Randomly Rounding Rationals with Cardinality Constraints and Derandomizations

Benjamin Doerr, MPI Saarbrücken, Germany

STACS 2007, Aaaaaachen

Abstract: If you can round $\{0, \frac{1}{2}\}$ vectors nicely, then all rational vectors.

Motivation: Unbiased Rounding (Statistics)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.231	41.846	12.517	12.502	33.021	151.117
Saarlouis	55.839	53.774	9.014	12.789	30.142	161.558
St. Wendel	50.048	50.000	6.925	10.127	25.286	142.386
Homburg	54.083	45.447	9.033	11.770	28.640	148.973
	211.201	191.067	37.489	47.188	117.089	604.034

Bundestagswahl 2005: Votes in the Saarland.

Problem: Round to multiples of 1.000

Why? – Increase readability
– Confidentiality protection

Motivation: Unbiased Rounding (Statistics)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.231	41.846	12.517	12.502	33.021	151.117
Saarlouis	55.839	53.774	9.014	12.789	30.142	161.558
St. Wendel	50.048	50.000	6.925	10.127	25.286	142.386
Homburg	54.083	45.447	9.033	11.770	28.640	148.973
	211.201	191.067	37.489	47.188	117.089	604.034

Bundestagswahl 2005: Votes in the Saarland.

Problem: Round to multiples of 1.000

How? – Unbiased: $\Pr(12.502 \to 13.000) = 0.502,$

$$\Pr(12.502 \to 12.000) = 0.498.$$

In CS: "Randomized rounding"

Motivation: Unbiased Rounding (Statistics)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.231	41.846	12.517	12.502	33.021	151.117
Saarlouis	55.839	53.774	9.014	12.789	30.142	161.558
St. Wendel	50.048	50.000	6.925	10.127	25.286	142.386
Homburg	54.083	45.447	9.033	11.770	28.640	148.973
	211.201	191.067	37.489	47.188	117.089	604.034

Bundestagswahl 2005: Votes in the Saarland.

Problem: Round to multiples of 1.000

How? – Unbiased/randomized rounding
– Preserve additivities (row/column sums, grand total)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.000	42.000	13.000	13.000	33.000	151.000
Saarlouis	56.000	54.000	9.000	13.000	30.000	162.000
St. Wendel	50.000	50.000	7.000	11.000	25.000	142.000
Homburg	54.000	45.000	9.000	12.000	29.000	149.000
	211.000	191.000	37.000	47.000	117.000	604.000

Motivation: Unbiased Rounding (Statistics)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.231	41.846	12.517	12.502	33.021	151.117
Saarlouis	55.839	53.774	9.014	12.789	30.142	161.558
St. Wendel	50.048	50.000	6.925	10.127	25.286	142.386
Homburg	54.083	45.447	9.033	11.770	28.640	148.973
	211.201	191.067	37.489	47.188	117.089	604.034

Bundestagswahl 2005: Votes in the Saarland.

Problem: Round to multiples of 1.000

How? – Unbiased/randomized rounding

– Hard constraints: Preserve additivities

– As independent as possible

$\Rightarrow$ low errors if entries are aggregated.

Motivation: Unbiased Rounding (Statistics)

	SPD	CDU	Grüne	FDP	Linke	
Saarbrücken	51.231	41.846	12.517	12.502	33.021	151.117
Saarlouis	55.839	53.774	9.014	12.789	30.142	161.558
St. Wendel	50.048	50.000	6.925	10.127	25.286	142.386
Homburg	54.083	45.447	9.033	11.770	28.640	148.973
	211.201	191.067	37.489	47.188	117.089	604.034

Bundestagswahl 2005: Votes in the Saarland.

Problem: Round to multiples of 1.000

How? – Unbiased/randomized rounding
– Hard constraints: Preserve additivities
– As independent as possible
$\Rightarrow$ low errors if entries are aggregated.

Corollary of main result: This can be done in time $O(mn)$.

Randomly Rounding Rationals with Cardinality Constraints and Derandomizations

1. Precise setting
   - Randomized rounding
   - Hard constraints
   - Negative correlation and large deviations
2. Previous work and this result
3. A proof ingredient

Randomized Rounding (RR)

- Let $x \in [0,1]$.[1]

- A random variable y is a *randomized rounding of* x, write $y \approx x$, if

$$\Pr(y = 1) = x,$$
$$\Pr(y = 0) = 1 - x.$$

- Note: $E(y) = x$. [In statistics: "Unbiased rounding"]

- $y = (y_1, \ldots, y_n)$ is a randomized rounding of $x \in [0,1]^n$ if $y_i \approx x_i$ for all i.[2]

[1] Restriction to $[0,1]$ for presentation purposes only. In general: y is a RR of x, if $y - \lfloor x \rfloor$ is a RR of $x - \lfloor x \rfloor$ as defined above.

[2] Traditionally "stochastically independent". But not in this talk.

Hard Constraints: Cardinality Constraints

- Cardinality Constraint: $\sum_{i \in I} x_i = k \quad (k \in \mathbb{N})$
  Exactly k of the x_i, $i \in I$, shall become one.

- In rounding: $\sum_{i \in I} y_i = \sum_{i \in I} x_i$.

- General form: $\sum_{i \in I} y_i \approx \sum_{i \in I} x_i$.

Hard Constraints: Cardinality Constraints

- Cardinality Constraint: $\sum_{i \in I} x_i = k \quad (k \in \mathbb{N})$
  Exactly k of the x_i, $i \in I$, shall become one.

- In rounding: $\sum_{i \in I} y_i = \sum_{i \in I} x_i$.

- General form: $\sum_{i \in I} y_i \approx \sum_{i \in I} x_i$.

- Do not go well with independent RR:
  - $\Pr(\sum_{i \in I} y_i = \sum_{i \in I} x_i)$ can be as small as $|I|^{-1/2}$.
  - "Repeat independent RR until constraint satisfied" does not work.
    Counter-example: $x = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $I = \{1, 2, 3\}$.

Negative Correlation

Def: A set of binary random variables $\{y_1, \ldots, y_n\}$ is negativly correlated if for all $b \in \{0,1\}$,

$$\Pr(\forall i \in [n] : y_i = b) \leq \prod_{i=1}^{n} \Pr(y_i = b).$$

Note: Independence is a special case of negative correlation.

Lemma:[1] Let $\{y_i | i \in S\}$ be negativly correlated for all $S \subseteq [n]$. Then the usual Chernoff-Hoeffding bounds hold. E.g., for all $c \in [0,1]^n$ and $\lambda > 0$,

$$\Pr(|c^\top y - E(c^\top y)| \geq \lambda) \leq \exp(-2\lambda^2/n).$$

[1] Panconesi, Srinivasan (1997)

RR with One Constraint

Problem: Given $x \in [0,1]^n$, generate a random $y \in \{0,1\}^n$ such that

(1) y is a randomized rounding of x.

(2) $\sum_{i=1}^n y_i$ is a randomized rounding of $\sum_{i=1}^n x_i$.

(3) For all $S \subseteq [n]$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

RR with One Constraint

Problem: Given $x \in [0,1]^n$, generate a random $y \in \{0,1\}^n$ such that

(1) y is a randomized rounding of x.

(2) $\sum_{i=1}^n y_i$ is a randomized rounding of $\sum_{i=1}^n x_i$.

(3) For all $S \subseteq [n]$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

Simple cases:

(a) $\sum_{i=1}^n x_i = 1$: Choose $i \in [n]$ with prob. x_i, set $y_i = 1$ and $y_j = 0$, $j \neq i$.

(b) $x_i = c/n$ for all i: Choose random $I \subseteq [n]$, $|I| = c$, and set $y_i = 1$ iff $i \in I$.

RR with One Constraint

Problem: Given $x \in [0,1]^n$, generate a random $y \in \{0,1\}^n$ such that

(1) y is a randomized rounding of x.

(2) $\sum_{i=1}^n y_i$ is a randomized rounding of $\sum_{i=1}^n x_i$.

(3) For all $S \subseteq [n]$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

Simple cases:

(a) $\sum_{i=1}^n x_i = 1$: Choose $i \in [n]$ with prob. x_i, set $y_i = 1$ and $y_j = 0$, $j \neq i$.

(b) $x_i = c/n$ for all i: Choose random $I \subseteq [n]$, $|I| = c$, and set $y_i = 1$ iff $i \in I$.

Srinivasan (FOCS 2001): Linear time solution (RAM model) for arbitrary x. No derandomization. Builds on simple case (a).

RR with One Constraint

Problem: Given $x \in [0,1]^n$, generate a random $y \in \{0,1\}^n$ such that

(1) y is a randomized rounding of x.

(2) $\sum_{i=1}^n y_i$ is a randomized rounding of $\sum_{i=1}^n x_i$.

(3) For all $S \subseteq [n]$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \le \prod_{j \in S} \Pr(y_j = b)$.

Simple cases:

(a) $\sum_{i=1}^n x_i = 1$: Choose $i \in [n]$ with prob. x_i, set $y_i = 1$ and $y_j = 0$, $j \neq i$.

(b) $x_i = c/n$ for all i: Choose random $I \subseteq [n]$, $|I| = c$, and set $y_i = 1$ iff $i \in I$.

Srinivasan (FOCS 2001): Linear time solution (RAM model) for arbitrary x. No derandomization. Builds on simple case (a).

D. (STACS 2005): Linear time solution (bit model) for x having finite binary expansion. Can be derandomized. Builds on simple case (b).

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in [0,1]^n$ having finite binary expansion if you can for all $x \in \{0, \frac{1}{2}\}^n$. [Constructive: Solve expansion length times a half-integral problem.]

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in [0,1]^n$ having finite binary expansion if you can for all $x \in \{0, \frac{1}{2}\}^n$. [Constructive: Solve expansion length times a half-integral problem.]

Advantage: The $\{0, \frac{1}{2}\}$ case typically is simpler and faster. E.g., unbiased rounding:
– $O(mn)$ for $\{0, \frac{1}{2}\}$–matrices,
– $O(mn \min\{m, n\})$ for arbitrary matrices.

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in [0,1]^n$ having finite binary expansion if you can for all $x \in \{0, \frac{1}{2}\}^n$. [Constructive: Solve expansion length times a half-integral problem.]

D. (STACS 2007): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in (\mathbb{Q} \cap [0,1])^n$ if you can for all $x \in \{0, \frac{1}{2}\}^n$.

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in [0,1]^n$ having finite binary expansion if you can for all $x \in \{0, \frac{1}{2}\}^n$. [Constructive: Solve expansion length times a half-integral problem.]

D. (STACS 2007): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in (\mathbb{Q} \cap [0,1])^n$ if you can for all $x \in \{0, \frac{1}{2}\}^n$.
[Constructive: If $px \in \mathbb{Z}^n$, $p \in \mathbb{N}$, then solve $O(p^2)$ times a half-integral problem.]

More Constraints

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in [0,1]^n$ having finite binary expansion if you can for all $x \in \{0, \frac{1}{2}\}^n$. [Constructive: Solve expansion length times a half-integral problem.]

D. (STACS 2007): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in (\mathbb{Q} \cap [0,1])^n$ if you can for all $x \in \{0, \frac{1}{2}\}^n$.
[Constructive: If $px \in \mathbb{Z}^n$, $p \in \mathbb{N}$, then solve $O(p^2)$ times a half-integral problem.]
[If $p = p_1 \ldots p_k$, then p^2 becomes $\sum_i p_i^2$. Good e.g. for p a power of 10.]

A Proof Ingredient from Discrepancy Theory

Let $\mathcal{H} = (V, \mathcal{E})$ be a unimodular[1] hypergraph. Then the following two assertions are equivalent.

(1) There is a $w : V \to \{0, 1/p, \ldots, (p-1)/p\}$ such that

- for all $E \in \mathcal{E}$, $w(E) := \sum_{v \in E} w(v) \in \mathbb{Z}$;
- w is not constant zero.

(2) $\mathcal{H}$ has a perfectly balanced non-trivial partial coloring, that is, there is a $\chi : V \to \{\text{red}, \text{blue}, \text{uncolored}\}$ such that

- each $E \in \mathcal{E}$ has the same number of red and blue vertices;
- χ is not constant uncolored.

---

[1] Its incidence matrix is totally unimodular[2].

[2] Each square submatrix has determinant -1, 0 or 1.

Summary

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): ... for $x \in [0,1]^n$ having finite binary expansion ...

D. (STACS 2007): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in (\mathbb{Q} \cap [0,1])^n$ if you can for all $x \in \{0, \frac{1}{2}\}^n$.
[Constructive: If $px \in \mathbb{Z}^n$, $p \in \mathbb{N}$, then solve $O(p^2)$ times a half-integral problem.]
[If $p = p_1 \ldots p_k$, then p^2 becomes $\sum_i p_i^2$. Good e.g. for p a power of 10.]

Summary

Let $C \in \{0,1\}^{m \times n}$ and $\mathcal{S} \subseteq 2^{[n]}$ closed under taking subsets.

Definition: y is an *independent-looking* $(C, \mathcal{S})$*-RR* of $x \in [0,1]^n$ if

(1) y is a randomized rounding of x.

(2) Cy is a randomized rounding of Cx.

(3) For all $S \in \mathcal{S}$, $b \in \{0,1\} : \Pr(\forall j \in S : y_j = b) \leq \prod_{j \in S} \Pr(y_j = b)$.

D. (STACS 2006): ... for $x \in [0,1]^n$ having finite binary expansion ...

D. (STACS 2007): You can generate and derandomize independent-looking $(C, \mathcal{S})$-RRs for all $x \in (\mathbb{Q} \cap [0,1])^n$ if you can for all $x \in \{0, \frac{1}{2}\}^n$.
[Constructive: If $px \in \mathbb{Z}^n$, $p \in \mathbb{N}$, then solve $O(p^2)$ times a half-integral problem.]
[If $p = p_1 \ldots p_k$, then p^2 becomes $\sum_i p_i^2$. Good e.g. for p a power of 10.]

Danke!