Asymptotically Relative Efficiencies of Goodness of Fit Tests for the Two-Sample Problem.

Sherzad A. Mirakhmedov
Tashkent Sate Institute of Communication.
Department of Probability Theory & Statistics
Amir Temur str.108.
700084.Tashkent.
Uzbekistan.
E-mail address sherzod@ulug.silk.org

1. Introduction.

Determination the quality of the goodness of fit tests is important statistical and applied problem. There are several approaches to relate (asymptotic) performance of two different sequences of tests. The asymptotic quality (or efficiency) of tests (based on statistic T_n) depend on asymptotic behavior (as sample size n tends to infinity) of alternatives H_1, type 1 error $\alpha_n(T_n)$ and type 2 error $\beta_n(T_n)$. In the Pitman's approach $\alpha_n(T_n)$ and $\beta_n(T_n)$ are fixed and H_1 is sent to null hypothesis H_0; in the Bahadur's concept $\beta_n(T_n)$ and H_1 are fixed and $\alpha_n(T_n)$ is sent to zero; in the Hodges- Leahman case $\alpha_n(T_n)$ and H_1 are fixed and $\beta_n(T_n)$ is sent to zero. One might say that the concept of Pitman on the one hand and Bahadur and Hodges-Leahman approach on the other hand are the extreme points of view. Between these extreme cases there are intermediate approaches, when H_1 tends to H_0 but "not too fast" and $\alpha_n(T_n)$ tends to zero (for intermediate between Pitman and Bahadur case) or $\beta_n(T_n)$ tends to zero (for intermediate between Pitman and Hodges-Leahman case) also "not too fast". There are two critic point of view for Bahadur and Hodges-Leahman approaches: there seems no need to use statistical methods in the case of alternatives far from the null hypothesis; on the other hand for investigate these approaches we be in need of probabilities of a strong deviation for test statistics under null hypothesis for Bahadur case and under alternative for Hodges-Leahman case. The proof of such kind theorems is very problematic as usually. At least by these reasons the intermediate approaches are more important for applications (see Kallenberg ,1983). We consider mentioned approaches for determinate of asymptotic relative efficiency for three differ class of goodness of fit tests for verifying the validity of the hypothesis on form of a distribution and the hypothesis on homogeneity of two section. Here we consider the hypothesis on homogeneity of a two parent population, as example.

2. The hypothesis of homogeneity of two section.

Let $U_1, .., U_{n-1}$ and $V_1, .., V_k$ be independent random samples from two continuous distributions F(x) and G(x) respectively. The problem is to test the null hypothesis that these two parent populations are identical. That is F(x) = G(x). Without loss of generality we suppose that the support of both the probability distributions is unit interval (0,1) and that the first distribution F(x) is uniform on (0,1). We wish to test null hypothesis H_0 : G(y) = y , $0 \le y \le 1$, versus sequence of alternatives H_1 : $G(y)= y + L_n(y)\delta(n)$, $0 \le y \le 1$, where $L_n(0) = L_n(1) = 0$ and $\delta(n) \to 0$, as $n \to 0$, is chosen rather differ by depending on considered problem.

Let $0=U_{0n} \le U_{1n} \le .. \le U_{n-1,n} \le U_{nn}=1$ be the ordered of U observations and $(\eta_1, .., \eta_n)$ be the vector of spacing – frequencies , i.e. η_j is the number of V observations falling into the interval $[U_{j-1,n} , U_{jn})$, j =1,..,n.

We consider tests based on the statistics of the form

(1) $$R_n = \sum_{j=1}^{n} \varphi(\eta_j, \frac{j}{n-1}),$$

where $\varphi(u,v)$ is defined for u = 0,1,2,..,.and $v \in (0,1)$.

Let P_i , E_i, V_i denoted probability, expectation and variance under H_i ,i = 0 , 1, and let A_{in} and B_{in}^2 be the asymptotically mean of E_iR_n and V_iR_n . We suppose that $A_{1n} > A_{0n}$.

Follow to sense (but not strong definition) of Bahadur and Hodges- Leahman approaches we defined the asymptotically efficiency of test (based on) R_n by asymptotically value of (slopes)

$$\alpha_n(R_n) = -\ln P_0\{R_n \geq A_{1n}\} = -\ln P_0\{\tilde{R}_{on} \geq B_{0n}\mu_{0n}\}$$

and / or

$$\beta_n(R_n) = -\ln P_1\{R_n \geq A_{0n}\} = -\ln P_0\{\tilde{R}_{on} \leq -B_{1n}\mu_{1n}\} ,$$

where $\tilde{R}_{in} = (R_n - A_{in}) / B_{in}$, $\mu_{in} = (A_{1n} - A_{0n}) / B_{in}^2$.

If alternative H_1 fixed (i.e. $\delta(n) = \delta$ constant, its correspondent to Bahadur's and Hodjes-Leahman's cases) then $\mu_{in} \geq C > 0$, i = 0, 1.If $H_1 = H_{1n}$ closed to H_0 such that $B_{in}\mu_{in} \to C > 0$ we dial with Pitman approach . If H_{1n} closed to H_0 such that

(2) $\mu_{in} \to 0, \quad B_{in}\mu_{in} \to \infty$

we speak about intermediate approaches: between Pitman and Bahadur approaches, in this case the asymptotically efficiency defined by asymptotic behavior of $\alpha_n(R_n)$; between Pitman and Hodjes- Leahman approaches, in this case the asymptotically efficiency is defined by asymptotic behavior of $\beta_n(R_n)$. Note that in the Pitman and intermediate cases $B_{1n} = B_{0n}(1+0(1))$. The Pitman efficiency for the tests based on statistics of (1) was consider by Holst and Rao (1981). We dial with intermediate approaches in detail. We consider three type of intermediate efficiencies which defined by additional to (2) conditions: weak intermediate efficiency, if $\sqrt{n}\,\delta(n) \leq c\sqrt{\ln n}$; middle intermediate efficiency, if $\sqrt{n}\,\delta(n) = o(n^{1/6})$; strong intermediate efficiency if $\sqrt{n}\,\delta(n) = o(\sqrt{n})$. Particularly we show that on the class of test based on non- symmetric statistics of type (1) the linear test is optimal in the all type of intermediate efficiency senses. But if we consider tests based on symmetric statistics of type (1) the Dixon's test, which optimum in Pitman sense is still optimum in the both weak and middle efficiency senses and loss optimality property in the strong intermediate efficiency sense.

REFERENCES

Kallenberg ,W.C.M. (1983). Intermediate efficiency , theory and examples. Ann.Statist.,**11,** 170-182.

Holst, L. & Rao, S. (1981). Asymptotic spacings theory with application to the two-sample problem. Canadian J. Statist.,9, 79-89.

Ivchenco,G.I. & Mirakhmedov Sh.A.(1995). Large deviations and intermediate efficiency of decomposable statistics in multinomial scheme. Math. Methods Statist.,4, 294 –311.