# Partial rounding and near-independence.

Thomas Pensyl*

Joint work with Khoa Trinh, David Harris, and Aravind Srinivasan.

August 15, 2016

**Abstract**

Rounding. Dependent rounding is useful. Negative correlation is useful. Near-independence used. Entropy maximization often yields useful distributions. Goal of this paper is to explore if entropy maximization can be used to obtain distributions with near-independence properties.

## 1 Introduction

A very common method for solving (or approximating) combinatorial optimization problems is to first formulate the problem as an integer program, and then solve its linear relaxation. We then must find some clever way to 'round' the fractional solution into an integer solution, which represents an actual solution to the original problem. Often a randomized approach will give a good objective value in expectation and yield a much simpler analysis than deterministic approaches. This paper studies distributions which may be useful for this purpose.

Let $\mathbf{x} \in [0,1]^n$ be a fractional vector and $\mathbf{X} \in [0,1]^n$ be a vector randomly drawn from a random distribution. Also let $\mathbf{a} \in \mathbf{R}_+^n$. Building off Srinivasan's dependent rounding sampling algorithm in [3], Byrka et. al. in [1] give a distribution with the following properties:

| | | |
|---|---|---|
| **(A0)** | Almost-integral: | $\mathbf{X}$ contains at most $O(\log \frac{a_{max}}{a_{min}})$ non-integral elements. |
| **(A1)** | Preserves Marginals: | $\forall i$, $\mathbf{E}[X_i] = x_i$ |
| **(A2)** | Preserves Weighted Sum: | $\mathbf{a} \cdot \mathbf{X} = \mathbf{a} \cdot \mathbf{x}$ |
| **(A3)** | Negative Correlation: | $\forall S \subseteq [n]$ $\mathbf{E}[\prod_{i \in S} X_i] \leq \prod_{i \in S} x_i$, and $\mathbf{E}[\prod_{i \in S}(1 - X_i)] \leq \prod_{i \in S}(1 - x_i)$. |
| **(A4)** | Near-Independence: | For *small*, disjoint $S, T \in [n]$: $\mathbf{E}[\prod_{i \in S} X_i \prod_{i \in T}(1 - X_i)] \approx \prod_{i \in S} x_i \prod_{i \in T}(1 - x_i)$ |

Notice we can get full integrality (stronger than (A0)), preserved marginals (A1), and full-independence (stronger than (A3) and (A4)), by simply independently setting each $X_i = 1$ with probability $x_i$ and $X_i = 0$ with probability $1 - x_i$. It is the linear constraint of (A2) which weakens the other properties. **In this paper we explore distributions with a direct tradeoff between (A0) and (A4).**

The reason we are okay with weakening property (A0) is that in some applications, it suffices to obtain a 'mostly-integral' vector, as long as there are at most a constant number of fractional elements, which can then be dealt with in a brute force way to obtain the final solution. This is exactly the approach used in [1] for $k$-median, where $O(1)$ fractional elements result in a solution with $k + O(1)$ facilities, and a preprocessing step from Li and Svensson[2] results in a valid solution with exactly $k$ facilities.

In Section 2, we consider two explicit distributions and prove they provide tradeoff between near-independence and number of fractional elements. In Section 3, we define the entropy of such distributions and give experimental evidence that suggests the maximum entropy distribution may have some of the near-independence properties.

*This work was supported in part by NSF Awards CNS-1010789 and CCF-1422569.

## 2 Symmetric Randomized Dependent Rounding

Suppose we are given vectors $\mathbf{x} = (x_1, \ldots, x_n) \in (0,1)^n$, $\mathbf{a} = (a_1, \ldots, a_n) \in (\mathbb{R} \setminus \{0\})^n$. (If $a_i$ is 0, we would exclude $x_i$ from this vector and just round it independently. Also if $\mathbf{x}$ has any integral elements already, we just run this algorithm on the subvector of fractional elements.) The following algorithm describes a single randomized rounding step.

**Algorithm 1** SRDR($\mathbf{x}, \mathbf{a}$)

1: $\delta \leftarrow \min_{i: a_i \neq 0} \{\frac{x_i}{|a_i|}, \frac{1-x_i}{|a_i|}\}$.
2: Choose random distinct $i^*, j^* \in [n]$.
3: With probability $\frac{1}{2}$, $\mathbf{X} \leftarrow \mathbf{x} + \delta(a_{i^*}\mathbf{e}_{i^*} - a_{j^*}\mathbf{e}_{j^*})$, else $\mathbf{X} \leftarrow \mathbf{x} - \delta(a_{i^*}\mathbf{e}_{i^*} - a_{j^*}\mathbf{e}_{j^*})$.
4: **return** $\mathbf{X}$

- The direction vectors in step 3 are chosen such that the weighted sum is preserved deterministically ($\sum_i a_i X_i = \sum_i a_i x_i$), and the marginals are preserved in expectation ($E[X_i] = x_i$).
- $\delta$ is chosen to be the maximum value such all possible values of $\mathbf{X}$ in step 3 remain in $[0,1]^n$. Because it is the maximum such value, at least one of the $2\binom{n}{2}$ values possible from steps 2 and 3 will have at least 1 fixed (integral) element. Therefore with probability at least $\frac{2}{n^2}$ SRDR fixes an element. If this step is repeated $(\log \frac{1}{2\epsilon}) \cdot n^2$ times, it fixes at least one component with probability at least $1 - \epsilon$.
- The random choice of $i^*, j^*$ in step 2 is the key difference from ordinary (weighted) dependent rounding. As the next two sections will show, this step suffices to provide additional near-independence properties.

### 2.1 Upper bound on near-independence.

**Theorem 2.1: SRDR: Near-Negative Correlation**

Let $\mathbf{X} = \text{SRDR}(\mathbf{x}, \mathbf{a})$ for some vectors $\mathbf{x} \in (0,1)^n$ and $\mathbf{a} \in (\mathbb{R} \setminus \{0\})^n$, $n \geq 2$. For any set $S \subseteq [n]$,

$$\mathbf{E}\left[\Big(\prod_{i \in S} X_i\Big)^p\right] \leq \Big(\prod_{i \in S} x_i\Big)^p \tag{1}$$

holds for $p = 1 - \frac{1}{n}$. Furthermore, if all (nonzero) weights in $\{a\}_{i \in S}$ have the same sign, the inequality holds for $p = 1$.

*Proof.* Fix a set $S \in [n]$. We will analyze the expected value of $\Lambda := \prod_{i \in S} X_i$ in terms of $\lambda := \prod_{i \in S} x_i$. Define

$$b_i := \begin{cases} a_i & i \in S \\ 0 & i \notin S \end{cases}, \qquad A_i := 1 + \frac{b_i \delta}{x_i}, \qquad B_i := 1 - \frac{b_i \delta}{x_i}.$$

Then the expected value of $\Lambda$ conditioned on particular $i^*, j^*$ being chosen is:

$$\mathbf{E}[\Lambda^p | \{i^*, j^*\} = \{i, j\}] = \frac{1}{2}\lambda^p \left(\frac{x_i + b_i\delta}{x_i}\right)^p \left(\frac{x_j - b_j\delta}{x_j}\right)^p + \frac{1}{2}\lambda^p \left(\frac{x_i - b_i\delta}{x_i}\right)^p \left(\frac{x_j + b_j\delta}{x_j}\right)^p \tag{2}$$

$$= \frac{1}{2}\lambda^p (A_i^p B_j^p + B_i^p A_j^p). \tag{3}$$

Observe that $\mathbf{E}[\Lambda | \{i^*, j^*\} = \{i, j\}] = \lambda(1 - \frac{b_i b_j \delta^2}{x_i x_j})$, and if $b_i$ *and* $b_j$ have the same sign (or one or both are 0), then this quantity is at most $\lambda$. Thus, if *all* of $\{a_i\}_{i \in S}$ have the same sign, SRDR preserves full negative correlation: $\mathbf{E}[\Lambda] \leq \lambda$. (This is equivalent to the known negative correlation property of positively-weighted dependent rounding.)

We now continue with the general case. What the proof boils down to is expanding all terms of the form $(1+x)^p$ using the generalized binomial theorem, multiplying everything to get a polynomial in $x$, and then showing (in Lemma 2.1) that $p = 1 - \frac{1}{n}$ is the precise value which causes the higher order terms to vanish (and the lower order terms to be negative). However, the algebra is much cleaner if we first massage the equation before doing any expansions.

In the following summations, $i, j \in [n]$.

$$\mathbf{E}[\Lambda^p] = \sum_{i<j} \Pr[\{i^*, j^*\} = \{i, j\}]\, \mathbf{E}[\Lambda^p | \{i^*, j^*\} = \{i, j\}] \tag{4}$$

$$\mathbf{E}[\Lambda^p] = \sum_{i<j} \frac{1}{\binom{n}{2}} \cdot \frac{1}{2} \lambda^p (A_i^p B_j^p + B_i^p A_j^p) \tag{5}$$

$$\frac{4}{\lambda^p}\binom{n}{2}\mathbf{E}[\Lambda^p] = \sum_{i<j} (A_i^p + B_i^p)(A_j^p + B_j^p) - (A_i^p - B_i^p)(A_j^p - B_j^p) \tag{6}$$

$$\frac{8}{\lambda^p}\binom{n}{2}\mathbf{E}[\Lambda^p] = \Big(\sum_i (A_i^p + B_i^p)\Big)^2 - \sum_i (A_i^p + B_i^p)^2 - \Big(\sum_i (A_i^p - B_i^p)\Big)^2 + \sum_i (A_i^p - B_i^p)^2 \tag{7}$$

$$\frac{8}{\lambda^p}\binom{n}{2}\mathbf{E}[\Lambda^p] \le n \sum_i (A_i^p + B_i^p)^2 - \sum_i (A_i^p + B_i^p)^2 + \sum_i (A_i^p - B_i^p)^2 \tag{8}$$

$$= \sum_i \big(n(A_i^{2p} + B_i^{2p}) + 2(n-2)A_i^p B_i^p\big). \tag{9}$$

To get (8) we applied the Cauchy-Schwarz inequality and the fact that any square is nonnegative. We will show that for the appropriate choice of $p$, $\mathbf{E}[\Lambda^p] \le \lambda^p$. We now expand $A_i$ and $B_i$ using the generalized binomial theorem.

$$\frac{8}{\lambda^p}\binom{n}{2}\mathbf{E}[\Lambda^p] \le \sum_i \left( n\Big(1 + \frac{b_i\delta}{x_i}\Big)^{2p} + n\Big(1 - \frac{b_i\delta}{x_i}\Big)^{2p} + 2(n-2)\Big(1 - (\frac{b_i\delta}{x_i})^2\Big)^p \right) \tag{10}$$

$$= \sum_i \left( n \sum_{k\ge 0} \binom{2p}{k}\Big(\frac{b_i\delta}{x_i}\Big)^k + n \sum_{k\ge 0}\binom{2p}{k}(-1)^k\Big(\frac{b_i\delta}{x_i}\Big)^k + 2(n-2)\sum_{k\ge 0}\binom{p}{k}(-1)^k\Big(\frac{b_i\delta}{x_i}\Big)^{2k} \right) \tag{11}$$

$$= \sum_i \left( 2n \sum_{\ell\ge 0}\binom{2p}{2\ell}\Big(\frac{b_i\delta}{x_i}\Big)^{2\ell} + 2(n-2)\sum_{k\ge 0}\binom{p}{k}(-1)^k\Big(\frac{b_i\delta}{x_i}\Big)^{2k} \right) \tag{12}$$

$$= 2\sum_i \sum_{k\ge 0}\left( n\binom{2p}{2k} + (n-2)\binom{p}{k}(-1)^k \right)\Big(\frac{b_i\delta}{x_i}\Big)^{2k}. \tag{13}$$

We now fix $p = 1 - \frac{1}{n}$ and apply Lemma 2.1 from below:

$$\frac{8}{\lambda^{1-1/n}}\binom{n}{2}\mathbf{E}[\Lambda^{1-1/n}] \le 2\sum_i \sum_{k\ge 0} f_k(n, 1-1/n)\Big(\frac{b_i\delta}{x_i}\Big)^{2k} \tag{14}$$

$$\le 2\sum_i f_0(n, 1-1/n)\Big(\frac{b_i\delta}{x_i}\Big)^0 \tag{15}$$

$$= 2n(n\cdot 1 + (n-2)\cdot 1(-1)^0) = 4n(n-1) \tag{16}$$

$$\mathbf{E}[\Lambda^{1-1/n}] \le \lambda^{1-1/n}. \tag{17}$$

□

**Lemma 2.1**

Define $f_k(n,p) := n\binom{2p}{2k} + (n-2)\binom{p}{k}(-1)^k$. Then $f_k(n, 1-1/n) \le 0$ for all integers $k \ge 1$ and $n \ge 2$.

*Proof.* First consider the $k=1$ case:

$$f_1\Big(n, 1-\frac{1}{n}\Big) = n\frac{2(1-\frac{1}{n})(2(1-\frac{1}{n})-1)}{2} - (n-2)(1-\frac{1}{n}) = \Big(1-\frac{1}{n}\Big)\Big(n\Big(2-\frac{2}{n}-1\Big)-n+2\Big) = 0. \quad (18)$$

We show the $k \geq 2$ case by induction: assume $f_{k-1}(n, 1-\frac{1}{n}) \leq 0$. It is straightforward to verify the following recurrence holds:

$$f_k(n,p) = \Big(1-\frac{p+1}{k}\Big)\Big(f_{k-1}(n,p) - n\frac{2p}{2k-1}\binom{2p}{2(k-1)}\Big). \quad (19)$$

When $k \geq 2$, $n \geq 1$, and $p = 1-\frac{1}{n}$, then $\frac{p+1}{k} \leq 1$. Also $2p = 2 - 2/n \in [1,2]$, so $\binom{2p}{2(k-1)} \geq 0$. Thus the recurrence implies $f_k(n, 1-\frac{1}{n}) \leq f_{k-1}(n, 1-\frac{1}{n}) \leq 0$. ☐

**Corollary 2.1: SRDR: Near-Independence, Upper Bound**

Let $\mathbf{X} = \text{SRDR}(\mathbf{x}, \mathbf{a})$ for some vectors $\mathbf{x} \in (0,1)^n$ and $\mathbf{a} \in (\mathbb{R} \setminus \{0\})^n$, $n \geq 2$. Let $S, T \subseteq [n]$ be disjoint subsets. Then

$$\mathbf{E}\Big[\Big(\prod_{i\in S} X_i \prod_{i \in T}(1-X_i)\Big)^p\Big] \leq \Big(\prod_{i\in S} x_i \prod_{i\in T}(1-x_i)\Big)^p, \quad (20)$$

holds for $p = 1-\frac{1}{n}$. Furthermore, if all $\{a_i\}_{i\in S}$ are positive (or $S = \emptyset$) and all $\{a_i\}_{i\in T}$ are negative (or $T=\emptyset$), or vice versa, the inequality holds for $p=1$.

*Proof.* Given sets $S, T$, define vectors $\mathbf{X}'$, $\mathbf{x}'$, and $\mathbf{a}'$:

$$X_i' := \begin{cases} 1-X_i & i \in T \\ X_i & i \notin T \end{cases}, \qquad x_i' := \begin{cases} 1-x_i & i \in T \\ x_i & i \notin T \end{cases}, \qquad a_i' := \begin{cases} -a_i & i \in T \\ a_i & i \notin T \end{cases} \quad (21)$$

It is not hard to see that if $\mathbf{X} = \text{SRDR}(\mathbf{x}, \mathbf{a})$, then $\mathbf{X}' = \text{SRDR}(\mathbf{x}', \mathbf{a}')$. Then

$$\mathbf{E}\Big[\Big(\prod_{i\in S} X_i \prod_{i\in T}(1-X_i)\Big)^p\Big] = \mathbf{E}\Big[\Big(\prod_{i\in S\cup T} X_i'\Big)^p\Big] \leq \Big(\prod_{i \in S\cup T} x_i'\Big)^p = \Big(\prod_{i\in S} x_i \prod_{i\in T}(1-x_i)\Big)^p \quad (22)$$

holds for $p = 1-\frac{1}{n}$. If all $\{a_i\}_{i\in S}$ are positive and all $\{a_i\}_{i\in T}$ are negative (or vice versa), then all $\{a_i'\}_{i\in S\cup T}$ have the same sign so (22) holds for $p=1$. ☐

Notice by Jensen's inequality, the above results also hold for $p$ smaller than the proved values.

## 2.2 Lower bound on near-independence.

We now prove a corresponding lower bound. The proof here is less elegant. We start directly with the binomial expansions. For $p \in [1,2]$, we can give simple lower bounds. (A more sophisticated proof could likely remove this restriction on $p$, but the current result suffices if we are interested in $p = 1+\epsilon$.)

**Lemma 2.2: Truncated Binomial Theorem**

For $x \in [-1,1]$ and $p \in [1,2]$,

$$(1+x)^p \geq 1 + px + \binom{p}{2}x^2 + \binom{p}{3}x^3$$

*Proof.* We start with the Maclaurin series (i.e. the generalized binomial theorem):

$$(1+x)^p = \sum_{k=0}^{\infty} \binom{p}{k} x^k.$$

The lemma follows if we can show positive (the case $x = 0$ is trivial) the sum of all terms for $k \geq 4$.

$$\sum_{k=4}^{\infty} \binom{p}{k} x^k = \sum_{j=2}^{\infty} \left( \binom{p}{2j} x^{2j} + \binom{p}{2j+1} x^{2j+1} \right) = \sum_{j=2}^{\infty} \binom{p}{2j} x^{2j} \left( 1 + \frac{p-2j}{2j+1} x \right)$$

Since $p \in [1, 2]$, $\binom{p}{2j}$ is positive for all $j \geq 2$. The term $1 + \frac{p-2j}{2j+1} x \geq 1 - \left| \frac{2j-p}{2j+1} x \right| \geq 1 - \frac{2j-1}{2j+1} |x| > 0$. Therefore, each term in the last sum is positive. □

**Lemma 2.3**

For $a, b \in [0, 1]$ and $p \in [1, 2]$, the following hold:

$$\frac{1}{2}((1+a)^p(1-b)^p + (1-a)^p(1+b)^p) \geq 1 - \frac{p}{2}(a^2 + b^2) \tag{23}$$

$$\frac{1}{2}((1+a)^p(1+b)^p + (1-a)^p(1-b)^p) \geq 1 + \binom{p}{2}(a^2 + b^2) \tag{24}$$

$$\frac{1}{2}((1+a)^p + (1-a)^p) \geq 1 + \binom{p}{2} a^2. \tag{25}$$

*Proof.* For (23), we apply Lemma 2.2 and get:

$$\begin{aligned}
&\frac{1}{2}\big((1+a)^p(1-b)^p + (1-a)^p(1+b)^p\big) \\
\geq\ & \frac{1}{2}(1 + pa + \binom{p}{2}a^2 + \binom{p}{3}a^3)(1 - pb + \binom{p}{2}b^2 - \binom{p}{3}b^3) \\
&+ \frac{1}{2}(1 - pa + \binom{p}{2}a^2 - \binom{p}{3}a^3)(1 + pb + \binom{p}{2}b^2 + \binom{p}{3}b^3) \\
=& 1 + \binom{p}{2}(a^2 + b^2) - p^2 ab - p\binom{p}{3}(ab^3 + a^3 b) + \binom{p}{2}^2 a^2 b^2 - \binom{p}{3}^2 a^3 b^3 \\
=& 1 + \left( \binom{p}{2} - \frac{p^2}{2} \right)(a^2 + b^2) + \frac{p^2}{2}(a-b)^2 - p\binom{p}{3}(ab^3 + a^3 b) + \binom{p}{2}^2 a^2 b^2 (1 - \frac{(p-2)^2}{3^2} ab) \\
\geq& 1 - \frac{p}{2}(a^2 + b^2) + \binom{p}{2}^2 a^2 b^2 (1 - \frac{1}{9}) \\
\geq& 1 - \frac{p}{2}(a^2 + b^2).
\end{aligned}$$

In particular we have used $ab = \frac{1}{2}(a^2 + b^2) - \frac{1}{2}(a-b)^2$, $a \leq 1$, $b \leq 1$, $p \geq 1$, and $\binom{p}{3} \leq 0$.

Again applying Lemma 2.2, expanding, and cancelling, we get:

$$\begin{aligned}
&\frac{1}{2}\big((1+a)^p(1+b)^p+(1-a)^p(1-b)^p\big)\\
\geq&1+\binom{p}{2}(a^2+b^2)+p^2ab+p\binom{p}{3}(ab^3+a^3b)+\binom{p}{2}^2a^2b^2+\binom{p}{3}^2a^3b^3\\
\geq&1+\binom{p}{2}(a^2+b^2)+p^2ab\Big(1+\frac{(p-1)(p-2)}{6}(a^2+b^2)\Big)\\
\geq&1+\binom{p}{2}(a^2+b^2)+p^2ab\Big(1+\frac{-(1/2)^2}{6}(2)\Big)\\
\geq&1+\binom{p}{2}(a^2+b^2).
\end{aligned}$$

(25) follows directly from (24) by setting $b=0$. □

Now we prove the full result. We show it only for positive weight vectors, but again conjecture the result should still hold when negative weights are allowed.

**Theorem 2.2: SRDR:Near-Independence, Lower Bound**

Let $X=\mathrm{SRDR}(\mathbf{x},\mathbf{a})$ for some vectors $\mathbf{x}\in[0,1]^n$ and $\mathbf{a}\in\mathbb{R}^n_{>0}$. For any pair of distinct subsets $S,T\subseteq[n]$,

$$E\Big[\Big(\prod_{i\in S}X_i\prod_{j\in T}(1-X_j)\Big)^p\Big]\geq\Big(\prod_{i\in S}x_i\prod_{j\in T}(1-x_j)\Big)^p,$$

for $p=1+\max\left\{\frac{|S|-1}{n-|S|},\frac{|T|-1}{n-|T|}\right\}$, *if* $p\leq 2$.

*Proof.* Fix sets $S,T$ and let $s,t$ be their respective cardinalities. Define $\lambda:=\prod_{i\in S}x_i^p\prod_{j\in T}(1-x_j)^p$ and $\Lambda:=\prod_{i\in S}X_i\prod_{j\in T}(1-X_j)$. Let $P=\{i,j\}$ be the random pair of distinct indices chosen in line 2 of SRDR. Then $E[\Lambda^p|P]$ takes one of several different forms depending on the intersections of $P$ with $S$ and $T$.

$$E[\Lambda^p]=\frac{1}{\binom{n}{2}}\sum_P E[\Lambda^p|P]=\frac{1}{\binom{n}{2}}\Big(\sum_{P\cap(S\cup T)=\emptyset}E[\Lambda^p|P]+\sum_{P\cap(S\cup T)=\{i\}}E[\Lambda^p|P]+\sum_{\substack{P\cap S=\{i\}\\ \text{and } P\cap T=\{j\}}}E[\Lambda^p|P]+\sum_{\substack{P\cap S=\{i,j\}\\ \text{or } P\cap T=\{i,j\}}}E[\Lambda^p|P]\Big) \tag{26}$$

Define $y_i:=x_i$ for $i\in S$, and $y_i:=1-x_i$ for $i\in T$. Observe $\lambda=(\prod_{i\in S\cup T}y_i)^p$. Also define $z_i:=\frac{a_i\,\delta}{y_i}$, and notice by choice of $\delta$ in line 1 of SRDR that $z_i\in[0,1]$. Then, for example, we may express expectations in the following form:

$$\begin{aligned}
\mathbf{E}[\Lambda^p|P=(i\in S,j\in T)]&=\frac{1}{2}\Big(\prod_{\ell\notin\{i,j\}}y_\ell\Big)^p(x_i+a_i\,\delta)^p(1-(x_j-a_j\,\delta))^p+\frac{1}{2}\Big(\prod_{\ell\notin\{i,j\}}y_\ell\Big)^p(x_i-a_i\,\delta)^p(1-(x_j+a_j\,\delta))^p)\\
&=\frac{1}{2}\Big(\prod_{\ell\notin\{i,j\}}y_\ell\Big)^p(y_i+a_i\,\delta)^p(y_j+a_j\,\delta)^p+\frac{1}{2}\Big(\prod_{\ell\notin\{i,j\}}y_\ell\Big)^p(y_i-a_i\,\delta)^p(y_j-a_j\,\delta)^p)\\
&=\frac{\lambda^p}{2}((1+z_i)^p(1+z_j)^p+(1-z_i)^p(1-z_j)^p)\\
&\geq\lambda^p(1+\binom{p}{2}(z_i^2+z_j^2)),
\end{aligned}\tag{27}$$

where we applied Lemma 2.3 in the last step. Similarly, we get $E[P\cap(S\cup T)=\emptyset]=\lambda^p$ and

$$\mathbf{E}\left[\Lambda^p|\ {}_{\text{or } P\cap T=\{i,j\}}^{P\cap S=\{i,j\}}\right]=\frac{\lambda^p}{2}((1+z_i)^p(1-z_j)^p+(1-z_i)^p(1+z_j)^p)\geq\lambda^p\big(1-\frac{p}{2}(z_i^2+z_j^2)\big), \tag{28}$$

$$\mathbf{E}[\Lambda^p|P\cap(S\cup T)=\{i\}]=\frac{\lambda^p}{2}((1+z_i)^p+(1-z_i)^p)\geq\lambda^p(1+\binom{p}{2}z_i^2). \tag{29}$$

Now we apply these to (26). ($i \neq j$ and $i, j \in [n]$ is implicit in all sums.)

$$\begin{aligned}
\binom{n}{2}\frac{E[\Lambda^p]}{\lambda^p} &\geq \sum_{\substack{i\notin(S\cup T)\\ j\notin(S\cup T)}} 1 + \sum_{\substack{i\in(S\cup T)\\ j\notin(S\cup T)}} (1+\binom{p}{2}z_i^2) + \sum_{\substack{i\in S\\ j\in T}} (1+\binom{p}{2}(z_i^2+z_j^2)) + \sum_{\substack{\{i,j\}\subseteq S\\ \text{or } \{i,j\}\subseteq T}} (1-\frac{p}{2}(z_i^2+z_j^2)) \\
&= \binom{n}{2} + \sum_{\substack{i\in(S\cup T)\\ j\notin(S\cup T)}} \binom{p}{2}z_i^2 + \sum_{\substack{i\in S\\ j\in T}} \binom{p}{2}(z_i^2+z_j^2) - \sum_{\substack{\{i,j\}\subseteq S\\ \text{or } \{i,j\}\subseteq T}} \frac{p}{2}(z_i^2+z_j^2) \\
&= \binom{n}{2} + (n-s-t)\sum_{i\in(S\cup T)} \binom{p}{2}z_i^2 + t\binom{p}{2}\sum_{i\in S} z_i^2 + s\binom{p}{2}\sum_{i\in T} z_i^2 - (s-1)\frac{p}{2}\sum_{i\in S} z_i^2 - (t-1)\frac{p}{2}\sum_{i\in T} z_i^2 \\
&= \binom{n}{2} + \left(\binom{p}{2}(n-s) - p\frac{s-1}{2}\right)\sum_{i\in S} z_i^2 + \left(\binom{p}{2}(n-t) - p\frac{t-1}{2}\right)\sum_{i\in T} z_i^2 \\
&= \binom{n}{2} + \frac{p}{2}\Big((p-1)(n-s)-(s-1)\Big)\sum_{i\in S} z_i^2 + \frac{p}{2}\Big((p-1)(n-t)-(t-1)\Big)\sum_{i\in T} z_i^2 \\
&= \binom{n}{2} + 0 + 0 \\
E[\Lambda^p] &\geq \lambda^p.
\end{aligned}$$

The penultimate step follows from $p-1 \geq \max\left\{\frac{s-1}{n-s}, \frac{t-1}{n-t}\right\}$. □

Again by Jensen's inequality, this result also holds for any $p$ larger than the stated value.

## 2.3 Comparison to random permutation method

In [ourkmedpaper], the authors obtain near-independence results by running dependent rounding in an order determined by a single random permutation. In the interest of comparison, we will convert our result into something more closely resembling theirs. Suppose we start with $n$ variables and run SRDR until only $k$ fractional variables remain. In the following bounds we assume $n$ and $k$ are sufficiently large and make rough approximations, ignoring constants. Let $\alpha = \min_i\{x_i, 1-x_i\}$, and let $t$ be the number of terms in our product (i.e. $|S\cup T|$). Now $\lambda^{1-1/(k+1)} \approx \lambda\Big(1-\frac{\log\lambda}{k}\Big) \leq \lambda\Big(1-\frac{\log\alpha^t}{k}\Big) = \lambda\Big(1+\frac{t\log\frac{1}{\alpha}}{k}\Big)$. So SRDR gives:

$$\lambda\Big(1-\frac{t^2\log\frac{1}{\alpha}}{k}\Big) \leq E[\Lambda] \leq \lambda\Big(1+\frac{t\log\frac{1}{\alpha}}{k}\Big)\lambda.$$

For dependent rounding on a random permutation, Theorem 2.10[?] leaves $k = O(\log\frac{a_{max}}{a_{min}})$ fractional elements and gives (roughly)

$$\lambda\Big(1-\frac{t^2}{n\alpha^3}\Big) \leq E[\Lambda] \leq \Big(1+\frac{t^2}{n\alpha^3}\Big)\lambda.$$

In the case that all $x_i$ are uniform, Theorem 2.11[] leaves $k = \frac{1}{\alpha}O(\log\frac{a_{max}}{a_{min}})$ fractional elements and gives

$$\lambda\Big(1-\frac{t^2}{n\alpha^2}\Big) \leq E[\Lambda] \leq \Big(1+\frac{t^2}{n\alpha^2}\Big)\lambda.$$

In the case that the weights $a_i$ are uniform, Theorem 2.12[] leaves $k \leq 1$ fractional elements and gives

$$\lambda\Big(1-\frac{t^2}{n\alpha^2}\Big) \leq E[\Lambda] \leq \Big(1+\frac{t^2}{n\alpha^2}\Big)\lambda.$$

The results are tricky to compare without looking at a specific application. Clearly SRDR has a much smaller dependence on $\alpha$, and in the weighted case avoids the dependence on $\frac{a_{max}}{a_{min}}$. However, it's likely $k$, and therefore $t$ can be at most a constant, whereas in the random permutation method, you can have $t = O(\sqrt{n})$ and still get a good bound, though in the weighted case, you must also be able to bound $\frac{a_{max}}{a_{min}}$.

## 2.4 Closest Edge

In this section we describe an alternate distribution which, experimentally, gives the same guarantees as SRDR, though we have proved less (and only for the unweighted case), as the proof is messier. The difference here is that we randomly move along one of $2n$ directions, where each complementary pair of directions is determined by the shortest distance toward each face of the polytope.

**Algorithm 2** ClosestEdge

1: $\delta \leftarrow \frac{1}{n-1} \min_i \{x_i, 1 - x_i\}$.
2: With probability $\frac{1}{2}$, $\delta \leftarrow -\delta$.
3: $i \leftarrow \text{Random}(1, \ldots, n)$.
4: $\mathbf{x} \leftarrow (x_1 + \delta, \ldots, x_i - (n-1)\,\delta, \ldots, x_n + \delta)$.

This step chooses uniformly at random from $2n$ symmetric vectors defined such that: no vector exits the unit cube; at least one vector hits a face (a component goes to 0 or 1); all vectors have equal magnitude; and all vectors lie along the hyperplane where $\sum x_i$ is constant.

**Theorem 2.3: Closest Edge: Near-negative correlation**

Let $\mathbf{x} \in (0,1)^n$, and $\mathbf{X} = \mathit{ClosestEdge}(\mathbf{x})$. For any distinct pair $(i, j)$,

$$\mathbf{E}[X_i X_j] < x_i x_j.$$

*Proof.* For brevity define $r := n - 1$, $x := \frac{\delta}{x_i}$, and $y := \frac{\delta}{x_j}$.

$$\begin{aligned}
E[X_i X_j] =& \frac{1}{2n}\big((x_i + r\,\delta)(x_j - \delta) + (x_i - \delta)(x_j + r\,\delta) + (n-2)(x_i + \delta)(x_j + \delta) \\
&+ (x_i - r\,\delta)(x_j + \delta) + (x_i + \delta)(x_j - r\,\delta) + (n-2)(x_i - \delta)(x_j - \delta)\big) \\
=& \frac{x_i x_j}{2n}\big((1 + rx)(1 - y) + (1 - x)(1 + ry) + (n-2)(1 + x)(1 + y) \\
&+ (1 - rx)(1 + y) + (1 + x)(1 - ry) + (n-2)(1 - x)(1 - y)\big) \\
=& \frac{x_i x_j}{n}\big((1 - rxy) + (1 - rxy) + (n-2)(1 + xy)\big) \\
=& \frac{x_i x_j}{n}\big(n + (n - 2 - 2r)xy\big) = \frac{xy}{n}\big(n + (n - 2 - 2n + 2)xy\big) \\
=& x_i x_j\big(1 - xy\big) \\
<& x_i x_j
\end{aligned}$$

□

**Theorem 2.4: Closest Edge: Near-positive correlation.**

Let $\mathbf{x} \in (0,1)^n$, and $\mathbf{X} = \mathit{ClosestEdge}(\mathbf{x})$. For any distinct pair $(i, j)$,

$$\mathbf{E}[(X_i X_j)^p] \geq (x_i x_j)^p,$$

when $p \geq 1 + 1/(n-2)$.

*Proof.* Again, let $r := n-1$, $x := \frac{\delta}{x_i}$, and $y := \frac{\delta}{x_j}$.

$$\begin{aligned}E[(X_iX_j)^p] =& \frac{1}{2n}\big((x_i + r\,\delta)^p(x_j-\delta)^p + (x_i - \delta^p)(x_j + r\,\delta)^p + (n-2)(x_i+\delta)^p(x_j+\delta)^p \\ &+ (x_i - r\,\delta)^p(x_j+\delta)^p + (x_i+\delta)^p(x_j - r\,\delta)^p + (n-2)(x_i-\delta)^p(x_j-\delta)^p\big) \\ =& \frac{x_i^p x_j^p}{2n}\big((1+rx)^p(1-y)^p + (1-x)^p(1+ry)^p + (n-2)(1+x)^p(1+y)^p \\ &+ (1-rx)^p(1+y)^p + (1+x)^p(1-ry)^p + (n-2)(1-x)^p(1-y)^p\big)\end{aligned}$$

We apply Lemma 2.2. The algebra is messy, so we will deal with each pair of terms which has nice cancellation properties. Due to the alternating signs, terms with net odd powers cancel out. The first and fourth terms are bounded below by:

$$\begin{aligned}& (1+rx)^p(1-y)^p + (1-rx)^p(1+y)^p \\ \geq\ & (1 + prx + \binom{p}{2}r^2x^2 + \binom{p}{3}r^3x^3)(1 - py + \binom{p}{2}y^2 - \binom{p}{3}y^3) \\ & + (1 - prx + \binom{p}{2}r^2x^2 - \binom{p}{3}r^3x^3)(1 + py + \binom{p}{2}y^2 + \binom{p}{3}y^3) \\ =& 2\Big(1 + \binom{p}{2}(r^2x^2+y^2) - p^2rxy - p\binom{p}{3}(rxy^3 + r^3x^3y) + \binom{p}{2}^2 r^2x^2y^2 - \binom{p}{3}^2 r^3x^3y^3\Big)\end{aligned}$$

The second and fifth terms are the same as first and fourth, but with $x, y$ swapped, and so similarly sum to:

$$2\Big(1 + \binom{p}{2}(x^2+r^2y^2) - p^2rxy - p\binom{p}{3}(r^3xy^3 + rx^3y) + \binom{p}{2}^2 r^2x^2y^2 - \binom{p}{3}^2 r^3x^3y^3\Big)$$

The third and sixth terms:

$$\begin{aligned}& (n-2)(1+x)^p(1+y)^p + (n-2)(1-x)^p(1-y)^p \\ \geq\ & (n-2)(1 + px + \binom{p}{2}x^2 + \binom{p}{3}x^3)(1 + py + \binom{p}{2}y^2 + \binom{p}{3}y^3) \\ & + (n-2)(1 - px + \binom{p}{2}x^2 - \binom{p}{3}x^3)(1 - py + \binom{p}{2}y^2 - \binom{p}{3}y^3) \\ & = 2(n-2)\Big(1 + \binom{p}{2}(x^2+y^2) + p^2xy + p\binom{p}{3}(xy^3 + x^3y) + \binom{p}{2}^2 x^2y^2 + \binom{p}{3}^2 x^3y^3\Big)\end{aligned}$$

Summing it all up (and cancelling out the factor of 2) we get

$$\begin{aligned}\frac{\mathbf{E}[(X_iX_j)^p]}{(x_ix_j)^p} \geq& \frac{1}{n}\Big(n + \binom{p}{2}(x^2+y^2)(n-2+r^2+1) + p^2xy(n-2-2r) \\ &+ p\binom{p}{3}xy(x^2+y^2)(n-2-r-r^3) + \binom{p}{2}^2 x^2y^2(n-2+2r^2) + \binom{p}{3}^2 x^3y^3(n-2-2r^3)\Big)\end{aligned}$$

Now substitute $r = n-1$, and divide everything through by $n$.

$$\frac{\mathbf{E}[(X_iX_j)^p]}{(x_ix_j)^p} \geq 1 + \binom{p}{2}(x^2+y^2)(n-1) - p^2xy \tag{30}$$

$$- p\binom{p}{3}xy(x^2+y^2)(n^2-3n+3) + \binom{p}{2}^2 x^2y^2(2n-3) - \binom{p}{3}^2 x^3y^3(2n^2-6n+5) \tag{31}$$

We apply $(x^2+y^2) = (x-y)^2 + 2xy \geq 2xy$ to terms in (30). This is a good relaxation since $x - y = 0$ in the tight case. This gives

$$\binom{p}{2}(x^2+y^2)(n-1) - p^2xy \geq xy(2\binom{p}{2}(n-1) - p^2) = xyp((n-1)(p-1)-p) = xyp(p(n-2)-(n-1)) \geq 0,$$

when $p \geq 1 + \frac{1}{n-2} = \frac{n-1}{n-2}$.

For (31), for the first term we use the $(x-y)^2$ bound again. (Notice $\binom{p}{3} \leq 0$ and $n^2 - 3n + 3 \geq 0$.) The second term is positive for $n \geq 2$ and we throw it out. For the third term, we apply $2n^2 - 6n + 5 < 2n^2 - 6n + 6$ and $x^3y^3 < x^2y^2$.

$$\begin{aligned}
\frac{\mathbf{E}[(X_iX_j)^p]}{(x_ix_j)^p} &\geq 1 - 2p\binom{p}{3}x^2y^2(n^2 - 3n + 3) - \binom{p}{3}^2 x^2y^2(2n^2 - 6n + 6) \\
&= 1 - 2p\binom{p}{3}x^2y^2(n^2 - 3n + 3)\left(1 + \frac{(p-1)(p-2)}{6}\right) \\
&\geq 1 - 2p\binom{p}{3}x^2y^2(n^2 - 3n + 3)\left(1 - \frac{1}{24}\right) \geq 1
\end{aligned}$$

□

## 2.5 Other 'good' distributions.

It is interesting that both ShortestEdge and SRDR have (as suggested experimentally) the same nice near-independence properties. Notice ShortestEdge is a distribution on $1/n$ points, and SRDR is a distribution on $O(1/n^2)$ points. In exploring other distributions experimentally, we find a curious pattern emerges.

**Conjecture 2.1**

Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be any set of orthonormal vectors, and let $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ be their respective orthogonal projections onto the $\mathbf{x} \cdot \mathbf{1} = 0$ plane. More explicitly, we have

$$\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}.$$

Then the uniform distribution on $\{\mathbf{x} \pm \delta\, \mathbf{w_1}, \mathbf{x} \pm \delta\, \mathbf{w_2}, \mathbf{x} \pm \delta\, \mathbf{w_3}\}$ has the same properties as SRDR from Theorems 2.2 and 2.1, for $\delta$ such that all points $\mathbf{x} \pm \delta\, \mathbf{w}_i$ remain in $[0,1]^3$.

Notice if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is $\{(1,0,0), (0,1,0), (0,0,1)\}$, then we get the shortest edge distribution. On the other hand if we choose $\left\{(\frac{1-\sqrt{3}}{3}, \frac{1}{3}, \frac{1+\sqrt{3}}{3}), (\frac{1}{3}, \frac{1+\sqrt{3}}{3}, \frac{1-\sqrt{3}}{3}), (\frac{1+\sqrt{3}}{3}, \frac{1-\sqrt{3}}{3}, \frac{1}{3})\right\}$, then we get SRDR. Furthermore, by for $v_1 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, we get $w_1 = (0,0,0)$. By carefully choosing $v_2$ and $v_3$ we can show that Conjecture 2.1 implies *any* pair of orthogonal vectors in the $\mathbf{x} \cdot \mathbf{1} = 0$ plane form an good distribution.

Unfortunately these conjectures do not hold true for nonuniform weights. They also does not hold true in the most obvious generalization to higher dimensions, although most distributions generated in this way do work for higher dimensions, so perhaps there is some slightly stricter condition which would make it remain true.

# 3 Finding the maximum-entropy distribution.

In the last section we explored several different distributions which all appear to have near-independence properties. In some contexts, the method of entropy-maximization oftens yields the 'best' distribution in some sense. In this section we will explore the technique of entropy maximization and see if it yields a distribution with near-independence properties.

**Definition 3.1: Entropy**

Given a distribution $D$ on some discrete set $S$, its (Shannon) entropy is defined as:

$$H(D) = \sum_{s \in S} \Pr[s] \ln(\Pr[s]),$$

where $(0 \ln 0)$ is taken to be 0, which is the value of $\lim_{x \to 0^+} x \ln x$. For brevity we define $h(x) := x \ln x$.

## 3.1 Example 1

As a warmup, consider the basic problem of finding a distribution $D$ which rounds $\mathbf{x} \in [0,1]^2$ to an integral solution while preserving the marginals. Here $D$ must range over $R = \{(0,0),(0,1),(1,0),(1,1)\}$, and may be defined by the set of probabilities $\{p_{00}, p_{01}, p_{10}, p_{11}\}$. The only constraints on $D$ are that the marginals must be preserved, and the values $p_i$ must be valid probabilities (between 0 and 1). Thus, we may write our entropy maximization problem as:

$$\begin{array}{ll} \text{maximize} & h(p_{00}) + h(p_{01}) + h(p_{10}) + h(p_{11}) \\ \text{subject to} & p_{10} + p_{11} = x_1 \\ & p_{01} + p_{11} = x_2 \\ & p_{00} + p_{01} + p_{10} + p_{11} = 1 \\ & 0 \le p_{00}, p_{01}, p_{10}, p_{11} \le 1 \end{array}$$

The solution to the above problem is: $\begin{cases} p_{00} &= (1-x_1)(1-x_2) \\ p_{01} &= (1-x_1)x_2 \\ p_{10} &= x_1(1-x_2) \\ p_{11} &= x_1 x_2 \end{cases}$. Notice this is exactly the distribution in which each variable is rounded independently! This is promising toward our goal of near-independence. Note that because of the linearity of entropy from independent sources, the maximum entropy is $h(x_1) + h(1-x_1) + h(x_2) + h(1-x_2)$.

## 3.2 Example 2

Now consider we change the problem to allow $D$ to range over all pairs $(x_1, x_2) \in [0,1]^n$ such that at least one of $x_1$ and $x_2$ are integral. In other words, instead of ranging just over the corners of the unit square, we are also ranging over the edges. As the range includes a continuous space, our definition of entropy is no longer useful; we could choose an arbitrarily large number of possible points for our distribution, and get an arbitrarilly large entropy.

Instead, suppose we define a new distribution $C(D)$ by first sampling $\mathbf{X}$ from $D$, and then sampling $\mathbf{Z}$ by setting $Z_i$ to 0 or 1 with probability $x_i$ or $1 - x_i$. Additionally, for simplicity of experiments, we will assume the distribution $D$ is over a finite set of discrete points. We now define the 'cube entropy' of $D$ to be $H(C(D))$:

**Definition 3.2: Cube Entropy**

Let distribution $D$ range over a finite, discrete set of points $R$, where $\mathbf{x} \in [0,1]^n$, $\forall\, \mathbf{x} \in R$. Define:

$$\begin{aligned} G(D) := H(C(D)) &= \sum_{\mathbf{z}\in\{0,1\}^n} h(\Pr[\mathbf{Z}=\mathbf{z}]) \\ &= \sum_{\mathbf{z}\in\{0,1\}^n} h\Big(\sum_{x\in R} \Pr[\mathbf{X}=\mathbf{x}]\Pr[\mathbf{Z}=\mathbf{z}\,|\,\mathbf{X}=\mathbf{x}]\Big) \\ &= \sum_{\mathbf{z}\in\{0,1\}^n} h\Big(\sum_{x\in R} \Pr[\mathbf{X}=\mathbf{x}] \prod_{i:\mathbf{z}_i=0}(1-\mathbf{x}_i)\prod_{i:\mathbf{z}_i=1}\mathbf{x}_i\Big) \end{aligned}$$

Notice, if $D$ is just a single vector with probability 1, $G(D)$ is exactly the entropy of independently rounding each element. Also notice if $D$ only ranges over $\{0,1\}^n$, then $G(D) = H(D)$.

We can now ask for a distribution $D$ over the edges of the unit square which maximizes $G(D)$. We find that there is no longer a single distribution obtaining the maximum value of $G(D)$. One such distribution is of course the one we already found in the previous problem. Another maximizing distribution is to choose $(1, x_2)$ or $(0, x_2)$ with probability $x_1$ or $1 - x_1$, respectively.

As it turns out, while preserving marginals, the independent rounding entropy is the maximum possible value for $G(D)$. As we randomly move $x$ within the polytope, we may decrease, but never increase, the total entropy. Namely, if $D$ changes only a single element, or is a linear combination of such distributions, or the composition of such distributions, then the entropy $G(D)$ will remain constant, and all variables remain independent.

## 3.3 Adding a linear constraint.

We now add the constraint $\mathbf{X} \cdot \mathbf{a} = \mathbf{x} \cdot \mathbf{a}$. Notice this limits the range of $D$ to the intersection of a hyperplane with the unit hypercube, and it is no longer possible to change only a single element at a time, so we do expect the entropy to decrease.

Since we do want at least one element to become integral, we will require that $D$ ranges over the 'edges' of the polytope $P(\mathbf{a}) := \{\mathbf{X} \in [0,1]^n \mid \mathbf{X} \cdot \mathbf{a} = \mathbf{x} \cdot \mathbf{a}\}$.

At this point we do not have a general theoretical solution to the maximization problem, and instead rely on experimentation. We make the simplifying assumption that $D$ ranges over at most 1 point for each edge of the polytope. In $n$ dimensions, there are up to $2n$ edges, corresponding to each of the $n$ coordinates being fixed to 0 or 1 (all configurations may not be possible depending on the weight vector $\mathbf{a}$). Because this assumption is not necessarily valid, we may not actually find the real distribution of maximum entropy.

We construct a maximization problem and attempt to solve it numerically, for $n = 3$. As it is nonlinear, we are not guaranteed optimality of our solutions. Here let $\mathbf{x}_0$ be the original vector, and $D$ be the distribution over points $\mathbf{x}_1, ..., \mathbf{x}_6$ (corresponding to the points on each of the 6 edges.) each chosen with probability $p_1, ..., p_6$, respectively.

$$
\begin{aligned}
\text{maximize } & G(D) \\
\text{subject to } & \mathbf{x}_1 \in \{\mathbf{X} \in P(\mathbf{a}) | X_1 = 0\} \\
& \mathbf{x}_2 \in \{\mathbf{X} \in P(\mathbf{a}) | X_2 = 0\} \\
& \mathbf{x}_3 \in \{\mathbf{X} \in P(\mathbf{a}) | X_3 = 0\} \\
& \mathbf{x}_4 \in \{\mathbf{X} \in P(\mathbf{a}) | X_1 = 1\} \\
& \mathbf{x}_5 \in \{\mathbf{X} \in P(\mathbf{a}) | X_2 = 1\} \\
& \mathbf{x}_6 \in \{\mathbf{X} \in P(\mathbf{a}) | X_3 = 1\} \\
& \sum_{i=1}^{6} p_i = 1 \\
& 0 \leq p_i \leq 1 \qquad\qquad \forall i, j
\end{aligned}
$$

In testing the above program for a large variety of values of $\mathbf{x}_0$, we found that for the unweighted case ($\mathbf{a} = \mathbf{1}$), the resulting distributions did appear to have the $1 - 1/n$ upper bound on near independence as in Corollary 2.1. However, there are simple examples for which the corresponding lower bound in Theorem 2.2 does not hold for any value of $p$.

For nonuniform weights, we encountered some examples for which solving the above program gave a distribution which violated the $1 - 1/n$ upper bound. However, by manually exploring nearby distributions, namely, by allowing for multiple points on an edge, we were always able to find a distribution with higher entropy for which the $1 - 1/n$ upper bound did hold.

# 4 Open Questions

The upper and lower bounds on near-independence given in Theorem 2.2 and Corollary 2.1 have a particularly nice clean form. We conjecture that they are the tightest bounds possible of this form. However, it appears there may be many distributions with the same property. It remains to characterize the family of distributions with this property and find the deeper principle which causes this property to hold.

The relationship between maximum-entropy distributions and near-independence holds some promise. A better theoretical or even numerical method of generating the entropy-maximizing distribution is necessary to continue this line of exploration. It's also possible that there may be better functions than $G(D)$ for measuring the entropy of partial roundings.

Perhaps the most applicable direction to explore would be attempting to generalize near-independence to multiple linear constraints. The distributions in Section 2 do not have an obvious generalization to multiple linear constraints. Entropy, on the other hand, should directly generalize and it would be very interesting to see if the maximum-entropy distributions have any near-independence properties for multiple linear constraints. We optimistically conjecture that the upper bound in Corollary 2.1 should hold for $p = 1 - \frac{k}{n}$, where $k$ is the number of constraints.

# References


[1] J. Byrka, T. Pensyl, B. Rybicki, A. Srinivasan, and K. Trinh. An improved approximation for $k$-median, and positive correlation in budgeted optimization. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 737–756, 2015.

[2] S. Li and O. Svensson. Approximating $k$-median via pseudo-approximation. In *STOC*, pages 901–910, 2013.

[3] A. Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *FOCS*, pages 588–597, 2001.