Avd. Matematisk statistik
KTH Teknikvetenskap

Sf 2955: Computer intensive methods :
ON THE BOOTSTRAP HYPOTHESIS AND BOOTSTRAP CONSISTENCY

Timo Koski

1 Preliminaries

Let X be a random variable and let a distribution function F on the real line be defined as

$$F(x) = \mathbf{P}(X \leq x), \quad -\infty < x < \infty.$$

We assume that the true distribution function is a member of a class of distribution functions $\mathcal{M}$. Let θ, the quantity of interest, be

$$\theta = T(F).$$

Let $x_1, \ldots, x_n$ be a sample of $X_1, \ldots, X_n$, I.I.D. with the distribution F. The empirical (cumulative) distribution function is

$$\widehat{F}_n(x) = \frac{1}{n} \times (\text{ the number of } X_i \leq x).$$

The plug-in estimator $\widehat{\theta}_n$ of $\theta = T(F)$ on basis of $X_1, \ldots, X_n$ is defined by

$$\widehat{\theta}_n = T\left(\widehat{F}_n\right) = \widehat{\theta}(X_1, \ldots, X_n). \tag{1.1}$$

Let $X_1^*, \ldots, X_n^*$ be the bootstrap random variables based on $x_1, \ldots, x_n$, i.e., $P\left(X_j^* = x_i\right) = \frac{1}{n}$ for any j and i. Then

$$\widehat{\theta}_n^* = \widehat{\theta}(X_1^*, \ldots, X_n^*). \tag{1.2}$$

is the plug-in estimator in terms of the the bootstrap random variables.

Example 1.1 The Mean Let $\theta = E[X]$. Then

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}, \tag{1.3}$$

and thus

$$\widehat{\theta}_n^* = \frac{1}{n}\sum_{i=1}^{n} X_i^* = \overline{X^*}, \tag{1.4}$$

■

2 The Bootstrap Hypothesis

For many statistical problems we are interested in finding the distribution function

$$F_{\widehat{\theta}_n}(x) = P_F\left(\widehat{\theta}_n - \theta \leq x\right), \quad -\infty < x < \infty.$$

This may be difficult or impossible to find analytically without some simplifying assumptions. The idea in bootstrapping is to study $F_{\widehat{\theta}_n}$ by using the bootstrap distribution

$$F_{\widehat{\theta}_n^*}(x) = P_{\widehat{F}_n}\left(\widehat{\theta}_n^* - \widehat{\theta}_n \leq x\right).$$

Definition 2.1 Bootstrap Hypothesis We think that

$$F_{\widehat{\theta}_n}(x) \approx F_{\widehat{\theta}_n^*}(x)$$

with high probability for large n and and uniformly in x.

■

One has to note that in this $F_{\widehat{\theta}_n^*}$ is a random variable as it is ultimately a function of $X_1, \ldots, X_n$.

We can make the statement of the bootstrap hypothesis precise by considering the quantity

$$K\left(F_{\widehat{\theta}_n}, F_{\widehat{\theta}_n^*}\right) \stackrel{\text{def}}{=} \sup_{x \in R} \mid F_{\widehat{\theta}_n}(x) - F_{\widehat{\theta}_n^*}(x) \mid .$$

In any given situation we should show, of course, that

$$\lim_{n\to\infty} K\left(F_{\widehat{\theta}_n}, F_{\widehat{\theta}_n^*}\right) \stackrel{a.s}{=} 0,$$

as $n \to \infty$.

3 The Bootstrap Hypothesis for Estimating the Mean

We shall consider the case in the example 1.1 above, or, estimator (1.3) of $\theta = E[X]$, when $\sigma^2 = V[X]$. We observe (recall) the following.

1. $$E_F\left[\widehat{\theta}_n\right] = \theta \tag{3.1}$$

2. $$V_F\left[\widehat{\theta}_n\right] = \frac{\sigma^2}{n}. \tag{3.2}$$

3. $$E_{\widehat{F}_n}\left[\widehat{\theta}_n^*\right] = \frac{1}{n}\sum_{i=1}^{n} E_{\widehat{F}_n}\left[X_i^*\right] = E_{\widehat{F}_n}\left[X_1^*\right] = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}. \tag{3.3}$$

4. $$\begin{aligned} V_{\widehat{F}_n}\left[\widehat{\theta}_n^*\right] &= \frac{1}{n^2}\sum_{i=1}^{n} V_{\widehat{F}_n}\left[X_i^*\right] \\ &= \frac{1}{n} V_{\widehat{F}_n}\left[X_1^*\right] \\ &= \frac{1}{n} E_{\widehat{F}_n}\left[\left(X_1^* - E_{\widehat{F}_n}\left[X_1^*\right]\right)^2\right] \\ &= \frac{1}{n} E_{\widehat{F}_n}\left[\left(X_1^* - \overline{x}\right)^2\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \frac{\widehat{\sigma}^2}{n}. \end{aligned} \tag{3.4}$$

   Here we have the plug-in estimate of σ^2,

   $$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2. \tag{3.5}$$

For study of bootstrap hypothesis we will consider the two sequences of random variables

$$H_n = \sqrt{n}\left(\widehat{\theta}_n - \theta\right) \tag{3.6}$$

and

$$H_n^* = \sqrt{n}\left(\widehat{\theta}_n^* - \overline{X}\right). \tag{3.7}$$

We set

$$K_n \stackrel{\text{def}}{=} \sup_{x \in R} \mid P_F\left(\frac{H_n}{\sigma} \leq \frac{x}{\sigma}\right) - P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq \frac{x}{\widehat{\sigma}}\right) \mid$$

and use $\Phi(t)$, the cumulative distribution function of $N(0,1)$, in the identity

$$= \sup_{x \in R} \mid P_F\left(\frac{H_n}{\sigma} \leq \frac{x}{\sigma}\right) - \Phi\left(\frac{x}{\sigma}\right)$$

$$+\Phi\left(\frac{x}{\sigma}\right) - \Phi\left(\frac{x}{\widehat{\sigma}}\right)$$

$$+\Phi\left(\frac{x}{\widehat{\sigma}}\right) - P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq \frac{x}{\widehat{\sigma}}\right) \mid .$$

By the triangle inequality we get the upper bound

$$K_n \leq A_n + B_n + C_n,$$

where

$$A_n = \sup_{x \in R} \mid P_F\left(\frac{H_n}{\sigma} \leq \frac{x}{\sigma}\right) - \Phi\left(\frac{x}{\sigma}\right) \mid \tag{3.8}$$

$$B_n = \sup_{x \in R} \mid \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(\frac{x}{\widehat{\sigma}}\right) \mid \tag{3.9}$$

$$C_n = \sup_{x \in R} \mid P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq \frac{x}{\widehat{\sigma}}\right) - \Phi\left(\frac{x}{\widehat{\sigma}}\right) \mid . \tag{3.10}$$

We shall now show that each of these sequences will converge to zero, as $n \to \infty$.

4 Proof of The Bootstrap Hypothesis for Estimating the Mean

We have derived the inequality

$$K_n \leq A_n + B_n + C_n,$$

Here A_n is a non-random quantity, which is shown to converge to zero by the famous Berry-Esseen bound. Then we shall show that B_n converges to zero almost surely by the fact that $\widehat{\sigma}^2$ converges (consistency of estimation) to σ^2 in view of the law of large numbers. Finally we use Berry-Esseen again to find a random upper bound for the nonnegative random variable C_n. Then it is established that this upper bound converges to zero by invoking the Zygmund -Marcinkiewicz strong law of large numbers.

4.1 The Berry-Esseen bound, $A_n \to 0$

We write by (1.3)

$$\frac{H_n}{\sigma} = \frac{\left(\sum_{i=1}^{n} X_i - n\theta\right)}{\sqrt{n}\sigma}$$

so that

$$A_n = \sup_{x \in R} \mid P_F \left(\frac{\left(\sum_{i=1}^{n} X_i - n\theta\right)}{\sqrt{n}\sigma} \leq \frac{x}{\sigma} \right) - \Phi\left(\frac{x}{\sigma}\right) \mid .$$

The **central limit theorem** tells that

$$\frac{\left(\sum_{i=1}^{n} X_i - n\theta\right)}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1),$$

as $n \to \infty$. But we know and need even more, we have the following inequality giving a kind of speed of convergence, see A. Gut: *An Intermediate Course in Probability. 2nd Ed.* p. 165, eq. (5.4),

$$\sup_{x \in R} \mid F_{\frac{\left(\sum_{i=1}^{n} X_i - n\theta\right)}{\sqrt{n}\sigma}}(x) - \Phi(x) \mid \leq \frac{c}{\sigma^3} \frac{E_F\left[\mid X_1 - \theta \mid^3\right]}{\sqrt{n}}. \tag{4.1}$$

Clearly we require also that $E_F\left[|X_1 - \theta|^3\right] < \infty$. The inequality in (4.1) is known as the **Berry-Esseen bound**, c.f. Esseen 1944. c is a universal constant (one can take $c = 0.8$, sharper bounds have been found recently) that does not depend on n. But clearly this shows that $A_n \to 0$, as $n \to \infty$.

4.2 $B_n \to 0$

This is the simple case. In view of (3.5)

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2,$$

we get by an algebraic identity that

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}^2$$

The strong law of large numbers gives, as $n \to \infty$, that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \overset{a.s.}{\to} E_F\left[X^2\right],$$

and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{a.s.}{\to} \theta$$

and thus

$$\widehat{\sigma}^2 \overset{a.s.}{\to} E_F\left[X^2\right] - 2\theta^2 + \theta^2 = E_F\left[X^2\right] - \theta^2 = \sigma^2. \tag{4.2}$$

In other words, $\widehat{\sigma}^2$ is a consistent estimator of σ^2. Thus

$$\Phi\left(\frac{x}{\widehat{\sigma}}\right) \overset{a.s.}{\to} \Phi\left(\frac{x}{\sigma}\right),$$

for every x, as $n \to \infty$, as $\Phi(x)$ is a continuous function. However, for the desired conclusion to hold we need that $\Phi(x)$ is a **uniformly continuous function** [1]. Then, by (3.9)

$$B_n \overset{a.s.}{\to} 0.$$

[1]**Uniform continuity**: We have by mean value theorem of differential calculus that

$$\Phi(x+h) - \Phi(x) = \Phi^{'}(\xi)\, h$$

where $\xi = (1-\lambda)x + \lambda(x+h) = x + \lambda h$, $0 < \lambda < 1$. Since

$$\Phi^{'}(\xi) = \phi(\xi) \leq \phi(0)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, we get

$$\mid \Phi(x+h) - \Phi(x) \mid \leq \phi(0) \cdot |h|.$$

Let us now fix an arbitrary $\epsilon > 0$. Hence, for $|h| < \frac{\epsilon}{\phi(0)}$

$$\mid \Phi(x+h) - \Phi(x) \mid \leq \epsilon,$$

and this bound is **the same for all** x for a given arbitrary $\epsilon > 0$.

4.3 The Berry-Esseen bound, $C_n \to 0$

Let us first condition on $X_1 = x_1, \ldots, X_n = x_n$. Then $E_{\widehat{F}_n}\left[\widehat{\theta}_n^*\right] = \overline{x}$ and we consider in (3.7) the bootstrap random variable

$$H_n^* = \sqrt{n}\left(\widehat{\theta}_n^* - \overline{x}\right). \tag{4.3}$$

Then we can write as in the preceding case

$$\frac{H_n^*}{\widehat{\sigma}} = \frac{\sum_{i=1}^n X_i^* - n\overline{x}}{\sqrt{n}\widehat{\sigma}}.$$

Now we apply the Berry-Esseen bound (4.1) again on the distribution of the variable in the right hand side, since the means and variances are given in (3.3) and (3.5), respectively, w.r.t $\widehat{F}_n$ are fixed. Thus

$$\sup_{x \in R} \mid P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq x\right) - \Phi(x) \mid$$

$$\leq \frac{c^*}{\widehat{\sigma}^3} \frac{E_{\widehat{F}_n}\left[\mid X_1^* - \overline{x} \mid^3\right]}{\sqrt{n}}.$$

We have, as above,

$$E_{\widehat{F}_n}\left[\mid X_1^* - \theta \mid^3\right] = \int_{-\infty}^{\infty} \mid x - \overline{x} \mid^3 d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mid x_i - \overline{x} \mid^3.$$

Thus we have obtained

$$\sup_{x \in R} \mid P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq x\right) - \Phi(x) \mid \leq \frac{c}{\widehat{\sigma}^3}\frac{1}{n^{3/2}}\sum_{i=1}^n \mid x_i - \overline{x} \mid^3. \tag{4.4}$$

We need an auxiliary inequality, or

$$\sum_{i=1}^n \mid x_i - \overline{x} \mid^3 \leq 2^3\left(\sum_{i=1}^n \mid x_i - \theta \mid^3 + n \mid \theta - \overline{x} \mid^3\right). \tag{4.5}$$

Thus we get in the right hand side of the inequality (4.4) that

$$\frac{c}{\widehat{\sigma}^3}\frac{1}{n^{3/2}}\sum_{i=1}^n \mid x_i - \overline{x} \mid^3 \leq \frac{c2^3}{\widehat{\sigma}^3}\left(\frac{1}{n^{3/2}}\sum_{i=1}^n \mid x_i - \theta \mid^3 + \frac{1}{n^{1/2}} \mid \theta - \overline{x} \mid^3\right) \tag{4.6}$$

Then, we recall that we obtained this by fixing the outcomes $X_1 = x_1, \ldots, X_n = x_n$. When we consider the upper bound (4.6) as a random variable, we get an inequality between two stochastic variables as

$$C_n \leq \frac{c2^3}{\widehat{\sigma}^3}\left(\frac{1}{n^{3/2}}\sum_{i=1}^{n} \mid X_i - \theta \mid^3 + \frac{1}{n^{1/2}} \mid \theta - \overline{X} \mid^3\right). \tag{4.7}$$

Here the researcher has needed to dig deep in the reservoirs of knowledge about probability theory. There one finds the **Zygmund -Marcinkiewicz strong law of large numbers**, as given in the next lemma.

Lemma 4.1 Let $Y_1, \ldots, Y_n, \ldots$ be I.I.D. random variables with distribution F. Suppose that for some $0 < \delta < 1$ it holds that $E\left[\mid Y \mid^{\delta}\right] < \infty$. Then it holds that

$$\frac{1}{n^{1/\delta}}\sum_{i=1}^{n} Y_i \stackrel{a.s}{\rightarrow} 0,$$

as $n \rightarrow \infty$.

■

A proof may be found on p. 122 in Y.S. Chow and H. Teicher:*Probability Theory. Independence. Interchageanbility. Martingales*. Springer-Verlag, 1978. Let us now apply this to

$$Y_i = \mid X_i - \theta \mid^3$$

and take $\delta = 2/3$. Thus

$$E\left[\mid Y \mid^{\delta}\right] = E\left[\mid X_i - \theta \mid^2\right] = \sigma^2 < \infty.$$

Hence we get in the right hand side of (4.6), where $3/2 = 1/\delta$, that

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n} \mid X_i - \theta \mid^3 = \frac{1}{n^{1/\delta}}\sum_{i=1}^{n} Y_i \stackrel{a.s}{\rightarrow} 0$$

as $n \rightarrow \infty$. As has been shown in (4.2), $\widehat{\sigma}^2 \stackrel{a.s.}{\rightarrow} \sigma^2$. In addition, the law of large numbers entails that

$$\mid \theta - \overline{X} \mid \stackrel{a.s.}{\rightarrow} 0,$$

and therefore

$$\frac{1}{n^{1/2}} \mid \theta - \overline{X} \mid^3 \stackrel{a.s.}{\rightarrow} 0.$$

Hence we have proved that the two terms in the right hand side of (4.7) converge to zero almost surely. This completes the proof of $C_n \rightarrow 0$.

5 Bootstrapping, Failures of Bootstrap, Berry-Esseen

In summary, we have in other words established that the bootstrap hypothesis holds for

$$P_F\left(\frac{H_n}{\sigma} \leq \frac{x}{\sigma}\right) - P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq \frac{x}{\widehat{\sigma}}\right).$$

The case under study, $\overline{X}$ as estimator of $\theta = E[X]$, is as such of no great particular interest as an application of bootstrapping, which is designed for analysis of more complicated estimators.

However, in a way the proof above indicates that $P_{\widehat{F}_n}\left(\frac{H_n^*}{\widehat{\sigma}} \leq \frac{x}{\widehat{\sigma}}\right)$ is a better approximation of $P_F\left(\frac{H_n}{\sigma} \leq \frac{x}{\sigma}\right)$ than $\Phi(x)$.

Babu & Rao (1993) state that bootsrap may fail to be consistent for

- extreme value statistics.
- when $E[X_1^2] = +\infty$, bootstrap distribution does not converge to any probability distribution.

For more on the Berry-Esseen bound one can study
`http://en.wikipedia.org/wiki/Berry−Esseen_theorem`
and
`http://sv.wikipedia.org/wiki/Carl−Gustav_Esseen`

6 Additional Sources

- G.J. Babu & C.R. Rao: Bootstrap Methodology. in C.R. Rao (ed.) *Handbook of Statistics*, Vol. 9,1993. pp. 627−659.
- A. DasGupta: *Asymptotic Theory of Statistics and Probability.* Springer 2008.
- C-G. Esseen: Fourier Analysis of Distribution Functions. A Mathematical Study of Laplace-Gaussian Law. *Inaugural Dissertation, October 14th, 1944*, Almqvist & Wiksells Boktryckeri, Uppsala, 1944.