THE SHANNON ENTROPY OF RANKED SET AND PARTIALLY RANK-ORDERED SET SAMPLES OF THE SAME SIZE: NO ORDERING RELATIONSHIP FOR THE UNIFORM PARENT

KOREA SUPERINTELLIGENCE LABS

Abstract. In partially rank-ordered set (PROS) sampling one fixes a set size S and a design parameter $D = \{d_1, \dots, d_n\}$ splitting the ranks $\{1, \dots, S\}$ into n consecutive blocks, and measures one unit drawn uniformly at random from each block; the measured value from block d_r has the mixture density $f_{(d_r)} = |d_r|^{-1} \sum_{u \in d_r} f^{(u:S)}$. Ranked set sampling (RSS) of size n is the case $S = n$. Hatefi and Jafari Jozani, who study the information content of such samples, prove $m^{-1} H_S(X_{\text{rss}}) \le H_n(X_{\text{pros}}) \le H_n(X_{\text{srs}})$ for the balanced design and then record that they "were not able to obtain an ordering relationship among the Shannon entropy of RSS and PROS data of the same size". We settle that comparison for the Uniform$(0,1)$ parent, and the answer is negative: both orders occur. For the balanced design PROS$(2,4)$ one has $H_2(X_{\text{pros}}) = \frac{9}{2} - 2\log 2 - \frac{27}{8}\log 3 < 1 - 2\log 2 = H_2(X_{\text{rss}})$, while for the design $D = (\{1,2\}, \{3\}, \dots, \{7\}, \{8,9\})$ at set size $S = 9$ and sample size $n = 7$ one has $H_7(X_{\text{rss}}) < H_7(X_{\text{pros}})$, the gap being $-\frac{997119143}{74118870} + \frac{79155653}{5764801}\log 2 + 10\log 3 - 2\log 5 - 2\log 7 = -0.06004185\ldots$. So no ordering relationship between the two entropies can hold. Every entropy here is exact — a rational number plus a rational combination of $\log 2, \log 3, \log 5, \log 7$ — and every closed form and every inequality stated as a theorem below is checked by the Lean 4 kernel over Mathlib, with no evaluation axiom and no floating-point step. Outside the formal development we report two measurements, labelled as such: an exhaustive search over every design of set size $S \le 10$ shows that no reversal occurs for $S \le 8$, so $S = 9$ is the smallest set size at which one exists; and over all 34 balanced designs with $n \ge 2$ and $S = mn \le 16$ the inequality $H_n(X_{\text{pros}}) \le H_n(X_{\text{rss}})$ never fails, which locates the obstruction in the general design parameter rather than in the design family as a whole.

1. Introduction

The two designs. Let $X_1, \dots$ be i.i.d. with density f, and write $f^{(u:S)}$ for the density of the u-th order statistic of a sample of size S.

Ranked set sampling (RSS), introduced by McIntyre [7], produces a sample of size n from n sets of n units each: from the i-th set it measures the unit judged to have rank i. The ranking is by inspection and costs nothing; only the measurement is expensive. The i-th observation has density $f^{(i:n)}$, and the n observations are independent.

Partially rank-ordered set sampling (PROS), introduced by Öztürk [8], weakens the demand on the ranker. One fixes a *set size* S and a *design parameter* $D = \{d_1, \dots, d_n\}$ partitioning $\{1, \dots, S\}$ into n mutually exclusive subsets; the units of a set of size S are partially rank-ordered, so that "each unit in subset h has a rank smaller than the ranks of units in subset h' for all $h' \ge h$" [3, §1], with no ordering asked for *inside* a subset. From the r-th set one unit is drawn uniformly at random from the block d_r and measured. Its density is the uniform mixture of the corresponding consecutive order-statistic densities,

$$f_{(d_r)}(x) \;=\; \frac{1}{|d_r|} \sum_{u \in d_r} f^{(u:S)}(x), \tag{1}$$

which is the display carried by [3, §2], there written with the common block size m of the balanced design; for a design whose blocks have different sizes the same formula, with $|d_r|$ in place of m, is that paper's own appendix lemma on unbalanced designs (Lemma 7 in the numbering of v1), stated there for judgment order statistics, which under perfect ranking are the order statistics. The sample $\{X_{(d_1)}, \ldots, X_{(d_n)}\}$ has size n and is drawn in one cycle. When every block is a singleton, so $S = n$, PROS is RSS; when there is a single block, so $n = 1$, it is simple random sampling of one unit. Both degeneracies are recorded in [3, §2]: "RSS and SRS can be expressed as special cases of the PROS(n, S) design when $S = n$ and $S = 1$, respectively."

Throughout, PROS(n, S) denotes, as in [3], the *balanced* design in which $d_r = \{(r-1)m + 1, \ldots, rm\}$ with $m = S/n$; a general design parameter D need not be balanced, and the distinction is what this paper turns on.

The differential (Shannon) entropies of the three designs at sample size n are

(2)
$$H_n(X_{\text{rss}}) = -\sum_{i=1}^{n} \int f^{(i:n)} \log f^{(i:n)}, \quad H_n(X_{\text{pros}}) = -\sum_{r=1}^{n} \int f_{(d_r)} \log f_{(d_r)}, \quad H_n(X_{\text{srs}}) = n\, H(X_1),$$

the displays of [3, §4.1]. "Data of the same size" means that all three carry n measured observations: RSS at set size n, PROS at set size $S = \sum_r |d_r| \geq n$.

The question. Hatefi and Jafari Jozani [3] prove, from the convexity of $t \log t$ applied to the mixture (1), that for the balanced design PROS(n, S) with $m = S/n$

$$\frac{1}{m} H_S(X_{\text{rss}}) \;\leq\; H_n(X_{\text{pros}}) \;\leq\; H_n(X_{\text{srs}}), \tag{3}$$

a lower bound in terms of the RSS entropy *at set size* S, not at sample size n. Their §4.1 then says, verbatim:

> "In the following lemma, we show that the Shannon entropy of PROS data is smaller than that of SRS data of the same size. **Unfortunately, we were not able to obtain an ordering relationship among the Shannon entropy of RSS and PROS data of the same size.** Instead, we obtain a lower bound for the Shannon entropy of a PROS(n, S) sample in terms of the Shannon entropy of an RSS data of size S when the set size is S."

The quotation is verbatim from the LaTeX source of arXiv:1504.07336v1, the version we read, where it stands in §4.1 immediately before the lemma carrying (3) (that version's Lemma 3). The published AStA text is behind a paywall and we did not read it; §6 sets out what can and cannot be said about it from outside.

Question 1.1. Is there an ordering relationship between $H_n(X_{\text{rss}})$ and $H_n(X_{\text{pros}})$ — the entropies of a ranked set sample and of a partially rank-ordered set sample carrying the same number n of measured observations?

The variance counterpart of Question 1.1 is settled. Frey and Feeman [2] write in their abstract that they "completely determine how PROSS schemes compare in the two-subset case", and "also prove a conjecture of Ozturk … about how the efficiency of the PROSS sample mean compares to that of the RSS sample mean". The entropy counterpart, so far as our search reached (§6), has not been taken up by anyone since [3] recorded the gap.

What is settled here. We answer Question 1.1 for the Uniform$(0, 1)$ parent, where $H(X_1) = 0$, so $H_n(X_{\text{srs}}) = 0$ and every quantity in (2) is a rational number plus a rational combination of logarithms of positive rationals. The answer is that *no ordering relationship exists*: both orders occur, at designs that both lie inside the definition of [3, §2].

- For the balanced design PROS$(2, 4)$, that is $D = (\{1, 2\}, \{3, 4\})$ with $S = 4$, $n = 2$,
$$H_2(X_{\text{pros}}) = \tfrac{9}{2} - 2\log 2 - \tfrac{27}{8}\log 3 \;<\; 1 - 2\log 2 = H_2(X_{\text{rss}}),$$
the gap being $-\frac{7}{2} + \frac{27}{8}\log 3 = +0.20781647\ldots$ (Theorem 5.1).

- For the unbalanced design $D = (\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8,9\})$ with $S = 9$, $n = 7$,
$$H_7(X_{\text{rss}}) = 21 - 4\log 2 - 4\log 3 - 3\log 5 - 7\log 7 \;<\; H_7(X_{\text{pros}}),$$
the gap being $-\frac{997119143}{74118870} + \frac{79155653}{5764801}\log 2 + 10\log 3 - 2\log 5 - 2\log 7 = -0.06004185\ldots$ (Theorem 5.2).

Together these give Theorem 5.3: neither $H_n(X_{\text{pros}}) \le H_n(X_{\text{rss}})$ nor $H_n(X_{\text{rss}}) \le H_n(X_{\text{pros}})$ holds for all designs. Of the two, the second inequality is the one that is not anticipated by (3), and it is the harder to find: the design had to be searched for (Remark 5.5).

Two further points shape the answer. First, the reversing design has *unequal* block sizes. Within the balanced notation $\text{PROS}(n,S)$ of [3] we could not make the order reverse at all, which is recorded as Conjecture 5.6; so what has no ordering is the general design parameter D of [3, §2], not the design family as a whole, and the authors' intuition was right about the case their notation covers. Second, the reversal is not generic: an exhaustive search over every design with $S \le 10$ finds none at any $S \le 8$, so $S = 9$ is the smallest set size at which one exists (Remark 5.5).

What is exact, and what is measured. Everything named as a theorem or proposition below is exact and machine-checked (§7): the entropies are closed forms, the comparisons are decided by rational bounds on $\log 2, \log 3, \log 5, \log 7$, and no numerical quadrature enters any proof. The decimal expansions printed alongside the closed forms are truncations, given for the reader. Two statements are *measurements* obtained outside the formal development — the extent of the search (Remark 5.5) and the balanced-design evidence (Conjecture 5.6) — and they are labelled as such wherever they appear.

Organisation. Section 2 fixes definitions and proves the two degeneracies. Section 3 is the exact-evaluation instrument, together with the published values it reproduces before being aimed anywhere new. Section 4 computes the PROS entropies and says why the route is confined to blocks of size at most two. Section 5 contains the main theorems, the negative controls, and the two measurements. Section 6 reports what was searched. Section 7 describes the machine verification.

2. Preliminaries

The parent is Uniform$(0,1)$ throughout, so $f = \mathbf{1}_{(0,1)}$ and
$$f^{(u:S)}(x) = \frac{S!}{(u-1)!\,(S-u)!}\,x^{\,u-1}(1-x)^{\,S-u}, \qquad x \in [0,1],$$
the Beta$(u, S-u+1)$ density. It is convenient to index a block by the numbers of ranks below and above it rather than by its endpoints. For $p, m, q \ge 0$ with $m \ge 1$ put $S = p+m+q$ and
$$g_{p,q}(x) = f^{(p+1:\,p+q+1)}(x) = \frac{(p+q+1)!}{p!\,q!}x^p(1-x)^q, \qquad g_{p,m,q} = \frac{1}{m}\sum_{j=0}^{m-1} g_{\,p+j,\;q+m-1-j}, \tag{4}$$
so that $g_{p,m,q}$ is the density (1) of the observation drawn from the block of m consecutive ranks lying above p ranks and below q ranks in a set of size S. For a density h carried by $[0,1]$ we write
$$H(h) = -\int_0^1 h(x)\log h(x)\,dx.$$

Definition 2.1. A *design* is a finite sequence $D = (m_1,\ldots,m_n)$ of positive integers, read as the block sizes of $D = \{d_1,\ldots,d_n\}$ in rank order; its *set size* is $S = \sum_r m_r$ and its *sample size* is n. Its entropy is
$$H_n(X^D_{\text{pros}}) = \sum_{r=1}^n H\big(g_{\,p_r,\,m_r,\,q_r}\big), \qquad p_r = \sum_{s<r} m_s, \quad q_r = \sum_{s>r} m_s,$$
and the RSS entropy at sample size n is $H_n(X_{\text{rss}}) = \sum_{i=1}^n H\big(g_{\,i-1,\,n-i}\big)$.

This is (2) written out; the blocks of a design are consecutive rank intervals, as the partial ordering of [3, §1] requires, and a design is determined by its sequence of block sizes.

Proposition 2.2 (the designs coincide)**.** *For every $n \geq 0$, the design $(1, 1, \ldots, 1)$ with n blocks satisfies $H_n(X_{\mathrm{pros}}^{(1,\ldots,1)}) = H_n(X_{\mathrm{rss}})$. For every $S \geq 1$ the one-block design (S) satisfies $g_{0,S,0} \equiv 1$, hence $H_1(X_{\mathrm{pros}}^{(S)}) = 0 = H_1(X_{\mathrm{rss}})$.*

Proof. For $m = 1$ the mixture (4) has a single term, $g_{p,1,q} = g_{p,q}$, so the two sums of Definition 2.1 agree term by term. For the one-block design,

$$g_{0,S,0}(x) = \frac{1}{S}\sum_{j=0}^{S-1} \frac{S!}{j!\,(S-1-j)!} x^j (1-x)^{S-1-j} = \sum_{j=0}^{S-1} \binom{S-1}{j} x^j (1-x)^{S-1-j} = \big(x+(1-x)\big)^{S-1} = 1,$$

by the binomial theorem, and $H(\mathbf{1}_{[0,1]}) = 0$. □

Proposition 2.2 is the exact form of the sentence of [3, §2] quoted in §1; the authors state the reduction and give no proof, none being needed. We record it as a theorem rather than a remark because it is also the coincidence control of the present development: it is stated for *every* n and *every* S, not at sample points, and both degeneracies must and do return the RSS and SRS values exactly.

Lemma 2.3 (reflection)**.** *$H\big(h(1-\cdot)\big) = H(h)$ for every h; consequently $H(g_{q,p}) = H(g_{p,q})$ and $H(g_{q,m,p}) = H(g_{p,m,q})$.*

Proof. The substitution $x \mapsto 1-x$ carries $[0,1]$ to itself and preserves Lebesgue measure, and $g_{p,q}(1-x) = g_{q,p}(x)$, $g_{p,m,q}(1-x) = g_{q,m,p}(x)$ directly from (4) (reversing the summation index $j \mapsto m-1-j$ in the second case). □

3. Exact evaluation

The reduction. Every density met here is a polynomial with rational coefficients that factors over $\mathbb{Q}$ into affine factors, each nonnegative on $[0,1]$: write it as

$$h(x) = c\,x^p(1-x)^q(u+vx)^r. \tag{5}$$

For such an h the logarithm splits, and it splits without a positivity hypothesis, because the identity is multiplied through by h and both sides vanish where h does.

Lemma 3.1.
(1) *For $k \geq 0$ and $0 \leq a \leq b$, $\int_a^b y^k \log y\,dy = G_k(b) - G_k(a)$ with $G_k(t) = \dfrac{t^{k+1}\log t}{k+1} - \dfrac{t^{k+1}}{(k+1)^2}$ and $G_k(0) = 0$; and $\int_a^b y^k\,dy = (b^{k+1} - a^{k+1})/(k+1)$.*
(2) *If h has the shape (5) and $u + vx \neq 0$ on $[0,1]$, then*

$$H(h) = -\Big(\log c \int_0^1 h + p\int_0^1 h(x)\log x\,dx + q\int_0^1 h(x)\log(1-x)\,dx + r\int_0^1 h(x)\log(u+vx)\,dx\Big).$$

(3) *Each of the four integrals in (ii) reduces to a finite rational combination of the integrals of (i): the first and second directly, the third after $x \mapsto 1-x$, the fourth after $y = vx + u$, provided $u \geq 0$ and $v > 0$.*

Proof. (i) is the fundamental theorem of calculus with the antiderivative extended by $G_k(0) = 0$; the extension is continuous at 0 because $t \mapsto t\log t$ is, and the integrand is interval-integrable on $[a,b]$ since log is. (ii) is the pointwise identity $h\log h = h\big(\log c + p\log x + q\log(1-x) + r\log(u+vx)\big)$, valid at every x (where $h(x) = 0$ both sides vanish), integrated term by term; the four integrands are interval-integrable, being continuous functions times log of an affine function. (iii) is the substitution rule, the polynomial in each case being expanded in the new variable. □

Hence every entropy in this paper is a rational number plus a rational combination of logarithms of positive rationals, and after factoring the arguments into primes, of $\log 2, \log 3, \log 5, \log 7$.

The published values, reproduced first. Before the instrument of Lemma 3.1 is aimed at anything new it is made to reproduce what is already in print.

Proposition 3.2 (order-statistic entropies)**.** *For the* Uniform$(0,1)$ *parent,*

$$\begin{aligned} &H(f^{(1:2)}) = \tfrac{1}{2} - \log 2, && H(f^{(1:7)}) = \tfrac{6}{7} - \log 7, && H(f^{(2:7)}) = \tfrac{1319}{420} - \log 42,\\ &H(f^{(3:7)}) = \tfrac{887}{210} - \log 105, && H(f^{(4:7)}) = \tfrac{319}{70} - \log 140, && H(f^{(3:9)}) = \tfrac{3107}{630} - \log 252,\\ &H(f^{(4:9)}) = \tfrac{7201}{1260} - \log 504, && H(f^{(5:9)}) = \tfrac{1879}{315} - \log 630, \end{aligned}$$

and the remaining values at $S = 7$ and $S = 9$ follow from Lemma 2.3.

These are not new. They are the Beta differential entropy

$$H(U_{(i)}) = \log B(i, n-i+1) - (i-1)\big[\psi(i) - \psi(n+1)\big] - (n-i)\big[\psi(n-i+1) - \psi(n+1)\big] \tag{6}$$

attributed to Ebrahimi, Soofi and Zahedi [1], evaluated at integer parameters. What is done here is only the evaluation from the integral, by Lemma 3.1: neither the Beta function nor the digamma function is used, and none is available in the library we verify against.

We take (6) in the form, and with the attribution, given by Jafari Jozani and Ahmadi [4, §2], who write "Ebrahimi et al. (2004) obtained an expression for $H(U_{(i)})$ which is given by" and then display it; [1] is not open access and we did not read it, so the attribution is theirs rather than ours. The identity itself is not in question: (6) agrees with the direct integration of $-\int_0^1 g \log g$ at every pair with $1 \le i \le n \le 7$, and it agrees with each of the eight closed forms of Proposition 3.2.

Proposition 3.3 (RSS entropies)**.** *For the* Uniform$(0,1)$ *parent,*

$$H_2(X_{\text{rss}}) = 1 - 2\log 2 = -0.3862943611\ldots,$$
$$H_7(X_{\text{rss}}) = 21 - 4\log 2 - 4\log 3 - 3\log 5 - 7\log 7 = -4.6167226576\ldots$$

Proof. By Definition 2.1, Lemma 2.3 and Proposition 3.2: $H_2(X_{\text{rss}}) = 2H(f^{(1:2)}) = 1 - 2\log 2$, and

$$H_7(X_{\text{rss}}) = 2H(f^{(1:7)}) + 2H(f^{(2:7)}) + 2H(f^{(3:7)}) + H(f^{(4:7)}),$$

whose rational part is $\frac{12}{7} + \frac{1319}{210} + \frac{887}{105} + \frac{319}{70} = 21$ and whose logarithmic part is $-2\log 7 - 2\log 42 - 2\log 105 - \log 140$, which factors as stated. ☐

Since $H_n(X_{\text{srs}}) = 0$ for the uniform parent, $H_n(X_{\text{rss}})$ is exactly the distribution-free difference $H(X_{\text{rss}}) - H(X_{\text{srs}}) = \sum_{i=1}^n H(U_{(i)})$ tabulated by Jafari Jozani and Ahmadi in Table 1 of [4] as

$$k(n) = -0.386,\ -0.989,\ -1.742,\ -2.611,\ -3.574,\ -4.616,\ -5.727,\ -6.897,\ -8.121 \quad (n = 2, \ldots, 10),$$

and their displayed closed form at $n = 2$ is $1 - 2\log 2 \approx -0.3863$. The values of Proposition 3.3 are $k(2)$ and $k(7)$; what is added here is the exact form and the proof, against their three printed decimals. (Table and values are quoted from the preprint arXiv:1301.4292, which is the version we read and which carries exactly one table; we could not check the table's number in the published version.)

Remark 3.4 (the reproduction gate, outside the formal development)*.* Before being used for anything new, the exact evaluator that produced the coefficient data for Propositions 3.2–3.3 was made to reproduce, from the densities alone, all nine tabulated values $k(2), \ldots, k(10)$ of [4] to within one unit in their last printed decimal — the largest deviation is $7.3 \cdot 10^{-4}$, at $n = 7$, their third decimals appearing to be truncated rather than rounded — and the closed form at $n = 2$ exactly. Evaluating (6) at $i = 1, \ldots, n$ and simplifying gives

$$H_n(X_{\text{rss}}) = \frac{n(n-1)}{2} - \log\Big(n! \prod_{i=1}^n \binom{n}{i}\Big),$$

which is a consequence of the published formula rather than a new statement; it is recorded here only because it makes the reproduction checkable at a glance. Both this identity and the numerical reproduction lie outside the machine-checked development, which contains the two values of Proposition 3.3 and no general n.

A rational bound on $\log 7$**.** The comparisons of §5 are decided by rational two-sided bounds on the four logarithms that occur. The library we verify against carries such bounds for $\log 2, \log 3, \log 5$ but not for $\log 7$, so we prove one.

Lemma 3.5. $1.9459 < \log 7 < 1.94592$.

Proof. Six terms of the series $-\log(1-x) = \sum_{k\geq 1} x^k/k$ at $x = \frac{1}{8}$ approximate $\log\frac{7}{8} = \log 7 - 3\log 2$ with error at most $\frac{1}{7\cdot 8^6} < 5.6\cdot 10^{-7}$; combining with the bounds $0.6931471803 < \log 2 < 0.6931471808$ gives the claim. (The series at $x = \frac{6}{7}$, which the analogous published bound for $\log 5$ suggests, would need 72 terms for the same accuracy.) □

4. The entropy of a PROS observation

Proposition 4.1 (block entropies)**.** *Let $d = \{1,2\}$ be the lowest two ranks.*

(1) *At set size $S = 4$ the density is $f_{(d)}(x) = 2(1-x)^2(1+2x)$ and*
$$H(f_{(d)}) = \tfrac{9}{4} - \log 2 - \tfrac{27}{16}\log 3 = -0.2970554177\ldots$$
(2) *At set size $S = 9$ the density is $f_{(d)}(x) = \frac{9}{2}(1-x)^7(1+7x)$ and*
$$H(f_{(d)}) = \tfrac{133307284}{37059435} - \log\tfrac{9}{2} - \tfrac{8388608}{5764801}\log 8 = -0.9328402161\ldots$$

Proof. The densities are (1) written out: $\frac{1}{2}\big(4(1-x)^3 + 12x(1-x)^2\big) = 2(1-x)^2(1+2x)$ and $\frac{1}{2}\big(9(1-x)^8 + 72x(1-x)^7\big) = \frac{9}{2}(1-x)^7(1+7x)$. Both have the shape (5) with $p = 0$, $r = 1$, $u = 1$ and $v = 2$ respectively $v = 7$, so Lemma 3.1 applies; the affine substitution $y = 1 + 7x$ in the fourth integral of Lemma 3.1(ii) maps $[0,1]$ to $[1,8]$ and is what puts the denominator $7^8 = 5764801$ into the second value. □

Proposition 4.2 (design entropies)**.** *For the* Uniform$(0,1)$ *parent,*

(1) *the balanced design* PROS$(2,4)$*, that is $D = (2,2)$ with $S = 4$, $n = 2$, has*
$$H_2(X_{\text{pros}}) = \tfrac{9}{2} - 2\log 2 - \tfrac{27}{8}\log 3 = -0.5941108354\ldots;$$
(2) *the design $D = (2,1,1,1,1,1,2)$, that is $D = (\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8,9\})$ with $S = 9$, $n = 7$, has*
$$H_7(X_{\text{pros}}) = \tfrac{2553615413}{74118870} - \tfrac{102214857}{5764801}\log 2 - 14\log 3 - \log 5 - 5\log 7 = -4.5566808041\ldots$$

Proof. In case (i) both blocks are $\{1,2\}$ up to reflection, so $H_2(X_{\text{pros}}) = 2H(f_{(d)})$ by Lemma 2.3 and Proposition 4.1(i). In case (ii) the two outer blocks contribute $2H(f_{(d)})$ with $f_{(d)}$ as in Proposition 4.1(ii), and the five singleton blocks contribute $2H(f^{(3:9)}) + 2H(f^{(4:9)}) + H(f^{(5:9)})$ by Lemma 2.3 and Proposition 3.2. Factoring $\frac{9}{2}$, 8, 252, 504 and 630 into primes and collecting gives the stated form. □

Remark 4.3 (why the blocks have size at most two). The logarithm of a mixture does not split, so the entropy of a PROS observation is $-\int_0^1 g\log g$ for a polynomial g of degree $S-1$, and Lemma 3.1 applies exactly when g factors into affine factors over $\mathbb{Q}$. For a block $\{a, a+1\}$ of a set of size S,
$$f_{(d)}(x) = \frac{S}{2}\,x^{a-1}(1-x)^{S-a-1}\Big[\binom{S-1}{a-1}(1-x) + \binom{S-1}{a}x\Big],$$
whose bracket is affine with a rational root, so it does factor. At block size 3 it need not: the block $\{1,2,3\}$ of a set of size 6 produces the irreducible factor $6x^2 + 3x + 1$, and the entropy then involves logarithms of algebraic numbers rather than of rationals. This is a limitation of the closed-form

route, not of computing power, and it is why every exact statement in this paper uses blocks of size at most two. It costs nothing here, because the reversal of §5 lives among such designs.

5. Both orders occur

Theorem 5.1 (PROS strictly below RSS)**.** *For the* Uniform$(0,1)$ *parent and the balanced design* PROS$(2,4)$*,*

$$H_2(X_{\mathrm{pros}}) < H_2(X_{\mathrm{rss}}), \qquad H_2(X_{\mathrm{rss}}) - H_2(X_{\mathrm{pros}}) = -\tfrac{7}{2} + \tfrac{27}{8}\log 3 = +0.2078164743\ldots$$

Proof. Subtract Proposition 4.2(i) from Proposition 3.3; the terms in $\log 2$ cancel and the difference is $-\frac{7}{2} + \frac{27}{8}\log 3$, which is positive because $\log 3 > 1.0986122885$ gives $\frac{27}{8}\log 3 > 3.7078\ldots > \frac{7}{2}$. □

This is the direction anticipated by (3) and by the convexity argument behind it. The other direction is not.

Theorem 5.2 (PROS strictly above RSS)**.** *For the* Uniform$(0,1)$ *parent and the design* $D = (\{1,2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8,9\})$ *with set size* $S = 9$ *and sample size* $n = 7$*,*

$$H_7(X_{\mathrm{rss}}) < H_7(X^D_{\mathrm{pros}}),$$

the difference being

$$H_7(X_{\mathrm{rss}}) - H_7(X^D_{\mathrm{pros}}) = -\frac{997119143}{74118870} + \frac{79155653}{5764801}\log 2 + 10\log 3 - 2\log 5 - 2\log 7 = -0.0600418535\ldots$$

Proof. Subtract Proposition 4.2(ii) from Proposition 3.3. The rational parts give $21 - \frac{2553615413}{74118870} = -\frac{997119143}{74118870}$, the coefficients of $\log 2$ give $\frac{102214857}{5764801} - 4 = \frac{79155653}{5764801}$, and those of $\log 3, \log 5, \log 7$ give $-4 + 14 = 10$, $-3 + 1 = -2$, $-7 + 5 = -2$. That this rational combination is negative follows from the two-sided rational bounds $0.6931471803 < \log 2 < 0.6931471808$, $1.0986122885 < \log 3 < 1.0986122888$, $1.6094379123 < \log 5 < 1.6094379126$ and Lemma 3.5, by a single linear arithmetic step: the coarsest of them, the bound on $\log 7$, leaves a slack of about $4 \cdot 10^{-5}$ after multiplication by its coefficient, some three orders of magnitude below the margin 0.06. □

Theorem 5.3 (no ordering relationship)**.** *For the* Uniform$(0,1)$ *parent there are designs* D_1 *and* D_2*, each with set size strictly larger than its sample size, such that*

$$H_{n_1}(X^{D_1}_{\mathrm{pros}}) < H_{n_1}(X_{\mathrm{rss}}) \qquad \text{and} \qquad H_{n_2}(X_{\mathrm{rss}}) < H_{n_2}(X^{D_2}_{\mathrm{pros}}), \qquad n_i = |D_i|.$$

Consequently neither $H_n(X^D_{\mathrm{pros}}) \le H_n(X_{\mathrm{rss}})$ *nor* $H_n(X_{\mathrm{rss}}) \le H_n(X^D_{\mathrm{pros}})$ *holds for all designs* D*: there is no ordering relationship between the Shannon entropy of RSS data and of PROS data of the same size.*

Proof. Take $D_1 = (2,2)$, of set size 4 and sample size 2, and $D_2 = (2,1,1,1,1,1,2)$, of set size 9 and sample size 7, and apply Theorems 5.1 and 5.2. □

Theorem 5.3 answers Question 1.1, negatively, for the uniform parent. Both designs are inside the definition of [3, §2]: each is a partition of $\{1,\ldots,S\}$ into consecutive blocks, with the units partially rank-ordered between blocks and unordered within them. The second is not covered by the balanced notation PROS(n,S), which the authors adopt for the rest of their paper, but it is covered by their definition — and its observation densities are the ones their own unbalanced-design lemma assigns, as recalled after (1). The difference between definition and notation is exactly where the ordering fails.

Proposition 5.4 (negative controls)**.** *The universally quantified claims*

$$\forall D \ne \emptyset :\ H_{|D|}(X^D_{\mathrm{pros}}) \le H_{|D|}(X_{\mathrm{rss}}) \qquad \text{and} \qquad \forall D \ne \emptyset :\ H_{|D|}(X_{\mathrm{rss}}) \le H_{|D|}(X^D_{\mathrm{pros}})$$

are both false. Moreover $H_2(X_{\mathrm{rss}}) \ne 0$*, and neither* $D_1 = (2,2)$ *nor* $D_2 = (2,1,1,1,1,1,2)$ *has set size equal to its sample size.*

Proof. The first two are Theorems 5.2 and 5.1 read as refutations; the third is $1 - 2\log 2 < 0$; the fourth is $4 \neq 2$ and $9 \neq 7$. □

The last two clauses are not decoration. For the uniform parent $H_n(X_{\mathrm{srs}}) = 0$, so a development in which some definitional error collapsed every entropy to 0 would satisfy every ordering statement above trivially and every proof would still go through; $H_2(X_{\mathrm{rss}}) \neq 0$ excludes that. The clause about set sizes excludes the other way the theorems could be vacuous, namely that the exhibited designs were secretly the coincidence case of Proposition 2.2, in which PROS *is* RSS.

Remark 5.5 (the extent of the search; a measurement, outside the formal development). The design of Theorem 5.2 had to be found, and where it lives is itself informative. Over *every* design of set size $S \leq 10$ — every composition of S into consecutive blocks — the quantity $H_n(X_{\mathrm{rss}}) - H_n(X^D_{\mathrm{pros}})$ was evaluated to 40 digits by panelled quadrature, and exactly, in rational arithmetic, over every design whose blocks all have size at most two for $S \leq 13$. The outcome:

- for $S \leq 8$ the difference is never negative; it is 0 exactly at the coincidence designs of Proposition 2.2, and the smallest strictly positive value at $S = 8$ is $+0.0066391870\ldots$, at $n = 6$, $D = (2, 1, 1, 1, 1, 2)$;
- at $S = 9$ the minimum is $-0.0600418535\ldots$, attained at $n = 7$, $D = (2, 1, 1, 1, 1, 1, 2)$, the design of Theorem 5.2;
- at $S = 10$ the minimum is $-0.1249503475\ldots$, at $n = 8$, $D = (2, 1, 1, 1, 1, 1, 1, 2)$.

So $S = 9$ is the smallest set size at which a PROS design has strictly larger Shannon entropy than a ranked set sample of the same size, and the reversal is not generic. The closest call anywhere in the search is a near miss in the other direction at $S = 9$, $n = 8$, $D = (1, 1, 1, 1, 1, 1, 1, 2)$, where the difference is exactly

$$-\tfrac{1071238013}{148237740} - \tfrac{38246987}{5764801}\log 2 + 14\log 3 - \log 5 - \log 7 = +2.7337059299 \cdot 10^{-6}.$$

These are computations carried out outside the machine-checked development. The scan over all designs with $S \leq 10$ is the 40-digit quadrature; the exact scan, in rational arithmetic and restricted to designs whose blocks have size at most two, extends it to $S \leq 13$ and agrees with it wherever both apply. The exact closed forms displayed here come from the exact evaluator, which carries out the same reduction as the proofs of §3 and is therefore a mirror of them rather than an independent oracle; the independent check is the quadrature. Only the two comparisons of Theorems 5.1 and 5.2 are part of the formal development. Formalising the near miss would require eleven further exact entropies — four order statistics at $S = 8$ and seven at $S = 9$ — which was priced and left out because it adds a data point, not a result.

Conjecture 5.6 (balanced designs; supported by measurement, not proved)**.** *For the* Uniform$(0, 1)$ *parent and the balanced design* PROS(n, S) *with* $m = S/n$*,*

$$H_n(X_{\mathrm{pros}}) \leq H_n(X_{\mathrm{rss}}),$$

with equality exactly when $m = 1$ *or* $n = 1$*.*

The evidence is a check of all 34 pairs (n, m) with $n \geq 2$ and $S = mn \leq 16$, by the same high-precision evaluation as Remark 5.5 and equally outside the formal development; among them the inequality is strict at exactly the 19 pairs with $m \geq 2$. (The remaining column $n = 1$ is the one-block design, where both sides are 0 by Proposition 2.2.) The two equality cases are proved, for all n and all S, in Proposition 2.2, and the first strict case $(n, m) = (2, 2)$ is Theorem 5.1; the general inequality is proved neither here nor, so far as we know, anywhere. Conjecture 5.6 sharpens Theorem 5.3: what has no ordering relationship is the general design parameter D of [3, §2], while the balanced notation PROS(n, S) that [3] uses throughout appears to have one — and it is the direction their own lower bound (3) points at. A proof would be a genuine theorem; we note that both sides are mixtures of the same S order-statistic densities. The PROS density is the uniform mixture over one consecutive block, and the RSS density is the hypergeometric mixture

$$f^{(r:n)} = \sum_{u=1}^{S} \frac{\binom{u-1}{r-1}\binom{S-u}{n-r}}{\binom{S}{n}} f^{(u:S)}, \qquad 1 \le r \le n \le S,$$

which holds because a sample of size n may be realised as a uniformly chosen n-subset of a sample of size S, and the r-th smallest member of the subset is the u-th smallest of the whole for exactly $\binom{u-1}{r-1}\binom{S-u}{n-r}$ of the $\binom{S}{n}$ subsets. So the question is whether those two weight vectors are comparable in a way the entropy respects. The identity is elementary and stands outside the machine-checked development; we verified it symbolically for all $1 \le r \le n \le S \le 9$.

6. Related work, and what was searched

The settled sibling of Question 1.1 is the *variance* comparison. Frey and Feeman [2] determine completely how PROSS schemes compare in the two-subset case and prove Öztürk's conjecture about how the efficiency of the PROSS sample mean compares with that of the RSS sample mean. So the question that this paper's question is modelled on was answered in 2017 for second moments; for entropy it was not, and [3] is where it is recorded as open. (We quote the abstract of [2]; we did not read its body, and nothing below depends on more than its abstract. Likewise we did not read [8], which enters only as the origin of the design and as a citation target counted below.)

The negative half of that last sentence is a report of a search, not a theorem, and we state what the search was. The searches were run in September 2026. We enumerated the forward citations of [3] in its preprint and its published form — four in OpenAlex, six in Semantic Scholar, eight in the Google Scholar citation cluster, eight in their union — and read them by title and abstract: none compares the two entropies; the entropy papers among them estimate the entropy of a parent distribution from ranked set data, which is a different problem. We enumerated both authors' complete OpenAlex work lists and found no entropy follow-up after 2016. We enumerated the 48 works citing Öztürk [8]; exactly one is about entropy and it is again entropy estimation. We queried the arXiv full-text search for *partially rank-ordered set*, *PROS sampling* and *rank-ordered set* (three statistics papers, all from 2014–15, one of them [3]) and for *ranked set* together with *entropy* in the abstract (seven, none about PROS). We grepped an 86 GB local full-text mirror of the mathematics arXiv for the same phrases: the hits are [3] itself, two bibliographies, and two papers that merely name the design. No paper we reached computes a PROS entropy for the uniform parent, or compares the two entropies of Question 1.1. This includes [3] itself, which never treats a uniform parent — the string "uniform" occurs twice in the source of v1, both times in commented-out lines and both times about the uniform random draw *within* a subset — and whose numerical work is Fisher-information relative efficiency for exponential, normal, logistic, extreme-value and gamma parents. The one place where the density (1) does appear in general form is the coherent-system literature discussed next.

Two named risks remain, and a referee with library access is better placed than we are to close them.

First, the published version of [3]*.* It is behind a paywall and its body was not read here; every pointer we make into [3] is to the arXiv version v1. The publisher's free front matter does show the abstract, the reference list and the appendix, and these say that the published version is an edited one, not a reprint of v1. The published abstract drops v1's closing sentence ("Several examples including the FI of PROS samples from the location-scale family of distributions as well as a regression model are discussed") and carries a few further copy-edits. The published reference list adds four items to v1's twenty-two — McIntyre [7], its 2005 reappearance in *The American Statistician*, and two papers of Stokes — all classical ranked-set-sampling references and none about entropy, so no entropy reference was added. But the published appendix numbers as Lemma 8, Lemma 9 and Theorem 4 the three results that v1 numbers Lemma 7, Lemma 8 and Theorem 4 (its Tables 9 and 10 keep their numbers), so the published version carries exactly one further numbered lemma somewhere before the appendix. Which subsection gained it cannot be determined from outside, and §4.1 cannot be excluded. Nothing in this paper depends on what [3] proves; the risk is to the claim that the comparison of Question 1.1 was still open, not to any theorem below.

Second, the coherent-system literature. The density (1) is, under another name, the lifetime density of a mixed coherent system of S i.i.d. components whose Samaniego signature is uniform on a consecutive block. Both papers we identified are paywalled and their bodies were not read; we read their abstracts. Toomaj and Doostparast [9] give an expression for the Shannon entropy of a mixed r-out-of-n system "in terms of the system signature, the distribution and density functions of the lifetime model, and the information measures of the beta distribution" — that is, a general formula for exactly the quantity $H(f_{(d_r)})$ of (1), which is why the block entropies of §4 are presented here as computations and not as discoveries. Toomaj, Di Crescenzo and Doostparast [10] give sufficient conditions for entropy orderings *between* coherent systems, i.e. between single mixture densities; Question 1.1 compares the two *sums* of (2) over a whole design, which is not a comparison of two systems, and neither abstract mentions sampling designs. We therefore do not believe either paper decides Question 1.1, but we did not read them.

7. Verification

Propositions 2.2, 3.2, 3.3, 4.1, 4.2, 5.4, Lemmas 2.3, 3.1, 3.5 and Theorems 5.1, 5.2, 5.3 are checked, in the form stated, by the kernel of the Lean 4 proof assistant [5] over the Mathlib library [6]: five modules, about a thousand lines, elaborated in 127 s. The development contains no `sorry` and declares no axiom of its own; of the thirty-six declarations audited with `#print axioms` — every declaration in the development that carries a provenance tag — thirty-five depend on `propext`, `Classical.choice` and `Quot.sound` and the remaining one on `propext` alone. In particular nothing is evaluated by Lean's compiled evaluator: there is no use of `native_decide`, and no floating-point or interval arithmetic anywhere, the comparisons of §5 being decided from rational two-sided bounds on $\log 2, \log 3, \log 5$ (from Mathlib) and $\log 7$ (Lemma 3.5, which fills a gap in that library). Densities are the polynomials (4) and entropies the interval integrals of §2, so what the kernel checks is the definition of (2) and not a numerical surrogate for it; blocks are indexed by (p, m, q) as in (4) so that no truncated natural-number subtraction occurs. Not formalised are Remarks 3.4, 4.3 and 5.5, Conjecture 5.6 together with the hypergeometric identity quoted after it, and the decimal truncations printed beside the exact values; the closed forms themselves are theorems. As an independent check outside the kernel, every exact value above was reproduced to 12 or more digits by 40-digit panelled quadrature of $-\int h \log h$, and the eight order-statistic entropies of Proposition 3.2 additionally by 50-digit quadrature. The repository is private; repository reference to be added at submission.

References

[1] N. Ebrahimi, E. S. Soofi and H. Zahedi, *Information properties of order statistics and spacings*, IEEE Transactions on Information Theory **50** (2004), no. 1, 177–183; `doi:10.1109/TIT.2003.821973`. The source, according to [4], of the Beta differential-entropy formula (6), which we quote as [4] quotes it; [1] is not open access and we did not read it.

[2] J. Frey and T. G. Feeman, *Efficiency comparisons for partially rank-ordered set sampling*, Statistical Papers **58** (2017), no. 4, 1149–1163; `doi:10.1007/s00362-016-0742-2`. The settled variance analogue of Question 1.1. Body not read; the sentences attributed to it in §1 and §6 are quoted from its abstract on the publisher's page.

[3] A. Hatefi and M. Jafari Jozani, *Information content of partially rank-ordered set samples*, arXiv:1504.07336v1 [math.ST], submitted 28 April 2015 (24 pages); published as AStA Advances in Statistical Analysis **101** (2017), no. 2, 117–149, `doi:10.1007/s10182-016-0277-9`. Every pointer to it above, and in particular the quoted sentence of §4.1, is to the arXiv version v1, which is the version we read; the arXiv record carries no journal reference, and the published bibliographic data are from Crossref and the publisher's page. In v1, Section 2 fixes the design and the mixture density (1); Section 4.1 carries the displays (2), the bound (3) and the quoted sentence. On the published version, which we did not read, see §6.

[4] M. Jafari Jozani and J. Ahmadi, *On uncertainty and information properties of ranked set samples*, Information Sciences **264** (2014), 291–301; `doi:10.1016/j.ins.2013.12.025`; preprint arXiv:1301.4292, which is the version we read. In the preprint, Section 2 quotes (6) and states that $H(X_{\mathrm{RSS}}) - H(X_{\mathrm{SRS}}) = \sum_i H(U_{(i)})$ is distribution-free; its Table 1, the only table it contains, carries the values $k(n)$, $n = 2, \ldots, 10$, reproduced in Remark 3.4. (The

volume and pages are those recorded by Crossref, OpenAlex and dblp. Several sources, including the reference list of [3] itself, instead give “Information Sciences **260**, 1–16”; that is wrong.)

[5] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: A. Platzer and G. Sutcliffe (eds.), Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, Cham, 2021, pp. 625–635; `doi:10.1007/978-3-030-79876-5_37`.

[6] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381; `doi:10.1145/3372885.3373824`.

[7] G. A. McIntyre, *A method for unbiased selective sampling, using ranked sets*, Australian Journal of Agricultural Research **3** (1952), no. 4, 385–390; `doi:10.1071/AR9520385`. The origin of ranked set sampling. (The journal is now *Crop and Pasture Science*, under which name [3] cites this paper; the article appeared again in The American Statistician **59** (2005), no. 3, 230–232, `doi:10.1198/000313005X54180`, and [3] lists both.)

[8] Ö. Öztürk, *Sampling from partially rank-ordered sets*, Environmental and Ecological Statistics **18** (2011), no. 4, 757–779; `doi:10.1007/s10651-010-0161-9`. The origin of the PROS design; cited as such by [3]. Not read here.

[9] A. Toomaj and M. Doostparast, *A note on signature-based expressions for the entropy of mixed r-out-of-n systems*, Naval Research Logistics **61** (2014), no. 3, 202–206; `doi:10.1002/nav.21577`. Body not read; the sentence quoted in §6 is from its abstract.

[10] A. Toomaj, A. Di Crescenzo and M. Doostparast, *Some results on information properties of coherent systems*, Applied Stochastic Models in Business and Industry **34** (2018), no. 2, 128–143; `doi:10.1002/asmb.2277`. Body not read; what is said about it in §6 is from its abstract.