Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 136 (2006) 4194–4203
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Estimating the center of symmetry: Is it always better to use larger sample sizes?

Sudip Bose

Department of Statistics, The George Washington University, Washington, DC 20052, USA

Received 13 January 2003; accepted 3 June 2005
Available online 7 July 2005

Abstract

We show that for symmetric stable distributions with index $\gamma < 1$, the performance of the sample mean, as an estimator of the center of symmetry, worsens as the sample size increases, where performance is measured by concentration around the estimand. We look at order statistics and show that for symmetric distributions (not necessarily stable), the performance of the sample median as an estimator of the center of symmetry improves as the sample size increases.

MSC: 60E07; 60E15; 62F10

Keywords: Stochastic concentration; Stochastic order; Peakedness; Majorization; Universal domination; Order statistics; Linear estimators; Stable distributions; Symmetric stable distributions; Symmetric distributions; Pitman closeness; Marginal Pitman closeness

1. Introduction

The laws of large numbers are among the most fundamental results in statistics. Khinchine's version of the WLLN tells us that if $X_1, X_2, \ldots, X_n$ is an *iid* sample from a distribution and $E(X_i) = \theta$ exists, then the sample mean $\bar{X}_n$ is consistent for θ, i.e.

$$\bar{X}_n \xrightarrow{P} \theta,$$

E-mail address: sudip@gwu.edu.

0378-3758/$ - see front matter © 2005 Published by Elsevier B.V.
doi:10.1016/j.jspi.2005.06.003

where $\xrightarrow{P}$ denotes convergence in probability. In other words, it pays to collect more data, and to use a 'central' statistic to estimate a measure of central tendency.

Suppose that a random variable X has distribution symmetric about its median θ. (Of course, if $E|X| < \infty$, its mean $E(X) = \theta$ also.) It is well-known that for certain thick-tailed distributions, the sample mean is not a consistent estimator of the center of symmetry θ. This holds for the Cauchy distribution, and more generally for symmetric stable distributions with exponent γ between 0 and 1. In addition to the above limiting results, one can prove exact results about centering (and using all the data) not being beneficial for estimating the median of symmetric stable distributions with exponent γ between 0 and 1. Thus, the quick answer to the question asked in the title of this paper is "no, if one compares sample means for varying sample sizes, but yes if one compares sample medians for varying sample sizes". We describe below the criterion for comparing performances of estimators and the majorization-based concept of "more centered" or "less centered".

1.1. Stochastic order and stochastic concentration

The random variable X (with c.d.f. F_X) is said to be stochastically as small as Y (with c.d.f. F_Y), denoted by $X \overset{\text{st}}{\leqslant} Y$ if for all $a \in \mathscr{R}$,

$$F_X(a) \geqslant F_Y(a),$$

and is said to be stochastically smaller than Y, denoted by $X \overset{\text{st}}{<} Y$ if in addition, $F_X(a_0) > F_Y(a_0)$ for some real a_0.

If $X \overset{\text{st}}{\leqslant} Y$ then for any real-valued non-decreasing function ϕ, (assuming the expectations exist,)

$$E[\phi(X)] \leqslant E[\phi(Y)].$$

This leads to the idea of stochastic concentration of estimators. Given two estimators $T(X)$, $U(X)$ of a parameter $\theta \in \Theta$, T is said to have greater stochastic concentration about θ than U does, if for all θ, the distribution of $|T(X) - \theta|$ is stochastically smaller than the distribution of $|U(X) - \theta|$. This stochastic concentration ordering is essentially the same as the peakedness order of Shaked and Shanthikumar (1994, p. 210), based on the notion of peakedness in Birnbaum (1948). A consequence of greater stochastic concentration (Hwang, 1985) is that if L is any loss function satisfying

$$L(a, \theta) = g(|a - \theta|),$$

where, for each θ, $g : [0, \infty) \to [0, \infty)$ is a non-decreasing function (g may depend on θ),

$$R(T, \theta) = E_\theta L(T(X), \theta) \leqslant E_\theta L(U(X), \theta) = R(U, \theta) \quad \forall \theta \in \Theta.$$

In other words not only is T at least as likely as U to be within a neighborhood $(\theta - a, \theta + a)$ of the true parameter, it is at least as good with regard to a broad class of loss functions which includes absolute error, squared error, weighted squared error, and any other symmetric loss function where the loss is at least as large when a is farther away from θ than when it is