SAMPLING THE DISK WITHOUT REJECTION

RENÉ RÜHR

Abstract. We study a novel disk-sampling algorithm. Similar to the rejection method, it uses a square to produce samples. Here the inscribed square is used. Symmetries are used to cover disk segments outside the square.

1. Introduction

Sampling inside the circle is one of the oldest sampling schemes, with variants discussed as early as [Neu51]. The two most common strategies are the polar method and the rejection method where the circumscribed square of the disk is sampled until a point falls inside the disk. Computationally, the rejection method suffers from branch divergence, while polar coordinates require the calculation of trigonometric functions and a square root.

The disk is useful for sampling other distributions. Box-Muller use the disk for sampling from the normal distribution [BM58]. Higher-dimensional spheres can also be bootstrapped from the disk [Mar72] resp. from the normal distribution [Mul59]. Specific domains like computer graphics have a particular interest in the disk driven by current advances in ray tracing [PJH23].

We present a Markov chain whose stationary measure is the uniform probability on the disk. Unlike rejection sampling, our method first samples from the inscribed square. We utilize symmetries to cover the remaining segments of the disk outside the square. The approach uses a fixed number of conditional statements. This makes it more amenable to vectorization than rejection sampling in environments where trigonometric functions are not hardware-accelerated.

2. New algorithm

We now present details to the algorithm. We shall sample from the disk of squared radius 2. We choose this particular radius since most constants from now on are integers.

Consider the square $S = [-1,1] \times [-1,1]$ inscribed in the disk of radius $\sqrt{2}$, $D = \mathrm{Disk}(0,\sqrt{2})$. Consider further the four disks $D_i = D + v_i$ which are translates of D where $v_1 = (2,0)$, $v_2 = (0,2)$, $v_3 = (-2,0)$ and $v_4 = (0,-2)$. Note that $D_i \cap D$ is a disk segment contained in S. Then $E_i = (D_i \cap D) - v_i$ is one of the disk segments of $D \setminus S$. In particular, $D = S \sqcup \bigsqcup_{i=1}^{4} E_i$. For every point $p \in D_i \cap D$ we let $p' = p - v_i \in E_i$, the *adopted variate* of p. Having the notation in place, the algorithm can be described as such:

- Let p be a uniform sample on S.
- If $p \in D_i$, cache $p' \in E_i$ as immediate output for the next functional call.
- Return p.

We give a pseudo-code implementation in Algorithm 1. The check $p \in D_i$ reads $\|p - v_i\|^2 < 2$, which has common expressions for all four cases. The conditional statements can be turned into arithmetic operations suitable for SIMD instruction sets, see [Rüh24b].

We call this algorithm adoption method, since we use the inner segments to find the adopted samples in E_i of S.

These adopted variates are not unlike antithetic variates where symmetry also plays a key role. Instead of using translations, one can easily envision to deploy other symmetries.

Algorithm 1 Adoption Algorithm

```
Output: Sample (x, y) in Disk(0, √2)
Static variables: Initialize x' ← 0, y' ← 0
procedure ADOPTION
    if x' ≠ 0 or y' ≠ 0 then
        x ← x', y ← y'
        x' ← 0, y' ← 0
        return (x, y)
    end if
    x ← random(−1, 1)
    y ← random(−1, 1)
    s ← x² + y² + 2
    if      s <  4x then x' ← x − 2, y' ← y
    else if s <  4y then x' ← x,     y' ← y − 2
    else if s < −4x then x' ← x + 2, y' ← y
    else if s < −4y then x' ← x,     y' ← y + 2
    end if
    return (x, y)
end procedure
```

3. Markov Chain

The algorithm can be understood as an ergodic Markov chain whose stationary measure is m, the uniform probability measure on D. We will deduce that a sequence of samples $\{x_i\}$ equidistributes.

Let C be the complement of the $E_i^S = D \cap D_i$ inside S. Label all the partition elements $C, \{E_j^S\}, \{E_k\}$ by P_i. Let m_i be the uniform probability measures on P_i. These are the conditional measures of m restricted to P_i.

The transition probability $\rho(dx, y)$ of x given y (for each fixed y this is a measure) is easily seen to be

$$\rho(dx, y) = m_{i_y}(dx) p_{i_x, i_y}$$

where i_y, i_x are the indices for which $y \in P_{i_y}, x \in P_{i_x}$ and p_{i_x, i_y} is the transition probability of a finite Markov chain on 9 symbols. Let $m(E)$ denote the area of any of the circular segments.

Let $a = m(E)/m(S)$ and note $1-4a = \frac{m(S)-4m(E)}{m(S)} = \frac{1-8m(E)}{m(S)} = m(C)/m(S)$. Then

$$P = (p_{ji}) = \begin{bmatrix} 1-4a & a & a & a & a & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1-4a & a & a & a & a & 0 & 0 & 0 & 0 \\ 1-4a & a & a & a & a & 0 & 0 & 0 & 0 \\ 1-4a & a & a & a & a & 0 & 0 & 0 & 0 \\ 1-4a & a & a & a & a & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Solving $\pi P = \pi$, we find the stationary distribution

$$\pi = (m(C), m(E), m(E), m(E), m(E), m(E), m(E), m(E), m(E)).$$

This chain is ergodic, and we can deduce convergence from that fact: Let x_i denote the outcomes of the sampling and let $f : D \to \mathbb{R}$. We need to show that, almost surely,

$$\frac{1}{n}\sum_{i<n} f(x_i) \to m(f).$$

Reorder the sum according to which D_j each x_i falls into. Relabel them x_{j_k} and let n_j be their cardinalities:

$$\frac{1}{n}\sum_{i<n} f(x_i) = \sum_j \frac{n_j}{n}\frac{1}{n_j}\sum_{k<n_j} f(x_{j_k})$$

For each fixed j, the sequence x_{j_k} are sampled from m_j. By the the law of large numbers for iid random variables, $\frac{1}{n_j}\sum_{k<n_j} f(x_{j_k}) = m_j(f) + o(1)$ as $n_j \to \infty$. By the ergodic theorem for Markov chains, $\frac{n_j}{n} = \pi_j + o(1)$ (in particular $n_j \to \infty$).

Combining these two facts,

$$\frac{1}{n}\sum_{i<n} f(x_i) = \sum_j \pi_j m_j(f) + o(1) = m(f) + o(1).$$

4. Alias sampling

We can remove the requirement to keep memory of the previous sample by applying ideas of the alias method [Wal77]. When sampling uniformly in S, the probability of sampling inside C is $\frac{m(C)}{m(S)}$. We wish to decrease this probability to be $m(C)$ by sampling again with probability q. This results in an equation for q:

$$m(C) = \frac{m(C)}{m(S)}(1-q) + \frac{m(C)}{m(S)} q \frac{m(C)}{m(S)}$$

which gives $q = \frac{2}{\pi} (= m(S))$. Additionally, each E_i^S gets E_i as its alias with alias probability 1/2 setting up the following rule: If $p \in E_i^S$, with probability 1/2 we take p, otherwise p'.

To see why we did not give C an alias to the E_i or E_i^S directly, we note that we first need to produce samples in such sets – by sampling S.

We give a pseudo-code implementation in Algorithm 2. We use that the condition $p \in C$ can be checked with $x^2 + y^2 + 2 \geq 4\max(|x|, |y|)$.

Algorithm 2 Adoption Algorithm without memory

```
Output: Sample (x, y) in Disk(0, √2)
procedure ADOPTION
    x ← random(−1, 1)
    y ← random(−1, 1)
    s ← x² + y² + 2
    if s ≥ 4 max(|x|, |y|) and random(0, 1) < 2/π then
        x ← random(−1, 1)
        y ← random(−1, 1)
        s ← x² + y² + 2
    end if
    if s < 4 max(|x|, |y|) and random(0, 1) < 1/2 then
        if      s <   4x then x ← x − 2
        else if s <   4y then y ← y − 2
        else if s < −4x then x ← x + 2
        else if s < −4y then y ← y + 2
        end if
    end if
    return (x, y)
end procedure
```

5. Monty Python sampling

One can also interpret this method as a variant of the Monty Python method introduced by Marsaglia and Tsang [MT98]. In this method, a probability distribution is cut and rearranged to fit a rectangle, and then rejection sampling is applied. By taking the inscribed square of the disk, we cut the disk segments that lie outside of the square and translate them inwards (adopted variates by translation). Alternatively, we fold the segments inside, leading to adopted variates by mirroring them at the boundary of the square. Then a strategy how to do sample in a one-to-many manner has to be devised for which we gave two possible solutions.

6. Application to low discrepancy sequences on the square

In computer graphics, there is often an interest in low-discrepancy sequences [Shi91]. These are abundant for the square, with classic examples including Sobol [Sob67] and Halton [Hal64]. For more modern approaches, see the Siggraph notes [Owe03]. If we are given a map from the square to the disk, we can warp these constructions to the disk. One such map is the concentric mapping, which is area-preserving with little distortion [SC97]. For an overview, see the Pixar Technical Memo [Chr18].

Clearly, instead of generating (x, y) randomly from the inscribed disk, we can take it from a low-discrepancy sequence. Since the map is locally isometric, the low-discrepancy property is preserved. To avoid seams that may appear at the edges of the inscribed square, one should prefer a low-discrepancy sequence on the 2-torus $\mathbb{R}^2/\mathbb{Z}^2$, where sides of the square are identified.

Acknowledgement. This work was supported by the ISF 264/22 while the author was at the Weizmann Institute of Science. I thank Omri Sarig for his hospitality and Won Chun for very helpful comments on the blog version [Rüh24a] of this paper.

References

[BM58] George EP Box and Mervin E Muller. A note on the generation of random normal deviates. *The annals of mathematical statistics*, 29(2):610–611, 1958.

[Chr18] Per Christensen. Progressive sampling strategies for disk light sources. *Pixar Animation Studios, Emeryville, CA, USA, Tech. Rep*, 2018.

[Hal64] John H Halton. Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM*, 7(12):701–702, 1964.

[Mar72] George Marsaglia. Choosing a point from the surface of a sphere. *The Annals of Mathematical Statistics*, 43(2):645–646, 1972.

[MT98] George Marsaglia and Wai Wan Tsang. The monty python method for generating random variables. *ACM Transactions on Mathematical Software (TOMS)*, 24(3):341–350, 1998.

[Mul59] Mervin E Muller. A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*, 2(4):19–20, 1959.

[Neu51] Von Neumann. Various techniques used in connection with random digits. *Notes by GE Forsythe*, pages 36–38, 1951.

[Owe03] Art B Owen. Quasi-monte carlo sampling. *Monte Carlo Ray Tracing: Siggraph*, 1:69–88, 2003.

[PJH23] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. MIT Press, forth edition, 2023.

[Rüh24a] René Rühr. Sampling the disk from a square: The adoption method. `https://rene.ruhr/gfx/adoption`, 2024.

[Rüh24b] René Rühr. Simd code, benchmark and visualisation for the adoption method. `https://github.com/reneruhr/adoption`, 2024.

[SC97] Peter Shirley and Kenneth Chiu. A low distortion map between disk and square. *Journal of graphics tools*, 2(3):45–52, 1997.

[Shi91] Peter Shirley. Discrepancy as a quality measure for sample distributions. In *Eurographics*, volume 91, pages 183–194, 1991.

[Sob67] Il'ya Meerovich Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

[Wal77] Alastair J. Walker. An efficient method for generating discrete random variables with general distributions. *ACM Trans. Math. Softw.*, 3(3):253–256, sep 1977.