



# Characterization of the Geometric Distribution Via Linear Combinations of Observations and of Records

Barry C. Arnold

*University of California Riverside, Riverside, USA*

Jose A. Villasenor

*Colegio de Postgraduados, Montecillo, Mexico*

---

## Abstract

In a sequence of independent identically distributed geometric random variables, the sum of the first two record values is distributed as a simple linear combination of geometric variables. It is verified that this distributional property characterizes the geometric distribution. A related characterization conjecture is also discussed. Related discussion in the context of weak records is also provided.

*AMS (2000) subject classification.* Primary 60E05; Secondary 62E10.

*Keywords and phrases.* Upper records, convolution, record spacings, weak records.

---

## 1 Introduction

Consider a sequence of independent identically distributed (i.i.d.) positive integer valued random variables  $\{X_n\}_{n=1}^{\infty}$ . Denote the corresponding sequence of upper records by  $\{X^{(n)}\}_{n=1}^{\infty}$ . Specifically, the first random variable in the sequence is identified as the first record, the second record is the first subsequent  $X_n$  which exceeds  $X_1$ . It is well known that the record value sequence corresponding to a sequence of geometric random variables has a simple distributional structure. If we define the record spacings sequence  $\{S_n\}_{n=1}^{\infty}$  by  $S_1 = X_1 = X^{(1)}$  and for  $n > 1$ ,  $S_n = X^{(n)} - X^{(n-1)}$ , then in the geometric case these spacings are independent random variables. Geometric characterizations based on the independence of the record spacings are well known. In the present paper we will consider a simple relationship between the distribution of the first two records and the distribution of the first two  $X_n$ 's. Two related conjectured characterizations are described. In addition parallel results are discussed in the case of weak records.

## 2 The Conjectured Characterizations

Consider a sequence of i.i.d. positive integer valued random variables  $\{X_n\}_{n=1}^\infty$  with corresponding upper record sequence  $\{X^{(n)}\}_{n=1}^\infty$ . If the  $X_i$ 's have a common geometric distribution, then because the record spacings are themselves geometrically distributed with homogeneous success probabilities, it follows that

$$X^{(1)} + X^{(2)} \stackrel{d}{=} X_1 + 2X_2. \tag{2.1}$$

After formulating this unusual relationship between the two sequences,  $\{X_n\}_{n=1}^\infty$  and  $\{X^{(n)}\}_{n=1}^\infty$ , it becomes plausible that this is a characteristic property of the geometric distribution. Two conjectures were considered.

**CONJECTURE 1.** *Suppose that  $X^{(1)} + X^{(2)} \stackrel{d}{=} X_1 + 2X_2$ , then  $p_x = P(X = x) = p(1 - p)^{x-1}$ , for each  $x = 1, 2, \dots$  for some  $p \in (0, 1)$ .*

**CONJECTURE 2.** *Suppose that, for some positive integer  $m > 2$ ,  $\sum_{i=1}^m X^{(i)} \stackrel{d}{=} \sum_{i=1}^m iX_i$ , then  $p_x = P(X = x) = p(1 - p)^{x-1}$ , for each  $x = 1, 2, \dots$  for some  $p \in (0, 1)$ .*

Both conjectures are judged to be plausible. Conjecture 2 would appear to be more difficult to resolve. In the next section we will provide a proof of Conjecture 1 under no regularity conditions. A proof of Conjecture 2 remains elusive.

## 3 Proof of Conjecture 1

Throughout this section we will employ the usual convention, when convenient, of denoting  $1 - p$  by  $q$  and denoting  $1 - p_x$  by  $q_x$ .

**THEOREM 1.** *If  $\{X_n\}_{n=1}^\infty$  are i.i.d. positive integer valued random variables with common discrete density function  $f(x) = p_x, x = 1, 2, \dots$  where  $p_x > 0 \forall x$  so that a record value sequence is well-defined, and if  $X^{(1)} + X^{(2)} \stackrel{d}{=} X_1 + 2X_2$ , then  $p_x = p(1 - p)^{x-1}, x = 1, 2, \dots$  for some  $p \in (0, 1)$ .*

**PROOF.** First note that set of possible values of  $X_1 + 2X_2$  and of  $X^{(1)} + X^{(2)}$  is the set  $\{3, 4, 5, \dots\}$ .

**Necessity** It is well-known that if the  $X_i$ 's are i.i.d. with a common Geometric ( $p$ ) distribution, then the record spacings  $X^{(m)} - X^{(m-1)}$  are also i.i.d. with a common geometric ( $p$ ) distribution. Since we can write  $X^{(1)} + X^{(2)} = (X^{(2)} - X^{(1)}) + 2X^{(1)}$ , the result follows.

**Sufficiency** As in the statement of the theorem we have  $P(X = x) = p_x, x = 1, 2, \dots$

Assuming that  $X^{(1)} + X^{(2)} \stackrel{d}{=} X_1 + 2X_2$ , we wish to prove that  $p_x = pq^{x-1}$ ,  $x = 1, 2, \dots$ . First note that

$$P(X_1 + 2X_2 = 3) = p_1p_1,$$

while

$$P(X^{(1)} + X^{(2)} = 3) = p_1 \frac{p_2}{q_1}.$$

Equating these expressions we may conclude that  $p_2 = p_1q_1$ . For simplicity of notation we will denote  $p_1$  by  $p$ . Thus far we have shown that  $p_1 = p = pq^{1-1}$  and  $p_2 = pq = pq^{2-1}$ . We now argue inductively. Suppose that for some positive even integer  $2k$ , we have  $p_j = pq^{j-1}$  for every  $j \leq 2k$ , we claim that in such a case because of Eq. 2.1, we will also have  $p_{2k+1} = pq^{2k+1-1}$ . To see this, consider

$$\begin{aligned} P(X_1 + 2X_2 = 2k + 2) &= \sum_{j=1}^k P(X_2 = j, X_1 = 2k + 2 - 2j) \\ &= \sum_{j=1}^k pq^{j-1}, pq^{2k+1-2j} \\ &= p^2 \sum_{j=1}^k q^{2k-j}, \end{aligned}$$

and

$$\begin{aligned} P(X^{(1)} + X^{(2)} = 2k + 2) &= \sum_{j=1}^k P(X^{(1)} = j, X^{(2)} = 2k + 2 - j) \\ &= \sum_{j=2}^k pq^{j-1} \frac{pq^{2k+1-j}}{q^j} + p \frac{p_{2k+1}}{q} \\ &= p^2 \sum_{j=2}^k q^{2k-j} + pp_{2k+1}/q. \end{aligned}$$

Since (2.1) holds, we may conclude that

$$pp_{2k+1}/q = p^2q^{2k-1}.$$

which implies that  $p_{2k+1} = pq^{2k} = pq^{(2k+1)-1}$ , as claimed.

A similar argument will show that if for some positive odd integer  $2k - 1$ , we have  $p_j = pq^{j-1}$  for every  $j \leq 2k - 1$ , then because of Eq. 2.1, we will also

have  $p_{2k} = pq^{2k-1}$ . For this, it is necessary to equate  $P(X_1 + 2X_2 = 2k + 1)$  and  $P(X^{(1)} + X^{(2)} = 2k + 1)$ .

It then follows by induction that  $p_x = pq^{x-1}$  for every  $x = 1, 2, \dots$ , i.e. that  $X$  has a geometric( $p$ ) distribution.

#### 4 Discussion Regarding Conjecture 2

The proof of Theorem 1 was less transparent than was expected. Although Conjecture 2 is eminently plausible, the book-keeping necessary to prove the result appears to be daunting and the conjecture remains open. However, if we consider the case in which  $m = 3$ , we may argue that the conjecture appears to be unlikely to be true based on the following observations.

The possible values of  $X_1 + 2X_2 + 3X_3$  and of  $X^{(1)} + X^{(2)} + X^{(3)}$  are  $\{6, 7, 8, \dots\}$ . If we assume that  $P(X_1 + 2X_2 + 3X_3 = 6) = P(X^{(1)} + X^{(2)} + X^{(3)} = 6)$ , this implies that

$$p_1^3 = p_1 \frac{p_2}{1 - p_1} \frac{p_3}{1 - p_1 - p_2},$$

from which we obtain

$$p_3 = \frac{p_1^2(1 - p_1)(1 - p_1 - p_2)}{p_2}.$$

Thus  $p_1$  and  $p_2$  appear to be unconstrained, except that their sum must be less than 1.

If we consider other possible values, i.e., consider equalities of the form

$$P(X_1 + 2X_2 + 3X_3 = y) = P(X^{(1)} + X^{(2)} + X^{(3)} = y),$$

then each new value of  $y$  will result in an expression for  $p_y$  in terms of  $p_1, p_2, \dots, p_{y-1}$ . However no obvious constraints on  $p_1$  or  $p_2$  appear to arise.

Of course, if  $p_2 = p_1(1 - p_1)$  then subsequent  $p_j$ 's appear to be of the geometric form (i.e.,  $= p_1(1 - p_1)^{j-1}$ ). However, other choices for  $p_2$  would seem to lead to non-geometric solutions.

Cases in which  $m > 3$ , exhibit similar problems and, in fact, would appear to admit an even wider variety of non-geometric solutions. It appears that only in the case  $m = 2$  is a characterization possible.

REMARK 1. We have carefully avoided stating that non-geometric solutions will exist in cases in which  $m > 2$ , because we have been unable to explicitly determine completely a convergent non-geometric sequence that satisfies the condition  $\sum_{i=1}^m X^{(i)} \stackrel{d}{=} \sum_{i=1}^m iX_i$ .

### 5 An Analogous Weak Record Result

When we turn to investigate record phenomena for sequences of i.i.d. non-negative integer valued random variables, the concept of weak records plays the role usually played by records. An observation in the sequence  $\{X_i\}_{i=1}^\infty$  is a weak record if it exceeds or equals all the preceding  $X_i$ 's in the sequence. In this setting geometric random variables with possible values  $\{0, 1, 2, \dots\}$  play a role analogous to that played by positive geometric variables in record value discussions. In this Section we will add asterisks to non-negative integer random variables and corresponding weak records to distinguish them from the positive random variables and ordinary records discussed in the previous Sections.

We thus will consider a sequence  $\{X_i^*\}_{i=1}^\infty$  of non-negative random variables with a corresponding weak record sequence denoted by  $\{X^{*(i)}\}_{i=1}^\infty$ . (an introduction to weak records can be found in Arnold et al. (1998)). We will say that a non-negative integer valued random variable  $X^*$  has a *geometric\** distribution if its discrete density is of the form  $P(X^* = k) = p(1 - p)^k$ ,  $k = 0, 1, 2, \dots$  and we write  $X^* \sim geo^*(p)$ . Parallel to the result for positive geometric variables, it is well-known that the weak record spacings corresponding to *geometric\**( $p$ ) are themselves i.i.d. with a common *geometric\**( $p$ ) distribution. It is consequently plausible that the following result, analogous to Theorem 1, might be true (this was suggested by a referee). The proof is a close parallel to the proof for ordinary (i.e., positive) geometric variables.

**THEOREM 2.** *If  $\{X_n^*\}_{n=1}^\infty$  are i.i.d. non-negative integer valued random variables with common discrete density function  $f(x) = p_x$ ,  $x = 0, 1, 2, \dots$  where  $p_x > 0 \forall x$  so that a weak record value sequence is well-defined, and if  $X^{*(1)} + X^{*(2)} \stackrel{d}{=} X_1^* + 2X_2^*$ , then  $p_x = p(1 - p)^x$ ,  $x = 0, 1, 2, \dots$  for some  $p \in (0, 1)$ .*

**PROOF.** First note that set of possible values of  $X_1^* + 2X_2^*$  and of  $X^{*(1)} + X^{*(2)}$  is the set  $\{0, 1, 2, \dots\}$ .

**Necessity** We use the fact that if the  $X_i^*$ 's are i.i.d. with a common *geometric\**( $p$ ) distribution, then the record spacings  $X^{*(m)} - X^{*(m-1)}$  are also i.i.d. with a common *geometric\**( $p$ ) distribution. Since we can write  $X^{*(1)} + X^{*(2)} = (X^{*(2)} - X^{*(1)}) + 2X^{*(1)}$ , the result follows.

**Sufficiency** As in the statement of the theorem we have  $P(X = x) = p_x$ ,  $x = 0, 1, 2, \dots$ , however it will be convenient to denote  $p_0$  by  $p \in (0, 1)$ .

For convenience we define  $V = X_1^* + 2X_2^*$  and  $W = X^{*(1)} + X^{*(2)}$ . Under the assumption that  $V \stackrel{d}{=} W$  we wish to prove that  $p_k = p(1 - p)^k$   $k = 0, 1, 2, \dots$  where  $p = P(X_1^* = 0)$ . Elementary computations yield the following expressions for the discrete densities of  $V$  and  $W$ , in which we use the notation  $q_j = P(X_1^* \geq j)$ .

$$\text{For } k \text{ odd, } P(V = k) = \sum_{j=0}^{(k-1)/2} p_j p_{k-2j}, \tag{5.1}$$

$$\text{For } k \text{ odd, } P(W = k) = \sum_{j=0}^{(k+1)/2} p_j p_{k-j} / q_j, \tag{5.2}$$

$$\text{For } k \text{ even, } P(V = k) = \sum_{j=0}^{k/2} p_j p_{k-2j}, \tag{5.3}$$

$$\text{For } k \text{ even, } P(W = k) = \sum_{j=0}^{k/2} p_j p_{k-j} / q_j. \tag{5.4}$$

Since  $V \stackrel{d}{=} W$ , we can equate (5.1) and (5.2) when  $k=2$  and conclude that  $p_1 = p(1 - p)$ . Next consider an arbitrary  $k > 2$  and assume that, for  $j < k - 1$ , it has been verified that  $p_j = p(1 - p)^j$  and  $q_j = (1 - p)^j$ . Then by equating (5.1) and (5.2), if  $k$  is odd, or by equating (5.3) and (5.4), if  $k$  is even, we may conclude that  $p_{k-1} = p(1 - p)^{k-1}$ . We may thus, by induction, conclude that  $P(X_1^* = k) = p_k = p(1 - p)^k$ ,  $k = 0, 1, 2, \dots$ , i.e., that  $X^{**1} \sim geo^*(p)$ .

### 6 Closing Observations

Conjecture 2 continues to be tantalizing. Our arguments in Section 4 strongly suggest that it will not prove to be true. One might try to use simulations to compare the distributions of  $X_1 + 2X_2$  and of  $X^{(1)} + X^{(2)}$  using a particular non-geometric distributions for the  $X_i$ 's. However, it is highly unlikely that any well-known choice for the distribution of the  $X_i$ 's will result in the desired equi-distribution of the two statistics. We believe that the best hope for resolving the problem lies in identifying a convergent non-geometric discrete density as outlined at the end of Section 4.

*Acknowledgments.* We are grateful for the careful reading and helpful suggestions provided by the reviewers of an earlier version of this paper. In particular, the suggestion that weak records might be considered led to the discussion in Section 5.

*Funding.* Funding for Jose Villasenor was provided by the Colegio de Postgraduados, Montecillo, Mexico.

*Compliance with Ethical Standards.* The authors declare that they have no conflict of interest

*Open Access.* This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

ARNOLD, B.C., BALAKRISHNAN, N. and NAGARAJA, H.N. (1998). *Records*. Wiley, New York.

*Publisher's Note.* Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

BARRY C. ARNOLD  
DEPARTMENT OF STATISTICS,  
UNIVERSITY OF CALIFORNIA RIVERSIDE,  
RIVERSIDE, CA, USA  
E-mail: barry.arnold@ucr.edu

JOSE A. VILLASENOR  
DEPARTMENT OF STATISTICS,  
COLEGIO DE POSTGRADUADOS,  
MONTECILLO, MEXICO

Paper received: 13 February 2020; accepted 18 November 2021.