Centrum voor Wiskunde en Informatica

REPORTRAPPORT

A note on bootstrapping the local time of the empirical process

R. Helmers

Department of Operations Research, Statistics, and System Theory

Report BS-R9422 July 1994

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications.
SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

A Note on Bootstrapping the Local Time of the Empirical Process

R. Helmers
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract

In this note we prove that Efron's bootstrap is asymptotically consistent in estimating the distribution of the local time of the empirical process, provided the underlying distribution of the observations is continuous. We employ the classical method of moments. It appears that our result is not easy to obtain from the general theory of bootstrapping (functionals of) empirical processes.

RÉSUMÉ

Dans cette note nous démontrons que le bootstrap d'Efron est convergent pour estimer la fonction de répartition du temps local du processes empirique, pourvue que la fonction de répartition des observations est continue. Nous utilisons la méthode des moments classique. Il appraraî̂t que notre résultat est difficile à partir de la théorie générale du bootstrap pour des (fonctionnels de) processus empiriques.

AMS Subject Classification (1991): 62G09, 60J55, 62G30, 62G20.
Keywords & Phrases: bootstrapping, local time, empirical processes, Brownian bridge, Raleigh distribution

1. INTRODUCTION AND MAIN RESULT

Let $X_1, X_2, \ldots$ be independent random variables, defined on a single probability space $(\Omega, \mathcal{A}, P)$, with common distribution function (*df*)F on the real line. Let F_n denote the empirical *df* based on $X_1, \ldots, X_n$; i.e. $F_n(x) = n^{-1}\Sigma_{i=1}^n I(X_i \leq x), -\infty < x < \infty$. Conditionally given $X_1, \ldots, X_n$, let $X_1^*, \ldots, X_n^*$ denote a random sample of size n, drawn with replacement from F_n. Let F_n^* denote the empirical *df* based on $X_1^*, \ldots, X_n^*$. Define the empirical process

$$U_n(x) = n^{\frac{1}{2}}(F_n(x) - F(x)),\ -\infty < x < \infty \tag{1.1}$$

and the bootstrapped empirical process

$$U_n^*(x) = n^{\frac{1}{2}}(F_n^*(x) - F_n(x)),\ -\infty < x < \infty \tag{1.2}$$

It is well-known that $U_n \underset{d}{\rightarrow} B(F)$ (cf. BILLINGSLEY (1968)) in the space (D, d), and, in addition, that $U_n^* \underset{d}{\rightarrow} B(F)$, with P-probability 1 (i.e., for almost all sequences $X_1, X_2, \ldots$) (cf. BICKEL & FREEDMAN (1981)). Here B denotes the Brownian bridge process, and $\underset{d}{\rightarrow}$indicates convergence in distribution. A beautiful and farreaching extension of this result was obtained by GINÉ & ZINN (1990).

The local time of U_n at zero up to 'time' $x_F = \sup\{x : F(x) < 1\}$ is given by

$$L_{x_F}^0(U_n) = n^{-\frac{1}{2}} \sum_{x \leq x_F} I_{\{o\}}(U_n(x)) \tag{1.3}$$

Report BS-R9422
ISSN 0924-0659
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Note that, if F is strictly increasing, then $L^0_{x_F}(U_n)$ is nothing but $n^{-\frac{1}{2}}$ times the number of zero-crossings of the U_n-process. If $U_n(x)$ happens to be zero for all x, which belong to a 'flat part' of F, then we simply count this as a single zero crossing. It is well-known that $L^0_{x_F}(U_n)$ is distributionfree when F is continuous. In this case, we may as well take F equal to the uniform *df* on (0, 1) and write $L^0_1(U_n)$ for the local time of the uniform empirical process.

It is already known for more than 30 years (cf. DWASS (1961)) that,

$$\lim_{n\to\infty} P(L^0_1(U_n) \le x) = 1 - e^{-\frac{1}{2}x^2},\ x \ge 0 \tag{1.4}$$

i.e. the local time of the uniform empirical process possesses asymptotically ($n \to \infty$) a Raleigh distribution. DWASS (1961) proved (1.4) by showing that all the moments of $L^0_1(U_n)$ converges to those of the Raleigh distribution. Because the Raleigh distribution is determined by its moments, this of course yields (1.4). At this point one should note that the limiting *df* in (1.4) can be identified with the *df* of the local time $L^0_1(B)$ of a Brownian bridge process B. In fact RÉVÉSZ (1982) has shown that, with P-probability 1, $L^0_1(B)$ is properly defined by:

$$L^0_1(B) = \lim_{\epsilon\to 0} \frac{\lambda\{s : 0 < s < 1, |B(s)| \le \epsilon\}}{2\epsilon} \tag{1.5}$$

where λ denotes Lebesgue measure on $(0,1)$. In addition, RÉVÉSZ (1982) also showed that, on a suitable probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$, there exists a sequence of Brownian bridges $\{\tilde{B}_n\}_{n\ge 1}$, and a sequence of uniform empirical processes $\{\tilde{U}_n\}_{n\ge 1}$, such that, for any $\epsilon > 0$,

$$|L^0_1(\tilde{U}_n) - L^0_1(\tilde{B}_n)| = \mathcal{O}(n^{-\frac{1}{4}+\epsilon}) \tag{1.6}$$

as $n \to \infty$, a.s. $[\tilde{P}]$.
As a simple consequence of all this the assertion (1.4) can now be replaced by

$$L^0_1(U_n) \underset{d}{\to} L^0_1(B) \tag{1.7}$$

This fact was also recognized by KHOSHNEVISAN (1992), who gave 'process versions' of (1.6) and (1.7). At the same time KHOSHNEVISAN (1992) was able to sharpen the a.s. order bound $\mathcal{O}(n^{-\frac{1}{4}+\epsilon})$ of RÉVÉSZ (cf. (1.6)) slightly to $o(n^{-\frac{1}{4}} \log n^{\frac{3}{4}+\epsilon})$, for any $\epsilon > 0$.

The aim of this note is to investigate whether $L^0_1(U_n)$ can be bootstrapped. I.e. we want to know whether $L^0_1(U^*_n) \underset{d}{\to} L^0_1(B)$, with P-probability 1, as $n \to \infty$? Our interest in this question comes from the fact that $L^0_1(\cdot)$ viewed as function of the uniform empirical process and of the Brownian bridge process is not at all continuous. So it appears that our problem cannot be settled easily by an application of an 'extended continuous mapping' theorem. Neither finding a suitable 'strong approximation' argument, like the one (cf. (1.6)) leading to (1.7), seems to be an easy task to perform. However, a direct approach - quite in the spirit of Dwass's 1961- paper - turns out to be feasible for the problem at hand. We shall in fact apply the classical method of moments to prove the following result:

THEOREM 1. *As $n \to \infty$, we have with P-probability 1, that*

$$L^0_1(U^*_n) \underset{d}{\to} L^0_1(B) \tag{1.8}$$

We refer to SHORACK & WELLNER (1986), p. 398-400 for a short introduction to local time for empirical processes and to WELLNER (1992) for an excellent recent survey on bootstrapping empirical processes.

2. PROOF OF THEOREM 1. To establish (1.8) we employ the method of moments; i.e. we shall prove that, with P-probability 1,

$$\lim_{n\to\infty} E^*(L_1^0(U_n^*))^r = 2^{\frac{r}{2}}\Gamma(\frac{r}{2}+1) \tag{2.1}$$

for any fixed positive integer r. Here E^* of course denotes conditional expectation w.r.t. the bootstrap resampling. Note that the expression on the r.h.s. of (2.1) is precisely equal to the rth moment of $L_1^0(B)$. The following simple identity (cf. TITCHMARSH (1960), p. 63, example 18) will facilitate our computations: If $\alpha > 0, \beta > 0$, then

$$\int_x^y (y-z)^{\alpha-1}(z-x)^{\beta-1}dz = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}(y-x)^{\alpha+\beta-1} \tag{2.2}$$

Let $X_{1:n} < \ldots < X_{n:n}$ denote the order statistics corresponding to $X_1, \ldots, X_n$. Because F is the uniform *df* on $(0,1)$ there are - with $P-$ probability 1 - no ties among the $X_{k:n}$ and $F_n(X_{k:n}) = \frac{k}{n}$, for $k = 1, \ldots, n$. Set $X_{0:n} = 0$.

First we verify (2.1) for $r = 1, 2$ and 3. Subsequently we shall treat (2.1) for general r. To begin with the case $r = 1$ we note that, with $P-$ probability 1,

$$E^* L_1^0(U_n^*) = n^{-\frac{1}{2}} \sum_{k=0}^{n} E^* I(F_n^*(X_{k:n}) = \frac{k}{n})$$

$$= n^{-\frac{1}{2}} \sum_{k=0}^{n} P^*(F_n^*(X_{k:n}) = \frac{k}{n})$$

$$= n^{-\frac{1}{2}} \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k (1-\frac{k}{n})^{n-k}$$

$$\sim (2\pi)^{-\frac{1}{2}} n^{-1} \sum_{k=1}^{n-1} \left[\frac{k}{n}(1-\frac{k}{n})\right]^{-\frac{1}{2}}$$

$$\sim (2\pi)^{-\frac{1}{2}} \int_{n^{-1}}^{1-n^{-1}} (x(1-x))^{-\frac{1}{2}} dx \sim \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \tag{2.3}$$

Here we have used Stirling's formula and the well-known fact that $\int_0^1 (x(1-x))^{-\frac{1}{2}}dx = \pi$ (a special case of (2.2)).

Next we compute the second moment of $L_1^0(U_n^*)$ in a similar fashion:

$$E^*(L_1^0(U_n^*))^2 =$$

$$= n^{-1} \sum_{k=0}^{n} \sum_{l=0}^{n} E^*(I(F_n^*(X_{k:n}) = \frac{k}{n}) \wedge I(F_n^*(X_{l:n}) = \frac{l}{n}))$$

$$= n^{-1} \sum_{k=0}^{n} \sum_{l=0}^{n} P^*(F_n^*(X_{k:n}) = \frac{k}{n} \wedge F_n^*(X_{l:n}) = \frac{l}{n})$$

$$\sim 2n^{-1} \sum_{k=1}^{n} \sum_{l=k+1}^{n} \frac{n!}{k!(l-k)!(n-k)!} (\frac{k}{n})^k (\frac{l-k}{n})^{l-k} (1-\frac{l}{n})^{n-l}$$

$$\sim \pi^{-1} n^{-2} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n-1} \left[\frac{k}{n} \frac{(l-k)}{n} (1-\frac{l}{n}) \right]^{-\frac{1}{2}}$$

$$\sim \pi^{-1} \int_{n^{-1}}^{1-n^{-1}} x^{-\frac{1}{2}} \int_{x}^{1-n^{-1}} (y-x)^{-\frac{1}{2}} (1-y)^{-\frac{1}{2}} dy dx$$

$$= \int_{n^{-1}}^{1-n^{-1}} x^{-\frac{1}{2}} dx \sim 2 \tag{2.4}$$

where we have applied identity (2.1), with $\alpha = \beta = \frac{1}{2}$, to find that $\int_x^{1-n^{-1}} (y-x)^{-\frac{1}{2}}(1-y)^{-\frac{1}{2}} dy \sim \pi$, as $n \to \infty$.

The computation of the higher moments of $L_1^0(U_n^*)$ is somewhat more involved. Let us first look at the case $r = 3$:

$$E^*(L_1^0(U_n^*))^3 =$$

$$= n^{-\frac{3}{2}} \sum_{k=0}^{n} \sum_{l=0}^{n} \sum_{m=0}^{n} E^*(I(F_n^*(X_{k:n}) = \frac{k}{n}) \wedge I(F_n^*(X_{l:n}) = \frac{l}{n}) \wedge I(F_n^*(X_{m:n}) = \frac{m}{n}))$$

$$= n^{-\frac{3}{2}} \sum_{k=0}^{n} \sum_{l=0}^{n} \sum_{m=0}^{n} P^*(F_n^*(X_{k:n}) = \frac{k}{n} \wedge F_n^*(X_{l:n}) = \frac{l}{n} \wedge F_n^*(X_{m:n}) = \frac{m}{n})$$

$$\sim 6n^{-\frac{3}{2}} \sum_{k=1}^{n} \sum_{l=k+1}^{n} \sum_{m=l+1}^{n} \frac{n!}{k!(l-k)!(m-l)!(n-m)!} (\frac{k}{n})^k (\frac{l-k}{n})^{l-k} (\frac{m-l}{n})^{m-l} (1-\frac{m}{n})^{n-m}$$

$$\sim 3 \cdot 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} n^{-3} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n-1} \sum_{m=l+1}^{n-1} [\frac{k}{n} (\frac{l-k}{n}) (\frac{m-l}{n}) (1-\frac{m}{n})]^{-\frac{1}{2}} \sim$$

$$\sim 3 \cdot 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} \int_{n^{-1}}^{1-n^{-1}} \int_{x}^{1-n^{-1}} \int_{y}^{1-n^{-1}} x^{-\frac{1}{2}} (y-x)^{-\frac{1}{2}} (z-y)^{-\frac{1}{2}} (1-z)^{-\frac{1}{2}} dx dy dz$$

$$= 3 \cdot 2^{-\frac{1}{2}} \pi^{-\frac{1}{2}} \int_{n^{-1}}^{1-n^{-1}} x^{-\frac{1}{2}} \int_{x}^{1-n^{-1}} (y-x)^{-\frac{1}{2}} dy dx =$$

$$= 6 \cdot 2^{-\frac{1}{2}}\pi^{-\frac{1}{2}} \int_{n^{-1}}^{1-n^{-1}} x^{-\frac{1}{2}}(1-x)^{\frac{1}{2}}dx \sim 3 \cdot 2^{-\frac{1}{2}}\pi^{\frac{1}{2}} \tag{2.5}$$

where we have applied (2.1) once more, now with $\alpha = \frac{1}{2}, \beta = \frac{3}{2}$. Note that the asymptotic values $(\frac{\pi}{2})^{\frac{1}{2}}, 2$ and $3 \cdot 2^{-\frac{1}{2}}\pi^{\frac{1}{2}}$ obtained for $\lim_{n\to\infty} E^*(L_1^0(U_n^*))^r$ for $r = 1, 2, 3$ respectively, indeed coincide with the r.h.s. of (2.1).

It remains to check (2.1) for $r \geq 4$. To do this, note that - similarly as in the case $r = 3$ - one easily deduces that

$$E^*(L_1^0,(U_n^*))^r \sim$$

$$\sim r!\,(2\pi)^{-\frac{r}{2}} \int_{n^{-1}}^{1-n^{-1}} \int_{x_1}^{1-n^{-1}} \cdots \int_{x_{r-2}}^{1-n^{-1}} \int_{x_{r-1}}^{1-n^{-1}} x_1^{-\frac{1}{2}}(x_2-x_1)^{-\frac{1}{2}} \ldots (x_r - x_{r-1})^{-\frac{1}{2}}(1-x_r)^{-\frac{1}{2}} \prod_{i=1}^{r} dx_i$$

$$= r!\,2^{-\frac{r}{2}}\pi^{-\frac{r}{2}+1} \int_{n^{-1}}^{1-n^{-1}} \int_{x_1}^{1-n^{-1}} \cdots \int_{x_{r-3}}^{1-n^{-1}} \int_{x_{r-2}}^{1-n^{-1}} x_1^{-\frac{1}{2}}(x_2-x_1)^{-\frac{1}{2}} \ldots (x_{r-1} - x_{r-2})^{-\frac{1}{2}} \prod_{i=1}^{r-1} dx_i$$

$$\sim r!\,(2\pi)^{-\frac{r}{2}+1} \int_{n^{-1}}^{1-n^{-1}} \int_{x_1}^{1-n^{-1}} \cdots \int_{x_{r-3}}^{1-n^{-1}} x_1^{-\frac{1}{2}}(x_2-x_1)^{-\frac{1}{2}} \ldots (x_{r-2} - x_{r-3})^{-\frac{1}{2}}(1-x_{r-2})^{\frac{1}{2}} \prod_{i=1}^{r-2} dx_i$$

$$\sim r!\,2^{-\frac{r}{2}}\pi^{-\frac{r}{2}+2} \int_{n^{-1}}^{1-n^{-1}} \int_{x_1}^{1-n^{-1}} \cdots \int_{x_{r-4}}^{1-n^{-1}} x_1^{-\frac{1}{2}}(x_2-x_1)^{-\frac{1}{2}} \ldots (x_{r-3} - x_{r-4})^{-\frac{1}{2}}(1-x_{r-3}) \prod_{i=1}^{r-3} dx_i$$

$$\sim r!\,2^{-\frac{r}{2}}\pi^{-\frac{r}{2}+2} \frac{\Gamma(\frac{1}{2})\Gamma(2)}{\Gamma(\frac{5}{2})} \int_{n^{-1}}^{1-n^{-1}} \cdots \int_{x_{r-5}}^{1-n^{-1}} x_1^{-\frac{1}{2}}(x_2-x_1)^{-\frac{1}{2}} \ldots (1-x_{r-4})^{\frac{3}{2}} \prod_{i=1}^{r-4} dx_i$$

$$\sim r!\,2^{-\frac{r}{2}}\pi^{-\frac{r}{2}}\pi^{-\frac{r}{2}+2} \frac{\Gamma(\frac{1}{2})\Gamma(2)}{\Gamma(\frac{5}{2})} \frac{\Gamma(\frac{1}{2})\Gamma(\frac{5}{2})}{\Gamma(3)} \cdots \frac{\Gamma(\frac{1}{2})\Gamma(\frac{r}{2})}{\Gamma(\frac{r+1}{2})} =$$

$$= \frac{r!2^{-\frac{r}{2}}\pi^{\frac{1}{2}}}{\Gamma(\frac{r+1}{2})} = 2^{\frac{r}{2}}\Gamma(\frac{r}{2}+1)$$

where we have used the well-known 'duplication formula' for the Γ-function (cf., e.g., TITCHMARSH (1960), p. 57) in the last line. □

3. CONCLUDING REMARKS

In this note we have shown that Efron's bootstrap is asymptotically consistent in estimating the distribution function of the local time of the empirical process, provided the underlying distribution of the observations is continuous.

What happens if we take the resample size m different from n? The answer is given in Remark 3.1.

REMARK 3.1. Theorem 1 remains valid if we allow the bootstrap resample size m to be different from n. Then, as $\min(m, n) \to \infty$, we have with P-probability 1, that

$$L_1^0(U_{m,n}^*) \xrightarrow[d]{} L_1^0(B)$$

where $U_{m,n}^*$ is U_n^* (cf. (1.2)) with F_n^* replaced by F_m^* and $L_1^0(U_{m,n}^*) = m^{-\frac{1}{2}} \sum_{x \le 1} I_{\{0\}}(U_{m,n}^*(x))$.

A well-known feature of Efron's bootstrap is that the number of original observations left out in a bootstrap sample of the same size can be quite large, namely $\sim ne^{-1}$ on the average. i.e. there will be many equal bootstrap observations in the resampling, while the original data set contains no ties at all, with P-probability 1. Perhaps somewhat surprisingly, this doet not affect the asymptotic behaviour of bootstrapped 'local time of the empirical process'. However, if one asks the question whether the probability that at 'time point' $X_{1:n}$ a zero-crossing occurs, can also be naively bootstrapped, one easily obtains the following negative result:

REMARK 3.2. Suppose F is continuous. Since $F_n(X_{1:n}) = \frac{1}{n}$, with P-probability 1, we have that

$$P(I_{\{0\}}(U_n(X_{1:n})) = 1) = 0$$

and, with P-probability 1,

$$P^*(I_{\{0\}}(U_n^*(X_{1:n})) = 1) \to e^{-1}, \text{ as } n \to \infty.$$

I.e. the probability of a zero-crossing of the U_n-process at 'time' $X_{1:n}$ is zero, while - on the other hand - the probability of a zero-crossing of the U_n^*- process at $X_{1:n}$ approaches e^{-1}, as $n \to \infty$. A similar phenomenon occurs at 'time' $X_{k:n}$, for any fixed k.

REFERENCES

1. P.J. BICKEL & D. FREEDMAN (1981), Some asymptotic theory for the bootstrap, Ann. Statist. **9**, 1196-1217.
2. P. BILLINGSLEY (1968), *Convergence of Probability Measures*, Wiley, New York.
3. M. DWASS (1961), Random crossings of cumulative distribution functions, Pacific J. Math. **2**, 127-134.
4. E. GINÉ & J. ZINN (1990), Bootstrapping general empirical measures, Ann. Probab. **18**, 851-869.
5. D. KHOSHNEVISAN (1992), Level crossings of the empirical process, Stoch. Proc. Appl. **43**, 331-343.
6. P. RÉVÉSZ (1982), On the local time of Brownian bridge, in: Trans. 9th Praque Conf. on Inform. Theory, Statist. Dec. Funct., and Random Processes, p. 67-76.
7. G.R. SHORACK & J.A. WELLNER (1986), *Empirical processes with applications to statistics*, Wiley, New York.
8. E.C. TITCHMARSH (1960), *The Theory of Functions*, Oxford University Press, London.
9. J.A. WELLNER (1992), Bootstrap Limit Theorems: a partial survey, in: *Nonparametric Statistics and Related Topics*, 313-329 (ed. A.K. Md. E. Saleh), Elsevier Science Publishers, Amsterdam.