Some Measures of Support for

Statistical Hypotheses

M.Emadi and N.R.Arghami

Department of Statistics, School of Mathematical Sciences,

Ferdowsi University, Mashhad, Iran.

Abstract. Inadequacy of p-value as a measure of support for a statistical hypothesis provided by data has been pointed out by many authors. Likelihood ratio as such a measure, proposed by Richard Royall (2000), has an intuitive appeal and the advantage of being very simple. In this paper, however, we first state some properties that we want such a measure to have and propose an optimality criterion. Then it is shown that although a continuous optimal measure does not exist, there exists a continuous measure that has many desirable properties and is nearly optimal.

Keywords. P-value, Likelihood Ratio, Law of likelihood, Evidential Statistics.

1 Introduction

P-value, if interpreted correctly, is a measure of evidence that the data provide in favor of the null hypothesis. The fact that p-value is very often misinterpreted as "the probability of type I error" and misused as "the probability that the null hypothesis is true", has been mentioned by many authors. Using p-value as "a measure of the amount of evidence provided by the data in favor of the null hypothesis", however, is very common, has been accepted by many notable statisticians (see for instance Lehman (1975)) and has been discussed by many authors (Robins etal (2000) Bayarri, M.J and Berger, J.O. (2000), Schervish, M.J. (1996), Berger, J.O. and Se- llke, T. (1987), Casella, G. and Berger, R.L. (1987), Degroot, M.H. (1973)).

Although in the case of point null hypothesis (against the complement hypothesis), p-value (suitably calibrated) is an acceptable measure of evidence (Berger, J.O. etal (2001)), it is not recommended for the cases in which both hypotheses are simple. For this case Richard Royall (2000), relying on the "law of likelihood" and "likelihood principle", forcibly argues that such a measure should only be the likelihood ratio (or a continuous and strictly increasing transformation of it). The argument is quite convincing, but since it is only based on a principle, one cannot rest assure unless its distributional properties are investigated and its optimality (in some sense) is compared with other intuitively appealing measures, even though they may not obey likelihood principle. In Section one, some desirable properties for such measures are mentioned and an optimality criterion is introduced and shown that there does not exist an optimal measure that satisfies the required conditions. In Section 3 a measure of support for the null hypothesis is introduced which is nearly optimal, has all the desirable properties and is more optimal than (a strictly increasing function of) likelihood ratio. In both sections it is assumed that both hypotheses are simple and R (the likelihood ratio) is a continuous random variable under both hypothesis. Our recomended measure of support is introduced in Section 4. Section 5 contains some figures.

2 Desirable properties of a measure of support

Let random vector $\underline{X}$ have densities $f_1(\underline{x})$ and $f_1(\underline{x})$ (both completely know) under H_1 and H_2 respectively. We are going to look for a function η of $r\left(=\frac{f_1(\underline{x})}{f_2(\underline{x})}\right)$ that can be used as a measure of support for H_1 against H_2 provided by observation $\underline{X}=\underline{x}$ and

(p_1) takes values in the unit interval [0, 1],

(p_2) is a continuous function of the likelihood ratio *r*,

(p_3) is strictly increasing in *r*,

(p_4) is symmetric with respect to H_1 and H_2,

(p_5) is consistent,

(p_6) has small probability of being strongly misleading,

Property p_1 is not really essential because one can always use a simple strictly increasing transformation to make the measure take values in the interval $[0,\infty]$ or $[-\infty,\infty]$, without loosing any of the other properties. Properties p_2 through p_4 are essential because their absence, we believe, would be counter intuitive.

We say η is consistent if under $H_1:\eta \xrightarrow{p} 1$ and under $H_2:\eta \xrightarrow{p} 0$.

Property p_4 means that

$$\begin{aligned}\eta\ (H_1,H_2) &= \textit{the amount support provided by the data for } H_1 \textit{ against } H_2\\ &= 1-\textit{the amount support provided by the data for } H_2 \textit{ against } H_1\\ &= \eta\ (H_2,H_1)\end{aligned}$$

We mention in passing that p-value has properties p_1 through p_3.

Property p_6 means that we want η to have small probability of taking small values (values near zero) under H_1 and also small probability of taking large values (values near 1) under H_2. In other words if we denote the density of η under H_i by k_i, $(i=1,2)$ and denote the cdf of η under H_i by k_i, $(i=1,2)$, we want their plots to be as in Figures 1 and 2 respectively.

It is evident that probability of strong misleading evidence is smaller, the farther the two graphs in Figure 2 are apart. So we take our optimality criterion to be the **a**rea **b**etween the **c**urves of cdf's of η under H_1, H_2 and denote it by $abc(\eta)$. We can write,

$$\begin{aligned} abc(\eta) &= \int_0^1 K_2(\eta)d\eta - \int_0^1 K_1(\eta)d\eta \\ &= \int_0^1 [1-K_1(\eta)]d\eta - \int_0^1 [1-K_2(\eta)]d\eta \\ &= \mathrm{E}_1(\eta) - \mathrm{E}_2(\eta), \end{aligned}$$

where $\mathrm{E}_i(\eta)$ denotes the expected value of η under H_i, $i=1, 2$.

As an example, the curves of the cdf's of p-value are shown in Fig.3.\\

Theorem 1 $abc(\eta)$ is maximized by η_0 where

$$\eta_0(r) = \begin{cases} 1 & r > 1 \\ 0 & r < 1 \end{cases}$$

and η_0 has properties p_1, p_4 and p_5.

To prove the above theorem we need the following

Lemma 1 $\frac{g_1(r)}{g_2(r)} = r,$

where g_i is the pdf of $R = \frac{f_1(\underline{X})}{f_2(\underline{X})}$ under H_i, and $r = \frac{f_1(\underline{x})}{f_2(\underline{x})}$ is the observed value of R, $i=1, 2$.

Proof: We can write

$$\frac{g_1(r)}{g_2(r)} = \lim_{\delta \to 0} \frac{G_1(r+\delta) - G_1(r)}{G_2(r+\delta) - G_2(r)} = \lim_{\delta \to 0} \frac{\int_A f_1(\underline{x})d\underline{x}}{\int_A f_2(\underline{x})d\underline{x}}$$

where $A = \left\{\underline{x} \middle| r < \frac{f_1(\underline{x})}{f_2(\underline{x})} \le r + \delta\right\}$.

But, noting that $f_1(\underline{x}) = \frac{f_1(\underline{x})}{f_2(\underline{x})} \times f_2(\underline{x})$ we have

$$r < \frac{\int_A f_1(\underline{x})d\underline{x}}{\int_A f_2(\underline{x})d\underline{x}} \le r + \delta \,.\mathrm{r}$$

So the result follows. ■

Proof of Theorem 1: We can write

$$abc(\eta) = \int_0^1 \eta(r)[g_1(r) - g_2(r)]dr.$$

The above integral is maximized if $\eta(r)$ takes its smallest value (zero) when $g_1(r) - g_2(r) < 0$ and its greatest value (one) where $g_1(r) - g_2(r) > 0$. Thus η_0 maximizes $abc(\eta)$ by Lemma 1. That η_0 has properties p_1, p_4 and p_5 is evident. ■

As is evident η_0 (taking only values 0 or 1) either completely supports H_1 or completely supports H_2, thus it is not suitable for our purpose. Fig.4 shows plot of η_0 against r and Fig.5. shows cdf of η_0 under H_1 and H_2.

In order to find a more acceptable measure, we impose the restriction of unbiased ness as defined below.

Definition 1 A measure of support η is called unbiased if

$$\eta \overset{s}{\geq} U(0,1) \textit{ under } H_1 \textit{ and } \eta \overset{s}{\leq} U(0,1) \textit{ under } H_2$$

or equivalently if

$$P_{H_1}(\eta \leq c) \leq c \quad and \quad P_{H_2}(\eta \leq c) \geq c \quad \forall c \in [0,1] \qquad ■$$

If η is distributed as $U(0,1)$ then it has equal probabilities of supporting either of the two hypotheses H_1 and H_2, while it has more probability of supporting $H_1(H_2)$ if it is stochastically greater (smaller) than $U(0,1)$. So unbiased ness may be considered a desirable property on its own.

Lemma 2 A continuous and strictly increasing measure of support η for H_1 against H_2 is unbiased if and only if

$$G_1(r) \leq \eta(r) \leq G_2(r) \quad , \qquad \forall r \geq 0$$

where (as before) $G_i(r)$ is the cdf of $R = \frac{f_1(\underline{X})}{f_2(\underline{X})}$ under H_i , $i = 1,2$.

Proof: For $\forall r \geq 0$ we can write

$$\begin{aligned} G_1(r) \leq \eta(r) &\Leftrightarrow \eta^{-1}(r) \leq G_1^{-1}(r) \\ &\Leftrightarrow G_1\left[\eta^{-1}(r)\right] \leq r \\ &\Leftrightarrow P_{H_1}\left[R \leq \eta^{-1}(r)\right] \leq r \\ &\Leftrightarrow P_{H_1}\left[\eta(R) \leq r\right] \leq r \end{aligned}$$

The proof for $\eta(r) \leq G_2(r)$, $\forall r \geq 0$ is similar. ■

Theorem 2 Among all unbiased functions of r, $abc(\eta)$ is maximized by

$$\eta_3(r) = \begin{cases} G_1(r) & r < 1 \\ G_2(r) & r > 1 \end{cases}$$

where $\eta_3(1)$ can be any number in the unit interval.

Proof: is similar to the proof of Theorem 1 except that values that η can take are restricted by Lemma 2. ■

The following theorem implies that η_3 is strictly increasing in r.

Theorem 3 $G_1(r) \le G_2(r)$, $\forall r \ge 0$.

Proof: we can write

$$\begin{aligned} G_1(r) &= P_1(R \le 1) \\ &= \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} \le r} f_1(\underline{x}) d\underline{x} \\ &= \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} \le r} \frac{f_1(\underline{x})}{f_2(\underline{x})} \times f_2(\underline{x}) d\underline{x} \\ &\le r \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} \le r} f_2(\underline{x}) d\underline{x} = rG_2(r) \end{aligned}$$

We can also write

$$\begin{aligned} \overline{G}_1(r) &= 1 - G_1(r) = P_{H_1}(R > r) \\ &= \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} > r} f_1(\underline{x}) d\underline{x} = \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} > r} \frac{f_1(\underline{x})}{f_2(\underline{x})} \times f_2(\underline{x}) d\underline{x} \\ &> \int_{\frac{f_1(\underline{x})}{f_2(\underline{x})} > r} f_2(\underline{x}) d\underline{x} = r[1 - G_2(r)] = r\overline{G}_2(r) \end{aligned}$$

So for $r \ge 1$we have $\overline{G}_1(r) > \overline{G}_2(r)$ that is for $r \ge 1$ we have $G_1(r) < G_2(r)$, $\forall r \ge 0$ ■

The fact that η_3 satisfies restriction p_4 is shown by the following

Theorem 4 $\eta_3(H_2, H_1) = 1 - \eta_3(H_1, H_2)$

Proof: We have

$$\eta_3(H_2, H_1) = \begin{cases} G_1^*(r') & r' < 1 \\ G_2^*(r') & r' > 1 \end{cases} , \quad where \quad r' = \frac{f_2(\underline{x})}{f_1(\underline{x})} = \frac{1}{r}$$

$$= \begin{cases} P_{H_2}(R' \le r') & r' < 1 \\ P_{H_1}(R' \le r') & r' > 1 \end{cases} , \quad where \quad R' = \frac{f_2(\underline{X})}{f_1(\underline{X})} = \frac{1}{R}$$

$$= \begin{cases} P_{H_2}(R \ge r) & r > 1 \\ P_{H_1}(R \ge r) & r < 1 \end{cases} = 1 - \eta_3(H_1, H_2).$$

where G_i^* is cdf of R', $i = 1, 2$. ■

Theorem 5 η_3 is consistent.

Proof: is similar to the proof consistency of η_2 in Theorem 8 bellow. ■

We can remedy the discontinuity of η_3 by suitably scaling it to give

$$\eta_4 = \begin{cases} \frac{1}{2G_1(1)} G_1(r) & r < 1 \\ 1 - \frac{1}{2G_2(1)}\left[1 - G_2(r)\right] & r > 1 \end{cases}$$

Since η_3 is continuous for all values of r except for $r = 1$, it is evident that η_4 is continuous for all values of r and $\eta_4(1) = \frac{1}{2}$. η_4 is however not optimal that is $abc(\eta_3) - abc(\eta_4) > 0$ (where the difference can be substantial for large sample size) and can be biased if $G_1(1) > \frac{1}{2}$ or $G_2(1) < \frac{1}{2}$ (which can happen for small sample sizes for some distributions such as exponential distribution). ■

The fact that maximizing $abc(\eta)$ has no solution among the class of continuous measures is shown by
the following
Theorem 6 Let

$$\eta_\delta = \begin{cases} G_1(r) & r < 1 - \delta \\ G_2(r) & r > 1 + \delta \\ \frac{G_2(1+\delta) - G_1(1-\delta)}{2\delta}\left[r - (1+\delta)\right] & 1 - \delta \le r < r + \delta \end{cases}$$

then $\forall \varepsilon > 0 \ \exists \delta > 0 \ s.t. \ 0 < abc(\eta_3) - abc(\eta_\delta) < \varepsilon$.

Proof:

Since $G_1(r)$ and $G_2(r)$ are continuous, as $\delta \to 0$, $abc(\eta_3) - abc(\eta_\delta) \to 0$.
So the result follows. ■

3 A measure of support for H_1 against H_2 based on p-values

The main drawbacks of p-value as a measure of support for H_1 (against H_2) are the fact that its value does not in any way depend on H_2 and that under H_1 it is distributed as $U(0,1)$, so that under H_1 it is equally likely to take any value in the unit interval. The second of the above is also responsible for the fact that, under H_1, p-value is stochastically much smaller than η_3. So it is conceivable that if a measure is defined as a function of both p-values (p-value when H_1 is the null hypothesis and p-value when H_2 is the null hypothesis), it would not only have a symmetric relation with H_1 and H_2, but it would also have greater abc. It turns out that η_2, as defined below, has all desirable properties and its abc is not far from optimal.

Theorem 7 Let

$$\eta_2 = \frac{G_1(r)}{G_1(r)+\overline{G}_2(r)}$$

where G_i is cdf of R under H_i, $i=1,2$. Then η_2 is unbiased.

Proof: We know from Theorem 3 that $G_1(r) \le G_2(r)$, $\forall r \ge 0$.

So we can write

$$G_1(r) \le 1-\overline{G}_2(r) \Rightarrow G_1(r)+\overline{G}_2(r) \le 1$$

$$\Rightarrow \forall c \in [0,1],\ P_1(\eta_2 \le c) = P_1\left(\frac{G_1(R)}{G_1(R)+\overline{G}_2(R)} \le c\right)$$

$$\le P_1(G_1(R) \le c) = c,\ since\ G_1(R) \overset{H_1}{\sim} U(0,1)$$

Since by p_4 $\eta_2(H_1, H_2) = 1-\eta_2(H_2, H_1)$, we have

$$P_2(\eta_2(H_1, H_2) \le c) = P_2(\eta_2(H_2, H_1) \ge 1-c)$$
$$= 1-P_2(\eta_2(H_2, H_1) \le 1-c) \ge 1-(1-c) = c$$

where the above inequality follows from the first part of the proof. ■

Theorem 8 η_2 has all properties p_1 through p_6.

Proof: That η_2 has properties p_1 and p_2 is obvious.

Assuming that G_1 and G_2 are differentiable we differentiate η_2 w.r.t r to get

$$\frac{d\eta_2}{dr} = \frac{g_1(r)\overline{G}_2(r)+g_2(r)G_1(r)}{\left(G_1(r)+\overline{G}_2(r)\right)^2} > 0\ ,\ \forall r$$

This proves p_3.

To prove p_4 we can write

$$\eta_2(H_2, H_1) = \frac{P_2(R' \le r')}{P_2(R' \le r')+P_1(R' > r')},\ R' = \frac{1}{R},\ r' = \frac{1}{r},\ R = \frac{f_1(\underline{X})}{f_2(\underline{X})},\ r = \frac{f_1(\underline{x})}{f_2(\underline{x})}$$

$$= \frac{P_2(R \ge r)}{P_2(R \ge r)+P_1(R < r)} = \frac{\overline{G}_2(r)}{\overline{G}_2(r)+G_1(r)} = 1-\eta_2(H_1, H_2).$$

To prove p_5 (consistency) we assume that the elements of the random vector $\underline{X}$ are iid random variables each distributed according to a distribution with density f_i^* under H_i, $i=1,2$. Furthermore we assume that f_1^* and f_2^* have common support and are distinct, that is the set

$\{x | f_1^*(x) \neq f_2^*(x)\}$ has nonzero Lobesgue measure. Since $G_1(r)$ and $\overline{G}_2(r)$, are p-values of the MP (Neyman Pearson) tests when H_1 and H_2 are regarded as the null hypothesis respectively and since under the above assumptions these tests are consistent (their power tends to 1 as $n \to \infty$), we conclude that as $n \to \infty$ $G_1(R) \xrightarrow{H_1} 1$, $G_1(R) \xrightarrow{H_2} 0$, $\overline{G}_2(R) \xrightarrow{H_1} 0$ *and* $\overline{G}_2(R) \xrightarrow{H_2} 1$ in probability. Therfore $\eta_2(R)$ tends in probability to 1 and 0 under H_1 and H_2 respectively.

It was proved in Theorem 7 that $G_1(r) + \overline{G}_2(r) \leq 1$, $\forall r \geq 0$, knowing that $G_1(R) \overset{H_1}{\sim} U(0,1)$ we can write

$$P_1\left(\eta_2(H_1, H_2) \leq \frac{1}{k}\right) \leq P_1\left(G_1(R) \leq \frac{1}{k}\right) = \frac{1}{k}\ ,\ \ \forall k > 1,,$$

and

$$P_2\left(\eta_2(H_1, H_2) \leq 1 - \frac{1}{k}\right) \leq P_2\left(\eta_2(H_1, H_2) \leq \frac{1}{k}\right)\ ,\ \ by\ \ p_4,$$

$$\leq P_2\left(\overline{G}_2(R) \leq \frac{1}{k}\right) = \frac{1}{k}.$$

■

4 Testing statistical Hypotheses by observing the value of η_2

Let

$$\varphi^\circ(r) = \begin{cases} 1 & \eta_2(r) < \dfrac{\gamma}{\gamma+1} \\ 0 & \eta_2(r) \geq \dfrac{\gamma}{\gamma+1} \end{cases}$$

be a test for testing H_1 (as the null hypothesis) against H_2 (as the alternative hypothesis), where H_1 and H_2 are both simple, then we have the following

Theorem 9 Let φ° be as defined above, then

i) $epr(\varphi^\circ) = \gamma$, where $epr(\varphi^\circ)$ is defined to be the error probability ratio ofs φ° that is

$$epr(\varphi^\circ) = \frac{probability\ of\ type\ I\ error\ of\ \varphi^\circ}{probability\ of\ type\ II\ error\ of\ \varphi^\circ}$$
$$= \frac{G_1(r)}{\overline{G}_2(r)}$$

ii) φ° minimizes both error probabilities among all members of the class

$$D_\gamma = \langle \varphi | epr(\varphi) = \gamma \rangle$$

Proof: (i) It is obvious that φ° can be written as

$$\varphi^\circ = \begin{cases} 1 & \lambda_2(r) < \gamma \\ 0 & \lambda_2(r) \geq \gamma \end{cases}$$

where $\lambda_2(r) = \dfrac{G_1(r)}{\overline{G}_2(r)}$. It is elementary that λ_2 takes values in the interval $[0, \infty]$ and is also a continuous and strictly increasing function of r. So we can write

$$epr(\varphi^\circ) = \frac{P_1[\lambda_2(R) \leq \gamma]}{P_2[\lambda_2(R) > \gamma]} = \frac{P_1[R \leq \lambda_2^{-1}(\gamma)]}{P_2[R > \lambda_2^{-1}(\gamma)]}$$
$$= \frac{G_1[\lambda_2^{-1}(\gamma)]}{\overline{G}_1[\lambda_2^{-1}(\gamma)]} = \lambda_2[\lambda_2^{-1}(\gamma)] = \gamma$$

(ii) Since $\lambda_2(r)$ is strictly increasing in r, φ° is MP of its size. Now let $\varphi \in D_\gamma$, that is let $epr(\varphi) = \gamma$ and suppose $\alpha^\circ = \alpha(\varphi^\circ) =$ type I error probability of φ°.

If $\alpha(\varphi) \leq \alpha^\circ$, then since φ° is MP of size α°, we have

$$\beta(\varphi) \geq \beta(\varphi^\circ) = type\ II\ error\ probability\ of\ \varphi^\circ,$$

and if $\alpha(\varphi) > \alpha(\varphi^\circ)$ then we can write

$$epr(\varphi) = epr(\varphi^\circ) \Rightarrow \frac{\alpha(\varphi)}{\beta(\varphi)} = \frac{\alpha(\varphi^\circ)}{\beta(\varphi^\circ)}$$
$$\Rightarrow \beta(\varphi) > \beta(\varphi^\circ)$$

So in either case we $\beta(\varphi) \geq \beta(\varphi^\circ)$.

That $\alpha(\varphi) \geq \alpha(\varphi^\circ)$ follows from the above inequality and the assumption that $epr(\varphi) = epr(\varphi^\circ)$. ■

If the researcher is supplied with the value of η_2, by the above theorem, he can perform a test of H_1 against H_2 in such a way that the error probability ratio of his test is equal to his desired value γ, and he can be sure that this test has smallest error probabilities of both kinds among all tests with $epr = \gamma$.

A desirable property of such a test is that both errors benefit from an increase in sample size (both decrease) and thus the "large sample problem", (which is the fact that in practice the null hypothesis is always rejected for large enough sample size) does not arise.

We would like to mention that in all specific cases (that is when f_1 and f_2 belong to some specific family such as normal, exponential, etc.) that we have considered, $abc(\eta_2)$ is greater than $abc(\eta_1)$. The following Figures show the above claim for the normal and the exponential cases. Also note that it is seen from the corresponding figures, that η_1 is not unbiased.

5 Some Plot

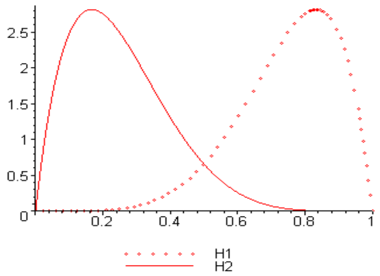

Fig.1

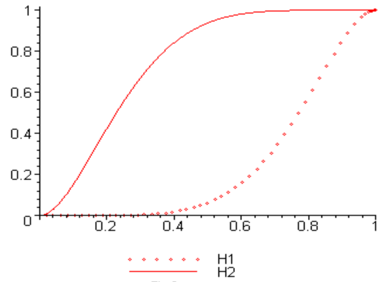

Fig.2

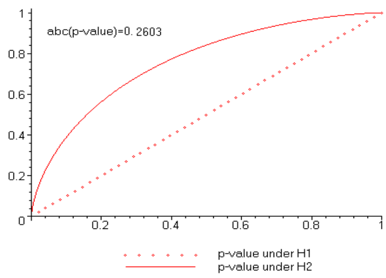

Fig. 3 (Normal, n=1)

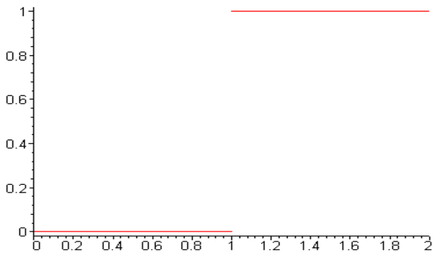

Fig.4 (Normal, n=1)

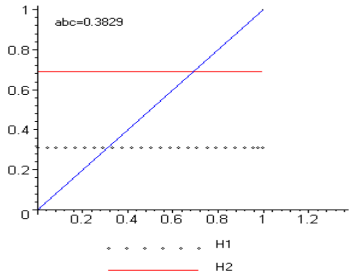

Fig.5 (eta0 under Normal, n=1)

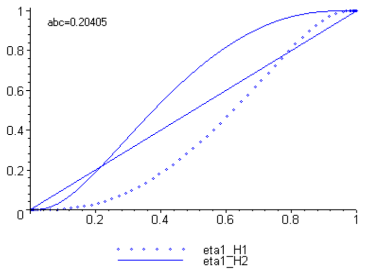

Fig.6 (Normal, n=1)

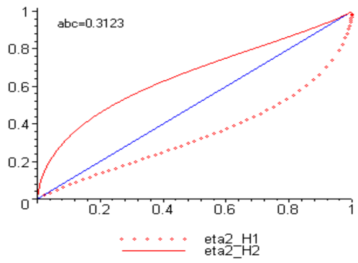

Fig.7 (Normal, n=1)

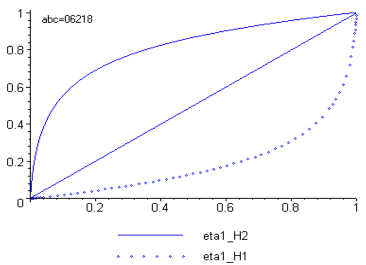

Fig. 8 (Normal, n=5)

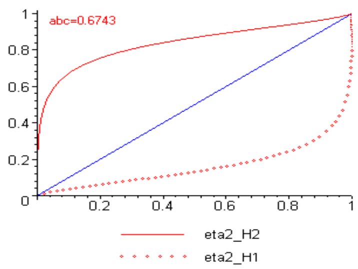

Fig. 9 (Normal, n=5)

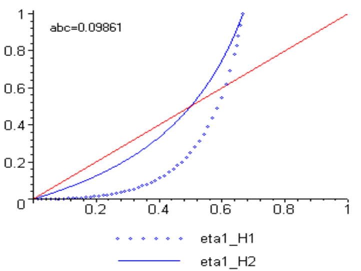

Fig. 10 (Exponential, n=1)

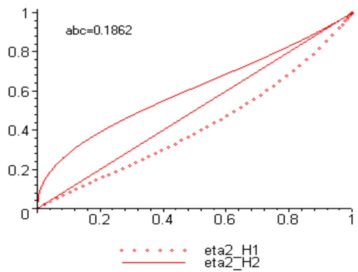

Fig. 11 (Exponential, n=1)

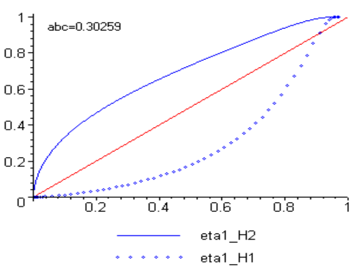

Fig. 12 (Exponential, n=5)

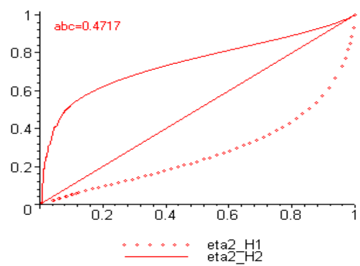

Fig. 13 (Exponential, n=5)

References

Blume, J.D. (2001) "*Likelihood Methods for Measuring Statistical Evidence*"

J.O. Berger, Thomas Sellke, M.J. Bayarri (2001) "*Calibration of P-values for Precise Null Hypotheses*" ,The American Statistician

Richard Royal, (2000) "*On the Prob. of Observing Misleading Statistical Evidence*", JASA

Robins etal (2000) "*Asymptotic Distribution of P-values in composite Null Models*", JASA

Bayarri, M.J and Berger, J.O. (2000) "*P-values for Composite Null Models*", JASA

Royall, R. (1999) "*Statistical Evidence*" CHAPMAN & HALL

Schervish, M.J. (1996) "*P-values: What They Are and what They Are Not*", The American Statistician

Blyth, C.R. and Staudte, R.G. (1995) "*Estimating Statistical a Hypothesis*" Statistics and Probability Letters

Berger, J.O. and Sellke, T. (1987) "*Irreconcilability of P-values and Evidence*", JASA

Casella, G. and Berger, R.L. (1987) "*Reconciling Baysian and Frequentist Evidence*", JASA

E.L. Lehman (1975) "*Nonparametrics: Statistical Methods Based on Ranks*", Holden and Day, San Francisco.
Degroot, M.H. (1973) "*Doing What Comes naturally …*", JASA