OPTIMIZING PORTFOLIO TAIL MEASURES: ASYMPTOTICS AND EFFICIENT SIMULATION OPTIMIZATION

Sandeep Juneja

Tata Institute of Fundamental Research
Mumbai, India

ABSTRACT

We consider a portfolio allocation problem where the objective function is a tail event such as probability of large portfolio losses. The dependence between assets is captured through multi-factor linear model. We address this optimization problem using two broad approaches. We show that a suitably scaled asymptotic of the probability of large losses can be developed that is a simple convex function of the allocated resources. Thus, asymptotically, portfolio allocation problem is approximated by a convex programming problem whose solution is easily computed and provides significant managerial insight. We then solve the original problem using sample average simulation optimization. Since rare events are involved, naive simulation may perform poorly. To remedy this, we introduce change-of-variable based importance sampling technique and develop a single change of measure that asymptotically optimally estimates tail probabilities across the entire space of feasible allocations.

1 INTRODUCTION

Tail events that occur rarely but whose occurrence results in catastrophic losses are increasingly important in managerial decision making. Investors holding a portfolio of assets closely monitor and control the value-at-risk (a specified tail-percentile of the loss distribution) of their investment portfolio. Insurance companies hold huge reserves to protect against the possibility of rare but catastrophic losses. A manager of a complex project is concerned about the overall project time to avoid enormous penalty and other costs associated with large delays. In communication networks huge buffers are maintained and scheduling of information packets is carefully managed to minimize the packet loss probability.

These problems maybe difficult to analyze as closed form expressions of tail probabilities or related measures such as value-at-risk are typically unavailable, and often are sufficiently complex functions of the controllable inputs that managerial insights for optimal decision making are not easily available. In this article we consider the optimization problem of allocating resources that minimize the tail event probabilities of large losses subject to minimal expected return requirement and other constraints. This objective function may be particularly important to risk managers concerned with unexpected tail losses. The distribution of losses from the underlying factors is critical to our analysis and we address this issue for super-exponentially, exponentially and polynomially decaying tail distribution of underlying factors. As will be apparent, the methodology that we develop is applicable in great generality, however to maintain focus, we limit our discussion to the portfolio optimization set-up.

We address this portfolio allocation problem using two broad approaches. First we develop simple asymptotic expressions for the tail probabilities that are useful surrogates to the probabilities in the associated resource allocation problem. We note that the resulting *approximate* problem is a convex programming problem that is easily solved *exactly*. We also note that an asymptotic analysis of other tail measures such as value-at-risk results in an identical optimization problem. Markowitz in his classic work on portfolio optimization considered the problem of selecting a portfolio of *n* securities with minimum variance for a given amount of expected return. We conduct a similar analysis with an asymptotic of the probability of large losses replacing variance as the risk measure. For simplicity, we do this analysis when the dependence between different securities has a multi-factor linear relationship. In particular, we note that in the special case when all factors have exponentially decaying tail distributions, the portfolio optimization problem reduces to a linear program. In other cases we get a convex programming problem with linear constraints.

In our second approach we solve the *exact* problem *approximately* via sample average simulation optimization. Note that the probability of an event is simply the expectation of indicator of that event. As is well known, the

sample average method addresses problems where for each feasible allocation, the objective function is an expectation of a random variable that is difficult to compute. In the associated optimization problem this expectation is replaced by an unbiased sample average of independent samples of the underlying random variable. In our setting there is an added difficulty that the objective function corresponds to a rare event probability, so enormous amount of samples are needed to get reliable estimates of the objective function at each feasible allocation. To remedy this drawback, we introduce a change-of-variable based importance sampling technique and develop a single importance sampling change of measure that asymptotically efficiently estimates tail probabilities across the entire space of feasible allocations. This idea was implemented for PERT networks with independent component distributions in Chauhan et. al. (2007). To the best of our knowledge these are the first successful implementations of importance sampling in simulation optimization problems. We note that the asymptotic solution obtained from our first approach may be a useful input in setting an initial point for the simulation optimization problem as well as for coming up with a good importance sampling measure.

Specifically, the main contribution of this article include:

1. Using the results in Juneja, Karandikar and Shahabuddin (2007), we develop asymptotics for the tail probabilities of portfolio of assets when the underlying random variables have tails that are superexponentially decaying, exponentially decaying or polynomially decaying. We develop an iterative methodology to solve the resultant asymptotic optimization problem. This involves checking feasibility of some linear constraints and solving a convex programming problem.

2. We develop an importance sampling methodology to efficiently conduct the simulation optimization of the exact non-asymptotic problem. We prove that it is asymptotically effective over the complete space of allocations. It can also be shown that as the number of samples generated increase, the simulation optimization algorithm converges to the solutions of the exact problem.

A word of caution is in order: The multi-factor linear model for losses from individual assets or securities may be reasonable to capture the central or near central portion of the resulting portfolio loss distribution. However, it could be significantly inaccurate in the far off tail distribution of losses. Therefore practitioners need to be careful of the potential modeling error. Having said that, the key ideas that we propose may in many cases be easily extended to other non-linear models that potentially better capture the loss tail distribution.

In the conference we will demonstrate the effectiveness of the proposed methodology through a few simple numerical examples.

In Section 2 we develop the mathematical framework useful to our analysis. In this section we also study the asymptotic optimization problem in light and heavy tailed settings. In Section 3, we consider the simulation optimization algorithms to solve the proposed portfolio optimization problem. Here we also develop provably efficient importance sampling based algorithms for the simulation optimization technique. Finally, we conclude with a brief conclusion and description of ongoing work, in Section 4.

2 MATHEMATICAL FRAMEWORK AND BACKGROUND

Let $\mathbf{L} = (L_i : 1 \leq i \leq m)$ denote the vector of losses from m assets or securities and $\theta = (\theta_i : 1 \leq i \leq m) \in \Re^m$ where θ_i denotes the investment in asset i. Then

$$L(\theta) = \sum_{i=1}^{m} \theta_i L_i$$

denotes the loss from the portfolio consisting of these assets. Further, assume that there exist n mutually independent factors $\mathbf{X} = (X_i : i = 1, \ldots, n)$ and a matrix $\mathscr{A} = \{a_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ such that each

$$L_i = \sum_{j=1}^{n} a_{ij} X_j.$$

In this paper we primarily consider the problem

$$\min_{\theta \in C} P(L(\theta) > u) \tag{1}$$

for a specified compact set C and for large values of u. Note that by selecting appropriate coefficients of $\mathscr{A}$, we can capture the popular settings where there are common and idiosyncratic factors involved in modeling losses from each asset.

We first formulate the asymptotic problem where $P(L(\theta) > u)$ is replaced by its suitable asymptotic surrogate that depends on the distribution of components of random vector $\mathbf{X}$ and that is a simple convex function of θ.

We need notation to specify the distribution of components of $\mathbf{X}$.

For rv X_j, let

- $f_j(\cdot)$ and $F_j(\cdot)$, respectively denote its probability density function and distribution function.
- $\tilde{\lambda}_j(\cdot)$ and $\Lambda_j(\cdot)$, respectively, denote its hazard rate and hazard function.

- $\tilde{\gamma}_j(\cdot)$ and $\Gamma_j(\cdot)$, respectively, denote the hazard rate and hazard function of $-X_j$.

Then, $f_j(x) = \tilde{\lambda}_j(x)\exp[-\Lambda_j(x)]$. As is well known, $\tilde{\lambda}_j(x) = \Lambda'_j(x) = f_j(x)/(1 - F_j(x))$, where for any function $h(\cdot)$, $h'(x)$ denotes its the derivative. Note that if $\frac{\Lambda_i(x)}{\Lambda_j(x)} \to \infty$ as $x \to \infty$, then X_i has a lighter tail compared to X_j, in the sense that

$$\frac{P(X_i > x)}{P(X_j > x)} = \exp(-\Lambda_i(x) + \Lambda_j(x)) \to 0.$$

Equivalently, then X_j has a heavier tail compared to X_i.

2.1 Light Tailed Settings

The following assumption and result in Juneja et al (2007) are needed to develop suitable asymptotic for $P(L(\theta) > u)$ in the light tailed settings.

Assumption 1 *For each $j \leq n$, there exists an $\alpha_j \geq 1$ and constants λ_j, $0 < \lambda_j < \infty$:*

$$\lim_{x \to \infty} \frac{\tilde{\lambda}_j(x)}{\lambda_j \alpha_j x^{\alpha_j - 1}} = 1. \tag{2}$$

(Note that this implies that

$$\lim_{x \to \infty} \frac{\Lambda_j(x)}{\lambda_j x^{\alpha_j}} = 1, \tag{3}$$

see Feller 1971, VIII.9, Theorem 1(b), pg 281). Furthermore, there exist constants $\eta_j \geq 1$ and γ_j, $0 < \gamma_j < \infty$:

$$\lim_{x \to \infty} \frac{\tilde{\gamma}_j(x)}{\gamma_i \eta_j x^{\eta_j - 1}} = 1. \tag{4}$$

Let $\alpha = \min(\alpha_j : j \leq n)$.
Let $\mathscr{H}$ contain all the indexes i such that $\alpha_i = \alpha$.
Define

$$\lambda^* = \frac{1}{(\sum_{j \in \mathscr{H}} 1/\lambda_j^{\frac{1}{\alpha - 1}})^{\alpha - 1}}, \tag{5}$$

for $\alpha > 1$ and

$$\lambda^* = \min_{j \in \mathscr{H}} \lambda_j$$

for $\alpha = 1$. Let $L = \sum_{j=1}^n X_j$.
Theorem 1 (Juneja et. al. 2007) *Under Assumption 1, when the components $\mathbf{X}$ are independent,*

$$\lim_{u \to \infty} \frac{\log P(L > u)}{u^\alpha} = -\lambda^*. \tag{6}$$

2.1.1 Solving the Approximate Problem Exactly

It follows from Assumption 1 that for any constant $b_i > 0$,

$$P(b_i X_i > x) = \exp[-\lambda_i (\frac{x}{b_i})^{\alpha_i}(1 + o(1))].$$

Set $\delta_j(\theta) = \sum_{i=1}^m \theta_i a_{ij}$, and

$$\beta(\theta) = \min_{j \leq n}\{\alpha_j I(\delta_j(\theta) > 0), \eta_j I(\delta_j(\theta) < 0)\},$$

where for any event A, $I(A)$ denotes its indicator function. Then $L(\theta) = \sum_{j=1}^n \delta_j(\theta) X_j$. Furthermore, for $j \leq n$, define

$$\lambda_j^*(\theta) = \lambda_j/\delta_j(\theta)^{\beta(\theta)} I(\delta_j(\theta) > 0, \alpha_j = \beta(\theta))$$
$$+ \gamma_j/|\delta_j(\theta)|^{\beta(\theta)} I(\delta_j(\theta) < 0, \eta_j = \beta(\theta)),$$

and set $\lambda_j^*(\theta) = \infty$ on the complement of set

$$\{\delta_j(\theta) > 0, \alpha_j = \beta(\theta)\} \cup \{\delta_j(\theta) < 0, \eta_j = \beta(\theta)\}.$$

Then, from Theorem 1 it follows that

$$\lim_{u \to \infty} \frac{\log P(L(\theta) > u)}{u^{\beta(\theta)}} = -\lambda^*(\theta), \tag{7}$$

where for $\beta(\theta) > 1$,

$$\lambda^*(\theta) = \frac{1}{(\sum_{j \leq n} 1/\lambda_j^*(\theta)^{\frac{1}{\beta(\theta) - 1}})^{\beta(\theta) - 1}},$$

and for $\beta(\theta) = 1$,

$$\lambda^*(\theta) = \min_{j=1,\dots,n} \lambda_j^*(\theta).$$

Therefore, for large u, $\exp(-\lambda^*(\theta)u^{\beta(\theta)})$ becomes a reasonable surrogate for $P(L(\theta) > u)$ and in (1) we may replace the objective function by it. To solve the resultant optimization problem, we first need to search for all θ that maximize $\beta(\theta)$. This can be found iteratively as follows:

- First identify all rv X_j or $-X_j$ with the smallest 'rarity index' α_j and/or η_j. That is, find rv with rarity index

$$\beta = \min_{j \leq n}(\alpha_j, \eta_j).$$

These are the heaviest tailed rv in our portfolio.
- We then introduce constraints on θ so that we only consider those portfolios where there is no exposure to these tails. Thus, if α_j equals β, we introduce the constraint $\sum_{i=1}^m \theta_i a_{ij} \leq 0$. If η_j equals

β, we introduce the constraint $\sum_{i=1}^{m} \theta_i a_{ij} \geq 0$. Any θ satisfying all these constraints has the property that $\beta(\theta) > \beta$. Let D denote the collection of newly introduced constraints.

- We, then check if the set with constraints C and D has a feasible solution.
- If the resultant set has a feasible solution, we repeat the above steps after augmenting the constraint set C to include constraints D so that the portfolios satisfying these constraints are not exposed to the relatively heavy tailed losses. In conducting the repeat steps, in recomputing β, we do not consider α_j (resp. η_j) for which we set $\sum_{i=1}^{m} \theta_i a_{ij} \leq 0$ (resp. $\sum_{i=1}^{m} \theta_i a_{ij} \geq 0$) in the earlier steps.
- Else, if the resultant set is infeasible, we stop.

It is easy to see that due to no-arbitrage considerations, under reasonable constraint set C, this algorithm will eventually stop with the resultant set of constraints infeasible (this is discussed later). (At that point, all $\theta \in C$ maximize the original $\beta(\theta)$ discussed earlier.) The asymptotic approximation to the solution of (1) is achieved by solving

$$\max_{\theta \in C} \lambda^*(\theta)$$

where now C is the set obtained after incorporating the constraints introduced in the above algorithm.

We now solve this problem separately for $\beta = 1$ and $\beta > 1$. We may take the initial set C to correspond to

$$C = \{\sum_{i=1}^{k} \theta_i = 1, \sum_{i=1}^{k} \theta_i r_i \geq \bar{r}\}$$

where $(r_i : 1 \leq i \leq k)$ correspond to expected rates of returns from each security and $\bar{r}$ corresponds to the minimum desired rate of return. This is more than the risk free rate of return. For simplicity we assume that all α_j and η_j equal β. This ensures that C is not augmented by additional constraints, as the no arbitrage considerations imply that augmenting the set C with constraints

$$\sum_{i=1}^{m} a_{ij} \theta_i = 0$$

for $j = 1, \ldots, n$ leads to an infeasible solution. (A feasible solution would give rate of return of at least $\bar{r}$ without any risk.)

Recall that $\delta_j(\theta) = \sum_{i=1}^{m} a_{ij} \theta_i$. In this case,

$$\lambda_j^*(\theta) = \lambda_j / \delta_j(\theta)^\beta,$$

if $\delta_j(\theta) > 0$, it equals

$$\eta_j / |\delta_j(\theta)|^\beta$$

if $\delta_j(\theta) < 0$, and ∞ otherwise.

2.1.2 Exponential Decay Rate, $\beta = 1$

In this case, rather than maximizing $\lambda^*(\theta)$, we may equivalently minimize

$$\max_{j \leq n} \left(\frac{\delta_j(\theta)}{\lambda_j}, -\frac{\delta_j(\theta)}{\gamma_j} \right)$$

subject to $\theta \in C$.

This problem is solved using the linear program:

$$
\begin{aligned}
\min \quad & \zeta \\
\text{s. t.} \quad & \sum_{i=1}^{m} a_{ij} \theta_i \leq \zeta \lambda_j \quad j = 1, \ldots, n \\
& -\sum_{i=1}^{m} a_{ij} \theta_i \leq \zeta \gamma_j \quad j = 1, \ldots, n \\
& \theta \in C.
\end{aligned}
$$

2.1.3 Super-exponential Decay Rate, $\beta > 1$

Again, rather than maximizing $\lambda^*(\theta)$, we may equivalently minimize

$$\sum_{j \leq n} 1 / \lambda_j^*(\theta)^{\frac{1}{\beta-1}}$$

subject to $\theta \in C$. This may be re-formulated as a convex programming problem with linear constraints as follows:

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{n} \left(\frac{(\xi_j^+)^{\beta/(\beta-1)}}{\lambda_j^{1/(\beta-1)}} + \frac{(\xi_j^-)^{\beta/(\beta-1)}}{\gamma_j^{1/(\beta-1)}} \right) \\
\text{s. t.} \quad & \xi_j^+ - \xi_j^- = \sum_{i=1}^{m} a_{ij} \theta_i \quad j = 1, \ldots, n. \\
& \xi_j^+, \xi_j^- \geq 0 \quad j = 1, \ldots, n. \\
& \theta \in C.
\end{aligned}
$$

This is easily solved using standard numerical techniques (see, e.g., Boyd and Vandenberghe 2004).

2.2 Heavy Tailed Settings

Recall that a rv W is said to be heavy or sub exponentially tailed if

$$\frac{P(W_1 + W_2 > u)}{2P(W_1 > u)} \to 1$$

as $u \to \infty$, where W_1 and W_2 have the same distribution as W (see, eg., Pakes 2004). We assume that there exists a heavy

tailed random variable Y, and constants $(\lambda_j > 0 : j \leq n)$, $(\gamma_j > 0 : j \leq n)$ and $\beta > 0$ such that

$$\lim_{u \to \infty} \frac{P(aX_j > u)}{P(Y > u)} = \lambda_j a^\beta, \qquad (8)$$

and

$$\lim_{u \to \infty} \frac{P(-aX_j > u)}{P(Y > u)} = \gamma_j a^\beta, \qquad (9)$$

for each $a > 0$. This would be true for instance if for each $j \leq n$,

$$P(X_j > u) \sim \lambda_j \frac{L(u)}{u^\beta},$$

$$P(-X_j > u) \sim \gamma_j \frac{L(u)}{u^\beta},$$

and $P(Y > u) \sim \frac{L(u)}{u^\beta}$, where $L(\cdot)$ is a slowly varying function, that is, $\lim_{x \to \infty} L(tx)/L(x) = 1$ for all $t > 0$.

In that case, under mild technical conditions,

$$P(L(\theta) > u) = P(\sum_{j=1}^n \delta_j(\theta)X_j > u) \sim \sum_{j=1}^n P(\delta_j(\theta)X_j > u),$$

and this in turn is asymptotically similar to $P(Y > u)$ times

$$\left(\sum_{j=1}^n (\delta_j(\theta)^\beta \lambda_j I(\delta_j(\theta) > 0) + |\delta_j(\theta)|^\beta \gamma_j I(\delta_j(\theta) < 0)) \right)$$

Therefore, the asymptotic optimization problem reduces to the simple convex programming problem

$$\begin{aligned}
\min \quad & \sum_{j=1}^n \left((\xi_j^+)^\beta \lambda_j + (\xi_j^-)^\beta \gamma_j \right) \\
\text{s. t.} \quad & \xi_j^+ - \xi_j^- = \sum_{i=1}^m a_{ij}\theta_i \ \ j = 1, \dots, n. \\
& \xi_j^+, \xi_j^- \geq 0 \ \ j = 1, \dots, n. \\
& \theta \in C.
\end{aligned}$$

Note that it is easy to generalize this analysis to include cases where the positive and negative tails of X_j have different decay rates and also where light as well as heavy tailed random variables are involved.

2.3 Optimizing Asymptotic Value-at-Risk

Suppose that the following asymptotic holds:

$$\lim_{u \to \infty} \frac{1}{u^\alpha} \log P(L(\theta) > u) = -\lambda^*(\theta).$$

Suppose that $L(\theta)$ has a continuous density so that the value-at-risk at level $p \in (0,1)$, $u_p(\theta)$ is the unique solution to

$$P(L(\theta) > u_p(\theta)) = p.$$

It follows that for $\varepsilon > 0$ and sufficiently small p lies between

$$\exp(-\lambda^*(\theta)u_p(\theta)^\alpha(1 \pm \varepsilon))$$

so that

$$\lim_{p \to 0} \frac{u_p(\theta)}{(-\log(p))^{1/\alpha}} = 1/\lambda^*(\theta)^{1/\alpha}.$$

So again, asymptotically, the θ that minimizes $u_p(\theta)$ corresponds to the θ that maximizes $\lambda^*(\theta)$.

Similar results can be easily shown in the heavy-tailed settings and for performance measures such as the expected shortfall.

3 SAMPLE AVERAGE OPTIMIZATION: SOLVING EXACT PROBLEM APPROXIMATELY

The basic idea behind the approach is fairly simple and intuitive and is presented, for instance, in Kleywegt and Shapiro (2001). Observe that the loss probability can be re-expressed as

$$q(\theta) = \mathbb{E}[I(L(\theta, \mathbf{X}) > u)]$$

where $I(\cdot)$ represents the indicator function, assumes value 1 if an event happens and 0 otherwise, and $\mathbb{E}[\cdot]$ denotes expectation under a measure P under which the vector $\mathbf{X}$ has joint density function $f_\mathbf{X}$. Typically, $q(\theta)$ is not known in closed form. However due to the law of large numbers, it can be approximated by the sample average of i.i.d. samples of $I(L(\theta, \mathbf{X}) > u)$. Concretely, suppose $\mathbf{X}^1, \dots, \mathbf{X}^N$ denote N independent, identically distributed samples from the distribution $f_\mathbf{X}$. Then an approximation $\hat{q}(\cdot)$ to the objective function may be obtained as

$$\hat{q}(\theta) = \frac{1}{N} \sum_{i \leq N} I(L(\theta, \mathbf{X}^i) > u).$$

$\hat{q}(\cdot)$ can now be used as a surrogate objective to solve the resultant deterministic optimization problem in Equation 1 and obtain an approximate solution. This sample average optimization methodology has certain drawbacks in our setting. First note that $\hat{q}(\theta)$ would typically not be a continuous or differentiable and hence gradient based methods are difficult to implement. Second, there is a problem of rare event simulation. If u is large, the probability $\hat{q}(\theta)$ requires enormous computational effort to estimate accurately. Fortunately, both these problems can be rectified by

using change of variables technique and then importance sampling.

3.1 Change of Variable Technique

Note that

$$q(\theta) = \int_{L(\theta,\mathbf{x})>u} f_{\mathbf{X}}(x_1,\ldots,x_n)dx_1 \cdots dx_n.$$

Recall that, $L(\theta,\mathbf{x}) = \sum_{j=0}^{n} \delta_j(\theta)x_i$. For notational convenience, we denote $\delta_j(\theta)$ by δ_j suppressing the dependence on θ. Also, without essential loss of generality we assume that each $\delta_j > 0$. If it is less than zero we can consider the rv $-X_j$ instead of X_j. If $\delta_j = 0$, then the term can be ignored.

Change the variables of integration to $y_j = \delta_j x_j$ for all $j \leq n$. of θ. Then,

$$q(\theta) = \int_{\sum_{j=1}^{n} y_i > u} f_{\mathbf{X}}\left(\frac{y_1}{\delta_1},\ldots,\frac{y_n}{\delta_n}\right)\frac{dy_1}{\delta_1} \cdots \frac{dy_n}{\delta_n}.$$

Let $g : \Re^n \to \Re^+$ denote a joint pdf of n random variables. Then RHS may be re-expressed as

$$\int_{\sum_{j=1}^{n} y_i > u} \frac{f_{\mathbf{X}}\left(\frac{y_1}{\delta_1},\ldots,\frac{y_n}{\delta_n}\right)}{g(y_1,\ldots,y_n)} g(y_1,\ldots,y_n) \frac{dy_1}{\delta_1} \cdots \frac{dy_n}{\delta_n}$$

with the caveat that $g(y_1,\ldots,y_n) > 0$ whenever $f_{\mathbf{X}}\left(\frac{y_1}{\delta_1},\ldots,\frac{y_n}{\delta_n}\right) > 0$.
Hence,

$$q(\theta) = \frac{1}{\prod_{j=1}^{n} \delta_j} E_g\left[\frac{f_{\mathbf{X}}\left(\frac{Y_1}{\delta_1},\ldots,\frac{Y_n}{\delta_n}\right)}{g(Y_1,\ldots,Y_n)} I(\sum_{j=1}^{n} Y_j > u)\right], \quad (10)$$

where E_g is the expectation operator associated with joint density g. Therefore, the sample average of samples of

$$\frac{1}{\prod_{j=1}^{n} \delta_j} \frac{f_{\mathbf{X}}\left(\frac{Y_1}{\delta_1},\ldots,\frac{Y_n}{\delta_n}\right)}{g(Y_1,\ldots,Y_n)} I(\sum_{j=1}^{n} Y_j > u)$$

genrated using g provides a surrogate for the objective function $q(\theta)$.

Also note that due to independence of random components: $f_{\mathbf{X}}(\mathbf{X}) = \prod_{j=1}^{n} f_j(X_j)$. Suppose that the rv remain independent under joint density g, i.e.,

$$g(y_1,\ldots,y_n) = \prod_{j=1}^{n} g_j(y_j).$$

Let

$$Z(\theta) = \frac{1}{\prod_{j=1}^{n} \delta_j} \frac{\prod_{j=1}^{n} f_j\left(\frac{Y_j}{\delta_j}\right)}{\prod_{j=1}^{n} g_j(Y_j)} I(\sum_{j=1}^{n} Y_j > u).$$

Let $\bar{Z}_n(\theta)$ denote the sample average of n i.i.d. samples of $Z(\theta)$. Deterministic optimization can be conducted using this $\bar{Z}_n(\cdot)$ as the objective function. The advantage is that now the resultant approximation to the objective function is continuous in each δ_j and hence in each θ_i. Further, we can similarly obtain an approximation to each partial derivative $\frac{\partial q(\theta)}{\partial \theta_i}$. To see this, note first that for $i = 1,\ldots,m$, $\frac{\partial}{\partial \theta_i}\delta_j = a_{ij}$, and that under mild regularity conditions that allow interchange of derivative with the expectation that $\frac{\partial}{\partial \theta_i} q(\theta)$ equals

$$-E_g\left[\sum_{j=1}^{n}\left(\frac{a_{ij}}{\delta_j} + \frac{a_{ij}f_j'\left(\frac{Y_j}{\delta_j}\right)Y_j}{\delta_j^2 f_j\left(\frac{Y_j}{\delta_j}\right)}\right) Z(\theta)\right]. \quad (11)$$

Therefore, the average of samples of

$$-\left[\sum_{j=1}^{n}\left(\frac{a_{ij}}{\delta_j} + \frac{a_{ij}f_j'\left(\frac{Y_j}{\delta_j}\right)Y_j}{\delta_j^2 f_j\left(\frac{Y_j}{\delta_j}\right)}\right) Z(\theta)\right]$$

for each i provide a surrogate for the gradient of $q(\theta)$. Note that the gradient estimate from n samples equals $\frac{\partial}{\partial \theta_i}\bar{Z}_n(\theta)$.

Also note that again changing variables by setting each $X_j = \frac{Y_j}{\theta_j}$ in (12) we get

$$-E_f\left[\sum_{j=1}^{n}\left(\frac{a_{ij}}{\delta_j} + \frac{a_{ij}f_j'(X_j)X_j}{\delta_j f_j(X_j)}\right) I(L(\theta,\mathbf{X}) > u).\right] \quad (12)$$

This then provides an alternative way to estimate the gradient of $q(\theta)$.

3.2 Importance Sampling

The analysis in the last section allowed us to get a continuous approximation for the objective function and its gradient. The problem of rarity due to large value of u remains. This can be addressed by selecting an appropriate importance sampling distribution g. We now discuss an asymptotically optimal choice of g that efficiently estimates $q(\theta)$ for all feasible θ for large values of u.

Again consider

$$q(\delta) = P(\sum_{j=1}^{n} \delta_j X_j > u)$$

where again we suppress the dependence on θ, and without loss of generality, consider all $\delta_j > 0$. Further consider a light tailed regime so that for $j = 1, \ldots, n$

$$f_j(x) = e^{-\lambda_j x^\alpha (1 + o(1))}.$$

As we discussed earlier, it then follows that

$$\lim_{u \to \infty} \frac{1}{u^\alpha} \log q(\delta) = -\lambda^*(\delta)$$

where

$$\lambda^*(\delta) = \left(\sum_{j=1}^n (\lambda_j \delta_j^{-\alpha})^{-(\alpha-1)^{-1}} \right)^{-(\alpha-1)}$$

if $\alpha > 1$ and $\lambda^*(\delta) = \min_{j \leq n} \lambda_j \delta_j^{-1}$ if $\alpha = 1$.

Then for any δ, if we generate $(Y_1, \ldots, Y_n)$ independently using densities $(g_j : j \leq n)$, respectively, and our importance sampling estimator for $q(\delta)$ is

$$\hat{q}_g(\delta) = \frac{1}{\prod_{j=1}^n \delta_j} \frac{\prod_{j=1}^n f_j(Y_j/\delta_j)}{\prod_{j=1}^n g_j(Y_j)} I(\sum_{j=1}^n Y_j > u).$$

Since $E_g[\hat{q}_g(\delta)^2] \geq q(\delta)^2$, it follows that

$$\liminf_{u \to \infty} \frac{1}{u^\alpha} \log E_g[\hat{q}_g(\delta)^2] \geq -2\lambda^*(\delta).$$

An importance sampling measure associated with g is said to be asymptotically optimal if the above inequality holds as an equality and liminf can be replaced by a lim. That is, if

$$\limsup_{u \to \infty} \frac{1}{u^\alpha} \log E_g[\hat{q}_g(\delta)^2] \leq -2\lambda^*(\delta).$$

We now identify an importance sampling distribution that achieves this for $\delta > 0$. Select

$$g_j(y) = \frac{c_j}{u} e^{-\frac{c_j}{u} y}$$

for $j = 1, \ldots, n$ where each c_j is a constant independent of δ. This implies that $E_g Y_j = u/c_j$.

We now show that with this choice of density functions we achieve asymptotic optimality for any $\delta > 0$. To see this note that $E_g[\hat{q}_g(\delta)^2]$ equals $\frac{1}{(\prod_{j=1}^n \delta_j)^2}$ times

$$\int_{\sum_{j=1}^n y_j > u} \left(\frac{\prod_{j=1}^n f_j(y_j/\delta_j)}{\prod_{j=1}^n g_j(y_j)} \right)^2 \prod_{j=1}^n g_j(y_j) dy_1 \ldots dy_n.$$

Plugging in the expressions for various density functions, the RHS equals $\frac{u^n}{(\prod_{j=1}^n \delta_j)^2}$ times

$$\int_{\sum_{j=1}^n y_j > u} \left(\frac{e^{-\sum_{j=1}^n \lambda_j \delta_j^{-\alpha} y_j^\alpha}}{\prod_{j=1}^n c_j e^{-\sum_{j=1}^n \frac{c_j}{u} y_j}} \right)^2 \prod_{j=1}^n g_j(y_j) dy_1 \ldots dy_n,$$

where to keep the discussion simple we have assumed that the $o(1)$ terms in the original density functions equal zero and the terms y_j are restricted to be non-negative. Since

$$\lim_{u \to \infty} \frac{1}{u^\alpha} \log \frac{u^n}{\prod_{j=1}^n \delta_j^2 c_j} = 0,$$

we ignore these from further analysis. To prove the result, it suffices to bound from above

$$\left(\frac{e^{-\sum_{j=1}^n \lambda_j \delta_j^{-\alpha} y_j^\alpha}}{e^{-\sum_{j=1}^n \frac{c_j}{u} y_j}} \right)^2 I(\sum_{j=1}^n y_j > u) \qquad (13)$$

by $e^{-2\lambda^*(\delta) u^\alpha (1 + o(1))}$. To this end, note that

$$\sum_{j=1}^n \lambda_j \delta_j^{-\alpha} y_j^\alpha \geq \lambda^*(\delta) (\sum_{j=1}^n y_j)^\alpha \qquad (14)$$

when $(y_j : j = 1, \ldots, n)$ are non-negative. This is obvious for $\alpha = 1$. To see this for $\alpha > 1$, consider a rv that takes value $(\lambda_j \delta_j^{-\alpha})^{1/(\alpha-1)} y_j$ for $j = 1, \ldots, n$ with probability

$$\frac{(\lambda_j \delta_j^{-\alpha})^{-1/(\alpha-1)}}{\sum_{i=1}^n (\lambda_i \delta_i^{-\alpha})^{-1/(\alpha-1)}}$$

and use the fact that α moment of a non-negative rv is greater than or equal to the first moment raised to power α for $\alpha > 1$. Using (14), we bound (13) from above by

$$e^{-2\lambda^*(\delta) (\sum_{j=1}^n y_j)^\alpha + 2 \frac{\max_{j=1,\ldots,n} c_j}{u} (\sum_{j=1}^n y_j)} I(\sum_{j=1}^n y_j > u)$$

and the result follows.

The proposed algorithm works for any choice of the positive constants $(c_j : j = 1, \ldots, n)$. In practice these may be chosen so that the importance sampling distribution puts significant mass around points where the asymptotically optimal θ (determined in previous section) concentrates mass. Under mild conditions it can be shown that the proposed importance sampling technique also efficiently estimates each of the gradient associated with $q(\delta)$.

Importance sampling can similarly seen to be uniformly effective when the rv $(X_j : j \leq n)$ are heavy tailed. In that case, ideas from Juneja et al. (2007), Asmussen and Kroese (2006) and Juneja (2007) may be used to get estimators that

are logarithmically efficient, have bounded relative error or have vanishing relative error, respectively.

4   CONCLUSIONS AND ONGOING WORK

In this article we considered the portfolio optimization problem extending the classic Markowitz framework to the case where the objective function corresponds to a tail measure such as the probability of large losses. We developed asymptotics for the probability of large losses as a function of the allocated resources and formulated the resultant non-linear programming problem. We also developed importance sampling based efficient simulation optimization procedures to solve the portfolio optimization problem. Our ongoing effort involves conducting experiments to validate the proposed algorithms. In a more elaborate version of this article we analyze in greater detail some of the issues alluded to in this article. We also study the efficient frontier associated with this tail risk problem.

REFERENCES

Asmussen, S., D. P. Kroese. 2006. Improved Algorithms for R are Event Simulation with Heavy Tails, *Adv. Appl. Probab.* 38(2): 545-558.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*, Cambridge University Press.

Chauhan, M., S. Juneja, H. Kalra, R. Karumanchi. 2007. Procurement of Multiple items through Auctions in PERT Networks, *Collaborative Report-GM Research and Development Center* ISL 134/CL-07-14.

Embrechts, P., C. Kluppelberg, T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*, Springer, Applications of Mathematics vol. 33.

Feller, W. 1971. *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York.

Juneja, S. 2007. Estimating Tail Probabilities of Heavy Tailed Distributions with Asymptotically Zero Relative Error, *Queueing Systems* 57: 115-127.

Juneja, S., R. L. Karandikar, P. Shahabuddin. 2007. Asymptotics and Fast Simulation for Tail Probabilities of Maximum of Sums of Few Random Variables, *ACM Transactions in Modelling and Simulation* 17(2)

Kleywegt, A.J., A. Shapiro. 2001. *Stochastic Optimization*, 3rd Edition of the Handbook of Industrial Engineering, Gavriel Salvendy, Ed., John Wiley, New York : 2625 - 2650.

Pakes, A. G. 2004. Convolution Equivalence and Infinite Divisibility, *J. Appl. Prob.* 41: 407-424.

AUTHOR BIOGRAPHY

SANDEEP JUNEJA received his Ph.D. in Operations Research from Stanford University in 1994. He is currently an Associate Professor in the School of Technology and Computer Science at the Tata Institute of Fundamental Research in India. His research interests include applied probability, computational finance, simulation and rare-event analysis. He serves on the editorial boards of Management Science and ACM TOMACS. His email address is juneja@tifr.res.in.