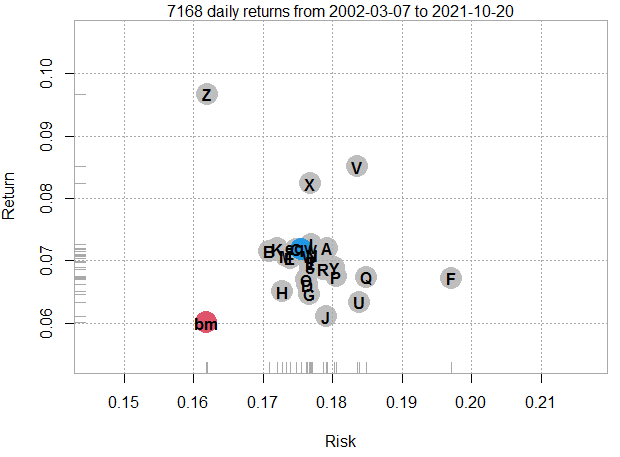
Issue: with increasing n, the cross-sectional distribution of asset scores gets narrower. Counter intuitive: the more funds we analyze, the higher the chance that some will outperform purely due to luck.

What speaks against a random portfolio which has uniform measure over the feasible domain is that in practice, managers device strategies based on certain beliefs which create systematic biases. I.e., those portfolios are a-typical.

Consider a sample of $n$ independent and identically distributed realizations $x = (x\_1, ..., x\_n)$ of a random portfolio $X$ having probability distribution $F$ defined on a space $\mathcal{X}$.

A random portfolio is a real-valued random vector having a probability distribution $F$ defined on a space $\mathcal{X}$.

//Idea: set rebalancing dates to index-review dates, i.e., whenever index composition change (buy-and-hold in between, effectively passive). Use Varsi to compute cross-sectional distribution of portfolio returns. Where is the BM-performance?



RP’s

Introduction – The need for a control

Any research conducted under scientifically acceptable conditions uses experimental methods. The crux of experimental designs lies in the inclusion of a control, i.e., a base case where a certain idea is not applied, where a variable of interest is not manipulated, in order to assess the cause-and-effect relationships and to determine the value or validity of the proposed idea. Such procedure is not part of the standard repertoire in the investment industry. Investment professionals implement all sorts of ideas, but it is rarely clear whether those ideas are indeed the source of their performances. Skill and luck are hard to disentangle.

Rather than by experimental designs, the way performances are evaluated in practice is by comparing the realizations of past profits and losses to those of other managers or products (peer group comparison) or to general market evolution, usually proxied by indices like the Standard and Poor's 500 (benchmark comparison). Such comparisons have their purpose, but this cannot be to declare one strategy better than another because relative performances crucially depend on the analyzed time period and on the selection of benchmark or peer group and without further information, nothing can be said about cause and effect.

The academic responses to the performance evaluation problem take various forms. The most direct approach comes in form of hypothesis tests for the difference in performance measures of two strategies (typically, the portfolio in question and a benchmark) like, for instance, Sharpe ratio tests building upon the work of \cite{bib:JobsonKorkie1981}, the correction of \cite{bib:Memmel2003} and the extension of \cite{bib:LedoitWolf2008} accounting for stylized facts of asset returns. Unfortunately, such tests suffer from low power, i.e., they are unlikely to identify superior performance in the data even if there is \cite{bib:KazakPohlmeier2018}.

Classical holdings- or transaction based performance attribution tools in the line of \cite{bib:Brinson1985} (but applied to equity-only portfolios) trying to identify sources of excess returns (i.e., relative to a benchmark) cannot solve the problem either since they only identify the \textit{Causa Proxima}, i.e., the immediate cause.

Tools are widely (ab-)used in the industry for their explanatory power to, at the same time, outlie the difference in the allocation structure between the portfolio and the benchmark by grouping stocks into easily interpretable categories like countries, sectors or currencies, and to quantify the individual contributions to overall performance coming from the allocation differences among and within the pre-defined categories. The \textit{determinants} of portfolio performance \cite{bib:BrinsonEtAl1986} are however not to be understood in an etiological sense. Unless the categories, according to which the attribution is done, correspond to active bets undertaken by the portfolio manager characterizing the strategy, allocation and selection effects do not \textit{cause} over- or underperformance. They are residuals in the sense that the strategy happens to generate the effects while pursuing another goal or goals only able to identify what one might call the \textit{Causa Proxima}, the immediate cause while the \textit{Causa Remota}, the remote and maybe indirect cause which is somehow encoded in the strategy that actually lead to the performance delta remains unknown.

Typically, attributions are done towards easily interpretable categories. The tools then give answers to questions of the sort: How much did over- or underweighting a certain country, sector or currency contribute to an observed performance difference? The \textit{determinants} of portfolio performance \cite{bib:BrinsonEtAl1986} are not to be understood in an etiological sense. Unless the categories, according to which the attribution is done, correspond to active bets undertaken by the portfolio manager characterizing the strategy, allocation and selection effects do not \textit{cause} over- or underperformance. They are residuals in the sense that the strategy happens to generate the effects while pursuing another goal or goals only able to identify what one might call the \textit{Causa Proxima}, the immediate cause while the \textit{Causa Remota}, the remote and maybe indirect cause that lead to the performance delta which is somehow encoded in the strategy remains unknown.

Say, for instance, that large companies in a given market perform particularly well over a certain period. A capitalization-weighted index will, by construction, perform strongly as well and any portfolio with a negative loading on large firms is likely to suffer from the underweight. While a performance attribution by size would identify the effect, generalizing that company size is indicative of future price developments would be a leap of faith. Albeit, in this case it

in this case, it is indicative, as has been shown repeatedly since the work of \cite{bib:Banz1981} (albeit the effect is reversed to our example, i.e., stocks with lower market capitalization tend to outperform stocks with a higher market capitalization in the future). But one has to be right for the right reasons.

Banz1981 uses an almost apologetic tone when documenting the empirically observed size anomaly because they have no general equilibrium model to explain it. Looking at the cross-section of portfolio returns it seems odd to consider the capitalization weighted benchmark as the average. It does indeed represent the average investor in a given market, but that is far from the average possible portfolio. Far in both, allocation as well as performance.

Does it really surprise that portfolios with a tilt to small companies outperform the benchmark when there are so many more smaller companies than larger ones, i.e., so many more portfolios with low size scores than ones with large size scores?

Let us engineer a silly example to make a point. Say a manager constructs a portfolio proportional to the number of times the letter “Z” appears in the company names of the constituents of the MSCI USA index. Common sense prohibits us from trusting the manager with our savings. But, what if the manager does not reveil his dubious approach? Comparing a backtest of the manager’s strategy over the last $20$ years (from 2002-03-07, the earliest available start date in our dataset to 2021-10-20) to the capitalization weighted parent index shows an annualized outperformance of $3.6$\% (based on geometric returns) and an alpha of YYY with t-statistic of $3.3$. The strategy shows a market beta of $0.9$ and rather small negative loadings on the remaining four factors of the Fama-French $5$ factor model \cite{bib:FamaFrench2015} with significant t-statistics for Value and XXX. Table \ref{tbl:ponzi\_stats} summarizes the descriptive statistics. Note that, despite the economically significant difference in Sharpe ratios between the portfolio and the benchmark of $0.60$ versus $0.37$, the Sharpe ratio test of \cite{bib:LedoitWolf2008}\footnote{The test accounts for time series structures in the data by employing heteroscedasticity and autocorrelation consistent (HAC) estimates of standard error.} barely rejects the null hypothesis of equal Sharpe ratios at the $1$\% tolerance.

What is going on? Is there anything special about the letter Z? Of course not. In fact, it happens to be the case that we could have chosen \textit{any} letter of the alphabet and the simulated out-of-sample backtest would have shown an outperformance and a positive alpha in every single case, and in one quarter of the times the Fama-French regression would even have shown a significant alpha at the $5$\% tolerance.

We could even take the nonsense to the extreme and invert the strategies by investing in all assets except those with a particular letter in the company name. Again, all backtests outperform the benchmark, this time even more pronounced with more than one third of significant positive alphas\footnote{The overlap in the allocations of the \textit{inverse} strategies is of course rather high and one should anyways adjust for multiple testing, but this is not the point here.}. Our comic example is similar to the seemingly paradoxical empirical results presented in \cite{bib:ArnottHsuKalesnikTindall2013} who find that the arguably nonsensical inverses of sensible investment strategies, i.e., strategies built from well-founded investment beliefs, which outperform the capitalization weighted benchmark, outperform even more. The cause is readily identified by a tilt towards the size and value factors meaning that both, sensible and senseless strategies outperform for the same (unintended) reasons. Even randomly generated strategies, so their finding, lead to outperformances for the same causes, letting the authors conclude that “value and size arise naturally in non-price-weighted strategies and constitute the main source of their return advantage” and that, therefore, “a simple performance measure becomes an unreliable gauge of skill”. Indeed, performance analysis must be made in a differentiated way using control. Also our strategies’ outperformances can be rationalized by significant betas for value and size (except for the high flyer “Z” strategy which happens to coincide with a handful of conservative and high-profitability stocks).

But one has to be careful. When dividing the time window into two decades, we observe that the outperformance was accumulated in the first decade while, in the second one, the picture changes drastically with over two thirds of the nonsense strategies underperforming the benchmark. Yet, factor loadings are comparable over the two subperiods. It should be kept in mind that the Fama-French factors are designed as long-short strategies and can therefore only have limited explanatory power for long-only strategies, since the long-only strategy can only exploit an overweighting of the long-leg and cannot exploit the short-leg (other than to avoid it), which often explains a large part of the anomalies\footnote{Add example of Low-Volatility Anomaly exemplified by Blitz on the long-term dataset for the U.S. (highest returns achieved by second or third quintiles), cite book of Blitz. Maybe, add text to argue that one should not readily conclude that, just because a long-short strategy (market neutral) is profitable, a long-only strategy loading on the long-leg would therefore outperform the capitalization weighted benchmark.}.

The point in all this is simple: We need to know what to control for. And for that, we need to know how stock features are distributed cross-sectionally. The distribution of cross-sectional factor scores is determinant for performance analysis for any deviation from the capitalization weighting will load on some characteristics.

The point that we want to make is that the usual statistical economics toolkit is of little service to detect the ponzi scheme (hopefully, good old Due Diligence would uncover the fraud). The main issue lies in the relative perspective vis-à-vis the benchmark which is actually a highly improbable portfolio from a random portfolio perspective.

What is going on here? How can we explain the robust outperformance over all 26 essentially random letters-strategies? Well, we can not. It is coincidence. Looking at the past $10$ rather than $20$ years, the picture changes drastically with over two thirds of the nonsense strategies underperforming the benchmark. Again, what is going on? Has letter-based investing come out of fashion?

We ex-post picked the winning characteristic from the control group.

However, the condition of ones’ lawn should not be secondary to the question: Is it greener on the other side? What good is to know that a benchmark lost 10\% more during a crisis than our own portfolio when in fact, within the restrictions the portfolio has to meet, a loss could have been avoided altogether? The reference point should thus, so we argue, not be a point but a set, i.e., the set of potential outcomes that could have been realized while adhering to the portfolio’s constraints, and performance should be evaluated in terms of the relative location in the domain of the achievable set. We call this achievable set the cross-section of portfolio returns, or more generally, the cross-section of portfolio scores.

Performance analysis is inherently relative. Whether we compare to peers, to a capitalization-weighted benchmark, to alternative indices, to the risk-free rate or to past performances or simply care about the sign of the performance (positive or negative), judgment is always made relative to a reference point.

The benchmark-relative perspective is deeply rooted in economic theory (\cite{bib:Sharpe1964}, \cite{bib:Lintner1965}, \cite{bib:Mossin1966}) and forms the blueprint for classical performance analysis who’s tools try to identify sources of \textit{excess} returns, i.e., relative to the benchmark, and to attribute them to active bets undertaken by the portfolio manager. The primary goal of the tools is explanatory in nature aiming to unravel the determinants of portfolio performance.

This is done either from a holdings- or transaction based perspective going back to \cite{bib:Dietz1966}, \cite{bib:BAI1968} and most famously to \cite{bib:Brinson1985} and \cite{bib:BrinsonEtAl1986}, asking questions like how much did over- or underweighting a certain country, sector or currency contribute to an observed performance difference? Or, from a return based perspective which uses regression on a set of factors which are recognized to explain security returns and are associated with a positive premium in order to assess whether the active performance is attributable to a particular investment style.

While the Brinson-type of performance attribution is purely descriptive, the regression setup goes beyond simple performance measurement as it allows for statistical inference in terms of significance testing of the out- or underperformance (alpha) and the loadings on other return drivers (betas) and therefore provides a basis for normative conclusions (assuming that the usual assumptions for linear regression are met). A skilled manager should be able to beat the market (factor) in a statistically significant manner after accounting for alternative betas. The skilled manager is expected to generate alpha with significant t-statistic. Hence, skill is reserved to the active manager who is able to harvest a positive return premium which cannot be explained by a known factor. Any benchmark replicating strategy, since it involves no active deviation from the capitalization-weighted allocation structure in a market, is therefore called passive and comes with performance expectations which should match the performance of a market indexes rather than trying to outperform it.

The regression approach poses a peculiar problem in that it controls for known factors. But what about unknown factors? Is skill but an indication of confounding variables? Or what about strategies based on Machine Learning (ML) algorithms which exploit non-linear relations between known factors and future returns. Are such skillful because they can not be explained be the linear model? A more appropriate control for such systematic strategies would be to generate a representation of the potential performance results over the parameter space of the ML model and investment constraints it needs to adhere to.

A more direct approach of performance analysis comes in form of hypothesis tests for the difference in performance measures of two strategies (typically, the portfolio in question and a benchmark). For instance, Sharpe ratio tests building upon the work of \cite{bib:JobsonKorkie1981}, the correction of \cite{bib:Memmel2003} and the extension of \cite{bib:LedoitWolf2008} accounting for stylized facts of asset returns. However, at least since \cite{bib:KazakPohlmeier2018} it is known that those tests suffer from low power, i.e., they are unlikely to identify superior performance in the data even if there is.

Irrespective of the statistical testing procedure, even if a powerful test can be found, issues remain with the limited control provided by benchmark, peer group or factor strategy comparisons since they are selected by the experimenter and necessarily introduce biases. Ideally, the control should be randomized.

A random portfolio forms an obvious choice of control as by design, it incorporates no penchant to any strategy; any one portfolio structure is just as likely to occur as any other. A random portfolio should represent the set of counterfactual results, i.e., the entirety of performances that could have been achieved over a certain period given a universe of investable assets subject to investors constraints.

In principle, this allows for a description of the probability distribution of cross-sectional performance measures under the null hypothesis of no skill which provides a means to statistically test the benefit of having a certain strategic portfolio feature.

Any approach to portfolio selection deemed skillful should then systematically outperform the bulk of random allocations, at least if outperformance is measured in terms of the objective seeked for by the strategy. Instead of comparing a particular investment strategy to the average investor in a certain market (i.e., a capitalization weighted market index) or to competitor products, performance analysis is vis-à-vis an exhaustive control. Of course, this is in a first instance a theoretical construct. The practical challenge is how to sample from the population of possible portfolios.

%In the most general formulation, random portfolios should be the outcome of a uniform sampling on the domain of feasible allocations. The feasibility and complexity of such a task depends on the dimensionality of the asset space and on the shape of the region induced by the constraints. In the following, we will present several solutions and solution procedures that are computationally tractable, each tailored to a specific problem type. Exact uniformity will not always be attainable but this is not per se a problem because our interest lies in a pragmatic application where near-uniformity may be gauged sufficient for certain statistical inference.

Clearly, the idea of creating random portfolios for performance evaluation is not new. \cite{bib:CohenPogue1967} used random portfolios in the analysis of mutual fund performances more than half a century ago and defended their approach against criticism in \cite{bib:CohenPogue1968}. Ever since the claim of the economist Burton Malkiel \cite{bib:Malkiel1973} that “a blindfolded monkey throwing darts at a newspaper’s financial pages could select a portfolio that would do just as well as one carefully selected by experts” random portfolios have been promoted to probe investment skill. Most prominently, The Wall Street Journal's Dartboard Contest, a monthly column published by the business newspaper between 1988 and 2002, put Malkiel's claim to the test by letting their staffers (acting as the allegoric monkeys) literally throw darts at a stock table, while investment experts picked their own stocks, always for a holding period of six months\footnote{Prior to January 1990, the holding period was for one month. The extension to six months was made to alleviate a possible bias from the price pressure resulting from the announcement effect.}.

If nothing else, the game added another animal symbolism to the jargon at Wall Street, emblematic in the ongoing debate on active versus passive management and the underlying hypothesis on market efficiency.

However, the result of the game is less interesting than the experimental setup, or, the deficiencies of the informal setup. Several academic studies have examined the game pointing out biases like expert’s tilt towards high risk stocks (\cite{bib:MetcalfMalkiel1994}), low dividend\footnote{Performances were computed on price series rather than on total return series, thus ignoring the effect of dividends and arguably incentivizing professionals to pick stocks with high growth opportunities.} yield stocks (\cite{bib:Liang1999}) and high momentum stocks (\cite{bib:PettengillClark2001}), questioning the initial finding of \cite{bib:BarberLoeffler1993} that observed short-term price increases after the announcement would result from information content (rather than price pressure). Relevant to our contribution is the approach of \cite{bib:Liang1999} to use the bootstrap technique cite{bib:Efron1979}, a randomized inference tool based on sampling, for performance analysis. As we will see, random portfolios in the spirit of \cite{bib:CohenPogue1967} or the dartboard portfolios is essentially a bootstrap exercise. Or otherwise put, the bootstrap distribution of portfolio returns constitutes a (rather naïve) form of a random portfolio.

The inconclusive results of the academic analysis of the dartboard game highlights the difficulty and inherent biases of simple performance comparisons. Discrepancies in conclusions usually emerge from methodological discrepancies. In a different but related context for instance, that of mutual fund performance evaluation, \cite{bib:KosowskiEtAl2006} and \cite{bib:FamaFrench2010} also leveraged on the power of randomization as a control for luck via bootstrap methods. Both studies, although investigating nearly the same dataset and employing similar Monte Carlo procedures, managed to find contradictory results. Thanks to the replication of \cite{bib:HarveyLiu2021} and their dissection of the employed methods the different outcomes have become understandable.

%the game has not met scientific standards, thorough studies have analyzed the contest in detail \cite{bib:BarberLoeffler1993}, \cite{bib:GreeneSmart1999} and

Section \ref{sec:literature\_review} gives a survey of the literature on random portfolios for performance evaluation. Although the topic is an old one, the list of articles devoted to it is rather short, which shows that the use of random portfolios for performance analysis is not standard. Furthermore, there is no standard definition of what a random portfolio should represent and how it should be constructed. Contrary to common belief, we argue that, under a rigorous definition, sampling from a random portfolio is not trivial.

Alternatively to forming random portfolios bottom-up by sampling weights, a common practice is to synthesize the price or return paths of a hypothetical random portfolio directly. However, such Monte Carlo simulation necessitate an assumption about the underlying stochastic process and thus disqualify as a model-free control.

In what follows we aim to give a detailed account of the details and methodological possibilities to devise a randomized control via a random portfolio. It is not our goal to advocate random portfolios as a panacea to performance measurement problems. Rather, we see it as an addition to the statistical economics toolkit of performance analysis instruments.

The remainder of this article is structured as follows. Section \ref{sec :generating\_random\_portfolios} gives a first glimpse into the technical aspects of random portfolio generation and defines the general problem statement as an algorithmic sampling exercise. Section \ref{sec:literature\_review} then documents the existing literature on the topic distinguishing between finance specific publications and research on volume computation by means of randomized sampling procedures, which (overlooking the semantic issues) can be applied one-to-one in finance. With section \ref{sec:deep\_dive} we then dive deeply into the topic of geometric random walk based sampling giving an overview of potential algorithms together with a description of their functioning and properties. Aspects regarding their programmatic implementation are treated in section \ref{sec:implementation}. Finally, in section \ref{sec:XXX}, we run empirical experiments using random portfolios, constructed with the aid of state-of-the-art techniques from Computational Geometry, applied to performance evaluation in global equity markets.\\

Concentration of measure

Section on fallacies of random portfolios like turnover 🡪 converges to constant because of concentration of norm phenomena \citeA{bib:AggarwalHinneburgKeim2001}. Can be handled by inclusion of L1-norm constraint. Diamond shape, absolute value constraint. Can be linearized but doubles the dimensionality of the problem.

Cross-section of portfolio returns 🡪 Distribution of portfolio returns under the Null of no skill \cite{bib:Lisi2011}. Not really so. It’s more the posterior of the mean of that distribution. Ok if we compare the mean of bunch of quant strategies to that distribution. But comparing one particular strategy to it seems unfair. Take the BM as an example. Almost surely an extreme tail event. Then, which tail depends on the market regime and the properties of the random portfolio characteristics (distribution of factor scores) and the typical performance pattern in either regime (bear / bull).

However, both approaches fail to capture the essence and inherit value of random portfolios as a model-free control. Monte Carlo simulations typically synthesize the price or return paths of a hypothetical random portfolio, thus necessitating an assumption about the underlying stochastic process. Although this is overcome by the random stock selection approach chosen by the dartboard game, thus allowing for an agnostic position with respect to (w.r.t.) the data generating process, random stock picking forms no acceptable control for it fails to account for real-world constraints that strategies built from theory typically need to satisfy. This is where previous work on random portfolios fell short and may explain why the great break trough has never materialized.

%However, despite the ubiquitous academic literature on alpha potential, many of which, nota bene, are empirical in nature, the aspect of a proper randomized control is missing across the board.

%Rather than as a random stock picking excercise (where hypothetical amounts are invested equally among the randomly selected firms), random portfolios should be understood as a random sampling from the feasible set of portfolio allocations. And importantly, the stochastically generated portfolios should obey the same constraints as the strategy built from theory.

%Even more should it surprise that the financial literature does not offer a ready solution for anyone wanting to conduct a comparison of an investment approach not only to a benchmark or peer group, but to a set of stochastically generated portfolios that are unbiased by design and contain no investment style, yet respect the same constraints as the strategy in question.

%\textit{cite papers on active management}. \textit{Maybe mention rise of ETF, Vanguard}

%https://econs.online/en/articles/economics/heads-or-tails/

%although their experiments have not always met scientific standards

%http://www.investorhome.com/darts.htm

%effectively disappears if you (1) account for the fact that the pros pick relatively riskier stocks and (2) measure returns from the day after the column appears (thereby eliminating the announcement effect).

First, close to no financial literature exists treating the mathematical challenges of generating random portfolios, which are in effect considerable. Everyone intuitively understands that random portfolios should obey certain properties like being equally likely to occur and being composed of components that are limited in some way to make economic sense, but it is not a priori clear how one should actually build them. Technically, what we search to do boils down to uniform sampling over some constrained domain, i.e., finding a set of uniformly distributed points in an n-dimensional space bounded by k linear or nonlinear constraints without knowing the boundaries of the solution space. This is a problem in its own right for which no general solution exists. Interestingly, although originally a well established problem in mathematical programming, most recent advances come from the field of bioinformatics and the modelling of metabolic networks. Overlooking the semantic issues….

%Current text treats the mathematical derivation, programmatic implementation and empirical application of randomized algorithms for near-uniform sampling from the feasible set. The feasible set is defned as the solution space to a portfolio optimization problem consisting of all possible values of the choice variables that satisfy the problem’s constraints. The complexity of such a sampling procedure depends on the number of variables, the combined size of domains of variables, and the structure formed by the constraints. I will restrict analysis to convex sets that can be expressed as a bounded intersection of a fnite number of closed halfspaces and/or ellipsoids (i.e. linear and quadratic constraints). The fundamental geometric bodies of study are thus polytopes and ellipsoids with the characterising property of being defned in high dimensional Euclidean space.

%After an optimal portfolio is located, remaining stochastic portfolios uniformly spreading the entire feasible set can be used for performance evaluation as by design, they do not display any penchant to any one portfolio strategy. This idea that random portfolios form an optimal control for performance analytics has long been recognized. Yet, to the best of my knowledge, none of the designs found in the literature were setup to uniformly cover the full range of candidate portfolios. Being able to do so would correct the inevitable biases encountered in traditional peer group comparisons and allow for proper hypothesis testing.

\section{Generating random portfolios}

\label{sec:generating\_random\_portfolios}

\subsection{A first glimpse}

\label{sec:a\_first\_glimpse}

The task of generating random portfolios begins with a definition of the asset space and the definition of randomness.

Maybe the most intuitive case to begin with is when asset weights are non-negative and sum to one\footnote{If weights sum to less than one, a portfolio is said to be under invested. When the sum exceeds one, then the portfolio includes a leverage.}, which are the two defining properties of long-only portfolios. H

An appropriate sample space for compositional data, i.e. non-negative bounded-sum data, is the simplex, denoted

The geometric body which describes the space of long-only portfolios is given by the standard simplex

$\Delta{n-1} = \{x \in \mathds{R}^n} | \sum\_{i=1}^n = 1 \}$

Typically, mutual funds are bound by financial markets regulation to always be fully invested and not to short sell (i.e., negative weights are not allowed). For those managers, the simplex constraint is a legal restriction which they cannot overcome. Since this is a very common case we will restrict all further analysis to it and require that the asset space will always be given by the standard simplex or a subset of the standard simplex induced when adding further constraints. However, the routines which we propose later on to generate random portfolios are by no means limited to the simplex case but readily extend to long-short applications.

As for the definition of randomness, we begin with the (arguably) most intuitive case of a uniformly distributed random portfolio, i.e., we focus on the uniform measure over the asset space. Hence, in the most basic case a random portfolio is defined as $\omega \sim U(\Delta)$. In words: A real-valued random vector uniformly distributed over the simplex.

As we will see, this very particular setup allows for an exact description of portfolio characteristics like the distribution of returns of a random portfolio which therefore makes it possible to infer conclusions purely analytically. In practice however, the simplex condition is usually not the only constraint that asset managers have to respect and once further constraints are imposed, analytical results are hard to come by. The solution is to use Monte Carlo methods, i.e., to sample realizations from the theoretical random portfolio and to base inference on numerical methods. The complexity of such algorithmic inference depends on the type of constraints one wishes to account for. Often, additional constraints coming from the regulators aim to limit (risk) exposure towards a single asset issuer or groups of issuers\footnote{Maybe the most common such limit is the UCITS 5/10/40 rule, stating that no single asset can represent more than 10\% of the fund's assets and that holdings of more than 5\% cannot, in aggregate, exceed 40\% of the fund's assets. For further information see <https://www.esma.europa.eu/databases-library/interactive-single-rulebook/ucits>.}, thus imposing upper bounds on the asset weights individually or collectively. The geometric representations for such linear restrictions are hyperplanes which slice the simplex to become a polytope. Other constraints may be self-imposed by the manager, for instance to prevent overly concentrated portfolios (to regularize the weights), to avoid large deviation from a benchmark in terms of weights, e.g., by setting country or sector lower and upper bounds relative to the benchmark allocation or in terms of variation of the return difference (tracking-errror), to limit transaction costs or turnover from a previously held allocation during a rebalancing of the portfolio or to control for certain portfolio characteristics like factor exposure, sustainability criteria, risk metrics, etc.). The decisive aspect of such constraints lies in it’s mathematical characteristic (linear, quadratic, convex, non-linear, etc.) Certain risk measures like Value-at-Risk (VaR) are non-linear, meaning that, if a random portfolio is subject to a maximum VaR constraint, it’s domain can no longer be represented by a common geometric body, which makes uniform sampling extremely hard if not impossible. Other measures of risk like variance or tracking error are quadratic and geometrically form ellipsoids. Sampling from ellipsoids is rather simple but if the simplex condition also has to be met, one has to sample from the intersection of the simplex with an ellipsoid, which is far from trivial (but we will get to that).

Before losing track in the abundance of possible constraints, let us resort to the base case where the asset space is given by the standard simplex and increase complexity afterwards step-by-step. Taking draws from a random portfolio in the base case means sampling uniformly from the simplex, which is straight forward. Although, one cannot just sample the marginals uniformly from the interval $(0, 1)$ and then standardize the vector to sum to one as this would produce a biased distribution with a peak at the centroid of the simplex and additional peaks at the center of the edges. However, the simple adjustment $\omega\_i = \frac{log x\_i}{\sum\_{i=1}^n log x\_i}$, where $x\_i$ are independent uniform random variables on $(0, 1)$, solves the problem \cite{bib:Devroye1986}.

Alternatively, the Dirichlet distribution with flat parametrization gives samples which are uniformly distributed over the simplex. The Dirichlet distribution is said to be flat when all parameter elements $\alpha\_i = 1. To sample from a Dirichlet distribution with general parameter vector $\alpha = (\alpha\_1, \alpha\_2, …, \alpha\_n)$ it is enough to sample the marginals from a Gamma distribution with shape parameter $a = \alpha\_i$ and fixed rate parameter $b = 1$ and then standardizing by the sum $\omega\_i = \frac{x\_i}{\sum\_{i=1}^n x\_i}.

The Dirichlet distribution can be obtained by normalizing a set of independent, equally scaled gamma random variables. Formally, if $W\_i \sim Ga(\alpha\_i, \beta)$, $i = 1,..., n$, then $X = C(W) ∼ Dir(\alpha)$ where $C$ means the closure operation. From: “Distributions on the Simplex Revisited”

The Dirichlet case, being a parametric model, has the nice feature that we can very easily change our understanding of randomness and deviate from the uniform measure simply by varying the parameter vector $\alpha$. Other distributions on the simplex could also be considered, in particular, generalizations of the Dirichlet family or the logratio normal family which are the two main ways to describe distributions for random compositions.

W.S. Rayens, C. Srinivasan, Dependence properties of generalized Liouville distributions on the simplex. J. Am. Stat. Assoc. 89(428), 1465–1470 (1994)

However, the flat Dirichlet model is of particular relevance in the context of random portfolio modelling since it allows for an analytical description of the portfolio distribution. Although, this is not due to the Dirichlet pdf, but because it has uniform distribution over the simplex. \cite{bib:CalesChalkisEmiris2021} have shown that the distribution of certain characteristics like the mean or a factor score of a random portfolio which has uniform distribution over the simplex can be computed exactly by solving the geometric problem of computing the ratio of volumes of the simplex sliced by a hyperplane over the volume of the entire simplex. The normal of the hyperplane is defined by the vector of single asset characteristics (i.e., the cross-section of asset characteristics) like the mean vector or the vector of asset’s factor scores. The distribution is then found by successively shifting the hyperplane over the simplex and re-computing the ratio of volumes. Note that, since the geometric representation involves hyperplanes, the procedure is limited to find the distribution of linear characteristics. \cite{bib:CalesChalkisEmiris2021} resort to an exact procedure initially proposed by \cite{bib:Varsi1973} and reviewed by \cite{bib:Ali1973} that computes the ratio of volumes in $O(n^2)$ steps. \cite{bib:CalesChalkisEmiris2021} note that a closed form solution to the geometric problem would exist, pointing to the work of \cite{bib:Lasserre2015}, but that the analytical solution is numerically unstable for $n>20$ and therefore not applicable for problems of practical relevance. \cite{bib:CalesChalkisEmiris2021} further derive a recursive algorithm to compute the moments of the distribution. Hence, for a otherwise unconstrained random portfolio, the distribution of its’ characteristics is available exactly, thus allowing for exact inference. The solution is however limited to the uniformly distributed random portfolio.

Building upon another geometric argument, \cite{bib:Bachelard2022} have generalized the computation of exact (central) moments of the distribution of a characteristic (defined by the inner product of a data vector (like the cross-sectional asset mean) and a random weights vector) based on the expected norm of the weights vector. The generalization was proven to hold for weights following a Dirichlet distribution with arbitrary $\alpha$. Given the exact moments, \cite{bib:Bachelard2022} suggest to use an expansion like that of Cornish-Fisher to arrive at an approximate but deterministic distribution. The context in \cite{bib:Bachelard2022} is a purely statistical one, that of exact bootstrap estimation, yet the author shows that the distribution of the random portfolio proposed in \cite{bib:CalesChalkisEmiris2021} is equivalent, technically, to the limiting\footnote{Limiting is to be understood in the sense that the bootstrap procedure would be run forever} distribution of the Bayesian bootstrap of \cite{bib:Rubin1981}. The author further proposed a grid structure as a geometric representation of the classical bootstrap procedure of \cite{bib:Efron1979} and variants thereof (like the $m$ out of $n$ bootstrap) and showed that the geometric properties of the weight space allows for an exact computation of the moments (of the distribution of a linear statistic) also in the discrete case. Their results translate one-to-one to the concept of a random portfolio with the distribution of portfolio characteristics corresponding to the limiting classical bootstrap distribution of the characteristics seen as a sample statistic written as a weighted sum where data are fixed whereas weights are random. This brings us back to the design of the dartboard game as the data generating process of a random portfolio.

Dart throwing monkeys are easily digitalized by drawing counts $c = (c\_1, ..., c\_n)$ from a Multinomial distribution $\text{Mult}(\pi\_1, ..., \pi\_n)$ with $n$ trials where all $\pi\_i = 1/n$. Normalizing each count produces weights $\omega\_i = c\_i / n$. Unlike for the Dirichlet case, the asset weight space is not given by the canonical simplex but consists of a discrete grid $G$ over the canonical simplex with nodes representing the outcome of the dart-based portfolio selection process. The node at the centroid of the simplex forms the equally weighted portfolio and vertices representing single-asset portfolios. Grid nodes located on the boundary of the simplex form sparse portfolios, i.e., allocations where one or more position is exactly zero.

Depending on the specifications of our version of the dartboard game, i.e., how many darts are thrown at a list of how many company names, and whether a name can be hit multiple times or not (i.e., sampling with or without replacement), the grid may look differently and importantly, the probability of activating a particular grid node can change. We say that a particular grid node is activated when the dart-based sampling procedure results in a portfolio allocation that corresponds to the (Euclidean) coordinates of that particular grid node. Computing the node-activation probabilities is a combinatorial exercise.

\textit{Discuss combinatorial node-activation probabilities?}\\

Abstracting from the interpretational aspects, Bayesian posterior distribution of the mean parameter of a random portfolio in the Dirichlet case or the approximate sampling distribution of the sample mean of the random portfolio in the classical bootstrap case, inferential differences are known to be small. However, the bridge to the bootstrap interpretation poses a question mark on the widely adopted view that a random portfolio would encompass all possible allocations and corresponding performance outcomes, i.e., “opportunity set”. While possible theoretically, the geometric properties of the sampling space, be it the simplex or the grid over the simplex, are such that it is overwhelmingly unlikely to draw an allocation which is not close (in terms of

$L\_p$ norm) to the equally weighted one. As such, the opportunity set forms a narrow bound around the equally weighted portfolio. In the continuous case, this has to do with how volume is distributed in the simplex, or more specifically, how volume concentrates with increasing dimension. In the discrete analog, it is due to a concentration of measure phenomenon.

The remedy is to define a distribution with more mass towards to boundaries of the simplex. In the Dirichlet case, this is easily achieved by letting $\alpha \leftarrow \infty$, whereas in the discrete case, the solution is to induce an altered grid structure corresponding to an $m$ out of $n$ bootstrap scheme with $m < n$. As shown in \cite{bib:Bachelard2022}, the two are connected by a mapping of the parameters $m$ and $n$ of the frequentist bootstrap to the parameter of the symmetric Dirichlet distribution implied by the Bayesian bootstrap.

In the Wall Street Journals’ dartboard column, the experimenters chose to operate four darts, i.e., $m=4$, while the list of stocks contained all ticker from the New York, American and NASDAQ exchanges which, according to \cite{bib:Leung1999} amounted to $n=8000. Clearly, under such a specification, the distribution of the weights of a random portfolio has nearly all mass concentrated on one of the components, i.e., the corners of the simplex\footnote{A setup with $m=1$ would result in a distribution of the weights with peaks around the $n$ possible Dirac delta distributions centered on one of the vertices of the simplex.}. Consequently, the distribution of the random portfolio characteristics is very similar to the distribution of the assets’ characteristic. Testing for superior stock picking ability in such a setup therefore poses a great challenge as

The research question posed by the Wall Street Journal, whether professional analysts would defeat chance in the stock picking challenge, was tackled with an ill-designed experimental setup. What can we learn from a

Compare distribution of asset scores to the portfolio score. E.g. 68% of single stock returns exceed the cap-weighted BM return. What does that mean? Only 45% exceed eqw-portfolio.

Add example of alphabet portfolios – distribution of RP means

As such, the distribution of a characteristic of a random portfolio like its mean is equivalent to the posterior distribution of the parameter. This should set warning bells ringing as it runs counter to the intuition that a random portfolio should cover all outcomes from all possible portfolio allocations. Clearly, not all outcomes will be covered.

\subsection{General Problem Statement}

\label{sec:general\_problem\_statement}

The fundamental problem of study underlying the application of generating realizations from a random portfolio is equivalent to the problem addressed in the seminal paper \cite{bib:Smith1984}: Given a bounded $k$-dimensional body $K \subset \mathds{R}^d$, where $k \leq d$, find a way to efficiently sample pseudo-random points $(X\_1, X\_2, ..., X\_n) \in K$ such that $\mathds{P}(X \in A \subset K) = V(A) / V(K)$ with $V$ denoting the $k$-dimensional content of $K$. While \cite{bib:Smith1984} considered the general case of sampling from a generic surface $S$, here, we restrict analysis to sampling from $K$ which is either a polytope $\mathcal{P}$, an ellipsoid $\mathcal{E}$ or their intersection.

Before presenting his generic solution to the problem \cite{bib:Smith1984} elaborates three conventional approaches which he calls transformation, composite and rejection techniques, and explains their merits and shortcomings. We briefly summarize his insights here because the methods are valid candidates to generate samples from a (exactly) uniformly distributed random portfolio, although clearly limited to low-dimensional setups.

%The challenges in finding an applicable sampler to generate random points in $K$ lie in transformations, intersections and efficiency issues: can we transform the geometric bodies to an affine subspace in order to improve the acceptance rate of a sampler? Is there an explicit form for the intersection of several bodies? Can we effectively sample from $K$ when the dimension of $K$ is large, i.e., $10^2 < d < 10^4$? \\

The \textit{transformation} technique maps uniformly distributed points from a hypercube $H$ with a smooth deterministic function $T$ onto $K$ where $T(x)$ has to preserve uniformity\footnote{A necessary and sufficient condition for this is that the Jacobian of $T$ is constant over all $x \in H$. \cite{bib:Smith1984}}. In principle, this is a highly efficient method since there exist very efficient pseudo-random number generator to sample from $H$. The problem is just that $T$ is known only for a very limited class of regions $S$ like spheres or simplices, not so though for polytopes. Theoretically, one could partition any bounded polytope into a finite union of simplices, and then apply the transformation from the hypercube to each simplex yielding the \textit{composite} technique. Although conceptually sound, the complexity of identifying the simplices is generally such that the approach is not tractable computationally.

The idea behind the \textit{Acceptance-rejection} technique is to first find an enclosing set $D \supset K$ for which efficient sampling algorithms exist, take samples from $D$ and either accept them if they lie within $K$ or reject them otherwise. Accepted points will be uniformly distributed in $K$ because if a point is uniformly distributed within $D$ then it is conditionally uniformly distributed in $K$ given it lies in $K$. The problem here is that even when $D$ is chosen based on certain optimality conditions like the smallest enclosing sphere, the number of trial points in $D$ needed to get a point in $K$ grows, as Smith calls it 'explosively'\footnote{As an example, \cite{bib:Smith1984} shows that when $K$ is a hypercube and $D$ is a circumscribed sphere the expected number of points generated in $D$ needed to find one in $K$ grows from 1.5 for $k = 2$ to $10^{30}$ for $k = 100$.}.\\

A further alternative (not mentioned in \cite{bib:Smith1984} which is of high theoretical relevance since it allows for exact uniform sampling, requires full identification of the boundary of the set. Let us consider the case when $K$ is a convex polytope: $K = \{\omega \in \mathds{R}^d | A\omega \leq b, A'\omega = b'\}$. This representation by the linear system is called H-representation. Alternatively, one may represent polytopes in terms of their vertices $V = \{v\_1, v\_2, ..., v\_n\}$, which is then called V-representation. If one can solve the vertex enumeration problem, i.e., determining the set $V$ of all vertices as a function of $A, A', b$ and $b'$, then samples from a uniform portfolio having domain $K$ would be simple linear combinations for the vertices. Hence, this may be the preferred sampling method if one is only exposed to a small number of linear constraints and as long as the number of assets is small ($d \leq 40$). Otherwise, the approach is practically infeasible as enumerating vertices is a problem that scales very badly\footnote{It remains in fact an open question in computational geometry whether the vertices or the convex hull of a polytope can be computed in total polynomial time for arbitrary dimensions.}\todo{Is that true?} mainly because the number of vertices $n$ given dimension $d$ can be very large. For instance, a $d$-dimensional cube has $n = 2^d$ vertices. A general polytope with $m$ non-redundant hyperplanes can have

${m - \frac{d}{2} \choose \frac{m}{2}} + {m - \frac{d}{2} - 1 \choose \frac{m}{2} - 1} \approx O(m^{\frac{n}{2}})$ vertices (see e.g., \cite{bib:HenkGebertZiegler2004}). So even in the best case szenario\footnote{The literature suggests (see e.g., \cite{bib:BremnerFukudaMarzetta1998}) that the primal-dual polytope algorithm underlying the Matlab function \textit{lcon2vert.m} which performs vertex enumeration has an efficiency of $O(nd^2)$.} where we would have a vertex enumerating algorithm with efficiency $O(n)$ (since vertices can only be found one by one), the number of required simple operations would exceed anything manageable in cases of practical relevance.

%One can check that doing $225$ times even very simple operations, like addition, takes approximately $13$ seconds. Hence $230$ simple operations take approximately $10$ hours and $250$ take more than $3$ years. Hence, we conclude that a vertex enumeration problem is not solvable for cases of practical relevance where the number of dimensions (assets) is typically in the hundreds.

%D. Bremner, K. Fukuda and A.Marzetta. Primal Dual Methods for Vertex and Facet Enumeration. Discrete and Computational Geometry, 20, (1998)

%M. Henk, J. R. Gebert, and G. M. Ziegler. "Basic Properties Of Convex Polytopes." Handbook of Discrete and Computational Geometry. Chapter 16. CRC Press. 2004.

Because of their impracticability in most cases of practical relevance, \cite{bib:Smith1984} proposed an alternative to the conventional techniques which is to sample the set $K$ from within via a geometric random walk routine that produces a Markov chain whose stable state converges to the uniform distribution. The idea is to generate a long sequence of points and randomly mix their order so that the sequence becomes i.i.d. uniform over $K$. For this to be true, the algorithm needs to fulfil certain properties. \cite{bib:Smith1984} first presents a generic mixing procedure before showing it's properties and proving that it leads to the desired goal of uniform random variates.

\textit{discuss MCMC. Ubiquituosly used in Bayesian analysis. Steady state distribution not limited to the uniform.}\\

In the following we give an overview of the existing finance-specific literature on random portfolios, one the one hand, an on the technical literature regarding sampling (and it's origins in volume estimation) on the other hand. Eventually, the goal is to deep-dive into some of the proposed Markov chain Monte Carlo (MCMC) procedures and to show that those are most suitable (and essentially the only viable option) to efficiently generate random portfolios from a wide variety of domains.

Literature & History

The paper highlights two main applications where an efcient solution to the problem would be useful. First, it names the possibility of generating multivariate random variates with arbitrary density function *f* over *S* by frst generating uniformly distributed points *Y* within the region under the graph of *f* over *S* and projecting *Y* onto *S*. As a second application it’s suggested to see the uniformly distributed points as feasible solutions to constrained mathematical programs and to use them as starting points for heuristic optimum seeking procedures or as stochastic probes to directly evaluate the value of the objective function at those points.

