Fall September 11, 2019

Order statistics on the spacings between order statistics

Iosif Pinelis, *Michigan Technological University*

ORDER STATISTICS ON THE SPACINGS BETWEEN ORDER STATISTICS FOR THE UNIFORM DISTRIBUTION

IOSIF PINELIS

Abstract. Closed-form expressions for the distributions of the order statistics on the spacings between order statistics for the uniform distribution are obtained. This generalizes a result by Fisher concerning tests of significance in the harmonic analysis of a series.

Contents

1. Statements of main results and discussion

For any natural $n \geqslant 2$, let $U_1, \dots, U_n$ be independent identically distributed (iid) random variables (r.v.'s) each having the uniform distribution on the interval $[0, 1]$. As usual, let $U_{n:1} \leqslant \ldots \leqslant U_{n:n}$ denote the corresponding order statistics. Consider

$$G_i := U_{n:i} - U_{n:i-1} \quad \text{for} \quad i \in \overline{1, n+1}, \tag{1.1}$$

with $U_{n:0} := 0$ and $U_{n+1:n+1} := 1$. Here and in what follows, $\overline{\alpha, \beta} := \{k \in \mathbb{Z} \colon \alpha \leqslant k \leqslant \beta\}$.

One may refer to the G_i's as the gaps or, as it is usually done in the literature, spacings between the consecutive order statistics. See e.g. the paper by Pyke [8], containing a review of known results for the spacings for the underlying uniform distribution and other distributions as well; see also [3] for later updates.

Let now $G_{n+1:1} \leqslant \ldots \leqslant G_{n+1:n+1}$ denote the ordered gaps $G_1, \dots, G_{n+1}$, so that the ransom sets $\{G_{n+1:1}, \dots, G_{n+1:n+1}\}$ and $\{G_1, \dots, G_{n+1}\}$ are the same. Let also $G_{n+1:0} := 0$ and $G_{n+1:n+2} := 1$, so that

$$0 = G_{n+1:0} \leqslant G_{n+1:1} \leqslant \ldots \leqslant G_{n+1:n+1} \leqslant G_{n+1:n+2} = 1. \tag{1.2}$$

The main result of this note describes the cumulative distribution function (cdf) of each of the ordered gaps $G_{n+1:1}, \dots, G_{n+1:n+1}$:

Date: Wednesday 11$^{\text{th}}$ September, 2019.

2010 *Mathematics Subject Classification.* 62E15, 62F03.

Key words and phrases. Order statistics, spacings, uniform distribution are obtained, tests of significance.

Theorem 1.1. *For all* $k \in \overline{1, n+1}$ *and* $x \in [0, 1]$

$$\mathsf{P}(G_{n+1:k} > x) = (-1)^{k+1}(n+1)\binom{n}{k-1}\sum_{r=0}^{k-1}\frac{(-1)^r}{n-r+1}\binom{k-1}{r}\big(1-(n-r+1)x\big)_+^n. \tag{1.3}$$

Everywhere here, $u_+ := 0 \vee u = \max(0, u)$.

The proof of Theorem 1.1 is based on

Theorem 1.2. *Take any* $k \in \overline{0, n+1}$ *and* $x \in (0, 1)$*. Then*

$$\mathsf{P}(G_{n+1:k} \leqslant x < G_{n+1:k+1}) = (-1)^k\binom{n+1}{k}\sum_{r=0}^{k}(-1)^r\binom{k}{r}\big(1-(n-r+1)x\big)_+^n. \tag{1.4}$$

The special case of (1.3) with $k = n + 1$ is

$$\mathsf{P}(G_{n+1:n+1} > x) = \sum_{s=1}^{n+1}(-1)^{s-1}\binom{n+1}{s}\big(1-sx\big)_+^n. \tag{1.5}$$

Remark 1.3. It is now a textbook fact (see e.g. [1, Exercise 20, page 103]) that the joint distribution of the gaps $G_1, \dots, G_{n+1}$ is the same as that of $R_1, \dots, R_{n+1}$, where

$$R_i := \frac{X_i}{X_1 + \dots + X_{n+1}} \tag{1.6}$$

and the X_i's are iid (say standard) exponential random variables. Moran [7, page 93] ascribes mentioning of this fact to Fisher [4], and a proof of it – without a specific reference – to Clifford.

In fact, Fisher [4] used geometric arguments to obtain the following formula for the distribution of $R_{n+1:n+1} = \max_{1\leqslant i\leqslant n+1} R_i$:

$$\mathsf{P}(R_{n+1:n+1} > x) = \sum_{j=1}^{n+1}(-1)^{j-1}\binom{n+1}{j}(1-jx)_+^n. \tag{1.7}$$

In view of Remark 1.3, (1.5) is equivalent to (1.7).

Accordingly, one may replace $G_{n+1:k}$ in (1.3) by $R_{n+1:k}$ and thus obtain a generalization of (1.7).

In Fisher's setting, the X_i's were certain test statistics in independent tests of significance in the harmonic analysis of a series; these tests are labeled by $1, \dots, n+1$ in this note and by $1, \dots, n$ in [4]. The test statistics X_i are then normalized by in by the observable quantity $X_1 + \dots + X_{n+1}$, yielding the ratios R_i, as in (1.6). Thus, $\mathsf{P}(R_{n+1:n+1} > x)$ is the probability that at least one of the $n+1$ normalized test statistics R_i will exceed the critical value x.

Accordingly, the probability $\mathsf{P}(G_{n+1:k} > x)$ in (1.3), which equals $\mathsf{P}(R_{n+1:k} > x)$, is the probability that at least $\ell := n - k + 2$ of the $n+1$ normalized test statistics R_i will exceed the critical value x; note that here ℓ can take any value in $\overline{1, n+1}$. Similarly, the probability $\mathsf{P}(G_{n+1:k} > x)$ in (1.4) is the probability that exactly $\ell - 1 = n - k + 1$ of the $n+1$ normalized test statistics R_i will exceed the critical value x; note that here $\ell - 1$ can take any value in $\overline{0, n}$.

Thus, Theorems 1.1 and 1.2 above provide useful additional information concerning the independent tests considered by Fisher.

The proofs in this note are based on certain geometric, combinatorial, and analytic considerations.

The proof of (1.7) in [4] involves geometric ideas, apparently rather different from the ones used here. That proof in [4] appears very informal and sketchy, and some parts of it seem unclear. In particular, it seems unclear how the polynomial f in t and the polynomial P in g, introduced on page 57 in [4], are related to each other. It is also unclear how the ultimate expression for the probability in question was obtained in [4]. For these reasons as well, it may be of use to now have a formal proof of Fisher's result and its generalizations given in the present note.

The following result is also based on Theorem 1.2.

Corollary 1.4. *For all* $k \in \overline{0, n+1}$

$$\mathsf{E}\, G_{n+1:k} = \frac{H_{n+1} - H_{n+1-k}}{n+1}, \tag{1.8}$$

where

$$H_j := 1 + \frac{1}{2} + \dots + \frac{1}{j}$$

is the j*th harmonic number, with* $H_0 := 0$.

In particular, since $\frac{1}{r} \sim \int_r^{r+1} \frac{dx}{x}$ as $r \to \infty$, it follows from (1.8) that

$$\mathsf{E}\, G_{n+1:k} \sim \frac{1}{n} \ln \frac{n}{n-k} \quad \text{if} \quad n-k \to \infty;$$

as usual, we write $a \sim b$ for $a/b \to 1$. Further, if $k = o(n)$, then

$$\mathsf{E}\, G_{n+1:k} \sim \frac{k}{n^2},$$

so that $\mathsf{E}\, G_{n+1:k}$ is asymptotically linear in k. Further, taking $k = 1$, we also see that the smallest among the gaps $G_0, \dots, G_n$ is $n+1$ times as small on the average as the average of these $n+1$ gaps.

On the other hand, the expectation of the largest among the gaps $G_0, \dots, G_n$ is

$$\mathsf{E}\, G_{n+1:n+1} = \frac{H_{n+1}}{n+1} \sim \frac{\ln n}{n}$$

as $n \to \infty$, so that the largest gap is about $\ln n$ times as large on the average as the average of the gaps.

2. Proofs

Proof of Theorem 1.2. We begin with the following simple observation. Let μ be a measure defined on the Borel σ-algebra over $\mathbb{R}^n$ with a finite joint tail function T_μ defined by the formula

$$T_\mu(x) := \mu\big(Q(x)\big), \tag{2.1}$$

where

$$Q(x) := \prod_{i=1}^{n} (x_i, \infty) \tag{2.2}$$

is the "tail" orthant with the vertex $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

For any $a=(a_1,\dots,a_n)$ and $b=(b_1,\dots,b_n)$ in $\mathbb{R}^n$ such that $a_i\leqslant b_i$ for all $i\in\overline{1,n}$, consider the parallelepiped

$$\Pi_{a,b}:=\prod_{i=1}^n(a_i,b_i]. \tag{2.3}$$

Also, let

$$h=(h_1,\dots,h_n):=b-a$$

and, for each $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)\in\{0,1\}^n$, let $a+\varepsilon h:=(a_1+\varepsilon_1h_1,\dots,a_n+\varepsilon_nh_n)$ and $|\varepsilon|:=\varepsilon_1+\dots+\varepsilon_n$; note that the ith coordinate $a_i+\varepsilon_ih_i$ of the vector $a+\varepsilon h$ equals a_i or b_i depending on whether ε_i equals 0 or 1. As usual, let I_A denote the indicator function of a set A. Then

$$\mu(\Pi_{a,b})=\int_{\mathbb{R}^n}d\mu\prod_{i=1}^n\big(\,\mathsf{I}_{(a_i,\infty)}-\mathsf{I}_{(b_i,\infty)}\,\big) \tag{2.4}$$

$$=\int_{\mathbb{R}^n}d\mu\sum_{\varepsilon\in\{0,1\}^n}(-1)^{|\varepsilon|}\,\mathsf{I}_{Q(a+\varepsilon h)}, \tag{2.5}$$

whence

$$\mu(\Pi_{a,b})=\sum_{\varepsilon\in\{0,1\}^n}(-1)^{|\varepsilon|}\,T_\mu(a+\varepsilon h). \tag{2.6}$$

In particular, for $y\in\mathbb{R}$, $\alpha\in[0,\infty)$, and $\gamma=(\gamma_1,\dots,\gamma_n)\in\mathbb{R}^n$, consider now the measure $\mu_{y,\alpha,\gamma}$ that has the density with respect to the Lebesgue measure on $\mathbb{R}^n$ given by the formula

$$\frac{d\mu_{y,\alpha,\gamma}}{dx}=(y-\gamma\cdot x)_+^\alpha \tag{2.7}$$

for $x=(x_1,\dots,x_n)\in\mathbb{R}^n$, where $\gamma\cdot x:=\sum_1^n\gamma_ix_i$ and (concerning the case $\alpha=0$) $0^0:=0$. Then, using induction on n or, more specifically, iterated integration, it is easy to see that

$$T_{\mu_{y,\alpha,\gamma}}(x)=\frac{(y-\gamma\cdot x)_+^{\alpha+n}}{\prod_1^n\big((\alpha+i)\gamma_i\big)} \tag{2.8}$$

for $x=(x_1,\dots,x_n)\in\mathbb{R}^n$.

Choosing now $\alpha=0$, we get

$$\operatorname{vol}_n(\Pi_{a,b}\cap H_{\gamma,y})=\frac{1}{n!\prod_1^n\gamma_i}\sum_{\varepsilon\in\{0,1\}^n}(-1)^{|\varepsilon|}\,\big(y-\gamma\cdot(a+\varepsilon h)\big)_+^n, \tag{2.9}$$

where vol_n denotes the volume in $\mathbb{R}^n$ and $H_{\gamma,y}:=\big\{x\in\mathbb{R}^n\colon\gamma\cdot x\leqslant y\big\}$. For $a=(0,\dots,0)$ and $b=(1,\dots,1)$, formula (2.9) was given in [2] and [6]. (The condition that $\gamma_i>0$ for all i was missing in [2].)

Next, note that the joint probability density function (pdf), say f, of the order statistics $U_{n:1},\dots,U_{n:n}$ is given by the formula

$$f(y_1,\dots,y_n)=n!\,\mathsf{I}\{0<y_1<\dots<y_n<1\} \tag{2.10}$$

for $(y_1,\dots,y_n)\in\mathbb{R}^n$; see e.g. [3, page 12]. Since, in view of (1.1), the r.v.'s $G_1,\dots,G_n$ are obtained from $U_{n:1},\dots,U_{n:n}$ by a linear transformation with determinant 1, we see that the joint pdf, say g, of the gaps $G_1,\dots,G_n$ is given by the

formula

$$g(z_1,\dots,z_n)=n!\,\mathsf{I}\{z_1>0,\dots,z_n>0,z_1+\dots+z_n<1\} \tag{2.11}$$

for $(z_1,\dots,z_n)\in\mathbb{R}^n$.

Take now any $j\in\overline{0,n}$, $x\in(0,1)$, and $y\in(0,1]$, and let

$$p_{n,j}(x,y):=\mathsf{P}\Big(G_i\leqslant x\ \forall i\in\overline{1,j},\ G_i>x\ \forall i\in\overline{j+1,n},\ \sum_1^n G_i<y\Big). \tag{2.12}$$

Then, by (2.11),

$$p_{n,j}(x,y)=n!\operatorname{vol}_n(\Pi_{a^{j,x},b^{j,x}}\cap H_{\gamma_1,y}), \tag{2.13}$$

where $a_i^{j,x}:=0$ and $b_i^{j,x}:=x$ for $i\in\overline{1,j}$, $a_i^{j,x}:=x$ and $b_i^{j,x}:=1$ for $i\in\overline{j+1,n}$, and $\gamma_1:=(1,\dots,1)\in\mathbb{R}^n$.

So, letting $|\varepsilon|_*:=\sum_1^j\varepsilon_i$ and $|\varepsilon|_{**}:=\sum_{j+1}^n\varepsilon_i$ for $\varepsilon\in\{0,1\}^n$ and using (2.9), we have

$$\begin{aligned}p_{n,j}(x,y)&=\sum_{\varepsilon\in\{0,1\}^n}(-1)^{|\varepsilon|}\big(y-(n-j)x-|\varepsilon|_*x-|\varepsilon|_{**}(1-x)\big)_+^n\\&=\sum_{\alpha=0}^{j}\sum_{\beta=0}^{n-j}(-1)^{\alpha+\beta}\binom{j}{\alpha}\binom{n-j}{\beta}\big(y-(n-j)x-\alpha x-\beta(1-x)\big)_+^n.\end{aligned}$$

Note also that $(n-j-\beta+\alpha)x\geqslant0$ for $\alpha\in\overline{0,j}$, $\beta\in\overline{0,n-j}$, and $x\in(0,1)$, so that $\big(y-(n-j)x-\alpha x-\beta(1-x)\big)_+=\big(y-\beta-(n-j-\beta+\alpha)x\big)_+=0$ for $y\in(0,1]$ and $\beta\in\overline{1,n-j}$. Therefore, the latter displayed expression for $p_{n,j}(x,y)$ greatly simplifies:

$$p_{n,j}(x,y)=\sum_{\alpha=0}^{j}(-1)^{\alpha}\binom{j}{\alpha}\big(y-(n-j+\alpha)x\big)_+^n \tag{2.14}$$

$$=\sum_{r=0}^{j}(-1)^{j-r}\binom{j}{r}\big(y-(n-r)x\big)_+^n \tag{2.15}$$

$$=\sum_{r=-\infty}^{\infty}(-1)^{j-r}\binom{j}{r}\big(y-(n-r)x\big)_+^n. \tag{2.16}$$

The latter equality holds because

$$\binom{j}{r}=0\quad\text{for}\quad j\in\overline{0,\infty}\quad\text{and}\quad r\in\overline{-\infty,-1}\cup\overline{j+1,\infty}, \tag{2.17}$$

with $\binom{j}{r}$ understood in the combinatorial sense, as the cardinality of the set $\binom{[j]}{r}$ of all subsets of the set

$$[j]:=\overline{1,j}$$

of cardinality r; for instance, $\binom{j}{-1}=0$ for all $j\in\overline{0,\infty}$ because any set has no subsets of cardinality -1. For another, analytic approach to generalized binomial coefficients $\binom{j}{r}$, which leads to the same results for $j\in\overline{0,\infty}$ and $r\in\overline{-\infty,-1}\cup\overline{j+1,\infty}$, see e.g. [5].

After these preliminary observations, we are ready to consider the probability in (1.4):

$$P_{n,k}:=\mathsf{P}(G_{n+1:k}\leqslant x<G_{n+1:k+1})=Q_{n,k}+R_{n,k}, \tag{2.18}$$

where

$$Q_{n,k} := \sum_{J\in\binom{[n]}{k}} \mathsf{P}\Big(G_i \leqslant x\ \forall i\in J,\ G_i > x\ \forall i\in[n]\setminus J,\ \sum_1^n G_i < 1-x\Big) \tag{2.19}$$

$$= \binom{n}{k} p_{n,k}(x, 1-x), \tag{2.20}$$

$$R_{n,k} := \sum_{J\in\binom{[n]}{k-1}} \mathsf{P}\Big(G_i \leqslant x\ \forall i\in J,\ G_i > x\ \forall i\in[n]\setminus J,\ 1 > \sum_1^n G_i \geqslant 1-x\Big) \tag{2.21}$$

$$= \binom{n}{k-1}\big(p_{n,k-1}(x,1) - p_{n,k-1}(x,1-x)\big); \tag{2.22}$$

here we used the definition of $p_{n,k}(x,y)$ in (2.12) and the fact that, in view of (2.11), the r.v.'s $G_1,\dots,G_n$ are exchangeable.

In Theorem 1.2, k may take any value in the set $\overline{0,n+1}$, whereas the expression in (2.16) for $p_{n,j}(x,y)$ was established only for $j\in\overline{0,n}$. However, $Q_{n,n+1}=0$ because $\binom{n}{n+1}=0$, and $R_{n,0}=0$ because $\binom{n}{-1}=0$, whereas $k-1\in\overline{0,n}$ for $k\in\overline{1,n+1}$. It follows that, for all $k\in\overline{0,n+1}$, we can replace all entries of $p_{n,\cdot}(x,\cdot)$ in the above expressions for $Q_{n,k}$ and $R_{n,k}$ by the corresponding expressions according to (2.16). Thus, letting now

$$a_r := a_{n,r}(x) := (-1)^r\big(y-(n-r)x\big)_+^n, \tag{2.23}$$

we have

$$(-1)^{k+1}P_{n,k} = \binom{n}{k}\sum_{r=-\infty}^{\infty}\binom{k}{r}a_{r-1} \tag{2.24}$$

$$+\binom{n}{k-1}\sum_{r=-\infty}^{\infty}\binom{k-1}{r}a_r \tag{2.25}$$

$$+\binom{n}{k-1}\sum_{r=-\infty}^{\infty}\binom{k-1}{r}a_{r-1} \tag{2.26}$$

$$=\sum_{r=-\infty}^{\infty}\left(\binom{n}{k}\binom{k}{r}+\binom{n}{k-1}\binom{k-1}{r-1}+\binom{n}{k-1}\binom{k-1}{r}\right)a_{r-1} \tag{2.27}$$

$$=\sum_{r=-\infty}^{\infty}\left(\binom{n}{k}\binom{k}{r}+\binom{n}{k-1}\binom{k}{r}\right)a_{r-1} \tag{2.28}$$

$$=\sum_{r=-\infty}^{\infty}\binom{n+1}{k}\binom{k}{r}a_{r-1} = \binom{n+1}{k}\sum_{r=-\infty}^{\infty}\binom{k}{r}a_{r-1}. \tag{2.29}$$

Now (1.4) immediately follows, in view of (2.18) and (2.23). $\square$

Proof of Theorem 1.1. For all $j \in \overline{0, n+1}$

(2.30)

$$\mathsf{P}(G_{n+1:j+1} > x) - \mathsf{P}(G_{n+1:j} > x) = \mathsf{P}(G_{n+1:j} \leqslant x < G_{n+1:j+1})$$

$$= P_{n,j} = (-1)^{j+1}\binom{n+1}{j}\sum_{r=0}^{j}\binom{j}{r}a_{r-1}, \tag{2.31}$$

the latter two equalities holding by virtue of (2.18), (1.4), and (2.23). Also, in view of (1.2) and because $x \in (0,1)$, we have $\mathsf{P}(G_{n+1:0} > x) = 0$. So, one can find $\mathsf{P}(G_{n+1:k} > x)$ by summation:

$$\mathsf{P}(G_{n+1:k} > x) = \sum_{j=0}^{k-1} P_{n,j}, \tag{2.32}$$

and this is how the expression of $\mathsf{P}(G_{n+1:k} > x)$ in (1.3) was actually found.

However, once that expression has been obtained, it is sufficient – and much easier – to verify (1.3) by checking the identity

$$V_{n,j+1} - V_{n,j} \overset{(?)}{=} P_{n,j} \tag{2.33}$$

for all $j \in \overline{0, n}$, where

$$V_{n,j} := (-1)^{j+1}(n+1)\binom{n}{j-1}\sum_{r=0}^{j-1}\frac{a_{r-1}}{n-r+1}\binom{j-1}{r}, \tag{2.34}$$

the right-hand side of (1.3) with j in place of k, taking also (2.23) into account; hence,

$$V_{n,j+1} := (-1)^{j}(n+1)\binom{n}{j}\sum_{r=0}^{j}\frac{a_{r-1}}{n-r+1}\binom{j}{r}. \tag{2.35}$$

Indeed, it will immediately follow from (2.31) and (2.33) that

$$\mathsf{P}(G_{n+1:j+1} > x) - \mathsf{P}(G_{n+1:j} > x) = V_{n,j+1} - V_{n,j} \tag{2.36}$$

for $j \in \overline{0, n}$. Since $\mathsf{P}(G_{n+1:0} > x) = 0 = V_{n,0}$, it will then follow by induction on k or, equivalently, by telescoping summation, that $\mathsf{P}(G_{n+1:k} > x) = V_{n,k}$ for all $k \in \overline{0, n+1}$, which will complete the proof of Theorem 1.1.

Turning now back to identity (2.33), we see that each side of it equals $-a_{-1}$ when $j = 0$.

Next, it is convenient to replace $\sum_{r=0}^{j}$ in (2.31) and (2.35), as well as $\sum_{r=0}^{j-1}$ in (2.34), by $\sum_{r=0}^{n}$; in view of (2.17) and the condition $j \in \overline{0, n}$, these replacements will not affect the values of the corresponding expressions for $P_{n,j}$, $V_{n,j+1}$, and $V_{n,j}$.

Thus, it suffices to check that the coefficients of the a_{r-1}'s on both sides of (2.33) are the same for all $j \in \overline{1, n}$ and $r \in \overline{0, n}$, which amounts to checking the identity

$$\frac{n+1}{n-r+1}\left(\binom{n}{j}\binom{j}{r} + \binom{n}{j-1}\binom{j-1}{r}\right) \overset{(?)}{=} \binom{n+1}{j}\binom{j}{r} \tag{2.37}$$

for such j and r, which in turn becomes immediately obvious on replacing $\binom{n}{j}$, $\binom{n}{j-1}$, and $\binom{j-1}{r}$ there by the corresponding equal expressions $\binom{n+1}{j}\frac{n+1-j}{n+1}$, $\binom{n+1}{j}\frac{j}{n+1}$, and $\binom{j}{r}\frac{j-r}{j}$. Theorem 1.1 is now proved. □

Proof of Corollary 1.4. Take any $j \in \overline{0,n}$. Then, in view of (1.2) and Theorem 1.2,

$$\mathsf{E}(G_{n+1:j+1} - G_{n+1:j}) = \mathsf{E}\int_0^1 dx\ \mathsf{I}\{G_{n+1:j} \leqslant x < G_{n+1:j+1}\} \tag{2.38}$$

$$= \int_0^1 dx\ \mathsf{P}(G_{n+1:j} \leqslant x < G_{n+1:j+1}) \tag{2.39}$$

$$= (-1)^j \binom{n+1}{j} \sum_{r=0}^{j} (-1)^r \binom{j}{r} \int_0^1 dx\, \big(1-(n-r+1)x\big)_+^n \tag{2.40}$$

$$= (-1)^j \binom{n+1}{j} \frac{S}{n+1}, \tag{2.41}$$

where

$$S := \sum_{r=0}^{j} (-1)^r \binom{j}{r} \frac{1}{n-r+1} \tag{2.42}$$

$$= \sum_{r=0}^{j} (-1)^r \binom{j}{r} \int_0^1 du\, u^{n-r} \tag{2.43}$$

$$= \int_0^1 du\, u^n \sum_{r=0}^{j} (-u)^{-r} \binom{j}{r} \tag{2.44}$$

$$= \int_0^1 du\, u^n (1-1/u)^j = (-1)^j \frac{(n-j)!j!}{(n+1)!}. \tag{2.45}$$

So,

$$\mathsf{E}\, G_{n+1:j+1} - \mathsf{E}\, G_{n+1:j} = \mathsf{E}(G_{n+1:j+1} - G_{n+1:j}) = \frac{1}{(n+1)(n+1-j)}. \tag{2.46}$$

Also, again in view of (1.2), $G_{n+1:0} = 0$ and hence $\mathsf{E}\, G_{n+1:0} = 0$. Thus,

$$\mathsf{E}\, G_{n+1:k} = \sum_{j=0}^{k-1} (\mathsf{E}\, G_{n+1:j+1} - \mathsf{E}\, G_{n+1:j}) = \sum_{j=0}^{k-1} \frac{1}{(n+1)(n+1-j)} = \frac{H_{n+1} - H_{n+1-k}}{n+1}, \tag{2.47}$$

which completes the proof of Corollary 1.4. □

The following alternative proof of Corollary 1.4 is more direct, as it does not rely on Theorem 1.2. Instead, it uses the more elementary Remark 1.3.

"Direct" proof of Corollary 1.4. Let $X_1, \dots, X_{n+1}$ be as in Remark 1.3, and then let $X_{n+1:1} \leqslant \dots \leqslant X_{n+1:n+1}$ be the corresponding order statistics. Then for the r.v.'s $R_1, \dots, R_{n+1}$ defined by (1.6) and the corresponding order statistics $R_{n+1:1} \leqslant \dots \leqslant R_{n+1:n+1}$ we have

$$R_{n+1:k} := \frac{X_{n+1:k}}{X_{n+1:n+1} + \dots + X_{n+1:n+1}}. \tag{2.48}$$

The joint pdf, say h, of $X_{n+1:1} \leqslant \dots \leqslant X_{n+1:n+1}$ is given by the formula

$$h(x_1, \dots, x_{n+1}) = (n+1)! e^{-w_{n+1}}\, \mathsf{I}\{0 < x_1 < \dots < x_{n+1}\} \tag{2.49}$$

for $(x_1,\dots,x_{n+1})\in\mathbb{R}^{n+1}$, where

$$w_j := x_1+\dots+x_j;$$

see e.g. [3, page 12] again.

We will also need the following very simple but crucial observation: for any real $u>0$

$$\frac{1}{u}=\int_0^\infty dt\, e^{-tu}. \tag{2.50}$$

One can view this as a decomposition of the inconvenient function $u\mapsto\frac{1}{u}$ into the nice "harmonics" $u\mapsto e^{-tu}$.

Now, introducing

$$L_{n+1,j} := \int_{S_{n+1}} dx_{1,n+1}\, x_j\, e^{-w_{n+1}},$$

where

$$S_j := \{(x_1,\dots,x_j)\colon 0<x_1<\dots<x_j\} \quad\text{and}\quad dx_{1,j} := dx_1\cdots dx_j$$

for natural j, and using Remark 1.3, (2.48), (2.49), and (2.50) (with $u=w_{n+1}$), we can write

$$\begin{aligned}
\mathsf{E}\, G_{n+1:k} &= \mathsf{E}\, R_{n+1:k}\\
&= (n+1)!\int_{S_{n+1}} dx_{1,n+1}\, e^{-w_{n+1}}\,\frac{x_k}{w_{n+1}}\\
&= (n+1)!\int_0^\infty dt \int_{S_{n+1}} dx_{1,n+1}\, e^{-w_{n+1}}\, x_k\, e^{-tw_{n+1}}\\
&= (n+1)!\int_0^\infty dt \int_{S_{n+1}} dx_{1,n+1}\, x_k\, e^{-(1+t)w_{n+1}}\\
&= (n+1)!\int_0^\infty \frac{dt}{(1+t)^{n+2}}\, L_{n+1,k} = n! L_{n+1,k}.
\end{aligned} \tag{2.51}$$

It remains to evaluate $L_{n+1,k}$. Toward this end, for positive real $t_1,\dots,t_{n+1}$ consider

$$\begin{aligned}
M(t_1,\dots,t_{n+1}) &:= \int_{S_{n+1}} dx_{1,n+1}\, e^{-t_1x_1-\dots-t_{n+1}x_{n+1}}\\
&= \frac{1}{t_{n+1}}\int_{S_{n-2}} dx_{1,n}\, e^{-t_1x_1-\dots-t_{n-1}x_{n-1}-(t_{n+1}+t_n)x_n}\\
&\;\;\vdots\\
&= \frac{1}{t_{n+1}(t_{n+1}+t_n)\dots(t_{n+1}+\dots+t_1)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
L_{n+1,k} &= -\frac{d}{dh}\, M(\underbrace{1,\dots,1}_{k-1}, 1+h, \underbrace{1,\dots,1}_{n+1-k})\Big|_{h=0} \\
&= -\frac{d}{dh}\,\frac{1}{(n+1-k)!(n+2-k+h)\cdots(n+1+h)}\Big|_{h=0} \\
&= \frac{1}{(n+1)!}\Big(\frac{1}{n+2-k}+\dots+\frac{1}{n+1}\Big) = \frac{H_{n+1}-H_{n+1-k}}{(n+1)!}.
\end{aligned}$$

Now Corollary 1.4 follows by (2.51). □

References

1. Barry C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A first course in order statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1992, A Wiley-Interscience Publication. MR 1178934
2. D. L. Barrow and P. W. Smith, *Spline Notation Applied to a Volume Problem*, Amer. Math. Monthly **86** (1979), no. 1, 50–51. MR 1538918
3. H. A. David and H. N. Nagaraja, *Order statistics*, third ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003. MR 1994955
4. R. A. Fisher, *Tests of significance in harmonic analysis*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **125** (1929), no. 796, 54–59.
5. David Fowler, *The binomial coefficient function*, The American Mathematical Monthly **103** (1996), no. 1, 1–17.
6. Jim Lawrence, *Polytope volume computation*, Math. Comp. **57** (1991), no. 195, 259–271. MR 1079024
7. P. A. P. Moran, *The random division of an interval*, Supplement to the Journal of the Royal Statistical Society **9** (1947), no. 1, 92–98.
8. R. Pyke, *Spacings*, Journal of the Royal Statistical Society. Series B (Methodological) **27** (1965), no. 3, 395–449.

Department of Mathematical Sciences, Michigan Technological University, Houghton, Michigan 49931

E-mail address: `ipinelis@mtu.edu`