

THE BOOTSTRAP IN HYPOTHESIS TESTING

PETER J. BICKEL¹ AND JIAN-JIAN REN²

University of California at Berkeley and Tulane University

We propose a unifying principle which identifies a very broad class of hypotheses and statistics for which a suitable application of the m out of n bootstrap yields asymptotically correct critical values and power for contiguous alternatives. We also show that this attractive principle can fail in situations which the m out of n bootstrap can deal with (Bickel, Götze and van Zwet, 1997)(BGvZ). We formalize the m out of n bootstrap theory for testing and show that under mild conditions, it provides correct significance level, asymptotic power under contiguous alternatives, and consistency. We conclude with simulation results supporting the asymptotics.

AMS subject classifications: 62G09, 62G10.

Keywords and phrases: Censored data, change points, empirical processes, goodness of fit, Hadamard differentiability, manifolds, U statistics.

1 Introduction

It is logically clear but not always noted that the usual nonparametric bootstrap (the n out of n bootstrap) fails in setting critical values for test statistics in hypothesis testing. The problem is that hypothesis restrictions are not reflected adequately by the empirical distribution when one is resampling as many observations as one has in the sample. For example, Freedman (1981) points out that in setting confidence intervals for the usual slope estimate for regression through the origin, one must resample not the residuals but the residuals centered at their mean. If one considers setting confidence bands as the dual of hypothesis testing, a moment's thought will show that not centering the residuals is tantamount to not imposing the model requirement for the hypothesis tests that the expectation of the error is 0. For more recent examples, see Härdle and Mammen (1993), Mammen (1992) and Bickel, Götze and van Zwet (1997) (BGvZ). BGvZ note that the m out of n bootstrap, $m \rightarrow \infty$, $\frac{m}{n} \rightarrow 0$, is in principle usable. In particular, Bickel and Ren (1996) study the following situation: testing for goodness of fit with doubly censored data where the usual bootstrap as usual fails and finding a distribution approximating the truth under H_0 is difficult. They propose using the m out of n bootstrap to set the critical value of the test

¹Research partially supported by NSF Grant DMS 9504955.

²Research partially supported by NSF Grants DMS 9510376 and 9626535/9796229.