Discrete distributions for which the regression of the first record on the second is linear

Fernando López-Blázquez
Departamento de Estadística e Investigación Operativa
Universidad de Sevilla, Spain

Jacek Wesolowski
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska, Poland

Abstract

The linearity of regression of the first record on the second is examined for discrete random variables. Both ordinary and weak records are considered. The analysis involves the determination of all possible linear relationships and all possible probability distributions. Several characterizations of geometric distributions are also shown.

Key Words: Discrete distributions, generalized geometric distribution, geometric distribution, linearity of regression, ordinary records, weak records.

AMS subject classification: 60E05, 62E10, 02E15.

1 Introduction

For a sequence $X = \{X_n\}_{n \geq 1}$ of independent identically distributed (iid) random variables let us define record times as $U(1) = 1$, $U(n) = \inf\{j > U(n-1) : X_j > X_{U(n-1)}\}$, for $n = 2, 3, \ldots$. Then $R_n = X_{U(n)}$ is called the $n$-th record of the sequence $X$. The linearity of the regression of $R_{n+1}$ given $R_n$ within the class of continuous distributions was studied for the first time in Nagaraja (1977), where a family of three distributions with this property was identified. Nagaraja (1988) also described a class of distributions for which the regression of $R_n$ on $R_{n+1}$ is linear, and observed that the exponential distribution is the only distribution for which both the regressions for the adjacent records are linear. All these results were obtained under the assumption that the common distribution of $X_j$ is continuous.

*Correspondence to: Fernando López-Blázquez. Dpto. de Estadística e I.O., Universidad de Sevilla, Sevilla, Spain. Email: lopez@cica.es

Received: January 2000; Accepted: October 2000
Instead of the regular records defined above, for the discrete distributions Vervaat (1973) proposed to use weak records, which are defined by weak record times \( V(1) = 1, \ V(n) = \inf\{j > V(n-1) : X_j \geq X_{V(n-1)}\} \), for \( n = 2, 3, \ldots \). Then, \( W_n = X_{V(n)} \) is called the \( n \)-th weak record. This definition seems to be much more natural in the discrete case, since it gives no priority to the index of the observation, which agrees with the intuition for the iid observations. Observe that in the case of continuous distributions \( R_n = W_n \text{ a.s.} \) Furthermore, in the discrete case weak records are also defined for distributions with bounded support, while for ordinary records this is not possible, without additional assumptions.

We restrict ourselves to supports of \( X_j \)'s of the form \( \{0, 1, \ldots, N\} \) with \( N \) possibly equal to infinity. The joint distributions for weak records can be easily derived

\[
P(W_1 = k_1, \ldots, W_n = k_n) = p_{k_n} \prod_{r=1}^{n-1} \frac{p_{k_r}}{q_{k_r}}, \quad 0 \leq k_1 \leq \cdots \leq k_n \leq N, \tag{1.1}\]

where \( p_k = P(X_1 = k) \) and \( q_k = \sum_{j \geq k} p_j, \ k \geq 0. \) Obviously, \( k_n < N \) if \( N = \infty \) in (1.1).

In the case of ordinary records, as a bounded support is not permitted, \( N = \infty \) and the joint distribution is

\[
P(R_1 = k_1, \ldots, R_n = k_n) = p_{k_n} \prod_{r=1}^{n-1} \frac{p_{k_r}}{q_{k_r+1}}, \quad 0 \leq k_1 < \cdots < k_n < \infty.
\]

Consequently, \( P(W_{n+1} = l|W_n = k) = p_l/q_k, \ 0 \leq k \leq l, \) and \( P(R_{n+1} = l|R_n = k) = p_l/q_{k+1}, \ 0 \leq k < l < \infty, \) so both conditional distributions have a simple form. Consequently, the problem of the linearity of the regression of \( R_{n+1} \) on \( R_n \) was solved in Korwar (1984), where the family of distributions consisting of the geometric tail and negative hypergeometric of the second type tail distributions was characterized - see also comments in the monograph of Arnold, Balakrishnan and Nagaraja (1998). The same problem for the regression of \( W_{n+1} \) on \( W_n \) was solved in Stepanov (1993) and Wesolowski and Ahsanullah (2000), where the question of the linearity of the regression of \( W_{n+2} \) on \( W_n \) was also completely resolved. More characterizations through weak records can be seen in Aliev (1998).

Nothing is known about regressions in the opposite direction, i.e. for \( E(R_n|R_{n+1}) \) or \( E(W_n|W_{n+1}) \) which seem to be much more complicated.