THEORY PROBAB. APPL.
Vol. 48, No. 4, pp. 718–723

LINEAR PREDICTION OF WEAK RECORDS: THE DISCRETE CASE*

F. LÓPEZ-BLÁZQUEZ†

Abstract. We characterize the family of discrete distributions for which the best mean square error predictor of a future weak record is a linear function of a past or observed weak record.

Key words. records, weak records, conditional expectation, mean square error, linear predictors negative hypergeometric distributions, geometric distribution

DOI. 10.1137/S0040585X97980816

1. Introduction.

A problem of interest is the prediction of a future record value, R_{i+s} $(s>0)$, with the information provided by a past or observed record R_i. A common procedure for obtaining a predictor is to minimize the mean square error (MSE). It is well known that the best MSE predictor is the conditional expectation $\mathbf{E}[R_{i+s}\,|\,R_i]$. The functional form of this predictor may be complicated — for that reason it is sometimes desirable to find a predictor which is best within a class of simple predictors. Of special interest, due to their simplicity, are the linear predictors of the form $\beta_0+\beta_1 R_i$.

A natural question arises: What are the distributions for which the best MSE predictor, $\mathbf{E}[R_{i+s}\,|\,R_i]$, is a linear function of R_i? Nagaraja [7] solved the problem of linearity of regression for the adjacent record values (i.e., $s=1$) in the case of continuous distributions. For further results see also [8]. The case $s=2$ was solved by Ahsanullah and Wesołowski [1]. López-Blázquez [5] solved the problem for an arbitrary s by imposing some smoothness conditions; see also [6]. Recently, Dembińska and Wesołowski [3] solved the problem in the continuous case without any smoothness assumptions.

A modification of the definition of records for discrete distributions was given by Vervaat [10] by introducing weak records. This modification basically permits ties between records. For this reason, weak records are well defined for discrete distributions with bounded support, while usual records are not. Stepanov [9] investigated linearity of regression for weak records in the adjacent case for discrete distributions. More general forms for the regression function in the adjacent case have been studied by Aliev [2]. Wesołowski and Ahsanullah [11] have characterized the discrete distributions with linear regression between weak records when the spacing is $s=2$.

The aim of this paper is to give the class of discrete distributions for which the best MSE predictor is linear for an arbitrary spacing $s>0$. Our main result states that such a class consists of geometric, beta-binomial and beta-negative-binomial distributions, which are just the families characterized in the adjacent case.

2. Weak records.

Let X be a random variable with a distribution concentration on a lattice of real numbers. Since the problems we are concerned with here are invariant under scale and shift transformations, without loss of generality we will restrict ourselves to the case in which the support of X is $\operatorname{supp}(X)=\{0,1,\dots,N\}$, where $N\leqq\infty$. If $N=\infty$, the symbol $\{0,1,\dots,N\}$ has to be understood as $\{0,1,2,\dots\}$. We denote $p_k=\mathbf{P}\{X=k\}$, $q_k=\mathbf{P}\{X\geqq k\}$, and $c_k=p_k/q_k$ for $k\in\{0,1,\dots,N\}$.

Let $\{X_n\}_{n\geqq 1}$ be a sequence of independent and identically distributed random variables with $X_n\stackrel{\rm d}{=}X$. Consider the sequence $\{U(i)\}_{i\geqq 1}$ defined recurrently by

$$U(1)=1,\quad U(i)=\min\bigl\{m\colon\ m>U(i-1)\ \text{and}\ X_m\geqq X_{U(i-1)}\bigr\},\qquad i\geqq 2.$$

*Received by the editors March 2, 2000.
http://www.siam.org/journals/tvp/48-4/98081.html
†Dpto. Estadistica e I.O., Universidad de Sevilla, Spain.

The random variable $R_i = X_{U(i)}$ is called the ith weak (upper) record of the sequence $\{X_n\}_{n\geqq 1}$.

The joint probability mass function (p.m.f.) of the first m weak records is

$$\mathbf{P}\{R_1 = j_1, \ldots , R_m = j_m\} = p_{j_m} \prod_{i=1}^{m} c_{j_i}, \qquad 0 \leqq j_1 \leqq \cdots \leqq j_m \leqq N \tag{2.1}$$

(obviously, if $N = \infty$, the last inequality is strict). From (2.1), it can easily be deduced that the conditional p.m.f. of R_{i+1} given $R_i = j$ is

$$\mathbf{P}[R_{i+1} = k \mid R_i = j] = \frac{p_k}{q_j}, \qquad 0 \leqq j \leqq k \leqq N. \tag{2.2}$$

Note that the conditional probability given in (2.2) does not depend on i. In the same way,

$$\mathbf{P}[R_{i+2} = k \mid R_i = j] = \frac{p_k}{q_j} \sum_{l=j}^{k} c_l, \qquad 0 \leqq j \leqq k \leqq N, \tag{2.3}$$

which is again a probability that does not depend on i. Following with this argument, it can be shown that for fixed $s > 0$, $\mathbf{P}[R_{i+s} = k \,|\, R_i = j]$ does not depend on i.

We denote $e_m(j) = \mathbf{E}[R_{i+m} \,|\, R_i = j]$, $m > 0$, $j \geqq 0$, $i \geqq 0$. So we have

$$e_1(j) = \frac{1}{q_j} \sum_{k=j}^{N} k p_k, \tag{2.4}$$

and changing summation symbols and using (2.4), we get

$$e_2(j) = \frac{1}{q_j} \sum_{k=j}^{N} k p_k \sum_{l=j}^{k} c_l = \frac{1}{q_j} \sum_{l=j}^{N} p_l e_1(l), \tag{2.5}$$

and by recurrence it can be shown that

$$e_{m+1}(j) = \frac{1}{q_j} \sum_{l=j}^{N} p_l e_m(l) \qquad \text{for any} \quad m \geqq 0 \quad \text{and} \quad j \geqq 0 \tag{2.6}$$

with $e_0(j) = j$ for all $j \geqq 0$.

Consider the following families of discrete distributions (see [4]):

(a) The negative hypergeometric distribution of the first kind (or beta-binomial), denoted by $\mathrm{nh}_I(\alpha, \beta, n)$ and with p.m.f.

$$p_k = \frac{\binom{\alpha+k-1}{k} \binom{\beta-\alpha+n-k}{n-k}}{\binom{\beta+n}{n}}, \qquad k = 0, 1, \ldots , n,$$

where α and β are real numbers such that $\beta + 1 > \alpha > 0$ and n is a natural number;

(b) the negative hypergeometric distribution of the second kind (or beta-negative-binomial), denoted by $\mathrm{nh}_{II}(\alpha, \beta, \gamma)$ and with p.m.f.

$$p_k = \frac{\gamma}{\gamma + k} \frac{\binom{\beta}{\gamma} \binom{\alpha+k-1}{k}}{\binom{\alpha+\beta+k}{k}}, \qquad k = 0, 1, 2, \ldots ,$$

where α, β, and γ are real numbers such that $\alpha > 0$, $\beta + 1 > \gamma > 0$;

(c) the geometric distribution, denoted by ge(p), with p.m.f. $p_k = (1-p)^k p$, $k = 0, 1, 2, \ldots$, where $p \in (0,1)$.

Let $\mathcal{C}_s$ be the class of discrete distributions with support on $\{0, 1, \ldots, N\}$, for certain $N \leqq \infty$, such that $\mathbf{E}[R_{i+s} \mid R_i]$ is linear. The following lemma gives a characterization of $\mathcal{C}_1$.

LEMMA 1. *Let X be a discrete distribution with support on $\{0, 1, \ldots, N\}$ such that $\mathbf{E}[R_{i+1} \mid R_i] = \gamma_0 + \gamma_1 R_i$, where γ_0 and γ_1 are some real numbers. Then γ_0 and γ_1 are positive and*

(i) *if $0 < \gamma_1 < 1$, then $\gamma_0/(1-\gamma_1)$ is a natural number and*

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

(ii) *if $\gamma_1 = 1$, then*

$$X \sim \mathrm{ge}\left(\frac{1}{1+\gamma_0}\right);$$

(iii) *if $\gamma_1 > 1$, then*

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

Proof. See the paper by Stepanov [9]. Note that in his result the case (i) is missing due to his unnecessary assumption of unboundedness of the support of X. This fact has been noted by Wesołowski and Ahsanullah [11].

Let X be a random variable such that $\mathbf{E}[R_{i+1} \mid R_i] = \gamma_0 + \gamma_1 R_i$ (i.e., X follows one of the distributions characterized in Lemma 1). Then $e_1(l) = \gamma_0 + \gamma_1 l$, for $l \in \{0, 1, \ldots, N\}$, or equivalently, from (2.4),

$$\sum_{k=l}^{N} k p_k = (\gamma_0 + \gamma_1 l)\, q_l \qquad \text{for all} \quad l = 0, 1, \ldots, N. \tag{2.7}$$

From (2.5), using (2.7), after some elementary algebra, we get

$$e_2(j) = \gamma_0(1+\gamma_1) + \gamma_1^2 j, \qquad j = 0, 1, \ldots, N,$$

and by recurrence, using (2.6), we obtain

$$e_s(j) = \mathbf{E}[R_{i+s} \mid R_i = j] = \gamma_0 \frac{1-\gamma_1^s}{1-\gamma_1} + \gamma_1^s j, \qquad j = 0, 1, \ldots, N.$$

(In order to have a compact notation, in the following we assume that $(1-\gamma_1^s)/(1-\gamma_1) = s$ if $\gamma_1 = 1$.)

Summing up, if X is a random variable such that $\mathbf{E}[R_{i+1} \mid R_i] = \gamma_0 + \gamma_1 R_i$, then, for any $s \geqq 1$, $\mathbf{E}[R_{i+s} \mid R_i]$ is also linear with

$$\mathbf{E}[R_{i+s} \mid R_i] = \beta_0 + \beta_1 R_i,$$

where $\beta_0 = \gamma_0(1-\gamma_1^s)/(1-\gamma_1)$ and $\beta_1 = \gamma_1^s$; in other words, $\mathcal{C}_1 \subseteq \mathcal{C}_s$.

Our aim is to prove the opposite inclusion. This is the main result of the following section.

3. The main result.

In order to prove that $\mathcal{C}_s \subseteq \mathcal{C}_1$ we need some previous results.

Let us denote by $\mathbf{v} = (v(0), v(1), \ldots, v(N))^t$ a generic vector in $\mathbf{R}^{N+1}$. (If $N = \infty$, $\mathbf{v}$ is a sequence of real numbers.) Let X be a random variable with support on $\{0, 1, \ldots, N\}$, $\{p_k\}_{k=0}^N$ its p.m.f., and $q_k = \mathbf{P}\{X \geqq k\}$. Let D_N be the subset of $\mathbf{R}^{N+1}$ formed by the vectors $\mathbf{v}$ such that $\sum_{k=0}^{N} |v(k)|\, p_k < \infty$. Note that, if N is finite, $D_N = \mathbf{R}^{N+1}$. Consider the linear operator defined on the elements of D_N as

$$T(\mathbf{v}) = \mathbf{w}, \qquad \mathbf{v} \in D_N, \tag{3.1}$$

where $w(j) = q_j^{-1} \sum_{k=j}^{N} v(k)\, p_k$, $j = 0, 1, \dots, N$.

The matrix (an infinite matrix if $N = \infty$) associated to the linear operator defined in (3.1) is

$$A = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots & p_N \\ 0 & p_1/q_1 & p_2/q_1 & \cdots & p_N/q_1 \\ 0 & 0 & p_2/q_2 & \cdots & p_N/q_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

which is an upper triangular matrix. If N is finite, the determinant of A is $\det A = \prod_{k=0}^{N}(p_k/q_k) > 0$, and this implies that the linear operator $T\colon \mathbf{R}^{N+1} \to \mathbf{R}^{N+1}$ is bijective. In the case $N = \infty$, the condition $p_k/q_k \neq 0$ for all $k \geqq 0$ implies that T is a bijective operator that maps D_∞ onto its image.

If $\mathbf{e}_m$ denotes the vector with components $e_m(j) = \mathbf{E}[R_{i+m} \,|\, R_i = j]$, $j = 0, 1, \dots, N$, from (2.6) we have $T(\mathbf{e}_m) = \mathbf{e}_{m+1}$ and, by recurrence,

$$\mathbf{e}_s = T^{s-1}(\mathbf{e}_1) \qquad \text{for all} \quad s > 1, \tag{3.2}$$

where $T^k = T \circ \cdots \circ T$ (k times).

THEOREM 1. *Let X be a discrete distribution with support on $\{0, 1, \dots, N\}$ such that*

$$\mathbf{E}[R_{i+s} \mid R_i] = \beta_0 + \beta_1 R_i, \tag{3.3}$$

where β_0 and β_1 are some real numbers and $s \geqq 1$. Then β_0 and β_1 are positive. Let γ_0 and γ_1 be unique positive solutions to the equations

$$\beta_1 = \gamma_1^s, \qquad \beta_0 = \gamma_0 \frac{1 - \gamma_1^s}{1 - \gamma_1}. \tag{3.4}$$

Then

(i) *if $0 < \beta_1 < 1$, then $\gamma_0/(1 - \gamma_1)$ is a natural number and*

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

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

$$X \sim \mathrm{ge}\left(\frac{1}{1 + \gamma_0}\right);$$

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

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

Proof. Observe that as X is nondegenerate, $e_s(l) = \mathbf{E}[R_{i+s} \,|\, R_i = l]$ is strictly increasing, so that β_1 must be positive. Also note that $\beta_0 = e_s(0)$; then β_0 is positive. Let γ_0 and γ_1 be unique positive solutions to (3.4). For $m = 1, \dots, s$, let us write

$$e_m(j) = \gamma_0 \frac{1 - \gamma_1^m}{1 - \gamma_1} + \gamma_1^m j + d_m(j), \qquad j = 0, 1, \dots, N. \tag{3.5}$$

Obviously, from (3.3), $d_s(j) = 0$ for $j = 0, 1, \dots, N$. Combining (2.6) and (3.5), after some algebra, we obtain

$$d_{m+1}(j) = \gamma_1^m d_1(j) + \frac{1}{q_j} \sum_{l=j}^{N} p_l d_m(l), \qquad j = 0, 1, \dots, N, \tag{3.6}$$

for $m = 1, \dots, s - 1$.

Consider the $(N+1)\times(N+1)$ matrix (an infinite matrix if $N=\infty$)

$$A = \begin{bmatrix} p_0 & p_1 & p_2 & \cdots & p_N \\ 0 & p_1/q_1 & p_2/q_1 & \cdots & p_N/q_1 \\ 0 & 0 & p_2/q_2 & \cdots & p_N/q_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and the $(N+1)$-dimensional vectors (infinite sequences if $N=\infty$)

$$\mathbf{d}_m = \big(d_m(0), d_m(1), \ldots, d_m(N)\big)^t, \qquad m = 1, \ldots, s.$$

With this notation, (3.6) can be written in the matrix form as $\mathbf{d}_{m+1} = \gamma_1^m \mathbf{d}_1 + A\mathbf{d}_m$, $m = 1, \ldots, s-1$, from which we get

$$\text{(3.7)} \qquad \mathbf{d}_m = B_m \mathbf{d}_1, \qquad m = 1, \ldots, s, \qquad \text{with} \quad B_m = \sum_{k=0}^{m-1} \gamma_1^{m-k-1} A_k.$$

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

$$\text{(3.8)} \qquad \mathbf{d}_1 = B_m^{-1}\mathbf{d}_m \qquad \text{for} \quad m = 1, \ldots, s,$$

by hypothesis $\mathbf{d}_s = 0$, and then from (3.8) we get $\mathbf{d}_1 = \mathbf{0}$, or equivalently, from (3.5)

$$\mathbf{E}[R_{i+1} \mid r_i = j] = \gamma_0 + \gamma_1 j, \qquad j = 0, 1, \ldots, N,$$

and the conclusion of the theorem follows immediately from Lemma 1.

REFERENCES

[1] M. Ahsanullah and J. Wesołowski, *Linearity of best predictors for nonadjacent record values*, Sankhyā Ser. B, 60 (1998), pp. 221–227.

[2] F. A. Aliev, *Characterization of distributions through weak records*, J. Appl. Statist. Sci., 8 (1998), pp. 13–16.

[3] A. Dembińska and J. Wesołowski, *Linearity of regression for nonadjacent records values*, J. Statist. Plann. Inference, 90 (2000), pp. 195–205.

[4] N. L. Johnson, S. Kotz, and A. Kemp, *Univariate Discrete Distributions*, Wiley, New York, 1992.

[5] F. López-Blázquez, *Caracterización de Distribuciones Mediante Valores Esperados de Estadísticos de Orden y Records*, Ph. D. Dissertation, Universidad de Sevilla, Sevilla, Spain, 1990.

[6] F. López-Blázquez and J. L. Moreno-Rebollo, *A characterization of distributions based on linear regression of order statistics and record values*, Sankhyā Ser. A, 59 (1997), pp. 311–323.

[7] H. N. Nagaraja, *On a characterization based on record values*, Austral. J. Statist., 19 (1977), pp. 70–73.

[8] H. N. Nagaraja, *Some characterizations of continuous distributions based on regressions of adjacent order statistics and record values*, Sankhyā Ser. A, 50 (1988), pp. 70–73.

[9] A. V. Stepanov, *A characterization theorem for weak records*, Theory Probab. Appl., 38 (1993), pp. 762–764.

[10] W. Vervaat, *Limit theorems for records from discrete distributions*, Stochastic Process. Appl., 1 (1973), pp. 317–334.

[11] J. Wesołowski and M. Ahsanullah, *Linearity of regression for nonadjacent weak records*, Statist. Sinica, 11 (2001), pp. 39–52.