Electron. Commun. Probab. **28** (2023), article no. 32, 1–6.
https://doi.org/10.1214/23-ECP537
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Generating discrete uniform distribution from a biased coin using number-theoretic method

Xiaoyu Lei*

Abstract

In this article, a new algorithm generating discrete uniform distribution on n elements from a binary random source is proposed by introducing the notion of coprimity and modulo function in number theory. The algorithm is a generalization of Von Neumann's algorithm for $n = 2$ [4] and Dijkstra's algorithm for prime n [1]. The proof for the validity of the algorithm introduces the notion of random variable in residual classes, and combines the ideas and techniques from both probability theory and number theory.

Keywords: number theory: probability theory; random number.
MSC2020 subject classifications: 11Z05; 60B99.
Submitted to ECP on March 7, 2023, final version accepted on July 17, 2023.

1 Background

Sampling a target distribution from a random physical source has many applications. Sometimes the random source is biased with unknown distribution, while we need a specific target distribution. Therefore, an efficient algorithm generating a specific distribution from a random source with unknown bias is of significant value in many aspects.

A simple method generating a fair binary distribution from a biased binary random source was first proposed in [4]. Von Neumann's algorithm was generalized to prime number in [1]. For a prime number p, Dijkstra's algorithm takes p flips as an input and generates a discrete uniform distribution on p elements by using the lexicographic ranking of the input among the orbit determined by rotating the input p times. Inspired by the modulo function in number theory, another method was proposed in [3] to generate a discrete uniform distribution on p elements using the rank sum. The usage of rank sum makes the method arguably easier than Dijkstra's algorithm. Because rotating the input and determining the lexicographic ranking among the orbit are replaced by a simple formula "rank sum $\mod p$" to return a number.

Based on the idea in [3], a generalization of Dijkstra's algorithm is proposed in this article to generate a discrete uniform distribution on n elements for any integer n. And a proof for the validity of the algorithm which combines probabilistic and algebraic techniques is provided in the article.

The article is organized as follows: In Section 2, the motivation of the work and the algorithm are heuristically explained and constructed. In Section 3, the algorithm is

*The University of Chicago, United States of America. E-mail: leixy@uchicago.edu

formally stated. The validity of the algorithm is proved by a novel method which needs knowledge from both probability theory and number theory.

2 Motivation of the work

Let $X \in \{H, T\}$ be an outcome of a biased coin flip with probability $a = \mathbb{P}(X = H)$ of getting a head and $b = \mathbb{P}(X = T) = 1 - a$ of getting a tail. Let $\{X_i : i \geq 0\}$ be i.i.d. copies of X representing independent flip outcomes of the same biased coin. For integer n, let random vector

$$\boldsymbol{X}^n = (X_0, \ldots, X_{n-1}) \in \{H, T\}^n$$

denote the outcome of n independent flips.

In [4], a simple algorithm was proposed to generate a fair binary distribution from a biased coin by noticing the fact $\mathbb{P}\left(\boldsymbol{X}^2 = (H, T)\right) = \mathbb{P}\left(\boldsymbol{X}^2 = (T, H)\right) = ab$. The algorithm consists of three steps: 1. flip the coin twice; 2. if the outcome is two heads or two tails, discard the result and return to step 1; 3. if the outcome is (H, T), return 0; if the outcome is (T, H), return 1. To generalize Von Neumann's algorithm to arbitrary integer n, define the event

$$A_i = \{\text{only one head appears in the } i\text{-th position in the flip sequence}\}\, (0 \leq i \leq n-1),$$

where we assume the position begins from 0 in the article. The generalization of Von Neumann's algorithm can be formally stated in the algorithm $\mathcal{A}_1$ by noticing the fact

$$\mathbb{P}(\text{the event } A_i \text{ happens}) = \mathbb{P}(\boldsymbol{X}^n \in A_i) = ab^{n-1} (0 \leq i \leq n-1).$$

The motivation of our work is to improve the algorithm $\mathcal{A}_1$ by making use of more flip outcomes to return a number. With a bit more thought, we can define the event

$$B_i = \left\{ \begin{matrix} \text{only one head appears in the } i\text{-th position in the flip sequence, or} \\ \text{only one tail appears in the } i\text{-th position in the flip sequence} \end{matrix} \right\} (0 \leq i \leq n-1).$$

By the fact that

$$\mathbb{P}(\text{the event } B_i \text{ happens}) = \mathbb{P}(\boldsymbol{X}^n \in B_i) = ab^{n-1} + a^{n-1}b (0 \leq i \leq n-1),$$

the algorithm $\mathcal{A}_1$ also returns a uniform distribution on the set $\{0, \ldots, n-1\}$ if A_i is replaced by B_i in the statement of the algorithm, by which more flip outcomes are used to return a number.

How can we make use of more flip outcomes to return a number? Before going beyond the above, more notations need to be introduced to make the explanation below clear. For a specific sequence of n flips $\boldsymbol{x}^n = (x_0, \ldots, x_{n-1}) \in \{H, T\}^n$, let

$$N(\boldsymbol{x}^n) = \sum_{i=0}^{n-1} 1_{\{x_i = H\}}$$

Algorithm 1 $\mathcal{A}_1$:

Input: A sequence of flips from a biased coin X

Output: Integer in $\{0, \ldots, n-1\}$

1: Flip the coin n times

2: If none of the events $\{A_i : 0 \leq i \leq n-1\}$ happens, then discard the outcome and return to step 1

3: Else if the event A_i happens, **return** i

denote the head count of $\boldsymbol{x}^n$. For $0 \le k \le n$, define

$$S_k = \{A \subset \{0, \dots, n-1\} : |A| = k\}$$

to be the set of all subsets of $\{0, \dots, n-1\}$ containing k elements, where $|A|$ means the cardinality of set A. Then there exists a bijection between set S_k and the set consisting of all flip sequences with k heads $\{\boldsymbol{x}^n \in \{H,T\}^n : N(\boldsymbol{x}^n) = k\}$ by the position of heads. For $\{i_1, \dots, i_k\} \in S_k$, it corresponds to a flip sequence with k heads where heads appear only in the i_j-th flip for $1 \le j \le k$. In the following derivation, we don't distinguish the elements in S_k and the flip sequences with k heads. Notice for $\boldsymbol{x}^n \in S_k$, the probability of getting $\boldsymbol{x}^n$ in n flips

$$\mathbb{P}(\boldsymbol{X}^n = \boldsymbol{x}^n) = \prod_{i=0}^{n-1} \mathbb{P}(X_i = x_i) = a^{N(\boldsymbol{x}^n)} b^{n-N(\boldsymbol{x}^n)} = a^k b^{n-k},$$

which only relies on the head count k.

Let (k,n) be the greatest common divisor of integers k and n. Let $\mathcal{K}$ consist of all integers which are less than and coprime to n,

$$\mathcal{K} = \{0 \le k \le n : (k,n) = 1\}.$$

The reason to consider coprimity here is due to Lemma A.1, which indicates if $(k,n) = 1$, the set S_k can be partitioned into n subsets of equal size. If for all $k \in \mathcal{K}$, each subset of flip sequences in the equal size partition of S_k is assigned a target value by an algorithm, the algorithm can return a uniform distribution on n elements. The idea can be formally stated in the following way.

Let sets $\{C_i : 0 \le i \le n-1\}$ be a partition of the disjoint union $\bigsqcup_{k\in\mathcal{K}} S_k$ satisfying that each S_k is partitioned into n subsets of equal size by $\{C_i \cap S_k : 0 \le i \le n-1\}$, which is formally formulated as

$$\bigsqcup_{k\in\mathcal{K}} S_k = \bigsqcup_{i=0}^{n-1} C_i, \text{ and } |C_0 \cap S_k| = \cdots = |C_{n-1} \cap S_k| = \frac{1}{n}|S_k| \quad (k \in \mathcal{K}). \tag{2.1}$$

Then the probability

$$\mathbb{P}(\boldsymbol{X}^n \in C_i) = \sum_{k\in\mathcal{K}} \mathbb{P}(\boldsymbol{X}^n \in C_i \cap S_k) = \sum_{k\in\mathcal{K}} |C_i \cap S_k| a^k b^{n-k} = \frac{1}{n} \sum_{k\in\mathcal{K}} |S_k| a^k b^{n-k}, \tag{2.2}$$

where the last equality is by the second part in (2.1). Notice the right side of (2.2) is independent from the choice of C_i, based on which the algorithm $\mathcal{A}_2$ below returns a uniform distribution on the set $\{0, \dots, n-1\}$.

What remains is to find the sets $\{C_i : 0 \le i \le n-1\}$ satisfying (2.1). By introducing the modulo function in number theory, there exists a simple way of designing

Algorithm 2 $\mathcal{A}_2$:

Input: A sequence of flips from a biased coin X
Output: Integer in $\{0, \dots, n-1\}$

1: Flip the coin n times
2: If $\boldsymbol{X}^n$ belongs to none of $\{C_i : 0 \le i \le n-1\}$, then discard the outcome and return to step 1
3: Else if $\boldsymbol{X}^n \in C_i$, **return** i

$\{C_i : 0 \le i \le n-1\}$ satisfying (2.1). For a flip sequence $\boldsymbol{x}^n \in \{H,T\}^n$, let

$$R(\boldsymbol{x}^n) = \sum_{i=0}^{n-1} i \cdot 1_{\{x_i = H\}}$$

be the rank sum of heads, which is the sum of positions of all heads in $\boldsymbol{x}^n$ beginning from 0. By the bijection between the set S_k and the set consisting of all flip sequences with k heads $\{\boldsymbol{x}^n \in \{H,T\}^n : N(\boldsymbol{x}^n) = k\}$, we can choose

$$C_i = \{\boldsymbol{x}^n : R(\boldsymbol{x}^n) = i \pmod n, N(\boldsymbol{x}^n) \in \mathcal{K}\},$$

which is shown in the next section.

3 Generating discrete uniform distribution

Based on the discussion above, the algorithm we propose can be formally stated as the algorithm $\mathcal{A}_3$ below.

The theorem below guaranteeing the validity of the algorithm $\mathcal{A}_3$ can be proved by a novel method involving both probabilistic and algebraic ideas.

Theorem 3.1. *Let X be a biased coin with probability a getting a head and probability b getting a tail. Then the algorithm $\mathcal{A}_3$ returns a discrete uniform distribution on the set $\{0,\dots,n-1\}$.*

Proof. Let $\{\boldsymbol{X}_i^n = (X_{in},\dots,X_{in+n-1}) : i \ge 0\}$ be i.i.d. outcomes of n flips and τ be the first time $(N(\boldsymbol{X}_i^n), n) = 1$. For $0 \le m \le n-1$, the probability the algorithm $\mathcal{A}_3$ returns m

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_3 = m) &= \mathbb{P}(R(\boldsymbol{X}_\tau^n) = m \pmod n) \\
&= \frac{\mathbb{P}(R(\boldsymbol{X}^n) = m \pmod n, (N(\boldsymbol{X}^n), n) = 1)}{\mathbb{P}((N(\boldsymbol{X}^n), n) = 1)} \\
&= \frac{\mathbb{P}(R(\boldsymbol{X}^n) = m \pmod n, N(\boldsymbol{X}^n) \in \mathcal{K})}{\mathbb{P}(N(\boldsymbol{X}^n) \in \mathcal{K})}.
\end{aligned}$$

What is left is to show

$$\begin{aligned}
&\mathbb{P}(R(\boldsymbol{X}^n) = m_1 \pmod n, N(\boldsymbol{X}^n) \in \mathcal{K}) \\
&= \mathbb{P}(R(\boldsymbol{X}^n) = m_2 \pmod n, N(\boldsymbol{X}^n) \in \mathcal{K}) \, (0 \le m_1, m_2 \le n-1),
\end{aligned} \tag{3.1}$$

by which $\mathbb{P}(\mathcal{A}_3 = m) = 1/n \quad (0 \le m \le n-1)$.

To show (3.1), consider the random variables taking value in the residual classes of integers modulo n, $\mathbb{Z}_n = \{\overline{0},\dots,\overline{n-1}\}$, where $\overline{i}$ represents the residual class of i modulo n. Regard $\overline{0}$ as tail and $\overline{1}$ as head, then the flip of a biased coin can be viewed as a random variable $Y \in \{\overline{0},\overline{1}\}$ with probability $\mathbb{P}(Y = \overline{1}) = a$ and probability $\mathbb{P}(Y = \overline{0}) = b$. Let random vector $(Y_0,\dots,Y_{n-1})$ represent the outcome of n independent flips, where Y_i's are i.i.d. copies of Y. Then there exist the following equivalences,

$$R(\boldsymbol{X}^n) = m \pmod n \Leftrightarrow \sum_{i=0}^{n-1} \overline{i} \cdot Y_i = \overline{m} \quad (0 \le m \le n-1), \tag{3.2}$$

Algorithm 3 $\mathcal{A}_3$:

Input: A sequence of flips from a biased coin X
Output: Integer in $\{0,\dots,n-1\}$

1: Flip the coin n times
2: If $(N(\boldsymbol{X}^n), n) > 1$, then discard the outcome and return to step 1
3: Else **return** $R(\boldsymbol{X}^n) \bmod n$

and

$$N\left(\boldsymbol{X}^n\right)=k \Leftrightarrow \sum_{i=0}^{n-1} Y_i=\overline{k} \quad (1\leq k\leq n-1),$$

by which we have the equivalence below,

$$N\left(\boldsymbol{X}^n\right)\in \mathcal{K} \Leftrightarrow \sum_{i=0}^{n-1} Y_i\in \overline{\mathcal{K}}, \tag{3.3}$$

where $\overline{\mathcal{K}}$ consists of the residual classes of all integers in $\mathcal{K}$, $\overline{\mathcal{K}}=\left\{\overline{k}: k\in\mathcal{K}\right\}$. Based on (3.2) and (3.3), equality (3.1) can be reformulated as

$$\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_1},\sum_{i=0}^{n-1} Y_i\in\overline{\mathcal{K}}\right)=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_2},\sum_{i=0}^{n-1} Y_i\in\overline{\mathcal{K}}\right)(0\leq m_1,m_2\leq n-1). \tag{3.4}$$

To prove (3.4), fix an element $\overline{k}\in\overline{\mathcal{K}}$, where $(k,n)=1$. Consider the probability

$$\begin{aligned}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{0},\sum_{i=0}^{n-1} Y_i=\overline{k}\right)&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i+\sum_{i=0}^{n-1} Y_i=\overline{k},\sum_{i=0}^{n-1} Y_i=\overline{k}\right)\\&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i+1}\cdot Y_i=\overline{k},\sum_{i=0}^{n-1} Y_i=\overline{k}\right).\end{aligned} \tag{3.5}$$

Let σ be the permutation on the set $\{0,\ldots,n-1\}$ which maps each number to its next except $n-1$, which is mapped to 0,

$$\sigma=\begin{pmatrix}0&1&\cdots&n-2&n-1\\1&2&\cdots&n-1&0\end{pmatrix}.$$

Because all Y_i's are i.i.d., the permutation doesn't change the distribution of $(Y_0,\ldots,Y_{n-1})$

$$(Y_0,\ldots,Y_{n-1})\overset{d}{=}(Y_{\sigma(0)},\ldots,Y_{\sigma(n-1)}). \tag{3.6}$$

Based on (3.5) and (3.6), the probability

$$\begin{aligned}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{0},\sum_{i=0}^{n-1} Y_i=\overline{k}\right)&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i+1}\cdot Y_i=\overline{k},\sum_{i=0}^{n-1} Y_i=\overline{k}\right)\\&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i+1}\cdot Y_{\sigma(i)}=\overline{k},\sum_{i=0}^{n-1} Y_{\sigma(i)}=\overline{k}\right).\end{aligned} \tag{3.7}$$

Since the permutation σ is a bijection on the set $\{0,\ldots,n-1\}$, it's obvious that

$$\sum_{i=0}^{n-1} Y_{\sigma(i)}=\sum_{i=0}^{n-1} Y_i. \tag{3.8}$$

Consider the term $\sum_{i=0}^{n-1}\overline{i+1}\cdot Y_{\sigma(i)}$,

$$\begin{aligned}\sum_{i=0}^{n-1}\overline{i+1}\cdot Y_{\sigma(i)}&=\sum_{i=0}^{n-2}\overline{i+1}\cdot Y_{i+1}+\overline{n}\cdot Y_{\sigma(n-1)}\\&=\sum_{i=1}^{n-1}\overline{i}\cdot Y_i+\overline{0}\cdot Y_0\\&=\sum_{i=0}^{n-1}\overline{i}\cdot Y_i.\end{aligned} \tag{3.9}$$

Combining (3.7), (3.8) and (3.9), we have

$$\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{0},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{k},\sum_{i=0}^{n-1}Y_i=\overline{k}\right).$$

By iterating the derivation above, we have

$$\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{0},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=s\cdot\overline{k},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)(s\geq 0). \tag{3.10}$$

According to Lemma A.2 in Appendix A, each $\overline{k}\in\overline{\mathcal{K}}$ generates $\mathbb{Z}_n$, therefore

$$\left\{s\cdot\overline{k}:s\geq 0\right\}=\mathbb{Z}_n. \tag{3.11}$$

By (3.10) and (3.11), we have the equality

$$\begin{aligned}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_1},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_2},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)\\(0\leq m_1,m_2\leq n-1,\overline{k}\in\overline{\mathcal{K}}).&\end{aligned} \tag{3.12}$$

Then fix $\overline{m_1}$ and $\overline{m_2}$, sum both sides of (3.12) over $\overline{k}\in\overline{\mathcal{K}}$, we can show (3.4),

$$\begin{aligned}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_1},\sum_{i=0}^{n-1}Y_i\in\overline{\mathcal{K}}\right)&=\sum_{\overline{k}\in\overline{\mathcal{K}}}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_1},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)\\&=\sum_{\overline{k}\in\overline{\mathcal{K}}}\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_2},\sum_{i=0}^{n-1}Y_i=\overline{k}\right)\\&=\mathbb{P}\left(\sum_{i=0}^{n-1}\overline{i}\cdot Y_i=\overline{m_2},\sum_{i=0}^{n-1}Y_i\in\overline{\mathcal{K}}\right)(0\leq m_1,m_2\leq n-1).\end{aligned}$$

□

A Lemmas in number theory

Lemma A.1. *For $0\leq k\leq n$, if $(k,n)=1$, then $n\big|\binom{n}{k}$.*

Proof. It's easy to see $n\big|k\binom{n}{k}$ by the combinatorial equality $k\binom{n}{k}=n\binom{n-1}{k-1}$. Since $(k,n)=1$, we have $n\big|\binom{n}{k}$. □

Lemma A.2. *Let $\mathbb{Z}_n$ denote the residual classes modulo n. The residual class of integer k modulo n generates the abelian group $\mathbb{Z}_n$ if and only if k and n are coprime:*

$$\langle\overline{k}\rangle=\mathbb{Z}_n\Leftrightarrow(k,n)=1.$$

Proof. See 6.16 Corollary in page 65 in [2] for proof. □

References

[1] Dijkstra, E. W.: Making a fair roulette from a possibly biased coin. *Information Processing Letters* **36**(4) (1990), 193.

[2] Fraleigh, J. B.: A First Course in Abstract Algebra: Pearson New International Edition. *Pearson Custom Library*, Pearson Education, 2013. MR0225619

[3] Lei, X.: An efficient method generating discrete uniform distribution from a biased coin, arXiv:2205.01664

[4] Von Neumann, J.: Various techniques used in connection with random digits. *Applied Math Series* **12** (1951), 36–38.