On the Kolmogorov-Smirnov statistic of certain dependent random variables

Einmahl, J.H.J.

Published in:
Publications de l'Institut de Statistique de l'Universite de Paris

Publication date:
1985

Citation for published version (APA):
Einmahl, J. H. J. (1985). On the Kolmogorov-Smirnov statistic of certain dependent random variables. *Publications de l'Institut de Statistique de l'Universite de Paris*, *30*(3), 53-57.

Pub. Inst. Stat. Univ.
XXX, fasc. 3, 1985, 53 à 57

ON THE KOLMOGOROV-SMIRNOV STATISTIC OF CERTAIN DEPENDENT RANDOM VARIABLES

By John H.J. Einmahl
Catholic University Nijmegen

SUMMARY

In this note a simple derivation of a Dvoretzky-Kiefer-Wolfowitz type exponential bound is given for the Kolmogorov-Smirnov statistic of a certain class of dependent random variables. Applications to spacings and Studentized observations are discussed.

AMS 1980 subject classifications: Primary 62H10; Secondary 60E15.

Key words and phrases: Kolmogorov-Smirnov statistic, exponential bound, k-step uniform spacings, Studentized observations.

INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be independent random variables with common distribution function F and let $\hat{F}_n$ be the empirical distribution function based on these random variables. The empirical process is defined by

$$H_n(t) = n^{\frac{1}{2}}(\hat{F}_n(t) - F(t)), \quad t \in \mathbb{R}. \tag{1}$$

Dvoretzky, Kiefer and Wolfowitz (1956) showed the existence of a universal constant C_1 such that for all $n \in \mathbb{N}$ and all $\lambda \geq 0$

$$P(\sup_{t \in \mathbb{R}} |H_n(t)| \geq \lambda) \leq C_1 \exp(-2\lambda^2). \tag{2}$$

In this note a similar result will be proved for a more general class of random variables which contains in particular uniform k-spacings and Studentized observations from a normal distribution.

THE MAIN THEOREM

In order to formulate our main theorem let us introduce some notation. Let us start with the random variables X_i of the introduction and define two new random variables Z_{1n} and Z_{2n} ($Z_{2n} > 0$ a.s.) by

$$Z_{1n} = f_1(X_1, \ldots, X_n) \textit{ and } Z_{2n} = f_2(X_1, \ldots, X_n), \tag{3}$$

where f_1 and f_2 are $(\mathcal{B}^n, \mathcal{B})$-measurable functions.

Define

$$Y_i = Z_{2n}^{-1}(X_i - Z_{1n}), \textit{ for } i = 1, 2, \ldots, n, \tag{4}$$

and for $a_{1n} \in \mathbb{R}$, $a_{2n} \in \mathbb{R}^+$

$$G_n(t) = F(ta_{2n} + a_{1n}), \quad t \in \mathbb{R}. \tag{5}$$

A useful choice for a_{1n} and a_{2n} is

$$a_{\ell n} = EZ_{\ell n}, \; \ell = 1, 2. \tag{6}$$

Let the empirical distribution function of $Y_1,\ldots,Y_n$ be denoted by $\hat{G}_n$ and define the corresponding empirical process by

$$K_n(t) = n^{\frac{1}{2}}(\hat{G}_n(t)-G_n(t)), \quad t \in \mathbb{R}. \tag{7}$$

Let

$$\begin{aligned} p_{1n} &= P(Z_{1n}\le a_{1n},\ Z_{2n}\le a_{2n}),\ p_{2n} = P(Z_{1n}\le a_{1n},\ Z_{2n}\ge a_{2n}) \\ p_{3n} &= P(Z_{1n}\ge a_{1n},\ Z_{2n}\le a_{2n}),\ p_{4n} = P(Z_{1n}\ge a_{1n},\ Z_{2n}\ge a_{2n}) \end{aligned} \tag{8}$$

THEOREM. *If* $(Y_1,\ldots,Y_n)$ *and* (Z_{1n},Z_{2n}) *are stochastically independent and* $\sup\{\sum_{j=1}^{4} p_{jn}^{-1} \mid n \in \mathbb{N}\} = C_2 < \infty$, *then there exists a universal constant* $C = C_1 \cdot C_2$ *such that for every* $n \in \mathbb{N}$ *and* $\lambda \ge 0$

$$P(\sup_{t\in\mathbb{R}} |K_n(t)| \ge \lambda) \le C \exp(-2\lambda^2). \tag{9}$$

PROOF. Define

$$\begin{aligned} V_n^1 &= \sup_{t\in\mathbb{R}\setminus\mathbb{R}^-} K_n(t), \quad V_n^2 = \sup_{t\in\mathbb{R}^-} K_n(t), \\ V_n^3 &= \sup_{t\in\mathbb{R}^-} (-K_n(t)), \quad V_n^4 = \sup_{t\in\mathbb{R}\setminus\mathbb{R}^-} (-K_n(t)). \end{aligned} \tag{10}$$

It's obvious that

$$P(\sup_{t\in\mathbb{R}} |K_n(t)| \ge \lambda) \le \sum_{j=1}^{4} P(V_n^j \ge \lambda). \tag{11}$$

Let us now observe that

$$\begin{aligned} &P(V_n^1 \ge \lambda) = \\ &= P(\sup_{t\in\mathbb{R}\setminus\mathbb{R}^-} n^{\frac{1}{2}}(n^{-1}\sum_{i=1}^{n} 1_{(-\infty,t]}\left(\frac{X_i-Z_{1n}}{Z_{2n}}\right)-G_n(t)) \ge \lambda \mid Z_{1n}\le a_{1n}, Z_{2n}\le a_{2n}) = \\ &= P(\sup_{t\in\mathbb{R}\setminus\mathbb{R}^-} (H_n(tZ_{2n}+Z_{1n})+n^{\frac{1}{2}}(F(tZ_{2n}+Z_{1n})-F(ta_{2n}+a_{1n}))) \ge \lambda \mid \\ &\qquad \mid (Z_{1n}\le a_{1n}, Z_{2n}\le a_{2n}) \le \\ &\le P(\sup_{t\in\mathbb{R}\setminus\mathbb{R}^-} H_n(tZ_{2n}+Z_{1n}) \ge \lambda \mid Z_{1n}\le a_{1n}, Z_{2n}\le a_{2n}) \le \\ &\le P(\sup_{s\in\mathbb{R}} |H_n(s)| \ge \lambda)/P(Z_{1n}\le a_{1n}, Z_{2n}\le a_{2n}) \le p_{1n}^{-1}\cdot C_1 \exp(-2\lambda^2). \end{aligned} \tag{12}$$

In the same way can be proved

(13) $$P(V_n^j \geq \lambda) \geq p_{jn}^{-1}.C_1 \exp(-2\lambda^2), \quad j = 2,3,4.$$

Combining (11)-(13) and the second condition in the Theorem yields

(14) $$P(\sup_{t\in\mathbb{R}} |K_n(t)| \geq \lambda) \leq \sum_{j=1}^{4} p_{jn}^{-1}.C_1 \exp(-2\lambda^2) \leq C \exp(-2\lambda^2).$$

Q.E.D.

APPLICATIONS

In the case of uniform spacings the Theorem can be applied, using the well-known fact that the vector of spacings is distributed as $(E_1/S_n,\ldots,E_n/S_n)$, where $S_n = \sum_{i=1}^{n} E_i$ and $E_1,\ldots,E_n$ are independent exponentially distributed random variables with any scale parameter $\lambda \in \mathbb{R}^+$ (see e.g. Pyke (1965), p.403). Also nonoverlapping k-step uniform spacings (see e.g. Del Pino (1979)) can be treated replacing the E_i's by independent Gamma (k,λ) random variables.

Another interesting example can be obtained by taking for the X_i's Normal (μ,σ^2) random variables (for any $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$), Z_{1n} the sample mean and Z_{2n} the square root of the sample variance.

Remark 1. In the cases above a_{1n} and a_{2n} are chosen according to (6).

Remark 2. From another point of view $\sup_{t\in\mathbb{R}} |K_n(t)|$ can be seen as the Kolmogorov-Smirnov statistic of the X_i's when the parameters are estimated.

Remark 3. That the conditions in the Theorem are fulfilled in all these cases is well-known, especially in all cases the independence condition can be shown in a nice way, using a result in Basu (1955, Theorem 2), which implies that a sufficient complete statistic and a statistic of which the distribution does not depend on the parameter(s) are stochastically independent.

ACKNOWLEDGEMENTS

I would like to thank Martien van Zuijlen for confronting me with the problem and Frits Ruymgaart for some helpful remarks.

REFERENCES

[1] BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā 15*, 377-380.

[2] DEL PINO, G.E. (1979). On the asymptotic distribution of k-spacings with applications to goodness-of-fit tests. *Ann. Statist. 7*, 1058-1065.

[3] DVORETZKY, A., KIEFER, J. & WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist. 27*, 642-669.

[4] PYKE, R. (1965). Spacings. *Journal Roy. Statist. Soc. Series B 27*, 395-436.

Reçu en Juin 1984

JOHN H.J. EINMAHL
Department of Mathematics
Cath. Un. Nijmegen
Toernooiveld
6525 ED Nijmegen
The Netherlands