Ann. Inst. Statist. Math.
Vol. 41, No. 4, 661-669 (1989)

SOME PROPERTIES OF RECORD VALUES COMING FROM THE GEOMETRIC DISTRIBUTION

A. C. DALLAS

Section of Statistics, Department of Mathematics, Athens University, Panepistemiopolis, Athens 157 81, Greece

(Received January 29, 1987; revised October 20, 1988)

Abstract. The independence between spacings of record values and of individual record values is well known when the records come from a geometric distribution. Here we examine the form a function of two record values must have if we require independence from a lower record value. Also similar questions are examined in relation to conditional expectation and conditional variance.

Key words and phrases: Geometric distribution, record values, independence, conditional expectation, conditional variance, power series.

1. Introduction

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables on the nonnegative integers, each with probability distribution $P(i) = P(X = i)$, $i = 0, 1, \ldots$. The sequence defined by $N(0) = 1$ and $N(n) = \min\{j: X_j > X_{N(n-1)}\}$ for $n = 1, 2, \ldots$ is called the sequence of (upper) record times, while the corresponding sequence $R_n = X_{N(n)}$, $n = 0, 1, \ldots$ is called the sequence of (upper) records. The joint probability distribution of $R_0, R_1, \ldots, R_n$ is given by

$$(1.1) \qquad P(R_0 = r_0,\ R_1 = r_1, \ldots,\ R_n = r_n) = \prod_{i=0}^{n} P(r_i) \Big/ \prod_{i=0}^{n-1} Q(r_i)\,,$$

$r_0 < r_1 < \cdots < r_n$ where $Q(r_i) = P(X > r_i)$.

Let X be a random variable on the nonnegative integers with probability distribution given by

$$(1.2) \qquad P(X = i) = pq^i, \quad i = 0, 1, \ldots, \quad q \in (0, 1), \quad p + q = 1\,.$$

This is the geometric distribution with parameter q, and with left end zero. Substituting (1.2) in (1.1), we find that R_0, $R_1 - R_0, \ldots, R_n - R_{n-1}$ are independent and that $R_i - R_{i-1}$, $i = 1, 2, \ldots, n$ have the same distribution as

$R_0 + 1$. Some direct consequences of these properties are: R_i and $R_n - R_j$, for $i \leq j \leq n$, are independent, $R_n - R_j$ has a constant conditional expectation and a constant conditional variance given R_i, and $R_n - R_j$ has the negative binomial distribution with parameter q and left end $n - j$. Some forms of these properties have been used to characterize (1.2) via order statistics properties, e.g., Pfeifer (1979), Deheuvels (1984) and Gupta (1984). Some distributional results in the discrete case are also given by Ahsanullah and Holland (1984) and several related questions are examined in a very general setup by Rao and Shanbhag (1986). Further information is found in the papers cited in the above papers.

In this paper, we examine the following problem: Suppose we are given an arbitrary function $g(R_j, R_n)$, when R_j and R_n are coming from (1.1), and assume that some of the above-mentioned properties hold for it. What, then, would be the form of $g(R_j, R_n)$? It is proved that, under some conditions, $g(R_j, R_n)$ would be either $h(R_n - R_j)$ or $R_n - R_j$ in some appropriate region, the function $h(\cdot)$ being arbitrary.

2. The results

Let $j < n$ be two fixed nonnegative integers and let R_j, R_n be the corresponding record values of the geometric distribution (1.2). The set of values taken by (R_j, R_n) is the set of pairs of integers $D = \{(r_j, r_n)\colon r_j = j, j+1, \ldots, r_n = r_j + n - j\}$. We start with a result related to the distribution of spacings of records.

THEOREM 2.1. *Let R_j and R_n, $j < n$, be two record values coming from the geometric distribution* (1.2) *and let $g(x, y)$ be a function strictly increasing in y. We assume that $g(R_j, R_n)$ has a negative binomial distribution with parameter q, with left end $n - j$ and q being fixed. Then $g(x, y) = x - y$ on D and it is arbitrary elsewhere.*

PROOF. Let $C_z = \{(x, y)$ such that $g(x, y) = z\}$ for $z = n - j, n - j + 1, \ldots$ be a partition of the set D. According to the assumption, we have

$$\sum_{(x,y)\in C_z} P(R_j = x,\ R_n = y) = \binom{z-1}{n-j-1} p^{n-j} q^{z-n+j}, \qquad z = n-j, n-j+1, \ldots . \tag{2.1}$$

In the above equation, we use the independence of R_j and $R_n - R_j$. After some calculations, equation (2.1) is transformed into

$$\sum_{(x,y)\in C_z} \binom{x}{j}\binom{y-x-1}{n-j-1} q^{y-n} = \binom{z-1}{n-j-1} p^{-j-1} q^{z-n+j}, \tag{2.2}$$

$$z = n - j, n - j + 1, \ldots .$$

Expanding p^{-j-1} in equation (2.2), we obtain

$$\sum_{(x,y)\in C_z} \binom{x}{j}\binom{y-x-1}{n-j-1} q^{y-n} = \sum_{x=j}^{\infty} \binom{x}{j}\binom{z-1}{n-j-1} q^{x+z-n}, \tag{2.3}$$

$$z = n - j, n - j + 1, \ldots .$$

Consider now equation (2.3) for $z = n - j$. It takes the form

$$\sum_{(x,y)\in C_{n-j}} \binom{x}{j}\binom{y-x-1}{n-j-1} q^{y-n} = \sum_{x=j}^{\infty} \binom{x}{j} q^{x-j}. \tag{2.4}$$

Suppose that there exists $(x_0, y_0) \in C_{n-j}$ such that $y_0 > x_0 + n - j$. As $g(x,y)$ is strictly increasing in y, we have $n - j = g(x_0, y_0) > g(x_0, x_0 + n - j)$. But $g(x_0, x_0 + n - j)$ is an admissible value of $g(R_j, R_n)$, therefore, $g(x_0, x_0 + n - j) \geq n - j$ and this is a contradiction. Hence, $(x,y) \in C_{n-j}$ implies $y = x + n - j$. Equation (2.4) implies the converse, hence one obtains $(x,y) \in C_{n-z}$ iff $y = x + n - j$. Working in exactly the same way for the other values of z, we find inductively $(x,y) \in C_z$ iff $y = x + z$ for $z = n - j, n - j + 1, \ldots$. This implies that $g(x,y) = y - x$ on D with an arbitrary extension elsewhere. This concludes the proof of the theorem.

In the previous theorem it was assumed that $g(x,y)$ is strictly increasing in y. This is not redundant if no further assumptions are imposed on j and n. We can see that, if $n = 2j + 1$, there are functions $g(x,y) \neq y - x$ which are not strictly increasing in y and at the same time, $g(R_j, R_n)$ has a negative binomial distribution with left end $n - j$. Suppose that $g(x,y)$ is nondecreasing in y; then either $g(r_j, r_n) = r_j + 1$ on D or $g(r_j, r_n) = r_n - r_j$ on D with $r_n \neq n + 1$ and is equal to $r_n + r_j - n$ for $r_n = n + 1$ are two examples where $g(R_j, R_n)$ follows the required negative binomial distribution. If nothing is assumed about $g(x,y)$, then $g(r_j, r_n) = r_n - r_j$ on D for $r_n \neq n + 2$ and is equal to $r_n + r_j - n - 1$ for $r_n = n + 2$ is an example of a function $g(x,y)$ not monotonous in y and at the same time, $g(R_j, R_n)$ has the required property. Further examples can be constructed in a similar way. Finally, we remark that in Theorem 2.1, if we further impose the condition that the assumptions hold for all $q \in (0,1)$, the result of the theorem remains the same while the counterexamples are still valid.

Now we proceed to examine the consequences of the independence assumption. We know that R_i and $R_n - R_j$ are independent for $i \leq j \leq n$, and this implies the independence of $h(R_n - R_j)$ and R_i, where $h(\cdot)$ is any function. We next prove that this is the only function of R_j and R_n for which this holds.

THEOREM 2.2. *Let $g(x,y)$ be a function such that R_i and $g(R_j, R_n)$ are independent for some fixed nonnegative integers, $i \le j \le n$ and for all $q \in (0,1)$. Then, $g(x,y) = h(y-x)$ on D and arbitrary elsewhere, where $h(\cdot)$ is any function.*

PROOF. The independence of $g(R_j, R_n)$ and R_i implies

$$E(s^{g(R_j,R_n)}t^{R_i}) = E(t^{R_i})E(s^{g(R_j,R_n)}), \tag{2.5}$$

for $|s| < 1$, $|t| < 1$. But

$$\begin{aligned} &P(R_n = r_n,\ R_j = r_j,\ R_i = r_i) \\ &\quad = P(R_n - R_j = r_n - r_j)P(R_j - R_i = r_j - r_i)P(R_i = r_i), \end{aligned}$$

for $r_n - r_j \ge n - j$, $r_j - r_i \ge j - i$ and $r_i \ge i$. Each difference and R_i have a negative binomial distribution with appropriate parameters. Now in equation (2.5) we set $r_i - i = x$, $r_j - j = y$ and $r_n - n = z$. Then after some algebra, we obtain

$$\begin{aligned} &\sum_{x=0}^{\infty}\sum_{y=x}^{\infty}\sum_{z=y}^{\infty}\binom{z-y+n-j-1}{n-j-1}\binom{y-x+j-i-1}{j-i-1} \\ &\qquad \cdot \binom{i+x}{x}p^{n+1}q^{z}s^{g(y+j,z+n)}t^{x} \\ &\quad = E(s^{g(R_j,R_n)})\sum_{x=0}^{\infty}\binom{i+x}{x}\cdot p^{i+1}q^{x}t^{x}, \end{aligned} \tag{2.6}$$

for $|s|, |t| < 1$. Equating coefficients of t^x in (2.6), we find

$$\begin{aligned} &\sum_{y=x}^{\infty}\sum_{z=y}^{\infty}\binom{z-y+n-j-1}{n-j-1}\binom{y-x+j-i-1}{j-i-1}q^{z-x}s^{g(y+j,z+n)} \\ &\quad = a(s,p) \quad \text{for} \quad x = 0,1,\ldots, \quad |s| < 1. \end{aligned} \tag{2.7}$$

In equation (2.7), we set $y - x = u$, $z - y = v$ and then subtract the equation corresponding to $x = 0$ from it. The resulting equation is

$$\begin{aligned} &\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\binom{v+n-j-1}{n-j-1}\binom{u+j-i-1}{j-i-1}q^{u+v} \\ &\qquad \cdot (s^{g(u+x+j,u+v+x+n)} - s^{g(u+j,u+v+n)}) = 0 \\ &\qquad\qquad \text{for} \quad x = 0,1,\ldots, \quad |s| < 1, \quad |q| < 1. \end{aligned} \tag{2.8}$$

Setting $m = u + v$ in equation (2.8), and working as before, we have

$$\sum_{v=0}^{m} \binom{v+n-j-1}{n-j-1}\binom{m-v-j-i-1}{j-i-1} \cdot \left(s^{g(m-v+x+j,m+x+n)} - s^{g(m-v+j,m+n)}\right) = 0 ,$$

$$x, m = 0, 1, \ldots, \quad |s| < 1 .$$

We consider the equation which corresponds to $m = 0$ and then use induction. This gives

$$g(x+j, m+x+n) = g(j, m+n) \quad \text{for} \quad x, m = 0, 1, \ldots .$$

Change now $m + x + n$ to y and $x + j$ to x. Substitution gives

$$g(x,y) = g(j, y - x + j) = h(y - x) \quad \text{on} \quad D ,$$

with an arbitrary extension elsewhere. This concludes the proof of the theorem.

The independence of R_i and $R_n - R_j$, $1 \leq i \leq j \leq n$ for the geometric records implies that the conditional expectation of any function of $R_n - R_j$, given $R_i = r_i$, does not depend on r_i. We prove now that the constancy of the conditional expectation leads to the same conclusion as in Theorem 2.2.

THEOREM 2.3. *Let $g(x,y)$ be any function such that $g(R_j, R_n)$ has a finite expectation. Suppose that for all $q \in (0,1)$*

$$E(g(R_j, R_n) \mid R_i = r_i) = d , \tag{2.9}$$

const. all $r_i = i, i+1, \ldots$, where $i \leq j \leq n$ are fixed. Then $g(x,y) = h(y-x)$ on D and it is arbitrary elsewhere.

PROOF. From the properties of record values, we have

$$P(R_n = r_n, R_j = r_j \mid R_i = r_i) = P(R_n - R_j = r_n - r_j) P(R_j - R_i = r_j - r_i) ,$$

for $r_n - r_j \geq n - j$, $r_j - r_i \geq j - i$, $r_i \geq i$. Making use of the above relation in (2.9) and using arguments similar to the ones used in the proof of Theorem 2.2, we arrive at the required result.

For the conditional variance we need some restrictions on the form of

$g(x,y)$ to get a result analogous to the previous ones. We have the following theorem:

THEOREM 2.4. *Let $g(x,y)$ be a function monotonous in y and such that $E(g^2(R_i, R_n))$ is finite for some fixed $i < n$. Then*

$$V(g(R_i, R_n) \mid R_i = r_i) = \upsilon \,, \tag{2.10}$$

υ const., all $r_i = i, i+1, \ldots$ for all $q \in (0,1)$, where $V(X)$ denotes the variance of X, implies $g(x,y) = a(y-x) + b(x)$ on D and arbitrary otherwise where $a(x)$ is monotonous and $b(x)$ is any function.

PROOF. For the conditional distribution, we have

$$P(R_n = r_n \mid R_i = r_i) = \binom{r_n - r_i - 1}{n - i - 1} p^{n-i} q^{r_n - r_i - n + i} \,, \tag{2.11}$$

$r_i = i, i+1, \ldots,\ r_n \geq r_i + n - i$. Using (2.11) in (2.10) and introducing the notation $r_i - i = x$, $r_n - r_i - n + i = z$, $n - i - 1 = k$ and $g(x+i, y+n) = h(x,y)$, equation (2.10) is written as

$$\sum_{z=0}^{\infty} h^2(x, x+z) \binom{z+k}{k} p^{n-i} q^z - \left(\sum_{z=0}^{\infty} h(x, x+z) \binom{z+k}{k} p^{n-i} q^z \right)^2 = \upsilon \,, \tag{2.12}$$

all $x = 0, 1, \ldots,\ q \in (0,1)$. In (2.12), set $x = 0$ and then subtract the resulting equation from (2.12). This gives

$$\sum_{z=0}^{\infty} (h^2(x, x+z) - h^2(0,z)) \binom{z+k}{k} q^z p^{-n+i} = \left(\sum_{z=0}^{\infty} h(x, x+z) \binom{z+k}{k} q^z \right)^2 - \left(\sum_{z=0}^{\infty} h(0,z) \binom{z+k}{k} q^z \right)^2 , \tag{2.13}$$

all $x = 0, 1, \ldots,\ q \in (0,1)$. In equation (2.13), we expand $p^{-n+i} = (1-q)^{-k-1}$. After the calculations have been made, we arrive at the following equation

$$\sum_{m=0}^{\infty} \left(\sum_{z=0}^{m} (h^2(x, x+z) - h^2(0,z)) \binom{z+k}{k} \binom{m-z+k}{k} \right) q^m$$
$$= \sum_{m=0}^{\infty} \left(\sum_{z=0}^{m} (h(x, x+z) h(x, x+m-z) \right.$$

$$- h(0,z)h(0, m - z))\binom{z+k}{k}\binom{m-z+k}{k}\Bigg) q^m ,$$

$x = 0, 1, \ldots$, all $q \in (0, 1)$ and equating coefficients of equal power, we have

$$\sum_{z=0}^{m} h(x, x+z)(h(x, x+z) - h(x, x+m-z))\binom{z+k}{k}\binom{m-z+k}{k}$$
$$= \sum_{z=0}^{m} h(0,z)(h(0,z) - h(0, m-z))\binom{z+k}{k}\binom{m-z+k}{k},$$

all $m, x = 0, 1, \ldots$. Now change the variable z to $m - z$ and add the resulting equation to the one above. By this, we find

$$\text{(2.14)} \qquad \sum_{z=0}^{m} (h(x, x+z) - h(x, x+m-z))^2 \binom{z+k}{k}\binom{m-z+k}{k}$$
$$= \sum_{z=0}^{m} (h(0,z) - h(0, m-z))^2 \binom{z+k}{k}\binom{m-z+k}{k},$$

all $m, z = 0, 1, \ldots$. From equation (2.14) for $m = 1$, using the monotonicity of $h(x,y)$ in y, we obtain

$$h(x,x) - h(x, x+1) = h(0,0) - h(0,1) \qquad x = 0, 1, \ldots .$$

For $m = 2$, in (2.14) we have

$$h(x,x) - h(x, x+2) = h(0,0) - h(0,2) \qquad x = 0, 1, \ldots ,$$

and by subtraction

$$h(x, x+1) - h(x, x+2) = h(0,1) - h(0,2) \qquad x = 0, 1, \ldots .$$

In the same way, proceeding inductively, one obtains

$$h(x, x+s) - h(x, x+s+t) = h(0,s) - h(0, s+t) ,$$

all $x, s, t = 0, 1, \ldots$. Now set $s = 0$ and $x + t = y$ to find

$$h(x,y) = h(0, y-x) + h(x,x) - h(0,0) ,$$

all $x = 0, 1, \ldots, y \geq x$. This in turn implies that $g(x,y)$ has the form

$$g(x,y) = a(y-x) + b(x) ,$$

on D with $a(\cdot)$ monotonous and arbitrarily defined elsewhere. This concludes the proof.

In Theorems 2.2 and 2.3 nothing was assumed about the form of the distribution of $g(R_j, R_n)$. If we further assume that it has the appropriate negative binomial distribution, then $g(R_j, R_n) = R_n - R_j$ on D. This is stated and proved in the theorem that follows.

THEOREM 2.5. *Suppose that either the conditions of Theorem 2.2 or of Theorem 2.3 hold for $g(R_j, R_n)$. Assume in addition that $g(R_j, R_n)$ has the negative binomial distribution with parameter q, with left end $n-j$ for all $q \in (0,1)$. Then $g(x,y) = y - x$ on D and arbitrary elsewhere.*

PROOF. From the conclusion of either theorem we have $g(x,y) = h(y-x)$ on D. The joint distribution of R_j and R_n is given by

$$P(R_j = r_j,\ R_n = r_n) = \binom{r_j}{j}\binom{r_n - r_j - 1}{n-j-1} p^{n+1} q^{r_n - n},$$

for $(r_j, r_n) \in D$. Using the above equation in the evaluation of $p_z = P(h(R_n - R_j) = z)$ and the assumptions, we find

$$\sum_{(x,y)\in C_z} \binom{x}{j}\binom{y-x-1}{n-j-1} p^{n+1} q^{y-n} = \binom{z-1}{n-j-1} p^{n-j} q^{z-n+j}, \tag{2.15}$$

$z = n-j, n-j+1, \ldots$ for all $q \in (0,1)$. Now take a fixed value of z. Then we can see at once that a fixed $(x,y) \in C_z$ implies $(x+u, y+u) \in C_z$ for all $u \geq j - x$. Therefore, $(j+k, y-x+j+k) \in C_z$ for $k = 0, 1, \ldots$. From the l.h.s. of equation (2.15) using summation over k and the binomial expansion, we arrive at the inequality

$$p_z \geq \binom{y-x-1}{n-j-1} p^{n-j} q^{y-x-n+j}, \tag{2.16}$$

$z = n-j, n-j+1, \ldots$. We denote by w_{y-x} the r.h.s. of the inequality (2.16). Since the above assumptions should be valid for all $q \in (0,1)$, we can choose q in such a way that the sequence P_z, $z = n-j, n-j+1, \ldots$ satisfies $p_{n-j+1} > p_{n-j} > p_{n-j+2} > \cdots$. For this to hold, it suffices to have $(n-j)^{-1} < q < 2((n-j)(n-j+1))^{-0.5}$. But the possible values of $y-x$, when $(x,y) \in D$, are $n-j, n-j+1, \ldots$. Therefore, a similar relation holds for the sequence w_{y-x}, $y - x = n-j, n-j+1, \ldots$ and the corresponding terms of the sequences are equal, that is $p_r = w_r$ for $r = n-j, n-j+1, \ldots$. Now consider the term w_{n-j+1}. Then the (x,y)'s which satisfy $y - x = n-j+1$ must belong to some C_z. Inequality (2.16) implies that $p_z \geq w_{n-j+1}$ and this is possible only if $z = n - j + 1$, because both terms are the unique maxima of the corresponding sequences. Hence, we have $(x, y) \in C_{n-j+1}$ iff $y - x = n - j + 1$. Next we consider w_{n-j}. Arguing in the same way we find an analogous result. By induction, we conclude $(x, y) \in C_z$ iff $z = y - x$ for all $z = n - j, n - j + 1, \ldots$. Hence, $g(x, y) = y - x$ on D and arbitrary elsewhere. This finishes the proof of the theorem.

References

Ahsanullah, M. and Holland, B. (1984). Record values and the geometric distribution, *Statist. Hefte*, **25**, 319-327.

Deheuvels, P. (1984). The characterization of distributions by order statistics and record values—a unified approach, *J. Appl. Probab.*, **21**, 326-334.

Gupta, R. C. (1984). Relationships between order statistics and record values and some characterization results, *J. Appl. Probab.*, **21**, 425–430.

Pfeifer, D. (1979). 'Record values' in einem stochastischen Modell mit nicht-identischen Verteilungen, Ph. D. Thesis, Aachen.

Rao, C. R. and Shanbhag, D. N. (1986). Recent results on characterization of probability distributions: A unified approach through extensions of Deny's theorem, *Adv. in Appl. Probab.*, **18**, 660-678.