Kybernetika

Marie Hušková

Rank statistics approach in generalized bootstrap

Kybernetika, Vol. 31 (1995), No. 3, 293--296

Persistent URL: http://dml.cz/dmlcz/124719

Terms of use:

© Institute of Information Theory and Automation AS CR, 1995

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

KYBERNETIKA — VOLUME 31 (1995), NUMBER 3, PAGES 293–296

RANK STATISTICS APPROACH IN GENERALIZED BOOTSTRAP

MARIE HUŠKOVÁ

Mason and Newton [8] introduced the generalized (exchangeably weighted) bootstrap procedure and proved its consistency for the empirical distribution function, the quantile function and the mean. Surprisingly, Hájek's [2] results on the asymptotic behavior of the simple linear rank statistics play the crucial role in the proof. Mason–Newton's work inspired several authors who applied this method successfully to other types of statistics.

The purpose of the paper is to point out how powerful tool is Hájek's rank statistics methodology in the proof of a.s. consistency for various resampling schemes and to survey the existing results.

MAIN RESULT

Let $X_1, \ldots, X_n$ be i.i.d. random variables defined on a probability space $(\Omega, \mathcal{A}, P)$. Denote by F the common distribution function(d.f.) and by F_N its empirical counterpart. Let $\theta(F)$ be a parameter of interest and let $\theta_N = \theta(F_N)$ be its estimator.

Efron [3,4] introduced a method that enable to estimate the distribution of θ_N, its bias, its variability etc. under quite mild assumptions – the method is known as the *Efron bootstrap.*

The desired estimators are based on the random sample of size N from the empirical distribution function F_N. The sample is called the bootstrap sample and the corresponding empirical d.f. is the bootstrap empirical d.f. (denoted by F_N^*) that can formally be expressed as follows:

$$F_N^*(x) = \sum_{i=1}^{N} I\{X_i \le x\} \frac{1}{N} M_i,$$

where $I\{A\}$ is the indicator of the set A and $(M_1, \ldots, M_N)$ has the multinomial distribution $\left(n; \frac{1}{n}, \ldots, \frac{1}{n}\right)$.

Mason and Newton [8] proposed to replace $(M_1, \ldots, M_N)$ by a vector of random weights $(w_{N1}, \ldots, w_{NN})$ defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P})$ independent of $X_1, \ldots, X_N$ and fulfilling

$$w_{N1}, \ldots, w_{NN} \text{ are exchangeable,} \tag{W.1}$$

$$w_{Ni} \geq 0, \quad i = 1, \dots, N, \sum_{i=1}^{N} w_{Ni} = 1, \tag{W.2}$$

$$N \sum_{i=1}^{N} \left(w_{Ni} - \frac{1}{N} \right)^2 \xrightarrow{\tilde{P}} c^2, \quad N \to \infty \quad \text{for some } c > 0, \tag{W.3}$$

$$N \max_{1 \leq i \leq N} \left(w_{Ni} - \frac{1}{N} \right)^2 \xrightarrow{\tilde{P}} 0, \quad N \to \infty, \tag{W.4}$$

The generalized bootstrapped empirical d. f. has then the form:

$$F_{Nw_N}(x) = \sum_{i=1}^{N} w_{Ni} I\{X_i \leq x\}.$$

Typical choices of weights include, among others, multinomial weights (resulting in classical Efron's bootstrap and Dirichlet weights (Bayesian bootstrap).

Since the weights $(w_{N1}, \dots, w_{NN})$ are exchangeable random variables, we have conditionally for given $(X_1, \dots, X_N)$:

$$\left\{ \sum_{i=1}^{N} w_{Ni} I\{X_i \leq x\}, \, x \in R \right\} \stackrel{\mathcal{D}}{=} \left\{ \sum_{i=1}^{N} w_{NR_i} I\{X_i \leq x\}, \, x \in R \right\},$$

where $(R_1, \dots, R_N)$ is a random permutation of $\{1, \dots, N\}$. Clearly, given $X_1, \dots, X_N$ and $w_{N1}, \dots, w_{NN}$ the statistic

$$\sum_{i=1}^{N} w_{NR_i} I\{X_i \leq x\},$$

is the simple linear rank statistic for arbitrary fixed x. Then the problem to verify whether the bootstrap "works" reduces to the investigation of the conditional limit distribution of the mentioned simple linear rank statistic.

Hájek ([2], see also Jurečková [7] – Theorem 2.1) formulated the necessary and sufficient conditions for asymptotic normality of the simple linear rank statistics. In our case these conditions ((3 – 5) and (7) in Jurečková [7]) have the form:

$$\max_{1 \leq i \leq N} \{W_{Ni}^2\} \xrightarrow{\tilde{P}} 0 \quad \text{as } N \to \infty \tag{A.1}$$

$$\max_{1 \leq i \leq N} \{Y_{Ni}^2\} \longrightarrow 0 \quad \text{as } N \to \infty \quad \text{a. s. } [P] \tag{A.2}$$

and for every $\varepsilon > 0$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} W_{Ni}^2 Y_{Ni}^2 I\{N W_{Ni}^2 Y_{Ni}^2 > \varepsilon\} \xrightarrow{\tilde{P}} 0 \quad \text{as } N \to \infty \quad \text{a. s. } [P], \tag{A.3}$$

where

$$Y_{Ni} = \frac{I\{X_i \le x\} - \frac{1}{N}\sum_{j=1}^{N} I\{X_j \le x\}}{\sum_{j=1}^{N}\left(I\{X_j \le x\} - \frac{1}{N}\sum_{v=1}^{N} I\{X_v \le x\}\right)^2}, \quad 1 \le i \le N,$$

$$W_{Ni} = \frac{(w_{Ni} - \overline{w}_N)}{\sum_{j=1}^{N}(w_{Nj} - \overline{w}_N)^2}, \quad 1 \le i \le N,$$

$$\overline{w}_N = \frac{1}{N}\sum_{v=1}^{N} w_{Nv}.$$

It is easy to realize that if $0 < P(X_i < x) < 1$ and the assumptions on the weights (W.1) – (W.4) are fulfilled then (A.1) – (A.3) hold true, Hájek's [2] theorem can be applied and as a consequence we receive that the bootstrap "works", i.e., as $N \to \infty$,

$$\sup_y \left| P\left(\sqrt{N}\,(F_{N,W}(x) - F_N(x)) \le y | X_1, \ldots, X_n\right) - \right.$$
$$\left. -P\left(\sqrt{N}\,(F_N(x) - F(x)) \le y\right)\right| \longrightarrow 0 \quad \text{a.s. } [P]$$

for every x fixed. Mason and Newton [8] showed that analogous assertions hold true for the respective processes $\{F_{N,w}(x),\ x \in R\}$ and $\{F_N(x),\ x \in R\}$, for the quantile processes and the mean $(\theta(F) = EX)$.

Hušková and Janssen [5, 6] extended these results to the $\mathcal{U}$-statistics with both nondegenerate and generate kernels. In this situation a new type of rank statistics with the $\mathcal{U}$-structure appeared and the assertions on their limit distributions were new even within the rank statistics theory.

Aerts and Janssen [1] studied the generalized bootstrap for the $\mathcal{U}$-quantiles.

Finally, the most general results were proved by Praestgaard and Wellner [9] who studied the generalized bootstrapping of the general function – indexed empirical processes.

All these results say that the bootstrap "works" in quite general bootstrap scheme or, in other words, the (conditional) d.f. of the generalized bootstrap version of the studied statistic provides an a.s. consistent estimator for the unknown d.f. of this statistic.

In the proofs of these results Hájek's [2] theorem played extremely important role and as a side results its sufficient part was extended to various classes of rank statistics.

ACKNOWLEDGEMENT

The paper was partially supported by the University grant GAUK 365.

(Received November 1, 1994.)

REFERENCES

[1] M. Aerts and P. Janssen: Weighted bootstrapping for $\mathcal{U}$-quantiles. Submitted.

[2] J. Hájek: Some extensions of the Wald–Wolfowitz–Noether theorem. Ann. Math. Statist. *32* (1961), 506–523.

[3] B. Efron: Bootstrap methods: Another look at the jackknife. Ann. Statist. *7* (1979), 1–26.

[4] B. Efron: The Jackknife, the Bootstrap, and Other Resampling Plans. SIAM, Philadelphia 1982.

[5] M. Hušková and P. Janssen: Generalized bootstrap for studentized $\mathcal{U}$-statistics: a rank statistic approach. Statist. Probab. Lett. *16* (1993), 225–233.

[6] M. Hušková and P. Janssen Consistency of the generalized bootstrap for degenerate $\mathcal{U}$-statistics. Ann. Statist. *21* (1993), 1811–1823.

[7] J. Jurečková: Jaroslav Hájek and asymptotic theory of rank tests. Kybernetika *31* (1995), 239–250.

[8] D. M. Mason and M. A. Newton: A rank statistic approach to the consistency of a general bootstrap. Ann. Statist. *20* (1992), 1611–1624.

[9] J. Praestgaard and J. Wellner: Exchangeably weighted bootstraps of the general empirical process. Ann. Probab. *21* (1993), 2053–2086.

Doc. RNDr. Marie Hušková, CSc., Matematicko-fyzikální fakulta Univerzity Karlovy (Faculty of Mathematics and Physics – Charles University), Sokolovská 83, 186 00 Praha 8. Czech Republic.