

Asymptotic Relative Efficiency in Testing

Ya. Yu. Nikitin, yanikit47@gmail.com

University of Saint-Petersburg

September, 2010.

Keywords: Pitman efficiency, Bahadur exact slope, Hodges-Lehmann index, Kullback-Leibler information, goodness-of-fit, testing of symmetry, independence test, large deviations.

Asymptotic relative efficiency of two tests

Making a substantiated choice of the most efficient statistical test of several ones being at the disposal of the statistician is regarded as one of the basic problems of Statistics. This problem became especially important in the middle of XX century when appeared computationally simple but "inefficient" rank tests.

Asymptotic relative efficiency (ARE) is a notion which enables to implement in large samples the quantitative comparison of two different tests used for testing of the same statistical hypothesis. The notion of the asymptotic efficiency of *tests* is more complicated than that of asymptotic efficiency of *estimates*. Various approaches to this notion were identified only in late forties and early fifties, hence, 20–25 years later than in the estimation theory. We proceed now to their description.

Let $\{T_n\}$ and $\{V_n\}$ be two sequences of statistics based on n observations and assigned for testing the null-hypothesis H against the alternative A . We assume that the alternative is characterized by real parameter θ and for $\theta = \theta_0$ turns into H . Denote by $N_T(\alpha, \beta, \theta)$ the sample size necessary for the sequence $\{T_n\}$ in order to attain the power β under the level α and the alternative value of parameter θ . The number $N_V(\alpha, \beta, \theta)$ is defined in the same way.

It is natural to prefer the sequence with smaller N . Therefore the relative efficiency of the sequence $\{T_n\}$ with respect to the sequence $\{V_n\}$ is specified as the quantity

$$e_{T,V}(\alpha, \beta, \theta) = N_V(\alpha, \beta, \theta) / N_T(\alpha, \beta, \theta),$$

so that it is the reciprocal ratio of sample sizes N_T and N_V .

The merits of the relative efficiency as means for comparing the tests are universally acknowledged. Unfortunately it is extremely difficult to explicitly compute $N_T(\alpha, \beta, \theta)$ even for the simplest sequences of statistics $\{T_n\}$. At present it is recognized that there is a possibility to avoid this difficulty by calculating the limiting values $e_{T,V}(\alpha, \beta, \theta)$ as $\theta \rightarrow \theta_0$, as $\alpha \rightarrow 0$ and as $\beta \rightarrow 1$ keeping two

other parameters fixed. These limiting values $e_{T,V}^P$, $e_{T,V}^B$ and $e_{T,V}^{HL}$ are called respectively the Pitman, Bahadur and Hodges–Lehmann asymptotic relative efficiency (ARE), they were proposed correspondingly in [15], [2] and [9].

Only close alternatives, high powers and small levels are of the most interest from the practical point of view. It keeps one assured that the knowledge of these ARE types will facilitate comparing concurrent tests, thus producing well-founded application recommendations.

The calculation of the mentioned three basic types of efficiency is not easy, see the description of theory and many examples in [17], [13] and [20]. We only mention here, that Pitman efficiency is based on the central limit theorem for test statistics. On the contrary, Bahadur efficiency requires the large deviation asymptotics of test statistics under the null-hypothesis, while Hodges-Lehmann efficiency is connected with large deviation asymptotics under the alternative. Each type of efficiency has its own merits and drawbacks.

Pitman efficiency

Pitman efficiency is the classical notion used most often for the asymptotic comparison of various tests. Under some regularity conditions assuming asymptotic normality of test statistics under H and A , it is a number which has been gradually calculated for numerous pairs of tests.

We quote now as an example one of the first Pitman's results that stimulated the development of nonparametric statistics. Consider the two-sample problem when under the null-hypothesis both samples have the same continuous distribution and under the alternative differ only in location. Let $e_{W,t}^P$ be the Pitman ARE of the two-sample Wilcoxon rank sum test with respect to the corresponding Student test. Pitman proved that for Gaussian samples $e_{W,t}^P = 3/\pi \approx 0.955$, and it shows that the ARE of the Wilcoxon test in the comparison with the Student test (being optimal in this problem) is unexpectedly high. Later Hodges and Lehmann in [9] proved that

$$0.864 \leq e_{W,t}^P \leq +\infty,$$

if one rejects the assumption of normality and, moreover, the lower bound is attained at the density

$$f(x) = \begin{cases} 3(5-x^2)/(20\sqrt{5}) & \text{if } |x| \leq \sqrt{5}, \\ 0 & \text{otherwise.} \end{cases}$$

Hence the Wilcoxon rank test can be infinitely better than the parametric test of Student but their ARE never falls below 0.864. See analogous results in [18] where the calculation of ARE of related estimators is discussed.

Another example is the comparison of independence tests based on Spearman and Pearson correlation coefficients in bivariate normal samples. Then the value of Pitman efficiency is $9/\pi^2 \approx 0.912$.

In numerical comparisons, the Pitman efficiency appears to be more relevant for moderate sample sizes than other efficiencies [8]. On the other hand, Pitman

ARE can be insufficient for the comparison of tests. Suppose, for instance, that we have a normally distributed sample with the mean θ and variance 1 and we are testing $H : \theta = 0$ against $A : \theta > 0$. Let compare two significance tests based on the sample mean \bar{X} and the Student ratio t . As the t -test does not use the information on the known variance, it should be inferior to the optimal test using the sample mean. However, from the point of view of Pitman efficiency, these two tests are equivalent. On the contrary, Bahadur efficiency $e_{t, \bar{X}}^B(\theta)$ is strictly less than 1 for any $\theta > 0$.

If the condition of asymptotic normality fails, considerable difficulties arise when calculating the Pitman ARE as the latter may not at all exist or may depend on α and β . Usually one considers limiting Pitman ARE as $\alpha \rightarrow 0$. In [21] Wieand has established the correspondence between this kind of ARE and the limiting approximate Bahadur efficiency which is easy to calculate.

Bahadur efficiency

The Bahadur approach proposed in [2], [3] to measuring the ARE prescribes one to fix the power of tests and to compare the exponential rate of decrease of their sizes for the increasing number of observations and fixed alternative. This exponential rate for a sequence of statistics $\{T_n\}$ is usually proportional to some non-random function $c_T(\theta)$ depending on the alternative parameter θ which is called the *exact slope* of the sequence $\{T_n\}$. The Bahadur ARE $e_{V,T}^B(\theta)$ of two sequences of statistics $\{V_n\}$ and $\{T_n\}$ is defined by means of the formula

$$e_{V,T}^B(\theta) = c_V(\theta) / c_T(\theta).$$

It is known that for the calculation of exact slopes it is necessary to determine the large deviation asymptotics of a sequence $\{T_n\}$ under the null-hypothesis. This problem is always nontrivial, and the calculation of Bahadur efficiency heavily depends on advancements in large deviation theory, see [6], [7].

It is important to note that there exists an upper bound for exact slopes

$$c_T(\theta) \leq 2K(\theta)$$

in terms of Kullback–Leibler information number $K(\theta)$ which measures the "statistical distance" between the alternative and the null-hypothesis. It is sometimes compared in the literature with the Cramér–Rao inequality in the estimation theory. Therefore the absolute (nonrelative) Bahadur efficiency of the sequence $\{T_n\}$ can be defined as $e_T^B(\theta) = c_T(\theta)/2K(\theta)$.

It is proved that under some regularity conditions the likelihood ratio statistic is asymptotically optimal in Bahadur sense [3], [20, §16.6], [1].

Often the exact Bahadur ARE is uncomputable for any alternative θ but it is possible to calculate the limit of Bahadur ARE as θ approaches the null-hypothesis. Then one speaks about the *local* Bahadur efficiency.

The indisputable merit of Bahadur efficiency consists in that it can be calculated for statistics with non-normal asymptotic distribution such as Kolmogorov–Smirnov, omega-square, Watson and many other statistics.

Consider, for instance, the sample with the distribution function (df) F and suppose we are testing the goodness-of-fit hypothesis $H_0 : F = F_0$ for some known continuous df F_0 against the alternative of location. Well-known distribution-free statistics for this hypothesis are the Kolmogorov statistic D_n and omega-square statistic ω_n^2 . The following table presents their local absolute efficiency in case of six standard underlying distributions:

| Statistic | Distribution | | | | | |
|--------------|--------------|----------|---------|-------------------|--------|--------|
| | Gauss | Logistic | Laplace | Hyperbolic Cosine | Cauchy | Gumbel |
| D_n | 0.637 | 0.750 | 1 | 0.811 | 0.811 | 0.541 |
| ω_n^2 | 0.907 | 0.987 | 0.822 | 1 | 0.750 | 0.731 |

Table 1: Some local Bahadur efficiencies.

We see from Table 1 that the integral statistic ω_n^2 is in most cases preferable with respect to the supremum-type statistic D_n . However, in the case of Laplace distribution the Kolmogorov statistic is locally optimal, the same happens for the Cramér-von Mises statistic in the case of hyperbolic cosine distribution. This observation can be explained in the framework of Bahadur local optimality, see [13, Ch.6].

See also [13] for the calculation of local Bahadur efficiencies in case of many other statistics.

Hodges-Lehmann efficiency

This type of the ARE proposed in [9] is in the conformity with the classical Neyman-Pearson approach. In contrast with Bahadur efficiency, let us fix the level of tests and let compare the exponential rate of decrease of their second-kind errors for the increasing number of observations and fixed alternative. This exponential rate for a sequence of statistics $\{T_n\}$ is measured by some non-random function $d_T(\theta)$ which is called the Hodges-Lehmann index of the sequence $\{T_n\}$. For two such sequences the Hodges-Lehmann ARE is equal to the ratio of corresponding indices.

The computation of Hodges-Lehmann indices is difficult as requires large deviation asymptotics of test statistics under the alternative.

There exists an upper bound for the Hodges-Lehmann indices analogous to the upper bound for Bahadur exact slopes. As in the Bahadur theory the sequence of statistics $\{T_n\}$ is said to be *asymptotically optimal in the Hodges-Lehmann sense* if this upper bound is attained.

The drawback of Hodges-Lehmann efficiency is that most *two-sided* tests like Kolmogorov and Cramér-von Mises tests are asymptotically optimal, and hence

this kind of efficiency cannot discriminate between them. On the other hand, under some regularity conditions the one-sided tests like linear rank tests can be compared on the basis of their indices, and their Hodges-Lehmann efficiency coincides locally with Bahadur efficiency, see details in [13].

Coupled with three “basic” approaches to the ARE calculation described above, intermediate approaches are also possible if the transition to the limit occurs simultaneously for two parameters at a controlled way. Thus emerged the Chernoff ARE introduced by Chernoff [5], see also [12]; the intermediate, or the Kallenberg ARE introduced by Kallenberg [11], and the Borovkov-Mogulskii ARE, proposed in [4].

Large deviation approach to asymptotic efficiency of tests was applied in recent years to more general problems. For instance, the change-point, “signal plus white noise” and regression problems were treated in [16], the tests for spectral density of a stationary process were discussed in [10], while [19] deals with the time series problems, and the empirical likelihood for testing moment conditions is studied in [14].

References

- [1] Arcones, M. (2005). Bahadur efficiency of the likelihood ratio test. *Mathematical Methods of Statistics*, **14**, 163-179.
- [2] Bahadur, R.R. (1960). Stochastic comparison of tests. *Ann. Mathem. Statist.*, **31**, 276–295.
- [3] Bahadur, R.R. (1967). Rates of convergence of estimates and test statistics. *Ann. Mathem. Statist.*, **38**, 303–324.
- [4] Borovkov, A. and Mogulskii, A. (1993). Large deviations and testing of statistical hypotheses. *Siberian Adv. Math.*, **2**(3, 4); **3**(1, 2).
- [5] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations, *Ann. Mathem. Statist.*, **23**, 493–507.
- [6] Dembo, A. and Zeitouni, O. (1998). Large deviations techniques and applications. 2nd ed. , Springer, New York.
- [7] Deuschel, J.-D. and Stroock, D.(1989). Large Deviations. Academic Press, Boston.
- [8] Groeneboom, P. and Oosterhoff, J. (1981). Bahadur efficiency and small sample efficiency. *Intern. Statist. Review*, **49**, 127–141.
- [9] Hodges, J. and Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t -test. *Ann. Mathem. Statist.*, **26**, 324–335.
- [10] Kakizawa, Y. (2005). Bahadur exact slopes of some tests for spectral densities. *J. Nonparametric Stat.*, **17**, 745-764.

- [11] Kallenberg, W.C.M. (1983). Intermediate efficiency, theory and examples. *Ann. Statist.*, **11**, 170–182.
- [12] Kallenberg, W.C.M. (1982). Chernoff efficiency and deficiency. *Ann. Stat.*, **10**, 583-594.
- [13] Nikitin, Ya.(1995). Asymptotic Efficiency of Nonparametric Tests. Cambridge University Press.
- [14] Otsu T. (2010). On Bahadur efficiency of empirical likelihood. *Journ. of Econometrics*, **157**, 248-256
- [15] Pitman, E.J.G. (1949). Lecture Notes on Nonparametric Statistical Inference. Columbia University: Mimeographed.
- [16] Puhalskii, A., Spokoiny, V. (1998). On large-deviation efficiency in statistical inference. *Bernoulli*, **4**, 203-272.
- [17] Serfling, R. (1980). Approximation Theorems of Mathematical Statistics. John Wiley and Sons, New York.
- [18] Serfling, R. (2010). Asymptotic relative efficiency in estimation. *International Encyclopedia of Statistical Sciences*, Springer.
- [19] Taniguchi, M. (2001). On large deviation asymptotics of some tests in time series. *Journ. Stat. Plann. Inference*, **97**, 191-200.
- [20] Van der Vaart, A.W.(1998). Asymptotic Statistics. Cambridge University Press.
- [21] Wieand, H.S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.*, **4**, 1003–1011.

Based on an article from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science. Heidelberg: Springer Science +Business Media, LLC.