THE m OUT OF n BOOTSTRAP AND GOODNESS OF FIT TESTS WITH DOUBLY CENSORED DATA

Peter J. Bickel[1,2] and Jian-Jian Ren
University of California-Berkeley and University of Nebraska-Lincoln

AMS* 1991 *subject classifications. Primary 62G10; secondary 62E20.

Key words and phrases. Bootstrap, Cramér-von Mises statistics, doubly censored data, goodness of fit tests.

Abstract. This paper considers the use of the m out of n bootstrap (Bickel, Göetze, and van Zwet, 1994) in setting critical values for Cramér-von Mises goodness of fit tests with doubly censored data. We show that, as might be expected, the usual n out of n nonparametric bootstrap fails to estimate the null distribution of the test statistic. We show that if the m out of n bootstrap with $m \to \infty$, $m = o(n)$ is used to set the critical value of the test, the proposed testing procedure is asymptotically level α, has the correct asymptotic power function for $\sqrt{n}$ alternatives and is asymptotically consistent.

1. Introduction.

It is logically clear but not always evident or appreciated that the usual nonparametric bootstrap (the n out of n bootstrap) should fail when one tries to estimate the distribution of test statistics under a semiparametric (restricted nonparametric) hypothesis and ignores the restrictions imposed by the hypothesis. For example, Freedman (1981) points out that in setting confidence intervals on 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 that the expectation of the error is 0.

[1]Supported by the Sonderforschungsbereich 343 'Diskrete Strukturen in der Mathematik' University of Bielefeld

[2]NATO Collaborative Research Grant 5-2-05/920650

For more recent examples, see Härdle and Mammen (1993), Mammen (1992), Bickel, Göetze and van Zwet (1994). The same phenomenon occurs when one considers hypothesis testing problems using a censored or incomplete sample. Imposing hypothesis restrictions for censored data can be very complicated given the unknown distributions of the censoring variables. In the situation we consider here, it is impossible to estimate the censoring distribution! Specifically, we consider Cramér-von Mises goodness of fit tests with doubly censored data. The problem is introduced in Section 2, where we show that the usual n out of n nonparametric bootstrap fails to estimate the null distribution of the test statistic. In Section 3, we propose that one uses the m out of n bootstrap to set the critical value of the test and show that the proposed testing procedure is asymptotically consistent and has correct power against $\sqrt{n}$ alternatives with the proofs deferred to Section 5. For the exponential distribution family, Section 4 presents a simulation result which compares the power functions using the true critical value and the one by the m out of n bootstrap.

One should note that the testing procedure proposed here includes the goodness of fit tests for proportional hazards model with right censored data (Wei, 1984) as a special case, and that the proposed procedure easily applies to many other models with censored data or with uncensored data. An alternative to the m out of n bootstrap method in the model of this paper, the Fredholm Integral Equation (FIE) method, has been constructed by Ren (1993). The m out of n bootstrap method, although less powerful to second order, has the advantage of easy implementation and generalization to essentially any semiparametric testing problem.

In collaboration with Göetze, van Zwet and others we are in the process of studying this broad applicability and the critical issue of choice of m. In the example we study in Section 4, $m = \sqrt{n}$ is chosen arbitrarily and evidently performs reasonably well.

We close the basic heuristics of this method which are evident in the classic problem of testing $H : \mu = 0$ vs $K : \mu > 0$ when $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, 1)$. The correct test is to reject if $\sqrt{n}\bar{X}_n \geq z(1-\alpha)$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $z(1-\alpha)$ is the $1-\alpha$ quantile of $\mathcal{N}(0,1)$. Suppose that we instead use the bootstrap critical value of $\sqrt{n}\bar{X}_n$. That is call $\hat{c}_{m,n}$ the $1-\alpha$ quantile of the distribution of $\sqrt{m}\ \bar{X}^*_m$, where $X^*_1, \ldots, X^*_m$ are a sample of size m with replacement from $X_1, \ldots, X_n$ and reject if $\sqrt{n}\bar{X}_n \geq \hat{c}_{m,n}$. It is easy to see that if $m \to \infty$ with $m = o(n)$, then $\hat{c}_{m,n} = z_{1-\alpha} + o_p(1)$ not only