Pitfalls and Problems in Random Portfolio Generation: Simplified Approaches and Their Uses

London Quant Group
26-2-2008

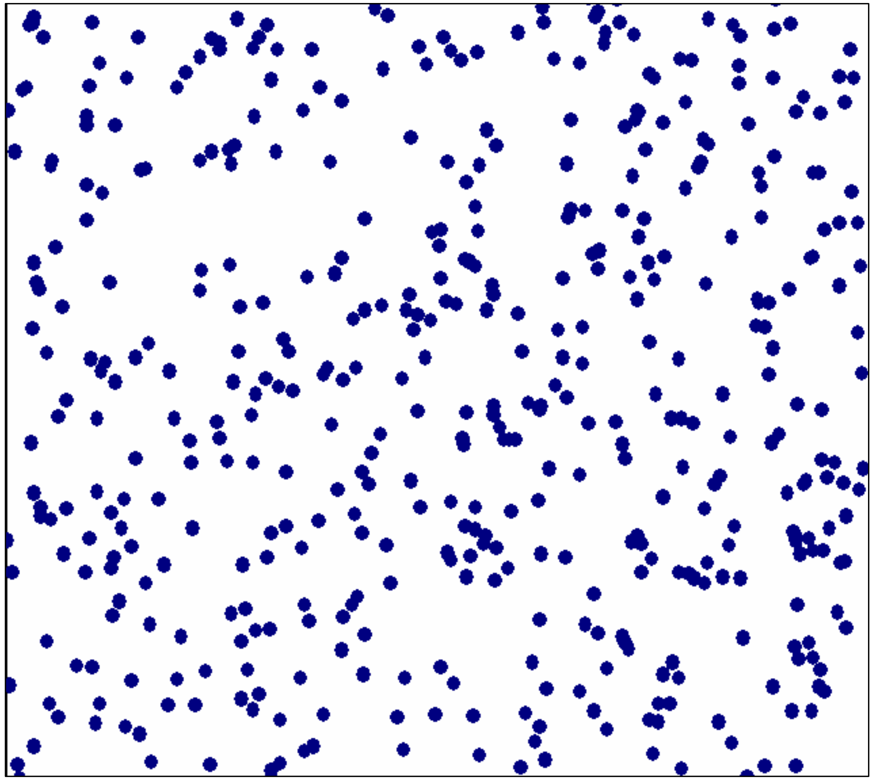
R-Squared Ltd

Richard Young

- Do you need true randomness?
- How can you tell when you have it
  - Example of common problem
- Acceptable (?) approximation techniques
  - Approximate distribution
  - Approximate boundary

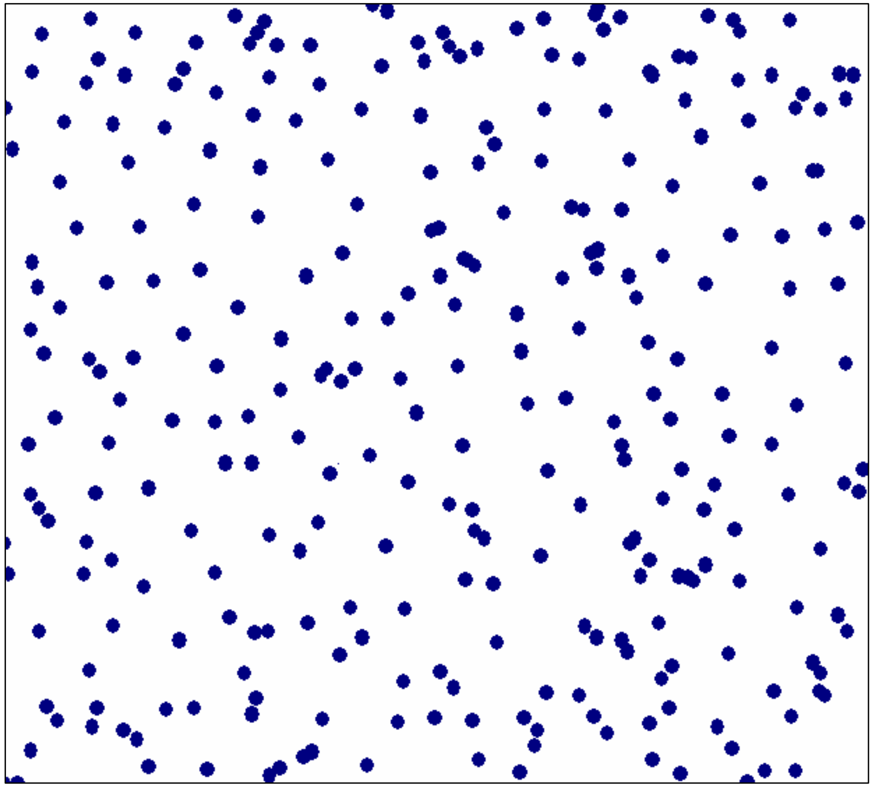
Not A Random Distribution

A



Random[1]

B



Quasi-random[2]

1. Well Pseudorandom actually
2. Well Pseudorandom modified to be like Quasi-random

Quasi-random Monte Carlo

- Well established for approximating high dimensional integrals
- Used in efficient MC methods for option pricing
- Share some properties of true random, but (by design) more efficient at space filling
- Orders of magnitude reduction in number of samples needed, but with the loss of the ability to calculate a confidence interval.
    - Efficiency improvements fall as dimensions increase
- But theory still underdeveloped for hyper-volume sampling (?)

((Ref 3) http://www.puc-rio.br/marco.ind/quasi_mc.html)

- Do you need true randomness?

- How can you tell when you have it
  - Example of common problem

- Acceptable (?) approximation techniques
  - Approximate distribution
  - Approximate boundary

Number of Assets

- Each asset in a portfolio is a new dimension in the search space
- So all portfolios with n-1 holdings are on the hyper-surface of the n-holding search space
- This makes the question of how many there in the portfolio integral to the question of a random distribution
  - Random (normal, equal)
  - Controlled by the constraints

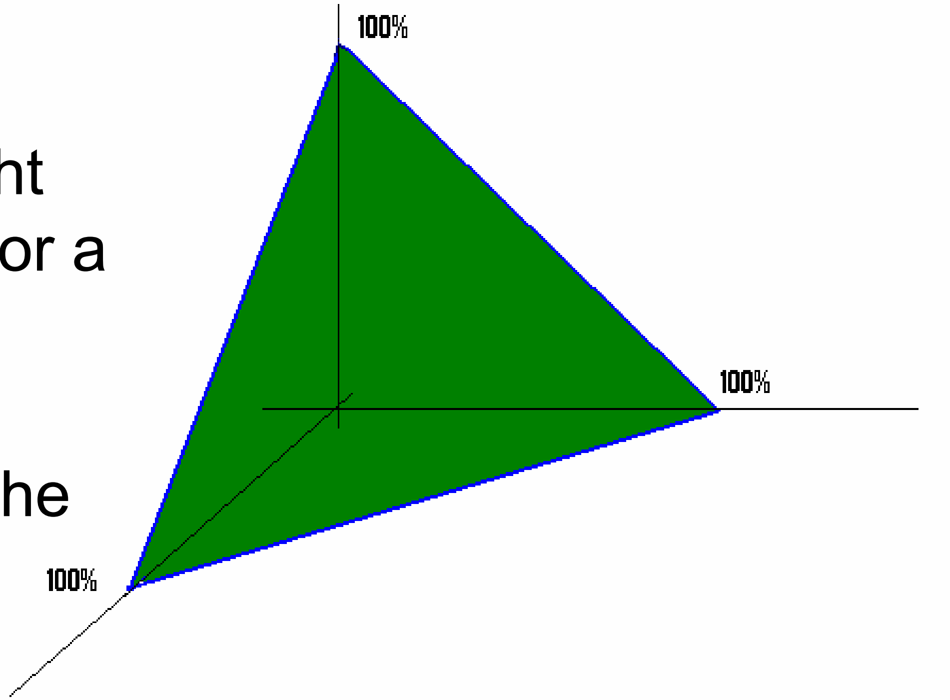
Testing an Algorithm

- It is very easy to describe random portfolios
- But in reality hard to test
  - People are notoriously bad at picking random patterns from non-random
- A true test of any given set of portfolios and constraints is almost as hard as the original problem
  - You need to fully define the boundary
- An approximation can be done by sample – Use a hyper-sphere, its very fast to test if a point is in the interior
  - Creates a sub-problem: how much of the hyper-sphere is outside the constraints
    - To a first approximation it is proportional to the fraction of surface portfolios that fall outside

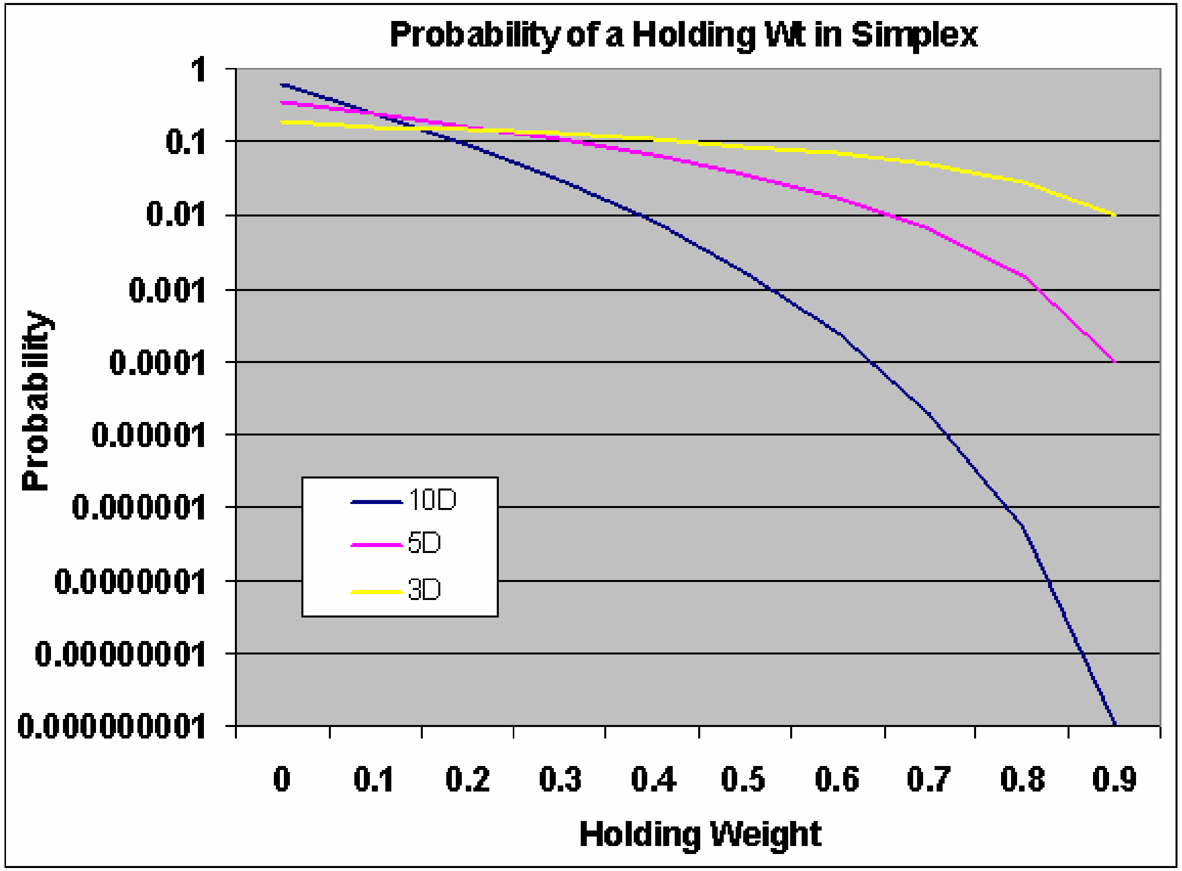
Testing an Algorithm
The Simplex as a Test Space

- Run over a known search space
  - Simplex
  - Probability of holding weight can be exactly calculated for a simplex (Ref 1)

    $P(r_1 < x < r_2) = (1-r_1)^{n-1} - (1-r_2)^{n-1}$
  - Plot, for a selected asset, the generated weights against theoretical



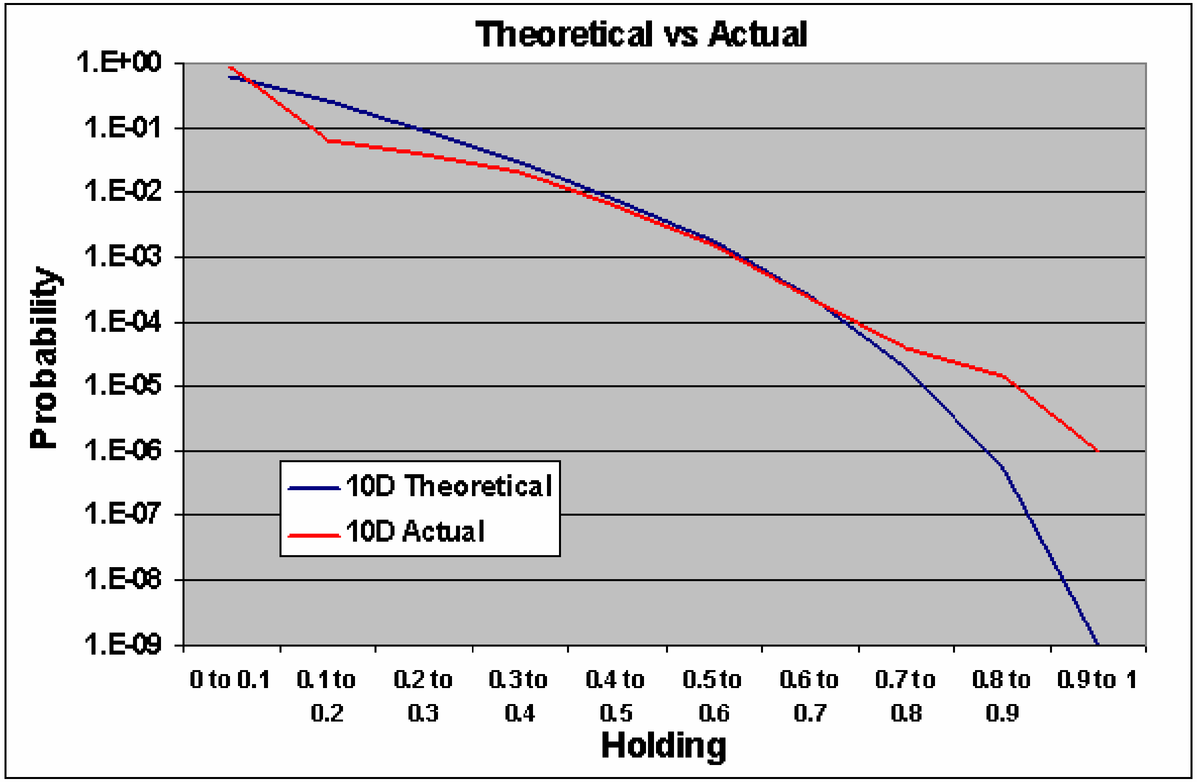
The Simplex as a Test Space



- This is one of the few true tests easily applied in any number of dimensions,
  - Only the portfolio constraint

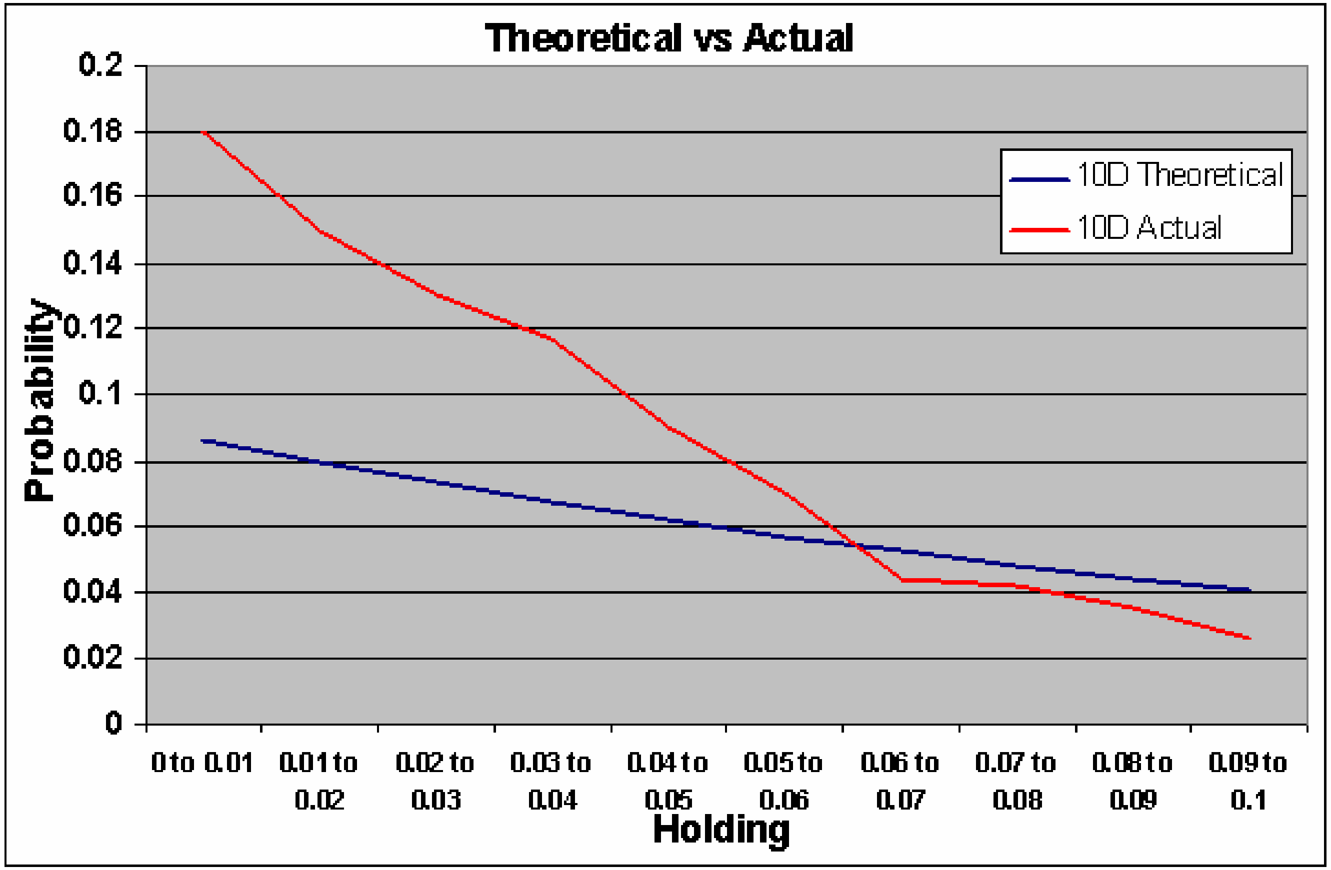
Boundary Hugging

- Common to may approaches
- Particular problem where the constraints are very asymmetric. Tight in some dimensions, wide in others.



Boundary Hugging

- Tends to create distribution bias towards extreme portfolios



- Do you need true randomness?
- How can you tell when you have it
  - Example of common problem
- Acceptable (?) approximation techniques
  - Approximate distribution
  - Approximate boundary

Making True Randomness Easier

- Compromise the accuracy of the constraints to get much better distribution
  - fill space with a set of 'small' sub-spaces, simplexes or hyper-cubes
    - Hyper-cubes lend themselves to quasi-random numbers generated by easily available math libraries (eg NAGs g05yac (Ref 4))
    - Two common approaches to sampling a simplex
      - Equivalent to sampling from a Dirichlet distribution
      - Based on Order Statistics (Ref 2)
  - The smaller the closer the approximation to the constraints, but the more work needed
    - Number of sub-spaces rises VERY fast with number of dimensions
  - Either repeat many times for each sub-spaces
    - True random distribution, but potential expensive
  - Or, fill one sub-space, and translate to the others
    - Much quicker, but NOT random above the scale of the sub-space, this may not matter for most financial problems

Random Walks from the Benchmark

- This is only a feasible short cut if the constraints are loose enough to be beyond the region that you wish to sample
  - Generally requires a well diversified portfolio
- The longer the walk the more likely to reach the boundary.
- Example of limited walk approach
  - Start with the benchmark
  - Choose number of holdings in portfolio, P
  - Remove random portfolio members until P holdings left
  - Pick random weights (say -0.1% to 0.2%) to add to randomly selected assets in portfolio until 100% holding achieved again

Reduce the Dimensionality!

- Group constraints are big saving
  - Eg All French Banks are identical
  - 1600 assets split into 10 regions and 20 sectors
    - One 200D problem
    - and 200 sub-problems with an average of 8 dimensions
- Solve the problem at the top level, then weight holdings within the peer group
  - mini-low dimensional problem, only with just the budget constraint of the region-sector weight and the individual holding constraints
- Not guaranteed random, but does reflect the constraints
- The difficult question here is how many assets within a peer group are used to fill the required holding

Small Dimension Analysis

- Some analysis can be revealing even at the sub-problem level
- Returns attribution
  - S&P500 Stock selection, do within sector.
    - Now 20ish problems with average of 25 dimensions. Far more tractable than one 500D problem
  - Then do sector weight attribution, one 20D problem
- Risk model testing, split into factor and stock specific risk
  - Done at portfolio level each factor in the risk model becomes a dimension

References

1. “Advances in Portfolio Construction and Implementation”, Chapter 7, R. Dawson & R. Young. Edited by S Satchell & A Scowcroft.
2. “Non-Uniform Random Variate Generation” Luc Devroye page 568
3. “An Introduction to Quasi-Random Numbers”, http://www.puc-rio.br/marco.ind/quasi_mc.html
4. NAGs Quasi-random number generator http://www.nag.co.uk/numeric/CL/manual/pdf/G05/g05yac.pdf