot{x}_{i,j}) \prod_{i,j} \theta_{i,j}{}^{\ddot{x}_{i,j}-1}$$

The (non-informative) uniform prior is given by $\dot{x} = \mathbf{1}$. $\ddot{S}$ denotes the sum $\ddot{S} = \sum_{i,j} \; \ddot{x}_{i,j}$.

Within this framework, the maximum a posteriori (MAP) estimates θ^* under diagonal and point symmetry hypotheses are:

- Diagonal Symmetry: $\theta^*_{i,j} = \ddot{x}_{i,j} + \ddot{x}_{j,i})/2(\ddot{S} - rc)$
- Point Symmetry: $\theta^*_{i,j} = (\ddot{x}_{i,j} + \ddot{x}_{r-i+1,c-j+1})/2(\ddot{S} - rc)$

Observe that, with the uniform prior, the MAP estimate θ^* is equal to the maximum likelihood (ML) estimate for θ under the assumption that the hypothesis is true, and in this work we denote θ^* indistinctly as a MAP and ML estimate.

The integration step of FBST may be performed by generating a set of of M points $\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(M)}\}$ with Dirichlet distribution of parameter vector $(\ddot{x})$, and computing the percentage of points with posterior density greater than $f_x(\theta^*)$:

$$\theta^{(k)} \;\sim\; D(\ddot{x}),\; k = 1, 2, \ldots, M$$

$$\overline{\text{Ev}}(H) \;=\; \sum_{k=1}^{M} I\left(f_x(\theta^{(k)}) > f_x(\theta^*)\right)/M$$

where $I(s) = 1$ if s is true and 0 otherwise. A more efficient Monte Carlo method for the integration step, with control of precision, is presented in [8].

3.2. The Power-Divergence statistic

In this work, we use the power-divergence statistic as a benchmark to evaluate the FBST performance. We briefly introduce this approach, in the context of contingency tables. We denote θ^* as the ML estimate for θ under the assumption that hypothesis is true, and $n = \sum_{i,j} x_{i,j}$.

Cressie and Read [4] introduced the *power-divergence family* of goodness-of-fit statistics,

$$2nI^{\lambda}(x/n : \theta^*) = \frac{2}{\lambda(\lambda+1)} \sum_{i,j} x_{i,j} \left[\left(\frac{x_{i,j}}{n\theta^*_{i,j}} \right)^{\lambda} - 1 \right]; \quad -\infty < \lambda < \infty \tag{1}$$

where λ is the family parameter. The term *power divergence* describes the fact that the statistic $2nI^{\lambda}(x/n : \theta^*)$ measures the divergence of x/n from θ^* through a (weighted) sum of powers of the terms $x_{i,j}/n\theta^*_{i,j}$ for $i = 1 \ldots r, j = 1 \ldots c$. Although the equation (1) is not defined for $\lambda = -1$ or $\lambda = 0$, the power-divergence statistic for these cases is defined by the continuous limits of Equation (1) as $\lambda \to -1$ and $\lambda \to 0$. Under the hypothesis H, power divergence statistics are Asymptotically distributed as Pearson's χ^2. Cases of particular interest are the following statistics:

$\lambda = 1$	$\chi^2 = \sum_{i,j}(x_{i,j} - n\theta^*_{i,j})^2/(n\theta^*_{i,j})$	chi-square
$\lambda = 0$	$G^2 = 2\sum_{i,j} x_{i,j} log(x_{i,j}/n\theta^*_{i,j})$	max. log-likelihood ratio
$\lambda = -1/2$	$F^2 = 4\sum_{i,j}(\sqrt{x_{i,j}} - \sqrt{n\theta^*_{i,j}})^2$	Freeman-Tukey
$\lambda = -1$	$GM^2 = 2\sum_{i,j} n\theta^*_{i,j} log(n\theta^*_{i,j}/x_{i,j})$	modified log-likelihood ratio
$\lambda = -2$	$NM^2 = \sum_{i,j}(x_{i,j} - n\theta^*_{i,j})^2/x_{i,j}$	Neyman-modified chi-square

Besides these standard values of λ, the numerical experiments described in next section include also $\lambda = 2/3$ (denoted by $L2/3$ in the figures), which is recommended as a good alternative, see [1, p.40-41].

4. NUMERICAL EXPERIMENTS AND DISCUSSION

In order to evaluate the performance of the alternative tests, we present in this section some numerical experiments. We use two examples for diagonal symmetry and two more for point symmetry. The two examples of diagonal symmetry are standard benchmarks in the statistical literature, first presented in [9, p.36]; they are also presented in

[2]. The two examples of point symmetry are adapted from mechanical vibration experiments at the engineering school of University of Sao Paulo. Tables 1 and 2 present these examples.

Table 1: Examples used for diagonal symmetry experiments

(a)

	Speckled wood		
Ringlet	Absent	Occasional	Common
Absent	10	4	5
Occasional	7	11	16
Common	3	8	43

(b)

	Speckled wood		
Ringlet	Absent	Occasional	Common
Absent	105	18	6
Occasional	27	5	5
Common	9	5	5

Table 2: Examples used for point symmetry experiments

(a)

	A	B	C	D	E
1	15	11	15	7	13
2	17	8	35	5	11
3	9	4	23	17	10

(b)

	A	B	C	D
1	87	33	32	63
2	104	18	118	39
3	47	124	32	112
4	82	27	49	77

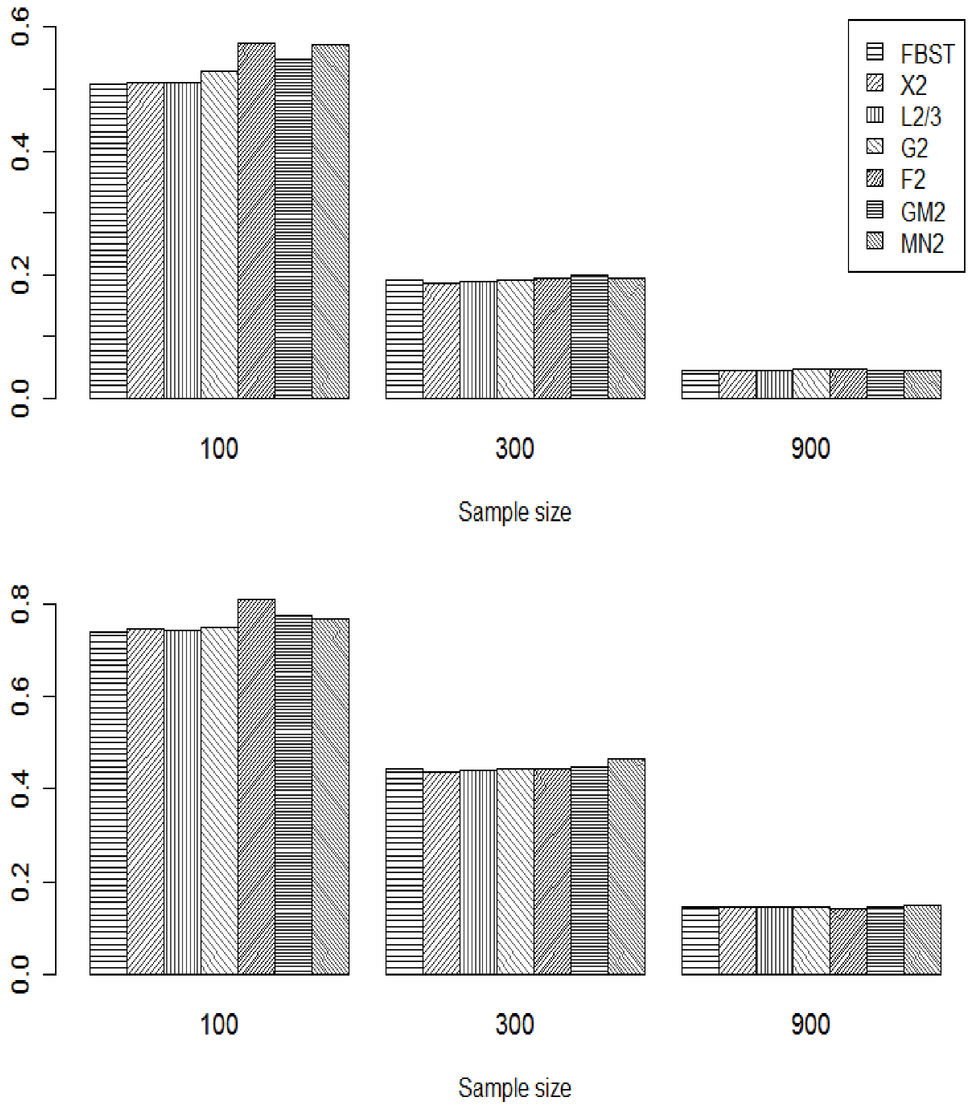

Figure 1: Type II errors of the alternative tests for diagonal symmetry experiments; original samples from Table 1a (top) and 1b (bottom).

In this work, we established an empirical Type I error (rejection rate for a true hypothesis) of 5%, thus calibrating the threshold for acceptance/rejection of the hypothesis for each alternative test. So, our interest is to compare the Type II errors (acceptance rate for a false hypothesis) for increasing sample sizes. In order to estimate the type I and type II errors, for each sample size n we simulated two collections of 1000 samples.

The first collection consists of samples drawn under the hypothesis, i.e., each sample is drawn with a multinomial distribution with parameters (n, θ^*).

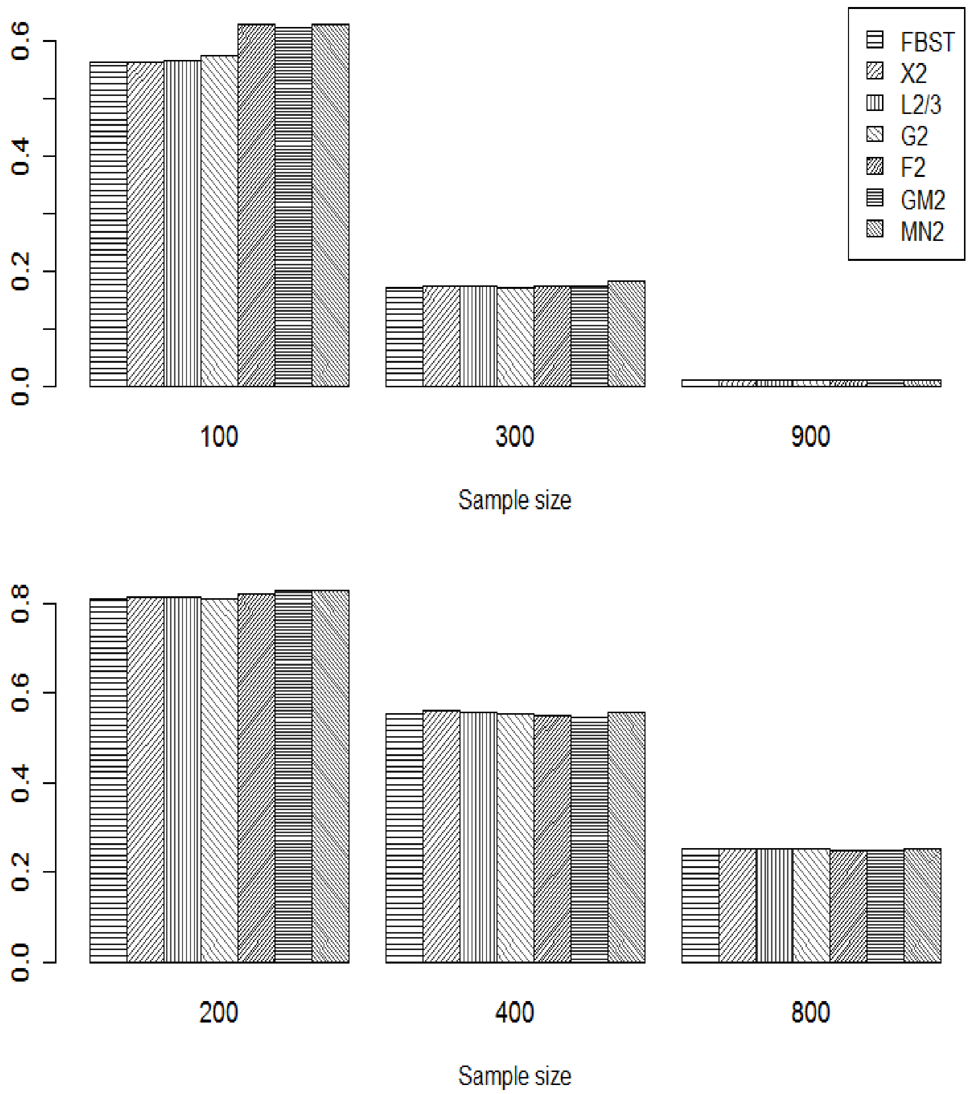

Figure 2: Type II errors of the alternative tests for point symmetry experiments; original samples from Table 2a (top) and 2b (bottom).

The second collection consists of samples drawn with a multinomial distribution with parameters $(n, \theta^{(h)})$, where $\theta^{(h)}$ is drawn from the posterior distribution, that is, each sampling iteration for the second collection is performed in two steps:
(a) draw $\theta^{(h)} \sim D(\ddot{x})$; (b) draw $x^{(h)} \sim M(n, \theta^{(h)})$.

The numerical results for the experiments with diagonal symmetry and point symmetry are presented in Figures 1 and 2, respectively. The sample sizes have been adjusted in log-linear scales. For the diagonal symmetry problems (Figure 1), the tests FBST, X2 and L2/3 are the best performers, in special for small sample sizes ($n = 100$). For the point symmetry problems (Figure 2), the relative performance of tests based on power-divergence are less stable: for small sample sizes ($n = 100, 200$), the best performers for examples 2a and 2b are, respectively: FBST and X2 (top); and FBST and G2 (bottom); for moderate sample sizes ($n = 300, 400$), the best performers are: FBST, G2 (top); and F2, GM2 (bottom). It is important to notice that, despite their slightly better performance in this case, F2 and GM2 perform much worse in other cases.

In the four examples presented in detail in our numerical experiments, and in many more we used as benchmarks, the FBST emerges as a very stable and strong performer for problems concerning tests of symmetry in contingency tables. The FBST seems to have a very stable behavior performing, at most benchmarks, better than many tests in the power divergence family, and as well as or very close to the best test in the power divergence family for the example at hand, that is, the optimal λ for that specific example. Since it is in general very hard to guess in advance the optimal parameter λ^* for a specific example or application, these conclusions suggest that the FBST is a very good alternative for this class of problems.

ACKNOWLEDGEMENTS

The authors are grateful for the support of EACH-USP and IME-USP - the School of Arts and Sciences and the Institute of Mathematics and Statistics of the University of São Paulo, CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, and FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo.

REFERENCES

1. T.R.C.Read and N.A.C.Cressie (1988). Goodness-of-Fit Statistics for Discrete Multivariate Data. NY: Springer-Verlag.
2. S.Tomizawa, K.Yamamoto and K.Tahata (2007). An Entropy Measure of Departure from Point-Symmetry for Two-Way Contingency Tables. Symmetry: Culture and Science; Journal of International Symmetry Association 18, 279-297.
3. K.D.Wall and G.A.Lienert (1976). A Test for Point-Symmetry in J-dimensional Contingency Cubes. Biom. J. 18, 259-264.
4. A.H.Bowker (1948). A test for symmetry in contingency tables. J Am Stat Assoc 43(244), 572-574.
5. N.A.C.Cressie and T.R.C.Read (1984). Multinomial Goodness-of-Fit Tests. J R Stat Soc B 46(3), 440-464.
6. C.A.B.Pereira and J.M.Stern (1999). Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy,* 1, 99–110.
7. C.A.P.Pereira, J.M.Stern and S.Wechsler (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis,* 3, 1, 79–100.
8. M.S.Lauretto, C.A.B.Pereira, J.M.Stern and S.Zacks (2003). Full Bayesian significance test applied to multivariate normal structure models. Braz J Prob Stat 17, 147-168.
9. G.J.G.Upton and B.Fingleton (1989). Spatial Data Analysis by Example, Vol. 2. Wiley, Chichester.