

# Commentary on Basu (1956)

Robert Serfling\*

*University of Texas at Dallas*

March 2010

Prepared for forthcoming

*Selected Works of Debabrata Basu* (Anirban DasGupta, Ed.),  
Springer Series on Selected Works in Probability and Statistics.

## Asymptotic relative efficiency of estimators

For statistical estimation problems, it is typical and even desirable that more than one reasonable estimator can arise for consideration. One natural and time-honored approach for choosing an estimator is simply to compare the sample sizes at which the competing estimators meet a given standard of performance. This depends upon the chosen measure of performance and upon the particular population distribution  $F$ .

For example, we might compare the sample mean versus the sample median for location estimation. Consider a distribution function  $F$  with density function  $f$  symmetric about an unknown point  $\theta$  to be estimated. For  $\{X_1, \dots, X_n\}$  a sample from  $F$ , put  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$  and  $\text{Med}_n = \text{median}\{X_1, \dots, X_n\}$ . Each of  $\bar{X}_n$  and  $\text{Med}_n$  is a *consistent* estimator of  $\theta$  in the sense of convergence in probability to  $\theta$  as the sample size  $n \rightarrow \infty$ . To choose between these estimators we need to use further information about their performance. In this regard, one key aspect is *efficiency*, which answers:

**Question A** *How concentrated about  $\theta$  is the sampling distribution of  $\hat{\theta}$ ?*

## Criteria for asymptotic relative efficiency

### Variance as a measure of performance

A simple and natural criterion relative to the above question is the *variance* of the sampling distribution: the smaller this variance, the more “efficient” is that estimator. In this

---

\*Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: [serfling@utdallas.edu](mailto:serfling@utdallas.edu). Website: [www.utdallas.edu/~serfling](http://www.utdallas.edu/~serfling). Support by NSF Grant DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

regard, let us consider “large-sample” sampling distributions. For  $\overline{X}_n$ , the classical central limit theorem tells us: if  $F$  has finite variance  $\sigma_F^2$ , then the sampling distribution of  $\overline{X}_n$  is approximately  $N(\theta, \sigma_F^2/n)$ , i.e., Normal with mean  $\theta$  and variance  $\sigma_F^2/n$ . For  $\text{Med}_n$ , a similar classical result [10] tells us: if the density  $f$  is continuous and positive at  $\theta$ , then the sampling distribution of  $\text{Med}_n$  is approximately  $N(\theta, 1/4[f(\theta)]^2 n)$ . On this basis, we consider  $\overline{X}_n$  and  $\text{Med}_n$  to perform equivalently at respective sample sizes  $n_1$  and  $n_2$  if

$$\frac{\sigma_F^2}{n_1} = \frac{1}{4[f(\theta)]^2 n_2}.$$

Keeping in mind that these sampling distributions are only approximations assuming that  $n_1$  and  $n_2$  are “large”, we define the *asymptotic relative efficiency (ARE)* of  $\text{Med}$  to  $\overline{X}$  as the *large-sample limit* of the ratio  $n_1/n_2$ , i.e.,

$$\text{ARE}(\text{Med}, \overline{X}, F) = 4[f(\theta)]^2 \sigma_F^2. \quad (1)$$

More generally, for any parameter  $\eta$  of a distribution  $F$ , and for estimators  $\hat{\eta}^{(1)}$  and  $\hat{\eta}^{(2)}$  which are approximately  $N(\eta, V_1(F)/n)$  and  $N(\eta, V_2(F)/n)$ , respectively, the *ARE of  $\hat{\eta}^{(2)}$  to  $\hat{\eta}^{(1)}$*  is given by

$$\text{ARE}(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, F) = \frac{V_1(F)}{V_2(F)}. \quad (2)$$

*Interpretation:* If  $\hat{\eta}^{(2)}$  is used with a sample of size  $n$ , the number of observations needed for  $\hat{\eta}^{(1)}$  to perform equivalently is  $\text{ARE}(\hat{\eta}^{(2)}, \hat{\eta}^{(1)}, F) \times n$ .

In view of the asymptotic normal distribution underlying the above formulation of ARE in estimation, we may also characterize the ARE given by (2) as the limiting ratio of sample sizes at which the *lengths of associated confidence intervals at approximate level  $100(1-\alpha)\%$* ,

$$\hat{\eta}^{(i)} \pm \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \sqrt{\frac{V_i(F)}{n_i}}, \quad i = 1, 2,$$

converge to 0 at the same rate, when holding fixed the coverage probability  $1 - \alpha$ . (In practice, of course, consistent estimates of  $V_i(F)$ ,  $i = 1, 2$ , are used in forming the CI.)

The treatment of ARE for consistent asymptotically normal estimators using the variance criterion had been long well established by the 1950s – see [1] for a string of references.

## Probability concentration as a measure

Instead of comparison of asymptotic variance parameters as a criterion, one may quite naturally compare the *probability concentrations* of the estimators in any  $\varepsilon$ -neighborhood of the target parameter  $\eta$ :  $P(|\hat{\eta}_n^{(i)} - \eta| > \varepsilon)$ ,  $i = 1, 2$ . When we have

$$\frac{\log P(|\hat{\eta}_n^{(i)} - \eta| > \varepsilon)}{n} \rightarrow \gamma^{(i)}(\varepsilon, \eta), \quad i = 1, 2,$$

as is typical, then the ratio of sample sizes  $n_1/n_2$  at which these concentration probabilities converge to 0 at the same rate is given by  $\gamma^{(1)}(\varepsilon, \eta)/\gamma^{(2)}(\varepsilon, \eta)$ , which then represents another ARE measure for the efficiency of estimator  $\widehat{\eta}_n^{(2)}$  relative to  $\widehat{\eta}_n^{(1)}$ . This entails approximation of the sampling distribution in the tails. Accordingly, instead of central limit theory the relevant tool is *large deviation theory*, which is rather more formidable. In the context of hypothesis testing, Chernoff [3] argued that when the sample size approaches infinity it is appropriate to minimize both Type I and Type II error probabilities, rather than minimizing one with the other held fixed. He developed an ARE index essentially based on tail probability approximations. See also [10, 1.15.4] for general discussion.

## How compatible are these two criteria?

Those who have been fortunate enough to observe D. Basu in action, as I was when we were colleagues at Florida State University in the early 1970s, know his talent for inquiring into the boundaries of any good idea. Relative to the present context, when the variance and probability concentration criteria were just becoming established criteria in the 1950s, stemming from somewhat differing orientations, it was Basu who thought of exploring their compatibility. Basu [2] provides an example in which not only do the variance-based and concentration-based measures disagree on which estimator is better. but they do so in the most extreme sense: one ARE is infinite at every choice of  $F$  in a given class  $\Omega$ , while the other ARE is zero for every such  $F$ .

Basu's construction is straightforward and worth discussing, so we briefly examine some details. For  $X_1, \dots, X_n$  an i.i.d. sample from  $N(\mu, 1)$ , put

$$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i \quad \text{and} \quad S_n = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Basu defines the estimation sequences for  $\mu$  given by  $T = \{t_n\}$  and  $T' = \{t'_n\}$ , with

$$t_n = (1 - H_n)\bar{X}_n + nH_n \quad \text{and} \quad t'_n = \bar{X}_{[\sqrt{n}]},$$

where  $H_n = 1$  if  $S_n > a_n$  and 0 otherwise, and  $a_n$  satisfies  $P(S_n > a_n) = 1/n$ . He shows that  $\sqrt{n}(t_n - \mu) \xrightarrow{d} N(0, 1)$ . Since also  $n^{-1/4}(t'_n - \mu) \xrightarrow{d} N(0, 1)$ , it follows that the ARE according to (2) is given by

$$\text{ARE}(t_n, t'_n, N(\mu, 1)) = \lim_{n \rightarrow \infty} \frac{n^{-1}}{n^{-1/2}} = 0. \quad (3)$$

He also shows that the corresponding ARE based on concentration probabilities for any fixed choice of  $\varepsilon$  is given by

$$\lim_{n \rightarrow \infty} \frac{n^{-1}}{o(n^{-1})} = \infty. \quad (4)$$

An immediate observation about this example is that it is not pathological. Rather, it employs ordinary ingredients characteristic of typical application contexts.

Another important aspect is that the disagreement between the two notions of ARE is as extreme as possible. Not merely differing with respect to whether the ARE is  $< 1$  or  $> 1$ , here one version is *infinite* at every choice of  $F$  in the class  $\Omega = \{N(\mu, 1) : -\infty < \mu < \infty\}$ , while the other version is *zero* for every such  $F$ .

The details of proof yield the interesting corollary that (4) also gives the *concentration probability* ARE of  $t_n$  versus simply  $\bar{X}_n$ . Thus the estimator which is *optimal* under the *variance* ARE criterion is *infinitely nonoptimal* under the concentration probability ARE criterion.

A slight variation on Basu's  $\{t_n\}$  provides an example of *superefficient estimator*, similar to that of J. L. Hodges (see Le Cam, 1953). discussed in Lehmann and Casella (1998). Put

$$t_n^* = A(1 - H_n)\bar{X}_n + nH_n$$

for some constant  $A \neq 1$ . Then we have that  $\sqrt{n}(t_n^* - \mu) \xrightarrow{d} N(0, A^2) + \lim_{n \rightarrow \infty} \sqrt{n}\mu(A - 1)$ , i.e.,  $\sqrt{n}(t_n^* - \mu)$  converges to  $\pm\infty$  if  $\mu \neq 0$  and otherwise converges to  $N(0, A^2)$ . Therefore, in the case that  $\mu = 0$  and  $A < 1$ , the estimator  $t_n^*$  outperforms the “optimal” estimator. See Lehmann and Casella (1998) for useful discussion of superefficiency.

We see that the content of Basu's example, like all of his contributions to statistical thinking, reflects great ingenuity and insight applied very productively to useful purposes.

## Subsequent developments

The impact of Basu [2] thus has been to motivate stronger interest in “large deviation (LD) approaches” to ARE. For example, Bahadur [1] follows up with a deep discussion of this approach along with many constructive ideas. Quite a variety of LD and related moderate deviation approaches are discussed in Serfling [10, Chap. 10]. More recently, Puhalskii and Spokoiny [9] provide an extensive treatment of the LD approach in statistical inference. For convenient elementary overviews on ARE in estimation and testing, see DasGupta [4], Serfling [11], and Nikitin [8], for example.

## Acknowledgment

Very helpful suggestions of Anirban DasGupta are greatly appreciated and have been used to improve the manuscript. Also, support by NSF Grant DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

## References

- [1] Bahadur, R. R. (1960). Asymptotic relative efficiency of tests and estimates. *Sankhyā* **22** 229–252.
- [2] Basu, D. (1956). On the concept of asymptotic relative efficiency. *Sankhyā* **17** 193–196.

- [3] Chernoff, H. (1952). A measure of asymptotic relative efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics* **23** 493–507.
- [4] DasGupta, A. (1998). Asymptotic relative efficiency. *Encyclopedia of Biostatistics, Vol. I*, 210–215, P. Armitage and T. Colton (Eds.). John Wiley, New York.
- [5] Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *University of Calif. Publ. in Statistics* **1** 277–330.
- [6] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. Springer.
- [7] Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press.
- [8] Nikitin, Y. (2010). Asymptotic relative efficiency in testing. *International Encyclopedia of Statistical Sciences* (M. Lovric, ed.). Springer.
- [9] Puhalskii, A. and Spokoiny, V. (1998). On large-deviation efficiency in statistical inference. *Bernoulli* **4** 203–272.
- [10] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.
- [11] Serfling, R. (2010). Asymptotic relative efficiency in estimation. *International Encyclopedia of Statistical Sciences* (M. Lovric, ed.). Springer.