DISTRIBUTIONS OF SPACINGS OF ORDER STATISTICS AND THEIR RATIOS

MOHAMED I. RIFFI DEPARTMENT OF MATHEMATICS ISLAMIC UNIVERSITY OF GAZA, PALESTINE

ABSTRACT. We study, in this paper, the distributions of spacings between successive order statistics both when the distribution of the random sample has support of the form $[a, b]$ or of the form $(-\infty, +\infty)$. We also study stopping times related to these spacings. We present explicit formulae for the expected values of these stopping times. We also study the distributions of the ratios of non-negative successive order statistics and the behavior of some related stopping times.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_n$ be independent random variables constituting a random sample from a distribution with absolutely continuous cumulative distribution function (cdf) F;i.e., F has probability density function (pdf) f with respect to the Lebesgue measure. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of the random variables $X_1, \ldots, X_n$. Let $D_i := X_{(i+1)} - X_{(i)}, i = 1, 2, 3, \ldots, n-1$; i.e., D_i is the difference between the two adjacent order statistics $X_{(i+1)}$ and $X_{(i)}$. The random variables $D_1, D_2, \ldots, D_{n-1}$ are called the spacings between the successive order statistics [1, p. 178]. These spacings evidently give some idea about the "spread" of a distribution. Also note that the sample Range,

---

1991 *Mathematics Subject Classification.* Primary 60A99. Secondary 28A99.

Key words and phrases. order statistics, distribution, stopping times, successive order statistics, spacings of order statistics.

$R = X_{(n)} - X_{(1)}$, is given by $R = \sum_{i=1}^{n-1} D_i$. They also can be used to construct confidence intervals for the corresponding population. Therefore, these spacings may be of interest in their own right.

Our work is inspired by the results of DasGupta, Y. Rinott, and B. Vidakovic about stopping times related to diagnostics and outliers [2]. They worked out problems regarding the sample range $X_{(n)} - X_{(1)}$ and the waiting time until the range becomes greater than a fixed constant.

In Section 2, we use the joint distribution of order statistics to derive the pdf of the random variables $D_1, \ldots, D_{n-1}$ in terms of F and f of the distribution of the random sample.

In Section 3, we study the smallest sample size needed to get the spacings between order statistics as small as a fixed but arbitrary positive constant. We define stopping times related to this sample size and study their behaviors in terms of their expectations and variances.

In Section 4, we study the distributions of ratios of the spacings between order statistics and study the smallest sample size required before these divisions become smaller than a fixed constant.

We will assume throughout the paper that, unless otherwise stated, we are sampling from a distribution with support $(-\infty, +\infty)$.

2. Distributions of D_is

Recall that the pdf of the ith order statistic $X_{(i)}$ is given by (Arnold et al [1])

$$f_i(x) = i\binom{n}{i} f(x)F^{i-1}(x)[1 - F(x)]^{n-i}, \quad -\infty < x < \infty. \tag{2.1}$$

We also recall that the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is given by (Arnold et al [1])

$$\begin{aligned} f_{i,j}(x,y) = & \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F^{i-1}(x) \\ & \times [F(y) - F(x)]^{j-i-1}[1 - F(y)]^{n-j} f(x) f(y), \end{aligned} \tag{2.2}$$

for $-\infty < x < y < \infty$.

Lemma 2.1. *Let $D_1, D_2, \ldots, D_{n-1}$ be the spacings between successive order statistics. Then, for any $0 < c < \infty$ and $i = 1, 2, \ldots, n-1$,*

$$F_{D_i}(c) = P(D_i \le c) = 1 - \int_{-\infty}^{\infty} i\binom{n}{i} f(x)F^{i-1}(x)\,[1 - F(x+c)]^{n-i}\,dx. \tag{2.3}$$

Proof. Let S be the region bounded by $X_{(i)} \le X_{(i+1)}$ and $X_{(i+1)} \le X_{(i)}+c$. It follows from (2.2) that

$$\begin{aligned}
F_{D_i}(c) &= \iint_S f_{i,i+1}(x,y)\,dy\,dx \\
&= \int_{-\infty}^{\infty}\int_x^{x+c} \frac{n!}{(i-1)!(n-i-1)!}[1-F(y)]^{n-i-1}F^{i-1}(x)f(x)f(y)\,dy\,dx \\
&= \int_{-\infty}^{\infty} \frac{n!f(x)}{(i-1)!(n-i-1)!}F^{i-1}(x)\left(\int_x^{x+c} f(y)[1-F(y)]^{n-i-1}\,dy\right)dx \\
&= \int_{-\infty}^{\infty} i\binom{n}{i} f(x)F^{i-1}(x)\left((1-F(x))^{n-i} - (1-F(x+c))^{n-i}\right)dx.
\end{aligned}$$

From (2.1), we have

$$\int_{-\infty}^{\infty} i\binom{n}{i} f(x)F^{i-1}(x)[1-F(x)]^{n-i}\,dx = 1.$$

Consequently, the result follows. ∎

Lemma 2.2. *Let $D_1, D_2, \ldots, D_{n-1}$ be the spacings between successive order statistics coming from a random sample from a distribution with support $[a, b]$, where $-\infty < a < b < \infty$. Then for any $0 < c < b$,*

$$F_{D_i}(c) = 1 - \int_a^{b-c} i\binom{n}{i} f(x)F^{i-1}(x)[1 - F(x+c)]^{n-i}\,dx. \tag{2.4}$$

Proof.

$P(D_i \leq c) = I_1 + I_2$, where

$$I_1 = \int_a^{b-c}\int_x^{x+c} \frac{n!}{(i-1)!(n-i-1)!}[1-F(y)]^{n-i-1}F^{i-1}(x)f(x)f(y)dydx,$$

$$I_2 = \int_{b-c}^{b}\int_x^{b} \frac{n!}{(i-1)!(n-i-1)!}[1-F(y)]^{n-i-1}F^{i-1}(x)f(x)f(y)dydx.$$

$$\begin{aligned} I_1 &= \int_a^{b-c} \frac{n!f(x)}{(i-1)!(n-i-1)!}F^{i-1}(x)\left(\int_x^{x+c} f(y)[1-F(y)]^{n-i-1}dy\right)dx \\ &= \int_a^{b-c} i\binom{n}{i} f(x)F^{i-1}(x)\left\{(1-F(x))^{n-i} - (1-F(x+c))^{n-i}\right\}dx \end{aligned} \tag{2.5}$$

$$\begin{aligned} I_2 &= \int_{b-c}^{b} \frac{n!f(x)}{(i-1)!(n-i-1)!}F^{i-1}(x)\left(\int_x^{b} f(y)[1-F(y)]^{n-i-1}dy\right)dx \\ &= \int_{b-c}^{b} i\binom{n}{i} f(x)F^{i-1}(x)[1-F(x)]^{n-i}dx. \end{aligned} \tag{2.6}$$

By adding (2.5) to (2.6) and noting that

$$\int_a^b i\binom{n}{i} f(x)F^{i-1}(x)[1-F(x)]^{n-i}dx = 1,$$

we get (2.4), and that completes the proof. ■

Lemma 2.3. *Suppose that the distribution of the sample has support of the form* $[a, b]$, *where* $-\infty < a < b < \infty$. *Let* $D_{n-1} = X_{(n)} - X_{(n-1)}$ *and*

$0 \le a \le b$. *Then*

$$P(D_{n-1} \le c) = 1 - \int_{a+c}^{b} nf(y)F^{n-1}(y-c)dy. \tag{2.7}$$

Proof. Let S be the region bounded by $X_{(n)} \le X_{(n-1)}$ and $X_{(n)} - c \le X_{(n-1)}$. It follows from

$$\begin{aligned} P(D_{n-1} \le c) &= \iint_{S} f_{n-1,n}(x,y)dxdy \\ &= \int_{a}^{a+c}\int_{a}^{y} n(n-1)f(y)f(x)F^{n-2}(x)dxdy \\ &\quad + \int_{a+c}^{b}\int_{y-c}^{y} n(n-1)f(y)f(x)F^{n-2}(x)dxdy \end{aligned} \tag{2.8}$$

We manipulate the integrals on the right-hand side of (2.8) by integrating first with respect to x to get

$$\begin{aligned} P(D_{n-1} \le c) &= \int_{a}^{a+c} nf(y)\left[\int_{a}^{y}(n-1)f(x)F^{n-2}(x)dx\right]dy \\ &\quad + \int_{a+c}^{b} nf(y)\left[\int_{y-c}^{y}(n-1)f(x)F^{n-2}(x)dx\right]dy \\ &= \int_{a}^{a+c} nf(y)F^{n-1}(y)dy \\ &\quad + \int_{a+c}^{b} nf(y)F^{n-1}(y)dy - \int_{a+c}^{b} nf(y)F^{n-1}(y-c)dy, \end{aligned}$$

Since $F^{n-1}(a) = 0$. Therefore,

$$P(D_{n-1} \le c) = \int_a^b nf(y)F^{n-1}(y)dy - \int_{a+c}^b nf(y)F^{n-1}(y-c)dy.$$

Then (2.7) follows since $\int_a^b nf(y)F^{n-1}(y)dy = 1$. ■

The following Lemma deals with the distribution of D_{n-1} when the support is of the form $(-\infty, \infty)$.

Lemma 2.4. *Let $D_{n-1} = X_{(n)} - X_{(n-1)}$ and $0 < c < \infty$. Then*

$$P(D_{n-1} \le c) = 1 - \int_{-\infty}^{\infty} nf(y)F^{n-1}(y-c)dy. \qquad (2.9)$$

Proof. The proof is similar to that of Lemma 2.3. ■

3. The Stopping Time N_1

We now turn to the study of the smallest sample size needed to keep the spacings between successive order statistics less than or equal to a fixed but arbitrary constant $0 < c < \infty$ in the infinite support cases. Then we study the expected value and the variance of the stopping time $N_1 = \inf\{n \ge 1 : D_{n-1} \le c\}$ when the underlying distribution of the random sample has support $[a, b]$, where $-\infty < a < b < \infty$ and $0 < c < b$.

Theorem 3.1. *Let D_{n-1} be the space between the nth and $(n-1)$th order statistics coming from a random sample from a distribution with support $[a, b]$, where $-\infty < a < b < \infty$. Then for any $0 < c < b$,*

$$\sum_{n=2}^{\infty} P(D_{n-1} \ge c) = \int_{a+c}^{b} f(y)[1 - F(y-c)]^{-2}dy + F(a+c) - 1. \qquad (3.1)$$

Proof. By Theorem 2.3, we see that

$$\sum_{n=2}^{\infty} P(D_{n-1} \geq c) = \sum_{n=2}^{\infty} \int_{a+c}^{b} nf(y)F^{n-1}(y - c)dy$$

$$= \int_{a+c}^{b} f(y) \left(\sum_{n=2}^{\infty} nF^{n-1}(y - c) \right) dy,$$

where we have used the Monotone Convergence Theorem to carry the infinite sum inside of the integral. Now, Note that

$$\sum_{n=2}^{\infty} nF^{n-1}(y - c) = \frac{d}{da} g(a),$$

where $g(a) = \sum_{i=0}^{\infty} a^i - a - 1$, $a = F(y - c)$, and $0 \leq F(y - c) \leq 1$. Therefore,

$$\sum_{n=2}^{\infty} nF^{n-1}(y) = [1 - F(y - c)]^{-2} - 1,$$

and

$$\sum_{n=2}^{\infty} P(D_{n-1} \geq c) = \int_{a+c}^{b} f(y)[1 - F(y - c)]^{-2}dy + F(a + c) - 1,$$

since $F(b) = 1$. ∎

Theorem 3.2. *Let D_{n-1} be the space between the nth and $(n-1)$th order statistics coming from a random sample from a distribution with support $[a, b]$, where $-\infty < a < b < \infty$. Let $N_1 = \inf\{n \geq 1 : D_{n-1} \leq c\}$, where $0 < c < b$. Then*

$$E(N_1) = 1 + \int_{a+c}^{b} f(y)[1 - F(y - c)]^{-2}dy + F(a + c). \qquad (3.2)$$

Proof. By definition of N_1, $P(N_1 \geq n) = P(D_{n-1} \geq c)$. Therefore,

$$\mathrm{E}(N_1) = \sum_{n=0}^{\infty} P(N_1 \geq n) = \sum_{n=2}^{\infty} P(N_1 \geq n) + 2$$

By (3.1), we see that

$$E(N_1) = 1 + \int_{a+c}^{b} f(y)[1 - F(y-c)]^{-2}\,dy + F(a+c).$$

■

Theorem 3.3. *Let D_{n-1} be the space between the nth and $(n-1)$th order statistics coming from a random sample from a distribution with support $[a,b]$, where $-\infty < a < b < \infty$. Let $N_1 = \inf\{n \geq 1 : D_{n-1} < c\}$, where $0 < c < b$. Then*

$$Var(N_1) = 2 - E(N_1) + 4\int_{a+c}^{b} f(y)[1 - F(y-c)]^{-3}\,dy + 4F(a+c) - (E(N_1))^2. \tag{3.3}$$

Proof. We use the identity [3, p. 33]

$$\begin{aligned} E(N_1(N_1 - 1)) &= \sum_{n=1}^{\infty} 2nP(N_1 \geq n) \\ &= \sum_{n=2}^{\infty} 2nP(D_{n-1} \geq c) + 2 \\ &= \sum_{n=2}^{\infty} \int_{a+c}^{b} 2n^2 f(y) F^{n-1}(y-c)\,dy, \end{aligned} \tag{3.4}$$

by (2.7). We now manipulate (3.4) by first using the Monotone Convergence Theorem to carry the infinite sum inside the integral, and then use the equation

$$\sum_{n=2}^{\infty} 2n^2 a^{n-1} = \frac{4}{(1-a)^3} - \frac{2}{(1-a)^2} - 2, \tag{3.5}$$

where $a = F^{n-1}(y-c)$. Therefore, using (3.5), we get

$$E(N_1(N_1 - 1)) = 2 - 2E(N_1) + 4\int_{a+c}^{b} f(y)[1 - F(y-c)]^{-3}\,dy + 4F(a+c).$$

Therefore,

$$\mathrm{Var}(N_1) = 2 - \mathrm{E}(N_1) + 4\int_{a+c}^{b} f(y)[1-F(y-c)]^{-3}dy + 4F(a+c) - (\mathrm{E}(N_1))^2.$$

■

Remark 3.1. *Equation (3.2) gives the expectation of the stopping time N_1. It simply says that when $c \Rightarrow 0$, then $E(N_1) \Rightarrow \infty$. And $E(N_1) \Rightarrow 2$ when $c \Rightarrow b$. In fact, $E(N_1) \geq 2$.*

Suppose that we are sampling from a uniform distribution over the interval $[0, 1]$. Then, in this case,

$$E(N_1) = 1 + \int_{c}^{1} \frac{1}{(1 - y + c)^2} dy + c = \frac{1}{c} + c.$$

For example, when $c = 1/2$, then $E(N_1) = 2.5$. When $c \Rightarrow 0$, we see that $E(N_1) \Rightarrow \infty$.

In the infinite support case $(-\infty, \infty)$ or $(0, \infty)$ the expectation of N_1 does not exist.

Example 3.1. *Suppose we are sampling from a uniform distribution over the interval $[0, 1]$. Then, we see that*

$$\begin{aligned} Var(N_1) &= 2 - (\frac{1}{c} + c) + 4(\frac{1}{c^2} - 1) + 4c - (\frac{1}{c} + c)^2 \\ &= \frac{3 - c - 4c^2 + 3c^3 - c^4}{c^2}. \end{aligned} \tag{3.6}$$

The Equation (3.6) says that $Var(N_1) \Rightarrow \infty$ as $c \Rightarrow 0$, and that $Var(N_1) \Rightarrow 0$ as $c \Rightarrow 1$. For example, $Var(N_1) = 7.25$ when $c = .5$.

4. The stopping time N_2

Here we let $R_n = X_{(n)}/X_{(n-1)}$ when the sample is drawn from a non-negative distribution with support $(0, \infty)$.

Lemma 4.1. *Assume that the support of X is $(0, \infty)$. Let R_n be the ratio of the nth and $(n-1)$th order statistics; that is $R_n = X_{(n)}/X_{(n-1)}$.*

Then, for $0 \leq c \leq \infty$, *we have*

$$P(R_n \leq c) = 1 - \int_0^\infty nf(y)F^{n-1}(y/c)dy. \tag{4.1}$$

Proof. Let S be the region bounded by $X_{(n-1)} \leq X_{(n)}$ and $cX_{(n-1)} \leq X_{(n)}$. Then

$$\begin{aligned} P(R_n \leq c) &= \iint_S f_{n-1,n}(x,y)dxdy \\ &= \int_0^\infty \int_{y/c}^y n(n-1)f(y)f(x)F^{n-2}(x)dxdy \\ &= \int_0^\infty nf(y)[F^{n-1}(y) - F^{n-1}(y/c)]dy \\ &= 1 - \int_0^\infty nf(y)F^{n-1}(y/c)dy. \end{aligned}$$

∎

Theorem 4.1. *Assume that the support of X is $(0,\infty)$. Let R_n be the ratio of the nth and $(n-1)$th order statistics; that is $R_n = X_{(n)}/X_{(n-1)}$. Let $N_2 = \inf\{n \geq 1 : R_n \leq c\}$. Then, there exists a constant α such that for $0 < \alpha < c < \infty$, we have*

$$EN_2 = 1 + \int_0^\infty f(y)[1 - F(y/c)]^{-2}dy. \tag{4.2}$$

Proof.

$$\sum_{n=2}^\infty P(R_n \geq c) = \int_0^\infty f(y)\left(\sum_{n=2}^\infty nF^{n-1}(y/c)\right)dy,$$

where we have used the Monotone Convergence Theorem to carry the infinite sum inside the integral. Since $0 \leq F^{n-1}(y/c) \leq 1, \forall y \geq 0$, we have

$$\sum_{n=2}^{\infty} nF^{n-1}(y/c) = [1 - F(y/c)]^{-2} - 1.$$

Using (4.1), we see that

$$\sum_{n=2}^{\infty} P\left(R_n \geq c\right) = \int_0^{\infty} f(y)[1 - F(y/c)]^{-2}dy - 1. \qquad (4.3)$$

Then (4.2) follows from (4.3)

$$\mathrm{E}(N_2) = \sum_{n=0}^{\infty} P(N_2 \geq n) = 1 + \int_0^{\infty} f(y)[1 - F(y/c)]^{-2}dy.$$

■

Theorem 4.2. *Assume that the support of* X *is* $(0, \infty)$. *Let* R_n *be the ratio of the nth and* $(n-1)th$ *order statistics; that is* $R_n = X_{(n)}/X_{(n-1)}$, *and let* $N_2 = \inf\{n \geq 1 ; R_n \leq c\}$. *Then, there exists a constant* β *such that for* $0 < \beta < c < \infty$, *we have*

$$Var(N_2) = 2 - E(N_2) + 4\int_0^{\infty} f(y)[1 - F(y/c)]^{-3}dy - (E(N_2))^2. \qquad (4.4)$$

Proof. The proof of Theorem 4.2 is similar to that of (3.3), and hence is omitted. ■

Remark 4.1. *It follows from* (4.2) *that* $E(N_2) \geq 2$ *and that* $E(N_2) \to \infty$ *as* $c \to 0$.

The constant α *depends on the distribution of the sample as one can see in the following example.*

In the following example, when the underlying distribution is exponential with parameter 1, we see that in order for $E(N_2)$ *to exist, we must have* $c > 2$.

If $f(y) = e^{-y}, thenF(y/c) = 1 - e^{-y/c}$, *where we suppose that* $c > 2$. *So,* $\alpha = 2$ *here. Then*

$$\int_0^{\infty} f(y)[1 - F(y/c)]^{-2}dy = \int_0^{\infty} e^{-y}e^{2y/c}dy = \frac{c}{c-2}.$$

Therefore, by (4.2) *we see that*

$$E(N_2) = 1 + \frac{c}{c-2}, \quad \text{for } c > 2.$$

For example, when $c = 4$, $E(N_2) = 3$.

The constant β *depends on the distribution of the sample as one can see in the following example.*

For the variance, we see that

$$\begin{aligned} Var(N_2) &= 2\left(1 + \frac{c}{c-2}\right) + \frac{4c}{c-3} - \left(1 + \frac{c}{c-2}\right)^2 \\ &= 2 + \frac{2(-12 + 17c - 8c^2 + c^3)}{(3-c)(2-c)^2}, \end{aligned} \tag{4.5}$$

for $c > 3$. *In this example* $\beta = 3$.

The Equation 4.5 says that $Var(N_2) \to \infty$ *as* $c \to 3$ *and that* $Var(N_2) \to 0$ *as* $c \to \infty$. *For example,* $Var(N_2) = 6$ *when* $c = 4$.

References

[1] B. C. Arnold, N. Balakrishnan, and Nagaraja H. N., *A First Course in Order Statistics*, John Wiley, New York, 1992.

[2] George Casella, Roger L. Berger, *Statistical Inference*, Duxbury Press, 1990.

[3] DasGupta, Y. Rinott, and B. Vidakovic, *Stopping Times Related to Diagnostics and Outliers*, Preprint, 1999.

[4] Richard Durret, *Probability: Theory and Examples*, Wadsworth, 1991.

Islamic University of Gaza, P.O.Box 108, Gaza, Palestine

E-mail address: miriffi@hotmail.com