Bootstrap

(notes by Pavel Chigansky)

Bootstrap is a method of estimating distribution of statistics. Classical applications, which require such estimation, include construction of confidence sets, choosing critical value of tests, estimation of bias and risk, etc.

1. The method

Let us start with a typical problem to which the bootstrap method applies. Suppose a functional $T(F)$ is estimated given an i.i.d. sample $X_1,\dots,X_n$ from an unknown distribution F by the plug-in estimator $T(\widehat{F}_n)$, where $\widehat{F}_n$ is the empirical distribution. In general, such estimator can be biased, and in some applications it would be of interest to estimate the bias

$$b_T(F) = \mathbb{E}_F T(\widehat{F}_n) - T(F). \tag{1.1}$$

Once such estimator becomes available, it can be compensated from $T(\widehat{F}_n)$ thus obtaining an estimator with a reduced bias[1].

The bias functional (1.1) is the solution to equation

$$\mathbb{E}_F \Psi(F,\widehat{F}_n,t) = 0, \quad t\in\mathbb{R}, \tag{1.2}$$

where $\Psi(F,\widehat{F}_n,t) = T(\widehat{F}_n) - T(F) - t$. There are many other problems, in which the estimated quantity can be defined as the solution to equation (1.2) with a suitably chosen functional Ψ.

The *bootstrap method* suggests to replace F and $\widehat{F}_n$ in (1.2) with $\widehat{F}_n$ and $\widehat{F}_n^*$ respectively, where $\widehat{F}_n^*$ is the empirical distribution of the *bootstrap sample*

$$X_1^*,\dots,X_n^* \big| X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \widehat{F}_n.$$

If there are no ties in the original sample, generating the bootstrap sample merely amounts to drawing n independent samples from the uniform distribution over the set $\{X_1,\dots,X_n\}$, considered as fixed. The solution $\widehat{t}_n$ to the obtained equation,

$$\mathbb{E}_{\widehat{F}_n} \Psi(\widehat{F}_n,\widehat{F}_n^*,t) = 0, \tag{1.3}$$

lecture notes for "Advanced Statistical Models B" course.

[1] this may or may not increase the MSE risk, but typically it remains of the same order with respect to the sample size n

is the *bootstrap* estimator for t. Here $\mathbb{E}_{\widehat{F}_n}$ stands for the conditional expectation, given the original sample.

REMARK 1.1. If the model is parametric, i.e. the distributions are of the form F_θ for some $\theta \in \Theta \subseteq \mathbb{R}^d$, then it makes sense to utilise the parametrisation and to replace $\widehat{F}_n$ in (1.3) with $F_{\widehat{\theta}_n}$ for some reasonable estimator $\widehat{\theta}_n$, and to sample X_j^*'s from $F_{\widehat{\theta}_n}$ rather than from $\widehat{F}_n$. This version of bootstrap is referred to as *parametric*. We will continue to use the notations $\widehat{F}_n$ and $\widehat{F}_n^*$ in both cases, whose precise meaning should be clear from the context.

REMARK 1.2. An important practical feature of the bootstrap method is that, whenever the expectation in (1.3) cannot be computed explicitly, it can be approximated within *arbitrary* accuracy using Monte Carlo averaging,

$$\frac{1}{B}\sum_{j=1}^{B}\Psi(\widehat{F}_n,\widehat{F}_n^{*j},t) \xrightarrow[B\to\infty]{\mathbb{P}_{\widehat{F}_n}} \mathbb{E}_{\widehat{F}_n}\Psi(\widehat{F}_n,\widehat{F}_n^{*},t) \tag{1.4}$$

where $\widehat{F}_n^{*j}$ denote i.i.d. replicas of $\widehat{F}_n^*$.

REMARK 1.3. Any statistic which depends on the sample through its order statistics is a function of the empirical distribution. Hence the bootstrap methodology applies to a variety of useful statistics, beyond plug-in estimators.

1.1. Bias estimation. As mentioned above, one of the classical applications of the bootstrap method is to bias reduction of estimators. Suppose a functional $T(F)$ is estimated by its plug-in estimator $T(\widehat{F}_n)$. The estimation bias solves the equation, cf. (1.2),

$$\mathbb{E}_F T(\widehat{F}_n) - T(F) - t = 0,$$

and hence the corresponding bootstrap estimator is

$$\widehat{t}_n = \mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*) - T(\widehat{F}_n). \tag{1.5}$$

The estimated bias can be compensated from the original estimator. This may of course increase the MSE risk, but in many cases it will remain to be of order $O(n^{-1})$, while the bias will be reduced to order $O(n^{-2})$.

EXAMPLE 1.4. Let us specify the formula (1.5) for the squared mean $T(F) = (\mu_F)^2$. In this case,

$$\mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*) = \mathbb{E}_{\widehat{F}_n}\Big(\frac{1}{n}\sum_{j=1}^{n} X_j^*\Big)^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}_{\widehat{F}_n} X_i^* X_j^* =$$
$$\frac{1}{n^2}\Big(n\mathbb{E}_{\widehat{F}_n}(X_1^*)^2 + n(n-1)\big(\mathbb{E}_{\widehat{F}_n}X_1^*\big)^2\Big) = \frac{1}{n}\overline{X_n^2} + \Big(1-\frac{1}{n}\Big)\overline{X}_n^2$$

and hence the bootstrap bias estimator is a multiple of the empirical variance

$$\widehat{t}_n = \frac{1}{n}\overline{X_n^2} + \Big(1-\frac{1}{n}\Big)\overline{X}_n^2 - (\overline{X}_n)^2 = \frac{1}{n}\big(\overline{X_n^2} - \overline{X}_n^2\big).$$

We can compensate the bias from the original estimator

$$\widehat{T}_n^B := \overline{X}_n^2 - \widehat{t}_n = \left(1+\frac{1}{n}\right)\overline{X}_n^2 - \frac{1}{n}\overline{X_n^2}.$$

The bias of the estimator after compensation is

$$\begin{aligned}
&\mathbb{E}\widehat{T}_n^B - \mu_F^2 = \mathbb{E}_F\left(1+\frac{1}{n}\right)\overline{X}_n^2 - \frac{1}{n}\mathbb{E}_F\overline{X_n^2} - \mu_F^2 = \\
&\left(1+\frac{1}{n}\right)\mathbb{E}_F(\overline{X}_n - \mu_F)^2 + \left(1+\frac{1}{n}\right)\mu_F^2 - \frac{1}{n}\mathbb{E}_F\overline{X_n^2} - \mu_F^2 = \\
&\left(1+\frac{1}{n}\right)\frac{1}{n}\operatorname{Var}_F(X_1) + \left(1+\frac{1}{n}\right)\mu_F^2 - \frac{1}{n}(\operatorname{Var}_F(X_1)+\mu_F^2) - \mu_F^2 = \frac{1}{n^2}\operatorname{Var}_F(X_1),
\end{aligned}$$

which is by an order of magnitude smaller than the bias of the original estimator,

$$\mathbb{E}_F T_n(\widehat{F}_n) - \mu_F^2 = \mathbb{E}_F(\overline{X}_n)^2 - \mu_F^2 = \mathbb{E}_F(\overline{X}_n - \mu_F)^2 = \frac{1}{n}\operatorname{Var}_F(X_1).$$

A different bias reduction technique, called the *Jackknife method*, yields zero bias in this example (Problem 5).

If the true distribution is normal $N(\mu,\sigma^2)$ with unknown parameters μ and σ^2, we can apply parametric bootstrap. To this end we can use the MLE

$$\widehat{\mu}_n = \overline{X}_n \quad \text{and} \quad \widehat{\sigma}_n^2 = S_n^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2,$$

in which case the samples X_j^* are conditionally i.i.d. from the normal $N(\widehat{\mu}_n, \widehat{\sigma}_n^2)$ distribution. Hence

$$\mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*) = \mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^*\right)^2 = \mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^* - \widehat{\mu}_n\right)^2 + \widehat{\mu}_n^2 = \frac{1}{n}\widehat{\sigma}_n^2 + \widehat{\mu}_n^2$$

and the parametric bootstrap bias estimator

$$\widehat{t}_n = \mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*) - \widehat{\mu}_n^2 = \frac{1}{n}\widehat{\sigma}_n^2 = \frac{1}{n}(\overline{X_n^2} - \overline{X}_n^2)$$

coincides with the nonparametric one.

Let us try another parametric family, for example, $X_j \sim \operatorname{Exp}(\lambda)$. The MLE of μ_F is $\widehat{\mu}_n = \overline{X}_n$ and the distribution $\widehat{F}_n^*$ is $\operatorname{Exp}(1/\overline{X}_n)$. Since for the exponential distribution the variance equals squared mean,

$$\mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*) = \mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^*\right)^2 = \mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^* - \widehat{\mu}_n\right)^2 + \widehat{\mu}_n^2 = \frac{1}{n}\operatorname{Var}_{\widehat{F}_n}(X_1^*) + \widehat{\mu}_n^2 = \frac{1+n}{n}\widehat{\mu}_n^2,$$

and hence parametric bootstrap bias estimator (1.5) is

$$\widehat{t}_n = \frac{1}{n}\widehat{\mu}_n^2 = \frac{1}{n}\overline{X}_n^2.$$

This formula is different from the parametric bootstrap in the normal case; it can be readily checked that the residual bias is of order $O(n^{-2})$ as well (check !). ■

1.2. Variance estimation. Another typical application of the bootstrap method is to estimation of variance. Let $T(F)$ be a functional of interest, $T(\widehat{F}_n)$ its plug-in estimator and its variance

$$\mathrm{Var}_F\left(T(\widehat{F}_n)\right) = \mathbb{E}_F T(\widehat{F}_n)^2 - \left(\mathbb{E}_F T(\widehat{F}_n)\right)^2.$$

This is easily put into the form (1.2) and the bootstrap variance estimator is

$$\widehat{t}_n := \mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*)^2 - \left(\mathbb{E}_{\widehat{F}_n} T(\widehat{F}_n^*)\right)^2. \tag{1.6}$$

EXAMPLE 1.5. Let us specify the bootstrap variance estimator (1.6) for $T(F) = \mu_F^2$. In this case (1.6) reads

$$\widehat{t}_n := \mathbb{E}_{\widehat{F}_n}\mu^4_{\widehat{F}_n^*} - \left(\mathbb{E}_{\widehat{F}_n}\mu^2_{\widehat{F}_n^*}\right)^2 = \mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^*\right)^4 - \left(\mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^*\right)^2\right)^2. \tag{1.7}$$

Define the centred empirical moments

$$\widehat{m}_p = \mathbb{E}_{\widehat{F}_n}(X_1^* - \overline{X}_n)^p = \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^p.$$

Then

$$\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^*)^2 = \mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n + \overline{X}_n)^2 = \mathrm{Var}_{\widehat{F}_n}\left(\overline{X}_n^*\right) + \overline{X}_n^2 = \frac{1}{n}\widehat{m}_2 + \overline{X}_n^2,$$

where the last equality holds, since the cross term vanishes

$$\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)\overline{X}_n = \overline{X}_n \frac{1}{n}\sum_{j=1}^n \mathbb{E}_{\widehat{F}_n}(X_j^* - \overline{X}_n) = \overline{X}_n \mathbb{E}_{\widehat{F}_n}(X_1^* - \overline{X}_n) = 0.$$

Similarly,

$$\begin{aligned}
&\mathbb{E}_{\widehat{F}_n}\left(\overline{X}_n^*\right)^4 = \mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n + \overline{X}_n)^4 = \\
&\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)^4 + 4\overline{X}_n\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)^3 + 6\overline{X}_n^2\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)^2 + \overline{X}_n^4,
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)^3 =& \mathbb{E}_{\widehat{F}_n}\left(\frac{1}{n}\sum_{j=1}^n (X_j^* - \overline{X}_n)\right)^3 = \frac{1}{n^3}\sum_{j=1}^n \mathbb{E}_{\widehat{F}_n}(X_j^* - \overline{X}_n)^3 = \\
& \frac{1}{n^2}\mathbb{E}_{\widehat{F}_n}(X_1^* - \overline{X}_n)^3 = \frac{1}{n^2}\widehat{m}_3 \\
\mathbb{E}_{\widehat{F}_n}(\overline{X}_n^* - \overline{X}_n)^4 =& \mathbb{E}_{\widehat{F}_n}\left(\frac{1}{n}\sum_{j=1}^n (X_j^* - \overline{X}_n)\right)^4 = \frac{1}{n^4}n\mathbb{E}_{\widehat{F}_n}(X_1^* - \overline{X}_n)^4 + \\
& \frac{1}{n^4}\binom{4}{2}\frac{n(n-1)}{2}\left(\mathbb{E}_{\widehat{F}_n}(X_1^* - \overline{X}_n)^2\right)^2 = \frac{1}{n^3}\widehat{m}_4 + 3\frac{n(n-1)}{n^4}\widehat{m}_2^2.
\end{aligned}$$

Plugging all these expressions into (1.7) we obtain

$$\begin{aligned}
\widehat{t}_n =& \frac{1}{n^3}\widehat{m}_4 + 3\frac{n(n-1)}{n^4}\widehat{m}_2^2 + 4\overline{X}_n\frac{1}{n^2}\widehat{m}_3 + 6\overline{X}_n^2\frac{1}{n}\widehat{m}_2 + \overline{X}_n^4 - \left(\frac{1}{n}\widehat{m}_2 + \overline{X}_n^2\right)^2 = \\
& \frac{4\overline{X}_n^2\widehat{m}_2}{n} + \frac{2\widehat{m}_2^2 + 4\overline{X}_n\widehat{m}_3}{n^2} + \frac{\widehat{m}_4 - 3\widehat{m}_2^2}{n^3}
\end{aligned} \tag{1.8}$$

A simple calculation shows that the actual variance satisfies

$$\mathrm{Var}_F(T(\widehat{F}_n)) = \mathrm{Var}_F(\overline{X}_n^2) = 4\mu_F^2\sigma_F^2\frac{1}{n} + O(n^{-3/2}),$$

and hence the estimator in (1.8) is exact at least to the leading asymptotic order. ■

1.3. Confidence intervals. An important application of the bootstrap method is to construction of confidence intervals. Let us first recall the standard pivot technique. Consider the equation

$$\mathbb{P}_F\Big(\big|T(F) - T(\widehat{F}_n)\big| \le t\Big) = 1 - \alpha, \tag{1.9}$$

where $1-\alpha$ is the required coverage probability. If $T(\widehat{F}_n)$ is a pivotal statistic, that is, if the distribution of $T(F) - T(\widehat{F}_n)$ under $\mathbb{P}_F$ does not depend on F, then the solution of (1.9), call it $t_{n,\alpha}$, will not depend on the unknown distribution F and hence the confidence interval with the endpoints $T(\widehat{F}_n) \pm t_{n,\alpha}$ will have the desired coverage probability.

This approach is rarely realizable exactly since typically $T(\widehat{F}_n)$ will not be pivotal. Instead, if $T(F)$ is differentiable one can use the nonparametric Delta method to construct *asymptotic* confidence interval using the approximation

$$\sqrt{n}\big(T(\widehat{F}_n) - T(F)\big) \xrightarrow[n\to\infty]{d(\mathbb{P}_F)} N\big(0, V_F\big),$$

where V_F is an explicit functional of F. If $V_{\widehat{F}_n}$ is consistent, the confidence interval

$$T(\widehat{F}_n) \pm \frac{1}{\sqrt{n}}\sqrt{V_{\widehat{F}_n}} z_{1-\alpha/2} \tag{1.10}$$

has asymptotic coverage probability of $1-\alpha$.

In this context, bootstrap is an alternative approach, which applies directly to equation (1.9), which fits the format of (1.2) regardless of whether $T(\widehat{F}_n)$ is pivotal or not! The corresponding bootstrap equation is

$$\mathbb{P}_{\widehat{F}_n}\Big(|T(\widehat{F}_n) - T(\widehat{F}_n^*)| \le t\Big) = 1 - \alpha, \tag{1.11}$$

to be solved for t. The obtained confidence interval

$$T(\widehat{F}_n) \pm \widehat{t}_{n,\alpha},$$

where $\widehat{t}_{n,\alpha}$ is the solution, will have asymptotic coverage probability $1-\alpha$. This follows from the analysis of the bootstrap approximation, to be explained below in some detail (see Theorem 2.1). In fact, quite remarkably, the convergence of the actual coverage probability of the bootstrap confidence intervals to $1-\alpha$ can be significantly faster than that of (1.10)!

Since $\widehat{F}_n$ is a discrete distribution, it will be typically impossible to solve equation (1.11) exactly. Instead we can find the least t for which the left hand side becomes greater or equal $1-\alpha$. The error thus introduced is asymptotically negligible if F is continuous. Another practical difficulty is that quantiles of the distribution of

$$S = |T(\widehat{F}_n) - T(\widehat{F}_n^*)|$$

under $\mathbb{P}_{\widehat{F}_n}$ are usually cumbersome to compute exactly (Problem 7 is a rare example where this is possible). This is where the Monte Carlo approximation, mentioned in Remark 1.2, comes to rescue. The quantile in question can be approximated by the quantile of the empirical distribution of i.i.d. samples of S under $\mathbb{P}_{\widehat{F}_n}$, within any desired precision.

2. Why does bootstrap work?

2.1. Consistency. General theory of bootstrap is quite involved and has to do with empirical processes. Here is a typical result, see [1], [3] and references therein.

THEOREM 2.1. *Suppose that functional $T(\cdot)$ is Hadamard differentiable with respect to the uniform norm with the influence function $0 < \int L_F^2(x)dF(x) < \infty$. Then*

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}_F\Big(\sqrt{n}\big(T(\widehat{F}_n)-T(F)\big)\le u\Big)-\mathbb{P}_{\widehat{F}_n}\Big(\sqrt{n}\big(T(\widehat{F}_n^*)-T(\widehat{F}_n)\big)\le u\Big)\right| \xrightarrow[n\to\infty]{\mathbb{P}_F} 0.$$

This theorem asserts that the asymptotic distribution of $T(\widehat{F}_n)$ is consistently estimated by bootstrap. In particular, this implies that the coverage probability of the bootstrap confidence interval converges to its target value $1-\alpha$ with the sample size. Indeed, note that instead of solving for t in (1.11), we can equivalently solve

$$\mathbb{P}_{\widehat{F}_n}\Big(\sqrt{n}|T(\widehat{F}_n)-T(\widehat{F}_n^*)|\le u\Big)=1-\alpha,$$

for $u=t/\sqrt{n}$. The confidence interval then has its endpoints at $T(\widehat{F}_n)\pm\widehat{u}_n/\sqrt{n}$, where $\widehat{u}_n$ is the solution to the latter equation (it depends on the sample!). The coverage probability is

$$\mathbb{P}_F\big(\sqrt{n}|T(F)-T(\widehat{F}_n)|\le\widehat{u}_n\big)$$

and by Theorem 2.1,

$$\begin{aligned}
&\Big|\mathbb{P}_F\big(\sqrt{n}|T(F)-T(\widehat{F}_n)|\le\widehat{u}_n\big)-(1-\alpha)\Big| = \\
&\Big|\mathbb{P}_F\big(\sqrt{n}|T(F)-T(\widehat{F}_n)|\le\widehat{u}_n\big)-\mathbb{P}_{\widehat{F}_n}\Big(\sqrt{n}|T(\widehat{F}_n)-T(\widehat{F}_n^*)|\le\widehat{u}_n\Big)\Big| \le \\
&\sup_{u\in\mathbb{R}}\Big|\mathbb{P}_F\big(\sqrt{n}|T(F)-T(\widehat{F}_n)|\le u\big)-\mathbb{P}_{\widehat{F}_n}\Big(\sqrt{n}|T(\widehat{F}_n)-T(\widehat{F}_n^*)|\le u\Big)\Big| \to 0.
\end{aligned}$$

The proof of Theorem 2.1 is beyond the scope of these notes. To demonstrate some ideas, let us give a proof for a particularly simple functional, the mean $\mu_F = \int xdF(x)$. To this end, we will need the following uniform estimate in the CLT.

THEOREM 2.2 (Berry-Esseen). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with $\mu = \mathbb{E}X_1$, $\sigma^2 = \mathbb{E}X_1^2$ and $\rho = \mathbb{E}|X_1-\mu|^3 < \infty$. Then*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\Big(\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}\le x\Big)-\Phi(x)\right|\le C\frac{\rho}{\sigma^3}n^{-\frac{1}{2}},$$

where C is an absolute constant[2].

For $T(F) = \mu_F$ we have

$$\mathbb{P}_{\widehat{F}_n}\left(\sqrt{n}\big(T(\widehat{F}_n^*) - T(\widehat{F}_n)\big) \le x\right) = \mathbb{P}_{\widehat{F}_n}\left(\sqrt{n}\frac{\overline{X}_n^* - \overline{X}_n}{\widehat{\sigma}_n} \le \frac{x}{\widehat{\sigma}_n}\right),$$

where we divided by $\widehat{\sigma}_n$, the empirical standard deviation of the original sample. Thus the B-E bound, applied to the bootstrap sample, implies

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\widehat{F}_n}\left(\sqrt{n}\big(T(\widehat{F}_n^*) - T(\widehat{F}_n)\big) \le x\right) - \Phi(x/\widehat{\sigma}_n)\right| \le C\frac{\widehat{\rho}_n}{\widehat{\sigma}_n^3}n^{-\frac{1}{2}},$$

where $\widehat{\rho}_n$ is the empirical third central moment. Thus by the triangle inequality

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\mathbb{P}_F\left(\sqrt{n}\big(T(\widehat{F}_n) - T(F)\big) \le x\right) - \mathbb{P}_{\widehat{F}_n}\left(\sqrt{n}\big(T(\widehat{F}_n^*) - T(\widehat{F}_n)\big) \le x\right)\right| \le \\ &C\frac{\rho}{\sigma^3}n^{-\frac{1}{2}} + \sup_{x\in\mathbb{R}}\big|\Phi(x/\widehat{\sigma}_n) - \Phi(x/\sigma)\big| + C\frac{\widehat{\rho}_n}{\widehat{\sigma}_n^3}n^{-\frac{1}{2}} \xrightarrow[n\to\infty]{\mathbb{P}_F} 0,\end{aligned} \tag{2.1}$$

where the uniform convergence of the second term is argued in Problem 8.

2.2. Coverage error of bootstrap intervals. Some insights into the large sample asymptotic properties of the bootstrap confidence intervals can be gained form the following simple example.

EXAMPLE 2.3. Let us construct parametric bootstrap confidence interval for the mean μ_F of normal distribution $N(\mu,\sigma^2)$. In this case the bootstrap sample is normal $N(\widehat{\mu}_n,\widehat{\sigma}_n^2)$, where $\widehat{\mu}_n$ and $\widehat{\sigma}_n^2$ are, e.g., the best unbiased estimators (UMVUE's) see Remark 1.1,

$$\widehat{\mu}_n = \overline{X}_n \quad \text{and} \quad \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{j=1}^n (X_j - \overline{X}_n)^2 =: S^2(\widehat{F}_n).$$

Hence $\mu_{\widehat{F}_n^*} \sim N(\widehat{\mu}_n, n^{-1}\widehat{\sigma}_n^2)$ under $\mathbb{P}_{\widehat{F}_n}$ and the equation (1.11) becomes

$$2\Phi\left(\sqrt{n}\frac{t}{\widehat{\sigma}_n}\right) - 1 = 1 - \alpha.$$

Hence the bootstrap quantile estimator is

$$\widehat{t}_n = \frac{\widehat{\sigma}_n}{\sqrt{n}}z_{1-\alpha/2},$$

and the corresponding confidence interval has endpoints at

$$\widehat{\mu}_n \pm \frac{\widehat{\sigma}_n}{\sqrt{n}}z_{1-\alpha/2}.$$

The actual coverage probability can be computed exactly,

$$\mathbb{P}_F\left(|\widehat{\mu}_n - \mu| \le \widehat{t}_n\right) = \mathbb{P}_F\left(\left|\sqrt{n}\frac{\overline{X}_n - \mu}{\sqrt{S_n^2}}\right| \le z_{1-\alpha/2}\right) = 2G_{n-1}(z_{1-\alpha/2}) - 1,$$

[2]the value of C is only known to lie between 0.4 and 0.475

where G_{n-1} is the Student distribution with $n-1$ degrees of freedom. Hence the error in coverage probability stems from the distance between Student and normal laws. A calculation shows that this deviation satisfies

$$G_{n-1}(z_{1-\alpha/2}) - \Phi(z_{1-\alpha/2}) = O(n^{-1}).$$

Note that this is, in fact, better than predicted by the rough estimates provided by the B-E bound in (2.1).

The bootstrap method can be also applied to the *pivotal* equation

$$\mathbb{P}_F\left(\left|(\mu_{\widehat{F}_n} - \mu_F)\Big/\sqrt{S^2(\widehat{F}_n)}\right| \le t\right) = 1-\alpha, \tag{2.2}$$

whose bootstrap analog is

$$\mathbb{P}_{\widehat{F}_n}\left(\left|(\mu_{\widehat{F}^*_n} - \mu_{\widehat{F}_n})\Big/\sqrt{S^2(\widehat{F}^*_n)}\right| \le t\right) = 1-\alpha. \tag{2.3}$$

Since $X_1^*, \dots, X_n^* \overset{\text{i.i.d.}}{\sim} N(\widehat{\mu}_n, \widehat{\sigma}_n^2)$, the statistic under the absolute value has Student distribution (scaled by $\sqrt{n}$) and therefore the parametric bootstrap in this case coincides with the Student confidence interval, and is therefore exact. ■

This simple example indicates that the coverage error of (at least, parametric) bootstrap confidence interval is of order $O(n^{-1})$, if based on the statistic $\mu - \widehat{\mu}_n$, but it can be even better, if based on the *studentized* statistic $(\mu - \widehat{\mu}_n)/\widehat{\sigma}_n$. Remarkably, this turns out to be true in greater generality. Calculation as in (2.1), based on the Berry-Esseen theorem, shows that the coverage error of nonparametric bootstrap confidence intervals based on $T(F) - T(\widehat{F}_n)$ is at least $O(n^{-1/2})$. For the symmetric interval (but not necessarily otherwise) the error is actually of order $O(n^{-1})$. This can be seen using the Edgeworth expansion of the normal law, which is yet another refinement of the CLT, see [**2**].

Yet even better rates are possible, if the bootstrap method is applied to studentized pivotal quantities. Let $\sigma_{T,n}(F)$ the standard deviation of $T(F)$ for a sample of size n,

$$\sigma_{T,n}(F)^2 = \mathbb{E}_F\big(T(\widehat{F}_n) - T(F)\big)^2.$$

The bootstrap interval constructed on the basis of equation

$$\mathbb{P}_F\left(\left|\big(T(\widehat{F}_n) - T(F)\big)\Big/\sigma_{T,n}(\widehat{F}_n)\right| \le t\right) = 1-\alpha$$

is called symmetric *studentized pivotal* interval, that is, t is found by solving the equation

$$\mathbb{P}_{\widehat{F}_n}\left(\left|\big(T(\widehat{F}^*_n) - T(\widehat{F}_n)\big)\Big/\sigma_{T,n}(\widehat{F}^*_n)\right| \le t\right) = 1-\alpha. \tag{2.4}$$

Note that (2.2) is a particular instance of the parametric bootstrap confidence interval of this kind. Remarkably, under certain technical conditions, studentized bootstrap interval (2.4) has coverage error of order $O(n^{-2})$, see [**2**].

This improvement however does not come for free. The probability in (2.4) is approximated by the average over B i.i.d. copies of $\widehat{F}^*_n$ as in (1.4). If the functional $F \mapsto \sigma_{T,n}(F)$ does not have a closed form expression, the quantity $\sigma_{T,n}(\widehat{F}^{*j}_n)$ for each $j = 1, \ldots, B$ is to be approximated using an additional averaging over i.i.d. copies of $\widehat{F}_n^{**}$. This may drastically increase the computational cost.

Computer experiment

Consider the nonparametric estimation problem for the median of the unknown distribution. Implement the following asymptotic confidence intervals.

(1) The confidence interval obtained by means of the Delta-method. To this end, estimate the unknown density f, needed for calculation of the limit variance, by k-neighbours estimator

$$\widehat{f}_n(x) = \frac{k}{n} \frac{1}{D\big(N_k(x)\big)}, \quad x \in \mathbb{R}.$$

where $N_k(x)$ is the set of k sample points, closest to x, $D(A) = \max(A) - \min(A)$ is the diameter of the set A and $k = [n^{4/5}]$. This particular choice of k will be explained later in the course.

(2) Bootstrap interval based on (1.11) (use $B = 1,000$ in (1.4)).

(3) Studentized bootstrap interval based on (2.4).

Generate a sample of size $n = 100$ from a known density and evaluate the above intervals. Compare the results and the running time of the code. Run $M = 10,000$ Monte Carlo trials to approximate the coverage probability and the mean length of the confidence intervals in (1) and (2). Repeat the experiment for the sample size of $n = 1,000$. Summarize your findings in view of the theory presented in this section.

Exercises

PROBLEM 1. Show that the bootstrap estimators of the mean μ_F, based on the equations $\mu_F - t = 0$ and $\mathbb{E}_F \mu_{\widehat{F}_n} - t = 0$, both equal to the empirical mean $\overline{X}_n$.

PROBLEM 2. For the sample of size $n = 3$ without ties, find the bootstrap distribution of the empirical mean.

PROBLEM 3. Specify the bootstrap variance estimator for the plug-in estimator of the functional

$$T(F) = \int \psi dF,$$

where ψ is a fixed function.

PROBLEM 4. Specify the bootstrap bias estimator for the plug-in estimator of variance

$$\sigma_F^2 = \int x^2 dF(x) - \Big(\int x dF(x) \Big)^2.$$

Argue that the estimator, obtained after compensation of the bias, is not unbiased, but its residual bias is of order $O(n^{-2})$.

PROBLEM 5 (Jackknife bias estimator). Suppose T_n is an estimator for some quantity $\theta(F)\in\mathbb{R}$, based on an i.i.d. sample of size n. Assume that its bias satisfies

$$\mathbb{E}_F T_n - \theta(F) = \frac{a_F}{n} + \frac{b_F}{n^2} + O\Big(\frac{1}{n^3}\Big) \tag{2.5}$$

with some functionals a_F and b_F. Let $T_{n/i}$ the value of the statistic T_n obtained by evaluating it on the sample, with the point X_i being removed. Define $\overline{T}_n = \frac{1}{n}\sum_{j=1}^n T_{n/j}$ and let the Jackknife bias estimator

$$\widehat{\beta}_n^J = (n-1)\big(\overline{T}_n - T_n\big).$$

(1) Show that

$$\mathbb{E}_F \widehat{\beta}_n^J = \frac{a_F}{n} + O(n^{-2})$$

and therefore the estimator $T_n^J := T_n - \widehat{\beta}_n^J$ estimates $\theta(F)$ with bias of smaller order than T_n, cf. (2.5),

$$\mathbb{E}_F T_n^J - \theta(F) = O(n^{-2})$$

(2) Show that the plug-in estimator $(\overline{X}_n)^2$ for the squared mean $\theta(F) = \big(\int x dF(x)\big)^2 = \mu_F^2$ satisfies the assumption (2.5) and specify the Jackknife estimator. Argue that it is, in fact, unbiased in this case.

PROBLEM 6. Construct studentized parametric bootstrap confidence interval for the mean of exponential distribution. Does the obtained interval have exact coverage?

PROBLEM 7.

(1) Specify the construction of $1-\alpha$ bootstrap confidence interval for $T(F) = F(x_0)$ where x_0 is a given point.

(2) Evaluate your interval for $\alpha = 0.2$ and $n = 10$, if $\widehat{F}_n(x_0) = 1/2$ was observed.

PROBLEM 8. Prove that the second term in the right hand side of (2.1) converges to zero at rate $n^{-\frac{1}{2}}$.

References

[1] Larry Wasserman, All of nonparametric statistics. Springer Science & Business Media, 2006.
[2] Peter Hall, The bootstrap and Edgeworth expansion, Springer, 1992
[3] Van der Vaart, Aad W. Asymptotic statistics. Vol. 3. Cambridge university press, 2000.

DEPARTMENT OF STATISTICS, THE HEBREW UNIVERSITY, MOUNT SCOPUS, JERUSALEM 91905, ISRAEL
E-mail address: Pavel.Chigansky@mail.huji.ac.il