INSTITUTE OF MATHEMATICS
OF THE POLISH ACADEMY OF SCIENCES

Sniadeckich 8, 00-956 Warsaw, P.O. Box 21, Poland, www.impan.pl, e-mail: im@impan.pl
tel. +48 22522 81 00, fax: +48 22629 39 97

Prof. dr habil. Tomasz Rychlik
Department of Mathematical Statistics

Report
of the member of the Dissertational Council on the thesis
Distributions of total number of upper and lower record values in nonstandard situations by Igor Vladimirovich Bel'kov
submitted for the granting procedure
of the Candidate degree in the Mathematical and Physical Sciences
in speciality 01.01.05: Probability Theory and Mathematical Statistics

The subject of the dissertation was a study of record values and record occurrence times in finite and infinite sequences of independent variables. The thesis is 76 pages long (I analyze only the English version). It contains 57 references, and it is based on 6 papers written by the candidate. Bel'kov is the only author of one of them, four papers were written in collaboration with the supervisor Professor Valery Nevzorov, and one has three authors: Bel'kov, Nevzorov, and Mohammad Ahsanullah. This last one was published in a foreign *Journal of Statistical Theory and Applications*, and the others were published in Russian journals and in a book collection of papers. I would encourage the candidate for submitting his further achievements to international journals which make them more accessible for a wide mathematical audience.

Chapter 1 was devoted to calculating lower and upper bounds on the total number of lower and upper records in finite sequences of independent random variables with two arbitrary and different continuous distributions. The results strongly depend on the sample size, numbers of observations with different distributions, and their order. Bel'kov derived sharp lower nad upper bounds only in the cases of three (with one outlier) and four observations (with one or two outliers) for arbitrary locations of outlying observations. For $n = 5$ random variables, he established some non-sharp general bounds, and determined sharp evaluations in a specific model with particularly constructed distributions: one arises when the density of the standard uniform distribution is partitioned into k pieces which are further spread over the interval $[0, 2]$, and the other is just the uniform distributions over the gaps in the support of the former one. Optimal partitions were determined numerically for various k. The same procedure was performed for $n = 6$ which gave radically different

09/2-02-484 от 18.09.2020

results. E.g., for $n = 6$ the transformed limiting distribution of the outliers proving the maximal number of records has atoms at the support ends unlike the one in the case $n = 5$. By the way, I wonder how it was possible to determine the limiting outlier distribution if the optimal one for each k was determined numerically.

Although I appreciate clever calculations in analysis of the above problems, I doubt if they contribute much to the theory of random records. They give solutions to very particular problems and do not allow getting more general conclusions. More apparent results were obtained under the assumption that the logarithms of distributions of consecutive observations are proportional (in so called F^α scheme).

Chapter 2 may be the most interesting part of the Bel'kov thesis for the mathematical community. The author focused on the record ranges there, i.e., the record values in the differences between the smallest and greatest observations. The first valuable result of the chapter is the claim that the record range values from the mixtures of exponential and negative exponential distributions (Laplace distributions in particular) have the distribution identical the upper records from the exponential populations, i.e., with the consecutive sums of i.i.d. exponential random variables. He also determined the expectation of the number of record ranges in sequences of finite sizes, which unlike the record range values depend on the mixing proportion. He further characterized the Laplace distribution by independence of two first record range spacings in the family of absolutely continuous symmetric distributions. Similar results were obtained for the geometric distribution. The increments of record ranges of an arbitrary mixtures of the geometric distributions and their negative reflections are independent identically geometrically distributed, whereas the uniform mixtures are characterized by independence of two first record range increments among all discrete symmetric distributions. The positive impression of Chapter 2 is somehow messed by numerous mistakes in its presentation. It looks as if the first draft of the chapter was never read and corrected by the author. I point out these mistakes at the end of my report.

The part of the dissertation I like most is Chapter 3 where several sequential optimization problems were solved. In all these problems it was assumed that we consecutively observe up to n uniform random variables and just after each observation we should to decide whether we reject the current observation (and all preceding ones) or start counting the quantities of interest from this very point. The quantities which were considered here include

- the sum of upper record values,
- the sum of lower records,

- the sum and difference of upper and lower records,
- the number of all (upper and lower) records (in this case the solution can be extended to i.i.d. sequences with arbitrary known continuous distribution).

The purpose was to maximize the expectation of the corresponding quantity. Decision procedures were based on intuitive arguments. For instance, in the first problem including large values of observations, especially at the preliminary stage, may exclude many further candidates for records. Hence a wise solution consists in rejecting relatively large values at each step with increasing acceptance region when the number of further observations decreases. For each above problem, and for each number of remaining observations, the author determined numerically two-element partitions of the interval $[0, 1]$ on the sets of rejection and start counting procedure for the current value of random observation.

I think that it is possible to prove that all the solutions of the problems presented in Chapter 3 are in fact the solutions in the most general family of sequential optimization procedures based on available random data. Each start of counting procedure can be treated as a stopping rule of the rejection process, and for the stopping rules with a finite horizon one can construct the optimal stopping time using the universal backward induction method (see, e.g., Chow, Y. S., Robbins, H., Siegmund, D. (1971). *Great Expectations: the Theory of Optimal Stopping.* Houghton Mifflin Co., Boston, Mass.) This is my suggestion for future research of Mr. Bel'kov.

I have several critical remarks concerning presentation of the results in the thesis. Except for Chapter 2, this is rather far from classic structures of mathematical texts: definitions of basic notions — theorem — proof —possible comments. It was sometimes difficult to distinguish the statements and arguments proving them. Some substantial arguments were omitted, and should be recovered by the readers. Some sequences of arguments were interrupted by comments not related to the respective reasonings. I suggest the candidate submitting further papers to international mathematical journals which require more professional presentations of achievements.

Conclusion. Despite of some negative comments of my report, I claim that he dissertation written by Igor Vladimirovich Bel'kov on the subject *Distributions of total number of upper and lower record values in nonstandard situations* fulfils essential requirements established in the decree *On the procedure of granting academic degrees in the St. Petersburg State University.* The applicant Igor Vladimirovich

Bel'kov deserves granting with the Candidate degree in the Mathematical and Physical Sciences in speciality 01.01.05: Probability Theory and Mathematical Statistics. Point 11 of the decree mentioned above has not been violated.

Member of the Dissertational Council

Prof. dr habil. Tomasz Rychlik
Department of Mathematical Statistics
Institute of Mathematics
Polish Academy of Sciences

Tomasz Rychlik

Warsaw, August 24, 2020

Mistakes found in Chapter 2:

p. 113, l. 10 bottom: X_{2m} with $m = 0$ does not exist.

p. 113, l. 6 bottom: $\mathbb{P}(X_1 = X_2) = \frac{(1-p)(1-r)}{1-pr}$.

p. 115, Proposition 2.1: no justification for the claims.

p. 117, l. 3: $S(n)$ is the value of maximum at the moment of nth record range occurrence.

p. 119, l. 1: $\mathbb{P}\{A(y, v, k, n)\} = (1 - pe^{-v} - qe^{-y})^k$ (formula automatically copied from [63]).

p. 119, lines 4 and 5: $X_{L(n-1)+k+1}$.

p. 119, l. 6: H_p (twice).

p. 120, l. 3: $0 < p = 1 - q < 1$,

p. 121, l. 13: $g(x, u)$.

p. 122, l. 9 bottom: continuous *decreasing* functions.

p. 122, l. 1 bottom: $s \in (0, 1)$.

p. 123, l. 7: $W(1) = |X_1|$.

p. 123, l. 8 bottom: notation X_m is wrong because m was just used for denoting the minimal value.

p. 123, l. 7 bottom: $(1 + \ldots + p^{m-1})r$.

p. 123, l. 4 bottom: qrp^{-x+1} and qsp^{x-1} in the consecutive numerators.

p. 124, l. 1: $\frac{q(rp^{-m+\ell-1}+sp^{M+\ell-1})}{rp^{-m}+sP^{M}}$.

p. 125, l. 7: $\mathbb{P}(W(2) = u + n | W(1) = u)$.

p. 126, l. 6 bottom: $m = 1$ (drop n).

Other minor mistakes:

p. 85, l. 6: kth records were introduced in [24] by Dziubdziela and Kopociński.

p. 85, l. 11: $X_{n-k,n-1} \leq X_{n-k+1,n} \leq \ldots$.

p. 139, l. 3 bottom: $\Big|_{(x_n)_1}^{(x_n)_2}$.

p. 139, l. 2 bottom: $(x_n)_2$.

p. 140, l. 3: $B_n(x)$.

p. 140, l. 1 bottom: $(1 - x_n)^{n+1}$ (subscript lacking).

p. 142, lines 7 and 9 and later on: x_1 and x are two notations for the value of U_1.

p. 145, l. 1 bottom: (3.7) instead of (3.2).

p. 113, l. 10 bottom: [40] instead of [34].