

Search the doc:

- scipy
- scipy.cluster
- scipy.constants
- scipy.datasets
- scipy.fft
- scipy.fftpack
- scipy.integrate
- scipy.interpolate
- scipy.io
- scipy.linalg
- scipy.misc
- scipy.ndimage
- scipy.odr
- scipy.optimize
- scipy.signal
- scipy.sparse
- scipy.spatial
- scipy.special
- scipy.stats

# Quasi-Monte Carlo submodule

## (scipy.stats.qmc)

This module provides Quasi-Monte Carlo generators and associated helper functions.

## Quasi-Monte Carlo

### Engines

<b>QMC</b> <b>Engine</b> (d, *[, optimization, seed])	A generic Quasi-Monte Carlo sampler class meant for subclassing.
<b>Sobol</b> (d, *[, scramble, bits, seed, optimization])	Engine for generating (scrambled) Sobol' sequences.
<b>Halton</b> (d, *[, scramble, optimization, seed])	Halton sequence.
<b>LatinHypercube</b> (d, *[, scramble, strength, ...])	Latin hypercube sampling (LHS).
<b>PoissonDisk</b> (d, *[, radius, hypersphere, ...])	Poisson disk sampling.
<b>MultinomialQMC</b> (pvals, n_trials, *[, engine, ...])	QMC sampling from a multinomial distribution.
<b>MultivariateNormalQMC</b> (mean[, cov, cov_root, ...])	QMC sampling from a multivariate Normal $N(\mu, \Sigma)$ .

### Helpers

<b>discrepancy</b> (sample, *[, iterative, method, ...])	Discrepancy of a given sample.
<b>geometric_discrepancy</b> (sample[, method, metric])	Discrepancy of a given sample based on its geometric properties.
<b>update_discrepancy</b> (x_new, sample, initial_disc)	Update the centered discrepancy with a new sample.
<b>scale</b> (sample, l_bounds, u_bounds, *[, reverse])	Sample scaling from unit hypercube to different bounds.

## Introduction to Quasi-Monte Carlo

Quasi-Monte Carlo (QMC) methods [1], [2], [3] provide an  $n \times d$  array of numbers in  $[0, 1]$ . They can be used in place of  $n$  points from the  $U[0, 1]^d$  distribution. Compared to random points, QMC points are designed to have fewer gaps and clumps. This is quantified by discrepancy measures [4]. From the Koksma-Hlawka inequality [5] we know that low discrepancy reduces a bound on integration error. Averaging a function  $f$  over  $n$  QMC points can achieve an integration error close to  $O(n^{-1})$  for well behaved functions [2].

Most QMC constructions are designed for special values of  $n$  such as powers of 2 or large primes. Changing the sample size by even one can degrade their performance, even their rate of convergence [6]. For instance  $n = 100$  points may give less accuracy than  $n = 64$  if the method was designed for  $n = 2^m$ .

Some QMC constructions are extensible in  $n$ : we can find another special sample size  $n' > n$  and often an infinite sequence of increasing special sample sizes. Some QMC constructions are extensible in  $d$ : we can increase the dimension, possibly to some upper bound, and typically without requiring special values of  $d$ . Some QMC methods are extensible in both  $n$  and  $d$ .

QMC points are deterministic. That makes it hard to estimate the accuracy of integrals estimated by averages over QMC points. Randomized QMC (RQMC) [7] points are constructed so that each point is individually  $U[0, 1]^d$  while collectively the  $n$  points retain their low discrepancy. One can make  $R$  independent replications of RQMC points to see how stable a computation is. From  $R$  independent values, a t-test (or bootstrap t-test [8]) then gives approximate confidence intervals on the mean value. Some RQMC methods produce a root mean squared error that is actually  $o(1/n)$  and smaller than the rate seen in unrandomized QMC. An intuitive explanation is that the error is a sum of many small ones and random errors cancel in a way that deterministic ones do not. RQMC also has advantages on integrands that are singular or, for other reasons, fail to be Riemann integrable.

(R)QMC cannot beat Bahkvalov's curse of dimension (see [9]). For any random or deterministic method, there are worst case functions that will give it poor performance in high dimensions. A worst case function for QMC might be 0 at all  $n$  points but very large elsewhere. Worst case analyses get very pessimistic in high dimensions. (R)QMC can bring a great improvement over MC when the functions on which it is used are not worst case. For instance (R)QMC can be especially effective on integrands that are well approximated by sums of functions of some small number of their input variables at a time [10], [11]. That property is often a surprising finding about those functions.

Also, to see an improvement over IID MC, (R)QMC requires a bit of smoothness of the integrand, roughly the mixed first order derivative in each direction,  $\partial^d f / \partial x_1 \cdots \partial x_d$ , must be integral. For instance, a function that is 1 inside the hypersphere and 0 outside of it has infinite variation in the sense of Hardy and Krause for any dimension  $d = 2$ .

Scrambled nets are a kind of RQMC that have some valuable robustness properties [12]. If the integrand is square integrable, they give variance  $vars_{NET} = o(1/n)$ . There is a finite upper bound on  $vars_{NET} / var_{MC}$  that holds simultaneously for every square integrable integrand. Scrambled nets satisfy a strong law of large numbers for  $f$  in  $L^p$  when  $p > 1$ . In some special cases there is a central limit theorem [13]. For smooth enough integrands they can achieve RMSE nearly  $O(n^{-3})$ . See [12] for references about these properties.

The main kinds of QMC methods are lattice rules [14] and digital nets and sequences [2], [15]. The theories meet up in polynomial lattice rules [16] which can produce digital nets. Lattice rules require some form of search for good constructions. For digital nets there are widely used default constructions.

The most widely used QMC methods are Sobol' sequences [17]. These are digital nets. They are extensible in both  $n$  and  $d$ . They can be scrambled. The special sample sizes are powers of 2. Another popular method are Halton sequences [18]. The constructions resemble those of digital nets. The earlier dimensions have much better equidistribution properties than later ones. There are essentially no special sample sizes. They are not thought to be as accurate as Sobol' sequences. They can be scrambled. The nets of Faure [19] are also widely used. All dimensions are equally good, but the special sample sizes grow rapidly with dimension  $d$ . They can be scrambled. The nets of Niederreiter and Xing [20] have the best asymptotic properties but have not shown good empirical performance [21].

Higher order digital nets are formed by a digit interleaving process in the digits of the constructed points. They can achieve higher levels of asymptotic accuracy given higher smoothness conditions on  $f$  and they can be scrambled [22]. There is little or no empirical work showing the improved rate to be attained.

Using QMC is like using the entire period of a small random number generator. The constructions are similar and so therefore are the computational costs [23].

(R)QMC is sometimes improved by passing the points through a baker's transformation (tent function) prior to using them. That function has the form  $1 - 2|x - 1/2|$ . As  $x$  goes from 0 to 1, this function goes from 0 to 1 and then back. It is very useful to produce a periodic function for lattice rules [14], and sometimes it improves the convergence rate [24].

It is not straightforward to apply QMC methods to Markov chain Monte Carlo (MCMC). We can think of MCMC as using  $n = 1$  point in  $[0, 1]^d$  for very large  $d$ , with ergodic results corresponding to  $d \rightarrow \infty$ . One proposal is in [25] and under strong conditions an improved rate of convergence has been shown [26].

Returning to Sobol' points: there are many versions depending on what are called direction numbers. Those are the result of searches and are tabulated. A very widely used set of direction numbers come from [27]. It is extensible in dimension up to  $d = 21201$ .

## References

[1]

Owen, Art B. "Monte Carlo Book: the Quasi-Monte Carlo parts." 2019.

[2](1,2,3)

Niederreiter, Harald. "Random number generation and quasi-Monte Carlo methods." Society for Industrial and Applied Mathematics, 1992.

[3]

Dick, Josef, Frances Y. Kuo, and Ian H. Sloan. "High-dimensional integration: the quasi-Monte Carlo way." Acta Numerica no. 22: 133, 2013.

[4]

Aho, A. V., C. Aistleitner, T. Anderson, K. Appel, V. Arno'l'd, N. Aronszajn, D. Asotsky et al. "W. Chen et al.(eds.), "A Panorama of Discrepancy Theory", Sringer International Publishing, Switzerland: 679, 2014.

[5]

Hickernell, Fred J. "Koksma-Hlawka Inequality." Wiley StatsRef: Statistics Reference Online, 2014.

[6]

Owen, Art B. "On dropping the first Sobol' point." [arXiv:2008.08051](#), 2020.

[7]

L'Ecuyer, Pierre, and Christiane Lemieux. "Recent advances in randomized quasi-Monte Carlo methods." In Modeling uncertainty, pp. 419-474. Springer, New York, NY, 2002.

[8]

DiCiccio, Thomas J., and Bradley Efron. "Bootstrap confidence intervals." Statistical science: 189-212, 1996.

[9]

Dimov, Ivan T. "Monte Carlo methods for applied scientists." World Scientific, 2008.

[10]

Caflish, Russel E., William J. Morokoff, and Art B. Owen. "Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension." Journal of Computational Finance: no. 1 27-46, 1997.

[11]

Sloan, Ian H., and Henryk Wozniakowski. "When are quasi-Monte Carlo algorithms efficient for high dimensional integrals?." Journal of Complexity 14, no. 1 (1998): 1-33.

[12](1,2)

Owen, Art B., and Daniel Rudolf, "A strong law of large numbers for scrambled net integration." SIAM Review, to appear.

[13]

Loh, Wei-Liem. "On the asymptotic distribution of scrambled net quadrature." The Annals of Statistics 31, no. 4: 1282-1324, 2003.

[14](1,2)

Sloan, Ian H. and S. Joe. "Lattice methods for multiple integration." Oxford University Press, 1994.

[15]

Dick, Josef, and Friedrich Pillichshammer. "Digital nets and sequences: discrepancy theory and quasi-Monte Carlo integration." Cambridge University Press, 2010.

[16]

Dick, Josef, F. Kuo, Friedrich Pillichshammer, and I. Sloan. "Construction algorithms for polynomial lattice rules for multivariate integration." Mathematics of computation 74, no. 252: 1895-1921, 2005.

[17]

Sobol', Il'ya Meerovich. "On the distribution of points in a cube and the approximate evaluation of integrals." Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki 7, no. 4: 784-802, 1967.

[18]

Halton, John H. "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals." Numerische Mathematik 2, no. 1: 84-90, 1960.

[19]

Faure, Henri. "Discrepance de suites associees a un systeme de numeration (en dimension s)." Acta arithmetica 41, no. 4: 337-351, 1982.

[20]

Niederreiter, Harold, and Chaoping Xing. "Low-discrepancy sequences and global function fields with many rational places." Finite Fields and their applications 2, no. 3: 241-273, 1996.

[21]

Hong, Hee Sun, and Fred J. Hickernell. "Algorithm 823: Implementing scrambled digital sequences." ACM Transactions on Mathematical Software (TOMS) 29, no. 2: 95-109, 2003.

[22]

Dick, Josef. "Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands." The Annals of Statistics 39, no. 3: 1372-1398, 2011.

[23]

Niederreiter, Harald. "Multidimensional numerical integration using pseudorandom numbers." In Stochastic Programming 84 Part I, pp. 17-38. Springer, Berlin, Heidelberg, 1986.

[24]

Hickernell, Fred J. "Obtaining O (N-2+e) Convergence for Lattice Quadrature Rules." In Monte Carlo and Quasi-Monte Carlo Methods 2000, pp. 274-289. Springer, Berlin, Heidelberg, 2002.

[25]

Owen, Art B., and Seth D. Tribble. "A quasi-Monte Carlo Metropolis algorithm." Proceedings of the National Academy of Sciences 102, no. 25: 8844-8849, 2005.

[26]

Chen, Su. "Consistency and convergence rate of Markov chain quasi Monte Carlo with examples." PhD diss., Stanford University, 2011.

[27]

Joe, Stephen, and Frances Y. Kuo. "Constructing Sobol sequences with better two-dimensional projections." SIAM Journal on Scientific Computing 30, no. 5: 2635-2654, 2008.