ASSESSING THE QUALITY OF BOOTSTRAP SAMPLES AND ESTIMATES OBTAINED WITH FINITE RESAMPLING

Yannis G. Yatracos
Yau Mathematical Sciences Center
Tsinghua University

AUGUST 2020

A Statistical Problem

- Statistical population (F), i.e. measurements, in **box # 1**. Unknown parameter θ : population's mean, variance etc
- Draw sample $X_1, \ldots, X_n$ from **box # 1** and use it to construct S_n, an estimate of θ. S_n will change values depending on the sample. Thus, there is variability in S_n.
- Interested in **the quality** of S_n : its variability around θ, i.e. its **bias and variance**, or **its distribution** (F_n) to calculate, e.g., the probability $P[1 < |S_n - \theta| < 2], \theta \in R$.
- Denote the **quantity of interest** by $\alpha_n(G, \theta), \; G = F, \; F_n$.
- When $\alpha_n(G, \theta)$ is not easy to obtain,the **Bootstrap** is used. **Question: Do we prefer S_n to be more variable or less variable?** S_n's variability will depend also on the variability of the population in **box # 1**.

The Bootstrap Method

- Introduced by Efron (1979) to determine either the accuracy of S_n or related quantity $\alpha_n(G, \theta),\ G = F,\ F_n$.
- Sample $X_1, \ldots, X_n$, in box # 2, is a mini-population. Compute the characteristic of S_n that is of interest by drawing B Bootstrap samples from box # 2 and plugging each in S_n to obtain $S^*_{n,1}, \ldots, S^*_{n,B}$; "$*$" denotes estimates from Bootstrap samples.
- **What are the sources of variability** in $S^*_{n,1}, \ldots, S^*_{n,B}$?
- Variability from **box # 1** **AND** from **box # 2**!!
- **How do you feel about the additional variability introduced by finite resampling?**

An Example where Bootstrap is used

- **Example 1:** Sample $\mathcal{X}_{\mathbf{n}} = (X_1, \ldots, X_n)$, independent, real $X_i \sim F(\text{or } F_\theta), \theta \in R^k$. Estimate: $S_n = S(\mathcal{X}_n)$. Interested in $\theta = \zeta_n = Var(S_n)$.
- Take B Bootstrap samples $\mathcal{X}^*_{n,1}, \ldots, \mathcal{X}^*_{n,B}$ of size $m_n = n$.
  How each of the B samples is drawn?
  Either with replacement from box # 2, OR by estimating $\hat{\theta}_n$ and drawing samples from $F_{\hat{\theta}_n}$.
  Let $S^*_{n,1} = S(\mathcal{X}^*_{n,1}), \ldots, S^*_{n,B} = S(\mathcal{X}^*_{n,B})$ and estimate $\zeta_n = Var(S_n)$ by

$$\hat{\zeta}^*_{n,B} = \frac{1}{B-1}\sum_{b=1}^{B}(S^*_{n,b} - \bar{S}^*_{n,B})^2, \qquad \bar{S}^*_{n,B} = \frac{1}{B}\sum_{b=1}^{B} S^*_{n,b}.$$

- **Example 2:** $\mathcal{X}_n = (X_1, \ldots, X_n)$, independent, real $X_i \sim F_\theta, \theta = EX_1, 1 \leq i \leq n$.
- Interested in the law of $S_n = \sqrt{n}(\bar{X}_n - \theta)$.
- The Bootstrap approach: approximate it by the law of $\sqrt{n}(\bar{X}_n^* - \bar{x}_n)$, using B bootstrap samples; $\bar{x}_n$ is the observed sample mean.

- **Extending Bootstrap's use:** The variance of S_n or its distribution are two of its possible characteristics to be estimated using the Bootstrap methodology.
- The method is now a "general purpose tool" (Young, 1994),
- an extension of the maximum likelihood "plug-in principle" (Efron and Tibshirani, 1993, denoted by E&T).
- **CLAIM:** ... if $n \geq N_1$ and $B \geq N_2$ one can estimate well a variance or ...; N_1, N_2 are predetermined.

Disturbing Bootstrap results

- Most theoretical assessments of Bootstrap methodology ignore the fact that (in practice) the number B of bootstrap samples is fixed. This is what is examined in the sequel.
- VERY RECENT Stanford Statistics Seminar, 11/1/2019
  Speaker: Larry Wasserman, Carnegie Mellon University
  Title: Model Free Predictive Inference
  Abstract: ... We start with high-dimensional regression. **First we show that the bootstrap is very inaccurate,** which motivates moving away from the usual focus ...
- **Bootstrap Claims in Asymptotics, with** $n \uparrow \infty$
  "The approximation of distributions using the Bootstrap methodology is asymptotically valid in many cases. "
  "The distribution of $T(X_1, \ldots, X_n, F) \approx$ distribution of $T(X_1^*, \ldots, X_n^*, \hat{F}_n)$".

Disturbing Bootstrap results

• **Counter-example** (Bickel and Freedman, 1981) $X_1, \ldots, X_n$ $i.i.d.$ $U(0, \theta)$, then for the asymptotic distribution of $X_{(n)} = \max\{X_1, \ldots, X_n\}$

$$\frac{n(\theta - X_{(n)})}{\theta} \xrightarrow{\mathcal{D}} \text{Exponential distribution}$$

BUT, the conditional distribution of

$$\frac{n(X_{(n)} - X^*_{(n)})}{X_{(n)}}$$

does not have a weak limit almost surely.

Disturbing Bootstrap results

- **Bootstraping regression models:** (Bickel and Freedman, 1981) "The Bootstrap fails quite generally when the number of parameters is too large".
- Liu (1988) and Mammen (1993) proposed and justified the use of "Wild Bootstrap" with better results than the Bootstrap when the errors are heteroscedastic.
- (Mammen, 1993) ' "The Bootstrap does not work in a satisfactory way because it has to mimic a complex stochastic structure of high dimensional distributions"
- Haerdle and Mammen (1990) provided Examples where Efron's Bootstrap is not consistent but Wild Bootstrap is.

Disturbing Results for Finite n, B

• (Devroye 1996, 2019) For convex loss $L(.\ ,\ .)$, and risk $R(T_n,\theta)=EL(T_n,\theta)$,

$$\inf_{T_n} R(T_n(\mathcal{X}_n^*),\theta) \geq \inf_{S_n} R(S_n(\mathcal{X}_n),\theta).$$

• We will examine questions a practitioner may ask when the dimension d of the observations or that of S_n is large.

• To what extent can the Bootstrap user be sure of "hitting" ζ (or ζ_n) (Young, 1994)?

• What are **measures for the loss in accuracy** of the bootstrap estimate?

• Is the precision of the Bootstrap estimate $\zeta^*_{n,B}$ influenced by the nature of ζ_n?

• How much can we learn from the bootstrap samples about the population F (or F_θ) and/or the distribution $F_{n,\theta}$ (or F_n) of S_n?

- Heuristics for the Examples, Geometries and Propositions:

a) *d* fixed: *B* bootstrap samples provide the Statistical Experiment with additional randomisation. Thus, $\hat{\zeta}^*_{n,B}$ is not the function of a sufficient statistic and can be improved by conditioning on the sufficient statistic.

b) *d* increases: *i)* the additional randomisation and its effect are increasing.

ii) 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 from $F_{n,\hat{\theta}_n}$ (or $\hat{F}_n$) "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 S_n is large.

• **The effect of the additional randomisation on $\hat{\zeta}^*_{n,B}$: d finite, or $\uparrow \infty$, $||\theta|| \uparrow \infty$.** Recall that for random variables U, V,

$$Var(U) = Var\ E(U|V) + EVar(U|V). \tag{1}$$

Let $U = \hat{\zeta}^*_{n,B}$, $V = \mathcal{X}_n$ the sample, $\hat{\zeta}_n = E(\hat{\zeta}^*_{n,B}|\mathcal{X}_n)$ which is an estimate, then

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

Since $E\hat{\zeta}^*_{n,B} = E\hat{\zeta}_n$ and $Var(\hat{\zeta}^*_{n,B}) > Var(\hat{\zeta}_n)$, the Mean Squared Error (MSE) of $\hat{\zeta}^*_{n,B}$ exceeds the MSE of $\hat{\zeta}_n$ by the *cushion-error* $E[Var(\hat{\zeta}^*_{n,B}|\mathcal{X}_n)]$, **thus $\hat{\zeta}^*_{n,B}$ is inadmissible**.

• By letting B increase to infinity before n, $E[Var(\hat{\zeta}^*_{n,B}|\mathcal{X}_n)]$ vanishes. However, this is not in accordance with the bootstrap methodology.

Example 3: Let $\mathbf{X} = (X_1, \ldots, X_d) : X_i \sim N(i, 1), i = 1, \ldots, d$ with the X's independent. Take n $i.i.d.$ copies of $\mathbf{X}$: $\bar{X}_i$ is the i-th column average, $\mathcal{X}_n = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. Interest in:

$$\zeta_{1,n} = (Var(\bar{X}_1), \ldots, Var(\bar{X}_d)) = (\frac{1}{n}, \ldots, \frac{1}{n});$$

$$\zeta_{2,n} = (Var(\frac{1}{n}\sum_{i=1}^{n} X_{i,1}^2), \ldots, Var(\frac{1}{n}\sum_{i=1}^{n} X_{i,d}^2))$$

$$= (\frac{6}{n}, \ldots, \frac{2+4j^2}{n}, \ldots, \frac{2+4d^2}{n}).$$

A difference: $\zeta_{1,n}$ is bounded but $\zeta_{2,n}$ increases with d.

In simulations, $\hat{\zeta}_{1,n}, \hat{\zeta}_{2,n}$ are the usual estimates of $\zeta_{1,n}, \zeta_{2,n}$; $\hat{\zeta}^*_{1,n,B}$, $\hat{\zeta}^*_{2,n,B}$ are Bootstrap estimates of $\zeta_{1,n}, \zeta_{2,n}$, using the function "bootstrap" from Statlib suggested in E&T (1993).

We use $n = m = 20$, due to the *cushion error w.l.o.g.* $B = 50$, we have 100 repeats and $d = 1, 5, 10, 20, 200$. Similar phenomena were observed for $n = m = 50$ and $B = 500$.
In simulations, **for each of the 100 repeats** calculate:

$$D_{n,1} = ||\hat{\zeta}_{1,n} - \zeta_{1,n}||^2 - ||\hat{\zeta}^*_{1,n,B} - \zeta_{1,n}||^2$$

$$D_{n,2} = ||\hat{\zeta}_{2,n} - \zeta_{2,n}||^2 - ||\hat{\zeta}^*_{2,n,B} - \zeta_{2,n}||^2.$$

• If $D_{n,j} > 0$, $\hat{\zeta}^*_{j,n,B}$ is more accurate and also reflects the quality of the bootstrap sample, $j = 1, 2$.
• The proportion of positive differences indicates how often good bootstrap samples and estimates are obtained.
• The size of the differences gives an idea about the expected estimation precision and the effect of the parameter values.

Fig. 1: odd numbered plots concern $\zeta_{1,n}$, even numbered, $\zeta_{2,n}$.

- The best situation for Bootstrap: $d = 1$; $\hat{\zeta}^*_{n,B}$, is better than $\hat{\zeta}_n$ roughly 50% of the time (see proposition 2(i)).
- For large d, $\hat{\zeta}^*_{n,B}$'s error is dramatically larger than that of $\hat{\zeta}_n$ depending on the parameter values; compare plots 9, 10.
- The proportion of dots sliding through the line at zero, with negative values, increases with d; compare plots 1,3,5,7,9.
- The *cushion-error* is "lying" below the line at zero and is clearly observed in plots 9 and 10.

- Fig. 2: simulations when $d = 1$ only, for $D_k, k = 1, 2$, $n = m = 20, B = 50, \sigma^2 = 1$. To show the dependence of *the cushion-error* on the size of the parameter values, the means are $10^j; j = 0, 2, 4, 6, 8$.
- The effect of the increase in $E[Var(\hat{\zeta}^*_{n,B,2}|\mathcal{X}_n)]$ is clear, from the range of the values in plots 2, 4, 6, 8, 10.

- For the target $\zeta_n (\in R^d)$ one is interested in the probability

$$P[||\hat{\zeta}^*_{n,B} - \zeta_n||^2 \leq ||\hat{\zeta}_n - \zeta_n||^2],$$

especially when $d \uparrow \infty$. You may ask what is $\hat{\zeta}_n$? There are competing estimates, e.g. Jackknife estimates. We use:

$$\hat{\zeta}_n = E[\hat{\zeta}^*_{n,B} | \mathcal{X}_n].$$

It will be shown that

$$\lim_{d \to \infty} P[||\hat{\zeta}^*_{n,B} - \zeta_n||^2 \leq ||\hat{\zeta}_n - \zeta_n||^2 + C_d] = 0,$$

with $C_d \uparrow \infty, \; C_d > 0.$

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

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

$$E||\hat{\zeta}^*_{n,B} - \zeta_n||^2 = E||\hat{\zeta}_n - \zeta_n||^2 + \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{d}} \mathbf{EVar}(\hat{\zeta}^*_{\mathbf{n,i,B}}|\mathcal{X}_{\mathbf{n}}).$$

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

$$P[||\hat{\zeta}^*_{n,B} - \zeta_n||^2 \leq ||\hat{\zeta}_n - \zeta_n||^2 + \alpha \sum_{i=1}^{d} Var(\hat{\zeta}^*_{n,i,B}|\mathcal{X}_n)]$$

$$\leq \frac{\max\{E(\hat{\zeta}^*_{n,i,B} - \zeta_{n,i})^4; i = 1, \ldots, d\}}{\mathbf{d}(1-\alpha)^2\sigma^4}.$$

The probability that $\hat{\zeta}^*_{n,B}$ is better than $\hat{\zeta}_n$ in estimating ζ_n decreases to 0 as d increases. In the context of Example 1,

$$\lim_{d\to\infty} P[||\hat{\zeta}^*_{n,B} - \zeta_n||^2 \leq ||\hat{\zeta}_n - \zeta_n||^2 + \alpha \sum_{i=1}^{d} Var(\hat{\zeta}^*_{n,i,B}|\mathcal{X}_n)] = 0.$$

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.

Remark 1. The term $\alpha\sum_{i=1}^{d} Var(\hat{\zeta}^*_{n,i,B}|\mathcal{X}_n)$ in Proposition 1 may be replaced by its expected value. Assuming that $0 < \sigma^2 < EVar(\hat{\zeta}^*_{n,i,B}|\mathcal{X}_n),\ i = 1, \ldots, d$,

$$P[||\hat{\zeta}^*_{n,B} - \zeta_n||^2 \leq ||\hat{\zeta}_n - \zeta_n||^2 + \alpha \sum_{i=1}^{d} EVar(\hat{\zeta}^*_{n,i,B}|\mathcal{X}_n)]$$

$$\leq \frac{Var[||\hat{\zeta}^*_{n,B} - \zeta_n||^2] + Var[\sum_{i=1}^{d} Var(\zeta^*_{n,i,B}|\mathcal{X}_n)]}{(1-\alpha)^2 d^2 \sigma^2}.$$

Remark 2. Proposition 1 holds also with θ, $E(\hat{\theta}^*_{n,B}|\mathcal{X}_n)$, $\hat{\theta}^*_{n,B}$ instead of $\zeta_n, \hat{\zeta}_n, \hat{\zeta}^*_{n,B}$; $\hat{\theta}^*_{n,B}$ 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 $||\hat{\theta}_n - \theta||$. From the Proposition 1 it is expected that $||\hat{\theta}^*_{n,B} - \hat{\theta}_n||$ would be large at least when $\hat{\theta}_n = E(\hat{\theta}^*_{n,B}|\mathcal{X}_n)$ and either d or the parameter values are large.

Question: Population density f. Assume we are interested in $E_f X$ or $Var_f X$. How well can one hit a target related with f using a sample from another density g? And if g is far from f?

The Bootstrap analogue: Obtain $\mathcal{X}^*_{n,1}, \ldots, \mathcal{X}^*_{n,B}$ from $f_{\hat{\theta}_n}$. How much can tell us about f_θ? **It will depend on the distance between the distributions of these samples and f_θ.**

Hellinger Distance

Definition: Let $f,\ g$ be densities on $\mathcal{Y}(\subset R^n)$. The Hellinger distance $H(f,g)$ is:

$$0 \leq H^2(f,g) = \int_{\mathcal{Y}} (\sqrt{f(y)} - \sqrt{g(y)})^2 dy = 2[1 - \rho(f,g)] \leq 2,$$

$$\text{the affinity } \ \rho(f,g) = \int_{\mathcal{Y}} \sqrt{f(y)g(y)} dy.$$

Notation: If $F,\ G$ are respectively c.d.fs. of f, g we may use $H(F,G)$ instead of $H(f,g)$.

- $H^2(f,g) = 2$ for f, g with supports that do not intersect: "$f,\ g$ are separated."
- Draw a graph of f, g with intersecting supports: ...

- **Property of** H : For $\Pi_{i=1}^n f,\ \Pi_{i=1}^n g$, that corresponds to $X_1, \ldots, X_n$ *i.i.d.* either f or g,

$$H^2(\Pi_{i=1}^n f, \Pi_{i=1}^n g) = 2[1 - \rho^n(f,g)] \underset{n\to\infty}{\longrightarrow} 2;$$

"asymptotically in n the densities separate" and we can distinguish if either f or g is true.

Example 4: $f_{\theta_1} \sim N(\theta_1, I),\ f_{\theta_2} \sim N(\theta_2, I)$. Then,

$$H^2(f_{\theta_1}, f_{\theta_2}) = 2(1 - e^{-\frac{||\theta_1 - \theta_2||^2}{8}})$$

which behaves like $||\theta_1 - \theta_2||^2$ when $||\theta_1 - \theta_2||$ is small.

- Thus, the H-geometry for $\{f_\theta, \theta \in R^k\}$ is Euclidean Geometry in R^k and small neighborhoods around f_θ are like spheres.

Bootstrap Geometry

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be *i.i.d.* $N(\theta, I), \theta \in R^2$, $\mathcal{X}_n = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$. Estimate θ by $\hat{\theta}_n$ and obtain B Bootstrap samples $\mathcal{X}^*_{n,i}$, each from $N(\hat{\theta}_n, I)$ and $\hat{\theta}^*_{n,i} = \hat{\theta}_n(\mathcal{X}^*_{n,i}),\ 1 \leq i \leq B.$

- In Figure 3: $F_\theta = N(\theta, I),\ F_{\hat{\theta}_n} = N(\hat{\theta}_n, I), * = N(\hat{\theta}^*_{n,i}, I).$

We examine heuristically what the bootstrap samples can tell about the model F_θ (or $F_{n,\theta}$).

• $H(F_\theta, F_{\hat{\theta}^*_{n,i}}),\ i = 1, \ldots, B$, reflect the quality of the bootstrap samples used to estimate $\zeta_n = \alpha(F_{n,\theta})$ and **indirectly** F_θ and θ.

• Among $H(F_\theta, F_{\hat{\theta}^*_{n,i}}),\ i = 1, \ldots, B$, **those smaller than** $H(F_\theta, F_{\hat{\theta}_n})$ determine, at least in some situations, **the better Bootstrap samples**.

• For several other distributions, $||\theta - \eta|| \approx H(F_{n,\theta}, F_{n,\eta})$ and the proportion of better Bootstrap samples,

$$P[H(F_\theta, F_{\hat{\theta}_n^*}) \leq H(F_\theta, F_{\hat{\theta}_n})] \approx P(||\hat{\theta}_n^* - \theta|| \leq ||\hat{\theta}_n - \theta||).$$

• The H-“sphere” with centre F_θ and radius $H(F_\theta, F_{\hat{\theta}_n})$ includes, for estimation purposes, **the most informative part of the Bootstrap World.**

• For a large class of models, the proportion of Bootstrap samples in this sphere is less than or equal to 50%. For the Uniform in $(0, \theta), \theta \in R^k$, this circle is almost surely empty when $\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$ is used to obtain Bootstrap samples.

• In a version of Figure 3 for high dimensional parameters, the ratio of the volumes of the small sphere with center F_θ and the sphere with centre $F_{\hat{\theta}_n}$ tends to zero when the dimension increases, as Proposition 1 suggests.

• To study the effect of F_θ($resp.F$) in the accuracy of $\zeta^*_{n,B}$ several θ-values are considered, and d or the dimension of S_n may increase to infinity; w.l.o.g. B is fixed, since $E[Var(\hat{\zeta}^*_{n,B}|\mathcal{X}_n)]$ increases to infnity for selected θ-values and it is decreasing as B increases.

• 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.

Example 5. Let $\mathbf{X} = (X_1, \ldots, X_d) : X_i \sim \mathcal{N}(\mu_i, 1), i = 1, \ldots, d$, with the X's independent. Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a sample of size n. Bootstrap samples are obtained using the MLE of the means. We are interested in a functional $T(\bar{X}_1, \ldots, \bar{X}_d)$.

• The conditional distribution of $(\bar{X}_1^*, \ldots, \bar{X}_d^*)$ and that of $(\bar{X}_1, \ldots, \bar{X}_d)$ are compared. Φ is the *c.d.f.* of $\mathcal{N}(0, 1)$. $\bar{X}_i \sim \mathcal{N}(\mu_i, n^{-1})$, conditionally on $\bar{X}_i = \bar{x}_i$, $\bar{X}_i^* \sim \mathcal{N}(\bar{x}_i, n^{-1})$ and $a_i = \bar{x}_i - \mu_i$ is a realization of $A_i; i = 1, \ldots, d$. It is hoped that the law $\mathcal{N}_d$ of $(\bar{X}_1, \ldots, \bar{X}_d)$ is not far from the conditional law $\mathcal{N}_d^*$ of $(\bar{X}_1^*, \ldots, \bar{X}_d^*)$.

For the infinite **X**-vectors and their distributions, $\mathcal{N} = \Pi_{i=1}^{\infty}\Phi[d(n^{.5}(y_i - \mu_i)]$ and $\mathcal{N}^* = \Pi_{i=1}^{\infty}\Phi[d(n^{.5}(y_i - \bar{x}_i)]$,

$$H^2(\mathcal{N}, \mathcal{N}^*) = 2(1 - exp\{-\frac{n}{8}\sum_{i=1}^{\infty} a_i^2\}).$$

For any $c(> 0)$, the probability $P(|\bar{X}_i - \mu_i| > cn^{-0.5}) = p_n > 0$, therefore $\sum_{i=1}^{\infty} P(|\bar{X}_i - \mu_i| > cn^{-0.5}) = \infty$. Independence of the coordinates of $(\bar{X}_1, \ldots, \bar{X}_d, \ldots)$ and Borel-Cantelli Lemma imply that $|\bar{X}_i - \mu_i| > cn^{-0.5}$ infinitely often, thus, $\sum_{i=1}^{\infty} A_i^2 = \infty$ and a.s. $H^2(\mathcal{N}, \mathcal{N}^*) = 2$.

• $\mathcal{N}$ and $\mathcal{N}^*$ are 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, the supports of $\mathcal{N}_d$ and $\mathcal{N}_d^*$ are quite different, explaining partially the problems of the bootstrap when the target depends on location parameters, as in Example 3 with $\zeta_{2,n}$.

Does this result hold only for the Normal model? NO!

• It is shown for location models that, as in Example 5, the Bootstrap world consisting of $\{\mathcal{X}_n^*\}$ and the statistician's original $\{\mathcal{X}_n\}$-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 seen in Fig. 1 and the Bootstrap Geometry, the percentage of better bootstrap samples for estimation purposes is **no more than 50%.** A result in Shepp (1965) for location families is used in the proof.

Proposition 2. Let $\mathbf{X} = (X_1, X_2, \dots, X_d, \dots)$ be a vector with independent components, X_i has distribution $F(x_i - \theta_i)$ and $\theta_i \in R$ 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 $\Pi_{i=1}^{\infty} F(x_i - \hat{\theta}_{n,i})$and let $\hat{\theta}^*_{n,i}$ be the corresponding Bootstrap estimate. Denote by $\theta, \hat{\theta}_n$ 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,

$$P(||\hat{\theta}^*_n - \theta|| \leq ||\hat{\theta}_n - \theta||) \leq 0.5;$$

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

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 $Q=\Pi_{i=1}^{\infty}F\{dy_i\}$ be the distribution when all θ_i are equal to zero, let $Q^\theta=\Pi_{i=1}^{\infty}F\{d(y_i-\theta_i)\}$ be the distribution of **X**, and let $Q^{\hat{\theta}_n}$ be the conditional distribution of $\mathbf{X}^*$ given **X**. Then,

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

(iii) If $H^2(Q^\theta,Q^0)$ is a function of $\sum_{i=1}^{\infty}\theta_i^2$ then $Q^{\hat{\theta}_n}$ and Q^θ are singular for any θ.

Remark 3. If the distributions of $\hat{\zeta}_n$ and $\hat{\zeta}^*_{n,B}$ are respectively $\Pi_{i=1}^{\infty}f(y_i-\zeta_i),\Pi_{i=1}^{\infty}f(y_i-\hat{\zeta}_{n,i})$ 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 $||\hat{\zeta}^*_{n,B}-\zeta_n||,||\hat{\zeta}_n-\zeta_n||$.

Discussion

• 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 $\mathcal{X}_n$ and $\hat{\zeta}_n$ and suggestions are given how to stay respectively near each other. Then, the chance to obtain better $\hat{\zeta}^*_{n,B}$ increases, but the potential of substantial inadmissibility remains.

• The most serious of the observed problems is due to finite resampling ($B < \infty$), 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 $\mathcal{X}_n$ and $E(\hat{\zeta}^*_{n,B}|\mathcal{X}_n)$.

• The explicit calculation of $E[Var(\zeta^*_{n,B}|\mathcal{X}_n)]$ and its estimation will help to determine a B-value that will bring $\zeta^*_{n,B}$ closer to $E(\zeta^*_{n,B}|\mathcal{X}_n)$ and reduce the loss in accuracy. When $E(\zeta^*_{n,B}|\mathcal{X}_n)$ is unbounded, the suggested B-value may be extremely large. For example, using Jackknife, $B_n = n,\ m_n = n - 1$ and and all samples of size $n - 1$ are considered.

• Selecting better bootstrap samples by comparison with $\mathcal{X}_n$ is suggested. Any additional information on $F_{n,\theta}$ (resp. F_n) should be used as in Hall and Presnell (1997), where the Bootstrap samples near $\mathcal{X}_n$ are only used. Proposition 2 suggests increasing the size m_n of the Bootstrap samples with the model dimension. This sampling is against the (traditional) Bootstrap philosophy, and may not always provide a pertinent estimate.

Schepp's Theorem

Theorem (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 and let $\mathbf{a} = \{a_1, a_2, \ldots\}$ be a numerical sequence. Let $Q = \Pi_{i=1}^{\infty} F\{dy_i\}$ and $Q^{\mathbf{a}} = \Pi_{i=1}^{\infty} F\{d(y_i - a_i)\}$ be the distributions of W and W + **a**, respectively.
(i) If $\sum_{i=1}^{\infty} a_i^2 = \infty$, then Q and $Q^{\mathbf{a}}$ are singular.
(ii) Assume that the Fisher information $I(F) < \infty$, then Q and $Q^{\mathbf{a}}$ are singular if $\sum_{i=1}^{\infty} a_i^2 = \infty$, and Q and $Q^{\mathbf{a}}$ are equivalent if $\sum_{i=1}^{\infty} a_i^2 < \infty$.
(iii) If Q and $Q^{\mathbf{a}}$ are equivalent for all **a** with $\sum_{i=1}^{\infty} a_i^2 < \infty$, then $I(F) < \infty$.