ON THE FOCUSING OF CRAMÉR - VON MISES TEST.

ALEJANDRA CABAÑA AND ENRIQUE M. CABAÑA

Abstract. The statistical bibliography frequently refers to *omnibus tests* intended to be sensitive to all or at least a wide variety of alternatives, and *focused or directional tests* directed to detect efficiently some specific alternatives.

In fact, the apparent opposition between omnibus and focused is artificial, and, for instance, K-S test is focused on changes in position of Double Exponential distribution, as well as Cramér - von Mises is focused on changes in position of the distribution with density $f(t) = 1/(2\cosh(\pi t/2))$.

We provide in this article a simple proof of this latter fact.

1. Introduction

In the statistical literature refering to a test as being omnibus or directional often implies opposite categories.

Omnibus tests are able to detect a wide bunch of alternatives, and no special ability to detect any particular one is intended.

When statistical practitioners wish to detect specific alternatives they can use directional tests. These ones focus their power in the direction of the interesting alternatives.

The former tests are not expected to be efficient in the detection of particular alternatives. On the other hand, it is generally claimed that the second ones have the drawback that they have a poor power against alternatives other that the ones on which they were focused.

Research partially supported by TIN2008-06582-C03-02/TIN, Ministerio de Ciencia y Tecnología.

Partially supported by CSIC-Udelar,Uruguay, Centre de Recerca Matemàatica, Barcelona, Spain and Carolina Foundation, Spain.

Notwhithstanding, it is well established that a test can be both omnibus and focused: this is the case of the well known omnibus Kolmogorov - Smirnov goodness-of-fit test, that is also focused to detect changes in position of samples of the Double - Exponential Distribution as shown by J. Capon ([3]) by computing lower bounds for the asymptotic efficiency of the test for several alternatives.

In this short note, we show that the well known Cramér - von Mises goodness-of-fit test, also reputed to be an omnibus test, is also focused to detect changes in position of random samples of another family of distributions obtained by changes in location and scale from the distribution with probability density

$$g(t) = \frac{1}{2\cosh(\pi t/2)}. \tag{1}$$

It is known (see [8]) that there is one direction with the highest asymptotic power that is possible for Cramér - von Mises test. We present here a straightforward computation of such direction.

The principal result is that the asymptotic power of the Cramér - von Mises test for those alternatives is almost optimal. This statement is made precise in §4, where the power of the test is compared with the power of the two-sided test based on the likelihood ratio.

This kind of quasi-optimal behaviour characterises several tests of goodness-of-fit developed by the authors in which a quadratic statistic of Watson type is employed in such a way that the resulting tests are consistent against any alternative, and also have a near optimum efficiency for some alternative of focusing arbitrarily selected by the user (see [1], [2] and references therein).

The tuning on the interesting alternatives is a part of the design of our tests, but the quasi-optimum efficiency is inherent to the statistic in use.

The efficiency of our tests is described in the already cited articles. But the fact that the efficiency of the classical Cramér - von Mises test share such kind of properties does not appear to us to be widely discussed in the statistical literature, and motivates this article.

The power of Cramér - von Mises test has been analysed by several authors, and is fully described by Durbin and Knott ([5]), for instance. We describe it from scratch in order to facilitate the reading, before identifying in §3 the alternatives for optimum power.

2. The Cramér - von Mises goodness-of-fit test.

The Cramér - von Mises statistic $\omega_n^2 = n\int_{-\infty}^{\infty}(F_n(t) - F_0(t))^2 dF_0(t)$ quantifies a quadratic distance between the probability distribution function F_0 and the empirical distribution function $F_n(t) = \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq t\}}$ of the sample of i.i.d. random variables $X_1, X_2, \ldots, X_n$ with probability distribution F.

By introducing the *empirical process* $b_n(t) = \sqrt{n}(F_n(t) - F_0(t))$, ω_n^2 is written as

$$\omega_n^2 = \int_{-\infty}^{\infty} b_n^2(t) dF_0(t).$$

We shall assume that F_0 is continuous, with density f_0, finite first- and second-order moments, and, with no loss of generality that $\int t dF_0(t) = 0$, $\int t^2 dF_0(t) = 1$.

Let the probability distribution of ω_n^2 be denoted by $P(t, F, n) = \mathbf{P}\{\omega_n^2 \leq t\}$.

The Cramér - von Mises test of the null hypothesis $\mathcal{H}_0$: "$F = F_0$", with confidence level α, rejects $\mathcal{H}_0$ when $\omega_n^2 > c_n(\alpha)$, where $c_n(\alpha)$ solves the equation $P(c_n(\alpha), F_0, n) = 1 - \alpha$, and its power for the alternative F is $1 - P(c_n(\alpha), F, n)$.

2.1. **The asymptotic law of ω_n^2 under $\mathcal{H}_0$.** Since b_n converges in law to a brownian bridge associated to F_0, that is, to a Gaussian centred process b^{F_0} with covariances $\mathbf{E}b^{F_0}(s)b^{F_0}(t) = F_0(s \wedge t) - F_0(s)F_0(t)$, then ω_n^2 has the asymptotic law of $\int (b^{F_0}(t))^2 dF_0(t) \sim \int_0^1 b^2(u)du$, where b denotes a standard Brownian bridge, because b^{F_0} has the same law as $b \circ F_0$.

In order to obtain the distribution of $Q_0 = \int_0^1 b^2(u)du = \|b\|^2$, the L^2 squared norm of the standard Brownian bridge b in $L^2(([0,1])$ with the Lebesgue measure, let us follow Durbin ([4]) and compute the Fourier expansion

$$b(u) = \sum_{j=1}^{\infty} \left(\int_0^1 b(v)\psi_j(v)dv \right) \psi_j(u) \tag{2}$$

of b in terms of the complete orthonormal system $\{\psi_j(u) = \sqrt{2}\sin j\pi u : j = 1, 2, \dots\}$ of eigenfunctions of the covariance kernel which admits the expansion

$$\mathbf{E}b(u)b(v) = u \wedge v - uv = \sum_{j=1}^{\infty} \frac{1}{j^2\pi^2}\psi_j(u)\psi_j(v).$$

The random coefficients in (2) are independent centred Gaussian variables vith variances

$$\mathbf{E}\left(\int_0^1 b(u)\psi(u)du\right)^2 = \int_0^1\int_0^1 (u \wedge v - uv)\psi(u)\psi(v)\,du\,dv = \frac{1}{j^2\pi^2}$$

and hence we may rewrite (2) as $b(u) = \sum_{j=1}^{\infty} \frac{B_j}{j\pi}\psi_j(u)$, by introducing the i.i.d. standard Gaussian variables $B_j = j\pi\int_0^1 b(u)\psi_j(u)du$, leading us to conclude

$$Q_0 = \|b\|^2 = \sum_{j=1}^{\infty} \frac{B_j^2}{j^2\pi^2}. \tag{3}$$

2.2. **The limiting law of ω_n under sequences of contiguous alternatives.** Let us assume now that for each n, the sample has a probability law $F^{(n)}$ with density $f_n(t)$ satisfying

$$\sqrt{\frac{f_n(t)}{f_0(t)}} = 1 + \frac{\delta k_n(t)}{2\sqrt{n}}$$

for a sequence of functions k_n such that

$$\int_{-\infty}^{\infty}(k_n(t) - k(t))^2 dF_0(t) \to 0, \quad \int_{-\infty}^{\infty} k^2(t)dF_0(t) = 1.$$

When this happens, we shall say that the alternative $\mathcal{H}(k, \delta)$ holds. These alternatives are contiguous to the null hypothesis (see [9]) and therefore the asymptotic law of b_n under $\mathcal{H}(k, \delta)$ is the same one corresponding to $\mathcal{H}_0 = \mathcal{H}(k, 0)$ plus a deterministic term, according to Le Cam Third Lemma ([6], [7]).

The limiting distribution of the empirical process under $\mathcal{H}(k,\delta)$, is obtained by noticing that the first term in the decomposition

$$b_n(t) = \sqrt{n}(F_n(t) - F^{(n)}(t)) + \sqrt{n}(F^{(n)}(t) - F_0(t)).$$

tends to $b^{(F_0)}(t)$, and the second one is written as

$$\sqrt{n}\int_{-\infty}^{t}(f_n(s) - f_0(s))ds = \sqrt{n}\int_{-\infty}^{t}\frac{\delta k_n(s)}{\sqrt{n}}dF_0(s) \to \delta\int_{-\infty}^{t}k(s)dF_0(s)$$

so that, with the change of variables $u = F_0(t)$ and the new function K defined by

$$K(u) = \int_0^u \kappa(v)dv, \kappa(F_0(t)) = k(t),$$

we get
(4)

$$b_n(t) \xrightarrow{\mathcal{L}} b^{(F_0)}(t) + \delta\int_{-\infty}^{t}k(s)dF_0(s) = b(u) + \delta\int_0^u \kappa(v)dv = b(u) + \delta K(u).$$

The assumptions on k imply that κ satisfies $\int_0^1 \kappa(u)du = 0$, $\int_0^1 \kappa^2(u)du = 1$, and, in particular, $K(0) = K(1) = 0$. The function κ shall be called *standardized shape* of the alternative $\mathcal{H}(k,\delta)$.

From (4), we obtain

$$\omega_n^2 \xrightarrow{\mathcal{L}} \int_0^1 (b(u) + \delta K(u))^2 du.$$

Let us notice that this expression of the limit law of ω_n leads to conclude that when the null hypothesis is replaced by $\mathcal{H}(k,\delta)$, then the asymptotic expectation of ω_n increases in the amount

$$\Delta(\delta) = \delta^2 \int_0^1 K^2(u)du. \tag{5}$$

It is reasonable to expect that larger values of $\Delta(\delta)$ be associated with larger powers of the tests comparing $\mathcal{H}_0$ with $\mathcal{H}(\delta,k)$. Therefore, we search in the next section the function K that maximises $\Delta(\delta)$ for given δ.

3. The focused alternatives.

3.1. **The standardized shape κ of the alternative that produces the largest increment in the asymptotic expectation of ω_n.** We shall obtain the function $K(u) = \int_0^u \kappa(s)ds$ that maximises $\int_0^1 K^2(u)du$ with the restrictions

$$\int_0^1 \kappa^2(u)du = 1, \int_0^1 \kappa(u)du = 0.$$

The associated Euler equations express that for each continuously differentiable g such that

$$g(0) = g(1) = 0, \int_0^1 K'(u)g'(u) = 0$$

the condition

$$\int_0^1 K(u)g(u)du = 0$$

must hold.

The condition $\int_0^1 K'(u)g'(u) = 0$ holds for every g such that $g(0) = g(1) = 0$ provided

$$\int_0^1 K'(u)g'(u)du = [g(u)K'(u)]_0^1 - \int_0^1 K''(u)g(u)du = 0.$$

Since the integrated term in the right-hand side vanishes, we find that when g is orthogonal to K'' in $L^2([0,1])$, it is also orthogonal to K, and this means that K and K'' are proportional, that is, for some constant $\pm\lambda^2$, K solves the differential equation $K'' = \pm\lambda^2 K$.

The solutions of $K'' = \pm\lambda^2 K$ in $[0,1]$ with border conditions $K(0) = K(1) = 0$, satisfying $\int_0^1 (K'(u))^2 du = 1$ are

$$K(u) = \frac{\sqrt{2}}{j\pi} \sin j\pi u, j = 1, 2, \ldots.$$

The solution with maximum norm is the one with $j = 1$, hence

$$\kappa(u) = \sqrt{2}\cos \pi u. \tag{6}$$

This is the standardized shape of the alternative that maximises (5) for given δ. The corresponding function $K(u) = \int_0^u \kappa(s)ds$ is proportional to the first function in the orthonormal system introduced in 2.1, that is, $K(u) = \psi_1(u)/\pi$.

3.2. **Alternatives of change in location.** When the alternative distributions specify a change in location

$$f_n(t) = f_0(t + \delta c/\sqrt{n})$$

we have

$$\sqrt{\frac{f_n(t)}{f_0(t)}} = 1 + \frac{\delta c}{2\sqrt{n}} \frac{f_0'(t)}{f_0(t)} + o(\tfrac{1}{\sqrt{n}})$$

so that $k(t) = c\frac{f_0'(t)}{f_0(t)}$. The constant c is introduced in order to be able to impose $\|k\|^2 = 1$.

It follows that $\kappa(u) = c\frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}$ and Equation (6) shows that the alternative shall be detected by the Cramér - von Mises statistic with maximum asymptotic increment of the expectation when

$$c\frac{f_0'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} = \sqrt{2}\cos \pi u.$$

In order to solve this differential equation in F_0, we return to the variable $t = F_0^{-1}(u)$, and get

$$cf_0'(t) = \sqrt{2} f_0(t) \cos \pi F_0(t),$$

which, integrated in $(-\infty, t]$ gives

$$cf_0(t) = \frac{\sqrt{2}}{\pi} \sin \pi F_0(t).$$

A further integration leads to

$$\frac{\sqrt{2}t}{c\pi} = \int_0^t \frac{dF_0(s)}{\sin \pi F_0(s)} = \int_{F_0(0)}^{F_0(t)} \frac{du}{\sin \pi u}$$

$$= \frac{1}{2\pi} \log \frac{(\cos \pi F_0(t) - 1)(\cos \pi F_0(0) + 1)}{(\cos \pi F_0(t) + 1)(\cos \pi F_0(0) - 1)}.$$

By imposing with no loss of generality that F_0 is centred in 0, follows the simpler expression

$$\gamma t = \log \frac{1 - \cos \pi F_0(t)}{1 + \cos \pi F_0(t)},$$

in which the parameter $\gamma = \frac{2\sqrt{2}}{c}$ determines the dispersion.

By solving in F_0 and choosing $\gamma = \pi$ to get a distribution with variance equal one, we conclude

$$F_0(t) = \frac{1}{\pi} \arccos \frac{1 - e^{\pi t}}{1 + e^{\pi t}}, \quad f_0(t) = \frac{1}{2\cosh(\gamma t/2)}. \tag{7}$$

3.3. **Asymptotic law of ω_n under changes in location for samples with the law of Equation (7), and power of the test.** The statistic ω_n has the asymptotic law of

$$Q(\delta) = \int_0^1 (b(u) + \delta K(u))^2 du = \int_0^1 \left(b(u) + \frac{\delta}{\pi} \psi_1(u) \right)^2 du.$$

Since $b(u) + \frac{\delta}{\pi}\psi_1(u) = \sum_{j=1}^{\infty} \frac{1}{j\pi} B_j + \frac{\delta}{\pi}\psi_1(u)$, then

$$Q(\delta) = \left\| b + \frac{\delta}{\pi}\psi_1 \right\|^2 = \frac{1}{\pi^2} \left[(B_1 + \delta)^2 + \sum_{j=2}^{\infty} \frac{1}{j^2} B_j^2 \right]$$

Cramér - von Mises test of $F^{(n)}(t) = F_0(t)$ against $F^{(n)}(t) = F_0(t + \frac{2\sqrt{2}\delta}{\pi\sqrt{n}})$ with significance level α is asymptotically equivalent to the test of $\mathcal{H}_0$:"$\delta = 0$" with critical region $Q(\delta) > c(\alpha)$ where $c(\alpha)$ solves $\mathbf{P}\{Q(0) > c(\alpha)\} = \alpha$. The power, that we have computed by a numerical convolution for the purposes discussed in next section, is

$$\Pi(\delta, \alpha) = \mathbf{P}\{Q(\delta) > c(\alpha)\}.$$

4. Comparison with the two-sided test based on Neymann and Pearson statistic.

The Neyman and Pearson test of $\mathcal{H}_0$ against the alternatives $\mathcal{H}_n$ that the true density of the sample distribution is $g_n(t) = f_0(t + \frac{\delta}{c\sqrt{n}})$

has critical region

$$\sum_{i=1}^{n} \log\left(f_0\left(X_i + \tfrac{\delta}{c\sqrt{n}}\right)/f_0(X_i)\right) \geq \text{constant},$$

asymptotically equivalent to

$$\frac{\delta}{\sqrt{n}} \sum_{i=1}^{n} \frac{f_0'(X_i)}{cf_0(X_i)} \geq \text{constant}.$$

When $\mathcal{H}_0$ holds, the variables $f_0'(X_i)/(cf_0(X_i))$ are centred, with variance 1, and therefore the asymptotic law of the statistic $T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f_0'(X_i)}{cf_0(X_i)}$ is standard normal.

If the sequence of alternatives $\mathcal{H}_n$ hold, then

$$\mathbf{E}T_n = \sqrt{n}\mathbf{E}f_0'(X_1)/(cf_0(X_1)) = \sqrt{n}\int \frac{f_0'(x)}{cf_0(x)} f_0(x + \tfrac{\delta}{c\sqrt{n}})dx$$

has limit δ, $\mathbf{E}(f_0'(X_i))/(cf_0(X_i)))^2$ tends to 1, hence T_n converges in law to $Z + \delta$, Z standard Gaussian.

As a consequence, the test of $\delta = 0$ against $\delta > 0$ with optimal asymptotic power is the one with critical region $T_n >$ constant.

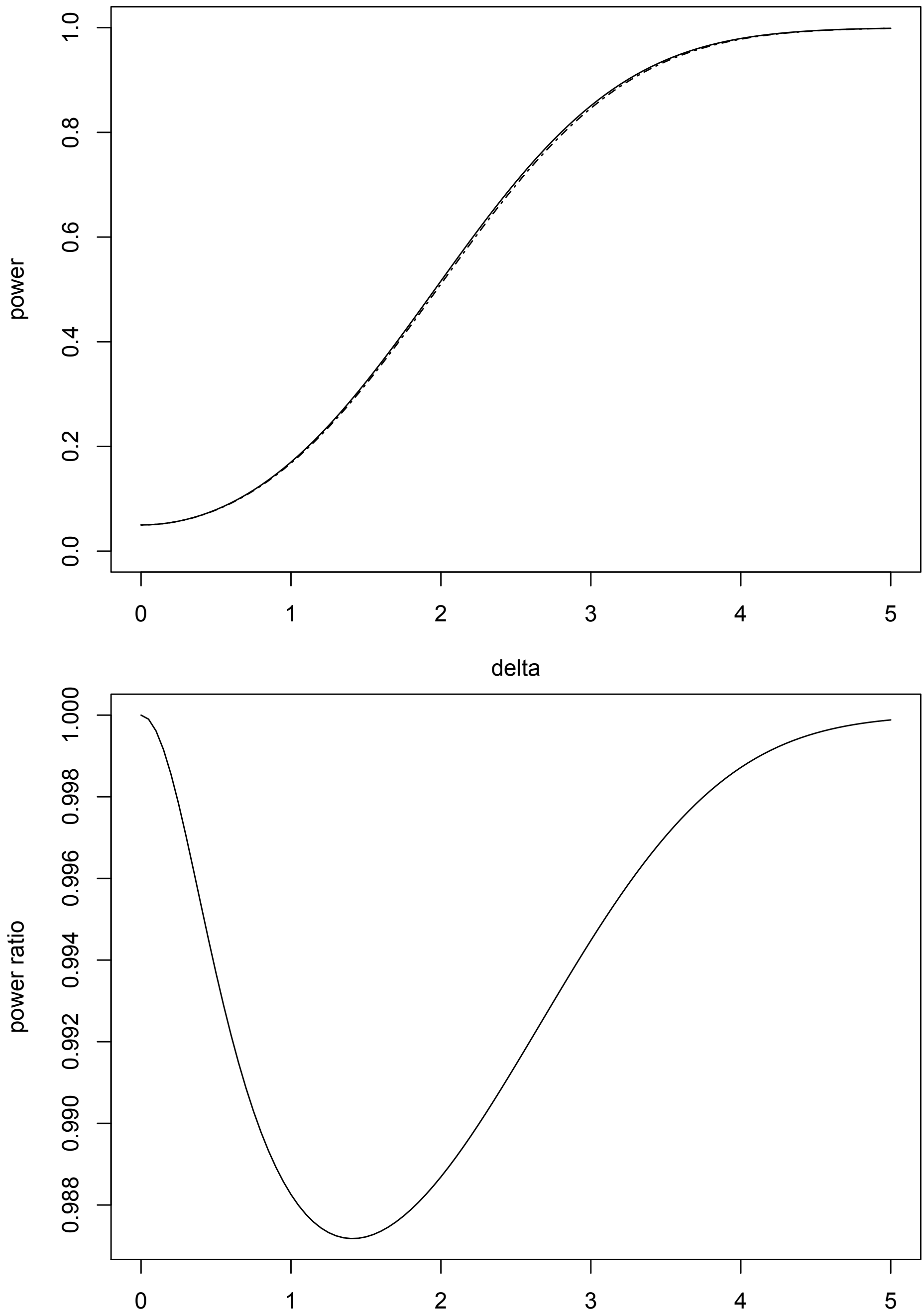
While there is no optimal test for $\delta = 0$ against $\delta \neq 0$, the usual practice if there are not significant differences between the cases $\delta > 0$ or $\delta < 0$ is to reject $\delta = 0$ when $|T_n| >$ constant. In that case, if Φ denotes as usual the standard normal cumulative distribution function, the asymptotic power of the two - sided test with asymptotic level α, is

$$\Pi^*(\delta, \alpha) = \mathbf{P}\{Z + \delta > \Phi^{-1}(1 - \tfrac{\alpha}{2})\} + \mathbf{P}\{Z + \delta < \Phi^{-1}(\tfrac{\alpha}{2})\}$$
$$= \Phi(\Phi^{-1}(\tfrac{\alpha}{2}) + \delta) + \Phi(\Phi^{-1}(\tfrac{\alpha}{2}) - \delta).$$

The practically coincident plots of the functions $\Pi(\delta,.05)$ and $\Pi^*(\delta,.05)$ in Figure 1 show that Cramér - von Mises test against the alternative of displacement of samples with distribution (7) is *almost optimal*, in the sense that its performance is *almost* asymptotically equivalent to the performance of the test with critical region $T_n >$constant.

The relationship between the asymptotic powers (and the intended meaning of "almost optimal") is better shown in the second diagram

FIGURE 1. Almost coincident asymptotic powers $\Pi^*(\delta,.05)$ and $\Pi(\delta,.05)$ of the two-sided test based on the Neymann and Pearson statistic (solid line) and of the Cramér - von Mises test (dotted line), respectively, for alternatives of change in position of a sample with distribution (7) (upper diagram) and ratio $\Pi(\delta,.05)/\Pi^*(\delta,.05)$ (lower diagram).



of Figure 1, where the ratio $\Pi(\delta,.05)/\Pi^*(\delta,.05)$ obtained by numerical computation is plotted.

References

[1] Cabaña, A. and Cabaña, E. M. (2005), *Goodness-of-fit to the exponential distribution, focused on Weibull alternatives*, Communications in Statistics, Simulation and Computation **34**, 711-724.

[2] Cabaña, A. and Cabaña, E. M. (2009), *Goodness-of-fit tests for continuous regression*, Methodology and Computing in Applied Probability, **11** 119-144.

[3] Capon, J. (1965) *On the asymptotic efficiency of the Kolmogorov - Smirnov test.* J. Amer. Statist. Assoc. **60**, 843-853.

[4] Durbin, J. (1973) *Distribution Theory for Tests Based on the Sample Distribution Funcion.* SIAM, 64 p.

[5] Durbin, J. & Knott, M. (1972) *Components of Cramér - von Mises Statistics. I.* J. Royal Statist. Soc. Series B. **34**, 290-307.

[6] Hajek, J. & Sidak, Z. (1967) *Theory of Rank Tests*, Academic Press, New York.

[7] Le Cam L. & Yang, G. L. (1990), *Asymptotics in Statistics. Some basic concepts.* Springer-Verlag, New York.

[8] Neuhaus, G. (1976), *Asymptotic Power Properties of the Cramér-von Mises Test Under Contiguous Alternatives* Jr. Multivariate Analysis **6**, 95-110.

[9] Oosterhoff, J & van Zwet, W. R., (1979), *A note on contiguity and Hellinger distance*, in Contributions to Statistics, 157-166, Reidel, Dordrecht.

Universitat Autònoma de Barcelona, Spain.

Universidad de la República, Montevideo, Uruguay.