Ann Inst Stat Math (2017) 69:417–428
DOI 10.1007/s10463-015-0545-6

Lower and upper bounds on the variances of spacings

Paweł Marcin Kozyra[1] · **Tomasz Rychlik**[1]

Received: 22 April 2015 / Revised: 23 June 2015 / Published online: 19 September 2015

Abstract For classic i.i.d. samples with an arbitrary nondegenerate and finite variance distribution, Papadatos (1995, Annals of the Institute of Statistical Mathematics, 47, 185–193) presented sharp lower and upper bounds on the variances of order statistics, expressed in population variance units. We provide here analogous results for spacings. Also, we describe the parent distributions which attain the bounds.

Keywords Order statistic · Spacing · Variance · Sharp bound · Bernstein polynomial

1 Introduction

We assume that a random variable X has a positive and finite variance. Let $X_1, \dots, X_n$ denote n i.i.d. copies of X, and $X_{1:n} \leq \cdots X_{n:n}$ stand for the respective order statistics. Their spacings are defined by $S_{i:n} = X_{i+1:n} - X_{i:n}$, $i = 1, \dots, n-1$. Under the above assumptions, we derive sharp lower and upper bounds for the variance ratios $\mathbb{V}ar\ S_{i:n}/\mathbb{V}ar\ X$ for all $1 \leq i < n < \infty$. Also, we describe the families of two-point distributions which attain the bounds, possibly in the limit.

Spacings play important roles in various problems of statistical inference and other branches of applied probability. Comprehensive discussions of their properties and applications, especially in constructing goodness-of-fit tests, are presented, e.g., in Pyke (1965, 1972) and David and Nagaraja (2003). Various evaluations of the expectations of linear combinations of order statistics, and spacings in particular, were

✉ Tomasz Rychlik
trychlik@impan.pl

Paweł Marcin Kozyra
pawel_m_kozyra@wp.pl

[1] Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00656 Warsaw, Poland

presented in the literature. The first ones are due to Moriguti (1953) who derived sharp bounds on expected spacings in population standard deviation units. Raqab (2003) presented optimal upper bounds on the expectations of spacings in more general scale units, generated by central absolute populations moments of various orders $p \geq 1$. Danielak (2004) extended these results to arbitrary quasi-ranges, i.e., differences of order statistics $X_{j:n} - X_{i:n}$, $1 \leq i < j \leq n$. Kozyra and Rychlik (2015) obtained tight lower and upper bounds on the differences of expected order statistics measured in Gini mean difference units. More stringent standard deviation bounds in the restricted families of decreasing density and decreasing failure rate distributions were determined by Danielak and Rychlik (2004). More general families of distributions with decreasing density and failure rate on the average were studied in Danielak and Rychlik (2003). Recently, Goroncy and Rychlik (2015a, b) presented analogous results for the distributions with increasing density and increasing failure rate functions, respectively.

By far less is known about evaluations of variances of order statistics and their functions. Papadatos (1995) determined sharp lower and upper bounds on variances of single order statistics, expressed in terms of single observation variance units. The upper bound for the special case of sample median was earlier presented in Yang (1982), and its tightness was proved by Lin and Huang (1989). Papadatos (1997) refined these results for the families of symmetric parent distributions. More precise solution to the problem was presented in Jasiński and Rychlik (2013). Much earlier, lower and upper bounds for the variances of sample extremes were delivered by Moriguti (1951).

This paper is a first attempt at evaluating the variances of nontrivial linear combinations of order statistics. Main results are presented in Sect. 2. Section 3 contains their proofs.

2 Results

Using the assumptions and notation of the first paragraph of Sect. 1 which shall hold throughout the whole paper, we state the following.

Proposition 1 *For arbitrary fixed* $1 \leq i < n < \infty$, *the bound*

$$\frac{\mathbb{V}ar\, S_{i:n}}{\mathbb{V}ar\, X} \leq \max_{0 \leq u \leq 1} g_{i,n}(u) \tag{1}$$

with

$$g_{i,n}(u) := \binom{n}{i} u^{i-1}(1-u)^{n-i-1} \left[1 - \binom{n}{i} u^i (1-u)^{n-i}\right] \tag{2}$$

is sharp. Let

$$F_u(x) := \begin{cases} 0, & \text{if } x < a, \\ u, & \text{if } a \leq x < b, \\ 1, & \text{if } x \geq b, \end{cases} \tag{3}$$

denote the family of two-point distribution functions with arbitrary $a < b$ *and* $0 < u < 1$. *If* $\max_{0 \leq u \leq 1} g_{i,n}(u) = g_{i,n}(u_*)$ *for some* $u_* = u_*(i, n) \in (0, 1)$, *then the* *upper bound in* (1) *is attained iff the parent distribution function is* (3) *with* $u = u_*$. *If* $\max_{0\le u\le 1} g_{i,n}(u) = g_{i,n}(0)$ ($g_{i,n}(1)$, *respectively), then this is attained in the limit by the parent distribution functions* (3) *with* $u \downarrow 0$ ($u \uparrow 1$, *respectively). For* $1 \le i < n \ge 3$, *the trivial inequality*

$$\frac{\mathbb{V}ar\, S_{i:n}}{\mathbb{V}ar\, X} \ge 0 \tag{4}$$

cannot be improved, and becomes equality in the limit for the parent distribution functions (3) *with* $u \downarrow 0$ *when* $i \ge 2$ *and* $u \uparrow 1$ *when* $i \le n-2$.

Note that

$$g_{i,n}(u) = \binom{n}{i} u^{i-1}(1-u)^{n-i-1} \sum_{\substack{j=0\\ j\ne i}}^{n} B_{j,n}(u),$$

where

$$B_{j,n}(u) := \binom{n}{j} u^{j}(1-u)^{n-j}, \qquad 0 \le j \le n,$$

denote the Bernstein polynomials of degree n. This implies that $g_{i,n}(u) > 0$ for all $n \in \mathbb{N}$, $i \in \{1, \dots, n-1\}$ and $u \in (0,1)$. Moreover, $g_{i,n}(u) = 0$ if either $i \ge 2$ and $u = 0$ or $i \le n-2$ and $u = 1$. This observation is intimately connected with the tight zero lower bound of Proposition 1. Also, relation $g_{i,n}(u) = g_{n-i,n}(1-u)$ together with (1) imply that the upper bounds for the variances of $S_{i:n}$ and $S_{n-i:n}$ coincide. The same conclusion for the lower bounds results from the last claim of Proposition 1.

In Lemmas 1 and 2, we describe maxima of (2) for various parameters i and n.

Lemma 1 *For every* $n \ge 3$

(i) *function* $g_{1,n}$ *has a unique maximum at* 0, *and* $g_{1,n}(0) = n$,
(ii) *function* $g_{n-1,n}$ *has a unique maximum at* 1, *and* $g_{n-1,n}(1) = n$.

Lemma 2 *Fix* $n \ge 4$ *and* $2 \le i \le n-2$. *Function* (2) *has either a unique local and global maximum or two local maxima and one local minimum between them. The local extreme arguments are the only zeros of the polynomial*

$$h_{i,n}(u) = [2(n-1)u - 2i + 1]B_{i,n}(u) - u(n-2) + i - 1. \tag{5}$$

Let $u_* = u_*(i,n)$ *denote the global maximum point. Then* $u_*(2,4) \in \left\{\frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}, \frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right\}$ *and* $g_{2,4}(u_*(2,4)) = \frac{2\sqrt{2}}{3} \approx 0.94281$. *For* $n > 4$ *yields*

(i) *if* $i < \frac{n}{2}$ ($i > \frac{n}{2}$), *then* $u_*(i,n) < \frac{1}{2}$ ($u_*(i,n) > \frac{1}{2}$, *respectively),*
(ii) *if* $n \ge 6$ *is even, then* $u_*\left(\frac{n}{2}, n\right) = \frac{1}{2}$ *and*

$$g_{\frac{n}{2},n}\left(\frac{1}{2}\right) = \binom{n}{\frac{n}{2}} \frac{1}{2^{n-2}} \left[1 - \binom{n}{\frac{n}{2}} \frac{1}{2^{n}}\right].$$

The proof of the first statement of Lemma 2 is based on the following variation diminishing property (VDP, for brevity) of Bernstein polynomials of Schoenberg (1959).

Lemma 3 [cf., e.g., Rychlik (2001), Lemma 14] *The number of sign changes of a non-zero linear combination of Bernstein polynomials $\sum_{k=0}^{m} b_k B_{k,m}$ of degree m on the interval $(0, 1)$ does not exceed the number of the sign changes of the sequence $(b_0, \ldots, b_m)$. Moreover, the signs of the combination at the right neighborhood of 0 and the left neighborhood of 1 coincide with the signs of the first and last non-zero elements of the sequence, respectively.*

We are not able to arbitrate theoretically which functions $g_{i,n}$, $2 \leq i \leq n-2$, $i \neq \frac{n}{2}$, have one and two local maxima. Also, in the latter case, we not not have tools for deciding if both the local maxima are located in the same half of the unit interval. Numerical analysis of functions (2) for small n shows that two maxima appear only for $i = 2$, $n = 4$ (see Lemma 2). If n increases, the possibility of two maxima becomes less likely. Note that (2) can be represented as a linear combination of Bernstein polynomials $B_{j,2n-2}$, $j = i-1, \ldots, 2i-2, 2i, \ldots, 2n-i-1$, with positive coefficients. The full such combination with $j = i-1, \ldots, 2n-i-1$ amounts to $\binom{n}{i} u^{i-1}(1-u)^{n-i-1}$ which is certainly unimodal. It seems that removing one component with $j = 2i-1$ does not violate the property, and becomes almost negligible, especially for large n.

Using Lemmas 1 and 2 we are able to specify general result of Proposition 1 for particular $i \in \{1, \ldots, n-1\}$ and $n \in \mathbb{N}$. Only case $i = 1$ and $n = 2$ described in Proposition 2 needs an additional justification. Propositions 3 and 4 are direct conclusions of Proposition 1 and Lemmas 1 and 2.

Proposition 2 *We have*

$$\frac{2}{3} \leq \frac{\mathbb{V}ar\, S_{1:2}}{\mathbb{V}ar\, X} \leq g_{1,2}(0) = g_{1,2}(1) = 2.$$

The lower inequality becomes equality iff X is uniformly distributed.

Writing here and later that $\frac{\mathbb{V}ar\, S_{i:n}}{\mathbb{V}ar\, X} \leq (\geq) g_{i,n}(u_*)$, we mean that the upper (lower, respectively) bound amounts to $g_{i,n}(u_*)$ and is attained by the two-point distribution (3) with $u = u_*$ if $0 < u_* < 1$, and in the limit by a sequence of F_u with $u \to u_*$ if $u_* = 0$ or $u_* = 1$. We use the convention for the sake of brevity.

Proposition 3 *If $n \geq 3$, then:*

$$0 = g_{1,n}(1) \leq \frac{\mathbb{V}ar\, S_{1:n}}{\mathbb{V}ar\, X} \leq g_{1,n}(0) = n,$$
$$0 = g_{n-1,n}(0) \leq \frac{\mathbb{V}ar\, S_{n-1:n}}{\mathbb{V}ar\, X} \leq g_{n-1,n}(1) = n.$$

Proposition 4 *If $n \geq 4$ and $2 \leq i \leq n-2$, then*

$$0 = g_{i,n}(0) = g_{i,n}(1) \leq \frac{Var\, S_{i:n}}{\mathbb{V}ar\, X} \leq g_{i,n}(u_*),$$

Table 1 Upper bounds on variances of spacings from samples of size $n = 20$

i	$u_*(i, 20)$	$g_{i,20}(u_*(i, 20))$	i	$u_*(i, 20)$	$g_{i,20}(u_*(i, 20))$
1	0	20	6	0.27038	0.75942
2	0.04347	3.25396	7	0.32794	0.67092
3	0.09792	1.71152	8	0.38537	0.61799
4	0.15502	1.17002	9	0.44271	0.58958
5	0.21270	0.90714	10	0.5	0.58061

where u_* is described in Lemma 2.

In particular, for even n and $i = \frac{n}{2}$, we have

$$0 = g_{2,4}(0) = g_{2,4}(1) \le \frac{\mathbb{V}ar\ S_{2:4}}{\mathbb{V}ar\ X} \le g_{2,4}\left(\frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}\right)$$
$$= g_{2,4}\left(\frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right) = \frac{2\sqrt{2}}{3} \approx 0.94281,$$
$$0 = g_{\frac{n}{2},n}(0) = g_{\frac{n}{2},n}(1) \le \frac{\mathbb{V}ar\ S_{\frac{n}{2}:n}}{\mathbb{V}ar\ X} \le g_{\frac{n}{2},n}\left(\frac{1}{2}\right)$$
$$= \binom{n}{\frac{n}{2}} \frac{1}{2^{n-2}} \left[1 - \binom{n}{\frac{n}{2}} \frac{1}{2^n}\right], \qquad n \ge 6.$$

Table 1 presents numerical values of upper bounds $g_{i,20}(u_*(i, 20))$ on variances of spacings $S_{i:20}$ for samples of size $n = 20$ and $1 \le i \le 10$, together with respective arguments $u_*(i, 20)$ which describe the two-point distribution functions (3) attaining the bounds. Respective values for $11 \le i \le 19$ are immediately deduced from the relations $u_*(i, n) = 1 - u_*(n - i, n)$ and $g_{i,n}(u_*(i, n)) = g_{n-i,n}(u_*(n - i, n))$. We can see that if i increases from 1 to 10, then $u_*(i, 20)$ increases from 0 to 0.5, whereas $g_{i,20}(u_*(i, 20))$ decreases from 20 to 0.58061. From Proposition 4 and the Stirling formula we deduce that the upper bounds for the central spacings with $i = \frac{n}{2}$ decrease to 0 at the rate $4\sqrt{\frac{2\pi}{n}}$ as n increases to infinity. By Proposition 3, the respective bounds for the extreme spacings tend to infinity faster.

3 Proofs

Proof of Propostion 1. If $X_1, \ldots, X_n$ are i.i.d. copies of X with a common cumulative distribution function F, then

$$\mathbb{V}ar\ X = \iint_{\mathbb{R}^2} F(x \wedge y)\bar{F}(x \vee y)\mathrm{d}x\mathrm{d}y, \qquad (6)$$

$$\mathbb{E}\, S_{i:n} = \binom{n}{i} \int_{\mathbb{R}} F^i(x)\bar{F}^{n-i}(x)\mathrm{d}x,$$
$$\mathbb{E}\, S_{i:n}^2 = \binom{n}{i} \iint_{\mathbb{R}^2} F^i(x \wedge y)\bar{F}^{n-i}(x \vee y)\mathrm{d}x\mathrm{d}y, \tag{7}$$

where $x \wedge y = \min\{x, y\}$, $x \vee y = \max\{x, y\}$, and $\bar{F}(x) = 1 - F(x)$ denotes the survival function of F. The first formula is ascribed to Hoeffding (1940). The others were presented by Pearson (1902) and Irwin (1925), respectively. These and many other similar formulas may be also found in Jones and Balakrishnan (2002). We also have

$$(\mathbb{E}\, S_{i:n})^2 = \binom{n}{i}^2 \iint_{\mathbb{R}^2} F^i(x)\bar{F}^{n-i}(x)F^i(y)\bar{F}^{n-i}(y)\mathrm{d}x\mathrm{d}y$$
$$= \binom{n}{i}^2 \iint_{\mathbb{R}^2} F^i(x \wedge y)\bar{F}^{n-i}(x \wedge y)F^i(x \vee y)\bar{F}^{n-i}(x \vee y)\mathrm{d}x\mathrm{d}y.$$

In consequence,

$$\begin{aligned}
\mathbb{V}ar\, S_{i:n} &= \iint_{\mathbb{R}^2} \left[\binom{n}{i} F^i(x \wedge y)\bar{F}^{n-i}(x \vee y)\right.\\
&\quad \left. - \binom{n}{i}^2 F^i(x \wedge y)\bar{F}^{n-i}(x \wedge y)F^i(x \vee y)\bar{F}^{n-i}(x \vee y)\right]\mathrm{d}x\mathrm{d}y\\
&= \iint_{\mathbb{R}^2} \binom{n}{i} F^{i-1}(x \wedge y)\bar{F}^{n-i-1}(x \vee y)\\
&\quad \times \left[1 - \binom{n}{i} F^i(x \wedge y)\bar{F}^{n-i}(x \vee y)\right] F(x \wedge y)\bar{F}(x \vee y)\mathrm{d}x\mathrm{d}y\\
&\leq \max_{0 \leq u = F(x\wedge y) \leq v = F(x \vee y) \leq 1} f_{i,n}(u, v) \iint_{\mathbb{R}^2} F(x \wedge y)\bar{F}(x \vee y)\mathrm{d}x\mathrm{d}y,
\end{aligned} \tag{8}$$

where

$$f_{i,n}(u, v) := \binom{n}{i} u^{i-1}(1 - v)^{n-i-1}\left[1 - \binom{n}{i} v^i(1 - u)^{n-i}\right], \quad 0 \leq u \leq v \leq 1.$$

Noting that for every fixed $1 \leq i \leq n - 1$ and $0 \leq u \leq 1$, we have

$$f_{i,n}(u, v) \leq f_{i,n}(u, u) = g_{i,n}(u)$$

for all $u \leq v \leq 1$, and recalling (6), we complete the proof of inequality (1).

Now we justify its sharpness. Suppose that $u_* = \arg\max g_{i,n}(u) \in (0, 1)$. Then we get the equality in (8) iff either $F(x \wedge y)\bar{F}(x \vee y) = 0$ or $F(x \wedge y) = F(x \vee y) = u_*$ almost surely with respect to the Lebesgue measure on $\mathbb{R}^2$. This is equivalent to the condition that 0, u_* and 1 are the only values of F. Assume now that $u_* = 0$, and

$X, X_1, \ldots, X_n$ are i.i.d. with parent distribution function F_u defined in (3) for some $0 < u < 1$. Then clearly

$$\mathbb{V}ar_u X = (a-b)^2 u(1-u).$$

The spacing $S_{i:n}$ has two-point distribution

$$\begin{aligned}\mathbb{P}_u(S_{i:n} = b-a) &= \mathbb{P}_u(X_{1:n} = \cdots = X_{i:n} = a, X_{i+1:n} = \cdots = X_{n:n} = b)\\ &= \binom{n}{i} u^i (1-u)^{n-i} = B_{i,n}(u),\\ \mathbb{P}_u(S_{i:n} = 0) &= 1 - B_{i,n}(u).\end{aligned}$$

Therefore

$$\mathbb{V}ar_u S_{i:n} = (a-b)^2 B_{i,n}(u)\big[1 - B_{i,n}(u)\big] = g_{i,n}(u)\mathbb{V}ar_u X,$$

and

$$\lim_{u\downarrow 0} \frac{\mathbb{V}ar_u S_{i:n}}{\mathbb{V}ar_u X} = \lim_{u\downarrow 0} g_{i,n}(u) = g_{i,n}(0),$$

by continuity of $g_{i,n}$. The proof for $u \uparrow 1$ is analogous.

It remains to verify attainability of the lower bounds (4). If $i \geq 2$, then $g_{i,n}(0) = 0$, and mimicking arguments of the previous reasoning we obtain

$$\lim_{u\downarrow 0} \frac{\mathbb{V}ar_u S_{i:n}}{\mathbb{V}ar_u X} = \lim_{u\downarrow 0} g_{i,n}(u) = g_{i,n}(0).$$

The similar claim is concluded if $i \leq n-2$ and $g_{i,n}(1) = 0$. □

Proof of Lemma 1. (i) We first focus on the case $i = 1$ and show that $g_{1,n}$ is strictly decreasing on the interval [0, 1]. Consider

$$g'_{1,n}(u) = n(1-u)^{n-3} h_{1,n}(u),$$

where

$$h_{1,n}(u) = n(1-u)^{n-1}\big(2(n-1)u - 1\big) - n + 2.$$

Observe that $h_{1,n}(0) = -2(n-1)$, $h_{1,n}(1) = -(n-2)$ and

$$h'_{1,n}(u) = n(n-1)(1-u)^{n-2}(3 - 2nu),$$

which implies that $h_{1,n}$ is increasing on $\left[0, \frac{3}{2n}\right]$ and decreasing on $\left[\frac{3}{2n}, 1\right]$. We show that

$$h_{1,n}\left(\frac{3}{2n}\right) = 2n\left(1 - \frac{3}{2n}\right)^n - (n-2) < 0, \qquad n \geq 3,$$

which means that

$$\frac{2n-3}{n-2} < \left(\frac{2n}{2n-3}\right)^{n-1}, \qquad n \geq 3.$$

By the Bernoulli inequality,

$$\left(\frac{2n}{2n-3}\right)^{n-1} > 1 + \frac{3(n-1)}{2n-3} = \frac{5n-6}{2n-3}, \qquad n \geq 2,$$

it remains to notice that $\frac{5n-6}{2n-3} \geq \frac{2n-3}{n-2}$, which is equivalent to $(n-1)(n-3) \geq 0$, $n \geq 3$, and verifies desired claim. Summing up, we have $h_n(u) < 0$ and $g'_{n,1}(u) < 0$ for all $0 < u < 1$ and $n \geq 3$, which implies that

$$\max_{u \in [0,1]} g_{1,n}(u) = g_{1,n}(0) = n, \qquad n \geq 3.$$

(ii) The conclusion for $i = n-1$ follows from the relation $g_{i,n}(u) = g_{n-i,n}(1-u)$ and the previous statement. □

Proof of Lemma 2. For given $n \geq 4$ and $2 \leq i \leq n-2$ we have:

$$\begin{aligned} g'_{i,n}(u) &= \binom{n}{i} u^{i-2}(1-u)^{n-2-i} h_{i,n}(u) \\ &= \binom{n}{i} \frac{u^{i-2}(1-u)^{n-2-i}}{n+1} \sum_{j=0}^{n+1} a_{j,n+1} B_{j,n+1}(u) \end{aligned}$$

[cf. (5)], where

$$a_{j,n+1} = \begin{cases} -2(n-i)i, & \text{if } j = i, \\ 2(n-i)i, & \text{if } j = i+1, \\ (i-1)(n+1) - j(n-2), & \text{otherwise.} \end{cases} \tag{9}$$

Since $2 \leq i \leq n-2$, the arithmetic sequence $\tilde{a}_{j,n+1} = (i-1)(n+1) - j(n-2)$, $j \in \{0, \dots n+1\}$, decreases from $\tilde{a}_{0,n+1} = (i-1)(n+1) > 0$ to $\tilde{a}_{n+1,n+1} = -(n+1)(n-1-i) < 0$. For any fixed $i \in \{2, \dots, n-2\}$, if we replace any pair $\tilde{a}_{i,n+1}$, $\tilde{a}_{i+1,n+1}$ by arbitrary $a < 0$ and $b > 0$, we obtain another sequence with consecutive signs $+-+-$ (we suppressed here multiple pluses and minuses, and dropped a possible zero at $j = \frac{(i-1)(n+1)}{n-2}$). This holds true for (9), in particular. By Lemma 3, $g_{i,n}$ is either first increasing and then decreasing or it is consecutively increasing, decreasing, increasing and ultimately decreasing.

We now treat the case $i = 2$, $n = 4$ with use of standard calculus tools. Then

$$
\begin{aligned}
g'_{2,4}(u) &= 6(2u-1)[18u^2(1-u)^2 - 1] \\
&= 6(2u-1)[3\sqrt{2}u(1-u) - 1][3\sqrt{2}u(1-u) + 1] \\
&= 108\left(u - \frac{1}{2} + \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}\right)\left(u - \frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right)\left(u - \frac{1}{2}\right) \\
&\quad \times \left(u - \frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}\right)\left(u - \frac{1}{2} - \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right).
\end{aligned}
$$

Hence the derivative $g'_{2,4}$ restricted to $[0, 1]$ has three zeros at $\frac{1}{2}$, $\frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}$, and $\frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}$. Moreover $g'_{2,4}(u) > 0$ iff either $u \in \left(0, \frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}\right)$ or $u \in \left(\frac{1}{2}, \frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right)$. By symmetry of the function about $\frac{1}{2}$, we get

$$
\begin{aligned}
\max_{u \in [0,1]} g_{2,4}(u) &= g_{2,4}\left(\frac{1}{2} - \frac{\sqrt{6}}{6} + \frac{\sqrt{3}}{6}\right) = g_{2,4}\left(\frac{1}{2} + \frac{\sqrt{6}}{6} - \frac{\sqrt{3}}{6}\right) \\
&= \frac{2\sqrt{2}}{3} \approx 0.94281.
\end{aligned}
$$

(i) Now we proceed to $n \geq 5$. Observe that

$$
\begin{aligned}
g_{i,n}(u) &= \binom{n}{i} u^{i-1}(1-u)^{n-1-i}\left[\sum_{\substack{j=0 \\ i \neq j \neq n-i}}^{n} \binom{n}{j} u^j (1-u)^{n-j} + \binom{n}{i} u^{n-i}(1-u)^i\right] \\
&= \binom{n}{i}[u(1-u)]^{i-1}(1-u)^{n-2i} \sum_{\substack{j=0 \\ i \neq j \neq n-i}}^{n} \binom{n}{j} u^j (1-u)^{n-j} \\
&\quad + \binom{n}{i}^2 [u(1-u)]^{n-1},
\end{aligned}
$$

and

$$
\begin{aligned}
g_{i,n}(1-u) &= \binom{n}{i}[u(1-u)]^{i-1}u^{n-2i} \sum_{\substack{j=0 \\ i \neq j \neq n-i}}^{n} \binom{n}{j} u^j (1-u)^{n-j} \\
&\quad + \binom{n}{i}^2 [u(1-u)]^{n-1}.
\end{aligned}
$$

In consequence,

$$g_{i,n}(u) - g_{i,n}(1-u) = \binom{n}{i}[u(1-u)]^{i-1} \sum_{\substack{j=0 \\ i\neq j\neq n-i}}^{n} \binom{n}{j} u^j (1-u)^{n-j} \times [(1-u)^{n-2i} - u^{n-2i}].$$

The sign of the difference is identical with that of the expression in square brackets. Therefore for $i < \frac{n}{2}$ this difference is positive on $(0, \frac{1}{2})$ and negative on $(\frac{1}{2}, 1)$. If $i > \frac{n}{2}$, the signs are reversed. This immediately implies our claims.

(ii) Suppose finally that $n \geq 6$ is even and $i = \frac{n}{2}$. Due to (2), $g_{\frac{n}{2},n}$ is symmetric about $\frac{1}{2}$, and $g_{\frac{n}{2},n}(\frac{1}{2})$ is a local extreme. We prove that this is a maximum, verifying that $g''_{\frac{n}{2},n}(\frac{1}{2}) < 0$. Using $i = \frac{n}{2}$ for simplicity of notation we have

$$g''_{i,2i}(u) = \frac{(2i)![u(1-u)]^{i-3}}{i!^4}\Big[i!^2(i-1)(4iu^2 - 4iu - 6u^2 + i + 6u - 2)$$
$$-2u^i(1-u)^i(2i)!(2i-1)(4iu^2 - 4iu - 3u^2 + i + 3u - 1)\Big],$$

$$g''_{i,2i}\left(\frac{1}{2}\right) = \frac{(2i)!}{2^{2i-1}i!^4}h(i),$$

where

$$h(i) = 4^{-i}(2i)!(2i-1) - i!^2(i-1)$$

determines the sign of $g''_{i,2i}\left(\frac{1}{2}\right)$. We shall prove that $h(i) < 0$ for $i \geq 3$ by induction. We check that $h(3) = -\frac{63}{4}$ and assume that $h(i) < 0$ for some $i \geq 3$ which is equivalent to $\binom{2i}{i}4^{-i}\left(2 + \frac{1}{i-1}\right) < 1$. We show that the relation holds for $i+1$ as well. Indeed,

$$\binom{2i+2}{i+1}4^{-i-1}\left(2+\frac{1}{i}\right) = \binom{2i}{i}4^{-i}\left(2+\frac{1}{i}\right)\frac{2i+1}{2(i+1)}$$
$$< \binom{2i}{i}4^{-i}\left(2+\frac{1}{i-1}\right)\frac{2i+1}{2i+2}$$
$$< \binom{2i}{i}4^{-i}\left(2+\frac{1}{i-1}\right) < 1,$$

by the inductive assumption. This ends the proof. □

Proof of Proposition 2. The upper bound is evident by Proposition 1, since

$$g_{1,2}(u) = 2[1 - 2u(1-u)] = 2 - 4u + 4u^2, \qquad 0 \leq u \leq 1,$$

attains its maximal value 2 at 0 and 1. In order to establish the lower one, we first note that

$$\mathbb{E}\, S_{1:2}^2 = \mathbb{E}\,(X_1 - X_2)^2 = 2\,\mathbb{V}ar\ X$$

[cf. also (6) and (7)]. Accordingly the problem of minimizing

$$\frac{\mathbb{V}ar\ S_{1:2}}{\mathbb{V}ar\ X} = 2 - \frac{\left(\mathbb{E}\, S_{1:2}\right)^2}{\mathbb{V}ar\ X} \tag{10}$$

is dual to that of maximizing $\frac{\left|\mathbb{E}\, S_{1:2}\right|}{\sqrt{\mathbb{V}ar\ X}}$. We focus on the later one. Suppose that X_1, X_2 are independent, and have a common distribution function F with mean μ and finite and positive variance σ^2. Then

$$\begin{aligned}\mathbb{E}\, S_{1:2} &= \mathbb{E}[F^{-1}(U_{2:2}) - \mu] - \mathbb{E}[F^{-1}(U_{1:2}) - \mu] \\ &= \int_{\mathbb{R}} [F^{-1}(x) - \mu][f_{2:2}(x) - f_{1:2}(x)]\mathrm{d}x,\end{aligned}$$

where $U_{1:2}$ and $U_{2:2}$ denote the minimum and maximum of two i.i.d. standard uniform random variables, and

$$f_{1:2}(x) = \begin{cases} 2(1-x), & \text{if } 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

$$f_{2:2}(x) = \begin{cases} 2x, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise,} \end{cases}$$

stand for the respective density functions. By Cauchy–Schwarz inequality,

$$\begin{aligned}|\mathbb{E}\, S_{1:2}| &= 2\left|\int_0^1 [F^{-1}(x) - \mu](2x-1)\mathrm{d}x\right| \\ &\le 2\sqrt{\int_0^1 [F^{-1}(x) - \mu]^2\mathrm{d}x \int_0^1 (2x-1)^2\mathrm{d}x} = \frac{2\sqrt{3}}{3}\,\sigma.\end{aligned}$$

This is a special case of the classic bounds on the expectation of sample ranges due to Plackett (1947), and together with (10), determine the lower variance bound. Observe that equality holds in the Cauchy–Schwarz inequality iff

$$F^{-1}(x) - \mu = \alpha(2x-1), \qquad 0 < x < 1, \tag{11}$$

for some real α. Since F^{-1} is nondecreasing and nonconstant function, α has to be positive. Condition $\int_0^1 [F^{-1}(x) - \mu]^2\mathrm{d}x = \sigma^2$ implies that $\alpha = \sqrt{3}\,\sigma$. Hence, equation (11) uniquely determines the quantile function of the uniform distribution on the interval $[\mu - \sqrt{3}\,\sigma, \mu + \sqrt{3}\,\sigma]$. Clearly, changing parameters μ and σ we obtain the uniform distribution on arbitrary intervals. These distributions attain the lower variance bound of Proposition 2. □

Acknowledgements The authors are grateful to the associate editor and referee for helpful comments that allowed them to correct mistakes and improve the presentation.

References

Danielak, K. (2004). Sharp upper bounds for expectations of differences of order statistics in various scale units. *Communications in Statistics—Theory and Methods*, *33*, 787–803.

Danielak, K., Rychlik, T. (2003). Sharp bounds for expectations of spacings from DDA and DFRA families. *Statistics and Probability Letters*, *65*, 303–316.

Danielak, K., Rychlik, T. (2004). Sharp bounds for expectations of spacings from decreasing density and failure rate families. *Applicationes Mathematicae (Warsaw)*, *31*, 369–395.

David, H. A., Nagaraja, H. N. (2003). *Order statistics* (3rd ed.). Hoboken, NJ: Wiley.

Goroncy, A., Rychlik, T. (2015a). Optimal bounds on expectations of order statistics and spacings from nonparametric families of distributions generated by convex transform order. *Metrika*, *78*, 175–204.

Goroncy, A., Rychlik, T. (2015b). Evaluations of expected order statistics and spacings based on the IFR distributions (submitted).

Hoeffding, W. (1940). Maßtabinvariante Korrelations-Theorie. *Schriften des Mathematischen Instituts and Instituts für Angewandte Mathematik der Universität Berlin*, *5*, 181–233.

Irwin, J. O. (1925). The further theory of Francis Galton's individual difference problem. *Biometrika*, *17*, 100–128.

Jasiński, K., Rychlik, T. (2013). Maximum variance of order statistics from symmetric populations revisited. *Statistics*, *47*, 422–438.

Jones, M. C., Balakrishnan, N. (2002). How are moments and moments of spacings related to distribution function. *Journal of Statistical Planning and Inference*, *103*, 377–390.

Kozyra, P. M., Rychlik, T. (2015). Sharp bounds on the expectations of L-statistics expressed in the Gini mean differences units. *Communications in Statistics—Theory and Methods* (to appear).

Lin, G. D., Huang, J. S. (1989). Variances of sample medians. *Statistics and Probability Letters*, *8*, 143–146.

Moriguti, S. (1951). Extremal properties of extreme value distributions. *Annals of Mathematical Statistics*, *22*, 523–536.

Moriguti, S. (1953). A modification of Schwarz's inequality, with applications to distributions. *Annals of Mathematical Statistics*, *24*, 107–113.

Papadatos, N. (1995). Maximum variance of order statistics. *Annals of the Institute of Statistical Mathematics*, *47*, 185–193.

Papadatos, N. (1997). A note on maximum variance of order statistics from symmetric populations. *Annals of the Institute of Statistical Mathematics*, *49*, 117–121.

Pearson, K. (1902). Note on Francis Galton's problem. *Biometrika*, *1*, 390–399.

Plackett, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika*, *34*, 120–122.

Pyke, R. (1965). Spacings. (With discussion). *Journal of the Royal Statistical Society: Series B*, *27*, 395–449.

Pyke, R. (1972). Spacings revisited. In L. M. Le Cam, J. Neyman, E. L. Scott (Eds.), *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 417–427) *Theory of Statistics*. Berkeley, CA: University of California Press.

Raqab, M. Z. (2003). Evaluating improvements for spacings of order statistics. *Extremes*, *6*, 259–273.

Rychlik, T. (2001). *Projecting statistical functionals*. Lecture Notes in Statistics (Vol. 160). New York: Springer.

Schoenberg, I. J. (1959). On variation diminishing approximation methods. In R. E. Langer (Ed.), *Proceedings of a symposium on numerical approximation, Madison, April 21–23, 1958* (pp. 249–274). Madison, WI: University of Wisconsin Press.

Yang, H. (1982). On the variances of median and some other order statistics. *Bulletin of the Institute of Mathematics Academia Sinica*, *10*, 197–204.