Sampling on Parametric Polynomial Curves — the error

with
Apostolos Chalkis
& Christina Katsamaki

Geometry, Probability and Computing Seminar
UT San Antonio

Josué Tonelli-Cueto
Inria Paris & IMJ-PRG

5/May/2022

Part I:

Sampling, how hard can it be?

The Problem:

Given a random vector $x \in \mathbb{R}^n$,

is there an algorithm

such that its output

is identically distributed to x?

Discrete case I

Classical source of randomness:

$X \in \{0,1\}$ uniform

$$\Downarrow$$

We can produce almost any "reasonable" discrete distribution on $\mathbb{N}$

All probabilities should be computable in the Turing sense

Discrete case II

How to produce $\underset{\text{uniform}}{y \in \{0,1,\dots,n-1\}}$ out of $\underset{\text{uniform}}{x \in \{0,1\}}$?

$y = 0$

For $i \in \{0, \dots, \lfloor \log(n) \rfloor + 1\}$

- Sample $x \in \{0,1\}$,
- $y \leftarrow 2y + x$

If $y \le n-1$, output y

Otherwise: restart

How much time?

$\mathcal{O}(\log(n))$ in expectation

Discrete case III

And if we don't have random sources?

BIG PROBLEM!

If we can do this,

then $P \neq NP$

Continuous Case I: Exact sampling

Which sources?

Given continuous random sources $x_1, \dots, x_k \in \mathbb{R}$, which random vectors $y \in \mathbb{R}^N$ can we sample using a BSS machine?

Like Turing machine but can handle real numbers and operations $(+, -, \times, /, \sqrt{\ }, \leq, =)$ at cost one

Example: Normals are more powerful than uniforms

Sample $x, y, z \sim N(0,1)$

$$u \leftarrow \frac{x}{\sqrt{x^2+y^2+z^2}}$$

Output u

THM. $u \sim \mathcal{U}[-1,1]$

Q: And the converse?

If we need exactness,
it looks as extremely hard.

If we allow approximation,
we definitely can
(at least in practice)

Continuous Case II: Approximate sampling

Def. Given $x, y \in \mathbb{R}^n$ random vectors, their TV distance is given by

$$\mathrm{dist}_{TV}(x,y) := \sup_{B \in \mathcal{B}} \left| \mathbb{P}(x \in B) - \mathbb{P}(y \in B) \right|$$

Which sources are reasonable?

Given random sources $x_1, \dots, x_k \in \mathbb{R}$ and a target $y \in \mathbb{R}^n$ is there a BSS machine that for $\ell \in \mathbb{N}$ computes fast $\tilde{y} \in \mathbb{R}^n$ s.t. $\mathrm{dist}_{TV}(y, \tilde{y}) < 2^{-\ell}$?

Should we allow off-line computations?

What does this mean? poly(ℓ)? Or polylog(ℓ)?

Bridging the discrete and the continuous I

FACT: BSS machines cannot be constructed!

We cannot store real numbers and operate them at cost one ;(

How can we deal with this?

Big issue!

$\mathrm{dist}_{TV}(\tilde{x}, x) = 0$

if $\tilde{x}$ cont & x disc.

Bridging the discrete and the continuous II

Idea: We get a discrete random variable $x \in \mathbb{R}^n$ with support $\{x_1, \dots, x_k\}$, the last step is to turn x in continuous as follows:

Sample x
If $x = x_i$, then sample e_i
Output $x + e_i$

where $e_1, \dots, e_k$ are continuous random vectors that are "easy to sample".

Bridging the discrete and the continuous II

What we can produce $\downarrow$

What we want $\downarrow$

$$X_e \in \mathcal{U}\{0, 1/2^e, \dots, \tfrac{2^e-1}{2^e}\} \longrightarrow x \in \mathcal{U}[0,1]$$

In which sense does this happen?

An universal method: Inverse Transform Sampling

Thm. Let $\varphi : I = [a,b] \longrightarrow (0,\infty)$ be a continuous function such that

$$\int_I \varphi = 1$$

Let $\Phi(t) := \int_a^t \varphi(s)\, ds$ and $u \in \mathcal{U}[0,1]$, then

$$x := \Phi^{-1}(u)$$

has density φ

Lots of questions...

- how to solve $\Phi(x) = u$?
- how to deal with errors and hard-to-compute Φ?

More questions than answers...

Q1. What is the hierarchy among randomness sources?

Q2. How fast can we approximate in general?

Q3. Is TV dist the right notion?

Q4. How good is inv. trans. sampling when all details are taken into account?

Part II:

Sampling on a parametric polynomial curve

THE PROBLEM

Given a parametric polynomial curve

$$\gamma : I := [-1,1] \longrightarrow \mathbb{R}^n,$$

can we sample $x \in \operatorname{im} \gamma$ uniformly with respect the arc-length?

$\Longleftrightarrow$ Sample $t \in I$ with density $\propto \|\gamma'\|_2 = \sqrt{\sum_{i=1}^{n} (\gamma_i')^2}$

Why do we care?

- TDA in algebraic geometry
- Bridging gap between
  what we assume
  and what we can produce
- Simplest case for sampling in algebraic geometry

Condition of sampling γ

$$C(\gamma) := \sup_{t \in I} \frac{\|\gamma\|_0}{\|\gamma'(t)\|_2} \in [1, \infty]$$

where $\|\gamma\|_0 := \sum_{i,j} j|\gamma_{i,j}|$ with $\gamma = (\sum \gamma_{i,j} T^j)$

Intuition: The nearer a curve is to have a point of zero speed, the harder is to sample

Main THM

Let $\gamma : I \to \mathbb{R}^n$ a parametric polynomial curve of degree d. Then there is a sampler for γ that, on e, runs on

$$\mathcal{O}\left(e^3 \left(1 + \log dC(\gamma)\right)^3 d^3 C(\gamma)^3 \right)$$

on-line arithmetic operations.

It produces $t \in [-1,1]$ with TV dist $\leq 2^{-e}$ to target random variable.

Method (Olver & Townsed)

- Approximate $\varphi(t) := \|\gamma'(t)\|_2$ with Chebyshev
- Inverse Transform Sampling with bijection

Q: How fast is this?

↳ Q: How fast can we approximate φ?
(via Chebyshev)

Inv. Transf. Sampling Through Bijection

Input: $\varphi : [-1, 1] \to (0, \infty)$, ℓ

Output: x_ℓ distributed according to φ approx.

- Sample $u \in [0, 1]$ uniformly
- Bisection of $[-1, 1]$ ℓ times until we get $J = [k/2^\ell, \frac{k+1}{2^\ell}]$ with solution of $\Phi(x) = u$
- Output $x \in J$ uniform

THM. $\mathrm{dist}_{TV}(x_\ell, x) \leq 2^{1-\ell} \max_{x \in I} |\varphi'(x)|$

Fact. $x \sim \varphi$ $y \sim \psi$ $\mathrm{dist}_{TV}(x, y) \leq \| \varphi - \psi \|_1$

We only have to control the L^1 norm.

Where condition appears...

$$\varphi'(t) = \left(\|\gamma'(t)\|_2\right)' = \frac{\langle \gamma', \gamma'' \rangle}{\|\gamma'\|_2^2} \leq \frac{\|\gamma''(t)\|_2}{\|\gamma'(t)\|_2} \leq \underbrace{d\, C(\gamma)}$$

Chebyshev Interpolation I

Chebyshev olynomials

$$\psi_k(x) = \cos(k \arccos(x)) \quad \text{for } x \in [-1,1]$$

$$= \sum_{i=0}^{k} \binom{k}{2i} (1-x^2)^i x^{k-2i}$$

$$Z(\psi_k) = \{ \xi_{0,k}, \dots, \xi_{k-1,k} \}$$

$f : I \to \mathbb{R}$ $\qquad \psi I_k(f)$ kth degree interpolation

$$\psi I_k(f)(\xi_{a,k+1}) = f(\xi_{a,k+1})$$

Chebyshev Interpolation II: Advantages

1st Advantage: Easy to compute

$$\Psi I_K(f) = \frac{c_0}{2} + \sum_{i=1}^{K} c_a \Psi_a \iff c_a = \frac{2}{K+1} \sum_{i=0}^{K} f(\xi_{i,K+1}) \Psi_a(\xi_{i,K+1})$$

2nd Advantage: Fast to evaluate.

If $p = \sum_{a=0}^{K} c_a \Psi_a$, then

$$p(x) = \frac{1}{2}\left(b_0(x) - b_2(x)\right)$$

where

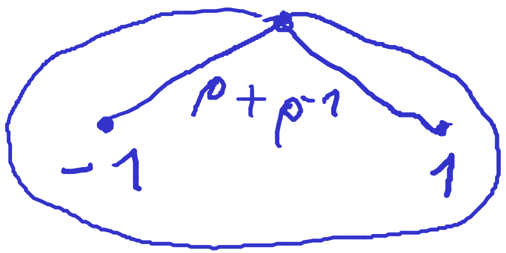
$$\begin{cases} b_{K+1}(x) = 0 = b_{K+2}(x) \\ b_a(x) = 2x\, b_{a+1}(x) - b_{a+2}(x) + c_a \end{cases}$$

3rd Advantage Fast to integrate

Chebyshev Interpolation III: Speed of convergence

THM. Let

$$E_\rho := \{ z \in \mathbb{C} \mid |z + \sqrt{z^2 - 1}| = \rho \} \qquad (\rho > 1)$$



and $f : \operatorname{int} E_\rho \to \mathbb{C}$ analytic

$$\| f - I_k(f) \|_\infty \leq \frac{4 \| f \|_{E_\rho}}{\rho - 1} \rho^{-k}$$

where $\| f \|_{E_\rho} := \max \{ |f(z)| \mid z \in E_\rho \}$

The speed I

$\varphi := \|\gamma'\|_2$ $\qquad\qquad$ $\varphi = \sqrt{p}$

Obs 1. φ^2 admits an. ext to the full complex plane

Obs 2. As long as $Z(p) \cap \overline{\operatorname{int} E_\rho} = \emptyset$, φ admits analytic ext. to $\overline{\operatorname{int} E_\rho}$

Prop. If p has no zeros inside E_ρ, then φ has an analytic extension $\tilde{\varphi}$ to int. of E_ρ and

over I

$$\|\tilde{\varphi}\|_{E_\rho} \leq \|\varphi\|_\infty \, \rho^d$$

The speed II

How big can ρ be?

THM We can pick

$$\rho \geq 1 + \frac{1}{3d\, C(\gamma)}$$

Open Questions

- Effect of subdivision in sampling

$$[-1, 1] \longrightarrow [-1, 0] \quad [0, 1]$$

- Why not Newton? (Assume $\varphi' > 0$)

$$\Phi(x) = u$$

$$\Psi_e(u) = \begin{cases} 1 & \text{if } e = 0 \\ N_{\Phi - u}(\Psi_{e-1}(u)) & \text{if } e > 0 \end{cases}$$

$\|\Psi_e - \Phi^{-1}\|_\infty$ small not good enough, we need $\|\Psi_e' - (\Phi^{-1})'\|_1$ small

- Complexity without $C(\gamma)$

Thank you!