AI505

Optimization

Sampling Plans
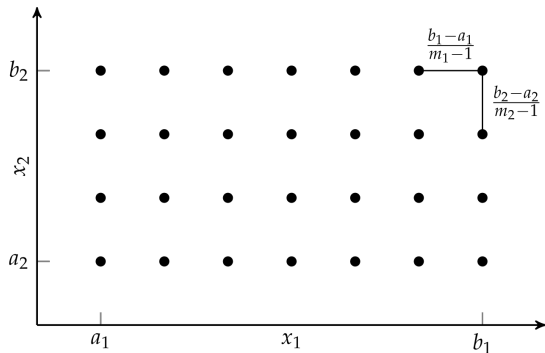
Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

Full Factorial

- Uniform and evenly spaced samples across domain

- Simple, easy to implement, and covers domain

- Sample count grows exponentially with dimension

Random Sampling

- Uses pseudorandom number generator to define samples according to our desired distribution

- If variable bounds are known, a common choice is a uniform distribution across domain of possible values

- Ideally, if enough points are sampled and the right distribution is chosen, the design space will be covered
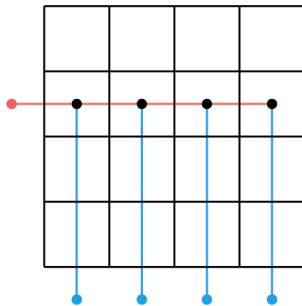
Uniform Projection Plans

- Sample locations are random, but their projection onto a particular subspace is uniform
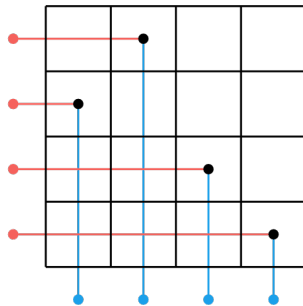
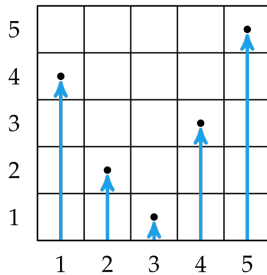too clustered
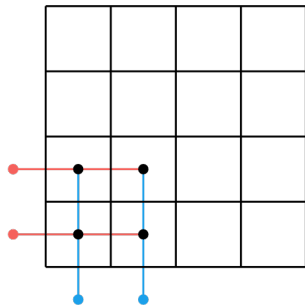
no variation in
one component

uniform projection

Uniform Projection Plans

- Examples: Random permutation, Latin square

$$p \quad = \quad 4 \quad 2 \quad 1 \quad 3 \quad 5$$
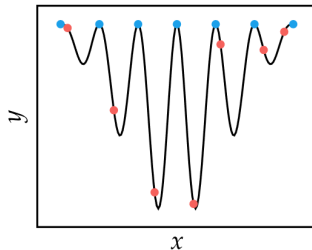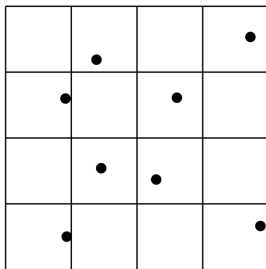
Stratified Sampling

- Each point is sampled uniformly at random within each grid cell instead of the center

- Can capture details that regularly-spaced samples might miss
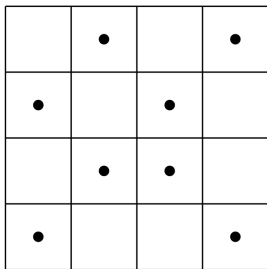
- $f(x)$
- sampling on grid
- stratified sampling

Space Filling Metrics

- A sampling plan may cover a search space fully, but still leave large areas unexplored

- Space-filling metrics quantify this aspect of sampling-plan performance

Example (Uniform Projection)
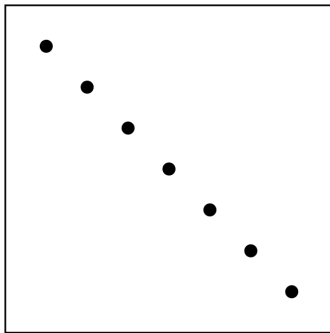
Space-Filling Metrics: Discrepancy

- The maximum difference between the fraction of samples in a hyper-rectangular subset H and that subset's volume

- Often very difficult to compute directly

$$d(X) = \underset{\mathcal{H}}{\text{supremum}} \left| \frac{\#(X \cap \mathcal{H})}{\#X} - \lambda(\mathcal{H}) \right|$$

Space-Filling Metrics: Pairwise Distances
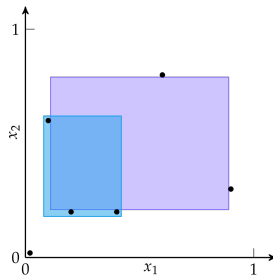
- Method of measuring relative space-filling performance of two-point sampling plans

- Better spread-out plans will have larger pairwise distances

- The pairwise distances of each set are sorted in ascending order

- The plan with the first pairwise distance exceeding the other is considered more space-filling

- Suggests simple algorithm of producing a set of randomly distributed sampling plans, then picking the one with greatest pairwise distances
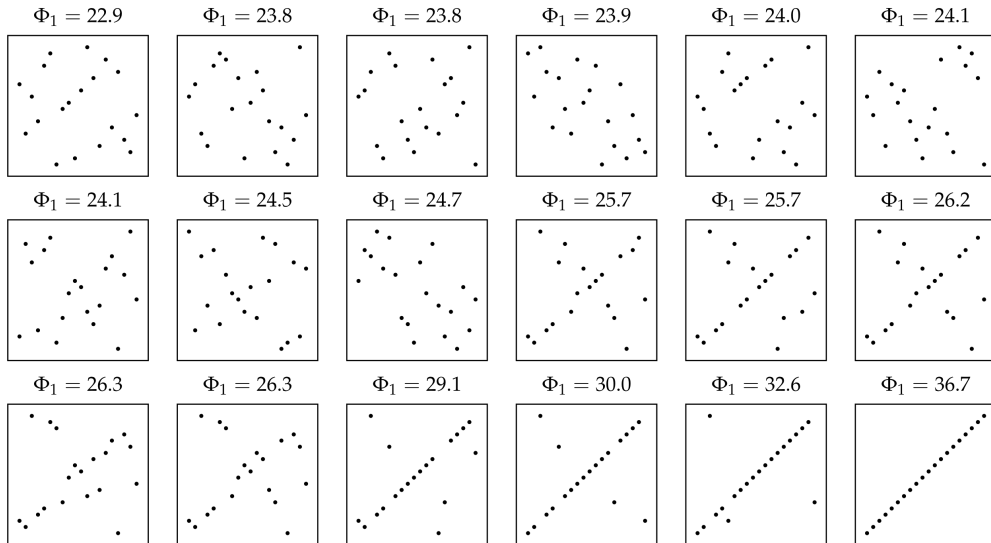
Space-Filling Metrics: Morris-Mitchell

- Alternative to previously suggested algorithm that simplifies optimization problem

$$\underset{X}{\text{minimize}} \quad \underset{q \in \{1,2,3,10,20,50,100\}}{\text{maximize}} \quad \Phi_q(X)$$

$$\Phi_q(X) = \left(\sum_i d_i^{-q} \right)^{\frac{1}{q}}$$

- is the th pairwise distance and is a tunable parameter

Space-Filling Metrics: Morris-Mitchell

$\Phi_1 = 22.9$ $\Phi_1 = 23.8$ $\Phi_1 = 23.8$ $\Phi_1 = 23.9$ $\Phi_1 = 24.0$ $\Phi_1 = 24.1$

$\Phi_1 = 24.1$ $\Phi_1 = 24.5$ $\Phi_1 = 24.7$ $\Phi_1 = 25.7$ $\Phi_1 = 25.7$ $\Phi_1 = 26.2$

$\Phi_1 = 26.3$ $\Phi_1 = 26.3$ $\Phi_1 = 29.1$ $\Phi_1 = 30.0$ $\Phi_1 = 32.6$ $\Phi_1 = 36.7$

Space-Filling Subsets

- Often, the set of possible sample points is constrained to be a subset of available choices

- A space-filling metric for a subset S within a finite set X is the maximum distance between a point in X and the closest point in S using a norm to measure distance

$$d_{\max}(X, S) = \underset{x \in X}{\text{maximize}} \; \underset{s \in S}{\text{minimize}} \; \|s - x\|_q$$
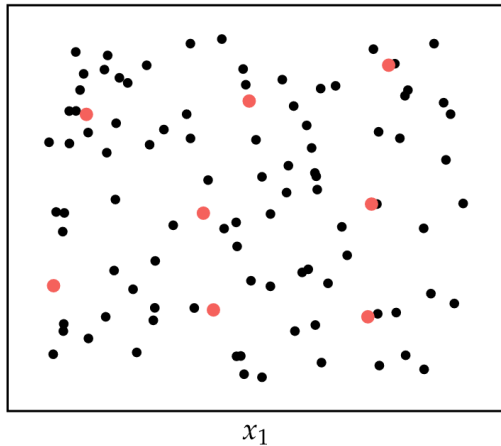
- A space-filling subset minimizes this metric

- Often computationally intractable, but heuristics like greedy search and exchange often produce acceptable results

Space-Filling Subsets
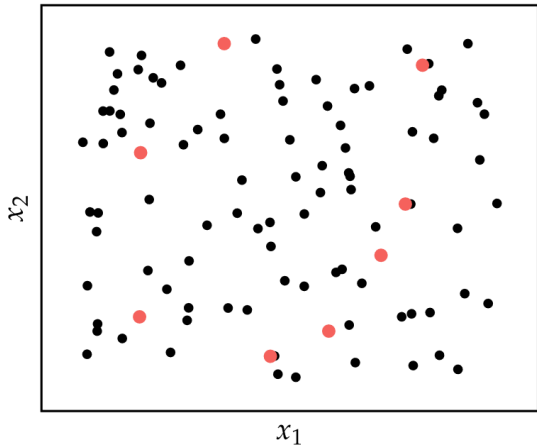
greedy local search   exchange algorithm

Quasi-Random Sequences

- Also called low-discrepancy sequences, quasi-random sequences are deterministic sequences that systematically fill a space such that their integral over the space converges as fast as possible

- Used for fast convergence in Monte Carlo integration, which approximates an integral by sampling points in a domain

Quasi-Random Sequences

- Additive Recurrence: recursively adds irrational numbers

- Halton Sequence: sequence of fractions generated with coprime numbers

- Sobol Sequence: recursive XOR operation with carefully chosen numbers

Quasi-Random Sequences
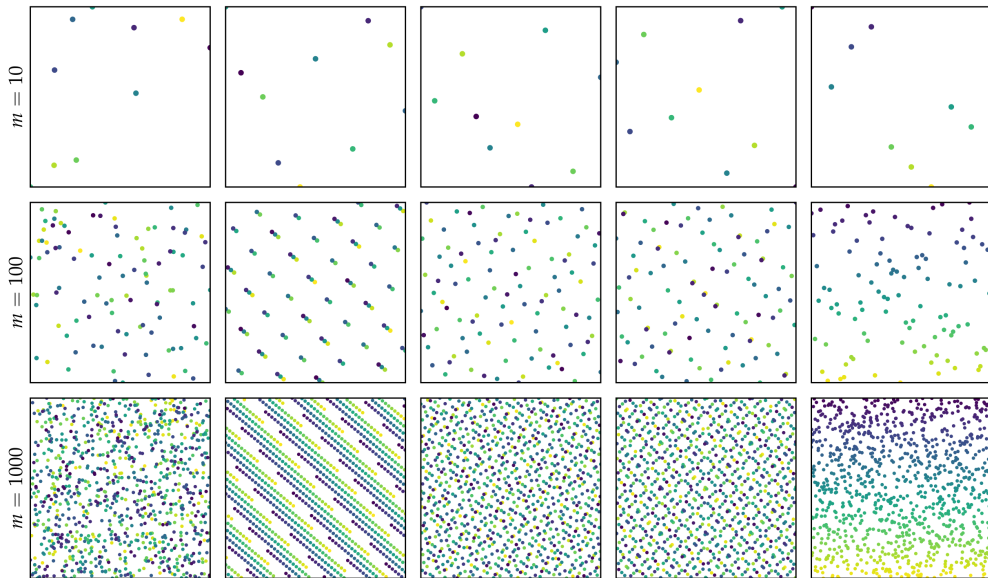
Summary

- Sampling plans are used to cover search spaces with a limited number of points

- Full factorial sampling, which involves sampling at the vertices of a uniformly discretized grid, requires a number of points exponential in the number of dimensions

- Uniform projection plans, which project uniformly over each dimension, can be efficiently generated and can be optimized to be space-filling

- Greedy local search and the exchange algorithm can be used to find a subset of points that maximally fill a space

- Quasi-random sequences are deterministic procedures by which space-filling sampling plans can be generated

- They can be used to create **surrogate models** to optimize in place of computationally costly functions