Asymptotic Relative Efficiency

A Tutorial*

(Prepared for the Encyclopedia of Biostatistics, John Wiley & Sons)

by

Anirban DasGupta
Purdue University

Technical Report #96-40

Department of Statistics
Purdue University

October, 1996

ASYMPTOTIC RELATIVE EFFICIENCY

Anirban DasGupta
Department of Statistics
Purdue University

Abstract

This is a brief tutorial on the topic of asymptotic relative efficiency and efficient estimation. Starting with common parametric problems, the article describes various common measures of relative efficiency, for both estimation and testing problems, and then discusses concepts of higher order efficiency and efficient estimation in complex models. The relevance of asymptotic efficiency in finite samples is also discussed. Several illustrative examples are given. The tone of this article is nontechnical.

*Research supported by NSF Grant DMS 9307727, NSA Grant 96G187.

1. ASYMPTOTIC RELATIVE EFFICIENCY

One of the most important problems of statistical practice is point estimation of an unknown parameter, say θ. In most cases, there are many apparently reasonable estimators of θ. For example, if one wants to estimate the mean of a normally distributed characteristic, it seems reasonable to estimate it by the mean of the characteristic from a sample. Since for normal random variables, the mean and median are the same, it also seems reasonable to use the median of the sample values as an estimate. Indeed, each estimate is a consistent estimate in this case; i.e., each estimate $\hat{\theta}$ satisfies $P(|\hat{\theta}-\theta|>\epsilon) \longrightarrow 0$ as $n \longrightarrow \infty$, for any given $\epsilon > 0$. Asymptotic efficiency is a common method to discriminate between two reasonable estimates when there is nothing to discriminate between them from the viewpoint of consistency.

Typically, two consistent estimates, say $\hat{\theta}_1$ and $\hat{\theta}_2$, will also have limiting normal distributions, i.e., $\sqrt{n}(\hat{\theta}_i-\theta) \longrightarrow N(0,\sigma_i^2(\theta)), i = 1,2$. In such a case, it is common to approximate the variance of $\hat{\theta}_i$ by $\sigma_i^2(\theta)/n$. The exact variance of $\hat{\theta}_i$ may be hard to calculate for a fixed sample size n, and thus the approximation really does become important. Since variance is a natural measure of accuracy of an estimate, it seems natural to define the efficiency of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ as the ratio $\sigma_2^2(\theta)/\sigma_1^2(\theta)$. In principle, the quantities $\sigma_1^2(\theta)$ and $\sigma_2^2(\theta)$ may depend on the unknown parameter θ. However, fortunately, in many important problems of statistics, they are just fixed positive constants not depending on θ, and therefore the Asymptotic Relative Efficiency (ARE) σ_2^2/σ_1^2 has the very appealing interpretation of being one number summarizing the performance of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$. The values of ARE are between 0 and ∞, and ARE > 1 corresponds to $\hat{\theta}_1$ being more efficient than $\hat{\theta}_2$.

Example 1. Suppose $X_1, X_2, \ldots X_n$ are independent observations from a $N(\theta, 1)$ population. Then, $\sqrt{n}(\bar{X}-\theta) \longrightarrow N(0,1)$ in distribution, and $\sqrt{n}(M-\theta) \longrightarrow N(0,\pi/2)$ in distribution, where M is the median of the sample data $X_1, X_2, \ldots, X_n$. Thus, according to our definition, the asymptotic relative efficiency of the sample median with respect to the sample mean for a normally distributed population is $2/\pi \approx .63$.

Interestingly, the situation reverses and the sample median becomes a more efficient estimate if the observations $X_1, X_2, \ldots, X_n$ are instead obtained from a population with a Double Exponential density, $1/2 \;\; e^{-|x-\theta|}$. In this case, $\sqrt{n}(\bar{X}-\theta) \longrightarrow N(0,2)$ and $\sqrt{n}(M-\theta) \longrightarrow N(0,1)$, and the asymptotic relative efficiency of the sample median with respect to the sample mean is 2.

Example 2. Sir Ronald Fisher, one of the founding fathers of much of statistics as we know it today, once had a communication with A. Eddington, a noted physicist, about how to estimate the standard deviation of a normal distribution. Thus, if $X_1, X_2, \ldots X_n$ are independent samples from the $N(\theta, \sigma^2)$ distribution where both parameters are unknown, the specific question was a comparison of the two estimates

$$\hat{\sigma}_1 = \frac{\Gamma\frac{n-1}{2}}{\sqrt{2}\Gamma\frac{n}{2}}\sqrt{\sum_{i=1}^{n}\ (x_i - \overline{x})^2}$$

$$\text{and} \quad \hat{\sigma}_2 = \frac{\sqrt{\pi}}{\sqrt{2(n-1)n}} \cdot \sum_{i=1}^{n} |x_i - \overline{x}|$$

where Γ (.) denotes the Euler Gamma function. Each estimate is unbiased for estimating σ. It is known that $\hat{\sigma}_1$ is the UMVUE (uniformly minimum variance unbiased estimate) of σ for each fixed sample size. So in fixed samples, $\hat{\sigma}_1$ is more efficient that $\hat{\sigma}_2$. An interesting question would be if even asymptotically, it has an ARE > 1. Using standard methods of large sample theory, it is seen that $\sqrt{n}(\hat{\sigma}_1 - \sigma) \longrightarrow N(0, \frac{\sigma^2}{2})$ and $\sqrt{n}(\hat{\sigma}_2 - \sigma) \longrightarrow N(0, \frac{\pi-2}{2}\sigma^2)$ in distribution as $n \longrightarrow \infty^2$. Thus, applying the definition of the ARE, the ARE of $\hat{\sigma}_1$ with respect to $\hat{\sigma}_2$ is $\pi - 2$, which is indeed larger than 1.

2. EFFICIENT ESTIMATES

A question of natural interest is the following: is there such a thing as a "most efficient" estimate, and how do we formulate such a concept? It turns out that in parametric estimation problems, it is indeed possible to easily formulate such a concept. Thus, suppose $X_1, X_2, \ldots, X_n$ are independent observations from a population with density $f(x|\theta)$. The quantity $I(\theta) = -E_\theta\left[\frac{d^2}{d\theta^2}f(x|\theta)\right]$, whenever the definition makes sense, is called the Fisher Information function. Let $\hat{\theta}$ be any estimate of θ that is consistent and asymptotically normal, i.e., $\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N(0, \sigma^2(\theta))$. Then, under some (frequently satisfied) regularity conditions on the density $f(x|\theta)$, it is true that $\sigma^2(\theta) \geq 1/I(\theta)$ (exceptions may occur at a "few" values of θ; this phenomenon is known as superefficiency, but we will not worry about this.) Thus, any estimate $\hat{\theta}$ of θ which actually attains the bound $\sigma^2(\theta) \equiv 1/I(\theta)$, can be legitimately called an EFFICIENT estimate of θ.

In a given problem, there are usually many efficient estimates of the unknown parameter θ. Standard methods of estimation typically result in efficient estimates, although in finite samples they may have different variances, biases, etc. This reinforces the fundamental issue that efficiency and ARE are intrinsically asymptotic indices in nature, but one hopes that if one estimate is more efficient than another according to the definition of ARE, in moderate samples it outperforms the other estimate as well. Among the standard methods of point estimation, maximum likelihood estimates and Bayes estimates typically are all efficient estimates, although exceptions to these general phenomena can and do occur. For instance, if the number of nuisance parameters grows with an increasing sample size, then maximum likelihood estimates of the most important parameter will usually not be efficient. Also, method of moment estimates may not be efficient even in very simple problems.

3. OTHER MEASURES OF EFFICIENCY

Besides point estimation, another very important problem of statistics is testing of hypothesis. As in point estimation, there are usually many reasonable tests of a specified hypothesis and it is useful to have a concept of efficiency of one test with respect to another. Various efficiency measures have been proposed here too, primarily among them Pitman efficiency and Bahadur efficiency.

The Pitman efficiency is defined in the following way: fix a type 1 error probability or level α, fix an alternative θ, and specify a desired power $1-\beta$ at this alternative. Let $n_i(\alpha, \beta, \theta)$ denote the minimum sample size required by the ith test, $i = 1, 2$, to achieve this goal. The Pitman efficiency of the first test with respect to the second is taken as the limit of the ratio $n_2(\alpha, \beta, \theta)/n_1(\alpha, \beta, \theta)$ as the alternative $\theta \longrightarrow \theta_0$ at a suitable rate, where θ_0 is exactly (the) boundary between the null and the alternative hypothesis.

Of course, there are a number of subtle issues involved here. Dependence of this limit on α and β would make universal interpretation of the efficiency value difficult. Also, the limit itself should exist for the definition to make any sense. Finally, the boundary value θ_0 may not be unique. In almost all problems that commonly occur, fortunately these subtleties do not cause any problems and one has a quite good efficiency measure.

Bahadur efficiency proceeds along the same lines, except one lets α tend to zero, keeping the alternative θ fixed. Thus the Bahadur efficiency can depend on both θ and the desired particular power $1-\beta$. Fortunately, again, usually dependence on β does not occur, although dependence on θ does. Thus, in contrast to the Pitman measure of efficiency, which is usually one single number, the Bahadur efficiency measure is a curve or a function, a function of the specified alternative θ. This is actually good in some sense, as one has an efficiency measure that discriminates between two competing tests based on which alternative values are really important in the given context. Bahadur [3]'s original approach was to compare the rates at which the P-Values corresponding to the two tests converge to zero at the specified θ. However, the two descriptions are equivalent.

Example 3. Suppose $X_1, X_2, \ldots X_n$ are independent observations from the Double Exponential density $1/2 \quad e^{-|x-\theta|}$ and we would like to test $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$. The following two tests appear to be reasonable:

Sign Test.	Count $N = \#$ sample values > 0.
	Reject H_0 if N is large.
Median Test.	Find the median M of the sample values.
	Reject H_0 if M is large (large positive).

The exact critical values for each test (i.e., what is to be regarded as a "large" value) can be found by large sample considerations (or even exactly, although it may involve numerical computing and randomization).

Now, it turns out that the Pitman efficiency of the Sign test with respect to the Median test is 1. So the Pitman measure does not discriminate between the two tests. Interestingly enough, the Bahadur efficiency does, and indeed, Sievers [27] shows that the Bahadur efficiency of the Sign test with respect to the Median test equals

$$e_B(\theta) = \log \frac{1}{\sqrt{4g(\theta)(1-g(\theta))}} \cdot [\log \ 2 + g(\theta) \log \ g(\theta) + (1-g(\theta)) \log(1-g(\theta))]^{-1}$$

where $g(\theta) = 1/2 \quad e^{-\theta}$

$e_B(\theta)$ is seen to be > 1 for any $\theta > 0$, establishing for Double Exponential data, the Sign test could be regarded as a better choice than the Median test. The following is a short table of the Bahadur efficiency at selected values of θ.

$\underline{\theta}$	0	.1	.25	.5	1	2	5
$e_B(\theta)$	1	1.003	1.017	1.057	1.182	1.545	3.214

4. RELEVANCE FOR FINITE SAMPLES

An important practical question is how closely the ARE approximates the ratio of the variances of two estimators in finite samples. It is difficult to give a very general answer to this, but in many examples the fixed sample relative efficiency seems to monotonically converge to the asymptotic relative efficiency, and the approximation becomes quite close for sample sizes ≥ 25. For example, for estimating the mean of a normal distribution, the ARE of the median with respect to the mean is off from the asymptotic value $2/\pi$ by at most 6.8% for sample sizes ≥ 20. Trimmed means are also common alternatives to usual sample averages as estimates of populations means. A certain amount of trimming of the smallest and the largest observations causes the effect of potential outliers to be decreased and has other nice advantages. Usually 5 or 10% trimming from each side is recommended; see Bickel and Lehmann [8] . For the 10% trimmed mean estimate for estimating a normal mean, the fixed sample relative efficiency with respect to the regular mean is at most 2.75% off from the asymptotic value .975 for sample sizes ≥ 20. Thus, there is some empirical evidence that the ARE reasonably approximates the fixed sample efficiency in moderate sample sizes. Expansions of the fixed sample quantity in which the asymptotic quantity is the leading term have also been attempted, frequently on a case by case basis. One can see Albers [2], Chandra and Ghosh [12], Groeneboom and Oosterhoff [18], Hodges and Lehmann [22] for such developments.

5. SENSITIVITY WITH RESPECT TO UNDERLYING DISTRIBUTION

It is entirely possible that one estimate or test is more efficient than another if samples are obtained from one distribution, but loses this advantage, may be drastically, for a fairly similar distribution. The comparison of the sample median and the sample mean is a good illustration. The mean has an efficiency of 1.57 with respect to the median if samples are known to come from a normal distribution, but this efficiency drops to .5 if data instead come from Double Exponential density, described before. Yet it is not easy to distinguish between the two distributions from moderate samples by using common methods, graphical or otherwise. Bickel and Lehmann [8] provide some concrete results in this direction. For example, they show that if samples are obtained from any density that is symmetric and unimodal about the mean, then the 5% trimmed mean has an ARE of at least .83 with respect to the mean for estimating the population mean, and of course, for a lot of particular such densities, the efficiency is substantially larger than 1. This may be used as an argument for using the 5% trimmed mean if concerns about the exact density from which one is sampling exist. More information on this can be found in Huber [23] and Staudte and Sheather [28].

6. CONCEPTS OF HIGHER ORDER EFFICIENCY

As stated before, in parametric estimation problems, it is customary to have many estimates which are fully efficient. It then becomes necessary, at least from a theoretical standpoint, to have a criterion to distinguish among them. The concept of Second order efficiency (now usually referred to as third order efficiency) was introduced to address this issue. See Rao [25, 26], and Akahira and Takeuchi [1], Efron [14], Ghosh, Sinha and Wieand [16] for later developments. The idea is to derive an expansion for n times the variance of a statistic, in which the leading term is the Fisher information function and subsequent terms decrease in reciprocals of powers of $1/n$. The second term is used as a comparison among different estimators, or simply for selecting an estimator which is first as well as second order efficient. In a peculiar result, Pfanzagl [24] showed that often first order efficiency automatically implies second order efficiency as well, making it necessary to consider higher order efficiencies as a basis for comparison and selection. Bickel, Chibisov, Van Zwet [6] and Ghosh [15] expand on these results and ideas.

7. COMPLEX MODELS

Parametric models using a given functional form for the density are often convenient choices, and perhaps restrictive. Similarly, the assumption that the sample observations are independent also often does not meet the criteria of realism. Real data often have a positive serial correlation or have a time series character. Models broader than parametric can be of various types; nonparametric models used to be the popular alternative. In standard nonparametric modelling, very little is assumed about the density besides some minimal features, mostly to do with shape and symmetry, unimodality, etc. Intermediate between fully parametric and fully nonparametric models, are the recent semiparametric models. It should be mentioned that complexity may arise not just from more complex models, but also because the quantity to be estimated is more complex than a simple thing like a mean or variance. For example, Bickel and Ritov [10], Hall and Marron [20] talk about estimating $\int(f'(x))^2dx$, the integrated squared derivative of a density.

Efficient estimation in complex models has a large literature, of a substantially more difficult nature, as expected. The literature includes Begun et. al. [4], Bickel et. al. [7] Chen [13], Groenboom and Wellner [19], Hasminskii and Ibragimov [21], and Van der Vaart [29], Efficiency for dependent samples also has a substantial literature, but is more scattered. Grenander and Rosenblatt [17] is a classic reference which established efficiency of the sample mean for estimating the mean of a stationary process under quite mild conditions. Brockwell and Davis [11] gives more information and discusses more problems. Efficient estimation in a relatively recent class of time series models known as long memory processes appears to be of a totally different qualitative nature. This can be seen in Beran [5].

References

[1] Akahira, M. and Takeuchi, K. (1981). Asymptotic Efficiency of Statistical Estimators: Concepts and Higher order Asymptotic Efficiency, Springer-Verlag, New York.

[2] Albers, W.(1974). Asymptotic Expansions and the Deficiency Concept in Statistics, Mathematical Centre Tract, **58**, *Amsterdam.*

[3] Bahadur, R. R. (1960). Stochastic comparison of tests, *Ann. Math Statist.*, **31**, 276-295.

[4] Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A. (1983). Information and Asymptotic Efficiency in Parametric- Semiparametric Models, *Ann. Statist.*, **11**, 432-452.

[5] Beran, J. (1995). Statistics for Long Memory Processes, Chapman and Hall, New York.

[6] Bickel, P. J., Chibisov, D. M., and van Zwet, W. R.(1981). On Efficiency of First and Second Order, *Int. Statist.Review*, **49**, 169- 175.

[7] Bickel, P. J., Klassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). Efficient and Adaptive Estimation for Semiparametric Models, Johns Hopkins Univ. Press.

[8] Bickel, P. J. and Lehmann, E. L. (1976). Descriptive Statistics for Nonparametric Models,*Ann. Statist.*, **4**, 1139 - 1158.

[9] Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the errors in variables model, *Ann. Statist.*,**15**, 513-540.

[10] Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates, Sankhya, Ser A, **50**, 381-393.

[11] Brockwell, P. J. and Davis, R. A. (1991). Time Series: Theory and Methods, Springer-Verlag, New York.

[12] Chandra, T. and Ghosh, J. K. (1978). Comparison of tests with same Bahadur efficiency, Sankhya , Ser. A, **40**, 253-277.

[13] Chen, H. (1995). Asymptotically efficient estimation in semiparametric generalized linear models, *Ann. Statist.* **23**, 1102- 1129.

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

[15] Ghosh, J. K. (1994). Higher Order Asymptotics, Institute of Mathmatical Statistics, Hayward, California.

[16] Ghosh, J. K., Sinha, B. K. and Wieand, H. S. (1980). Second order efficiency of mle with respect to a bounded bowl-shaped loss function, *Ann. Statist.*, **8**, 506-521.

[17] Grenander, U. and Rosenblatt,M.(1957). Statistical Analysis of Stationary Time Series, Wiley, New York.

[18] Groenboom, P. and Oosterhoff, J. (1980). Bahadur efficiency and Small Sample Efficiency: A Numerical Study, Mathematische Centrum , Report SW 68-80, Amsterdam.

[19] Groenboom, P. and Wellner, J. A. (1992). Information Bounds and Nonparametric Maximum Likelihood Estimation, Birkhauser, Boston.

[20] Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives, Statist. Prob. Letters, **6**, 109-115.

[21] Hasminskii, R. Z. and Ibragimov, I. A. (1983). On asymptotic efficiency in the presence of an infinite dimensional nuisance parameter, USSR - Japan Symp, 195- 229, Springer - Verlag, New York.

[22] Hodges, J. L. Jr. and Lehmann, E. L. (1970). Deficiency, *Ann. Math. Statist.*, **41**, 783-801.

[23] Huber, P. J. (1981). Robust Statistics, Wiley, New York.

[24] Pfanzagl, J. (1979). First order efficiency implies second order efficiency, Contributions to Statistics (J. Hajek Memorial Volume, J. Jureckova ed.), pp 167-196, Academia, Prague.

[25] Rao, C. R. (1961). Asymptotic efficiency and limiting information, Proc. Fourth Berkeley Symp. ,I, 531-545.

[26] Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion), JRSS, Ser B, **24**, 46-72.

[27] Sievers, G. L. (1969). On the probability of large deviations and exact slopes, *Ann. Math. Statist.*, **40**, 1908-1921.

[28] Staudte, R. G. and Sheather,S. S. (1990). Robust Estimation and Testing, Wiley, New York.

[29] van der Vaart, A. (1996). Efficient maximum likelihood estimation in semiparametric mixture models, *Ann. Statist.*, **24**, 862-878.

(ASYMPTOTIC RELATIVE EFFICIENCY
EFFICIENT ESTIMATES
PITMAN EFFICENCY
BAHADUR EFFICIENCY
HIGHER ORDER EFFICIENCY)

Anirban DasGupta