

Open-Source Software for Science

From Science to Sofware to Science

Pamphile T. Roy



tupui



PamphileRoy

Pamphile who?

>>> 2019 - PhD from Cerfacs - UQ in CFD

- Statsmodels

>>> 2020 - Microsoft Flight Simulator 2020 - Backend and ML team

- SciPy

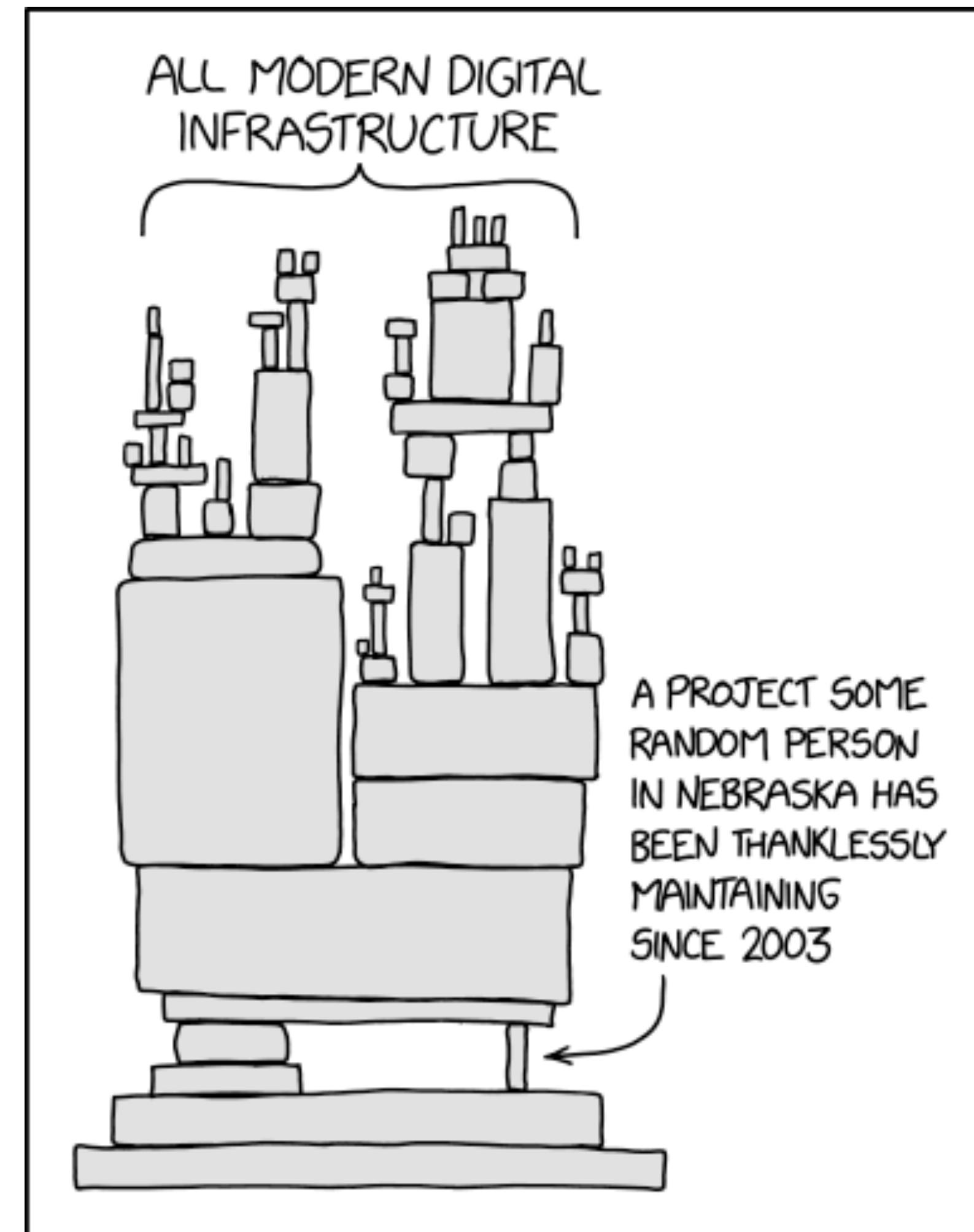
>>> 2021 - iTranslate – Backend and ML team

- Scientific Python, NumFOCUS

>>> Since 2022 - Quansight - Consultant

- SALib, ..., SimDec 😊

What does Open-Source mean?



What does Open-Source mean?

>>> Source code is available

But wait, there is more!

>>> Community

>>> Open Development

>>> Accountability

Still, Why? I have *Secrets*

- >>> Reproducible science
- >>> New collaborations
- >>> Code Quality
- >>> Help from random people like me 😅

Papers about Software?

nature methods

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature methods](#) > [perspectives](#) > article

Perspective | [Open Access](#) | Published: 03 February 2020

SciPy 1.0: fundamental algorithms for scientific computing in Python

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, SciPy

[1.0 Contributors](#) + Show authors

[Nature Methods](#) 17, 261–272 (2020) | [Cite this article](#)

171k Accesses | 11k Citations | 868 Altmetric | [Metrics](#)

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [review articles](#) > article

Review Article | [Open Access](#) | Published: 16 September 2020

Array programming with NumPy

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, ... Travis E. Oliphant + Show authors

[Nature](#) 585, 357–362 (2020) | [Cite this article](#)

328k Accesses | 6851 Citations | 1827 Altmetric | [Metrics](#)

Python the new R or Matlab

- >>> Python is still one of the fastest growing language
- >>> Most developers/researchers doing machine learning and AI are using Python
- >>> More than 50% of developers using C-like are not interested in continuing to do so
- >>> Python is fast enough... really. And if not there are solutions (Cython, Pythran, Numba, or even Rust)

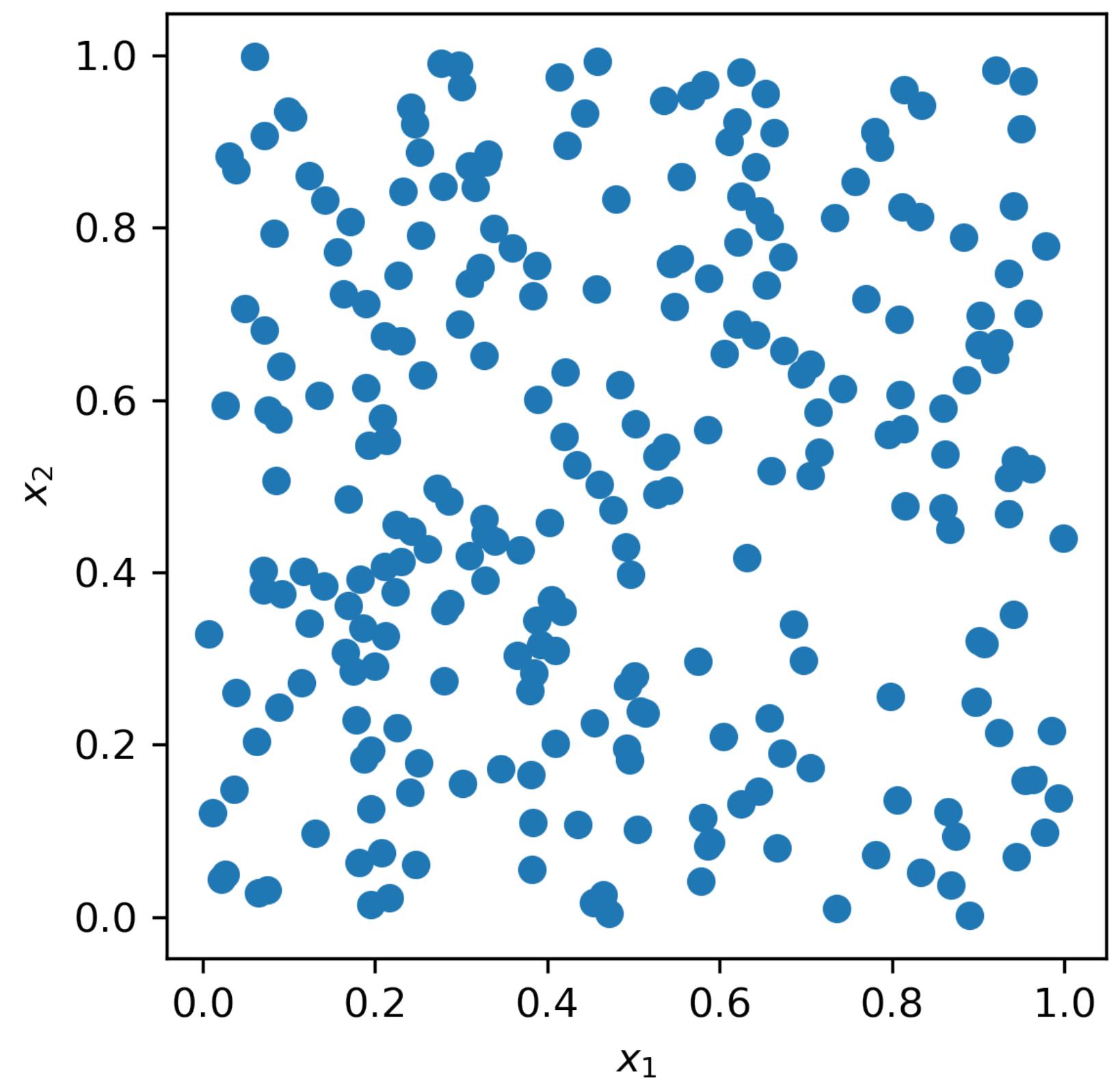
SciPy

Statistics, optimization,
interpolation,
integration, *



Random numbers you said?

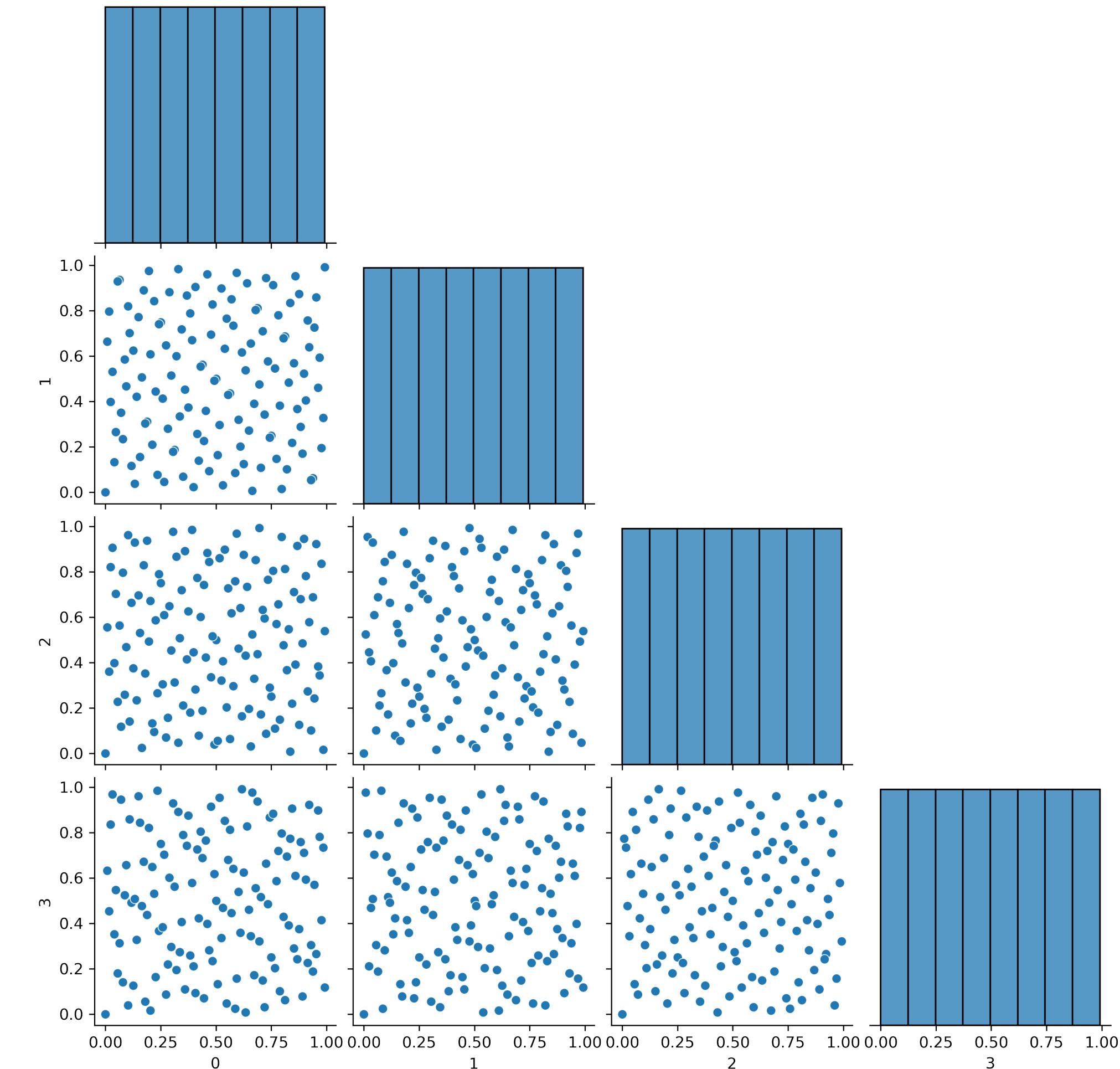
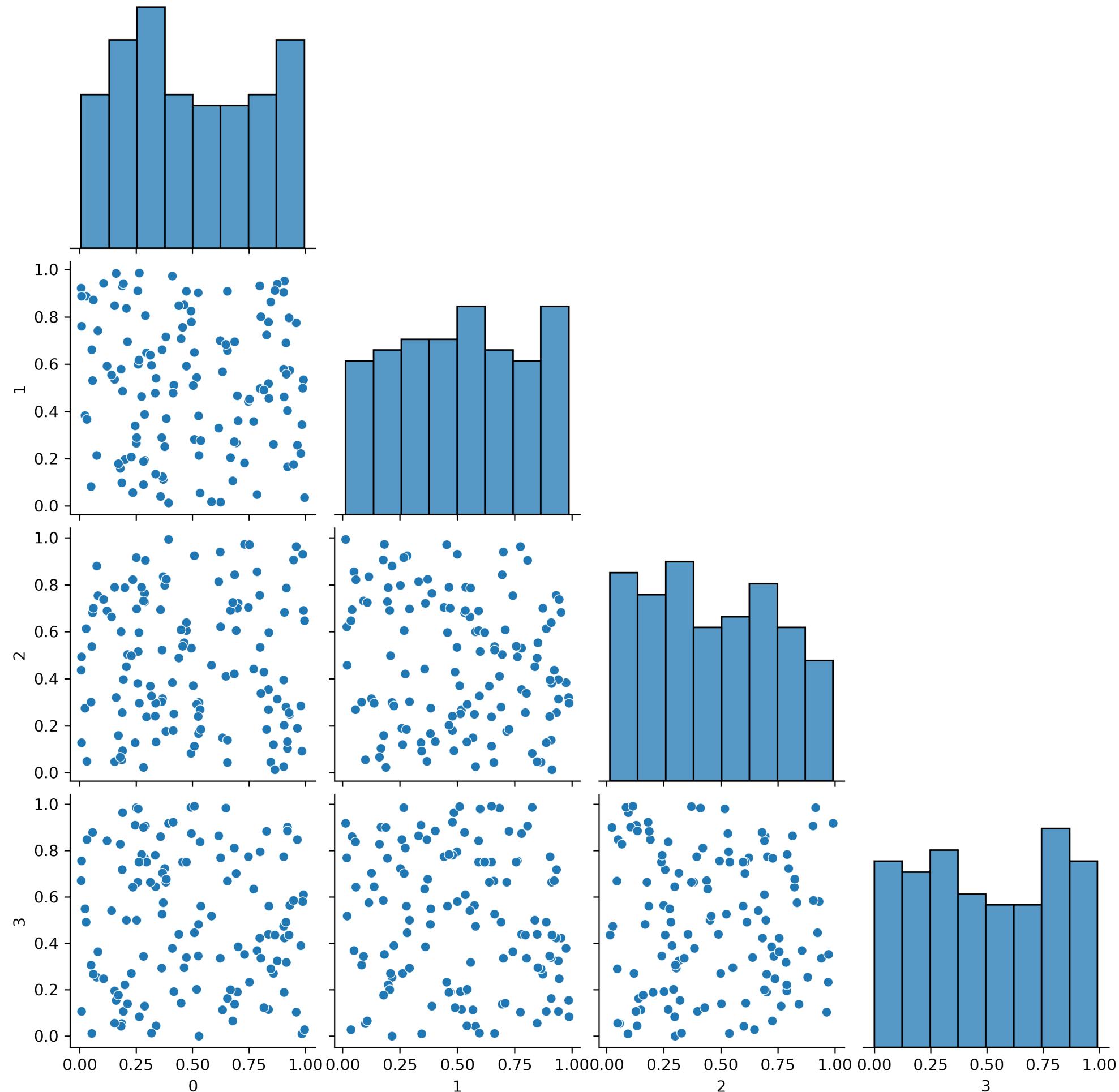
$\mathbf{x} = (x_1, \dots, x_n)$
 $x \sim \mathcal{U}(0,1)$



What is Quasi-Monte Carlo?

- >>> Deterministic methods
- >>> Not **i**ndependent and **i**dentially **distributed (IID)**
- >>> Better space filling
- >>> Better convergence

What is Quasi-Monte Carlo?



What is Quasi-Monte Carlo?

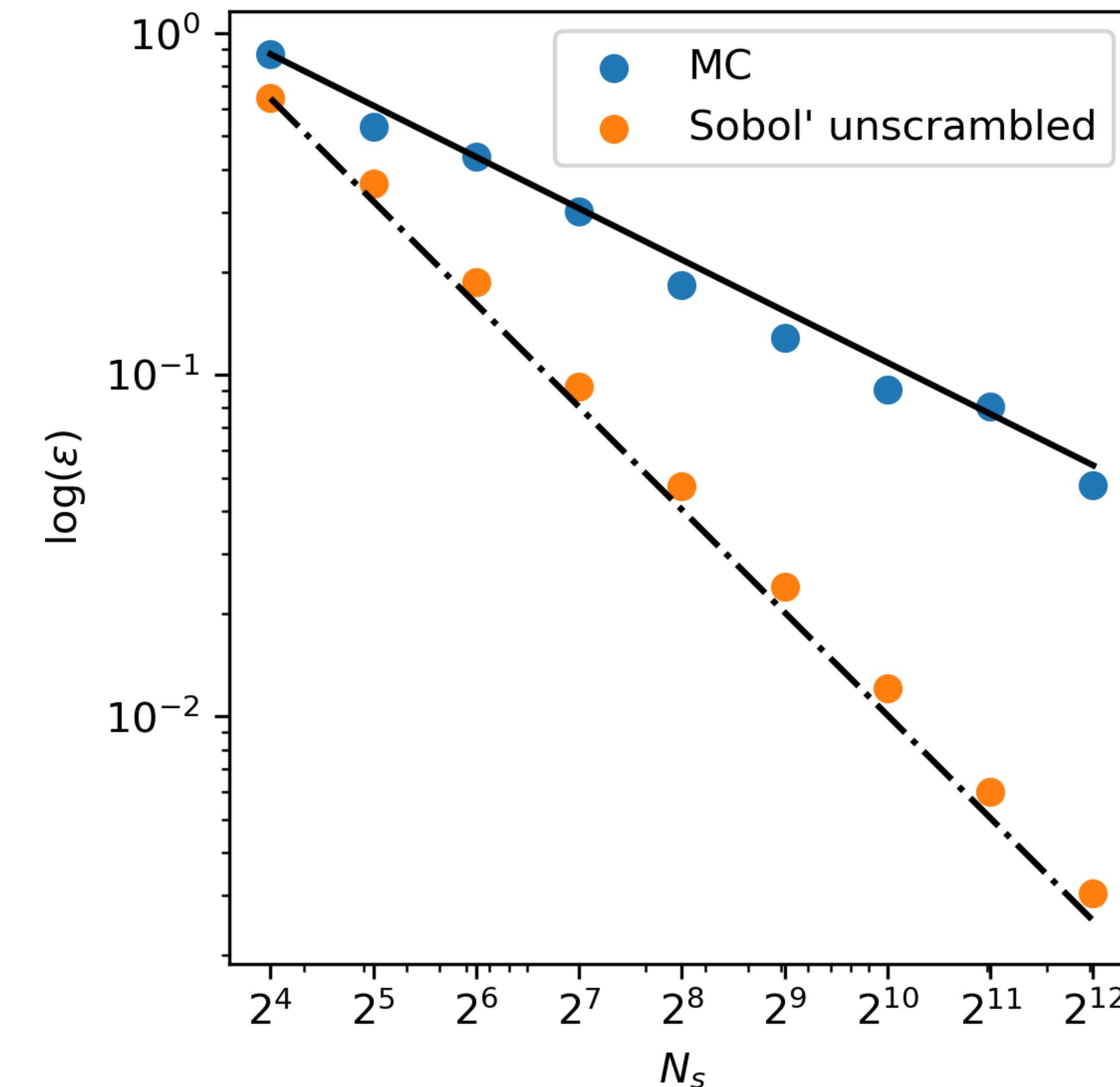
$$f(\mathbf{x}) = \left(\sum_{j=1}^5 x_j \right)^2$$

$$x_j \sim \mathcal{U}(0,1)$$

$$\mu = 5/3 + 5(5 - 1)/4$$

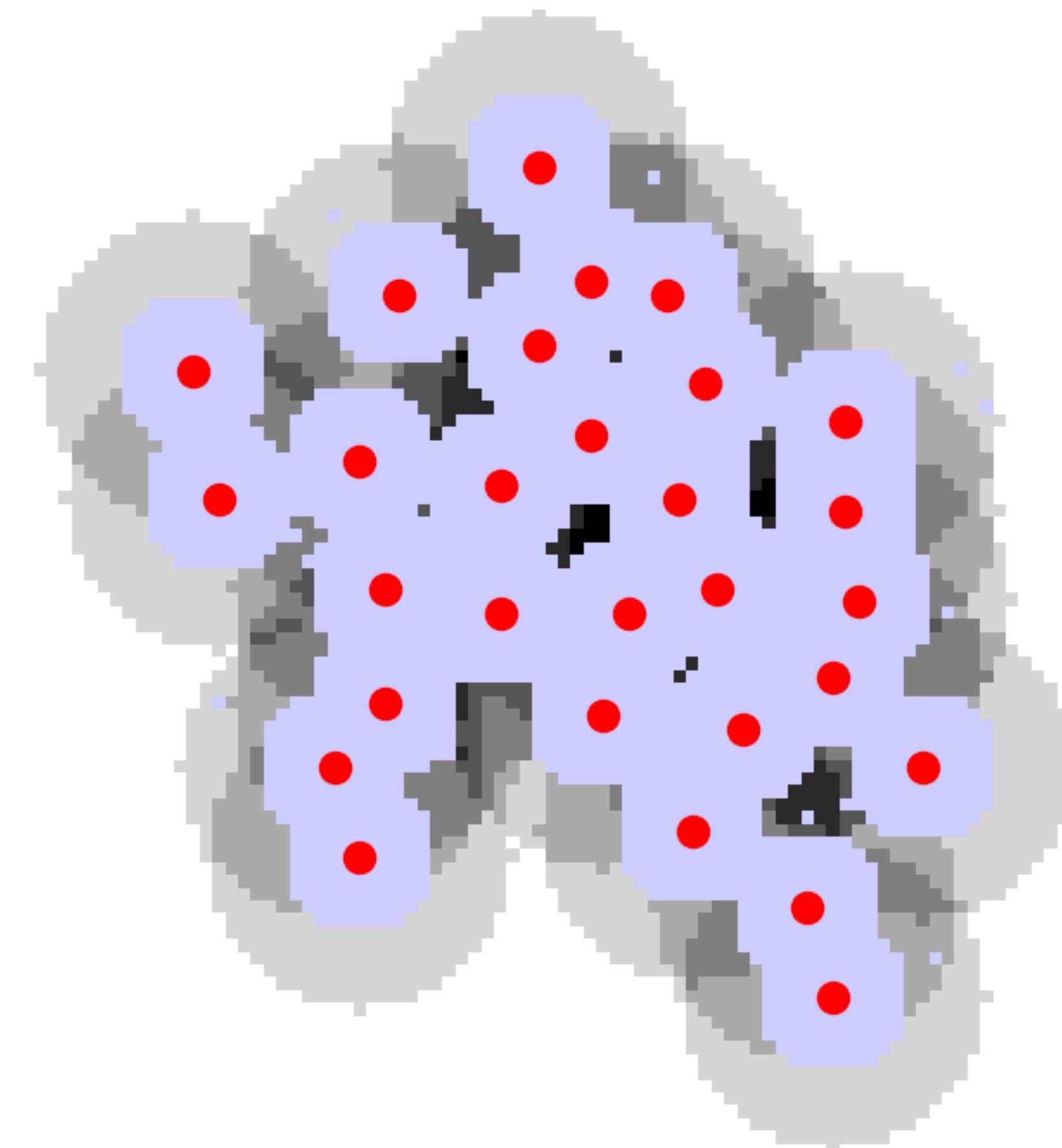
>>> Faster convergence
integration

... $\mathcal{O}(n^{-1/2})$ to $\mathcal{O}(n^{-1})$



```
>>> from scipy.stats import qmc
```

```
>>> Sobol'  
>>> Halton  
>>> LHS and Orthogonal LHS  
>>> Poisson disk  
>>> Multinomial  
>>> Multivariate normal  
>>> Discrepancy
```



Sample various distributions with QMC

```
>>> Easy API
```

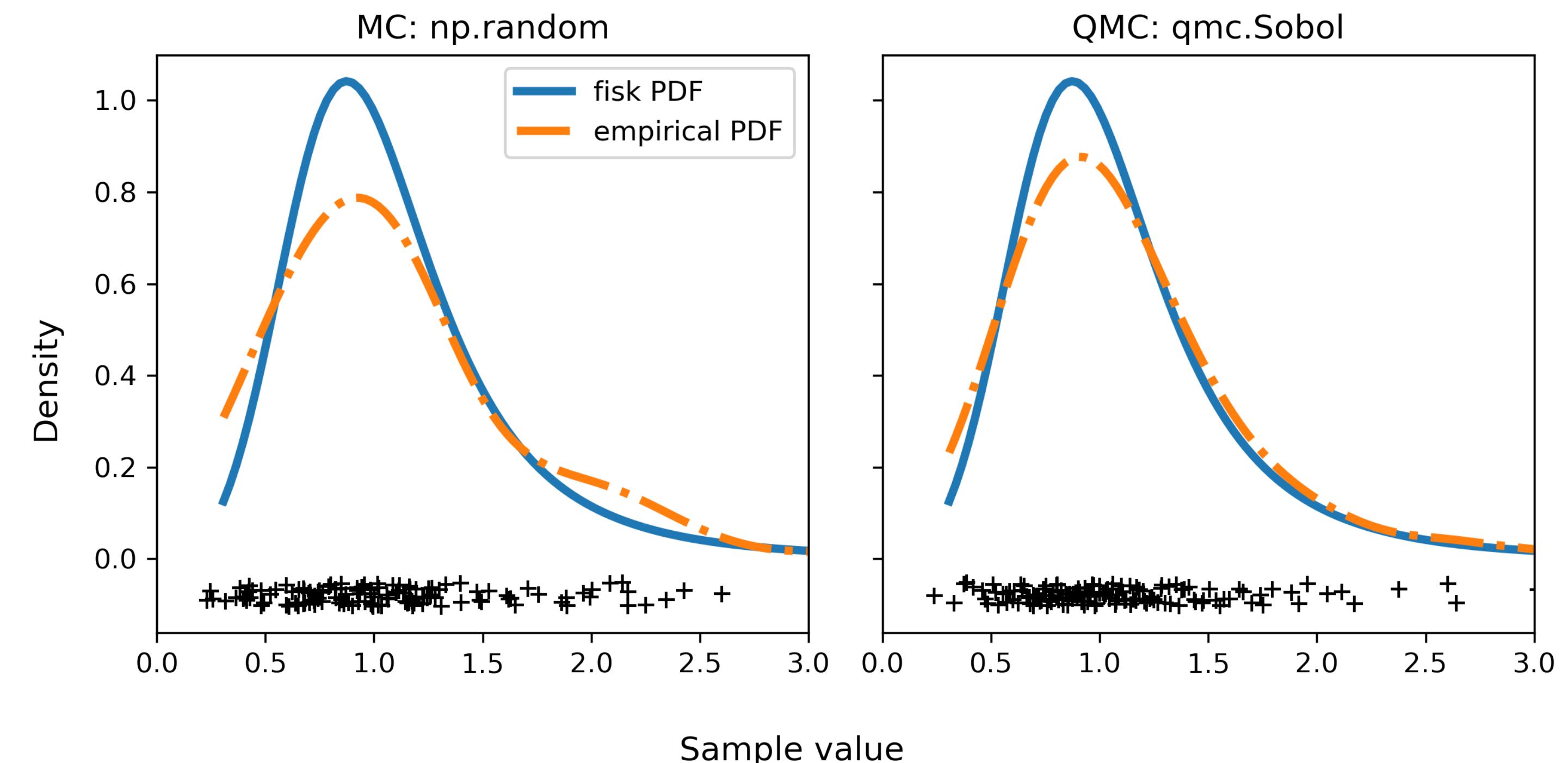
```
>>> sampling
```

```
>>> import scipy.stats as stats
```

```
>>> dist = stats.fisk(c=3.9)
```

```
>>> rng_dist = stats.sampling.NumericalInverseHermite(dist)
```

```
>>> sample_qmc = rng_dist.qrvs(128)
```



Do's and don'ts

- >>> Know where speed matters: sampling/init
- >>> Use QMC!
- >>> Don't modify the sequences
- >>> Don't use a seed (ready to die on this hill)

Collaboration at it's Best

>>> Hundreds of e-mails

>>> Hundreds of comments

>>> Some papers a long the way

>>> And new friends

ENH: add stats.qmc module with quasi Monte Carlo functionality
#10844

Merged by rgommers [scipy:master](#) ← [tupui:enh_qmc](#) on Jan 30, 2021

Conversation 613 Commits 86 Checks 0 Files changed 13 +3,172 -1

tupui on Sep 19, 2019 • edited Member

Add `scipy.stats.qmc`.

Reference issue

Closes #9695

What does this implement/fix?

Provide a set of functions to create and assess quasi-Monte Carlo Design of Experiments.

It provides a generic class `scipy.stats.qmc.QMCEngine` which defines a QMC engine/sampler. An engine is state aware: it can be continued, advanced and reseted. 4 base samplers are available:

- `scipy.stats.qmc.Sobol`: the well known Sobol' low discrepancy sequence. Several warnings have been added to guide the user into properly using this sampler. The sequence is scrambled by default.
- `scipy.stats.qmc.Halton`: Halton low discrepancy sequence. The sequence is scrambled by default.
- `scipy.stats.qmc.LatinHypercube`: plain LHS design.
- `scipy.stats.qmc.OrthogonalLatinHypercube`: Orthogonal version of LHS which is more uniform than plain LHS.

ENH: add Sobol' indices #17628

Merged by mdhaber [scipy:main](#) ← [tupui:sobol_indices](#) on Jan 18 v1.11.0rc1

Conversation 181 Commits 92 Checks 14 Files changed 5 +1,016 -0

tupui on Dec 19, 2022 Member

What does this implement/fix?

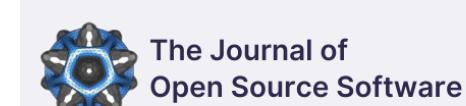
This PR adds a function to compute Sobol' indices.

Additional information

The function would compute Sobol' indices [1,2]. Being able to compute these indices allows to reduce the dimensionality of a problem, better understand the importance of each factors and also see how parameters are interacting with each other. As such, it's an important engineering tool. If you have 2 variables and only have the budget to improve your knowledge on one of them, this can help to make a choice. There are a lot of successful usage of Sobol' indices in the literature and in real world applications. The EU (through the JRC), is now requiring to conduct uncertainty analysis when evaluating a system. As such, they are recommending the use of Sobol' indices.

Upon previous discussion with maintainers of the SALib library, they expressed their interest in getting this into SciPy so they can rely on our implementation and focus on advanced post processing. We would also help the method gain visibility in the Python sphere (R, Matlab, even Excel have great first class support)

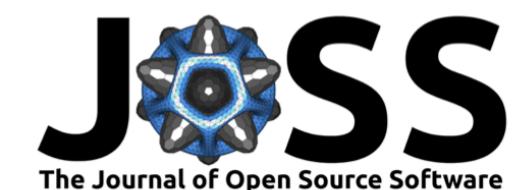
Software to Papers



About Papers Docs Blog Submit

Quasi-Monte Carlo Methods in Python

Starlark Python Meson C Submitted 17 February 2023 • Published 23 April 2023



Quasi-Monte Carlo Methods in Python

Pamphile T. Roy¹, Art B. Owen², Maximilian Balandat³, and Matt Haberland⁴

1 Quansight 2 Stanford University 3 Meta 4 California Polytechnic State University, San Luis Obispo, USA ¶ Corresponding author

DOI: [10.21105/joss.05309](https://doi.org/10.21105/joss.05309)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Summary

NumPy random number generators and SciPy distributions are widely used to generate random numbers. However, challenges might arise when sampling in high dimensions. Quasi-Monte Carlo (QMC) methods provide an answer to these problems but are arguably hard to use. Thanks to new developments in SciPy, a new submodule was introduced in version 1.7.0 making state-of-the-art QMC methods available: `scipy.stats.qmc`.

Editor: Mehmet Hakan Satman

Reviewers:

- [@yangyushi](#)
- [@ptmerz](#)

Statement of need

International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing
↳ MCQMC 2020: [Monte Carlo and Quasi-Monte Carlo Methods](#) pp 71–86 | [Cite as](#)

[Home](#) > [Monte Carlo and Quasi-Monte Carlo Methods](#) > Conference paper

On Dropping the First Sobol' Point

Art B. Owen [✉](#)

Conference paper | [First Online: 21 May 2022](#)

654 Accesses | 3 Citations

Part of the [Springer Proceedings in Mathematics & Statistics](#) book series (PROMS, volume 387)

Review

Editor: [@jbytecode](#) (all papers)
Reviewers: [@yangyushi](#) (all reviews),
[@ptmerz](#) (all reviews)

Authors

Pamphile T. Roy (0000-0001-9816-1416),
Art B. Owen (0000-0001-5860-3945),
Maximilian Balandat (0000-0002-8214-
8935), Matt Haberland (0000-0003-
4806-3601)

Citation

Roy et al., (2023). Quasi-Monte Carlo Methods in Python. *Journal of Open Source Software*, 8(84), 5309, <https://doi.org/10.21105/joss.05309>

Who is funding that?

- >>> Grants from private organizations: CZI,
NumFOCUS, Moore, ...
- >>> Companies: Quansight, Anaconda, NVIDIA, META, ...
- >>> Volunteers - Mostly

<https://tidelift.com/subscription/pkg/pypi-scipy>

<https://opencollective.com/scipy>

<https://chanzuckerberg.com/eoss/proposals/a-solid-foundation-for-statistics-in-python-with-scipy>

Is it sustainable?

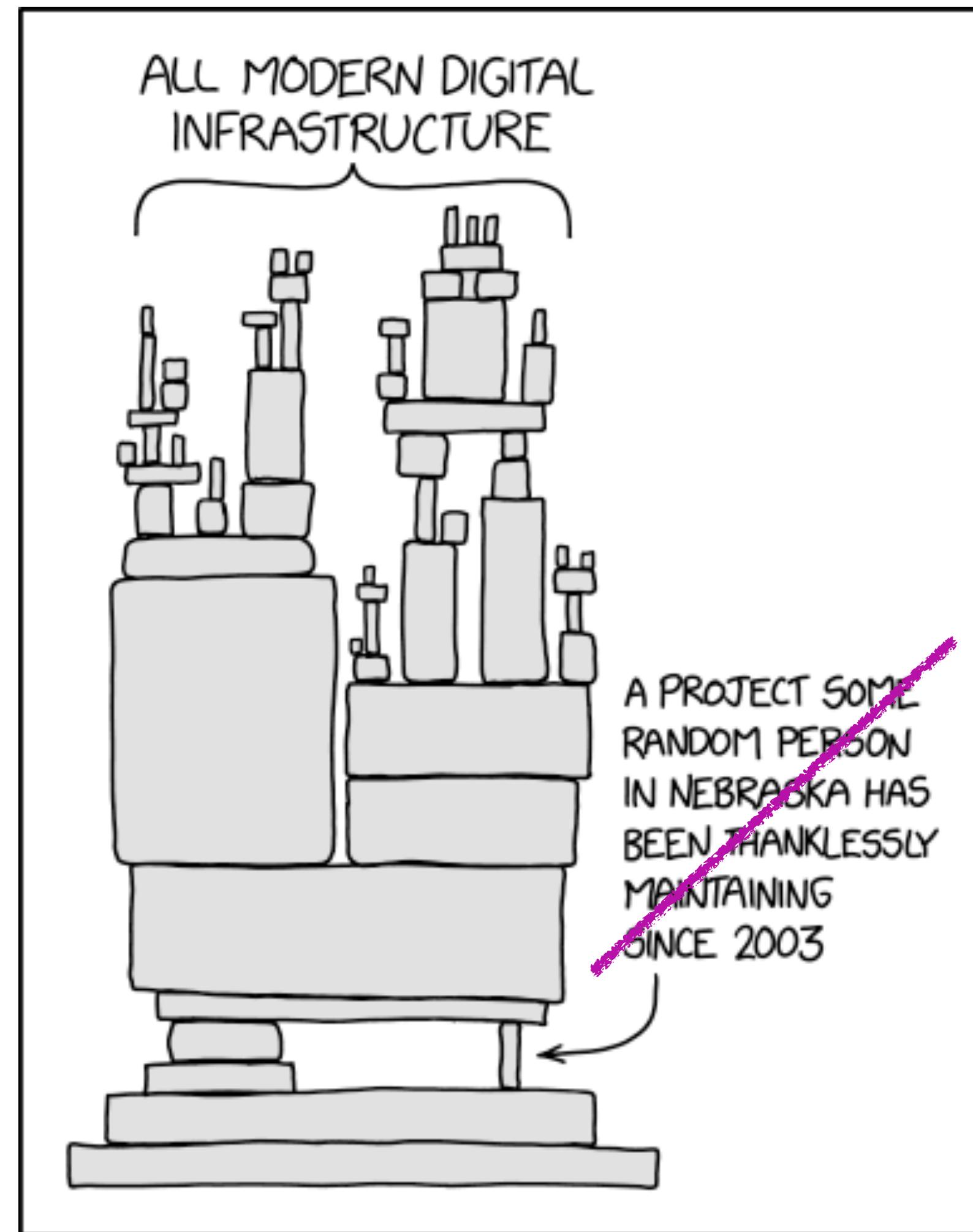
- >>> Still relies on volunteers
- >>> Science in Python is still very “academic”
- >>> Governance
- >>> Communication
- >>> Responsibilities

Scientific Python

Community | developed
| owned



What does Open-Source mean?

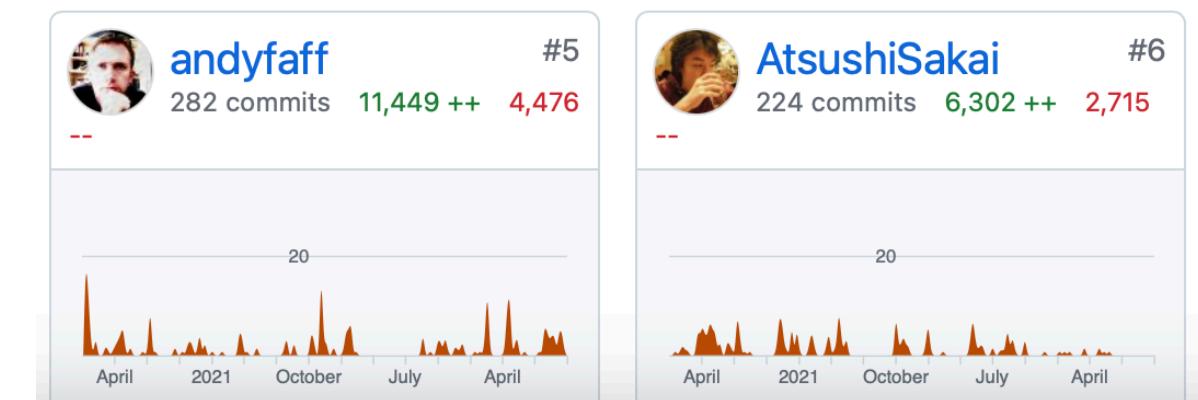
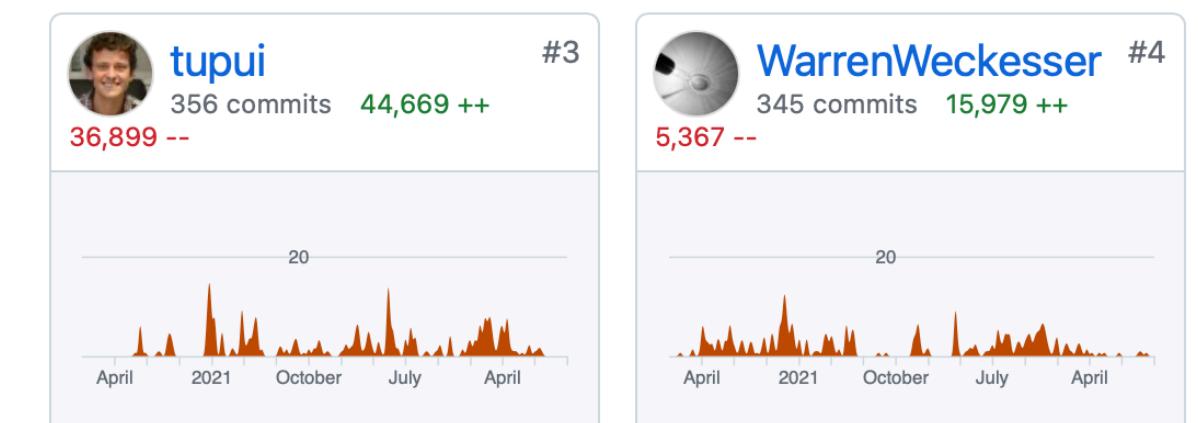
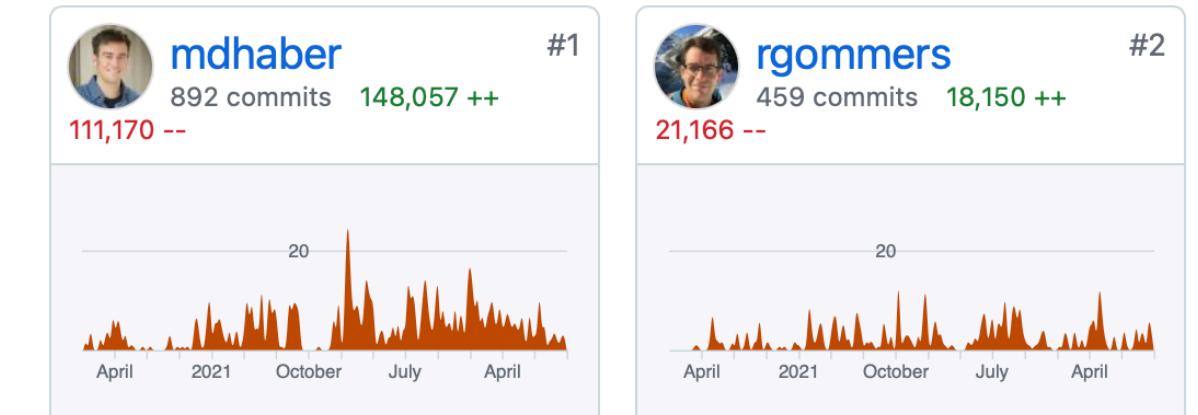
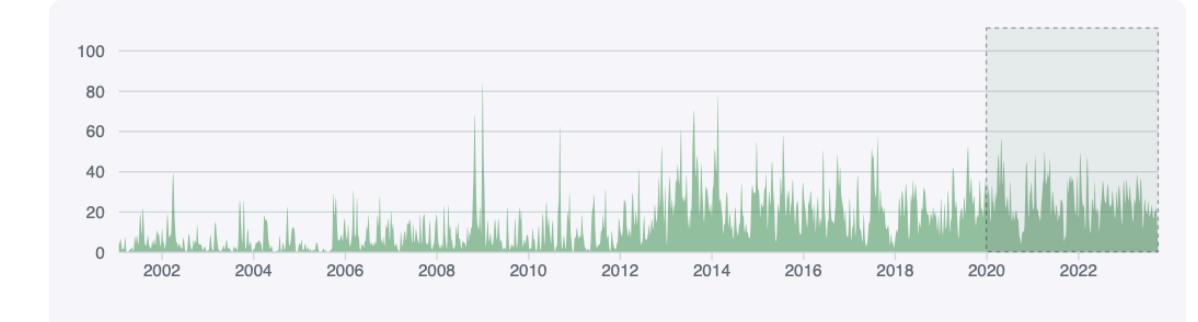


Open-Source Community!

Jan 1, 2020 – Oct 2, 2023

Contributions: Commits ▾

Contributions to main, excluding merge commits and bot accounts



You help[ed] make it happen...

Issues, PR, reviews, discussions, issue management: *every contribution matters, even a star on the repository...* Join us!

Thank You!

SimDec

Simulation Decomposition

