RT-MAE-8801

ON THE CONCEPT OF P-VALUE

by

Carlos A.B. Pereira

and

Sérgio Wechsler

Palavras Chaves: Bayés factor; likelihood principle, null
(Key words) and alternative hypotheses, p-value, P-value, significance tests, weighted likelihood ratio.

Classificação AMS: 62A15
(AMS Classification) 62F03

- Janeiro de 1988 -

ON THE CONCEPT OF P-VALUE

Carlos Alberto de Bragança Pereira &
Sérgio Wechsler
Universidade de São Paulo, BRAZIL

Present Address:
Operations Research Center
University of California
3115 Etcheverry Hall
Berkeley, CA 94720

ABSTRACT: Simple examples illustrate how misleading a p-value constructed with no regard to the alternative hypothesis can be. A p-value which regards the alternative hypothesis, called here P-value, is precisely defined. It is shown that the use of the P-value avoids the kind of inconsistencies illustrated by the examples. Although P-values could be considered useless by Bayesians, the use of prior distributions (to obtain weighted likelihoods) is a way by which classical statisticians could regard alternative composite hypotheses when performing significance tests.

KEY WORDS: Bayes Factor, Likelihood Principle, Null and alternative hypotheses, p-value, P-value, significance tests, weighted likelihood ratio.

1. INTRODUCTION

Significance testing is a widespread statistical procedure in scientific studies. It consists of the computation of a p-value and then of the judgement of the plausibility of a null hypothesis H. Unfortunately, some usual constructions of p-values completely disregard the alternative hypothesis, A. The present note aims to show, by simple examples, how wrong the conclusions based on such misleading p-values can be. A p-value which regards A, called here P-value, will be precisely defined in Section 4. The P-value is a well-defined (even if not a useful) quantity under the Bayesian view. It is also well-defined under Fisherian and Neyman-Pearson-Wald schools when both H and A are simple hypotheses.

Discussions on the concept of p-values have an extremely large reference list - see Good (1983). Most recently, Casella & Berger (1987) and Berger & Sellke (1987) contributed to this list but little additional light was brought to the subject, as pointed out by Pratt (1987) in the comments following these papers. Obviously, after observing data x and having defined a subset C of the parameter space, the numbers $Pr\{C|x\}$ and $Pr\{x|C\}$ have completely different meanings. Therefore, comparison of values of posterior probabilities to p-values (as done by Casella & Berger, 1987) seems to be a very confusing matter. Another confusing aspect of significance tests appears when relations between sample sizes and p-values are made. The recipe of Lindley & Scott (1986) is contrary to the recipe of Peto et al. (1976), as pointed out by Royall (1986). This is briefly discussed in Section 6.

The object of the present note is to illustrate practical aspects of the philosophical point of view of I.J. Good on p-values (Good, 1983). The authors believe that Good's work completely illuminates the matter. However, p-values are still not being carefully presented for users of Statistics (see Miettinen, 1985 ch.9 and Peto et al., 1976).

References supporting the use of significance tests are Cox (1977), Cox & Hinkley (1974), and Kempthorne & Folks (1971). That one should disregard A when constructing significance tests is not suggested in these texts. In fact, as

Spjøtvoll (1977) remarks in his discussion on Cox (1977), explicit regard to A should be emphasized whenever p-values are being defined.

In many cases the p-value (or even the P-value) turns out to be a tail area. This fact is so common that p-values are often called tail areas. That this nomenclature is unfortunate is shown in Section 4 where we obtain a P-value which is the sum of three areas, one of which is central while the remaining are tail. A P-value that is the sum of three areas was already presented by Good (1983) and called "triple tail" since none of those areas were central.

2. DEFINITION OF p-VALUE

The intuitive notion of p-value is captured by the following definition. Consider an experiment producing data **x**, an observation of **X**, to test **H** versus **A**.

> **Definition 1:** The p-value is the probability, under **H**, of the event composed by all sample points that favor **A** (against **H**) at least as much as **x** does.

However, many textbooks and papers present definitions which make no explicit reference to A (Cox, 1977) or definitions which completely ignore A (Freedman, Pisani, & Purves, 1978 or Pratt & Gibbons, 1981). Typical examples of such definitions are presented next, respectively. Definition 2 is in the direction of Mosteller & Rourke (1973) and Definition 3 is given in Berger & Sellke (1987).

> **Definition 2:** The p-value is the probability, under **H**, of the event composed by all sample points at least as "extreme" as **x** is.

Let now T=T(X) be a statistic for which large values cast doubt on **H** and at point **x** it takes the value t=T(x).

Definition 3: The p-value at point t is the probability

$$p(t) = Pr\{T \geq t \,|H\}.$$

Definition 2 presupposes an ordering of the sample points which gives meaning to "extreme." If the ordering regards A, Definitions 1 and 2 are essentially the same. However, if the ordering disregards A, Definitions 2 and 3 are essentially the same and the p-values can completely ruin the statistical analysis as shown by simple examples such as the ones presented in the following section. It must be remarked that although Berger & Wolpert (1984) present Definition 3, in their comments they prescribe that, at least informally, A must be regarded. A formal working definition of the kind of Definition 1 is introduced in Section 4.

3. EXAMPLES

This section presents three examples of misleading conclusions obtained by the use of Definition 3, the one that completely disregards A. All the examples concern simple H versus simple A. To stress the point of the present note simple examples were chosen. However, more standard examples could be constructed at the price of sacrificing simplicity. Section 5 discusses how to deal with composite hypotheses when constructing P-values.

Example A: Consider an urn containing exactly three marbles: one black, one white, and one green. Three marbles were randomly selected from this urn. Consider the following two hypotheses:

H: The selection was done with replacement.

versus

A: The selection was done without replacement.

Suppose the data consist of the vector $X=(X_1,X_2)$, where X_1= number of black marbles in the sample and X_2= number of white marbles in the sample. Note that the null probability function of X (multiplied by 27) can be displayed as

x_2 \ x_1	0	1	2	3
3	1			
2	3	3		
1	3	6	3	
0	1	3	3	1

Table 1: Null Probability Function
(times 27) of X in Example A

Suppose that the point (1,1) is observed. A person who disregards A will construct a p-value using Definition 3 (or Definition 2 under an ordering which disregards A). For any statistic T which disregards A, such a p-value at $t = T(1,1)$ will be $p(t)=1$. For instance, if the χ^2 statistic $T=(X_1-1)^2+(X_2-1)^2$ is used, then the p-value will be $p(0)=1$ since $T(1,1)$. By disregarding A, no point is more supportive of H than $x=(1,1)$, the vector of expected frequencies under H. Also note that, under H, all other sample points have smaller probability (are more "extreme") than $x=(1,1)$ has. Hence, any ordering equivalent to the probability ordering under H (e.g., the χ^2 statistic above) will produce the unity as the p-value. H would be rejected only upon observation of a point other than $x=(1,1)$.

Yet the probability of $x=(1,1)$ under A is one! The conclusion here is that one rejects H (accepts A) only when observing points which are impossible under A!

Example B: Let X be a normal variable with mean zero and unknown variance σ^2 producing data x. A minimal sufficient statistic here is $T=T(X)=X^2$ with $t = x^2$. Using Definition 3 to test H: $\sigma=2$, one will compute the p-value as the tail area

$$p(t) = Pr\{T \geq t \mid \sigma=2\} = 2\Phi(-\frac{\sqrt{t}}{2})$$

where Φ is the standard normal distribution function.

Now let the alternative hypothesis be $A{:}\sigma=1$. Clearly, the tail area under A is smaller than the one under H. Indeed,

$$p_A(t) = Pr\{T \geq t \mid A\} = 2\Phi(-\sqrt{t}) < 2\Phi(-\frac{\sqrt{t}}{2}) = p(t)\,.$$

Hence, small values of p(t) correspond to even smaller values of $p_A(t)$ and favor H, not A. For instance, by following the decision procedure prescribed by Burdette & Gehan (1970), if $t = x^2 = 16$, one may wrongly state that there is moderate evidence against H since

$$p(16) = .0454 = < .05\,.$$

However, since $p_A(16) \cong 0$, the evidence against A is much stronger (see Figure 1).

Example C: Let f be the density of a random variable X from which an observation x is obtained. Consider the significance test of "H: f is normal with zero mean and variance 4" versus "A: f is the standard Cauchy density." Suppose again that $t = x^2 = 16$. Then, as in Example B, the p-value is p(16) = .0454. Note that even with Example B having a different hypothesis A, the p-value has exactly the same value. Of course, this is because the alternative hypotheses were disregarded.

It must be pointed out that contrary to Example B, here the tail area under A is $p_A(16) = .156$, which is greater than the p-value. However, as we shall see in Section 5, the value t=16 again highly supports H, <u>not</u> A.

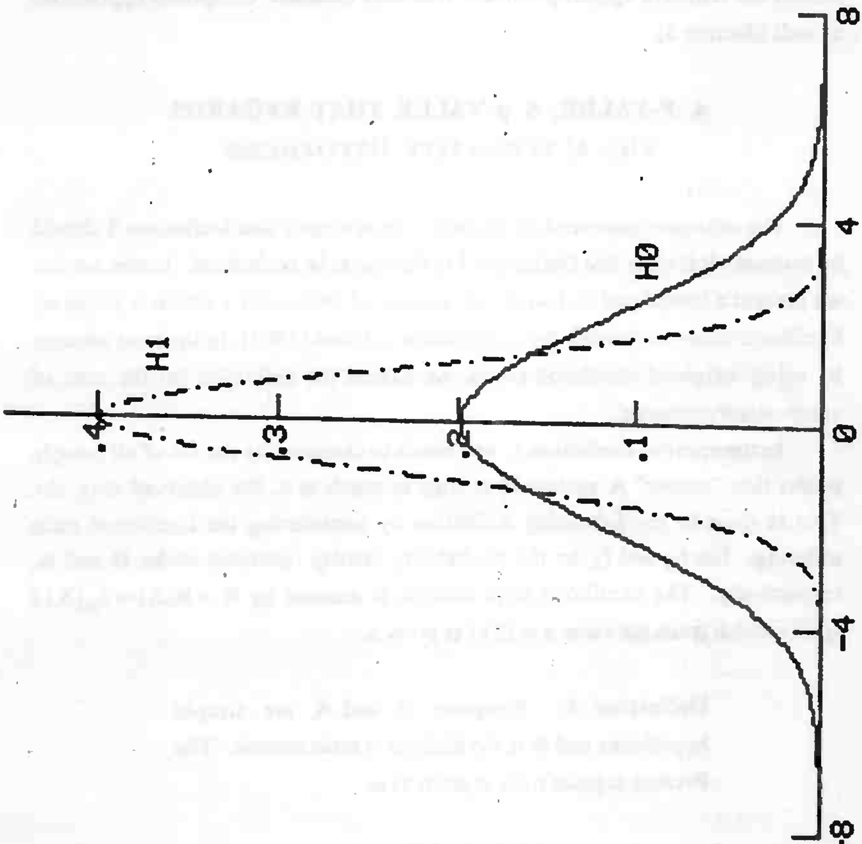

FIGURE 1: Standard Normal density (H1) and Normal density with parameters 0 and 4 (H0)

In both Examples B and C, the sample size is n=1. This is not a restriction of our point and it was used just for simplicity. The same inconsistencies occur in more realistic situations where n is large. Yet, simple hypotheses do not restrict the criticism against p-values. One may consider composite hypotheses as well (Section 5).

4. P-VALUE, A p-VALUE THAT REGARDS THE ALTERNATIVE HYPOTHESIS

The examples presented in Section 3 show clearly that Definition 3 should be avoided. It is clear that Definition 1 is the one to be considered. In this section we present a formalized and workable version of Definition 1 which is based on likelihood ratios, following the suggestions of Good (1983). In the next section, by using weighted likelihood ratios, we extend the definition for the case of composite hypotheses.

In the spirit of Definition 1, one needs to characterize the set of all sample points that "favour" A against H at least as much as x, the observed data, do. This is done in the following definition by considering the likelihood ratio ordering. Let f_H and f_A be the probability density functions under H and A, respectively. The likelihood ratio statistic is denoted by $R = R(X) = f_A(X) / f_H(X)$ which takes the value $r = R(x)$ at point x.

Definition 4: Suppose H and A are simple hypotheses and R is the likelihood ratio statistic. The P-value at point r (or at point x) is

$$P(r) = \Pr\{R \geq r \mid H\}.$$

The following result shows the consistency of P-values. On the other hand, p-values lack this kind of consistency.

Lemma: For any positive **r**,

$$P_A(r) = Pr\{R \geq r \mid H\} \geq P(r).$$

A proof of this result would follow the standard steps of Theorem 1 on page 443 of DeGroot (1986).

Now we apply the P-value to the examples of Section 3. The conclusions will now be coherent and contrary to those obtained when p-values were used.

Example A: (continuation) The statistic **R** takes the value 27/6 at point (1,1) and 0 at any other point (x_1,x_2). If point (1,1) has been observed, the P-value is

$$P(27/6) = Pr\{R \geq 27/6 \mid H\} = 6/27 < 1 = p(1,1).$$

Note that for any other sample point, the P-value is the unity, supporting **H** and rejecting, indeed, **A**.

Example B: (continuation) Recall that **H**: $\sigma = 2$ and **A**: $\sigma = 1$, and note that $\{x;\ R(x) \geq R(4)\} = \{x;\ T(x) \leq 16\} = \{x;\ -4 \leq x \leq 4\}$. Hence the P-value at point $x=4$ (or $x = -4$) is

$$P(2e^{-6}) = 1-2\Phi(-2) = .9546,$$

the central area, <u>not</u> the tail area p-value. In fact, here P = 1-p .

Example C: (continuation) Recall that **H**: Normal (0,2) and **A**: Cauchy (0,1). These two densities are illustrated in Figure 2. Figure 3 introduces the possible values of R(x). From both figures one understands why $t = x^2 = 16$ is less "extreme" than $t = .04$, although the later has a much higher density than the former.

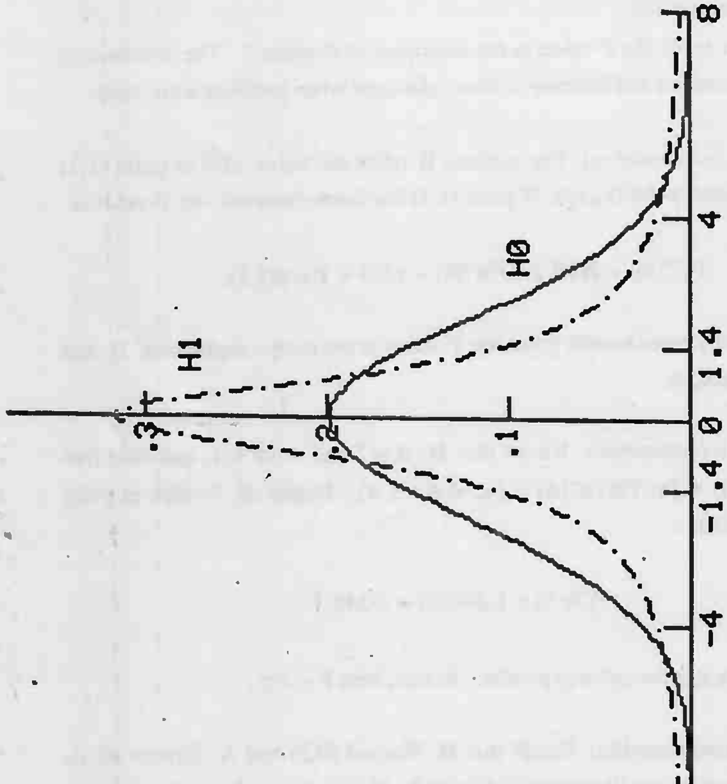

FIGURE 2: Standard Cauchy density (H1) and Normal density with parameters 0 and 4 (H0)

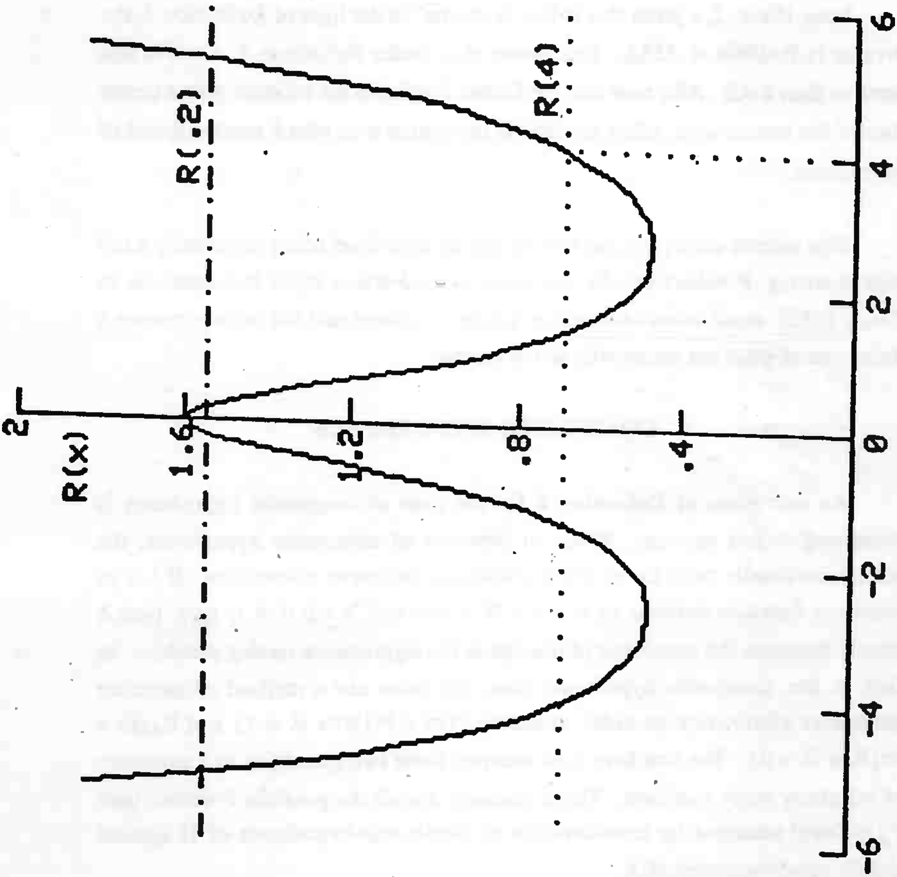

FIGURE 3: Likelihood Ratio, R(x).

If the observed sample point is $x = 4$ (or $t = 16$), then the set $\mathfrak{T} = \{x; R(x) \geq R(4)\}$ is the union of the following three intervals: $(-\infty, -4]$, $[-1.388, 1.388]$, and $[4, \infty)$.

The P-value (area under the Normal (0,2) at set $\mathfrak{T}$) at $x = 4$ is then $P = .5553 > .0454 = p(16)$.

Now, if $x = .2$, a point that is less "extreme" in the light of Definition 3, the P-value is $P=.0904 < .5553$. This shows that, under Definition 4, $x = 4$ is less extreme than $x =.2$. Also note that the former P-value is the tail area plus a center slice of the central area, while the later is the central area plus a small sub-tail of the tail area.

This section shows that the idea of significance level being necessarily a tail area is wrong. P-values can be tail areas, central areas, triple tail areas (as in Good, 1983), triple areas obtained as a sum of central and tail areas, or even a finite sum of areas not necessarily tail or central.

5. COMPOSITE HYPOTHESES

An extension of Definition 4 for the case of composite hypotheses is presented in this section. When in presence of composite hypotheses, the statistician usually faces the problem of nuisance parameter elimination. If λ is an indicator function defined as $\lambda = 1$ if H is true and $\lambda = 0$ if A is true, then λ clearly becomes the parameter of interest in the significance testing problem. In fact, in the composite hypotheses case, one must use a method of nuisance parameter elimination in order to define $P(r) = \Pr\{R \geq r \mid \lambda = 1\}$ and $P_A(r) = \Pr\{R \geq r \mid \lambda = 0\}$. The aim here is to interpret these two quantities as a summary of infinitely many numbers. Those numbers are all the possible P-values (and P_A-values) obtained by consideration of simple sub-hypotheses of H against simple sub-hypotheses of A.

The partial likelihood approach is a celebrated way of treating composite hypotheses. For example, the Fisher exact test is an important example of the use of such an approach. However, as shown in Irony & Pereira (1986), it is possible to improve Fisher's solution by considering a Bayesian tool for a correct P-value evaluation. This Bayesian tool is now used to extend Definition 4 to the composite hypotheses case.

Over the two sets of parametric points that characterize H and A consider respectively two probability measures Π_H and Π_A. These two measures define two weighting systems that can be interpreted as orderings of preference among parametric points in H and A, respectively. Now, for every possible observation x, define the weighted likelihood values $f_H(x)$ and $f_A(x)$ which are the weighted averages of the likelihood function under H and A, respectively. Also, define the weighted likelihood ratio statistic $\underline{R} = \underline{R}(X) = f_A(X) / f_H(X)$ which takes the value $\underline{r} = \underline{R}(x)$ at point x.

Definition 5: Suppose H and A are composite hypotheses and $\underline{R}$ is the weighted likelihood ratio statistic. The P-value at point $\underline{r}$ (or at point x) is

$$P(\underline{r}) = \Pr\{\underline{R} \geq \underline{r} \mid H\} = \int I_{\underline{r}}(x) f_H(x) \partial x,$$

where $I_{\underline{r}}(x)$ is the indicator function of the set $\{x; \underline{R}(x) \geq \underline{r}\}$.

It is interesting to note that, by considering the ordering defined by $\underline{R}$, it would be possible to compute the probability of $\{\underline{R} \geq \underline{r}\}$ (a p-value) for all elements of H. The above P-value is the weighted average (using Π_H) of these probabilities. Analogously, we could characterize the P_A-values using A and Π_A. Consequently, the property included in the Lemma of Section 4 still holds if we consider f_H and f_A as the sample models under H and A, respectively. Now, the situation of Example B (and C) is considered under a more realistic situation of composite hypotheses.

Example D: In Example B, let A be composite. It is not difficult to see that if A: $\sigma < 2$, then the P-value is a central area. Analogously, if A: $\sigma > 2$, then the P-value is a tail area. The interesting question is what form the P-value has when A: $\sigma \neq 2$. For simplicity consider that Π_H is degenerated at point 2 and Π_A is characterized by a one-degree-of-freedom χ^2 (prior) density for $1/\sigma^2$. Simple integration shows that f_A is the standard Cauchy density and therefore the computation of the P-value is reduced to the computation of the P-value of Example C where a triple area can be obtained.

To end this section, our version of the Fisher significance test for comparing proportions is presented. Note that, although with different dimensions, both hypotheses are composite. Basu (1979) and Pereira & Lindley (1987) present examples questioning the partial likelihood method used in the Fisher significance test. The same examples could be used to support the P-value of Definition 5.

Example E: Let $\mathbf{X} = (X_1,X_2)$ where X_1 and X_2 are independent binomial with probability of success p_i and sample size n_i, i=1,2. After observing data $\mathbf{x} = (x_1,x_2)$ suppose one wants to evaluate the P-value for **H:** $p_1 = p_2$ versus **A:** $p_1 \neq p_2$. These two hypotheses are illustrated in Figure 4.

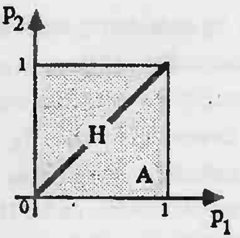

Figure 4: Hypotheses H and A in Example D

If Π_H is uniform in the line H and Π_A is uniform in the set A, then, considering $y = x_1+x_2$ and $N = n_1+n_2$, we have

$$f_H = \frac{\binom{y}{x_1}\binom{N-y}{n_1-x_1}}{\binom{N}{n_1}} \frac{1}{(N+1)},$$

and

$$f_A = \frac{1}{(n_1+1)(n_2+1)} .$$

For the case of $n_1 = n_2 = 5$, Table 2 displays all possible values of $\underline{R}(x_1,x_2)$ divided by 77. To obtain f_H it is enough to divide the inverted values in the table by 2772. Suppose that $\mathbf{x} = (4,1)$ is observed. Then $\underline{r} = 77/25$ and P = .0577. Using now Fisher significance procedure we have p = .2063. Note that p is 4 times the value of P and that Irony & Pereira (1986) show by simulation that the frequency of more extreme (ordering of Table 2) points than $\mathbf{x}$, under H, is indeed much smaller than p. On the other hand, this frequency is very close to P.

x_2 \ x_1	0	1	2	3	4	5
0	1/252	1/126	1/56	1/21	1/6	1
1	1/126	1/140	1/105	1/60	1/25	1/6
2	1/56	1/105	1/120	1/100	1/60	1/21
3	1/21	1/60	1/100	1/120	1/105	1/56
4	1/6	1/25	1/60	1/105	1/140	1/126
5	1	1/6	1/21	1/56	1/126	1/252

Table 2: Values of $\underline{R}(\mathbf{x})$ (Divided by 77) in Example E

Note that Definition 5 does make use of a prior probability over the parametric space. Then, by considering the full Bayesian probability space, P and P_A are in fact well-defined conditional probabilities. Therefore the computation of these quantities is a non-problem for a Bayesian. However, its use in significance testing may be questioned by Bayesians since it is believed to violate the Likelihood Principle (Berger & Wolpert, 1984, p. 105).

6. P-VALUES: A BAYESIAN LOOK

In this section we discuss the inappropriateness of the use of P-values, as a "measure of improbability" of H, for Bayesians.

Rejecting H whenever the Bayes' factor, B, (the ratio of the posterior probabilities of A and H) is large, is equivalent to rejecting H whenever the P-value is small. In other words, for every $c > 0$, there exists an $\alpha > 0$ such that the event $\{B > c\}$ is equivalent to the event $\{P < \alpha\}$. The constant α depends on the prior probability of H, on the constant c, and on the sampling model. In fact it can be proved that

$$\alpha = 1 - F_R\left(\frac{\pi c}{1-\pi}\right) = P\left(\frac{\pi c}{1-\pi}\right),$$

where F_R is the conditional distribution function of the statistic R given H, and π is the prior probability of H.

In this sense, comparison of the actual P-value to α can be viewed as a mere computational option in the implementation of a Bayes' test. Note that the P-values and consequently α may change under a different sampling model, even when the Bayes' factor remains unchanged. Therefore, the use of the P-value as a "measure of improbability (Jeffreys ,1961) of H on the actual data" is clearly a violation of the Likelihood Principle. Furthermore, using P-values in hypothesis testing without any regard to the corresponding Bayes' test (which produces the value of the corrected significance level $P(\frac{\pi c}{1-\pi})$) again violates the Likelihood Principle and is consequently unacceptable to Bayesians. One should also keep in mind that p-values (not P-values) are unacceptable even as computational tools, for the above kind of equivalence with Bayes' tests no longer holds. Recall the examples of Section 3.

To add more confusion to the subject, Royall (1986) pointed out that it is listed in important literature (Peto et. al, 1976 and Lindley & Scott, 1984) two opposite prescriptions for the important role of sample sizes in significance testing.

> *"... A given P-value in a large trial is usually stronger evidence that the treatments differ than the same value in a small trial of the same treatments would be."* (Peto et. al, 1976, p. 593)

> *" All significance tests are dubious because the interpretatio to be placed on the phrase 'significant at 5% ' depends on the sample size: it is more indicative of the falsity of the null hypothesis with a small sample than with a large one."* (Lindley & Scott, 1984, p. 1)

To explicate their statement Peto et. al (1976) used a prior probability for H and computed $\frac{\Pr(H \mid \text{significant})}{\Pr(A \mid \text{significant})}$, where significant means that the event $\{T \geq t\}$ has ocurred. Hence, in their argument, they did not use the whole information given by the event $\{T=t\}$. Had they used the correct Bayes' Factor, $\frac{\Pr(H \mid T=t)}{\Pr(A \mid T=t)}$, their conclusion would trivially be the opposite (the one of Lindley & Scott, 1984), as demostrated by DeGroot (1986, pp. 380-1).

7. CONCLUSION

The consideration of A when defining significance levels is not new (although p-values have always been largely used) and we can refer here to de Finetti (1972, p. 163), Good (1983, p. 140), Jeffreys (1961, p. 383), Lindley (1978), and Neyman (1981). The present note just shows how dangerous it is not to consider A in the definition of significance levels. We also present a working definition that does regard A, even in the case of composite hypotheses. Using prior distributions (or equivalently weighting systems), the definition of P-value for the case of composite hypotheses is an extension of the one introduced for the case where both hypotheses are simple. Such a P-value is therefore well-defined for Bayesians who can compute it with no discomfort. Classical statisticians may also consider the P-value (and even use it in the way Burdette and Gehan (1970) uses p-values in significance testing) if they accept the idea of replacing, in the likelihood ratio, maximum of likelihoods with corresponding weighted average likelihoods; i.e., if they accept to use averages for maximas.

Yet, the use of P-values other than for computational purposes is unacceptable for Bayesians, since a procedure based exclusively on them would violate the likelihood principle. Consequently, it would be incoherent for Bayesians. An inferential procedure that does not violate the Likelihood Principle ought to be based only on the observed sample point, not on others (more "extreme") points that could be observed but were not (Basu, 1975). As pointed out in Section 6, however, the computation of a Bayes' Factor can be replaced- probably disavantageously - by the computation of the corresponding P-value, which is to be appreciated only in the light of the Bayes' Factor scale determined by the loss function. Still, this fact does not qualify P-values as Bayesian quantities. Note that Bayes' tests are based only on Bayes' Factors and these do not change with proportional likelihoods. Since P-values do change, we may well have two Bayes' Factors (related to two different models) with equal values corresponding to different P-values.

We close this article with the following quotation from Professor Dennis V. Lindley (1978, p. 5):

> ***One can only judge something in relation to the alternatives - a principle that is often not appreciated either in statistics or in politics. It was a great achievement of Neyman and Pearson to recognize this.***

BIBLIOGRAPHY

BASU, D. (1975). Statistical information and likelihood (with discussion). *Sankhyā*, A 37: 1-71.

BASU, D. (1979). Discussion of Joseph Berkson's paper "In dispraiseof the exact test." *Journal of Statistical Planning and Inference*, 3: 189-92.

BERGER, J.O. & SELLKE, T. (1987). Testing a point null hypothesis: the irreconcilability of P values and evidence. *Journal of the American Statistical Association*, 82: 112-22.

BERGER, J.O. & WOLPERT, R.L. (1984). *The likelihood principle*. Hayward, California: Institute of Mathematical Statistics Monography Series.

BURDETTE, W.J. & GEHAN, E.A. (1970). *Planning and analysis of clinical studies*. Springfield, Illinois: Charles C. Thomas.

CASELLA, G. & BERGER, R.L. (1987). Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association*, 82: 106-11.

COX, D.R. (1977). The role of significance testing (with discussion). *Scandinavian Journal of Statistics*, 4: 49-70.

COX, D.R. & HINKLEY, D.V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

DE FINETTI, B. (1972). *Probability, Induction, and Statistics*. New York, John Wiley.

DEGROOT, M.H. (1986). *Probability and Statistics* (2nd ed.). London: Addison-Wesley.

DICKEY, J.M. & LIENTZ, B.P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Stat.*, 41: 214-66.

FREEDMAN, D., PISANI, R., & PURVES, R. (1978). *Statistics*. New York: Norton.

GIBBONS, J.D. & PRATT, J.W. (1975). P-values: interpretation and methodology. *American Statistician*, 29: 20-25.

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

GOOD, I.J. (1987). Comment on Casella & Berger (1987) and Berger & Sellke (1987). *Journal of the American Statistical Association*, 82: 125-8.

IRONY, T.Z. & PEREIRA, C. A. de B. (1986). Exact tests for equality of two proportions: Fisher v. Bayes. *Journal of Statistical Computation and Simulation*, 25: 93-114.

JEFFREYS, H. (1961). *Theory of Probability* (3rd ed.). London: Oxford University Press.

KEMPTHORNE, O. & FOLKS, L. (1971). *Probability, Statistics, and Data Analysis*. Ames: The Iowa State University Press.

LINDLEY, D.V. (1978). The Bayesian approach (with discussion). *Scandinavian Journal of Statistics*, 5: 1-26.

LINDLEY, D.V. & SCOTT, W.F. (1984). *New Cambridge Statistical Tables*. London: Cambridge University Press.

MIETTINEM, O.S. (1985). *Theoretical Epidemiology: principles of occurrence research in medicine*. New York: John Wiley.

MOSTELLER, F.R. & ROURKE, R.E. (1973). *Sturdy Statistics: nonparametrics and order statistics*. Mass.: Addison-Wesley.

MORRISON, D.E. & HENKEL, R.E. (1970). *The significance test controversy*. Chicago: Aldine Publishing Company.

NEYMAN, J. (1981). Egon Pearson (August 11, 1895 - June 12, 1980). An appreciation. *The Annals of Statistics*, 9: 1-2.

PEREIRA, C.DE B. & LINDLEY, D.V. (1987). Examples questioning the use of partial likelihood. *The Statistician* (to appear).

PETO, R., PIKE, M. C., ARMITAGE, P., BRESLOW, N. E., COX, D. R., HOWARD, S.V., MANTEL, N., MCPHERSON, K., PETO, J., & SMITH, P.G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient, I. Introduction and design. *British Journal of Cancer*, 34: 585-612.

PRATT, J.W. (1987). Comment on Casella & Berger (1987) and Berger & Sellke (1987). *Journal of the American Statistical Association*, 82: 123-5.

PRATT, J.W. & GIBBONS, J.D. (1981). *Concepts of nonparametric theory*. New York: Springer-Verlag New York Inc.

ROYALL, R.M. (1986). *The effect of sample size on the meaning of significance tests*. Technical Report #587. Department of Biostatistics, John Hopkins University. Washington, DC.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; Inverse Gaussian Distribution: A Bayesian Approach, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; On the Hardy-Weinberg Equilibrium in Generalized ABO Systems, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; Generalizing Importance of Components for Multistate Monotone Systems, São Paulo, IME-USP, 1987, 12p.