Kybernetika

Marek Omelka

The behavior of locally most powerful tests

Kybernetika, Vol. 41 (2005), No. 6, [699]--712

Persistent URL: http://dml.cz/dmlcz/135687

KYBERNETIKA — VOLUME 41 (2005), NUMBER 6, PAGES 699–712

THE BEHAVIOR OF LOCALLY MOST POWERFUL TESTS

MAREK OMELKA

The locally most powerful (LMP) tests of the hypothesis $H : \theta = \theta_0$ against one-sided as well as two-sided alternatives are compared with several competitive tests, as the likelihood ratio tests, the Wald-type tests and the Rao score tests, for several distribution shapes and for location, shape and vector parameters. A simulation study confirms the importance of the condition of local unbiasedness of the test, and shows that the LMP test can sometimes dominate the other tests only in a very restricted neighborhood of H. Hence, we cannot recommend a universal application of the LMP tests in practice. The tests with a high Bahadur efficiency, though not exactly LMP, also seem to be good in the local sense.

Keywords: testing statistical hypothesis, locally most powerful tests

AMS Subject Classification: 62F03

1. INTRODUCTION

Let $X_1, \dots, X_n$ be independent identically distributed (i.i.d.) random variables with a common density $f(x, \theta)$ (with respect to Lebesgue measure), where $\theta \in \Omega \subset \mathbb{R}$. Consider the problem of testing the simple null hypothesis $H : \theta = \theta_0$ against the one-sided alternative $K : \theta > \theta_0$ ($K : \theta < \theta_0$) or against the two-sided alternative $K : \theta \neq \theta_0$. A uniformly most powerful (UMP) test or a uniformly most powerful unbiased (UMPU) test of H exists e. g. when the density f belongs to the exponential family of distributions (Lehmann [8]). In the case of nonexistence of a UMP test, the task of a statistician is to find a suitable and reasonable test. One possibility is to use a locally most powerful (LMP) test, which maximizes the slope of a power function in a neighborhood of the null hypothesis. Another natural way how to derive the locally most powerful test is to restrict the attention to the tests which are locally admissible in the Brown and Marden [1] sense. While the form of locally most powerful tests is well known in many situations, surprisingly there are not many studies showing up to what extent, in which neighborhood, the test is really optimal, and whether the neighborhood of the hypothesis is large enough to justify the use of the locally most powerful test.

This is just the goal of the present paper: In several specific cases, we shall compare the power functions of the LMP tests and of some competitive tests in order to show how far their optimalities go. Our results supplement the theoretic asymptotic results of Chibisov [2], Kallenberg [7], as well as the simulation study of Ramsey [11].

Section 2 describes some basic concepts pertaining to the locally optimal tests, and introduces some alternative tests to LMP ones. Section 3 studies the tests based on sample $X_1, \dots, X_n$ for three specific models:

1. the *double-exponential distribution* with the location parameter θ and density $f(x) = \frac{1}{2}\, e^{-|x-\theta|}$ and testing $H\colon \theta = 0$ against $K\colon \theta > 0$;
2. the *Weibull distribution* with the shape parameter a and density function $f(x) = ax^{a-1}\exp(-x^a)$ and testing $H : a = 1$ against $K : a > 1$ and also $H\colon a = 1$ against $K\colon a \neq 1$;
3. the *normal distribution* $\mathcal{N}(\mu, \sigma^2)$ with both parameters unknown and testing $H\colon (\mu, \sigma^2)^\mathsf{T} = (0,1)^\mathsf{T}$ against $K\colon (\mu, \sigma^2)^\mathsf{T} \neq (0,1)^\mathsf{T}$ (sometimes called the test of full specification of the normal distribution).

The aim of this paper is to provide an insight into the performance of tests rather than to give exact values of power functions. That is why our results are often illustrated on the figures, which give a better picture of the situation than the tables.

2. BASIC CONCEPTS AND DEFINITIONS

Consider a random sample $\boldsymbol{X} = (X_1, \dots, X_n)$ from a distribution with density $f(x,\theta)$ which depends on the unknown parameter θ. Let H be a null hypothesis about this parameter and Φ be a test function defined on the sample space which gives the probability of rejecting H when the sample $\boldsymbol{X} = \boldsymbol{x}$ is observed. Denote $\beta_\Phi(\theta) = \boldsymbol{E}_\theta \Phi(\boldsymbol{X})$ the power function of this test.

Definition 1. Let d be a measure of the distance of an alternative $\theta \in K$ from a given hypothesis H. A level α test Φ_o is said to be locally most powerful (LMP) if, given any other level α test Φ, there exists $\Delta > 0$ such that $\beta_{\Phi_\mathrm{o}}(\theta) \geq \beta_\Phi(\theta)$ for all $\theta \in K$ with $0 < d(\theta) < \Delta$.

We shall restrict ourselves to the real θ and the null hypothesis $H\colon \theta = \theta_0$; then it is natural to take $d(\theta) = \theta - \theta_0$ as a measure of distance for one-sided alternatives $K\colon \theta > \theta_0$ and $d(\theta) = |\theta - \theta_0|$ for two-sided alternatives.

In typical cases the LMP test can be found as a test maximizing the first derivative of the power function at the point of the null hypothesis θ_0. Computing the derivative of the power function $\beta_\Phi(\theta) = \int \Phi(x_1, \dots, x_n) \prod_{i=1}^n f(x_i, \theta)\,\mathrm{d}x_1, \dots, \mathrm{d}x_n$ of an arbitrary test Φ, we are often allowed to differentiate under the integral sign (the differentiability of power functions is closely connected with L_1-derivatives and the precise mathematical theory can be found e.g. in Witting [12]). Let $\dot{f}(x,\theta)$ denote the derivative of $f(x,\theta)$ with respect to θ. Then

$$\frac{\partial \beta_\Phi(\theta)}{\partial \theta} = \int \Phi(x_1, \dots, x_n) l(\boldsymbol{x}, \theta) \prod_{i=1}^n f(x_i, \theta)\,\mathrm{d}x_1, \dots, \mathrm{d}x_n,$$

where $l(\boldsymbol{x}, \theta) = \sum_{i=1}^{n} \frac{\dot{f}(x_i,\theta)}{f(x_i,\theta)}$ is the well-known Fisher *score function* (calculated as the logarithmic derivative of the likelihood $L(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$). From the Neyman–Pearson lemma we get that the LMP test has the critical region $l(\boldsymbol{x}, \theta_0) \geq C_\alpha$ where C_α is appropriately chosen constant to reach the prescribed level α.

Remark. Notice that if the LMP test is not simultaneously the UMP test, then typically (with an exception of the finite sample space) there does not exist a universal neighborhood over which the LMP maximize the power uniformly. To see it, it suffices to compare the power of LMP test with the power of the Neyman–Pearson test for an arbitrarily close simple alternative θ_1.

As we may look at a LMP test as at a one-sided version of *Rao score test* (Lagrange multiplier test) for a one-dimensional parameter, it seems natural to compare this test with the *Wald (W) test* and with the *likelihood ratio (LR) test.* The W test for a one-dimensional parameter has a simple critical region $\hat{\theta} \geq C_\alpha$, where $\hat{\theta}$ is an efficient (e. g. maximum likelihood) or at least a consistent estimate of the parameter θ. We shall modify the LR test, originally constructed as a two-sided test, in a one-sided version in the following way: The test rejects the null hypothesis when $\hat{\theta} > \theta_0$ and at the same time $LR = 2\{\log L(\boldsymbol{x}, \hat{\theta}) - \log L(\boldsymbol{x}, \theta_0)\} \geq C_\alpha$. The third alternative test, which will be considered in the case of one-sided hypothesis, is the test maximizing the power for the simple alternative $\theta_1 = \theta_0 + 2/\sqrt{n\, I_f}$, where I_f is the Fisher information at θ_0. Efron [3] suggested this test as an alternative to the LMP test for the distribution with the "big statistical curvature". We denote this test as the EFR test.

A natural condition imposed on tests of the simple hypothesis $H\colon \theta = \theta_0$ against two-sided alternatives $K\colon \theta \neq \theta_0$ is that of the local unbiasedness:

Definition 2. A level α test $\Phi_\circ$ is said to be locally unbiased, if there exists $\Delta > 0$ such that $\beta_{\Phi_\circ}(\theta) \geq \alpha$ for all θ with $0 < d(\theta) < \Delta$.

If the score function $l(\boldsymbol{X}, \theta_0)$ has a symmetric distribution under the null hypothesis ($\theta = \theta_0$), the locally most powerful locally unbiased (LMPLU) test has a simple critical region $|(\boldsymbol{x}, \theta_0)| \geq C_\alpha$. But as Jurečková [6] pointed out, the test with such a critical region is not locally unbiased if the distribution of $l(\boldsymbol{X}, \theta_0)$ is asymmetric. To find the LMPLU test in this situation, we put the first derivative of the power function equal to zero at θ_0 and under that condition we maximize the second derivative of power function at θ_0. Under sufficiently smooth densities, we get the critical region using the generalized Neyman–Pearson lemma:

$$\dot{l}(\boldsymbol{x}, \theta_0) + [l(\boldsymbol{x}, \theta_0)]^2 \geq C_1\, l(\boldsymbol{x}, \theta_0) + C_2, \tag{1}$$

here C_1, C_2 are constants determined by the conditions of size and local unbiasedness and $\dot{l}(x, \theta)$ denotes the derivative of $l(x, \theta)$ with respect to θ.

The situation is much more complicated when we consider simple hypothesis $H\colon \boldsymbol{\theta} = \boldsymbol{\theta}_0$ for a vector parameter ($\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$). We can easily see that the Definition 1 is not very useful here and a different approach has to be adopted. Suppose that the power function is twice continuously differentiable and let $\{\ddot{\beta}_\Phi(\boldsymbol{\theta}_0)\}$ be the matrix of the second derivatives of the power function of a test Φ at $\boldsymbol{\theta}_0$. Isaacson [5] proposed the **type D test** which maximizes the determinant of the matrix $\{\ddot{\beta}_\Phi(\boldsymbol{\theta}_0)\}$ subject to the conditions of size and unbiasedness. But the disadvantage of this test is that it is very difficult to construct. To overcome this inconvenience, Gupta and Vermaire [4] came up with the test which is also locally unbiased but which maximizes the trace of the matrix $\{\ddot{\beta}_\Phi(\boldsymbol{\theta}_0)\}$. Brown and Marden [1] showed that the test of this type is locally admissible. We shall refer to this test as the LMMPU (Locally Most Mean Powerful Unbiased) test.

Remark. From a geometrical point of view the type D test maximizes the Gaussian curvature of the power surface at $\boldsymbol{\theta}_0$ and the LMMPU test maximizes the mean curvature among all locally unbiased level α tests. This implies that while type D test locally minimizes the volume of an infinitesimal ellipse with a given power $\beta > \alpha$ (subject to the conditions of size and unbiasedness), the LMMPU test locally maximizes the average power over a spherical neighborhood of the null hypothesis.

3. EXAMPLES – MONTE CARLO STUDY

3.1. Tests on the location parameter of the double-exp. distribution

Let $X_1, \dots, X_n$ be a sample from the double-exponential distribution with the density $\frac{1}{2}e^{-|x-\theta|}$ and consider testing $H\colon \theta = 0$ against $K\colon \theta > 0$. We take the modest sample size $n = 10$ and prescribe the size $\alpha = 0.0546875 (\doteq 0.055)$. In this situation the maximum likelihood estimate is the median $\tilde{X} = (X_{(5)} + X_{(6)})/2$ (where $X_{(1)} \le \ldots \le X_{(n)}$ is the ordered sample). We consider the following tests:

1. *The sign test* : $\sum_{i=1}^n 1_{\{X_i>0\}} \ge k$ $(k = 8)$, which is the LMP test in this case (e. g. Lehmann [8]).
2. *The EFR test* : $\sum_{i=1}^n \left\{|X_i| - |X_i - \theta_|\right\} \ge C_\alpha$, where $\theta_1 = 0 + 2/\sqrt{10} \doteq 0.63$ and $C_\alpha = 1.26$.
3. *The Wald test* : $\tilde{X} \ge C_\alpha$ $(C_\alpha = 0.625)$.
4. *The LR test* :
   $\tilde{X} \ge 0 \ \&\ LR = 2\sum_{i=1}^n \left\{|X_i| - |X_i - \tilde{X}|\right\} \ge C_\alpha \quad (C_\alpha = 2.86)$.

The power function of the sign test can be easily computed, as the statistic of this test has the binomial distribution $Bi(n,p)$ with parameters $n = 10$ and $p = 1 - 0.5\,e^{-\theta}$. The power functions of the other tests are estimated by means of Monte Carlo simulation (in this example, as well as in the following ones, more than 1 000 000 "pseudorandom" samples were generated).

The Figure 1, which includes the envelope power function (also calculated using the Monte Carlo simulation), illustrates the differences of power functions, using the LMP test as a standard. Thus the power function for the LMP test appears as a "zero" straight line and if $\beta_\Phi(\theta)$ is the power function for another test, it is illustrated as $\beta_\Phi(\theta) - \beta_{\mathrm{LMP}}(\theta)$.

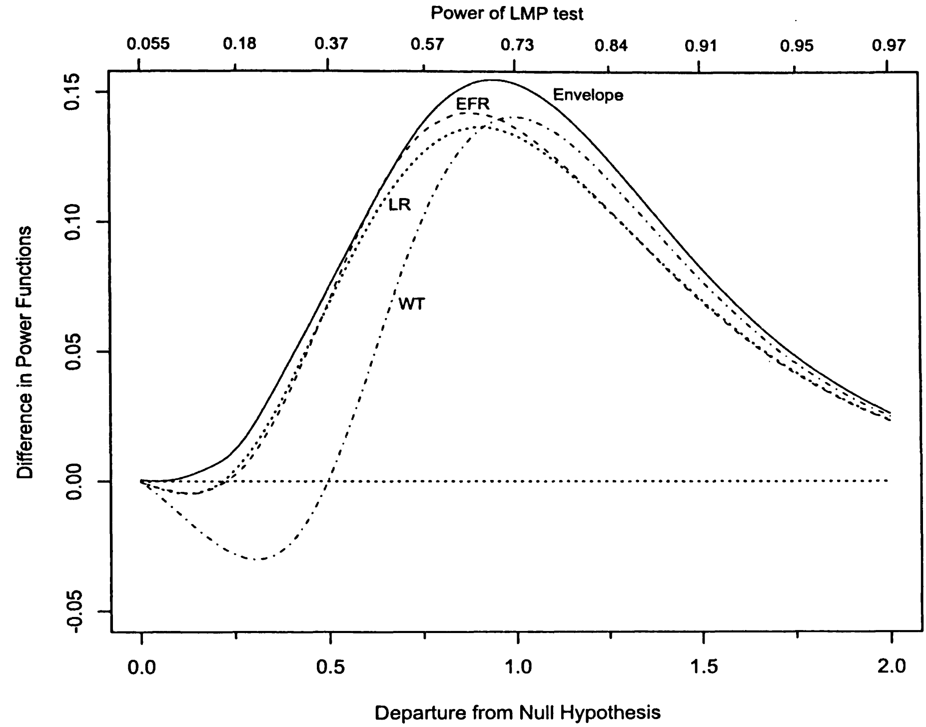

Fig. 1. The differences of power functions with the LMP test as a standard; *solid*: $\beta_{\mathrm{Env}} - \beta_{\mathrm{LMP}}$, *dashed*: $\beta_{\mathrm{EFR}} - \beta_{\mathrm{LMP}}$, *dotted*: $\beta_{\mathrm{LR}} - \beta_{\mathrm{LMP}}$, *dotdashed*: $\beta_{\mathrm{WT}} - \beta_{\mathrm{LMP}}$.

We can see that we pay quite a high price for the local optimality of the sign test. For more distant alternatives, the power of this test is considerably smaller than that of any other test. Hence, if our priority is not only the local sensitivity of the test, the EFR test and the LR test seem to be more preferable, since their power functions are quite near to the envelope power function over the whole alternative. We can also conclude that the W test is convenient when we look for a test strong against more distant alternatives. On the other hand, the sign test can be convenient in the practice, but it calls for a randomization if the prescribed size is not a natural level of the test. The W test is also quite simple to provide, since the formula $C_\alpha = -\log\{2(1 - \mathrm{B}_n(\alpha))\}$, where $\mathrm{B}_n(\alpha)$ is the $100(1-\alpha)$ per cent quantile of the beta distribution $\mathrm{B}(p,q)$ with parameters $p = \frac{n+1}{2}$, $q = \frac{n+1}{2}$, gives us the critical value which is exact for n odd and approximate for n even. For large n the critical value of the LR test can be approximated as well. However, the author does not know any simple approximation for the critical value of the EFR test in this case. Hence, the critical values should be tabulated to facilitate the practical application of the Efron test.

3.2. Tests on the shape parameter of the Weibull distribution

Let $X_1, \dots, X_n$ be a sample from the Weibull distribution with the density $f(x) = ax^{a-1}\exp(-x^a)$. Consider first testing the hypothesis $H\colon a = 1$ against $K\colon a > 1$. We take the modest sample size $n = 10$ and $\alpha = 0.05$ again. Let $\hat{a}$ be the maximum likelihood estimate of the parameter a. In this situation this estimate almost surely exists and is unique. Consider the following tests:

1. *The LMP test* – $\sum_{i=1}^{n} \{(1 - X_i)\log X_i + 1\} \geq C_\alpha \quad (C_\alpha = 5.76)$.
2. *The EFR test* – $\sum_{i=1}^{n} \{(\theta_1 - 1)\log(X_i) - (X_i)^{\theta_1} + X_i\} \geq C_\alpha$, where in our case $\theta_1 = 1 + 2/\sqrt{10 \cdot 1.82} \doteq 1.47$ and $C_\alpha = -2.33$.
3. *The Wald test* – $\hat{a} \geq C_\alpha \quad (C_\alpha \doteq 1.64)$.
4. *The LR test* – $\hat{a} \geq 1$ & $LR = n\log(\hat{a}) + \sum_{i=1}^{n} \left[(\hat{a} - 1)\log(X_i) - (X_i)^{\hat{a}} + X_i\right] \geq C_\alpha \quad (C_\alpha = 3.22)$.

Still another test, based on the extreme quotient $Q = \frac{X_{(n)}}{X_{(1)}}$, was proposed by Wong and Wong [13]. This test is easy to apply, scale-invariant and the critical values are easily computed. But we have found out that the power of this test is significantly smaller (relative loss is about 20 per cent) than the power of any of four considered tests. Thus this test is not considered in the sequel.

Figure 2 illustrates the differences in power functions, again using the LMP test as a standard. We can see how much we must pay for the local optimality. As well as in the test on the location parameter of the double-exponential distribution, we can reach the similar conclusions about power functions of the tests. But a closer look at Figures 1 and 2 shows that in this case the differences of the power functions are five times smaller. This is in agreement with the asymptotic theory which tells us that the shortcoming of the LMP test depends on the functional γ which Efron [3] called the statistical curvature. For the Weibull distribution with the shape parameter the curvature is $\gamma_a^2 \doteq 0.7$ (and in fact does not depend on the parameter $a > 0$). Efron pointed out that LMP tests work quite well if $\gamma_a^2/n < 1/8$, which is our case. Of course, such a simple rule is convenient for users, but it does not make sense if we are interested in a finer comparison of the tests based for example on the asymptotic deficiency. Unfortunately, we were not able to make such analysis in the case of the double-exponential distribution, since the density of this distribution is not smooth enough to define the Efron statistical curvature.

The smoothness of the Weibull density allows us to make an asymptotic expansions according to Chibisov [2]. He showed, that for large, n the W test behaves similarly as the Neyman–Pearson (NP) test for the simple alternative $\theta_W = \theta_0 + 2\,u_{1-\alpha}/\sqrt{n\,I_f}$; this can give us some intuition for which alternative is this test suitable. As $\theta_W = 1.77$ in our special situation ($n = 10$), we can see from Figure 2 that the approximation is not yet very precise. We have also checked that the approximation of critical values of an arbitrary NP test derived by Chibisov [2] is very accurate but unfortunately rather complicated.

Consider now the two-sided alternative $K\colon a \neq 1$. The sample size n and level α remain the same. We will investigate the following tests:

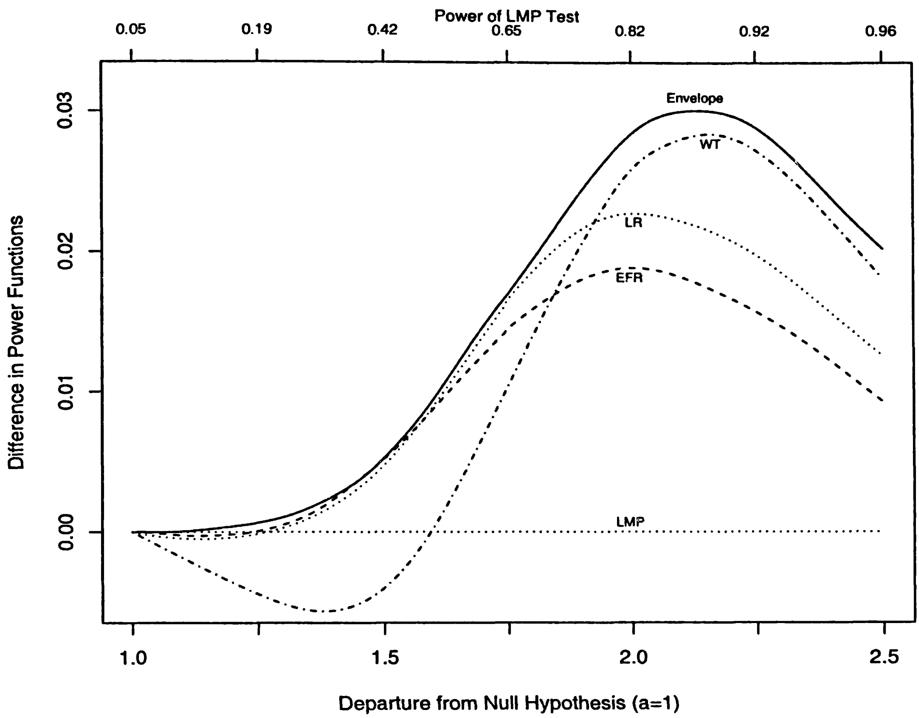

Fig. 2. The difference in power functions with the LMP test as a standard; *solid:* $\beta_{\mathrm{Env}} - \beta_{\mathrm{LMP}}$, *dashed:* $\beta_{\mathrm{EFR}} - \beta_{\mathrm{LMP}}$, *dotted:* $\beta_{\mathrm{LR}} - \beta_{\mathrm{LMP}}$, *dotdashed:* $\beta_{\mathrm{WT}} - \beta_{\mathrm{LMP}}$.

1. *The locally most powerful locally unbiased test* which has the critical region after a slight rearrangement of (1):

$$-\sum_{i=1}^{n} X_i (\log X_i)^2 + \left\{\sum_{i=1}^{n}(1 - X_i)\log X_i)\right\}^2 + \lambda_1 \left\{\sum_{i=1}^{n}(1 - X_i)\log X_i\right\} \geq \lambda_2, \tag{2}$$

   where $\lambda_1 = 24.07$ and $\lambda_2 = -80.37$ in our special case.

2. *The LR test* with the critical value 3.93 (to achieve the size $\alpha = 0.05$)

3. *The Wald test:* $\hat{a} \leq C_1$ or $\hat{a} \geq C_2$, where C_1, C_2 are found subject to the conditions of size and unbiasedness ($C_1 = 0.66$, $C_2 = 1.76$ in our case).

Similarly as in the previous examples, Figure 3 illustrates the difference in power functions with the LMPLU test as a standard. Although comparing with the one-sided test the region on which the LMPLU test is more powerful than the other tests

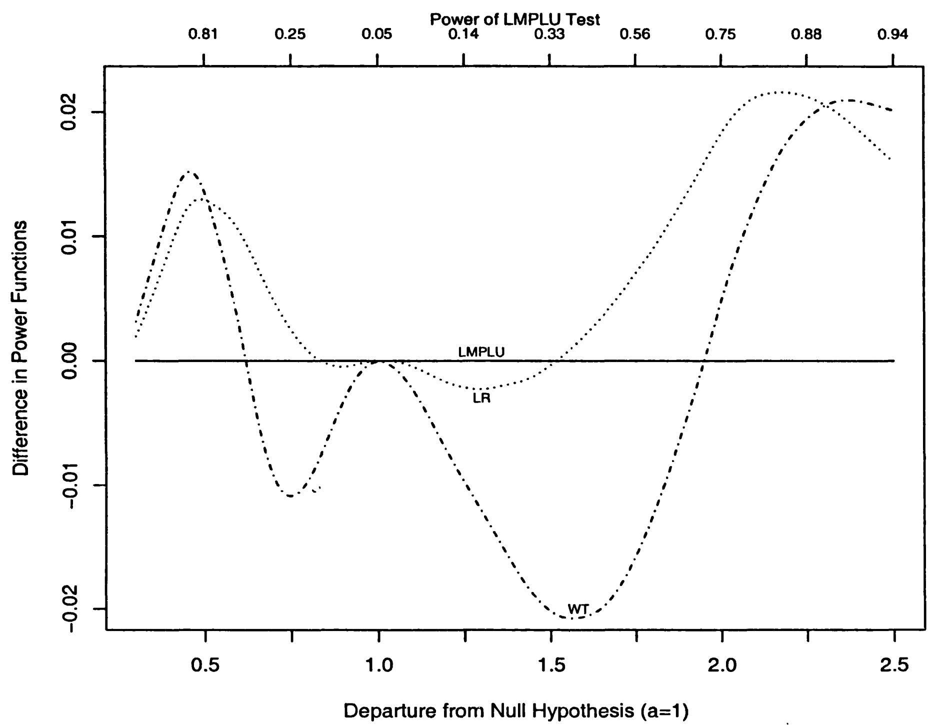

Fig. 3. The difference in power functions with the LMPLU test as a standard; *dotted*: $\beta_{\mathrm{LR}} - \beta_{\mathrm{LMPLU}}$, *dotdashed*: $\beta_{\mathrm{WT}} - \beta_{\mathrm{LMPLU}}$.

is considerably larger, the good behavior of the LR test is apparent again. The LR test is also locally unbiased by its nature and its critical value is well approximated by its asymptotic version. We notice that generally LR tests are efficient in the Bahadur sense under some mild conditions. The only disadvantage of LR tests is usually in calculating the maximum likelihood estimates. From this point of view, if our highest priority is not a local sensitivity, it is not worth using the LMPLU test whose construction is laborious (no approximations of constants λ_1, λ_2 in (2) are known to the author) and whose critical region is inscrutable. As in the previous examples, the W test can be recommended if we look for a test powerful especially against more distant alternatives. However, besides the rather poor local behavior, the computation of critical values can be difficult. As the distribution of $\hat{a}$ is rather skewed, the normal approximation does not work well for modest sample size and, moreover, if we insist on local unbiasedness, the critical values should be tabulated, because the acceptance region is not symmetric. So we can conclude that the W test is not very convenient for practical usage in this case.

3.3. Tests on the two-dimensional parameter of the normal distribution

Let $X_1, \ldots, X_n$ be a sample from the normal distribution with the density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, where both parameters μ and σ are unknown. Set $\boldsymbol{\theta} = (\mu, \sigma^2)^\mathsf{T}$ and consider testing $H: \boldsymbol{\theta} = (0,1)^\mathsf{T}$ against $K: \boldsymbol{\theta} \neq (0,1)^\mathsf{T}$. Let $\hat{\boldsymbol{\theta}} = (\bar{X}, s^2)^\mathsf{T}$ be the maximum likelihood estimate of the parameter $\boldsymbol{\theta}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$. The Fisher score function is

$$\boldsymbol{l}(\boldsymbol{X}, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta}) \log L(\boldsymbol{X}, \boldsymbol{\theta}) = \left(\sum \frac{X_i - \mu}{\sigma^2},\ -\frac{n}{2\sigma^2} + \sum \frac{(X_i - \mu)^2}{2\sigma^4}\right)^\mathsf{T}$$

and the Fisher information matrix is

$$\boldsymbol{J}_n(\boldsymbol{\theta}) = \boldsymbol{E}_{\boldsymbol{\theta}} \boldsymbol{l}(\boldsymbol{X}, \boldsymbol{\theta}) \boldsymbol{l}(\boldsymbol{X}, \boldsymbol{\theta})^\mathsf{T} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}.$$

We shall consider the following tests:

1. *Likelihood ratio (LR) test* with the test statistic
$LR = 2\{\log L(\boldsymbol{X}, \hat{\boldsymbol{\theta}}) - \log L(\boldsymbol{X}, \boldsymbol{\theta}_0)\} = n\bar{X}^2 + n\,(S^2 - 1 - \log(S^2))$.

2. *Wald (W) test* with the test statistic

$$WT = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\mathsf{T} \boldsymbol{J}_n(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \frac{n\bar{X}^2}{S^2} + \frac{n\,(S^2-1)^2}{2\,S^4}. \tag{3}$$

The matrix $\boldsymbol{J}_n(\hat{\boldsymbol{\theta}})$ is sometimes replaced with $\boldsymbol{J}_n(\boldsymbol{\theta}_0)$. But this would lead to a test which would be almost identical with the approximate D type test introduced later.

3. *Rao score (RS) test* with the test statistic
$RST = \boldsymbol{l}(\boldsymbol{X}, \boldsymbol{\theta}_0)^\mathsf{T} \boldsymbol{J}_n(\boldsymbol{\theta}_0)^{-1} \boldsymbol{l}(\boldsymbol{X}, \boldsymbol{\theta}_0) = n\,\bar{X}^2 + \frac{n}{2}\left(\bar{X}^2 + S^2 - 1\right)^2$.

4. *Approximate type D (AD) test* – with the test statistic

$$AD = n\bar{X}^2 + \frac{n-1}{2}\left(\frac{nS^2}{n-1} - 1\right)^2.$$

This test was proposed Isaacson [5], as he was not able to construct the exact type D test.

5. *Kolmogorov–Smirnov (KS) test* – this test was suggested by one of the referees.

6. *LMMPU test* ([4]) with the critical region $\left(\bar{X}^2 + S^2 - C\right)^2 + 4\bar{X}^2 \geq K^2$, where constants C, K are determined subject to the conditions of size and unbiasedness.

7. *Fisher test*: Let Φ stand for a distribution function of a standard normal variable and G_p for a distribution function of a variable with a χ^2-distribution with p degrees of freedom. The Fisher test is based on the statistic

$$Fisher = -2\,\log\left\{2\left[1-\Phi(|\sqrt{n}\bar{X}_n|)\right]\right\} - 2\,\log\left\{1-G_2[-2\log(H_n)]\right\},$$

where $H_n = 2\,G_{n-1}(S)$ if $S \leq \operatorname{med} G_{n-1}$ or $H_n = 2\,[1-G_{n-1}(S)]$ otherwise, and $\operatorname{med} G_{n-1}$ stands for the median of the distribution G_{n-1}.

This construction is known as Fisher's method of combining independent test statistics. Under the null hypothesis the statistic *Fisher* has χ^2-distribution with 4 degrees of freedom. The test is a one sample analogue of the test of Littel and Folks [9] who were dealing with the two sample problem. Analogously as [9], it can be shown that our test is optimal in the sense of Bahadur efficiency.

We prescribe the size $\alpha = 0.05$. It is well known that under the null hypothesis the statistics LR, WT, RS and also AD have asymptotically χ^2-distribution with 2 degrees of freedom. In practice we mostly approximate the critical values of these tests by the asymptotic ones. Let $\alpha_n = \boldsymbol{P}_{\boldsymbol{\theta}_0}(T_n > \chi^2_2(1-\alpha))$, where T_n is one of the mentioned statistics for a fixed sample size n. The Table 1 gives the true sizes of these tests when the asymptotic critical $\chi^2_2(0.95) \doteq 5.99$ value is used. As the true sizes are not always 0.05, in the sequel the estimates of the true critical values are used ensuring that all the test have approximately the size 0.05. We can also see that the W test defined in (3) is not very advisable unless the sample size is extremely large.

Table 1. True sizes of the tests when using the asymptotic critical value. The sign † (††) used when the true size is larger than 0.055 (0.10).

Test	α_{20}	α_{50}	α_{100}	α_{500}
LR test	$0.057^{\dagger}$	0.053	0.051	0.050
W test	$0.135^{\dagger\dagger}$	$0.087^{\dagger}$	$0.069^{\dagger}$	0.054
RS test	0.049	0.049	0.049	0.050
AD test	0.051	0.050	0.050	0.050

Let us now look at local properties of the proposed tests. To calculate the derivatives of power functions at $\boldsymbol{\theta}_0$ we can easily differentiate the power function

$$\beta_\Phi(\boldsymbol{\theta}) = \int \Phi(x_1,\ldots,x_n) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\, \mathrm{d}x_1,\ldots,\mathrm{d}x_n \tag{4}$$

under the integral sign. After some algebra we get

$$\left.\frac{\partial \beta_\Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{E}\,\Phi(\boldsymbol{X})\, \boldsymbol{l}(\boldsymbol{X},\boldsymbol{\theta}_0) \quad \text{and} \quad \left.\frac{\partial^2 \beta_\Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathsf{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{E}\,\Phi(\boldsymbol{X})\, \mathbf{A}(\boldsymbol{X},\boldsymbol{\theta}_0),$$

where

$$\mathbf{A}(\boldsymbol{X},\boldsymbol{\theta}_0) = \begin{bmatrix} n^2\bar{X}^2 - n, & \frac{n}{2}\sum_{i=1}^n X_i^2\bar{X} - \frac{n(n+2)}{2}\bar{X} \\ \frac{n}{2}\sum_{i=1}^n X_i^2\bar{X} - \frac{n(n+2)}{2}\bar{X}, & \frac{1}{4}\left(\sum_{i=1}^n X_i^2 - n\right)^2 - \sum_{i=1}^n X_i^2 + \frac{n}{2} \end{bmatrix}$$

Table 2. The derivatives of the power functions of the tests at $\boldsymbol{\theta}_0$ for two sample sizes.

	n = 20					n = 100				
Test	$\dot\beta_2$	$\ddot\beta_{11}$	$\ddot\beta_{22}$	$\det\{\ddot\beta_\Phi\}$	$\mathrm{tr}\{\ddot\beta_\Phi\}$	$\dot\beta_2$	$\ddot\beta_{11}$	$\ddot\beta_{22}$	$\det\{\ddot\beta_\Phi\}$	$\mathrm{tr}\{\ddot\beta_\Phi\}$
LR	0	2.78	1.59	4.44	4.38	0	14.5	7.5	108.3	22.0
W	−0.23	0.87	1.31	1.15	2.19	−0.45	11.1	6.6	73.3	17.7
RS	0.30	3.07	1.39	4.26	4.46	0.33	15.0	7.3	109.2	22.3
AD	0.27	2.87	**1.67**	**4.81**	4.55	0.30	14.9	**7.6**	**113.6**	22.5
KS	0.05	3.36	0.12	0.4	3.48	0.04	16.7	0.9	15.0	17.6
LMMPU	0	**3.71**	0.97	3.61	**4.69**	0	**20.5**	3.8	77.2	**24.3**
Fisher	0	2.98	0.05	0.15	3.03	0	14.9	0	0	14.9

and the expectation is taken under the null hypothesis. The expectation in (4) can be now easily estimated by means of the Monte Carlo simulation. For convenience we will denote the first and second derivatives of power functions of tests at $\boldsymbol{\theta}_0$ as

$$\dot\beta_i = \left.\frac{\partial\beta_\Phi(\boldsymbol{\theta})}{\partial\theta_i}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \text{and} \quad \ddot\beta_{ij} = \left.\frac{\partial^2\beta_\Phi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

The derivatives are given in Table 2. The table does not include the values of derivatives $\dot\beta_1$, $\ddot\beta_{12}(=\ddot\beta_{21})$, since their values are zero for each of considered tests.

Firstly, we should note that the W, RS, AD test and also KS test are not locally unbiased. In agreement with the asymptotic results of Peers [10], the W test is more powerful when $\sigma^2 < 1$, and the RS test is more powerful when $\sigma^2 > 1$. We can also see that the LMMPU really maximizes the trace of the matrix $\{\ddot\beta_\Phi\}$ and the AD test maximizes the determinant of this matrix although it is only an approximation of the type D test. Another apparent fact is that the ratio of the second derivatives tends to one for any pair of the LR test, W test, RS test and AD test. But this is not true for the KS, LMMPU tests and for the Fisher test, whose local performance seems to be completely different. The LMMPU test seems being extremely sensitive to small departures of μ from the null hypothesis and much less sensitive to small departures of σ^2 than these four tests. Also, the KS has a better-than-average sensitivity to the change in the location parameter. The sensitivity of the Fisher test to small departures of μ is even only average, and this test together with the KS test seem to be quite insensitive to small changes of the scale parameter σ.

Global properties

After the local considerations, let us shortly consider also the global behavior of the considered tests. Some of the results for the sample size $n = 20$ can be found in the Figure 4, where we can see the contour plot of the difference of the power function of the LR test with respect to the power functions of the other tests. The axes shows the true value of parameters. For the scale parameter the logarithmic transformation is used. Because of the symmetry of the power functions in the parameter μ, only $\mu \geq 0$ are considered. We see immediately that the LR test has a very good performance. Although this test is not uniformly most powerful, the

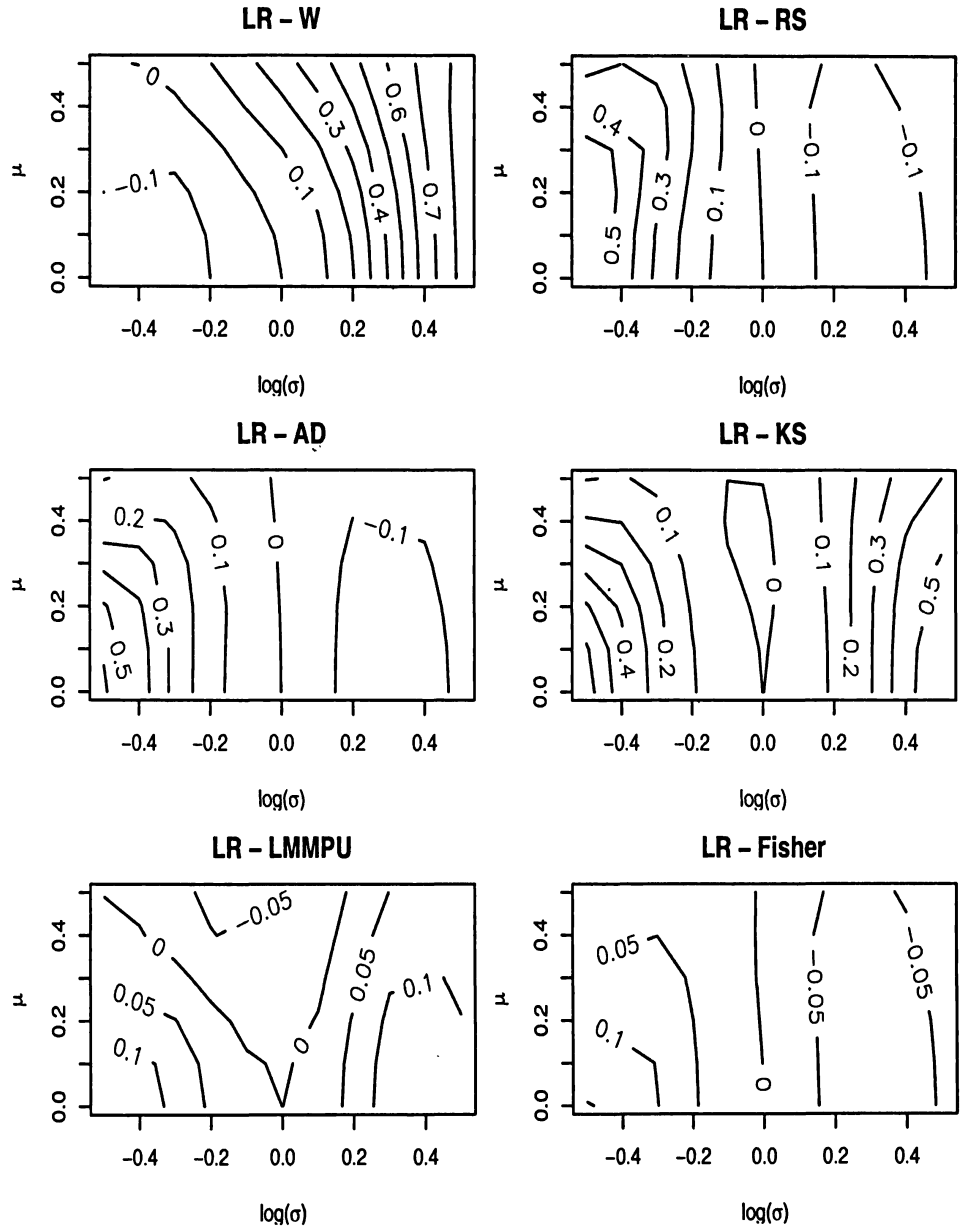

Fig. 4. The contour plot of the difference of the power function of the LR test with respect to the power functions of the other tests.

lack of power in the part of the parameter space is small in comparison with the excess of the power in the rest. This is especially true for the W, RS, AD and KS test which are not locally unbiased. From the figures we can easily deduce that for each of the tests there exists a region where the test is doing very badly and the area of this region is not negligible, especially when the sample size n is modest. This is particularly true for the W test which is doing very badly for $\sigma^2 > 0$. But also the KS test has a rather poor performance and it was completely outperformed by the LMMPU test. The only tests which are comparable with the LR test are the locally unbiased tests. Moreover, it is interesting that the Fisher test behaves very well, despite the small values of the derivative of the power function. This fact confirms the observations made in the previous examples, showing that the value of the derivative of the power function at the point of the null hypothesis gives only rather local information about the performance of the corresponding test.

As a conclusion, we recommend to choose the LMMPU test if the sensitivity to the changes in the location parameter is our main interest. However, it is difficult to compute the constants C, K of its critical region. On the other hand, the preference between the LR and Fisher tests might be a matter of taste. While the LR test does better for $\sigma^2 < 1$, the Fisher test is preferable in the opposite case. Table 2 also confirms a better local sensitivity of the LR test. However, the exact knowledge of the null distribution of the statistic Fisher strongly speaks in favour of this test, while the asymptotic critical value of the LR test, being used for small sample sizes, leads to size exceeding 0.05 (see Table 1). Therefore, the Fisher test may be a slightly more convenient. Nevertheless, our conclusions are in a good accordance with the theory of hypotheses testing, because both the most advisable tests are also optimal in the Bahadur sense.

ACKNOWLEDGEMENT

This work was partially supported by the Ministry of Education, Youth and Sports of the Czech Republic under Project MSM 113200008. The author thanks both referees for their careful reading of the manuscript and their valuable comments and proposals which led to a considerable improvement of the original text. The author also thanks Professor Jana Jurečková for suggesting the problem and for useful advice and recommendations during the preparation of his MSc Thesis.

(Received March 19, 2004.)

REFERENCES

[1] L. D. Brown and J. M. Marden: Local admissibility and local unbiasedness in hypothesis testing problems. Ann. Statist. *20* (1992), 832–852.

[2] D.M. Chibisov: Asymptotic expansions for some asymptotically pptimal tests. In: Proc. Prague Symp. on Asymptotic Statistics, Volume II (J. Hájek, ed.), Charles University, Prague 1973, pp. 37–68.

[3] B. Efron: Defining the curvature of a statistical problem (with application to second order efficiency). Ann. Statist. *3* (1975), 1189–1242.

[4] A.S. Gupta and L. Vermeire: Locally optimal tests for multiparameter hypotheses. J. Amer. Statist. Assoc. *81* (1986), 819–825.

[5] S. L. Isaacson: On the theory of unbiased tests of simple statistical hypothesis specifying the values of two or more parameters. Ann. Math. Statist. *22* (1951), 217–234.
[6] J. Jurečková: L_1-derivatives, score function and tests. In: Statistical Data Analysis Based on the L_1-Norm and Related Methods (Y. Dodge, ed.), Birkhäuser, Basel 2002, pp. 183–189.
[7] W. C. M. Kallenberg: The shortcomming of locally most powerful test in curved exponential families. Ann. Statist. *9* (1981), 673–677.
[8] E. L. Lehmann: Testing Statistical Hypothesis. Second edition. Chapman & Hall, New York 1994.
[9] R. C. Littel and J. L. Folks: A test of equality of two normal population means and variances. J. Amer. Statist. Assoc. *71* (1976), 968–971.
[10] H. W. Peers: Likelihood ratio and associated test criteria. Biometrika *58* (1971), 577–587.
[11] F. L. Ramsey: Small sample power functions for nonparametric tests of location in the double exponential family. J. Amer. Statist. Assoc. *66* (1971), 149–151.
[12] H. Witting: Mathematische Statistik I. Teubner–Verlag, Stuttgart 1985.
[13] P. G. Wong and S. P. Wong: A curtailed test for the shape parameter of the Weibull distribution. Metrika *29* (1982), 203–209.

Marek Omelka, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Prague 8. Czech Republic.
e-mail: omelka@karlin.mff.cuni.cz