УДК 510.21

CONDITIONAL DISTRIBUTIONS OF RECORD VALUES

A.V. Stepanov*

*BFU n. by I. Kant, Russia, 236041, Kaliningrad, A. Nevsky Street, 14
E-mail: alexeistep45@mail.ru

Conditional distribution functions of record values given maximum are studied in the paper. The result can be further used for obtaining characterizations based on records.

records, order statistics, conditional distributions, Stiring's numbers

1. INTRODUCTION

Let X_1, $X_2, \ldots$ be a sequence of independent identically distributed random variables with a common distribution function $F(x) = P\{X < x\}$ with support $\Im$. The sequence of record times $L(n)$ and record values $X(n)$ is defined as follows:

$$L(1) = 1, \quad X(1) = X_1,$$

$$L(n+1) = \min\left\{ j : j > L(n), \quad X_j > X_{L(n)} \right\},$$

$$X(n) = X_{L(n)} \quad (n \geq 1).$$

The distribution function of $X(n)$, when F is continuous was found in Tata's paper [1]

$$P\{X(n) < x\} = \frac{1}{(n-1)!} \int_0^{-\log(1-F(x))} v^{n-1} e^{-v} dv \qquad (n \geq 1)$$

Define M_n $(n \geq 1)$ as the maximum of $X_1, X_2, \ldots, X_n$. It is a well-known fact (see M. Ahsanullah [2]; B. Arnold, N. Balakrishnan, H. Nagaraja [3] ; H. David [4]; V. Nevzorov [5] and others) that

$$P\{M_n < x\} = F^n(x).$$

The density function of M_n is

$$p_{M_n}(x) = np(x)F^{n-1}(x), \tag{1}$$

when the initial F is absolutely continuous and $p(x) = F'(x)$.

The paper of H. Nagaraja and V. Nevzorov [6] was dedicated to characterizations of distributions through properties of both record values and order statistics. It was found there as an auxiliary result (see lemma 5) that for any $x, y \in \Im$

$$P\{X(2) < y | M_2 = x\} = \begin{cases} 0 & y < x \\ \dfrac{1}{2} + \dfrac{F(y) - F(x)}{2(1 - F(x))} & y \geq x \end{cases}$$

The goal of our paper is to extend the last formula. We will find $P\{X(k)<y|M_n=x\}$ for any $k\ge 1, n\ge 1$. This formula could be used for further characterizations.

2. DEPENDENCE STRUCTURE OF M_n AND $X(2)$

Theorem 1 It is true for any x, y in $\Im$

$$P\{X(2)<y|M_n=x\}=\begin{cases} 0 & n=1,2 \quad y<x \quad (A) \\ \frac{1}{n}\sum_{l=2}^{n-1}\frac{n-l}{(l-1)l}\left(\frac{F(y)}{F(x)}\right)^l & n\ge 3 \quad y<x \quad (AA) \\ \frac{n-1}{n}+\frac{1}{n}\frac{F(y)-F(x)}{1-F(x)} & n\ge 1 \quad y\ge x \quad (AAA) \end{cases}$$

Proof Assume first that $F(x)$ is absolutely continuous and $p(x)=F'(x)$ for any $x\in\Im$.

(A) The result is obvious. Indeed, the event

$$\{X(2)<y, x<M_n\le x+dx\}$$

can't occur for $n=1,2$ and any $y<x$. That is why

$$P\{X(2)<y|M_n=x\}=0 \qquad (n=1,2, \quad y<x\in\Im).$$

(AA) Define following events $U_{L(1)}, U_1, U_{L(2)}, T, U_{M_n}, Q$ as

$$U_{L(1)}=\{X_1<y\}, \qquad U_1=\prod_{i=2}^{i_2-1}\{X_i<X_1\},$$

$$U_{L(2)}=\{X_1<X_{i_2}<y\}, \qquad T=\left\{\prod_{i=i_2+1}^{t-1}\{X_i<x\}\right\},$$

$$U_{M_n}=\{X_t\in[x,x+dx)\}, \qquad Q=\prod_{i=t+1}^{n}\{X_i<x\}.$$

We can write for $y<x$ when $p(x)>0$

$$P\{X(2)<y, x\le M_n<x+dx\}=\sum_{t=3}^{n}\sum_{i_2=2}^{t-1}P\{U_{L(1)}U_1U_{L(2)}TU_{M_n}Q\}. \qquad (2)$$

The events $U_{L(1)}$ and $U_{L(2)}$ give correspondingly the first and the second record values. The event U_{M_n} produces the maximum M_n , here t is the index such that $X_t=M_n$. The event U_1 consists of some trials that do not exceed the first record value (this set could be empty). The event T formed with the trials $X_{L(2)+1}, X_{L(2)+2}, ..., X_{t-1}$. It could contain some record values and could be empty. The event Q consists of the trials $X_{t+1}, X_{t+2}, ..., X_n$ (could be empty). The product of the events $U_{L(1)}U_1U_{L(2)}TU_{M_n}Q$ tells us that there are at least two record values belonging to $(-\infty, y)$ and $M_n\in[x,x+dx)$. It follows from (2), that for those x for which $p(x)>0$

$$P\{X(2)<y, x\le M_n<x+dx\}$$
$$=\sum_{t=3}^{n}\sum_{i_2=2}^{t-1}\left(\int_{-\infty}^{y}(F(y)-F(u))F^{i_2-2}(u)p(u)du(p(x)+0(dx))F^{n-i_2-1}(x)\right)=$$
$$=(p(x)+0(dx))F^{n-1}(x)dx\cdot\sum_{i_2=2}^{n-1}\sum_{t=i_2+1}^{n}\frac{1}{(i_2-1)i_2}\left(\frac{F(y)}{F(x)}\right)^{i_2}=$$
$$=(p(x)+0(dx))F^{n-1}(x)dx\cdot\sum_{i_2=2}^{n-1}\frac{n-i_2}{(i_2-1)i_2}\left(\frac{F(y)}{F(x)}\right)^{i_2}.$$

The joint density-distribution function $G(y,x)$ of the vector $(X(2),M_n)$ can be got from the last formula, where the density function is taken for M_n and the distribution for $X(2)$

$$G(y,x)=p(x)F^{n-1}(x)\sum_{l=2}^{n-1}\frac{n-l}{(l-1)l}\left(\frac{F(y)}{F(x)}\right)^{l},$$

where $y<x\in\Im$. Taking into account (1) we complete part (AA).

Remark 1 We supposed at the beginning of the proof that $F(x)$ is absolutely continuous for $x\in\Im$. The result holds true if F is continuous.

Part (AAA) of the theorem follows from more general result .

Presentation It is true for any $x\le y\in\Im$

$$P\{X(k)<y \mid M_n=x\}=\begin{cases}\sum_{i=1}^{n}P\{N(n)=i,\mu(x.y)\ge k-i\} & (n\le k) \quad (B)\\ P\{N(n)\ge k\}+\sum_{i=1}^{k-1}P\{N(n)=i,\mu(x,y)\ge k-i\} & (n>k) \quad (BB)\end{cases}$$

Here $N(n)$ is the total number of record values amongst $X_1,X_2,...,X_n$, $\mu(x,y)$ is the total number of record values fallen on (x,y). It is known (see A. Renyi [7], or R. Shorrock [8]) that

$$P\{N(n)=i\}=\frac{|S_n^i|}{n!},$$

where S_n^i are Stiring's numbers, i.e.

$$x(x-1)...(x-(n-1))=\sum_{i=0}^{\infty}S_n^i x^i.$$

Thus, in particular, $S_n^0=0$, $S_n^1=(-1)^{n-1}(n-1)!$

Random variables μ , taken for different intervals, reflect numbers of record values fallen on corresponding intervals. They are independent for different non-overlapping intervals (see R. Shorrock [8]) and

$$P\{\mu(x,y)=i\}=\frac{e^{-\lambda}\lambda^i}{i!}, \qquad (i\ge 0)$$

where

$$\lambda = -\log\left(\frac{1-F(y)}{1-F(x)}\right). \tag{3}$$

Let's resume proving the theorem. Formula (AAA) of the theorem should be considered for two cases: a) $n > k = 2$. Here we take part (BB) of the presentation . It is true

$$P\{X(2) < y | M_n = x\} = P\{N(n) \geq 2\} + P\{N(n) = 1, \mu(x,y) \geq 1\}$$
$$= 1 - P\{N(n) = 1\} + P\{N(n) = 1\} P\{\mu(x,y) \geq 1\} \qquad (n \geq 3) \tag{4}$$

The last product can be written, because the variables N and μ are independent. Indeed, the number of record values $N(n)$ amongst $X_1, X_2, \ldots, X_n$ is, in fact, the number of record values registered for the interval $(-\infty, x]$, because $(x < M_n < x + dx)$. According to Shorrock's result , the number of records registered for $(-\infty, x]$ and the number of records fallen on (x, y) are independent variables. Now from (4) we find

$$P\{X(2) < y | M_n = x\} = 1 - \frac{1}{n} + \frac{1}{n}\left(1 - P\{\mu(x,y) = 0\}\right);$$

b) $k = 2$, $n = 1,2$ For proving this part of the theorem formula (B) of the presentation might be exploited

$$P\{X(2) < y | M_n = x\}$$

$$= \begin{cases} P\{\mu(x,y) \geq 1\} & n = 1, k = 2 \quad y \geq x \\ P\{N(2) = 1, \mu(x,y) \geq 1\} + P\{N(2) = 2, \mu(x,y) \geq 0\} & n = 2, k = 2 \quad y \geq x \end{cases}$$

That is why

$$P\{X(2) < y | M_n = x\} = \begin{cases} \dfrac{F(y) - F(x)}{1 - F(x)} & n = 1, k = 2 \quad y \geq x \\ \dfrac{F(y) - F(x)}{2(1 - F(x))} + \dfrac{1}{2} & n = k = 2 \quad y \geq x \end{cases}$$

Remark 2 It can be also stressed that the conditional distribution function $P\{X(2) < y | M_n = x\}$ has the atom $\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{i}$ $(n \geq 2)$ at the point x .

3. DEPENDENCE STRUCTURE OF $X(k)$ AND M_n

Theorem 2 *It is true for any* $x, y \in \Im$

$$P\{X(k) < y | M_n = x\}$$

$$
=\begin{cases} 0 & 1\le n\le k \quad x>y \quad (C) \\ \sum_{i_2=2}^{n-k+1}\sum_{i_3=i_2+1}^{n-k+2}\cdots\sum_{i_k=i_{k-1}+1}^{n-1}\frac{n-i_k}{(i_2-1)\dots(i_{k-1}-1)(i_k-1)i_k}\left(\frac{F(y)}{F(x)}\right)^{i_k} & n>k \quad x>y \quad (CC) \\ \frac{e^{-\lambda}}{n!}\cdot\sum_{i=1}^{n}\left|S_n^i\right|\sum_{j=k-i}^{\infty}\frac{\lambda^j}{j!} & 1\le n\le k \quad x\le y \quad (B) \\ \frac{1}{n!}\sum_{i=k}^{n}\left|S_n^i\right|+\frac{e^{-\lambda}}{n!}\sum_{i=1}^{k-1}\left|S_n^i\right|\sum_{j=k-i}^{\infty}\frac{\lambda^j}{j!} & n>k \quad x\le y \quad (BB) \end{cases}
$$

where λ was determined by (3) and S_n^i are Stirling's numbers. Here parts (B), (BB) are taken from the presentation.

Proof (C) The proof is the same as it was for (A). In fact, we should prove (CC), because (B) and (BB) have been considered before. Assume as it was before that $p(x)=F'(x)>0$ for any $x\in\Im$. Take events

$$
U_{L(1)}=\{X_1<y\},\quad U_{L(2)}=\{X_1<X_{i_2}<y\},\dots,U_{L(k)}=\{X_{i_{k-1}}<X_{i_k}<y\},
$$

$$
U_1=\bigcap_{j=2}^{i_2-1}\{X_j<X_1\},\quad U_2=\bigcap_{j=i_2+1}^{i_3-1}\{X_j<X_{i_2}\},\dots,\quad U_{k-1}=\bigcap_{j=i_{k-1}+1}^{i_k-1}\{X_j<X_{i_{k-1}}\},
$$

$$
T=\bigcap_{j=i_k+1}^{t-1}\{X_j<x\},\quad U_{M_n}=\{x\le X_t<x+dx\},\quad Q=\bigcap_{j=t+1}^{n}\{X_j<x\}
$$

The event

$$
U_{L(1)}U_1U_{L(2)}U_2\dots U_{k-1}U_{L(k)}TU_{M_n}Q
$$

guaranties that there are at least k record values belonging to $(-\infty,y)$ and $M_n\in[x,x+dx)$. So, for $y<x$ such that $p(x)>0$ the following equality is valid

$$
P\{X(k)<y,x\le M_n<x+dx\}
$$

$$
=\sum_{t=k+1}^{n}\sum_{i_2=2}^{t-k+1}\sum_{i_3=i_2+1}^{t-k}\cdots\sum_{i_k=i_{k-1}+1}^{t-1}P\{U_{L(1)}U_1U_{L(2)}U_2\dots U_{k-1}U_{L(k)}TU_{M_n}Q\}
$$

$$
=\sum_{t=k+1}^{n}\sum_{i_2=2}^{t-k+1}\sum_{i_3=i_2+1}^{t-k}\cdots\sum_{i_k=i_{k-1}+1}^{t-1}\left(\int_{-\infty}^{y}F^{i_2-2}(x_2)p(x_2)dx_2\int_{x_2}^{y}F^{i_3-i_2-1}(x_3)p(x_3)dx_3\dots\right.
$$

$$
\left.\dots\int_{x_{k-1}}^{y}F^{i_k-i_{k-1}-1}(x_k)(F(y)-F(x_k))p(x_k)dx_k\{p(x)+0(dx)\}F^{n-i_k-1}(x)dx\right).
$$

From here we get the density-distribution function $G(y,x)$ of the vector $(X(k),M_n)$

$$
G(y,x)=p(x)F^{n-1}(x)\sum_{i_2=2}^{n-k+1}\sum_{i_3=i_2+1}^{n-k+2}\cdots\sum_{i_k=i_{k-1}+1}^{n-1}\frac{n-i_k}{(i_2-1)\dots(i_{k-1}-1)(i_k-1)i_k}\left(\frac{F(y)}{F(x)}\right)^{i_n},
$$

where $y<x$, x such, that $p(x)>0$. Using (1) we conclude the proof.

Remark 3 The result holds true if we suppose F is continuous.

REFERENCES

1. Tata, M. On outstanding values in a sequence of random variables // Z. Wahrscheinlichkeitheor. verw. Geb. – 1969. – Vol. 12. - P. 9 - 20.

2. Ahsanullah, M. Record Statistics / Nova Science Publishers / Inc. Commack, NY, 1995. – 217 p.

3. Arnold, B. Balakrishnan, N. Nagaraja, H. Records / Wiley, NY, 1998.- 321 p.

4. David, H. Order Statistics / Wiley, NY, 1970. - 298 p.

5. Nevzorov, V. Records: math theory / Fazis, Moscow / 2000. - 241 p.

6. Nagaraja, N. Nevzorov, V. On characterizations based on record values and order statistics // Journal of Statistical Planning and Inference – 1997. Vol. 63. - P. 271 - 284.

7. Renyi, A. On the extreme elements of observations // In: Selected papers of Alfred Renyi, Budapest, Akademiai Kiado. - 1976. - Vol. 3 - P. 50 - 65.

8. Shorrock, R. A limit theorem for inter–record times // Journal of Applied Probability. – 1972. - Vol. 9 - P. 543 - 555.

УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ РЕКОРДНЫХ ВЕЛИЧИН

А.В. Степанов

В данной работе изучаются условные распределения рекордных величин при фиксированном значении максимальной. Результаты этой работы могут использоваться в дальнейшем для получения характеризаций, основанных на рекордах.

рекорды, порядковые статистики, условные распределения, числа Стирлинга