Effects of Smoothing on Distribution Approximations

Peter Hall
Australian National University

and

Xiao-Hua Zhou
Indiana University School of Medicine

Abstract

We show that a number of apparently disparate problems, involving distribution approximations in the presence of discontinuities, are actually closely related. One class of such problems involves developing bootstrap approximations to the distribution of a sample mean when the sample includes both ordinal and continuous data. Another class involves smoothing a lattice distribution so as to overcome rounding errors in the normal approximation. A third includes kernel methods for smoothing distribution estimates when constructing confidence bands. Each problem in these classes may be modelled in terms of sampling from a mixture of a continuous and a lattice distribution. We quantify the proportion of the continuous component that is sufficient to "smooth away" difficulties caused by the lattice part. The proportion is surprisingly small — it is only a little larger than $n^{-1} \log n$, where n denotes sample size. Therefore, very few continuous variables are required in order to render a continuity correction unnecessary. The implications of this result in the problem of sampling both ordinal and continuous data are discussed, and numerical aspects are described through a simulation study. The result is also used to characterise bandwidths that are appropriate for smoothing distribution estimators in the confidence band problem. In this setting an empirical method for bandwidth choice is suggested, and a particularly simple derivation of Edgeworth expansions is given.

Keywords: Bandwidth, bootstrap, confidence band, confidence interval, continuity correction, coverage error, Edgeworth expansion, kernel methods, mixture distribution.

1 Introduction

1.1 Smoothing in distribution approximations

Rabi Bhattacharya has made very substantial contributions to our understanding of normal approximations in statistics and probability. None has been less important and influential than his exploration and application of smoothing as it is related to distribution approximations. For example, his development of ways of smoothing multivariate characteristic functions lies at the heart of his pathbreaking work on Berry-Esseen bounds and other measures of rates of