Taylor & Francis
Taylor & Francis Group

Statistics
A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: http://www.tandfonline.com/loi/gsta20

Linearity of regression for weak records, revisited

Rafał Karczewski & Jacek Wesołowski

To cite this article: Rafał Karczewski & Jacek Wesołowski (2017) Linearity of regression for weak records, revisited, Statistics, 51:4, 878-887, DOI: 10.1080/02331888.2017.1301940

To link to this article: https://doi.org/10.1080/02331888.2017.1301940

STATISTICS, 2017
VOL. 51, NO. 4, 878–887
https://doi.org/10.1080/02331888.2017.1301940

Taylor & Francis
Taylor & Francis Group

Linearity of regression for weak records, revisited

Rafał Karczewski and Jacek Wesołowski

Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland

ABSTRACT

Since many years characterization of distribution by linearity of regression of non-adjacent weak records $\mathbb{E}(W_{i+s}|W_i) = \beta_1 W_i + \beta_0$ for discrete observations has been known to be a difficult question. López-Blázquez [Linear prediction ofweak records. The discrete case. Theory Probab Appl. 2004;48(4):718–723] proposed an interesting idea of reducing it to the adjacent case and claimed to have the characterization problem completely solved. We will explain that, unfortunately, there is a flaw in the proof given in that paper. This flaw is related to fact that in some situations the operator responsible for reduction of the non-adjacent case to the adjacent one is not injective. The operator is trivially injective when $\beta_1 \in (0, 1)$. We show that when $\beta_1 \geq 1$ the operator is injective when $s = 2,3,4$. Therefore in these cases the method proposed by López-Blázquez is valid. We also show that the operator is not injective when $\beta_1 \geq 1$ and $s \geq 5$. Consequently, in this case the reduction methodology does not work and thus the characterization problem remains open.

ARTICLE HISTORY
Received 18 October 2015
Accepted 5 July 2016

KEYWORDS
Weak records; linearity of regression; injective operator; characterization; structure theory

AMS SUBJECT CLASSIFICATIONS
62E10; 62G32; 62E15

1. Introduction

The issue of characterization of the common distribution of a sequence $(X_n)_{n\geq 1}$ of iid variables by linearity of regression of records $\mathbb{E}(R_m \mid R_n) = \beta_1 R_n + \beta_0$ for $m \neq n$ has attracted the attention of researchers since the seventies in the last century, when Nagaraja [1], assuming that the common distribution of X_n's is continuous and following methods developed by Ferguson [2] for order statistics, characterized the triplet of exponential, power and Pareto type distributions in the case $m = n+1$. In Nagaraja [3] the case $m = n-1$ for continuous distribution was solved by reducing the problem to the one for order statistics. As a result another triplet of distributions was characterized. The characterization in the case $m = n+2$ was done in [4] through reducing the problem to second-order ordinary differential equation and a careful look at its probabilistic solutions. The characterization issue for continuous distributions was finally resolved in the general case of linearity of regression for non-adjacent records in [5] by using integrated Cauchy functional equation in case $m > n$ and in case $m < n$ by reducing the problem to an analogous problem for order statistics, the latter being solved by a similar method earlier in [6]. Since that time the case of continuous parent distribution has been studied further, for example, for generalized order statistics and for other patterns of regression functions. For these and related issues see, for example, [7–16].

In the case of discrete distribution instead of records (R_n), which are defined through a strict inequality, it is more natural to consider weak records (W_n), which are defined by '$\geq$' relation. That is, a repetition of the last weak record is the next weak record, while for regular records repetitions of records are discarded. In this case, the issue of characterization of the distribution of X_n's through

CONTACT Jacek Wesołowski wesolo@mini.pw.edu.pl Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Koszykowa 75, 00-662 Warszawa, Poland

linearity of regression $\mathbb{E}(W_m \mid W_n) = \beta_1 W_n + \beta_0$ for $m \neq n$ seems not to be related to the methods developed in the continuous case. In particular, under natural assumption that the support of the common law of X_n's is a set of the form $\{0, 1, \ldots, N\}$ with $N \le \infty$ we see that in the case of $m < n$, due to monotonicity of (W_n) sequence, we have $\beta_0 = 0$. To the best of our knowledge, under this assumption ($m < n$) the characterization was obtained only in two special cases: $\mathbb{E}(W_1 \mid W_2) = \beta_1 W_2$ for $\beta_1 > 0$ in [17] and $\mathbb{E}(W_m \mid W_n) = (m/n) W_n$ in [18].

For the case $m > n$ the characterization of distribution of X_n's was first given in [19] for $m = n+1$ with an improvement in [20] – see also related papers [21–23]. The case $m = n+2$ for $\beta_1 = 1$ was considered in [24] and for general β_1 in [20], where an approach via solution of a nonlinear difference equation was applied. In this way a triplet of geometric and negative hypergeometric distributions of the first and second kind was characterized. For the general case $m > n$, López-Blázquez in [25] (we refer to this paper by LB in the sequel) proposed an intriguing idea of reduction of the problem to the adjacent case of $m = n+1$, for which the solution has been already known. However, as it will be explained below, this interesting approach is not as universal as it is claimed in that paper. It appears that there are some inaccuracies in the proof in the case $\beta_1 \ge 1$, that is when $N = \infty$. When we encountered these inaccuracies we were rather confident that it would be possible to overcome them while preserving this brilliant idea of reduction to the adjacent case $m = n+1$. As we will see, this can be done only if $0 < m - n \le 4$. Unfortunately, for higher distances between m and n the idea introduced in LB does not work. Therefore the characterization in the case $m > n+4$ and $\beta_1 \ge 1$ still remains an open problem.

Finally, let us mention that the issue of characterization of discrete distributions by linearity of regression of ordinary records $\mathbb{E}(R_m \mid R_n) = \beta_1 R_n + \beta_0$ has also been considered in the literature. If $m > n$ only characterizations of tails o distribution were eventually obtained, see, for example, [26–32]. If $m < n$ no elegant characterization seems to be possible, see [18], except the case $m = 1$, $n = 2$, see [17,33].

2. Passing from the non-adjacent case to the adjacent one is problematic

We consider a sequence (X_n) of iid random variables having the common distribution $\mathbf{p} = (p_k)$ supported on $\{0, 1, \ldots, N\}$, $N \le \infty$. That is, $p_k = \mathbb{P}(X_1 = k)$, and we also write $q_k = \mathbb{P}(X_1 \ge k)$, $k = 0, 1, \ldots, N$. For such a sequence, we consider the respective sequence of weak records (W_n) which is defined as follows: Let $T_1 = 1$ and $T_n = \inf\{k > T_{n-1} : X_k \ge X_{T_{n-1}}, n > 1$. Then $W_n = X_{T_n}, n \ge 1$. The joint distribution of the first n weak records can be easily derived as

$$\mathbb{P}(W_1 = k_1, \ldots, W_n = k_n) = p_{k_n} \prod_{j=1}^{n-1} \frac{p_{k_j}}{q_{k_j}}, \quad 0 \le k_1 \le \cdots \le k_n.$$

Weak records were introduced in [34] and since then are one of the basic models for ordered discrete random variables. Their basic properties can be found in any monograph on records, for example, Ch. 2.8 of [35], Ch. 16 of [36] or in Ch. 6.3. of relatively recent monograph [37]. It is well-known that weak records form a homogeneous Markow chain with the transition probability of the form

$$\mathbb{P}(W_n = k_n \mid W_{n-1} = k_{n-1}) = \frac{p_{k_n}}{q_{k_{n-1}}}, \quad k_n \ge k_{n-1} \ge 0.$$

Therefore, for $m < n$

$$\mathbb{P}(W_n = k_n \mid W_m = k_m) = \sum_{k_m \le k_{m+1} \le \cdots \le k_{n-1} \le k_n} \prod_{i=m}^{n-1} \frac{p_{k_{i+1}}}{q_{k_i}}, \quad k_n \ge k_m \ge 0.$$

For fixed positive integers i,s we will be interested in conditional expectation $\mathbb{E}(W_{i+s} \mid W_i)$. Therefore we need to assume that $\mathbf{p}$ is such that this conditional expectation is finite. Since the conditional distribution of $W_{i+s} \mid W_i$ does not depend on i, we will denote the set of distributions $\mathbf{p}$ for which $\mathbb{E}(W_{i+s} \mid W_i)$ is finite by $\mathcal{M}_s$.

Let us consider a family C_s of discrete distributions $\mathbf{p} = (p_k)_{k \geq 0} \in \mathcal{M}_s$, concentrated on $\{0, 1, 2, \ldots, N\}$ ($N \leq \infty$) with property that if the common law of iid random variables $(X_n)_{n \geq 1}$ belongs to C_s then the regression of weak records $\mathbb{E}(W_{i+s} \mid W_i)$ is linear. It is known that $C_1 \subseteq C_s$ for all $s \geq 1$. We are interested in the opposite inclusion. In LB it is claimed that the opposite implication holds true, however the proof of this inclusion given in there is not correct. We will explain why it is not correct, then improve the method proposed in LB to show that the inclusion holds true for $s = 2,3,4$ and finally we will show that the method fails for $s \geq 5$.

Before we state the result from LB we need to introduce some notation. Let $\mathbf{v} = (v(0), v(1), \ldots, v(N)) \in \mathbb{C}^{N+1}$ (for $N = \infty$, $\mathbf{v} = (v(0), v(1), \ldots)$). Let us define a linear operator:

$$A : D(A) \longrightarrow \mathbb{C}^{N+1}; \quad Av(l) = \frac{1}{q_l} \sum_{k=l}^{N} v(k) p_k, \quad l = 0, 1, \ldots, N,$$

where

$$D(A) = \left\{ \mathbf{v} \in \mathbb{C}^{N+1} : \sum_{k=0}^{N} |v(k)| p_k < \infty \right\}.$$

We also define the domain of composition of operator A with itself since we will need that later on:

$$D(A^m) = \{\mathbf{v} \in D(A) : A^k \mathbf{v} \in D(A) \quad \text{for } k = 1, \ldots, m-1\} \text{ for } m \geq 2,$$

where

$$A^0 \mathbf{v} = \mathbf{v} \quad \text{and} \quad A^m \mathbf{v} = A(A^{m-1} \mathbf{v}) \quad \text{for } m \geq 1.$$

Below we present matrix representation of the operator A (which is an infinite matrix when $N = \infty$):

$$A = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots & p_N \\ 0 & \frac{p_1}{q_1} & \frac{p_2}{q_1} & \cdots & \frac{p_N}{q_1} \\ 0 & 0 & \frac{p_2}{q_2} & \cdots & \frac{p_N}{q_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{p_N}{q_N} \end{bmatrix}.$$

Note that A is an upper-triangular matrix with non-zero diagonal entries. Let $e_m(l) = \mathbb{E}(W_{i+m} \mid W_i = l)$. Then, directly from the form of the conditional distribution it follows that

$$\mathbf{e}_{m+1} = A\mathbf{e}_m \quad \text{for } m = 1, 2, \ldots \tag{1}$$

In particular $\mathbf{e}_m$ is in the domain of A, given that $\mathbf{e}_{m+1}$ exists. Now we can state the theorem proposed in LB.

Theorem 2.1: *Let X be a random variable with discrete distribution with support* $\{0, 1, 2, \ldots, N\}$ ($N \leq \infty$). *Let (W_n) be the sequence of weak records built on a sequence (X_n) of iid random variables* *having the same distribution as X. Assume that for some $i, s \geq 1$*

$$\mathbb{E}(W_{i+s} \mid W_i) = \beta_0 + \beta_1 W_i, \tag{2}$$

where $\beta_0, \beta_1 \in \mathbb{R}$. Then $\beta_0, \beta_1 > 0$. Let γ_0, γ_1 be unique solutions of

$$\beta_1 = \gamma_1^s, \quad \beta_0 = \gamma_0 \frac{1-\gamma_1^s}{1-\gamma_1}. \tag{3}$$

Then

(1) *if $0 < \beta_1 < 1$, then $\gamma_0/(1-\gamma_1) \in \mathbb{N}$*

$$X \sim \mathrm{nh}_I\left(1, \frac{\gamma_1}{1-\gamma_1}, \frac{\gamma_0}{1-\gamma_1}\right),$$

(2) *if $\beta_1 = 1$, then*

$$X \sim \mathrm{geo}\left(\frac{1}{1+\gamma_0}\right),$$

(3) *if $\beta_1 > 1$, then*

$$X \sim \mathrm{nh}_{II}\left(1, \frac{\gamma_0+1}{\gamma_1-1}, \frac{\gamma_0}{\gamma_1-1}\right).$$

The symbols of distributions above have the following meaning: nh_I is for the negative hypergeometric distribution of the first kind, geo is for the geometric distribution, nh_{II} is for the negative hypergeometric distribution of the second kind (more details on nh_I and nh_{II} laws can be found, e.g., in [20]).

We will now recall basic steps in the proof given in LB. Observe, that since $e_s(l) = \mathbb{E}(W_{i+s} \mid W_i = l)$ is strictly increasing, we have $\beta_1 > 0$ and $\beta_0 = e_s(0) > 0$. Let γ_0, γ_1 be unique solutions of (3). Now, for $m = 1, \ldots, s$ we define $\mathbf{d}_m$ through the equality

$$e_m(j) = \gamma_0 \frac{1-\gamma_1^m}{1-\gamma_1} + \gamma_1^m j + d_m(j), \quad j = 0, 1, \ldots, N. \tag{4}$$

Directly from the definition of $\mathbf{d}_m$ and the assumption that $\mathbf{e}_s$ exists we obtain that $\mathbf{d}_m$ is in the domain of A for $m = 1, \ldots, s-1$. From Equation (2) we have that $\mathbf{d}_s = 0$. After easy algebra we obtain

$$\mathbf{d}_{m+1} = \gamma_1^m \mathbf{d}_1 + A\mathbf{d}_m, \quad m = 1, \ldots, s-1. \tag{5}$$

From Equation (5) we can obtain that $\mathbf{d}_m$ is in the domain of A^2 for $m = 1, \ldots, s-1$ and by iterating (5) we get that $\mathbf{d}_1$ is in the domain of A^{s-1}. This can be iterated and, consequently,

$$\mathbf{d}_m = B_m \mathbf{d}_1, \quad m = 1, \ldots, s, \quad \text{where } B_m = \sum_{k=0}^{m-1} \gamma_1^{m-1-k} A^k. \tag{6}$$

Let us note that A and, consequently, B_m depends on the unknown distribution $\mathbf{p} = (p_n)_{n \geq 0}$. To emphasize this fact, sometimes we will write $B_m^{(\mathbf{p})}$ instead of B_m.

At this stage of argument we read in LB:

Note that B_m is an upper-triangular matrix with non-zero diagonal entries; then B_m has an inverse (even in the infinite case)

This is why, besides the case $N<\infty$ (equivalent to $\gamma_1\in(0,1)$), the proof is incorrect. The above statement is false in the case $N=\infty$ (that is $\gamma_1\geq 1$). Infinite matrices represent linear operators between linear spaces. In general such transformations, which are represented by infinite upper-triangular matrices with non-zero diagonal entries, do not have to be invertible and even injective. As an example consider a linear transformation $B:\mathbb{R}^\infty\to\mathbb{R}^\infty$ represented by the matrix

$$B=[b_{ij}]_{i,j=0}^{\infty};\quad b_{ij}=\begin{cases}1 & \text{for } i=j \text{ or } j=i+1,\\ 0 & \text{otherwise.}\end{cases}$$

Obviously, B is an upper-triangular matrix with non-zero diagonal entries. Let $v=(1,-1,1,-1,\dots)$. Then $Bv=0$ and thus B is not injective in $\mathbb{R}^\infty$, consequently, it cannot be invertible. However, if we consider B as a linear operator on the space of sequences convergent to 0, then B is invertible with B^{-1} being also upper-triangular with nth row of the form $(0,\dots,0,1,-1,1,-1,\dots)$, where the first 1 is at the position n, $n\geq 1$.

In the next section we will discuss in detail injectivity of the operator B_s defined in Equation (6), which is of crucial importance since the rest of the argument from LB lies in plugging $m=s$ in Equation (6). Since, as it was observed before, $\mathbf{d}_s=0$, it follows that

$$B_s^{(\mathbf{p})}\mathbf{d}_1=0.$$

So if $B_s^{(\mathbf{p})}$ was injective for any $\mathbf{p}\in\mathcal{M}_s$ we would get $\mathbf{d}_1=0$ and, consequently,

$$e_1(j)=\gamma_0+\gamma_1 j\quad \text{for } j=0,1,\dots,N.$$

That is, the crucial problem for validity of the proof as suggested in LB is a question of injectivity of $B_s^{(\mathbf{p})}$ for any $\mathbf{p}\in\mathcal{M}_s$.

3. Is the operator $B_s^{(\mathbf{p})}$ injective?

In this section we will show how injectivity of $B_s^{(\mathbf{p})}$ defined on $D(B_s^{(\mathbf{p})})=D(A^{s-1})$ depends on s. Let us recall that we are considering here only such distributions $\mathbf{p}$ for which $\mathbb{E}(W_{i+s}\mid W_i)<\infty$ and, as it has already been mentioned, this condition depends only on s and not on i. First, we will consider operators with domains being subsets of $\mathbb{C}^\infty=\{(x_0,x_1,\dots):x_k\in\mathbb{C}\}$, the linear space of sequences of complex numbers.

Theorem 3.1: *Let $B_s^{(\mathbf{p})}$ be the operator defined by Equation (6) with $N=\infty$ on a domain $D(B_s^{(\mathbf{p})})\subset\mathbb{C}^\infty$. Then $B_s^{(\mathbf{p})}$ is injective for any distribution $\mathbf{p}\in\mathcal{M}_s$ iff $s\in\{2,3,4\}$.*

Proof: Since $\sum_{k=0}^{s-1}z^k=\prod_{k=1}^{s-1}(z-\lambda_k)$, where $\lambda_k=\cos(2k\pi/s)+i\sin(2k\pi/s)$, $k=1,\dots,s-1$, $s\geq 2$, we can represent the operator B_s in the following way

$$B_s=\prod_{k=1}^{s-1}(A-\gamma_1\lambda_k I),\tag{7}$$

where I is the identity operator. Thus if $\gamma_1\lambda_\ell$ is an eigenvalue of A for some $\ell=1,2,\dots,s-1$, then B_s is not injective. Indeed, if $\mathbf{x}_\ell\in D(B_s)$ is a respective non-zero eigenvector of $\gamma_1\lambda_\ell$, then (note that $(A-\gamma_1\lambda_j I)$ and $(A-\gamma_1\lambda_k I)$ commute)

$$B_s\mathbf{x}_\ell=\left[\prod_{\substack{k=1\\k\neq\ell}}^{s-1}(A-\gamma_1\lambda_k I)\right](A-\gamma_1\lambda_\ell I)\mathbf{x}_\ell=0.$$

Consequently, B_s is not injective.

Assume now that none of $\gamma_1\lambda_\ell$, $\ell = 1, 2, \dots, s-1$ is an eigenvalue of A. Then B_s is a composition of injective operators, so B_s must also be injective.

Finally, we conclude that B_s is injective if and only if all $\gamma_1\lambda_\ell$, $\ell = 1, 2, \dots, s-1$, are not eigenvalues of A.

We will now examine eigenvalues of A which are of the form $\lambda = \gamma_1\lambda_\ell$. Let $\lambda \in \mathbb{C}, \mathbf{x} \in D(A), \mathbf{x} \neq 0$, be such that $A\mathbf{x} = \lambda\mathbf{x}$ which is equivalent to

$$\sum_{j=i}^{\infty} p_j x_j = \lambda x_i q_i \quad \forall i \geq 0. \tag{8}$$

After subtracting the equality for i and $i+1$ sidewise we obtain

$$x_i p_i = \lambda(x_i q_i - x_{i+1} q_{i+1}).$$

Hence we have

$$x_{i+1} = \frac{\lambda q_i - p_i}{\lambda q_{i+1}} x_i \quad \forall i \geq 0.$$

Expanding this recursion gives

$$x_{i+1} = \frac{\lambda q_i - p_i}{\lambda q_{i+1}} \frac{\lambda q_{i-1} - p_{i-1}}{\lambda q_i} \cdots \frac{\lambda q_0 - p_0}{\lambda q_1} x_0 \quad \forall i \geq 0.$$

We assumed that $\mathbf{x} \in D(A)$ and $\mathbf{x} \neq 0$ which now, given the expression above and (8) for $i=0$, implies

$$\lambda \text{ is an eigenvalue of } A \iff \sum_{k=1}^{\infty} |b_k(\lambda)| p_k < \infty \quad \text{and} \quad 0 = p_0 - \lambda q_0 + \sum_{k=1}^{\infty} b_k(\lambda) p_k, \tag{9}$$

where

$$b_k(\lambda) = \prod_{i=0}^{k-1} \frac{\lambda q_i - p_i}{\lambda q_{i+1}}.$$

Let us denote

$$S_n(\lambda) = \begin{cases} p_0 - \lambda q_0 & \text{for } n = 0 \\ p_0 - \lambda q_0 + \displaystyle\sum_{k=1}^{n} b_k(\lambda) p_k & \text{for } n \geq 1 \end{cases} \quad \text{and} \quad S_n^*(\lambda) = \sum_{k=1}^{n} |b_k(\lambda)| p_k.$$

By an easy induction argument we obtain the following product representation of $S_n(\lambda)$ for $n \geq 1$

$$S_n(\lambda) = (p_0 - \lambda q_0) \prod_{k=1}^{n} \frac{\lambda q_k - p_k}{\lambda q_k} = (p_0 - \lambda q_0) \prod_{k=1}^{n} \left(1 - \frac{c_k}{\lambda}\right), \quad \text{where } c_k = \frac{p_k}{q_k} \in (0, 1).$$

As observed in Equation (9), we have to consider the situation when $\lim_{n\to\infty} S_n^*(\lambda) < \infty$ and $\lim_{n\to\infty} S_n(\lambda) = 0$. Note that

$$\lim_{n\to\infty} S_n(\lambda) = 0 \iff \lim_{n\to\infty} |S_n(\lambda)|^2 = 0 \iff \lim_{n\to\infty} |p_0 - \lambda q_0|^2 \prod_{k=1}^{n} \left|1 - \frac{c_k}{\lambda}\right|^2 = 0.$$

Since $\lambda = \gamma_1\lambda_\ell$ for some $\ell = 1, \dots s-1$, $p_0 - \lambda \neq 0$. Furthermore, we observe that $1/\lambda_\ell = \bar{\lambda}_\ell = \lambda_{s-\ell}$. Consequently

$$\lim_{n\to\infty} S_n(\gamma_1\lambda_\ell) = 0 \iff \prod_{k=1}^{\infty}\left|1 - \frac{\lambda_{s-\ell}}{\gamma_1}c_k\right|^2 = 0.$$

Since the single factor in the product has the form

$$\left|1 - \frac{\lambda_{s-\ell}}{\gamma_1}c_k\right|^2 = 1 - 2\frac{c_k}{\gamma_1}\cos\left(\frac{2(s-\ell)\pi}{s}\right) + \left(\frac{c_k}{\gamma_1}\right)^2 = 1 - 2\frac{c_k}{\gamma_1}\cos\left(\frac{2\ell\pi}{s}\right) + \left(\frac{c_k}{\gamma_1}\right)^2,$$

we see that for all $k \geq 1$ it assumes the minimum for $\ell = 1$. With that and Equation (9) in mind, we can conclude that $\gamma_1\lambda_1$ is not an eigenvalue iff $\gamma_1\lambda_\ell$ are not eigenvalues for any $\ell = 1, 2, \dots, s-1$ which leads to

$$B_s^{(\mathbf{p})} \text{ is not injective} \iff \prod_{k=1}^{\infty}\left|1 - \frac{\lambda_1}{\gamma_1}c_k\right|^2 = 0 \quad \text{and} \quad \lim_{n\to\infty} S_n^*(\gamma_1\lambda_1) < \infty. \tag{10}$$

Thus we need to examine if the condition

$$0 = \prod_{k=1}^{\infty}\underbrace{\left(1 - 2\frac{c_k}{\gamma_1}\cos\left(\frac{2\pi}{s}\right) + \left(\frac{c_k}{\gamma_1}\right)^2\right)}_{a_{k,s}}$$

is satisfied.

Note that

- $a_{k,2} = 1 + 2(c_k/\gamma_1) + (c_k/\gamma_1)^2 > 1$,
- $a_{k,3} = 1 + c_k/\gamma_1 + (c_k/\gamma_1)^2 > 1$,
- $a_{k,4} = 1 + (c_k/\gamma_1)^2 > 1$.

Thus, in these three cases the above condition does not hold. Consequently, for any distribution $\mathbf{p}$ the operator $B_s^{(\mathbf{p})}$ is injective for $s = 2,3,4$.

Now consider $s \geq 5$ and a geometric distribution $\mathbf{p}$ with parameter $p \in (0,1)$. We choose the parameter p in such a way that $\cos(2\pi/5) > p/2$. Note that for geometric distribution $c_k = p$, $k = 1, 2, \dots$ Then

$$2\cos\left(\frac{2\pi}{s}\right) \geq 2\cos\left(\frac{2\pi}{5}\right) > p \geq \frac{p}{\gamma_1},$$

and thus

$$2\frac{p}{\gamma_1}\cos\left(\frac{2\pi}{s}\right) > \left(\frac{p}{\gamma_1}\right)^2,$$

which yields

$$1 > \underbrace{1 - 2\frac{p}{\gamma_1}\cos\left(\frac{2\pi}{s}\right) + \left(\frac{p}{\gamma_1}\right)^2}_{a_{k,s}} = \text{const}.$$

Furthermore

$$S_n^*(\gamma_1\lambda_1) = \sum_{k=1}^{n}\prod_{i=0}^{k-1}\left|\frac{\gamma_1\lambda_1 q_i - p_i}{\gamma_1\lambda_1 q_{i+1}}\right| p_k = \sum_{k=1}^{n}\frac{q_0 p_k}{q_k}\prod_{i=0}^{k-1}\left|\frac{\gamma_1\lambda_1 q_i - p_i}{\gamma_1\lambda_1 q_i}\right|$$

$$= p\sum_{k=1}^{n}\prod_{i=0}^{k-1}\sqrt{a_{k,s}} = p\sum_{k=1}^{n}(\sqrt{a_{1,s}})^k.$$

Since $a_{k,s} = a_{1,s} < 1$, we obtain that $\lim_{n\to\infty} S_n^* < \infty$. Therefore Equation (10) yields that in the case of geometric distribution $\mathbf{p}$ with the parameter p satisfying the inequality as above $B_s^{(\mathbf{p})}$ for $s \geq 5$ is not injective. ■

In Theorem 3.1 we examined injectivity of $B_s^{(\mathbf{p})}$ with domain $D(B_s^{(\mathbf{p})}) \subset \mathbb{C}^\infty$, but for the purposes of the problem we should only consider the injectivity or its lack on $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$. Of course, injectivity on $D(B_s^{(\mathbf{p})}) \subseteq \mathbb{C}^\infty$ implies injectivity on $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$. Consequently, it follows from Theorem 3.1 that for any $\mathbf{p}$ the operator $B_s^{(\mathbf{p})}$ is injective in $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$ for $s = 2,3,4$. The opposite implication may not be true in general but in the case we consider it turns out that it holds.

Theorem 3.2: *For any $\mathbf{p}$ the operator $B_s^{(\mathbf{p})}$ is injective on $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$ iff $s \in \{2, 3, 4\}$.*

Proof: As already mentioned the implication '$\Leftarrow$' is an immediate consequence of the same implication from Theorem 3.1.

We will prove the opposite implication by contradiction, that is, we will show that there exists a distribution $\mathbf{p}$ such that for $s \geq 5$ the operator $B_s^{(\mathbf{p})}$ is not injective in $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$. Let $s \geq 5$ and $\mathbf{p} = (p_k)_{k=0}^\infty$ be geometric distribution with parameter $p \in (0, 1)$ such that $\cos(2\pi/5) > p/2$. Then, from the proof of Theorem 3.1, we obtain that $\gamma_1\lambda_1, \gamma_1\bar{\lambda}_1$ are eigenvalues of A. We denote a non-zero eigenvector attached to $\lambda = \gamma_1\lambda_1$ by $\mathbf{x} = (x_0, x_1, \ldots)$ and the vector with conjugate entries by $\bar{\mathbf{x}} = (\bar{x}_0, \bar{x}_1, \ldots)$.

We will first note that $\mathbf{x}$ cannot be of the form $\mathbf{x} = i\mathbf{y}$ for a vector $\mathbf{y} \in \mathbb{R}^\infty$. Indeed, in such a case we would have $A\mathbf{y} = \lambda\mathbf{y}$ which is impossible since λ is not a real number.

Note that $\bar{\mathbf{x}}$ is an eigenvector of A attached to the eigenvalue $\bar{\lambda}$, because

$$A\bar{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}},$$

where the first equality holds since A is a matrix with real entries. Consider

$$B_s^{(\mathbf{p})}\mathbf{x} = \prod_{k=1}^{s-1}(A - \gamma_1\lambda_k I)\mathbf{x}.$$

Now, due to the fact that $(A - \gamma_1\lambda_k I)$ and $(A - \gamma_1\lambda_\ell I)$ commute, we obtain:

$$B_s^{(\mathbf{p})}\,\mathbf{x} = \left(\prod_{k=2}^{s-1}(A - \gamma_1\lambda_k I)\right)(A - \gamma_1\lambda_1 I)\mathbf{x} = \prod_{k=2}^{s-1}(A - \gamma_1\lambda_k I)\mathbf{0} = \mathbf{0}.$$

The fact that $\bar{\lambda}_1 = \lambda_{s-1}$ and that $\bar{\mathbf{x}}$ is an eigenvector of A respective to $\bar{\lambda}$ yield

$$B_s^{(\mathbf{p})}\bar{\mathbf{x}} = \prod_{k=1}^{s-2}(A - \gamma_1\lambda_k I)(A - \gamma_1\lambda_{s-1} I)\bar{\mathbf{x}} = \prod_{k=1}^{s-2}(A - \gamma_1\lambda_k I)\mathbf{0} = \mathbf{0}.$$

Fix $\mathbf{z} = \mathbf{x} + \bar{\mathbf{x}} \in \mathbb{R}^\infty$. Note that $\mathbf{z} \neq \mathbf{0}$, because as it has already been observed, $\mathbf{x}$ cannot have all entries which are purely imaginery. Finally,

$$B_s^{(\mathbf{p})}\mathbf{z} = B_s^{(\mathbf{p})}(\mathbf{x} + \bar{\mathbf{x}}) = B_s^{(\mathbf{p})}\mathbf{x} + B_s^{(\mathbf{p})}\bar{\mathbf{x}} = \mathbf{0}.$$

Hence $B_s^{(\mathbf{p})}$ is not injective in $D(B_s^{(\mathbf{p})}) \cap \mathbb{R}^\infty$ for $s \geq 5$. ■

4. Conclusion

The above considerations on injectivity of $B_s^{(\mathbf{p})}$ lead to the following correction to the result proposed in LB and recalled in Theorem 2.1.

Proposition 4.1: *The assertion of Theorem* 2.1 *holds true when* $\gamma_1 < 1$ (*that is,* $N < \infty$) *for any* $s \geq 1$. *For* $\gamma_1 \geq 1$ (*that is,* $N = \infty$) *the assertion of Theorem* 2.1 *holds true for* $s \in \{1, 2, 3, 4\}$.

Proof: It is well known (see Section 1) that the result for $s = 1$ holds true. The proof in the case $\gamma_1 < 1$ (which implies $N < \infty$) given in LB is correct since in this case the operator $B_s^{(\mathbf{p})}$ is invertible for any $s \geq 2$. For $\gamma_1 \geq 1$ (which implies $N = \infty$) due to Theorem 3.2 we have injectivity of $B_s^{(\mathbf{p})}$ for $s \in \{2, 3, 4\}$ therefore the method of the proof proposed in LB is correct and thus the respective part of the assertion from Theorem 2.1 holds true. ■

Finally, let us emphasize that for $\beta_1 \geq 1$ and $s \geq 5$ it follows from Theorem 3.1 that $B_s^{(\mathbf{p})}$ may not be injective for some distributions $\mathbf{p} \in \mathcal{M}_s$, even such that appear in the conclusion of Theorem 2.1 (the geometric law was identified as such in the proof of Theorem 3.1) and thus the argument used in LB is no longer valid. *Therefore the problem of characterization of the parent distribution of the sequence of iid observations from the discrete distribution by the condition* $\mathbb{E}(W_{i+s} \mid W_i) = \beta_1 W_i + \beta_0$ *for* $\beta_1 \geq 1$ *and* $s \geq 5$ *remains open* !

Acknowledgements

We are very thankful to referees for their remarks which helped us a lot in improving presentation of the material of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Nagaraja HN. On a characterization based on record values. Austral J Statist. 1977;19:70–73.
[2] Ferguson TS. On characterizing distributions by properties of order statistics. Sankhya Ser A. 1967;29:265–278.
[3] Nagaraja HN. Some characterizations of discrete distributions based on linear regressions of adjacent order statistics. J Statist Plann Inference. 1988;20:65–75.
[4] Ahsanullah M, Wesołowski J. Linearity of best predictors for non-adjacent record values. Sankhya Ser B. 1998;60:221–227.
[5] Dembińska A, Wesołowski J. Linearity of regression for non-adjacent record values. J Statist Plann Inference. 2000;90:195–205.
[6] Dembińska A, Wesołowski J. Linearity of regression for non-adjacent order statistics. Metrika. 1998;48:215–222.
[7] Beg MI, Ahsanullah M, Gupta RC. Characterizations via regressions of generalized order statistics. Statist Meth. 2013;12:31–41.
[8] Ahsanlullah M, Hamedani GG. Characterizations of continuous distributions based on conditional expectation of generalized order statistics. Comm Statist Theory Methods. 2013;42(19):3608–3613.
[9] Bairamov I, Ahsanullah M, Pakes AG. A characterization of continuous distributions via regression on pairs of record values. Aust N Z J Stat. 2005;47(4):543–547.

[10] Bieniek M. On characterizations of distributions by regression of adjacent generalized order statistics. Metrika. 2007;66(2):233–242.
[11] Bieniek M. A note on characterizations of distributions by regressions of non-adjacent generalized order statistics. Aust N Z J Stat. 2009;51(1):89–99.
[12] Bieniek M, Szynal D. Characterizations of distributions via linearity of regression of generalized order statistics. Metrika. 2003;58:259–271.
[13] Cramer E, Kamps U, Keseling C. Characterizations via linear regression for ordered random variables – a unifying approach. Comm Statist Theory Methods. 2004;33(12):2885–2911.
[14] Gupta RC, Ahsanullah M. Some characterization results based on the conditional expectation of a function of non-adjacent order statistic (record value). Ann Inst Statist Math. 2004;56(4):721–732.
[15] López-Blázquez F, Moreno-Rebollo JL. A characterization of distributions based on linear regressions of order statistics and record values. Sankhya Ser A. 1997;59:311–323.
[16] Yanev GP. Characterization of exponential distribution via regression of one record value on two non-adjacent record values. Metrika. 2012;75:743–760.
[17] López-Blázquez F, Wesołowski J. Discrete distributions for which the regression of the first record on the second is linear. Test. 2001;10:121–131.
[18] López-Blázquez F, Wesołowski J. Linearity of regression for the past weak and ordinary records. Statistics. 2004;38(6):457–464.
[19] Stepanov AV. A characterization theorem for weak records. Theory Probab Appl. 1994;38:762–764.
[20] Wesołowski J, Ahsanullah M. Linearity of regression for non adjacent weak records. Statist Sinica. 2001;11:39–52.
[21] Aliev FA. Characterizations of distributions through weak records. J Appl Statist Sci. 1998;8:13–16.
[22] Danielak K, Dembińska A. On characterizing discrete distributions via conditional expectations of weak record values. Metrika. 2007;66(2):129–138.
[23] Dembińska A, López-Blázquez F. A characterization of geometric distribution through *k*th weak records. Comm Statist Theory Methods. 2005;34(12):2345–2351.
[24] Aliev FA, Characterization of geometric distribution through weak records. In: Balakrishnan N, Ibragimov IA, Nevzorov VB, editors. Asymptotic methods in probability and statistics with applications. Birkhäuser: Boston; 2001. p. 299–307.
[25] López-Blázquez F. Linear prediction of weak records. The discrete case. Theory Probab Appl. 2004;48(4):718–723.
[26] Ahsanullah M, Holland B. Some properties of the distribution of record values from the geometric distribution. Handbook of Statist. 1984;25:319–327.
[27] Huang W-J, Su J-S. On certain problems involving order statistics – a unified approach through order statistics property of point processes. Sankhya Ser A. 1999;61:36–49.
[28] Kirmani SNUA, Alam SN. Characterization of the geometric distribution by the form of a predictor. Comm Statist Theory Methods. 1980;9:541–547.
[29] Korwar RM. On characterizing distributions for which the second record value has a linear regression on the first. Sankhya Ser B. 1984;46:108–109.
[30] Nagaraja HN, Sen P, Srivastava RC. Some characterizations of geometric tail distributions based on record values. Statist Papers. 1989;30:147–155.
[31] Rao RC, Shanbhag DN. Recent result on characterizations of probability distributions: a unified approach through extensions of Deny's theorem. Adv in Appl Probab. 1986;18:660–678.
[32] Srivastava RC. Two characterizations of the geometric distribution by record values. Sankhya Ser B. 1979;40:276–278.
[33] Franco M, Ruiz JM. Characterizations of discrete distributions based on conditional expectations of record values. Statist Papers. 2001;42:101–110.
[34] Vervaat W. Limit theorems for records from discrete distributions. Stoch Proc Appl. 1973;1:317–334.
[35] Arnold BC, Balakrishnan N, Nagaraja NH. Records. New York: Wiley; 1998.
[36] Nevzorov VB. Records: mathematical theory. Providence: AMS; 2001.
[37] Ahsanullah M, Nevzorov VB. Records via probability theory. Amsterdam: Atlantis Press; 2015.