ELSEVIER

Statistics & Probability Letters 59 (2002) 281–292

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Assessing the quality of bootstrap samples and of the bootstrap estimates obtained with finite resampling

Yannis Yatracos

Department of Statistics and Applied Probability, The National University of Singapore, 3 Science Drive 2, 117543 Singapore, Singapore

Received January 2001; received in revised form May 2002

Abstract

It is seen in simulations and confirmed theoretically that: (i) the loss in accuracy of the Monte Carlo approximation of the bootstrap estimate can be infinite, due to the additional uncertainty introduced by finite resampling, and (ii) the dimension of the data or the estimate of interest affect drastically the quality of the bootstrap samples and estimates.

Based on the findings, directions are provided to improve the bootstrap methodology.

MSC: primary 62C15; 62C99; 62F10; 62G09; secondary 62G30; 62H12

Keywords: Bootstrap estimate; Bootstrap geometry; Bootstrap sample; Inadmissibility; Jackknife; Model dimension

1. Introduction

The bootstrap was introduced by Efron (1979) to determine either the accuracy or the distribution of an estimate T, when this was not achieved with classical statistical methods. The method has gained much popularity and is considered a "general purpose tool" in statistics (Young, 1994), and an extension of the maximum likelihood "plug-in principle" (Efron and Tibshirani, 1993, denoted by E&T, in the sequel). Nevertheless, there has not been much discussion in the bootstrap literature on simultaneous estimation and the associated simulations (Hu and Kalbfleisch, 1999), and as a referee mentioned "most theoretical assessments of bootstrap methodology ignore the fact that (in practice) the number of bootstrap samples is finite".

E-mail address: yatracos@stat.nus.edu.sg, stayy@nus.edu.sg (Y. Yatracos).

PII: S0167-7152(02)00196-7

In applications, the bootstrap is used to estimate either the bias, or the variance or other parameters of the distribution F_T of a statistic T, all of which depend on the unknown distribution $F_{n,\theta}$ (resp. F_n) of the original sample $\mathscr{X}_{n\,x\,d}$ (denoted by $\mathscr{X}_d$). Thus, the bootstrap is used to estimate a quantity $\zeta = \beta(T) = \alpha(F_{n,\theta})$ (resp. $\alpha(F_n)$), and related questions are examined when a finite number B of bootstrap samples is obtained. How much can we learn from these samples about $F_{n,\theta}$ (resp. F_n)? To what extent can the user of the bootstrap be sure of "hitting its target" ζ (Young, 1994)? What measures the loss in accuracy of the bootstrap estimate $\hat{\zeta}^*_{n,B}$, $B < \infty$?

It is seen that $\hat{\zeta}^*_{n,B}$ is inadmissible with respect to the mean square error (MSE) compared with $\hat{\zeta}_n = E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)$, due to the additional randomisation introduced by finite resampling. The amount of inadmissibility $E[\mathrm{Var}(\zeta^*_{n,B}|\mathscr{X}_d)]$ depends on $F_{n,\theta}$ (resp. F_n) and therefore can be substantial. Often $\hat{\zeta}_n$ is the theoretical bootstrap estimate, and then $E[\mathrm{Var}(\zeta^*_{n,B}|\mathscr{X}_d)]$ measures the loss in accuracy using $\hat{\zeta}^*_{n,B}$ instead. It is also seen from different angles and for a large class of probability models that, when the dimension d of the observations or that of T increases, the quality of the bootstrap sample deteriorates compared with the original sample, and the chance $\hat{\zeta}^*_{n,B}$ will be more and more erroneous than $\hat{\zeta}_n$ increases to one. These phenomena appear in simulations in Example 1, estimating a vector of variances; $\hat{\zeta}^*_{n,B}$ is compared with a jackknife estimate $\hat{\zeta}_{n,J}$ (Quenouille, 1949) for which it always holds $\mathrm{Var}(\hat{\zeta}_{n,J}|\mathscr{X}_d) = 0$.

The results indicate that rules of thumb in the choice of B should be used with caution. Also, that the bootstrap samples and $\hat{\zeta}^*_{n,B}$ cannot often provide more information than the original sample $\mathscr{X}_d$ and $\hat{\zeta}_n$, and suggestions are given how to stay respectively near each other. Hall and Presnell (1997) lead in the same direction, selecting bootstrap samples by comparison with $\mathscr{X}_d$. Then, the chance to obtain better $\hat{\zeta}^*_{n,B}$ increases, but the potential of substantial inadmissibility remains.

E&T provide an introduction to the bootstrap also containing advanced material; Hall (1992) provides a rigorous mathematical foundation on the subject.

2. The background, the motivation and the approach

2.1. The background

The usual elements in the bootstrap methodology are: the dimension d of each observation (called *model dimension*); the size n of the original sample $\mathscr{X}_d$; the size m and the number B of the bootstrap samples; in parametric models, the estimate $\hat{\theta}_n$ used to generate the bootstrap samples, and its dimension k; the model $F_{n,\theta}$ (resp. F_n) and $F_{n,\hat{\theta}_n}$ (resp. the c.d.f. $\hat{F}_n$), which constitute the statistician's original and bootstrap worlds (E&T, Beran, 1994); the nature and the dimension of the estimate T, and the quantity of interest $\zeta = \beta(T)$ with its bootstrap estimate and $\hat{\zeta}^*_{n,B}$.

The steps to apply the bootstrap methodology are:

(i) a distribution is chosen, that is usually either $F_{n,\hat{\theta}_n}$ or $\hat{F}_n$, or F_T.
(ii) B samples of size $m = n$ are subsequently drawn from this distribution, and
(iii) an estimate $\hat{\zeta}^*_{n,B}$ is obtained based on these samples.

Bootstrap optimality results require B to increase to infinity before n, while either d or k stays fixed. In linear regression, when k is large, the bootstrap faces problems (Bickel and Friedman, 1981, p. 1211, l. 2–3). Comparing bootstrap with wild bootstrap in linear models with random design, Mammen (1993) let d increase to infinity with n to preserve asymptotically the model's dimensionality, and argued that the rates of convergence are different because the bootstrap "has to mimic a complex stochastic structure of high-dimensional distributions".

2.2. The motivation and the approach

We examine problems a practitioner with n observations faces with finite resampling, when d or the dimension of T is large. The examples, the geometries and the propositions are motivated by the heuristics that follow; the word distance denotes the Hellinger distance H (see Section 5; for its properties see Le Cam and Yang, 1990):

(a) For fixed n, as d increases, it is expected that the distance between $F_{n,\theta}$ (resp. F_n) and $F_{n,\hat{\theta}_n}$ (resp. $\hat{F}_n$) will increase, and asymptotically (in d) these distributions become singular. Thus, when d is large, the chance to obtain a bootstrap sample "near" $F_{n,\theta}$ (resp. F_n) and a "good" bootstrap estimate of a location parameter is small. A similar situation occurs when the dimension of T is large.

(b) Most important, drawing B bootstrap samples provides the statistical experiment with additional randomisation. Thus, $\hat{\zeta}^*_{n,B}$ is not the function of a sufficient statistic and can be improved.

A new element is introduced to confirm (a): the estimates $\hat{\theta}^*_{n,i}, i = 1,\dots,B$ obtained by plugging the bootstrap samples in $\hat{\theta}_n$. The distances $H(F_{n,\theta}, F_{n,\hat{\theta}^*_{n,i}})$, $i = 1,\dots,B$, reflect the quality of the bootstrap samples used to estimate $\zeta = \alpha(F_{n,\theta})$, and thus indirectly $F_{n,\theta}$ and θ; those smaller than $H(F_{n,\theta}, F_{n,\hat{\theta}_n})$ determine the better bootstrap samples. For several distributions, the Euclidean distance $\|\theta - \eta\| \sim H(F_{n,\theta}, F_{n,\eta})$, and the proportion of better bootstrap samples is measured by the probability $P[H(F_{n,\theta}, F_{n,\hat{\theta}^*}) \leqslant H(F_{n,\theta}, F_{n,\hat{\theta}_n})] = P[\|\hat{\theta}^* - \theta\| \leqslant \|\hat{\theta}_n - \theta\|]$.

The effect of the additional randomisation on the quality of $\hat{\zeta}^*_{n,B}$ as estimate of ζ is observed through the variance identity:

$$\mathrm{Var}(\hat{\zeta}^*_{n,B}) = \mathrm{Var}[E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)] + E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)] = \mathrm{Var}(\hat{\zeta}_n) + E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]. \quad (1)$$

$\hat{\zeta}^*_{n,B}$ and $\hat{\zeta}_n = E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)$ have the same bias. $E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]$ is the expected *cushion error* due to finite resampling; its value depends on $F_{n,\theta}$ (resp. F_n). In some cases, $\sup_\theta \{\mathrm{MSE}(\hat{\zeta}^*_{n,B}, \zeta) - \mathrm{MSE}(\hat{\zeta}_n, \zeta)\} = \infty$; $\hat{\zeta}_n$ may be the theoretical bootstrap estimate, as for example when ζ is an expectation. By letting B increase to infinity before n, the term $E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]$ vanishes, since $\zeta^*_{n,B}$ becomes, solely, the function of $\mathscr{X}_d$.

To study the effect of $F_{n,\theta}$ (resp. F_n) in the accuracy of $\zeta^*_{n,B}$, several θ-values are considered, and d or the dimension of T may increase to infinity; w.l.o.g. B is fixed, assuming $E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]$ increases to infinity for selected θ-values and since it is decreasing in B. The findings "at the limit" provide an idea about the effects of the model and the large dimension. Some results are proved assuming independence of the coordinate vectors and are not expected to improve under dependence.

3. Examples

In Example 1, the effects of the values of the model parameters and of d are revealed by comparing in simulations $\hat{\zeta}^*_{n,B}$ with a jackknife estimate $\hat{\zeta}_{n,J}$. In Example 2, as d increases, the distribution of a minimal sufficient statistic separates more from the distribution of the same statistic evaluated at a bootstrap sample.

Example 1. Let $X=(X_1,X_2,\ldots,X_d)$ be a normal d-dimensional vector with uncorrelated components, means $\mu_j=j$ and variances $\sigma_j^2=1$, $j=1,\ldots,d$. Consider n independent copies $(X_{i,1},X_{i,2},\ldots,X_{i,d})$ of X, and let $\bar{X}_j=n^{-1}\sum_{i=1}^n X_{i,j}$, $\zeta_{1,j}=\mathrm{Var}(\bar{X}_j)=n^{-1}$, $\zeta_{2,j}=\mathrm{Var}(n^{-1}\sum_{i=1}^n X_{i,j}^2)=2n^{-1}+4n^{-1}j^2$, $\zeta_1=(\zeta_{1,1},\ldots,\zeta_{1,d})$ and $\zeta_2=(\zeta_{2,1},\ldots,\zeta_{2,d})$. The quantities of interest ζ_1, ζ_2 are chosen such that one of them (that is, ζ_2) depends also on the location parameters. In each column j, the sample variance of the observations $\{X_{i,j},\ i=1,\ldots,n\}$ and $\{X_{i,j}^2,\ i=1,\ldots,n\}$ is computed, and is used to estimate $\zeta_{1,j},\zeta_{2,j}$, $j=1,\ldots,d$; these jackknife vector-estimates are denoted by $\hat{\zeta}_{n,k}$, $k=1,2$. The bootstrap estimates of $\zeta_{1,j}$ and $\zeta_{2,j}$ are obtained using, in column j, the function "bootstrap" in E&T (p. 404, 411); let $\hat{\zeta}^*_{n,B,1}$, $\hat{\zeta}^*_{n,B,2}$ denote these vector-estimates of ζ_1,ζ_2. Computations are made using $n=m=20$, w.l.o.g. $B=50$, and are repeated 100 times for each of the values $d=1,5,10,20,200$. Similar phenomena were observed for $n=m=50$ and $B=500$.

The difference of the squared norms $Dk=\|\hat{\zeta}_{n,k}-\zeta_k\|^2-\|\hat{\zeta}^*_{n,B,k}-\zeta_k\|^2$, $k=1,2$, is computed for each repetition. A positive difference indicates that $\hat{\zeta}^*_{n,B,k}$ is more accurate than $\hat{\zeta}_{n,k}$, and also reflects on the quality of the bootstrap sample. The proportion of positive differences measures how often good bootstrap samples and estimates are obtained. The size of the differences gives one an idea about the expected estimation precision and the effect of the parameter values.

The results appear in Fig. 1; odd numbered plots concern ζ_1 and those even numbered, ζ_2. In the best situation, that is for $d=1$, $\hat{\zeta}^*_{n,B,k}$ is better than $\hat{\zeta}_{n,k}$ roughly 50% of the time (see proposition 2(i)). For a given d value, the error of $\hat{\zeta}^*_{n,B,k}$ may be dramatically larger than that of $\hat{\zeta}_{n,k}$, and depends on the values of the population parameters in the estimand; compare plots 9, 10. The proportion of points sliding through zero, taking negative values, increases with d; compare plots 1, 3, 5, 7, 9. An interesting phenomenon is the *cushion-error* "lying" on the line at zero, towards the side of the negative differences, which is clearly observed in plots 9 and 10; most negative differences are at a distance larger than the *cushion-error* from the line.

In Fig. 2, simulation results are presented for Dk, $k=1,2$, with $d=1$, $n=m=20$, $B=50$, $\sigma^2=1$ and to show the dependence of the *cushion-error* on the parameter values, the means are 10^j, $j=$ 0, 2, 4, 6, 8. The effect of the increase in $E[\mathrm{Var}(\hat{\zeta}^*_{n,B,2}|\mathscr{X}_d)]$ is clear, from the range of the values in plots 2, 4, 6, 8, 10.

Example 2. In the set-up of Example 1 with known common variance $\sigma^2=1$, the bootstrap samples are obtained using the normality assumption. The conditional and the unconditional distributions of the bootstrap sample $(\bar{X}^*_1,\bar{X}^*_2,\ldots,\bar{X}^*_d)$ and of $(\bar{X}_1,\bar{X}_2,\ldots,\bar{X}_d)$ will be examined. Φ is the distribution of a standard normal, Π and Σ will denote respectively $\Pi_{i=1}^{\infty}$ and $\Sigma_{i=1}^{\infty}$.

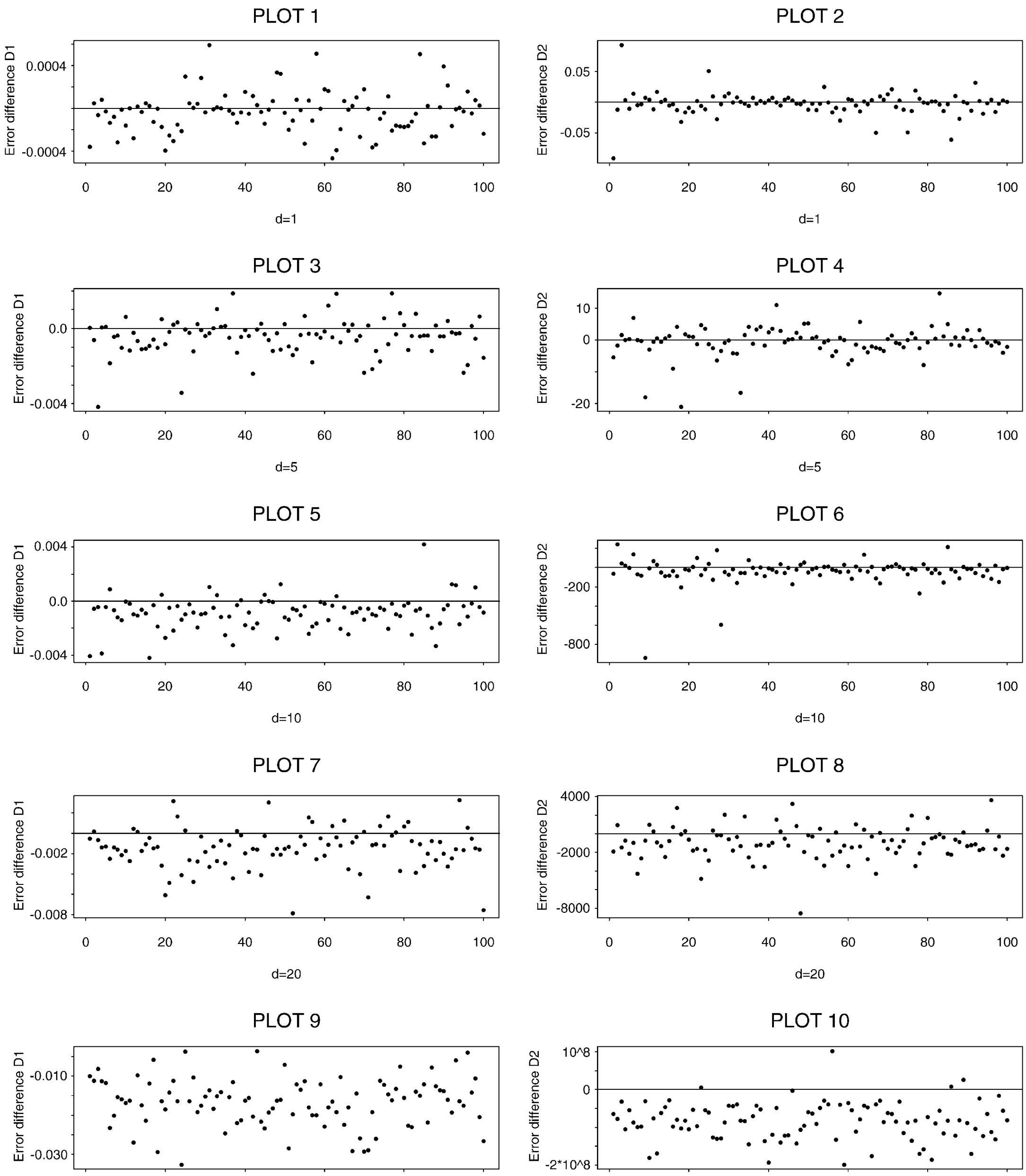

Fig. 1. $n = m = 20, B = 50$.

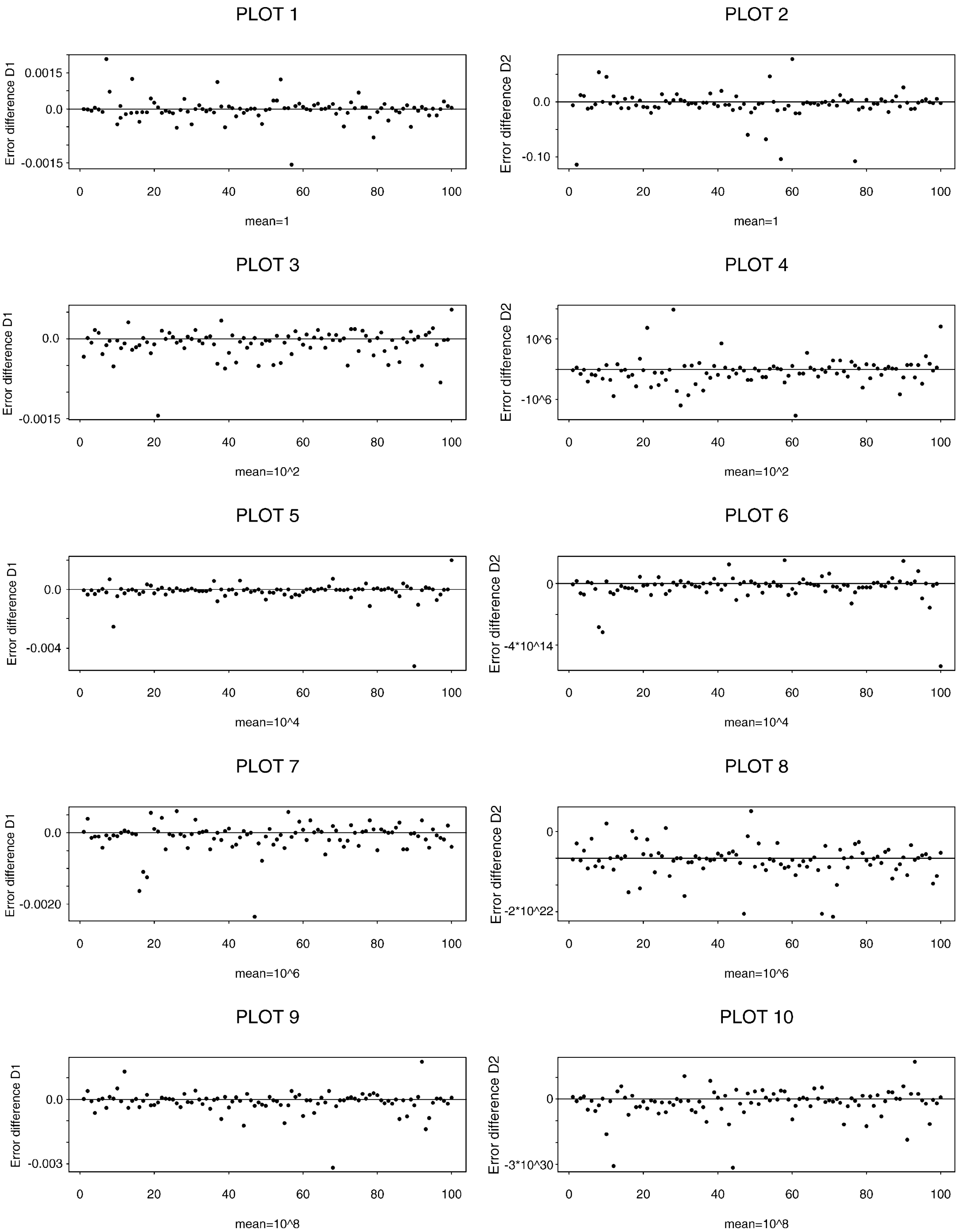

Fig. 2. $n = m = 20,\ B = 50,\ d = 1.$

$\bar{X}_i$ follows a normal distribution with mean μ_i, $\bar{X}_i^*$ conditionally on $\bar{X}_i = \bar{x}_i$ follows a normal distribution with mean $\bar{x}_i$, both have variance n^{-1}, and $a_i = \bar{x}_i - \mu_i$ is a realisation of A_i, $i = 1, \ldots, d$. For the distributions $N = \Pi\Phi[d\{n^{0.5}(y_i - \mu_i)\}]$ and $N^* = \Pi\Phi[d\{n^{0.5}(y_i - \bar{x}_i)\}]$ of the infinite vectors $H^2(N, N^*) = 2(1 - \exp\{-\frac{n}{8}\sum a_i^2\})$. For any $c(>0)$, the probability $P(|\bar{X}_i - \mu_i| > cn^{-0.5}) = p_n > 0$, therefore $\sum P(|\bar{X}_i - \mu_i| > cn^{-0.5}) = \infty$. The independence of the coordinates of $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_d, \ldots)$ and Borel–Cantelli Lemma imply that $|\bar{X}_i - \mu_i| > cn^{-0.5}$ infinitely often; thus, $\sum A_i^2 = \infty$ and a.s. $H^2(N, N^*) = 2$. The same holds asymptotically for the distributions of $(\bar{X}_1, \ldots, \bar{X}_d)$ and $(\bar{X}_1^*, \bar{X}_2^*, \ldots, \bar{X}_d^*)$, since their components are i.i.d..

N and N^* are asymptotically singular, when the size of the bootstrap sample remains the same in each coordinate of the observation-vector. Thus, for a large but fixed d, their supports are quite different, explaining partially the problems of the bootstrap when the target depends on location parameters, as in Example 1 with ζ_2.

4. The geometries

4.1. A geometry for the Bootstrap and the original samples

We examine heuristically what the bootstrap samples can "tell" about the model. Let $X_1, \ldots, X_n$ be a sample from a distribution F_θ with density $f(x, \theta)$, $\theta \in \Theta \subseteq R^k, k \geqslant 1$; $\mathscr{P} = \{F_\theta, \theta \in \Theta\}$ is metrised with the Hellinger distance H. $\hat{\theta}_n$ is an estimate of θ and $\hat{\theta}_n^*$ is the corresponding estimate using a bootstrap sample, which includes information about θ. In Fig. 3, to examine the relative positions of F_θ, $F_{\hat{\theta}_n^*}$ and $F_{\hat{\theta}_n}$ in $\mathscr{P}$, neighbourhoods are drawn as circles for bivariate normal densities, with common variance, means close to each other, and zero correlation. One sees here F_θ, the ("smooth empirical") distribution $F_{\hat{\theta}_n}$, and the bootstrap samples represented by the $F_{\hat{\theta}_n^*}$'s as stars in a circle with centre $F_{\hat{\theta}_n}$ and radius $O(n^{-1/2})$ in probability; this circle represents the Bootstrap World (BW).

The small circle with center F_θ and radius its distance from $F_{\hat{\theta}_n}$ represents, for estimation purposes, the most informative part of the bootstrap world. For a large class of models, the probability of observing samples from this circle is less than or equal to 0.5; the samples being in one of the two halves determined by the line (or hyperplane in higher dimensions) AB. This probability is maximised when F_θ is in the circumference of the circle with center $F_{\hat{\theta}_n}$. As expected, bootstrap samples are better than $\mathscr{X}_d$ when the latter is far from F_θ, but we will not know about it. For the uniform in $(0, \theta), \theta \in R$, the small circle is almost surely empty when $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$ is used to obtain the bootstrap samples and the same holds for $\theta \in R^k$. In a version of Fig. 3 for higher dimensional parameters, the ratio of the volumes of the small sphere with (unknown) centre F_θ and the sphere with centre $F_{\hat{\theta}_n}$ tends to zero as the dimension increases, thus providing the intuition for Proposition 1. Therefore, the best one can expect for $F_{\hat{\theta}_n^*}$ is to stay near $F_{\hat{\theta}_n}$.

4.2. A geometry for $\hat{\zeta}_n$, $\hat{\zeta}_{n,B}^*$ and ζ

The geometry presented for the bootstrap samples is also valid for the estimates $\hat{\zeta}_n$, $\hat{\zeta}_{n,B}^*$ and ζ. With abuse of notation regarding F, let F_ζ, $F_{\hat{\zeta}_n}$ and $F_{\hat{\zeta}_{n,B}^*}$ denote respectively the distributions of $\hat{\zeta}_n$, of $\hat{\zeta}_{n,B}^*$ conditionally on $\hat{\zeta}_n$, and of an estimate of F_ζ based on B bootstrap samples. The estimate

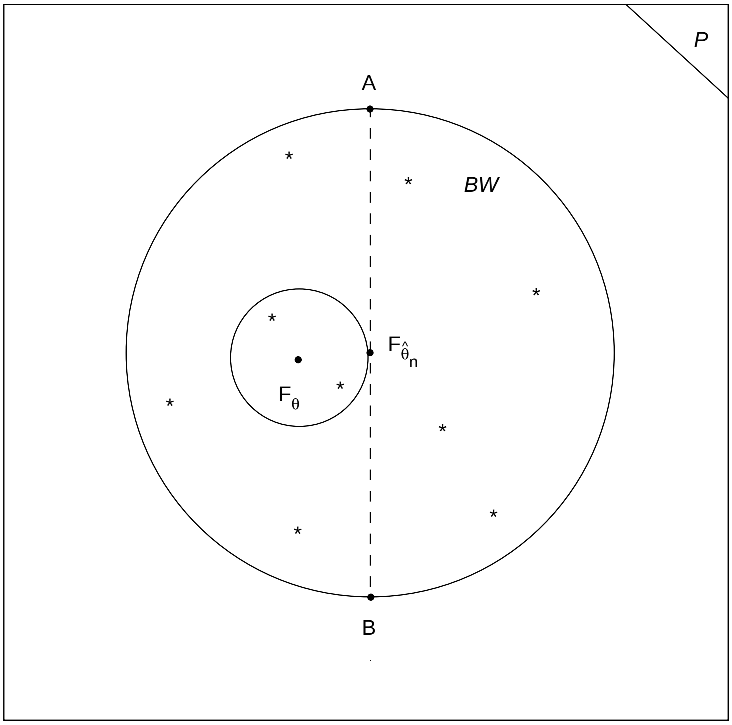

Fig. 3. F_θ: true distribution, $F_{\hat{\theta}_n}$: smooth empirical distribution *: smooth boostrap empirical distributions. BW: the Bootstrap World, P: the World of Measures.

$\hat{\zeta}_n = E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)$ is the parameter to be estimated by $\hat{\zeta}^*_{n,B}$. The heuristics in Section 4.1 indicate that, as the dimension of the estimand ζ increases, the probability $F_{\hat{\zeta}^*_{n,B}}$ will stay in the Hellinger ball with centre F_ζ and radius $H(F_\zeta, F_{\hat{\zeta}_n})$ decreases. In Proposition 1, it is seen that the same holds for a larger radius. The probability is maximised when F_ζ is on the surface of the larger ball representing the Bootstrap World. That is, the bootstrap estimate of F_ζ (or of ζ) has higher chance to beat the classical estimate when the latter deteriorates; as with the bootstrap samples, we will not know about it. In Example 1, the comparison of the differences Dk, $k = 1, 2$, was motivated by these heuristics.

5. Confirming the heuristics and the examples

Proposition 1. *Let $\hat{\zeta}^*_{n,B,i}$ be a bootstrap estimate of $\zeta_i (\in R)$, and let $\hat{\zeta}_{n,i} = E[\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d], =1,\ldots,d$; $\zeta, \hat{\zeta}_n, \hat{\zeta}^*_{n,B}$ are the corresponding d-vectors. Then,*

(a) *the estimate $\hat{\zeta}^*_{n,B}$ is inadmissible*:

$$E\|\hat{\zeta}^*_{n,B} - \zeta\|^2 = E\|\hat{\zeta}_n - \zeta\|^2 + \sum_{i=1}^{d} E\,\mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d).$$

(b) *If the estimates* $\hat{\zeta}^*_{n,B,i}$ *are independent, have uniformly bounded fourth moments and* $0 < \sigma^2 < \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$ *for* $i = 1, \ldots, d$, *then, for any* $0 < \alpha < 1$,

$$P\left[\|\hat{\zeta}^*_{n,B} - \zeta\|^2 \leqslant \|\hat{\zeta}_n - \zeta\|^2 + \alpha \sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)\right] \leqslant \frac{\max\{E(\hat{\zeta}^*_{n,B,i} - \zeta_i)^4; i = 1, \ldots, d\}}{d(1-\alpha)^2\sigma^4}.$$

The probability that $\hat{\zeta}^*_{n,B}$ *is better than* $\hat{\zeta}_n$ *in estimating* ζ *decreases to* 0 *as* d *and the cushion-error* $\alpha \sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$ *increase. In the context of Example* 1

$$\lim_{d\to\infty} P\left[\|\hat{\zeta}_n - \zeta\|^2 - \|\hat{\zeta}^*_{n,B} - \zeta\|^2 \geqslant -\alpha \sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)\right] = 0.$$

Both $\hat{\zeta}_n$ *and* $\hat{\zeta}^*_{n,B}$ *are affected by the curse of dimensionality but for the latter there is in excess the cushion error.*

Proof. (a) Since $\hat{\zeta}_{n,i} = E(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$,

$$E[(\hat{\zeta}^*_{n,B,i} - \zeta_i)^2|\mathscr{X}_d] = \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d) + (\hat{\zeta}_{n,i} - \zeta_i)^2, \quad i = 1, \ldots, d.$$

Taking expectations and adding for all i, the result is obtained.

(b) To obtain the probability bound, a conditional argument is used at first:

$$\begin{aligned} P\Bigg[&\|\hat{\zeta}^*_{n,B} - \zeta\|^2 - E(\|\hat{\zeta}^*_{n,B} - \zeta\|^2|\mathscr{X}_d) \\ &\leqslant \|\hat{\zeta}_n - \zeta\|^2 - E(\|\hat{\zeta}^*_{n,B} - \zeta\|^2|\mathscr{X}_d) + \alpha \sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)|\mathscr{X}_d\Bigg] \\ &\leqslant P[|\,\|\hat{\zeta}^*_{n,B} - \zeta\|^2 - E(\|\hat{\zeta}^*_{n,B} - \zeta\|^2|\mathscr{X}_d)| \geqslant (1-\alpha)d\sigma^2|\mathscr{X}_d] \\ &\leqslant \frac{\mathrm{Var}[\|\hat{\zeta}^*_{n,B} - \zeta\|^2|\mathscr{X}_d]}{(1-\alpha)^2 d^2\sigma^4}. \end{aligned}$$

The result follows by taking expectations in both sides of the last relation and using the independence of the bootstrap estimates and the fact that

$$E\mathrm{Var}[\|\hat{\zeta}^*_{n,B} - \zeta\|^2|\mathscr{X}_d] \leqslant \mathrm{Var}[\|\hat{\zeta}^*_{n,B} - \zeta\|^2]. \qquad \square$$

Remark 1. The term $\alpha \sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$ in Proposition 1 may be replaced by its expectation. Assuming that $0 < \sigma^2 < E\,\mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$, $i = 1, \ldots, d$, it holds

$$\begin{aligned} P&\left[\|\hat{\zeta}^*_{n,B} - \zeta\|^2 \leqslant \|\hat{\zeta}_n - \zeta\|^2 + \alpha \sum_{i=1}^{d} E\,\mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)\right] \\ &\leqslant \frac{\mathrm{Var}[\|\hat{\zeta}^*_{n,B} - \zeta\|^2] + \mathrm{Var}[\sum_{i=1}^{d} \mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)]}{(1-\alpha)^2 d^2\sigma^4}. \end{aligned}$$

If $\mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$, $i=1,\ldots,d$, are independent and have variances uniformly bounded from above: the probability bound decreases to 0 as d increases; the condition $0<\sigma^2$ is justified; and $\alpha \sum_{i=1}^{d} E\,\mathrm{Var}(\hat{\zeta}^*_{n,B,i}|\mathscr{X}_d)$ increases to infinity with d.

Remark 2. Proposition 1 holds with $\theta, E(\hat{\theta}^*_{n,B}|\mathscr{X}_d), \hat{\theta}^*_{n,B}$ instead of $\zeta, \hat{\zeta}_n, \hat{\zeta}^*_{n,B}$; $\hat{\theta}^*_{n,B}$ reflects the quality of the bootstrap sample, which deteriorates as d increases. In many applications of the bootstrap, one is more interested in comparing $\hat{\theta}^*_{n,B}-\hat{\theta}_n$ with $\hat{\theta}_n-\theta$. By writing $\hat{\theta}^*_{n,B}-\hat{\theta}_n=(\hat{\theta}^*_{n,B}-\theta)-(\hat{\theta}_n-\theta)$, one expects it could take large values in view of (1) and Proposition 1, at least when $\hat{\theta}_n=E(\theta^*_{n,B}|\mathscr{X}_d)$.

Definition 1. For densities f,g, with respect to a σ-finite measure λ on the space $\mathscr{X}$, their Hellinger distance $H(f,g)$ is defined as

$$H^2(f,g)=\int_{\mathscr{X}}(f^{1/2}(x)-g^{1/2}(x))^2\lambda(\mathrm{d}x).$$

The affinity of f,g is

$$\rho(f,g)=\int_{\mathscr{X}} f^{1/2}(x)g^{1/2}(x)\lambda(\mathrm{d}x).$$

$H^2(f,g)=2(1-\rho(f,g))$; $H^2(f,g)=2$ if and only if $\rho(f,g)=0$ or if f and g are singular ($f(x)g(x)=0$, λ-a.s.). For normal densities with vector-means μ_1 and μ_0 and variance-covariance matrix $\sigma^2 I$, their square Hellinger distance is $2(1-\exp\{-\|\mu_1-\mu_0\|^2/8\sigma^2\})\sim\|\mu_1-\mu_0\|^2$ when the latter is small, thus justifying Fig. 3.

It is shown below for location models that, as in Example 2, the bootstrap world and the statistician's original world become more distinct as d increases, and the bootstrap sample size remains the same in each coordinate of the observation-vector; the same holds for the distribution of $\hat{\zeta}_n$ and the conditional distribution of $\hat{\zeta}^*_{n,B}$, given $\hat{\zeta}_n$. It is also confirmed that, as argued in Section 4.1 and as been seen in Fig. 1, the percentage of better bootstrap samples for estimation purposes is no more than 50%. A remarkable result in Shepp (1965) for location families is used in the proof.

Theorem 1 (Shepp, 1965, p. 1108). *Let $W=(W_1,W_2,\ldots)$ be a vector of i.i.d. random variables with probability distribution $F=F\{dw\}$ on R (the real numbers), and $a=\{a_1,a_2,\ldots\}$ be a numerical sequence. Let $Q=\Pi F\{dy_n\}$ and $Q^a=\Pi F\{d(y_n-a_n)\}$ be the distributions of W and $W+a$, respectively.*

(i) *If $\sum a_n^2=\infty$ then Q and Q^a are singular.*
(ii) *Assume that the Fisher information $I(F)<\infty$: then Q and Q^a are singular if $\sum a_n^2=\infty$ and Q and Q^a are equivalent if $\sum a_n^2<\infty$.*
(iii) *If Q and Q^a are equivalent for all a with $\sum a_n^2<\infty$ then $I(F)<\infty$.*

Proposition 2. *Let $\mathbf{X}=(X_1,X_2,\ldots,X_d,\ldots)$ be a vector with independent components, X_i has distribution $F(x_i-\theta_i)$ and θ_i is real for all i. Consider n independent copies of $\mathbf{X}$, and let*

$\hat{\theta}_{n,i}$ be an estimate of θ_i; draw a bootstrap sample of size n from $\Pi F(x_i - \hat{\theta}_{n,i})$ and let $\hat{\theta}^*_{n,i}$ be the corresponding estimate. Denote by θ, $\hat{\theta}_n$ and $\hat{\theta}^*_n$ the infinite vectors. Then,

(i) *If the conditional distribution of $\hat{\theta}^*_n$ is symmetric around $\hat{\theta}_n$ then for all θ*

$$P[\|\hat{\theta}^*_n - \theta\| \leqslant \|\hat{\theta}_n - \theta\|] \leqslant 0.5;$$

in several models, this probability is equal to $P[H(F_{n,\theta}, F_{n,\hat{\theta}^*}) \leqslant H(F_{n,\theta}, F_{n,\hat{\theta}_n})]$.

Assume that, for any positive constant c, as i increases, the sequence of probabilities $\{P(n^\beta|\hat{\theta}_{n,i} - \theta_i| > c)\}$ does not converge to zero, $\beta > 0$. Let $Q = \Pi F\{dy_i\}$ be the distribution when all θ_i's are equal to zero, $Q^\theta = \Pi F\{d(y_i - \theta_i)\}$ be the distribution of **X**, *and $Q^{\hat{\theta}_n}$ be the conditional distribution of* $\mathbf{X}^*$ *given* **X**. *Then,*

(ii) *When $\theta_1 = \theta_2 = \cdots = \theta_d = \cdots = \theta_0$, the distributions $Q^{\hat{\theta}_n}$ and Q^θ are singular.*

(iii) *If $H^2(Q^\theta, Q)$ is a function of $\sum \theta_i^2$, then $Q^{\hat{\theta}_n}$ and Q^θ are singular for any θ.*

Proof. Part (i) follows from symmetry by looking at the probability conditionally first and then unconditionally.

For part (ii) one can follow the steps of the proof in Example 2. Note that $Q^\theta = \Pi F\{d(y_i - \theta_0)\}$, $Q^{\hat{\theta}_n} = \Pi F\{d(y_i - \hat{\theta}_{n,i})\}$ and let $A_i = \hat{\theta}_{n,i} - \theta_0$, $i = 1, 2, \ldots$. From the assumption on the sequence of probabilities $\sum P(n^\beta|\hat{\theta}_{n,i} - \theta_0| > c) = \infty$. From Borel–Cantelli $\sum A_i^2 = \infty$ a.s., part (i) of Shepp's theorem applies and $H^2(Q^{\hat{\theta}_n}, Q^\theta) = 2$.

For part (iii) note that $H^2(Q^\delta, Q^{\theta-\delta}) = H^2(Q^\theta, Q)$ which implies that $H^2(Q^\theta, Q^{\hat{\theta}_n}) = H^2(Q^{\hat{\theta}_n - \theta}, Q)$. The result follows from (ii) applied to the case $\theta_0 = 0$. □

Remark 3. If the distributions of $\hat{\zeta}_n$ and $\hat{\zeta}^*_{n,B}$ are respectively $\Pi f(y_i - \zeta_i), \Pi f(y_i - \hat{\zeta}_{n,i})$, the latter is symmetric around $\hat{\zeta}_n$, and the sequence $\{P(n^\beta|\hat{\zeta}_{n,i} - \zeta_i| > c)\}$ does not converge to zero, then Proposition 2 holds for $Q^\zeta, Q^{\hat{\zeta}_n}$ and $\|\hat{\zeta}^*_{n,B} - \zeta\|, \|\hat{\zeta}_n - \zeta\|$.

6. Discussion

The most serious of the observed problems is due to finite resampling, which is inherent in the bootstrap methodology. The results suggest to keep the bootstrap sample and $\hat{\zeta}^*_{n,B}$ as near as possible, respectively, to the original sample $\mathscr{X}_d$ and $E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)$.

The explicit calculation of $E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]$ and its estimation will help to determine a B-value that will bring $\hat{\zeta}^*_{n,B}$ closer to $E(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)$ and reduce the loss in accuracy. When $E[\mathrm{Var}(\hat{\zeta}^*_{n,B}|\mathscr{X}_d)]$ is unbounded, the suggested B-value may be extremely large.

Selecting better bootstrap samples by comparison with $\mathscr{X}_d$ is suggested; any additional information on $F_{n,\theta}$ (resp. F_n) should be used as in Hall and Presnell (1997). Shepp's (1965) theorem and Proposition 2 suggest to increase the size m of the bootstrap samples with the model dimension in order to reduce the separation of the statistician's original and bootstrap worlds. This sampling is against the (traditional) bootstrap philosophy, and may not always provide a pertinent estimate.

Rather than using finite resampling, one may use all samples of size n obtained from $\mathscr{X}_d$ with c.d.f. within a given distance from $F_{n,\hat{\theta}_n}$ (resp. $\hat{F}_n$). Then, there is no additional randomisation, these samples are near $\mathscr{X}_d$ but are not bootstrap samples.

Acknowledgements

Many thanks are due to an anonymous referee for various constructive remarks, and suggestions that helped to improve the presentation of this work.

References

Beran, R., 1994. Seven stages of bootstrap. In: Dirschedl, P., Ostermann, R. (Eds.), Computational Statistics. Physica-Verlag, Wurzburg, pp. 143–158.

Bickel, P.J., Friedman, D.A., 1981. Some asymptotic theory of the bootstrap. Ann. Statist. 9, 1196–1217.

Efron, B., 1979. Bootstrap methods: another look at the jackknife Ann. Statist. 7, 1–26.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Hall, P., Presnell, B., 1997. Intentionally-biased bootstrap samples. J. Roy. Statist. Soc. B 1, 143–158.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer, New York.

Hu, F., Kalbfleisch, J.D., 1999. The estimating function bootstrap. Canad. J. Statist. 3, 1–56.

Le Cam, L., Yang, G., 1990. Asymptotics in Statistics: Some Basic Concepts. Springer, New York.

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Ann. Statist. 21, 255–285.

Quenouille, M., 1949. Approximate tests of correlation in time series. J. Roy. Statist. Soc. B 11, 18–44.

Shepp, L.A., 1965. Distinguishing a sequence of random variables from a translate of itself. Ann. Math. Stat. 36, 1107–1112.

Young, G.A., 1994. Bootstrap: more than a Stab in the Dark? Statist. Sci. 9, 382–415.